



**Politecnico
di Torino**

Politecnico di Torino

Corso di Laurea

A.A. 2021/2022

Sessione di Laurea Marzo/Aprile 2022

**Metodo di segmentazione automatica
del canale naso-palatino tramite
l'utilizzo di reti basate
sull'apprendimento profondo.**

Relatore:

Prof. Molinari Filippo

Co-relatore:

Prof. Salvi Massimo

Candidati:

Vitale Giuseppe, mat. 265794

Indice

Abstract	1
1. Introduzione	2
1.1. Il dotto incisivo	3
1.2. Il canale naso palatino dal punto di vista odontoiatrico	5
1.3. Dataset e segmentazioni manuali	6
1.4. Reti neurali	12
1.4.1. Le CNN e il deep learning.....	14
1.4.2. Le Faster RCNN	14
1.4.3. Le U-Net.....	15
1.5. Operatori morfologici	17
1.6. Modelli deformabili	18
1.7. Stato dell'arte	20
1.8. Strumenti utilizzati	23
2. Metodo proposto	24
2.1. Flow chart e descrizione	25
2.2. Il dataset e l'acquisizione dati	27
2.3. Le reti Faster RCNN.....	28
2.3.1. Preprocessing	29
2.3.2. Allenamento delle Faster RCNN	31
2.3.3. Classificazione delle soluzioni.....	32
2.4. Ripescaggio Faster RCNN.....	34
2.4.1. Preprocessing	34
2.4.2. Allenamento delle Faster RCNN	35
2.4.3. Classificazione delle soluzioni.....	35
2.5. U-Net	35
2.5.1. Preprocessing	36
2.5.2. Allenamento delle U-Net.....	37
2.5.3. Classificazione delle soluzioni.....	38
2.6. PostProcessing.....	39
2.6.1. Eliminazione degli artefatti dovuti al rumore.....	39
2.6.2. Operatori morfologici	41

2.6.3.	Level set	42
2.6.4.	Miglioramento delle performance	43
3.	Risultati	45
3.1.	Metriche utilizzate	46
3.1.1.	Object level	46
3.1.2.	Voxel level.....	47
3.1.3.	Morfometriche	48
3.2.	Risultati delle Faster RCNN	49
3.3.	Risultati delle UNet.....	52
3.4.	Risultati dovuti al post processing.....	56
3.5.	Le caratteristiche morfometriche.....	59
4.	Conclusioni	61
	Bibliografia.....	64

Abstract

Questa è una tesi il cui fine è proporre un metodo di segmentazione automatica del canale naso-palatino conservando le caratteristiche morfologiche così che si possa utilizzare in implantologia odontoiatrica e incrementarne l'utilizzo per migliorare l'efficienza delle protesi.

In odontoiatria le caratteristiche del canale naso-palatino sono molto importanti per quel che riguarda l'impiantistica dentale. Se dovessero mancare i riferimenti che indicano come posizionare la protesi dentale, viene utilizzata l'inclinazione del canale per rispettare la fisiologia originaria del dente ed indica la corretta applicazione dell'impianto. Anche se molto utilizzata come tecnica, non è mai stata dimostrata una simmetria o almeno dipendenza tra l'asse del canale rispetto a quello dei denti.

Il metodo propone una segmentazione automatica del canale naso-palatino a partire da immagini di tomografia computerizzata a bassa potenza per odontoiatria, in modo da conservare le caratteristiche morfologiche principali (lunghezza, inclinazione e diametro) e la forma tridimensionale. In questo modo è possibile analizzare gli aspetti morfometrici e saperne di più sulla dipendenza tra il canale e i denti.

Il metodo richiede l'utilizzo di tecniche di segmentazione tridimensionale attraverso reti neurali basate sull'apprendimento profondo. Prima riduce il volume ad una regione d'interesse di volume fisso utilizzando delle reti neurali convoluzionali regionali che agiscono sulla proiezione bidimensionale delle immagini TC e successivamente viene effettuata la segmentazione semantica sul volume ridotto attraverso l'utilizzo di una UNet. Per conservare le caratteristiche principali del canale, alla segmentazione semantica viene affiancata una segmentazione che utilizza il level-set.

Quindi il risultato è una segmentazione del canale, che ne conserva le caratteristiche morfologiche.

I risultati sono successivamente stati confrontati con delle segmentazioni manuali in modo da ricavare l'efficienza e le criticità dell'algoritmo. L'analisi è stata fatta tenendo conto di tre metriche di validazione: una fa riferimento all'oggetto, una fa riferimento alla fedeltà del volume segmentato e una fa riferimento conservazione delle caratteristiche fondamentali necessarie successivamente per l'utilizzo in campo odontoiatrico, come l'inclinazione del canale rispetto alla segmentazione manuale.

1. Introduzione

In questo trattato si propone un metodo di segmentazione automatica del dotto naso-palatino che opera velocemente e nella maniera più fedele possibile rispetto segmentazioni fatte manualmente, provando a conservarne le caratteristiche morfologiche principali.

Il canale naso-palatino è un canale osseo che collega il setto nasale al palato, all'interno del quale passano vasi sanguigni e nervi. Dal punto di vista odontoiatrico viene utilizzata l'inclinazione del canale come punto di riferimento nell'implantologia di protesi dentali in mancanza di altri riferimenti anatomici. Tra i vari usi è posto in evidenza l'utilizzo del dotto per ricavare l'inclinazione degli incisivi superiori qualora si dovessero impiantare protesi dentali in mancanza di riferimenti dovuto al riassorbimento del canale dentale.

Nonostante la dipendenza tra l'inclinazione dell'asse degli incisivi superiori e l'inclinazione del dotto incisivo sia spesso utilizzata in implantologia dentale, non ci sono documentazioni scientifiche di un effettivo parallelismo tra i due.

Sarebbe importante dal punto di vista medico dimostrare un parallelismo o una dipendenza lineare che c'è tra l'asse del canale e l'asse degli incisivi superiori.

Lo scopo di questa tesi è proporre un metodo di segmentazione automatica del canale naso-palatino e una fedele ricostruzione tridimensionale permettendone un miglior riconoscimento delle caratteristiche morfometriche e anatomiche. L'utilizzo di questo metodo sarebbe di fondamentale interesse per proporre nuovi metodi di implantologia dentale o per poter studiare nuovi protocolli, che innanzitutto semplificherebbero il lavoro dell'odontoiatra e in secondo luogo aumenterebbero la compatibilità delle protesi dentali rispettando le peculiari caratteristiche anatomiche del soggetto, anche in caso di riassorbimento.

Inizialmente sono trattati gli argomenti serviti per lo svolgimento di questa tesi, a partire dalle caratteristiche del canale naso palatino dal punto di vista anatomico e medico, fino agli strumenti più comunemente utilizzati per l'elaborazione di immagini mediche odontoiatriche e concetti base sull'acquisizione di immagini radiografiche.

Sono trattati gli argomenti da letteratura utilizzati come base da cui si è sviluppato il metodo proposto, analizzando i metodi più innovativi di elaborazione automatica di immagini in campo odontoiatrico,

che tutt'ora sono allo stato dell'arte come auto estrazione di strutture anatomiche dentali. Saranno spiegate la struttura e il funzionamento delle reti neurali usate.

È mostrata anche una descrizione del dataset di partenza su cui si è sviluppato l'algoritmo, partendo dalle caratteristiche principali al lavoro che è stato necessario fare sullo stesso per renderlo nel miglior modo utilizzabile per l'allenamento delle reti e per la validazione.

In seguito è spiegato il metodo proposto in tutti i suoi passaggi, dall'acquisizione dei dati alla maschera di segmentazione finale. È mostrato anche tutto ciò che è stato utilizzato per l'allenamento delle reti e le strategie per aumentare l'efficienza.

I risultati sono mostrati elencando le metriche utilizzate per caratterizzare l'efficienza dell'algoritmo con una breve trattazione dei valori ottenuti. In più sono descritte tutte le scelte fatte per ottenere i risultati e le tecniche scartate in quanto portavano a risultati peggiori.

Nel nell'ultimo capitolo, sono presenti le conclusioni con relative criticità dell'algoritmo e peculiarità. Sono anche presente le idee su sviluppi futuri e dati parziali di un proseguo del lavoro finora svolto, ovvero calcolare il grado di parallelismo tra l'asse del dotto incisivo e l'asse degli incisivi superiori.

1.1. Il dotto incisivo

Il canale incisivo (indicato anche come canale nasopalatino o canale palatino anteriore) è un condotto osseo dell'osso mascellare che collega le cavità nasali e orali[1]. Contiene strutture neurovascolari quali il nervo nasopalatino, che fornisce input sensoriali al ganglio pterigopalatino, la mucosa del palato duro, la gengiva intorno agli incisivi fino ai canini e infine un'anastomosi vascolare tra il ramo posteriore dell'arteria sfenopalatina e l'arteria palatina maggiore. Il canale incisivo termina alla fossa incisiva nella cavità orale, posteriormente agli incisivi mascellari e sotto la papilla incisiva. Superiormente, si estende alla cavità nasale su ciascun lato del setto nasale come forame nasale o forame di Stenson, circa 2 cm posteriormente al margine interno di ciascuna narice [2].

Si è studiato che il dotto nasopalatino ha delle terminazioni superiori bilaterali ed un punto in cui sfocia nella fossa incisiva. Questo ha una forma a “Y” o “V” nella maggior parte della popolazione (60%) [3], ma si sono riscontrate anche altre forme.

Song et al. ha documentato che anche le strutture vascolari che passano all'interno del canale dipendono dal tipo di forma assunta dal canale (in precisione se la terminazione superiore ha 1 o più aperture).

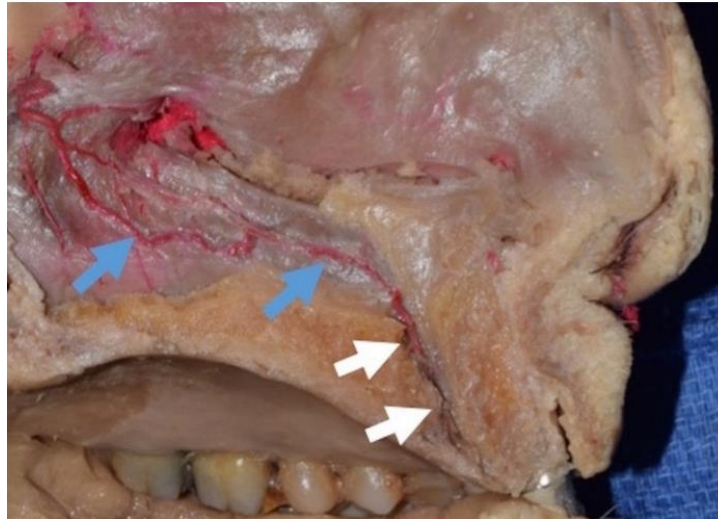


Figura 1 Immagine rappresentante la sezione del canale naso palatino dove è possibile vedere le strutture neurovascolari che passano al suo interno.

La radiografia craniofacciale è essenziale per la valutazione della patologia macroscopica o per conoscere l'anatomia indispensabile per la pianificazione di interventi chirurgici craniofacciali o interventi odontoiatrici restaurativi. Le modalità di imaging della regione maxillo-facciale includono raggi X con viste periapicali o panoramiche, tomografia computerizzata (TC) e risonanza magnetica (MRI). Di queste tecniche appena elencate, la tomografia computerizzata a bassa energia (detta Cone-beam computed tomography CBCT) è il modo migliore per valutare il canale naso-palatino dando migliori immagini ad alta risoluzione, eliminando la sovrapposizione delle immagini, offrendo una minore esposizione alle radiazioni e fornendo una migliore analisi della qualità dell'osso[4].

La morfometria del canale considera la lunghezza, i diametri alle aperture nasali e orali, cioè rispettivamente forami nasali e fossa incisiva e lo spessore della placca ossea buccale sopra il nasale e aperture orali. La lunghezza del canale incisivo è misurata come la massima distanza verticale tra le aperture nasali e orali ed è in media di 10 mm [5]. È stato dimostrato che la lunghezza del canale varia anche in base all'altezza dell'osso mascellare. Il diametro del foro incisivo è in media inferiore a 6 mm nella mascella dentula.

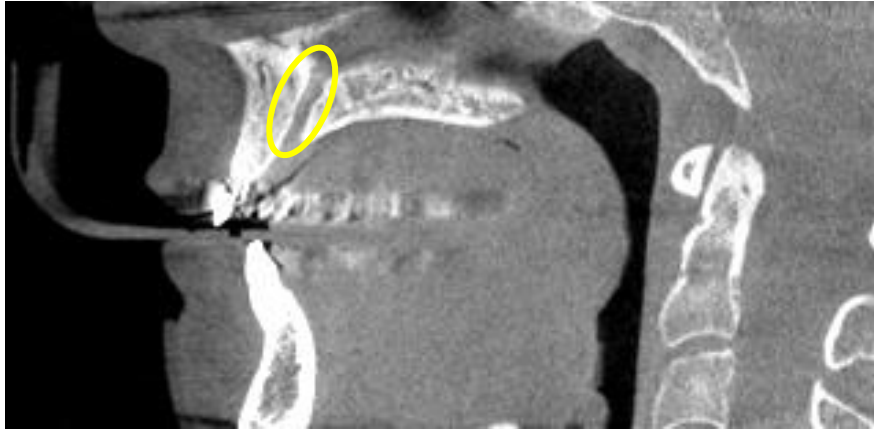


Figura 2 pone una slice di un'immagine CBCT sul piano trasversale dove si può notare il canale naso-palatino evidenziato in giallo.

1.2. Il canale naso palatino dal punto di vista odontoiatrico

In odontoiatria il posizionamento di un impianto dentale nella mascella anteriore è molto impegnativo. L'anatomia, l'estetica e il supporto delle labbra sono direttamente influenzati dal riassorbimento della cresta dopo l'estrazione dei denti anteriori [6] in più è stato dimostrato che l'osso alveolare mascellare anteriore si riassorbe in media del 34% dopo l'applicazione dell'impianto [7].

Il canale incisivo diventa di fondamentale importanza in quanto fornisce informazioni sull'angolazione del dente estratto dopo un parziale riassorbimento dell'osso alveolare.

La presenza del canale incisivo può essere una limitazione per il posizionamento ideale dell'impianto, ma questa può essere superata curando e innestando il canale incisivo [8] e successivamente è possibile inserire un impianto nell'area innestata. Un'altra soluzione è quella di inserire l'impianto proprio nel canale incisivo, infatti si è dimostrato che in una popolazione di 30 TAC l'applicazione dell'impianto è fallimentare solo nel 4% dei casi.

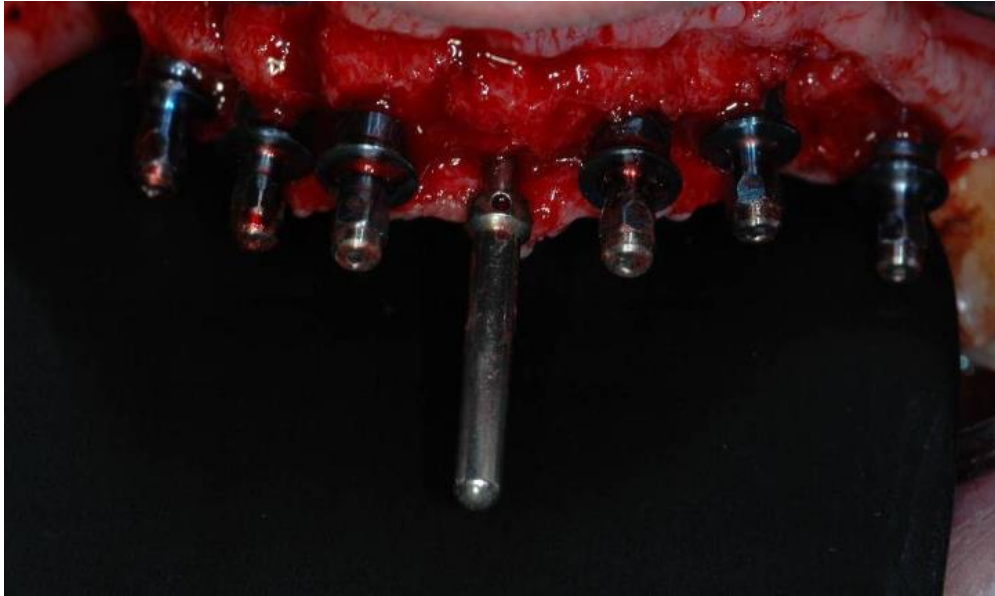


Figura 3 Esempio di utilizzo del dotto incisivo in un intervento di impianto dentale, il canale viene utilizzato per avere un riferimento anatomico.

1.3. Dataset e segmentazioni manuali

Per questo lavoro è stato utilizzato un dataset di 808 immagini da tomografia computerizzata per odontoiatria a bassa energia (in inglese Cone-beam computed tomography, CBCT).

La CBCT è una variante della tomografia computerizzata (TC) tradizionale usate dagli odontoiatri dove la macchina gira intorno al paziente per catturare le forme dei denti tramite raggi X a bassa energia. Come tutte le immagini di tipo radiografico sfruttano delle onde ionizzanti create da un tubo radiogeno al fine di creare una rappresentazione anatomica delle strutture del corpo umano. I raggi X attraversando il corpo, vengono parzialmente assorbiti dai tessuti ed è proprio grazie all'assorbimento delle onde che si possono estrarre informazioni sul tipo di tessuto che ha attraversato.

Dunque il fine è sapere qual sia il coefficiente di assorbimento del tessuto attraversato dalle onde ionizzanti, ciò rende possibile il riconoscimento dello stesso in quanto ogni organo e tessuto ha un assorbimento delle onde peculiare e ben distinguibile.

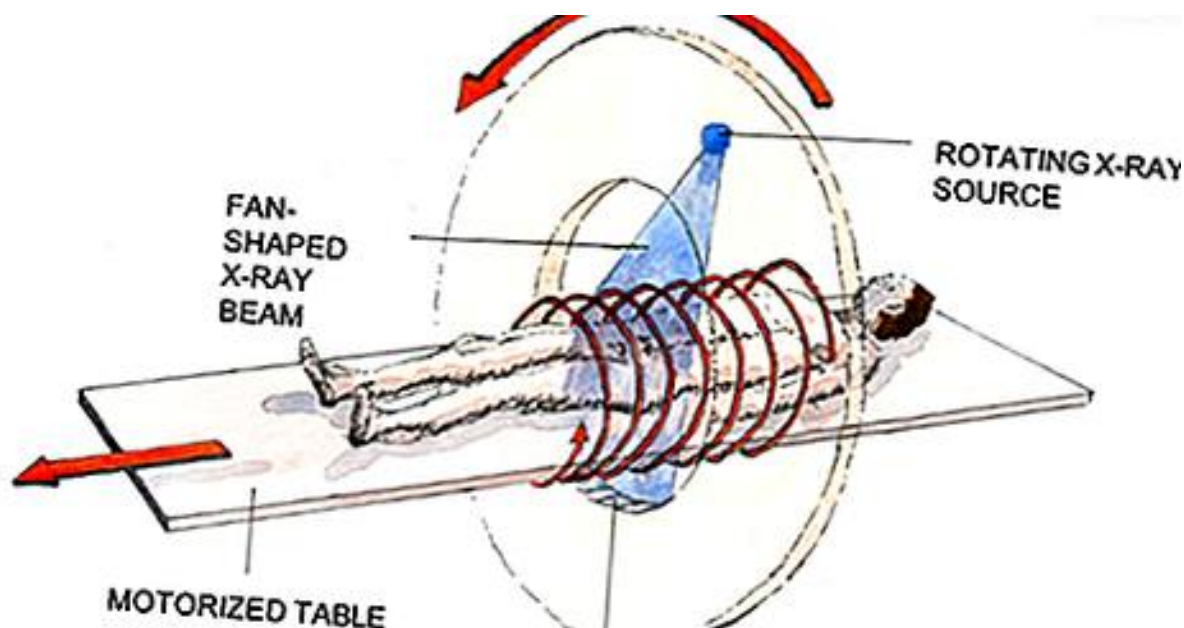


Figura 4 Rappresentazione del funzionamento di una scansione effettuata da una macchina per TC.

Nella TC la scansione non è effettuata lungo un solo piano, ma si ha un insieme di radiografie su diversi piani. Le scansioni vengono eseguite girando intorno ad un asse a varie angolazioni e permettono il superamento delle limitazioni delle immagini radiografiche tradizionali. Permettono la ricostruzione dei tessuti in tre dimensioni. Grazie all'esecuzione dell'algoritmo di retroproiezione filtrata, eseguito offline tramite un computer, è possibile elaborare tutte le acquisizioni effettuate alle varie angolazioni ricostruendo un'immagine tridimensionale con un'elevata risoluzione spaziale.

Come precedentemente detto, le immagini di tomografia computerizzata in campo odontoiatrico prendono il nome di CBCT, esse sfruttano delle onde coniche ionizzanti a minore energia rispetto alle onde ionizzanti usate in radiologia tradizionale (riducendo la dose assorbita dal paziente) per descrivere tramite il coefficiente di assorbimento la regione dentale, orale e maxillofacciale. Questa tecnica rende possibile la rappresentazione tridimensionale dei denti, e delle altre strutture ossee in quanto hanno un coefficiente di assorbimento alto e ben distinguibile.

Queste danno un'elevata risoluzione e rappresenta fedelmente le caratteristiche anatomiche della zona studiata, rendendo possibile una differenziazione dei tessuti nella zona maxillofacciale, dentale e mandibolare.

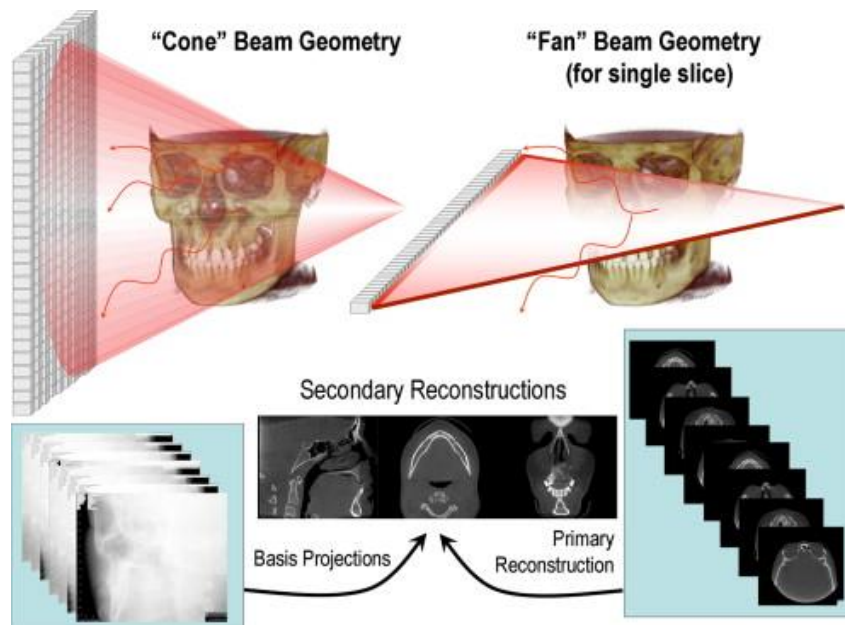


Figura 5 Schema di funzionamento di una CBCT.

Il dataset di immagini necessarie per lo svolgimento di questa tesi sono stati forniti dallo Studio Dentistico Romano Dr. Biagio e le immagini sono state acquisite da una Cone Beam CT Vatech PaX-I 3d Smart e Vatech PaX-Flex 3d. Ringrazio lo studio per la disponibilità ed in particolar modo il dott. Romano Davide per tutta la disponibilità data per lo svolgimento di questa tesi.

Le immagini sono state esportate in formato DICOM. Il formato DICOM è lo standard con cui vengono archiviate le immagini RX, TC o RM in campo biomedico. A differenza dei formati JPEG e PNG, non viene applicata una compressione dell'immagine ma vengono direttamente archiviate. Questo è solo un metodo per incapsulare l'immagine e definire il metodo con cui debbano essere codificate o interpretate. Il formato prevede il salvataggio, oltre all'immagine, di numerose informazioni come: nome e cognome del paziente, tipo di scansione, posizione, dimensione dell'immagine e tante altre. Vengono dunque memorizzate nell'intestazione e catalogate nel gruppo elementi, detti anche TAG DICOM.

Dall'analisi dei dati presenti nelle TAG DICOM è mostrata una varietà di caratteristiche all'interno del dataset che lo rendono molto disomogeneo e vario. All'interno del dataset le immagini differiscono per dimensione, slice-rate su tutti e 3 gli assi, energia applicata dal tubo radiogeno, posizione e quantità di informazioni contenute nell'immagine stessa (alcune immagini rappresentano l'intero cranio, mentre la maggior parte rappresentano totalmente o parzialmente la zona maxillo-palatina). Inoltre la grande variabilità è stata incrementata data dalla presenza di oggetti con un alto coefficiente di assorbimento (protesi dentali, viti, apparecchi, ecc...) che creano rumore sull'immagine e causano artefatti radiografici.

La variabilità del dataset non è data solo dalle caratteristiche dell'immagine, infatti si può facilmente notare un grade variabilità delle caratteristiche del canale che dipendono dal soggetto. Infatti si è dimostrato che le caratteristiche del canale cambiano in base all'età, etnia e sesso[3].

Nel dataset sono presenti immagini di pazienti in età avanzata che presentano un coefficiente di assorbimento minore rispetto ai pazienti più giovani o casi più rari in cui pazienti aventi giovane età devono ancora sviluppare i denti permanenti, essi infatti presentano 2 arcate dentali, una esterna (denti da latte) e la seconda (corrispondente ai denti permanenti) ancora sotto la gengiva. Sono presenti anche pazienti che non presentano nessun dente della propria arcata dentale, questi aumentano la variabilità del dataset in quanto i denti portano a un artefatto radiografico nella zona ossea intorno alla radice dei denti ed incrementano l'eterogeneità del dataset.

Mancanza di denti, presenza di protesi o otturazioni dentali causano degli artefatti radiografici.

Per quel che riguarda la prima causa di eterogeneità delle immagini, ovvero dovuta all'acquisizione delle immagini vi si può porre rimedio andando, nel preprocessing, ad uniformare queste caratteristiche ricampionando le immagini.

Le caratteristiche del dataset che lo rendono disomogeneo sono un fattore favorevole per l'allenamento delle reti neurali in deep learning, in quanto le rende stabili e coerenti.

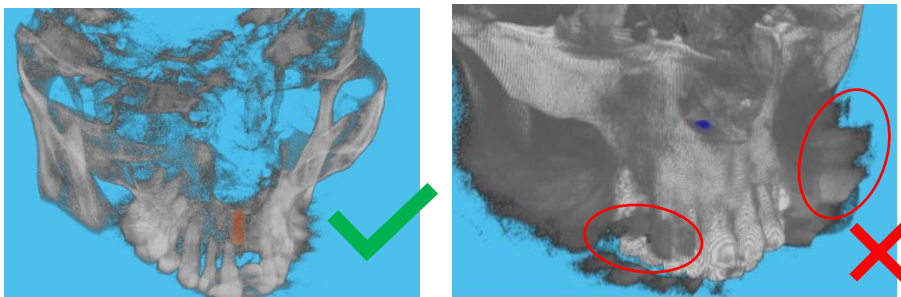


Figura 6 le due immagini sono delle immagini prese dal dataset. La prima a sinistra è un'immagine che soddisfa i criteri, il canale è presente e visibile. Nell'immagine centrale invece sono presenti degli artefatti radiografici (cerchiati) dovuti alle protesi dentali che non permette la corretta visione del dotto naso-palatino.

In realtà le immagini fornite dagli studi odontoiatrici sono più di 1000, molte di queste non contenevano il dotto naso palatino, ma soltanto la zona mandibolare oppure avevano scarsissima qualità dovuta a rumore radiografico dovuto a protesi dentali che ne rendevano irriconoscibile il canale anche ad occhio nudo o ancora con una scarsissima definizione dovuta sia all'acquisizione che ad artefatti da movimento. Il primo processo è stato escludere dal dataset le immagini non contenenti

il canale o con una qualità troppo bassa in modo che soddisfino i criteri di presenza e rappresentazione con discreta qualità del dotto, così da allenare le reti nella maniera con dati il più coerenti possibili.

Per sviluppare le reti neurali è stato anche necessario l'applicazione di maschere manuali in modo da creare un delle immagini di veri positivi usate come riferimento, sia come supervisione della rete e sia per avere dei criteri di efficienza della rete nella fase di validazione e prova.

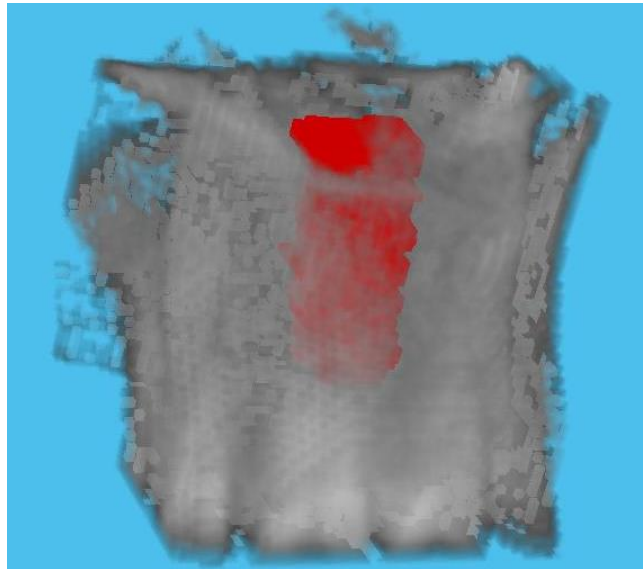


Figura 7 Rappresentazione dell'elemento segmentato manualmente in tre dimensioni.

Quindi per ogni immagine che soddisfa i criteri è stata applicata una segmentazione manuale creata da un operatore. Non essendo esperto l'operatore può aver creato dei problemi in entrambi i processi che sfruttano le maschere, creando un errore sia nell'apprendimento della rete che nella fase di validazione. In più è presente anche un errore intra-operatore dovuto al cambiamento nel tempo dei criteri seguiti per effettuare la segmentazione manuale, in quanto le segmentazioni sono state effettuate in più mesi.

La segmentazione manuale è stata effettuata tramite un algoritmo dotato di un'interfaccia creata appositamente per questo scopo. L'interfaccia mette a video 2 schermate, entrambe visualizzano delle slice contenenti il canale, ma l'applicazione della maschera manuale è stata effettuata sulla prima schermata che pone a video le slice rappresentanti il piano trasversale e quindi inizialmente bidimensionali. Solo infine sono riuniti tutti i piani al fine di creare una maschera tridimensionale combinando le diverse slice. La seconda schermata pone a video una slice sul piano sagittale in modo da notare se la segmentazione presenta una continuità spaziale.

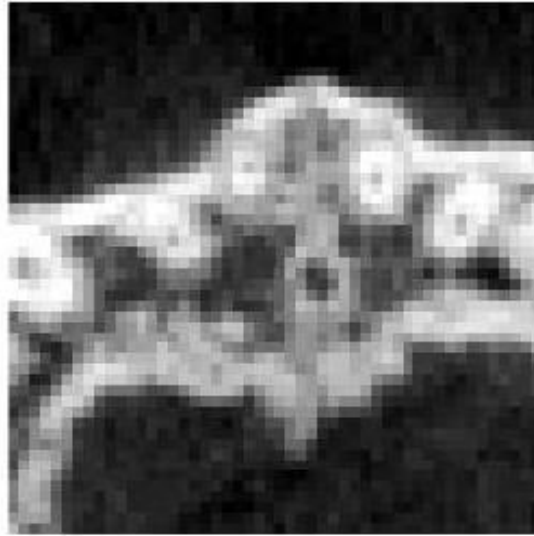


Figura 8 Esempio di slice contenente il canale

Il canale è stato, quasi sempre, facilmente rintracciabile perché presenta dei bordi ben definiti dovuti alla terminazione dell'osso dotata di maggiore densità (che si converte con un'alta luminosità) e l'interno meno denso (meno luminoso) creando un elevato contrasto. La minor densità è dovuta alla presenza di tessuti molli (nervo naso-palatino e vasi sanguigni) all'interno del dotto, ben distinguibili nella maggior parte delle immagini. È possibile notare all'interno del dataset che il canale, per quanto quasi sempre individuabile, ha delle caratteristiche che variano tra i vari pazienti sia per dimensione che per forma e non è raro anche che possa terminare separandosi e dividendosi, in quanto contenente alcune escrescenze ossee.

Sono stati proposti due metodi di segmentazione manuale: uno con una singola classe in cui vengono create le maschere contenenti sia il canale, ed un secondo metodo in cui sono presenti due classi, una contenente solo la parte interna del canale (dove sono posti i tessuti molli) e una seconda classe contenente la parte del bordo del canale dove vi è una zona con maggiore luminosità che rappresenta la terminazione della zona ossea caratterizzata da una maggiore densità.

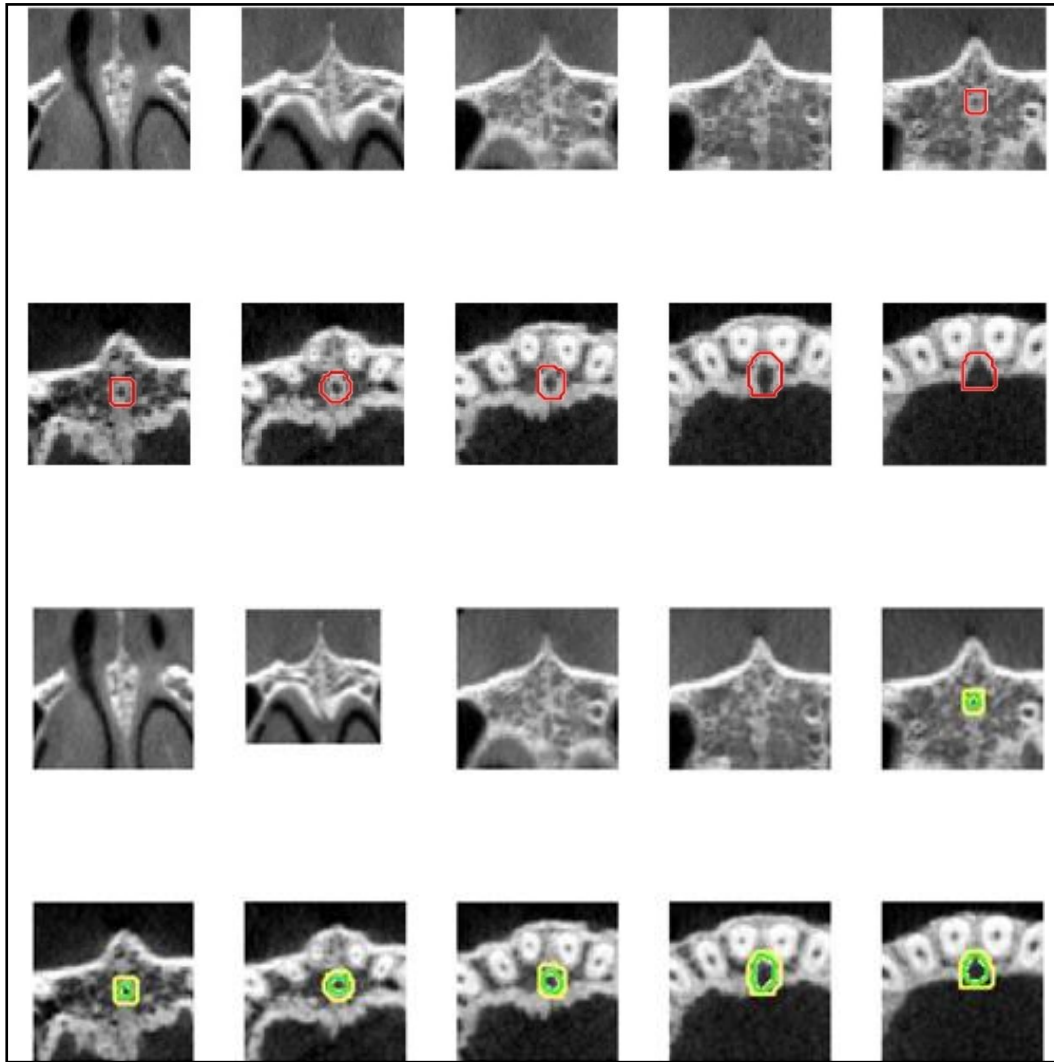


Figura 9 sono degli esempi di maschere manuali, in rosso le maschere manuali con il primo metodo descritto, in verde (canale interno) e giallo(bordo del canale) il secondo metodo.

1.4. Reti neurali

Secondo i principali articoli in capo odontoiatrico, i metodi di segmentazione più utilizzati negli ultimi anni sono delle reti neurali con il metodo di apprendimento profondo, nello specifico in questo metodo, si è scelto di utilizzare le Faster-RCNN per il rilevamento dell'oggetto sulla proiezione bidimensionale delle immagini e U-Net per la segmentazione semantica tridimensionale.

La rete neurale è un modello matematico ed informatico di calcolo strutturato in maniera non lineare. Esse possono essere utilizzate per simulare relazioni complesse tra ingressi ed uscite.

Una rete neurale riceve segnali esterni su uno strato di neuroni (unità di elaborazione) di ingresso, ciascuno dei quali è collegato a numerosi neuroni organizzati in più livelli. Ogni neurone elabora i

segnali ricevuti e trasmette ai nodi successivi. I neuroni sono divisi in più livelli, il primo livello è detto in ingresso (input) che prende i dati e li comunica ai livelli successivi, i livelli nascosti, che rielaborano i dati per poi trasmetterli all'ultimo livello, il livello d'uscita (output) che dà il risultato della segmentazione.

Il risultato della rete neurale dipende dall'elaborazione “n” funzioni che avvengono ai nodi. Nella fase di apprendimento la rete modifica i pesi che dà alle funzioni dei vari nodi in modo da ottimizzare l'efficienza di classificazione stessa della rete.

Esistono due sottoinsiemi di reti neurali che dipendono dal loro apprendimento: le reti supervisionate e non supervisionate.

L'apprendimento non supervisionato è basato su algoritmi di addestramento che modificano i pesi della rete facendo esclusivamente riferimento ad un insieme di dati che include le sole variabili di ingresso. Tali algoritmi tentano di raggruppare i dati di ingresso e di individuare pertanto degli opportuni cluster rappresentativi dei dati stessi.

Le supervisionate, dove si hanno per l'allenamento sia i dati in ingresso che le corrispettive uscite. Tramite un algoritmo (backpropagation) ridispone i pesi sui vari nodi in modo da ottimizzare la rete andando a lavorare sulla differenza tra output della rete e l'uscita che si ha inizialmente (veri positivi) rendendola minore possibile.

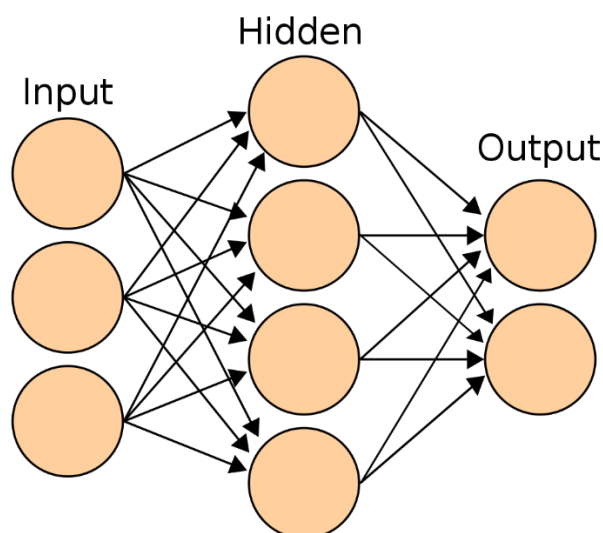


Figura 10 Esempio di rete neurale, i cerchi sono i neuroni. Le frecce sono le connessioni tra i vari livelli.

1.4.1. Le CNN e il deep learning

Tutte le reti utilizzate in questo trattato sono estensioni di modello di rete neurale convoluzionale (CNN). Queste sono delle reti neurali supervisionate capaci, a differenza delle reti per machine learning, di estrapolare le caratteristiche necessarie per la classificazione in automatico nella fase di apprendimento.

Per fare ciò le reti sfruttano le convoluzioni, ovvero dei filtri che passano sull'immagine e ne ricavano le caratteristiche comuni. Questa classe di filtri serve per ricavare un significato dall'immagine. In seguito, è applicata una seconda classe di filtri: i pooling, che riducono le dimensioni dell'immagine, infatti raggruppano una popolazione di pixel per poi rifiltrarli. Nella fase di allenamento vengono applicati entrambi i filtri in maniera casuale ed iterativa finché non individua il valore dei filtri che permette un matching tra i valori di input e di output che causa meno errori. [27]

Quando vengono applicati i filtri le dimensioni dell'immagine verranno ridotte a discapito delle caratteristiche. Verranno ridotte fino all'estrapolazione di tutte le caratteristiche determinati necessarie per una successiva corretta classificazione.

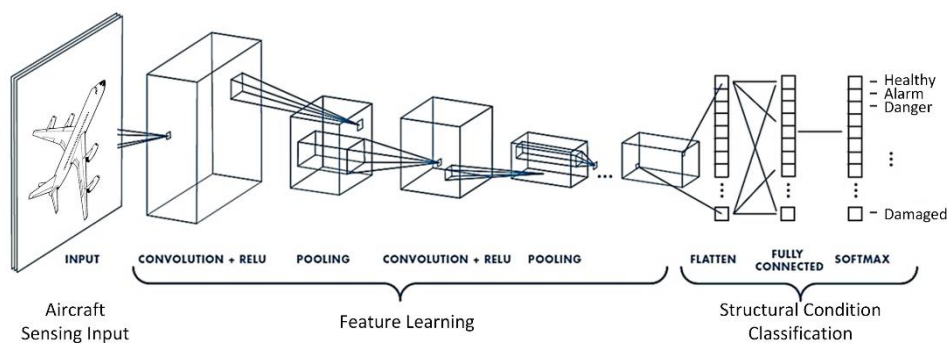


Figura 11 Esempio di vari livelli di una rete CNN.

1.4.2. Le Faster RCNN

La Faster-RCNN è una rete neurale il cui scopo è quello di dividere l'immagine acquisita e porre in output i riquadri che delimitano un oggetto e successivamente classificarlo in una categoria che dipende dall'oggetto che contiene al suo interno. Questa rete, in particolare, genera all'interno della rete neurale le ROI e le classifica come output.[11]

La Faster-RCNN è una rete per la ricerca di una regione d'interesse (regional purpose network RPN), a partire da un'immagine di qualsiasi dimensione, genera una serie di rettangoli associando un valore ad ognuno che indica la probabilità che il riquadro comprenda l'oggetto. Quindi ci saranno due livelli

fratelli che comunicano il risultato dei riquadri. Uno strato di regressione e uno strato di classificazione del riquadro in modo da modificare i contorni del riquadro in base ai risultati di classificazione dell'oggetto.

Il risultato della rete sarà una ROI che contiene, con la migliore probabilità, l'oggetto cercato dalla rete.

Le Faster RCNN sono le reti più performanti e veloci attualmente utilizzate per la ricerca di oggetti anche in campo medicale. In odontoiatria questo tipo di reti sono utilizzate per la ricerca e successiva classificazione dei denti.

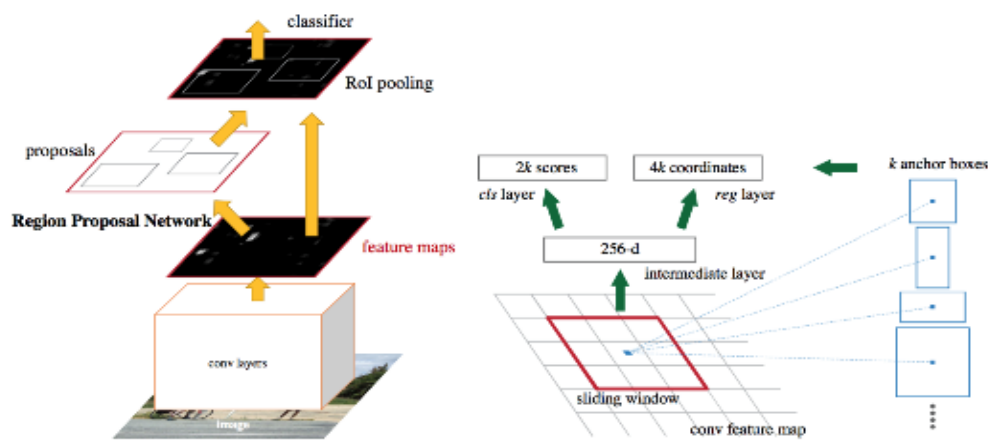


Figura 12 è una classica struttura di Faster RCNN

1.4.3. Le U-Net

L'U-Net è una rete neurale convoluzionale sviluppata per le immagini biomediche. Sostanzialmente è una rete neurale convoluzionale, ma è formata da più livelli successivi di apprendimento dove si applicano delle operazioni di pooling e sovracampionamento. Questi livelli sono utili per aumentare la risoluzione dell'output, portando l'allenamento del livello successivo ad essere basato sull'output del livello precedente. [20]

La U-Net si basa sul riferimento di una FCN (fully convolutional network) la cui architettura è stata ottimizzata per funzionare anche con poche immagini usate nella fase di allenamento, ottenendo comunque una precisione elevata delle segmentazioni. [24]

La rete è formata da due rami: uno contrattivo e uno espansivo. Questa struttura è più o meno simmetrica creando una forma di U (da cui la rete prende il nome). Il ramo contrattivo adopera, su

tutti i canali, un insieme di filtri convoluzionali usati per estrazione automatica delle caratteristiche.[20] I vari livelli sono collegati tra loro per permettere una maggiore e migliore propagazione dell'informazione. Nel ramo espansivo, il fine è quello di aumentare la risoluzione, per fare ciò si applicano le operazioni di sovracampionamento, che permettono un aumento delle dimensioni e tramite dei collegamenti diretti con il ramo contrattivo permette la trasmissione della sola la parte valida dell'informazione. Questa tecnica serve per ridurre le dimensioni delle immagini (quindi poter trattare immagini più grandi), in caso contrario le dimensioni sarebbero difficili da gestire e la l'utilizzo della rete sarebbe limitato dalla memoria della GPU.[24]

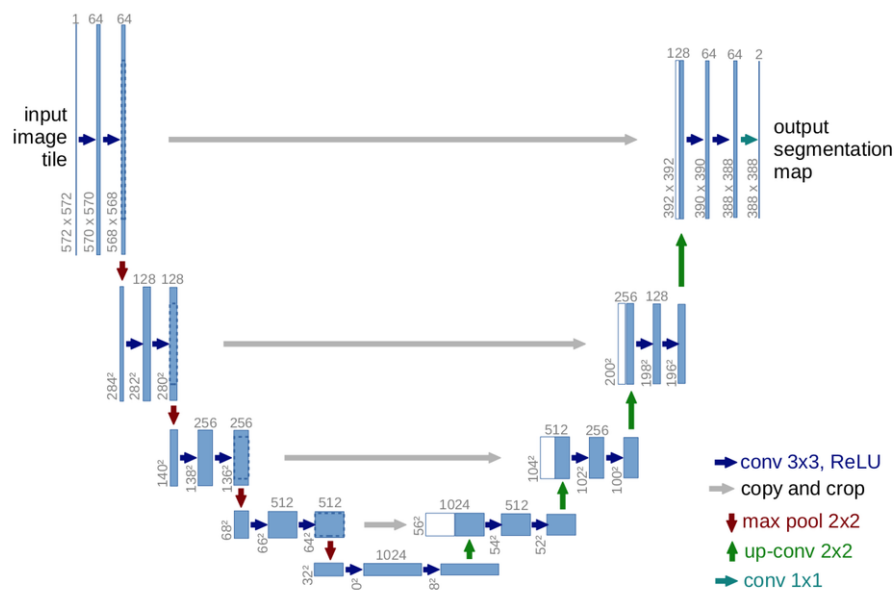


Figura 13 esempio layout di una generica U-Net; blu) convoluzioni classiche e ReLU, bianco) canali per fully coonected, rosso) operazioni di sottocampionamento, verde) operazioni di sovraccampionamento, azzurro) convoluzioni 1x1

Il percorso è quindi formato da più livelli, dove ogni livello è una tipica CNN, ovvero un'applicazione ripetuta di filtri convoluzionali dove ciascuna è seguita da un'unità ReLU e da un'operazione di max pooling. Tramite questo processo vengono ridotte le dimensioni spaziali, ma aumentano le dimensioni delle caratteristiche.

Le reti neurali descritte sono entrambe delle DNN (Deep Neural Network) ovvero delle reti neurali profonde che in automatico estraggono le feature d'interesse.

Per l'utilizzo delle UNet, è stata utilizzata una segmentazione semantica. La segmentazione semantica utilizza la rete per assegnare ad ogni pixel o voxel una classe di appartenenza, non riesce a distinguere all'interno di una segmentazione diverse istanze.

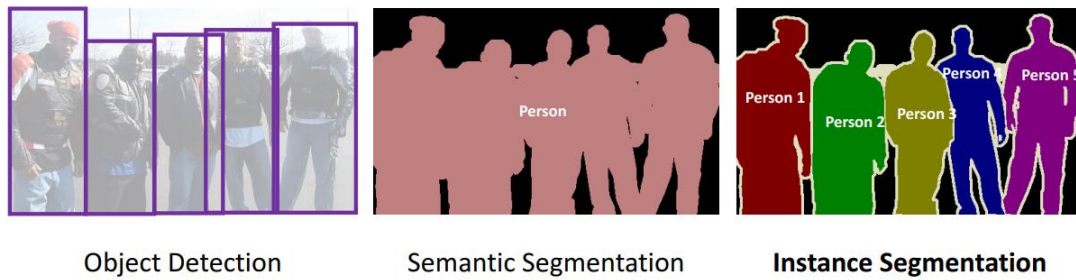


Figura 14 L'immagine descrive la differenza tra segmentazione semantica e segmentazione delle istanze.

1.5. Operatori morfologici

Nella proposta del metodo saranno usati anche degli operatori morfologici. Gli operatori morfologici sono degli operatori il cui fine è quello di estrarre degli elementi rispettando determinate morfologie. Essi possono essere applicati sia a immagini in toni di grigio che immagini in bianco e nero.

Fondamentale è l'elemento strutturale, questo è una struttura tridimensionale di dimensioni e forma ben precise e delimitate. Gli insiemi connessi si formano specificando i punti direttamente connessi (adiacenze) e i punti connessi in obliquo(contorni). Ci sono un sacco di forme che è possibile applicare: cerchio, sfera, diamante, ecc...

Una volta creato l'elemento strutturale questo, come una finestra, passa sull'immagine e ne segue una determinata operazione.

Le principali operazioni sono l'**erosione** dove passando sull'immagine, come un filtro considera dei pixel o voxel selezionati, solamente il più piccolo. Invece, la **dilatazione** opera conservando il pixel o voxel più grande.

Importanti sono anche gli operatori di apertura e chiusura. **Apertura** è una combinazione di erosione seguita da dilatazione il cui fine è quello di preservare maggiormente lo sfondo eliminando gli elementi più piccoli. Mentre la **chiusura** è una dilatazione seguita da erosione il cui fine è quello di chiudere i buchi rimuovendo lo sfondo compreso tra 2 elementi.

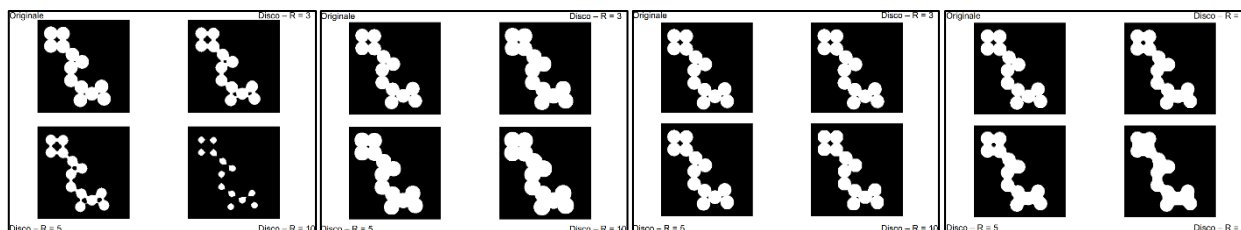


Figura 15 Le immagini sono degli esempi di applicazione di operatori morfologici: a)erosione, b)dilatazione, c)apertura e d)chiusura.

1.6. Modelli deformabili

I modelli morfologici sono degli elementi utilizzati a rappresentare forme, usando dei vicoli che dipendono dalle caratteristiche dell'immagine e utilizza la teoria dell'ottimizzazione per adattare maggiormente le forme ai bordi. Questo tipo di metodo serve, a partire da un contorno iniziale, per adattare iterativamente le il contorno agli elementi dell'immagine proposta estraendone la forma.

L'adattamento ai contorni è portato da un bilanciamento delle caratteristiche dell'immagine. Il bilanciamento è tra le forze esterne, che portano al movimento della curva tramite un'attrazione o repulsione, e dalle forze interne, che garantiscono la continuità della curva e impediscono che la stessa si possa arrotondare portando ad un adattamento migliore. Il bilanciamento tra queste forze porta ad un progressivo adattamento della curva che rappresenta il contorno alla specifica forma degli elementi che appartengono all'immagine.

Ci sono 2 tipi di modelli deformabili: gli snake che sfruttando un movimento parametrico della curva e il **level-set** detto anche modello deformabile geometrico che sfrutta una geometria tridimensionale creata iterativamente per adattarsi al contorno proposto.

Il level set proposto da Osher e Sethian nel 1988, consiste nel rappresentare un contorno C in uno spazio di dimensione superiore in cui il contorno stesso non è altro che il livello di zero della curva nello spazio di ordine superiore.[25]

Quindi a partire da un bordo iniziale il level set ad ogni iterazione (istante di tempo) prende le coordinate del punto e la distanza segnata dal contorno iniziale. La funzione 3D su cui è definito il contorno deriva dalla modellizzazione geometrica dell'immagine. Variando i livelli del contorno sulla funzione 3D o variando la funzione nel tempo, questa restituisce un contorno dinamico che si adatta alle forme e i profili dell'immagine. La porzione di curva inizialmente all'interno degli oggetti si espande fino al bordo superiore, la porzione all'esterno collassa sui bordi stessi, sdoppiandosi se del caso.

I matematici Chan e Vese proposero nel 2001 un metodo per rendere i modelli geometrici indipendenti dai bordi e dal gradiente. In seguito è descritta la formula. [26]

$$F_1(C) + F_2(C) = \int_{inside} |u_0(x, y) - c_1|^2 + \int_{outside} |u_0(x, y) - c_2|^2 dx dy$$

Dove F_1 è la funzione della curva che indica se l'oggetto è all'interno della curva se questo è maggiore di 0, F_2 indica se positivo se la curva è interno all'oggetto. C_1 indica la media dei valori dei pixel all'interno della curva, c_2 indicata media dei valori dei pixel all'esterno della curva e u_0 rappresenta l'immagine originale.

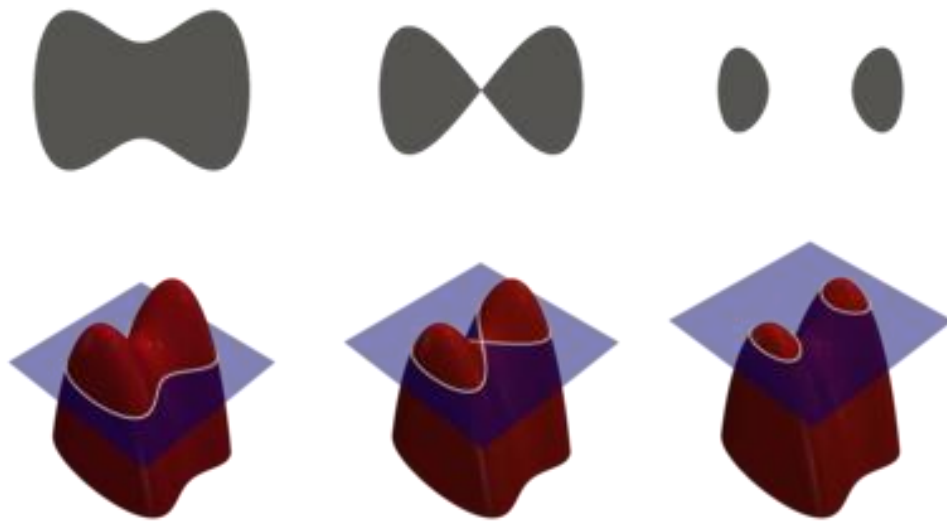


Figura 16 Esempio dell'adattamento dei contorni attraverso diversi livelli, nel metodo del level set.

1.7. Stato dell'arte

In letteratura sono stati presi come riferimenti gli articoli relativi all'elaborazione, segmentazione e detect object di oggetti in campo odontoiatrico utilizzando immagini radiografiche (sia CT che RX classiche).

Chiarezza per quanto riguarda l'elaborazione di immagini dal punto di vista odontoiatrico è stato fatto dall'articolo "A Review of Machine Learning Methodologies for Dental Disease Detection" [9] che descrive le tecniche di auto detect più sviluppate su immagini radiografiche. Il fine di queste reti è ricercare malattie odontoiatriche. Si analizza che le più utilizzate, nonché le più performanti sono le Mask – RCNN. Arrivano ad un'accuratezza del 70% all'80% per il detect delle malattie dentali. L'articolo conclude affermando che non c'è una rete definitiva dal punto di vista odontoiatrico, ma se ne utilizzano molte personalizzate e adatte ad uno specifico campo. Le più utilizzate in questo momento sono le RCNN, e le sue varianti (Fast, Faster e Mask RCNN), e le AlexNet pre-allenate.

Quindi cercando tra i più recenti articoli che presentassero buoni risultati dal punto di vista di detect object, i migliori risultati li porta l'articolo "Tooth detection and numbering in panoramic radiographs using convolutional neural networks" [10] il quale propone un metodo di segmentazione dei denti e successivamente una classificazione utilizzando le Faster RCNN su delle immagini radiografiche odontoiatriche. Per quanto riguarda il detect dei denti il sistema porta ad un ottenimento dei seguenti risultati: una sensibilità pari a 0.9941 e una precisione pari a 0.9945.

Da qui si è scelto di utilizzare per lo svolgimento della tesi, almeno inizialmente una Faster RCNN, una delle reti più performanti e computazionalmente meno gravose. "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks" [11] descrive la struttura, le caratteristiche e i risultati ottenuti del metodo di object detect della Faster RCNN. Questo articolo pone la Faster RCNN come stato dell'arte delle reti neurali region proposal. Importanti sono le caratteristiche di velocità che ha questa rete oltre la precisione e la possibilità di classificazione nell'output.

In ambito biomedico degna di nota è anche “Dental Material Detection based on Faster Regional Convolutional Neural Networks and Shape Features” [12] il quale propone un approccio per trovare il materiale dentale dei pazienti su delle immagini radiografiche di panoramica. Il metodo proposto si divide in 2 step, il primo che pone l’attenzione sulla selezione della ROI che contiene solo l’arcata dentale, utilizzando le Faster RCNN e successivamente un metodo per classificare il materiale di cui sono composti i denti (nel caso di protesi o otturazioni dentali). Interessante è la il metodo con cui divide l’arcata dentale dallo sfondo.

Anche dal punto di vista diagnostico le reti neurali hanno un buon rendimento, infatti “Detection and Classification of Dental Pathologies using Faster-RCNN in Orthopantomogram Radiography Image” [13], descrive un modello di detect dei denti e rilevazione di possibili patologie dentali a partire da immagini radiografiche ortopantomiche per odontoiatria. Il modello proposto sfrutta le Faster RCNN sulle immagini di ortopantomica e pone appunto le Faster RCNN allo stato dell’arte per quel che riguarda il rilevamento di patologie.

Infine negli ultimi 2 anni si sono fatte strada delle varianti della Faster RCNN che permettono una segmentazione con creazione finale di una maschera. Queste sono le Mask-RCNN per il bidimensionale, e le Mesh-RCNN per i volumi. Nell’articolo “Tooth Detection from Panoramic Radiographs Using Deep Learning” [14] si descrive l’efficienza dell’auto segmentazione e successiva classificazione dei denti nelle immagini radiografiche rappresentanti le panoramiche dentali tramite la Mask-RCNN.

Per quanto riguarda invece le applicazioni di elaborazione di immagini radiologiche in letteratura si citano spesso le U-Net per la segmentazione. L’articolo in assoluto più citato è “Convolutional neural networks: an overview and application in radiology” [15] dove si descrivono le CNN e l’utilizzo di queste per il detect di vari task, inclusa la radiologia. Oltre ai metodi sfruttati dalla CNN per adattarsi automaticamente all’allenamento spaziale, ovvero sfruttando le strategie di backpropagation, convolutional layer, pooling layer e i fully connected layers su porzioni di immagini che lo stesso algoritmo seleziona. In particolare l’articolo descrive le criticità e le possibili applicazioni in campo radiografico, con proposte di possibili tecniche per l’utilizzo in campo diagnostico. Inoltre sono anche trattati gli approcci dal punto di vista di segmentazione e di classificazione.

Nel campo odontoiatrico queste reti sono utilizzate soprattutto per una segmentazione dei denti a partire da immagini radiografiche rappresentanti panoramiche come si può evincere dall'articolo "Accurate segmentation of dental panoramic radiographs with u-nets" [16]. Esso propone un metodo sfruttando una Fully Connect Network, basato sull'architettura della U-Net. I risultati sono di tutto rispetto e la pongono allo stato dell'arte come rete grezza per la segmentazione. Di fatto sono partiti da 1500 immagini, utilizzandone 1200 per il training, ottenendo un DICE pari 0,934. E' Importante il metodo sfruttato in "UNet Architecture Based Dental Panoramic Image Segmentation" [17] per la riduzione del rumore radiografico, causato da otturazioni e protesi dentali. L'articolo in totale tratta lo sviluppo di un algoritmo che a partire da delle immagini radiografiche classiche della zona mascello-mandibolare prova ad effettuare una segmentazione utilizzando delle reti neurali in deep learning. Il metodo proposto sfrutta le UNet come modello di rete neurale. Si arriva un'accuratezza del 97% e una DICE di 94%.

Ma in quanto l'obiettivo della tesi è quello di sviluppare una segmentazione volumetrica, indispensabile è "3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation" [18]. In questo articolo si introduce l'uso delle UNet per una segmentazione volumetrica allenate su immagini volumetriche. Si descrivono 2 metodi per la segmentazione: il primo è un metodo semi automatico e sfrutta delle annotazioni rare su delle slice per poi provvedere alla segmentazione tridimensionale, il secondo metodo è un metodo automatico che allena su un dataset volumetrico dove sono annotate la classe dell'oggetto e lo sfondo.

"U-Net – Deep Learning for Cell Counting, Detection, and Morphometry" [19] tratta dello sviluppo di una UNet preallenate riallenata per il detect di un nuovo task per la segmentazione di cellule. Molto importante è il modo in cui il metodo bypassa il problema delle piccole dimensioni del dataset necessario per l'allenamento delle reti utilizzando un algoritmo di aumento del dataset, o anche "U-net: Convolutional networks for biomedical image segmentation" [20] che presenta sia delle reti e strategie di allenamento che sfruttano un forte aumento del dataset iniziale a partire da un efficiente dataset con segmentazioni manuali. Si sfruttano quindi delle UNet, e le si allenano con poche immagini.

Cercando metodi di segmentazione tridimensionale che sfruttano CBCT, allo stato dell'arte si pone "Improve 3d cone-beam ct reconstruction by slice-wise deep learning" [21] il quale presenta un metodo di segmentazione volumetrica sfruttando UNet su delle CT odontoiatriche

a bassa energia e descrive i vantaggi dal punto di vista diagnostico. Si propone un algoritmo di ricostruzione basato su una CNN dove questa è sviluppata per portare una proiezione dei denti e delle protesi dentali per ridurre il rumore sulla mappa d'attenuazione. Qui si passa da un approccio tridimensionale ad un problema bidimensionale, per poi ricostruire l'oggetto.

Tutte i metodi sopracitati sfruttano una segmentazione semantica delle strutture anatomiche, che può portare ad una scarsa fedeltà della segmentazione. In “Automatic dental root CBCT image segmentation based on CNN and level set method” [22] per ridurre il problema, oltre ad una rete con un'architettura CNN, si utilizza il level set per rispettare meglio i contorni delle strutture dentali. Precisamente si usano CNN per la segmentazione precisa dei denti, soprattutto nella zona sotto radice dove il rumore è maggiore, attraverso una struttura end-to-end in grado di segmentare automaticamente i denti dalle immagini TC odontoiatriche. In particolare si sfruttano reti con architettura UNet. Successivamente, viene utilizzato un modello di level-set per perfezionare gli output della CNN che possono migliorare ulteriormente i risultati della segmentazione.

Infine unico articolo trovato in letteratura per la segmentazione tridimensionale automatica di un canale in immagini CBCT è “Automatic mandibular canal detection using a deep convolutional neural network” [23] il cui obiettivo è la segmentazione del canale mandibolare sfruttando modelli basati su 2D SegNet, 2D e 3D U-Net. I risultati migliori di segmentazione del canale sono delle 3D U-Net che ha mostrato la migliore accuratezza globale di 0,99.

Quindi si può dedurre che le UNet sono molto performanti e utilizzate. Non si può dire che siano allo stato dell'arte in quanto è un campo in costante e rapido sviluppo, ma di sicuro offrono delle ottime soluzioni per la segmentazione in campo biomedico sfruttando delle immagini radiografiche. In “Fully Convolutional Networks for Semantic Segmentation” [24] descrive l'architettura fully connected network applicata ad una UNet, ovvero la base delle architetture viste in letteratura.

1.8. Strumenti utilizzati

Per lo svolgimento dell'algoritmo è stato utilizzato dal punto di vista hardware un laptop ThinkPad della Lenovo con le seguenti caratteristiche:

- Processore intel i7-8550U con una CPU 1,8GHz a 1,99GHz
- RAM da 32GB
- Windows 10 professional
- Scheda video condivisa Intel(R) UHD Graphics 620
- SSD M.2

Mentre dal punto di vista software è stati utilizzati il programmi forniti dal Politecnico di Torino ovvero Matlab per la programmazione e Excel per la gestione dei dati. La versione di Matlab utilizzata è quella riferita al 2021a. Sono state sfruttate in più le seguenti librerie:

- Signal processing toolbox
- Image processing toolbox
- Computer vision toolbox
- Deep learning toolbox
- Parallel computing toolbox

Si precisa che in mancanza di una workstation e con un elaboratore dotato di memoria e hardware limitato il metodo sviluppato è stato fortemente condizionato dal potere computazionale a disposizione e dal tempo necessario per l'esecuzione dei codici. Infatti sono state sfruttate reti neurali “leggere” nell'allenamento.

2. Metodo proposto

2.1. Flow chart e descrizione

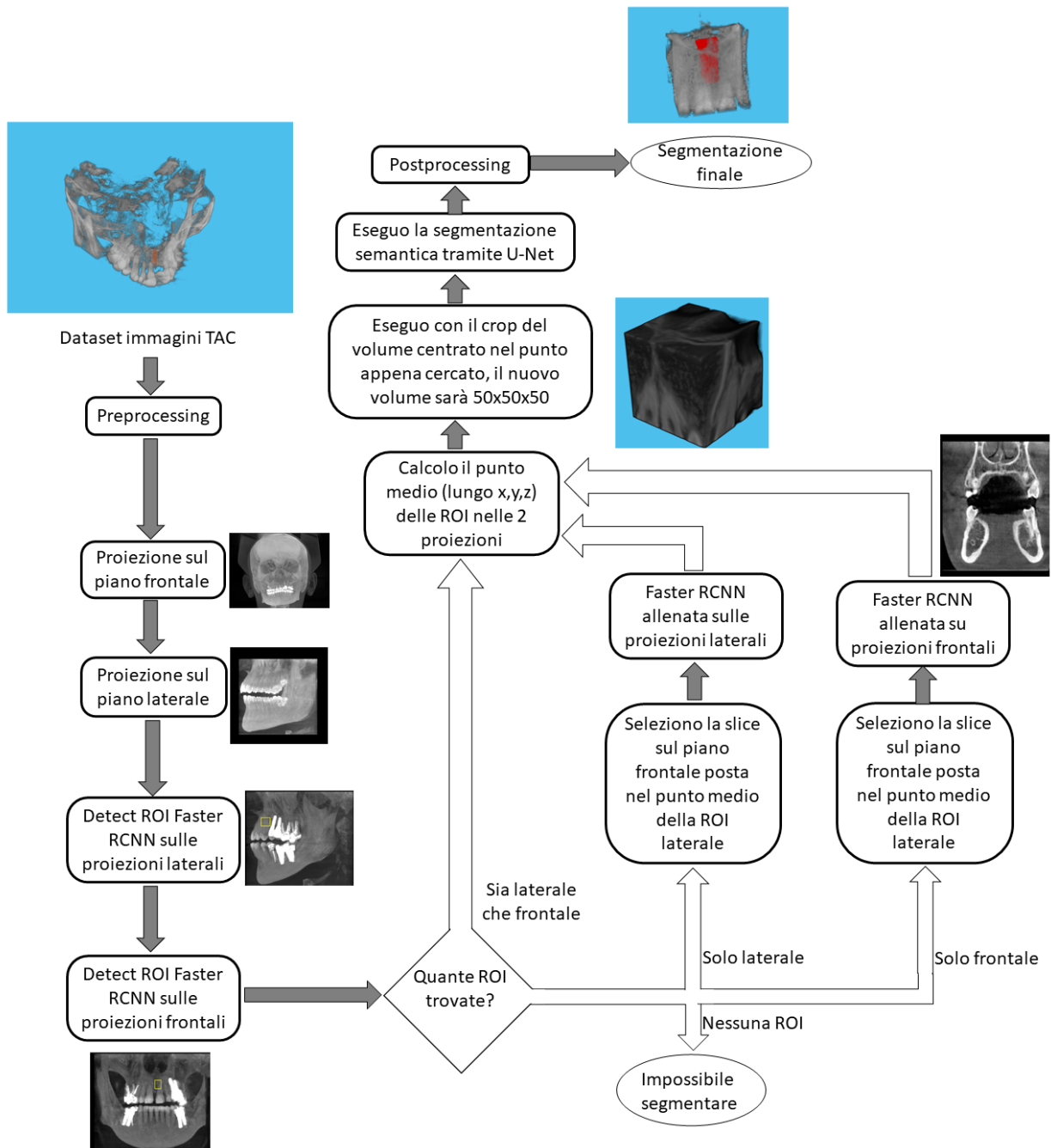


Figura 17 L'immagine rappresenta il flowchart del metodo proposto.

Il metodo proposto in questa tesi per la segmentazione automatica del dotto naso palatino, si basa su una segmentazione semantica data da delle reti neurali UNet.

Inizialmente i dati vengono caricati e resi omogenei nel preprocessing, in quanto molto spesso questi hanno dimensioni e risoluzioni diverse.

Qui vengono proiettate sul piano frontale e sagittale per applicare le reti che estraggono le regioni d'interesse (ROI), nello specifico le Faster RCNN. Qui in output avremo delle ROI che prevedono la posizione del canale. È possibile che in output ci siano più ROI. E' quindi necessario classificare i diversi output ed estrarre solo la ROI con più alta probabilità che rappresenti il canale. A parità di valori si estrae solo la ROI che rispetta determinati criteri (più centrale lungo l'asse trasversale).

Nel caso non dovesse essere trovata la regione solo su una delle 2 proiezioni, viene fruttata la ROI trovata nella proiezione complementare per ottenere una slice sul piano dove non è stato effettuato il detect ed effettuare un recupero della ROI singola slice. Questa deve contenere necessariamente il canale, quindi la si passa in una Faster RCNN allenata sulle singole slice. Per avere la certezza che la slice scelta contenga il canale se ne sceglie una appartenente parzialmente alla ROI trovata dalla sull'altro piano.

Questa viene utilizzata solo per aumentare le performance di ricerca della regione contenente il canale e non si applica direttamente la ricerca delle ROI sulle singole slice perché sarebbe troppo dispendioso dal punto di vista computazionale nonostante questa in teoria sia più affidabile in quanto il canale è ben visibile.

Una volta trovata la zona contenente il canale, viene estratta un volume di riferimento tridimensionale a partire dalla lavorazione delle ROI appena analizzate. Qui si applica un crop del volume creando dei volumi più piccoli pronti per passare nella rete neurale di segmentazione semantica.

Quindi per la segmentazione viene utilizzata una 3D U-Net che in uscita porterà 2 classi, ovvero: sfondo e oggetto (canale).

A quest'uscita viene applicato un post processing per migliorare il risultato della segmentazione.

Il post processing è formato innanzitutto da una classificazione usata per l'eliminazione dei corpi troppo piccoli e successivamente vengono applicati degli operatori morfologici classici per migliorare la qualità della segmentazione.

Infine viene applicato un operatore dinamico parametrico (level set) in modo non perdere l'effetto dei bordi.

Il risultato è la segmentazione del dotto naso-palatino accurata e in grado di conservare il più possibile le caratteristiche anatomiche e morfologiche del canale.

2.2. Il dataset e l'acquisizione dati

Le immagini sono TC odontoiatriche a bassa energia e sono salvate in formato DICOM. Le immagini del dataset presentano delle caratteristiche molto diverse tra loro, le cui le più rilevanti sono: la distanza tra due piani (che determina la differente risoluzione) e la dimensione.

Cambiando la distanza tra due piani (slice-rate), cambia anche la risoluzione spaziale e questo può essere un grave problema per la ricerca automatica degli oggetti, dovendo trovare canali rappresentati in modo molto diverso.

Quindi si rende necessario un preprocessing il cui scopo è quello di omogeneizzare le immagini in modo da avere sempre la stessa risoluzione, e dunque anche la stessa proporzione px/cm.

Per fare ciò si acquisiscono le immagini DICOM con le relative caratteristiche, in base alla caratteristica Slice-Rate X, Slice-Rate Y e Slice-Rate Z. Successivamente le immagini vengono ricampionate impostando uno slice-rate fisso. In questa tesi è stato impostato lo slice-rate pari a : 0.5 cm/px su tutte e tre le dimensioni.

Quindi le immagini vengono convertite prima di essere salvate da formato int16 (ovvero il formato con cui le immagini vengono acquisite) in formato uint8, ciò viene fatto perché anche non riducendo molto la quantità di informazione riduce sensibilmente lo spazio occupato sul disco e riduce il tempo di allenamento delle reti.

2.3. Le reti Faster RCNN

L'idea principale è quella di triangolare il volume contenente il canale sfruttando 2 reti neurali Faster RCNN, che eseguono la ricerca portando in uscita le regioni d'interesse contenenti il canale. La ricerca delle ROI viene applicata su due proiezioni dell'immagine TC tridimensionale, ovvero la proiezione sul piano frontale e sul piano laterale.

Una volta triangolata la ROI d'interesse è possibile ricavarne un volume tridimensionale attraverso la segmentazione semantica della UNet. Triangolando la posizione del canale è possibile ridurre il volume attraverso un crop dell'immagine 3D intorno al punto medio delle ROI appena trovate. In questo modo si aumenta la proporzione tra volume da cercare e sfondo, migliorando l'efficienza della segmentazione e in più riducendo i tempi di ricerca. Inoltre questo metodo, riduce la profondità della rete neurale UNet, portando un notevole riduzione del tempo necessario per l'allenamento.

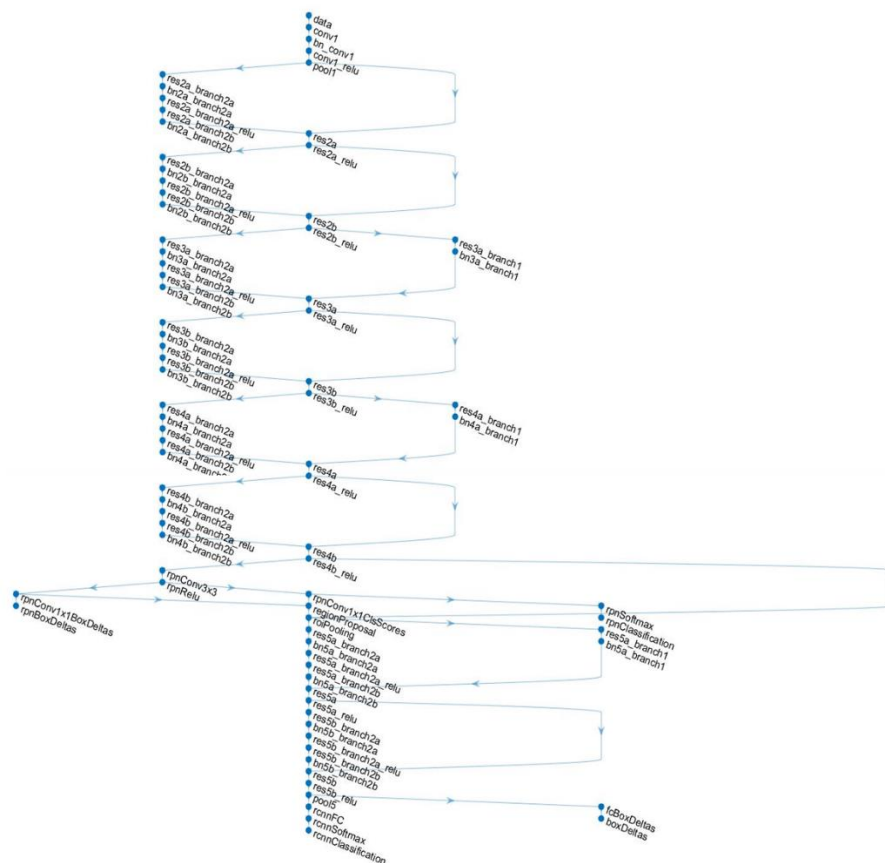


Figura 18 Struttura della Faster RCNN.

Come algoritmo di estrazione delle caratteristiche si è utilizzato una resnet18 e come livello di caratteristiche si è utilizzato un res4b_relu come consigliato dagli esempi matlab ideale per l'algoritmo resnet18.

2.3.1. Preprocessing

Il preprocessing è sostanzialmente il metodo per portare le immagini CBCT tridimensionali in immagini bidimensionali.

Per la ricerca delle ROI si sfrutteranno le due proiezioni sul piano frontale e laterale.

Si utilizza il metodo dei massimi per ottenere le proiezioni, quindi si utilizza l'operazione di massimo per eliminare una dimensione, e questa viene applicata sul vettore dei piani paralleli che rappresenta il pixel della panoramica. Non si utilizza l'algoritmo inverso della retroproiezione filtrata tramite l'antitrasformata di Radon, in quanto aumenterebbe ancor di più il rumore delle immagini che, non essendo già prive di rumore ed artefatti radiografici, rischiano di essere troppo distorte.

Successivamente si opera sulla dimensione delle immagini. Le Faster RCNN funzionano anche su formati diverse di immagini, ma si è preferito uniformare le immagini con una uguale dimensione per ridurre la variabilità e aumentare l'efficienza di ricerca delle ROI.

Per adeguare tutte le immagini si utilizza la tecnica dello zeropadding. Si portano le immagini ad una dimensione di 550x450 pixel dove nella porzione centrale è presente l'immagine, e se questa è minore delle dimensioni prefissate si riempiono le zone vuote con il valore di 0.

In fine il preprocessing si conclude con una modifica sull'istogramma delle luminosità delle stesse immagini. La regolazione del contrasto rimappa i valori di intensità dell'immagine sull'intero intervallo di visualizzazione del tipo di dati, in modo da incrementare i livelli di grigio utilizzati e aumentarne il contrasto.

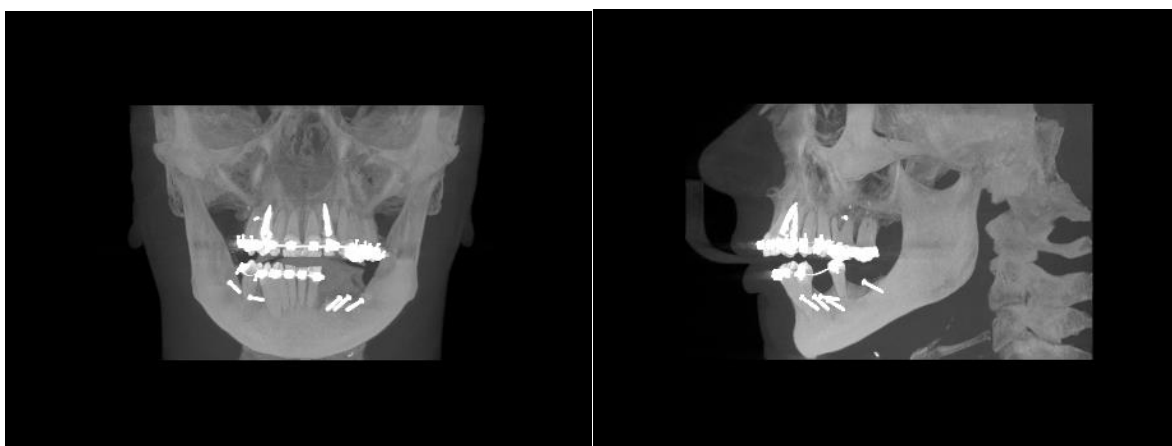


Figura 19 Un esempio di pre processing, in alto sono rappresentate le immagini in input e in basso le immagini dopo il postprocessing. A sinistra sono le immagini rappresentanti il piano frontale, a destra il piano laterale.

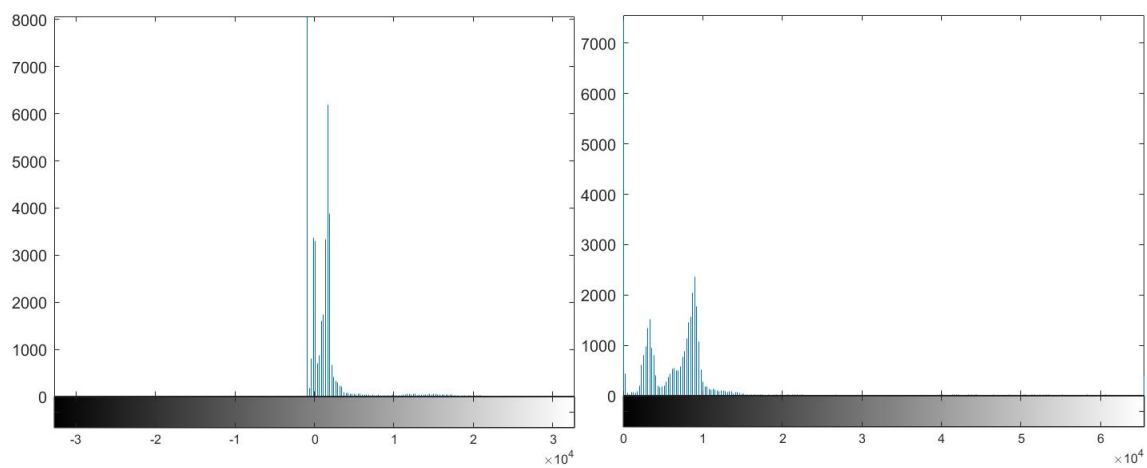


Figura 20 A sinistra l'istogramma delleluminosità dell'immagine prima del preprocessing e a destra l'istogramma delle luminosit  dopo il preprocessing

2.3.2. *Allenamento delle Faster RCNN*

E' necessario il corretto allenamento delle 2 reti neurali. Per l'allenamento si utilizza il dataset diviso in 3 parti: training set, validation set e test set.

Nell'input della rete, nella fase di allenamento, si devono fornire le immagini e le coordinate della ROI contenente il canale. Questo deriva dalla segmentazione manuale, e la rete in base all'input dato si allenerà, in quanto rete supervisionata. Per l'allenamento si utilizza il training set e il validation set.

La struttura della rete è come vista in precedenza e sarà allenata per trovare la ROI di una sola classe ovvero la ROI contenente il canale.

Per l'allenamento è stato utilizzato un incremento del dataset iniziale. Già nei seguenti articoli [19-20] valido per l'allenamento delle UNet, è stato introdotto in maniera analoga un metodo per migliorare l'allenamento delle reti sfruttando un aumento del dataset iniziale. Nel metodo descritto si utilizza un aumento del dataset dovuto ad una manipolazione delle immagini, in cui le rovescia facendone lo speculare su più assi, oppure l'estrazione dell'immagine di una dimensione appena più piccola ripetuta più volte in maniera casuale.

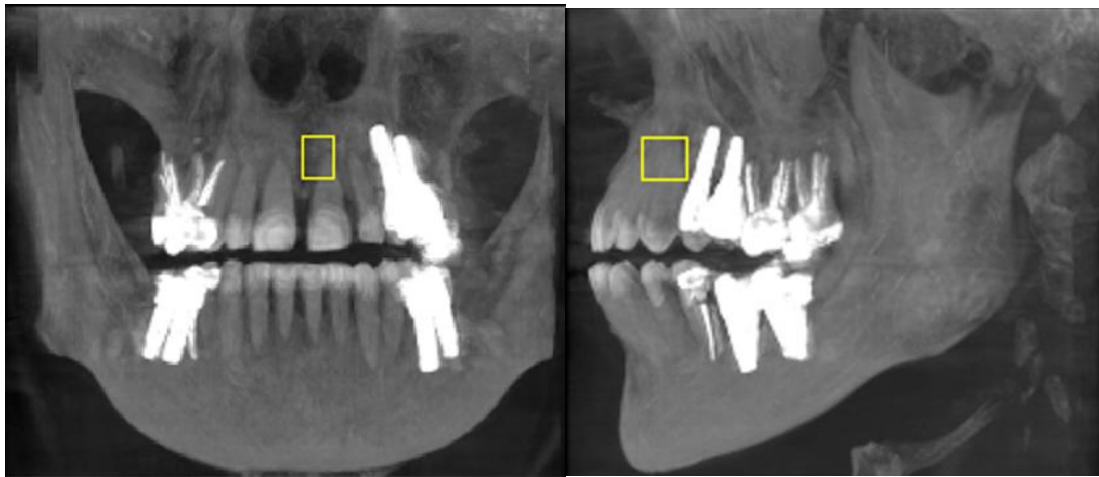


Figura 21 In giallo un esempio di ROI date in input alla rete contenenti il canale.

Le Reti sono state allente su 50 epoche, un MiniBarchSize di 4, un LearnRate iniziale pari a $1e-3$.

Per l'allenamento è stato utilizzato un dataset di 608 immagini. 100 sono state utilizzate per il processo di validazione.

2.3.3. *Classificazione delle soluzioni*

L'output delle Faster RCNN non darà necessariamente una singola soluzione, ma una popolazione di soluzioni. A questo punto è necessaria una classificazione per escludere i falsi positivi o veri positivi meno precisi e considerare soltanto le ROI più attendibili.

Per quanto riguarda l'estrazione delle ROI nelle proiezioni frontali innanzitutto si analizza lo score di attendibilità fornito nell'output della stessa rete. Si prendono solo le ROI i cui scores sono maggiori del valore massimo con una tolleranza delle 2%.

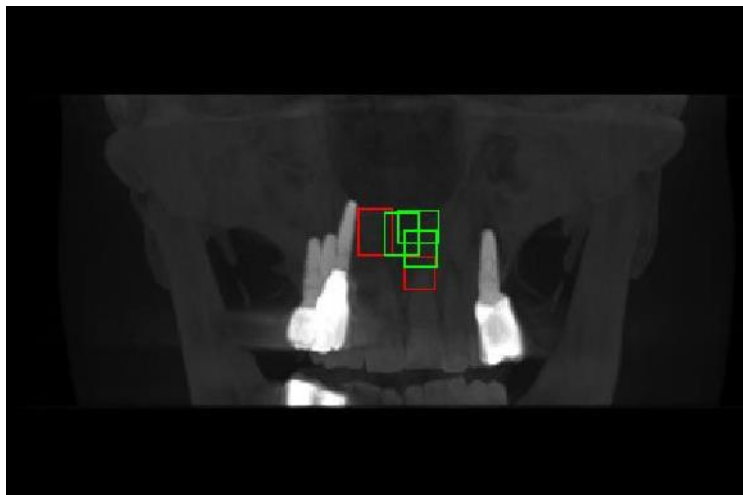


Figura 22 Esempio di prima classificazione, in verde i massimi con la tolleranza del 2%; in rosso le ROI aventi score minore.

Tra le ROI con maggiore attendibilità si considera quella più vicina al piano di simmetria della proiezione del cranio e mandibola. Il piano di simmetria è ottenuto tramite un global thresholding che separa i tessuti molli dal tessuto osseo. A questo punto si calcola il centroide dell'immagine e si prende la coordinata sull'asse trasversale come sistema di riferimento del piano di simmetria. La ROI più vicina al piano di simmetria è considerata corretta.

Se si presentano più ROI ad eguale distanza, si estrae la ROI con area più grande come esatta in quanto c'è più probabilità che contenga interamente il canale.

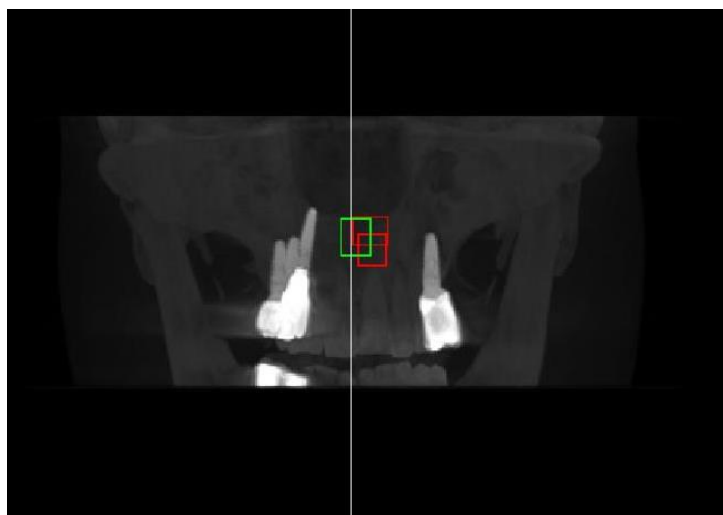


Figura 23 Esempio di seconda classificazione. La linea bianca rappresenta il piano di simmetria. Il riquadro verde è quello più vicino e rappresenta la ROI considerata giusta, i rossi rappresentano le ROI scartate.

Le ROI ottenute dalle proiezioni laterali, rispettano dei criteri simili a quelli utilizzati nelle proiezioni frontali, eccetto per il criterio della minore distanza rispetto al piano di simmetria. Si considera la ROI con score maggiore e se più di una condivide lo score più alto, tra loro si considera solo quella avente un'area maggiore.

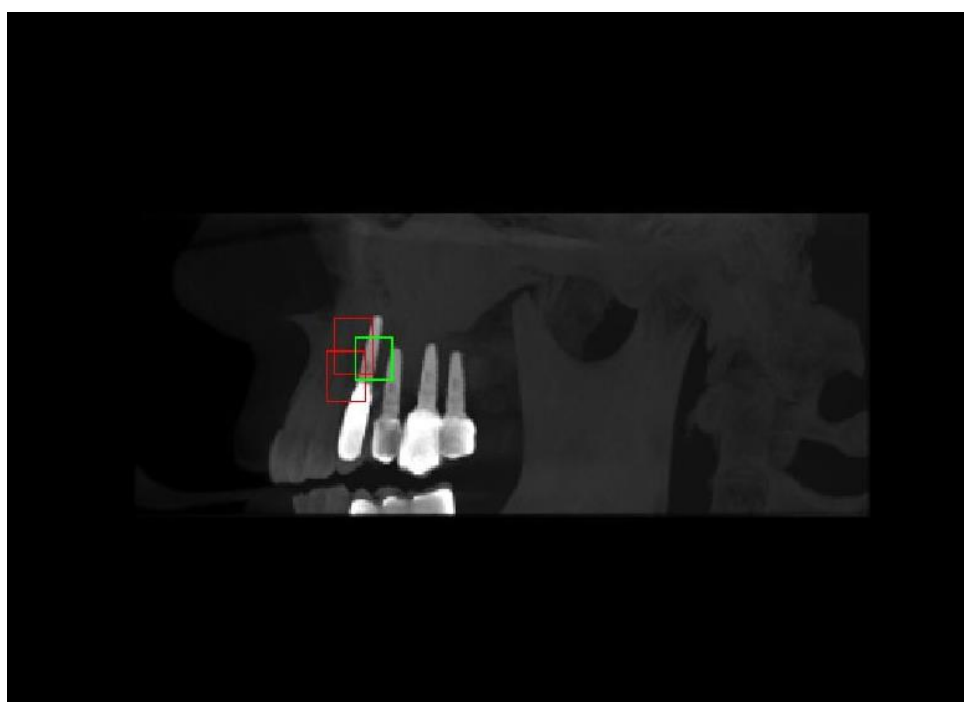


Figura 24 In verde la ROI con lo score maggiore, in rosso le ROI scartate avente score minore.

2.4. *Ripescaggio Faster RCNN*

Nel caso in cui delle prime due reti solo una riesce ad estrarre la ROI, allora il metodo prevede l'utilizzo di altre due reti di supporto (chiamate di "ripescaggio") che riprovano ad estrarre le ROI a partire dalle singole slice e non della proiezione dell'immagine TC.

Si sfrutta il punto medio della ROI complementare a quella su cui è avvenuto il mancato detect, per selezionare la slice contenente il canale e sulla quale avviene il secondo detect.

2.4.1. *Preprocessing*

Il preprocessing di queste reti prevede la selezione della slice, l'omogeneizzazione dell'immagine e l'aggiustamento della stessa operando sull'istogramma.

Per estrarre la slice si prende la quella più vicina alla coordinata trasversale del punto medio della ROI estratta sulla proiezione del piano complementare.

Il ripescaggio non è possibile farlo nel caso in cui non è stata estratta alcuna ROI dalle due prime reti.

Il preprocessing che sussegue l'estrazione della slice è simile a quello descritto precedentemente. Ovvero si adeguano le immagini ad uno stesso formato tramite la tecnica dello zeropadding (550x450 px) e si lavora sull'istogramma delle luminosità attraverso la funzione di rimappatura dell'istogramma dell'immagine.

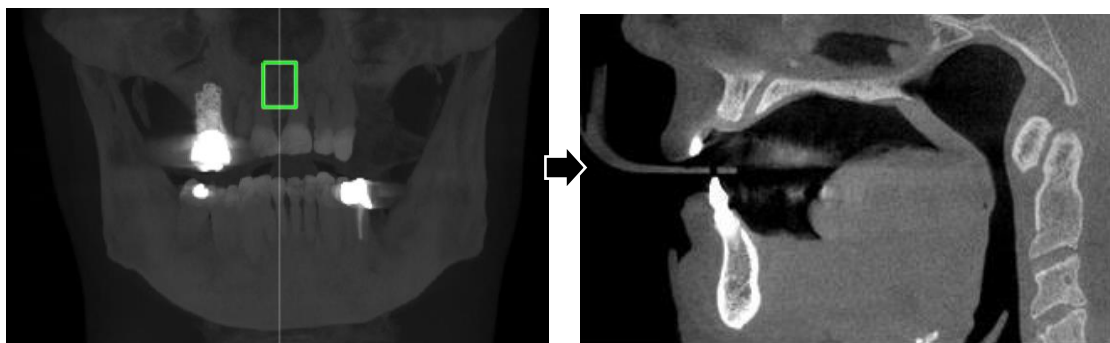


Figura 25 A sinistra l'immagine rappresenta la proiezione su cui si è effettuata l'estrazione della ROI da cui si è ricavato la slice contenente il canale. La slice selezionata è rappresentata dalla linea bianca e poi riportata a destra.

2.4.2. *Allenamento delle Faster RCNN*

Il dataset necessario per l'allenamento di queste reti di "ripescaggio" è minore al dataset totale in quanto se nessuna delle due prime reti restituisce alcuna ROI allora quell'immagine viene direttamente scartata e non prende parte al dataset delle nuove reti.

Anche in questo caso si utilizzeranno all'ingresso della rete oltre le immagini (slice) che contengono le coordinate delle ROI associate alla segmentazione manuale.

La struttura e le caratteristiche dell'allenamento delle reti è identica a quelle descritte in precedenza.

2.4.3. *Classificazione delle soluzioni*

All'uscita delle Faster RCNN nel caso siano presenti più ROI estratte, allora se ne selezionano, in entrambe le reti, solo la ROI che ha lo score maggiore. Nel caso ci siano più ROI che hanno lo score massimo, si prende la ROI con l'area maggiore.

2.5. *U-Net*

Una volta trovate le due ROI si effettua il crop dell'immagine. I volumi vengono ridotti selezionando solo una porzione cubica, centrata nel punto medio delle ROI precedentemente trovate attraverso le Faster RCNN.

I volumi vengono a questo punto lavorati e posti in input della UNet.

In uscita si ottengono più mappe di probabilità, una per classe e una che contiene la classificazione automatica dell'algoritmo che comprende tutte le classi.

Per la segmentazione tramite UNet si utilizza una segmentazione semantica.

Il livello di input è strutturato per ricevere delle immagini di dimensione 50x50x50 px, ed è strutturato per la ricerca di tre classi (sfondo, contorno del canale e canale).

Si utilizza la classica struttura totalmente simmetrica delle reti UNet per elaborare in 3 dimensioni. Quindi si ha una simmetria totale dei due rami della UNet, quello che permette una compressione delle dimensioni e quello che permette l'espansione. Si utilizza un solo sottolivello convoluzionale, questo permette un tempo di elaborazione minore e meno gravoso dal punto di vista computazionale. La grandezza dei filtri applicati è fissata a 2 ed il numero dei primi filtri applicati è fissata a 36. Nel livello di input non è stata applicata alcuna normalizzazione.

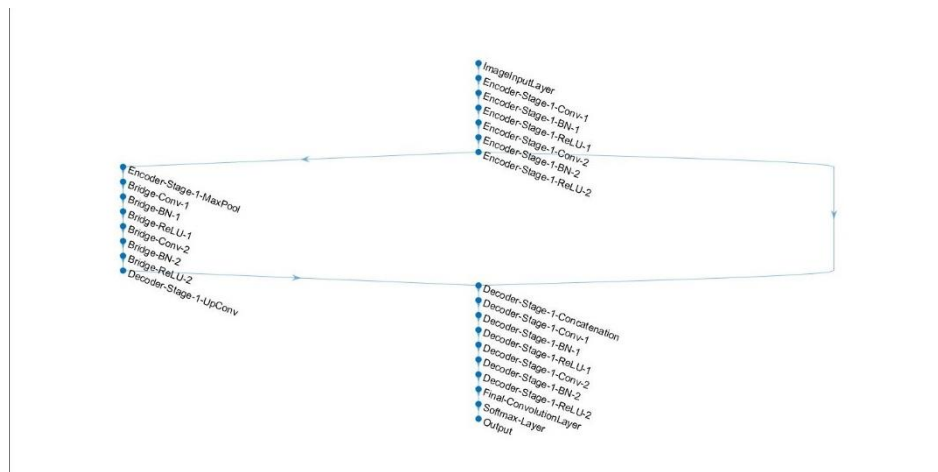


Figura 26 Struttura della 3D UNet.

2.5.1. Preprocessing

Il preprocessing per il passaggio nelle reti UNet è sostanzialmente il crop dell'immagine dove si sfruttano le ROI precedentemente trovate tramite le Faster RCNN per individuare il punto intorno al quale effettuare l'estrazione del volume.

Si applica l'estrazione del volume. A partire dalle coordinate fornite dai punti medi delle ROI precedentemente trovate, si applica un volume centrato in quelle coordinate così da avere un'alta probabilità che esso contenga il canale.

Nel metodo esposto si utilizza un volume estratto di 50x50x50 px.

I vantaggi della riduzione del volume sono molteplici: riducono il tempo di ricerca e di allenamento della rete, ed avendo una maggiore proporzione tra oggetto e sfondo aumenta l'efficienza dell'allenamento e la capacità di cercare il canale della rete stessa.

Anche in questo caso il preprocessing si conclude con una modifica sull'istogramma delle luminosità. Si effettua, come in precedenza, la regolazione del contrasto e la rimappatura dei valori di intensità sull'intero intervallo disponibile per la visualizzazione dei dati.

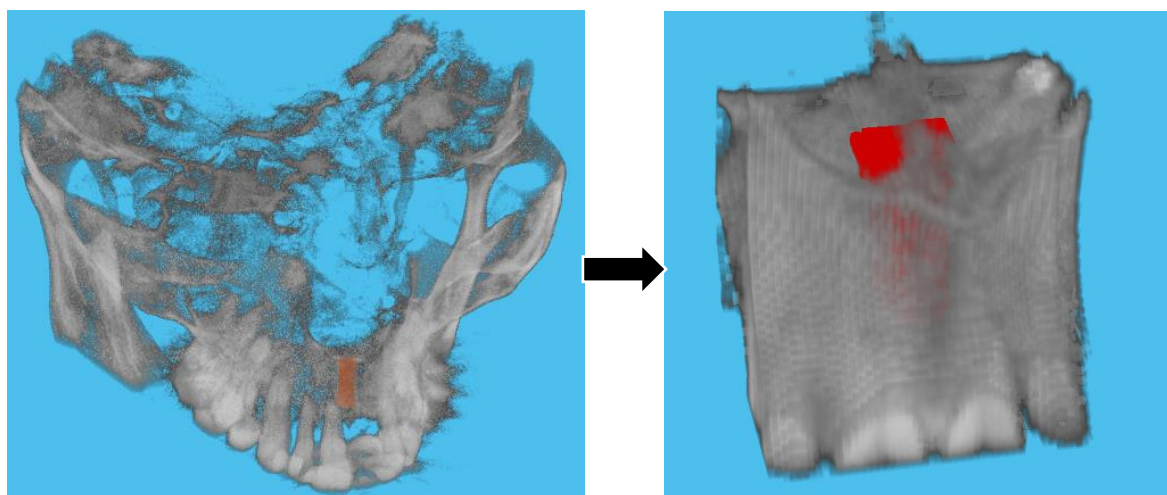


Figura 27 A sinistra l'immagine prima del crop (550x550x450 px), a destra l'immagine dopo il crop (50x50x50 px). In rosso è rappresentato il canale segmentato manualmente.

2.5.2. Allenamento delle U-Net

L'allenamento della rete è effettuato sul dataset diviso in tre parti: training set, validation set e test set.

Il dataset dell'allenamento è formato da 2 volumi della stessa dimensione (50x50x50 px) dove un volume contiene l'immagine croppata, ed è centrato nel baricentro del canale segmentato manualmente e il secondo contiene il medesimo volume centrato ma contenete le maschere che descrivono le classi a cui ogni voxel appartiene.

L'allenamento la rete è effettuato su tre classi: sfondo, contorno e canale.

Per aumentare l'efficienza dell'allenamento è stato utilizzato un metodo che prende ispirazione dall'articolo citato già nel precedente capitolo. Questo metodo tende ad aumentare il dataset, utilizzando un numero di immagini maggiori prese in punti random diversi vicino il canale e a

sua volta ruotarle e capovolgerle. Per ogni immagine tac vengono acquisite 3 immagini, in questo modo il dataset iniziale viene ampliato del 300%. Di ogni immagine acquisita, per aumentare maggiormente il dataset, vengono salvate le immagini speculari rispetto il piano orizzontale e piano verticale; di tutte queste immagini vengono salvate le immagini ruotate di 90°. Quindi il dataset finale sarà aumentato del 2300% rispetto il dataset iniziale.

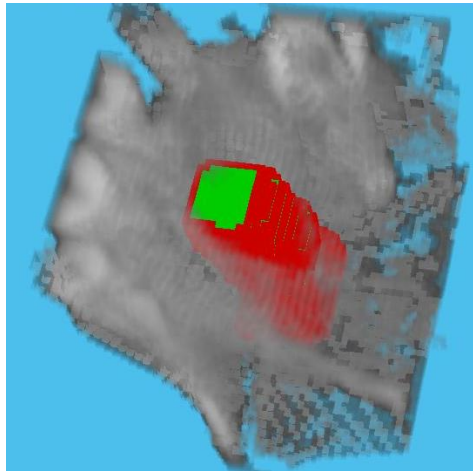


Figura 28 In rosso è rappresentata la classe bordo, in verde la classe canale.

La rete è stata allenata fissando un massimo di 50 epoche, il fattore di allenamento iniziale fissato a 0,000005, e la frequenza di analisi fissata a 5 epoche, la frequenza di validazione è fissata a 400. I parametri d'allenamento e la struttura della rete permettono un allenamento veloce e computazionalmente "leggero". È possibile allenare la rete non solo da workstation, ma anche da pc portatili con caratteristiche standard.

2.5.3. Classificazione delle soluzioni

In uscita della rete si avrà delle mappe tridimensionali di dimensioni identiche al volume dato in input della rete (ovvero 50x50x50 px).

La prima mappa è una classificazione automatica dove ad ogni voxel viene assegnato il valore della classe a cui la rete ritiene appartenga.

Parallelamente in output vengono fornite 3 mappe di probabilità (una per classe) dove ogni mappa contiene la probabilità che i voxel appartengano alla classe rappresentata dalla mappa.

Il metodo proposto sfrutta la mappa la cui classificazione è implementata automaticamente dall'algoritmo stesso della rete, dove si prende per ogni voxel la classe con la probabilità più alta.

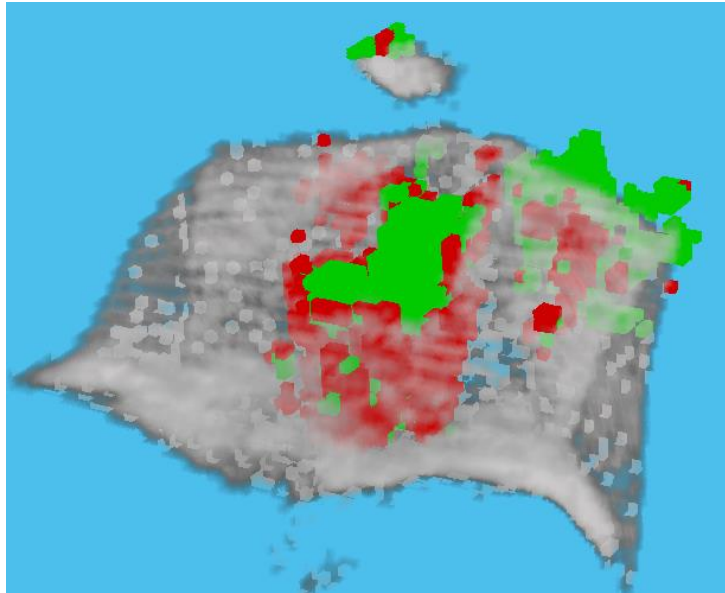


Figura 29 Esempio di output, in verde la classe canale, in rosso la classe bordi.

2.6. *PostProcessing*

Il postprocessing proposto è un postprocessing che agisce su più livelli, ovvero sfrutta una classificazione dei corpi trovati dalla rete per eliminare i corpi più piccoli che sono quasi sempre dovuti al rumore.

Successivamente utilizza operatori morfologici per adattare meglio il corpo e in fine utilizza solo su alcuni corpi una risegmentazione sfruttando degli operatori morfologici parametrici.

2.6.1. *Eliminazione degli artefatti dovuti al rumore*

In primo luogo si nota che la segmentazione semantica presenta alcuni problemi dovuti ad una classificazione errata di corpi molto piccoli, tutto ciò è dovuta appunto alla struttura della segmentazione semantica su immagini affette da rumore.

Per prima cosa si analizzano i corpi della segmentazione, di questi se ne calcolano le superfici esterne in quanto si può notare che anche nelle segmentazioni manuali, dove non è stato possibile trovare la continuità nel corpo, sono presenti più corpi ma tutti caratterizzati da una grande superficie esterna.

Una volta trovate le superfici esterne si assegna ad ogni corpo una classe, il criterio di assegnazione delle classi è basato su una soglia. Si considerano artefatti dovuti al rumore tutti i corpi la cui superficie è minore alla metà della superficie del corpo più grande e successivamente eliminati.

Questo metodo consente di ricavare alla fine uno o due corpi più influenti trovati dalla UNet.

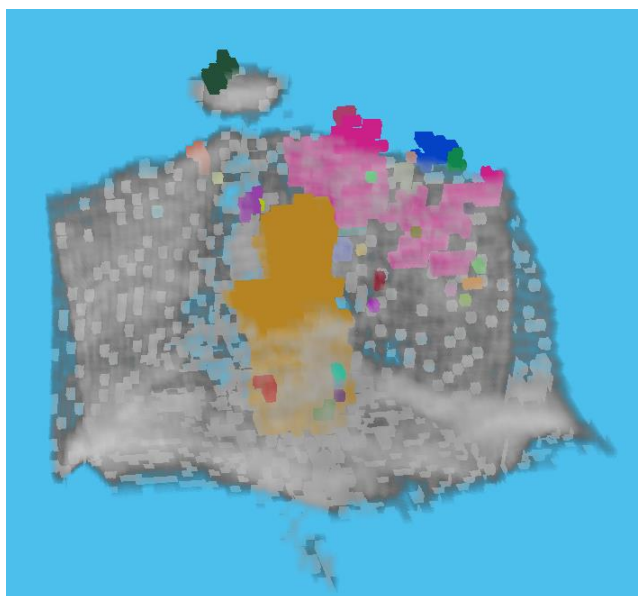


Figura 30 Risultato della segmentazione semantica, dove sono stati descritti con colori diversi i diversi corpi trovati

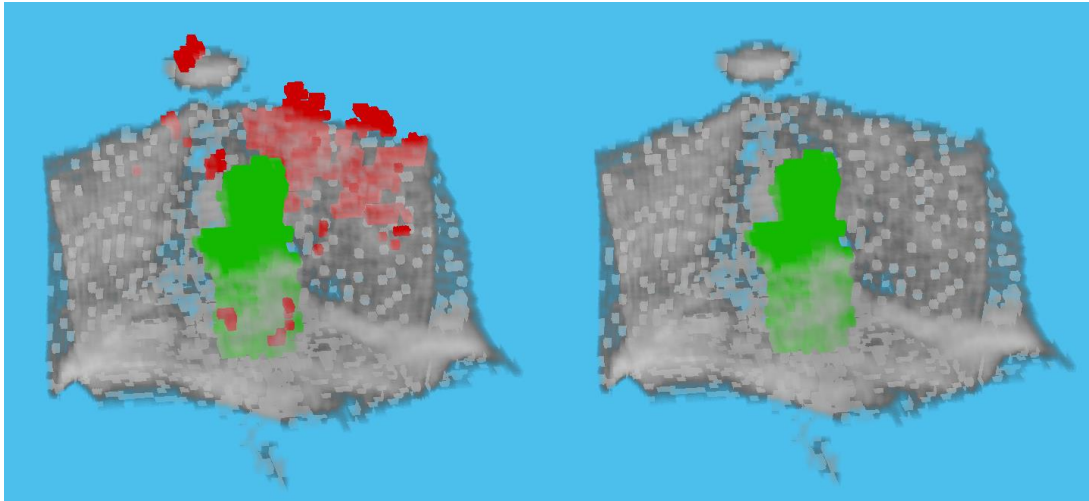


Figura 31 A sinistra la classificazione dei corpi in base alla dimensione della superficie, in rosso sono i corpi piccoli classificati come FP. A destra il corpo considerato correttamente segmentato, ovvero con la maggiore superficie.

2.6.2. Operatori morfologici

I corpi segmentati semanticamente hanno per lo più forme non precise e frastagliate, dovute al detect del solo canale senza considerarne il contorno. Per delimitarne maggiormente il contorno e per dargli una forma più lineare che rappresenti maggiormente il canale si è optato per utilizzare un operatore morfologico di chiusura utilizzando un kernell sferico di diametro 15 voxel. Il risultato è un oggetto da una forma quasi regolare ma che molto spesso tende a sovrasegmentare il canale.

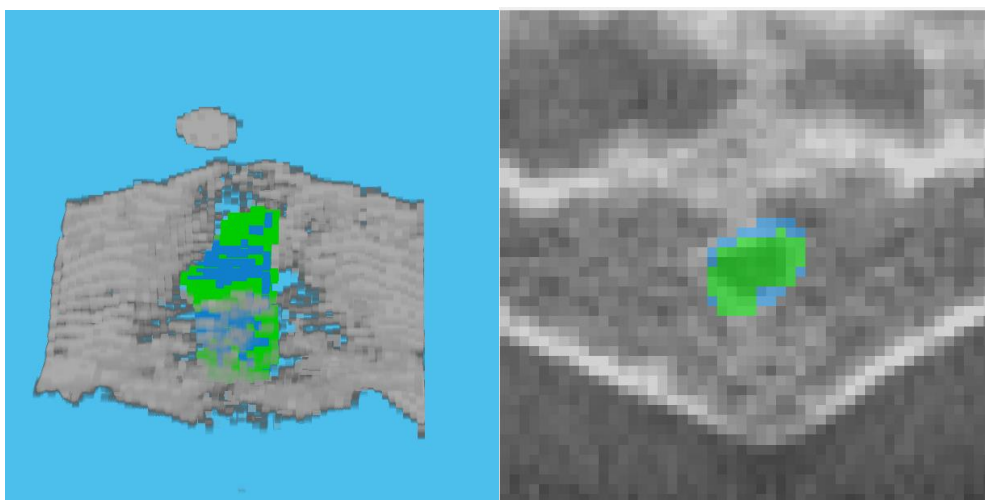


Figura 32 In verde la parte dell'oggetto dovuta agli operatori morfologici. A sinistra il volume tridimensionale, a destra un esempio su una slice del piano trasversale.

2.6.3. *Level set*

A questo punto il risultato è una segmentazione abbastanza accurata del canale, ma per migliorarne la forma dei bordi si utilizza il metodo del Level Set, in quanto la segmentazione dell'UNet essendo semantica non rispetta propriamente le caratteristiche morfologiche dei bordi. Il Level Set viene implementato tramite l'applicazione di una funzione presente nelle librerie di matlab che permette di applicare i contorni attivi attraverso vari metodi (activecontour).

Viene utilizzato il metodo di level-set proposto da Chan-Vese[26] in modo che non dipenda dai gradienti, in quanto spesso il gradiente in questo tipo d'immagini non è molto accurato a causa dell'elevato rumore, e l'utilizzo dei filtri contro il rumore rischia di far perdere informazione, fondamentale per la ricerca di oggetti piccoli.

Il Level Set necessita di una maschera di partenza da fornire in ingresso, in questo caso si utilizza la maschera ottenuta nel passaggio precedente. Le iterazioni del level set sono fissate a pochi cicli per non andare a modificare troppo il risultato tridimensionale dato dalla segmentazione delle reti UNet. Sono stati impostati dei valori che bilanciano energia interna ed esterna in modo che il volume del volume segmentato si riduca in modo da adattarsi al bordo.

Alla maschera posta in uscita, dopo aver applicato i contorni attivi, si utilizza un operatore morfologico di "maggioranza". In questo caso si utilizza un insieme 26-connesso intorno al pixel preso in esame e questo viene conservato solo se l'insieme contiene maggiormente l'oggetto cercato.

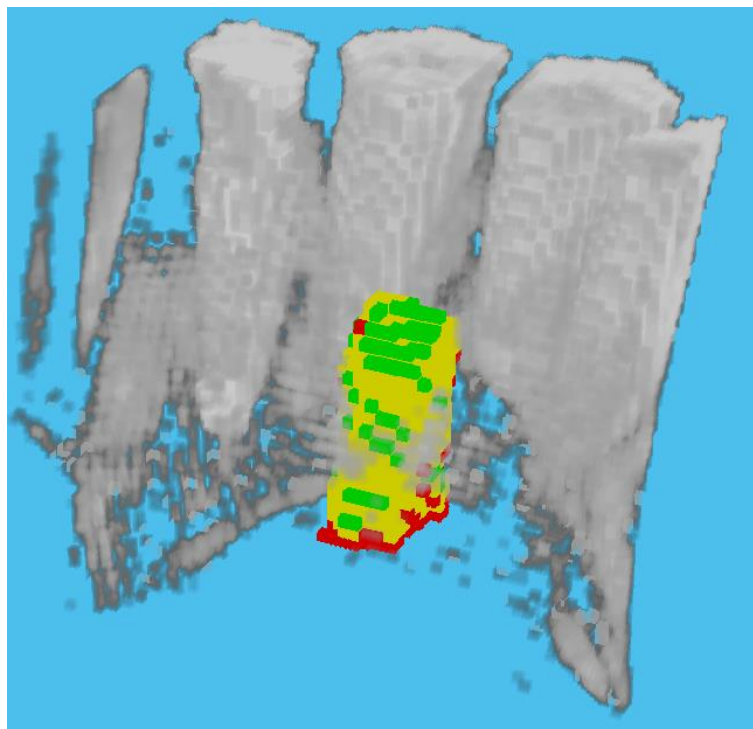


Figura 33 Rappresenta la segmentazione del canale dopo il Level Set. In verde la parte segmentata dal level set, in rosso la parte del corpo eliminata dalla segmentazione.

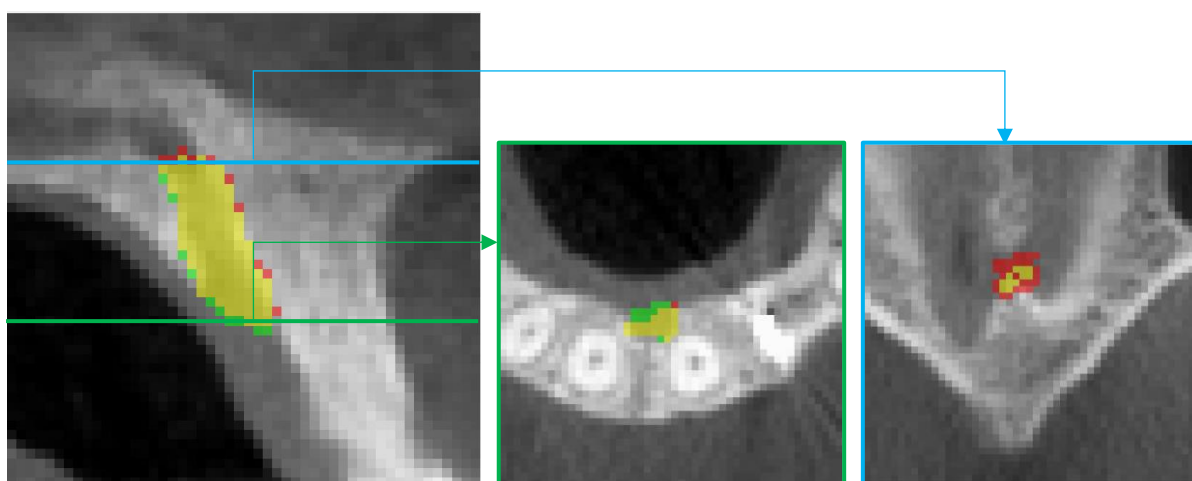


Figura 34 In quest'immagine si rappresenta l'effetto del level set. In verde la sono le zone neosegmentate e in rosso le parti eliminate dalla segmentazione. A sinistra è rappresentata una slice nel piano sagittale. A destra invece 2 slice sul piano trasversale.

2.6.4. *Miglioramento delle performance*

Il risultato ottenuto fin ora è la segmentazione tridimensionale del canale. Per migliorare ancora le performance dell'algoritmo si sono utilizzate delle tecniche di miglioramento della

segmentazione relative alle caratteristiche bidimensionali del canale. Infatti vengono analizzate le singole slice del canale e vengono classificate se corrette o errate in base alla misura del diametro. In base ai diametri ottenuti nella segmentazione manuale, viene impostata una soglia di lunghezza del diametro. Se il diametro della slice è inferiore alla soglia allora il canale segmentato nella slice viene eliminato.

Questo passaggio permette di evitare che nelle terminazioni siano presenti delle slice dove viene trovato erroneamente il canale. È possibile classificare le slice errate in quanto l'area e il diametro dell'oggetto trovato in queste slice è molto piccolo. Dunque se il diametro della slice è minore ad un valore di soglia allora viene considerata come classificata errata e successivamente eliminato il canale trovato in quella slice.

È anche stato creato un criterio di continuità. Infatti se il corpo in una slice non è stato trovato, oppure non ha superato il criterio del diametro, mentre le slice adiacenti contengono il canale, viene applicato l'operatore contorno attivo, dove come maschera iniziale viene utilizzata la maschera contenente l'oggetto segmentato in uno dei due piani adiacenti. In questo modo si ricerca il corpo provando a rispettarne la fedeltà dei contorni e in modo da rispettarne la continuità tridimensionale nello spazio.

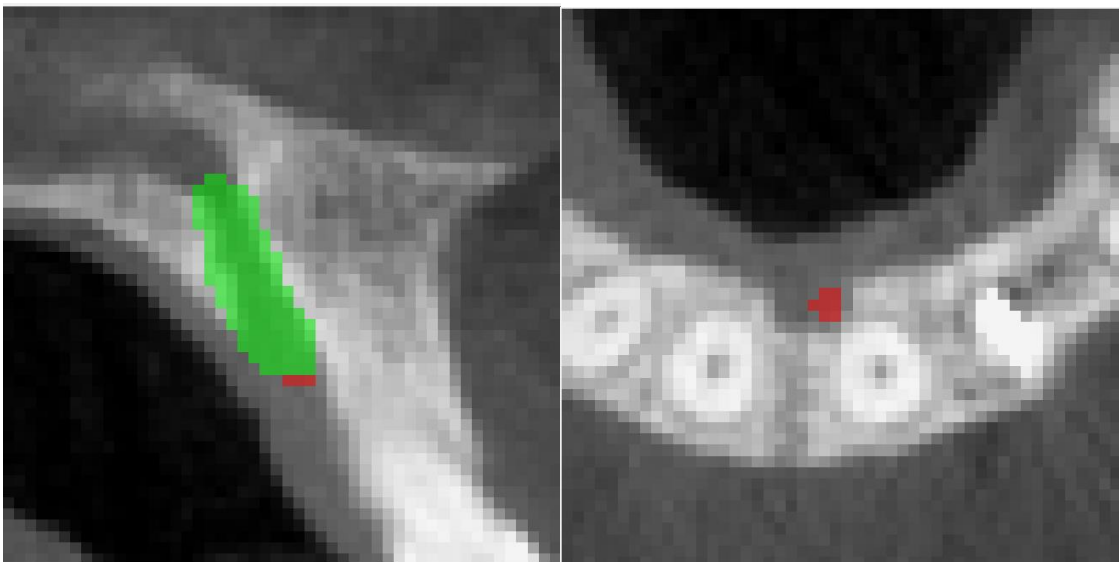


Figura 35 Esempi del processo di miglioramento delle performance. In verde è la segmentazione finale, e in rosso i piani classificati errati.

3. Risultati

I risultati sono ottenuti applicando le metriche alle segmentazioni ottenute dall'algoritmo e alle segmentazioni effettuate manualmente.

Il dataset è stato diviso in tre parti: training-set, validation-set e test-set.

Il training-set è formato da 608 immagini prese casualmente dal dataset totale. Queste immagini sono state utilizzate per l'allenamento della rete, difatti i risultati potrebbero essere corrotti in quanto sono le stesse immagini utilizzate per l'allenamento.

Il validation-set è formato da 100 immagini prelevate casualmente dal dataset totale. In questo caso la porzione del dataset contenente le 100 immagini è utilizzato nella fase di validation, che permette un maggiore allenamento della rete. Queste immagini sono utilizzate nell'allenamento ogni intervallo di iterazioni al fine di controllare se la rete sta apprendendo bene. Infatti queste vengono utilizzate nella fase di apprendimento per dirigere l'allenamento, ma non sono utilizzate frequentemente quindi la rete è meno abituata a riconoscere questa tipologia.

Il test-set, è formato da 100 immagini prese anche esse casualmente dal dataset totale. In questo caso le immagini non vengono impiegate nell'allenamento della rete. I risultati sono relativi a delle immagini che la rete non ha mai visto.

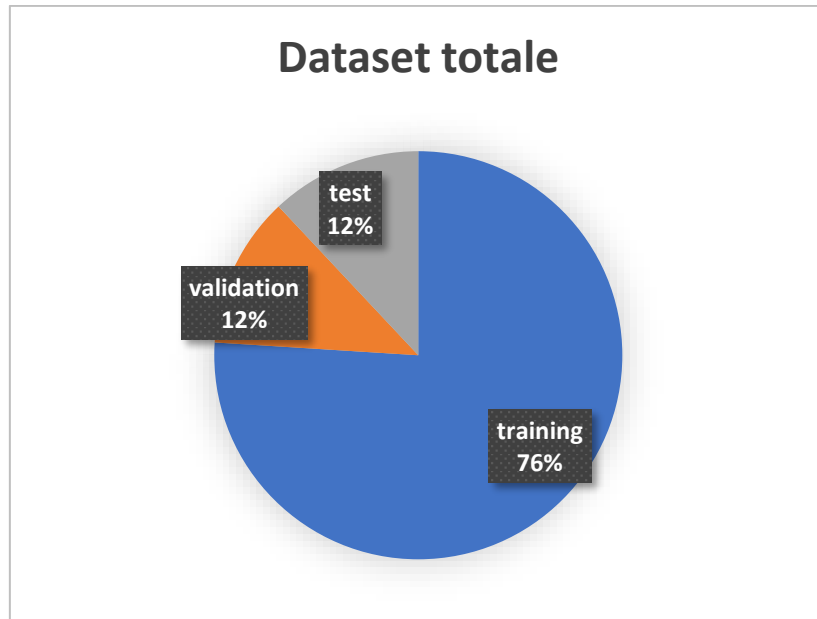


Figura 36 Grafico rappresentante la divisione del dataset.

3.1. *Metriche utilizzate*

Per analizzare i risultati sia delle varie reti che dell'intero algoritmo sono state utilizzate sostanzialmente 2 metriche diverse, una object level e una voxel level.

La prima viene utilizzata sulle reti Faster RCNN ed indica come l'algoritmo è capace di trovare gli oggetti, mentre le seconde metriche indicano la qualità della segmentazione effettuata e descrive la corrispondenza dell'assegnazione della classe canale sui voxel rispetto alla segmentazione manuale.

Infine si è scelto di utilizzare l'inclinazione, lunghezza e diametro medio come metriche per descrivere la conservazione delle caratteristiche morfologica ed anatomiche del canale.

3.1.1. *Object level*

Come metrica per descrivere l'efficienza della ricerca corretta degli oggetti che vengono trovati tramite l'uscita della Faster RCNN, si utilizza il criterio in cui la ROI bidimensionale deve contenere almeno il 75% la proiezione del canale fatta manualmente. Questa metrica va effettuata sulle varie proiezioni su cui si applicano le Faster RCNN.

Invece come metrica per le UNet si descrive che almeno il 50% del volume segmentato contenga almeno il 50% dell'oggetto d'interesse.

3.1.2. *Voxel level*

Per quanto riguarda invece la qualità della segmentazione, si utilizzano delle metriche che operano sulla caratteristica di ogni voxel.

Essi vengono divisi in tre classi: “veri positivi” ovvero i voxel classificati come canale sia dall'algoritmo che dalla segmentazione manuale; “falsi positivi” sono i voxel classificati come positivi dall'applicazione del metodo proposto ma non dalla segmentazione manuale, caratterizzano la sovrasedgmentazione; “falsi negativi” i voxel che la segmentazione finale ha considerato come background, ma la segmentazione manuale considera far parte della classe canale.

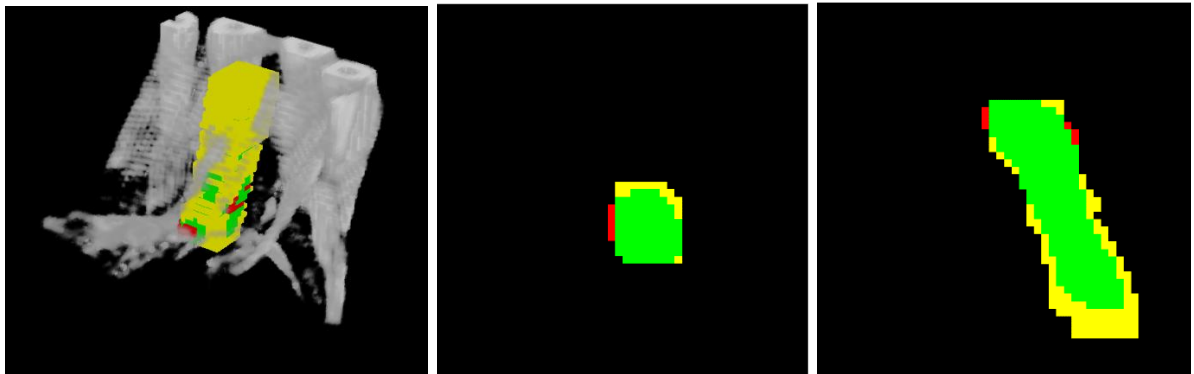


Figura 37 Esempio delle classi: in verde i Veri Positivi, in rosso i Falsi Positivi e in giallo i Falsi Negativi. A sinistra la visione 3D, al centro il piano trasversale e a destra il piano sagittale.

DSC

Descrive il rapporto tra sovrasedgmentazione e sottosedgmentazione. Indica in totale la qualità di segmentazione dell'algoritmo. Può essere espressa come:

$$DSC = \frac{2 * TP}{2 * TP + FP + FN}$$

Recall

Descrive come la segmentazione riesca in maniera efficace a classificare in maniera corretta la classe che si sta ricercando. Può essere espressa come:

$$recall = \frac{TP}{TP + FP}$$

Precision

Descrive il livello di oggetto correttamente classificato, rispetto all'oggetto affetto da sovrasegmentazione. Questa metrica appunto indica, se bassa, che la segmentazione è affetta da sovrasegmentazione. Può essere espressa come:

$$precision = \frac{TP}{TP + FN}$$

3.1.3. Morfometriche

In questa sezione, come metriche, si sono considerate le caratteristiche del canale e come vengano conservate dalla segmentazione automatica applicata.

Sono state considerate come caratteristiche: la lunghezza, il diametro medio e differenza tra l'angolo trovato dalla segmentazione automatica e quello della segmentazione manuale.

Lunghezza

La differenza tra la lunghezza del corpo segmentato manualmente ed automaticamente è stata presa come parametro per evidenziare una conservazione delle caratteristiche dell'algoritmo di ricerca automatica.

In questo caso si calcola la lunghezza dell'asse del canale calcolato come proiezione sull'asse Z e poi tramite l'angolo compreso tra l'asse stesso ed il sistema di riferimento.

Diametro medio

Anche il diametro è stato preso come indicatore della conservazione delle caratteristiche morfologiche. In questo caso si calcola la differenza del diametro medio tra segmentazione manuale ed automatica.

Inclinazione asse

Ultimo, ma più indicativo come indicatore è la differenza di inclinazione rispetto al sistema di riferimento del canale trovato tramite segmentazione automatica e quello ottenuto nella segmentazione manuale.

3.2. Risultati delle Faster RCNN

Per quanto riguarda le Faster RCNN applicate sulle proiezioni sono stati provati più metodi per l'allenamento. Si sono provati sia la struttura di estrazione delle caratteristiche come ResNet50, che la ResNet18. La ResNet50 non ha portato risultati (valori di DSC pari a 0,28). Non riuscendosi ad allenare considerava nella gran parte dei casi l'intera immagine o nessuna ROI per ridurre l'errore.

Mentre soddisfacenti sono i risultati portati dalla ResNet18. Sono state provate reti a diverse epoche di allenamento. Come prevedibile è stato riscontrato un incremento delle caratteristiche di detect in base ad un maggior numero di epoche per l'allenamento.

Sono stati provati allenamenti con differente numero di epoche ovvero 5, 15 e 50. Non si sono superate le 50 epoche in quanto sarebbe stato troppo impegnativo come tempo per l'allenamento. Come suggerito dalla teoria, si è facilmente dedotto che l'efficienza della rete è direttamente proporzionale al numero di epoche con cui la si allena.

Il cui più grande problema è la corretta classificazione delle soluzioni qualora ci fossero più ROI come risultato. Si sono provati più criteri per classificare correttamente nel caso ci fossero più ROI. Il primo considera solo la probabilità trovata in automatico dalla stessa rete (parametro "score"), che indica l'attendibilità che ha la ROI di contenere realmente il canale. Questo viene utilizzato sia sul piano frontale che sul piano laterale. Qualora dovessero esserci più ROI a parità di probabilità allora si predilige la ROI con un'area più grande (maggiore probabilità che contenga il canale).

Faster RCNN sul piano laterale				
	training+validation+ test	training	validation	test
N ROI sbagliate	0	0	0	0

N ROI non trovate	35	9	9	16
N ROI corrette	773	599	91	84
%ROI corrette	95,67	98,52	91,00	84,00

Soltanto per la proiezione frontale, avendo un numero elevato di classificazioni errate, si è adottato un altro criterio, dove si prendono le ROI con score dal valore maggiore alla metà del massimo e di queste se sono più di una si utilizza un criterio di vicinanza della ROI trovata rispetto al piano di simmetria della proiezione del cranio che è rappresentato nell'immagine. Anche in questo caso se le ROI estratte dovessero essere più di una, allora se ne considererà solo quella con distanza minore rispetto al piano di simmetria.

Faster RCNN sul piano frontale				
	training+validation+ test	training	validation	test
N ROI sbagliate	0	0	0	0
N ROI non trovate	17	12	3	2
N ROI corrette	791	596	97	98
%ROI corrette	97,9	98,03	97,00	98,00

Per aumentare le corrette segmentazioni, qualora dovessero essere estratte soltanto le ROI di una proiezione, sono state sfruttate delle reti dette di “ripescaggio” che agiscono sulle singole slice. Il dataset di queste reti è minore delle reti precedenti in quanto creato tramite uscita delle reti precedenti (si seleziona la slice più vicina al punto medio della ROI trovata), quindi se non dovessero essere estratte ROI nelle reti che sfruttano le proiezioni allora quelle immagini sono eliminate dal nuovo dataset.

Come criterio, se dovessero essere presenti più ROI estratte si utilizza solo quello dello score più alto.

	Faster RCNN ripescaggio sul piano frontale				Faster RCNN ripescaggio sul piano laterale			
	training+validation+ test	training	validation	test	training+validation+ test	training	validation	test
N ROI sbagliate	40	27	4	9	28	17	10	1
N ROI non trovate	20	9	7	4	16	3	4	9
N ROI corrette	748	572	89	87	764	588	86	90
%ROI corrette	92,57	94,08	89,00	87,00	94,55	96,71	86,00	90,00

Si può notare che le prestazioni delle reti di ripescaggio sono inferiori alle reti sulle proiezioni. Questo è dovuto ad una propagazione dell'errore tra le due reti in quanto fortemente connesse.

Combinando i risultati delle reti proposte, secondo il metodo descritto nel capitolo precedente (qualora non si trovino le ROI nella prima segmentazione si sfruttano le reti di “ripescaggio” come supporto) i risultati di detect delle ROI diventano:

	combinazione tra le reti principali e di ripescaggio sul piano frontale			
	training+validation+ test	training	validation	test
N ROI sbagliate	0	0	0	0
N ROI non trovate	5	3	2	0
N ROI corrette	803	605	98	100
%ROI corrette	99,38	99,51	98,00	100,00

	combinazione tra le reti principali e di ripescaggio sul piano laterale			
	training+validation+ test	training	validation	test
N ROI sbagliate	2	0	2	0
N ROI non trovate	9	4	1	4
N ROI corrette	797	604	97	96
%ROI corrette	98,64	99,34	97,00	96,00

I risultati della triangolazione, quindi della somma delle quattro reti portano sono:

Somma dei risultati delle 4 Faster RCNN				
	training+validation+test			
		training	validation	test
N ROI sbagliate	2	0	2	0
N ROI non trovate	14	7	3	4
N ROI corrette	792	601	95	96
%ROI corrette	98,02	98,85	95,00	96,00

È necessario che i valori sui singoli piani siano alti per essere utilizzati perché essendo molte reti neurali a cascata, si rischia in che l'errore si propaghi e diventi eccessivo.

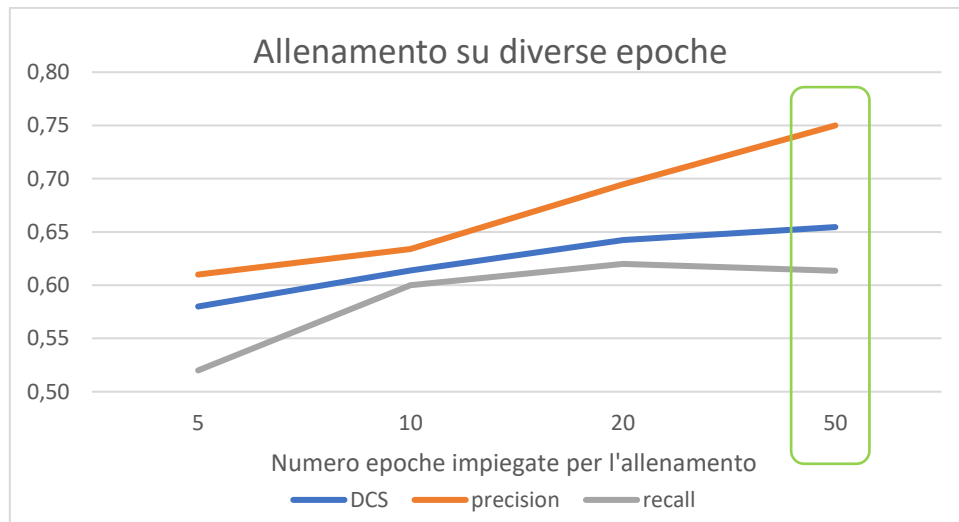
Lo scopo delle reti Faster RCNN è quello ottenere le coordinate per centrare il volume su cui sarà effettuato il crop dell'immagine TC.

È necessario che il volume risultante rispetti le metriche sopra descritte, ovvero che il volume croppato contenga totalmente il canale segmentato manualmente.

3.3. *Risultati delle UNet*

Dato che la rete neurale UNet sarà legata alle reti Faster RCNN, in quanto è necessario prima triangolare e tagliare al fine di ridurre le dimensioni dell'immagine, i risultati delle reti UNet porteranno un errore di fondo dovuto alle prime reti utilizzate. Quindi i risultati riportati in seguito sono affetti inevitabilmente da un errore dovuto alle reti messe in serie. L'errore dipende sostanzialmente da 2 fattori: il caso in cui non è possibile ricavare la ROI (ricerca non riuscita delle Faster RCNN) oppure dovuto a una ROI trovata errata, che non comprende il canale. In seguito a delle prove di validazione, si è appurato che delle 808 immagini 12 immagini non sono utilizzabili in quanto le Faster RCNN in uscita non trovano la ROI e rappresenteranno un errore collaterale dell'algoritmo.

L'allenamento delle reti UNet dipende fortemente dalle epoche con le quali si allenano. Sono stati provati 4 periodi di allenamento: 5 epoche, 10 epoche, 20 epoche e 50 epoche. Come prevedibile le reti tendono a migliorare la ricerca in maniera proporzionale alle epoche con cui si allenano. Quindi per l'allenamento si sono utilizzate 50 epoche.



Anche per le reti UNet è necessario fornire in input le segmentazioni manuali al fine dell'allenamento. Si sono provati più metodi di segmentazione e se n'è scelto quello che aveva maggiore efficienza per la corretta segmentazione del canale. Nel primo metodo, come classe dell'oggetto ricercato, è considerato il canale più i bordi. Nel secondo la classe che comprende solo il canale e nel terzo invece sono considerate 2 classi distinte, ovvero canale e bordo.

Nelle tabelle riportate in seguito sono presenti i risultati delle segmentazioni ottenute dalle tre metodi di allenamento, ma è stata necessaria l'applicazione della classificazione dei corpi segmentati per eliminare il rumore molto influente che potrebbe corrompere i risultati. Quindi i risultati sono riferiti alla segmentazione con l'eliminazione dei corpi più piccoli causati dal rumore. Si nota che il rumore è indipendente dal tipo di allenamento con cui si allena la rete, sia per le immagini fornite, sia per le epoche di allenamento.

I risultati sono riportati come media e deviazione standard sull'intero dataset e sulle divisioni dello stesso (training set, validation set e test set).

	ALLENAMENTO SUL CANALE			
	DCM	Precision	Recall	Corpi trovati(%)
media	0,65	0,68	0,67	76,73
devst	0,13	0,17	0,17	
med Tr	0,67	0,72	0,67	80,36
devst Tr	0,12	0,16	0,17	
med Val	0,61	0,56	0,70	67
devst Val	0,13	0,16	0,15	
med Test	0,60	0,56	0,68	66
devst Test	0,17	0,18	0,21	

	ALLENAMENTO SUL CANALE + BORDO (1 CLASSE)			
	DCM	Precision	Recall	Corpi trovati(%)
media	0,61	0,63	0,64	67,20
devst	0,12	0,17	0,16	
med Tr	0,62	0,67	0,63	68,81
devst Tr	0,11	0,15	0,16	
med Val	0,59	0,52	0,71	62
devst Val	0,13	0,15	0,14	
med Test	0,58	0,53	0,68	64
devst Test	0,16	0,16	0,19	

	ALLENAMENTO SUL CANALE E SUL BORDO (2 CLASSI)			
	DCM	Precision	Recall	Corpi trovati(%)
media	0,65	0,78	0,60	69,06
devst	0,15	0,18	0,18	
med Tr	0,66	0,83	0,59	66,01
devst Tr	0,14	0,15	0,18	
med Val	0,63	0,65	0,64	79
devst Val	0,13	0,18	0,15	
med Test	0,62	0,66	0,61	79
devst Test	0,17	0,20	0,19	

Tra i tre dataset da fornire nell'allenamento si è scelto l'allenamento su due classi. La scelta è ricaduta su questo allenamento perché dà il miglior compromesso tra le varie metriche. Si nota che le reti allenate su due classi sono molto specifiche, infatti presenta una precision alta, che indica poca sovrasedgmentazione, e una recall media, che indice una buona capacità di ricercare veri positivi.

Infine sul tipo di dataset da fornire in ingresso alla rete nella fase di allenamento, si è provata un'altra strategia per diversificare e provare il dataset che permette migliori risultati. E' stata provata la strategia di aggiustamento l'istogramma delle immagini del dataset per l'allenamento delle reti, in questo caso l'aggiustamento dell'istogramma della luminosità viene effettuato nella fase di preprocessing rispetto all'allenamento della rete stessa.

In seguito sono presenti i risultati sulla rete a due classi, la prima tabella è senza preprocessing, nella seconda invece si è effettuato un aggiustamento dell'istogramma delle luminosità nelle immagini fornite nella fase di allenamento.

	ALLENAMENTO SUL CANALE E SUL BORDO (2 CLASSI), DATASET SENZA PREPROCESSING			
	DCM	Precision	Recall	Corpi trovati(%)
media	0,65	0,78	0,60	69,06
devst	0,15	0,18	0,18	
med Tr	0,66	0,83	0,59	66,01
devst Tr	0,14	0,15	0,18	
med Val	0,63	0,65	0,64	79
devst Val	0,13	0,18	0,15	
med Test	0,62	0,66	0,61	79
devst Test	0,17	0,20	0,19	

	ALLENAMENTO SUL CANALE E SUL BORDO (2 CLASSI), DATASET CON PREPROCESSING			
	DCM	Precision	Recall	Corpi trovati(%)
media	0,65	0,77	0,61	70,79
devst	0,17	0,18	0,21	
med Tr	0,66	0,81	0,60	67,33
devst Tr	0,18	0,17	0,21	
med Val	0,66	0,67	0,67	83
devst Val	0,13	0,16	0,15	
med Test	0,64	0,67	0,64	81
devst Test	0,18	0,19	0,20	

Anche in questo caso il criterio per scegliere la strategia da applicare è stata dettata esclusivamente dal miglior compromesso delle metriche. Nel metodo si è scelto di utilizzare il dataset di immagini preprocessate da fornire per l'allenamento delle reti UNet.

3.4. Risultati dovuti al post processing

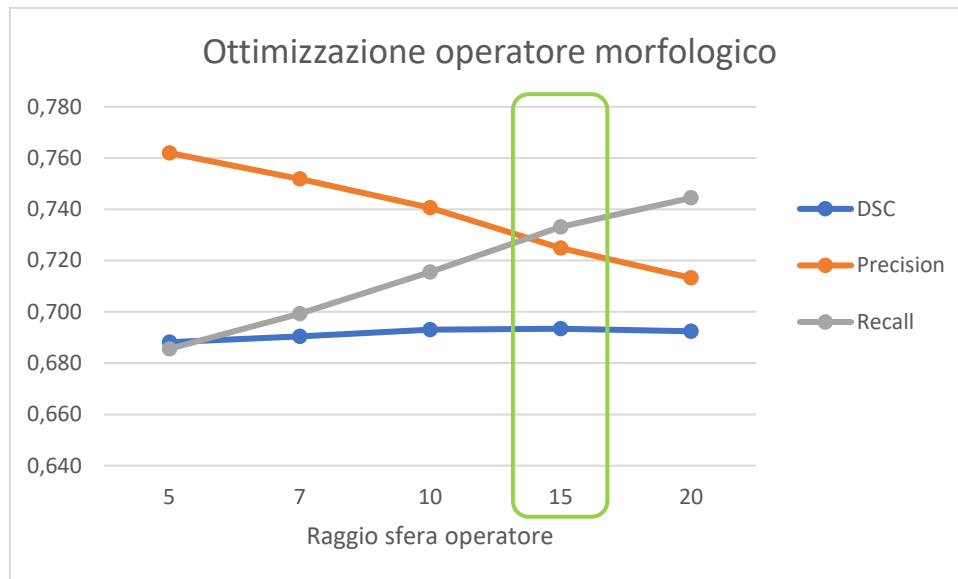
La combinazione delle 2 reti portano ad una leggera modifica delle caratteristiche rispetto all'allenamento, in quanto non sempre il canale è perfettamente centrato, e quando non vengono trovate le ROI la segmentazione non può procedere con la segmentazione semantica tridimensionale. Si sono tenute in considerazione anche l'eliminazione dei corpi più piccoli, in quanto la differenza dell'allenamento delle reti UNet porta ad una quantità diversa di corpi dovuti al rumore.

A questo punto si è voluto lavorare sulla rete al fine di migliorare le performance dell'algoritmo. Per incrementare il valore di recall si utilizza un operatore morfologico di chiusura, al fine di aumentare il volume segmentato e provare ad includere anche le zone dei falsi negativi.

Per quanto riguarda le caratteristiche dell'operatore utilizzato si è optato per uno di forma sferica per rispettare maggiormente la forma (altre forme davano dei contorni frastagliati e non continui). Per il valore del diametro si è ottenuta un'ottimizzazione tra 4 diversi valori, ovvero 5, 7, 10, 15 e 20 px.

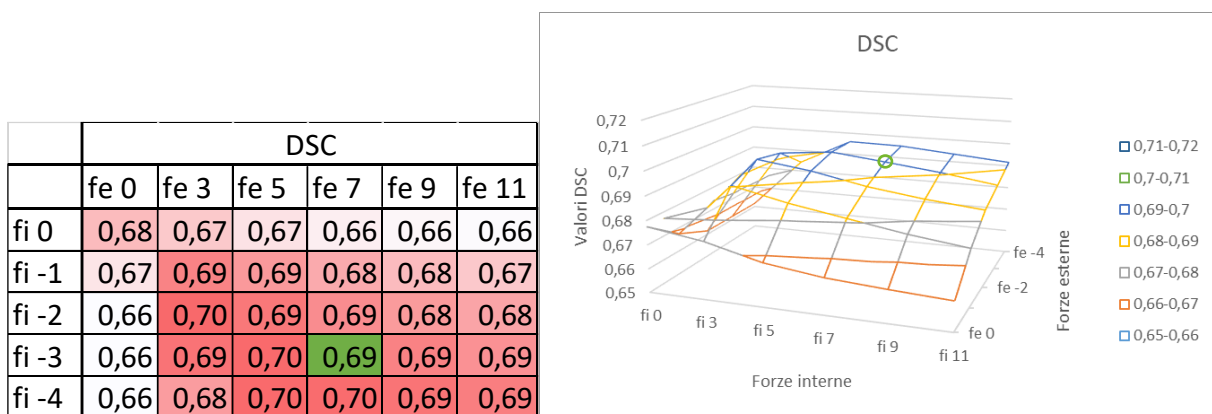
In seguito è rappresentato il grafico e la tabella dell'ottimizzazione. Evidenziato in verde è il valore successivamente scelto.

	Diametro sfera				
	5	7	10	15	20
DSC	0,688	0,690	0,693	0,693	0,692
Precision	0,762	0,752	0,741	0,725	0,713
Recall	0,686	0,699	0,715	0,733	0,744

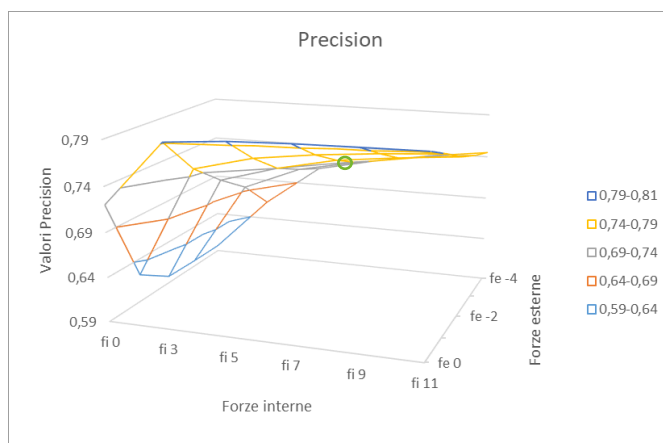


Il criterio di scelta è concorde agli altri criteri. Si è scelto il valore che portava il miglior compromesso delle metriche, ricordando che il fine di questo passaggio è quello di incrementare la metrica recall a discapito della precision. Dunque il valore scelto è: 15.

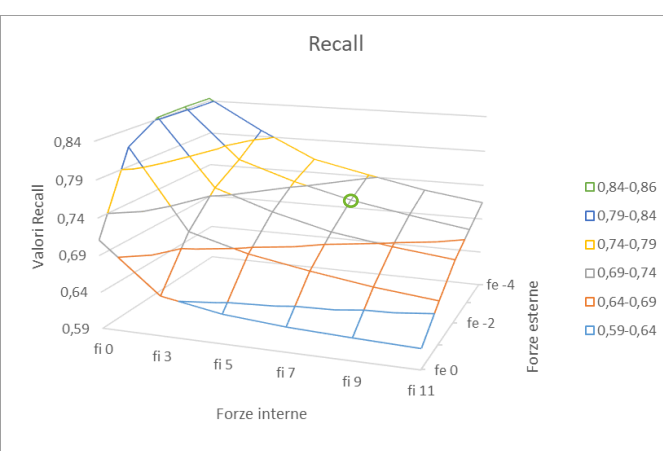
Quindi si prosegue con il metodo del level set per rilevare maggiormente i bordi. In questo caso si è utilizzato il metodo del Chan-Vese. Per impostare i valori che si devono fornire in input che andranno a gestire il bilanciamento tra forze interne ed esterne è stata fatta un'ottimizzazione dei parametri. In seguito è riportata la tabella che contiene il parametro di corretta ricerca dell'oggetto.



	Precision					
	fe 0	fe 3	fe 5	fe 7	fe 9	fe 11
fi 0	0,72	0,79	0,80	0,80	0,80	0,80
fi -1	0,62	0,75	0,77	0,77	0,78	0,79
fi -2	0,60	0,72	0,74	0,75	0,76	0,77
fi -3	0,60	0,70	0,72	0,74	0,75	0,76
fi -4	0,60	0,66	0,71	0,72	0,74	0,75



	Recall					
	fe 0	fe 3	fe 5	fe 7	fe 9	fe 11
fi 0	0,71	0,64	0,63	0,62	0,62	0,62
fi -1	0,82	0,71	0,68	0,67	0,66	0,65
fi -2	0,84	0,75	0,72	0,70	0,69	0,68
fi -3	0,84	0,77	0,74	0,72	0,71	0,70
fi -4	0,84	0,80	0,76	0,74	0,73	0,71



Sono stati scelti i valori evidenziati in verde, in quanto rappresentano il miglior compromesso. Questi sono un valore pari a -3 per le forze interne e 7 per le forze esterne. In questo modo la componente di espansione è ben bilanciata sulla componente di contrazione, il risultato sarà comunque una componente contrattiva, il cui scopo è quello di incrementare la metrica precision cercando di contenere le il decremento sulla metrica recall.

Ultima parte del postprocessing è l'aggiustamento dell'oggetto trovato seguendo un'analisi slice by slice. In questo caso si imposta un diametro minimo dell'oggetto contenuto in una slice, se questa soglia non viene superata allora si elimina l'oggetto contenuto in quella slice in quanto considerato errato.

In seguito sono riportati i risultati relativi quest'ultimo processo di aggiustamento e quindi saranno anche i risultati finali.

Sono riportati i risultati medi con le deviazioni standard sia sull'intero dataset che su ogni divisione dello stesso.

	Postprocessing con valori ottimizzati			
	DCM	Precision	Recall	Corpi trovati
media	0,69	0,74	0,70	81,56
devst	0,17	0,20	0,22	
med Tr	0,69	0,77	0,68	81,85
devst Tr	0,17	0,19	0,22	
med Val	0,69	0,63	0,79	84
devst Val	0,14	0,17	0,15	
med Test	0,67	0,63	0,75	79
devst Test	0,18	0,20	0,22	

3.5. *Le caratteristiche morfometriche*

Per vedere se le caratteristiche morfometriche sono state conservate. Sono stati utilizzati 3 indicatori descritti in precedenza, ovvero: differenza tra le lunghezze, differenza tra i diametri medi e differenza dell'inclinazione dell'asse.

Per quanto riguarda la conservazione del dimensione del diametro medio e della lunghezza i risultati sono:

	differenza in [mm]	Differenza relativa in %
Lunghezza del canale	$1,41 \pm 0,91$ mm	$6,52 \pm 4,23$ %
Diametri medi	$1,44 \pm 0,95$ mm	$9,6 \pm 5,34$ %

Si nota che le differenze in mm sono simili, questo errore può dipendere dalla risoluzione spaziale bassa. Infatti tutte le immagini sono state sotto campionate per ridurre il peso computazionale totale dell'algoritmo di ricerca e ridurre il peso dell'allenamento in quanto hardware a disposizione non permette di allenare immagini più grandi. Questo processo può aver causato la perdita di informazione o una leggera corruzione dell'immagine che porta ad una differenza fissa tra le misure sopracitate di circa 1,5mm.

La conservazione dell'inclinazione dell'asse rispetto il sistema di riferimento invece:

	Differenza angolo dente/canale medio
Rispetto l'asse X	$0,091 \pm 2,41^\circ$
Rispetto l'asse Y	$0,14 \pm 5,85^\circ$

I risultati dal punto di vista morfometrico indicano che la segmentazione automatica tende a conservare i dati con un errore del 9% circa. Mentre la differenza di inclinazione è pressoché paragonabile a 0° .

4. Conclusioni

L'obiettivo di questa tesi magistrale è stato quello di creare un metodo per la segmentazione del dotto naso-palatino ricostruendo fedelmente il canale al fine di conservare le caratteristiche morfologiche e morfometriche utili in campo odontoiatrico in implantologia dentale utilizzando delle reti neurali basate sull'apprendimento profondo. Prendendo spunto da articoli più recenti sul ciò che riguarda il processing di immagini mediche odontoiatriche, si sono sviluppate delle reti neurali convoluzionali per trovare il dotto naso-palatino.

Le Faster RCNN sono tra queste reti, già ampiamente utilizzate per il detect di denti nelle immagini di panoramica dentale, nella tesi sono state utilizzate per cercare la ROI bidimensionale tramite la proiezione delle immagini tridimensionali lungo il piano frontale e laterale. Sono state utilizzate anche per la ricerca del canale all'interno delle singole slice, nel caso che il primo metodo non dovesse essere sufficiente. Dall'uscita di queste reti sono presenti le coordinate della ROI che contiene il canale. È stata sfruttata la combinazione delle varie uscite per andare ad individuare una ROI tridimensionale contenente il canale, in modo da andare a ridurre il volume e poterlo più facilmente andare a processare. Infatti la segmentazione tridimensionale di un oggetto richiede un grande impiego di tempo e di memoria. Riducendo il volume di ricerca si riduce sensibilmente anche il carico computazionale e di tempo necessari per la corretta segmentazione.

Le UNet sono le altre reti sviluppate al fine della segmentazione del canale naso palatino. Queste reti sono ampiamente utilizzate in campo medico ed odontoiatrico per quel che riguarda la segmentazione di oggetti. È vastamente utilizzata soprattutto il caso di segmentazioni tridimensionali, in quanto la particolare architettura a forma di U permette una "facile" segmentazione anche nel caso delle immagini tridimensionali. Nell'elaborato si è utilizzato il volume ridotto (regione d'interesse) contenente il canale e tramite questo tipo di rete segmentare.

La trattazione di entrambe le reti è stata analizzata dalla fase di costruzione e training fino alla fase di utilizzo e postprocessing. Il risultato è stata la trattazione di un metodo che consiste nell'utilizzare delle tecniche di detect e segmentazione già ampiamente utilizzate nelle immagini mediche, in campo odontoiatrico per ottenere una segmentazione efficace e precisa

del canale naso-palatino. Il postprocessing proposto è stato ottimizzato ed utilizzato proprio per aumentare la precisione della segmentazione del canale. I risultati sono pari ad 82% di canali correttamente trovati e ricostruiti con un accuratezza espressa con DCS pari a 0,7.

Importante è la conservazione delle caratteristiche morfometriche, in quanto queste sono utilizzabili efficacemente in ambito odontoiatrico per poter ridurre il rischio di fallimento in implantologia dentale. Già attualmente il canale naso-palatino viene utilizzato in ambito di implantologia dentale come strumento per una corretta applicazione di impianti dentali in pazienti che presentano un parziale o totale ritiro di tessuto osseo presso il canale incisivo. Attualmente le caratteristiche di inclinazione e direzione del canale sono trovate manualmente inserendo uno strumento simile ad un'asta per facilitare gli odontoiatri la corretta angolazione del dente. È stato dimostrato, attraverso le metriche di conservazione delle caratteristiche morfometriche, che le stesse vengono trovate fedelmente nella segmentazione automatica. L'errore è di circa 1,4mm sia per quanto riguarda la lunghezza del canale che la dimensione del diametro medio, mentre per l'inclinazione

Il metodo ha ampi margini di miglioramento. Garantirebbero dei risultati migliori ampliare la profondità della rete UNet a 3 livelli (renderla troppo profonda potrebbe far perdere l'efficacia) ed aumentare le epoche di allenamento; anche per le Faster RCNN sarebbe ideale aumentare l'allenamento della rete almeno a 500 epoche. Altri margini di miglioramento sono una revisione fatta da un esperto alla segmentazione manuale e un ampliamento del dataset.

Una delle possibili applicazioni di questo metodo è l'affiancamento ad un metodo di auto segmentazione dei denti per poter trovare una relazione scientificamente dimostrata tra l'asse del canale e l'asse degli incisivi superiori. Questa aprirebbe una nuova strada per quel che riguardano i metodi e protocolli in odontoiatria per l'implantologia dentale. Un'altra applicazione è la ricerca dell'angolazione del canale rispetto al canale incisivo per poter strutturare un intervento e simularne la corretta applicazione e poter prevedere eventuali criticità dovute all'impianto.

Parallelamente alla tesi è stato sviluppato un algoritmo che oltre al metodo proposto implementa una segmentazione tridimensionale automatica dei denti al fine di ricavare una relazione tra il canale e i vari denti dell'arcata superiore. I risultati fin ora sono grezzi e spesso corrotti dall'elevato errore, ma in maniera minima si può dedurre che è presente un

parallelismo tra il canale e gli incisivi superiori (circa $1/3^\circ$). Anche la comunità scientifica odontoiatrica si sta muovendo in questo campo per migliorare l'efficienza delle protesi dentali. Si può notare effettivamente che sono stati pubblicati vari articoli sulla conformazione ed utilità del canale e sul suo utilizzo per impianti.

Dunque questo trattato vuole portare in evidenza la possibilità di un'accurata segmentazione del canale naso-palatino la possibilità di sfruttare un algoritmo per estrarre da immagini TC dati morfometrici in modo automatico su una grande popolazione di pazienti. È evidente che il metodo sia stabile e che i risultati siano simili, anche se migliorabili, rispetto le segmentazioni manuali.

Bibliografia

1. **Radlanski RJ, Emmerich S, Renz H.** Prenatal morphogenesis of the human incisive canal. *H. Anat Embryol.* 2004, 208:265-271. 10.1007/s00429-004-0389-y. 2004.
2. **Miwa Y, Asaumi R, Kawai T, Maeda Y, Sato I.** Morphological observation and CBCT of the bony canal structure of the groove and the location of blood vessels and nerves in the palatine of elderly human cadavers. 40:199-206. 10.1007/s00276-017-1952-6. s.l. : Surg Radiol Anat. 2018, 2018.
3. **Fukuda M, Matsunaga S, Odaka K, Oomine Y, Kasahara M, Yamamoto M, Abe S.** Three-dimensional analysis of incisive canals in human dentulous and edentulous maxillary bones. 1:12. 10.1186/s40729-015-0012-4. s.l. : Int J Implant Dent. 2015, 2015.
4. **Sasha Lake, Joe Iwanaga , Shogo Kikuta , Rod J. Oskouian , Marios Loukas , R. Shane Tubbs.** The Incisive Canal: A Comprehensive Review. *The Incisive Canal: A Comprehensive Review.* s.l. : Cerus, 30 luglio 2018.
5. **Khojastepour L, Haghnegahdar A, Keshkhtar M.** Morphology and dimensions of nasopalatine canal: a radiographic analysis using cone beam computed tomography. *J Dent Shiraz Univ Med Sci.* 2017, 18:244-250.
6. **Rosenfeld AL, Mecall RA.** The use of interactive computed tomography to predict the esthetic and functional demands of implants-supported prostheses. *Compendium* 1996, 17; 12: 1125-1144.
7. **Carlsson GE, Bergman B, Hedegard B.** Changes in contour of the maxillary alveolar process under immediate dentures. . *Acta Odontologica Scandinavica* 1967; 43: 44-75.
8. **Sonick M, Abrahams J, Faiella R.** A comparison of the accuracy of periapical, panoramic, and computerized tomographic radiographs in locating the mandibular canal. *JOMI* 1997. 9; 4: 455-460.
9. **Gautam Chitnis, Vidhi Bhanushali, Aayush Ranade, Tejasvini Khadase, Vaishnavi Pelagade, Jitendra Chavan.** A Review of Machine Learning Methodologies for Dental Disease Detection. DOI: 10.1109/INDISCON50162.2020.00025. s.l. : 2020 IEEE India Council International Subsections Conference (INDISCON), 22 gennaio 2020.
10. **Dmitry V. Tuzoff, Lyudmila N. Tuzova, Michael M. Bornstein, Alexey S. Krasnov, Max A. Kharchenko, Sergey I. Nikolenko, Mikhail M. Sveshnikov and Georgiy B. Bednenko.** Tooth detection and numbering in panoramic radiographs using convolutional neural networks. s.l. : British Institute of Radiology, 31 gennaio 2019.
11. **Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun.** Faster R-CNN: Towards Real-Time Object. *arXiv:1506.01497v3.* s.l. : arXiv, 06 gennaio 2016.
12. **Andaç Imak, Adalet Çelebi, Muammer Türkoglu, Abdulkadir ,Sengür.** Dental Material Detection based on Faster Regional Convolutional Neural Networks and Shape Features. <https://doi.org/10.1007/s11063-021-10721-5>. s.l., Turchia : Springer, 18 dicembre 2021.
13. **Anuradha Laishram, Khelchandra Thongam.** Detection and Classification of Dental Pathologies using Faster-RCNN in Orthopantomogram Radiography Image. s.l. : 2020 7th International Conference on Signal Processing and Integrated Networks (SPIN), 2020.

14. **Abraham, Shweta Shirsat and Siby.** Tooth Detection from Panoramic Radiographs Using Deep Learning. [aut. libro] Jerry Chun-Wei Lin, Raj Bhatnagar, Sonali Agarwal, P. Krishna Reddy Satish Narayana Srirama. *Big Data Analytics*. s.l. : Springer, 2021.
15. **Rikiya Yamashita, Mizuho Nishio, Richard Kinh Gian Do & Kaori Togashi.** Convolutional neural networks: an overview and application in radiology. <https://doi.org/10.1007/s13244-018-0639-9>. s.l. : Springer, 22 giugno 2018.
16. **Thorbjørn Lourcing Koch, Mathias Perslev, Christian Igel e Sami Sebastian Brandt.** ACCURATE SEGMENTATION OF DENTAL PANORAMIC RADIOGRAPHS WITH U-NETS. Venezia, Italia : 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019), 8-11 aprile 2019.
17. **S. Sivagami, P. Chitra, G. Sri Ram Kailash, and S.R. Muralidharan.** UNet Architecture Based Dental Panoramic Image Segmentation. s.l. : IEEE, 05 ottobre 2020.
18. **Ozgun C. i. cek, Ahmed Abdulkadir, Soeren S. Lienkamp, Thomas Brox, and Olaf Ronneberger.** 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. DOI: 10.1007/978-3-319-46723-8 49. s.l. : Springer International Publishing AG 2016, 2016.
19. **Thorsten Falk, Dominic Mai, Robert Bensch, Özgün Çiçek, Ahmed Abdulkadir, Yassine Marrakchi, Anton Böhm, Jan Deubner, Zoe Jäckel, Katharina Seiwald, Alexander Dovzhenko, Olaf Tietz, Cristina Dal Bosco, Sean Walsh, Deniz Saltukoglu, Tuan Leng Tay.** U-Net: deep learning for cell counting, detection, and morphometry. *Published: 17 December 2018*. s.l. : Nature Methods, 2019. Vol. volume 16, pages67–70.
20. **Olaf Ronneberger, Philipp Fischer, and Thomas Brox.** U-Net: Convolutional Networks for Biomedical Image Segmentation. <https://lmb.informatik.uni-freiburg.de/people/ronneber/u-net/>. s.l. : Uni. Freiburg, Dept. of Computer Science, 2015.
21. **Hongkai Yang, Kaichao Liang, Li Zhang, Kejun Kang, Yuxiang Xing.** Improve 3D cone-beam CT reconstruction by slice-wise deep learning. s.l. : IEEE 2018, 2018.
22. **Jun Ma, Xiaoping Yang.** Automatic dental root CBCT image segmentation based on CNN and level set method. *Proceedings Volume 10949, Medical Imaging 2019: Image Processing; 109492N (2019)* <https://doi.org/10.1117/12.2512359>.
23. **Gloria Hyunjung Kwak, Eun-Jung Kwak, Jae Min Song, Hae Ryoung Park, Yun-Hoa Jung, Bong-Hae Cho, Pan Hui & Jae Joon Hwang.** Automatic mandibular canal detection using a deep convolutional neural network. *SCIENTIFIC REPORTS (2020) 10:5711*.
24. **Evan Shelhamer, Jonathan Long and Trevor Darrell.** Fully Convolutional Networks for Semantic Segmentation. DOI 10.1109/TPAMI.2016.2572683. s.l. : 2016 IEEE, 2016.
25. **Stanley Osher, Ronald P. Fedkiw.** Level Set Methods: An Overview and Some Recent Results. 2000.
26. **Getreuer, Pascal.** Chan–Vese Segmentation. <https://doi.org/10.5201/ipol.2012.g-cv>. 2012. Vol. Published in Image Processing On Line on 2012–08–08.
27. **Ciresan, Dan, et al.** "Flexible, High Performance Convolutional Neural Networks for Image Classification". *Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence*. 17-11-2011. Vol. 2.

5. Ringraziamenti

A conclusione di questo elaborato, desidero menzionare tutte le persone, senza le quali questo lavoro non sarebbe potuto esistere.

Prima di tutto vorrei ringraziare il professor Filippo Molinari, relatore di questa tesi, per avermi dato la possibilità di credere in questo suo progetto. Ringrazio anche il mio co-relatore il professor Massimo Salvi per avermi accompagnato e sostenuto in questi mesi. È stato un punto di riferimento, mi ha guidato e consigliato per tutta l'arco della tesi invitandomi gentilmente a prendere sempre la strada più coerente.

Vorrei anche ringraziare il dottor Davide Romano dello Studio Dentistico Dr. Romano Biagio per la collaborazione, riguardo l'acquisizione del dataset e consigli pratici sulla corretta segmentazione del canale, senza il suo aiuto sarebbe stato impossibile svolgere questa tesi magistrale.

Un particolare e sentito ringraziamento va ai miei genitori, Antonio Vitale e Giuseppa Ruggiero, per avermi sempre sostenuto durante tutto il mio percorso. Sono stati un'ancora anche quando la marea portava a largo e la spinta quando faticavo a muovermi, grazie. Un enorme ringraziamento va anche a mio fratello, Marco Vitale, perché mi è sempre stato vicino ed è stato per me d'ispirazione con la sua forza ed energia. Lo ringrazio anche perché nella parte del lavoro più duro svolto in questo elaborato, ovvero le segmentazioni manuali, mi ha dato sostegno a non arrendermi e mi ha aiutato a svolgerne alcune. Un sentitissimo ringraziamento è dedicato anche alla dott. Federica Antezza, che come un faro nella notte ha illuminato la mia strada.

Un ringraziamento vorrei darlo alla mia famiglia, i nonni, gli zii e i cugini sono stati sempre vicini e mi hanno dato la motivazione per raggiungere questo obiettivo, grazie.

Un super ringraziamento va anche ai miei amici, dai miei colleghi biomedici, che sono stati un supporto unico nei momenti difficili e uno svago nei momenti di gioia, ai miei coinquilini ed ex coinquilini che sono stati a tutti gli effetti la mia famiglia, mi hanno donato affetto e sostegno anche nei momenti difficili; e gli amici tutti che mi son sempre stati vicini.