POLITECNICO DI TORINO

Master's Degreer in Biomedical Engineering

Master's Degree Thesis

Segmenting Breast Regions in Thermal Images Exploiting Deep Learning-Based Algorithms

Supervisors

Prof. Agostini VALENTINA

Prof. Faraci FRANCESCA DALIA

Dr. Fiorillo LUIGI

Candidate

Riccardo SANTARELLI

2026

Summary

Abstract

Thermography is an imaging technique harmless, low-cost and holding promise in detecting breast cancer at early stages. However, only with modern advances in infrared cameras and computational analysis, as machine learning based detection systems, it has achieved a sufficient accuracy to fulfill the role of adjunctive tool for breast abnormality detection. The automated segmentation of breast thermograms is crucial in improving the diagnostic utility of infrared thermography, as it precise localize regions of medical interest.
This project propose a modular framework to address multiclass segmentation, exploiting open-source codebase deep-learning networks, in a smooth pipeline. This project aims to automatic detect and segment left and right breast, left and right nipples. For this first iteration of this work, were deployed You Only Look at Once (YOLO) model, as fast ancor-free object detector, and Segment Anything Model (SAM), a promptable segmentator known for its strong zero-shot segmentation if assisted by a good prompting.
The proposed methods were evaluated on a new annotated small-sized database of 244 thermograms, provided by a private company. The workflow can be summarized in 3 step: assessment of YOLO and SAM separately, YOLO training as SAM prompt generator, assessment of SAM supported by YOLO.
Employing YOLO11 an overall mAP50 of 84.4±4.5% and an overall mAP5090 of 49.8±4.2% at the end of the training was achiveved. However with YOLO generating bounding boxes to point out anatomic regions, SAM has only partially demonstrated to be a powerful and precise segmentation tool, reaching, for the breast classes, a mean IoU of 84.7±3.9% and a Hausdorff distance of 22.8±4.8 on the proposed dataset. The obtained results align with those reported in previous similar studies.
Automatic identification of object and surface boundary of breast thermal images remains a difficult and challenging task. Future steps may focus on enhancing the performance, by the mean of preprocessing tecniques, change version or architecture of the models or expand the dataset.

Acknowledgements

It is difficult to write satisfactory acknowledgments for many reasons; first of all, so many people have contributed to shaping the person I am today, that it is impossible to list them all in just a page or so. Secondly, acknowledgments mark the conclusion —not only of this thesis, but of my entire university journey— and alongside the immense joy of reaching such an important milestone as graduation, there is always a lingering fear of losing something.

First and foremost, I would like to thank Professor Agostini, for accepting the role of supervisor for my thesis, and Professor Faraci, for her kindness and professionalism during these months of work. I would also like to sincerely thank Dr. Fiorillo Luigi, my co-supervisor, for the time he dedicated to me and for his willingness to clarify my doubts and guide me throughout the research process and the writing of this thesis.

I wholeheartedly thank my parents for always being by my side, supporting my choices and my dreams and, when necessary, reproaching me for my mistakes. Thank you for your advice and your criticism, which helped me grow. I hope that this achievement of mine may also be, as much as possible, a reward for you and for the sacrifices you have made.

Thank you to my grandparents —those who are still with us and those who are no longer here— because through their simplicity and affection, I have learned to recognize and appreciate the importance of the little things.
I thank my brother, who is my point of reference whenever I have doubts, especially when it comes to cooking.
Thank you to my sister, because I cannot imagine my life without her.

I would also like to thank the 'Gruppo Giovani del Don Bosco', which I consider a second family and which, just like my own family, I have never truly had the opportunity to thank, even though I have been part of it for many years now. In particular, I want to warmly thank Marcella Marcelli, who, like a second mother,

has supported and put up with me, both morally and materially. The support and the constant presence I received by this group gave me the strength I needed to overcome many difficult moments. My heartfelt thanks to you all.

Finally, I would like to thank the members of the film club "Absolute Cinema." Even though it was founded only recently, the shared moments we created together have truly shaped me. I could not have reached the end of this long and winding journey without you.

Table of Contents

Chapter 1

Introduction

1.1 Context of application

Cancer is a umbrella term used to describe a number of diseases associated with uncontrolled cell overgrowth resulting by tumor mutations. Breast cancer is the most commonly diagnosed cancer among women in western world [1]. Although the mortality rate in high-income countries dropped by 40% between the 1980s and 2020, due to improved access to effective treatment and cancer diagnosis [2], globally, has the second highest cancer-related mortality rate after skin cancer, with 2.3 million new cases and 670,000 women died from breast cancer in 2022 .

Early detection is essential to decrease mortality rates and improve survival rates. Mammography is the primary modality recommended for women over 40 years old by the Food and Drug Administration(FDA) and is widely used in wealthier countries [3]. For younger women, FDA entrust to clinical and self breast exam, that are manual exams performed by the clinician or the patient themselves respectively. Although they are very practical and easy to perform, breast tactile examinations have a poor ability to detect breast cancer, in particular early-stage ones. Self breast exam sensitivity is not even comparable to the sensitivity of mammography. Mammography screening test, in randomized controlled trials involving the general population, has been proven to reduce death rates [4]. Its strong performances in screening and diagnostic have led mammography to take the role of main routine screening tool since the late 1980s [5]. However, it is less effective for dense breasts, has a higher false positive rate [6], causes discomfort, and utilizes ionizing radiation which slightly increases cancer risk with repeated exposures [7] [8].

Since the presence of a broad spectrum of situations in which mammography is not applicable has, over the past decades, prompted the development and investigation of alternative methodologies, ables to become the new gold standards in this

uncovered niches. The one addressed in this thesis is thermography, a method first developed in the 1950s that has recently re-emerged as a focus of research due to advances in technology. Medical infrared imaging is the recording of temperature distribution of the patient's body using infrared radiation (IR) emitted by the skin surface at wavelengths between 0.8 μm and 1.0 μm.[9] Infrared thermography is a non-invasive, non-contact, passive, radiation-free technique that can reveals functional information binded to temperature changes.
It has been demonstrated that thermography can detect angiogenesis due to the increased demand of blood to supply in the new vessels and to the increased metabolic activity in cancerous tissue.
Only with modern advances in IR cameras and computational tools, as machine learning based detection systems, thermography has achieved a sufficient accuracy to fulfill the role as an effective tool for breast abnormality detection. These technological improvements were essential, as bare visual inspection alone lacks the precision required to reliably identify pathological patterns. In order to provide a useful tool to support the job of radiologists, a computer-aided diagnosis and detection (CAD) system is needed to analyze breast thermograms. Images analysis system, like this, usually consists 4 main steps: preprocessing of thermogram, regions of interest segmentation, extraction of features and classification.
Between the above-mentioned steps, this thesis project will focus in the image processing techniques for segmentation of the regions of interest (ROI). The extraction of features in ROI is necessary for an efficient use of machine learning classifier, which detects patterns with anomalies and without anomalies.
Thus, extract the relevant areas is considered critical, in order to provide sufficiently informative data for later stages.

Many authors have provided their own method to address the segmentation problem. This work uses a deep-learning based approach.

The motivation for research in this scope stems from the recognition that thermal imaging holds promise in detecting breast cancer early, and its success depend on the precision of the segmentation process. To optimize thermal imaging segmentation, is fundamental leveraging on advanced deep learning techniques.

1.2 Outline of the study

This thesis project, falling within the scope of breast cancer patients follow-up, have the aim to use pretrained deep learning models to address the segmentation of breast thermograms, without relying on resource-intensive tools. To achieve this goal a new pipeline is proposed, which involve the use of two computational

components: the Segment Anything Model and the You Only Look at Once model. The first is a powerful tool in zero-shot segmentation, but need the right prompt, while the second is an architecture for object detection. YOLO was chosen with the intention of being a proper support for SAM. The segmentation was performed on a novel dataset of only 244 images, consisting in frontal view thermograms of already diagnosed breast cancer patients.

The novelty and contributions of the study are summarized as follows:

- Implementing a newfound and promising segmentation methodology, in CAD system, that obviates intricate preprocessing procedures. Fine-tuning and validating the proposed approach on a new small dataset.

- Investigates the capacity of pre-trained CNN and Visual Trasformer architectures used in combination to simplify breast segmentation of thermal images.

The manuscript is structured in five chapters. The second chapter collects the foundational knowledge required for the understanding of subsequent presented work. It is organized into three main sections. The first section provides an overview of breast cancer, outlining its clinical and pathological aspects. The second section examines screening techniques, with particular emphasis on infrared thermography. The final section discusses computer vision methodologies for object detection and segmentation, which form the technical basis for the proposed approach.
The third chapter presents the design of the study, focusing on material and method. The first section details the dataset employed. The second section describes the models architectures implemented. Finally the third one discloses the evaluation metrics adopted to assess performance. Additionally, it provides a review of the most relevant and closely related studies in the existing literature.
The results and their discussion are the focus of the fourth chapter. Furthermore, a comparison with the experimental outcomes from the selected research in the previous chapter is conducted.
In the last chapter, conclusion and future scope to research close the manuscript.

Chapter 2

Background

2.1 Breast cancer

2.1.1 Physiopathology

Breast cancer is a disease in which abnormal epithelial breast cells grow out of control and form tumours. If left unchecked, the tumours can spread throughout the body and become fatal[10].

Inside a woman's breast are 15 to 20 mammary glands, or lobes. Each lobe is made of many smaller sections called lobules. Each lobule is connected to the nipples by a lactiferous duct, that drains milk and other secretions into the outside. Fibrous tissue and fat fill the spaces between the lobules and lactiferous ducts, in a structure called stroma. Breast cancer cells begin inside the milk ducts and/or the milk-producing lobules of the breast. The detailed structure of female breast is shown in Fig. 2.1

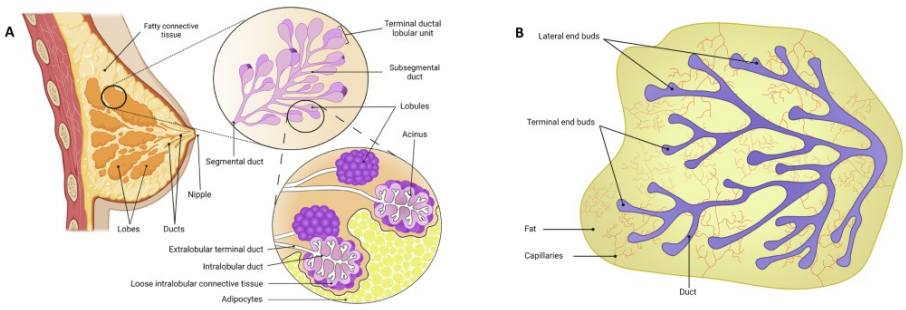

Figure 2.1: (A)Section of a breast, (B)Stroma

Breast cancer occurs when cells in the breast grow out of control and form a growth or tumor. Tumors may be cancerous (malignant) or not cancerous (benign), according to tumor area, growth rate and metastasis capacity [11]. The earliest form (in situ) is not life-threatening and can undergo to effective treatments [11]. In advance stage (invasive) cancer, cells can spread into nearby breast tissue. This creates tumours that cause lumps or thickening.

Invasive cancers can metastasize, spread to nearby lymph nodes or other organs. The major mechanisms that enable its progression include evasion of apoptosis, limitless capacity to divide, enhanced angiogenesis, resistance to anti-growth signals and induction of own growth signals, as well as the capacity to metastasize [12]. Metastasis can be life-threatening and fatal.

Like many other type of solid cancer, breast cancer staging is diagnosed by TNM staging. While cancer's grade describes the appearance of cancer cells and tissue compared to healthy ones, a cancer's stage account for the size and the spread of the primary tumor in the patient's body. TNM staging system allow to report crucial information in one of most coincise ways [13]. Each letter indicates a category: T is for the primary tumor, N is for regional lymph nodes and M is for distant metastasis. When a patient's cancer is staged with TNM, a number will follow each letter, that signifies the extent of the disease in each category. The higher the following number, the greater the severity of the disease[13].

There are many different types of breast cancer. Identify the right cancer types and subtypes is crucial to tailor the treatment to be as effective as possible with the fewest possible side effects. Nevertheless exist many ways to classify tumours, due to its heterogeneous nature. From a histologic point of view, the most common breast cancers, such as in situ ductal carcinoma (DCIS) and invasive carcinoma (IDC), are adenocarcinomas, since the cancers start in gland cells, in the milk ducts or lobules. According to many oncological textbook [14] [11] common types of breast cancer include:

- Invasive (infiltrating) ductal carcinoma (IDC): This cancer arises in milk ducts and spreads to basement membrane and nearby breast tissue. The toumorus cells show varying degrees of atypia, due to accumulation of mutations. It's the most common type of breast cancer, with an incidence percentage up to 70%.

- Lobular carcinoma: This breast cancer starts in the lobules and it develops in a "thread-like" shape rather than nodular one, typical of IDC. This results in tumors that typically remain clinically occult, escaping detection on mammography or physical examination until the disease becomes extensive. It's the second most common breast cancer in the United States.

- Ductal carcinoma in situ (DCIS): Like IDC, this breast cancer forms in epithelial cells of lactiferous ducts, but DCIS rarely spread beyond. DCIS is considered pre-invasive breast cancer, because cells can continue to undergo mutations, that may leading into invasive form.

Other kinds of cancers can grow in the breast, like angiosarcoma or sarcoma, but are not considered breast cancer, since they start in cells do not belong to the glandular tissue.

From a molecular level, breast cancers are also classified on the specific proteins they produce or active genes, called markers.
Estrogen receptors and progesterone receptors are the first biomarker checked, as they trigger metabolic pathways that stimulate the cancer cells to grow and divide. Cancers are called hormone receptor-positive or hormone receptor-negative based on whether or not they have these receptors.
Others important tumor markers is HER2/neu gene amplification, mutations and protein overexpression[15]. Human Epidermal growth factor Receptor 2 (HER2) is a protein that promotes normal cell growth, but in some cancers, includig breast cancer, an overabundance of HER2 can lead to faster cancer cell growth and spread. This type of tumor is indicated as HER2-positive. Breast cancer that doesn't have estrogen or progesterone receptors and also has low levels of the HER2 protein are called Triple-negative. These cancers tend to be more common in women younger than 40 years of age, who are Black, or who have a mutation in the BRCA1 gene [16].

Although several genetic mutations were reported to be highly associated with an increased risk of breast cancer, only in rare cases (5% to 10%) the disease occurs as part of a hereditary cancer susceptibility syndrome [17]. From this type are BRCA1 and BRCA2, mainly inherited in an autosomal dominant manner and characterized by a high penetrance (likelihood that a person with a specific gene will express the associated trait). Both BRCA1 and BRCA2 work to preserve chromosome structure, yet the precise nature of their contribution has proven difficult to define [18]. BRCA1 and BRCA2 germline mutation are primarily linked to the increased risk of breast carcinogenesis [19], moreover they predispose to contralateral breast, ovarian, and fallopian tube cancers as well [20].

After a biopsy is performed, breast cancer receptor cell status is tested with immunohistochemistry test. This thecnique uses antibodies to detect antigens in a tissue sample. The antibodies are usually linked to an enzyme or a fluorescent dye[15].

Treatment is based on the person, the type of cancer and its spread. Factors such as histology, stage, tumor markers, and genetic abnormalities guide individualized treatment decisions. Treatment can combine surgery, radiation therapy and medications.

2.1.2 Epidemiology

The annual report of World Health Organization indicates breast cancer as the most common invasive cancer in women worldwide, accounting for 30% of cancer cases in women[21] and confirming its top position since 2012. From the Global Cancer Observatory's biennal report, it is estimated 2.3 million women were diagnosed with breast cancer (11.7% of incidence), and about 680,000 died of the disease (6.9% of mortality) [1], making breast cancer the fourth leading cause of cancer death. The incidence of breast cancer is rising by around 3% per year from 2003 through 2021, as populations in many countries are getting older and screening programs getting capillar [21].

The incidence rate of breast cancer increases with age, from 1.5 cases per 100,000 in women aged 20 to 24 to a peak of 421.3 cases per 100,000 in women aged 75 to 79; 95% of new cases occur in women aged 40 years or older. The median age of women in 2022 of breast cancer diagnosis was 61 years.[22]
The 2020 GLOBOCAN data shows that age-standardized incidence rates (ASIR) of breast cancer are strongly and positively associated with the Human Development Index (HDI) [1]. In other words, wealthier countries women have higher prevalence of breast cancer than aesian, african or latinia counterpart. In the opposite direction, deaths due to breast cancer are more prevalently reported in transitioning countries (Melanesia, Western Africa, Micronesia/Polynesia, and the Caribbean) compared to the transitioned ones (Australia/New Zealand, Western Europe, Northern America, and Northern Europe). The difference, between the two economic categories, in incidence rate reaches approximately 88% [23].

A rapid increase in the incidence of breast cancer was first noted in western countries during the twentieth century's last decade, primarily due to increased screening, changes in reproductive patterns, and increased use of menopausal hormone therapy [24]. This trend persists until 2000, after which the incidence began to decline, especially in women younger than 50 years. With early detection and significant advances in treatment, breast cancer death rates have decreased over the past 25 years in North America and parts of Europe.[25]

Regarding our country, Italy is in line with the european average. According to data reported in the report "Cancer in Italy: Numbers 2024," published by the Italian Association of Tumor Registries (AIRTUM), the Italian Association

of Medical Oncology (AIOM), the AIOM Foundation, and PASSI, breast cancer remains the most common cancer in Italy. With over 53,000 new diagnoses in 2024, this cancer represented 30.3% of all cancers affecting women in Italy that year and 14.6% of all cancers diagnosed in the country [26].
The incidence, or number of new cases in a given period, of breast cancer is slightly increasing, especially among younger women. Despite this, the overall mortality rate for this cancer among adult women (aged 20 to 49) was 16.2% between 2006 and 2021, hence breast cancer remains the leading cause of cancer death in women (31%) [26]. The AIOM report confirms the results of previous year, on survival at 5 years from diagnosis (88%) and on the probability of living for a further 4 years conditional on having survived the first year after diagnosis (91%).

Identifying factors associated with an increased incidence of breast cancer development is important in general health screening for women. The number of risk factors of breast cancer is significant and includes both modifiable factors and non-modifiable ones. Risk factors for breast cancer are divided in reproductive-hormonal and lifesyle. They include: age, obesity, alcohol intake, family history of breast cancer, history of radiation exposure, reproductive history (such as age that menstrual periods began and age at first pregnancy), tobacco use and post-menopausal hormone therapy [27]. Approximately half of breast cancers develop in women who have no identifiable breast cancer risk factor other than gender (female) and age (over 40 years). Genetic factors cause 5% to 10% of all breast cancer cases, but may account for 25% of cases in women younger than 30 years [20]. BRCA1 and BRCA2 are the most important genes responsible for increased breast cancer susceptibility.

Breast cancer survival rates have been increasing and the number of people dying of breast cancer is steadily going down. In Italy, for instance, between 2007 and 2019, over 10,000 deaths related to this disease were avoided, a reduction corresponding to 6% [26]. Much of this is due to the widespread support for breast cancer awareness and funding for research.
Advances in breast cancer screening allow healthcare professionals to diagnose breast cancer earlier. Finding the cancer earlier makes it much more likely that the cancer can be cured. Even when breast cancer can't be cured, many treatments exist to extend life. New discoveries in breast cancer research are helping healthcare professionals choose the most effective treatment plans.

2.1.3 Screening techniques

Proper identification of breast abnormality, prior to the beginning of a cancerous growth, is the only effective way of reducing mortality due to breast cancer. Breast

cancer is diagnosed through physical examination, breast imaging, and tissue biopsy. Treatment options include surgery, chemotherapy, radiation, hormonal therapy, and, more recently, immunotherapy.

Mammography is considered the gold standard examination for cancer early detection, since the early 1960s, and is commonly used for screening and diagnosis [4] [5]. Mammography use X-ray radiation to obtain breast anatomical structures imaging. During the procedure, the breast is compressed using parallel-plate unit. This compression increase image quality by reducing the thickness of tissue that X-rays must penetrate, decreasing the amount of scattered radiation, reducing the required radiation dose, and preventing motion blur.

Screening mammography use low-dose X-ray compared to diagnostic mammography, which provides higher quality imaging with several views. Tumors can appear unusually dense within the breast, distort the shape of surrounding tissue, or cause small dense flecks called microcalcifications. Mammography is the ideal method for detecting DCIS and clusters of calcifications [28]. Carcinoma in situ was diagnosed in 78.9% and 68.4% of patients using mammography and MR mammography, respectively [29]. In randomized controlled trials involving the general population, the mammography screening test has been proven to reduce death rates[8]. Mammogram images have demonstrated to be technically more suitable for screening and, as a result, they can be employed for routine screening [30].

The sensitivity of mammography in the general populationis reside between 75% to 90%, while positive predictive value is at least 25% [5] [31]. The great difference between this two values indicates high tendency in false positive of mammography. It is estimated that, over 10 years of annual screening in the US, 1 in 2 women will have at least 1 false–positive mammogram result, and 1 in 5 women will have at least 1 false–positive clinical breast examination result [32]. In Europe, false positive mammography results are likely less common, as European countries report many fewer abnormal mammogram results and, on the whole, have similar breast cancer detection rates [33] [31]. Since only 61% of women who had a false-positive result that required a short-interval follow-up exam (repeat diagnostic mammogram within 6 months) and 67% of women who required a biopsy returned to routine screening [34], the high false positive rate constitute the great drawback of mammography.

Moreover mammography is of limited utility in many patients:

- ones with dense breasts, because on mammograms, both dense breast tissue and cancer appear white, making it difficult to distinguish between the two tissue types[35] [36]. Breast density can reduce mammography's detection rates up to 30% [37]. Greater density of breasts is observed in subject of

younger age and lower BMI, during pregnancy or the breastfeeding period, as well as during the intake of hormonal replacement therapy [38].

- in younger patients, due to breast density and the use of ionizing radiation. Younger tissue tend to be denser, thus a higher dose is required to obtain good image quality [36]. Moreover young women are more radiosensitive, due to high presence in their tissue of undifferentiated cells and hormones (estrogen increases radiation susceptiblty by 10%). Shifting of screening to the age range below 40 is undesirable, since it will lead to the increase of patients' radiation exposure.
- in those who cannot tolerate the breast compression that is required.

Breast ultrasound or magnetic resonance imaging (MRI) with contrast may be utilized in such cases.

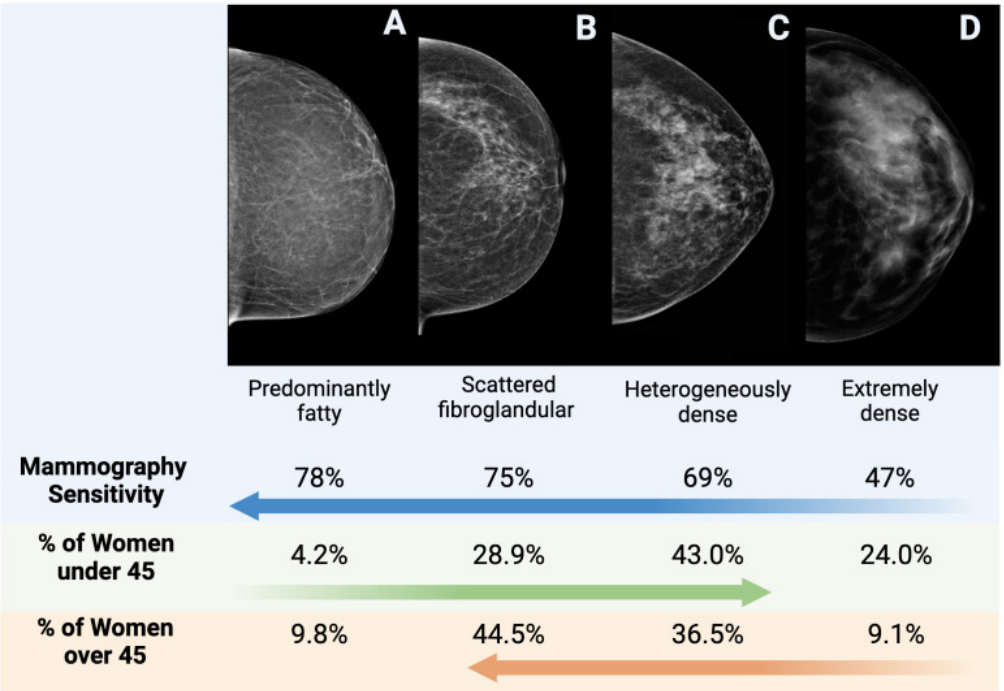

Figure 2.2: Comparison of BI-RADS classification to mammography sensitivity and percent of women under/over 45. The BI-RADS uses mammographic density for classification of breasts based on percent of fibroglandular tissue. The four categories are (A) predominantly fatty ($\leq 25\%$), (B) scattered fibroglandular (26-50%), (C) heterogeneously dense (51-75%), and (D) extremely dense (76-100%).

Breast imaging findings are classified in 7 categories by Breast Imaging Reporting

and Data System (BI-RADS), as show in following table Tab 2.1. This reporting system correlates imaging findings with their probability of underlying malignancy and recommends a broad treatment strategy [39]. To better understanding of mammography issues, figure Fig.2.2 has been added.

Category	Description	Recommendation
0	Incomplete	Additional imaging or comparison with prior exams is needed to render a final assessment
1	Negative	Symmetric and no masses, no architectural distortion, or no suspicious calcifications are present.
2	Benign	Findings are benign (e.g., simple cysts, fibroadenomas), with a 0% probability of malignancy.
3	Probably Benign	Findings have a very low likelihood of malignancy ($\leq$ 2%), and a short-interval follow-up is recommended.
4	Suspicious for Malignancy	Findings have a moderate to high suspicion for malignancy (2-95%), and biopsy is typically recommended.
5	Highly Suggestive of Malignancy	Findings have a very high probability of malignancy (>95%), and appropriate action (like biopsy) should be taken.
6	Proven Malignancy	The finding is confirmed to be cancer by biopsy, and further treatment (such as surgery) is recommended.

Table 2.1: BI-RADS classification system. Radiologists use it to describe results from breast imaging tests like ultrasound, mammography and MRI. They also use it to help determine next steps after an imaging test. The vast majority of screening mammograms fall into BI-RADS 1 or 2[40]. Screening mammograms with suspicious findings should generally be assigned BI-RADS 0 to indicate a callback for diagnostic evaluation, meaning additional views to confirm and further evaluate the finding.

Ultrasound differentiates the various types of tissue using reflecting index in the ultrasonic spectrum (usually between 2 and 12 MHz). By bouncing sound waves of the surface of the tissue and interpreting their reflection, it can be possible determine the boundaries between health tissue and cancer, assuming a structural distinction and an anatomical variation of the tumor from the surrounding breast tissue [41].

Ultrasounds is a non invasive harmless screening tecnique used in conjunction with mammography and clinical breast exam for supplemental screening in subsets of patients with dense breasts. This technique is normally used to further investigate suspicious areas of the breast found in the mammogram or during a breast exam [42]. It can also help distinguish between cysts (non-tumorous sacks filled with fluid) and solid masses.

Breast ultrasonography is similar in sensitivity to mammography and can be employed in image-guided biopsy. It was found that sensitivity of mammography declines with decreasing tumor size and increasing breast density, while ultrasound remained effective regardless of tumor size. However, the sensitivity of ultrasound declines in detecting non-palpable tumors such as microcalifications[43].

When used as a supplement to mammography, ultrasonography can improve sensitivity of screening at the expense of decreased specificity and increased biopsy rate [44]. Ultrasound is an attractive supplement to mammography because it is widely available, relatively inexpensive and does not inconvenience the patients.

On the opposite, MRI is time-consuming, has limited availability, and is expensive, but it can guarantee the highest imaging sensitive for a study. Magnetic resonance imaging is highly sensitive but slow, sometimes invasive (due to the intravenous administration of contrast agents), and its economic cost is even higher than that of mammography.

MRI uses a strong magnetic field along with pulsing radio waves to get a high resolution image of the breast at different crosssections. A contrast agent is added to help better image the breast. This procedure is used to screen women who are at a high risk of developing breast cancer or to better image tumors found in other tests [45]. This procedure is very expensive and time consuming and hence is only used as an adjunct to mammography for high risk asymptomatic and symptomatic women. Screening breast MRI has been found to be more sensitive but less specific than mammography for the detection of invasive breast cancers in high-risk women in both retrospective and prospective studies [46] [47].

Figure Fig.2.3 compares three mammography imaging technique.

For a better prognosis, minimizing the aggressiveness of the treatment and a decrease in mortality, early diagnosis assumes a decisive role. Thus breast thermography could be another breast disease screening tool. It is based on the principle that damaged breast tissues behave thermally different from healthy ones.

2.2 Infrared thermography for healtcare

Infrared thermography (IRT) is a imaging technique that measure the infrared radiation deriving from the objects'surface. Just like other radiation of the electromagnetic spectrum, IR has the emitted and reflected component. The first one, the thermal emission from the surface of the object, depends on its temperature and emissivity. While the reflected component depends on its reflectance and surrounding sources that irradiate the object.

Although IRT began to be employed in the medical research during the 60s, only after the recent advances in IR camera technology and calibration methods, it increasingly became a successful tool for studying the correlation between thermal physiology and skin temperature [48].
This technique use an infrared camera to detect the infrared heat energy that is emitted from the skin and create a visual map of the distribution of temperatures on the surface, called thermogram [49]. Between wavelengths of 2 and 14 μm, the emissivity of human skin is more or less constant at a value of 0.98 ± 0.01, which is close to that of a perfect black body [50]. Skin tone has negligible effect on skin's thermal emission.[51]
The black body behaviour allows to assert: skin surface temperature is uniquelly determined by the rate of heat exchange between the outside, which is easily controlled during mesurement, and the body core, which is held more or less at a constant temperature and acts as a thermal reservoir. In other words there is a direct connection between skin temperature and metabolic processes.

It is proved, IRT can usefully find appliction in diagnosis of breast cancer, diabetes neuropathy and peripheral vascular disorders [52].

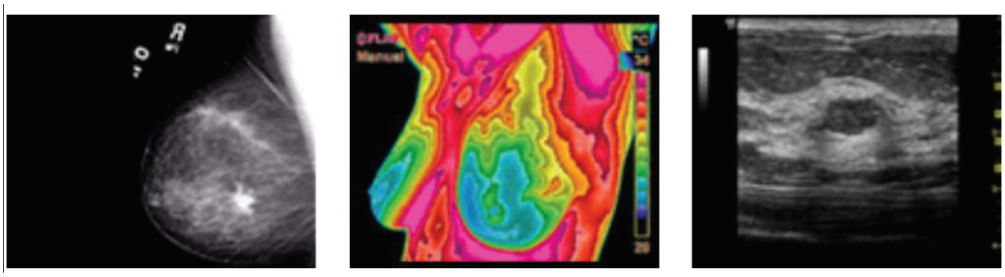

Figure 2.3: Example of imaging technique results. From the left to the right are rispectevly: traditional (x-ray) mammography, breast thermography (IRT) and ultrasound

2.2.1 Infrared thermography for breast cancer

Breast cancer is a heat sources identifiable by IRT.
Since tumors are clusters of cells which multiply in an uncontrolled manner, the metabolic heat generation rate and the blood perfusion rate of the tumor are, theoretically, higher than normal tissues. The increased tumor heat generation is dissipated to the surrounding tissue and, if not too deep, can be seen as a temperature spike at the surface of the body part; in our case, breast.

Although thermography has been around since the late 1950s, the biological rationales for thermal changes, indicating an underlying pathology, and the mechanisms of heat transfer in tumorous tissue have yet to be well described, due to the amount of components that take part and the complexity of their interactions. The main ones are:

- **Angiogenesis**
  Cancer stimulate an irregular development of blood vessels to deliver the necessary nutrients and oxygen to support its growth. Arised in a pathologic angiogenic situation, these disorganized network of blood vessels lack in flat muscle cells, making them unable to perform the vasoconstriction that would normally occur. Hypervascularization in the region leads an increase in local temperature [53].

- **Nitric oxide**
  Nitric oxide is an endothelium-derived relaxing factor, with strong vasodilatory proprieties [54], produced in high levels of ADP accumulation or hypoxia condition. These are the typical condition of tumourus extracellular matrix. Nitric oxide increas blood flow and therefore elevates the temperature [55]

- **Local Hyperestrogenism**
  Estrogen is highly involved in breast development, regulating ductal component, metabolic rate and fat deposition [56]. Estrogen, also, mediates vasodilation by increasing the local production of nitric oxide, therefore estrogen overproduction could result in vasodilation of the estrogen-sensitive tissues, leading to localized temperature changes [57]. Some estrogen metabolites have mutagenic proprieties.

In spite of numerous mechanisms underlying cancer heat, the temperature difference between healthy and diseased breasts is not always so evident(Contrary to figure Fig.4.6). Often, temperature variation is quite subtle in the early stages of malignancy, and changes may go unnoticed by direct visual interpretation of thermographies [58]. Considering how crucial early detection and timely intervention are in reducing breast cancer-related deaths, any false negative avoided by the

employment of the right equipment is heavly strategic.

Such small difference in radiation levels demand very sensitive detectors as well as standardized procedures [59]. The early cameras were cooled with liquid nitrogen ($-196\,^{\circ}\mathrm{C}$) to reduce noise, heavily limiting the tecnique for long time. Nowaday modern IR cameras can achieve sensitivities below 20 mK, thanks to the development of cooled quantum detectors.

Besides, normal breasts with increased vases are generally warmer and may be misinterpreted as abnormal. Therefore the subjective interpretation of thermographies often leads to false diagnoses, so intelligent detection systems are required [60]. For this reasons, despite the technological advances made in IR cameras, thermographic imaging largely remained qualitative in nature until recently [48], when computational tools, able to extract informations from images and automatic process them, emerge.

2.2.2 Thermic mammography

As previous discuss in the dedicated section, the most widely used tool for breast cancer detection is mammography, but it has some significant limitations including radiation exposure, cost, patient discomfort, and more importantly, a high false positive rate.

Breast infrared thermography is a noninvasive procedure that does not involve compression of the breast tissue or exposure to ionizing radiation, and able to assess breast physiological function, through high resolution temperature measurements of breast tissue. Moreover breast infrared thermography is suitable for women in all ages, including pregnant or nursing women, with all sizes and density of breast, with or without breast implants and fibrocystic breasts [61].

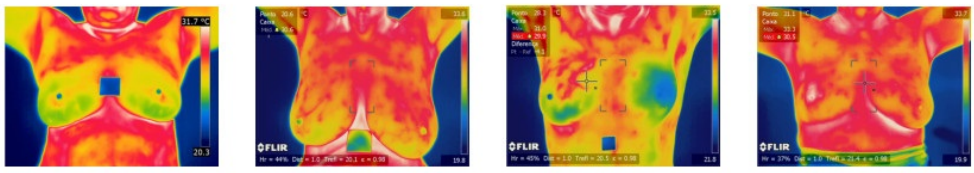

Figure 2.4: Sample images depict the breasts of several patients. Tumors exhibiting higher temperatures are visualized in shades of red or orange, whereas cooler tissues are represented in shades of green.

Estrogen dominance, ductal congestion, lymphatic congestion, and angiogenesis are all breast health risk factors that breast thermal imaging can help to identify.

To diagnose breast cancer using IR thermography, specific features on the surface temperature of the breasts are identified. The most common features are highly asymmetric temperature distributions between breasts, hyper thermic vascular patterns, localized hot spots, atypical complexity of the vascular pattern, temperature differences in the entire breast of more than (2 °C) and areolar and periareolar heat patterns [62].
These abnormal behavior of skin temperature can be divided in pathological changes in the spatial distribution of heat and pathological changes in the dynamic of the stimuli response [63]. To observe the first is sufficient a static passive breast imaging, while the second requires real time recording of cold stimulation based imaging procedures [64].

The last approach, called Dynamic IRT Mammography or Dynamic angiothermography, is more realiable and ensure good results [65]. Blood vessels, produced by cancerous tumors are simple endothelial tubes devoid of a muscular layer. Such blood vessels fail to constrict in response to sympathetic stimulus like a sudden cold stress and show a hyperthermic pattern due to vasodilatation. Dynamic angiothermography induced evaporation enhances thermographic contrast in case of tumors underneath the skin [66]. A non standardized and systematic approach and the dependency on the size and depth of the tumor are the major disadvantages of DIT, that not allow to reach accurate and meaningful results. Another major problem of DIT is the discomfort of the patient during the application of cold stress, as the cooling time can range from 2 to 6 min at temperatures below 15 °C [67]. The results of various studies show a very high variation in the accuracy of dynamic angiothermography, so that it is not preferred in practice despite its high accuracy.

To improve the performance of IRT using modern IR detectors are not enough, minimizing the sources of interference is mandatory. Minimizing unwanted interference can be achieved by using a smooth, non-reflective background, covering IR reflective surfaces, and blocking sunlight through windows and incandescent or halogen light sources [68].

Over time the IRT was overwhelmed by a long debate between those reasercher who praised the tecnique, for the grate promises as a diagnostic tool, and those who discredited it, for the high degree of subjectivity as well as a high rate of false positives and false negatives. Despite nowadays breast thermography has achieved, thanks to the advanced cameras, the standardized protocol and powerful computer

vision algorithms developed over decades, an average sensitivity and specificity of 90% [68], is still an additional tool to complement mammography. This academic brawl is reflected by the ambivalence of institutions. For example thermography was approved as a breast cancer risk assessment in 1982 by the US Food and Drug Administration, but after few years retracted its position and now the FDA does not recommend it as the sole tool for breast cancer detection, but suggest IRT mammography as adjunct tool along with another diagnostic technique [3].Both IR imaging and mammography technologies are of the complimentary nature. Neither used alone is sufficient, but when combined, each can counteract the deficiencies of the other.[69]

The figure Fig.2.5 below summarizes the benefit and drawback of breast infrared thermography.

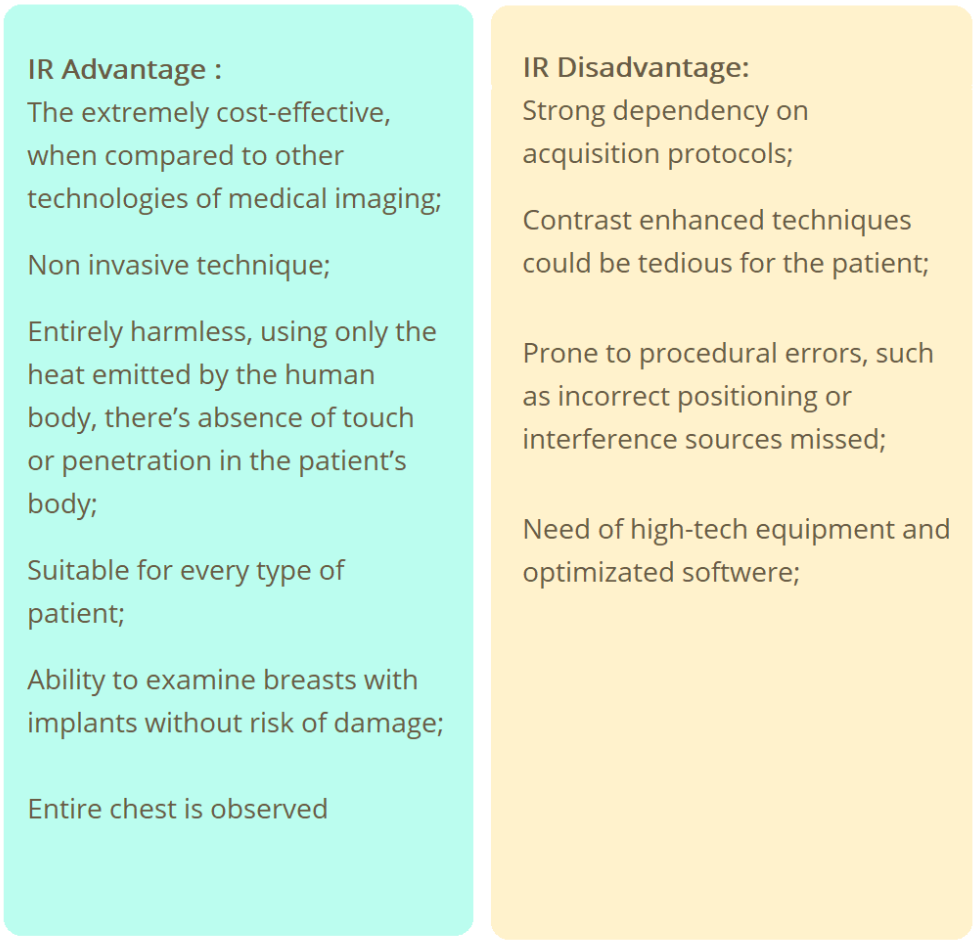

Figure 2.5: benefit and drawback of IRT for mammography

2.3 Deep Learning

2.3.1 Introduction

The recent development of artificial intelligence methodologies, notably deep learning and machine learning (ML), initiated a transformation epoch in medical diagnosis, particularly within medical image processing, allowing to automate and standardize essential tasks such as segmenting regions of interest.
Unlike conventional programming paradigms, that rely on explicitly defined instructions, ML algorithms learn patterns directly from data, by mapping mathematical functions to inputs, for predictive or classification purposes. This make ML algorithms suitable for complex and non-linear task, but require training and testing procedures.
Deep Learning is the branch of machine learning based on multilayered neural network models (NN).
This breakthrough in technology possess the capacity to surpass the constraints of conventional screening techniques, introducing a new era of accuracy and efficacy in medical diagnostics. In the past few years, computer-aided diagnosis and detection systems employing deep learning algorithms have attained significant success in diagnosing a range of cancers, including skin cancer [70] [71], cervical cancer [72], leukemia[73], gastrointestinal related cancer [74], eye diseases[75], and breast cancer[76] [77]. CAD systems provide substantial assistance to clinicians by facilitating expedited and autonomous diagnoses, particularly for ailments like breast cancer, while also aiding in detection, monitoring, and reporting procedures.
In general medical image analysis by CAD system consists in 4 step:

1. **Preprocessing** - set of transformations aimed at enhancing image quality, consistency, and compatibility with model expectations. Effective preprocessing ensures that image recognition models (used in segmentation) can learn more efficiently, generalize better, and achieve higher performance across diverse visual conditions.

2. **Segmentation** - apportionment of a digital image into meaningful segments, i.e. discrete groups of pixels connected to each other. Conventional image segmentation algorithms process high-level visual features of each pixel, like color or brightness, to identify object boundaries and background regions.

3. **Feature Extraction** - process to transform image data into a concise set of numerical features, while preserving essential visual information. This process, by providing a more manageable and informative dataset, reduces data complexity, improves the efficiency and accuracy of machine learning algorithms in the next step. Identify the minimum set of key characteristics, that best represent my starting image could be very challenging.

4. **Classification** - cataloguing the images, by associasson to each a label. Each label identify a class among a set of predetermined ones. For example in medical field common class identifiers are healthy and disease or cancer stages.

The sucess of deep learning lies in its good ability in extracting local and global relationships or structure even in raw format data. In other words, provide relatively unprocessed imaging, skipping the feature extraction step, to the learning system does not affect classification results. Between deep learning algorithms, convolutional neural networks (CNNs) have achieved notable success in segmentation tasks, including lesions recognition from background and anatomical structures separation. However, Transformer-based algorithms have recently emerged as a prominent paradigm in medical image segmentation, primarily due to their superior ability to model long-range dependencies and capture global contextual information[78].

Note that with CAD, the role of the computer analysis is not to replace the specialist (radiologist/thermologist/etc.), but rather to aid the physicians in their image interpretation and/or decision making. The final medical decision is and will be made by the doctor, not the computer.

2.3.2 Convolutional Neural Network

NNs are composed of nodes (also called perceptron), that try to mimic neuron behaviour, portioned in an input layer, an output layer, and some hidden layers in between. Each node of a hidden layer is associated with a set of weights and threshold function, which rule perceptron's input and output, respectively. Connetivity and theshold function determine the type of hidden layer. Besides the heavily increase in the number of hidden layers, CNNs have 2 peculiar types of layer: Convolutional layer and Pooling layer.

Convolutional neural networks, acknowledged as a prevalent deep learning architecture for image analysis, have demonstrated significant effectiveness in medical image tasks, including the identification of breast cancer from thermogram images [79] [80]. The efficacy of CNNs is primarily attributable to their convolutional functions and the deepness of their layers, where deeper architectures provide enhanced parameter capacity (neural scaling law), but also introduce challenges, including greater computational cost and increased chances of overfitting [81]. The generation of a public databases, containing breast thermograms and medical records of both sick and healthy patients, contributed to an exponential rise in the number of scientific publications [82], proposing advanced image-processing algorithms for thermography, since the begining of the century.

2.3.3 Object detection

The most important development in an intelligent CAD system is the segmentation of ROI and the subsequent selection and extraction of the information. Accurate object detection meliorate this process by isolating diagnostically significant regions while reducing irrelevant background information, thereby improving the reliability, efficiency, and interpretability of automated decision-making systems.

Object detection is a computer vision technique that recognizes and delineates individual objects in an image according to specified categories. The objects, which the algorithm has been trained to identify, are usually highlighted in the original image with a labeled box, called bounding box.

Given an input image, an object detection model identifies regions whose features are similar to those present in the training dataset and assigns them to the same object class. In this sense, object detection can be regarded as a form of pattern recognition. Such models do not recognize objects per se; instead, they operate on aggregates of visual properties— i.e., size, shape, and color—and classify image regions according to patterns learned from manually annotated training data.[83]

Bounding boxes are used to localize objects within the image by defining rectangular regions that enclose target instances. Each bounding box is parameterized by its center coordinates, width, and height relative to the image or feature map resolution. During training, the model learns to regress these parameters by minimizing a localization loss between predicted and ground-truth boxes. At inference time, the predicted bounding boxes are filtered using confidence thresholds and post-processing techniques, such as non-maximum suppression, to remove redundant detections and improve localization accuracy.[84]

Detectors can be primarily distinguished into anchor-based and anchor-free models, depending on how candidate bounding boxes are generated and refined. Anchor-based detectors rely on a predefined set of anchor boxes (also called priors) with fixed scales and aspect ratios, tiled across the feature maps. Priors are essentially templates for various shapes and sizes, tailored for the objects in. For each anchor, the detector predicts: offsets and confidence score. The first adjustes the anchor to better fit the target object, while the classification score indicates the probability of object's presence.
During training, ground-truth boxes are assigned to anchors based on an overlap criterion (e.g., Intersection over Union, IoU). The model learns to regress from anchors to object boxes. While this approach has been widely adopted (e.g., Faster

R-CNN, SSD, RetinaNet), it requires careful design of anchor sizes and ratios and can introduce a large number of negative samples, increasing computational cost and class imbalance.[84]

Anchor-free detectors eliminate predefined anchor boxes and instead predict object locations directly from image features. Typically, the model identifies key points such as object centers or corners and regresses the bounding box dimensions relative to these points. Training labels are assigned to pixels or feature locations rather than anchors.
This design simplifies the detection pipeline and reduces the need for manual hyperparameter tuning. Anchor-free methods often achieve better generalization and efficiency, especially for objects with diverse scales and shapes. Examples include FCOS, CenterNet, and YOLO.[84]

While different model families use different architectures, deep learning models for object detection follow a general structure. They consist of a backbone, neck, and head.

The backbone extracts features from an input image and is typically derived from a pretrained classification model. This feature extraction process generates multiple feature maps at varying spatial resolutions, which are forwarded to the neck. Breaking it down into hierarchical components, from fine (high-resolution) to coarse (low-resolution), allows to capture image's features across different sizes. The neck aggregates and concatenates these feature maps. The resulting multi-scale representations are, then, passed to the head, which predicts bounding boxes and classification scores.

In the head, object localization and classification can be performed either separately or not, respectively improving localization accuracy or execution speed. Two-stage detectors follow the former approach, while single-stage detectors adopt the latter.[85]

2.3.4 CNN in Object Detection

Recent object detection research and development, has focused largely on convolutional neural networks. Here will be presented the two CNNs models most discussed in object detection research.

Region-based Convolutional Neural Network (R-CNN) is a two-stage detector that uses a method called region proposals to generate 2000 ROI predictions per image. R-CNN then warps the extracted regions to a uniform size and forward

those regions through separate networks for feature extraction and classification. Each region is ranked according to the confidence score of its classification. The non-overlapping and top-ranking classified regions are the model's output.[86] This architecture results computational expensive and slow. Fast R-CNN and Faster R-CNN represent later developments, introducing architectural and computational optimizations, that thereby reduce inference time while increasing accuracy.[87] [88]
YOLO is a family of single-stage detection architectures based in Darknet, an open-source CNN framework. First developed in 2016, the YOLO architecture prioritizes speed, making it preferable for real-time applications. YOLO differs from R-CNN in several aspect. First of all, instead of processing region proposals through multiple independent networks, YOLO formulates object detection as a single end-to-end regression problem implemented within one neural network. Thus feature extraction and classification tasks are jointly performed. Secondly, compared to R-CNN's approach, YOLO makes less than 100 bounding box predictions per image. In addition to being faster than R-CNN, YOLO also produces less background false positives, although it has a higher localization error.[89] There have been many updates to YOLO since its inception, generally focusing on speed and accuracy.[90]

It is important to note that many other model architectures exist beyond R-CNN and YOLO. SSD and Retinanet are two additional models that use a simplified architecture similar to YOLO.[91] [92] DETR is an architecture developed by Meta that combines CNN with a transformer model and shows performance comparable to Faster R-CNN.[93]

2.3.5 Image segmentation

The segmentation step in CAD systems is the core of this thesis. Accurately identifying and delineating suspicious regions or regions of interest in various medical images (e.g., hot spots and lesions in thermograms), is essential for many clinical applications, including disease diagnosis, treatment planning, and monitoring of disease progression [94] [95]. Although manual segmentation remains the gold standard for delineating anatomical structures and pathological regions, it is inherently labor-intensive,time-consuming, and highly dependent on expert knowledge. In contrast, semi- and fully automated segmentation methods can markedly reduce workload, improve consistency, and facilitate the analysis of large-scale datasets [96].

While object detection aims only to localize objects with bounding boxes and classifying localized objects, object segmentation take a further step, adding the task of precisely demarcating object boundaries at pixel level. There are two main

ways to address this task[97]:

- according to proximity and visual similarity of pixels, adopted by **instance segmentation** algorithms.
- by classifying every pixel to an object class or the background, employed by **semantic segmentation** algorithms.

The results are presented in a set of new images, derived from the original image, one for each semantic categories or one for each object instance, depending on the segmentation approach.

Instance segmentation involves partitioning an image into sub-sections that are internally similar and distinct from the rest of the image, according to one or more features. The output is a set of masks that largely depends on the measurements accuracy of the features.[98]
Semantic segmentation is a form of dense prediction, that produces a label map of the same size as the input image, where each pixel is assigned an integer representing an object class label, corresponding to the predicted object type, at the corresponding pixel location in the input image.

If multiple instances of the same object class exist in an image, semantic segmentation not distinguish each individual instance with a different label. Thus when multiple object instances belonging to the same class are spatially adjacent or overlapping, semantic segmentation methods typically group them into a single contiguous region. In contrast, instance segmentation identify and delineates individual objects, providing richer object-level information. Instance segmentation models, therefore, are required to generate distinct segmentation masks for each object. However, it is computationally more complex and typically requires more annotated data. Consequently, semantic segmentation is often preferred for efficiency, while instance segmentation is favored for detailed object analysis.

Some of the most important segmentation techniques are presented in Table 2.2

A unified third approach called **panoptic segmentation**, try to overcome the limitations of segmentation. The word panoptic means “including everything visible in one view”. By assigning a semantic label as well as an instance ID to each pixel in the image, this segmentation aims to generate a coherent and unified scene representation for visual understanding.[99]

Semantic segmentation and instance segmentation have their distict applications cases in medical imaging. For example in some cases it may be useful to classify

image pixels into anatomical structure, such as bones, muscles, and blood vessels, while in others into pathological regions, such as cancer, tissue deformities, and multiple sclerosis lesions. In some studies the goal is to divide the entire image into subregions such as the white matter, gray matter, and cerebrospinal fluid spaces of the brain[101], whereas in others one specific structure has to be extracted, for example, breast tumors from infrared thermography images[102].

Finding accurate boundaries of the breast region is challenging due to the inherent limitations of thermal images (including the low contrast, low signal-to-noise ratio[103], and lack of clear edges), as well as great variations in breast shapes and sizes. Previous approaches proposed for breast region segmentation in thermal images included thresholding methods, region-based methods, Snake algorithms[104] and level sets[105]. In contrast, more recent research has shifted toward NN–based methods, driven by advances in deep learning and improved computational capabilities. A literature review on some of these new studies will be discussed later, in a dedicate section.

Name of method	Description
Edge detection	It relies on identifying abrupt gray-level changes to locate object boundaries, miming human approach. Edge detection achive the best results on picture with disproportionate areas; however, it performs poorly when edges are weak, noisy, or excessively numerous.
Thresholding	segments images based on intensity values alone and is computationally simple, requiring no prior knowledge of the image. Despite its efficiency, thresholding is ineffective for images without distinct intensity distributions and fails to consider spatial relationships, which often leads to inaccurate segmentation of continuous regions.
Region-based methods	these methods group neighboring pixels with similar characteristics through region growing, splitting, or merging, making them more robust to noise than edge-based approaches, though at the cost of higher computational and memory requirements. As clustering algorithms, Region growing begins from chosen pixels, called seeds region. The selection of appropriate seed points is a critical aspect of region growing, significantly influencing the efficacy and accuracy of the segmentation process.
Fuzzy-based method	It introduces fuzzy logic to model uncertainty, allowing pixels to belong to multiple regions with varying degrees of membership; this improves flexibility and realism in segmentation but increases computational complexity and requires careful design of membership functions.
Neural network–based methods	They perform segmentation through clustering or classification, further extend this adaptability by learning patterns directly from data and leveraging parallel processing capabilities. Compared to traditional methods, neural networks eliminate the need for handcrafted rules but demand extensive training time and are sensitive to initialization.

Table 2.2: Segmentation techniques description adapted from Pradeep et al.[100]. Overall, simpler methods such as thresholding and edge detection are efficient and suitable for well-structured images, whereas last 3 approaches provide greater robustness and accuracy for complex images at the expense of increased computational cost.

Chapter 3

Material and Method

3.1 Predikta database

The infrared images used in this work are gently provided by Predikta Research Solutions Ltd. The dataset consists of 244 thermal images of breast cancer patients undergoing chemotherapy treatment. The thermograms have been collected in the Hospital Universitário of the University of São Paulo (HU-USP) during the 2021.

It was used the static protocol proposed by Silva et al. [82] At each data session, the participant has been stood in front of the recording equipment, about 1 m far, with her back naked and her breasts exposed, and any objects that overlap her face, neck, and breasts has been removed (mask, goggles, ecc). Hair has been tied back, reducing noise on the thermal image. To help the patient in the placement relative to the camera, a floor pad has been employed. A plain, non-reflective background was placed behind the pad and IR reflective surfaces was covered to minimize undesired reflections. The collection room had no windows, and, in order to minimize the sources of IR interference, incandescent, halogen or natural illumination were avoided. Humidity was kept below 65% and temperature between 20°C and 25°C and the patient was acclimatized for, at least, 15 minutes in the room. During this period, the patient could wear a soft apron and it was necessary to avoid any contact with the breasts (ie. stay with crossed arms). It was crucial that thermography be the first procedure to be performed, after proper acclimatization, to avoid thermal artifacts that could altered the result of the thermography exam.

The equipment, used to recorded the dataset, consists of an embedded hardware system (Predikta Station) and a FLIR T530 professional camera was employed. The FLIR is a uncooled camera commonly used in scientific work, having a thermal reference point (thermocouple type K/arduino) attached. Other sensor's technical specifications are [106]: minimum images resolution of 320 by 240 pixels, accuracy

±2° C,thermal sensitivity less than 40 mK, 24° FOV, coupled with an automated analysis software.

3.2 Segmentation pipeline

3.2.1 Segment Anything Model

SAM is an open source image segmentation model developed by Meta AI.[107] This model can identify the precise border of specific objects in an image, given a prompt that indicates the location of desired object. Suitable prompt to specify the segmentation targets could be points or bounding boxes.
Many studies have yet applied out-of-the-box SAM models to typical medical image segmentation tasks, achieving satisfactory segmentation outcomes primarily on targets characterized by distinct boundaries.
SAM was selected for this project for its strong zero-shot performance in a variety of segmentation tasks.
The Segment Anything Model network architecture contains three crucial components: the Image Encoder, the Prompt Encoder, and the Mask Decoder.
As Figure 3.1 show, the two encoders embed their own inputs, image and prompts, respectively, and then the two information sources are combined in a lightweight mask decoder, that predicts segmentation masks.

SAM creates multiple masks for each positional input information sampled in a grid over the image. Then, masks are filtered for quality and deduplicated using non-maximal suppression.

The Image Encoder is the most powerful and essential components of SAM. It is built upon a vision transformer model (ViT), pre-trained with a dataset of 11 million images. Transformer is a deep learning model designed for sequential data, like sentences, focused on long-range dependencies research in data, which is essential for natural lenguage process, and useful in object detection. The idea behind vision transformer is decomposing an image into a set of fixed-size patches, each of which represents as a token, in an analogous way to a word in a sentence. The image encoder processes the entire image once and can be applied prior to prompting the model.
Prompt encoder incorporates points and bounding boxes using positional encodings and sum them with learned embeddings. For masks as prompts, after downsampling happens through convolutional layers, the embedding is summed element-wise with the input image embedding. For the prompt encoder, points, boxes, and text act as sparse inputs, and masks act as dense inputs.

Mask decoder resemble a transformer decoder with a dynamic mask prediction head.

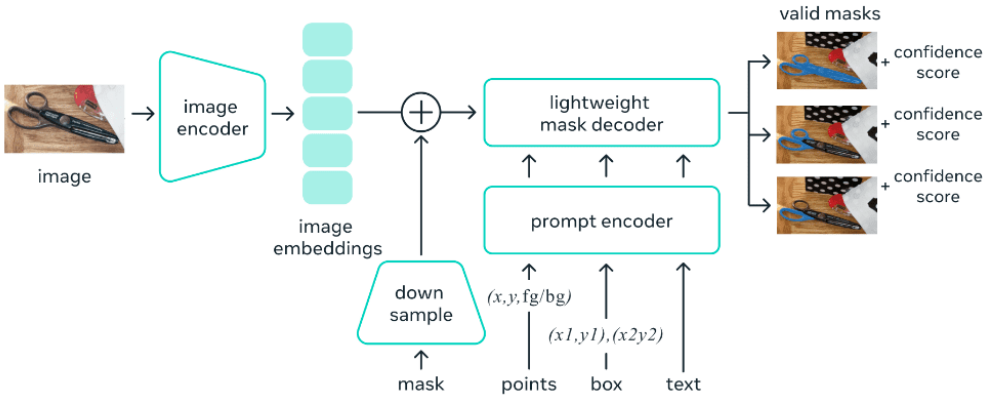

Figure 3.1: The model architecture of Segment Anything when giving it an image as input.

3.2.2 You Only Look Once

YOLO is a popular object detection and image segmentation model developed by Ultralytics. It falls into the category of a single-staged deep convolutional neural network detector, with the peculiarity of having a triple head, for dense prediction at different scale. It predicts bounding boxes and class probabilities for objects directly from the input image, in a single forward pass (hence the name), making it significantly faster.
YOLO model configurations cover a wide range of scenarios, from simple object detection to more complex tasks, like instance segmentation and object tracking. It is also designed to run efficiently on a variety of hardware platforms, from CPUs to GPUs.

In this work three versions of YOLO were used:

- YOLOv8, one of the latest (released in 2023) and solid installment of the YOLO architecture.

- YOLOv5u, an advancement from the original fifth version. YOLOv5u integrates the anchor-free, objectness-free split head, a feature previously introduced in the YOLOv8 models. This adaptation refines the model's architecture, leading to an improved accuracy-speed tradeoff in object detection tasks.

- YOLOv11 is the next major evolution among the Ultralytics products, released in 2024.

Every version builds on the architecture of YOLOv5, but introduces deeper improvements to both model structure and efficiency. The specific architectures are shown in figures Fig.3.2-3.4.
Being designed for fast inference, user-friendly and PyTorch-native with great amount of documentation, YOLOv5 is often considered what made relevant its model family. Although recent versions outperform YOLOv5, on GitHub, its Ultralytics repository still has more commitment and engagement compared to some other YOLO variant repository, indicating strong community usage and interest.[108]

Despite the number of layers varies depending on the version, YOLO follows the general structure of a CNN-based object detector:

- Backbone: This is the convolutional neural network responsible for extracting features from the input image.
- Neck: Also known as the feature extractor, it merges feature maps from different stages of the backbone to capture information at various scales.
- Head: It is responsible for making predictions. YOLO employs multiple detection modules that predict bounding boxes, objectness scores, and class probabilities for each grid cell in the feature map. These predictions are then aggregated to obtain the final detection.

In addition to the number of layers, which is 24, 22 and 23 respectively for v5u, v8 and v11, each version has its own particularity.

For example YOLOv5 uses a custom CSPDarknet53 module in the backbone, which is based on Darknet53 and employs cross-stage partial (CSP) connections to improve information flow between layers, boosting accuracy with respect to the previous versions. The CSP structure enhances the feature representation by partitioning the input feature map into two parts, one is directly passed to the next layer, the other undergoes convolution operation, and then merges them through a cross-stage hierarchical structure.[109]
Furthermore in the neck of YOLOv5 the feature fusion is mainly handled via the feature pyramid network (FPN). It fuses feature maps of different scales through up-sampling and splicing operations to enhance the detection of objects of different sizes. YOLOv5 utilizes a Path Aggregation Network (PAN) concepts to construct the feature pyramid, enhancing the entire feature map hierarchy.

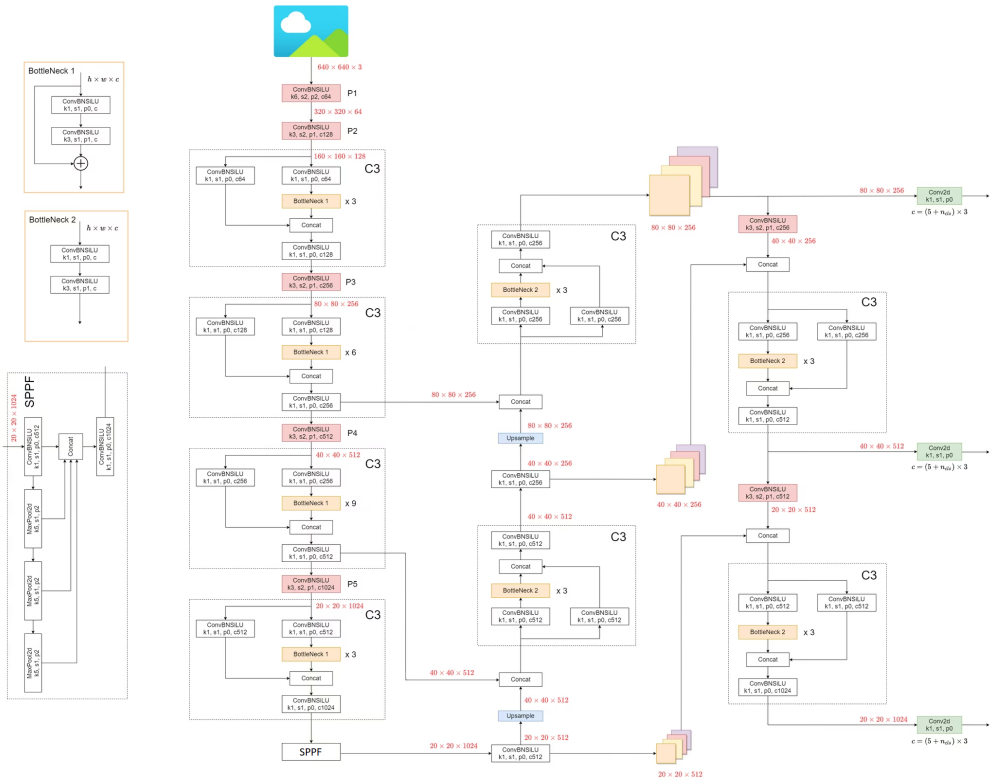

Figure 3.2: The model architecture of You Only Look Once version v5.

Although introduced for the first time in version 5, all the YOLO models in this project use the Spatial Pyramid Pooling Fast (SPPF) module, which facilitate the integration of key information from targets of various sizes by combining feature maps through pooling operations at different scales.

While YOLOv8 architecture introduces a C2f (CSPDarknet53 to 2-Stage Feature Pyramid Network)module in the place of the traditional FPN. This module combines high-level semantic features with low-level spatial information, leading to improved detection accuracy, especially for small objects.[110] The head network of YOLOv8 employs decoupled detection heads, using two parallel convolution branches to compute the regression and classification losses separately.

YOLOv11 adopts a feature extraction module based on Transformer, which can better capture globally dependent features than traditional CNN and it is especially suitable for handling complex scenes and long-distance-dependent features, which makes the model perform better in detecting fewer sample categories and complex

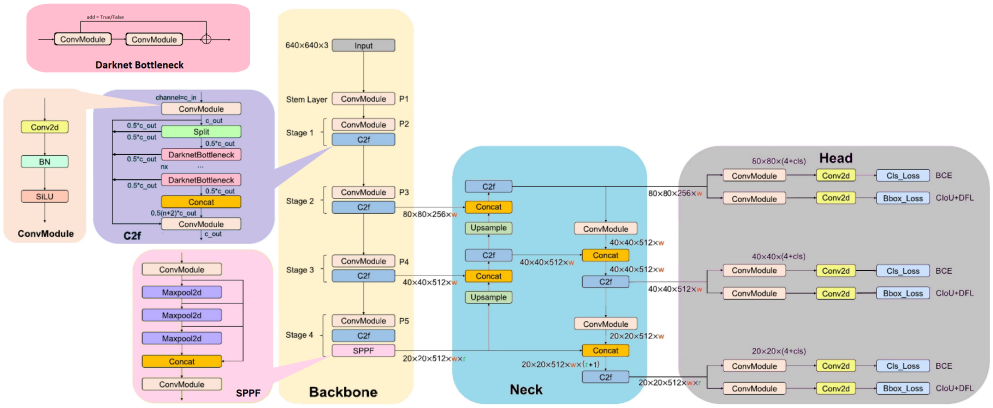

Figure 3.3: The model architecture of You Only Look Once version v8.

environments.[111]
While YOLO11 retains the Spatial Pyramid Pooling - Fast block from previous versions, it incorporates a redesigned Cross Stage Partial module augmented with a Spatial Attention component (C2PSA) after SPPF. This new module redefines how spatial information is processed within the network. By applying spatial attention, the model can selectively emphasize the most relevant regions of the feature maps while suppressing less informative areas. Through spatial pooling operations, C2PSA enables YOLO11 to better localize meaningful regions in an image, potentially improving detection accuracy for objects of varying sizes and positions.

Each version used in this project is an anchor-free detector. Instead of relying on a predefined set of anchor boxes to predict object locations, these models identify objects directly. This is often achieved by predicting an object's center point and its dimensions, or by identifying keypoints like corners. Despite these methods simplify the label assignment process during model training, they need to incorporate advanced techniques, like sophisticated loss functions, into the model. YOLO does indeed utilize advanced loss functions to optimize the model, including:

- Bounding box loss: is associated with the bounding box prediction error. This regression task is address by computing the IoU loss (CIoU by default), called 'box_loss' during training.
- DFL loss (Distribution Focal Loss), which is directly reported as dfl_loss during training. It helps the model to better estimate object categories, namely it is the loss associated with the error in the classification task. It uses a modified Binary Cross Entropy to support multi-label classification.

- VFL loss (Varifocal Loss), which is not separately shown but is incorporated within cls_loss (class loss) in the training logs. VFL is designed to address imbalances and uncertainties in classification tasks.

Each of these loss components plays a vital role in honing the model's accuracy, each focusing on a different aspect of the detection task (localization, classification, etc.).Training logs display these as box_loss, cls_loss, and dfl_loss, corresponding to model performance in respective areas. During training these losses are computed for each prediction layer and then summed up. Each loss component is weighted to control its contribution (by tunable hyperparameters). Additionally, the DFL loss has an extra weight that varies for each prediction layer to ensure predictions at different scales contribute appropriately to the total loss. The goal during training is to minimize these losses for better model performance.

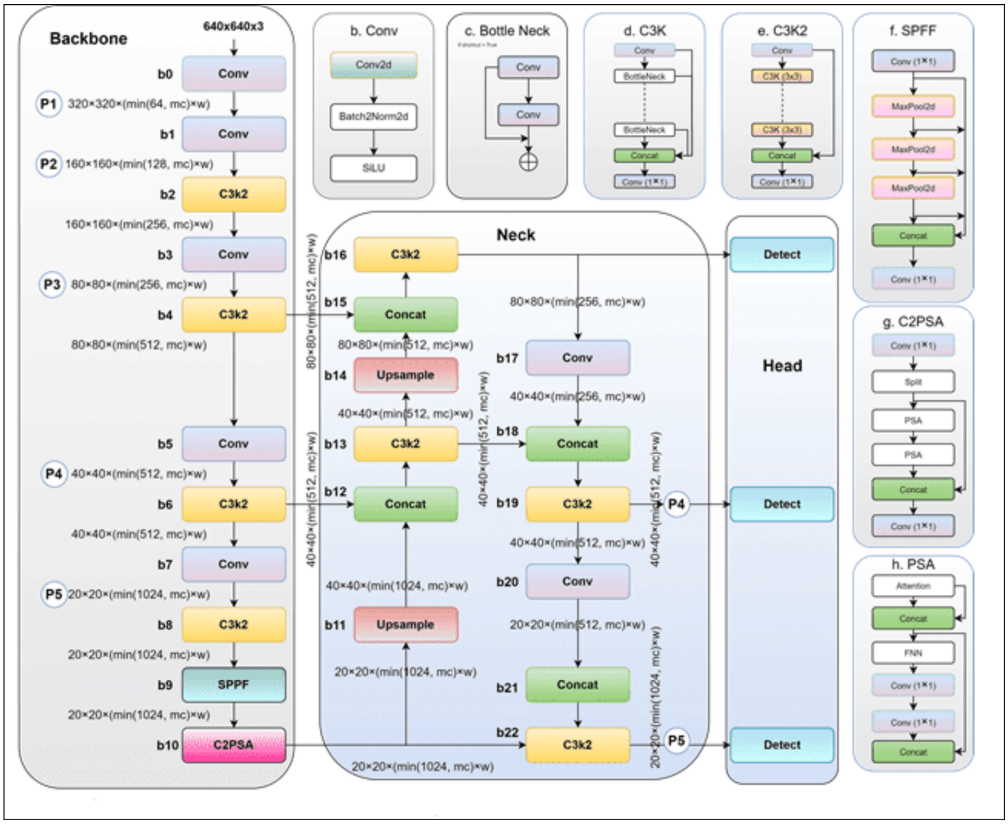

Figure 3.4: The model architecture of You Only Look Once version v11.

3.3 Design of experiment

3.3.1 Context

To establish a comprehensive understanding of the current state of research in this domain, it is essential to examine existing studies that have investigated the application of CNN–based object segmentation techniques inside CAD system for breast cancer detection in thermograms. The following literature review synthesizes prior work.
Regarding the usage of CNNs for the breast cancer IRT classification, this topic has been widely explored by many authors [77] [79] [80]. But it is really uncommon to find studies employing CNNs-base techniques for breast thermogram segmentation. Thermal images share similarities with a standard gray-scale or colored image; thus, mostly of the researchers works focus on trying to identify texture and statistical features directly from the thermal images, rather than search and extract ROI.
CNN-based models have emerged as a major advancement in the field, as their hierarchical feature learning provides strong generalization, unlike manually designed features, which often fail to adapt to variations in breast morphology and lighting conditions.

Çevik et al.[112] present a comprehensive evaluation of the first five YOLO family architectures for the detection and classification of breast lesions in thermal breast images. This study show the plausibility in exploit YOLOv5 architecture for breast reagions detection.

Tayel and Elbagoury[113] proposed an automated segmentation technique to extract the breast outline from breast infrared thermography images, based on the use of Fully Convolutional Networks (FCN). They use transfer learning to initialize the model with the parameters of pre-trained CNN-AlexNet from natural images of the database PASCAL VOC 2011 segmentation challenge. On a database with 285 breast thermography images (in grayscale), with 320x240 pixels resolution, their FCN reached as accuracy, sensitivity and specificity of 96.4%, 97.5%, and 97.8% respectively.

Rosli and Habaebi[114] compared three segmentation models U-Net, U-Net with Spatial Attention, and U-Net++, whose performance was assessed on DMR-IR, a public accessible database, through k-fold cross-validation with many metrics, including: IoU, Dice coefficient, ROC-AUC and PR-AUC. U-Net is a powerful medical image segmentation network with a deeply-supervised encoder-decoder architecture. The other two models are variants of the original U-Net design, expanding the architecture to increase accuracy. SAU-Net add a spatial attention

mechanisms to help the model focus on important spatial regions, learning a weight map for each feature map. While UNet++ (also called Nested U-Net) present a structural redesign of skip connections to reduce the semantic gap between encoder and decoder feature maps, through a nested, dense network pathways.The optimizer handles easier learning tasks when feature maps from decoder and encoder sub-networks are semantically similar. Rosli and Habaebi used five optimization algorithms to fine-tune the three networks: ADAM, NADAM, RMSPROP, SGDM, and ADADELTA. The ADAM optimizer consistently outperformed the others, yielding superior accuracy and reduced loss. Among the models, the baseline U-Net, despite being less complex, demonstrated the most effective performance, with precision of 0.9721, recall of 0.9559, specificity of 0.9801, ROC-AUC of 0.9680, and PR-AUC of 0.9472. The study shows that the original U-Net model, particularly when trained with the ADAM optimizer, outperforms more complex variants in robustness, breast region overlap, and noise handling. The results suggest that increased architectural complexity does not guarantee better performance and that standard U-Net remains an effective and efficient choice for medical image analysis.

S.Guan et.al[115] applied an autoencoder-like convolutional deconvolutional neural network (C-DCNN) to segment breast regions in a thermal mammography database of 132 gray-scale images. C-DCNN acts as a U-Net, applying an encoding, to decrease the spatial dimensions of the image, while also capturing relevant information, and a decoding, to upsample the feature map and produce a relevant segmentation map. They achieved good performance with an overall average IoU about 0.834 with 0.081 of standard deviation.
The same research team in a new study[116] developed a MultiResUnet, a true U-Net for deep-learning segmentation, that exhibited an average accuracy of 91.47%, surpassing their previous work by about 2%. However, limitations in segmentation of small breast, IoU errors, data augmentation, and manual challenges were identified by the authors, suggesting room for improvement.

The Mirasbekov et al. research[117] investigated the efficacy of machine learning methodologies, particularly Bayesian networks integrated with CNNs, to enhance early-stage breast cancer diagnosis. Despite its primary task is classification, it is remarkable the use of a LIME for explainable AI, that depicted areas that contributed most when classifying images, as a segmentation tool for Bayesian networks.

A 4D U-Net segmentation on digital IRT imaging is proposed in the P. Gomathi et.al's manuscript[118], which aims to the diagnosis of breast cancer, feeding a Binarized Spiking Neural Network (BSNN) the segmented ROIs. The digital infrared thermal images are taken from DMR-IR dataset. This study classify the pathology stage as No spread, Early Stage, Localized, Regional and Distant. The overall

accuracy and sensibility reported are about 98% and 95%.The 4D U-Net weight parameters are optimized with Glowworm Swarm Optimization Algorithm (GSOA). Initially the input dataset is pre-processed to compress the dynamic range of image, maintaining the local features, with an Altered Phase Preserving Dynamic Range Compression (APPDRC) approach, that remove the speckle noise.

Despite SAM has recently gained popularity in the medicine field for image segmentation, due to its impressive capabilities in various segmentation tasks, SAM lacks of domain-specific knowledge, as it is pretrained on natural images, limiting its effectiveness. For completeness, two research have to be mention. With the aims to fulfill the role of a foundation model for universal medical image segmentation, Jun Ma et al.[119] introduced MedSAM, a fine-tuned SAM on a large-scale dataset, with more than one million medical image-mask pairs. This dataset, assembled by the authors them-selves, comprehending most diverse imaging techniques, such as MRI, X-ray, endoscopy, except thermography.
Jun Ma et al. reported an overall median Dice Similarity Coefficient of about 82% for MedSAM.
Wu et al.[120] adopted an Adapter technique to fine-tune the SAM model and enhance its medical specific knowledge, reaching a final Dice of 89.8% with bounding box overlap of 75%. Also this study excluded IRT mammography from their dataset.
Y.Huang et al.[121] accomplished an extensive and comprehensive validation on zero-shot SAM's performance, coming to the conclusion that, as the number of scenarios increase, SAM behavior became inconsistent, exhibiting good performance only in some specific anatomical structures. Exploring the ViT-L and ViT-H's segmentation capabilities for the Everything mode and 5 prompt modalities on COSMOS 1050K (a huge medical dataset of a total 1050K 2D images and 6033K masks, collected and sorted expressly for this purpose),the research group found that ViT-H achieved the best results with box prompting.

Yan et al.[122] combined YOLOv3 with U-Net++ in an automatic multi-scale breast mass segmentation, demonstrating the benefit of modular specialization through stage-wise optimization. The first model was employed to localize potential lesions in X-ray mammogram, while the second precisely refine their boundaries.

The first to implement our method are Pandey et al.[123] They incorporate YOLOv8 with SAM for instance segmentation of lungs, showing that integrating large-scale models into detection–segmentation pipelines offers potential advantages in medical contexts, if the appropriate model was carefully selecting based on specific requirements.

These last two works illustrate that coupling robust detection backbones such as YOLO with advanced segmentation networks such as U-Net variants or SAM can effectively decouple localization and boundary modeling, thereby addressing limitations of single-stage designs.

3.3.2 Proposed Method

This project falls within the scope of breast cancer patients follow-up. This work has the main goal of using pretrained state-of-the-art model to address a multiclass segmentation of breast thermograms, while simplify the procedure. Secondly it investigates whether a two-stage segmentation framework can reach good performance with a limited dataset of only 244 images, with the minimum training. A further challenge in this project is represented by the difference between class sizes. The proposed method is based on the combination of two recent deep learning models: Segment Anything Model and You Only Look Once. The first is based upon visual transformer, while the last is a CNN-based model. Both selected architectures enable fast implementation and demonstrated good performance in previous studies, even though never used in combination in this specific field.

In contrast to numerous prior studies that depend extensively on intricate and multi-phase preprocessing techniques, this pipeline completely obviates the necessity for these procedures. The thermic images are directly fed to the models. By means of YOLO, bounding boxes are generated and used to point out the regions of interest to SAM, that complete the operation, segmenting the anatomical structure inside them. The training methodologies of these models and the evaluation of their performance will be discussed in a dedicated section.
Although, in general, it is preferable to fine-tune a single architecture, a light pipeline, such as the one proposed, ensures modularity and flexibility, without compromise performance results. Thus it is able to exploit the strengths of both types of model, adapt and evolve over time, as new versions or learner/architecture are released.
Indeed, on one hand transformers emphasize long-range dependencies, at the expense of capturing fine-grained local details, which can reduce accuracy when distinguishing between background and target regions. In contrast, CNNs rely on receptive fields that effectively identify fine structures, but frequently fail to capture broader anatomical context, leading to over- or under-segmentation; in particularly in dense breast tissues with indistinct lesion boundaries, where preserving subtle structural details is crucial.
A cleverly combined approach can address the limitation of both models.

Below a schematic representation of the proposed method is shown (Fig 3.5).

The novelty and contributions of the study are summerised as follows:

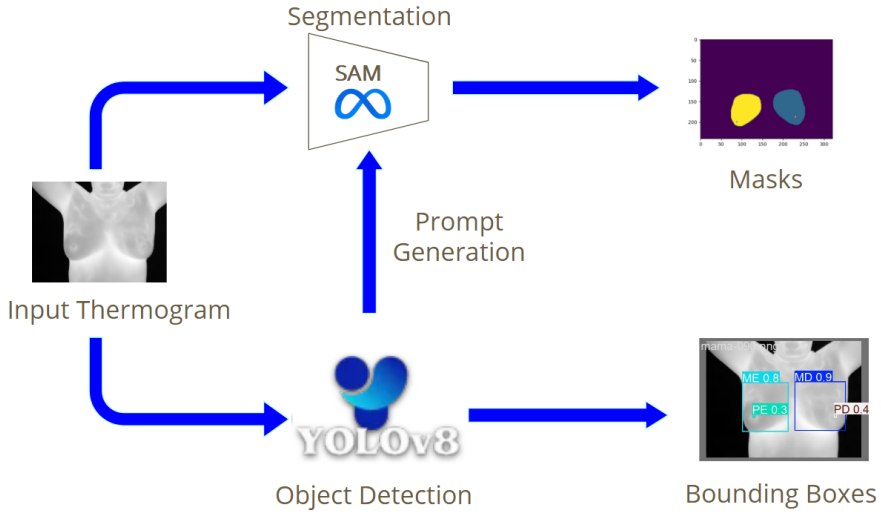

Figure 3.5: Schematic representation of the proposed method

- Implementing a modular double-stage pipeline, that could be insert in a CAD system, with the aim to obviate intricate segmentation and preprocessing procedures.

- Investigates the effectiveness of combining pre-trained CNN and ViT architectures to segment breast thermography images, employing a small dataset for fine-tuning, validation, and testing.

3.3.3 Evaluation Metrics

Evaluation metrics in computer vision problems aims to provide an informative way to measure the similarity between predicted ROI and ground-true. Generally, metrics for computer vision fall into two categories: region-based, if they assess the overlap, and boundary-based, if they assess the boundary proximity [124].
To evaluate the masks predicted by SAM, two metric of the first type were chosen, and two of the second. All of them are selected among typical metrics in the field of medical image segmentation [125].

- The Intersection over Union (IoU) score, also known as Jaccard's Index, measures the relative overlap between two figures. It is a standard performance

measure for image segmentation problems. The IoU is calculated by taking the intersection area of two bounding boxes and dividing it by the union area of these boxes.

- Dice coefficient (F1 score), represents the similarity between two samples. It is typically expressed as a fraction between 0 and 1, where a higher value indicates better model performance. It can be directly bond to IoU.
- HD95, refers to the 95% Hausdorff distance, which quantifies the maximum distance between two sets at the 95th percentile. A smaller value indicates a higher similarity between the two sets. The Hausdorff distance is the maximum of all the shortest distances between a point in one edge and any point in the other edge. It is considered a complementary metric to Dice score, since more sensitive to local differences, while F1 measures the global overlap. Recommended for segmentation tasks with complex boundaries and small thin segments.

Most favorites by many author, mAP50 and mAP50-95 were chosen to evaluate the bounding boxes generated by YOLO. These are metrics derived by IoU, used as threshold to distinguish between positive and negative prediction. Mean Average Precision (mAP) is a widely used metric for evaluating the performance of object detection models in computer vision. mAP is the area under the Precision-Recall curve, but, while mAP50 is computed with a fixed IoU threshold, seted to 0.5, mAP50-95 is the mean value over a range of thresholds (from 0.50 to 0.95). The last one is a more stringent evaluation metric as it measures the model's performance over a variety of IoU thresholds.
mAP provides a single value that represents the accuracy of a model across all classes and thresholds, making it a comprehensive measure of a model's ability to correctly identify and locate objects within an image or video. Unlike simpler metrics such as accuracy, mAP considers both the precision and recall of the model, offering a more nuanced understanding of its effectiveness.

Recall (also known as sensitivity) is the real positive percentage correctly captured relevant objects in the image. It assesses the model's completeness in identifying objects of interest. A high recall score indicates that the model effectively identifies most of the relevant objects in the data.

Precision is the proportion of true positive among total of predicted. In essence, precision provides insight into the model's ability to make positive predictions that are indeed accurate. A high precision score indicates that the model is skilled at avoiding false positives and provides reliable positive predictions.

3.4 Workflow

The aim of this thesis is automatic detect and segment four specific ROI inside gray scale thermogram, conjunctly using innovative large learning models, i.e. YOLO and SAM. These ROI are: left breast(ME), right breast(MD), left nipple(PE) and right nipple(PD).

The project workflow involves several steps, which are summarized in the Figure 3.6.
The two tools were initially tested separately and subsequently in combination. In the official SAM GitHub repository are provided three types of pre-trained models, named ViTB, ViT-L, and ViT-H, each with ascending sizes. Their model parameters range from small to large. The last one, ViT-H, was chosen for its substantial performance improvements over the others. However, it requires multiplied testing time due to its increased complexity.
Since the papers by Jun Ma et al. and Wu et al. explain that, making SAM a powerful segmentation model for the medical domain requires training on extremely large datasets- an approach that is not feasible for thermic mammography- relying on SAM's strong zero-shot segmentation ability was considered a preferable option. While for YOLO a k-fold cross-validation, with k equals to 10, was carry out. Consequently the dataset of 244 images was divided into a Trainset of 198 images, a Validset of 22 images and a Testset of 24 images.
As described in the previous 'Evaluation Metrics' section of this chapter, mAP50

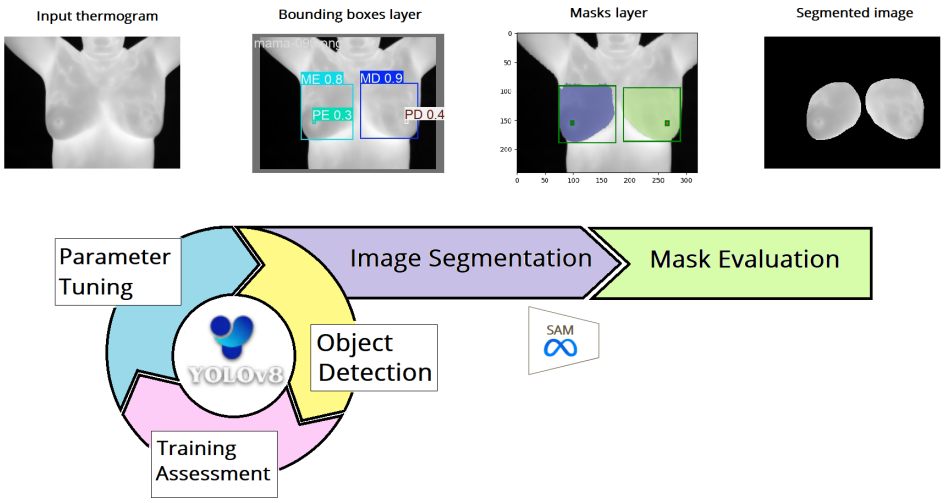

Figure 3.6: Flowchart used in this project

and mAP50-90 were used to verify the CNN's performance.
The data augmentation was limited as possible to force stability in the output: for each image YOLO had to produce exclusively 4 bounding boxes, one per class. The only augmentation admit are on hue, saturation, and value(brightness) to be more capable of handling real-world scenarios with varying lighting conditions and contrasts. Since the dataset already contains patient at different distances from the camera, scaling augmentation was also implemented. Ultralytics makes a large number of YOLO models available. Between these the version '5 small'(v5), '8 nano'(v8) and '11 nano' (v11) were chosen, primarily for low computational and memory requirements. In previous studies, both model had demonstrated to ensuring good performance and accuracy, despite their lightweight structure. Transfer learning is a key point for using small datasets and in medical research is really common, since images from novelty technique are impossible to collect in vast quantities. A great deal of data, power and time is required to train deep learning models from scratch. Thus, pre-trained models and fine tuning are used to solve these problems. Ultralitics offers diverse pretrained models, each specialized for specific tasks in computer vision, each trained on COCO dataset. The figures (Fig 3.7 - Fig 3.9) below show characteristics and performances of YOLO models.

See Detection Docs for usage examples with these models trained on COCO, which include 80 pretrained classes.

Model	size (pixels)	mAP^{val} 50-95	Speed CPU ONNX (ms)	Speed A100 TensorRT (ms)	params (M)	FLOPs (B)
yolov5nu.pt	640	34.3	73.6	1.06	2.6	7.7
yolov5su.pt	640	43.0	120.7	1.27	9.1	24.0
yolov5mu.pt	640	49.0	233.9	1.86	25.1	64.2
yolov5lu.pt	640	52.2	408.4	2.50	53.2	135.0
yolov5xu.pt	640	53.2	763.2	3.81	97.2	246.4

Figure 3.7: Characteristics and performances of YOLOv5 pretrained models provided by Ultralitics. Chosen model is highlighted in red.

See Detection Docs for usage examples with these models trained on COCO, which include 80 pretrained classes.

Model	size (pixels)	mAP^{val} 50-95	Speed CPU ONNX (ms)	Speed A100 TensorRT (ms)	params (M)	FLOPs (B)
YOLOv8n	640	37.3	80.4	0.99	3.2	8.7
YOLOv8s	640	44.9	128.4	1.20	11.2	28.6
YOLOv8m	640	50.2	234.7	1.83	25.9	78.9
YOLOv8l	640	52.9	375.2	2.39	43.7	165.2
YOLOv8x	640	53.9	479.1	3.53	68.2	257.8

Figure 3.8: Characteristics and performances of YOLOv8 pretrained models provided by Ultralitics. Chosen model is highlighted in red.

See Detection Docs for usage examples with these models trained on COCO, which include 80 pretrained classes.

Model	size (pixels)	mAP^{val} 50-95	Speed CPU ONNX (ms)	Speed T4 TensorRT10 (ms)	params (M)	FLOPs (B)
YOLO11n	640	39.5	56.1 ± 0.8	1.5 ± 0.0	2.6	6.5
YOLO11s	640	47.0	90.0 ± 1.2	2.5 ± 0.0	9.4	21.5
YOLO11m	640	51.5	183.2 ± 2.0	4.7 ± 0.1	20.1	68.0
YOLO11l	640	53.4	238.6 ± 1.4	6.2 ± 0.1	25.3	86.9
YOLO11x	640	54.7	462.8 ± 6.7	11.3 ± 0.2	56.9	194.9

Figure 3.9: Characteristics and performances of YOLOv11 pretrained models provided by Ultralitics. Chosen model is highlighted in red.

Chapter 4

Results and discussion

4.1 YOLO models comparison

The same procedure was followed to train the three models. Firstly it was estimate the YOLO's nominal responses, keeping as much default setting values as possible. YOLOv11, YOLOv8 and YOLOv5's initial performance are summarize in the Table4.1 .
In them, it is already revealed a consistent trend that persisted throughout the entire experimental series: a marked divergence between breasts and nipples, in all the three versions. This is primarily observable through precision and recall data. Since recall relates to effectiveness of detect ability and precision relates to the credibility of predictions, a low recall value and a high precision value means that the model misses the majority of targets, but the few detected are perfectly located. Given that, normally, in the medical domain a high recall is preferred over high precision, the training was conducted with the aim of achieving this type of performance, starting from these point.

The clear difference between classes may be caused by the smallness and elusive nature of nipples, which can confuse the models. Indeed as an object passes through the convolutional layers, its size is progressively reduced. Consequently, small objects may disappear after several layers and become undetectable. These initial data were adopted as a reference point for comparison, in this beginning phase of the study.

As second step, a partial progressive unfreezing is applied to the model to optimize pre-learned features during fine-tuning, to minimize the computational cost and time for later test. Thereafter, an optimization of the hyperparameters

YOLO		MD	ME	PD	PE	tot
v5	mAP50	47.24±8.11	66.30±4.88	48.24±6.62	61.5±4.45	55.82±6.02
	mAP5090	28.71±8.28	33.20±4.55	23.30±4.01	24.00±1.58	27.3±4.61
v8	mAP50	50.39±7.53	67.28±6.54	51.00±6.67	61.27±7.06	57.49±6.96
	mAP5090	32.58±5.03	36.44±5.84	23.11±3.51	20.93±4.11	28.27±4.62
v11	mAP50	64.95±6.24	65.61±6.5	56.25±6.55	58.2±5.91	61.25±6.3
	mAP5090	46.61±5.42	47.59±3.52	31.52±6.6	29.78±5.11	38.88±5.16

Table 4.1: nominal response in percentage: mean and standard deviation (%) of mAP50 and mAP50-90 computed on Validset with 10-fold cross-validation

was carried out, in particular the focus was in bounding boxes loss and dropout.
Freezing procedure seems to not benefit the learning of the models, as it can be easily perceivable by graphics of mAPs (reported below at the end of the section). As expected, the values of mAP50-90 are lower than mAP50. While between v5 and v8 no significative difference are observed, but for v11, where the mean of mAP50-90 seems be higher. It is possible to notice very similar trends as the number of frozen layers increase, in almost all cases, in all YOLO versions, in particular for total means, that generally decrease.

Observing the performance on trainset the disparity between classes is undeniable (graphs on figure 4.1). The mAP50 for right and left breast has constant mean close to 1, with a variance close to zero. While the mAP50 for nipples decrease, even by about 10 percentage point, as the network progressively stop from updating its weights. This decrement can be found in all the three versions. An analogous situation arises in mAP50-90, with even comparable reduction values, except for breasts trend, that have a slightly decrease.
PE and PD's mAP50 and mAP50-90 variance is generally higher than ME and MD's one. Despite a similar tendency, v5 and v11 seem have higher precision than v8 on trainset, especially for PE and PD classes. This is not repeated on validset, thus, from this point of view, there no relevant difference in the response at freezing between versions, save for v11 in keeping an higher detecting mean.

Analyzing the performance on validation set, breast-nipple disparity seem disappear in first instance, but rise again, with the implementation of dropout (graphs on figure 4.1-4.2). Generally mAP50 and mAP50-90 mean, for each one of the four classes, tends to lower its value, as the number of frozen layers increase. In the opposite, except for dropout null case, variability of results keep comparable even between mAP50 and mAP50-90.
As previous mentioned, YOLOv8 shows a slight more difficult to identify PE and

PD on high freezing numbers, with respect to other versions. This difference can vary from 0.1 point on mAP50-90 to 0.2 point on mAP50.

Since the values of average precision are higher on the trainset than on validset, we can assume that the models overfit the data. For this reason a dropout tuning was necessary.
Thanks to dropout the network can focussing on important feature and generalizing new data better. In fact, performance for each versions increased following dropout hyper-parameter, until 0.4 value, after that a plateau is reached. Even in this case a similar trend between versions was registered and none of three obtained a fulsome advantage over the others, as is possible see in the graph on figure Fig.4.4.
On the training set, the outcomes remain quite unchanged, except for a moderate improvement in classes of nipples for both precision score, and a small increase in variance, above all on v11's results of the same previous 2 classes. This was expected, as dropout is a method used for regularization in CNNs, with a stochastic approach. Furthermore, small sized objects are more susceptible to the effects of dropout.

Regarding, performance on Validset, a remarkable improvement, that involved all the classes, was found.
For each versions, an increase of the two metrics total mean and total variance can be observed, indeed the first one obtained a neat rise, while the variance underwent to a subtler one. It is interesting notice that, where performance was already strong, only modest improvements were achieved, whereas in cases with poorer initial results, dropout led to substantial improvements. For example, for the version 8, the mAP50 mean went from 0.57 to 0.82, a growth of 44%, and mAP50-90 mean went from 0.27 to 0.48, growing about of 70%. Instead the mAP50 standard deviation went from 0.0695 to 0.0736 and mAP50-90 standard deviation went from 0.046 to 0.054, growths of about 6% and 17%.
Comparing the previous table (Tab 4.1) with the ones further down (Tab 4.2 and Tab 4.3), all this claims are clearly visible. For further clarification graphs in the figures 4.1-4.3, at the end of this section, show the effect of dropout rate on the performance's variation.
An interesting fact is the stability of mAP50-90s, which mean and variance keep roughly constant as the dropout rate increase. May this is due to its nature, as mAP50-90 is an average computed on several confidence score.
The top performance on validation set is 0.8323 ± 0.063 in mAP50 and 0.5282 ± 0.052 in mAP50-90, reached for 0.5 of dropout by YOLOv11. Following immediately after the two remaining versions, whom for the same parameters, achieved 0.821 ± 0.049 in mAP50 and 0.470 ± 0.034 in mAP50-90, and 0.823 ± 0.060 in mAP50 and 0.4798 ± 0.043 in mAP50-90, reached by v5 and v8 respectively.

mAP 50

YOLO	dropout	MD	ME	PD	PE	tot
v5	0.5	87.24±7.44	91.31±2.62	71.16±6.17	78.56±3.64	82.07±4.97
	0.7	77.39±8.95	86.04±7.44	76.41±6.31	74.71± 6.67	78.63±7.34
v8	0.5	91.22±5.97	89.08±7.48	75.43±4.95	73.47±4.65	82.3±5.76
	0.7	78.32±9.64	83.63±4.03	79.54±6.77	72.86±6.02	78.59±6.62
v11	0.5	88.24±6.24	89.14±6.50	76.43±6.55	79.08±5.91	83.23±6.3
	0.7	86.55±5.91	86.29±6.75	75.62±4.82	77.87±6.16	81.63±5.91

Table 4.2: model response to dropout summarization: mean and standard deviation (%) of mAP50 computed on Validset with 10-fold cross-validation for the main 2 dropout rate values: 0.5 and 0.7

mAP 50-90

YOLO	dropout	MD	ME	PD	PE	tot
v5	0.5	56.07±5.16	65.10±3.28	33.38±4.36	33.33±2.76	46.97±3.89
	0.7	56.21±8.40	62.65±5.43	42.06±4.98	41.58±6.53	50.62±6.34
v8	0.5	59.86±5.83	64.56±3.86	35.79±4.44	31.69±3.23	47.98±4.34
	0.7	56.57±9.66	63.25±5.02	42.08±3.19	39.55±3.56	50.36±5.36
v11	0.5	63.33±5.42	64.66±3.52	42.83±6.60	40.46±5.11	52.82±5.16
	0.7	61.46±3.05	62.42±5.02	40.15±5.24	40.62±2.32	51.16±3.91

Table 4.3: model response to dropout summarization: mean and standard deviation (%) of mAP50-90 computed on Validset with 10-fold cross-validation for the main 2 dropout rate values: 0.5 and 0.7

To complete the YOLO model training a fine tune of bounding box loss weight was carried out. This hyper-parameter controls the prediction error of box positioning. It was found out, the optimum value for bounding box loss weight is 9, since outperform precedent results and further increase do not bring more improvement. Although rising it, from the default value of 7.5 to 9, does not yield any benefit to mAP50, a growth in AMP50–95 can be observed for the versions v5 and v8. For v11, instead, none progress was registered.
Despite the persistence of difference between breast and nipples classes performance, it is less pronounced compared to the previous results. At this point the testset was used to quantify the actual generalization ability of the models. The following

table (Tab4.6-4.7) shows the outcomes, reached by models on testing set.

The tuning of box loss did not provide clear guidance regarding the optimal combination to select; therefore, all configurations were evaluated to the testing phase. The data on test set strongly reveal the improvement yielded by increase box loss to the model performances, regardless to the versions. The best performances on testing set were obtained by YOLOv11, achieving an mAP50 of 0.844±0.045 and a mAP5090 of 0.498±0.042.

mAP 50

YOLO	b.loss	MD	ME	PD	PE	tot
v5	7.5	87.24±7.44	91.31±2.62	71.16±6.17	78.56±3.64	82.07±4.97
	9	80.86±6.51	89.92±6.80	78.66±5.47	77.15±7.98	81.65±6.69
v8	7.5	91.22±5.97	89.08±7.48	75.43±4.95	73.47±4.65	82.3±5.76
	9	80.82±9.99	85.68±7.04	75.18±13.16	74.24±7.35	78.98±9.38
v11	7.5	88.24±6.24	89.14±6.50	76.43±6.55	79.08±5.91	83.23±6.3
	9	83.59±8.15	85.28±10.36	72.29±7.95	74.77±8.61	78.98±8.77

Table 4.4: model results to boxloss(b.loss) summarization: mean and standard deviation (%) of mAP50 computed on Validset with 10-fold cross-validation for the main 2 boxloss values: 7.5 and 9

mAP 50-90

YOLO	b.loss	MD	ME	PD	PE	tot
v5	7.5	56.07±5.16	65.10±3.28	33.38±4.36	33.33±2.76	46.97±3.89
	9	59.12±7.24	62.55±7.90	42.43±6.38	42.71±2.95	51.7±6.12
v8	7.5	59.86±5.83	64.56±3.86	35.79±4.44	31.69±3.23	47.98±4.34
	9	56.91±10.02	62.82±7.06	39.31±9.96	40.57±4.91	49.9±7.99
v11	7.5	63.33±5.42	64.66±3.52	42.83±6.60	40.46±5.11	52.82±5.16
	9	60.82±6.37	61.73±6.32	38.75±4.32	39.81±5.29	50.28±5.58

Table 4.5: model results to boxloss(b.loss) summarization: mean and standard deviation (%) of mAP50-90 computed on Validset with 10-fold cross-validation for the main 2 boxloss values: 7.5 and 9

Finally, the fine-tuned models with best result on testing set were utilized in combination with SAM. Because there isn't a YOLO version that outperform the others, each of them were transferred to the following procedure.

mAP 50

YOLO	b.loss	MD	ME	PD	PE	tot
v5	7.5	85.33±9.49	90.11±4.86	69.72±9.57	78.42±3.0	80.89±6.73
	9	87.77±5.58	90.82±3.9	76.41±3.62	75.97±3.5	82.74±4.15
v8	7.5	80.91±8.88	85.32±6.32	69.42±7.81	74.87±7.28	77.63±7.57
	9	84.61±12.5	88.44±5.86	73.95±7.92	73.78±5.96	80.19±8.07
v11	7.5	89.99±4.52	90.60±6.20	69.64±3.53	72.34±3.72	80.64±6.93
	9	93.74±6.00	92.67±6.39	73.89±8.25	77.89±7.09	84.44±4.49

Table 4.6: model results : mean and standard deviation (%) of mAP50 computed on Testset with 10-fold cross-validation for the main 2 boxloss values: 7.5 and 9

mAP 50-90

YOLO	b.loss	MD	ME	PD	PE	tot
v5	7.5	53.03±6.45	62.35±4.62	32.93±4.51	32.50±3.37	45.2±4.73
	9	56.08±3.76	63.71±4.62	37.68±4.21	32.04±1.58	47.38±3.54
v8	7.5	55.52±5.80	62.76±5.86	31.16±3.79	31.18±3.61	45.15±4.76
	9	58.24±8.21	64.49±4.09	33.98±4.84	30.11±3.82	46.70±5.24
v11	7.5	60.11±5.01	64.91±4.22	31.11±5.46	30.77±2.53	46.72±4.3
	9	63.45±6.65	68.83±4.11	34.54±3.19	32.21±2.87	49.76±4.21

Table 4.7: model results to boxloss(b.loss) summarization: mean and standard deviation (%) of mAP50-90 computed on Testset with 10-fold cross-validation for the main 2 boxloss values: 7.5 and 9

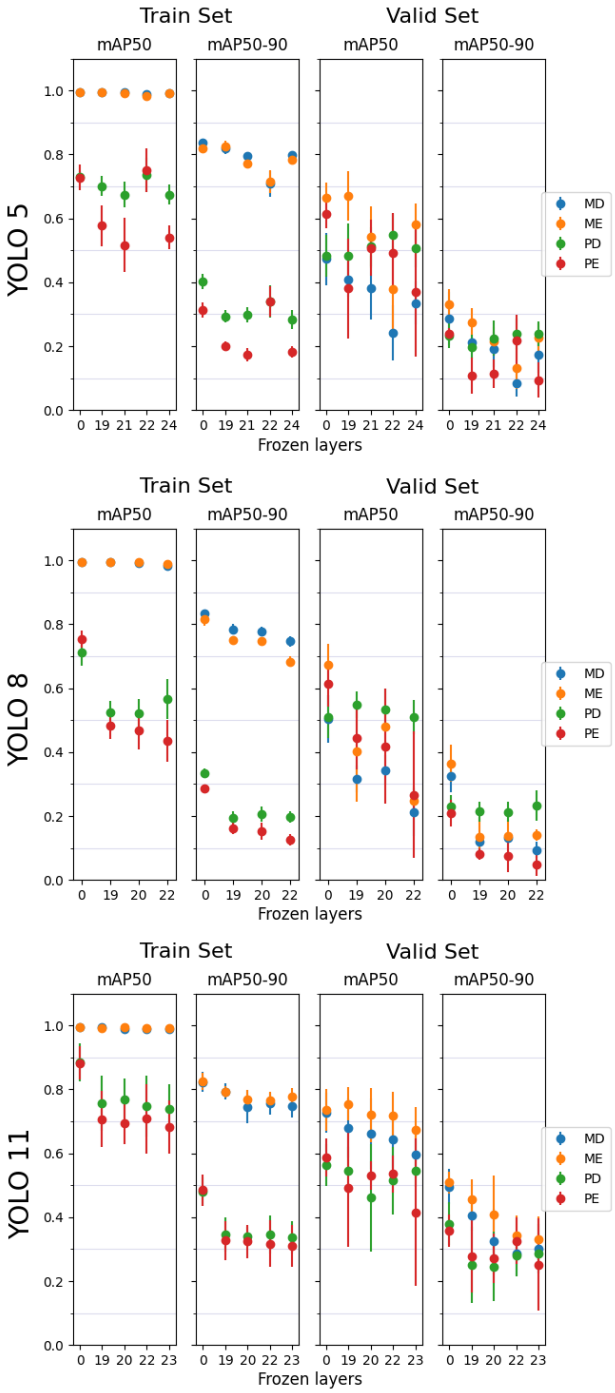

Figure 4.1: This figure shows the influence of freezing on the YOLO's performance (express as mean and variance of mAP50 and mAP50-90) on the Train and the Validation set, without appling the dropout

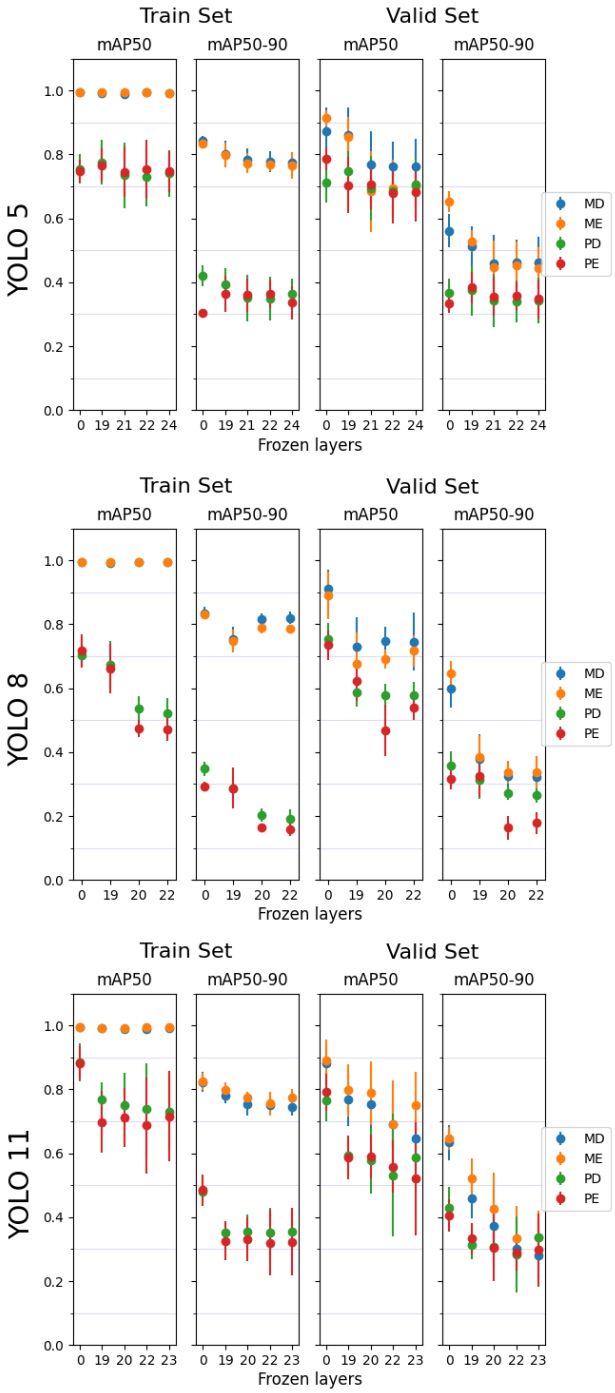

Figure 4.2: This figure shows the influence of freezing on the YOLO's performance (express as mean and variance of mAP50 and mAP50-90) on the Train and the Validation set, appling a dropout of 0.5

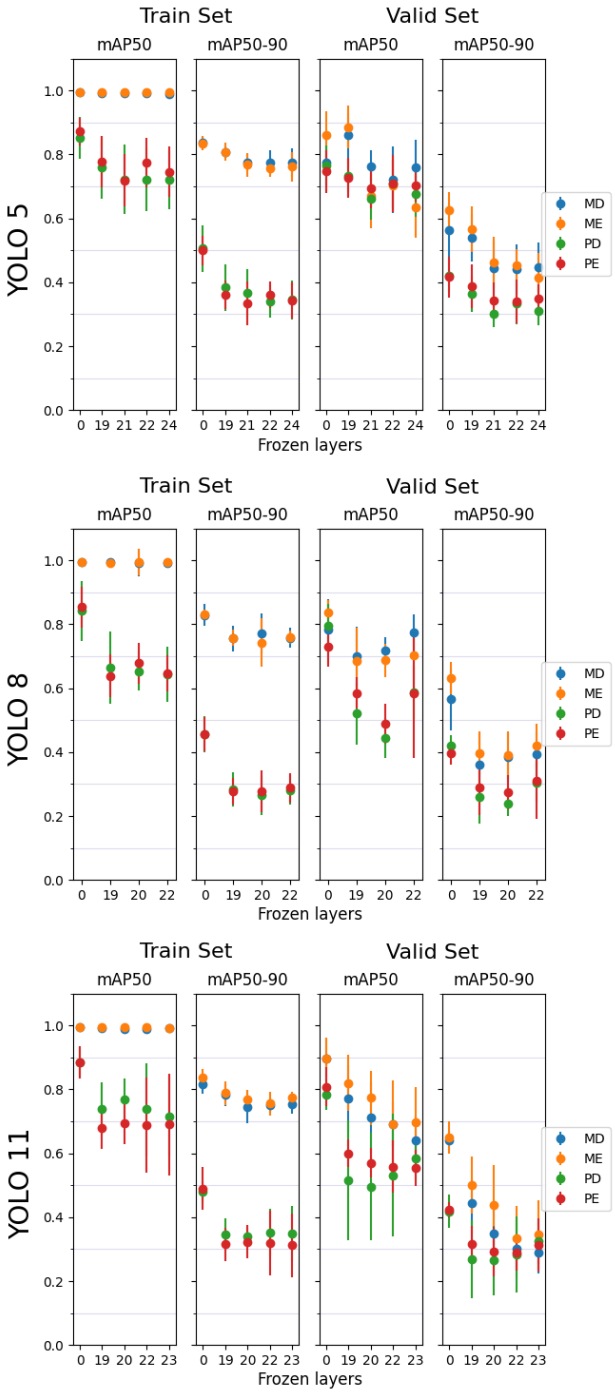

Figure 4.3: This figure shows the influence of freezing on the YOLO's performance (express as mean and variance of mAP50 and mAP50-90) on the Train and the Validation set, appling a dropout of 0.7

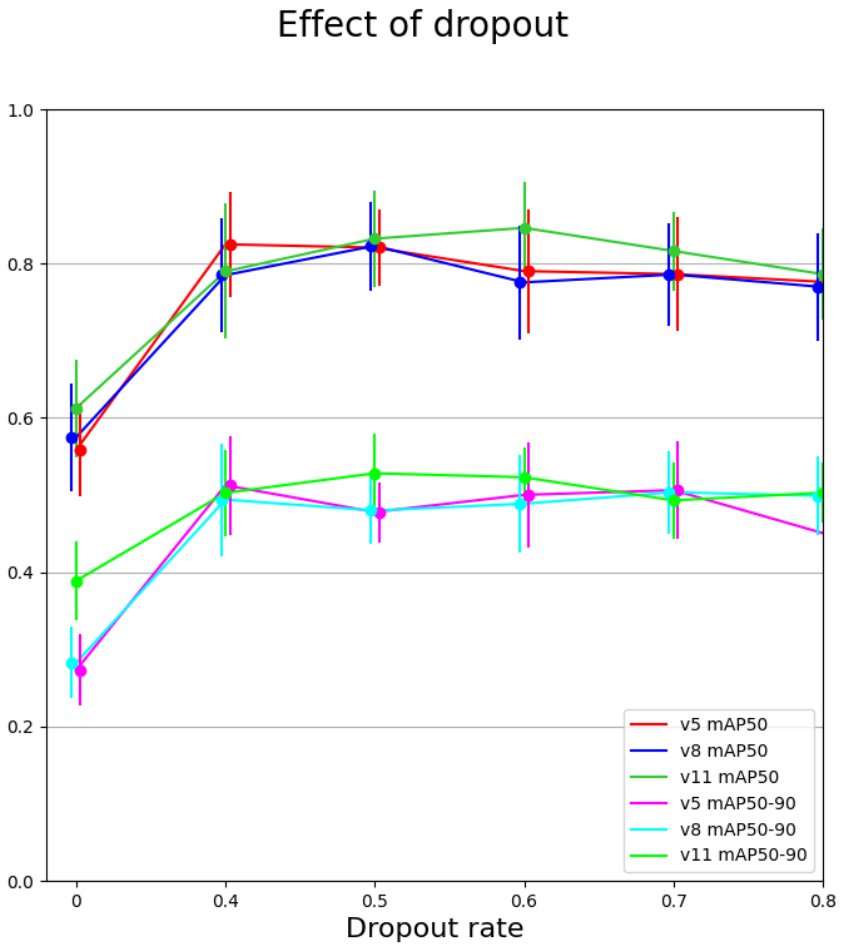

Figure 4.4: The influence of dropout on the performance of the three YOLO models, computed over the validation set

4.2 SAM results

Regarding SAM, the model was initially tested with ground-truth boxes as prompt to obtain the overall performance. These are resumed in the table below Tab4.8 and they were treated as a reference point for further experiments. Obviously this results were execute on the testing set. In order to obtain the boundaries and compute the Hausdorff Distances (HD and HD95), a laplacian edge detection filter was applied to the masks generated by SAM. This technique highlights regions of rapid intensity change, regardless of the direction, by calculating the second-order derivative of an image, resulting in an edge detector.
Although SAM achieved a satisfactory level in every proposed scores, a great variance was registered in every class. Also in this case there are a divergence between breasts values and nipples values. The top left graph in Fig.4.5 illustrates it perfectly.
Comparing these initial results with the work of Jun Ma et al., emerge the potentialities that SAM, in the general setting of this study, could have without a profound structural change.

Although accuracy (Acc) was not initially considered a valid metric, due to the nature of the task, ultimately the decision was made to include it, in order to enable a more comprehensive comparison with other studies.
In our case, the objective include also to identify the presence of very small objects, relative to the overall dimensions of the image, as nipples. This disproportion negatively affect the reliability of accuracy.
As it was expecting, the model obtains the higher value on breasts and a reduced accuracy on nipples.

Then the combination of the two computational tools(i.e. the proposed method) was tested, feeding SAM the bounding boxes predicted by YOLO models. All of the three fine-tuned model was implemented in this phase, since the little performance difference manifested in the previous experiments.
These last outcomes are reported in Tab4.9 and Tab4.10. The discussion on results can be divided between the two metrics categories. The first aspect that stands out is the quite overlap between the results of v5 and those of v8, for Jaccard's Index and Dice coefficient, indicating that using either of these versions as a prompter makes no real difference. In contrast, v11 exhibits improved statistics, which, with regard to breast classes, are comparable to those reported for the ground truth.
A similar argument can be made for Hausdorff Distances, with the exception that the models perform better on nipples classes. This is rather implausible and is likely attributable to an internal bias of Hausdorff metrics, since nipples have a shorter perimeter compared to breast. For this reason, HD and HD95 will parsimony be

taken into account in the decision-making process.

Being more specific for each class: with the respect to the values obtained by the ground-truth boxes, generated boxes of breasts reported a reduction of region-based metrics mean by about 15% and an increase of boundary-based by slightly less than threefold, for v5 and v8. In opposite, breast variability for all the metrics, do not diverge substantially from the initial benchmark.
While for the nipples classes, SAM model returned a substantial different performance, especially for the left nipple.
For example, the total collapse of Jaccard's Indexes and Dice coefficients means is evident, especially in v5 and v8; the average IoU in v8, for PD and PE classes, reached respectively about the 69% and 24% of the desirable maximum value, while the average F1 reached respectively about the 71% and 25%. Although with less strong reductions, similar trend is observable in v11.
The same general behavior appears even with the Hausdorff distances. This is clearly shown comparing the graphs in figure Fig.4.5.

The great difference in the results obtained from the two type of metrics, could be explained by the generation of an irregular border of the mask. Indeed, a very high IoU indicates that the two areas sufficiently overlap, whereas a very high HD is correlated with a large average distance between the two boundaries. Furthermore, the large variability observed in HDs measurements suggests low accuracy, from the model, in the delineation of the masks boundaries. Jun Ma et al. reported that SAM struggles with targets of weak border or low contrast, which make it prones to under or over-segmentation. This could be a similar case.
To support this argumentation some mask, where a jagged edge is clear visible, are shown. YOLO11, having achieved the best statistics in all the four metrics, seems less affected by this problem.

Metrics	MD	ME	PD	PE	tot
IoU	86.58±5.58	86.19±5.64	73.12±8.9	73.99±10.46	79.97±7.64
F1	93.71±3.29	93.48±3.31	84.16±5.97	84.64±6.92	88.95±4.87
HD	12.26±5.05	13.99±6.06	1.4±0.53	1.44±0.54	7.27±3.05
HD95	9.44±4.34	10.88±5.02	1.33±0.43	1.33±0.44	5.75±2.56

Table 4.8: SAM results with ground-truth boxes as prompt: mean and standard deviation of the four metrics computed on Testset. IoU and F1-score are expressed in %

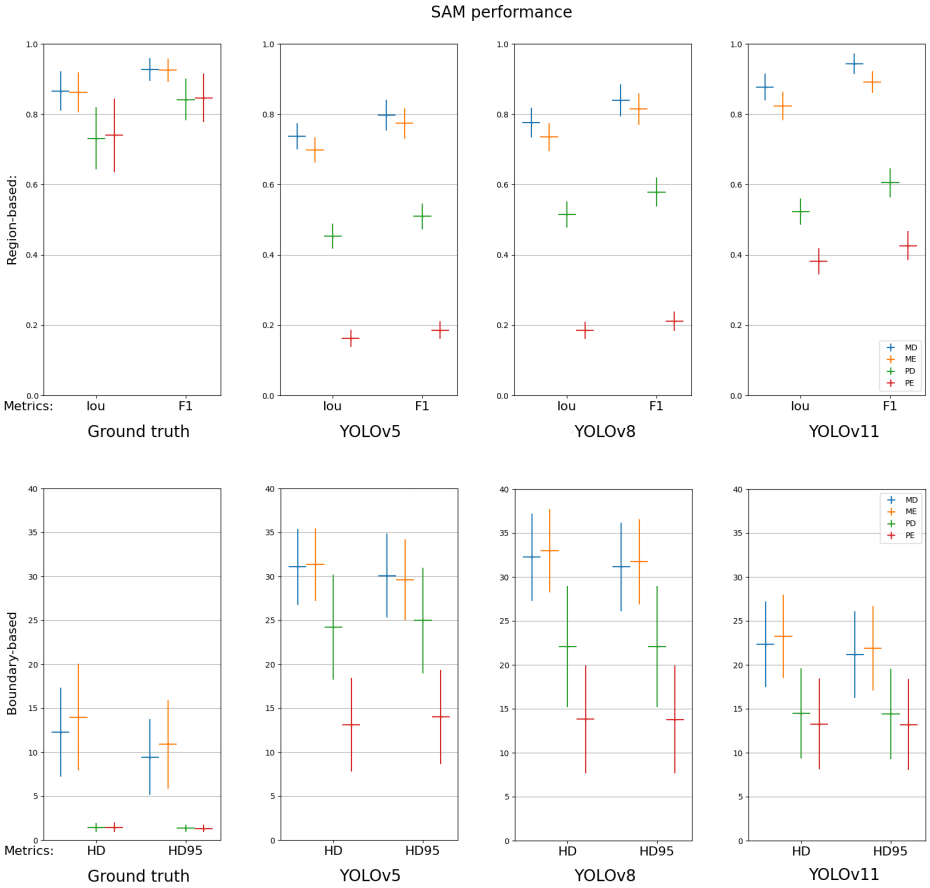

Figure 4.5: Comparison between zero-shot performance of SAM model with different prompt geneator: humans and YOLOs. Region-based evaluation method are above, Boundary-based evaluation method are below.

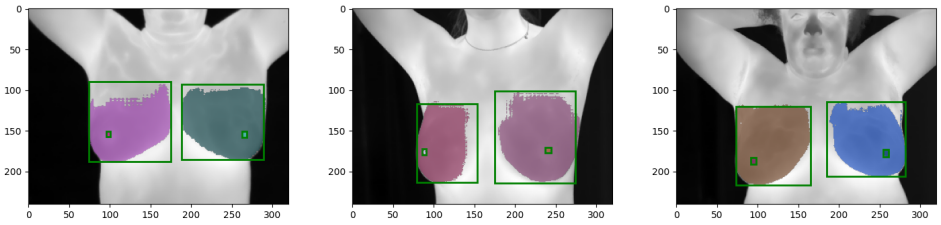

Figure 4.6: Examples of masks with irregular border, generated by SAM.

Region-based metrics

YOLO	Metrics	MD	ME	PD	PE	tot
v5	IoU	73.67±3.72	69.81±3.54	45.29±3.54	16.26±2.39	51.26±3.3
	F1	79.74±4.38	77.36±4.26	51±3.61	18.55±2.47	56.66±3.68
v8	IoU	77.55±4.23	73.49±4.02	51.46±3.69	18.48±2.49	55.25±3.61
	F1	83.93±4.57	81.43±4.43	57.84±4.10	21.09±2.81	61.07±3.98
v11	IoU	87.09±3.88	82.33±4.03	52.28±3.70	38.14±3.74	65.12±3.84
	F1	93.94±2.17	89.11±4.32	60.47±4.17	42.58±4.13	71.78±4.2

Table 4.9: SAM results with YOLO generated boxes as prompt: mean and standard deviation (%) of the two region-based metrics computed on Testset

Boundary-based Metrics

YOLO	Metrics	MD	ME	PD	PE	tot
v5	HD	31.08±4.32	31.36±4.13	24.22±5.99	13.12±5.33	24.95±4.92
	HD95	30.09±4.79	29.62±4.59	24.98±5.98	14.0±5.33	24.67±5.17
v8	HD	32.27±4.97	33.01±4.74	22.09±6.88	13.82±6.13	25.3±5.68
	HD95	31.14±5.04	31.74±4.83	22.08±6.88	13.78±6.13	24.68±5.72
v11	HD	22.35±4.87	23.25±4.74	14.48±5.14	13.26±5.15	18.33±4.97
	HD95	21.16±4.92	21.9±4.80	14.39±5.13	13.19±5.15	17.66±5.0

Table 4.10: SAM results with YOLO generated boxes as prompt: mean and standard deviation of the two boundary-based metrics computed on Testset

4.3 Discussion

This section confronts the proposed method with approaches from the existing literature presented in the previous chapter. Studies relying on different evaluation metrics were excluded, as they do not allow for a direct comparison.
Among the collected documentation, the following works emerge:

- the Tayel and Elbagoury's manuscript[113]. They use a FCN with pre-trained parameters to segment the breast outline, including inside the ROI both breast, armpits and shoulders.
- the assessment work on U-Net and its variations by Rosli and Habaebi[114].
- the two studies by S.Guan et.al[115] [116] on C-DCNN and MultiResUnet respectively. They reach a good results despite the small dataset.
- the paper on 4D U-Net of Goumati et.al [118]. They pre-processed the

image with APPDRC method, removing the speckle noise; then the net was optimized by glowworm swarm algorithm to exact disease affected portion of thermograms.

The performance values reported in these works are presented in Tab4.11. Additionally, the studies by Jun Ma et al.[119], Y.Huang et al.[121] and Wu et al.[120] were taken into consideration, as they employ SAM in the medical domain. Finally the undoubtedly necessary confrontation with the Pandey's article[123], where the combination of YOLO and SAM was originally proposed.

Although the method proposed by this thesis reports inferior performance, it tackles a more complex multi-class problem. Whereas the majority of the listed studies address a simpler binary segmentation task (foreground vs. background). Thus the reduced accuracy, IoU, and F1 are in agreement with the increased class discrimination difficulty.
Indeed, when focusing only on breast classes, the results are comparable to those reported in the literature, as shown in Tab.4.11.
Comparing tables Tab4.8 and Tab4.11 is logically justified that the method under consideration begins with an inherent disadvantage. The YOLO+SAM pipeline's outcomes is particularly affected by the detection of the nipples classes, probably because they consist of small objects. Each convolutional layer compresses spatial information, so objects with small footprints can be lost after multiple layers and fail to be detected.

Furthermore, four out of five analyzed studies train their models on databases composed of images with a resolution twice than the one used in this project.

Model	Author	**Acc**	**IoU**	**F1**
FCN	Tayel et al.	96.4	95.4	97.6
U-Net	Rosli et al.	96.37	92.92	96.30
C-DCNN	Guan et al.	89.7	83.4	90.9
MultiResU-Net	Guan et al.	91.47	85.07	//
4D U-Net	Gomathi et al.	98	//	//
YOLOv5+SAM		90.05	71.75	78.55
YOLOv8+SAM		90.13	75.52	82.26
YOLOv11+SAM		94.06	84.71	91.53

Table 4.11: Performances comparison between our proposed method and other segmentation methods. Reported evaluation considers only breast classes.

Jun Ma et al. and Wu et al. reported an overall median Dice Similarity Coefficient of about 82% and 89.8% .
The first group developed MedSAM, a fine-tuned SAM on a large-scale medical image dataset of 1,570,263 images, covering 10 imaging modalities and over 30 cancer types. While the second, in order to introduce domain-specific medical knowledge into SAM, devised two adaptation techniques: one for close volumetric correlation and the other for visual prompt-conditioned adaptation. Wu et al. trained their model on 5 different database, containing up to 5000 images.

Although our method not required an intensive training of SAM architecture, since based upon prompt functionality, the Dice score achieved is 91.53%, outperforming the two teams'models. It is worth pointing out, neither Jun Ma et al. nor Wu et al. employed IRT among their selected imaging modalities, nor did they examine breast cancer.
Even Y. Huang et al., that reached a general Dice coefficients of 84.86% and 84.49% on ViT-H and ViT-L, respectively, excluded thermography from their data. Albeit, both this and their study employed SAM for zero-shot segmentation with box prompting, they manually specified the bounding boxes, whereas we generated them automatically.

Pandey et al. used the combination of YOLOv8 and SAM for lung segmentation, obtaining a Dice score of 0.9012 ± 0.0633 in frontal chest radiography and 0.8799 ± 0.131 in transversal computed tomography. Since our data are largely consistent with those of the previous study, it can be reasonably concluded that this methodology is independent of the imaging technique employed. Full confrontation is expressed in the table Tab4.12 below

Model	**Tecnique**	**Acc**	**IoU**	**F1**
YOLOv8+SAM	X-ray	//	82.02	90.12
YOLOv8+SAM	CT	//	78.55	87.99
YOLOv5+SAM	IRT	90.05	71.75	78.55
YOLOv8+SAM	IRT	90.13	75.52	82.26
YOLOv11+SAM	IRT	94.06	84.71	91.53

Table 4.12: Performances comparison between our work on IRT images and the segmentation work of Pandey et al. The reported data are expressed in percentage.

Chapter 5

Conclusion

5.1 Conclusion

The objective of this work was to explore the potentiality of a new paradigm for the segmentation of thermal images of breast cancer using pre-trained convolutional models. A novel IR dataset was used for this purpose.
The pipeline consists of two architectures, YOLO and SAM, each fulfilling a distinct role: the first is an object detector, while SAM performs the actual segmentation. This last one is know as a powerful tool in zero-shot segmentation.
The framework exhibits an initial potential in breast localization; however, its performance was affect by various limitations. With regard to object detection the best result was achieved for YOLOv11 with an mAP50 of $84.44 \pm 4.49\%$ and a mAP50-90 of $49.76 \pm 4.21\%$ over all the four class. While in segmentation task, SAM achieved overall metrics of IoU and F1-score of $65.12 \pm 3.84\%$ and $71.78 \pm 4.2\%$ rispectevely, and Hausdorff distance of 18.33 ± 4.97 and an 95th-percentile Hausdorff distance of 17.66 ± 5.0 on the proposed dataset.
Although YOLO ultimately achieved results that were overall in line with expectations, SAM performance is highly dependent on the version of YOLO used and, considering the complete task, did not perform greatly. At present, there is no definitive evidence explaining why one version outperforms the others; however, it can be hypothesized that this difference may be attributable to the recently introduced attention mechanism in v11.
Considering the clear discrepancy between classes results, the pipeline demonstrating its reliability in differentiating right and left breast, but struggling in recognize nipples. Indeed breast classes obtained segmentation score at least 19% better on detection task while at max 80% better on segmentation score with respect to nipples.
Since this difference persist, albeit to a lesser extent, even using ground-truth as

prompting, it can be reasonably assumed that is an inherent SAM difficulty.
However, the main limitation of the conducted study may lie in the dataset. The limited number of available thermograms could have influenced the results, and the quality of them has undoubtedly posed a significant obstacle.
In conclusion the paradigm presented in this project has shown good promise as a step toward a simpler breast segmentation system for thermographic image, because, despite the problematicness, it allows to avoid costly training on a new architecture.

Future directions could include exploring advanced augmentation techniques, updating the pipe-line either with newer version or a complete different architectures. Nonetheless, the need for additional work to further refine the pipeline, improve generalizability, and reduce any biases introduced by the dataset or methodology, is take into account for further improvements.

List of Tables

List of Figures

Appendix A

A1

```
!pip install ultralytics
!pip install pyyaml

from random import shuffle, seed as randseed, randint
from appoggio import YAM, mytestfunc, myboxplot
import matplotlib.pyplot as plt
from ultralytics import YOLO
from pathlib import Path
import numpy as np
import torch
import os
import yaml

#- INITIALIZATION -
# Prepare all the variable for the cross-validation
# Note: inside the folder 'thesis_prj1', there are
# five objects:
#   -   dataset, folder storing all the images
#   -   train, folder that will contain the training set (198)
#   -   valid, folder that will contain the validation set (22)
#   -   test, folder that will contain the testing set (24)
#   -   data.yaml, network configuration file

HOME = os.getcwd()
print(HOME) #check your current directory
dataset_location = f"{HOME}/thesis_prj1"
Images_name_list = os.listdir(f"{dataset_location}/dataset")
randseed(0)
shuffle(Images_name_list)

#get a random 10\% for test
```

```
    foldlen = 22
    imgs2test = Images_name_list[-foldlen-2:]
    file_path = Path(f"{dataset_location}/dataset")
    #print(imgs2test)  #check your testset

    train_results = np.zeros([10, 2, 4])
    valid_results = np.zeros([10, 2, 4])
    test_results = np.zeros([10, 2, 4])

    # - CROSS - VALIDATION -
    # Taken the image names list from the 'dataset' folder,
    # the images are distributed and copied in their respective
    # folders.

    testFlag = True
    for k in range(10): #10-fold validation
        imgs4valid = Images_name_list[foldlen*k:foldlen*(k+1)]
        for img in Images_name_list:

            if img in imgs2test :
                if testFlag==True:
                    folder = 'test'
                else: continue #jump to next img
            elif img in imgs4valid:
                folder = 'valid'
            else:
                folder = 'train'

            dest_path = Path(f"{dataset_location}/{folder}/images")
            if not dest_path.exists(): dest_path.mkdir()
            os.system(f"cp {file_path/img} {dest_path/img}")

        testFlag = False



        # -OBJECT DETECTION-
        # After reorganizing the images, in the following section
        # the configuration file 'data.yaml' is set, through YAM.
        # Then, the native (pretrained) model is loaded, and the
        # various settings for training are adjusted.
        # In the section titled 'Data Augmentation Settings',
        # the variables related to data augmentation are defined. In
        # this case, the majority of augmentation parameters have
        # been set to zero.

        f = 0    # specifies the number of the first N layers
                 # to be frozen.
        B = 16   # indicates the batch size.
```

```
        drop=0.5 # denotes the dropout rate
        bl = 9   # indicates the influence of 'box_loss' inside total
                 # loss function
        YAM(dataset_location)
        # change below with 'yolov5su.pt' or 'yolo11.pt' if necessary
        model = YOLO(f'yolov8n.pt')
        model.info()
        Th0  = randint(0,100)
        namef = f"f={f}_k={k}_{Th0}"
        results = model.train(
                        data=dataset_location+"/data.yaml",
                        imgsz=(320,240),
                        seed=Th0,
                        single_cls=False,
                        name= f"train_{namef}",

        # -Optimizier's Hyperparameter-
                        optimizer='AdamW',  #choose between SGD, Adam,
                                            #RAdam, AdamW, NAdam, RMSprop
                        epochs= 150,
                        patience= 30,
                        batch=B,
                        freeze=f,
                        dropout=drop,
                        box = bl,
                        #lr0=0.01,
                        #lrf=0.01,
                        #weight_decay=0.0005, #L2 regularization term
                        #momentum=0.9,

        # -Data Augumentation settings-
                        scale=0.3,
                        fliplr=0.0,
                        mosaic=0,
                        erasing=0.3,
                        crop_fraction=0.3,
                        hsv_h=0,
                        hsv_s=0,
                        hsv_v=0,
        )

    # yolo output is a tensor that contains the bounding box
    # coordinates, objectness score, and class probabilities.

        train_results[k, 0, :] = results.box.ap50
        train_results[k, 1, :] = results.box.ap

    # -Evaluation-
```

```
129         mydevice = torch.device('cuda:0' if torch.cuda.is_available()
    else 'cpu')
130         results = model.val(data=dataset_location+"/data.yaml",
131                             name=f"valid_{namef}",
132                             imgsz=640, batch=16,
133                             conf=0.6, iou=0.6, device=mydevice)
134         valid_results[k,0,:] = results.box.ap50
135         valid_results[k,1,:] = results.box.ap
136 
137 
138     # - TESTING -
139         test_results[k,:, :] = mytestfunc(f"train_{namef}")
140 
141         # destroy train and valid folder
142         for folder in ['train','valid']:
143             dest_path = Path(f"{dataset_location}/{folder}/images")
144             os.system(f"rm -rf {dest_path}")
145 
146 
147     # sum up the list's results
148     print('mean_train =',np.mean(train_results, axis=0))
149     print('std_train =',np.std(train_results, axis=0))
150     print('mean_valid =',np.mean(valid_results, axis=0))
151     print('std_valid =',np.std(valid_results, axis=0))
152     print('mean_test =',np.mean(test_results, axis=0))
153     print('std_test =',np.std(test_results, axis=0))
154 
155     myboxplot(train_results)
156     myboxplot(valid_results)
157     myboxplot(test_results)
```

Bibliography

[1] Sung H, Ferlay J, Siegel R.L., et al. «Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries». In: *CA Cancer J Clin* 71 (May 2021), pp. 209–249. DOI: 10.3322/caac.21660 (cit. on pp. 1, 7).

[2] Wilkinson L. and Gathani T. «Understanding Breast Cancer as a Global Health Concern». In: *The British Journal of Radiology* 95 (Dec. 2022), p. 1130. DOI: 10.1259/bjr.2021103 (cit. on p. 1).

[3] *Breast Cancer Screening: Thermogram No Substitute for Mammogram*. US Food and Drug Administration. Available online: https://www.fda.gov/consumers/consumer-updates/breast-cancer-screening-thermogram-no-substitute-mammogram, 2017 (cit. on pp. 1, 17).

[4] Valerie P. Jackson. «Diagnostic mammography». In: *Radiologic clinics of North America* 42 (Sept. 2004), pp. 853–70. DOI: 10.1016/j.rcl.2004.06.002 (cit. on pp. 1, 9).

[5] Pace L.E. and Keating N.L. «A systematic assessment of benefits and risks to guide breast cancer screening decisions». In: *JAMA* 311 (Apr. 2014), pp. 1327–35. DOI: 10.1001/jama.2014.1398 (cit. on pp. 1, 9).

[6] Kim J., Harper A., et al. «Global Patterns and Trends in Breast Cancer Incidence and Mortality across 185 Countries». In: *Nature Medicine* 31 (Feb. 2025), pp. 1154–1162. DOI: 10.1038/s41591-025-03502-3 (cit. on p. 1).

[7] Kakileti S.T., Manjunath G., et al. *New Perspectives in Breast Imaging*. London, UK: IntechOpen, Oct. 2017. ISBN: 9789535135586 (cit. on p. 1).

[8] Heywang-Köbrunner S.H., Hacker A., and Sedlacek S. «Advantages and Disadvantages of Mammography Screening». In: *Breast Care* 6 (May 2011), pp. 199–207. DOI: 10.1159/000329005 (cit. on pp. 1, 9).

[9] Feig SA. «Mammography equipment: principles, features, selection». In: *Radiologic clinics of North America* 25 (Sept. 1987), pp. 897–911. DOI: 3306772 (cit. on p. 2).

[10] Verbeek, André L. M., et al. «Mammographic screening: Keeping women alive». In: *Women's health (London, England)* 7 (June 2011), pp. 631–3. DOI: 10.2217/whe.11.73 (cit. on p. 4).

[11] DeVita, Hellmann, Lawrence, and Rosenberg. *Oncologia*. Piccin, 2017 (cit. on p. 5).

[12] Hanahan D and Weinberg R.A. «The hallmarks of cancer». In: *Cell* 100 (Jan. 2000), pp. 57–70. DOI: 10.1016/s0092-8674(00)81683-9 (cit. on p. 5).

[13] Rosen RD. and Sapra A. *TNM Classification. [Updated 2023 Feb 13]*. Treasure Island (FL): StatPearls Publishing, Jan. 2025. ISBN: Available from: https://www.ncbi.nlm.nih.gov/books/NBK553187/ (cit. on p. 5).

[14] Harbeck N., Penault-Llorca F., Cortes J., et al. «Breast cancer». In: *Nature Reviews Disease Primers* 5 (July 2019), p. 66. DOI: \url{https://doi.org/10.1038/s41572-019-0111-2} (cit. on p. 5).

[15] Lin F. and Chen Z. «Standardization of diagnostic immunohistochemistry: literature review and geisinger experience». In: *Archives of pathology & laboratory medicine* 138 (Dec. 2014), pp. 1564–1577. DOI: 10.5858/arpa.2014-0074-RA (cit. on p. 6).

[16] Irvin J. William Jr. and Lisa A. Carey. «What is triple-negative breast cancer?» In: *European Journal of Cancer* 44 (Dec. 2008), pp. 2799–2805 (cit. on p. 6).

[17] Collaborative Group on Hormonal Factors in Breast Cancer. «Familial breast cancer: collaborative reanalysis of individual data from 52 epidemiological studies including 58,209 women with breast cancer and 101,986 women without the disease». In: *Lancet* 389 (Oct. 2001), pp. 1389–1399. DOI: 10.1016/S0140-6736(01)06524-2 (cit. on p. 6).

[18] Venkitaraman A.R. «Cancer susceptibility and the functions of BRCA1 and BRCA2». In: *Cell* 108 (Feb. 2002), pp. 171–182. DOI: 10.1016/s0092-8674(02)00615-3 (cit. on p. 6).

[19] Shiovitz S. and Korde L.A. «Genetics of breast cancer: a topic in evolution». In: *Annals of Oncology* 26 (July 2015), pp. 1291–1299. DOI: 10.1093/annonc/mdv022 (cit. on p. 6).

[20] Zweemer R.P., Verheijen R.H., and Menko F.H. et al. «Differences between hereditary and sporadic ovarian cancer». In: *European Journal of Obstetrics and Gynecology and Reproductive Biology* 82 (Feb. 1999), pp. 151–153. DOI: 10.1016/s0301-2115(98)00218-8 (cit. on pp. 6, 8).

[21] *World health statistics 2025: monitoring health for the SDGs, Sustainable Development Goals*. licence: CC BY-NC-SA 3.0 IGO. World Health Organization. Genevra, 2025 (cit. on p. 7).

[22] Youlden D.R., Cramb S.M., Dunn N.A., Muller J.M., Pyke C.M., and Baade P.D. «The descriptive epidemiology of female breast cancer: an international comparison of screening, incidence, survival and mortality». In: *Cancer Epidemiology* 36 (June 2012), pp. 237–48. DOI: 10.1016/j.canep.2012.02.007 (cit. on p. 7).

[23] Ginsburg O., Bray F., Coleman M.P., et al. «The global burden of women's cancers: a grand challenge in global health». In: *Lancet* 389 (May 2017), pp. 847–860. DOI: 10.1016/S0140-6736(16)31392-7 (cit. on p. 7).

[24] Althuis M.D., Dozier J.M., Anderson W.F., Devesa S.S., and Brinton L.A. «Global trends in breast cancer incidence and mortality 1973-1997». In: *Int Journal Epidemiology* 34 (Apr. 2005), pp. 405–12. DOI: 10.1093/ije/dyh414 (cit. on p. 7).

[25] Carol E. DeSantis, Jiemin Ma, Mia M. Gaudet, et al. «Breast cancer statistics, 2019». In: *CA: A Cancer Journal for Clinicians* 69 (Nov. 2019), pp. 438–451 (cit. on p. 7).

[26] *I numeri del cancro in Italia 2023*. Associazione italiana di oncologia medica. Roma, 2023 (cit. on p. 8).

[27] Ruder E.H., Dorgan J.F., Kranz S., Kris-Etherton P.M., and Hartman T.J. «Examining Breast Cancer Growth and Lifestyle Risk Factors: Early Life, Childhood, and Adolescence». In: *Clin Breast Cancer* 8 (2008), pp. 334–42. DOI: 10.3816/CBC.2008.n.038 (cit. on p. 8).

[28] Salvatorelli L., Puzzo L., et al. «Ductal Carcinoma In Situ of the Breast: An Update with Emphasis on Radiological and Morphological Features as Predictive Prognostic Factors». In: *Cancers* 12 (Mar. 2020), p. 609. DOI: https://doi.org/10.3390/cancers12030609 (cit. on p. 9).

[29] Malur S., Wurdinger S., Moritz A., and other. «Comparison of written reports of mammography sonography and magnetic resonance mammography for preoperative evaluation of breast lesions with special emphasis on magnetic resonance mammography». In: *Breast Cancer Research* 3 (Dec. 2000), p. 55. DOI: https://doi.org/10.1186/bcr271 (cit. on p. 9).

[30] Sardanelli F., Aase H.S., Alvarez M., Azavedo E., Baarslag H.J., Balleyguier C., et al. «Position paper on screening for breast cancer by the European Society of Breast Imaging (EUSOBI)». In: *Eur Radiol.* 27 (July 2017), pp. 2737–43. DOI: 10.1007/s00330-016-4612-z (cit. on p. 9).

[31] Hofvind S., Ponti A., Patnick J., and other. «False-positive results in mammographic screening for breast cancer in Europe: a literature review and survey of service screening programmes». In: *Journal of Medical Screening* 311 (2012), pp. 57–66. DOI: 10.1258/jms.2012.012083 (cit. on p. 9).

[32] Elmore J.G., Barton M.B., Moceri V.M., and other. «Ten-year risk of false positive screening mammograms and clinical breast examinations». In: *The New England journal of medicine* 338 (Apr. 1998), pp. 1089–1096. DOI: 10.1056/NEJM199804163381601 (cit. on p. 9).

[33] Yankaskas B.C., Klabunde C.N., Ancelle-Park R., and other. «International comparison of performance measures for screening mammography: can it be done?» In: *Jounal of Medical Screening* 11 (Apr. 2004), pp. 187–193. DOI: 10.1258/0969141042467430 (cit. on p. 9).

[34] Miglioretti D.L., Abraham L., Sprague B.L., and other. «Association Between False-Positive Results and Return to Screening Mammography in the Breast Cancer Surveillance Consortium Cohort». In: *Annals Intern Med* 117 (Oct. 2024), pp. 1297–1307. DOI: 10.7326/M24-0123 (cit. on p. 9).

[35] Osako T., Iwase T., Takahashi K., and other. «Diagnostic mammographyand ultrasonography for palpable and nonpalpable breast cancerin women aged 30 to 39 years». In: *Breast Cancer* 14 (Mar. 2007), pp. 255–259 (cit. on p. 9).

[36] Carney P.A., Miglioretti D.L., Yankaskas B.C., and other. «Individual and combined effects of age, breast density, and hormone replacement therapy use on the accuracy of screening mammography». In: *Ann Intern Med* 138 (Mar. 2003), pp. 168–75 (cit. on pp. 9, 10).

[37] Østerås B.H., Martinsen A.C.T., Gullien R., and other. «Digital Mammography versus Breast Tomosynthesis: Impact of Breast Density on Diagnostic Performance in Population-based Screening». In: *Radiology* 293 (2019), pp. 60–68 (cit. on p. 9).

[38] Checka C.M., Chun J.E., Schnabel F.R., Lee J., and Toth H. «The relationship of mammographic density and age: implications for breast cancer screening». In: *AJR. American Journal of Roentgenology* 198 (Mar. 2012), pp. 292–295. DOI: 10.2214/AJR.10.6049 (cit. on p. 10).

[39] Magny S.J., Shikhman R., and Keppke A.L. *Breast Imaging Reporting and Data System*. Reading, MA: StatPearls Publishing, 2023 (cit. on p. 11).

[40] Lazarus E., Mainiero MB., Schepps B., and other. «BI-RADS lexicon for US and mammography: interobserver variability and positive predictive value». In: *Radiology* 239 (May 2006), pp. 385–391. DOI: 10.1148/radiol.2392042127 (cit. on p. 11).

[41] J.R. Keyserlingk, P.D. Ahlgren, E. Yu, et al. «Functional infrared imaging of the breast». In: *IEEE Engineering in Medicine and Biology* (May 2000), pp. 30–41 (cit. on p. 11).

[42] Crystal P., Strano S.D., Shcharynski S., et al. «Using sonography to screen women with mammographically dense breasts». In: *AJR Am J Roentgenol* 19 (May 2003), pp. 177–182. DOI: 10.1109/51.844381 (cit. on p. 12).

[43] Osako T., Iwase T., Takahashi K., et al. «Diagnostic mammographyand ultrasonography for palpable and nonpalpable breast cancerin women aged 30 to 39 years». In: *Breast Cancer* 14 (Mar. 2007), pp. 255–259. DOI: 10.2325/jbcs.14.255 (cit. on p. 12).

[44] W.A. Berg et al. «Combined screening with ultrasound and mammography compared to mammography alone in women at elevated risk of breast cancer: results of the first-year screen in ACRIN 6666». In: *JAMA J. Am. Med. Assoc.* 18 (June 2008), pp. 2151–63 (cit. on p. 12).

[45] D. Saslow et al. «American cancer society guidelines for breast screening with MRI as an adjunct to mammography». In: *CA. Cancer J. Clin.* 57 (Feb. 2007), pp. 75–89 (cit. on p. 12).

[46] E. Warner et al. «Surveillance of BRCA1 and BRCA2 mutation carriers with magnetic resonance imaging, ultrasound, mammography, and clinical breast examination». In: *JAMA* 292 (Nov. 2007), pp. 1317–1325 (cit. on p. 12).

[47] E.A. Morris et al. «MRI of occult breast carcinoma in a high-risk population». In: *Am. J. Roentgenol.* 181 (Mar. 2003), pp. 619–626 (cit. on p. 12).

[48] A. Mashekova, Y. Zhao, E.Y.K. Ng, et al. «Early detection of the breast cancer using infrared technology – A comprehensive review». In: *Thermal Science and Engineering Progress* 27 (July 2022) (cit. on pp. 13, 15).

[49] N.A. Diakides, M. Diakides, J.C. Lupo, et al. «Advances in medical infrared imaging». In: *Medical Infrared Imaging* (July 2007). CRC Press, Boca Raton, FL, pp. 1-1–13 (cit. on p. 13).

[50] Kurt Ammer and Francis J. Ring. «Standard procedures for infrared imaging in medicine». In: *Medical Infrared Imaging* (July 2007). CRC Press, Boca Raton, FL, pp. 22-1–9 (cit. on p. 13).

[51] Steketee J. «Spectral emissivity of skin and pericardium». In: *Physics in medicine and biology* 18 (1973), pp. 686–694. DOI: 10.1088/0031-9155/18/5/307 (cit. on p. 13).

[52] Ghayoumi zadeh Hossein and Haddadnia Javad. *The New Trends in the Application of Thermography Science for Diagnostic Purposes*. Raseda, CA: Supreme Century, May 2016 (cit. on p. 13).

[53] Gamagami P. «Indirect signs of breast cancer: angiogenesis study». In: *Atlas of Mammography* (1996). Blackwell Science, Cambridge, MA, pp. 231–258 (cit. on p. 14).

[54] S.M. Barman K.E. Barrett S. Boitano and H.L. Brooks. *Ganong's Review of Medical Physiology*. 23rd ed. LANGE medical book. San Francisco, CA: The McGraw-Hill, 2010 (cit. on p. 14).

[55] Anbar M., Brown C., Milescu L., et al. «The potential of dynamic area telethermometry in assessing breast cancer». In: *IEEE Eng Med Biol Magazine* 19 (May 2000), pp. 58–62. DOI: 10.1109/51.844381 (cit. on p. 14).

[56] Coad J. and Duntsall M. *Anatomy and Physiology for Midwives*. 3rd ed. Churchill Livingstone. Edinburgh: Elsevier, 2011 (cit. on p. 14).

[57] Norman AW and Henry HL. *Hormones*. 3rd ed. ISBN 978-0-08-091906-5. Academic Press, 2014 (cit. on p. 14).

[58] Gullino P.M. «Influence of blood supply on thermal properties and metabolism of mammary carcinomas». In: *Annals of the New York Academy of Sciences* 335 (July 1980), pp. 1–21 (cit. on p. 14).

[59] J.R. Keyserlingk, P.D. Ahlgren, E. Yu, and other. «Functional infrared imaging of the breast». In: *IEEE Eng. Med. Biol. Magazine* 67 (Mar. 2000), pp. 30–41 (cit. on p. 15).

[60] Francis S.V., Sasikala M., Bharathi G.B., and other. «Breast cancer detection in rotational thermography images using texture features». In: *Infrared Physics Technology* 67 (Mar. 2014), pp. 490–496 (cit. on p. 15).

[61] Foster K.R. «Thermographic Detection of Breast Cancer». In: *IEEE Engineering in Medicine and Biology Society* 17 (1998), pp. 10–14 (cit. on p. 15).

[62] Gautherie Michel and Charles M. Gros. «Breast thermography and cancer risk prediction». In: *Cancer - American Cancer Society* 45 (July 1980), pp. 51–56 (cit. on p. 16).

[63] Anbar M. «Clinical thermal imaging today». In: *IEEE Eng Med Biolo Magazine* 17 (July 1998). Blackwell Science, Cambridge, MA, pp. 25–33 (cit. on p. 16).

[64] Amalu W., Hobbins W., Head J., and Elliot R. *The Biomedical Engineering Handbook*. 3rd ed. Infrared imaging of the breast—An overview. Baton Rouge, La: CRC Press, 2006 (cit. on p. 16).

[65] Anbar Michael. «Dynamic Thermal Assessment». In: *Medical Infrared Imaging; Diakides, N.A., Bronzino, J.D.* (July 2007). CRC Press, Boca Raton, FL, pp. 8-1–22 (cit. on p. 16).

[66] Amalu W.C. «Non destructive Testing of the Human Breast: The Validity of Dynamic Stress Testing in Medical Infrared Breast Imaging». In: *IEEE Engineering in Medicine and Biology Society* 1 (2004), pp. 1174–1177 (cit. on p. 16).

[67] Li Jiang, W.Zhan, and M.H. Loew. «Modeling static and dynamic thermography of the human breast under elastic deformation». In: *Physics in Medicine I& Biology* 56 (Dec. 2010), p. 187. DOI: 10.1088/0031-9155/56/1/012 (cit. on p. 16).

[68] E.Y.-K. Ng. «A review of thermography as promising non-invasive detection modality for breast tumor». In: *International Journal of Thermal Sciences* 48 (May 2009), pp. 849–859 (cit. on pp. 16, 17).

[69] R. Omranipour, A. Kazemian, et al. «Comparison of the accuracy of thermography and mammography in the detection of breast cancer». In: *Breast Care* 11 (Aug. 2016), pp. 260–264. DOI: 10.1159/000448347 (cit. on p. 17).

[70] Attallah O.A. «Hybrid Trio-Deep feature fusion model for improved skin Cancer classification: merging dermoscopic and DCT images». In: *Technologies* () (cit. on p. 19).

[71] Pacal I., Ozdemir B., et al. «A novel CNN-ViT-Based deep learning model for early skin Cancer diagnosis.» In: *Biomedical Signal Processing and Control* 140 (June 2025). DOI: https://doi.org/10.1016/j.bspc.2025.107627 (cit. on p. 19).

[72] Attallah O.A. «CerCan-net: cervical Cancer classification model via Multi-Layer feature ensembles of lightweight CNNs and transfer learning.» In: *Expert Systems with Applications* 229 (Part B) (Nov. 2023). DOI: https://doi.org/10.1016/j.eswa.2023.120624 (cit. on p. 19).

[73] Elsayed Basel et al. «Deep learning enhances acute lymphoblastic leukemia diagnosis and classification using bone marrow images». In: *Frontiers in oncology* 13 (Dec. 2023). DOI: 10.3389/fonc.2023.1330977 (cit. on p. 19).

[74] Aslan M. F., Sabanci K., and Attallah O. «A framework for lung and Colon cancer diagnosis via lightweight deep learning models and transformation methods». In: *Diagnostics* 12 (Dec. 2022). DOI: https://doi.org/10.3390/diagnostics12122926 (cit. on p. 19).

[75] Bhimavarapu U. and Battineni G. «Deep Learning for the Detection and Classification of Diabetic Retinopathy with an Improved Activation Function». In: *Healthcare (Basel, Switzerland)* 11 (Dec. 2022), p. 29. DOI: 10.3390/healthcare11010097 (cit. on p. 19).

[76] Altameem A., Mahanty C., et al. «Breast Cancer detection in mammography images using deep convolutional neural networks and fuzzy ensemble modeling techniques». In: *Diagnostics* 12 (2022), p. 1812 (cit. on p. 19).

[77] Iqbal M. S., Ahmad W., et al. «Breast Cancer dataset, classification and detection using deep learning». In: *Healthcare (Basel, Switzerland)* 11 (Dec. 2022). DOI: https://doi.org/10.3390/healthcare10122395 (cit. on pp. 19, 34).

[78] Dosovitskiy A., Beyer L., Kolesnikov A., et al. «An image is worth 16x16 words: Transformers for image recognition at scale». In: *arXiv* (June 2021). DOI: arXiv:2010.11929v2 (cit. on p. 20).

[79] Torres-Galván J. C. et al. «Deep convolutional neural networks for classifying breast Cancer using infrared thermography». In: *Quant. InfraRed Thermography Journal* 19 (2022), pp. 283–294. DOI: https://doi.org/10.1080/17686733.2021.1918514 (cit. on pp. 20, 34).

[80] Nogales A., Perez-Lara F., and García-Tejedor Á. J. «Enhancing breast Cancer diagnosis with deep learning and evolutionary algorithms: A comparison of approaches using different thermographic imaging treatments». In: *Multimedial Tools Application* 83 (2024), pp. 42955–71 (cit. on pp. 20, 34).

[81] Alzubaidi L. et al. «Review of deep learning: concepts, CNN architectures, challenges, applications, future directions.» In: *Journal of Big Data* 8 (Mar. 2021), pp. 1–74 (cit. on p. 20).

[82] Silva L., Saade D., Sequeiros O., and Bravo R. «A New Database for Breast Research with Infrared Image». In: *Journal of Medical Imaging and Health Informatics* 4 (Oct. 2014), pp. 92–100. DOI: 10.1166/JMIHI.2014.1226. (cit. on pp. 20, 27).

[83] A. Disante and C. Disante. *Handbook of Image Processing and Computer Vision*. 4th ed. Vol. 3. Sonka, Vaclav Hlavac and Roger Boyle. Image Processing, Analysis, and Machine Vision. Milan: Springer, 2020 (cit. on p. 21).

[84] L.Jiao, F.Zhang, and other. «A survey of deep learning-based object detection». In: *IEEE Access* 7 (Sept. 2019), pp. 837–868. DOI: https://doi.org/10.1109/ACCESS.2019.2939201 (cit. on pp. 21, 22).

[85] Richard Szeliski. *Computer Vision: Algorithms and Applications*. 2nd ed. https://szeliski.org/Book. Springer, 2021 (cit. on p. 22).

[86] R.Girshick, J.Donahue, T.Darrell, and J.Malik. «Rich feature hierarchies for accurate object detection and semantic segmentation». In: *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR)* (Dec. 2014). DOI: https://arxiv.org/abs/1311.2524 (cit. on p. 23).

[87] Ross Girschick. «Fast R-CNN». In: *IEEE International Conference on Computer Vision (ICCV)* (2015), pp. 1440–48. DOI: https://arxiv.org/abs/1504.08083 (cit. on p. 23).

[88] S.Ren, K.He, R.Girshick, and J.Sun. «Fast R-CNN: Towards Real-Time Object Detection with Region Proposal Networks». In: *Advances in Neural Information Processing Systems (NIPS)* 28 (Jan. 2015). DOI: https://doi.org/10.48550/arXiv.1506.01497 (cit. on p. 23).

[89] J.Redmon, S.Divvala, R.Girshick, and A.Farhadi. «You Only Look Once: Unified Real-Time Object Detection». In: *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* 28 (May 2016), pp. 779–788. DOI: https://arxiv.org/abs/1506.02640 (cit. on p. 23).

[90] C.Y.Wang, A.Bochkovskiy, and H.Y.Mark Liao. «YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors». In: (June 2022). DOI: https://arxiv.org/abs/2207.02696 (cit. on p. 23).

[91] W.Liu, D.Anguelov, and other. «SSD: Single Shot MultiBox Detector». In: *Proceedings of the European Conference on Computer Vision (ECCV)* (June 2016), pp. 21–37. DOI: https://arxiv.org/abs/1512.02325 (cit. on p. 23).

[92] T.Y.Lin, P.Goyal, and other. «Focal Loss for Dense Object Detection». In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 42 (Nov. 2020), pp. 318–327. DOI: https://arxiv.org/abs/1708.02002 (cit. on p. 23).

[93] N.Carion, F.Massa, and other. «End-to-End Object Detection with Transformers». In: *Proceedings of the European Conference on Computer Vision (ECCV)* (Aug. 2020), pp. 213–229. DOI: https://doi.org/10.48550/arXiv.2005.12872 (cit. on p. 23).

[94] De Fauw J., Ledsam J.R., et al. «Clinically applicable deep learning for diagnosis and referral in retinal disease». In: *Nature Medicine* 24 (Sept. 2018), pp. 1342–50 (cit. on p. 23).

[95] Ouyang D., He B., Ghorbani A., et al. «Video-based ai for beat-to-beat assessment of cardiac function». In: *Nature* 580 (Feb. 2020), pp. 252–256 (cit. on p. 23).

[96] Wang G., Zuluaga, M.A. Li W., et al. «Deepigeos: a deep interactive geodesic framework for medical image segmentation». In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 41 (July 2018), pp. 1559–72 (cit. on p. 23).

[97] A. Disante and C. Disante. *Handbook of Image Processing and Computer Vision*. 4th ed. Vol. 1. Sonka, Vaclav Hlavac and Roger Boyle. Image Processing, Analysis, and Machine Vision. Milan: Springer, 2020 (cit. on p. 24).

[98] Sonawane M. and Dhawale C. «A brief survey on image segmentation methods.» In: *IJCA Proceedings on National Conference on Digital Image and Signal Processing* 1 (Apr. 2015), pp. 1–5. DOI: https://www.ijcaonline.org/proceedings/disp2015/number1/20475-3002 (cit. on p. 24).

[99] Kirillov A., K. He, R. Girshick, C. Rother, and P. Dollár. «Panoptic segmentation.» In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (Sept. 2019b), pp. 9404–13 (cit. on p. 24).

[100] Pradeep N., Girisha H., Sreepathi B., and Karibasappa K. «Feature extraction of mammograms». In: *International Journal of Bioinformatics Res* 4 (Jan. 2014), pp. 241–244 (cit. on p. 26).

[101] Lim Kelvin O. and Pfefferbaum Adolf. «Segmentation of MR Brain Images into Cerebrospinal Fluid Spaces, White and Gray Matter». In: *Journal of Computer Assisted Tomography* 13 (July 1989), pp. 588–593 (cit. on p. 25).

[102] Dihmani Hanane, Bouattane Omar, and Grief Ouafae Serrar. «A review on Suspicious-Regions Segmentation Methods in Breast Thermogram Image». In: *2nd International Conference on Innovative Research in Applied Science, Engineering and Technology (IRASET)* (Mar. 2022), pp. 1–5. DOI: 10.1109/IRASET52964.2022.9738265 (cit. on p. 25).

[103] Zhou Q., Li Z., and Aggarwal J. K. «Boundary extraction in thermal images by edge map.» In: *In Proceedings of the 2004 ACM symposium on Applied computing* (Mar. 2004), pp. 254–258 (cit. on p. 25).

[104] Ng E. Y. K. and Chen Y. «Segmentation of breast thermogram: improved boundary detection with modified snake algorithm». In: *Journal of mechanics in medicine and biology* 6 (Feb. 2006), pp. 123–136 (cit. on p. 25).

[105] Golestani N., Etehad T.M., and Ng E. Y. K. «Level set method for segmentation of infrared breast thermograms». In: *EXCLI journal* 13 (Mar. 2014), p. 241 (cit. on p. 25).

[106] *FLIRT500-Series Professional Thermal Imaging Cameras*. Drunen Nederland: SENSOR BV (cit. on p. 27).

[107] Kirillov et al. «Segment Anything». In: *IEEE International Conference on Computer Vision* (2023), pp. 4015–4026. DOI: arXiv:2304.02643 (cit. on p. 28).

[108] *Statistics for topic yolo*. 10 January 2026: RepositoryStats (cit. on p. 30).

[109] Zhu X., Lyu S., Wang X., and Zhao Q. «TPH-YOLOv5: Improved YOLOv5 Based on Transformer Prediction Head for Object Detection on Drone-captured Scenarios.» In: *In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops* (Oct. 2021), pp. 12778–2788 (cit. on p. 30).

[110] Jocher G., Chaurasia A., and Qiu J. *Version 8.0.0*. Ultralytics YOLO. Available online:https://github.com/ultralytics/ultralytics (accessed on 12 October 2025)., 2023 (cit. on p. 31).

[111] Khanam R. and Hussain M. «Yolov11: An Overview of the Key Architectural Enhancements». In: *arXiv* (Oct. 2024), pp. 12778–2788. DOI: arXiv:2410.17725v1 (cit. on p. 32).

[112] K.K.Çevik, S.Çivilibal, A.Bozkurt, and E.Dandil. «Enhanced Lesion Classification Based on YOLO Architectures Using Thermal Breast Images on a Patient by Patient Basis». In: *Traitement du Signal* 41 (Dec. 2024), pp. 2989–99. DOI: https://doi.org/10.18280/ts.410617 (cit. on p. 34).

[113] Tayel M.B. and Elbagoury A.M. «Automatic Breast Thermography Segmentation Based on Fully Convolutional Neural Networks». In: *International Journal of Research and Review* 7 (Oct. 2020), pp. 4–10. DOI: 10.2174/157340561566619050314203l (cit. on pp. 34, 57).

[114] R.S.Rosli, M.H.Habaebi, et al. «Analysis of breast region segmentation in thermal images using U-Net deep neural network variants». In: *Frontiers in Bioinformatics* 5 (Oct. 2025). DOI: https://doi.org/10.3389/fbinf.2025.1609004 (cit. on pp. 34, 57).

[115] Shuyue Guan, Nada Kamona, and Murray Loew. «Segmentation of Thermal Breast Images Using Convolutional and Deconvolutional Neural Networks». In: *Conference: 2018 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)* 85 (Oct. 2018), pp. 1–7. DOI: 10.1109/AIPR.2018.8707379 (cit. on pp. 35, 57).

[116] A. Lou, S.Guan, N.Kamona, and M.Loew. «Segmentation of infrared breast images using MultiResUnet neural networks». In: *Conference: 2019 IEEE Applied Imagery Pattern Recognition Workshop (AIPR)* 85 (Oct. 2019), pp. 1–6. DOI: 10.1109/AIPR47015.2019.9316541 (cit. on pp. 35, 57).

[117] Y.Mirasbekov, N.Aidossov, V.Zarikas, et al. «An Integrated Intelligent System for Breast Cancer Detection at Early Stages Using IR Images and Machine Learning Methods with Explainability». In: *SN Comput Science* 4 (Jan. 2023), p. 184. DOI: 10.1007/s42979-022-01536-9.9316541 (cit. on p. 35).

[118] P. Gomathi, C. Muniraj, and P.S. Periasamy. «Digital infrared thermal imaging system based breast cancer diagnosis using 4D U-Net segmentation». In: *Biomedical Signal Processing and Control* 85 (Aug. 2023). DOI: https://doi.org/10.1016/j.bspc.2023.104792 (cit. on pp. 35, 57).

[119] Jun Ma, Yuting He, et al. «Segment Anything in Medical Images». In: *Medical Image Analysis* (Apr. 2023). DOI: arXiv:2304.12306v3 (cit. on pp. 36, 58).

[120] Junde Wu, Wei Ji, et al. «Medical SAM Adapter: Adapting Segment Anything Model for Medical Image Segmentation». In: *Medical Image Analysis* (Apr. 2023). DOI: arXiv:2304.12620 (cit. on pp. 36, 58).

[121] Y.Huang, X.Yang., et al. «Segment Anything Model for Medical Images?» In: *Medical Image Analysis* 4 (Jan. 2024), pp. 92–100. DOI: arXiv:2304.14660v7 (cit. on pp. 36, 58).

[122] Yan Y., Conze P.H., Quellec G., et al. «Two-stage multi-scale breast mass segmentation for full mammogram analysis without user intervention». In: *Biocybernetics and Biomedical Engineering* 41 (Apr. 2021), pp. 746–757. DOI: https://doi.org/10.1016/j.bbe.2021.03.005 (cit. on p. 36).

[123] Pandey S., Chen K. F., and Dam E.B. «Comprehensive multimodal segmentation in medical imaging: Combining yolov8 with sam and hq-sam models». In: *In Proceedings of the IEEE/CVF international conference on computer vision* (Oct. 2023), pp. 2592–98. DOI: 10.1109/ICCVW60793.2023.00273 (cit. on pp. 36, 58).

[124] A. Celaya, B. Riviere, and D. Fuentes. «A Generalized Surface Loss for Reducing the Hausdorff Distance in Medical Imaging Segmentation». In: 5 (Jan. 2024). DOI: https://arxiv.org/html/2302.03868v3#bib.bib15 (cit. on p. 38).

[125] W.R. Crum, O.Camara, and D.L.G. Hill. «Generalized overlap measures for evaluation and validation in medical image analysis». In: *IEEE transactions on medical imaging* 25 (Nov. 2006), pp. 1451–1461. DOI: https://arxiv.org/html/2302.03868v3#bib.bib15 (cit. on p. 38).