Master of Science Thesis

for the degree of

Use of apriori information for anatomical segmentation of medical images

by

Emilio Ippoliti
S265752

Under the supervision of

Prof.ssa Santa Di Cataldo
Prof. Edoardo Patti
Dr.Giovanni Valbusa

Politecnico di Torino

Department of Mechanical and Aerospace Engineering
Politecnico di Torino
Torino, Italia - 10122
a.a. 2020-2021

DECLARATION BY THE SCHOLAR

I certify that:

- The work contained in this thesis is original and has been done by me under the guidance of my supervisor.
- The work has not been submitted to any other Institute for any degree or diploma.
- I have followed the guidelines provided by the Institute in preparing the thesis.
- I have conformed to the norms and guidelines given in the Ethical Code of Conduct of the Institute.
- Whenever I have used materials (data, theory and text) from other sources, I have given due credit to them by citing them in the text of the thesis and giving their details in the reference section.
- The thesis has been checked by anti-plagiarism software.

Emilio Ippoliti

Acknowledgement

I would like to express my deep and heartfelt gratitude to my respected guide and mentor, Dr. Giovanni Valbusa, for his evergreen blessings, valuable advice, support, good wishes, and encouragement all along the journey of this work. Without his undisputed support, this work would not have been true in the light of daylight. I am very grateful to my supervisors of the thesis Committee (Prof.ssa Santa Di Cataldo , Prof. Edoardo Patti) with particular recognition to Prof.ssa Santa di Cataldo for the attention and interest shown and for her willingness and comprehension demonstrated during the fulfillment of the work. Finally, I would also like to thank the company Bracco imaging S.P.A. for giving me the opportunity to work on this project and for the instrumentation granted to me.

Place: Polytechnic University of Turin,
Date: July 8, 2021

Emilio Ippoliti

Abstract

In this thesis, we investigated how in medical image segmentation, apriori information about objects' shape knowledge can be used to guide the lungs segmentation from chest X-ray images. We propose a deep convolutional neural network for lung segmentation, where shape information is represented by a statistical representation, based on principal components analysis of lung masks computed using the training dataset and described by their Signed Distance Functions. In this regard we provided a nonlinear extension, considering the expediency of KernelPCA and the related kernel functions. As a result, the proposed network learns to predict shapes instead of learning pixel-wise classification. The segmentation method named Shape Predictor Network (SP-Net) was applied to chest X-ray images of Covid-19 positive patients where shape could be of utmost importance. Results showed that SP-Net and Kernel SP-Net could constrain the predicted shape to resemble a pulmonary-like structure in those cases where the pixel-wise segmentation methods failed in the presence of a significant image artifact. Finally, we compared the proposed methods with state-of-the-art methods. U-net, a convolutional neural network particularly effective for medical image segmentation.

Keywords: Deep Learning , SP-NET,U-NET, A priori information , Shape Analysis ,PCA , KernelPCA

List of Tables

List of Figures

Abbreviations

AAM	Active Appearance Model
AI	Artificial Intelligence
ASM	Active Shape Model
BRATS	Brain Tumor Segmentation
CAD	Computer-aided diagnosis
CDI	Centro Diagnostico Italiano
COPD	Chronic obstructive pulmonary disease
Covid-19	CoronaVIrus Disease 19
CXR	Chest X-Ray
CT	Computed Tomography
DL	Deep Learning
DNN	Deep Neural Network
ETL	Extracting-Transform-Loading
FCM	Fuzzy C-Means
FCN	Fully Convolutional Network
GPU	e Graphics processing unit
GT	Ground Truth
JSRT	Japanese Society of Radiological Technology
KPCA	Kernel Principal Component Analysis
HVPP	Horizontal and Vertical Projection Profile
IPF	idiopathic pulmonary fibrosis
LR	Learning Rate
MCU	Montgomery
ML	Machine Learning
MRF	Markov Random Field
MRI	Magnetic Resonance Imaging
PCA	Principal Component Analysis
PDM	Point Distribution Model
PET	Positron Emission Tomograph
RBF	Radial Basis Function
Sars-Cov-2	Severe acute respiratory syndrome Coronavirus
SDF	Signed Distance Function
SIFT	Scale-invariant feature transform
SPA	Società Per Azioni
SP-NET	Shape Predictor Network
SSM	Statistical Shape Modelling
SVD	Singular Value Decomposition

Contents

CHAPTER 1

Introduction

This section explains the principal aspects and motivations that led to the elaboration of this work, it also intends to illustrate its main contributions. For this reason, we provide an introduction that gives an overview of how the idea of *a priori* information can be applied with convolutional neural networks and in the context of deep learning, with particular attention to its application in lung fields segmentation of CXR images of Covid-19 positive patients. The following section, called 'Motivation of thesis', provides a brief report on why *apriori* information can improve segmentation results but especially on the importance that it defines in delineating the lung field in CXR imaging for screening and improving the diagnosis and treatment of Sars-Cov-2 patients. In the final part are reported the foremost contributions that the thesis wants to give with the development and study of the proposed method.

1.1 Introduction

Apriori information is an essential component of the human vision system, allowing us to be more selective and effective in a visual recognition task. In computer vision, the need to include this knowledge about the task to be realized is increasingly recognized. Apriori shape information can be considered an exciting point in a computer vision problem.

The significant progress in computer vision induced us to make extensive medical imaging data to diagnose better, treat, and predict diseases. In this context, computer vision can exploit shape, contour as apriori information from an image sequence and provide an anatomical segmentation that helps with better human understanding. On these grounds, the thesis focus on the ailment and integration of apriori information in a specific branch of computer vision like image segmentation, which aims to extract the semantic objects lying in images by dividing any given image into meaningful contiguous regions or extracting one or more critical objects in images. In this perspective, it is a fundamental component of automated vision systems and uses in medical applications. Accurate segmentation of 2-D, 3-D medical images to isolate anatomical objects of interest for analysis is essential in almost any computer-aided diagnosis system or other medical imaging areas. However, the problem remains challenging, with no general and unique solution, due to a large and constantly growing number of different objects of interest, notable variations of their properties in images, and different medical imaging modalities. In this context, using apriori information, integrated with the segmentation algorithms, could improve the results, leading to better performance. Nowadays, image segmentation has become an essential tool and support for medical diagnosis and treatment. One method for initial screenings of Covid-19 positive patients is using Chest X-rays. Due to its rapidity, cost-effectiveness, and a low dose of radiation, CXR imaging is revealed as a suitable modality in particular situations as in emergency departments and outpatient clinics. . Indeed, the automatic analysis and segmentation of anatomical compartments in CXR images could alleviate the tremendous amount of work for the clinicians and researchers involved in the unprecedented challenge of the pandemic crisis. Therefore improvements in the segmentation carried out in this area are of utmost relevance. In particular one of the most critical steps in the automatic analysis of CXRs is to detect the lung boundaries accurately. The boundary extraction is the key to identifying shape irregularity and lung volume and thus provides insight into the identification of pneumonia and others symptoms related to Covid-19 and aids in reducing computational complexity for other lung-related disease identification algorithms by performing computations only on lung regions. These factors prompted us to explore the benefits of apriori information develop the proposed method within the thesis on lung segmentation in CXR images to aid and

support the R&D and academic community in the battle against Covid-19. Indeed during the pandemic, the research community has been finding ways to combat the Covid-19 disease. It has progressed rapidly in image processing tasks like segmentation and classification over the past year, including biomedical imaging and collecting medical images and data of positive patients. Although COVID-19 medical images are still not available in large quantities. Our experiments are completed on one of the few available datasets. The dataset used was released by Centro Diagnostico Italiano in Milan; clinical data and chest radiographs are collected in six Italian hospitals at the time of hospitalization of symptomatic patients with Covid-19 during the first wave of emergency in the country (March-June 2020) [2]. The segmentation methods implemented with these data were developed by considering artificial intelligence a brilliant solution and deep learning the mainstream of medical image segmentation. Especially convolutional Neural Network (CNN) [5] has attracted tremendous attention because of Its high efficiency for feature extraction and particular classification. In this setting, our thesis intends to explore how apriori shape information can be integrated within deep convolutional neural networks to improve their results and make their segmentations more reliable, especially on low contrast images due to the severe conditions of Covid-19 patients.

1.2 Motivation of thesis

Due to the significant advances in imaging devices and technologies, digital images play a more and more critical role in the hospital sector. An image records a scene of one anatomical and structural part of the patient in a numeric representation that can be stored, transmitted, and studied afterward. The expression "a picture is worth a thousand words" indicates that an image has a powerful ability to describe the rich information it carries. In the medical field, images are acquired through various modalities, providing invaluable access to see the human body's interior and allowing physicians to make a more accurate diagnosis. During the pandemic, especially CXR imaging has proved to be, as already anticipated, a suitable tool to give more information on the medical state of Covid -19 patients' lungs. In this domain, the extraction of useful information from CXR images has become an essential task. Although radiologists and researchers can naturally and easily solve this task (at least in the 2D natural interpretation of scenes), it is difficult for a computer to automatically interpreting an image. To analyze an image or understand the scene, one of the first assignments consists of partitioning the figure into several significant parts. Since an image is composed of many pixels, the segmentation problem can be treated as a labeling problem that assigns each pixel a label indicating a particular component in the scene. Alternatively, the segmentation task is considered as extracting the boundaries between different objects so that the image is partitioned into meaningful regions according to the edges. As we will comprehend, the accuracy of the segmentation is an absolute necessity in medical applications and has assumed, in recent times, an essential significance in the ways of diagnosing Coronavirus diseases. In most cases, manual segmentation by experts can provide the best and the most reliable result. Still, it is tedious and incredibly time-consuming, a burdensome factor for analysis and diagnosing Covid-19 disease. Moreover, manual segmentation is subjective to operator variability. These facts motivate researchers to develop automatic segmentation methods to deal with large datasets while achieving the accuracy of manual segmentation. All this led us in our thesis to investigate how segmentation methods can be applied to CXR images and how they can be enhanced for lung boundaries detection. However, lung field segmentation remains a strenuous accomplishment for a computer. The challenge is due to the considerable variability of anatomical shapes in different patients. To deal with this challenge, we have studied how to incorporate the apriori shape information on image segmentation methods. So the optimal segmentation map is constrained by the set of valid shapes learned from the object of interest. In this manner, segmentation is more robust to noise, while using the global shape of the object can also alleviate the ambiguity of non-visible boundaries between different organs due to similar tissue properties or noise and sampling ar-

tifacts introduced during the CXR acquisition process. In this respect, our analysis intends to deepen how apriori information of relative order can be applied to train a convolutional neural network effectively and how such a concept can induce better performance about lung field segmentation than state-of-art. Above all, the motivation of the thesis lies in the hope that the following work can be helpful on how artificial intelligence and its applications can support research and studies related to the ongoing global pandemic of coronavirus diseases.

1.3 Contribution

The thesis explores and investigates how apriori shape information can be integrated into leading deep learning algorithms, failing to learn high-level topological shape knowledge and often failing to constrain the object segmentation results. We represent a novel, deep fully convolutional neural network that can predict in the shape domain instead of classifying each pixel separately, constraining the segmented region to a class of learned shapes. The principal components analysis constructs the shape space to realize the drawbacks and advantages of a linear dimension reduction method in capture and retrieving the total variation in the training dataset. Furthermore, we consider the feasibility of a nonlinear version of the proposed method based on Kernel PCA, perceiving as a nonlinear estimation of apriori shape information may or not yield better performances. We also inspect the training process of the proposed model by emphasizing the optimization of the hyperparameters and regularization techniques. In this context, we explore how early stopping and data augmentation can reduce the network overfitting and whether or not they can give better results, although all two, as we will see, lead to complications of the algorithm. In conclusion, We apply the proposed method to Chest X-ray images of positive Covid-19 patients, where the shape can be information of paramount importance to guide the lung segmentation process in a misleading imaging modality, being confident that this work will help the research community in the relevant Covid-19 studies. To sum up, the main contributions of this thesis are the following:

- Investigation of a deep fully convolutional neural network that predicts shapes instead of segmentation maps
- Study and understanding of the principal component analysis in estimating the apriori shape information by the training set
- A nonlinear extension of the proposed method based on kernel principal components analysis.
- The applicability of the proposed method in a challenging context as lung fields segmentation in CXR images
- A small step on how artificial intelligence methods can aid in the diagnosis of Covid-19 diseases.

CHAPTER 2

Background

The 'Background' section provides a brief overview of the imaging modality, outlining the history, major applications and highlighting its fundamental use in screening and detecting covid-19. In the chapter, we also discuss the relevance of lung field segmentation in this medical field and the related methods used to date. In the concluding part, the use of apriori information in shape context and its derivations in lung fields segmentation are reviewed with particular attention to their application in deep learning.

2.1 Introduction

Image segmentation is one of the most exciting and challenging problems in computer vision generally and medical imaging applications specifically. Segmentation partitions an image area or volume into a non-overlapping and connected section, and this process plays an essential role in extracting regions of interest in medical images, providing more intuitive medical information than raw images. Indeed it is a vital aspect in computer-aided diagnosis systems for different applications. Medical segmentation is considered, in fact, a crucial imaging process as it extracts anatomical structures from a medical image with a semiautomatic or automatic process. It splits an image into areas based on a specified description, such as segmenting organs/tissues in the medical images for border detection, tumor, or mass detection. In this context, it represents one of the most exigent medical analysis tasks: to supply critical information about the volumes and shapes of these organs and assist doctors in diagnosing or making decisions. For this reason, the vast investment and development of medical imaging modalities such as microscopy, X-ray, computed tomography (CT), magnetic resonance imaging (MRI), and positron emission tomography (PET) attract researchers to implement new medical imaging methods in segmentation. However, X-ray imaging has been revealed to be one of the most commonly prescribed medical imaging procedures, often more than other imaging modalities. Indeed, these examinations are recognized as a valuable medical tool for various procedures since they are used noninvasively and painlessly to help diagnose disease, monitor therapy, and support medical and surgical treatment planning. Therefore, the X-ray modality is a low-cost imaging technique that has been used for over a century for medical imaging and continues to apply nowadays. Due to low cost and a low dose of radiation, hundreds to thousands of X-rays are generated in a typical hospital daily. This large amount of images creates critical diagnostic workloads for radiologists and medical practitioners. Consequently, it is paramount to develop computer-aided detection methods for X-rays to support these clinical issues where a helpful answer is organ segmentation. Nowadays, in the setting of the coronavirus pandemic, chest X-ray imaging, in particular, plays a significant role in the early diagnosis and treatment planning for patients with suspected or confirmed Covid-19 chest infections. Chest x-rays show characteristic radiographic findings in patients with Covid-19 pneumonia, demonstrating that lung segmentation is functional for Covid-19 identification and stratify disease severity.[2]

2.2 Chest X-Rays imaging

A chest radiograph or chest X-ray (CXR) is an imaging modality to diagnose conditions and diseases affecting the chest and nearby structures. The Chest X-ray (CXR) is widespread in everyday clinical practice, aiming to visualize disorders of the bony thorax and the organs of the thoracic cavity. It is responsible for approximately 30–40% of all X-ray examinations performed, regardless of the level of healthcare delivery.[6] Even though this imaging modality is not costly, readily available, and technically simple, it can reveal a significant amount of information regarding the patient's health, thus providing essential clues for an accurate diagnosis. Over the years, CXR has inevitably faced the rapid progress of medical imaging technology and has been subjected to criticism regarding its poor diagnostic sensitivity, which has to be counterbalanced by an accurate, detailed, and time-consuming interpretation by a radiologist. Indeed it has the advantages of being fast, easy, non-invasive, and inexpensive. Although easy to acquire, a chest radiograph (CXR) is one of the most complex imaging modalities to interpret. Its acknowledged interpretation heavily depends on the level of training and experience of the person interpreting it. Prior studies indicate that medical practitioners in an emergency or out-patient departments frequently miss radiographic abnormalities, and there is considerable discrepancy among their interpretations and those of trained radiologists[7]. Since clinical decisions depend on the accurate understanding of CXRs, it is highly desirable to develop computer-aided diagnosis (CADx) systems that can automatically interpret CXRs and assist medical practitioners in decision-making.

2.2.1 Chest X-rays Imaging : History and applications

The traditional technique used in acquiring and developing a chest radiograph is based on methods similar to those discovered by Roentgen in 1895. Indeed chest X-ray imaging employs electromagnetic waves or better X-rays to create pictures of the inside of the thoracic region. X-rays or X-radiations are referred to as Röntgen radiation after the German scientist Wilhelm Conrad Röntgen discovered it on November 8, 1895. Röntgen brought to light their medical use when he made a picture of his wife's hand on a photographic plate provided due to X-rays. The photograph was the first image of a human body part using X-rays and the incipit of future medical applications. Even the CXR examination involves positioning the patient's chest between the machine that produces the X-rays and a plate containing an X-ray detector that creates the image digitally or with X-ray film. The X-rays occur in a defined room with a movable camera attached to a large metal arm. When this latter is turned on, x-rays travel through the thoracic area and are absorbed in different amounts by internal organs and tissues. Indeed dense bone or materials absorbs most of the radiation. Whereas soft tissue, such as fat, muscles, and organs, allows more x-rays to advance through them. Consequently, bones appear white on the x-ray, soft tissue such as lung tissue shows up in shades of gray, and air arises black. Chest X-rays exploit small amounts of ionizing radiation to produce images. The risks of radiation exposure are related to the increase in imaging tests used rather than immediate consequences. X-radiations could cause a slightly increased risk of cancer in a lifetime but concern shallow risks of short-term side effects. In this light, radiologists use the "effective dose" to measure the risk of dose radiation; the effective dose is a weighted sum of organ doses. It is considered the best dosimetric quantity for estimating the risk of exposure to ionizing radiation [8]. The scientific unit of this measurement is the millisievert (mSv), but there are other units such as radium and ren. This measurement considers how different sensitive tissues are to radiation and enables us to compare it to familiar exposure sources, such as natural background radiation. A single chest x-ray exposes the patient to about 0.1 mSv It is about the same amount of radiation people are exposed to naturally over about ten days. While a chest CT delivers seven mSv,70 times as much.[9] The American College of Radiology advises limiting lifetime diagnostic radiation exposure to 100 mSv, equivalent to about 10,000 chest X-rays, but only 25 chest CT scans [9]. However, radiologists follow the ALARA principle because of accumulated radiation dose over a person's lifespan, which says patients should always receive radiation doses such as "As Low As Reasonably Achievable." [10]

Standard medical radiographs are generally considered very safe with minimal risk, but the technique also has some significant limitations despite these apparent advantages of low dose. These are mainly related to limited spatial resolution and that all anatomical structures within the chest, lying in the path of the x-ray, are projected onto each other. This issue, intended as 'overprojection, involves that some pulmonary lesions are challenging to see or analyze on a chest radiograph, and thus, the typical response is to acquire a CT scan for further analysis. Diseases that are usually easy to diagnose with a chest radiograph are, in fact, pneumonia, pneumothorax, symptomatic pleural effusion, cardiac enlargement with vascular congestion, and symptomatic tumors [11]. In contrast, small nodules, pulmonary fibrosis, and complex disease affecting the mediastinum and lung are often not fully visible and require CT imaging.

2.2.2 Chest X-rays Views

Chest X-Rays image is ordinarily captured with the patient in a standing position. However, different views or projections of the chest can be employed by changing the relative orientation of the patient and the direction of the X-ray beam. The most common positions are posteroanterior, lateral, and anteroposterior. In the posteroanterior (PA) view, the X-ray source is positioned so that the X-ray beam enters through the posterior (back) of the chest and exits out of the anterior (front) part, where the beam is detected. In this view, the patient stands facing a flat surface behind which is an X-ray detector. On the other hand, the lateral view is taken to complement the frontal image. A lateral position is helpful to localize a lesion, together with a frontal one. It allows a 3-dimensional analysis, and in this modality, the patient is asked to turn and lean one shoulder on the plate and raise hands over their head. An anterior-posterior chest X-Ray is generally less preferred because the image of the heart and mediastinum is less clear and focused in this projection. In this view, the patient is standing with their back against the film, and the x-ray source and detector are reversed compared to PA projection. AP chest x-rays are more complicated to read than PA x-rays and are generally reserved for situations where it is arduous for the patient to stand an ordinary chest x-ray, such as the patient is bedridden. CXR can be achieved in bedridden patients using portable machines, which can be transferred to intensive care units. In this context, portable CXR is essential in evaluating critically ill patients who frequently suffer from severe pulmonary diseases caused by pneumonia, heart failure, or respiratory distress requiring mechanical ventilation[11].

In conclusion, chest radiography is still in everyday use. The technique has been constantly improved over the last few decades, and it serves as a baseline examination for lung diseases worldwide and, as such, if wisely used by radiologists, is of great value.

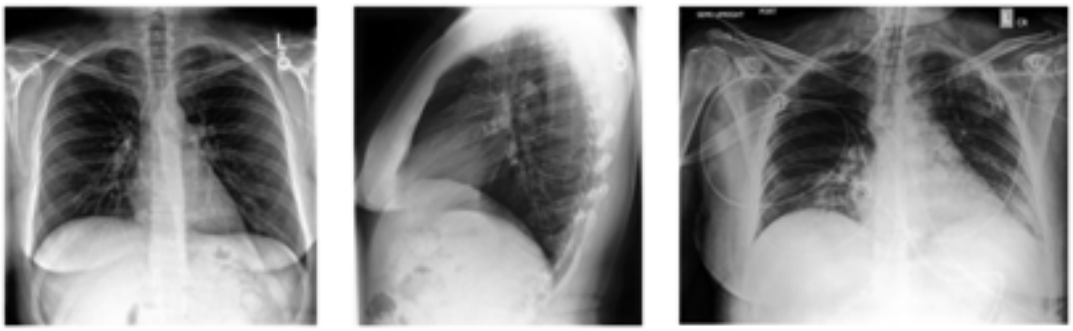


Figure 2.1: : Left: posterior-anterior (PA) view frontal chest radiograph. Middle: lateral chest radiograph. Right: Anterior-posterior (AP) view chest radiograph

2.2.3 Chest X-rays Imaging for Covid-19

The Coronavirus disease pandemic (COvid-19), originated in 2019 by the Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) virus, has become the most significant public health crisis our society has faced recently.Sars-Cov-2 affects mainly the respiratory system and, in extreme cases, causes a massive inflammatory response that reduces the total lung capacity. Because of that, the chest X-ray (CXR) and computed tomography or computerized X-ray imaging (CT) scan became the standard imaging tests for pneumonia and, consequently, Covid-19. These imaging examinations play, in fact, a pivotal role in different settings. Indeed, these imaging modalities are used during triage in unavailability, delay of, or the first negative result of reverse transcriptase-polymerase chain reaction (RT-PCR) and then stratifying disease severity[2]. However, The CT scan is the gold standard for lung disease diagnosis since it generates very detailed images. CXR is very useful in particular scenarios since they render the resulting images faster, exposes the patient to much less radiation, and be more widespread in the emergency care unit. Moreover, X-ray devices are easily accessible and reduce Coronavirus contamination during the imaging acquisition [12]. Chest X-ray (CXR) imaging is employed as an assistive tool in COVID-19 prognosis, and it is reported to have a potential diagnosis capability in recent studies [13]. It can be used as a possible chance for the early identification of Covid-19 patients at risk of severe outcomes. Furthermore, chest radiography of seriously ill people with respiratory symptoms when they present to the hospital can help identify those with Covid-19 pneumonia. Hence this screening is essential for evaluating the status of the patient and treatment. Like other cases of respiratory diseases, Covid-19 pneumonia causes the density of the lungs to increase. It reveals whiteness on radiography but depends on the severity of pneumonia. It obscures the lung markings that are customarily observed; however, this may be delayed in appearing or absent. When the increased whiteness partially covers lung markings, a ground glass pattern (ground-glass opacity, fig 2.2) occurs. This issue can be subtle and might need confirmation from a radiologist in CXR examinations. [14] On this premise, covid-19 pneumonia can be classed as atypical pneumonia because of the radiographic appearances of multifocal ground-glass opacity, linear opacities, and consolidation. Indeed these changes are also seen in other atypical cases of pneumonia, including other coronavirus infections (severe acute respiratory system, Sars, and Middle East respiratory syndrome, Mers). A quantitative meta-analysis covering 2847 patients in China and Australia, and a global descriptive analysis of 39 case report articles summarising 127 patients, found that Covid-19 pneumonia changes are mostly bilateral on chest radiographs (72.9%, 95% confidence interval 58.6 to 87.1) and have ground-glass opacity in 68.5% of cases (95% CI 51.8 to 85.2).[15]

On this basis, several studies were proposed to investigate the development of segmentation models for the opacity regions for the Covid-19 positive chest X-rays and the impact of lung segmentation in identifying pneumonia in CXR images. In this setting, a relevant study demonstrated how, in chest X-rays of the Covid-19 positive patients, the percentage of the predicted and segmented opacity region over the total lung area could indicate the patient severity regarding ICU admission and mortality. [16] While other studies, in a real-world application, showed how segmentation is an essential step since it removes background information, reduces the chance of data leak, and forces the model to focus only on meaningful image areas. Segmentation in this way improves the classification performance, as it increases the model's reliability and quality. Indeed . in these methods, the best identification rate of COVID-19, among other types of pneumonia, is achieved using segmented CXR images instead of raw images, demonstrating the importance of lung segmentation in CXR image classification algorithms[1]. Also, Further studies proposed to perform lung segmentation as the first step in their approaches[17],[18]. They evidenced that narrowing the region of interest down to the lungs' areas increased the trustworthiness of their methods. Thus, segmentation can aid in more reliable detection of Covid-19 by extracting the region of the lungs and left out of analysis areas that do not belong to the region of interest (ROI). In this context, our thesis work aims to investigate the lung segmentation process to contribute and support research against Covid-19. Because of that, we implemented the proposed method on a series of COVID-19 positive patients' CXR images to see if our approach could lead to improvements over the state-of-the-art in the Covid-19 research field.

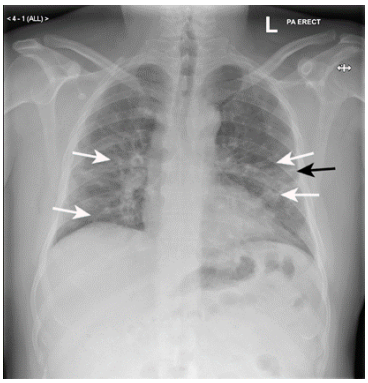


Figure 2.2: CXR Image of Covid-19 patient : glass opacity

2.2.4 Chest anatomy in CXR images

Accurate information of the lung shape and its anatomical variations is very noticeable in medical imaging and is paramount for proper segmentation. Therefore, an in-depth description of lung shape and their anatomical structures in CXR imaging appears needful and explanative in this perspective. The lungs are pyramid-shaped, paired organs connected to the trachea by the right and left bronchi; on the inferior surface, by the diaphragm. The diaphragm is the flat, dome-shaped muscle situated at the base of the lungs and thoracic cavity. Furthermore, The lungs are enclosed by the pleurae, which are connected to the mediastinum. The right lung is shorter and wider than the left lung, and this latter takes a smaller volume than the right. Indeed due to the elevated position of the liver, the left lung is slightly more extended than the right one. Furthermore, the left lung shares space with the heart and has an indentation in its wall called the cardiac notch of the left lung. In addition, the front and outer sides of the lungs face the ribs, which form light indentations on their surfaces. The medial surfaces face towards the center of the chest and lie against the heart, great vessels, and the carina, where the trachea branches into the two main bronchi. Moreover, both lungs have a recession termed as the hilum at the root of the lung, where the blood vessels and airways permeate the lungs.[17]

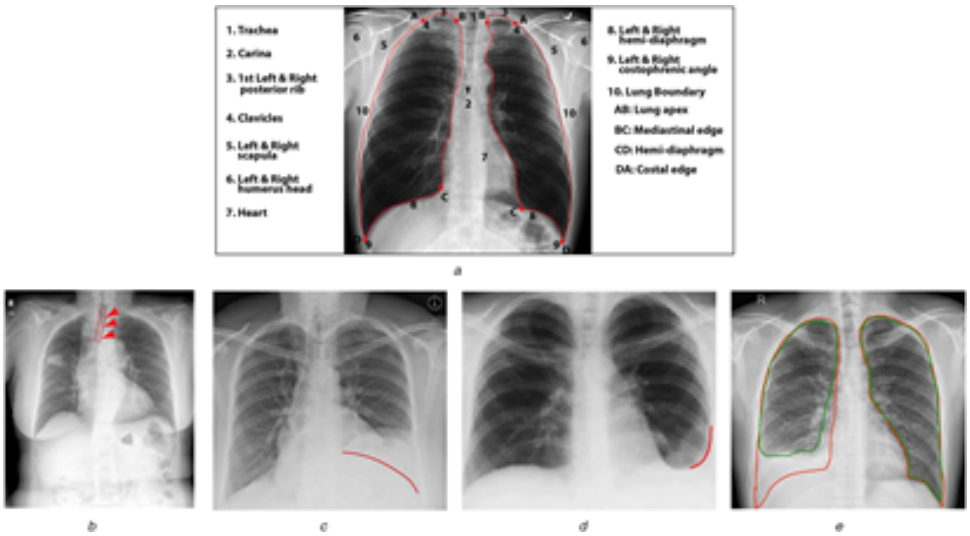


Figure 2.3: (a) Normal labeled CXR, (b) CXR indicating the tracheal shift toward the right, (c) Raised left hemidiaphragm due to left phrenic nerve palsy, normal position delineated in red, (d) Blunt CP angle indicating pleural effusion, delineated in red, (e) Two different ways of delineating lung boundary in abnormal cases

In CXR images, anatomical structures such as the heart and lungs are seemingly visible (Fig. 2.3 a), but others are obscured or occluded. Because of that, to effectively understand and develop computerized methods for processing and segmenting CXRs, firm knowledge about the visual characteristics of anatomical compartments is essentially required. The following section provides a brief overview of the visual appearance of anatomical structures on a CXR, In medical literature, many different mnemonics have been described for looking at a CXR, but the commonly used 'ABCDE'[18] is as follows:

A. Airways: the airways contain air and are of lower density as compared with the surrounding soft tissues. Thus, they appear darker on a CXR. The trachea is the vastest airway and is present midline in a normal CXR. Tracheal shift (Fig. 2.3 b) indicates abnormalities such as pleural effusion, pneumothorax, atelectasis, and pleural fibrosis. Thus, several CADx systems [20,21,22] locate the upper end of the trachea and its centerline and delineate it to detect abnormalities. The trachea divides into the left and right bronchus at the carina, The carina is a ridge of cartilage surrounding the area posterior to where the aortic arch crosses to the left of the trachea. Its detection is critical for CADx systems designed for pulmonary embolism. Although lung landmarks such as the trachea and carina are of great significance in a chest radiographic CADx system, generally remove the mediastinal region containing them before producing the final segmented output as a binary mask for visualization to prevent the two lungs from touching each other.

B. Bones: Bones are dense and thus appear brighter than soft tissues on a CXR. The bones visible on a CXR include clavicles, ribs, part of the spine, scapula, and the proximal humeri. Many CADx systems use bones (mainly the ribs and the clavicles) as markers to automatically detect rotations in CXRs [24,24], to accurately register temporal CXRs [27], and to assess the level of inspiration in an image. Bones can be segmented using dual-energy subtraction radiography [28] or using the software methods described in [26,27]. Various CADx systems perform bone segmentation and suppression to significantly improve radiologists' accuracy in detecting lung nodules [28,29,30] and focal pneumonia [32] in a CXR image.

C. Cardiomediastinal region: Cardiomediastinal region is an area between the lungs that consists of significant blood vessels, trachea with stem bronchi, muscular esophagus, thymus gland, and the heart. This region is denser as compared with the surrounding air-filled lungs and thus appears brighter on a CXR. Although most of the constituents of the mediastinal region, except the heart, are not visible on a normal CXR, the cardio-mediastinal profile plays a vital role in diagnosing various diseases. Various CADx systems look for subtle deviation in the mediastinal profile for detecting cardiomegaly and pericardial effusion [34]. Mediastinal contours are also traced by various rule-based methods for the effective segmentation of the lung field [33].

D. Diaphragm: The diaphragm separates the relatively denser abdominal region below it from the less dense lung region. Each hemidiaphragm should appear as a crisp dome-shaped contour on a normal CXR. Obscured, flattened, or raised hemidiaphragm (Fig. 2.3 c) indicates abnormalities such as consolidation of lower lobes, congenital diaphragmatic hernia, phrenic nerve palsy, and lung hyper-expansion. The angle between the diaphragm and the chest wall at the bottom of the lung, known as the costophrenic angle, also provides critical information. Blunt costophrenic angle (Fig. 2.3 d) helps detect pleural effusion due to pneumonia, cirrhosis, pulmonary embolism, and tuberculosis (TB). [36,37,38] angle is crucial for many CADX systems to detect pleural effusion in CXRs.

E. Expanded lung: CXRs are generally acquired after the patient inhales and holds the breath. The expanded air-filled lungs are less dense and thus appear darker on a CXR. In standard cases, a lung boundary (labeled as 10 in Fig. 1a) consists of lung apices (AB in Fig. 1a), mediastinal edges (BC in Fig. 2.3 a), hemidiaphragm (CD in Fig. 2.3a), and coastal borders (DA in Fig. 2.3 a). The mediastinal region, heart, aorta, and the structures below the diaphragm are not included in the lung field. However, there is a difference in opinion amongst radiologists about what constitutes the lung field in abnormal cases. Some consider only the air cavities as a part of the lung field (as shown in green in Fig. 2.3 e), while others believe the air cavities and the area obscured by the pleural fluid as a part of the lung field (as shown in red in Fig. 2.3 e) [37]. Most of the CADx systems and Lung Fields Segmentation methods segment only the air cavity as the lung field and use its shape as a feature to detect the abnormality.

2.3 Segmentation methods of lung fields in chest radiographs

2.3.1 Introduction

Lung field segmentation is undoubtedly the first and mandatory step of an automated system aimed at any computer analysis on chest radiographs. Indeed it takes priority importance as it precisely defines ROI in which specific radiologic signs such as pulmonary opacities, cavities, Kerley B lines for pulmonary edema, and lung nodules are searched[33,39]. Moreover, the segmentation algorithms used to identify these abnormalities being applied to the lung area result in more sensitivity. Besides, lung field segmentation outlines other quantitative parameters such as size, shape, texture, and volume are directly used as an indicator of various medical conditions. Because of that, a great deal of research has focused on lung field segmentation in postero-anterior chest radiographs. However, several theoretical studies in the past decades demonstrated that the automatic segmentation of the lung fields is a complicated problem from a computer vision point, and related reasons have been highlighted. For example, research by Vyborny [40] explains how chest radiographs are among the most challenging radiographs to produce technically and interpret diagnostically. Firstly, they reported how the habitus, the position, and finally, the level of inspiration of the person in the moment of acquiring the radiographs could originate considerable variations in radiographs of the same patient; it significantly impacts the visibility of some parts of the lung. Secondly, the setting of the chest unit, particularly the peak tube voltage, determines the visibility of several structures such as the bones and the vessels [41]. Finally, the last but most important cause of problems is that radiographs are projection images and represent all superimposed structures. Consequently, the segmentation task is then misleading if anatomical compartments are superimposed to the natural borders of the lungs. Examples are the hilum, where the arteries and veins enter the lungs, bony structures such as the clavicles, the spine, and the shoulder-blade, that are often better visible than the rib cage boundaries, the heart, the diaphragm, that is constantly hiding the costophrenic angles, and the intestinal gases in the left lung area. As a result, knowledge of these themes is strictly needed to recognize all the listed problems and solve or avoid the errors caused by their effects. Another critical issue about lung segmentation is the definition of the lung area to be detected and delineated. All methods presented in the literature aim to define the most visible parts of the lungs, those not hidden behind the diaphragm, the heart, and the spinal column. Indeed the segmentation process identifies an optimal area when the abnormalities or pathologies detected by the computerized system, such as tuberculosis, are restricted to the visible lung area. It is better to limit the processing to this search area since

it is less extended hence needing less computational costs. Nevertheless, this visible area is not enough for lung nodules detection systems since tumors may be present in the not visible parts of the lungs, and in this context, they would be lost by the cad system. Examples of the visible lung area extracted by the algorithm presented by Ginneken [43]. they are thorax images of patients with lung tumors, evidencing that some nodules are outside the segmented area. In this perspective, innovative studies are focused on methods to include the retro-cardiac region in the CXR-based lung segmentation for accurate lung capacity estimation. [42]

2.3.2 Segmentation methods

The main approaches that have been presented in the literature for solving lung segmentation at state of the art are:

- Rule-based methods
- Pixel-classification methods
- Model-based methods
- Hybrid methods
- Deep learning methods

According to this classification, based on Shi et al. [45], these categories consider deep learning methods, although they could be listed in pixel classification, in a separate group due to their surprising performance in computer vision. The rule-based systems are composed of algorithms performing consecutive steps, each containing specific processing with certain adjustable parameters, and employ heuristic assumptions to locate the lung region. On the other hand, Pixel classification methods are based on the classification of the pixels of the image into two classes, "lung pixels" and "not lung pixels"; this process is based on a set of features calculated for each pixel itself. Model-based segmentation is less common and uses both low-level appearance and shape analysis to identify the lung area. In hybrid methods, the best parts of the schemes are combined to produce a better approach to overcome the challenges of lung boundary detection. In the end, deep learning methods are based on deep neural networks, which process image data through convolutional layers and learn feature representation hierarchically, starting from low-level to more abstract representations. Nowadays, These are the most prominent state-of-the-art approaches.

2.3.3 Rule-Based Methods

The algorithms in this category set sequential steps and heuristic assumptions on imaging characteristics to segment the lung field. Several methods have been proposed that segment the lung field by using the rules based on its position, intensity, texture, shape, and structural relationship to other anatomical structures, Generally, several rules are applied in sequence to obtain the desired result. The beneficial aspect of these methods is that the constituent regulations can be used in different orders, giving notable freedom [46]. The general workflow pipeline of the rule-based Lung field segmentation methods is shown in Fig. 2.4. According to the applied rules, the rule-based methods are further classified as (i) intensity thresholding-based methods, (ii) edge-detection-based methods.

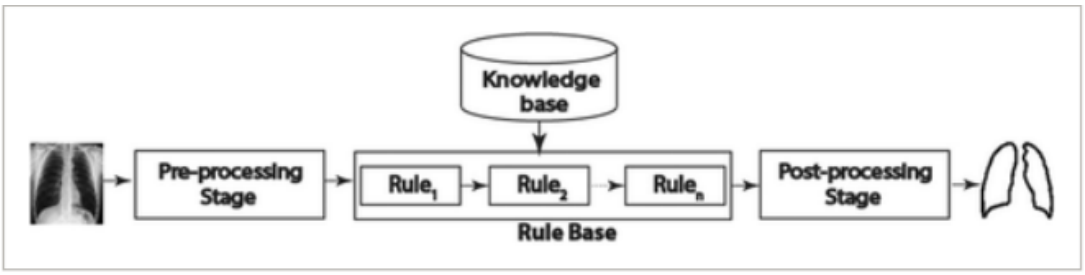


Figure 2.4: General Pipeline of Rule-based Methods

Intensity thresholding-based methods

Intensity thresholding-based methods classify each pixel of a CXR image into the lung or non-lung region based on a grey-scale threshold. The most considerable challenge in the successful application of these methods is determining a reasonable threshold value. One of the earliest methods belonging to this section is suggested by Cheng and Goldberg.[47] It involves the sequential implementation of three rules. First, horizontal and vertical projection profiles (HVPPs)(note) of a CXR image are taken to determine minimum rectangular frames enclosing the left and the suitable lung regions. After that, a grey-level histogram thresholding technique that uses a pre-determined static threshold value is applied to the rectangular frames to delineate lung boundaries. Ultimately, the top and side segments of the lung boundaries so achieved are refined, employing parabolic curves and straight lines. This method is swift and computationally inexpensive but lacks adaptiveness due to a static global threshold. For the same reason, it fails to perform reasonably well on CXR images with different contrasts. Instead of using one static global threshold value, Armato et al. [49] performed grey-level histogram analysis to initialize a range of threshold values for iterative global thresholding. The initial lung contours deduced by global thresholding are refined using more local thresholding and the large-scale aberrations like depressions or protrusions in the lung contours by applying morphological operations with a circular structuring element. The segmentation accuracy of the method improved because of the usage of an iterative thresholding process. However, iterative thresholding makes the approach computationally inefficient and slower. In conclusion, the intensity thresholding-based methods are fast and do not require further information about the morphology and location of lungs in a CXR. However, these procedures are elementary, fragile, and do not perform well on CXRs with poor contrast.

[1]

[1]HVPP: Horizontal and Vertical Projection profile is calculated separately for different axis. Projection profile along vertical axis is called Vertical Projection profile. Vertical projection profile is calculated for every column as sum of all row pixel values inside the column. Horizontal Projection profile is the projection profile of a image along horizontal axis. Horizontal Projection profile is calculated for every row as sum of all column pixel values inside the row.

Edge-based Methods

These methods leverage contrast differences between the lung and non-lung regions and determine lung boundaries by detecting abrupt changes in the intensity values of CXR image pixels. However, numerous pre-and post-processing operations are applied in a quest to obtain smooth and continuous boundaries. Li et al. [50] analyzed the first derivatives of HVPP to determine lung apex, costal, mediastinal, and hemidiaphragm edges. The reliance of the derivative approach on local information generates several false lung boundary points leading to discontinuous and disorganized lung boundary instead of global. The detected border is then processed using t and edge tracing and iterative boundary point adjustment procedures to form a smooth continuous boundary. Despite all the heuristics hypotheses, the method encounters difficulties in detecting mediastinal and hemidiaphragm and mediastinal edges due to their complex patterns and structure of grey-scale distribution. As in [48], Duryea and Boone [51] also analyzed HVPP to determine the potential locations of mediastinal and lateral edges. Nevertheless, they reduced the image resolution by a factor of 32 to limit the number of data pixels while at the same time preserving enough resolution to identify the lung field successfully. It has been subjectively evaluated that the accuracy of the just discussed method is at par with hand-drawn segmentation done by a human observer. However, the utmost limit of the method is that the segmentation uncertainties increase when the segmented regions are expanded to complete resolution. Iakovidis and Papamichalis [52] applied a similar approach to detect lung boundaries in portable CXRs. As in [49,50], edge points on the anatomical structures around the lung field are delineated using HVPP analysis and intensity thresholding. However, instead of the edge tracing technique, Bézier curve interpolation approximates the lung contours. The lung field segmented by this approach includes overlapped heart region, which is typically excluded by other methods. Pietka [53] performed histogram analysis to find a threshold value adopted to eliminate the bright sub-diaphragm, mediastinum, and parts of the thorax. HVPP gradient analysis is then performed to find the lung boundary. Finally, the detected lung border is smoothed using cubic spline interpolation and morphologic erosion procedure. These two methods exploit a single global threshold and fail to correctly delineate lung boundary in a chest radiograph with poor contrast between the lungs and other anatomic structures.

The processes relying on HVPP gradient analysis tend to poorly segment hemidiaphragm edges, especially the left one, due to the complicated stomach gas structure and cardiac boundaries. Xu and Doi [54] presented an improved method for detecting hemidiaphragm edges and lung field segmentation. They eliminated false edges identified by the gradient analysis using structural relationships between seven categories of anatomic landmark points, whereas the remaining edges are smoothed with polynomial functions. Instead of using fragile edge tracing, false edge elimination, and edge smoothing procedures in the latter stages, Ahmad et al. [88] used the fuzzy C-means (FCMs) clustering method to refine the lung boundary. In this method, a Gaussian derivative filter is applied at seven different orientations to identify a rough lung outline. First, the lung outline is filled with global thresholding and later is refined using FCM clustering. Instead of basing on a static threshold, the method relies on an adaptive threshold determined by the Otsu method [55].

Edge-detection-based methods have the advantage that they do not require any knowledge of lung anatomy. Nevertheless, these methods are sensitive to noise. Therefore, even if noise is filtered using a Gaussian or a similar mask during the pre-processing stage, the outcome is highly dependent on the size of the filter and the threshold value used.

2.3.4 Model-based Methods

Model-based methods are segmentation algorithms that search image-specific characteristics as closed, smooth contour or determined curves that respect used models' shape constraints. The latter intends to shape an object or anatomical part as a flexible two-dimensional (2D) curve or a three-dimensional surface that can be deformed under the influence of internal forces, external forces, and user-defined conditions. The internal forces, which retain the model smooth during the deformation, are defined within the curve or surface. The external forces, which direct the model toward the object boundary or other desired features within the images, are computed from the training or image data. Constraints about the appearance, position, shape of the object are manually specified or derived from the training data. These models offer promising segmentation results even under dreadful variability in object shape and image quality and are incredibly successful when the structure is delineated similar shapes across many subjects in the training data [56] These methods are classified as (i) parametric models and (ii) geometric models. A general pipeline is illustrated in the following Fig 2.5

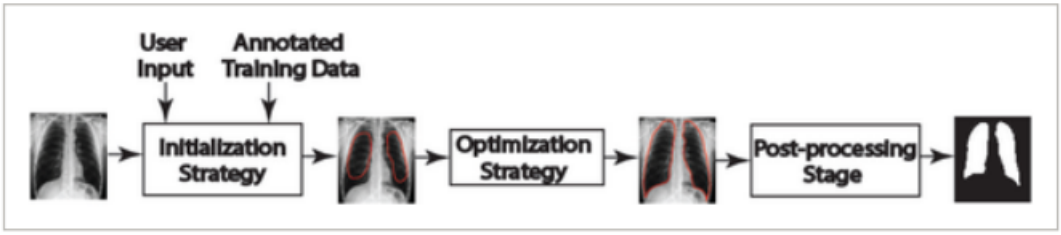


Figure 2.5: General Pipeline of Model-based Methods

Geometric Models

Geometric models represent surfaces or curves implicitly as a level set higher-dimensional scalar function. Two-dimensional curves are characterized by a level set of a function of three variables, i.e., a surface in 3D space. A plane, for instance, a 'zero level set,' intersects this surface and gives the desired contours. A significant advantage of this model over the parametric model is that a single function can describe more than one curve (see Fig. 6 b). Besides, these models are numerically stable, do not yield self-intersections, and naturally handle topological changes. Geometric models based on the level set method [56] have considerably succeeded in lung field segmentation. The advantages of using the level set method are that it is accurate, robust, and can handle sharp corners and cusps in evolving curves effectively. Furthermore, the process provides increased accuracy with the computational cost at par with other methods. Lee et al. [58] presented an unsupervised method based on a multi-resolution fractal feature vector where HVPP analysis is executed to gather ROI. Features from the ROI are extracted using the Gabor wavelet transform and fractal dimension calculation method. Fuzzy C-means clustering is performed on the extracted features to obtain the initial lung boundary, refined by applying the regularised level set evolution method. The approach is tested on a small dataset, and its performance is even lower than the default ASM method. Graph cuts is another optimization method in which objective function is minimized, similar to the level set method. For segmentation, the objective function is defined in terms of the boundary, region, and object model properties. Candemir et al. [59,86] presented a registration-driven Lung segmentation method based on graph cuts. In this method, a content-based image retrieval approach is used to find training images with lung shapes similar to that in the patient's image and create lung atlas. SIFT-flow non-rigid registration technique is then adopted o create an initial lung model using the patient-specific lung atlas. Final segmentation is achieved using graph cuts discrete optimization method with modified energy function. The technique is one of the state of the art methods for lung segmentation and can successfully segment challenging regions such as the apical and the clavicle. However, its performance near the costophrenic area can be further improved. Indeed Candemir et al. [59] have generalized this technique for pediatric CXRs as well.

Parametric Models

These techniques, conceived by Kass, Witkin, and Terzopoulos in 1987 and known as "snakes" [60], are based on parameterizing the contour and then evolving each element according to image and internal terms. Such techniques are fast and efficient. However, the original "purely parametric" formulation is mainly criticized for its limitations regarding the choice of sampling strategy, the internal geometric properties of the curve, topology changes (curve splitting and merging), and addressing problems in higher dimensions. A parametric model or, more simply, snake involves an energy minimizing, deformable spline influenced by constraint and external forces, resumed by the image, that pulls it towards object contours and internal forces that resist deformation. Snakes may be understood as a particular case of the general technique of matching a deformable model to an image employing energy minimization. In two dimensions, the active shape model and the active appearance model constitute a discrete version of this approach, taking advantage of the point distribution model to constrain the shape or appearance range to an explicit domain acquired from the training set. These methods employ a small set of parameters to form a model representing the preliminary analysis of the geometrical shape and appearance of the object to be delineated. The parametric models iteratively interact with the image features by adjusting their parameters according to the image forces. When the internal and external forces balance each other, an equilibrium is attained that delineates the segmented output. The convenience of these models is that they exploit compact representations and are very fast in their convergence. Nevertheless, their major limitation is that the adaptation of the model topology, such as merging or splitting parts during the deformation, is impossible. Hence, the model can only detect a single object. For multi-object segmentation, multiple models have necessarily to be initialized, one for each object. The parametric model-based methods are further categorized as (i) Active shape models and (ii) Active appearance models.

Active shape models

Active shape models (ASMs) are statistical models of the shape of objects which iteratively deform to fit an example of the object in a new image, developed by Tim Cootes and Chris Taylor in 1995. Regarding lung fields segmentation, in [61], Van Ginneken et al. modified the ASM method [62] and reported 'ASM with optimal features' for segmenting lung fields CXR images and cerebellum and corpus callosum from magnetic resonance imaging brain images. To segment the lung field, they constructed an initial shape model by manually marking a collection of landmark points on CXR. The landmarks are iteratively shifted to their best positions such that one side of the curve, formed by interpolating the landmarks, contains local optimal image features belonging to the lung region, and the other side does not. The local optimal image features are obtained from the training images using feature selection and a nonlinear kNN classifier. Choosing to use the optimal local features over the normalized first-order derivative profile used in the traditional ASM method appreciably improved the method's performance. However, its computational complexity is much higher than that of the traditional ASM method.

Yuan et al. [63] reported an image segmentation method in which gradient vector flow (GVF) is applied to steer the shape of a randomly initialized curve. The control points are updated using the magnitude and direction of the gradient vectors. Indeed, since the GVF of an image is independent of the evolving curve, the curve does not get trapped in local optima. Xu et al. [64] adapted the method proposed in [63] for lung field segmentation. They found that the point evolution equation employed in the GVF method is linear and inefficient for segmenting blurred lung boundaries. They revised the GVF-based ASM method by using a nonlinear global GVF function for model fitting. Because of this, adaption the overlap measure of the proposed method increased by 3% from that of the GVF-based ASM approach and by 5% from that of the ASM-based method. In the ASM-based methods, the landmark points used to create the initial lung boundary are manually selected. The authors of the following ASM-based methods[66] applied additional pre-processing steps to obtain the landmark points automatically. Iakovidis and Savelonas [65] presented a method in which the landmark points are automatically identified using rib cage endings. Subsequently, these points are interpolated using Bézier curves to set the initial lung boundary. The ASM model is then implemented to evolve the lung boundary and segment the lung field. The method has the advantage of automatically detecting landmark points of the initial lung boundary. However, this detection is highly challenging and erroneous, especially in low contrast CXR images. Furthermore, the method gives poor segmentation results when the lung is hugely deformed or collapsed because the initial lung boundary approximated using rib cage endings will not be near the actual lung

boundary.

Lee et al. [67] presented an innovative way to initialize the lung boundary by fitting the mean shape estimated over a set of training images inside the lung ROI given by exploiting HVPP analysis. The initial lung boundary is adapted by a variant of the ASM technique that uses k-means clustering and shape-based cluster validation technique. This modification does not provide any significant performance improvement. Wu et al. [68] proposed a method in which the initial lung boundary is estimated similarly, where the ASM's objective function is redesigned and consists of a weighted grey-scale appearance model's cost function, an edge constraint, and a distance constraint. The weighted component gives significance to the neighbors along with the directions perpendicular to the lung contour. Distance and edge constraints help improve convergence and prevent the model from getting trapped in the stomach gas region. The method accuracy on the JSRT dataset is 94% and is 4.4% better than the traditional ASM- based method and 2.7% better than ASM with local invariant features. The mean shape used to initialize the lung boundary in the method is population-based and not patient-specific. Thus, the method does not perform well when a specific CXR has considerable lung deformations. Shi et al. [69] introduced a deformable model that applies both population-based and patient-specific shape statistics to segment the lung field. In fact, in the initial phase, the model relies more on population-based statistics, but as it evolves, the patient-specific statistics start playing a predominant character in constraining it. Another central feature of this method is that the deformable model uses scale-invariant feature transform (SIFT) features more distinctive than gradient and intensity features. Nevertheless, the usage of SIFT features increases the computational complexity. Dawoud [70] proposed a method that considers both the intensity-based thresholding and shape information in an iterative framework. Initially, a parametric model for lung shape is derived from the training images. The model is then used to optimize iterative thresholding-based segmentation of the lung field. For a test image, the outputs of iterative thresholding are compared with the initial model based on Mahalanobis distance to achieve the most similar output. At last, ASM is applied to refine the output. The characteristic of the method is that it maximizes the information utilization by combining intensity information with shape one. The method has revealed better results than ASM. Nevertheless, a limiting requirement of the approach is that the resolution of the test image should be the same as that of training images. Zhang et al. [68] proposed a similar method in which intensity information and shape analysis are used to initialize the lung outline, modified using the ASM method. In setting initial landmark points sufficiently close to the target contour, Xu et al. [66] used global edge and region force (ERF) analysis to initialize the lung boundary. The optimization strategy also uses the global ERF field to evolve landmark points and pulls the

landmark points out of local optima. The algorithm's convergence is individuated by specifying either the number of iterations or the Euclidean distance threshold between two consecutive iterations. The method's accuracy on the JSRT dataset is 95.2 and is five better than the traditional ASM-based method. The reliance of the method on the global ERF field rather than local features during the evolution stage is advantageous as it removes the prerequisite of placing initial landmark points close to the target lung contour, and it is unnecessary a manual adjustment of pose parameter.

Appearance-based models

A limitation of ASM is that it considers only the shape of an object and not the appearance, texture, or grey-level variations across the object. This deficiency is filled by an extension of ASM, known as AAM, which concerns both the appearance and shape of the object while constructing preliminary analysis. They are, in fact, more robust than ASM but are more computationally expensive. . Li et al. [72] proposed a segmentation method based on shape and appearance models. In this method, the shape model is initialized by performing HVPP analysis and is iteratively evolved using multi-scale and multi-step-size strategies. This strategy enhances the searchability of the approach and prevents the evolving curve from getting stuck in local maxima. The appearance model (AM) employs multiple features with different weights to describe the lung boundary and increases segmentation accuracy. The multi-scale and multi-step-size strategy augmented with multiple features gives favorable results. However, in some cases, it creates a sham border near the real one due to overlapping ribs inside the lung field. Shao et al. [73] presented a joint shape and appearance-based method for lung field segmentation. A robust shape initialization approach using a learning-based landmark detection method is used to achieve the initial lung model. About 14 landmark points on lung boundary are initialized, and for each landmark point, a specific landmark detector is learned via cascade learning [76]. Instead of relying on a single shape model, the method divides the lung shape model into segments according to local shape variations and spatial positions to form local shape models. Similarly, the lung boundary is split into several segments according to appearance cues and spatial distances to establish a local appearance model. In the end, a hierarchical deformable-based framework is used to integrate scale-dependent shape and appearance information for robust and accurate segmentation.

In conclusion, ASM and AAM do not perform well at widely varying shapes, require proper initialization for a successful convergence, and the output boundary strongly relies on tuning the parameters. Moreover, for lung region segmentation, the algorithm can get trapped at local minima due to strong rib cage and clavicle bone edges.

2.3.5 Pixel classification methods

In these algorithms, each pixel is labeled and considered as a lung or a non-lung pixel using a classifier, like support vector machines and shallow neural networks trained with CXRs and their corresponding lung masks. Since the early 1990s, this kind of machine learning technique has been extensively used to segment various anatomical structures such as the brain, spine, lungs, liver, kidney, and colon [77], and these approaches are based on training a binary or multi-class classifier using a large annotated dataset. Specifically, the training is based on anatomical structure's distinctive shape, appearance, size, pose, location, and structural relationship with other anatomical structures [79]. Therefore the trained classifier then exhaustively classifies each pixel as either belonging to the background or anatomical structure. However, a properly trained classifier responds to the test data with the same error as the training data. Thus, if training data are the correct representation of the test cases, the properly trained classifier labels the pixels accurately and segments the anatomical compartments precisely. Nevertheless, training a classifier properly for anatomical structure segmentation is surprisingly tricky. Apart from the difficulties in achieving sufficient annotated medical training data of variety and high quality, the process of taking relevant features and their representation is cumbersome and resource-consuming.Genarl workflow pipeline is shown in Fig 2.6.

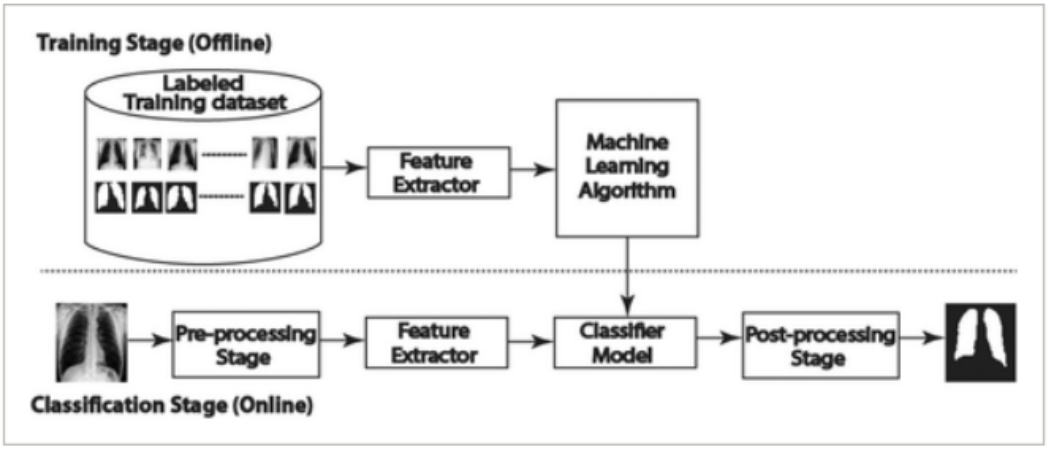


Figure 2.6: General Pipeline of Pixel Classification Methods

In these methods, the feature extraction process is automatic, and the segmentation is based on the possibility of learning from the characteristics of the images themselves. The primary challenge with these methods is determining the appropriate class of features to be extracted and robustly extracting them. McNitt-Gray et al. [77] used linear discriminant analysis (LDA), k-nearest neighbors (kNN), and feed-forward back-propagation neural network classifiers to classify each pixel of a CXR into one of several anatomic classes, namely heart, upper mediastinum sub-diaphragm, lungs, axilla, the base of head/neck, and background. Each classifier is trained on a small custom dataset of 17 CXRs exploiting 59 grey-level-based, local texture-based, and local difference-based features and finally tested on a separate set of 16 CXR images. The percentage of pixels correctly classified by KNN, LDA, and NN classifiers are 70, 70, and 76, respectively. In [78], McNitt-Gray et al. investigated the effect of the quality and number of features on the performance of different classifiers. In order to determine the impact, they individualized eight best features from the 59 implying stepwise discriminant analysis and eight features (not including the best) randomly. Moreover, the classifiers are re-trained using the selected feature sets, and their performance is evaluated.

They realized that the performance of classifiers trained using eight best features is comparable with their respective classifier trained with the complete feature set, while the classifier's performance using eight randomly selected features is substandard. The feature selection step also saved a substantial amount of training time. These findings stress the profound impact of features on the machine learning algorithm's ability to create a good classifier model. Instead of using these methods, Tsujii et al. [79] used adaptive-sized hybrid NN (ASHNN) along with relative pixel addresses, normalized density, and histogram equalized entropy features. The network is trained on 14 images and evaluated on 71 images acquiring 92% accuracy. The employment of a small training set led to the overfitting of the classifier. Nevertheless, the use of ASHNN is advantageous as it can learn quickly and automatically adjust the number of layers. Similar to [79], Vittitoe et al. [80] segmented a CXR into the lung or non-lung region using spatial and textural information. However, their approach is diverse regarding how this information is modeled. They used Markov random field (MRF) modeling to incorporate spatial and textural information via potential functions parameterized, implying spatially varying probability distributions. These distributions are empirically deduced from a training set consisting of 115 CXRs. In conjunction with MRF modeling, iterative conditional modes techniques are used to classify each pixel of the CXR optimally. Then when evaluated on a separate test set of 115 CXRs classified 94.8% pixels correctly, which is better than [79, 81].

It performed better because the MRF model provides a flexible and powerful way to capture textural and local spatial information [83]. Vittitoe et al. [80] demonstrated the flexibility of the approach by segmenting the CXR into six anatomical regions lung, sub-diaphragm, heart, subdiaphragm, body, mediastinum, and background. In multi-class anatomical segmentation, the approach classified 90% of the pixels correctly. Although MRF modeling makes the approach robust and flexible, it is computationally very complex and costly. The aforementioned PC-based methods are supervised and need an extensive annotated training set for generating a classifier model. In place of performing supervised classification, Shi et al. [84] adopted an unsupervised approach to segment the lung field in a CXR. They employed Gaussian kernel-based FCM with spatial constraints to segment the lung field. The method is tested on 52 CXRs taken from the JSRT dataset and attained 97.8% accuracy. Although FCM clustering is ductile and has a solid mathematical background, it can give inaccurate results when applied to medical images in the presence of noise, artifacts, and intensity inhomogeneity.

In contrast to deep-learning-based methods, the feature engineering for these approaches is more intuitive. The performance of methods entirely relies on the quality of distinctive features and is highly variable.

2.3.6 Hybrid methods

In these methods, the best parts and features of the previous approaches are combined to produce a better algorithm to overcome the challenges of lung boundary detection. For instance, in [83], deformable models and pixel classification approaches are combined with majority voting, and a better boundary detection performance is reported. In [84], an atlas-based system is used in which the model atlases are registered to the patient CXR using the SIFT-flow algorithm and combined with graph cut boundary detection. Furthermore, Tuan Anh Ngo and Gustavo Carneiro[87] proposed a new methodology for lung segmentation in CXR using a hybrid method based on a combination of distance regularized level set and deep structured inference. Using the publicly available Japanese Society of Radiological Technology (JSRT) dataset, they show that this approach produced accurate lung segmentation results in the field. Wan Ahmad, W Mimi Diyana, and W Zak [88] realized a novel method based on oriented Gaussian derivatives filter with seven orientations Fuzzy C-Means (FCM) clustering and thresholding to refine the lung region. However, there are other hybrid methods for lung segmentation, but they can be summarized in their intent to combine the advantageous characteristics of the previously described approaches with improving state-of-art performances. Therefore, their consideration is paramount for accurate lung segmentation.

2.3.7 Deep learning Methods

Thanks to advancing computational power and large amounts of data, deep learning has become the default machine-learning method [89] used because it can determine significantly higher complex patterns than traditional machine-learning methods or segmentation methods. Deep learning has been a determinant and vital asset to foster Artificial Intelligence (AI) in the recent few years. It has performed remarkable or yet superior human-level performance on image classification [90], speech identification [91], and reading knowledge [93]. Therefore, this is particularly essential for the field of medical imaging analysis. In recent years, with the warming up of deep learning research, Convolutional Neural Network (CNNs) have attracted researchers tremendously in image processing. Its high efficiency in the patterns extraction model generated by learning makes it highly accurate for image segmentation. In all deep learning approaches, CNNs are of prime concern. By utilizing confined connectivity patterns, such as those employed in the ImageNet competition Krizhevsky et al.[94], CNN's have fast enhanced the state-of-the-art approach for image processing. The current CAD systems must provide accurate and efficient diagnosis and deal with various medical images.

With advances in GPU technology, computer vision systems designed with deep neural networks trained on an amount of data have been shown to produce more accurate results than conventional approaches. In deep learning algorithms, input data are processed through deep convolutional layers, which learn feature representation hierarchically, starting from low-level to more abstract models. Especially, convolutional neural networks (CNNs) have received considerable attention in image analysis problems since they preserve the spatial relationship between the image pixels. Deep neural networks and convolutional neural networks use, de facto, multiple linear and non-linear processing layers to perform pixel classification . However, no consensus has been among the researchers on the minimum number of layers that a network should prevent from being classified as a shallow network rather than a deep network. Schmidhuber [97]labels networks with a depth greater than two hidden layers as 'deep networks ' and more profound than ten as 'very deep networks.'

For a given problem, the depth of a deep network or the number of hidden layers is established intuitively and is one of the paramount hyper-parameters, and it is a quantity that cannot be learned from the data. Fig 6b illustrates a deep network, with a depth of four hidden layers, in which each successive layer of the network learns features from the previous layer to learn more complex features. It is practically impossible to extract such complex features intuitively, and thus, deep-learning-based feature extractors have effectively replaced hand-crafted feature extractors. The most commonly used deep learning model, the deep convolutional neural network (D-CNN), has outperformed state of the art in many visual recognition tasks [2], and has been used in the medical image analysis domain for brain tumor segmentation [3, 4], pancreas segmentation [5, 6], detection of mitotic cells [7-8], knee cartilage segmentation [9], pulmonary embolism, and detection of chest pathology [10, 11] during the recent years. Despite the popularity and acknowledgment of deep learning algorithms in medical imaging, not so many studies have been reported in the literature for lung field segmentation in CXRs. A recent study uses a semantic segmentation approach [12] in which the input is a CXR image and the output is a map indicating the lung region of each pixel. In [13], researchers proposed using fully convolutional networks (FCN) [14] for segmenting lung and heart regions. A full convolution network is an encoder-decoder architecture. The encoder models and patterns the semantic information in the image; the decoder recovers the location information lost during the pooling process and produces a map containing the lung region probability of each pixel. FCN makes, in fact, a rough map due to its elementary decoder architecture. Therefore, researchers [13] applied architectural modifications and adjustments by adding a drop-out layer after every convolutional layer, re-ordering the feature maps, and replacing pooling layers with convolutional layers. Although the application of CNN for lung field segmentation dates back to 1994 [91], Kalinovsky and Kovalev [110] recently employed SegNet, a semantic pixel-wise image labeling network, for lung boundary detection. They used a neural network is a four-layered encoder-decoder CNN (ED-CNN), having four encoders and four decoders. The encoders map high-resolution images to lower resolution feature maps, while decoders perform vice versa. In comparison, a fully connected softmax layer performs the classification. When tested on a dataset consisting of 354 images from two sources, the network gives DSC 96.2%. The most significant advantage and aspect of the method are that the generated output is the same size as the input image. Long et al. [112] retained the advantageous feature of SegNet and built a fully convolutional network (FCN) that takes an input of the arbitrary size and produces an output of the same size. In this application, the contracting layers in FCN are supplemented by successive layers, in which upsampling operators replace pooling operators. The proposed network architecture has multi-resolution layer combinations that appreciably improve the state of the art while speed-

ing up and simplifying the learning and inference. Ronneberger et al. [113] modified FCN by bringing symmetry in contracting and expanding paths. In this way, they created a U-shaped architecture and named it as U-Net. U-Net vastly augments the training images by applying elastic deformations, drastically reducing the number of images required for training and the dimension of the training set. It outperforms the prior best available method in segmenting neuronal structures in electron microscopic stacks. In recent years another interesting deep learning approach based on generative models was investigated for lung field segmentation. Indeed, in [114], researchers proposed using a generative adversarial network (GAN) for lung field segmentation in CXRs. GANs concerns two networks: a generator and a discriminator. For segmentation problems, the generator produces artificial lung masks using manually delineated lung regions; the discriminator produces an extrapolated probability if the map is synthetic or from a ground-truth mask set. According to the probability, the discriminator guides the generator to generate masks more similar to the ground truth masks. This generative approach achieves the appreciable dice coefficient percentage to 95,1% on the JSRT dataset. All proposed DNN-based approaches perform as well as inter-observer performance for lung region detection. However, these methods need an enduring training process and require high computational complexity. Moreover, the satisfactory performance of a deep learning approach is given thanks to the large set of annotated data. It is a critical issue, especially for the acquisition and manual segmentation of CXR images.

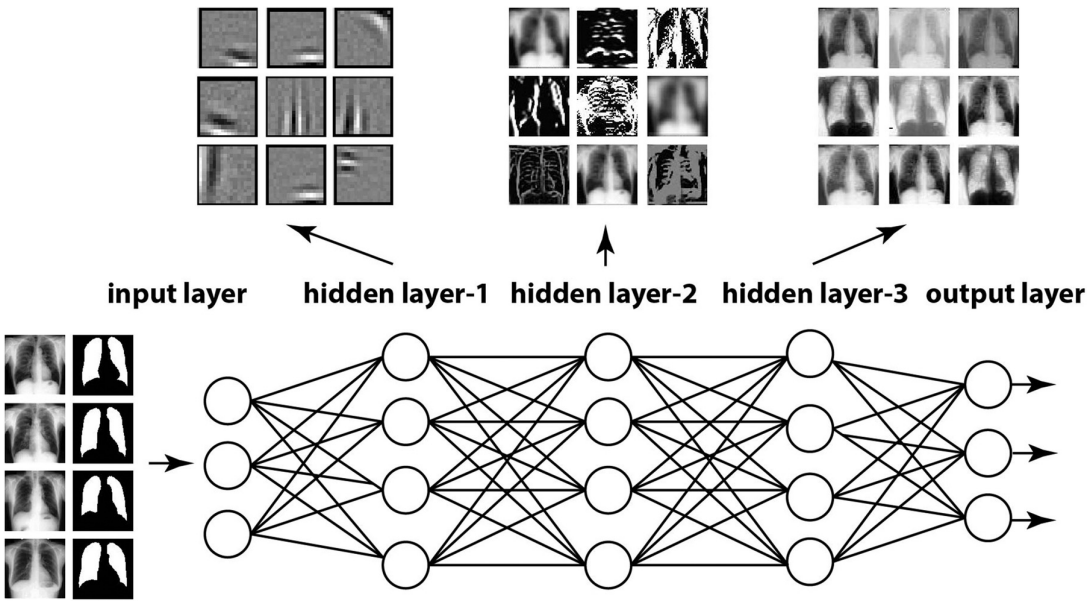


Figure 2.7: Example of hierarchical feature learning by a deep neural network

2.4 Use of apriori information

2.4.1 Introduction

In Latin, apriori means "what comes first." A priori understandings are the assumptions that come before the rest of the assessment, argument, or analysis. Apriori information encompasses all that knowledge that is extrapolated or learned before any cognitive or heuristic process. In the image processing and segmentation context, "Biocybernetics and Biomedical Engineering" [121] journal defines apriori information or knowledge as the combination of anatomical information about the shape, appearance, and gray-level distribution of the pixels, completed by annotations about the parameters used to tune or guide the segmentation method. A priori knowledge is an essential part of the development of pattern recognition systems and segmentation methods. The proper use of a priori knowledge allows a higher recognition algorithm than the typical usage level. Indeed this involves preventing errors and correcting them based on the apriori information acquired. Apriori information can take many forms: user interaction; appearance models; boundaries and edge polarity; shape models; topology specification; moments (e area/volume and centroid constraints); and distance prior between different regions/labels; and atlas or pre-known models. We will explore more significant attention segmentation frameworks that can incorporate apriori knowledge on expected shape in our context. The latter is also intended also, shape prior information or apriori shape information. When segmenting or localizing an anatomical structure, apriori information about the expected shape of that structure can significantly aid in the segmentation methods. Indeed these can benefit from using this information regarding the object's shape to be segmented to increase robustness against noise and occlusions. The shape knowledge may be incorporated by first constructing a statistical shape model from training cases and then constraining the segmentation results within the learned shape space. The shape space can be constructed with statistical analysis, e.g., with tools such as principal component analysis (PCA), isomap or Local Linear Embedding (LLE), Laplacian Eigenmaps. Usually, this shape statistic is taken into account in standard active contour algorithms. Fortunately, with the successful application of artificial intelligence (AI) in image processing, many challenging augmentation tasks are easily solved with deep convolutional neural networks. However, it has been shown that incorporation of shape prior information significantly improves the performance of the deep learning methods involved in segmentation tasks. Nevertheless, in these segmentation techniques like CNNs, it is tricky to incorporate such prior knowledge. On this premise, in the following sections, related works based on a priori shape information were illustrated and detailed.

2.4.2 Apriori shape information

The shape of an object is a geometrical description of the object boundary, and it plays an essential role in computer vision tasks such as recognition, segmentation, and tracking. In the context of image interpretation, apriori knowledge of the object of interest (color, texture, shape) is very useful in distinguishing the object from the background. In particular, incorporating the shape information in the segmentation task shows significant advantages of improving the accuracy and robustness of the algorithms when the appearance features of the object in the image alone are not sufficient due to noise, background clutter, or partial object occlusions. Therefore, it is indispensable and critical to building a shape model representing the prior shape knowledge of the object class, and it should be easily incorporated in solving the vision tasks. Shape modeling is a very challenging task due to the shape variability of the object of interest. A simple way to represent the apriori information is to use a shape template, a typical shape example of the object of interest. However, it is not specific to describe the object class with considerable natural variability. For example, the shape of the human organ shows both inter-individual variability (differences within populations) and intra-individual variability (differences of the same subject in tests taken at a different time or in other conditions). to construct a reliable shape model with information about the typical variation of the learned class, a straight direction is to gather this information by statistical means from several observations of the object (as many as possible), which leads to statistical shape models (SSMs). In this light, statistical shape analysis or statistical shape models had applications in various fields, including medical imaging, computer vision, computational anatomy as it analyzes the geometrical properties of a given set of shapes by statistical methods. Paramount aspects of this analysis are obtaining a measure of distance between shapes, estimating mean shapes from (possibly random) samples, shape variability, performing clustering, and testing differences between shapes. Although statistical shape modeling is a well-studied problem, generally, it consists of two critical components: A mathematical definition of the shape representation for statistic analysis and constructing a statistical model that describes the observed shape variations from a training set. For this reason, Various statistical models have been proposed to learn the apriori shape knowledge or information from a training set of the shape instances. Indeed, these models aim to represent the shape variation by linear or non-linear representations and globally or locally. Therefore most of the developed methods will be reported in the next section.

2.4.3 Use of apriori information in segmentation methods

A priori shape information has been used for medical image analysis for over two decades. The first approaches were based on the simple observation that a two-dimensional anatomical shape usually has a boundary that a smooth curve can well represent. In this regard, Kass et al. [60] introduced Active Contours or "snakes" based on this simple shape. Indeed the curves describing the shape information have been assumed to satisfy additional criteria based on minimizing bending energy in these methods. However, the idea to learn the properties of a shape from examples was initially formalized in the field of statistics by Dryden and Mardia [120] and later applied for image analysis by Cootes et al. [121, 126, 62]. In these models, known as the Active Shape models, the shape is relatively crudely represented as several manually selected landmark points in two dimensions. Only later, Blanz and Vetter extended Active Shape Models to three dimensions and used a dense set of points to define the shapes [122]. These models are known as 3D deformable models. In this context, active shape models and deformable models have since been used successfully for many tasks in computer visiion, computer graphics , and medical imaging, and in current image analysis literature, Active shape models and deformable Models are often not distinguished anymore. The term statistical shape models summarize, in fact, all these models. In general, the Active Shape Model (ASM) [12] is one of the most well-known methods in statistical shape modeling. This approach models the shape variations in a Point Distribution Model (PDM), which is learned in the following steps: Firstly, given a training set, each shape sample is represented by several points with correspondences across the training set; Secondly, an eventual alignment of the training shapes in a standard reference frame using Procrustes Analysis [127] has to be performed to eliminate pose variations. Third, the mean shape of all aligned training samples can be computed, and the variation modes concerning the mean shape are calculated by linear analysis such as principal component analysis (PCA). At last, the shape model is represented by a linear combination of the most prominent modes of variation. Furthermore, using bounded coefficient parameters, new shape instances can be generated to remain in the allowable shape domain (ASD) to look like those in the training set. This strategy was successfully applied to various types of shapes (faces, hands, organs, anatomical compartments), and for this reason, the point distribution model has become a standard in statistical shape modeling, particularly in the context of medical image segmentation. However, the application of the PDM strategy involves some disadvantages. One main drawback of this model is that it represents the shape variations globally since each mode of variation simultaneously influences all the shape variables. This global effect limits the shape model to have the flexibility of controlling local variations, which

is one of the desired properties in shape analysis or diagnostic purposes. To obtain variation modes that only affect a limited number of local landmarks, the orthomax method employed by [123] rotates the PCA modes to increase sparsity while maintaining the orthogonality of components. Another solution is Sparse PCA [123] which obtains the sparse modes and produces near-orthogonal components. Independent component analysis (ICA) [125] does not assume a Gaussian distribution and delivers statistically independent projections without orthogonality. These PCA-variant methods, however, provide no natural ordering for the various modes. Thus, different techniques have to be employed. Another practical issue in the PDM approach is that it is problematic to represent the full range of shape variations in high-dimensional space from a small training set. The size of the training set is always relatively small since the available images, and their required manual segmentation is not easy to obtain, especially in 3D cases, while the maximum number of eigenmodes can not exceed the number of the training examples minus one. [133] addressed this problem by introducing additional synthetic variance and covariance directly to the covariance matrix and coupling the movements of the neighboring points along the boundary. On the other hand, focusing on the shape representation, point-based or landmark-based representation (PDM) [128, 134], implicit representations [135], superquadric model [136], medial model [137], Fourier surface [138] are different examples to embody shape apriori information. Indeed, based on the shape representation of these methods, the statistical model is mainly represented either by a mean shape and the principal modes of variations or by a probability density function. The most popular method used in the former case is, as seen above, principal component analysis (PCA), and it is a variant [139], which approximates the shape by a linear combination of the most significant modes of variations. In the second case, Gaussian probability density function [139], Gaussian mixture models [141], kernel density [142], as well as manifold learning [143] have also been employed Nevertheless, the PCA-based shape modeling methods are the most referenced and the most intuitive way to represent the shape variations even if these methods can be applied directly to a point-based representation and other shape descriptions. [143] used Fourier parameterization of the object surface [144] and implemented principal component analysis (PCA) in the Fourier coefficient space. Closely, [145] expanded surface representation into a series of spherical harmonics (SPHARMs) and calculated shape eigenmodes in the shape parameter space. [146] used wavelet transform of the object contour and built a hierarchical shape model via PCA on the coefficients in each band, and [177, 148] used spherical wavelets to extend this approach to 3D cases. Moreover, implicit shape representation given by the level set framework is another popular and essential choice in shape representation, where the contour of the object is embedded as the zero level set of a higher dimensional surface. The embedding function is often chosen as signed distance

maps. Indeed [142], Leventon-Grimson-Faugeras were the first authors to incorporate shape information into the segmentation process based on active contours intrinsically represented by a level set function. The shape model used by the authors is the principal components analysis (PCA) that aims at capturing the main variations of a training set while removing redundant information. In particular [149] represented each training shape as a signed distance map sampled at regular intervals and performed PCA on the signed distance maps to build the shape model. [146] also applied PCA to collect signed distance representations of the training set and optimized the shape parameters directly in the segmentation process. Based on this global statistical representation, [137] considered a more challenging shape model that accounts for local variation. A common criticism of performing PCA on signed distance maps is that it can lead to invalid shapes since distance maps do not form a linear space. [179] addressed this problem by embedding the signed distance map manifold into the linear LogOdds vector space and applying PCA on the latter space. Furthermore, as a linear model, PCA cannot adequately model non-linear shape variations such as bending and shape variations of an articulated object. [133] estimates the probability density function of the distribution of shapes as a mixture of Gaussians to deal with this problem. Kernel Principal Component Analysis (KPCA) [144] introduces a non-linear mapping of the data to a feature shape, and PCA is performed in the feature space, and it was applied by [152] to the task of constructing non-linear ASMs. KPCA has become suitable for implicit shape representation [142] since it solves the problem that signed distance maps do not form a linear vector space. Manifold learning is also used to model non-linear shape prior. For example, [153] modeled a category of shapes as a shape apriori information manifold using Diffusion maps which generate a mapping from the original shape space into a low-dimensional space. In conclusion, this section constitutes a brief overview of methods and their critical issues about statistical shape modeling and a priori shape information. Additionally, we recommend a review [169] of apriori shape information for medical image segmentation reported over fifty projects, which use some statistical shape models for medical image analysis

2.4.4 Use of apriori information in deep learning methods

Deep learning has revolutionized image classification , segmentation [, and many other aspects of computer vision. In particular, recently proposed deep, fully convolutional neural networks show excellent performance in segmentation tasks. In this context, the use of apriori shape information could still improve, especially in some anatomical segmentation, the already promising results of deep neural networks. Moreover, the shape-awareness could overcome the limitations of previous deep learning approaches, which lack strong shape priors and the limits of active shape models, which miss, on the other hand, advanced pattern recognition capabilities. In image segmentation, the shape has been widely used in segmentation methods, like the active shape model, to constrain a segmentation result to a class of learned shapes. Indeed, to date, the shape information has been underutilized in deep segmentation networks. These methods do not use any statistical shape model and rely only on the fact that the sizeable receptive field of the convolutional neural network will perceive the anatomy of interest all at once, and therefore improbable shapes will be predicted only rarely in these modalities. Specifically, the deep and convolutional neural networks are trained with a pixel-wise loss function based on pixel-wise. Therefore it fails to learn high-level topological shape information and often fails to constrain the object segmentation results to possible shapes. However, combining shape information in a deep segmentation network is not straightforward, and especially, incorporating a priori shape information in deep segmentation networks is an ambitious challenge. Fei Chen [45] was one of the first researchers to propose a deep learning model to capture shape prior information. This latter was incorporated into two stages. The first stage uses a deep Boltzmann machine to learn the hierarchical structure of shapes in the training set. In the second state, the learned global and local shape variations are utilized in an energetic form to data-driven variational methods for making the final predictions. In addition to the training data, Mohammad Tofighi et al. [148] used a set of canonical shapes obtained via domain experts. They utilize the expected behavior via regularization term, which penalizes false positives results, which are not inside the boundary of the prior shape object. Zahra Mirikharaji et al. [149] came up with a new loss term for incorporating shape priors into an end-to-end fully convolutional network, penalizing all non-star shape segments in the output of FCN as in their task. All the training examples have the property of star shape. This method requires expensive optimization steps at inference time or user input information about the object center compared with energy-based approaches. However, it is a highly task-specific solution. In [151], authors propose novel cost functions in which they try to minimize the distance to the ground truth and the learned shape. They used convolutional autoencoders to project the input image into a shape space. These approaches

propose to utilize the prior shape information fully automatically, without user interaction and additional computational costs during the inference time. A different approach intended as a multi-network approach for incorporating shape information with the segmentation results was proposed in [153]. It uses a convolutional network to localize the segmentation object, an autoencoder to infer the object's shape, and finally uses deformable models, a version of SSM, to achieve segmentation of the target object. Another method for localizing shapes using a deep network is proposed in [8], where the final segmentation is performed using SSM. In this context, S. M. Masudur Rahman Al Arif et al. [152] proposed a novel, deep, fully convolutional neural network based on the SSM technique. Their approach can predict shapes instead of classifying each pixel separately, where a novel loss function computes the error directly in the shape domain compared to the other deep networks where errors are computed in a pixel-wise manner. In this paper, apriori shape information is captured from the training set using PCA, similar to well-known and aforementioned Statistical shape methods. To demonstrate the potentiality of the proposed method, the authors applied this neural network to a series of cervical vertebrae proving the effectiveness of this implementation. Fausto Milletari et al . proposed in [164] to integrate statistical shape apriori knowledge obtained through PCA into a deep neural convolutional network. Where PCA layer incorporates the modes of variation of the data at hand and produces predictions as a linear combination of these, it employed this approach to two different ultrasound dataset depicting the human heart and segmenting the left ventricle area, verifying that the proposed architecture improves the robustness and validity of the segmentation results and multiple measurements. On the other hand, Adrian V. Dalca et al. [165] introduced a generative probabilistic model that employs the learned shape prior through a convolutional neural network to compute segmentation in an unsupervised setting. In this method, the integration of anatomical priors can facilitate CNN-based anatomical segmentation in a range of novel clinical problems, where few or no annotations are available and thus standard networks are not trainable. From these methods, it appears that the information can improve the accuracy and robustness of some deep learning methods and increase the computational cost and complexity of the latter.

2.4.5 Use of apriori shape information in lungs fields segmentation

Lung shape analysis

Lung shape and variations are among the most riveting topics in the medical-radiological and CADX systems applying segmentation algorithms. The shape information can be used in the diagnosis of disease, determine the stage of illness, motion prediction, motion tracking in treatment planning, registration for better treatment and segmentation, Usually, the regular variations of the lung shape are interpreted as a normal lung and their physical functioning during the breathing. In contrast, abnormal variations of the lung shape can result from severe pulmonary diseases such as Sars-Cov-2. Specifically, lung shape variation is significant to radiologists who segment, interpret and analyze the radiological images [2]. In this regard, Hayashi et al. [154] showed that the knowledge of the variation of the lung is necessary to identify different images of related abnormalities. Aldur et al.[155] concluded that a surgeon should always recognize the variation of the lung with the support of the CAD system before committing lobectomies and segmental resection for patients. Moreover, in clinical trials, lung shape analysis can provide valuable qualitative and quantitative information from pulmonary disease patients .Since anatomical lung shape variation may lead to misunderstanding in different imaging techniques or misleading the diagnosis, sufficient knowledge is necessary for detecting related abnormalities in various imaging techniques and associated segmentation processes.

Indeed, lung shape analysis intended as a quantitative description of normal or abnormal lung shape, inter-subject variability, and age-related or disease-related differences is essential for several reasons. First, diseases with an age-related prevalence that affect the lung tissue can develop with a regional preference; for example, preferentially apical for emphysema as a component of chronic obstructive pulmonary disease (COPD) or subpleural basal for idiopathic pulmonary fibrosis (IPF). In IPF, this has been proposed to be related to locally high shear stress[167], which depends on the degree of lung inflation and change in the shape of posture and chest wall expansion during breathing. Detailed testing of this hypothesis has been limited by a lack of description of the ordinary and pathological lung shape. Second, mechanical changes to the lung tissue and chest wall accompany COPD or IPF affect the functional deformation of the lung tissue during breathing or breath-hold. In this context, lung shape and the related spatial distribution of abnormal tissue could hold prognostic information for staging or stratifying patients with these conditions. Finally, information on shape would improve automated image processing and segmentation methods, for instance, for detecting pulmonary fissures, where fissure integrity is a paramount predictor of outcome for endobronchial valve treatment of severe emphysema.

All these considerations provide insight into how shape information is essential in detecting and diagnosing certain diseases. Moreover, the previous sections evinced how it could be appropriate to incorporate shape knowledge as apriori information in the main segmentation approaches. As illustrated in [116], the effective modeling of apriori shape information is challenging because shape variation is complex and cannot always be modeled by a parametric distribution, and a shape instance derived from image appearance may have gross errors, Last but not least, local details of the input shape are difficult to preserve if they are not statistically significant in the training data. Biological shape variations in the lung compartment are often complex and highly challenging to interpret or explicate, In this perspective, an SSM provides a mathematical description of an object's shape and how its shape varies. It captures the global shape of the object of interest instead of using fixed geometric measurements such as lengths and angles. The statistical shape model (SSM) technique analyzes geometrical properties based on some given set of shapes via statistical methods with some crucial aspects, such as estimating the distance between the shapes, qualifying and quantifying the differences, and representing the mean shape among samples and their variability. SSM has many applications in medical image analysis to interpretation, recognition, classification, and especially segmentation. The active shape model (ASM) is one of the best model-based medical image segmentation approaches and bases its implementation and operation on statistical shape modeling. In recent years, ASM is widely used to model appearance, shape variations, and detect anatomical structures [45-47]. The ASM technique, which Cootes et al. [48, 49] developed, is limited to medical objects and can be used in a non-medical context. [46]. Based on the importance of this method for the analysis of apriori shape information, this technique is briefly reviewed as follows: A shape model, a gray level appearance model, and a matching algorithm are the essential components of the ASM technique [45, 48]. The shape model, a template of the region of interest, is represented as a distribution of points known as a point distribution model (PDM). Statistical variation of the shape in the training set encodes via PDM. Subsequently, PCA is applied to the shape vectors so that the shape model can be approximated and inferred, This process quantifies the average lung shape deriving a statistical average model and its principal modes of shape variation, such that the weighted sum of these components can retrieve back the object's shape. However, this approach allows extrapolating only the linear variations, not sometimes involving a complete description of lung shape information.

Segmentation methods

Segmentation is one of the most critical problems in medical imaging literature and is a foremost step of the computer-aided diagnosis (CAD) of lung diseases is the detection of the lung field. Accurate shape analysis of the lung can be an essential component in the segmentation process involving classification or clinical quantifications. Indeed lung shape analysis from different imaging systems for various applications has been studied in recent years. Klinder et al. [180] presented a prediction framework using 4DCT datasets for patient-specific modeling via a statistical motion model to predict respiratory lung motion. El Baz et al. [159] studied 3D lung shape analysis using 3D low dose CT datasets via spherical harmonic shape analysis to diagnose malignant lung nodules early. Zhang et al. [161] presented robust and effective lung shape modeling using CXR datasets via sparse learning-based shape prior modeling for lung localization and segmentation. The goal of this study is twofold: (1) represent two lung shape models which are different at the reference points in the registration process considering to show their impact on estimating the inter-patient 2D lung shape variations and (2) using the obtained models in lung field segmentation by utilizing active shape model (ASM) technique. Awais Mansoor et al. [160] introduced a hybrid principal component analysis (PCA)-DL-based approach for including apriori shape information for deformable object segmentation. In this method, a module named marginal shape deep learning (MaShDL) transforms the iterative process of the conventional SSM-based segmentation methods to a recursive marginal refinement approach and, precisely, begins by learning the mode of shape deformation in the eigenspace of the most prominent variation and then marginally increases the dimensionality of eigenspaces. The framework was evaluated using a comprehensive CXR dataset and demonstrates outstanding potential for generic applicability. Li et al. [143] presented a 2D lung inter-patient variation from CXR datasets for automatic lung field segmentation using statistical shape and appearance models. Lopes et al. [167] studied lung shape analysis from CXR datasets via a pre-trained convolutional neural network as feature extractors for tuberculosis detection. Gang et al. [168] studied lung shape analysis from CXR datasets to dimensionality reduction in deep learning to analyze lung cancer. Cheimariotis et al. [165] presented automatic lung segmentation in ventilation/perfusion lung SPECT datasets via active shape model trained on reference lung shapes from CT datasets. Afzali et al. [166] studied lung shape analysis using Fourier descriptors (FDs) to estimate inter-patient 2D lung shape variations from CXR imaging datasets. Nakao et al. [176] studied lung shape analysis from CT datasets via model-based shape matching (Laplacian-based shape registration) for lung surface deformation analysis. Souza et al. [177] proposed a different view of lung shape analysis using deep neural net work on CXR images to

perform an automatic CXR lung segmentation. Gaal et al. [178] presented lung shape analysis from CXR datasets using a novel deep learning-based approach to achieve a CXR lung segmentation. Usman et al. [179] proposed a novel semi-automated system using a deep U-Net architecture and lung nodule shape analysis to present 3D segmentation of lung nodules. Lung shape, therefore, potentially provides a prognostic marker of lung health or accelerated aging; it could provide a straightforward means for staging or stratification of lung disease. However, to distinguish between normal and abnormal lung shapes, a quantitative description of the standard or not- standard shape and variability is the first step required in the segmentation process. Nevertheless, the method or strategy for accomplishing this analysis is of paramount importance.

CHAPTER 3

Materials & Methods

Materials & Methods

This section is one of the most significant as it illustrates the main methods and materials availed in this thesis. Materials define the datasets handled to train and test the neural network with the preprocessing operations apt to conform the data or better the images to the used methods. In addition, attention is paid to the AIforCovid dataset, as it is one of the few datasets available that includes data and images from covid-19 positive patients. Regarding the methodology, we detail the proposed method termed "SP-net" in its main aspects and fundamental concepts to extract and process the apriori information extracted from the training dataset . In this respect, the section deals with the comparison method considered as the state of art and identified as deep neural network U-net.

3.1 Introduction

Datasets are an essential part of deep learning research and play a crucial role in developing any segmentation method based on a deep neural network. However, datasets consist primarily of images for this task and require label data for supervised and semi-supervised learning algorithms, which involves considerable effort and time for labeling operation. Datasets for unsupervised learning can also be onerous and costly to produce, although they do not need to be labeled. The aim of the proposed method is the segmentation of lung field X-ray images. In this perspective, for neural networks training based on supervised learning, two datasets have been used: Just and Montgomery county datasets. These are two publicly available datasets with lung field annotations and come from two different countries with different lung diseases. Besides, their widespread application in the relevant literature allows comparing this method with the current state-of-the-art. To evaluate the proposed SP-NET network, we used the AIforCOVID imaging archive as a test set that hosts an extensive archive of medical images of COVID-19 patients.

3.2 Datasets

3.2.1 JSRT Dataset

The JSRT dataset[174] was released in 1998 by the Japanese Society of Radiological Technology. Since then, it has been used by several researchers in the world for various research purposes such as image processing, image compression, and computer-aided diagnosis (CAD), and training and testing in Deep learning applications. In 2008, the JSRT database was made available online and free of charge due to that. The dataset consists of 247 posteroanterior (PA) chest x-ray images are acquired with a standard X-ray machine and collected from 13 different Japanese institutions and one in the USA. Of all,154 have lung nodules, and 93 have no lung nodule. The CXR images are scanned from films to a resolution of 2048x2048 pixels, with a grey-scale depth of 12 bits and pixel size of 0.175 mm, and released as disc image files with img extension. Also, the dataset contains additional information such as the patient's age which is between 16 to 89 years, gender, X and Y coordinates of the lung nodule.

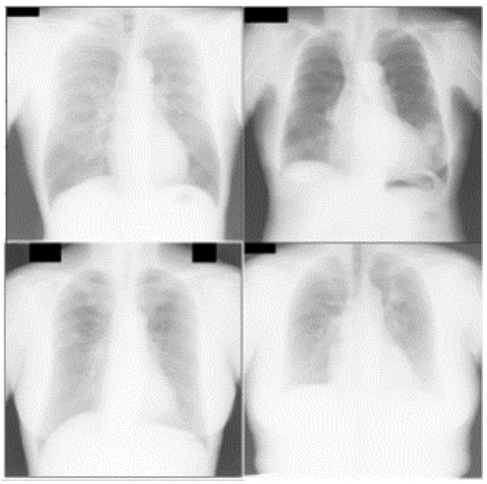


Figure 3.1: Four CXR images from the JSRT Dataset

The manually annotated masks of CXR. images are available and retrieved from Segmented Chest Radiograph (SCR) dataset[182] this is created by Image Sciences Institute, University Medical Centre of Utrecht, and it has been established to facilitate comparative studies on segmentation of the lung fields, the heart, and the clavicles in standard posterior-anterior chest radiographs. Indeed it is built on the JSRT dataset by manually delineating the relative anatomical structures' boundaries (right lung, left lung, heart, and clavicles) in the chest area. Then, the segmented masks are released as separate mask images for all anatomical structures in gif extension and exact resolution as their corresponding images in the JSRT dataset. It is important to emphasize that this dataset represents mostly regular lung masks. The reason is that, even though the majority of patients have lung nodules, these do not alter the contour of the lungs.

In the vast majority of literature, the segmentation is evaluated on the Japanese Society of Radiological Technology (JSRT) data. However, this dataset covers a far narrower set of abnormalities. It means that a method developed to segment CXRs of this dataset might not generalize well for many irregularities. Furthermore, X-rays in this dataset appear smoother intensity-wise due to an older imaging system. For all these reasons, the Montgomery County dataset was also considered representative for developing and evaluating our method. Manual masks separately for the left and right lung are also provided; these are binary images depicting white the lung and black backgrounds.

3.2.2 Montgomery Dataset

The Montgomery Dataset[173] was released in 2014 by the USUS. National Library of Medicine (USNLM) collaborates with the Department of Health and Human Services of Montgomery County, MD, USA. The dataset consists of 138 posterior-anterior x-rays, of which 80 x-rays are regular, and 58 x-rays are abnormal with manifestations of tuberculosis. The images were collected as part of Montgomery County's Tuberculosis screening program. The radiographs were acquired with a Eureka stationary X-ray machine and stored at a spatial resolution of 4020x4892, or 4892 x 4020 pixels, with a grey-scale depth of 12 bits and pixel size as 0.0875 mm, available in DICOM format. In addition, the dataset consists of separate mask images for both left and right lungs with png extension and clinical readings for each CXR. in the form of txt extension, which concerns the patient's age, gender, and abnormality seen in the lung, if any. Lungs were segmented under the supervision of a radiologist, following anatomical landmarks, such as the boundary of the heart and pericardium, excluding the area behind the heart and diaphragm. Furthermore, an inferred boundary was drawn when the pathology is severe and affects the lungs' morphological appearance. The binary lung masks were saved using the same naming convention as used for the X-ray images and clinical readings and separately for the left and right lung, in folders leftMask and rightMask, respectively The US National Library of Medicine has made The Montogomery dataset available free of charge to foster research in computer-aided diagnosis of pulmonary diseases with a particular focus on pulmonary tuberculosis.

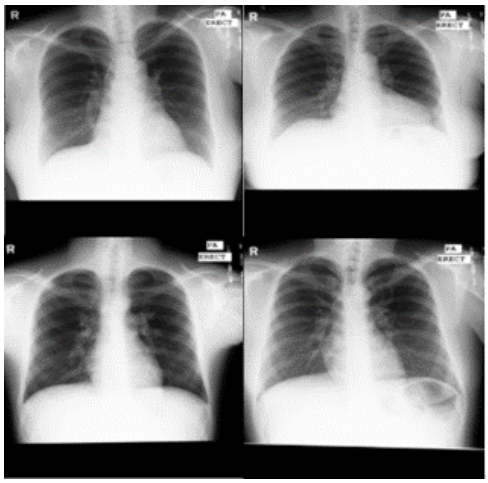


Figure 3.2: Four CXR images from the Montgomery Dataset

3.2.3 AIforCovid Dataset

The AIforCovid imaging archive[2] is a novel Dataset, released in December 2020, and introduced thanks to a recent project promoted by CDI Centro Diagnostico Italiano (Milan) in partnership with Università Campus Bio-Medico (Rome), and Istituto Italiano di Tecnologia (Genova) and several hospitals in Italy. The AiforCovid Dataset includes clinical data and CXR images from 820 patients by six Italian hospitals in 2020 during the first Covid-19 emergency. Moreover, each patient is associated with prognostic information related to the clinical outcome. This dataset has been created primarily to investigate artificial intelligence's potential to predict such patients' prognosis, distinguishing between severe and mild cases and offering a baseline reference for other researchers and practitioners. To this end, the data repository is made publicly available to encourage research on AI-based prognostic models and thus support healthcare systems in the fight against the Covid-19 pandemic. All Chest X-rays collected in this dataset were performed in Covid-19-positive adult patients at the time of hospital admission, and all the patients resulted positive for SARS-CoV-2 infection.

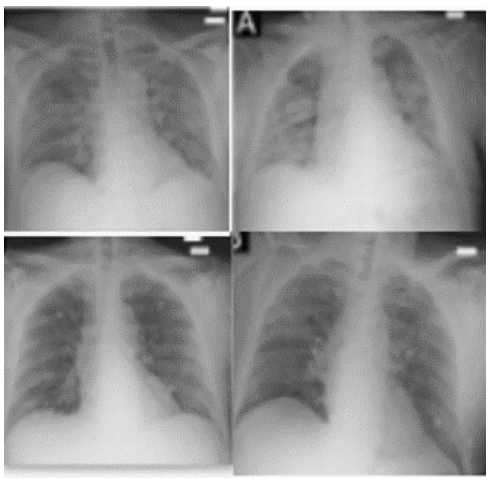


Figure 3.3: Four CXR images from the AIforCovid Dataset

On the other hand, the clinical outcome assigns each patient to the mild or the severe target. The former consists of patients intended to domiciliary isolation or hospitalized without any ventilatory support. In contrast, the latter includes the patients who required non-invasive ventilation support or intensive care unit (ICU). CXR images and clinical data were performed in different hospital centers using analog and digital units, and the execution parameters were settled according to the patient conditions. Therefore, they include common real-world issues such as outliers, different imaging devices, poorly standardized, and missing clinical data (identified with string "NaN"). The AIforCovid dataset contains images of patients in both lying and standing positions. These radiographs were acquired with isotropic pixel spacing ranging from 0.1 mm to 0.2 mm and stored in DICOM format using 16 bits. Finally, regarding image resolution, the CXR images have different sizes in pixels: Most of them are between 2336×2836 and 3520×4280 pixels. The others have a number of rows ranging from 1396 up to 4280, while the number of columns ranges from 1676 up to 4280.

Dataset	**Source**	**Content**	**Resolution**	**Bit Depth**	**Format**	**Public**
JSRT	J.S.R.T.	247 CXRs	2048x2048	12	.img	✓
Montgomery	U.S.N.L.M.	138 CXRs	4092x4892	12	.png	✓
AIforCovid	C.D.I.	821 CXRs	floating	12	.nrrd	✓

Table 3.1: CXR datasets used in the proposed method

3.3 Preprocessing

3.3.1 ETL-preprocessing

ETL (Extraction-Transformation-Loading) processing is a well-known process that extracts data from different sources, transforms those into the required format, and finally loads it into the target data destination. In this respect, the Extraction-Transformation-Loading (ETL) process is a preliminary and paramount step to collect data from datasets and train, verify, and tune the new deep learning algorithms and applications. For this reason, this step was essential to customize the images and masks for the proposed method. Regarding the JSRT Dataset, the Japanese Society of Radiological Technology has given free access by creating a user name and password to access the download site. It allowed us to download the CXR images as disc image files with an IMG extension which implied specific functions to access and convert them later to a more convenient file type. The corresponding masks were retrieved from the SCR dataset have been downloaded from the official website: SCR database, Segmentation in Chest Radiographs. The annotated masks have been released in gif extension and so extracted with "scikit-image" based functions. In addition, the database provides for each image of the JSRT dataset a single mask respectively for the left lung and the right lung. Because of that, it was necessary to reunite the segmented masks into a single image during the extraction process. The Montgomery Dataset is publicly available on the official website of the USUS. National Library of Medicine. Indeed, the image files were extracted by downloading a zip file containing both CXR images and CXR masks. These are released in png format, and because of that, typical image processing functions were used to extract them from the source file. Also, in this case, the masks are present separately for the two lungs. Therefore min the extraction step, they are reassembled in a single mask for each CXR image of the dataset. The Extraction Process for the AIforCovid dataset was a bit more complicated. The CXR images and masks are preserved in Nrrd format to support scientific visualization and image processing. It was possible to download the image files from Centro Diagnostico Italiano's official website, receiving the supervisor's credentials. After that, the images and masks were extracted from the zip file using an unusual package in Python based on ITK and formatted in the desired file type. Image data extracted from the source are raw and not usable in their original form. Therefore, in the data transformation stage, a series of rules or functions are applied to the extracted data to prepare it for loading finally into the end target. The transformation process consists of a precious pipeline to carry out image processing operations to customize each mask and image according to the requirements of the proposed method. The pipeline consists of the following

steps: image resizing, normalizing, and specifying the photometric interpretation. The masks, instead, were just resized and normalized. The procedure has been applied to all datasets to homogenize all images and masks according to the same scheme. This part is very crucial for the outcome of the proposed neural network. Indeed it will be discussed in more detail in the next section. The load phase places the data into the end target, and it is the last step of the ETL process. Our application provides the loading of images and masks within a training set, validation set, and test set, intending to place the JSRT and Montogomery dataset files in the former and the Aiforcovid image files in the latter.

3.3.2 Images Preprocessing

Introduction

Training a convolutional neural network on raw images will probably lead to harmful or inconsistent performances. Hence the image processing is an essential preliminary task before training the proposed model. Image preprocessing refers to all the transformations on the raw data before it is fed to the machine learning or deep learning algorithm. Moreover, this process aims to improve the quality of the image so that it can be better analyzed. In this case, this procedure is limited to making sure that all the datasets' images have the exact specifications. Indeed, the pipeline concerns only re-dimensioning images to the same resolution, normalization of them, and adjusting their photometric interpretation. The idea is that the neural network can be trained on images as much as possible identified by common real-world issues. In this sense, image preprocessing is defined by a series of operations on images at the lowest level of abstraction. These operations do not increase image information content, but they decrease if entropy is an information measure. Thus, preprocessing aims to improve the image data that suppresses undesired distortions or enhances some image features relevant for further processing and analysis tasks.

Resizing

Resizing images is a critical preprocessing step in computer vision. Principally, deep learning models train faster on smaller images. However, an input image that is twice as large requires the network to learn from four times as many pixels, and that time adds up. Moreover, many deep learning model architectures require an exact resolution, and raw collected images may vary in size. Therefore it emerges as a process of priority importance to ensure the images do not involve a computationally elevated cost for the network's training and respect the standard input of this last one. In particular convolutional neural networks need to resize the images in the dataset to a unified dimension. It implies that raw images must be preprocessed and scaled to have identical widths and heights before fed to the learning algorithm. In this regard, We establish a base size of 128x128 pixels for all images provided into our deep convolutional neural network so that this resolution does not presuppose significant time and effort in the training process and without excessive reduction in image quality. However, the resizing operation needs to perform interpolation of pixel values to up-size or down-size N-dimensional images. In this perspective, the bilinear interpolation method was applied to decrease the dimension of the images for all datasets. This method is widespread in computer vision and image processing applications. Indeed, it can be used where perfect image transformation with pixel matching is impossible, and so can calculate and assign appropriate intensity values to pixels. Unlike other interpolation techniques such as nearest-neighbor interpolation and bicubic interpolation, bilinear interpolation takes values of only the four nearest pixels, located in diagonal directions from a given pixel, to find the appropriate color intensity values of that pixel. Bilinear interpolation considers the closest two $\times$ two neighborhoods of known pixel values surrounding the unknown pixel's computed location. It then takes a weighted average of the four pixels to arrive at its final, interpolated value. This algorithm reduces the visual distortion caused by resizing an image to a non-integral zoom factor instead of nearest-neighbor interpolation, which will make some pixels appear more significant than others in the resized image. Moreover, it is computationally less expensive than bicubic interpolation. The JSRT dataset's images were resized from a resolution of 2048x2048 to a standard of 128x128 pixels. Likewise, the Montogomery dataset and Aiforcovid dataset images acquired to a spatial resolution of 4020x4892, or 4892 x 4020 pixels, were resampled to this resolution.

Normalization

Image normalization is a typical process in image processing that changes the range of pixel intensity values. Its purpose is to convert an input image into a range of pixel values that are more familiar or normal to the senses, hence the term normalization. The pixel values in images must be scaled before providing the images as input to a deep learning neural network model during the training or evaluation of the model.

The reason behind this process is that by training our network, these initial inputs will be multiplied with weights and added bias to cause activations that then back-propagated with gradients to train the model. Similar ranges for input images are crucial in this perspective so that gradients do not go out of control. Indeed, during the image preprocessing, each image's pixel values in our datasets were reduced to the range of 0-1 make computation more efficient. In our application, normalization branches into two different cases. The former involves linear normalization. It is performed according to the formula.

However, this process causes a loss of contrast in normalized images. Since some pixels with high brightness in the normalization process result in clipping the others to low values. In most cases, these are due to text or letters' presence inside the images that bring no information for network training purposes. As a result, they can be considered outliers pixels. In this context, the second approach offers an arguable alternative to standard normalization, and it avoids image degradation due to the latter. The following process is defined in our image preprocessing as "percentile normalization." It applies standard normalization considering maximum pixel value 95th percentile of pixel distribution and, as the minimum, the 5th one, keeping the pixel values between 0 and 1. It is better indicated from the following formula :

Consequently, this method allows us to exclude pixel outliers from the normalization process and not alter image contrast. Therefore percentile normalization was applied on all images of the considered datasets.

Photometric Interpretation

The photometric interpretation specifies how the pixel data should be interpreted for the correct image display as a monochrome or color image. To define if color information is stored in the image pixel values, we introduce the concept of samples per pixel (also known as the number of channels). Monochrome images have one sample per pixel and no color information stored in the image. Nevertheless, a scale of shades of gray from black to white is used to display these images. The number of shades of gray depends clearly on the number of bits used to store the sample that coincide with the pixel depth. Hence one discerns how it is essential to use a single photometric interpretation and the same pixel depth for all the images in our datasets. Clinical and radiological images, like chest X-ray images, have a grayscale photometric interpretation. Indeed the JSRT dataset's images were acquired by displaying the lungs field with gray tones tending to black while the remaining part of the chest is rendered with complementary colors, tending to white. This photometric interpretation was considered as a reference for all datasets. The Montogomery dataset presents images with the same photometric interpretation, and therefore, no adjustment was necessary. However, for the Aiforcovid images, the lung fields are rendered in gray shades close to white. Thus colors of this region are inverted with those of the background to respect the other dataset's specifications. Finally, all datasets' images were stored in 32 bits to properly image quality and ensure that the training and neural network weights are accurate.

3.3.3 Masks pre-processing

The segmentation masks of the datasets were rescaled to the exact resolution as the respective images. Nevertheless, for this assignment, the nearest neighbor interpolation has been employed. It results computationally the most efficient in resizing the segmentation maps. Indeed this is the default method for scaling boolean matrices. Moreover, it requires the least processing time of all the interpolation algorithms because it only considers one pixel and the closest one to the interpolated point. In the next phase, labeled masks are normalized with standard normalization. As follows, the pixel values are in the range of 0 and 1, and the segmentation maps are not altered in any way.

General Scheme of ETL Preprocessing

The following figure (Fig 3.4) resume the principal process carried out to bring the raw images and data to final form and employed in our algorithms.As the illustration shows, the essential steps described in the previous paragraphs consist of Extracting, Transforming and Loading the images and data into the final sets.

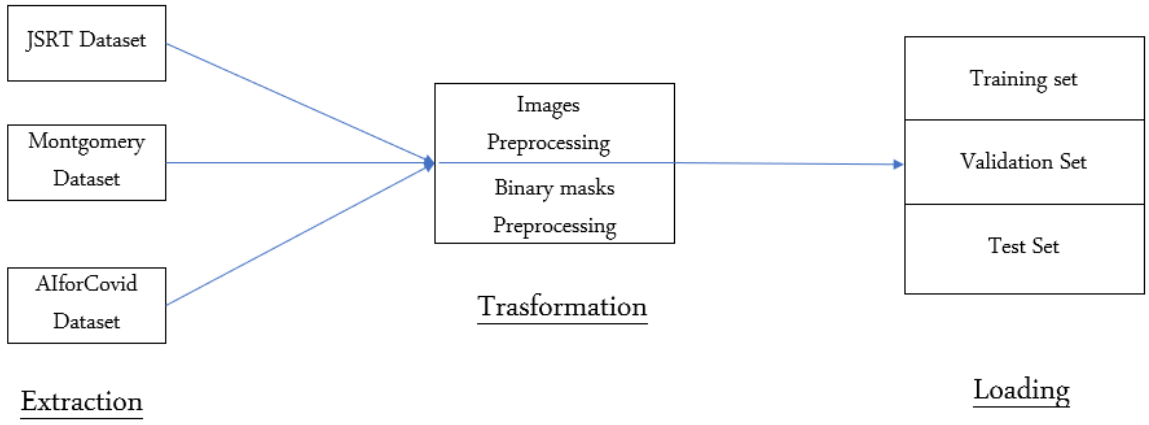


Figure 3.4: ETL Preprocessing

3.3.4 Training set, Validation set & Test Set configuration

After the ETL-preprocessing , as described above, the images and masks are divided into appropriate sets, according to the established schemes and specifications. As a matter of fact, to realize the training dataset and the validation set, all the images of the JSRT Dataset and the Montogomery Dataset have been used, making up by 299 and 86 images, respectively. While for the test set it consists of 43 images sampled by AIforCovid. The subdivision described here, is better reported by the following table:

Dataset	**Source**	**Content**	**Resolution**
Training Set	JSRT & Montgomery Dataset	299 CXRs	128x128
Validation Set	JSRT & Montgomery Dataset	86 CXRs	128x128
Test Set	AIforCovid Dataset	43 CXRs	128x128

Table 3.2: CXR datasets for Training , Validation & Testing

3.4 Methods

3.4.1 Introduction

As we discussed in the previous chapter, a wide range of challenges involves lung field segmentation processes and incorporating apriori information in them. Our goal is to investigate the benefit of integrating the apriori shape knowledge in deep neural networks. In this light, we extend the idea and method presented by Al Arif et al. [152] to the troublesome but compelling challenge of segmentation of lungs field in CXR images. The method presented in this paper is based fundamentally on principal components analysis used on the training dataset to capture the apriori shape information. In this respect, we also provide a nonlinear extension of PCA, applying the kernel PCA on the training dataset and integrating the apriori information derived in the neural network, renamed Kernel-SPnet. To understand these methods' actual capabilities and advantages, we will compare them with the convolutional neural network U-net, a segmentation method broadly used in the relevant literature and considered state-of-the-art for biomedical segmentation. The following paragraphs will describe and illustrate the convolutional neural network U-net and the proposed network termed SP-net and the nonlinear form Kernel Sp-net.

3.4.2 U-net

Introduction

In recent years, deep learning has predominated medical image segmentation. Especially encoder-decoder architectures, such as U-Net, can be included in state-of-the-art models with robust designs achieved by implementing skip connections to propagate local information from an encoder path to a decoder path to give detailed spatial information lost pooling operations. UNet is a Convolutional Neural network created in 2015 at the University of Freiburg, Germany [113]. Olaf Ronneberger, Philipp Fischer, and Thomas Brox purposely introduced it to analyze biomedical images. Indeed the method was very successful in the medical context, to the extent that various applications of U-Net in biomedical image segmentation, such as liver image segmentation ("sliver07" [176]), and brain image segmentation ("BRATS" [177]) obtaining excellent results. Furthermore, the network, freshly designed, won the ISBI cell tracking challenge in 2015. On this premise, in the following paragraphs, we will examine the main aspects of this neural network and its impact on biomedical segmentation.

U-net: An improvement of fully connected networks

Although Olaf Ronneberger et al.[1113] created U-Net, in 2015 in the paper "U-Net: Convolutional Networks for Biomedical Image Segmentation," it can be considered an improvement and development of a fully convolutional neural network, proposed in 2014 by Evan Shelhamer, Jonathan Long, Trevor Darrell[178]. In the paper "Fully convolutional networks for semantic segmentation," the authors define, in fact, a new fully convolutional net (FCN) for segmentation that combines layers of the feature hierarchy and refines the output's spatial precision, resulting in accurate and detailed segmentation. The paper has presented three different neural networks -FCN-32s, FCN-16s, and FCN-8s. The final number in each abbreviation indicates the pixel stride at the final prediction layer limits the scale of detail in the upsampled output. In this context, The FCN-8s (Figure 4.1) are proven to have the best performance. Hence, it achieves similar U-net results. Indeed it would not be wrong to say that U-Net is an extended, deeper version of FCN-8s. The main idea in [113] was to supplement a usual contracting network by successive layers, where upsampling operators replace pooling operators. Hence, these layers increase and recover the resolution of the output. To localize, high-resolution features from the contracting path are combined with the upsampled result. A subsequent convolution layer can then learn to assemble a more accurate output based on this information. One crucial modification in Unet architecture is that it also has many feature channels in the upsampling part, which enable the network to propagate context information to higher resolution layers. Consequently, the expansive path is more or less symmetric to the contracting one and yields an u-shaped architecture. The network does not include any fully connected layers and only uses the right part of each convolution; the segmentation map only contains the pixels. The whole context is available and retrieved from the input image. This strategy allows the seamless segmentation of arbitrarily large images by an overlap-tile approach (see Figure 2). Briefly, Unet exploits symmetric structure charactered by skip connections to predict the pixels in the border region of the image, so the missing context is extrapolated by mirroring the input image. This tiling strategy is essential to apply the network to large images since otherwise, the resolution would be limited

U-net Architecture

As evidenced, the U-Net architecture is built upon the FCN and modified in a way that results in a better segmentation in biomedical imaging. However, compared to FCNs, it presents two significant differences.

- UNet is symmetric to ensure the input images has the exact resolution of outputs;
- the skip connections between the downsampling path and the upsampling path apply a concatenation operator instead of a sum.

These skip connections play pivotal importance as they provide local information to the global information when upsampling, thereby recovers the background and spatial details. Due to its symmetry, the network holds a high number of feature maps in the upsampling 23path, which supports working with high-resolution layers the UNet architecture is constituted and divided into three parts :

- The contracting/downsampling path;
- Bottleneck
- The expanding/upsampling path

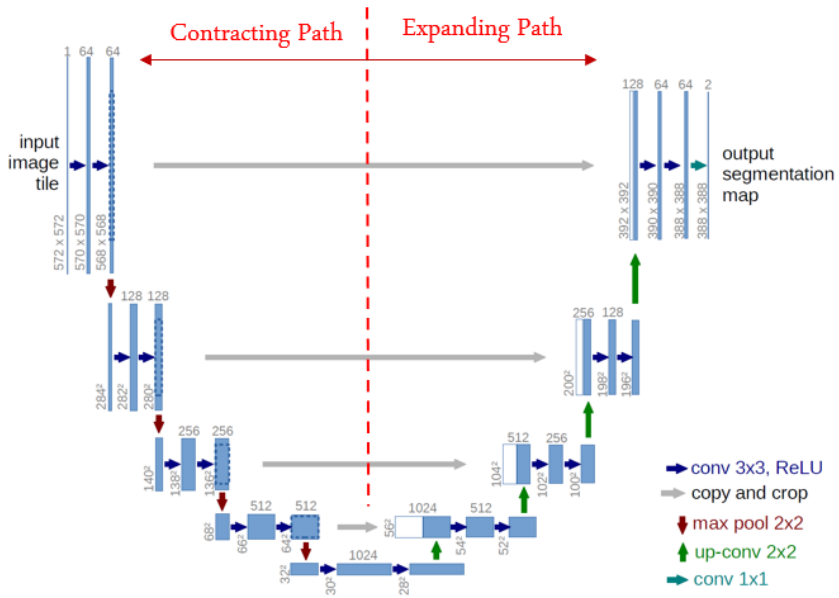


Figure 3.5: U-net architecture

As illustrated in Figure 3.5, the network architecture consists of a contracting path (left side) and an expansive path (right side). The contracting branch follows the typical architecture of a convolutional network, and its goal is to capture the context of the image to perform segmentation. In this step, spatial information was reduced, and features increased. The downsampling path contains many contraction blocks. Every block accepts two 3 by 3 convolution layers and a 2 by 2 max pooling, each followed by a ReLU. The number of filters doubles after each block. The bottleneck is between the contracting and expanding paths operating as a bridge. It composes of two 3 by 3 CNN layers and 2 by 2 up convolution layers. The most important part of this architecture is the expanding path. Like a contraction section, it is made of several expansion blocks. Each block transmits the input to two 3 by 3 CNN layers followed by a 2 by 2 upsampling layer. After each block, the number of kernels that the convolutional layer uses is halved to keep the symmetry. Nevertheless, the input is joined to the feature maps of the corresponding contraction layer. This step ensures that the features learned when contracting the image will be utilized to remodel it back. Finally, the produced mapping is sent to another 3 by 3 CNN layer with the number of feature maps equal to the segments targeted.

U-net in our thesis

The U-net architecture attains outstanding performance on very different biomedical segmentation applications and contexts. Thanks to the feasibility of data augmentation with elastic deformation, it only needs very few annotated images and has a reasonable training time. These concepts significantly impact the medical area, where the availability and accessibility of manually segmented images is a crucial and rare resource. However, in Unet, despite its strong power of representation, skip connections tend to use redundant information in low-level encoder features in a multi-scale approach. Furthermore, the contextual information of the encoder feature at the beginning of the network is insufficient, leading to poor performance for pixel-wise recognition when concatenating with the corresponding high-level decoder feature map. Addressing these issues by either modifying the architecture or changing the training scheme could improve the network capability in object detection and segmentation. Unet has been employed in our work as a reference standard and comparison network method on this premise. Moreover, the proposed method constitutes a modification of Unet architecture to integrate the apriori information retrieved from the training set images.

3.4.3 The proposed method: SP-net

Introduction

Apriori shape information has been widely used in medical image segmentation algorithms to bridle a segmented region to a class of learned shapes. Perceiving the idea of segmenting or localizing an anatomical structure, with apriori information about the expected shape of that, can significantly aid in the segmentation process. However, recent methods for medical segmentation primarily use deep learning algorithms, but the majority of state-of-the-art deep segmentation networks are trained with loss functions and defined in a pixel-wise manner that is unsuitable for learning or integrating topological shape information and constraining segmentation results. In this respect, the proposed deep fully convolutional neural network aims to integrate the apriori shape knowledge into the model to predict shapes instead of pixel-wise classification. In this perspective proposed method inspired by the reference paper by Al Arif et al[158]. and termed as Shape predictor network (abbreviated in SP-NET) modifies the Unet architecture, employing in the final layers specific operations to incorporate the apriori information and predict, therefore, in the shape domain. Hence, unlike the pixel-wise loss function-based segmentation networks, the loss for the proposed network is computed with the shape parameters, priorly calculated by principal component analysis in the ground truth generation. In our work, considering the medical importance of lung field segmentation, the proposed neural network focuses on integrating apriori knowledge of lung shape information in this context. Indeed, to demonstrate the potential and effectiveness of this method, we applied it to a series of CXR lung images of covid-19 positive patients and characterized by real-world issues, caused mainly by the consequences of the disease and the patient conditions render lung segmentation more challenging.

Methodology

As illustrated in fig1, SP-net is based on an architectural adaptation of Unet in which the final layer involves tthe predicted images are converted , by a principal components ,more specifically their eigenvectors, to shape parameters compactly represented in a lower-dimensional space.

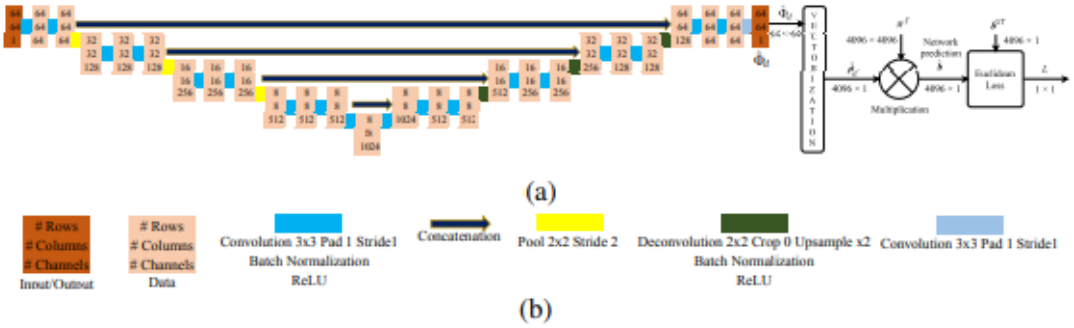


Figure 3.6: SP-net architecture

Specifically, our proposed network, SPNET, uses the same network architecture [11]. It takes a CXR image patch as input and produces its related shape representation intended as SDF, which is also defined over the same pixel space. Finally, the last layer outputs the difference signed distance function (ϕ_d), which is then sent to the final layer, where the integration of the information through the principal components sorted in a matrix W allowed to convert the predicted segmentation map into a shape parameter vector and compared it with the ground truth (b_{gt}).

In line with this approach, the development and unfolding of the method consist of the following steps and arguments. The overall pipeline of our proposed method is demonstrated in Figure 2.

1. Estimation of apriori shape information
2. Dataset and ground truth generation
3. Application of shape predictor network

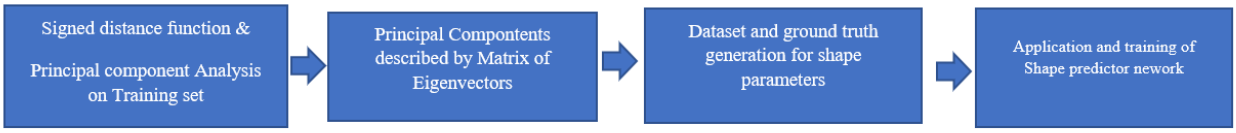


Figure 3.7: SP-net pipeline

Estimation of apriori shape information

The estimation of apriori information is a preliminary analysis performed before the deployment or training of the neural network. It computes apriority on shape variation given a set of training instances. In our context, it encompasses all those operations carried out to retrieve apriori information relating to the lung shape from the training set. In this regard, this analysis employs techniques used to extrapolate the shape information relatable to the object to be segmented of interest. Following Leventon's approach[149] and as referenced in [1], it primarily uses a defined shape representation intended as a "signed distance function" to collect richer information about structural and anatomical features from training images. Subsequently, it compacts the information using the principal components analysis to induce the segmentation maps to comply with apriori-shaped knowledge learned. On this premise, the leading aspects being investigated in the following paragraphs are the signal distance function and the principal components analysis, which are the real characters involved in shape learning.

Signed Distance Function

The estimation of apriori information aims to collect information about the lungs' shape in CXR images in our thesis. In this light, the shape representations focus the informative content of the lung images on the pulmonary curves. In this regard, implicit models such as signed distance functions become an essential element of shape analysis. A signed distance function (SDF) is a level set function that embeds points and contours in a higher-dimensional space and encoding information about structural and shape features of the concerned anatomical part. Indeed it represents the image's pixels by the distance between the point and the closest boundary of the target object or organ. In contrast, the sign of pixels denotes whether the pixel is inside the border of the target organ (negative), outside the organ (positive), or part of the boundary(zero). The Signed Distance Function (SDF) is defined by :

$$\phi(\mathbf{x}) = \begin{cases} 0, & \mathbf{x} \in \mathcal{C} \\ -d_\Omega(\mathbf{x}, \mathcal{C}), & \mathbf{x} \in \Omega_{in} \\ +d_\Omega(\mathbf{x}, \mathcal{C}), & \mathbf{x} \in \Omega_{out} \end{cases}$$

As evidenced by the formula, to every point x in , the shortest Euclidean distance d(x, C) to the contour is assigned. Function values are negative inside the shape in and positive in the outside region out. The contour is then defined implicitly as the zero levels set of its SDF: (x) = 0 This representation allows obtaining more contextual information than landmark samples employed in point model distribution, constituting a congenial asset for medical segmentation since it also requires background details. Moreover, manually localizing landmarks is tedious, while automatic landmark detection is prone to errors, especially when dealing with 3D objects. For this reason, SDF is a viable alternative to point model distribution. Several works have explored the Signed Distance Function (SDF) applications in computer vision and graphics. (Perera et al. 2015) uses truncated SDF to reconstruct volumetric surfaces on RGB-D images better. [166] treats the linearly shifted saliency map as the SDF and refines the predicted saliency map in multiple training stages with level set smoothness terms. (Park et al. 2019) learns the continuous 3D SDF directly from point samples by a network containing series of fully connected layers and an L1 regression loss. The learned SDFs are used for obtaining state-of-the-art shape representation and completion results. Our reference paper [152] proposes to use an unsigned distance map as an intermediate step for 2D organ shape prediction. Subsequently, PCA is involved in converting from distance map to shape parameter vector, and the segmentation map is not engaged. More recently, [181] use distance map prediction as a regularizer during

training for organ segmentation. Since they predict segmentation maps and distance maps in different branches, correspondences between projections of the separate segmentation and SDF branches are not guaranteed. Their method differs from ours, but in both cases, segmentation maps and SDFs involved in our model are connected by a differentiable Heaviside function and can be predicted and reconstructed as a whole. These considerations evince, like the signed distance function is an appropriate representation to provide further prominence to the shape information about the anatomical part. On these grounds, this shape representation is the primary operation computed in our method.

Our application

In our thesis, the signed distance function is employed as shape representation to convert the images or better binary masks belonging to the training set. Indeed the manual annotation for each training mask is transformed into a signed distance function (SDF). To convert the lungs shapes into an SDF (), The pixels lying on the manually annotated boundary curve have been assigned zero values. a All other pixels are assigned intensity based on the infimum of the euclidean distances between the corresponding pixel and the set of pixels with zero values. In this perspective, after processing and loading the images from the Montgomery and JSRT datasets within the training set, they are split into two groups consisting of 299 ground truth CXR images and the corresponding binary masks, respectively. In this way, the 299 masks, once detached from the training set, were converted into SDFs so that later the principal component analysis can process the shape of apriori knowledge. This procedure gives more relevance to the shape of the organ in order to improve the segmentation in our interest. Furthermore, the point-wise localization of contours or landmarks is replaced by this automatic process that makes the algorithm more computationally intensive, easier to manage and implement. The procedure is illustrated in the following figure (fig3.8).

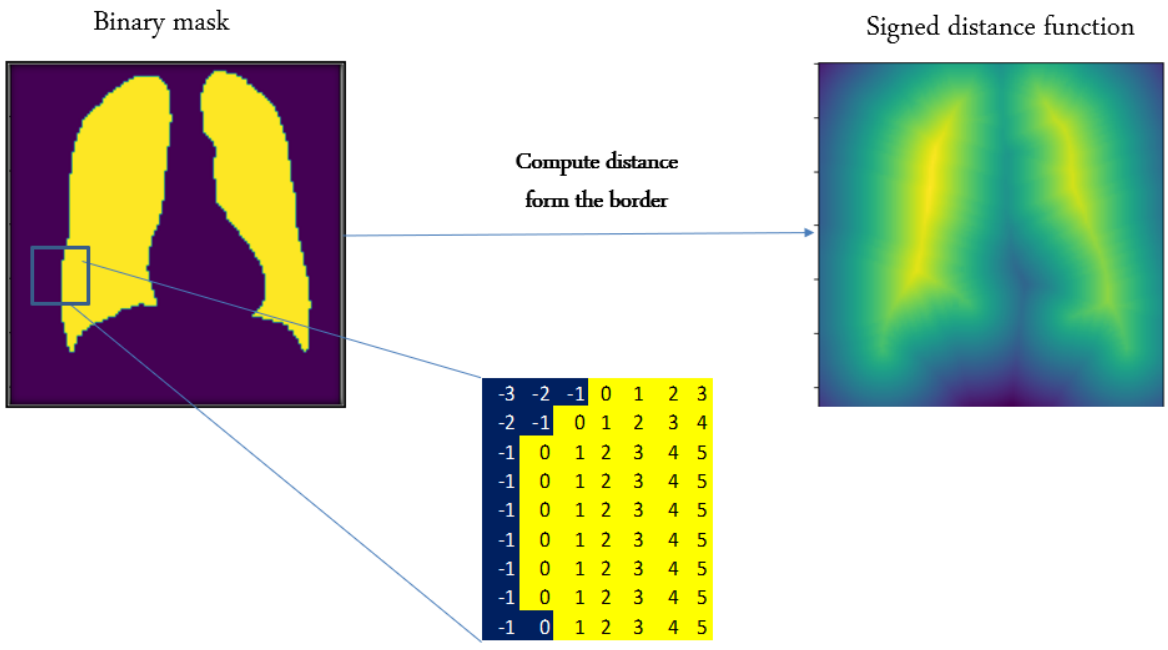


Figure 3.8: SDF of pulmonary binary mask

Principal Component Analysis

Principal component analysis constitutes the second and most determining process in the preliminary shape analysis related to estimating apriori shape information. It aims to capture the main shape variations of a training set while removing redundant information to include statistical apriori knowledge obtained into the deep neural convolutional network. PCA was introduced by Pearson in 1901 and became a popular approach in multivariate statistics. It aspires to find a linear subspace of lower dimensionality than the original feature space, where the new features have the largest variance To this end, PCA defines a new orthogonal coordinate system represented by principal components that optimally describe the variance in a single dataset and compress the latter onto a lower-dimensional feature subspace to maintain the most relevant information. In this context, it is considered one of the main algorithms involved in dimensional reduction.

Principal Component Analysis: Mathematical Steps

Mathematically, the idea of Principal Component Analysis (PCA) concerns extracting the principal modes of variation by computing the eigenvectors of the covariance matrix and therefore requires necessarily the following steps. Considering a dataset {x1,x2,x3,....xi } where i = { 1, 2, · · ·, N,} and each xi is a D-dimensional vector, PCA, this way, is involved in projecting the data onto an M-dimensional subspace, where the M dimension is lower than D. In the first instance, the data are centered at the origin by subtracting the mean observation formula1. It is intended as "mean centering" to perform PCA to ensure that the first principal component describes the direction of maximum variance. Indeed if mean subtraction is not performed, the first principal component might correspond more or less to the mean of the data. Hence, considering the origin as zero is needed for finding a basis, minimizing the mean square error of the approximation of the data. In this way, the data are zero average :

$$\frac{1}{N}\sum_{i=1}^{N} x_i = 0 \tag{3.1}$$

Secondly, by stacking the residual data in data matrix X column-wise, the covariance matrix is obtained as illustrated by the formula (3.2):

$$C = \frac{1}{N} X^T X \tag{3.2}$$

The covariance matrix is a symmetric matrix where its members describe the correlation between different data or parameters used in the analysis. Indeed The diagonal elements of the covariance matrix stand for the variance of each variable itself, and off-diagonal elements in the covariance matrix represent how variables are correlated with each other. Therefore the matrix fully describes the variability and the correlations inside the dataset. Subsequently, variance maximization is achieved by diagonalizing the covariance matrix by eigendecomposition leading to the eigenvectors v_i, coined principal components, represent the covariance maximizing basis while corresponding eigenvalues but constitute the eigenspectrum.

$$\lambda v = Cv \tag{3.3}$$

This process embodies the "core" of the principal components analysis as the eigenvectors (principal components) determine and report the directions of the new feature space, and the eigenvalues assess their magnitude. Solving the eigenproblem, the principal components assigned by PCA are decreasing variance, which is reflected by the respective eigenvalues.

Singular value decomposition

The principal component analysis is used to compute the dominant vectors representing a given data set and provide an optimal basis for reconstructing the given data. The computational basis of PCA is the calculation of the Singular value decomposition of the data matrix, which is equivalent to the eigenvalues and eigenvectors decomposition of the data covariance matrix. Singular Value Decomposition, or SVD, is a computational method applied to calculate principal components for a dataset. Using Singular value decomposition infact is sed to perform PCA is numerically robust and efficient. Futhermore, the intimate relationship between them can guide our intuition about what Principal component Analysis does and help us gain additional insights into this process. Eugenio Beltrami and Camille Jordan discovered the singular value decomposition independently, in 1873 and 1874, respectively. The singular values of the bilinear forms, represented as a matrix, considt of a complete and full set of invariants for bilinear forms under orthogonal substitutions. James Sylvester also arrived at the singular value decomposition for real square matrices in 1889, apparently independently of Beltrami and Jordan. The fourth mathematician to discover the singular value decomposition is Autonne in 1915, who arrived on the other hend via the polar decomposition. The first application of the singular value decomposition for rectangular and complex matrices seems to be reported by Carl Eckart and Gale J. Young in 1936;[25] they saw it as a generalization of the principal axis transformation for Hermitian matrices. However, practical methods for computing SVD date back to Ervand Kogbetliantz between 1954 and 1955 and Magnus Hestenes in 1958 and have a similar implementation to Jacobi's method, which uses plane rotations or Givens rotations. These approaches were replaced by the method of Gene H. Golub and William Kahan, published in 1965 (Golub Kahan 1965), which is based on Householder transformations or reflections. In 1970, Golub and Christian Reinsch published a variant of the Golub/Kahan algorithm that is still one of the most widely employed of the Golub/Kahan algorithm is still one of the most commonly used. Nowadays, SVD plays a predominant role in principal components analysis as the PCA includes numerous and computationally expensive operations, especially in analyzing large-size datasets such as image processing. In this light, the singular value decomposition is a more direct and expedient approach as it associates the principal components transformation a matrix factorization of the original dataset matrix.

In linear algebra, this process is a factorization of a real or complex matrix that generalizes the eigendecomposition of a standard square matrix as an extension of the polar decomposition. This technique provides a robust computational framework and enhances our understanding of what principal components are . It lets us compute them accurately for different datasets. The singular value decomposition (SVD) decompose the considered dataset X as :

$$\boldsymbol{X = U\Sigma W^T} \tag{3.4}$$

Where Σ is an n-by-p rectangular diagonal matrix of positive numbers (k), called the singular values of X; U is an n-by-n matrix, the columns of which are orthogonal unit vectors of length n called the left singular vectors of X; and W is a p-by-p whose columns are orthogonal unit vectors of length p and reached the right singular vectors of X. In terms of this factorization, the covariance matrix X^TX can be considered as :

$$\begin{aligned}\mathbf{X}^T\mathbf{X} &= \mathbf{W}\boldsymbol{\Sigma}^\mathsf{T}\mathbf{U}^\mathsf{T}\mathbf{U}\boldsymbol{\Sigma}\mathbf{W}^\mathsf{T} \\ &= \mathbf{W}\boldsymbol{\Sigma}^\mathsf{T}\boldsymbol{\Sigma}\mathbf{W}^\mathsf{T} \\ &= \mathbf{W}\hat{\boldsymbol{\Sigma}}^2\mathbf{W}^\mathsf{T}\end{aligned}$$

It underlines the eigenvector factorization of X^TX establishes that the right singular vectors W of X are equivalent to the eigenvectors of X^TX. In contrast, the singular values $\varsigma(k)$of X are equal to the square root of the eigenvalues λk) of X^TX. The SVD is revealed an efficient algorithm of X without having to form the matrix X^TX, so computing the SVD is now the standard procedure to obtain a principal components analysis from a data matrix[citation needed], unless only a handful of components are required. Indeed the SVD calculates the respective eigenvectors of X without having to form the covariance matrix X^TX, So computing the SVD now represents the standard to calculate a principal components analysis from a data matrix[citation needed], as it requires fewer steps and is more intuitive.

Full singular value decomposition

As previously discussed, singular value decomposition is based on a theorem from linear algebra which says that a rectangular matrix A can be decomposed into the product of three matrices : an orthogonal matrix U, a diagonal matrix S, and the transpose of an orthogonal matrix V :

$$\boldsymbol{M = U_r \Sigma V^T} \tag{3.5}$$

This process is also denoted as full SVD w where the matrix of right singular values, corresponding to the matrix of eigenvectors, is, according to linear algebra, square and composed of n columns and n rows, respectively. As a result, the Full SVD in the principal component analysis derives from the initial dataset, specifically the square matrix V of the eigenvectors of the covariance matrix M^TM. However, the eigenvectors retrieved are not orthonormal among them d therefore, it is linearly dependent, evidencing like the whole procedure reports principal components are unnecessary and consequently redundant in the lower-space. Hence, despite the directness of this approach compared to canonical PCA, the Full SVD process requires a considerable computational cost and time, especially for large datasets.

Reduced singular value decomposition

In the principal applications and image processing, it is unusual for the full SVD. Instead, it is often sufficient and faster, more economical for storage to compute a reduced version of the SVD, basing the dimension of the eigenvectors on the rank of the original dataset matrix.Van den Broek et al. [28] employed the SVD to reveal the rank and reduce the data dimension of multivariate images. This paper presents a feature reduction method using the singular value decomposition (SVD). Indeed SVD is applied in a new way and in drastically reducing computer processing. Reduced singular value decomposition is a mathematical technique. The insight underlying the use of SVD for these tasks is that it takes the original data, usually consisting of elevated dimensions, and breaks it down into linearly independent components. These components are, in some sense, an abstraction away from the noisy correlations found in the original data to sets of values that best approximate the underlying structure of the dataset along each dimension independently. The reduced, or economy-sized, SVD of the data matrix, indeed, is given by[21]:

Where r constitutes the rank of the original data matrix and corresponds to the number of linearly independent eigenvectors with the correspondent eigenvalues, by calculating only the latter, it excludes redundant information not strictly necessary for the projection of the dataset and retrieving of principal components, making the algorithm more flowing. The matrices U_k and V_k contain only the first k columns of U and V, and k contains only the first k singular values from . The matrix U_k is thus m×kr ,Σ_k is r×r diagonal, and V_k is k×n. The reduced SVD uses significantly less space and computation time if r" max(m, n). As a result, it is employed in many algorithms concerning the PCA transformation over dimensionally extensive datasets and our application.

$$\boldsymbol{M_r = U \Sigma_r V_r^T} \tag{3.6}$$

Principal Component Analysis: Denoising and Recostruction

Principal component analysis (PCA) is a powerful dimension reduction technique widely used in the data mining field. PCA aims to project the data into a lower-dimensional space while preserving the intrinsic information hidden in the data as much as possible. As previously discussed, it computes and finds eigenvectors of the covariance matrix defined as principal axes and sorts them by their eigenvalues according to the amount of explained variance. The centered data can then be projected onto these principal axes to yield principal components. For dimensionality reduction, one can keep only a subset of principal components and discard the rest in reconstructing the process of the data. Basically, PCA is an orthogonal linear transformation used to find a projection of all data into lower dimensions, whereas these termed principal components are those of the highest variance transformation. In this perspective, this transformation is just considered a matrix multiplication (3.7):

$$\mathbf{Y=XW} \tag{3.7}$$

Y is the lower-subspace based on retrieved principal components, X is input data, and W is a matrix of eigenvectors constituted by the principal axes delineating the maximum variance that can be used to recover the original matrix.

$$\boldsymbol{X_{rec} = YW^T + X_{mean}} \tag{3.8}$$

where X_{rec} is the reconstructed original space.

In summary, one can think of PCA projecting data onto a lower-dimensional subspace and reconstructing data back to the original dimension; the result of this process is called reconstruction. It occurs finding the projection that the best linear reconstructions of the data are as close as possible to the original data. The reconstructed data usually differs from the original one; the difference is measured as Euclidean distance between the original and reconstructed data and defined as the reconstruction error, which is the sum of eigenvalues of the ignored subspace and often associated with not relevant information. On this basis, the dimensionality reduction based on principal components analysis may also be an appropriate option when the variables in a dataset are noisy. Hence, discarding eigenvectors with small eigenvalues often result from noise, delineating PCA as a denoising technique. It removes from the reconstruction process the components with the lowest variance assessable to that corrupt data and related information. Consequently, most of the total variance is concentrated in the first few principal components than the same noise variance. The proportionate effect of the noise is negligible. The first few components a higher signal-to-noise ratio in reporting and back-projecting the original dataset. PCA thus can have the effect of noise reduction, concentrating much of the signal or information into the first few principal components, which can usefully be captured by dimensionality reduction. In contrast, the later principal components may be dominated by noise and disposed of without significant loss.

Our application

In our work, the principal component analysis aims to extract the a priori information related to the shape of the lung fields belonging to the training set. Indeed according to this idea introduced by Leventon, the PCA is not applied on the parametric geometric curves but on the signed distance functions (SDFs) of pulmonary contours, which are implicit and parameter-free shape representations. Following the steps involved in the principal component analysis, each SDF is vectorized in a single array with a size equal to the number of pixels in the image. This process allows the realization of a complete data matrix composed of 299 rows representing the number of samples or observations in the training set and 16384 columns that delineate the resolution of images. To compute the variance in shape using Principal Component Analysis (PCA), the mean shape is subtracted from each signed distance function to create a mean-offset map. In the ensuing stage, Singular Value Decomposition (SVD), the considered dataset is decomposed as:

$$\boldsymbol{X} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{W}^{\boldsymbol{T}} \tag{3.9}$$

where W is a matrix whose column vectors are equivalent to the eigenvectors representing the principal components of shape variation reported by the training dataset. Despite the size of the dataset considered, the last step allows obtaining efficiently and less expensively the eigenvectors derived from the covariance matrix that describe the main shape variations in the training set, interpreting the essential tool to integrate the shape apriori knowledge in tho the proposed network. Nevertheless, the deployment of the reduced version of SVD will entail some advantages , we use full SvD as descibed in the reference paper. This last step allows us to extract principal components describing the total variation captured by the training dataset. The eigenvector matrix also enables us to project the SDFs images in the shape space and ensure that the proposed network can be trained and predict in the latter. The principal component analysis also allows us to back-project the images into the original space by obtaining the respective SDF predicted and using the Heaveside function, the predicted binary mask. It is evident from Each SDF (Φ) in the training data to be represented by a mean SDF ($\overline{\Phi}$), matrix of eigenvectors (W), and a vector of predicted shape parameter b.

$$\boldsymbol{\Phi} = \overline{\boldsymbol{\phi}} + \boldsymbol{W}\boldsymbol{b} \tag{3.10}$$

In this perspective, it stands to reason the investigation of the denoising aptness bout the PCA algorithm can be exploited not only to capture shape information but also eventually refine

from noise effects the original data. In our algorithm, the reconstruction process based on PCA allows us to reconvert shape parameters into SDFs and then return to the original image space. We can reduce the effect of some artifacts by decreasing the number of components. This process referred to as 'component number optimization', employs a progressive reduction in the number of components to remove those associated with noise and to explore how this in the prediction of shape parameters and the subsequent reconstruction of the signed distance function and its binary mask may or may not improve the segmentation map.

In other words, the PCA in our algorithm, both in its standard version and in the SVD extension, embraces how tail features can be discarded as redundant or noisy information, thus reducing the dimensionality of the initial dataset while the most relevant data has been retained as a constituent and integral part of the apriori information.

Dataset & Ground Truth Generation

In this thesis, the Unet architecture is modified to generate a signed distance function (SDF) from the input image. Where predicted SDF is converted to shape parameters compactly represented and projected on the basis of principal components in shape space, in which finally the loss is computed. From a geometric point of view, this principal component analysis determines the best orthonormal basis Rm to represent a set of n-m dimensional samples of the training set. Therefore the input data can be projected to the new basis by the set of retained eigenvectors. Indeed the data can be synthesized by the leading principal components. In practice, only the first principal modes are necessary to model the biggest variations present in our training set. These p principal components are sorted in an eigenvectors matrix Wp. Then, the PCA can produce a new data set based on the training set φ_j:

$$\boldsymbol{\varphi} = \overline{\boldsymbol{\varphi}} + \boldsymbol{W_p X_{pca}} \tag{3.11}$$

where φ is the reconstructed sample and $\overline{\varphi}$ is the mean value and x_{pca} is called vectors or the eigenmodes of variation. This synthesis is considered in our application as "Shape parametrization". It employed as output in the proposed network constrain the latter to learn the shape information of the target organ and enforce the model output the global shape parameters of the anatomical curves, implicitly introducing continuity and smoothness terms into the segmentation process. In this light, the final layer of the neural networks integrates the shape information retrieved from the principal components, which allowed to project the predicted image into a shape parameter vector and compared it with the ground truth (b_{gt}). It implies this implies the apparent need to convert and generate the respective shape parameters (annotated as b ground truth) of SDF images so that they can be compared with those predicted and therefore proceed to train the neural networks, This essential stage is defined in the reference paper as the ground truth generation.

Ground Truth Generation

Principal component analysis allows each SDF (ϕ) in the training data to be represented by a mean SDF ($\overline{\phi}$), matrix of eigenvectors (W) and a vector of shape parameters, b: $\phi = \overline{\phi} + \mathrm{Wb}$, (1) where ϕ and $\overline{\phi}$ are the vectorized form of ϕ and $\overline{\phi}$, respectively. For each training example, in our method we can compute b as: $\mathrm{b} = W^T\ (\phi - \overline{\phi}) = W^T\phi_d$, (2) where ϕ_d is the vectorized difference SDF, $\phi_d = \phi - \overline{\phi}$. These parameters are used as the ground truth (b_{gt}) for training the proposed network. Hence they are computed subsequently to the eigenvector matrix.

$$\boldsymbol{b = W^T(\phi - \overline{\Phi}) = W^T\Phi_d} \tag{3.12}$$

Considering a training set composed of 299 training images and the respective binary masks, from the resulting SDF, the correspondent 299 shape parameters are generated to be used as output and reference of the network. Then the final prediction is computed as b: $W^T\phi_d$ or in the element-wise form: $b_i = \sum_{j=1}^{k} \mathrm{w}_{ij}\phi_d$, i = 1, 2, $\cdots$, k; (3) where w_{ij} is the value at the i-th row and j-th column of the transposed eigenvector matrix (W^T) and k is the number of shape parameters. Finally, the loss is defined as:

$$L = \sum_{i=1}^{k} L_i \text{ where } L_i = \frac{1}{2}(\hat{b_i} - b_i^{GT})^2.$$

Although this process concurs a simplification of the calculation of the loss function, it constitutes an additional step of our algorithm involving statistical variations of the shape parameters according to the implementation representing a feeble boundary with the concept of ground truth.

SP-net Architecture

SP-net is a fully convolutional deep neural network for shape prediction, which bases the configuration of most of its layers on the Unet architecture and employs the final to predict the outcome in the shape domain. SP-net provided a variant of Unet and presented a symmetrical structure consisting of a downsampling path involved in the general convolutional process and upsampling composed by transposed 2d convolutional layers to retrieve the original resolution input images processed in the learning phase. All these will be deepened and explored in the following paragraphs to be conceived more favorably the operation of the proposed network. The network has a basic foundation that looks like:

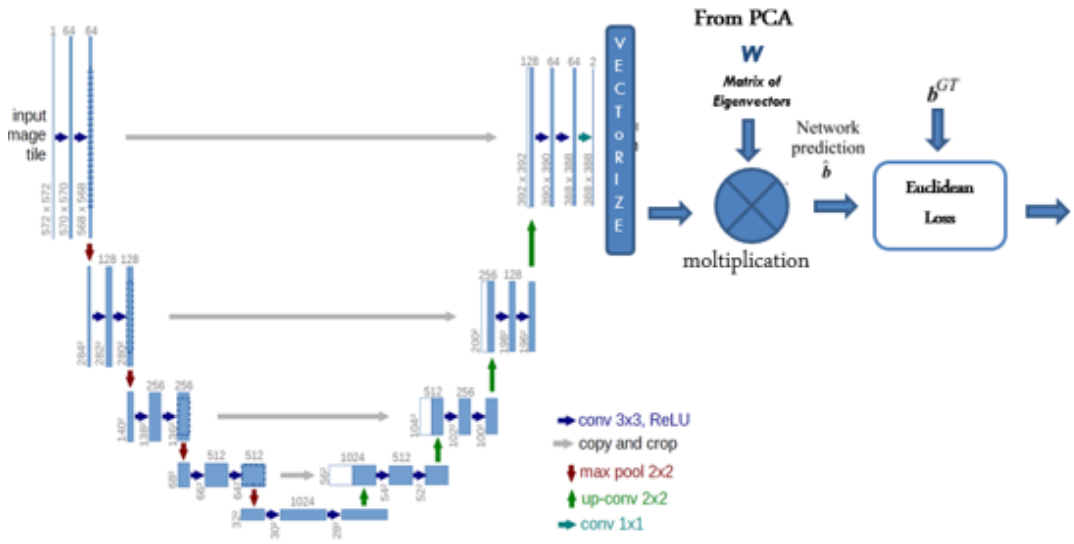


Figure 3.9: SP-net Architecture

The Encoder Path

In the first instance, the contraction path (also called the encoder path) is used to capture and process the contextual information in the image. It is constituted by a traditional stack of convolutional and max-pooling layers. Indeed the input images, composed of 128x128 pixels and in grayscale, thereby with a single channel, are convoluted by the first layers as the convolution process will increase the depth of the image, extracting 64 distinct features according to different kernels of filters. Subsequently, the consecutive process induces the elaboration of others 64 parts at the same resolution of input images. The max-pooling process halves down the size of the image, preserving maximum pixel value and thus obtain a pooled feature map and maximize its information content. The size or dimension of the filter and strides are two paramount hyper-parameters in this max pooling operation. The size is also reduced due to padding issues, but the implementation here uses padding= "same." A significant point to note the considered block using convolution operation and especially the pooling operation reduces the image's size. It is intended as a down-sampling operation. At the bottom-most part still, two convolutional layers are built but with no max pooling and constitute the network's bottleneck. The functions involved in the first block are repeated for other constituents of the encoder. They increase the number of channels and thus the number of filters, applying greater abstraction of the image information and reducing the spatial resolution. It implies a much more compact representation of the received image, and subsequent steps are used to recover contextual and background information lost during the down-sampling stage.

The Decoder path

The operations collected in the decoder path definition aim to recover the high resolution of the original images. Indeed, in the expansive path, the image data will be upsized to its original resolution with a series of transposed convolution. Transpose convolution (sometimes called deconvolution or fractionally convolution) is a technique for sampling an image with learnable parameters. On a high level, transposed convolution is precisely the opposite of a standard convolution, i.e., the input volume is a low-resolution image, and the output volume is a high-resolution image. It is nicely explained how a regular convolution can be expressed as a matrix multiplication of input image and filter to produce the output image. It involves the taking of the transpose of the filter matrix, we can reverse the convolution process. Hence the name transposed convolution. After the transposed convolution, the image is upsized. Then, this image is concatenated with the corresponding image from the contracting path and together makes an image a larger size. Here is to combine the information from the previous layers to get a more precise prediction. Same as before, this process is repeated for the three remaining blocks to restore the original resolution. Indeed the uppermost of architecture reshape the image to satisfy our prediction requirements.

Final layers

In conclusion, the latest layers are designed to integrate the shape apriori information derived from principal components analysis and converting the predicted image into the respective shape space where loss is computed. In this perspective, a specific dot layer is involved in multiplicating the network outcome and projecting it along the novel orthobasis represented by eigenvectors matrix and retrieved from previous principal component analysis. As a result, the neural networks are trained on the correspondent shape parameters, taking into account the information and restraint in the shape domain. It is deduced that the neural network has an image, an input, and a projection of the predicted image on the principal components, described as shape parameter (b) in the output. However, as evidence from the article and since the analysis of the principal components is an orthogonal transformation, it is possible to reconstruct the predicted segmentation map by the latter .

3.4.4 Kernel SP-net

Introduction

Principal component analysis (PCA) has demonstrated a popular tool for linear dimensionality reduction, primarily feature extraction. Indeed in the previous sections, we focused on how principal component analysis can extract apriori information from a training dataset consisting of CXR images and masks, exploiting the SDF transformation on the latter to incorporate the shape-awareness in the prediction of segmentation maps within the proposed network . However, if the given dataset has more complicated structures which cannot be well represented in a linear subspace, standard PCA will not be very convenient. One cannot assert that linear PCA will always detect all structures in the training set. By the use of suitable nonlinear components, one can extract more information. Hence we intend to take a further step forward by investigating a nonlinear extension of linear PCA and study how it could be integrated with the proposed method. In this connection, kernel Principal Component Analysis (Kernel PCA)[144] is used to extract nonlinear features and reduce dimensionality. It is revealed as a nonlinear form of PCA, which better exploits the complicated spatial structure of high-dimensional data. Kernel PCA was implicitly performing a linear PCA in some high-dimensional kernel feature space that was nonlinearly related to input space by using a suitable nonlinear kernel function mapping. Indeed, the kernel implemented induces KPCA to correlate the dataset nonlinear features ad compute principal components describing the variance of the data according to this interpolation. As a kernel method, kernel PCA suffers from the problem of kernel choice. Therefore, a procedure is also required to ensure the selection of the best kernel function, which is very suited is very well suited to extract interesting nonlinear structures in the training dataset . Against the background, we have tried different kernel provided in the relevant literature on the training dataset to understand which is the best in extracting the more relevant shape information and integrating into the shape predictor network. The kernel choice is based on comparable reconstruction errors of images in the original space, assessing the latter on the validation set constituted by images of the same dataset of the training and the test set, consisting of AIforcovid Images. For this reason, considering the applicability of the kernel methods in the SP-net network, we coined the term "Kernel SP-net" to underline the pertinency of these techniques.

Kernel Principal Component Analysis

kernel-PCA is a prominent nonlinear extension of the classical dimensionality reduction technique as it conducts principal component analysis for the feature vectors. More precisely, given data points $x_i \in x$, i = 1, 2, . . . , n, kernel PCA outputs a set of principal functions by the following two-step procedure: (i) transform the data nonlinearly into the feature space H, i.e., Xi $\rightarrow$, (ii) solve the linear PCA problem for the feature vectors, i.e., solve the directions in H for which the variance of (Xi) along those directions is maximized.The algorithm of kernel PCA is described as follows for the detail, see Scholkopf et al.[144]) $\tilde{\Phi(x)}$:=$\Phi(x)$-$\frac{1}{n}\sum_{j=1}^{n}\Phi(x_{\mathrm{j}})$.

The estimated covariance matrix is given by; C = $\sum_{j=1}^{n}\Phi(x_{\mathrm{j}})\Phi(x_{\mathrm{i}}))$ with the centered feature vectors. The principal directions $v \in C$ are given by the unit eigenvectors corresponding to the largest eigenvalues, and thus the problem is converted to solving the eigenproblem:
$Cv{=}\lambda v$

By using the kernel trick, this problem is reduced to the generalized eigen value problem that finds: v=$\sum_{j=1}^{n}\alpha_i\Phi(x_{\mathrm{i}})$ such that:

$K\alpha = n\lambda\alpha$ subject to: $\alpha^T K\alpha = 1$
where K is the n $\times$ n centered Gram matrix with $K_{ij} = k(x_i, x_j)$ Let $\lambda_1 \leq \lambda_2 ..... \leq \lambda_n \leq 0$ denote the ordered eigenvalues of K with associated eigenvectors $\alpha_i, ...\alpha_n$, where $\alpha_j = (\alpha_{1j}...\alpha_{nj})$ which are normalized. The j-th principal direction $v_j \in C$ is then given by:

$$\boldsymbol{v_j = \frac{1}{\sqrt{\lambda_j}}\sum_{i=1}^{N}\alpha_{ij}\Phi(X)} \tag{3.13}$$

the j-th principal component of the data point X_i is given by:

$$\boldsymbol{(v_j, \Phi(X)) = \frac{1}{\sqrt{\lambda_j}}\sum_{i=1}^{N}K\alpha_{ij} = \sqrt{\lambda}\alpha_{ij}} \tag{3.14}$$

For a test point X out of the sample, the j$_t$*hprincipalcomponentissimilarlygivenby*;

$$\boldsymbol{(v_j, \Phi(X)) = \frac{1}{\sqrt{\lambda_j}}\sum_{i=1}^{N}K(X, X)\alpha_{ij}} \tag{3.15}$$

where $k(x, y)$ is the centered kernel:

$$\tilde{k}(x,y) = k(x,y) - \tfrac{1}{n}\textstyle\sum_{i=1}^{n}k(x,\mathbf{X}_i) - \tfrac{1}{n}\sum_{i=1}^{n}k(\mathbf{X}_i,y) + \tfrac{1}{n^2}\sum_{i,j=1}^{n}k(\mathbf{X}_i,\mathbf{X}_j)$$

Kernel Choice and Hyperparameters Selection

Kernel PCA suffers from the problem of kernel choice. Therefore, a procedure is also required to ensure the selection of the best kernel function, which is very suited is very well suited to extract interesting nonlinear structures in the training dataset. Different kernels, provided in the relevant literature, are tried on the training dataset to understand which is the best in extracting the more relevant shape information and integrating it into the shape predictor network. The kernel choice is based on comparable reconstruction errors of images in the original space, assessing the latter on the training, validation, and test set constituted by binary masks. in this context kernel functions are explored based on their potential advantages in analyzing such training dataset, therefore the following functions are explored:

- Cosine Similarity kernel
- Radial Basis Function Kernel
- Laplacian Kernel
- Polynomial Kernel

Kernel on Cosine Similarity

It is called cosine similarity because Euclidean (L2) normalization projects the vectors onto the unit sphere. The dot product is then the cosine of the angle between the points denoted by the vectors. Cosine similarity considers the extent to which two vectors point in opposite direction, in the same direction,, or whether they are orthogonal.

$$\boldsymbol{K(x,y) = \frac{xy^T}{||x||||y||}} \tag{3.16}$$

Radial Basis Function Kernel

The RBF kernel function for two points or vectors computes the similarity based on euclidean distance , perceving how close they are to each other. This kernel also named Gaussian Kernel can be mathematically represented as follows:

$$\boldsymbol{K(x,y) = exp(-\gamma||x-y||^2)} \tag{3.17}$$

Laplacian Kernel

The Laplacian kernel is a variant based on the radial basis function kernel defined as follows, where x and y are the input points or vectors, and the manhattan distance is measured instead of the euclidean distance.

$$\boldsymbol{K(x,y) = exp(-\gamma||x-y||_1)} \tag{3.18}$$

Polynomial Kernel

The function polynomial kernel computes the degree polynomial kernel between two vectors. The polynomial kernel represents in fact the similarity between two vectors. Conceptually, these kernels consider the similarity between vectors under the same dimension and across dimensions. When used in machine learning or deep learning algorithms, this allows accounting for feature interaction.

$$\boldsymbol{K(x,y) = (1 + x * y)^d} \tag{3.19}$$

Kernel PCA in our application

In our application the kernel PCA allows us to perform the same steps as those performed by the approach provided by the principapl componet analysis. Through the use of the Kernel trick, the latter allows us to retrieve a covariance matrix defined according to the internal product established by the kernel function chosen. Then it is possible to obtain the respective eigenvector matrix and project the corresponding Shape parameters. This process involves and allows the reconstruction of the masks and segmentation maps predicted by the network called Kernel SP-net and to implement an upstream process for the dateset and ground truth generation. In addition, the process of projection and rear projection, assumed as reconstruction, allows us to adequately estimate the most appropriate kernel.

Kernel SP-net Architecture

The Kernel SP-net architecture (fig 3.10), as it constitutes an extension of the Shape Predictor Network method, structurally appears identical to the latter. Except for the fact that the integration of the apriori information includes the eigenvectors matrix retrieved from the Kernel Principal Component Analysis. The figure shows that it has the same number of layers as the previous networks and the same characteristics. However, in the final layers, the network includes the shape knowledge obtained through the application of the Kernel PCA on the training set composed by Signed distance functions.

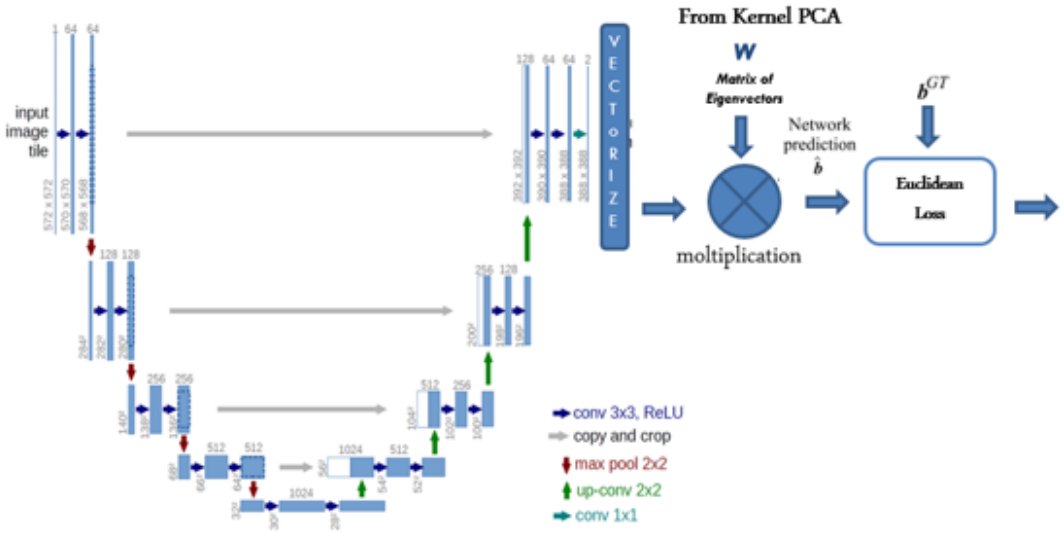


Figure 3.10: Kernel SP-net Architecture

3.5 Evaluation metrics

Evaluation metrics measure the quality of a Deep or machine learning model by explaining its performance in terms of figures. In this context, the segmentation methods are evaluated using appropriate metrics, which are divided into qualitative and quantitative. A qualitative metric involves visual inspection and rating of the segmented image by a domain expert. On the other hand, a quantitative index compares the predicted segmentation with the available ground truth of the medical image, annotated manually by a group of medical experts. Indeed, comparing images to evaluate segmentation quality is an essential part of measuring progress in this research area. In this perspective, the Dice coefficient and Jaccard index are the most functional and commonly used metrics for evaluating segmentation tasks in medical imaging and convolutional neural networks trained for image segmentation. Additionally, using two or more evaluation metrics is paramount to prevent a model from performing well on one evaluation metric but poorly using another measurement. The Dice coefficient, also known as the Sørensen–Dice index or simply Dice score, is a statistical tool that measures the similarity between two sets of data or, in this case, two sample images. This index has become the most broadly used tool in validating image segmentation algorithms created with AI. It is determined considering twice the overlapping area between the ground truth image and the segmented one divided by the total sum of pixels in both images. The Jaccard Index, also known as Intersection-Over-Union (IoU), is one of the most widespread metrics in semantic segmentation. It is calculated by the overlap between the predicted segmentation and the ground truth divided by the area of union between the predicted segmentation and the ground truth. Both metrics range from 0–1, with 0 signifying no overlap and 1 signifying perfectly overlapping segmentation.

CHAPTER 4

Design of Experiment

Design of Experiment

The following section expresses the strategy used to train the proposed neural network, emphasizing the optimization of hyper-parameters such as the learning rate and the policy adopted. In addition, the techniques of regularization employed in this method are illustrated to avoid the overfitting of the neural network, giving more prominence to the enforceability of the Earlystopping technique and Data Augmentation well as to the feasibility of its use. In conclusion, it will be reported with which hyperparameters the experiment will be conducted, and thus SP-net will be trained so that it can be compared under the appropriate conditions with U-net.

.

4.1 Introduction

In the design and accomplishment of an experiment, the training process is the most challenging part of deep learning techniques; in general, it is by far the most time-consuming, both in terms of effort required to configure and set the process and computational complexity necessary . Furthermore, it is dependent on the dataset size and available processing power; therefore constitutes a process that is not well defined in those terms. The purpose of training is to minimize the difference between the network's output and ground truth data. It is also required that the model would produce accurate results when presented with data that was not used during the training phase. The training phase of a neural network from an example is usually conducted by considering the difference or error between the predicted output of the network and target output, intended as the error. The network then adjusts its weighted associations according to a learning rule and using the respective error value. Other adjustments will cause the neural network to produce output that is increasingly similar to the target output. After a sufficient number of these attempts, the training can be terminated based upon specific criteria. The training process is solved using an optimization algorithm or method that searches through a space of possible values for the neural network model weights for a set of weights that results or brings in a good performance on the training dataset. Deep learning convolutional neural network models learn to map inputs to outputs given a training dataset of examples. Specifically, the training process involves finding a set of weights in the network that proves to be good enough to solve the specific problem.

In this connection it provides an iterative scheme, meaning that it progresses step by step with slight updates to the model weights each iteration and, in turn, a change in the performance of the model each iteration, solving an optimization problem that finds for parameters that result in a minimum loss when evaluating the samples in the training dataset. Indeed optimization is a directed search procedure, and the problem that we wish to solve when training a convolutional network model is complicated. The best general method or algorithm known for solving this issue is stochastic gradient descent, where model weights or are updated each iteration using the backpropagation of the error algorithm. The backpropagation algorithm compares the network output with the expected output or ground truth and calculates an error measure. From that result, it backtracks the change or gradient of every step in the computational process using the chain rule partial derivatives. Against this background, training has been the object of studies and investigations in the proposed method. We consider two essential arguments as optimizing hyperparameters and decreasing the chance of overfitting the training data. Overfitting is a problem that arises in neural network training. When a model is overfitted

to the training data, it loses its generalization capability, and further training decreases the model's performance on unseen data. On the other hand, hyperparameter optimization or tuning is the problem of choosing a set of optimal hyperparameters for a learning algorithm. A hyperparameter is a parameter whose value or used to handle the learning phase. By contrast, the other parameters (typically node weights) are learned. In the former case, we investigate the feasibility of regularizer techniques like Data Augmentation and Early stopping. In the latter case, we find a combination of hyperparameters that returns an optimal model which reduces a predefined loss function and in turn increases the accuracy on given independent data, paying particular attention to the initialization of learning rate and the choice of related policy involved in the backpropagation of the proposed neural network. In the following section, we report the principal techniques involved in reducing overfitting and the test conducted to retrieve optimal hyperparameters.

4.2 Regulaziter Tecniques

One of the most critical aspects when training neural networks are avoiding overfitting. This issue is fundamental in supervised machine learning, which prevents us from perfectly generalizing the models to fit well-observed data on training data and unseen data on the testing set. In other words, The model learned specific patterns specific to the training data, which are irrelevant in other cases. The network fails to generalize the features/pattern for the testing data. There are several manners in which overfitting is reduced in deep learning models. However, the proposed method in this content uses Functional solutions such as dropout regularization, batch normalization, which are fully integrated into the network layers. Nonetheless, the best option is to get more training data. Unfortunately, in real-world situations, often there is no possibility due to time, budget, or technical constraints. For this reason, we explore the use of the data augmentation technique turned into an artificial way to increase the size of the training dataset. Moreover, overfitting is intensified by looking at validation metrics, like loss or accuracy. Ordinarily, the validation metric stops improving after several epochs and begins to decrease afterward. On the other side, The training metric improves because the model seeks to find the best fit for the training data, resulting in worse performance on testing sets. On this basis, we investigate the viability of the Early stopping regularization, which guides how many iterations can be run before the learner begins to over-fit, observing the improvements of the validation loss function. In this perspective, we presented A brief description of these overfitting solutions and their relevance and adaptation in the proposed neural network.

4.2.1 Dropout technique

Dropout is a regularization technique based on the random drop units (along with their connections) from the neural network during training, proposed by Srivastava et al. in 2014 by their paper[182] Dropout: A Simple Way to Prevent Neural Networks from Overfitting. Dropout is implemented per layer in a neural network. It can be employed with most layers, such as convolutional layers, dense fully connected layers, and recurrent layers such as the long short-term memory network layer. In simple terms This technique, during training, some number of layer outputs are randomly ignored or better "dropped out." This implies making the layer look like and being treated like a layer with a different number of nodes and connectivity to the last layer. As a result, each update to a layer during training is performed with a different "view" of the configured layer. It means that their contribution to the activation of downstream neurons is temporally removed in the forward pass, and any update of the weights is not applied to the neuron in the backward pass. As a neural network learns, the neurons' weights are established in their context within the network. Neuron weights are tuned for specific features that provide some specialization. Neighboring neurons rely on this specialization, which, if taken too broadly, can result in a fragile and over-specialized model for the training data. The reliance on context for a neuron during training relates to complex co-adaptations. Imagine that neurons are randomly deleted from the network during training. Other neurons will have to step in and handle the required representation to make predictions for the missing neurons. It is believed to result in multiple independent internal representations learned by the network. From this perspective, the network becomes less sensitive to the specific weights of neurons. It, in turn, results in a network that is capable of better generalization and is less likely to overfit the training data. On these grounds, we adopted the dropout layer as an integral part of the proposed network to reduce the probability of overfitting.

4.2.2 Batch Normalization

Batch normalization is one of the most widely used techniques in regularization as it normalizes the set of activations in a layer. Normalization works by subtracting the batch mean from each activation and dividing it by the batch standard deviation. This normalization technique, along with standardization, is a standard technique in the preprocessing of pixel values, which is used to render artificial neural networks faster and more stable by normalization of the layers' inputs by re-centering and re-scaling. This method was proposed by Christian Szegedy and Sergey Ioffe in 2015.[183] Whereas the outcome of batch normalization is conspicuous, the reasons behind its effectiveness remain under discussion. It was believed that it could mitigate the problem or issue of internal covariate shift, where parameter initialization and changes in the distribution of each layer's inputs affect the network's learning rate.More recently, some scholars have argued that batch normalization does not reduce internal covariate shift but rather smooths the objective function, improving performance.On the other hand, others sustain that batch normalization achieves length-direction decoupling and thereby accelerates neural networks[184]. On this premise, we embrace the batch normalization in the convolutional and deconvolutional layer, which constitute the core of the Shape predictor network, in order to regulate the operations in image learning and processing,

4.2.3 Data Augmentation

One of the more widespread strategies to avoid overfitting is to increase the size of the training dataset. As illustrated, when the training data size is small, the network tends to have greater control over the training data. Nevertheless, gathering large amounts of data in real-world scenarios is a tedious time-consuming task; hence, collecting new data is not viable. Data augmentation provides techniques to increase or better augment the size of the existing training dataset without any support of external addition. If our training data consists of images, image augmentation techniques like horizontal vertical flipping, translation, rotation, or adding noise, cutouts can be applied to the existing training images to increase the number of instances. Data Augmentation is used to artificially expand the size of a training set by creating edited data from the existing one. It is a common practice to use DA to prevent overfitting, or the initial dataset is too small to train on, or even o squeeze better performance from the proposed model. In this light, we investigate the possibility of using the overcited technique in or context to increase the training dataset. We consider basic operations carried out on the training set, belonging to the Jsrt Dataset and the Montgomery Dataset. These operations, as mentioned above, constitute geometric transformations so that the size of the initial training dataset can be increased and the proposed method can be generalized to the test set. On this premise, we investigated geometrical mechanisms and eventually added noise to render the training images similar and characterized by the artifacts and misalignments of the Aiforcovid Dataset image acquisition process. The data augmentation proposed in our application to reflect the characteristics present in the test set makes use of rotation operations of angles between 0 and 90 degrees, horizontal and vertical tilts, and slightly shear misalignments. Moreover, we include the artificial corruption by Gaussian Noise in every image in order to represent the poor contrast of test images

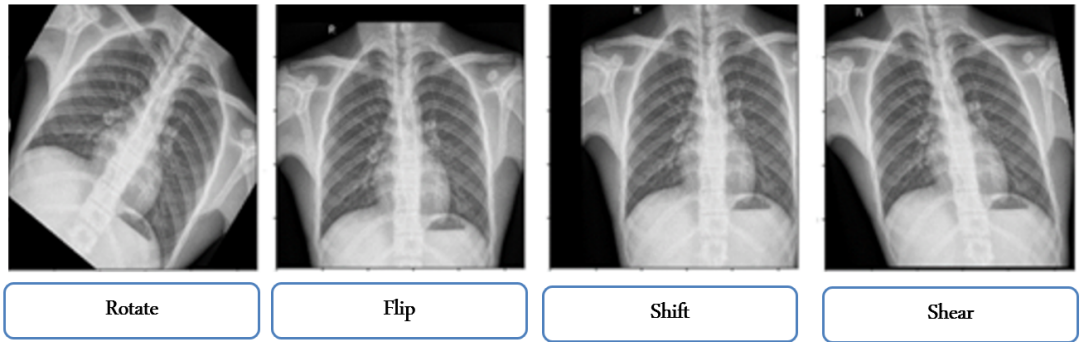


Figure 4.1: Data Augmentation : Geometrical transformartions

Dimension of Augmented Dataset

In the Deep Learning field, the performance and results of a model improve with the amount of data that has been used to train it. As previously discussed, Data Augmentation artificially increases the size of the training set by generating a new variant of each training instance. In this sense, It is a widespread and very well-known and technique for all computer vision problems that allows the generation of new training instances belonging to the same class as the base instance. In this manner, data augmentation forces the model to be more tolerant, increasing its capacity to generalize. This technique is usually during the training phase, inducing the network trains on slightly different images, artificially generated by the training set, in each epoch by implementing an image data generator employed in the training process. Nevertheless, in the proposed method, The training set is significant for extracting and processing apriori information retrieved by a shape analysis conducted before the training. From this assumption, it follows that data augmentation shall be a preliminary process to the network training since applied at this stage, it would also change the content and size of the matrix of the eigenvectors and, therefore, the principal components used by the network to predict in the shape domain. By understanding this order of processes, it is essential to determine the size of the augmented dataset and the number of artificially generated images from which the respective shape parameters are derived. In this respect, we investigate the relevance of Data Augmentation. We have realized several Datasets of different sizes based on Jsrt and Montgomery masks as the training set using this technique. After deriving the respective SDFs and eigenvectors Matrix, Twenty lung masks belonging to different classes (AiforCovid Dataset and an outsider part of JSRT Montgomery Dataset) were reconstructed employing the reconstruction, or better the retro projecting introduced by the principal component analysis. In this way, it is possible to demonstrate how this process could improve the Dice Coefficient, and Jaccard Index measured between reconstructed masks and real ones. As one can infer from the following, 4000 images and masks is the optimal dimension.

4.2.4 Earlystopping

In the broad set of strategies against this curse of overfitting is called regularization, early stopping is one paramount technique. It has a long history and background, which can be dating back to the 1970s. Also, it is widely used in iterative algorithms, especially in the first neural networks starting from the 1990s. Early stopping offers the considerable benefit to reduce overfitting without compromising on model accuracy. The main idea behind early stopping is to stop training before a model starts to overfit. This strategy is known as probably the most commonly used form of regularization in deep learning. Its popularity is due to both its effectiveness and its simplicity. Early stopping is an unobtrusive form of regularization in that it requires almost no change in the underlying training procedure, the objective function, or the set of allowable parameter values. It means that it is easy to use early stopping without damaging the learning dynamics. It contrasts with weight decay, where one must be careful not to use too much weight decay and trap the network in a wrong local minimum corresponding to a solution with pathologically small weights. Early stopping may be used either alone or in conjunction with other regularization strategies. Even when using regularization strategies that modify the objective function to encourage better generalization, it is rare for the best generalization to occur at a local minimum of the training objective. Moreover, it is also helpful because it reduces the computational cost of the training procedure. Besides the noticeable reduction in cost due to limiting the number of training iterations, it also has the benefit of providing regularization without requiring the addition of penalty terms to the cost function or the computation of the gradients of such additional term. However, Early stopping requires a validation set, which means some training data is not fed to the model. Indeed, it achieves A compromise to train on the training dataset and stop training when performance on a validation dataset starts to degrade. These early stopping rules work by splitting o the original dataset into a new training set and a validation set. The error on the validation set is evaluated as a priority for the generalization error in establishing when overfitting has begun. These methods are most usually employed in the training of neural networks. The validation set is used to determine and verify how well the model generalizes to unseen data. When the error on the training set begins to deviate from the validation loss, a threshold can be set to determine the early stopping condition and the correct number of epochs improvement after which training will be stopped as known as patience parameter It outlines how the validation set used in conjunction with early stopping is essential to determine the optimal training zone for a model.

Given these observations, Earlystopping emerges as an unavoidable technique in the learning phase of deep neural networks. Indeed we employed it in our method by realizing a validation set consisting of images and masks of the Montgomery and JSRT Dataset has been created. However, this issue shows that it is unfeasible to use the shape parameters obtained from the respective SDFs to calculate the validation loss on which to base the Earlystopping process. The shape vectors are different from those obtained from the training dataset a and also cannot be projected based on the information received from the latter . and depend on the statistical processes used during the algorithm, making fleeting the meaning and considering ground truth for shape parameter vectors. It leads us to consider the masks of the validation set as ground truth on which to calculate a new loss on Dice coefficient intended and termed as Dice loss.

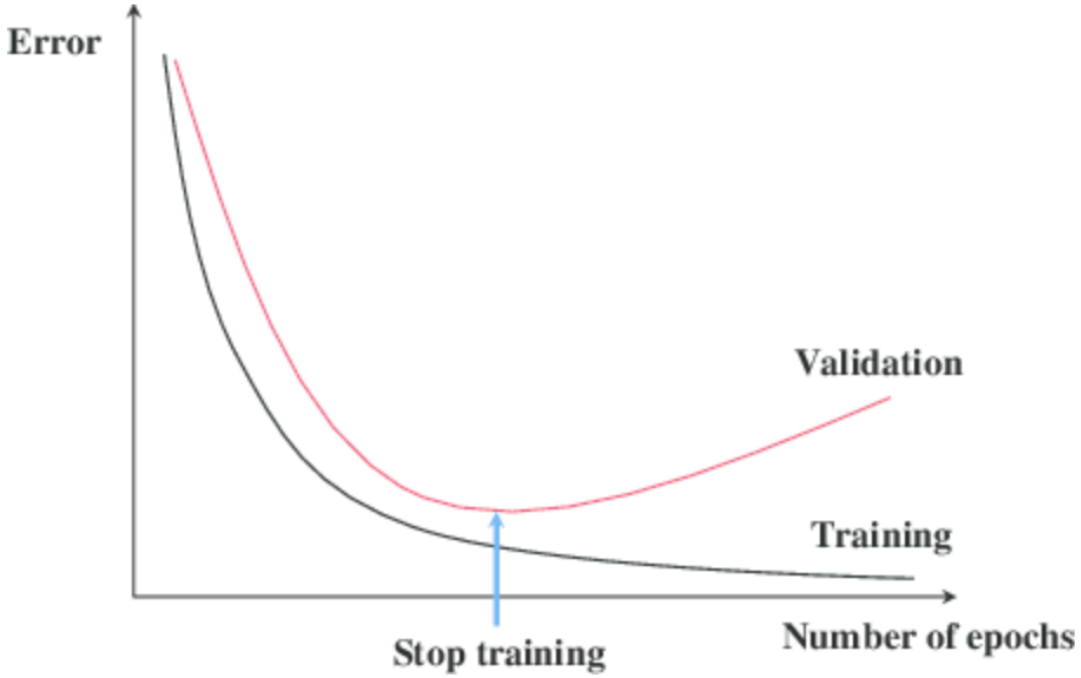


Figure 4.2: Earlystopping Techniques

4.3 Hyperparameters Optimization

4.3.1 Introduction

The development and training of neural networks are still unpredictable and challenging procedures. The difficulty of tuning these models does training and developing is more of an art than a science, based on the researcher's knowledge and experience. One of the reasons for this difficulty is that the training procedure of deep learning models includes several hyperparameters that affect how the training phase fits the model to the data. These parameters directly control the behaviors of training algorithms and significantly affect the performance of the resulting deep learning models. Unlike the internal model parameters, such as the neuron's weights, which can be learned from the model training phase, hyperparameters are set before the learning process. . Hyperparameters are the variables that determine the network structure) and the variables which set how the network is trained(E Learning Rate). However, while the former depends on the network architecture used, the latter requires careful and necessary optimization. In machine learning, hyperparameter optimization or tuning is the problem of choosing a set of optimal hyperparameters for a learning algorithm. In this context, the Learning rate might be the most crucial hyperparameter in deep learning, as the learning rate decides how much gradient to be backpropagated. It, in turn, decides by how much we move towards minima. The small learning rate makes the model converge slowly, while the significant learning rate makes the model diverge. So, the learning rate needs to be correct and honed. When training convolutional neural networks, it is often helpful to reduce the learning rate as the training progresses. It can be done by using predefined learning rate schedules or adaptive learning rate methods. These methods involve learning rate reduction during the training process are better known as Learning rate policies, and since they also depend on the architecture of the network and the initial data set, it is necessary to determine which beseemeth the proposed method. In this context, the number of epochs is revealed as a hyperparameter of the utmost importance as it defines the number of times that the learning algorithm will work through the entire training dataset. In this regard, The batch size and steps_per_epochs need to be configured to allow appropriate network training. The batch size is identified as several samples processed before the model is updated, whereas steps_per_epoch indicates the number of batch iterations before a training epoch is considered finished. It may be functional with vast data set or generating random data augmentations. As a result, both hyperparameters shall be set according to the training set size. In the following paragraphs, we will explain which criteria will be used to select and optimize the hyperparameters listed above and involved in the network's training.

4.3.2 The Learning Rate Optimization

The learning rate may be the most critical hyperparameter when configuring the neural network. Therefore, it is crucial to know how to investigate the effects of the learning rate on model performance and build an intuition about the dynamics of the learning rate model behavior. The learning rate is a parameter that controls and handles how much to change the model in response to the estimated error each time the model weights are updated. Setting the learning rate is arduous as a value too small may result in a lengthy training process that could get stuck, whereas a value too large may involve learning a sub-optimal set of weights too fast or an unstable training process. Indeed the learning rate controls how quickly the model is adapted to the problem. Lower learning rates require more training epochs given the more minor changes made to each update's weights, whereas more significant learning rates result in rapid changes and require fewer epochs.

In this light the challenge of training deep learning neural networks involves carefully selecting the learning rate as it may be crucial or the model. As suggested by Ian Goodfellow, "The learning rate is perhaps the most important hyperparameter. If you have time to tune only one hyperparameter, tune the learning rate." [185] In this perspective, we consider the training application strategies and experiments involved in selecting the initial value learning rate and the related policy used to reduce scheduler or adaptively learning rate as the training progresses.

Learning Rate Test: estimation of reasonable interval

There are multiple ways to choose a good starting point for the learning rate. A naive and trivial approach is to try a few different values and see which one gives the best loss without sacrificing training speed. Leslie N. Smith proposes a powerful technique to select a range of learning rates for a neural network in the 2015 paper "Cyclical Learning Rates for Training Neural Networks."[186] The trick is based on training a network starting from a low learning rate, increase the learning rate exponentially for every batch, and record the learning rate and training loss for every step. This is a simple way to estimate the reasonable minimum and maximum boundary values with one training run of the network for a few epochs. It is a "Learning rate range test". This test is enormously valuable whenever facing a new architecture or dataset. For learning rates that are too low, the loss may decrease but at a very shallow rate. When entering the optimal learning rate zone, we observe a quick drop in the loss function. Increasing the learning rate further cause an increase in the loss as the parameter updates to cause the loss to "bounce around" and even diverge. In this respect, we operate different learning rate strategies to apply the "Learning Test" to the proposed method, employing some epochs equal to ten, in order that the learning rate can be increased and reaches the set values in its exploration. In the successive section, we present and describe the learning rate policies, the rate of learning engaged in our training process.

Learning Rate Policies

As previously introduced, a highly non-trivial parameter in training convolutional neural networks is the optimization step size intended as the learning rate. Indeed, proper tuning of the learning rate can significantly reduce training time at different learning processes while ensuring that the model converges to a wider local minimum generalizing better to unseen data[167]. This pursuit of optimizing learning rates has yielded different methods; several proposing various policies that modify the learning rate according to a predefined function or performance metric(s) evaluated during training [185], while others using adaptive methods that modify learning rates according to the local first-order information. These methods, also referred to as "optimizers" stochastic gradient descent, typically train deep learning models. However, there are many variations of stochastic gradient descent: Adam, RMSProp, Adagrad, SGD with decay, etc. Thie latter can be divided into learning rate schedules or adaptive learning rate methods. Learning rate schedules seek to adjust the learning rate during training by reducing the learning rate according to a predefined schedule. Standard learning rate schedules include time-based decay, step decay, and exponential decay. Adaptive gradient descent algorithms such as Adagrad, Adadelta, RMSprop, Adam provide an alternative to classical SGD, as they set the goal of minimizing the objective function of a network by using and considering the gradient. Nevertheless, there has been no clear winner so far among these different strategies. Indeed, each method achieving different performance levels under varying settings like the network architecture, dataset, and other hyperparameters. It leads us to consider the use of various learning rate policies, pairs on the training dataset. Inspired by the "A systematic evaluation of learning rate policies in training CNNs for brain tumor segmentation"[188] we employed different optimizers in the training process. We evaluated their effectiveness based on the performance metric on the test set images, accordingly choosing the most impressive. The following sections report a brief description of the main characteristics of the methods just recently introduced, outlining the motivations that have induced us to employ them.

Stochastic gradient descent : Constant Learning Rate

The constant learning rate is the default learning rate schedule in the SGD optimizer[187], and it provides no reduction in the learning rate during the training phase. It is tricky to choose the correct learning rate, and often, this method does not lead to any convergence resulting in poor performance. .

Momentum is another argument in SGD optimizer which we could tweak to obtain faster convergence. Unlike classical SGD, momentum method helps the parameter vector to build up velocity in any direction with constant gradient descent so as to prevent oscillations.

$$v_t = \gamma v_{t-1} + \eta \nabla_\theta L(\theta)$$
$$\theta = \theta - v_t$$

SGD optimizer also has an argument called nesterov which is set to false by default.

Nesterov momentum is a different version of the momentum method which has stronger theoretical converge guarantees for convex functions.The Momentum method does not take into account direction it is going in, the Nesterov Accelerated Gradient method computes an approximation of the next position of the parameters. The update rule is given in :

$$v_t = \gamma v_{t-1} + \eta \nabla_\theta L(\theta - \gamma v_{t-1})$$
$$\theta = \theta - v_t$$

Adaptive Learning Rate Methods

Adagrad

The Adagrad [187] method adapts the updates to the slope of the error function. The algorithm adapts the learning rate to the parameters, so the size of the updates for each parameter depends on its importance. The Adagrad algorithm gives more significant updates for infrequent parameters and more minor updates for frequent parameters; the update rule is given in the equation. Here Gt is a diagonal matrix containing the sum of squares of past gradient concerning . The main advantage of Adagrad is that one does not need to tune the learning rate manually. However, Adagrad faces problems due to the accumulation of squared gradients in the denominator. Since every additional term is positive, the accumulated sum keeps growing. Therefore the learning rate becomes smaller as training progresses.

$$g_{t,i} = \nabla_\theta L(\theta_i)$$
$$\theta_{t+1,i} = \theta_{t,i} - \frac{\eta}{\sqrt{G_{t,ii} + \epsilon}} \cdot g_{t,i}$$

Adadelta

Adadelta [187] is an extended version of Adagrad which reduces the problem of the decreasing learning rate. It restricts the range of accumulated squared gradients to a certain fixed size.

$$\Delta\theta_t = -\frac{RMS[\Delta\theta]_{t-1}}{RMS[g]_t} g$$
$$\theta_{t+1} = \theta_t + \Delta\theta_t$$

RMSprop

RMSprop [187] is also an adaptive learning rate method that tackles the accumulation of squared gradients in Adagrad. RMSprop divides the learning rate indeed by an exponentially decaying average of squared gradients. It is an unpublished algorithm by G. Hinton.

$$E[g^2]_t = 0.9E[g^2]_{t-1} + 0.1g_t^2$$
$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{E[g^2]_t + \epsilon}} g_t$$

Adam

The Adaptive Moment Estimation (Adam) optimizer [187] also determines an adaptive the learning rate for each parameter. Adadelta and RMSprop store an exponentially decaying average of past squared gradients v_t , but Adam also considers an exponentially decaying average of past gradients m_t . Vectors v_t and mt are estimates of the mean and the uncentered variance of the gradients, respectively, which are biased towards zero.

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t$$
$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2$$

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t} + \epsilon} \hat{m}_t$$

Nadam

Nadam is an acronym for Nesterov and Adam optimizer[187]. As we have seen before, Adam can be considered as a combination of RMSprop and momentum: RMSprop contributes, in fact, the exponentially decaying average of past squared gradients.

, while on the other hand, momentum accounts for the exponentially decaying average of past gradients

. We have also proved that Nesterov accelerated gradient is superior to standard momentum.

Nadam (Nesterov-accelerated Adaptive Moment Estimation) [187] thus combines Adam and NAG. In order to incorporate NAG into Adam, its momentum is modified. . In order to add Nesterov momentum to Adam, we can thus similarly replace the previous momentum vector with the current momentum vector. the Nadam update rule is the following :

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t$$
$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t}$$
$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t} + \epsilon} \hat{m}_t$$

Expanding the second equation with the definitions of m̂ t and m t in turn gives us:

$$\theta_{t+1} = \theta_t - \frac{\eta}{\sqrt{\hat{v}_t} + \epsilon} \left(\frac{\beta_1 m_{t-1}}{1 - \beta_1^t} + \frac{(1 - \beta_1) g_t}{1 - \beta_1^t} \right)$$

Learning rate policies in the proposed method

In the training phase of deep neural networks, it is often valuable to reduce the learning rate as the training progresses. Even for a constant learning rate baseline, it is non-trivial to choose and implement for a DNN. Dynamic learning rates involve multi-step tuning of LR values at various training processes and offer high accuracy and fast convergence. We inspect principal learning rate policies in this idea, adopting predefined learning rate schedules or adaptive learning rate methods. Therefore, based on the dominant papers in the literature, we have selected six different learning rate strategies: Sgd with constant Learning rate, Sgd with Moment, and Nestorov Moment. Adam optimizer has been one of the overall choices in our experiment as it dam is an optimization algorithm that can be employed instead of the classical stochastic gradient descent procedure to update network weights iterative based on training data and in an adaptive way. Moreover, empirical results demonstrate that Adam works well in practice and compares favorably to other stochastic optimization methods. Adam can be looked at as a combination of Stochastic Gradient Descent with Momentum amìnd RMS prop. On the other hand, RMSprop is defined as a refinement of the Adagrad technique or a variant of Adadelta. Therefore it was employed in our algorithm than the latter. Insofar, RMSprop and Adam are very similar algorithms that do well in similar circumstances. [187] its bias correction helps Adam slightly outperform RMSprop towards the end of optimization as gradients become sparser. In practice, Adam is currently recommended as the algorithm to use and often works slightly better than RMSProp. However, it is usual also worth trying SGD+Nesterov momentum as an alternative. While momentum accelerates our search in the direction of minima, RMSProp impedes our search in the direction of oscillations. Finally, considering the importance of the Nesterov moment, we also examine the feasibility of the Nadam optimizer in our deep learning model. Adam and Nadam are particularly being adapted for benchmarks in deep learning papers. Good default settings for the tested machine learning problems are β_1=0.9, β2=0.999 and ϵ=10e-8 Momentum is set to a value greater than 0.0 and less than one for the learning rate schedules, where shared values such as 0.9 and 0.99 are used in practice. However, the relevant literature suggests the value of momentum is usually set to 0.9 in this concept, we evaluated the different learning rate policies on the test dataset, considering the pre-eminent the one with higher performance.

4.3.3 Number of Epochs

A significant challenge when training a deep learning model is deciding how many epochs to run. Too few might not lead to model convergence, while too many steps could lead to overfitting. The number of epochs is the measure of complete passes through the training dataset. The model should be trained for an optimal number to mitigate overfitting and increase the neural network's generalization capacity. It is as high as possible to ensure the convergence of the model but, on the other side, end training based on the error rates. Indeed, one should set this parameter as high as possible and terminate the training when the validation error increases. The number of epochs can be selected to an integer value between one and infinity. However, the optimal value can be determined only heuristically, training the model for different epochs, denoting the state of convergence of the training loss and the error or the increment or the validation loss. Hence we have trained the proposed network, varying such parameters and choosing the optimal configuration experimentally.

4.3.4 Batch Size

Batch size is a hyperparameter used in machine learning and refers to the number of training examples utilized in one iteration or stepper epoch; it defines the number of samples propagated through the networks. A batch of samples is propagated through the model and then backward propagated to estimate gradients for every instance in every single training step. The gradients of all samples then are averaged. The use of Mini-batches and, therefore, batch sizes smaller than the number of images that make up the dataset brings considerable advantages. It requires less memory. Since one trains the network using fewer samples, the overall training procedure requires less memory. It is essential if it is unfeasible to fit the whole dataset in the machine's memory- Furthermore, typically, networks train faster with mini-batches because the weights are updated after each propagation.

Batch size has a critical and impressive impact on the convergence of the training process and the resulting accuracy of the trained model. Typically, there is an optimal measure or range of values for batch size for every neural network and dataset.

Large batch sizes may cause wrong generalization (or even get stuck in a local minimum), which underlines overfitting and means that the neural network will perform poorly on samples outside the training set. On the other hand, small-batch sizes may lead to slow convergence of the learning algorithm. Consequently, for the proposed network, a batch size parameter is established respecting the constraints of the GPU memory, and the considerations reported previously. We have chosen a bath size parameter of 40 as an acceptable compromise.

4.3.5 Steps per Epoch

The steps per epoch or number of batch iterations before a training epoch is considered finished. It may be helpful for a vast data set or generate random data augmentations. Traditionally, the steps per epoch are calculated as the quotient of training dataset length and batch size since this will use all of the data points, one batch size worth. On these grounds in our application, considering the dataset and batch size, we appraised thirty steps per epochs, a practical choice to cover the whole dataset in the standard case and one hundred in Data augmentation.

4.4 Experimennts

4.4.1 Introduction

In this section, we describe the main experiments conducted in exploring the applicability of the proposed methods. Therefore we describe the optimizations to set the parameters of training of the neural networks in question but, above all, analysis to verify and to deepen the methodologies of estimation of the apriori information and how they are introduced in the neural networks. Moreover before considering implementation and training, it is necessary to determine what is the number of components needed in the denoising version of SP-net and Sp-net Kernel and especially what is the most appropriate kernel for this application. These processes require further experiments based on the process of image reconstruction through the PCA and Kernel PCA based on different numbers of components. With these premises, all the experimentations have been effected in the same conditions in order to be able to report a correct comparison between the proposed methods (Kernel SP-net and SP-net) mentioned above, considering the neural network U-net as reference standard.

4.4.2 Kernel PCA & PCA Experiments

In these experiments, the applicability and reliability of Kernel PCA and PCA in the process of projection and rear projection of binary mask images belonging to different sets are evaluated. This process is intended as reconstruction and is used to evaluate the most suitable number of components in the denoising applications of the proposed networks and especially which kernel is better in this operation. In this regard, image reconstruction processes will be performed using the above mentioned kernels for Kernel PCA and different numbers of components in the reconstruction process. This procedure will be performed under the same conditions for the linear PCA.

4.4.3 Training Optimization Experiments

In the first instance, one might wonder about the hyperparameters to be used to train the model, both triggering or not data augmentation, This request is essential to ensure that the proposed method does not fail due to an underfitting or overfitting, which would lead to a deviation of the results. Indeed, the first experiments are aimed at optimizing the understanding of what are the best parameters. As described above, these experiments will see the deployment of various techniques to adapt or not the value of the learning rate to train the model. In addition, various training has been performed to understand which are the best number of epochs and how Early stopping can be optimally applied. The value of batch size and steps for epochs are established depending on the capacity of the GPU Memory and the size of the dataset . The results produced led us to train the network using the following as a learning rate policy , Nadam technique employing a starting learning rate=0,001. For the number of epochs has been established a number equal to 300 because as observable by the loss and validation loss is suitable to avoid cases of underfitting and overfitting.Batch size and steps per epoch are estabilished according dataset features . Earlystopping is implemented considering a patience maesure of 30 epochs as suggested by reference paper[190] .

4.4.4 SP-Net Experiments

These experiments are intended to capture the drawbacks and advantages of the proposed method "Shape predictor network" . On this basis, SP-net is explored both with and without implementation of Data Augmentation. The latter concerns to perceive how geometrical transformation of the original masks and therefore of the SDFs can lead or not performance improvements, assuming that these transformations lead to a wider distribution of pixels constituting the masks transformed into signed distance functions, hoping that they can then cover the misalignment present within the images of AIforcovid dataset. Moreover, to understand if a reduction of the principal components can lead to advances effectively, the following reduced version is examined.

4.4.5 Kernel SP-Net Experiments

In this case, the experiments performed have the role of investigating an extension of the proposed method, trying to perceive how PCA kernel may or may not correct the defects expressed in the previous method. Therefore, this application has been considered and used in the same SP-Net conditions by exploring here a reduced version and not of Kernel SP-Net and the application of Data Augmentation. Regarding the kernel used, this was chosen based on the reconstruction errors in the set of validation and testing.

4.4.6 U-net Experiments

In latter phase, U-net being considered as the state of the art method, the related experiments were carried out with the intent to provide a comparison of the proposed methods and to understand the importance of the results especially in relation to the methods usually used for the segmentation of medical images. n this regard Unet has been implemented with Data Augmentation generating the same number of images expected in Sp-net and Kernel SP-net through the use of the same geometric transformations.

4.4.7 Comparison of Methods

In conclusion, the results obtained from these experiments were compared in order to understand the actual performance . Therefore a comparison on the application of the Data Augmentation has been carried out for what concerns the above mentioned neural networks and the application of the latter with the original dataset. While in the version of the reduced components' number, the results of SP-net of its non-linear extension Kernel SP-net are reported, compared in turn with the performance of the U-net network.

CHAPTER 5

Discussion & Results

This chapter is the most prominent of the thesis as it reports and describes the main findings following the application of the methods described above. These results are also discussed in order to expose and perceive the main advantages and disadvantages in the segmentation of the images that are part of the test dataset compared to the standard method assumed as a reference, such as Unet. Therefore, in addition to the graphs showing performance measurement indices, a qualitative analysis was also carried out to gain better the detriments and benefits realized through such methodologies.

5.1 Introduction

The results are the conclusive and biting part of this work. In this regard, they have been organized to follow a logical thread that leads to the exploration and application of the methods previously mentioned. Consequently, we have reported all the results concerning the training of the neural networks used. Since architecturally, SP-net and Kernel SPnet result to be identical. The verdict of the results for the training optimization is used in the same way for both neural networks, while U-net will be used under the same conditions to provide a comparison. In the rest of the chapter, the results regarding SP-net have been described in the first instance, both using the images of the original training set and those artificially generated through La Data Augmentation. In addition, the results obtained following the use of fewer components regarding the projection of images predicted by the network have been exposed to observe whether this application can bring improvements. Similarly, these results are reported in parallel to demonstrate whether effectively a nonlinear version of the method can behave better in capturing and perceiving the apriori information obtained nonlinearly through the predefinition of a kernel function. The procedure used in the choice of such a function has also been described. Finally, as anticipated, the results of Unet are expressed as a method of references and so with data augmentation. Wrapping up the resultant of this thesis is represented only from the dice coefficient and jaccard index. They also come crowned from a qualitative analysis picturing the best and worst cases of the respective methods.

5.2 Principal Component Analysis Results

The results shown here expose the criteria and findings retrievied after the exploration of principal component analysis and its use within the proposed neural network. In this regard, the cumulative explained variance is evaluated to assess the optimal number of components to establish a denoising end reduced version of SP-NET.

5.3 PCA: Cumulative explained variance

A vital part of using PCA is the ability to estimate how many components are needed to describe the data. it can be determined by looking at the cumulative explained variance ratio as a function of the number of components:

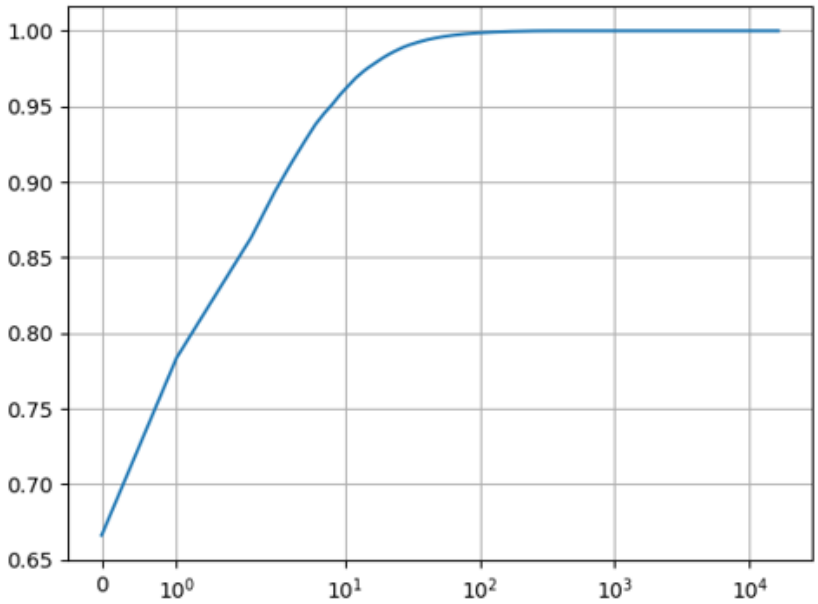


Figure 5.1: Cumulative explained variance by PCA

5.3.1 PCA: Choosing Number of Components

As illustrated in the previous paragraphs, we have sampled some numbers of components of the cumulative explained variance, starting from a minimum threshold of 100 components as the minimum number of components to obtain an adequate reconstruction of the masks and representing the 98% explained variance ad used as criteria of choice in the reference paper[152] of "Shape predictor Network". . The other numbers in the esperiment have been considered increasing the order of magnitude(therefore sampling 100, 1000, 10000, 15000 , total components).

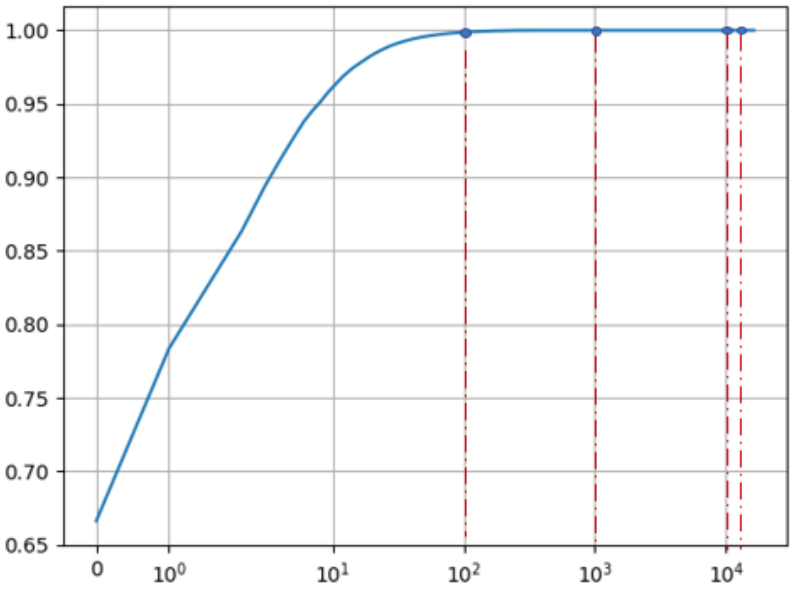
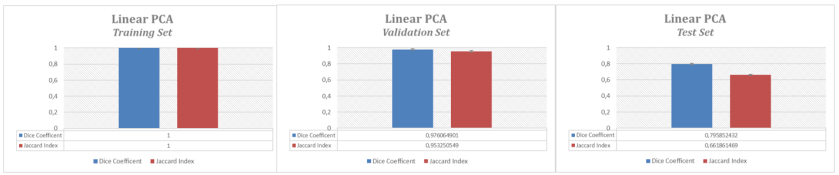


Figure 5.2: Sampled cumulative explained variance by PCA

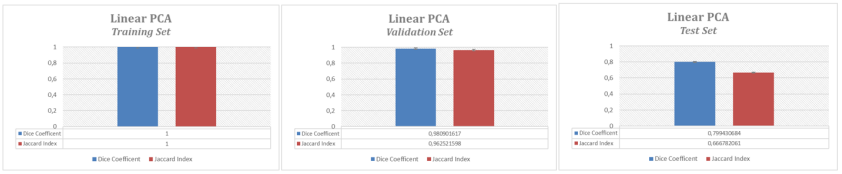
5.3.2 PCA : Reconstruction Process

The reconstruction process of masks and segmentations was performed and evaluated with the number of components sampled and validated on the training set, validation set and finally on the test set, exploiting the reconstruction of 10 images for each set and calculating the respective dice coefficent and jaccard index of the reconstructed lung curves.

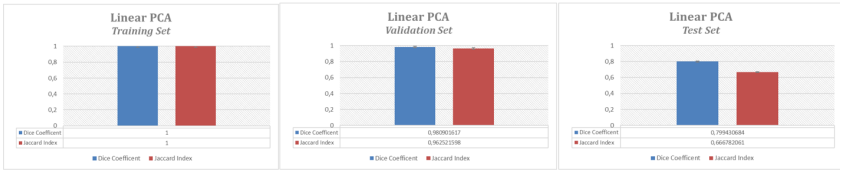
Total Componenents



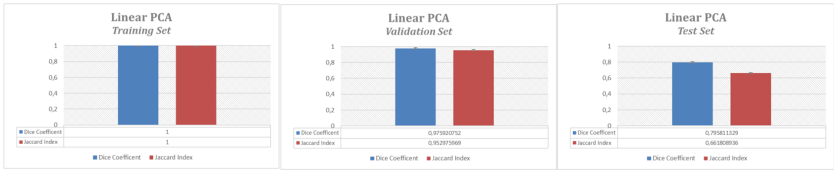
15000 Componenents



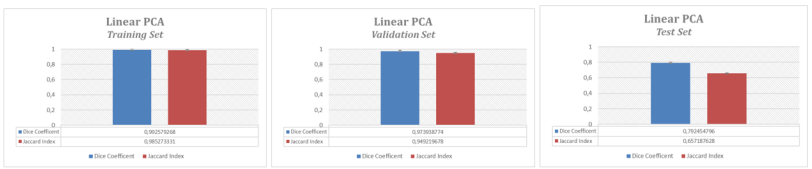
10000 Componenents



1000 Componenents



100 Componenents



Qualitative Analysis

These results express a qualitative analysis of the first images from the training set, validation set and test set reconstructed with the chosen component numbers, respectively.

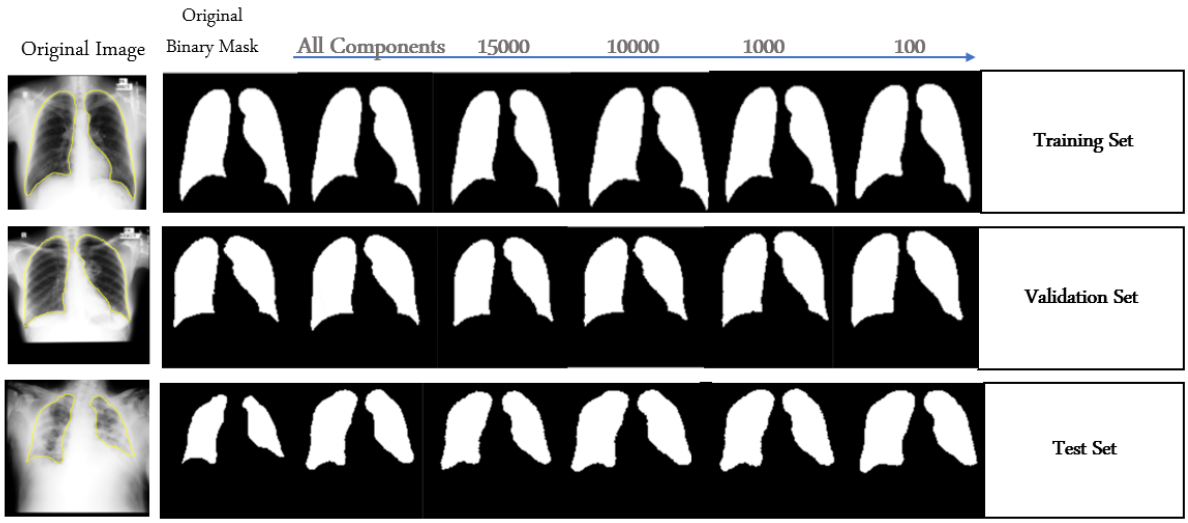


Figure 5.3: Qualitative analysis by PCA

5.4 Kernel Principal Analysis Results

As already shown the application of the Kernel PCA requires a correct choice of the Kernel and of the related hyperparameters corresponding to the training dataset on which it is used. As expressed in the previous chapters, the process to decide the best kernel function has been established and decided according to the process of reconstruction of the masks coming from the training set, validation set and Test set. The kernels that have been investigated are respectively kernel on cosine similarity , radial basis function kernel , laplacian kernel and polynomial kernel. For the latter three since their expression considers the use of hyperparameter , optimization process based on binary mask reconstruction was adopted to choose correct parameters.

5.4.1 Kernel PCA: Radial Basis Function Kernel

Since the results and the application of RBF kernels resulted to be very dependent on the γ parameter, three different values have been chosen respectively in the reconstruction process, two of them randomly taken (γ=0.00001, γ=1000) and one related to the average variance present in the traning dataset (γ = 0.010705), as reported in the paper [189]. From these findings we see that the application and explained variance of Kernel PCA depends on the chosen γ parameter

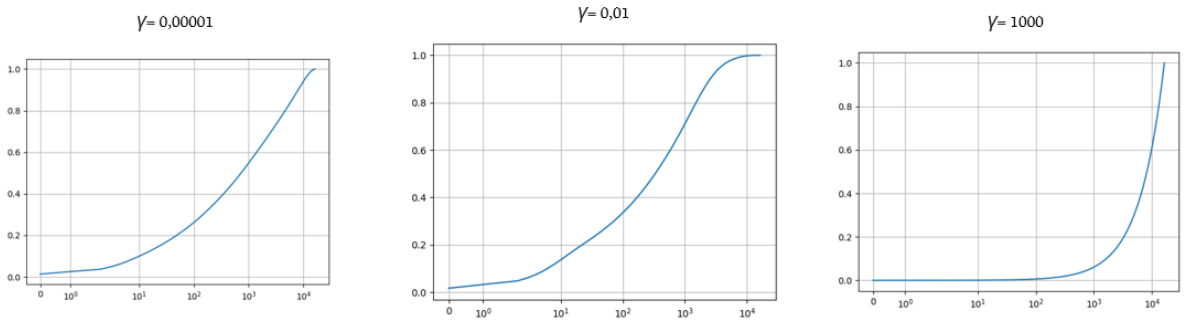
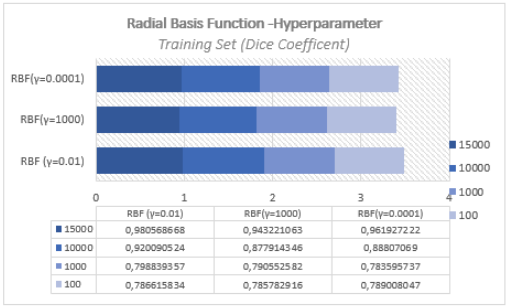


Figure 5.4: Cumulative Explained Variance by RBF Kernel (dependent on γ)

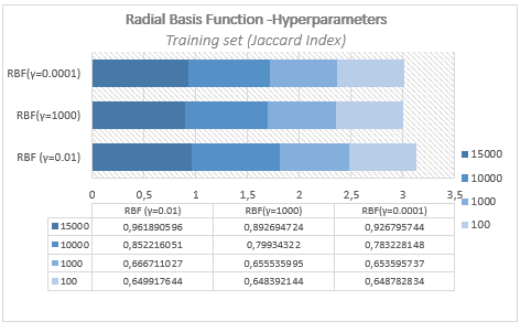
5.4 Kernel Principal Analysis Results

Also analyzing the results of the reconstruction processes on the training set , valdation set and test set , it is evident that the gamma parameter must be proportional to the variance of the investigated dataset.

Training Set :

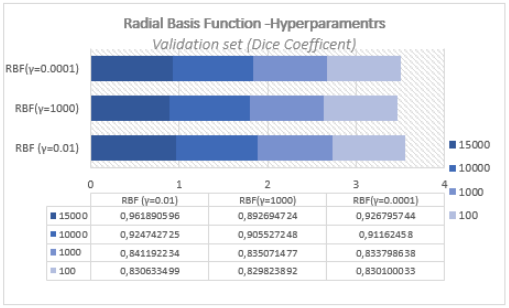


	RBF (γ=0.01)	RBF(γ=1000)	RBF(γ=0.0001)
15000	0,980568668	0,943221063	0,961927222
10000	0,920090524	0,877914346	0,88807069
1000	0,798839357	0,790552582	0,783595737
100	0,786615834	0,785782916	0,789008047

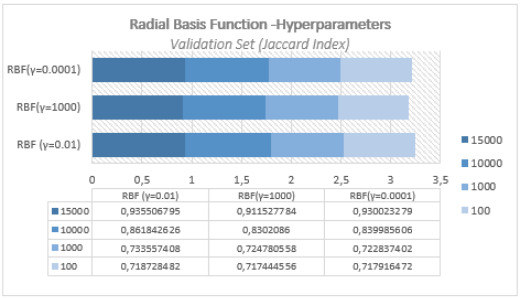


	RBF (γ=0.01)	RBF(γ=1000)	RBF(γ=0.0001)
15000	0,961890596	0,892694724	0,926795744
10000	0,852216051	0,79934322	0,783228148
1000	0,666711027	0,655535995	0,653595737
100	0,649917644	0,648392144	0,648782834

Validation Set :

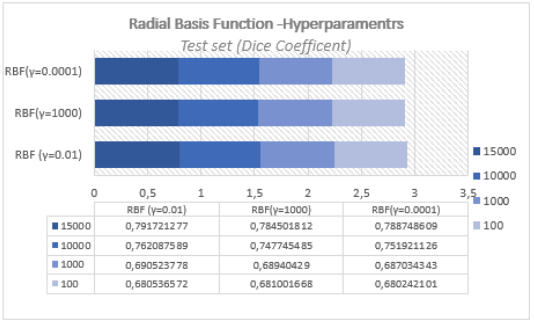


	RBF (γ=0.01)	RBF(γ=1000)	RBF(γ=0.0001)
15000	0,961890596	0,892694724	0,926795744
10000	0,924742725	0,905527248	0,91162458
1000	0,841192234	0,835071477	0,833798638
100	0,830633499	0,829823892	0,830100033

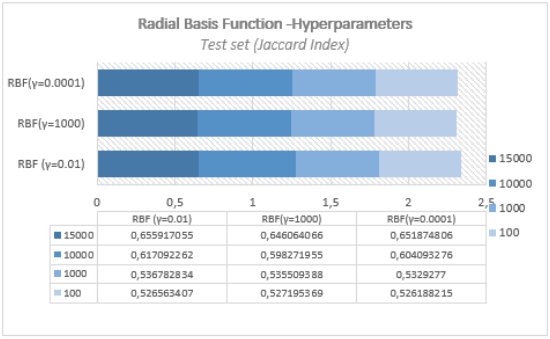


	RBF (γ=0.01)	RBF(γ=1000)	RBF(γ=0.0001)
15000	0,935506795	0,911527784	0,930023279
10000	0,861842626	0,8302086	0,839985606
1000	0,733557408	0,724780558	0,722837402
100	0,718728482	0,717444556	0,717916472

Test Set :



	RBF (γ=0.01)	RBF(γ=1000)	RBF(γ=0.0001)
15000	0,791721277	0,784501812	0,788748609
10000	0,762087589	0,747745485	0,751921126
1000	0,690523778	0,68940429	0,68703443
100	0,680536572	0,681001668	0,680242101

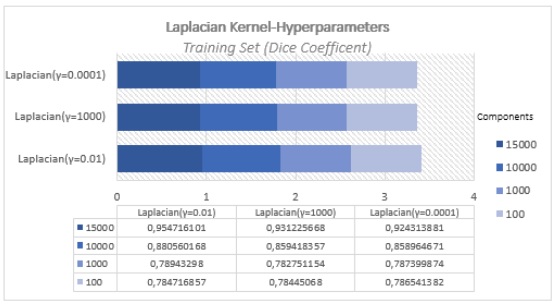


	RBF (γ=0.01)	RBF(γ=1000)	RBF(γ=0.0001)
15000	0,655917055	0,646064066	0,651874806
10000	0,617092262	0,598271955	0,604093276
1000	0,536782834	0,535509388	0,5329277
100	0,526563407	0,527195369	0,526188215

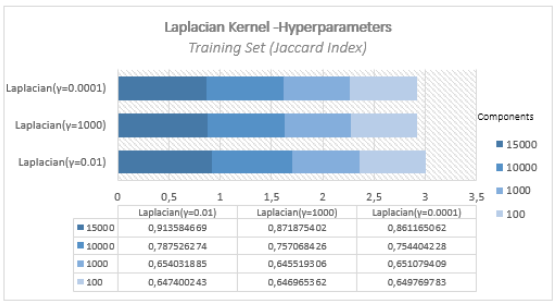
5.4.2 Kernel PCA: Laplacian Kernel

The laplacian kernel is considered as a variant of the Radial Basis Function kernel , in this perspective were used the same criteria and the same parameters gamma, in fact this kernel was verified in the reconstruction of 10 images of the training set, validation set and test according to the jaccard and Dice coefficient with the number of components sampled.

Training Set:

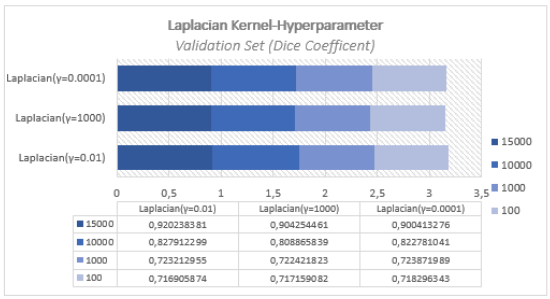


	Laplacian(γ=0.01)	Laplacian(γ=1000)	Laplacian(γ=0.0001)
15000	0,954716101	0,931225668	0,924313881
10000	0,880560168	0,859418357	0,858964671
1000	0,7894329 8	0,782751154	0,787399874
100	0,784716857	0,78445068	0,786541382

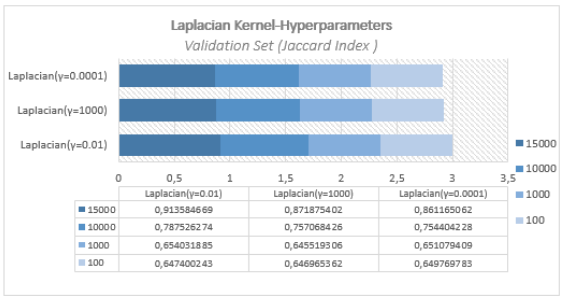


	Laplacian(γ=0.01)	Laplacian(γ=1000)	Laplacian(γ=0.0001)
15000	0,913584669	0,871875402	0,861165062
10000	0,787526274	0,757068426	0,754404228
1000	0,654031885	0,645519306	0,651079409
100	0,647400243	0,646965362	0,649769783

Validation Set:

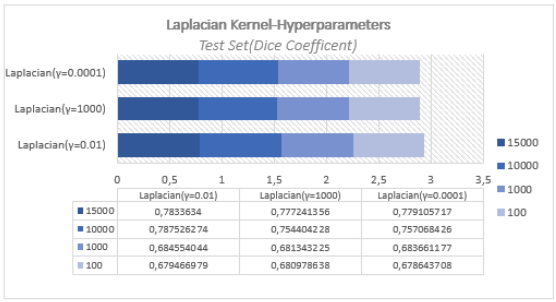


	Laplacian(γ=0.01)	Laplacian(γ=1000)	Laplacian(γ=0.0001)
15000	0,920238381	0,904254461	0,900413276
10000	0,827912299	0,80886539	0,822781041
1000	0,723212955	0,722421823	0,723871989
100	0,716905874	0,717159082	0,718296343

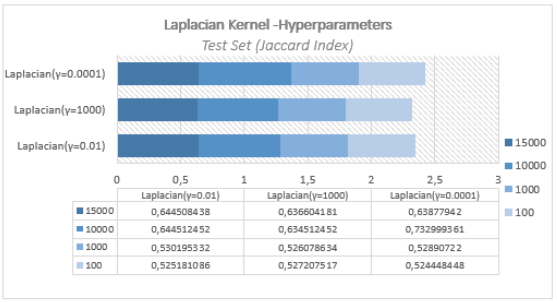


	Laplacian(γ=0.01)	Laplacian(γ=1000)	Laplacian(γ=0.0001)
15000	0,913584669	0,871875402	0,861165062
10000	0,787526274	0,757068426	0,754404228
1000	0,654031885	0,645519306	0,651079409
100	0,647400243	0,646965362	0,649769783

Test Set:



	Laplacian(γ=0.01)	Laplacian(γ=1000)	Laplacian(γ=0.0001)
15000	0,7833634	0,777241356	0,779105717
10000	0,787526274	0,754404228	0,757068426
1000	0,684554044	0,681343225	0,683661177
100	0,679466979	0,680978638	0,678643708

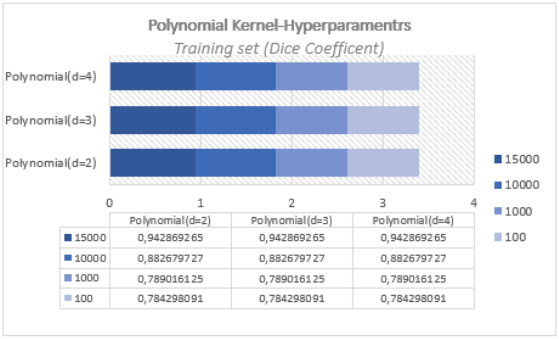


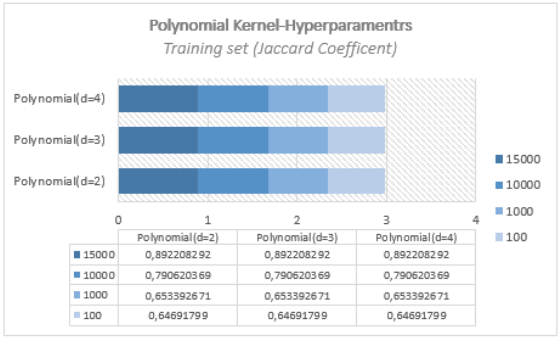
	Laplacian(γ=0.01)	Laplacian(γ=1000)	Laplacian(γ=0.0001)
15000	0,644508438	0,636604181	0,63877942
10000	0,644512452	0,634512452	0,732999361
1000	0,530195332	0,526078634	0,52890722
100	0,525181086	0,527207517	0,524448448

5.4.3 Kernel PCA: Polynomial Kernel

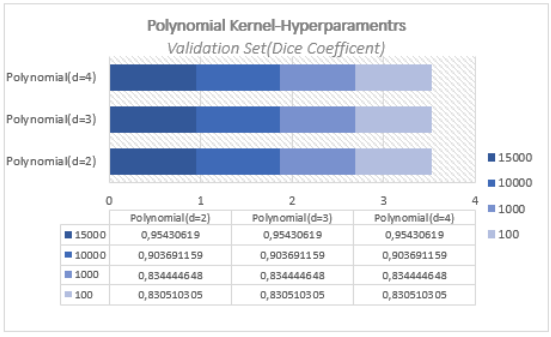
The polynomial kernel was adopted using different parameters, considering as second, third and fourth degree polynomial

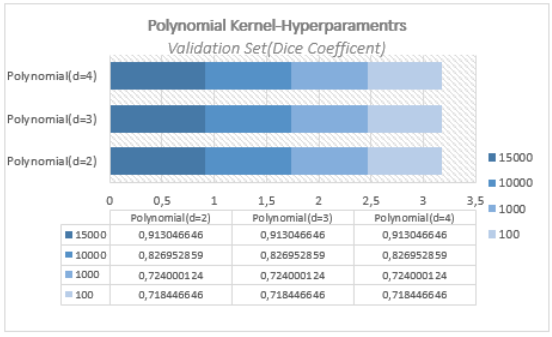
Training Set:



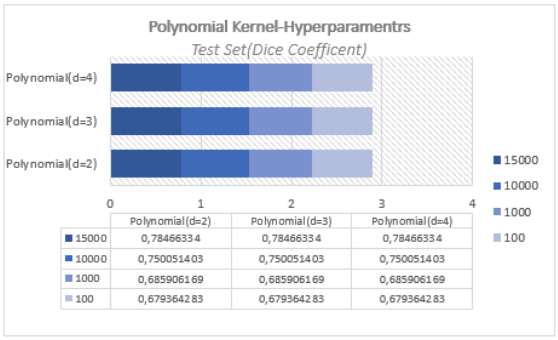


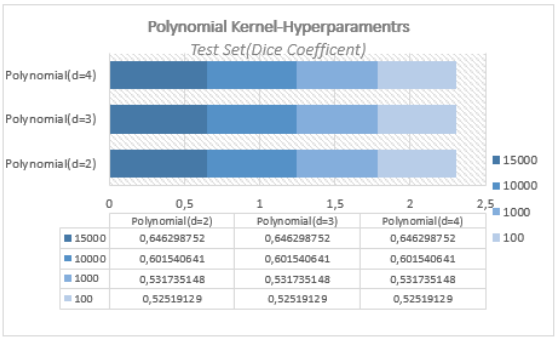
Validation Set:





Test Set:

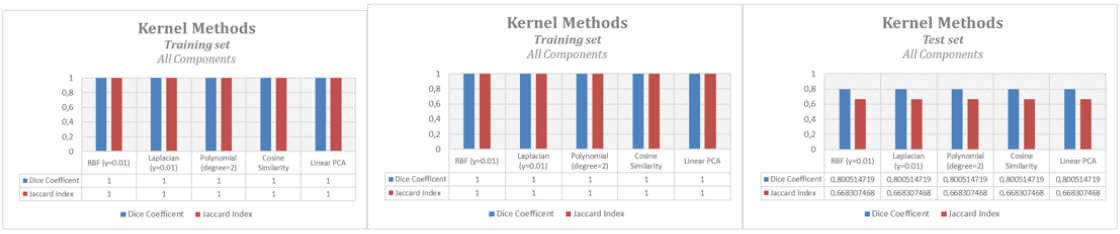




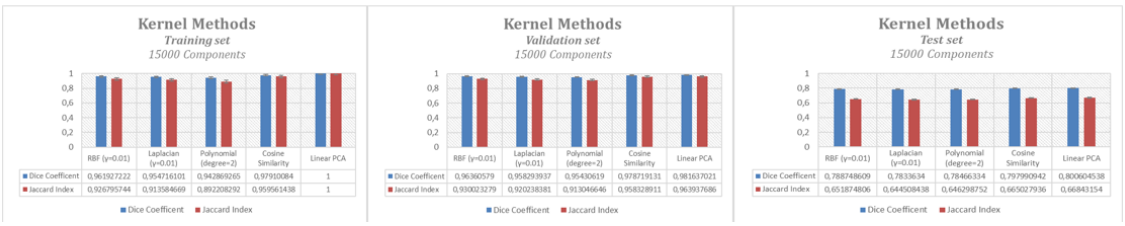
5.4.4 Comparison of Kernel Methods

Assigned the values of the parameters of the laplacian and RBF Kernel and establishing, as also from the reference paper, a degree equal to two for the polynomial kernel, these can be compared with both the cosine similarity kernel and linear PCA.

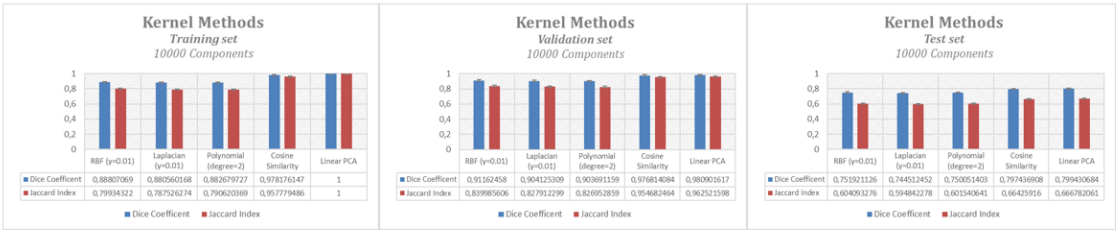
All Components



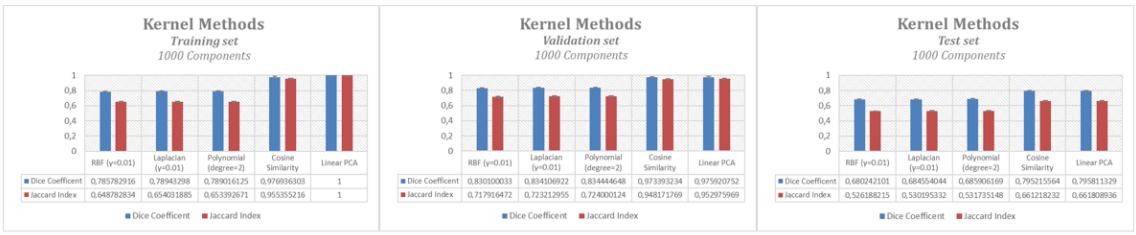
15000 Components



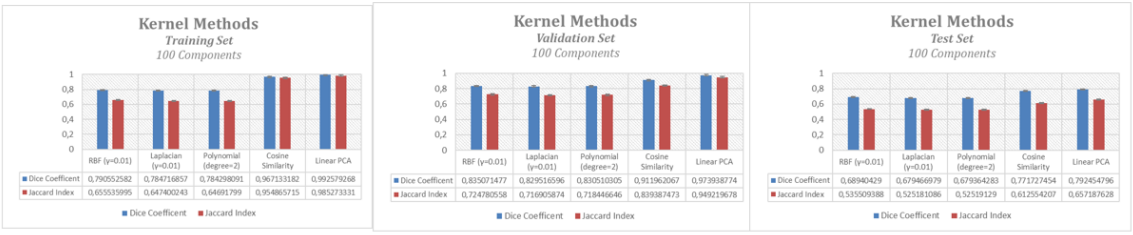
10000 Components



1000 Components



100 Components



5.4.5 Kernel Methods: Qualitative Analysis

This section demonstrates how even qualitatively the kernel based on cosine similarity is more robust and efficient in the reduction of components assuming results similar to the Linear PCA and for this reason it can be used as a Kernel function in the non-linear extension of SP-net.

Qualitative Analysis: Training set

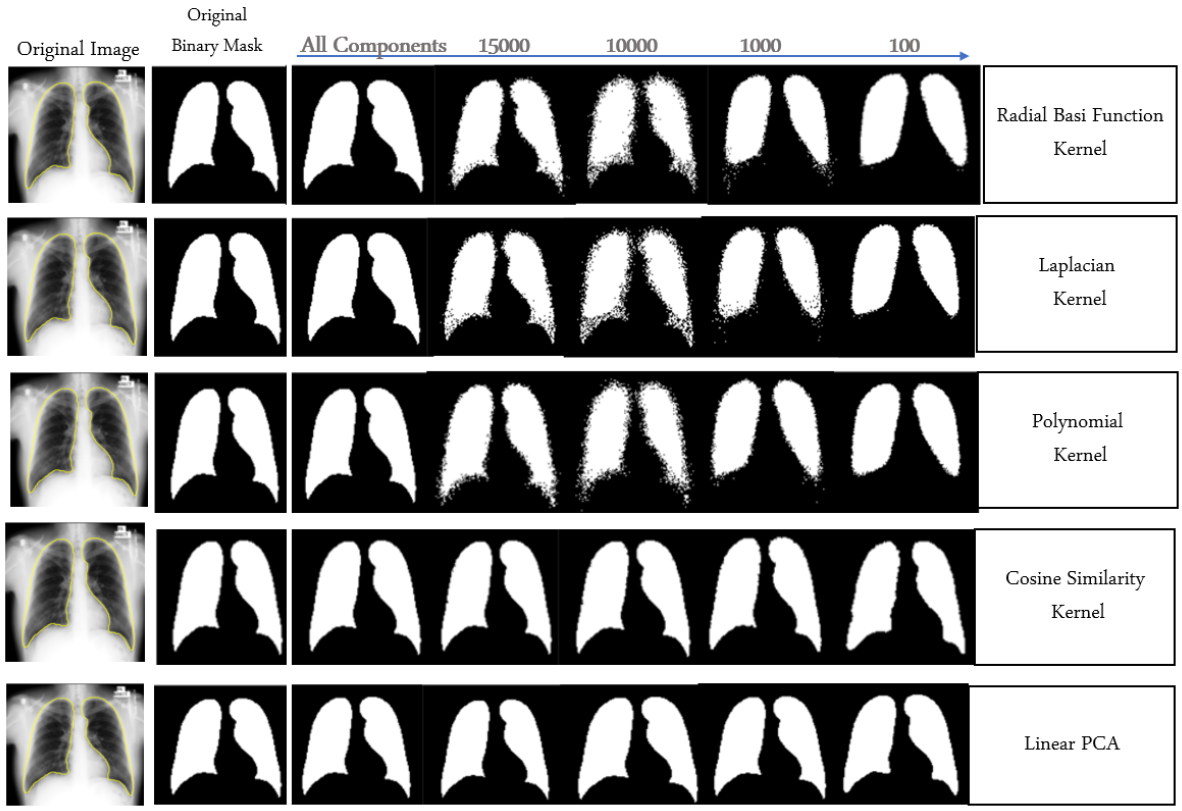


Figure 5.5: Qualitative Analysis : Training set

Qualitative Analysis: Validation set

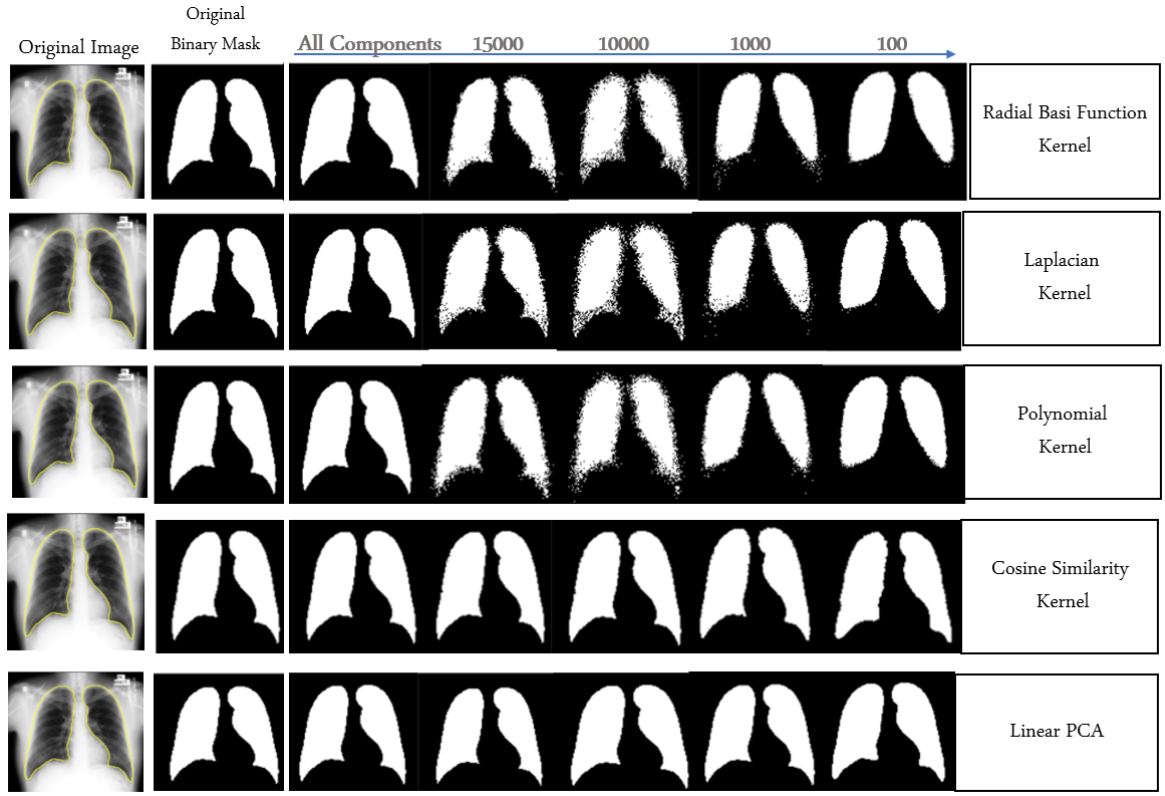


Figure 5.6: Qualitative Analysis : Validation set

Qualitative Analysis:Test set

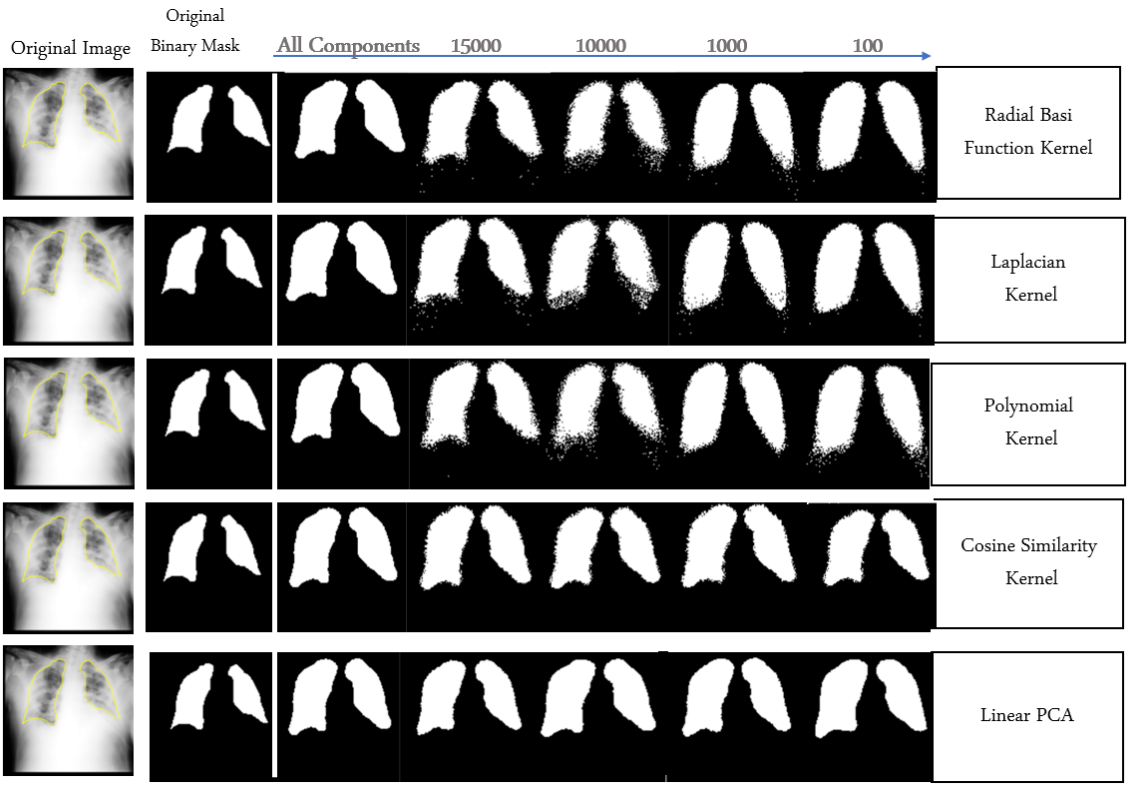


Figure 5.7: Qualitative Analysis : Test Set

Kernel on Cosine similarity for kernel choice

From a first visual analysis and from the reconstruction of the images it is perceived as the kernel based on cosine similarity is very robust especially in the reconstruction of the masks with smaller and smaller number of components. In fact, it can be observed how the other kernels lose the reconstruction of the initial lung shape below a certain number of components. Probably this reason is deduced from the concept that the cosine similarity through the redefinition in the inner product applies operations similar to those computed by Linear PCA on the original SDFs, although in reality it is a non-linear extension.Therefore, given this evidence and also on the basis of the cumulative variance explained from this application, one understands how this kernel is the most suitable in our algorithm. Therefore, given this evidence and also on the basis of the cumulative variance from this application, we understand how this kernel is the most suitable in our algorithm. Therefore it will also be used in the same way as Linear PCA in the denoising activity of the proposed methods.

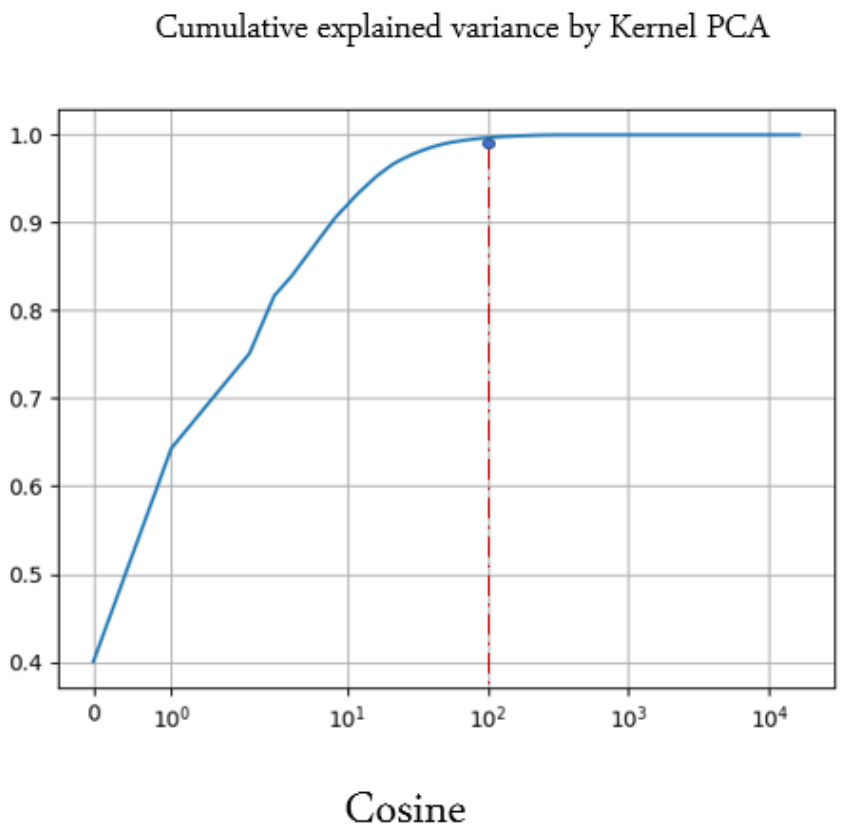


Figure 5.8: Cumulative Explained Variance by Kernel PCA on Cosine similarity

5.5 Training Optimization Results

5.5.1 Learning Rate Optimization

The suitable initial learning rate and the related Learning rate policy has been established considering six different optimization strategies, both adaptive and scheduler. In the first instance for both of these techniques some Learning rate tests were performed in order to obtain a confidence interval of values for which to test the proposed method .

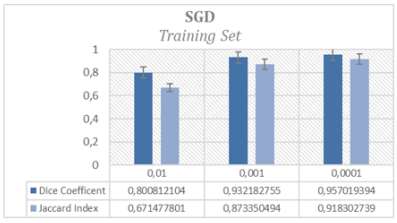
Systematic evaluation of learning rate policies

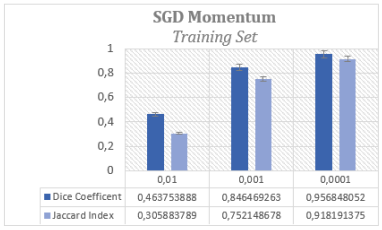
As per the refereed paper[188] , the authors establish and report on several strategies to train a convolutional U-net and carry out an in-depth study on which optimizer is the best to achieve loss convergence. Similarly, once we established the confidence interval of the learning rates, we implemented the following strategies on the proposed neural network .

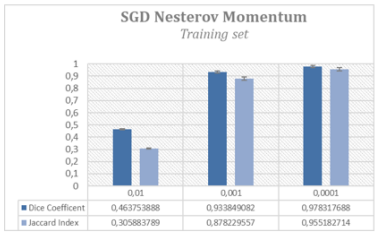
5.5 Training Optimization Results

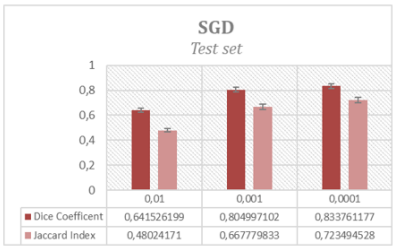
Scheduler leaning rate policies

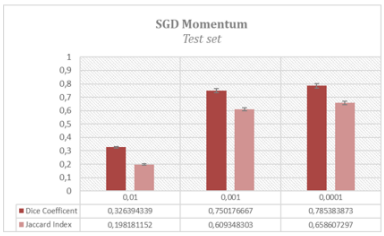
This section represents the related results of Dice coefficent and Jaccard Index reported on the training set and test set by evaluating the following learning rate policies: Stochastic gradient descent , Stochastic gradient descent with momentum, Stochastic gradient descent with Nesterov momentum

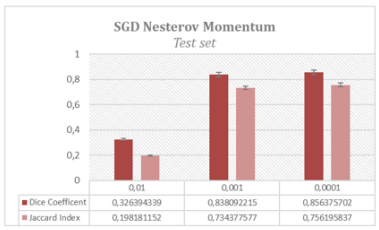






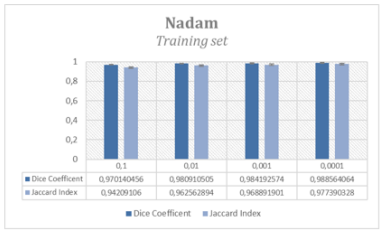


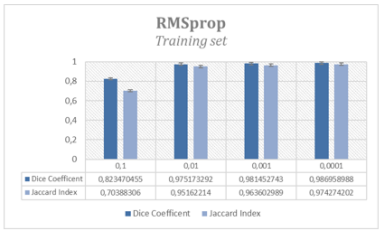


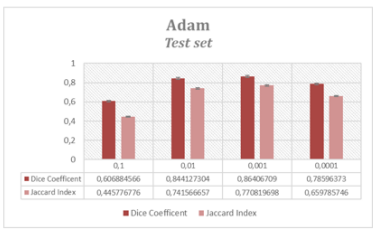


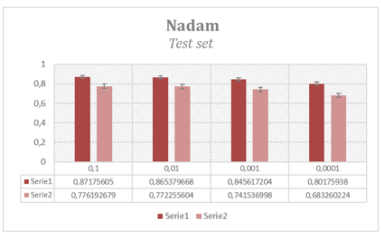
Adaptive leaning rate policies

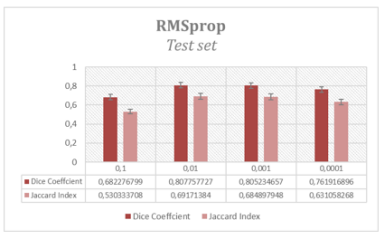
On the other hand in this case we represent the related results of Dice Coefficent and Jaccard Index reported on the training set and test set by evaluating the adaptive learning rate policies: Adam , Nadam and RMSprop.





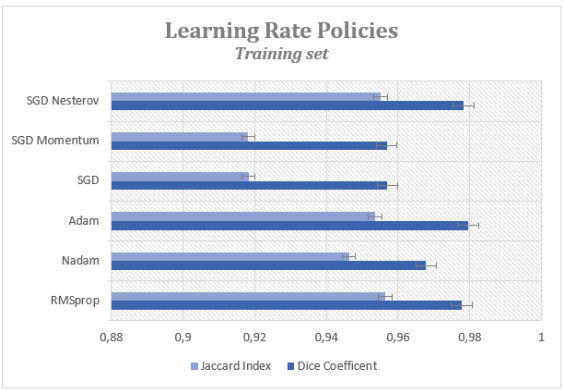


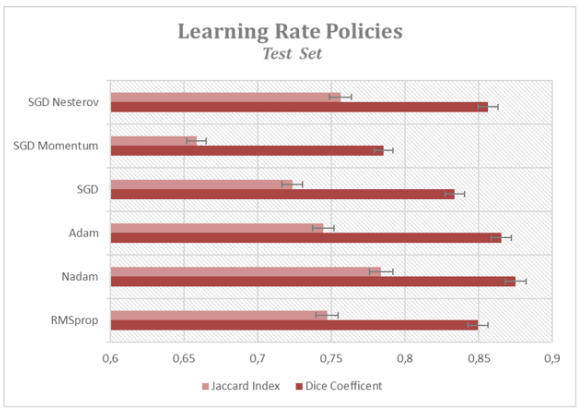




Comparison of learning rate policies

the following results show an efficient comparison between the various strategies estimated on the training set and test, demonstrating like the Nadam learning rate policy is the more adapted in the training of the net, since it previews like also in the literature a greater robustness in the falling in the local minimums of the loss function regarding the others. Assessing the results, for RMSprop, Stochastic gradient descent and Stochastic gradient descent with Mommemntum and Nesterov Momemntum the congenial intial learning rate is 0.0001 , while for Nadam and Adam the suitable learning rate is 0.001.



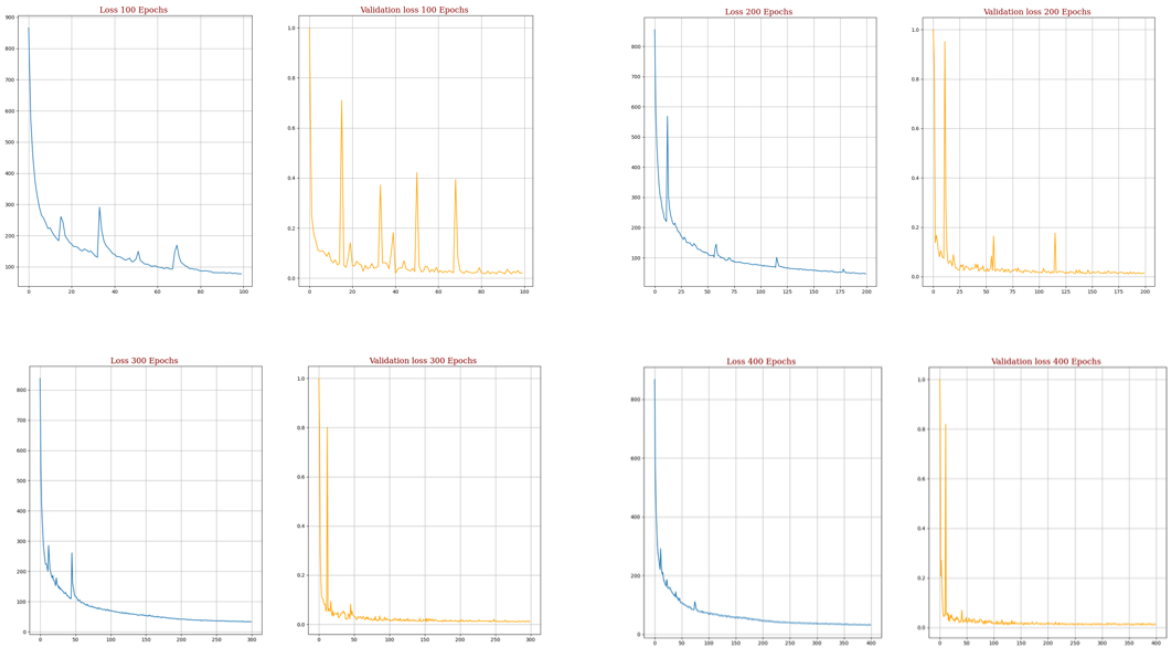


5.5.2 Number of Epochs

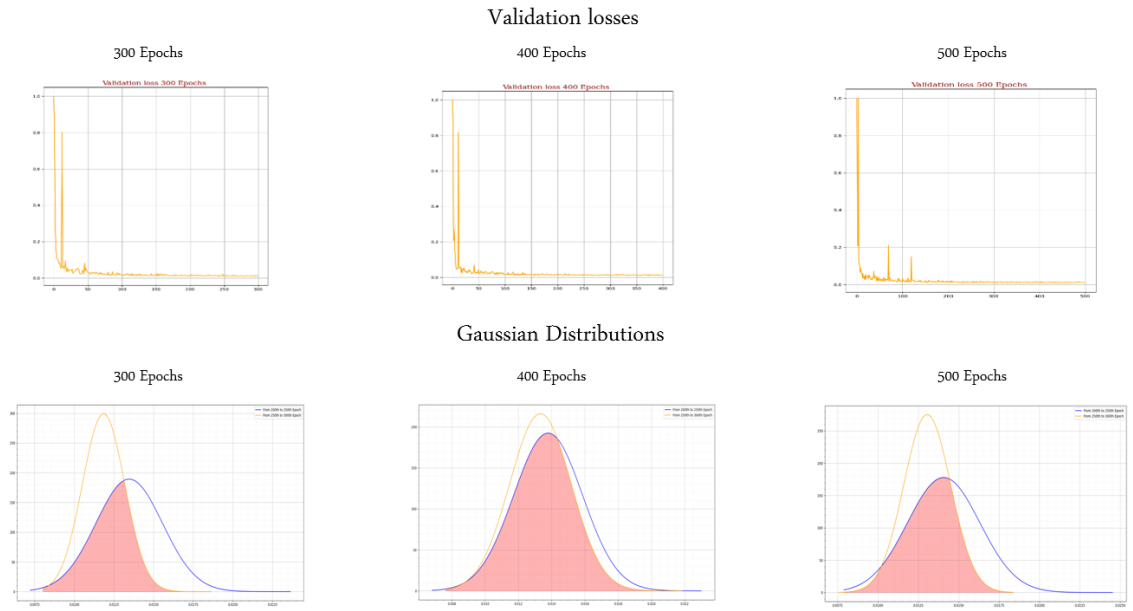
Assumed the initial value of learning rate the appropriate strategy of optimization (Nadam) , several training of the proposed networks were conducted , to establish the correct number of epochs. In particular both Kernel SP-net and SP-net , have been from trained for 100 , 200 , 300 , 400 , 500 epochs.On the basis of which is consulted the validation loss as it allows to avoid the overfitting or underfitting of the network. As can be seen from the results of the training, from the 200th epoch, there are no considerable oscillations in the validation loss, in fact considering the interval from 200 to 300 epochs, the normal distribution of values between the 200th and 250th epochs overlaps with the distribution of values of the remaining epochs. The gaussian distributions are take considering a validation loss of SP-net ,kernel SP-net and Unet training with 300 epochs , 400 epochs , 500 epochs.

5.5.3 SP-net : Number of Epochs

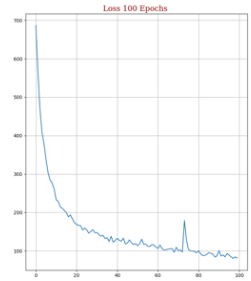
SP-net without Data Augmentation : Lass and validation loss for epochs

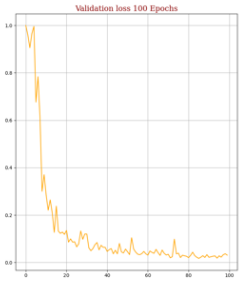


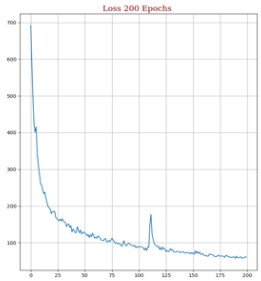
SP-net without Data Augementation : Validation losses and Normal distributions

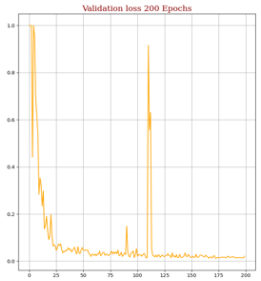


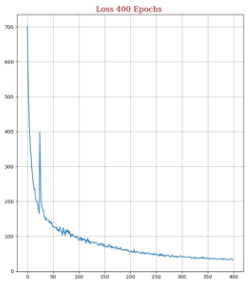
SP-net with Data Augmentation : Lass and validation loss for epochs

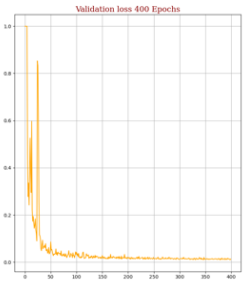


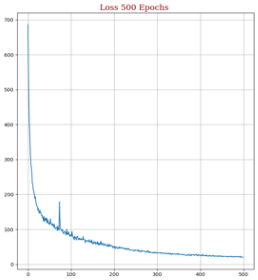


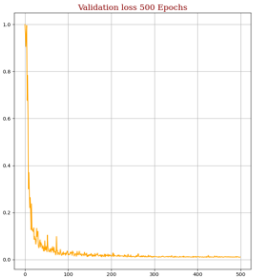








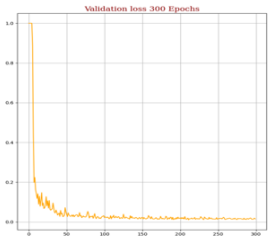




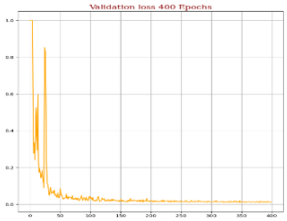
SP-net without Data Augmentation : Validation losses and Normal distributions

Validation losses

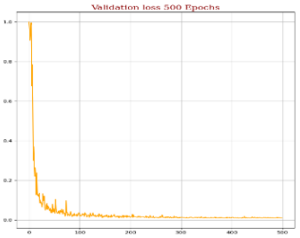
300 Epochs



400 Epochs



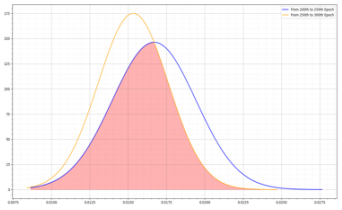
500 Epochs



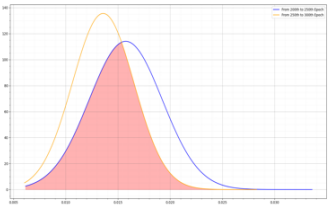
Gaussian Distributions

300 Epochs

400 Epochs

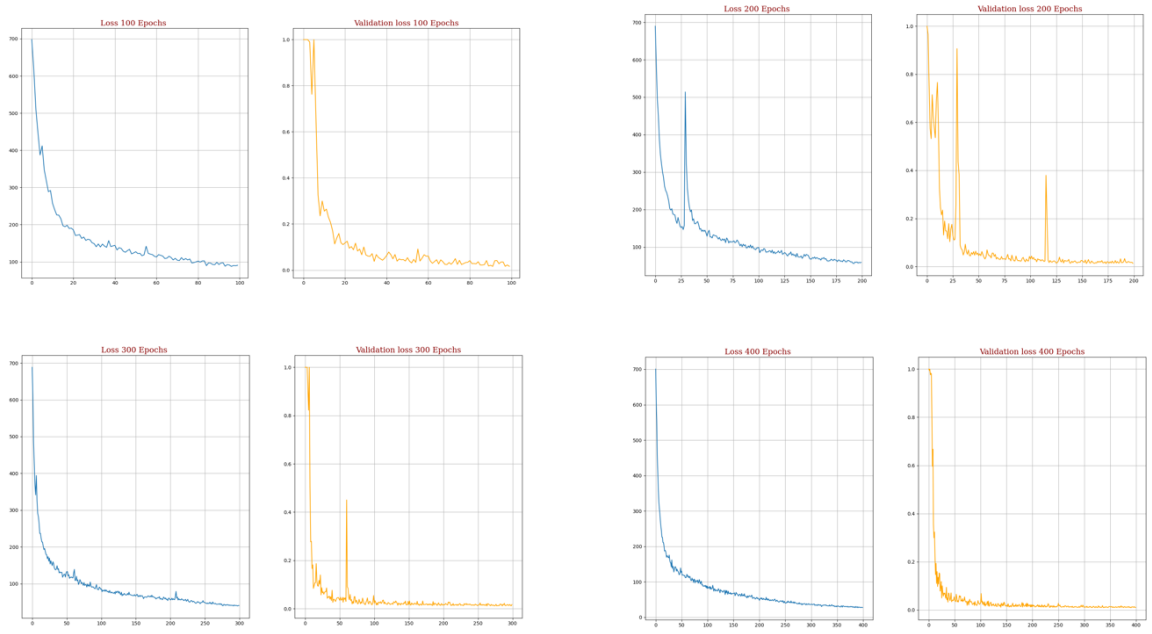


500 Epochs

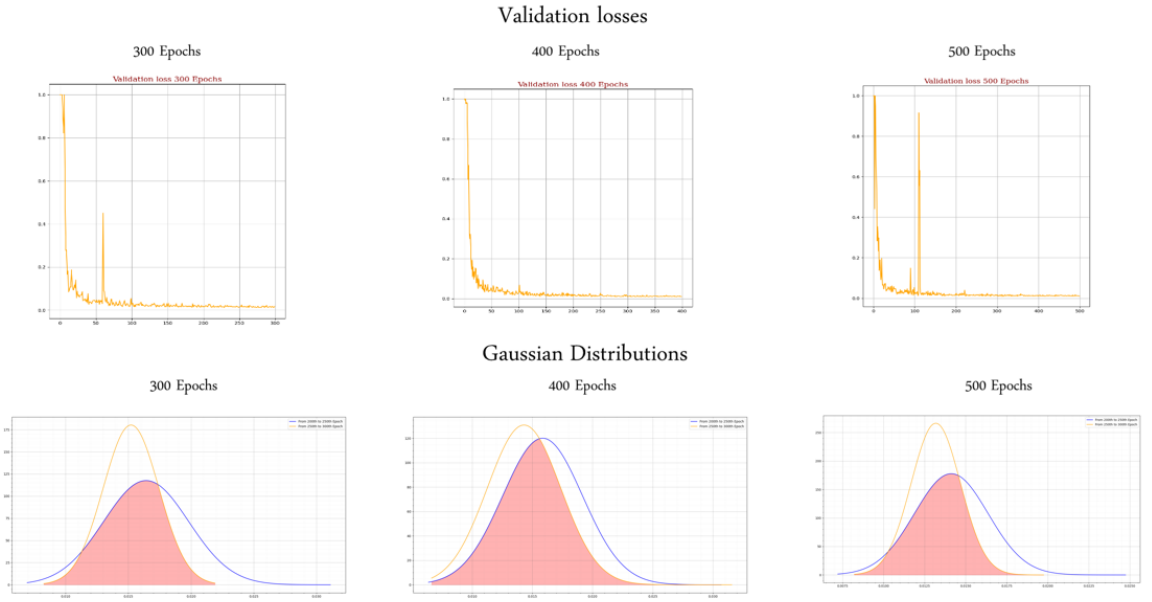


5.5.4 Kernel SP-net:Number of Epochs

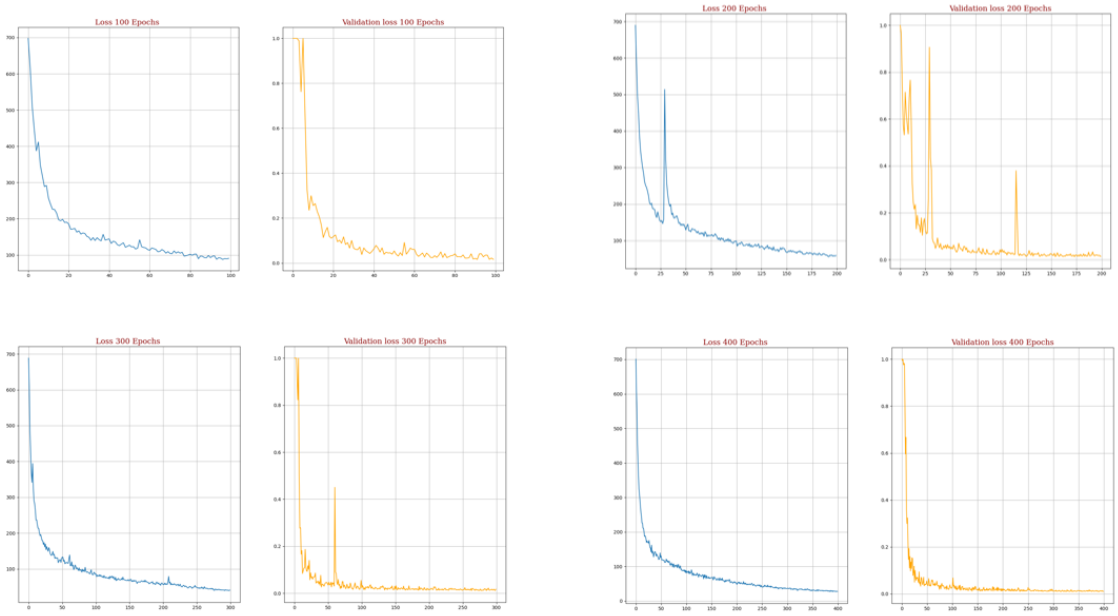
Kernel SP-net without Data Augmentation : Lass and validation loss for epochs



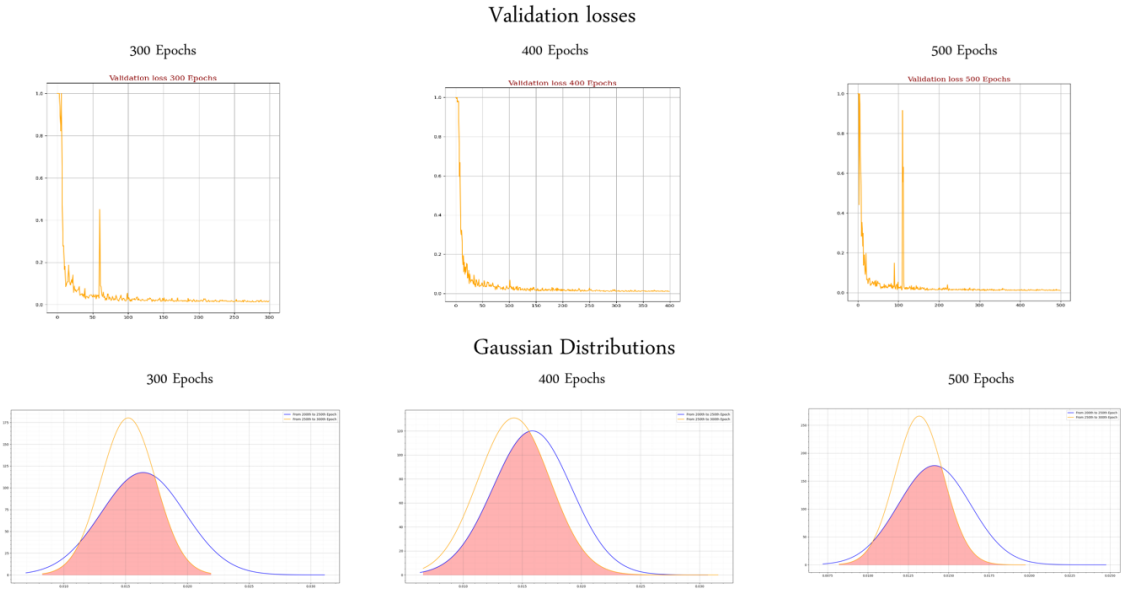
Kernel SP-net without Data Augmentation : Validation losses and Normal distributions



Kernel SP-net with Data Augmentation : Lass and validation loss for epochs

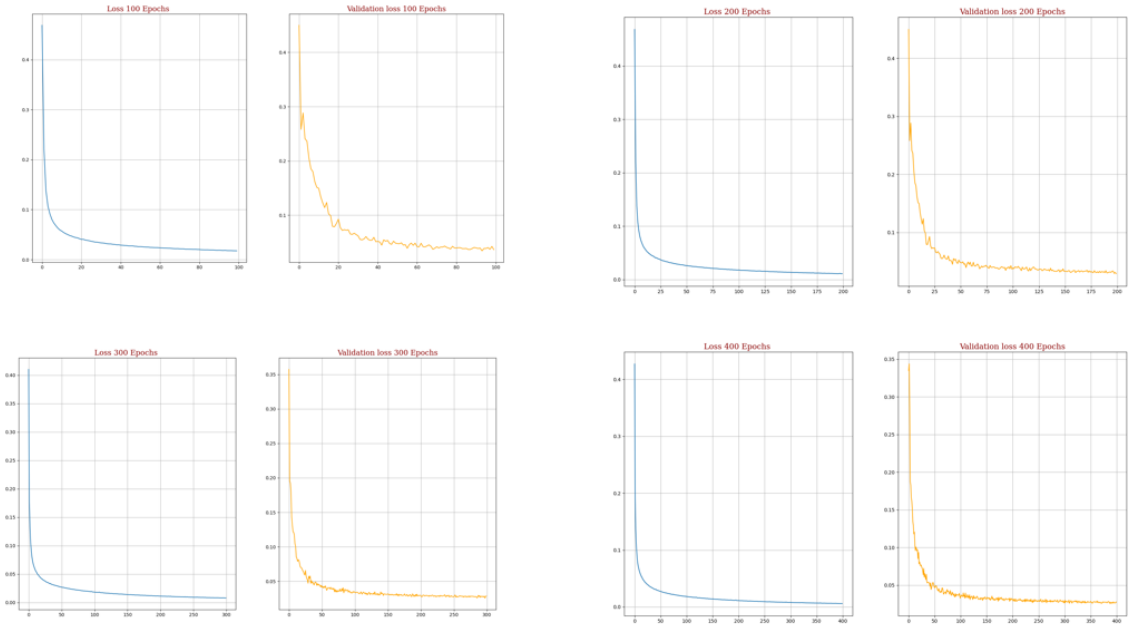


Kernel SP-net with Data Augmentation : Validation losses and Normal distributions

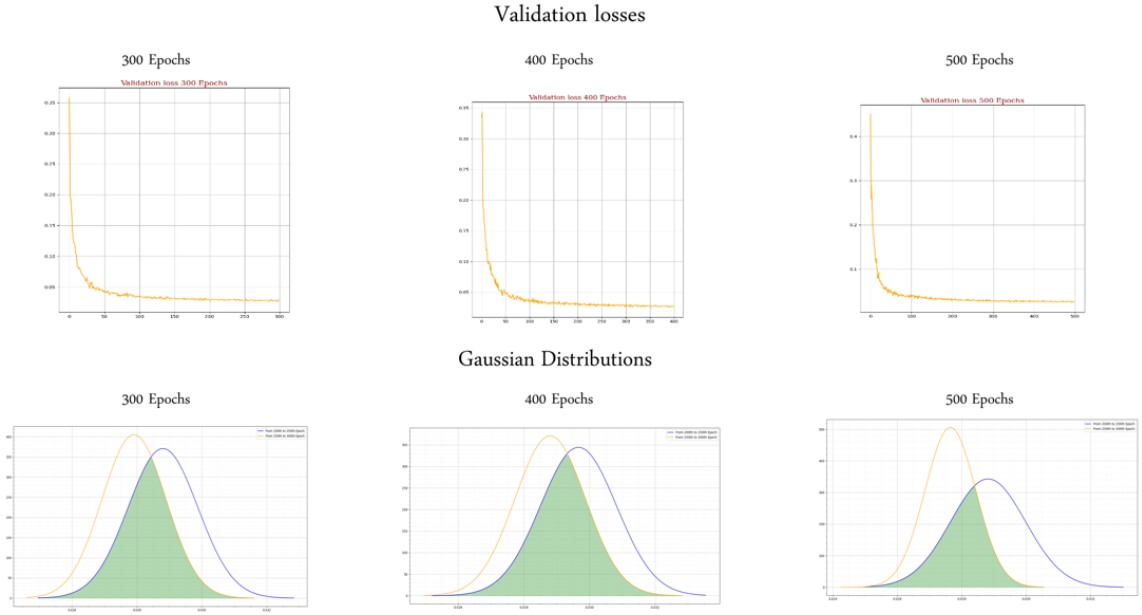


5.5.5 U-net:Number of Epochs

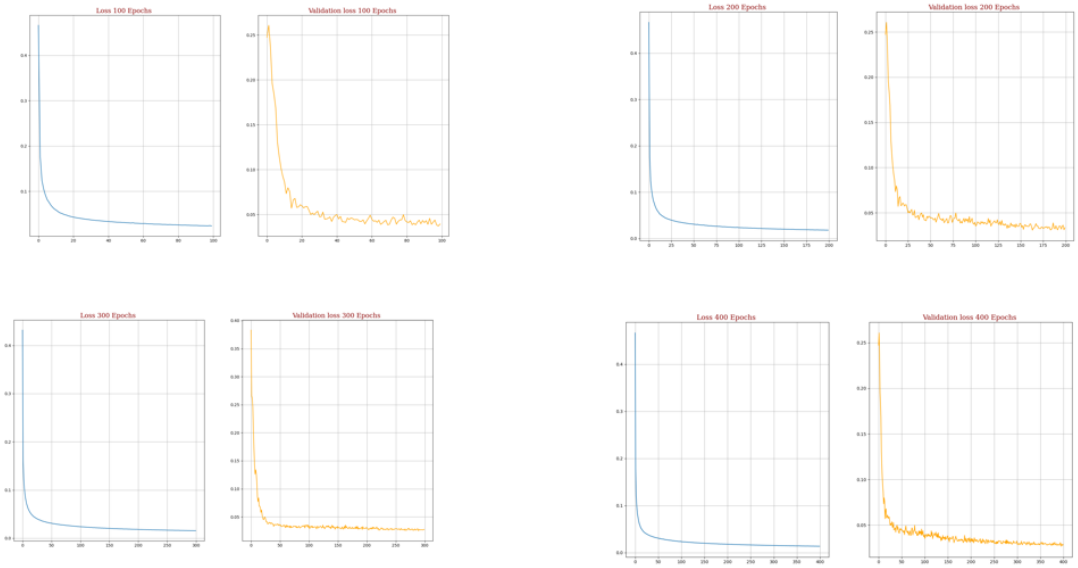
Unet without Data Augmentation : Lass and validation loss for epochs



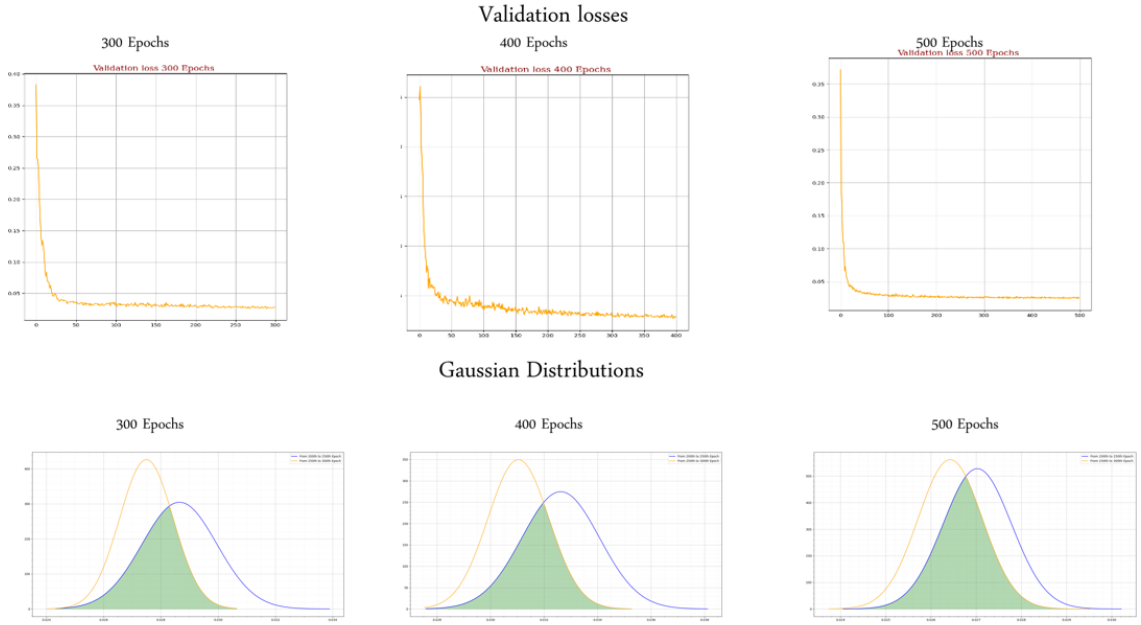
U-net without Data Augmentation : Validation losses and Normal distributions



Unet with Data Augmentation : Lass and validation loss for epochs



U-net without Data Augmentation : Validation losses and Normal distributions



Therefore, denoting slight shifts of the superpositions of the normal distributions, noted in the previous pages, was an optimal number of epochs equal to 300 in the training of our neural networks.

5.5.6 Batch size

The batch size is determined by the condition and capacity of the GPU memory and the size of the training dataset. In the proposed methods, it is observed that a value of 40 is ideal.

5.5.7 Steps per epochs

steps per:epoch can be calculated by total_samples/batch_size but considered larger to ensure that all images are implemented in the network . On the other a few times wider when data augmentation is used.In this perspective for standard case of dataset 10 iterations were used ,in data augmentation case 100 steps were used..

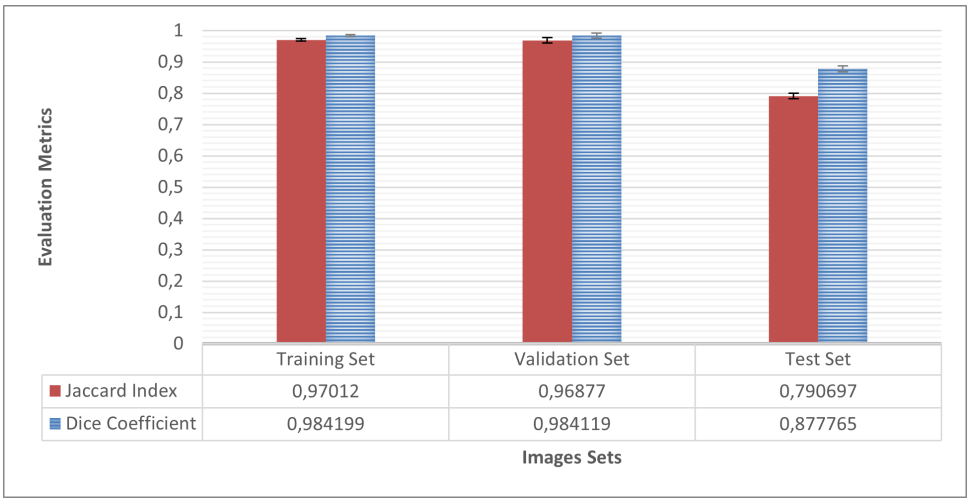
5.6 Final Results

The following paragraphs discuss and expose the final results obtained based on the training of the proposed neural networks. The results are evaluated principally considering the units of measurement previously mentioned, that is the jaccard index and the dice coefficient, performed and calculated on the total images belonging to the test and therefore on 43 images of the validation and training set. Given the conditions established in the previous sections, the experiments for each method were performed by training the network for 300 epochs, using an initial learning rate of 0.001.However Unet, as reported also in the literature, was trained using the strategy of Adam . Regarding the methods derived from SP-net they were trained according to the results obtained and therefore with the optimizer Nadam. Finally, the results reported here are referred to 10 repetitions for the training of each network.

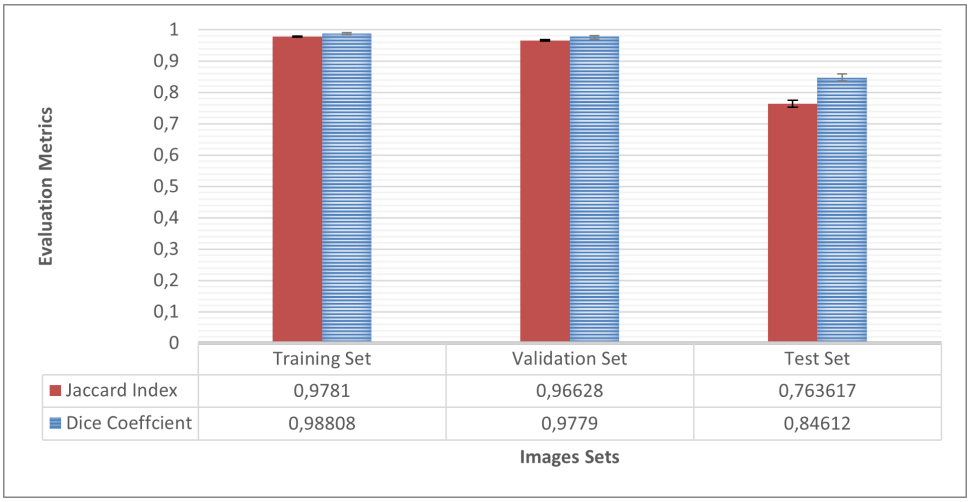
5.6.1 SP-NET Results

his section presents the results obtained for the training set, validation set and test set through the application of SP-NET considering the mean Jaccard Index and mean Dice Coefficient.

SP-NET without Data Augmentation:

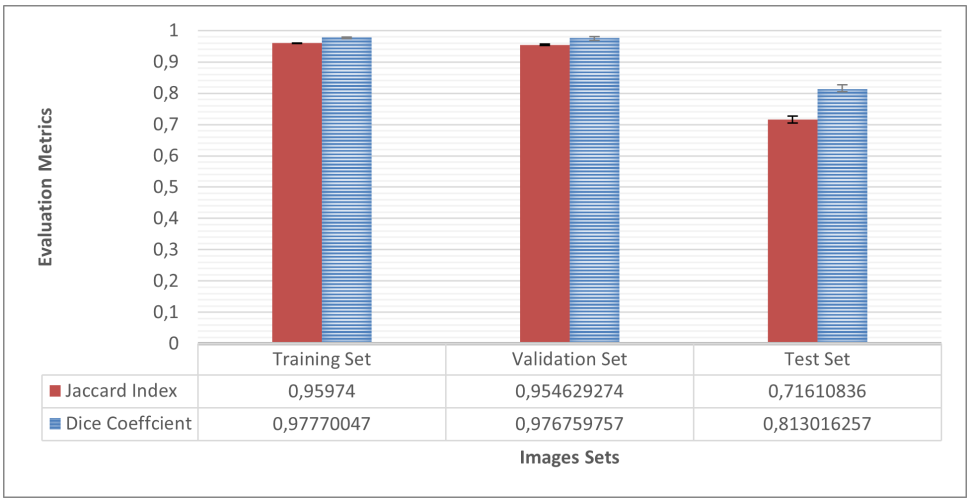


5.6.2 SPNET with Data Augmentation



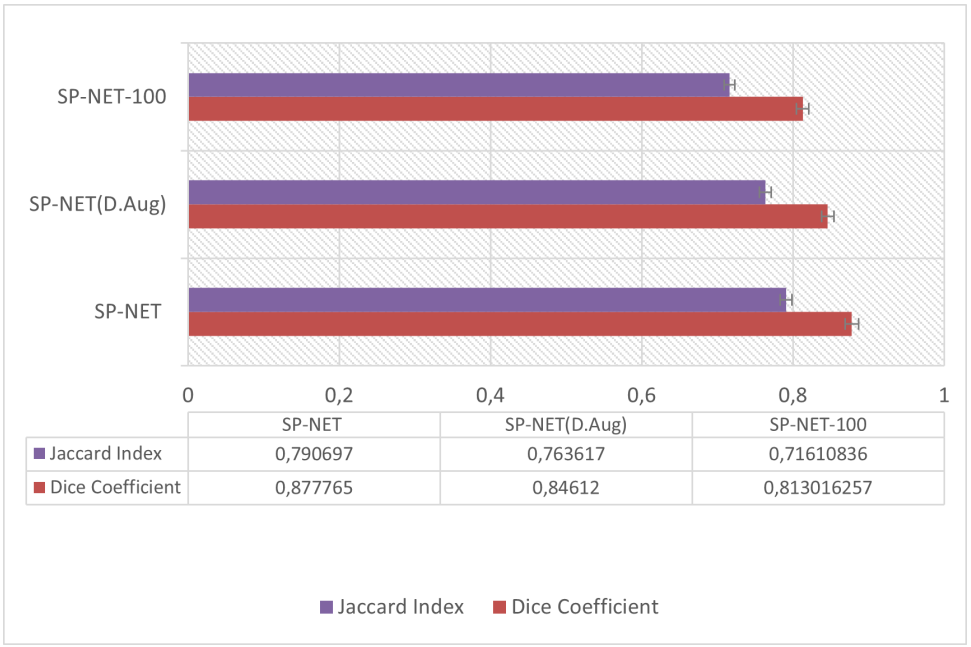
SP-NET 100

These results show the efficiency of the neural network SP-NET in its reduced version, employing less components in the reconstruction of the image and therefore applying an operation of denoing in the reconstruction of the segmentation.Indeed 100 indicates the number of components used.



Comparison of SP-NET methods

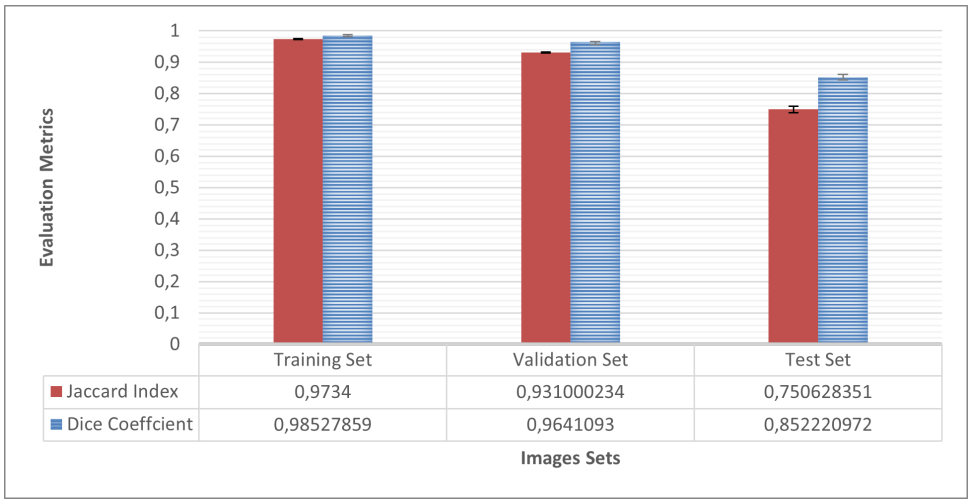
The chart below shows an actual comparison between the results defined with SP-NET and its variants on the test set images .



5.6.3 Kernel SPNET Results

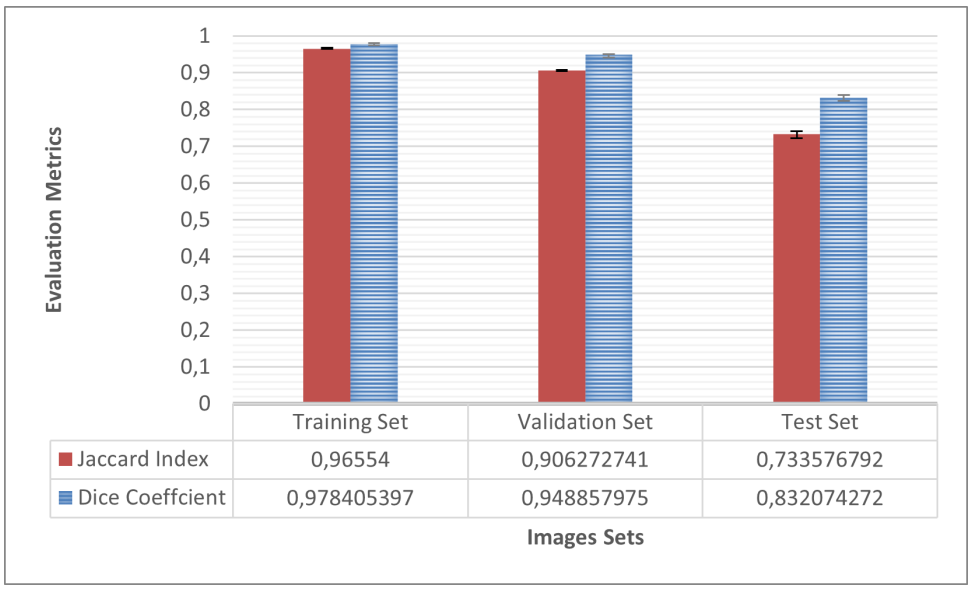
Similarly, this section introduces the key results achieved with a non-linear extension of the method in question and intended as the Kernel SPNET .

Kernel SP-NET without Data Augmentation:



	Training Set	Validation Set	Test Set
Jaccard Index	0,9734	0,931000234	0,750628351
Dice Coeffcient	0,98527859	0,9641093	0,852220972

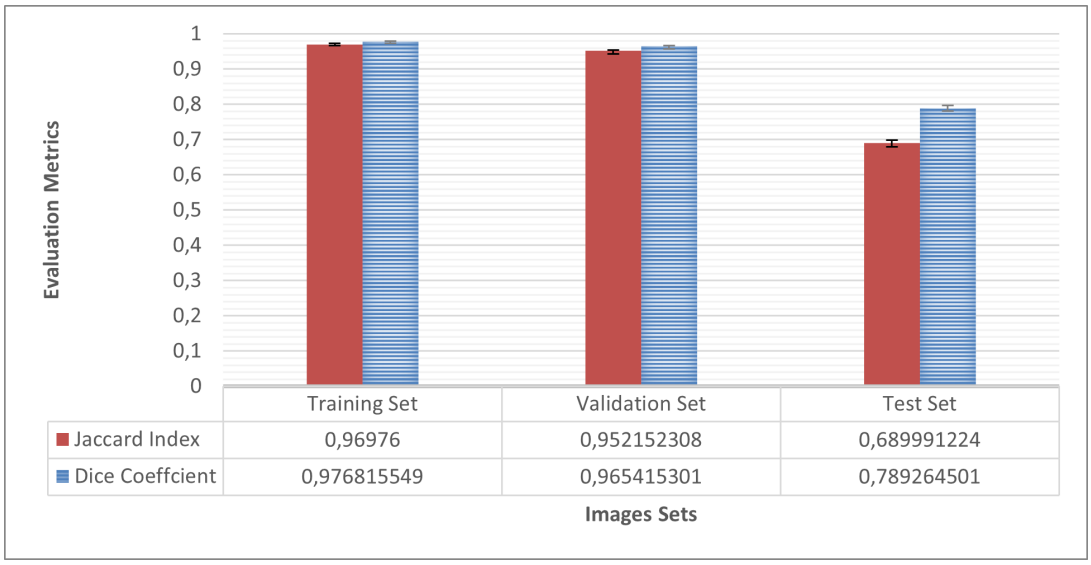
Kernel SP-NET with Data Augmentation:



	Training Set	Validation Set	Test Set
Jaccard Index	0,96554	0,906272741	0,733576792
Dice Coeffcient	0,978405397	0,948857975	0,832074272

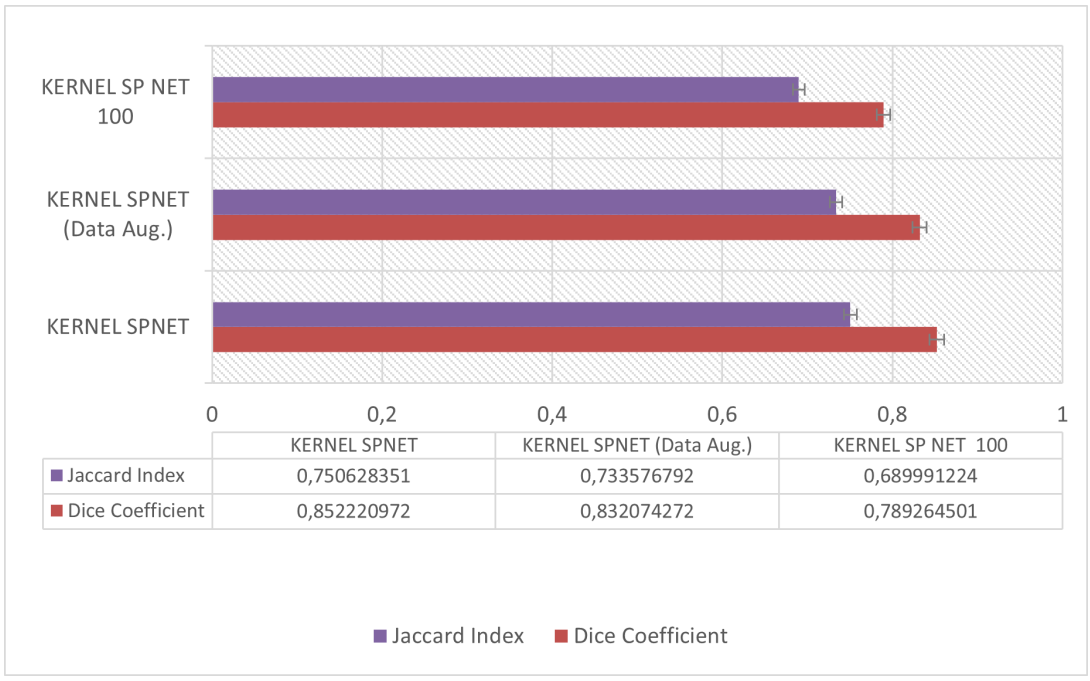
Kernel SP-NET 100

SImilar to the Denoisng process applied by SPNET 100 , in this we investigate results with the same procedure in the nonlinear version.



	Training Set	Validation Set	Test Set
Jaccard Index	0,96976	0,952152308	0,689991224
Dice Coeffcient	0,976815549	0,965415301	0,789264501

Comparison of KERNEL SPNET Methods:

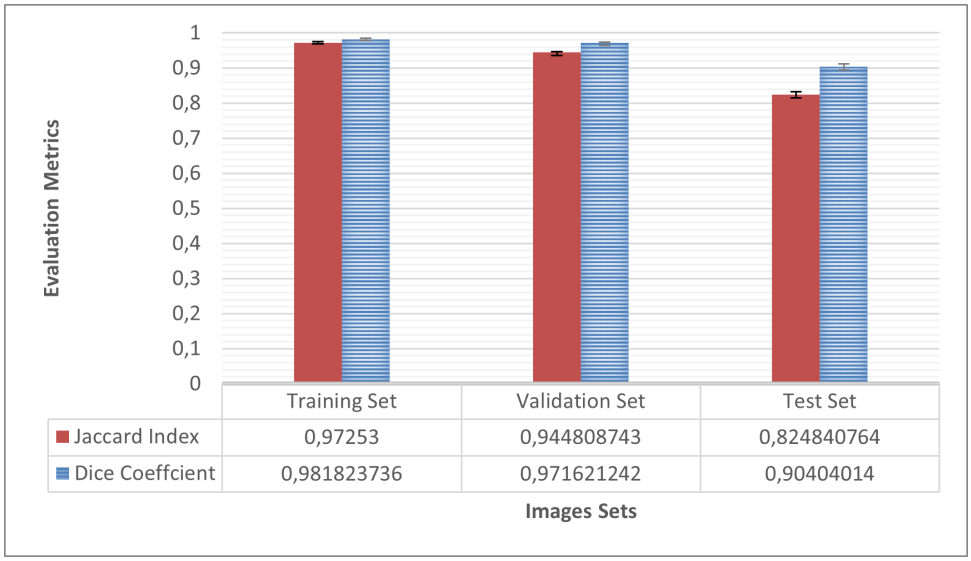


	KERNEL SPNET	KERNEL SPNET (Data Aug.)	KERNEL SP NET 100
Jaccard Index	0,750628351	0,733576792	0,689991224
Dice Coefficient	0,852220972	0,832074272	0,789264501

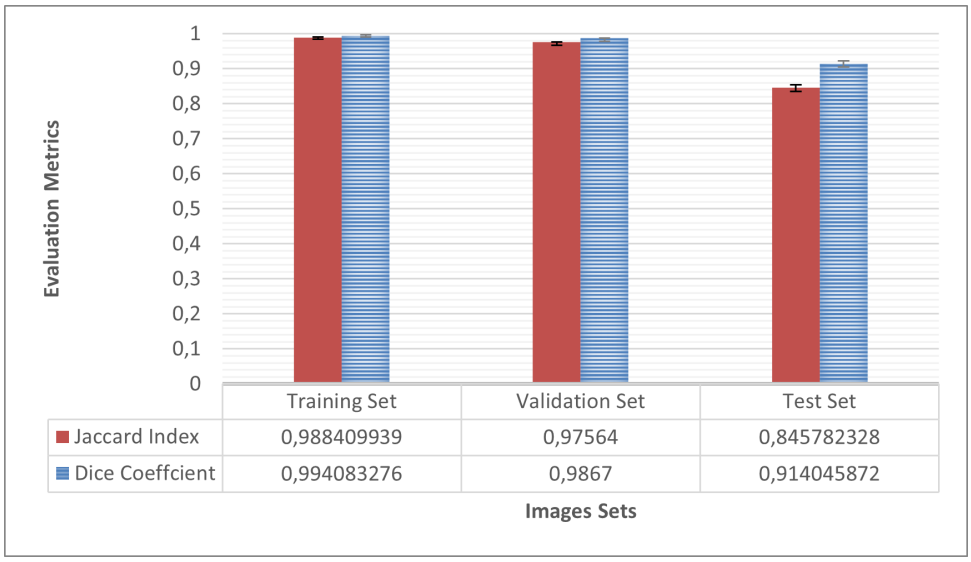
5.6.4 U-NET Results

The following bar graphs represent the results achieved from the referenced method ,applying or not Data Augmentations.

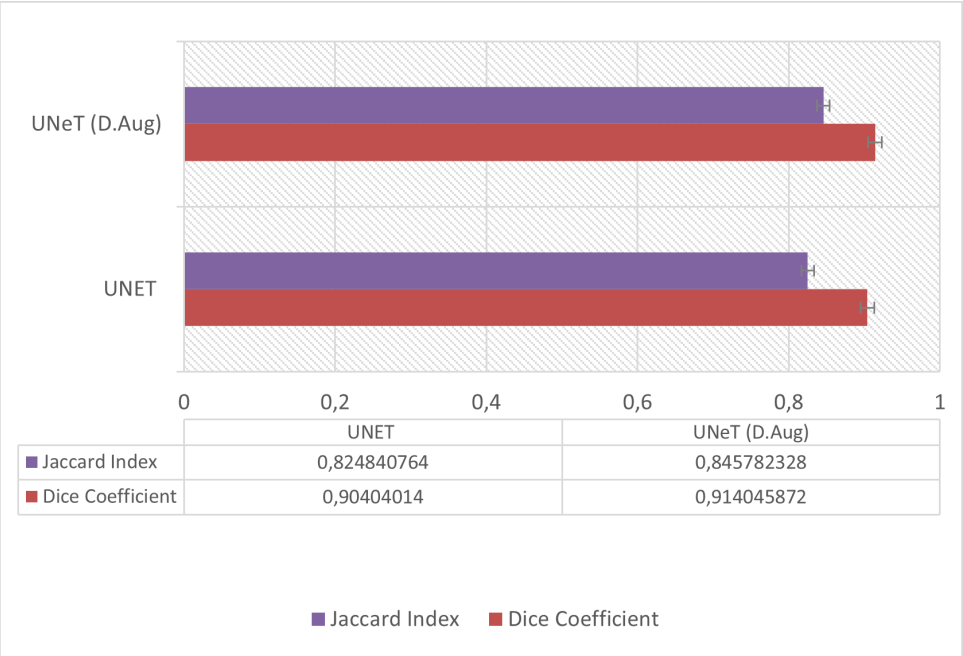
5.6.5 U-NET without Data Augmentation;



5.6.6 U-NET with Data Augmentation:



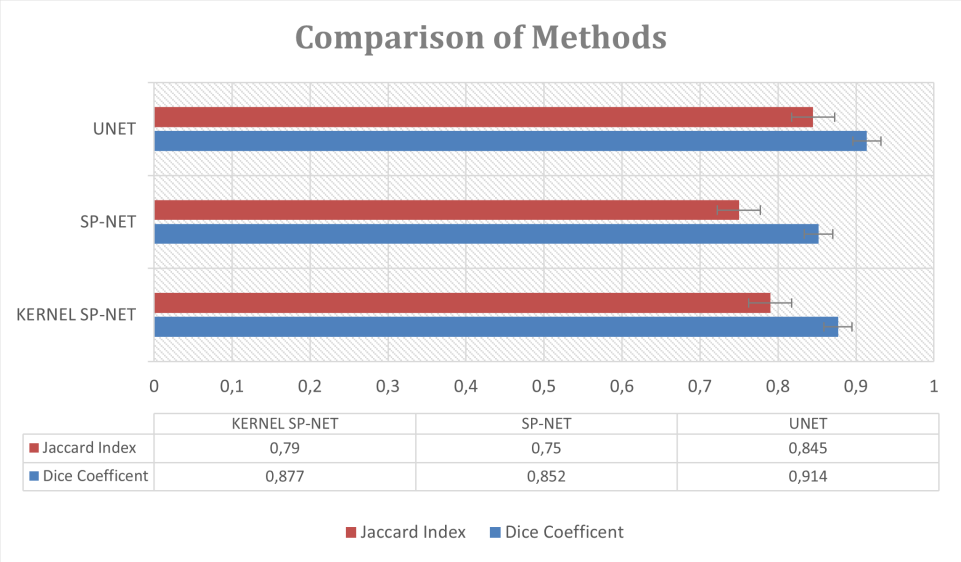
Comparison of Unet Methods :



	UNET	UNeT (D.Aug)
Jaccard Index	0,824840764	0,845782328
Dice Coefficient	0,90404014	0,914045872

5.6.7 Comparison of Methods :Final Results

As a consequence of the results achieved, we realize how the data augmentation in the proposed methods involves only a reduction of performance without improving effectively the learning. Therefore we agree that SP-NET and Kernel SP-NET can be used without the application and use of such technique and therefore in this mode are compared with Unet, that in the reduced version and so employing less components have not given better outcomes.

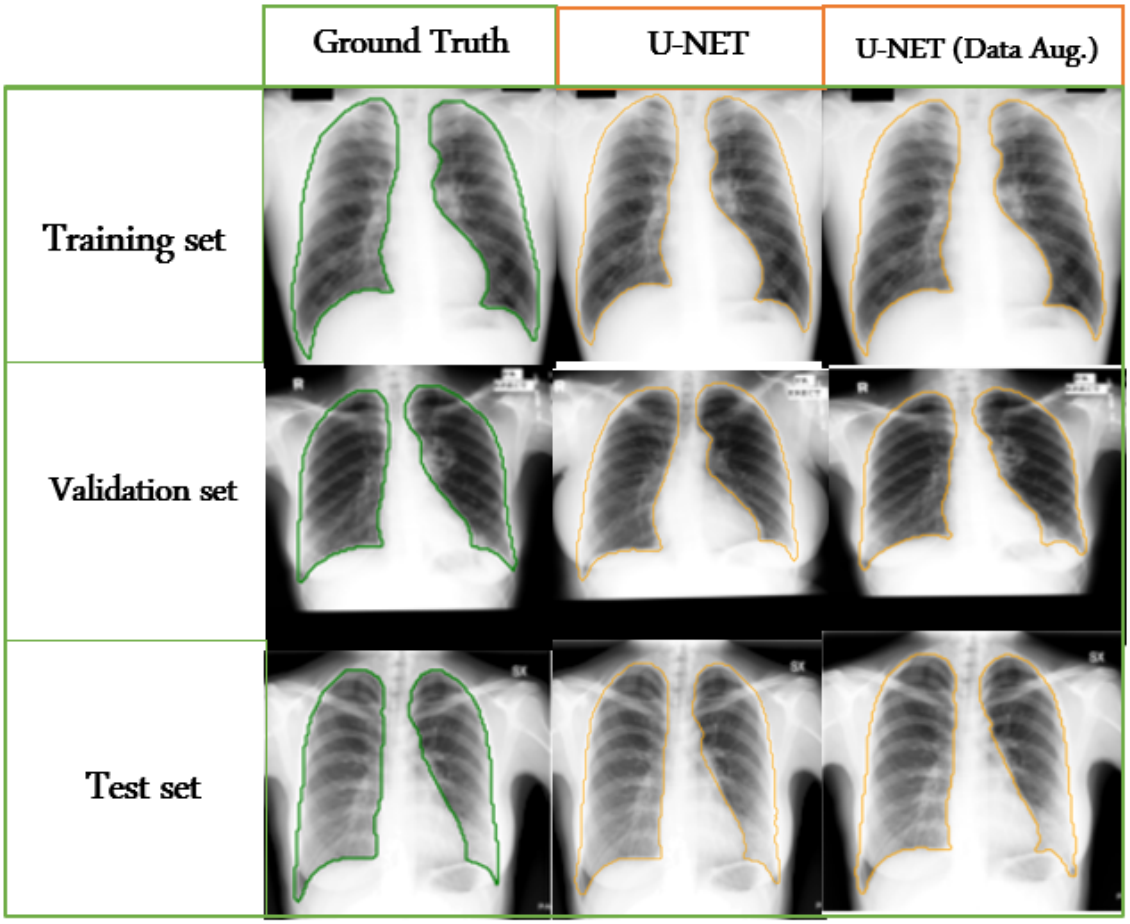


5.7 Qualitative Analysis

The qualitative analysis has the purpose and the desire to make the reader fully understand the results obtained so that he perceive and appreciate the advantages as well as the drawbacks. To this end, this section illustrates the segmentation maps predicted by our algorithms and compares them with the real ones. In the first place, the results are reported with reference to the sets of images given and only then will a qualitative analysis of the best and worst cases be carried out.

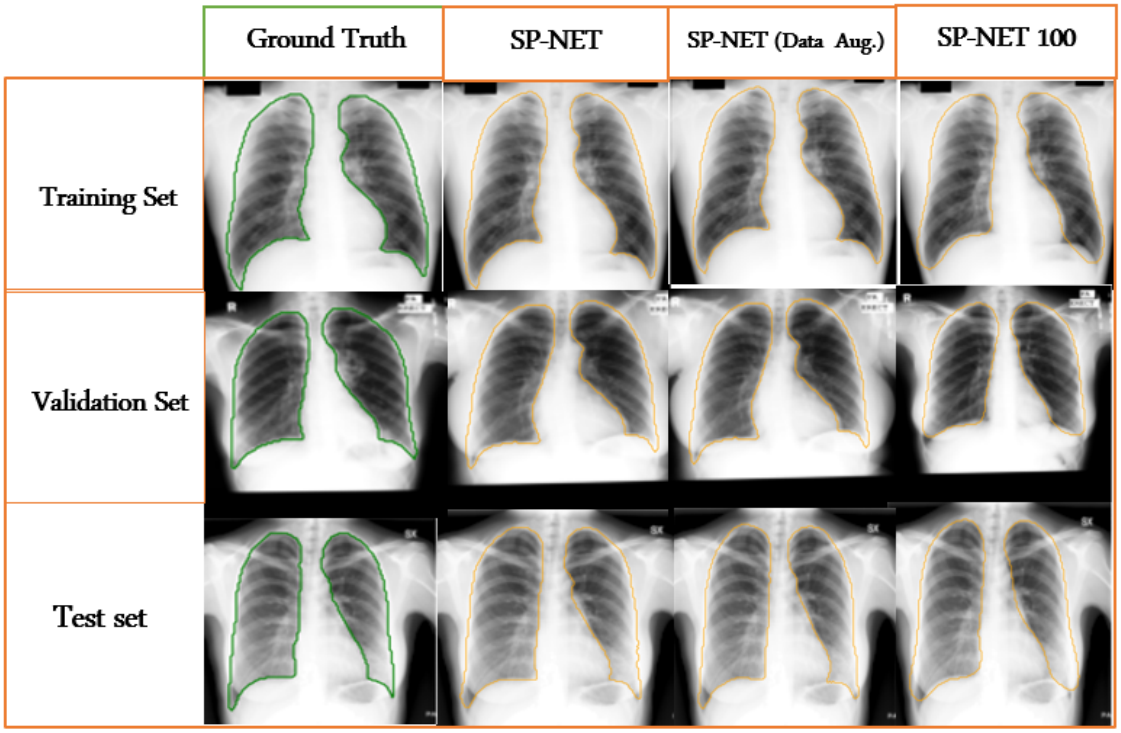
5.7.1 Qualitative Analysis : UNET

Finally, it is possible to appreciate how Unet is more able to pick up and delineate the lung boundary than the previous methods. This fact is better illustrated in the following analysis, which represents in more detail the segmentation of the images of the Aiforcovid dataset.



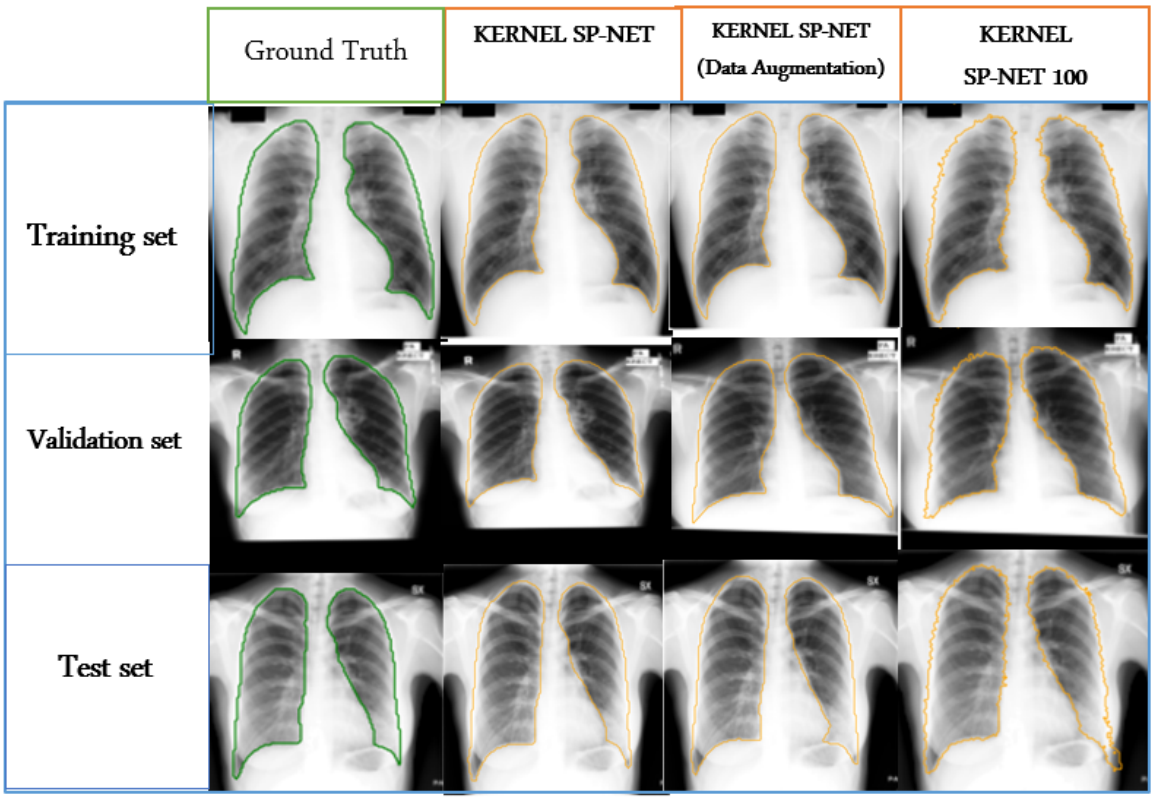
5.7.2 Qualitative Analysis:SP-NET

As it is refined also from an analysis of the first images belonging to the various sets, SP-NET results to be pariticularly effective both on the training and validation set, while on the test set SP.NET results to be able to segment the lungs summarily well.



5.7.3 Qualitative Analysis : Kernel SP-NET

Kernel SPNET in its versions, similarly to SP-NET, succeeds in defining the lung boundaries of both the images belonging to the validation and the training set, but also those belonging to the test set, even if the latter are not exhaustive.



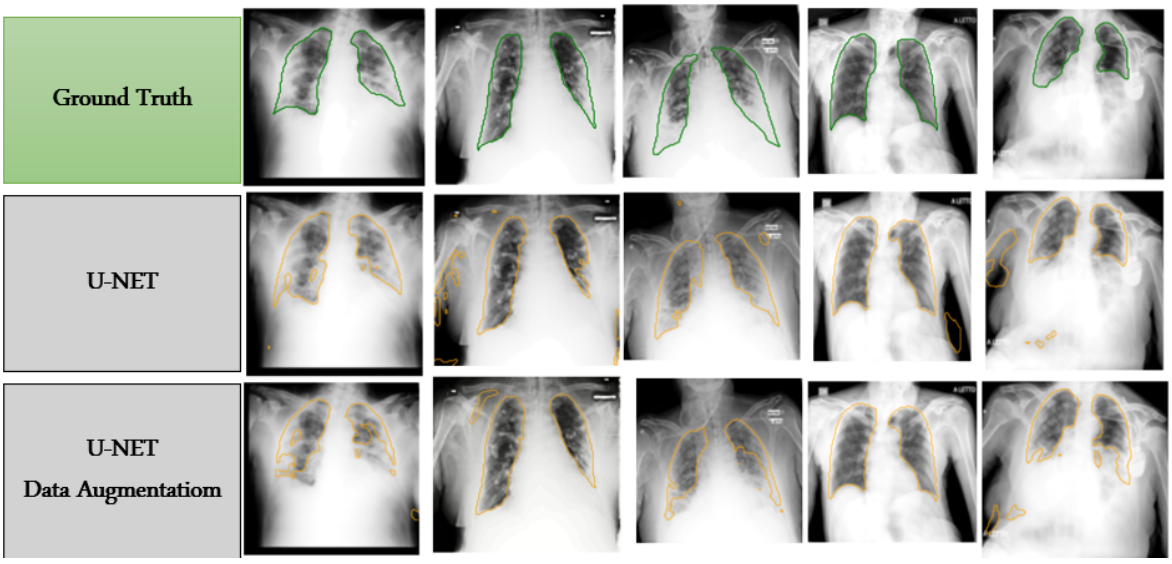
5.7.4 Qualitative Analysis of Aiforcovid Images

This last part is appropriate to give greater attention and prominence to the images that make up the test set, for this reason the section below has the objective of highlighting the harder cases on which it is particularly ardous for the methods discussed to implement a segmentation and the best cases in which they are very effective. In particular, since the methods exposed here have the purpose of improving the current state of the art, the worst cases are those in which U-net performs segmentation maps that are not completely reliable and the best are the images of the Aiforcovid Dataset on which the segmentation is very faithful.

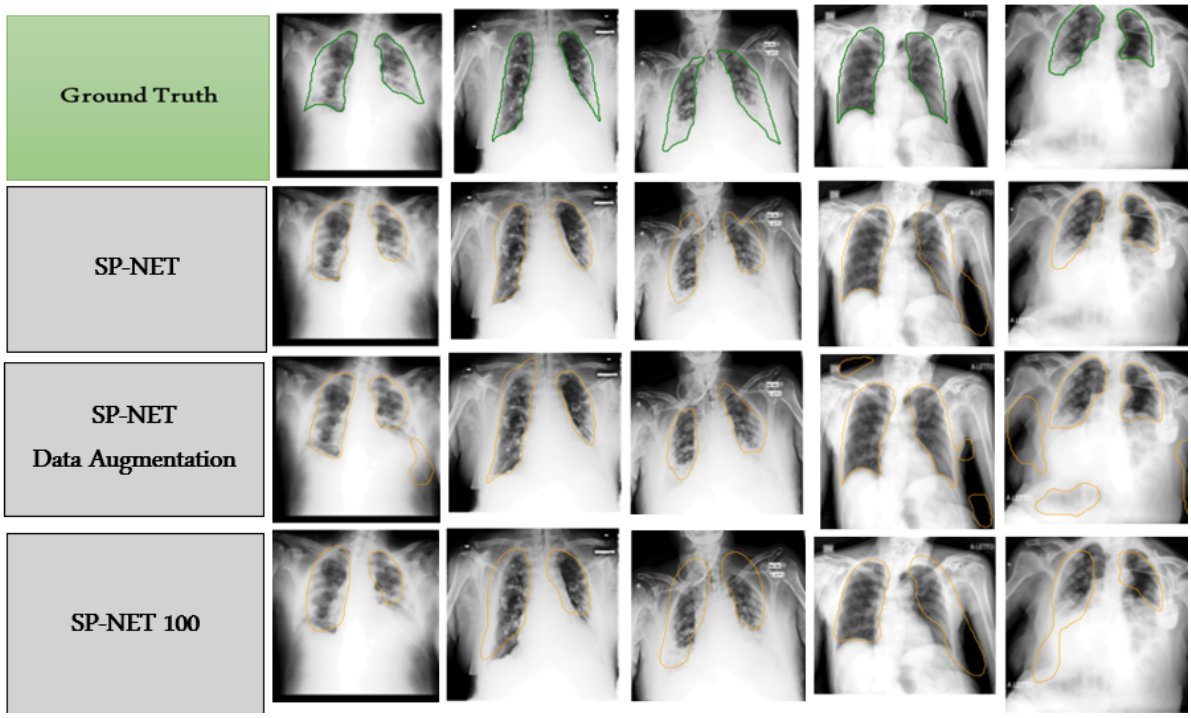
Harder Cases

In this context, the purpose of this analysis is to shed light on the shortcomings and deficiencies of the algorithms used in this work. Therefore, we carry out a qualitative evaluation of the cases extracted from AIforcovid dataset and that proved to be the most challenging to segment
.

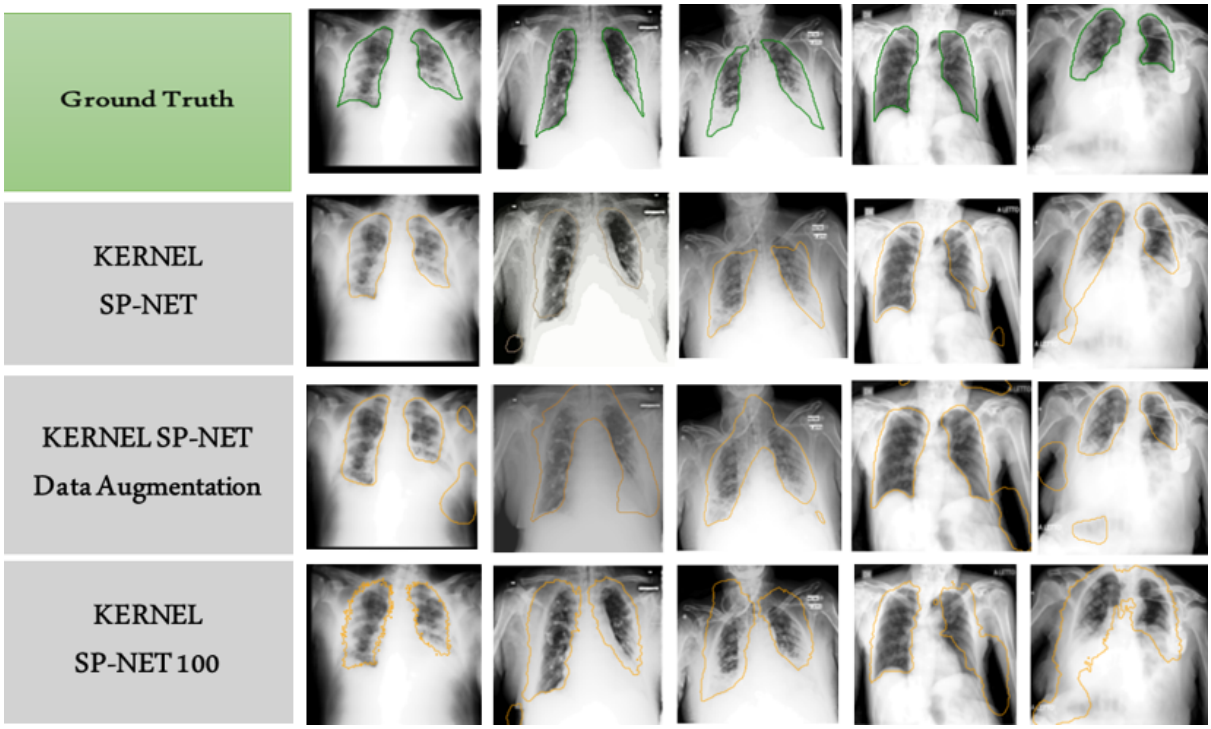
1.Unet :



2.SP-NET:

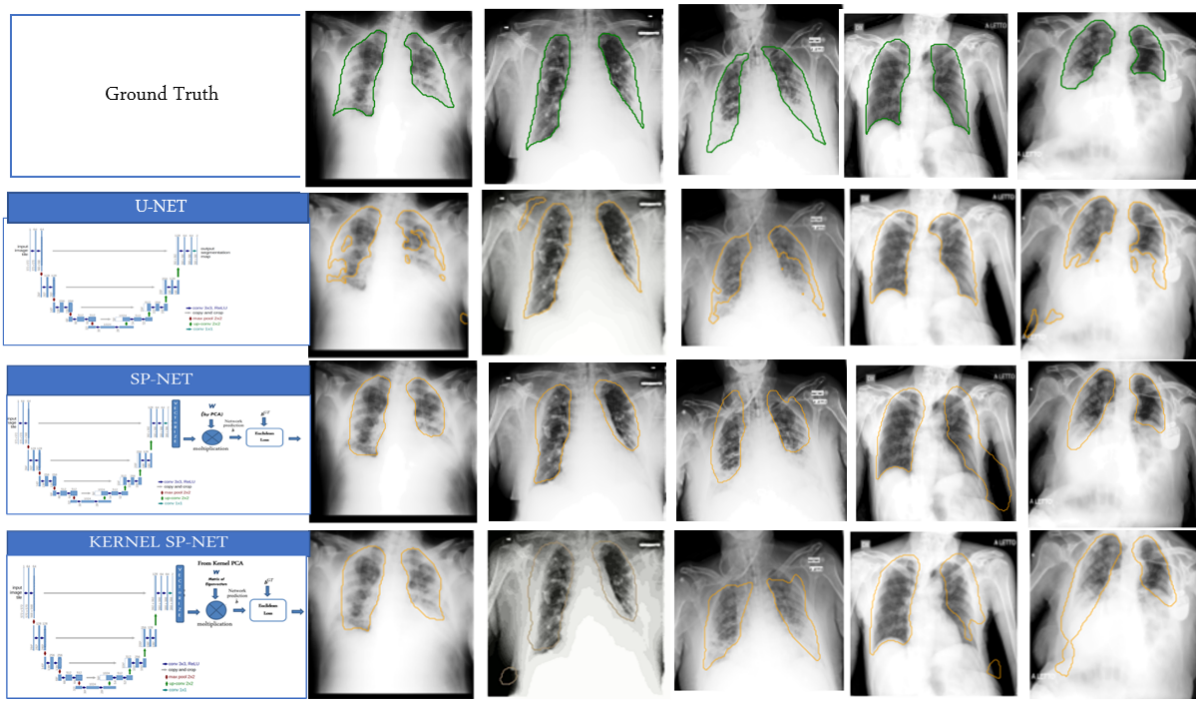


3.KERNEL SP-Net:



Harder Cases : Comparison of Methods

The last page of this chapter intends to perform among the methods and configurations considered best based on the results obtained. Starting from this, the segmentation maps obtained by SP-NET and Kernel SP-NET are compared with those of U-NET, and evaluated on the cases in which the latter fails.



5.8 Discussion of the Results

From the results, it emerges how SPnet, both in its original and the non-linear form (Kernel SPnet), fails to give qualitatively and numerically better performances than Unet . Although the dice coefficient and the jaccard index of the segmentations predicted by both the networks mentioned above are close to those of the reference method, they have significant differences. It is explained by some flaws and detriments reported by the applied methodologies and techniques used to acquire and capture apriori information based on the expected shape by the training dataset. From the qualitative analysis of proposed methods, it is possible to observe a limitation in the segmentation of some of the lungs present in the AIforcovid Dataset in many cases. Thus, it is possible to notice a defect called and intended as the "bridge effect2 corresponding to the conjunction of the right lung segmentation map and the left one at the top of the tracheal compartment. This imperfection is justified because the analysis of principal components, both in the standard version and with a kernel function, finds it challenging to analyze and perceive the variance of the pixels located in that area because they assume negative and positive values and so are very discontinuous. This problem could be solved by applying a specific threshold to analyze the pixels that make up this area or applying unique methodologies, but this would involve a greater complexity of the algorithm used to recognize lung boundaries. This problem could be solved by applying a specific threshold to analyze the pixels that make up this area or applying unique methodologies, but this would involve a greater complexity of the algorithm used to recognize lung images. However, in the images with poor contrast, the algorithm gives approximate segmentation, more faithful on some occasions than those predicted from Unet. This result is also found with the use of the Kernel SPnet, and the same applies unfortunately also to the defects described. Other problems are evident from the Data Augmentation. Even if, in particular cases, it gives more optimal segmentation for a dataset of greater dimensions, in most images, it originates the detection of false lung fields from the outcome of the network. his effect is since the geometric transformations used in the generation of artificial images carried out during the data augmentation introduce a distribution and, the variance of pixels where usually in chest radiographs where there are no lung areas. Therefore it is captured by the first principal components and reported in the reconstruction of the predicted masks. It follows that even using fewer components and therefore a reduced matrix of eigenvectors, this error would still be present. As a result, the appliance of data augmentation renders the algorithm more robust. On the other hand, it introduces artifacts in the prediction of the masks due to rotations and translations on the original pulmonary curves. The following result is also observed in the extension provided by the Kernel PCA since it performs steps analogous

to the principal component analysis. Extracting in the first components the pixels with more significant variance and corresponding to areas typically recognized as lung fields, Kernel SP-net with Data Augmentation predicts regions of segmentation outside the lung contours indicated by the manually annotated masks. The outcome of this segmentation depends on the shape that you want to annotate and of which you want to get the apriori information. In fact, in the reference article from which the following methods draw inspiration, it is noted that instead, the data augmentation leads to better performance due to the context of this application. In the case of the paper, in fact, the proposed method had the objective of segmenting vertebrae, and therefore any misalignment implemented in the generation of a larger dataset does not lead to a wrong distribution of pixels but a performance improvement. It is clear from the discussion of these results that the principal component analysis in the standard version and nonlinear integration and capture of apriori information and applicability depends very much on the anatomical shape in question and the desired segmentation.

CHAPTER 6

Conclusion

The section described here concerns the conclusion of the work accomplished so far. Indeed it composes the final thesis retrieved as a result of the results described and discussed previously. Therefore this section exposes in the first part a real conclusion of the results obtained from the proposed methods, focusing on how they are given principally by the techniques used to treat the apriori information acquired. In addition, this section also provides a description of possible solutions that could be proposed on the basis of the proposed algorithm in order to improve it. In fact, this part of the article is described as "Future Work."

6.1 Conclusion

Following the results obtained and discussed in the previous chapter, it is possible to conclude that the work achieved, the deepening, and the application of the proposed methods can give significant improvements only in the analysis of medical images in specific contexts. It was observed mainly in lung field segmentation as it is of vital importance in the development of methods addressed, the techniques used to acquire apriori information about the expected anatomical shape. Although the principal component analysis itself and its nonlinear form are used in many shape analysis cases, it has some limitations in this area. Although Spnet and Kernel Spnet, thanks to the principles and steps of the principal component analysis used for a priori information, can obtain a reliable segmentation in low-quality images compared to the unet, in the remaining part of the cases considered difficult, have significant defects not shown in state of the art. Since the latter is also present in the nonlinear version, we conclude that even using the Kernel PCA to explore the characteristics and nonlinear correlations do not significantly benefit. From this perspective, we perceive how incorporating apriori information can be helpful only in some cases, namely to ensure that we have approximate segmentations with a lung-like structure in the low-contrast images provided by the Aiforcovid dataset. While in the remaining cases, using the information by a network like Unet is not indispensable. Besides, it has been demonstrated that using different datasets, thanks to data augmentation, there is no significantly better performance. It also implies that with such applications to exploit the advantages, should be used other different techniques to capture the information apriori, because as also exposed in the previous section, the use and the basic concepts of principal component analysis and kernel PCA involves the introduction of some artifacts and false segmentations not present in the original version of the methods. In conclusion, the apriori information regarding the expected anatomical shape can be an additional faculty of the convolutional neural networks, especially when the network cannot give an effective segmentation but different from the natural mask manually annotated and reported by the ground truth. This concept will be expressed in the next section, describing the future work that can be developed possibly based on the results retrieved in this thesis.

6.2 Future Work

As already anticipated in the conclusion of the thesis, as far as the results reported in this work are concerned, the a priori information has a crucial role only in the segmentation of some images because thanks to this, they can delineate reliable pulmonary profiles compared to the Unet segmentation. We understand how the use and integration of a priori information can be essential in providing reliable discrimination in very complex images, while in others, it is superfluous or even harmful. Moreover, as can be seen from the use of Kernel PCA, an alternative form of principal component analysis is not decisive in this context. From this point, it follows that rather than investigating different techniques to capture the a priori information, it is more helpful to investigate in future works a method to consider when to use or not the a priori information in the prediction of the segmentation predicted by convolutional networks. As suggested by the conclusive part of the thesis, the shape knowledge is valuable only in those images in which unet succeeds in not distinguishing the areas belonging to the lungs from those forming the background, and in this case, the use of principal components could be decisive. To this end, for future work, we propose to approach the constitution of neural networks described in the thesis and seek a method that can distinguish those scenarios in which it is necessary to consider the apriori knowledge and those in which the use of these could lead to a worsening of performance as it would excessively affect the segmentation resulted from the convolutional neural network in question. In this way, integrating the a priori information could be a real solution and an additional opportunity to obtain realistic segmentations even in the most complex cases. However, on the other hand, we realize that introducing this mechanism certainly involves a greater complexity in the development and training of the neural network, and it was an exciting challenge given the potential benefits described above.

References

[1] Lucas Teixeira, Rodolfo M. Pereira, Diego Bertolini , Luiz S. Oliveira, Loris Nanni, George D. C. Cavalcanti and Yandre M. G. Costa
Impact of lung segmentation on the diagnosis and explanation of COVID-19 in chest X-ray images. . 2021.

[2] Paolo Soda , Natascha Claudia D'Amico, Jacopo Tessadori , Giovanni Valbusa et al. *AIforCOVID: predicting the clinical outcomes in patients with COVID-19 applying AI to chest-X-rays. An Italian multicentre study.* 2020

[3] Joanne Cleverley, James Piper, Melvyn M Jones
The role of chest radiography in confirming covid-19 pneumonia BMJ 2020

[4] Hayit Greenspan, Raúl San José Estépar, Wiro J Niessen, Eliot Siegel, and Mads Nielsen
*Position paper on COVID-19 imaging and AI: From the clinical needs and technological challenges to initial AI solutions at the lab and national level towards a new era for AI in healthcare.*Medical Image Analysis, 66:101800, 2020

[5] Jiuxiang Gua, , Zhenhua Wangb, , Jason Kuenb , Lianyang Mab , Amir Shahroudyb , Bing Shuaib , Ting Liub , Xingxing Wangb , Li Wangb , Gang Wangb , Jianfei Caic , Tsuhan Chenc
Recent Advances in Convolutional Neural Networks 2015

[6] Cornelia Schaefer-Prokop, Ulrich Neitzel, Henk W. Venema, Martin Uffmann,and Mathias Prokop
Digital chest radiography: an update on modern technology, dose containment and control of image quality European Radiology 2008

[7] Bruno Petinaux , Rahul Bhat, Keith Boniface and Jaime Aristizabal
Accuracy of radiographic readings in the emergency department The American Journal of Emergency Medicine Volume 29, Issue 1, January 2011, Pages 18-25

[8] Aaron Sodickson , Graham I. Warden, Cameron E. Farkas, Ichiro Ikuta, Luciano M. Prevedello, Katherine P. Andriole, Ramin Khorasani
Exposing Exposure: Automated Anatomy-specific CT Radiation Exposure Extraction for Quality Assurance and Radiation Monitoring. Radiology Journal 2012.

[9] American College of Radiology
Understanding Radiation Risk from Imaging Tests. Radiological Society of North America. Children and Radiation Safety. Accessed at www.radiologyinfo.org/en/info.cfm?pg=safety-rad-children 2012

[10] Myles L. Mowery and Vikramjeet Singh.
X-ray Production Technical Evaluation . StatPearls Publishing 2021

[11] European Society of Radiology
Breathe Easy: How radiology help to find and fight lung diseases . 2013

[12] Rubin GD, Ryerson CJ, Haramati LB, Sverzellati N, Kanne JP, Raoof S, Schluger NW, Volpi A, Yim JJ, Martin IB, et al.
*The role of chest imaging in patient management during the covid-19 pandemic: a multinational consensus statement from the fleischner society.*Radiology Journal 2020

[13] Feng Shi , Jun Wang, Jun Shi , Ziyan Wu , Qian Wang, Zhenyu Tang, Kelei He , Yinghuan Shi , and Dinggang Shen
Review of Artificial Intelligence Techniques in Imaging Data Acquisition, Segmentation, and Diagnosis for COVID-19..IEEE REVIEWS IN BIOMEDICAL ENGINEERING, VOL. 14, 2021

[14] Yang, W., Sirajuddin, A., Zhang, X. et al
The role of imaging in 2019 novel coronavirus pneumonia (COVID-19) Eur Radiol 30, 4874–4882 (2020)

[15] Mingchun Ou , Jieyun Zhu , Pan Ji , Hongyuan Li , Zhimei Zhong , Bocheng Li , Jielong Pang , Jianfeng Zhang and Xiaowen Zheng
Risk factors of severe cases with COVID-19: a meta-analysis Cambridge University Press: 12 August 2020

[16] Haiming Tang, Nanfei Sun and Yi Li
Segmentation model of the opacity regions in the chest X-rays of the Covid-19 patients in the us rural areas and the application to the disease severity BMJ 2020

[17] Arman Haghanifar, Mahdiyar Molahasani Majdabadi, Younhee Choi, S. Deivalakshmi and Seokbum Ko
COVID-CXNet: Detecting COVID-19 in Frontal Chest X-ray Images using Deep Learning Electrical Engineering and Systems Science , 2020

[18] Rob Wonnacott MSN, RN, CCRN
X-Ray Interpretation University of Michigan Medical Center ,Michigan Society for Respiratory Care , Fall Conference 2015

[19] Zahangir Alom, M M Shaifur Rahman, Mst Shamima Nasrin, Tarek M. Taha, Vijayan K. Asari
COVID_MTNet: COVID-19 Detection with MultiTask Deep Learning Approaches Electrical Engineering and Systems Science, Image and Video Processing 2020

[20] Yusuf S. Khan; David T. Lynch
Histology, Lung, The National Center for Biotechnology 2021

[21] Stephan Voigt
How to Read a Chest X-ray – A Step By Step Approach Southern Sudan medical journal 2008

[22] Jaeger S., Karargyris A., and Candemir S., et al
'Automatic screening for tuberculosis in chest radio graphs: a survey', Quant. Imaging Med. Surg., 2013, 3, (2), pp. 89–99

[23] Nakamori N., Doi K., and Sabeti V., et al
Image feature analysis and computer aided diagnosis in digital radiography: automated analysis of sizes of heart and lung in chest images Med. Phys., 1990, 17, (3), pp. 342–350

[24] Yao J., Dwyer A., and Summers R.M., et al
Computer-aided diagnosis of pulmonary infections using texture analysis and support vector machine classification Acad. Radiol., 2011, 18, (3), pp. 306–314

[25] Santosh K.C., Candemir S., and Jaeger S., et al.
Rotation detection in chest radiographs based on generalized line histogram of rib-orientations 2014 IEEE 27th Int. Symp. Computer-Based Medical Systems, May 2014, pp. 138–142

[26] Santosh K.C., Candemir S., and Jaeger S., et al.
Automatically detecting rotation in chest radiographs using principal rib-orientation measure for quality control Int. J. Pattern Recognit. Artif. Intell., 2015, 29, (02), p. 155

[27] Zhao H., Lo S.C.B., and Freedman M.T
Directional edge registration for temporal chest image subtraction 30th Applied Imagery Pattern Recognition Workshop, AIPR 2001, October 2001, pp. 32–36

[28] Kuhlman J.E., Collins J., and Brooks G.N., et al.
Dual-energy subtraction chest radiography: what to look for beyond calcified nodules Radiographics, 2006, 26, (1), pp. 79–92

[29] Soleymanpour E., Pourreza H.R., and Ansaripour E.,
Fully automatic lung segmentation and rib suppression methods to improve nodule detection in chest radiographs .Med. Signals Sens., 2011, 1, (3), pp. 191–199

[30] Manji F., Wang J., and Norman G., et al. ,
Comparison of dual energy subtraction chest radiography and traditional chest X-rays in the detection of pulmonary nodules Quant. Imaging Med. Surg., 2016, 6, (1), pp. 1–5

[31] Freedman M.T., Lo S.C., and Seibel J.C., et al:
Lung nodules: improved detection with software that suppresses the rib and clavicle on chest radiographs radiology, 2011, 260, (1), pp. 265–273

[32] Li F., Engelmann R., and Pesce L., et al.
Improved detection of focal pneumonia by chest radiography with bone suppression imaging Eur. Radiol., 2012, 22, (12), pp. 2729–2735

[33] Tsubamoto M., Johkoh T., and Kozuka T., et al.
Temporal subtraction for the detection of hazy pulmonary opacities on chest radiography AJR Am. J. Roentgenol., 2002, 179, (2), pp. 467–471

[34] Marano R., Liguori C., and Savino G., et al.
Cardiac silhouette findings and mediastinal lines and stripes: radiograph and CT scan correlation Chest, 2011, 139, (5), pp. 1186–1196

[35] M S Brown 1, L S Wilson, B D Doust, R W Gill, C Sun
Knowledge-based method for segmentation and analysis of lung boundaries in chest X-ray images Comput Med Imaging Graph . Nov-Dec 1998;22(6):463-77

[36] Armato S.G., Giger M.L., and MacMahon H
Detecting costophrenic angles in chest radio graphs' US Patent 6,483,934, 19 November 2002

[37] Maduskar P., Philipsen R.H., and Melendez J.
Automatic detection of pleural effusion in chest radiographs Med. Image Anal., 2016, 28, pp. 22–32

[38] Maduskar P., Philipsen R.H., and Melendez J.
Automated localization of costophrenic recesses and costophrenic angle measurement on frontal chest radiographs 2013

[39] Karargyris A., Siegelman J., and Tzortzis D. et al.
Combination of texture and shape features to detect pulmonary abnormalities in digital chest X-rays Comput. Assist. Radiol. Surg., 2016, 11, (1), pp. 99–106

[40] Cj Vyborny
The aapm/rsna physics tutorial for residents: Image quality and the clinical radiographic examination Radiographics, 17:479–498, 1997

[41] M Edwards , Z Lawson, S Morris, A Evans, S Harrison, R Isaac, J Crocker, C Powell
The presence of radiological features on chest radiographs: how well do clinicians agree? Clin Radiol. . 2012 Jul;67(7):664-8.

[42] Wouter J H Veldkamp ,Lucia J.M. Kroft and Jacob Geleijns
Dose and perceived image quality in chest radiography European Journal of Radiology 72(2):209-17 ,2009

[43] B.van Ginneken
Computer-aided diagnosis in chest radiographs P.h.D. dissertation, Utrecht Univ., Utrecht, The Nederlands, 2001.

[44] Shaoyong Guo, Xinwang Zhang , and Lijie Zhao
*Automatic Lung Segmentation Based on Texture and Deep Features of HRCT Images with Interstitial Lung Disease*2019

[45] Y Shi , F Qi, Z Xue, L Chen, K Ito, H Matsuo, D Shen
*Segmenting lung fields in serial chest radiographs using both population-based and patient-specific shape statistics*2019

[46] Gados D., and Horvath G.
Using heuristics for the lung fields segmentation in chest radiographs Springer, Berlin, Heidelberg, 2007), pp. 802–805

[47] Cheng D., and Goldberg M.
An algorithm for segmenting chest radiographs . Visual Communications and Image Processing'88: Third in a Series, 1988, pp. 261–268.

[48] Semmlow J.L., and Griffel B.
Biosignal and medical image processing, third edition . CRC Press, 2014

[49] Armato S.G., Giger M.L., and MacMahon H.
*Automated lung segmentation in digitized posteroanterior chest radiographs*cad. Radiol., 1998, 5, (4), pp. 245–255

[50] Li L., Zheng Y., and Kallergi M., et al
*Improved method for automatic identification of lung regions on chest radiographs*Acad. Radiol., 2001, 8, (7), pp. 629–638

[51] Duryea J., and Boone J.M.
*A fully automated algorithm for the segmentation of lung fields on digital chest radiographic images*ed. Phys., 1995, 22, (2), pp. 183–191

[52] Iakovidis D.K., and Papamichalis G.:
*Automatic segmentation of the lung fields in portable chest radiographs based on Bézier interpolation of salient control points*ST 2008 – IEEE Workshop on Imaging Systems and Techniques Proc., 2008, vol. 217139, pp. 82–87

[53] Pietka E.:
*Lung segmentation in digital radiographs*J. Digit. Imaging, 1994, 7, (2), pp. 79–84

[54] Xu X.W., and Doi K
*Image feature analysis for computer-aided diagnosis: detection of right and left hemidiaphragm edges and delineation of lung field in chest radiographs*Biomed. Eng.Online, 2015, 14, p. 20

[55] Otsu N.:
A threshold selection method from gray-level histograms Automatica 1975, 11, (285–296), pp. 23–27

[56] C Xu, DL Pham and JL Prince
Medical Image Segmentation Using Deformable Models: Medical Image Processing and Analysis Handbook of Medical Imaging, volume 2, chapter

[57] Stanley Osher and Ronald P.Fedkiw
Level Set Methods: An Overview and Some Recent Results Journal of Computational Physics Volume 169, Issue 2, 2001, Pages 463-502

[58] Lee W.-L., Chang K., and Hsieh K.-S
Unsupervised segmentation of lung fields in chest radiographs using multiresolution fractal feature vector and deformable models Med. Biol. Eng. Comput., 2015, pp. 1–14

[59] Candemir S., Antani S., and Jaeger S., et al
Lung boundary detection in pediatric chest x-rays PIE Medical Imaging, 2015, p. 94180Q

[60] M. Kass, A.Witkin, and D. Terzopoulos
Snakes: Active Contour Models International Journal of Computer Vision, 321-331 (1987) KIuwer Academic

[61] Van Ginneken B., Frangi A.F., and Staal J.J., et al
Active shape model segmentation with optimal features IEEE Trans. Med. Imaging, 2002, 21, (8), pp. 924–933

[62] Cootes T.F., Taylor C.J., and Cooper D.H. et al.
Active shape models-their training and application Comput. Vis. Image Underst., 1995, 61, (1), pp. 38–59

[63] Yuan X., Giritharan B., and Oh J.H.
Gradient vector flow driven active shape for image segmentation 2007 IEEE Int. Conf. Multimedia and Expo, 2007, pp. 2058–2061

[64] Xu T., Mandal M., and Long R., et al
*Gradient vector flow based active shape model for lung field segmentation in chest radiographs*Conf. Proc.:. . . Annual Int. Conf. IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society Annual Conf., 2009, vol. 2009, p. 3561

[65] Iakovidis D.K., and Savelonas M
*Active shape model aided by selective thresholding for lung field segmentation in chest radiographs*Final Program and Abstract Book – Ninth Int. Conf. Information Technology and Applications in Biomedicine, ITAB 2009, 5–7 November 2009

[66] Xu T., Mandal M., and Long R et al.
*An edge-region force guided active shape approach for automatic lung field detection in chest radiographs*Comput. Med. Imaging Graph., 2012, 36, (6), pp. 452–463

[67] Lee J.-S., Wu H.-H., and Yuan M.-Z
Lung segmentation for chest radiograph by using adaptive active shape models Biomed. Eng. Appl. Basis Commun., 2010, 22, (02), pp. 149–156

[68] Wu G., Zhang X., and Luo S., et al.
Lung segmentation based on customized active shape model from digital radiography chest images J. Med. Imaging Health Inf., 2015, 5, (2), pp. 184–191

[69] Shi Y., Qi F., and Xue Z., et al.
Segmenting lung fields in serial chest radiographs using both population-based and patient-specific shape statistics IEEE Trans. Med. Imaging, 2008, 27, (4), pp. 481–494

[70] Dawoud A.:
Lung segmentation in chest radiographs by fusing shape information in iterative thresholding', IET Comput. Vis., 2011, 5, (3), p. 185

[71] Zhang G., Cong L., and Wang L., et al.:
Lung fields segmentation algorithm in chest radiography IET 2014, pp. 137–144

[72] Li X., Luo S., and Hu Q., et al:
*Automatic lung field segmentation in x-ray radiographs using statistical shape and appearance models*J. Med. Imaging Health Inf., 2016, 6, (2), pp. 338–348

[73] Shao Y., Gao Y., and Guo Y:
*Hierarchical lung field segmentation with joint shape and appearance sparse learning*IEEE Trans. Med. Imaging, 2014, 33, (9), pp. 1761–1780

[74] Viola P., and Jones M.
*Robust real-time face detection*IEEE Trans. Med. Imaging, Int. J. Comput. Vis., 2004, 57, (2), pp. 137–154

[75] Wang S., and Summers R.M.
Machine learning and radiology Med. Image Anal., 2012, 16, (5), pp. 933–951

[76] Kevin Zhou S., Zhang J., and Zheng Y.
Discriminative learning for anatomical structure detection and segmentation Springer US, Boston, MA, 2012), pp. 273–306

[77] McNitt-Gray M.F., Sayre J.W., and Huang H.K., et al.
A pattern classification approach to segmentation of chest radiographs Proc. Spie 1898, 1993, vol. 1898, pp. 160–170

[78] McNitt-Gray M.F., Huang H.K., and Sayre J.W.
Feature selection in the pattern classification problem of digital chest radiograph segmentation IEEE Trans. Med. Imaging, 1995, 14, (3), pp. 537–547

[79] Tsujii O., Freedman M.T., and Mun S.K.
Automated segmentation of anatomic regions in chest radiographs using an adaptive-sized hybrid neural network Med. Phys., 1998, 25, (6), pp. 998–1007

[80] Vittitoe N.F., Vargas-Voracek R., and Floyd C.E. Jr.
Identification of lung regions in chest radiographs using Markov random field modeling Med. Phys., 1998, 25, (6), pp. 976–985

[81] Aristeidis Diplaros 1, Nikos Vlassis, Theo Gevers,
A spatially constrained generative model and an em algorithm for image segmentation IEEE Trans. Neural Netw., 2007, 18, (3), pp. 798–808

[82] van Ginneken B., and ter Haar Romeny B.M,
Automatic segmentation of lung fields in chest radiographs Med. Phys., 2000, 27, (10), pp. 2445–2455

[83] Arya S., and Mount D.M.,
Approximate nearest neighbor queries in fixed dimensions ODA, 1993, vol. 93, pp. 271–280

[84] Shi Z., Zhou P., and He L., et al:
Lung segmentation in chest radiographs by means of Gaussian kernel-based FCM with spatial constraints Sixth Int. Conf. Fuzzy Systems and Knowledge Discovery, 2009 FSKD'09, 2009, vol. 3, pp. 428–432

[85] Bram van Ginneken 1, Mikkel B Stegmann, Marco Loog
Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database Med Image Anal . 2006 Feb;10(1):19-40

[86] Sema Candemir, Stefan Jaeger, Kannappan Palaniappan, Jonathan P Musco, Rahul K Singh et al.
Lung segmentation in chest radiographs using anatomical atlases with nonrigid registration IEEE Trans Med Imaging . 2014 Feb;33(2):577-90.

[87] T. A. Ngo and G. Carneiro ,
Lung segmentation in chest radiographs using distance regularized level set and deep-structured learning and inference 2015 IEEE International Conference on Image Processing (ICIP), 2015, pp. 2140-2143

[88] Wan Siti Halimatul Munirah Wan Ahmad, W Mimi Diyana W Zaki, Mohammad Faizal Ahmad Fauzi,
Lung segmentation on standard and mobile chest radiographs using oriented Gaussian derivatives filter Biomedical engineering online , Volume 1 , p.1-26 ,2015

[89] LeCun, Y., Bengio, Y., Hinton, G.
Deep learning Nature, 521 , 436–444

[90] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun
Deep Residual Learning for Image Recognition arXiv preprint 2015

[91] Feifei Xiong , Jon Philip Barker
Deep Learning of Articulatory-Based Representations and Applications for Improving Dysarthric Speech Recognition Conference: ITG Conference on Speech CommunicationAt: Oldenburg, Germany

[92] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova.
Deep Learning of Articulatory-Based Representations and Applications for Improving Dysarthric Speech Recognition Conference: ITG Conference on Speech CommunicationAt: Oldenburg, Germany

[93] Yann LeCun, Leon Bottou, Yoshua Bengio, and Patrick Haffner.
Gradient-based learning applied to document recognition Proceedings of the IEEE, 86(11):2278–2324, 1998.

[94] A. Krizhevsky, I. Sutskever, and G. Hinton
ImageNet classification with deep convolutional neural networks In NIPS, 2012.

[95] K. Simonyan and A. Zisserman
Very deep convolutional networks for large-scale image CoRR, abs/1409.1556, 2014

[96] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich.
Going deeper with convolutions CoRR, abs/1407.1556, 2014

[97] Schmidhuber J.
Deep learning in neural networks: an overview Neural Netw. 2015 Jan;61:85-117. doi: 10.1016/j.neunet.2014.09.003. Epub 2014 Oct 13

[98] Roth H.R., Lu L., and Liu J., et al
Improving computer-aided detection using convolutional neural networks and random view aggregation IEEE Trans. Med. Imaging, 2016, 35, (5), pp. 1170–1181

[99] Pereira S., Pinto A., and Alves V., et al
Brain tumor segmentation using convolutional neural networks in MRI images IEEE Trans. Med. Imaging, 2016, 35, (5), pp. 1240–1251

[100] Havaei M., Davy A., and Warde-Farley D., et al
Brain tumor segmentation with convolutional neural networks in MRI images CoRR, 2015, abs/1505.03540

[101] Roth H.R., Lu L., and Farag A., et al.
Deeporgan: multi-level deep convolutional networks for automated pancreas segmentation CoRR, 2015, abs/1506.06448

[102] Cireşan D.C., Giusti A., and Gambardella L.M., et a
Mitosis detection in breast cancer histology images with deep neural networks Int. Conf. Medical Image Computing and Computer-assisted Intervention, 2013, pp. 411–418

[103] Chen H., Dou Q., and Wang X., et al
Mitosis detection in breast cancer histology images via deep cascaded networks 30h AAAI Conf. Artificial Intelligence, 2016

[104] Prasoon A., Petersen K., and Igel C., et al.
Deep feature learning for knee cartilage segmentation using a triplanar convolutional neural network Med. Image Comput. Comput. Assist. Int., 2013, 16, (Pt 2), pp. 246–253

[105] Y. Bar, I. Diamant, L. Wolf, S. Lieberman, E. Konen and H. Greenspan
Chest pathology detection using deep learning with non-medical training 2015 IEEE 12th International Symposium on Biomedical Imaging (ISBI), 2015, pp. 294-297

[106] Gao M, Bagci U, Lu L, Wu A, Buty M, Shin HC, Roth H, Papadakis GZ, Depeursinge A, Summers RM, Xu Z, Mollura DJ
Holistic classification of CT attenuation patterns for interstitial lung diseases via deep convolutional neural networks Comput Methods Biomech Biomed Eng Imaging Vis. 2018;6(1):1-6

[107] B. Zhao, Jiashi Feng and S. Yan
A survey on deep learning-based fine-grained object classification and semantic segmentations Computer Science International Journal of Automation and Computing,2017

[108] A. A. Novikov, D. Lenis, D. Major, J. Hladůvka, M. Wimmer and K. Bühler,
*Fully Convolutional Architectures for Multiclass Segmentation in Chest Radiographs*IEEE Transactions on Medical Imaging, vol. 37, no. 8, pp. 1865-1876, Aug. 2018, doi: 10.1109/TMI.2018.2806086.

[109] Akira Hasegawa, Shih-Chung Benedict Lo, Matthew T. Freedman M.D., Seong Ki Mun
*Convolution neural-network-based detection of lung structures*1994

[110] Kalinovsky, A. Kovalev, V
*Lung image Segmentation using deep learning methods and convolutional neural networks*Publishing Center of, BSU,2016

[111] Vijay Badrinarayanan, Alex Kendall, Roberto Cipolla,
*SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation*University of Cambridge , 2016

[112] Jonathan Long, Evan Shelhamer, Trevor Darrell
*Fully Convolutional Networks for Semantic Segmentation*Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2015, pp. 3431-3440

[113] Olaf Ronneberger,Philipp Fischer and Thomas Brox
*U-Net: Convolutional Networks for Biomedical Image Segmentation*MICCAI 2015: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015 pp 234-24

[114] OlWei Dai, Joseph Doyle, Xiaodan Liang, Hao Zhang, Nanqing Dong, Yuan Li, Eric P. Xing
*SCAN: Structure Correcting Adversarial Network for Organ Segmentation in Chest X-rays*Petuum Inc. fwei.dai,joe.doyle,xiaodan.,hao.zhang,nanqing.dong,christy.li,eric.xingg@petuum.com

[115] Andrea Aprovitola and Luigi Gallo
Knee bone segmentation from MRI: A classification and literature review Biocybernetics and Biomedical Engineering Volume 36, Issue 2, 2016, Pages 437-449

[116] Joshua B. Tenenbaum,Vin de Silva and John C. Langford
A Global Geometric Framework for Nonlinear Dimensionality Reduction SCIENCE VOL 290 22 DECEMBER 2000

[117] S. T. Roweis and L. K. Saul,
Nonlinear Dimensionality Reduction by Locally Linear Embedding Science Vol 290, 22 December 2000, 2323–2326.

[118] Belkin, Mikhail; Niyogi, Partha (2001).
Laplacian Eigenmaps and Spectral Techniques for Embedding and Clustering. Advances in Neural Information Processing Systems. MIT Press. 14: 586–691.

[119] G. Zheng; S. Li; G. Szekely (2017).
Statistical Shape and Deformation Analysis Academic Press. ISBN 9780128104941.

[120] Dryden, Ian L. ; Mardia, Kanti V Chichester etc
Statistical Shape Analysis : Wiley, 1998

[121] Cootes T.F.Taylor ,C.J.Cooper and D.H.Graham.
Active Shape Models-Their Training and Application* Computer Vision and Image Understanding Volume 61, Issue 1, January 1995, Pages 38-59

[122] V. Blanz and T. Vetter.
A morphable model for the synthesis of 3D faces. In Alyn Rockwood, editor, Siggraph 1999, Computer Graphics Proceedings,

[123] Mikkel B Stegmann, Karl Sjöstrand and Rasmus Larsen.
Sparse modeling of landmark and texture variability using the orthomax criterion In Medical Imaging, pages 61441G–61441G. International Society for Optics and Photonics, 2006. 39

[124] K Sjostrand, Egill Rostrup, Charlotte Ryberg, Rasmus Larsen, Colin Studholme, Hansjoerg Baezner, Jose Ferro, Franz Fazekas, Leonardo Pantoni, Domenico Inzitariet al.
Sparse decomposition and modeling of anatomical shape variation. IEEE Transactions on Medical Imaging, vol. 26, no. 12, pages 1625– 1635, 2007. 39

[125] Aapo Hyvärinen and Erkki Oja.
Independent component analysis: algorithms and applications. Neural networks, vol. 13, no. 4, pages 411–430, 2000.

[126] Timothy F Cootes and Christopher J Taylor.
Data driven refinement of active shape model search In British Machine Vison Conference, pages 383–392, 1996. 40

[127] Timothy F. Cootes, Gareth J. Edwards and Christopher J. Taylor.
Active appearance models IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 23, no. 6, pages 681–685, 2001. 7, 15, 38

[128] D. Cremers, M. Rousson and R. Deriche.
A review of statistical approaches to level set segmentation: integrating color, texture, motion and shape International Journal of Computer Vision, vol. 72, no. 2, pages 195–215, 2007. 8, 38, 39

[129] Dimitris Metaxas and Demetri Terzopoulos. *Shape and nonrigid motion estimation through physics-based synthesis.* IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 15, no. 6, pages 580–591, 1993. 38

[130] Stephen M. Pizer, Daniel S. Fritsch, Paul A. Yushkevich, Valen E. Johnson and Edward L. Chaney. *Segmentation, registration, and measurement of shape variation via image object shape.* IEEE Transactions on Medical Imaging, vol. 18, no. 10, pages 851–865, 1999. 3

[131] Lawrence H Staib and James S Duncan. *Model-based deformable surface finding for medical images* IEEE Transactions on Medical Imaging, vol. 15, no. 5, pages 720–731, 1996. 8, 38, 39

[132] Ian T Jolliffe. *Principal component analysis.* Springer, 2002

[133] M. Rousson and N. Paragios. *Shape priors for level set representations* pages 416–418. Springer, 2002. 38, 39

[134] Timothy F Cootes and Christopher J Taylor. *A mixture model for representing shape variation.* Image and Vision Computing, vol. 17, no. 8, pages 567–573, 1999. 38, 40

[135] D. Cremers, S.J. Osher and S. Soatto. *Kernel density estimation and intrinsic alignment for shape priors in level set segmentation.* International Journal of Computer Vision, vol. 69, no. 3, pages 335–351, 2006. 38, 69

[136] Gábor Székely, András Kelemen, Christian Brechbühler and Guido Gerig. *Segmentation of 2-D and 3-D objects from MRI volume data using constrained elastic deformations of flexible Fourier contour and surface models* Medical Image Analysis, vol. 1, no. 1, pages 19–34, 1996. 38

[137]] Lawrence H Staib and James S Duncan. *Model-based deformable surface finding for medical images* IEEE Transactions on Medical Imaging, vol. 15, no. 5, pages 720–731, 1996. 8, 38, 39

[138] András Kelemen, Gábor Székely and Guido Gerig. *Elastic model-based segmentation of 3-D neuroradiological data sets Medical Imaging* IEEE Transactions on, vol. 18, no. 10, pages 828–839, 1999. 39

[139] Christos Davatzikos, Xiaodong Tao and Dinggang Shen. *Hierarchical active shape models, using the wavelet transform.* Medical Imaging, IEEE Transactions on, vol. 22, no. 3, pages 414–423, 2003.

[140] Delphine Nain, Steven Haker, Aaron Bobick and Allen Tannenbaum. *Multiscale 3-d shape representation and segmentation using spherical wavelets.* IEEE Transactions on Medical Imaging, vol. 26, no. 4, pages 598–618, 2007

[141] Peng Yu, P Ellen Grant, Yuan Qi, Xiao Han, Florent Ségonne, Rudolph Pienaar, Evelina Busa, Jenni Pacheco, Nikos Makris, Randy L Buckneret al. Cortical surface shape analysis based on spherical wavelets. IEEE Transactions on Medical Imaging, vol. 26, no. 4, pages 582–597, 2007. 39

[142] M.E. Leventon, W.E.L. Grimson and O. Faugeras. Statistical shape influence in geodesic active contours. In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), volume 1, pages 316–323. IEEE, 2000. 39

[143] Andy Tsai, Anthony Yezzi Jr, William Wells III, Clare Tempany, Dewey Tucker, Ayres Fan, W Eric Grimson and Alan Willsky. Model-based curve evolution technique for image segmentation In IEEE Conference on Computer Vision and Pattern Recognition (CVPR), volume 1, pages I–463. IEEE, 2001. 39

[144] Bernhard Schölkopf, Alexander Smola and Klaus-Robert Müller. *Nonlinear component analysis as a kernel eigenvalue problem. Neural computation*, vol. 10, no. 5, pages 1299–1319, 1998. 40

[145] Carole J Twining and Christopher J Taylor. *Kernel principal component analysis and the construction of non-linear active shape models.* In British Machine Vision Conference (BMVC), volume 1, pages 23–32, 2001. 40

[146] Patrick Etyngier, Florent Segonne and Renaud Keriven. *Shape priors using manifold learning techniques* In IEEE International Conference on Computer Vision (ICCV), pages 1–8. IEEE, 2007. 38, 40

[147] Pandia Rajan Jeyaraj Mepco Schlenk S.Edward Rajan and Mepco Schlenk *Deep Boltzmann Machine Algorithm for Accurate Medical Image Analysis for Classification of Cancerous Region* August 2019 Cognitive Computation and Systems 1(3) DOI:

[148] Mohammad Tofighi, Tiantong Guo, Jairam K. P. Vanamala Vishal Monga
Prior Information Guided Regularized Deep Learning for Cell Nucleus Detection January 2019IEEE Transactions on Medical Imaging PP(99):1-1 DOI:10.1109/TMI.2019.2895318 Project: Deep networks with shape priors for nucleus detection

[149] Zahra Mirikharaji, Ghassan Hamarneh
Star Shape Prior in Fully Convolutional Networks for Skin Lesion Segmentation Zahra Mirikharaji, Ghassan Hamarneh MICCAI 2018

[150] Ozan Oktay et al.
Anatomically Constrained Neural Networks (ACNN): Application to Cardiac Image Enhancement and Segmentation In: CoRR (2017).

[151] Agostina J. Larrazabal, Cesar Martinez, Enzo Ferrante
Anatomical Priors for Image Segmentation via Post-Processing with Denoising Autoencoders Agostina J. Larrazabal, Cesar Martinez, Enzo Ferrante MICCAI 2019

[152] S. M. Masudur Rahman Al Arif, Karen Knapp, Greg Slabaugh SPNet: Shape Prediction Using a Fully Convolutional Neural Network Publisher: Springer International Publishing 2018

[153] Quadros, L.S., R. Palanichamy, and A.S. D'souza,
Variations in the lobes and fissures of lungs-a study in South Indian lung specimens Eur J Anat, 2014. 18(1): p. 16-20.

[154] Aziz A., Ashizawa K., Nagaoki K., Hayashi K.
High resolution CT anatomy of the pulmonary fissures J. Thorac. Imaging 19: 186-191.

[155] Aldur, M., et al.,
An accessory fissure in the lower lobe of the right lung. Morphologie: Bulletin de l'Association des anatomistes, 1997. 81(252): p. 5-7.

[156] El-Baz, A., et al.
*3D shape analysis for early diagnosis of malignant lung nodules*in Biennial International Conference on Information Processing in Medical Imaging. 2011. Springer.

[157] Meenakshi, S., K. Manjunath, and V. Balasubramanyam,
Morphological variations of the lung fissures and lobes Indian Journal of Chest Diseases and Allied Sciences, 2004. 46: p. 179-182.

[158] Mahyar Osanlouy, Alys R. Clark, Haribalan Kumar, Clair King, Margaret L. Wilsher, David G. Milne, Ken Whyte, Eric A. Hoffman GEP ita
Lung and fissure shape is associated with age in healthy never-smoking adults aged 20–90 years Mahyar Osanlouy, Alys R. Clark, Haribalan Kumar, Clair King, Margaret L. Wilsher, David G. Milne, Ken Whyte, Eric A. Hoffman GEP ita

[159] Ayman El-Baz, Garth M. Beache,2 Georgy Gimel'farb, Kenji Suzuki ,and Kazunori
Lung Imaging Data Analysis
A Generic Approach to Lung Field Segmentation from Chest

[160] Awais Mansoor Juan J. Cerrolaza, Geovanny Perez, Elijah Biggs, Kazunori Okada, Gustavo Nino, Marius George Linguraru.
Chest Radiographs using Deep Space and Shape Learning .

[161] Zhu N, Zhang D, Wang W et al.
A novel coronavirus from patients with pneumonia in Chinaé, 2019. N Engl J Med 382:727–733. https://doi.org/10.1056/NEJMoa2001017

[162] Tobias Heimann 1, Hans-Peter Meinzer
Statistical shape models for 3D medical image segmentation: a review, 2009 Aug;13(4):543-63. doi: 10.1016/j.media.2009.05.004. Epub 2009 May 27.

[163] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton
ImageNet Classification with Deep Convolutional Neural Networks , 2013 University of Toronto

[164] Fausto Milletari, Alex Rothberg, Jimmy Jia, Michal Sofka
Integrating statistical prior knowledge into convolutional neural networks , Catalyzer Corporation 2017

[165] Adrian V. Dalca, Guha Balakrishnan, John Guttag, Mert R. Sabuncu
Unsupervised Learning for Fast Probabilistic Diffeomorphic Registration, MICCAI 2018. Proceedings: LNCS 11070, pp 729-73

[166] Xuechen Li,Suhuai Luo,Qingmao Hu and Jiaming Li
*Automatic Lung Field Segmentation in X-ray Radiographs Using Statistical Shape and Appearance Models*April 2016Journal of Medical Imaging and Health Informatics 6(2):338-348

[167] UK Lopes and J F Valiati
Pre-trained convolutional neural networks as feature extractors for tuberculosis detection Computers in Biology and Medicine . 2017 Oct 1;89:135-143.

[168] Yu.Gordienko, Yu.Kochura, O.Alienin, O. Rokovyi, S. Stirenko, Peng Gang, Jiang Hui, Wei Zeng
Dimensionality reduction in deep learning for chest X-ray analysis of lung cancer Conference: 2018 Tenth International Conference on Advanced Computational Intelligence

[169] Megumi Nakao, Junko Tokuno, Toyofumi Chen-Yoshikawa, Hiroshi Date Tetsuya Matsuda
Surface deformation analysis of collapsed lungs using model-based shape matching International Journal of Computer Assisted Radiology and Surgery volume 14, pages1763–1774 (2019)

[170] Johnatan Carvalho Souza, João Otávio Bandeira Diniz, Jonnison Lima Ferreira, Giovanni Lucca and França da Silva,
An automatic method for lung segmentation and reconstruction in chest X-ray using deep neural networks Computer Methods and Programs in Biomedicine 177:285-296 2019

[171] Gusztáv Gaál, Balázs Maga, András Lukács,
Attention U-Net Based Adversarial Architectures for Chest X-ray Lung Segmentation Computer Methods and Programs in Biomedicine 177:285-296 2019

[172] Muhammad Usman1 Byoung-Dai Lee,Shi Sub Byon1 , Sung Hyun Kim1 and Byung-il Lee
Volumetric Lung Nodule Segmentation using Adaptive ROI with Multi-view Residual Learning
Center for Artificial Intelligence in Medicine and Imaging 2019

[173] J. Shiraishi, S. Katsuragawa, J. Ikezoe, T. Matsumoto, T. Kobayashi, K. Komatsu, M. Matsui, H. Fujita, Y. Kodera, and K. Doi.
*Development of a digital image database for chest radiographs with and without a lung nodule: receiver operating characteristic analysis of radiologists*AJR, 174:71–74, 2000.

[174] Stefan Jaeger,corresponding author1 Sema Candemir,1 Sameer Antani,1 Yì-Xiáng J. Wáng,2 Pu-Xuan Lu,3 and George Thoma1
*Two public chest X-ray datasets for computer-aided screening of pulmonary diseases*Quant Imaging Med Surg. 2014 Dec; 4(6): 475–477.

[175] B. van Ginneken, M.B. Stegmann, M. Loog,
"Segmentation of anatomical structures in chest radiographs using supervised methods: a comparative study on a public database", Medical Image Analysis, 2006, vol. 10, pp. 19-40

[176] Tobias Heimann, Bram van Ginneken, and Martin Styner.
"SLIVER07 : Home." www.sliver07.org. Retrieved 2018-12-24

[177] Theophraste Henry Alexandre , Marvin Lerousseau Theo Estienne, Charlotte Robert, Nikos Paragios , and Eric Deutsch,
Brain tumor segmentation with self-ensembled, deeply-supervised 3D U-net neural networks: a BraTS 2020 challenge solution. BraTS 2020 proceedings (LNCS)

[178] Jonathan Long, Evan Shelhamer, Trevor Darrell
Fully Convolutional Networks for Semantic Segmentation IEEE Conference on Computer Vision and Pattern Recognition (CVPR) 2015

[179] Kilian Pohl
Logarithm Odds Maps for Shape Representation Conference: Proceedings of the 9th international conference on Medical Image Computing and Computer-Assisted Intervention - Volume Part II 2012

[180] *Prediction Framework for Statistical Respiratory Motion Modeling* Philips Research Europe - Hamburg, Germany,
Institut fur Informationsverarbeitung, Leibniz University of Hannover, Germany 2013

[181] Shusil Dangi
A distance map regularized CNN for cardiac cine MR image segmentation Medical Physics Journal February 2019

[182] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, Ruslan Salakhutdinov
Dropout: A Simple Way to Prevent Neural Networks from Overfitting 15(56):19291958, 2014.

[183] Sergey Ioffe, Christian Szegedy
Batch Normalization: Accelerating Deep Network Training by Reducing Internal Covariate Shift Proceedings of the 32nd International Conference on Machine Learning, PMLR 37:448-456, 2015.

[184] Kohler, Jonas; Daneshmand, Hadi; Lucchi, Aurelien; Zhou, Ming; Neymeyr, Klaus; Hofmann, Thomas
Exponential convergence rates for Batch Normalization: The power of length-direction decoupling in non-convex optimization arXiv:1805.10694 2018

[185] Ian Goodfellow, Yoshua Bengio and Aaron Courville.
Deep Learning MIT Press book 2018

[186] Leslie N. Smith
Cyclical Learning Rates for Training Neural Networks Presented at WACV 2017.

[187] Sebastian Ruder
An overview of gradient descent optimization algorithms Insight Centre for Data Analytics, NUI Galway Aylien Ltd., Dublin 2017

[188] Syed Talha Bukhari and Hassan Mohy-ud-Din
A systematic evaluation of learning rate policies in training CNNs for brain tumor segmentation Institute of Physics and Engineering in Medicine, Physics in Medicine Biology, Volume 66, Number 10, 2021

[189] Sebastian Mika, Bernhard Scholkopf, Alex Smola ¨ Klaus-Robert M ¨uller, Matthias Scholz, Gunnar Ratsch
Kernel PCA and De-Noising in Feature Space GMD FIRST, Rudower Chaussee 5, 12489 Berlin, Germany 2013

[190] Yoshua Bengio
Practical recommendations for gradient-based training of deep architectures Machine Learning 2012