



POLITECNICO DI TORINO

DIPARTIMENTO DI INGEGNERIA MECCANICA E AEROSPAZIALE

CORSO DI STUDIO MAGISTRALE IN
STRUMENTAZIONE BIOMEDICA

**Development of an Automatic
Algorithm for Needle Segmentation
for Multi-Modal Optical Imaging
Modalities**

Relatore:

Prof.ssa Kristen Meiburger

Co-relatore:

Ing. Francesco Marzola

Tesi di Laurea di:

Federica Lemma

Matricola 288228

Anno accademico:

2022/2023

Indice

Abstract	1
1 Introduzione	3
1.1 Tecniche di Imaging	3
1.1.1 Optical Coherence Tomography	4
1.1.2 Imaging Fotoacustico	6
1.2 Stato dell'arte delle tecniche di tracking	8
1.2.1 Approcci tradizionali	8
1.2.2 Approcci di Deep Learning	12
2 Materiali	13
2.1 Dataset OCT	13
2.1.1 Acquisizione dei dati	13
2.1.2 Ricostruzione volumetrica	16
2.1.3 Pre Processing dei dati	18
2.1.4 Costruzione del Dataset 2D	20
2.2 Dataset PAT	21
2.2.1 Acquisizione dei dati	21
2.2.2 Pre Processing dei dati	23
2.2.3 Costruzione del Dataset 2D	26
3 Metodi	27
3.1 Deep Learning Neural Network	27
3.1.1 ResNet-34	31
3.1.2 OCT model	34
3.1.3 PA model	35
3.2 Inference	36
3.2.1 Pipeline di riduzione dei tempi computazionali	36
3.3 Post Processing	44
3.4 Metriche di valutazione e loss	47
4 Risultati	50
5 Conclusione	60
5.1 Limiti e Sviluppi Futuri	61
Bibliografia	63
Ringraziamenti	64

Elenco delle figure

1.1	Schema OCT [7]	4
1.2	Piani di scansione OCT	5
1.3	Principio di funzionamento delle immagini fotoacustiche [10]	6
1.4	Set-Up di validazione	11
2.1	Sistema di acquisizione OCT/PAT	14
2.2	Fabricated phantom. Top view (a) and lateral view of a section of the phantom (b). OCT B-scan of the phantom (c).	15
2.3	Phantom setup for OCT acquisition	16
2.4	Algoritmo di ricostruzione OCT basato sul metodo [15]	16
2.5	Visualizzazione volumetrica in Slicer3D del VolumeOCT_Needle2_pos1	18
2.6	Pipeline di Pre-Processing OCT	18
2.7		19
2.8	Visualizzazione volumetrica in Slicer3D volume NeedleInBeef_10s. In rosso è evidenziata la sezione utilizzata nell'addestramento.	22
2.9	MIP nelle tre direzioni di acquisizione PAT di un ago inserito nel petto di un pollo	22
2.10	MIP nelle tre direzioni di acquisizione PAT del palmo di una mano	23
2.11	MIPs in the three directions of a skin-needle merged volume.	23
2.12	Pipeline di pre processing volumi PAT	23
2.13	Istogramma di riferimento medio calcolato su due volumi OCT	24
2.14		25
3.1	Architettura rete CNN che mostra i layer di convoluzione + ReLu (attivazione)+ layer di Pooling + Layer Fully Conncted	27
3.2	Layer Convolutionale:	28
3.3	Esempi di funzioni di attivazione	28
3.4	Esempio: a sinistr l'input è un volume 4×4. A destra l'effetto del max pooling con stride pari a 1 e pari a 2. Questa applicazione riduce le dimensioni spaziali dell'input.	29
3.5	FCN Dopo un upsampling per recuperare la risoluzione completa dell'immagine all'ultimo livello, ogni pixel viene classificato utilizzando la softmax per generare le segmentazioni [21].	30
3.6	Architettura di una rete U- Net	30
3.7	Esempio del blocco di connessione residua	31
3.8	Architettura ResNet-34 con connessioni residue (la linea tratteggiata indica l'aumentare delle dimensioni con il padding). Il Downsampling è effettuato con layer convoluzionali con passo pari a 2.	32

3.9	A sinistra CH0: pixel ago a 1 e sfondo 0. A destra CH1: pixel ago a 0 e sfondo a 1	33
3.10	Pipeline di Inference sui volumi OCT	38
3.11	A sinistra: istogramma calcolato su immagine contenente ago. A destra: istogramma dell'immagine che non contiene l'ago . .	38
3.12	A sinistra: immagine estratta dal volume OCT con ago. A destra: immagine estratta dallo stesso volume solo sfondo . . .	39
3.13	Andamento della varianza calcolato per ogni frame del volume OCT	39
3.14	Pipeline di inference sui volumi PAT	41
3.15	A sinistra: immagine estratta dal volume PAT dopo HM. A destra: immagine binarizzata	41
3.16	Ricerca dei contorni all'interno dell'immagine	42
3.17	Contorno dell'ago trovato in base alla numerosità dei pixel . .	42
3.18	A destra: Slice estratta dal volume OCT N6P4 senza applicazione di post processing. A sinistra: Slice estratta dal volume OCT a cui sono state applicate le tecniche di post processing.	44
3.19	A sinistra: Slice estratta senza applicazione di Post Processing. A destra: Slice con applicazione di tecniche di Post Processing	45
4.1	Prestazioni della rete OCT. A destra: valore medio della loss calcolata per ciascuna epoca. A sinistra:valore medio del Dice calcolato per ogni epoca sul set di validazione.	51
4.2	Prestazioni della rete PAT	53
4.3	TEST OCT. Da sinistra verso destra: Immagine grayscale estratta dal volume OCT, Segmentazione Manuale ottenuta con Slicer3D, Segmentazione Automatica ottenuta con la rete neurale.	56
4.4	TEST PAT Da sinistra verso destra: Immagine grayscale estratta dal volume PAT, Segmentazione Manuale ottenuta con Slicer3D, Segmentazione Automatica ottenuta con la rete neurale.	56
4.5	MIP calcolata su volume PA needle+vessels e relativo istogramma	57
4.6	Slice estratta da volume OCT utilizzata per il training e relativo istogramma	58
4.7	Immagine estratta dal volume PA 'needle' DOPO Histogram Matching	58
4.8	Da sinistra verso destra: Immagine MIP calcolata dal volume PAT needle+vessels, Segmentazione Manuale ottenuta con Slicer3D, Segmentazione Automatica ottenuta con la rete neurale	59

Elenco delle tabelle

2.1	Parametri sistema OCT	13
2.2	Parametri sistema PAT	21
2.3	Volumi del Dataset PAT	26
3.1	Tempi di predizione Volumi OCT	40
3.2	Tempi di predizione Volumi PAT	43
3.3	Risultato dell'applicazione del post processing sui volumi OCT	45
3.4	Risultato dell'applicazione del post processing sui volumi PAT	46
4.1	Prestazioni calcolate sul training set del dataset OCT	52
4.2	Prestazioni calcolate sul validation set del dataset OCT	52
4.3	Prestazioni calcolate sul test set del dataset OCT	52
4.4	Performance calcolate sul training set del dataset PAT	53
4.5	Performance calcolate sul validation set del dataset PAT	53
4.6	Performance calcolate sul test set del dataset PA	54
4.7	Valore medio e deviazione standard delle performance calcolate sul Training set di ciascun dataset	55
4.8	Valore medio e deviazione standard delle performance calcolate sul Validation set di ciascun dataset	55
4.9	Valore medio e deviazione standard delle performance calcolate sui test set di ciascun dataset	55
4.10	Valore medio e deviazione standard delle performance calcolate sulle MIP	57

Abstract

Le tecnologie di imaging medico svolgono un ruolo fondamentale nelle procedure di intervento che richiedono una guida visiva per l'inserimento di sonde o aghi, come la biopsia, la chirurgia e alcune terapie. Una volta che l'ago è inserito all'interno dei tessuti, approcci di tracking basati su imaging possono essere usati per guidare l'ago nel sito di interesse [1].

Il tracking consiste nella localizzazione di uno strumento in real time in un dataset di lavoro spesso tridimensionale. Esso risulta utile nel caso di scarsa visibilità dello stesso, localizzazioni o anatomie complesse del sito di interesse o che richiedono un accesso altamente preciso, e per ridurre i tempi delle procedure, migliorando contemporaneamente la sicurezza. [2]

Il lavoro di tesi si pone come obiettivo quello di realizzare un algoritmo di tracking al fine di riconoscere e localizzare l'ago in immagini diagnostiche. Le principali tecniche di imaging utilizzate in questo lavoro di tesi sono la Tomografia a Coerenza Ottica e l'imaging fotoacustico. Su di esse si è lavorato per l'implementazione di un algoritmo di Deep Learning, che possa essere poi esteso su ulteriori modalità.

Il dataset utilizzato per lo scopo si compone di 9 volumi OCT di dimensioni 512x896x896 acquisiti su fantoccio e 13 volumi PAT di dimensioni 950x288x288 acquisiti in supporto Gel, in fantoccio e in carne. In entrambe le tipologie di dataset, i volumi si distinguono per la diversa posizione dell'inserzione dell'ago. I dati sono stati poi opportunamente divisi in training set, validation set e test set.

La pipeline di pre-processing si compone di diversi step. In primis, da ciascun volume sono state estratte le singole immagini bidimensionali, sulle quali è stata effettuata la normalizzazione tra il 1° e il 99° percentile, al fine di riportare in modo uniforme i valori dei pixel nel range 0-255, con conseguente aumento del contrasto. Segue il ridimensionamento delle stesse e l'applicazione di un filtro passa basso gaussiano per la rimozione del rumore.

Si è poi passati al training della rete ResNet-34 per segmentazione automatica utilizzando il toolkit Monai in Google Colab e sfruttando l'algoritmo di accelerazione della GPU, con le singole tipologie di volume. Al training set è stata applicata la tecnica di Data Augmentation, al fine di migliorare le performance della rete. La funzione di loss utilizzata è la DiceLoss mentre la metrica di valutazione è il Mean Dice.

Al fine di ridurre i tempi computazionali, in fase di inference è stata applicata una strategia che permettesse di selezionare ed effettuare la predizione sulle sole immagini del volume in cui fosse presente l'ago. Nel caso dei volumi OCT

tale tecnica si basa sul calcolo della varianza dell'immagine, mentre nel caso dei volumi PA, sulla ricerca del contorno dell'ago.

La metrica utilizzata nella validazione, il coefficiente Dice, raggiunge valori superiori all'80% in più del 60% dei campioni, con la presenza di qualche outlier dovuto anche al fatto che la segmentazione di riferimento non fosse estremamente accurata. L'implementazione della strategia in fase di predizione ha comportato una significativa riduzione dei tempi computazionali. Prima dell'implementazione della strategia, infatti, la predizione di ciascun volume richiedeva centinaia di secondi, ora ridotti a poche decine di secondi. È stata applicata la tecnica di post-processing di rimozione di oggetti piccoli e isolati, aumentando la continuità e migliorando i risultati.

Infine, il lavoro si concentra anche sull'integrazione di informazioni multi-modali, ovvero l'integrazione nello stesso modello di immagini OCT e immagini PAT, al fine di elaborare un algoritmo generale con la prospettiva di poter espandere l'applicazione dello stesso su una pluralità di modalità di imaging. Lavori futuri potrebbero inoltre includere un ulteriore miglioramento in termini di riduzione dei tempi computazionali per rendere la procedura real-time.

Capitolo 1

Introduzione

1.1 Tecniche di Imaging

Lo sviluppo di innovative tecniche di imaging nel corso degli ultimi decenni ha reso possibile la visualizzazione intra-operatoria come parte integrante dello strumento chirurgico. In particolare, ricerche sul tema hanno messo in luce l'importanza e la necessità di stimare e tracciare la posizione e la traiettoria dell'ago in procedure quali la biopsia nel tessuto esaminato.

Al fine di evitare il danneggiamento del tessuto durante le procedure percutanee, si necessita di una precisa detection dello strumento. Ruolo fondamentale gioca quindi la visibilità di quest'ultimo, strettamente collegata alla modalità di imaging utilizzata come guida. In questo lavoro vengono esaminate due modalità di imaging in particolare: la tomografia a coerenza ottica e l'imaging fotoacustico. L'OCT è una tecnologia di imaging basata principalmente su laser utilizzata per diagnosticare in primo luogo malattie della retina, in quanto garantisce un controllo preciso dello strumento chirurgico in profondità.

In particolare, l'OCT è utilizzato come supporto visivo durante la procedura chirurgica di cheratoplastica lamellare anteriore profonda ex vivo (DALK). La DALK viene eseguita principalmente con la tecnica definita 'Big Bubble', in cui il chirurgo inserisce un ago all'interno dello stroma corneale e inietta una bolla d'aria per separare la membrana di Descemet dallo stroma.

Tale procedura è estremamente complessa, per tale motivo, nasce l'esigenza di monitorare l'ago all'interno di volumi OCT al fine di assistere il chirurgo, apportando benefici in termini di durata della procedura e precisione nel raggiungimento del target. Diversi studi hanno messo in luce l'impatto dell'imaging OCT durante le operazioni chirurgiche o per esaminare i pazienti sottoposti ad anestesia, in quanto questo permette di osservare l'interazione dinamica tra tessuto e strumento [3].

L'imaging fotoacustico, invece, risulta promettente in questo campo; difatti, supera i limiti dell'imaging ad ultrasuoni in strutture caratterizzate da scattering e attenuazione sonora (come le ossa), che complicano la guida nelle procedure chirurgiche invasive in quanto riducono la qualità delle immagini. L'imaging fotoacustico usa luce laser pulsata per incitare l'espansione termica, la quale genera segnali acustici che possono essere rilevati con trasduttori. L'intensità relativa di questi segnali (cioè il contrasto dell'immagine) è determinata dall'assorbimento ottico dipendente dalla lunghezza d'onda della

regione dell'immagine. Ad esempio, il sangue e il metallo tendono ad avere un assorbimento ottico maggiore rispetto al tessuto circostante, il che li rende ottimi bersagli per l'imaging fotoacustico. Di conseguenza, l'imaging fotoacustico ha dimostrato il potenziale per rilevare impianti metallici, monitorare il flusso dei vasi, mappare la struttura dei vasi, determinare la presenza di aterosclerosi e guidare gli interventi chirurgici minimamente invasivi. [4]

1.1.1 Optical Coherence Tomography

La Tomografia a Coerenza Ottica (OCT) è una tecnica di imaging diagnostico non invasiva. E' applicata in diversi campi quali, in modo particolare, l'oftalmologia e la dermatologia. Il principio di funzionamento si basa sull'utilizzo della luce nella gamma spettrale del vicino infrarosso che, urtando contro un ostacolo nel suo percorso, determina un cambiamento delle sue proprietà in termini di direzione, ampiezza, lunghezza d'onda. [5].

La luce retrodiffusa viene misurata con un set-up interferometrico per ricostruire il profilo di profondità del campione nella posizione selezionata. La risoluzione assiale e il campo di imaging di un sistema OCT sono determinati dalla sorgente luminosa e dalle caratteristiche del rivelatore. [6]

Lo schema di acquisizione dello strumento è quello riportato in *Figura 1.1* e si compone di una sorgente luminosa a bassa coerenza, di un Beam Splitter, ovvero un dispositivo che divide la luce proveniente dalla sorgente in due raggi separati, uno trasmesso e uno riflesso. Questi sono chiamati raggio di riferimento e raggio campione. Le luci retrodiffuse dei due raggi si ricombinano tra loro nel Beam Splitter formando un nuovo raggio (raggio detector) che viene diretto al rivelatore.

Vi è poi uno specchio di riferimento, che riflette la luce del raggio di riferimento; il campione, che riflette la luce incidente del raggio campione a seconda delle sue proprietà; e infine il detector, che rileva le informazioni di interferenza trasportate dal raggio detector.

L'interferenza, ovvero la somma delle onde coinvolte, può essere costruttiva (quando sono in fase) o distruttiva (quando sono in controfase).

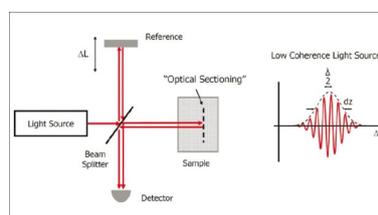


Figura 1.1: Schema OCT [7]

L'output di un sistema di acquisizione OCT è un tomogramma della porzione di tessuto considerata.

Un tomogramma è un'immagine bidimensionale che mostra una slice di un oggetto tridimensionale. In particolare, si definisce:

A-scan la sezione assiale nella direzione z , quindi una singola scansione unidimensionale del campione.

B-scan la composizione di *A-scans* adiacenti lungo la direzione x , quindi si ottiene una slice nel piano x - z .

Infine, *B-scans* adiacenti acquisiti lungo la direzione y formano la struttura tridimensionale, anche detta *C-scan*.

Ciascun piano che si forma con una coppia di assi, forma un'immagine bidimensionale che corrisponde ad un tipo diverso di visualizzazione. In particolare il piano xy corrisponde alla visualizzazione *En Face*, nel piano yz si ha la visualizzazione *Cross Section* e, come precedentemente citata, nel piano xz quella *B-scan*.

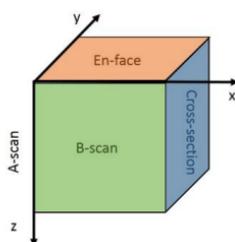


Figura 1.2: Piani di scansione OCT

Tale tecnica è in grado di performare immagini cross-section di tessuti biologici con una risoluzione assiale inferiore a 10 μm usando le onde luce. A differenza degli US, questa tipologia di imaging non richiede il contatto diretto con i tessuti ma può essere effettuata in aria. Uno svantaggio di questa tecnica è che il fenomeno di backscattering e la luce di assorbimento limitano l'imaging ai primi 2-3 mm di tessuto. [8] Ci sono diverse tecniche mediante le quali è possibile eseguire un OCT: nel dominio del tempo, nel dominio spettrale e nel dominio spettrale 'swept'.

Per quanto riguarda l'OCT nel dominio temporale, si sfrutta una luce monocromatica. L'interferenza costruttiva è osservata al massimo dell'intensità quando il percorso ottico di entrambe le onde è esattamente lo stesso. Muovendo lo specchio è possibile quindi osservare diversi *layer*, uno per volta. Si tratta infatti di una tecnica lenta, basata sul movimento meccanico dello specchio, causando anche l'introduzione di artefatti.

L'OCT nel dominio spettrale è basata sulla trasformazione da tempo in frequenza, per mezzo della Trasformata veloce di Fourier (FFT). Se sono presenti più onde luminose, caratterizzate da lunghezze e frequenze diverse, facendole passare insieme si può ottenere una combinazione nel tempo. Quindi, qualsiasi onda risulta composta da diverse onde e, applicando la trasformata di Fourier, è possibile mostrare quante onde ci siano e la riflessione di ciascuna frequenza separatamente. Facendo così, una combinazione di onde può essere divisa in onde separate. Quindi, in questo caso, non vi è una dipendenza dal movimento dello specchio ma dalla velocità della luce, pertanto questa tecnica

di imaging è molto rapida.

Infine l'OCT spettrale con sorgente Swept, ha un meccanismo simile alla precedente ma la separazione non avviene a livello del detector ma a quello della sorgente della luce stessa. Tramite questa tecnica si possono osservare strutture più profonde .

1.1.2 Imaging Fotoacustico

L'imaging fotoacustico è una nuova metodica di imaging non ionizzante e non invasiva basata sull'effetto fotoacustico, ovvero il risultato dell'irradiazione di un oggetto con impulsi laser che produce una risposta termoacustica. La luce assorbita in una specifica regione è convertita in calore, portando poi all'espansione termoelastica e quindi all'emissione di onde acustiche nella banda degli ultrasuoni. Tali onde US generate, vengono rilevate da trasduttori a US e analizzate per produrre immagini. [9].

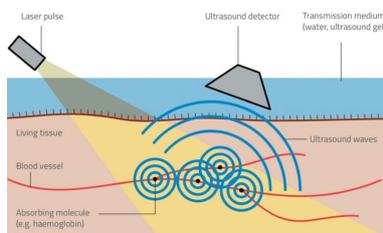


Figura 1.3: Principio di funzionamento delle immagini fotoacustiche [10]

Grazie alla natura ibrida di tale modalità di imaging, ovvero eccitazione ottica e detection acustica, presenta numerosi benefici quali alta risoluzione spaziale e basso scattering.

Difatti, poichè le immagini vengono generate a partire da un sistema con detector ad ultrasuoni, la risoluzione spaziale dipende dal sistema stesso e dalle caratteristiche in frequenza delle onde acustiche generate. Per ottenere buoni risultati si preferisce sfruttare durate degli impulsi attorno alle decine di nanosecondi, in modo da avere una risposta acustica a larga banda, fino ai megahertz [9] Variando quindi la durata e l'energia degli impulsi è possibile ottenere frequenze acustiche diverse, influenzando sulla risoluzione spaziale finale. Inoltre, la profondità di illuminazione ha un grande impatto sulle onde acustiche, data l'attenuazione dei tessuti, limitando la frequenza massima dell'onda fotoacustica e quindi dando un limite superiore alla risoluzione spaziale ottenibile.

Per sfruttare l'effetto fotoacustico al fine di ottenere immagini, è necessario utilizzare sistemi per l'imaging a ultrasuoni. [11]. Gli ultrasuoni sono onde acustiche che si propagano con una frequenza propria maggiore di 20000 Hz, limite massimo di percezione dell'orecchio umano. La generazione degli ultrasuoni avviene mediante effetto piezoelettrico, mediante il passaggio di una corrente elettrica in elemento cristallino che ne induce la vibrazione ad una certa frequenza.

In particolare, nell'imaging a US vengono impiegate sonde formate da array di cristalli piezoelettrici che sono in grado, allo stesso tempo, di generare e rilevare onde acustiche. A seconda del materiale in cui l'onda si propaga, bisogna considerare la non linearità del cammino di un impulso ad ultrasuoni, ovvero un'attenuazione esponenziale che dipende dalle caratteristiche del materiale. Nel caso di passaggio tra un materiale ed un altro, l'impulso presenta, oltre che attenuazione, anche effetti di trasmissione e riflessione.

Alla base della realizzazione di imaging ad ultrasuoni vi sono difatti questi due concetti sovraccitati, i quali vengono generati e rilevati dalla stessa sonda [11]: l'immagine viene ricostruita a partire dai diversi echi generati dalle discontinuità dei tessuti. Esistono diverse modalità con cui è possibile rappresentare le immagini e la più comune è la cosiddetta *B-mode*, cioè ogni punto nell'immagine viene rappresentato con un'intensità diversa a seconda dell'ampiezza dell'eco prodotto in quel punto, scandendo con tutto l'array di cristalli nella sonda.

1.2 Stato dell'arte delle tecniche di tracking

Diversi approcci sono stati proposti al fine di effettuare la procedura di tracking all'interno di immagini da biopsia, da quelli modellistici geometrici ad algoritmi implementati sfruttando l'intelligenza artificiale. La complessità di questo tipo di algoritmi risiede nelle caratteristiche delle immagini su cui operano:

In particolare, i volumi OCT possono essere influenzati da diversi artefatti, quali quello di proiezione e l'artefatto di movimento. Il primo è causato da ombre fluttuanti proiettate dalle strutture anatomiche presenti, che comporta una variazione sia dell'ampiezza che della fase. Il secondo si verifica quando c'è movimento del paziente durante la fase di acquisizione e causa l'offuscamento e la distorsione dell'immagine. Fattori che devono essere considerati per l'artefatto di movimento sono il tempo di acquisizione, che se più lungo aumenta la difficoltà a mantenere la posizione ferma per il paziente, e l'età o le condizioni mediche del paziente, che possono aumentare la probabilità di artefatti da movimento in presenza per esempio di tremori o altri disturbi del movimento. Mentre l'artefatto di movimento può essere superato con algoritmi di correzione del movimento relativamente semplici, gli artefatti di proiezione richiedono un'elaborazione più complessa algoritmi per ottenere un'immagine più pulita.

Anche i volumi PAT risentono di problematiche dovute ad artefatti di movimento, che possono essere causati dal paziente o dal sistema di acquisizione, che possono influenzare la corretta localizzazione dell'ago nel tempo, oltre che di distorsioni acustiche e geometriche: infatti, a causa delle proprietà ottiche e acustiche dei tessuti, tali distorsioni possono compromettere la qualità delle immagini e influenzare l'accuratezza del tracciamento dell'ago.

1.2.1 Approcci tradizionali

Gli approcci tradizionali si basano spesso su modelli matematici al fine di determinare la posizione dell'ago all'interno dell'immagine.

Si riportano di seguito alcuni dei principali metodi riscontrati in letteratura.

Principal Component Analysis

Il primo approccio si basa sulla *Principal Component Analysis* [12], una trasformazione ortogonale per convertire un insieme di variabili correlate in un insieme di variabili non correlate chiamate componenti principali.

Tale metodo si basa sulla scomposizione in autovalori e autovettori della matrice di covarianza. Gli autovalori, in ordine decrescente, rappresentano l'ammontare della variabilità totale osservata sulle variabili originarie, espressa da ciascuna componente principale; gli autovettori invece rappresentano le corrispondenti direzioni (ortogonali) di massima variabilità estratte dalle compo-

nenti principali.

Generalmente, la varianza della prima componente principale è massima. Pertanto, si calcola la varianza dell'immagine la quale viene sogliata opportunamente per distinguere pixel appartenenti all'ago e quelli appartenenti allo sfondo creando un'immagine binaria. Si creano così nell'immagine cluster di pixel luminosi appartenenti all'ago.

Ciascun cluster di pixel viene processato usando la PCA per distinguere l'ago dagli altri oggetti rettilinei, ovvero si determinano le componenti principali di ogni cluster. Si utilizzano le coordinate cartesiane di ogni pixel all'interno del cluster (x_i, y_i) per ricavare la direzione dell'asse principale del cluster risolvendo il primo autovettore della matrice di covarianza R_i , definita come :

$$R_i = \begin{pmatrix} \delta_{xx}^2 & \delta_{xy}^2 \\ \delta_{yx}^2 & \delta_{yy}^2 \end{pmatrix}$$
$$\delta_{xy} = \frac{1}{B} \sum_{i=1}^B (\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{y}_i - \bar{\mathbf{y}})$$

dove B è il numero dei pixel all'interno del cluster.

La varianza della componente principale è calcolata come:

$$\delta_{pc}^2 = \frac{\delta_{xx}^2 + \delta_{yy}^2 \pm \sqrt{(\delta_{xx}^2 - \delta_{yy}^2)^2 + 4(\delta_{xy}^2)^2}}{2}$$

L'ago è infatti individuato come il cluster con l'asse maggiore più lungo, ovvero quello con varianza maggiore tra tutti i cluster, la cui direzione è data da:

$$\alpha = \tan^{-1} \left(\frac{\sigma_{xy}^2}{\sigma_{yx}^2 - \sigma_{yy}^2} \right)$$

Il limite di tale metodo risiede nel fatto che viene riprodotta la linea su cui l'ago giace ma non identifica la posizione precisa dello stesso.

Model Fitting

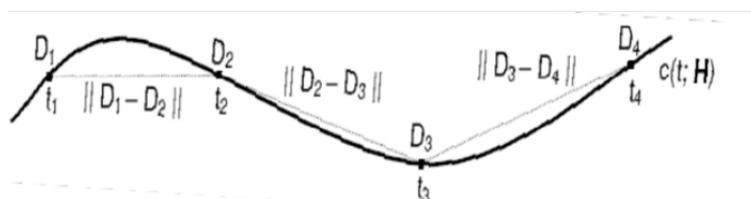
Un secondo approccio è quello di *Model fitting* basato sull'algoritmo *RANSAC*. [13] Questa metodologia sviluppa un modello matematico i cui parametri vengono ricavati iterativamente a partire da un insieme di dati contenente outliers e inliers e una funzione costo e fornisce in output i parametri del modello che massimizzano la funzione costo e gli inliers consistenti con tale modello.

Si basa sull'assunzione che l'intensità dei voxel dello strumento sia maggiore di quella del tessuto circostante e che la sua forma sia un cilindro sottile, lungo e possibilmente curvo.

Segue poi una fase di thresholding: si distinguono i voxel appartenenti all'ago (inliers) e quelli appartenenti allo sfondo in base all'intensità (outliers).

Infine, la fase di localizzazione dell'asse: l'asse dell'ago è rappresentato come una curva polinomiale parametrica spaziale $c(t, H) : \mathbb{R} \rightarrow \mathbb{R}^3$ di ordine $n-1$. H è una matrice di coefficienti della curva, t è un parametro della curva stessa che permettono di definire 4 punti di controllo D che giacciono su di essa (legati dalla relazione $H = DT^{-1}$)

$$c(t; H) = \begin{pmatrix} h_{11} & h_{12} & h_{13} & h_{14} \\ h_{21} & h_{22} & h_{23} & h_{24} \\ h_{31} & h_{32} & h_{33} & h_{34} \\ h_{41} & h_{42} & h_{43} & h_{44} \end{pmatrix} \cdot \begin{pmatrix} 1 \\ t \\ t^2 \\ t^3 \end{pmatrix}$$



Lo svantaggio di tale metodo risiede nel fatto che il valore ottimale della soglia è dedotto dal modello statistico stimato, che descrive l'intensità dei voxel. I parametri del modello sono stati stimati usando immagini in cui l'ago è presente, ma questo non è sempre possibile in immagini in vivo e in real time. Inoltre le ipotesi sulla forma e sull'intensità dell'ago risultano stringenti perché nelle immagini reali l'intensità dei pixel dello strumento non è facilmente distinguibile da ciò che lo circonda.

Fibra Ottica aggiuntiva

Infine l'approccio proposto da Shubert et al. [14] richiede che una fibra ottica sia inserita nell'ago, al fine di indurre la generazione di segnali acustici alla punta di esso mentre l'altra estremità della fibra è accoppiata ad una sorgente di impulsi laser.

L'algoritmo di tracking prevede due step:

- Il primo è la segmentazione della punta dell'ago mediante thresholding dell'immagine con un valore di soglia scelto dinamicamente in base all'intensità; viene applicata l'erosione binaria e dilatazione; segue il labeling dei componenti connessi; si calcola poi l'area dei pixel appartenenti a ciascun label e tali valori vengono rappresentati con un istogramma. L'area appartenente all'ago è quella maggiore, solo se questo valore rappresenta un outlier dell'istogramma. In questo caso, la posizione dell'ago per quel frame è il baricentro dell'area appartenente a quel label. Se non è presente alcun outlier, allora l'immagine non ha l'ago.

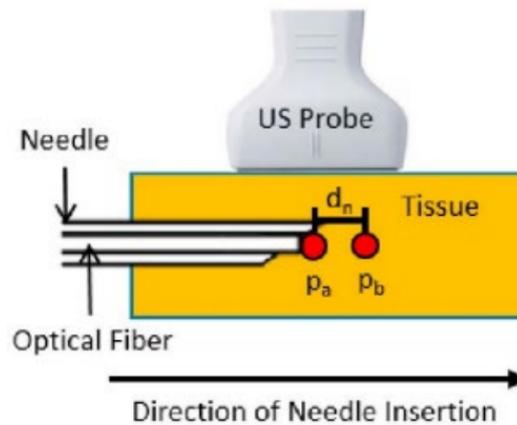


Figura 1.4: Set-Up di validazione

- Il secondo è il centraggio della sonda: ricavata la posizione p della punta dell'ago per ogni frame, si ricava la posizione media per tutti i frame.

Lo svantaggio di tale tecnica risiede tuttavia nella strumentazione aggiuntiva, ovvero l'inserimento della fibra ottica, in quanto potrebbe limitarne l'utilizzo in alcune procedure chirurgiche.

1.2.2 Approcci di Deep Learning

Il tracking automatico dello strumento è stato precedentemente affrontato in letteratura con approcci diversi, manuali (con il posizionamento di markers alla base dello strumento) e morfologici. Il problema è stato poi affrontato anche con metodi basati sul Deep Learning. Tale tecnica si è dimostrata essere promettente in primis per la caratteristica distintiva, insita del deep learning, di apprendere automaticamente le caratteristiche rilevanti dei dati; poi per la capacità di generalizzazione e adattabilità, punto chiave di questo lavoro di tesi.

Occorre tuttavia evidenziare che tale tecnica richiede una significativa quantità di dati per l'addestramento, oltre che risorse computazionali elevate.

In generale il Needle Tracking basato su rete neurale, parte da volumi grezzi di cui si effettua la segmentazione manuale e che verranno poi utilizzati come dati di training per la stessa al fine di identificare e delineare il target all'interno delle immagini. Tra le architetture maggiormente utilizzate per la segmentazione di immagini con questo scopo, è opportuno citare la Convolutional Neural Network (CNN) e la U-Net con le sue varianti, le cui performance risultano allo stato dell'arte efficienti in termini di accuratezza e robustezza.

Capitolo 2

Materiali

Per questo lavoro di tesi sono state utilizzate due tipologie di dataset, corrispondenti a due modalità di imaging differenti. L'addestramento della rete è stato effettuato nel primo caso con volumi OCT, successivamente un secondo modello con volumi PAT, infine il testing di immagini PAT con il modello OCT ha permesso di integrare entrambe le tipologie. Questa scelta si presta allo scopo del progetto di voler estendere il tracking dello strumento a diverse procedure di diagnostica, seguendo un approccio multimodale.

2.1 Dataset OCT

2.1.1 Acquisizione dei dati

I dati OCT sono stati forniti dal *Center for Medical Physics and Biomedical Engineering of the Medical University of Vienna*. Per l'acquisizione dei volumi è stato utilizzato un sistema combinato OCT/PAT, il quale prevede una sorgente di luce Swept che campiona continuamente (DC =100 %) con una frequenza pari a 400kHz. Ulteriori parametri del sistema utilizzati durante l'acquisizione sono mostrati nella Tabella 2.1.

Tabella 2.1: Parametri sistema OCT

Parametri OCT	
Central Wavelength	1310 nm
Bandwidth	29 nm
Swept Frequency	400 kHz
Source Power	10mW
Duty Cycle	100%
Coupler ratio	75:25
Axial pixel size	13.7 μ m
Lateral pixel size	9.8 μ m
Signal to noise ratio (SNR)	105.3 dB
Field of View	1 cm \times 1 cm

Nella figura seguente viene riportato lo schema circuitale

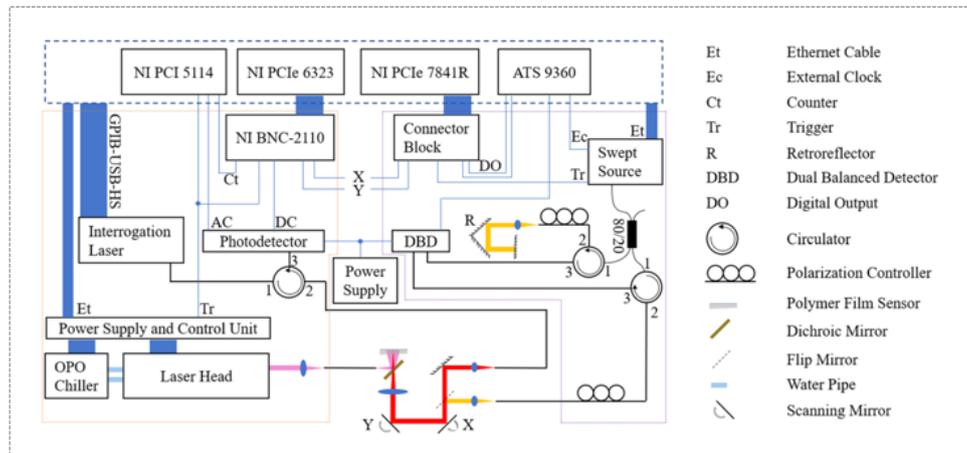


Figura 2.1: Sistema di acquisizione OCT/PAT

Tissue-mimicking phantom

Al fine di ricreare una situazione realistica di inserzione durante una procedura da biopsia di un ago in un tessuto, sono state effettuate diverse prove su fantocci per ricreare il tessuto e il target.

In particolare, per quanto riguarda l'emulazione del tessuto, è stato opportunamente scelto il materiale del fantoccio in base a proprietà meccaniche, stabilità della struttura tridimensionale e proprietà ottiche controllabili.

Il silicone Dragon Skin™ 30, con un rapporto di miscelazione di 1:1, una durezza di 30A e un tempo di polimerizzazione di 16 ore, è stato scelto per la sua tenacità, la compatibilità con gli agenti di diffusione e assorbimento e le sue proprietà meccaniche controllabili. Inoltre, l'indice di rifrazione del silicone, pari a 1.4, corrispondeva perfettamente a quello dei tessuti biologici (1,35 ÷ 1,55).

Per migliorare le proprietà di diffusione e assorbimento del fantoccio, al silicone sono stati aggiunti il biossido di titanio (TiO₂) e il pigmento nero Silc Pig™. Tuttavia, il silicone ha incontrato difficoltà nell'ottenere una distribuzione omogenea senza aggregati, sedimentazione o bolle d'aria. Per ridurre al minimo questi problemi sono state utilizzate tecniche come la sonicazione e il degasaggio sotto vuoto.

L'attrezzatura necessaria per la fabbricazione comprendeva:

- 60mL di matrice siliconica;
- 60mL di catalizzatore in silicone;
- TiO₂;
- pigmento nero
- 3 tubi conici Falcon™ da 50 ml;
- 3x siringhe da 50mL;
- 1 capsula di Petri;
- Camera a vuoto;
- 1x pennello.

Il processo di fabbricazione è iniziato preparando tre miscele di catalizzatore con diverse concentrazioni di TiO₂ e pigmento nero:

1. 8mg/mL di TiO₂ e 0,2% in peso di pigmento nero in 20mL di catalizzatore;
2. 0,2% in peso di pigmento nero in 20mL di catalizzatore;
3. 1% in peso in 20mL di catalizzatore.

Le miscele sono state poste in tubi conici Falcon™ e sottoposte a sonicazione in un bagno a ultrasuoni per rompere gli aggregati. Dopo la sonicazione, la matrice di silicone è stata aggiunta alle miscele in rapporto 1:1.

La tecnica di fabbricazione prevedeva di versare uno strato sottile della prima miscela, che rappresentava l'epidermide, sul fondo della piastra di Petri. Uno strato più sottile della seconda miscela è stato versato sopra di esso per simulare la disomogeneità del tessuto. La piastra di Petri è stata posta in una camera a vuoto per rimuovere le bolle d'aria ed è stato applicato uno strato più spesso della terza miscela, che rappresenta l'elevato assorbimento del tessuto.

Dopo l'indurimento, il phantom è stato rimosso dalla piastra di Petri e rivestito con un sottile strato di silicone trasparente per simulare lo strato corneo presente nelle immagini OCT della pelle.

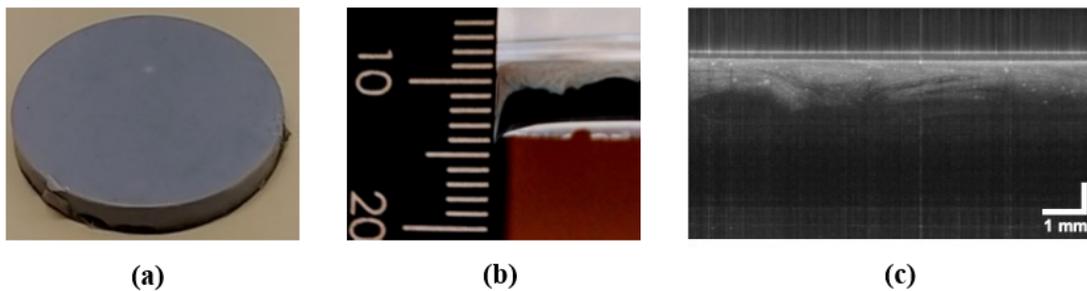


Figura 2.2: Fabricated phantom. Top view (a) and lateral view of a section of the phantom (b). OCT B-scan of the phantom (c).

All'interno del fantoccio è stata inserita una piccola sfera in plastica per mimare la regione di tessuto da essere esportata tramite biopsia. Per il progetto è stato utilizzato un ago di calibro 20 al fine di simulare al meglio la biopsia percutanea in vivo. Sono state registrate 9 simulazioni e per ogni simulazione tre posizioni dell'ago consecutive.

Dopo aver inserito l'ago nella posizione desiderata, il fantoccio è stato fissato con dei perni ad un blocco di polistirene, al fine di evitare movimenti durante l'inserzione.

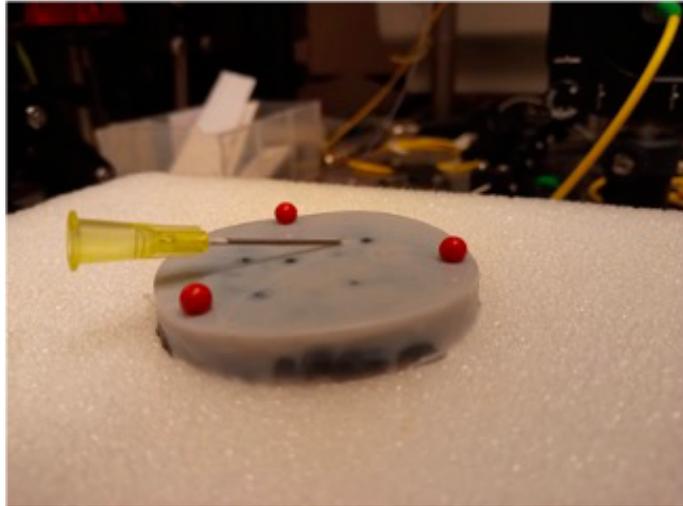


Figura 2.3: Phantom setup for OCT acquisition

2.1.2 Ricostruzione volumetrica

Diversi algoritmi di ricostruzione sono stati individuati in letteratura. In particolare, Watanabe e Itagaki [15] hanno sviluppato un algoritmo accelerato da GPU per la ricostruzione volumetrica di OCT, il cui sviluppo è schematizzato nella figura sottostante.

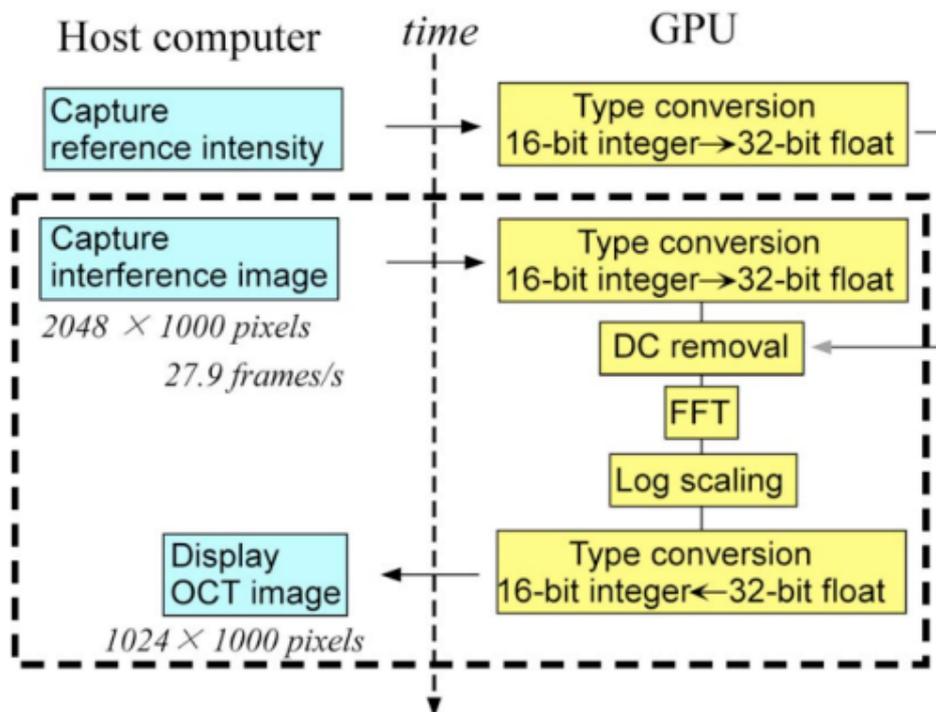


Figura 2.4: Algoritmo di ricostruzione OCT basato sul metodo [15]

Il processo si sviluppa in 4 operazioni principali:

1. **Conversione da 16 bit int a 32 bit float:** le immagini PA acquisite sono state salvate in file binari a 16 bit senza segno. Poichè si vuole eseguire la FFT, vengono convertiti a 32 bit con virgola mobile.

2. **Rimozione del DC:** viene effettuata un'operazione di denoising al fine di ridurre gli artefatti, sottraendo il segnale relativo allo sfondo dall'immagine di acquisizione. Il segnale di sfondo è infatti prodotto dal fascio di riferimento e si evidenzia quando si blocca il braccio del campione.

3. **FFT:** L'interferogramma si ottiene mediante la trasformata di Fourier dello spettro di ampiezza della luce del campione retrodiffusa.

4. **Conversione da 32 bit float a 16 bit int:** i dati vengono convertiti in un formato visualizzabile. a seguito di questa operazione è necessario un ricampionamento.

L'immagine interferente viene acquisita e salvata dal sistema OCT come un file binario che si presenta come un vettore di 1342177280 32-bit int ($1280 \times 1024 \times 1024$). L'operazione di resizing ha poi trasformato il vettore in un array numpy tridimensionale.

Poichè la presenza dell'ago aumenta il rumore legato all'artefatto di bordo, è stata applicata una finestra gaussiana a ogni A-line dell'immagine interferente.

Per quanto riguarda la FFT, è stata applicata su un numero di punti pari alla lunghezza della A-line e tutti i valori complessi dei pixel sono stati riportati nello spazio reale mediante valore assoluto. La replica spettrale è stata poi rimossa riducendo la lunghezza della A-line a 512 pixel. Infine gli artefatti di 'mirror' sono stati rimossi tagliando 64 B-scan da ogni lato del volume. La dimensione finale del volume è di 512x896x896.

Segue un'operazione di scaling logaritmico al fine di sfruttare l'intera dinamica dell'immagine, che funziona come un'equalizzazione dell'immagine.

A seguito dell'operazione di ricostruzione volumetrica, per ciascun ago e ciascuna posizione di esso, si è otterrà un volume finale NXPY (X numero che indica l'ago, Y numero che indica la posizione) con estensione .nrrd e di dimensioni 512x896x896.

Di seguito viene riportato il volume caricato in Slicer3D di un volume OCT nelle tre visualizzazioni. In rosso è evidenziata la sezione dell'ago utilizzata per l'addestramento dell'algoritmo.



Figura 2.5: Visualizzazione volumetrica in Slicer3D del VolumeOCT_Needle2_pos1

2.1.3 Pre Processing dei dati

Al fine di addestrare un modello di rete neurale, è opportuno lavorare con immagini di alta qualità per facilitare l'apprendimento. Si è scelto di addestrare la rete utilizzando la sezione sagittale, in cui l'inserzione dell'ago risulta maggiormente visibile, pertanto è necessario estrarre dal volume singole immagini bidimensionali. Sono state estratte 896 slices di dimensione 512x896, che verranno poi processate.

In particolare, la pipeline di pre processing si compone di diversi step, illustrati nel flowchart che segue.

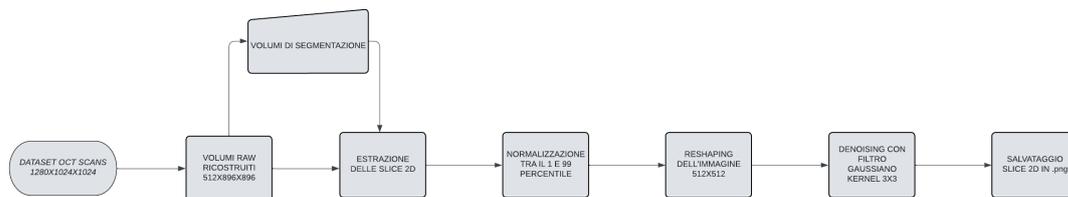


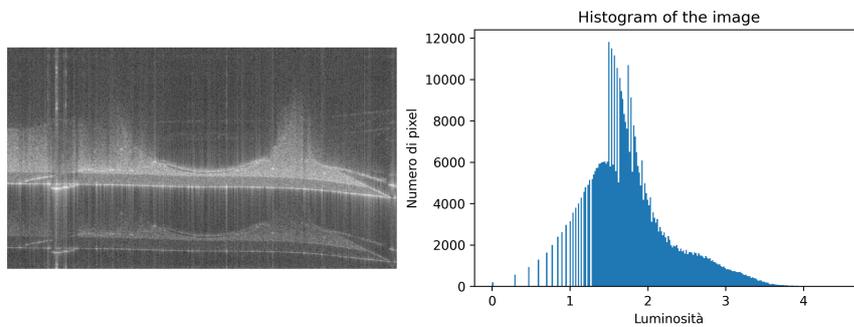
Figura 2.6: Pipeline di Pre-Processing OCT

1. **Normalizzazione:** le immagini vengono normalizzate tra il 1° e il 99° percentile, al fine di riportare i valori dei pixel, inizialmente compresi tra 0 e 5 circa, nel range 0-255.

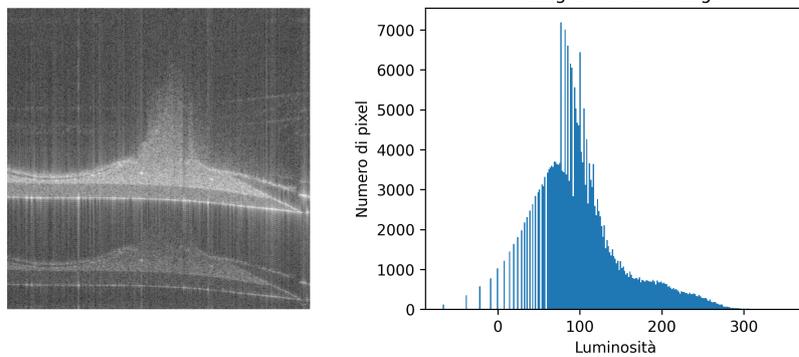
2. **Ridimensionamento:** La dimensione di ciascuna slice è stata modificata in 512x512. Avendo l'obiettivo di ricerca dell'ago, essendo le slice contenenti

quest'ultimo di numero inferiore rispetto alle slice che invece non lo contengono, si è scelto per le immagini utilizzate per il training set di ridimensionare l'immagine non randomicamente ma con un crop che permettesse, nel caso in cui fossero presente, di includere tutti i pixel appartenenti all'ago e meno background possibile.

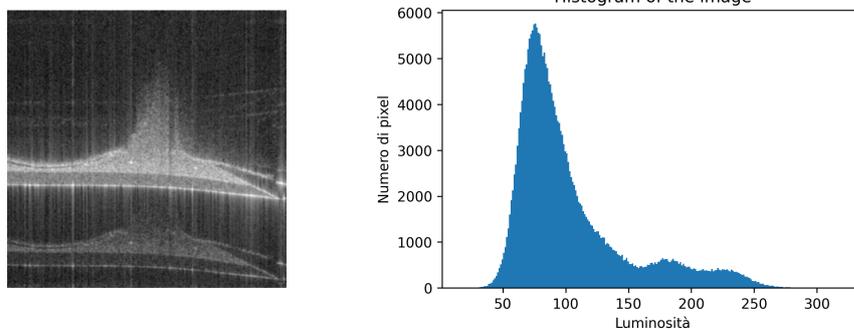
3. Riduzione del rumore: Al fine di ridurre il rumore presente nelle immagini, è stato implementato un filtro passa basso con kernel gaussiano. Il filtro gaussiano si basa sulla convoluzione dell'immagine con il kernel gaussiano, calcolando la media pesata dei pixel circostanti. La forma del kernel è determinata dal valore della deviazione standard, che controlla la distribuzione dei pesi dei pixel, e dalla sua dimensione. La dimensione del kernel scelta è 3×3 con sigma pari a 5. A seguito di questa modifica, lo sfondo risulta più omogeneo e anche il contrasto aumentato.



(a) Immagine originale



(b) Immagine normalizzata e ridimensionata



(c) Immagine filtrata

Figura 2.7

Infine, ciascun volume è stato caricato su Slicer 3D, al fine di creare le maschere binarie di segmentazione. Si associa ai pixel corrispondenti all'area dell'ago il valore 1, mentre allo sfondo il valore 0. Anche in questo caso, dal volume vengono estratte maschere bidimensionali e ridimensionate.

2.1.4 Costruzione del Dataset 2D

Il dataset OCT si compone complessivamente di 9 volumi, che si distinguono in base al tipo di ago e alla tipologia di inserzione:

- Needle2Pos1, Needle2Pos2, Needle2Pos3
- Needle6Pos2, Needle6Pos3, Needle6Pos4
- Needle7Pos1, Needle7Pos2, Needle7Pos3.

Al fine di addestrare una rete neurale, è stato calcolato il numero di immagini che presentassero la feature di interesse. I dati in possesso (nel complesso di 8064 immagini) sono stati suddivisi come segue:

- TRAINING SET – 4301 elementi, di cui 3828 non presentano l'ago.
- VALIDATION SET- 1075 elementi, di cui 972 non presentano l'ago. Si è scelto che la dimensione del validation test sia circa il 20 % del training set.
- TEST SET – 2688 elementi, di cui 2200 non presentano l'ago.

Nella divisione si è quindi tenuto conto delle due possibili occorrenze (presenza ago/ assenza ago) al fine di costruire un dataset bilanciato. L'algoritmo prevede quindi il caricamento dei volumi con le rispettive maschere di segmentazione manuali con estensione .nrrd e, iterando lungo la seconda dimensione, si sono estratte le slice bidimensionali e su ciascuna di esse è stata applicata la pipeline di pre-processing precedentemente descritta. Tali immagini sono state salvate poi nelle rispettive cartelle in formato .png.

2.2 Dataset PAT

2.2.1 Acquisizione dei dati

Il sistema utilizzato per le acquisizioni PAT è incluso nel sistema OCT/PAT usato anche per le acquisizioni OCT. I parametri applicati per queste acquisizioni sono riportati nella tabella sottostante.

Tabella 2.2: Parametri sistema PAT

Parametri PAT	
Excitation Wavelength	808 nm
Excitation Bandwidth	680-980 nm
Repetition rate	50 Hz
Pulse width	4 ns
Source Energy	33 mJ
Fluence at the sample	2.1 mJ/cm^2
Axial resolution	31 μm
Lateral resolution	62 μm
Field of View	1.5 cm \times 1.5 cm

Nel sottosistema PAT viene utilizzato un metodo di rilevamento completamente ottico che impiega un interferometro Fabry-Perot per rilevare gli impulsi fotoacustici. Il sistema prevede l'uso di un laser di eccitazione (SpitLight 600 OPO, INNOLAS, Krailling, Germania) e di un laser di interrogazione (Tunics T100S-HP-CL, Yenista Optics, Lannion, Francia).

La lunghezza d'onda di eccitazione è stata impostata a 808 nm in accordo al picco di assorbimento del sangue. Il processo di scanning nei volumi PAT è piuttosto lento e l'acquisizione di ciascun volume richiede circa 8 minuti.

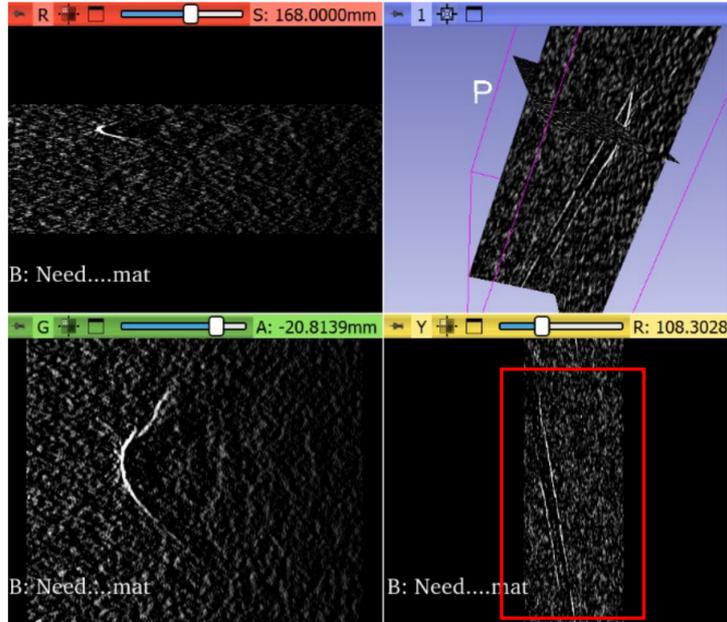


Figura 2.8: Visualizzazione volumetrica in Slicer3D volume NeedleInBeef_10s. In rosso è evidenziata la sezione utilizzata nell'addestramento

Volumi PAT combinati artificialmente

Al fine di ottenere un dataset realistico per l'algoritmo di tracking sono stati considerati volumi PAT combinati artificialmente. Alcuni dati sono stati acquisiti con un ago da biopsia calibro 20 in gel a ultrasuoni. Successivamente, l'ago è stato inserito in un pollo ex-vivo e nella carne, al fine di ricreare segnali realistici. Tuttavia da queste acquisizioni è risultato visibile solo il segnale proveniente dall'ago.

Alcuni dati sono stati acquisiti anche in vivo sulla pelle di alcuni volontari. L'obiettivo è quello di combinare artificialmente i volumi PAT che forniscono il solo segnale dell'ago e i volumi che forniscono solo il segnale della rete vascolare.

Questa operazione è stata effettuata in Matlab seguendo diversi step:

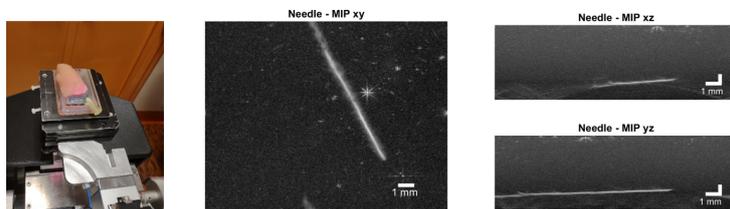


Figura 2.9: MIP nelle tre direzioni di acquisizione PAT di un ago inserito nel petto di un pollo

- I volumi PAT dell'ago sono stati flippati così che l'oggetto è allo stesso livello di profondità della rete vascolare nelle acquisizioni in pelle.
- La luminosità dei due volumi è stata regolata così che l'ago risulta più luminoso ma la rete vascolare sia ancora visibile.
- I due volumi sono stati sommati così che la rete vascolare e l'ago risultino sovrapposti.

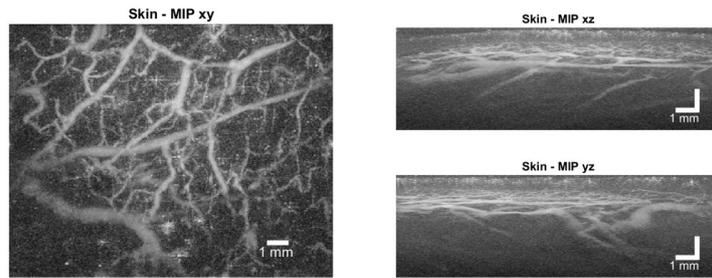


Figura 2.10: MIP nelle tre direzioni di acquisizione PAT del palmo di una mano

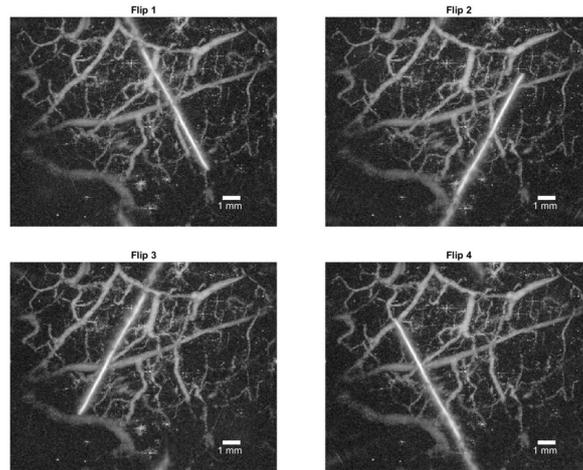


Figura 2.11: MIPs in the three directions of a skin-needle merged volume.

2.2.2 Pre Processing dei dati

Anche nel caso delle immagini fotoacustiche è risultata necessaria l'implementazione di una pipeline di pre processing per migliorare la qualità delle immagini di training ed uniformare il dataset.

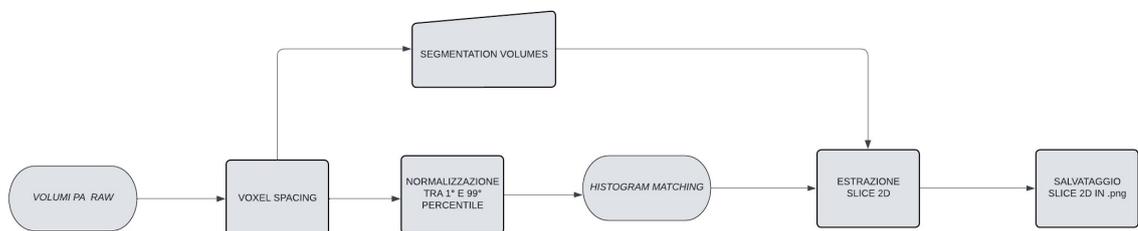


Figura 2.12: Pipeline di pre processing volumi PAT

In primis, i volumi utilizzati per il training del modello sono stati ricampionati. Questa scelta si deve al fatto che il dataset risulta piuttosto eterogeneo dal punto di vista dello spacing, con valori di voxel spacing [0.35278 mm, 0.35278

mm,1.0 mm]. Al fine di ottenere un numero intero di slice visualizzabile nelle tre dimensioni, si è scelto di riportare i valori dei voxel spacing a 1.0 mm in tutte le dimensioni, così da uniformarli.

Successivamente è stata effettuata la normalizzazione tra il 1° e il 99° percentile per riportare il valore dei pixel tra 0-255 e l'applicazione di un filtro gaussiano per la rimozione del rumore. In questo caso non è stato necessario effettuare il crop dell'immagine in quanto si tratta di patches di dimensioni 288x288. I volumi PAT sono poi stati sottoposti ad Histogram Matching.

L'histogram matching (HM) consiste nel modificare l'istogramma di un'immagine in base a quello di un'altra immagine di riferimento, quindi, è utile per rendere più simili i dati dal punto di vista del contrasto e della luminosità. Questa operazione è stata effettuata con l'obiettivo di uniformare l'istogramma delle immagini PAT a quello delle immagini OCT, al fine di poter effettuare il training della rete con immagini che abbiano caratteristiche simili.

Tale tecnica si basa sul presupposto che l'immagine da modificare e quella di riferimento abbiano le stesse dimensioni. Avendo assunto come dataset di riferimento quello dei volumi OCT con dimensioni 512x896x896, il quale presenta un istogramma delle luminosità maggiormente distribuito in maniera più uniforme, è stato necessario effettuare un reshaping dello stesso per uguagliarlo ai volumi PAT. Pertanto le dimensioni del volume di riferimento sono 288x896x288.

Per quanto riguarda la scelta del riferimento, diversi approcci sono utilizzabili. E' possibile sia scegliere arbitrariamente un'immagine dal dataset, oppure si mediano i volumi di interesse e si ricava un istogramma medio, al fine di avere un risultato maggiormente significativo. In questo caso è stato seguito il secondo approccio, ricavando l'istogramma medio di due volumi del dataset OCT.

Viene mostrato di seguito l'istogramma di riferimento calcolato sui volumi OCT ottenuto.

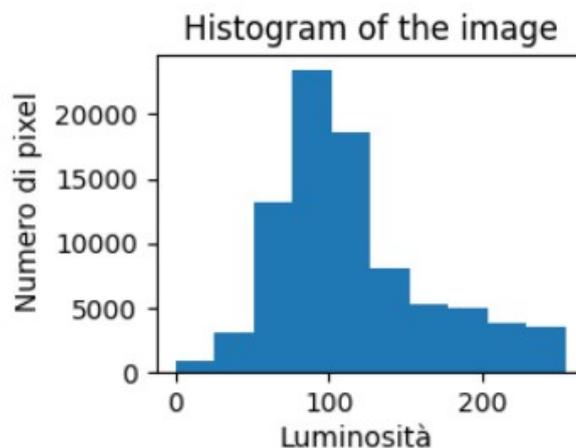
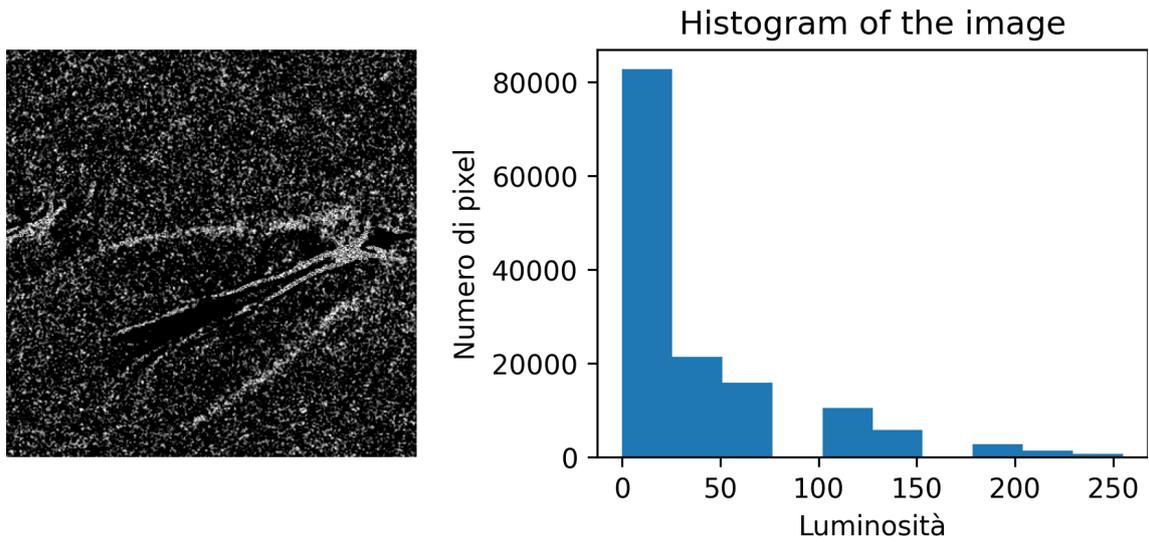


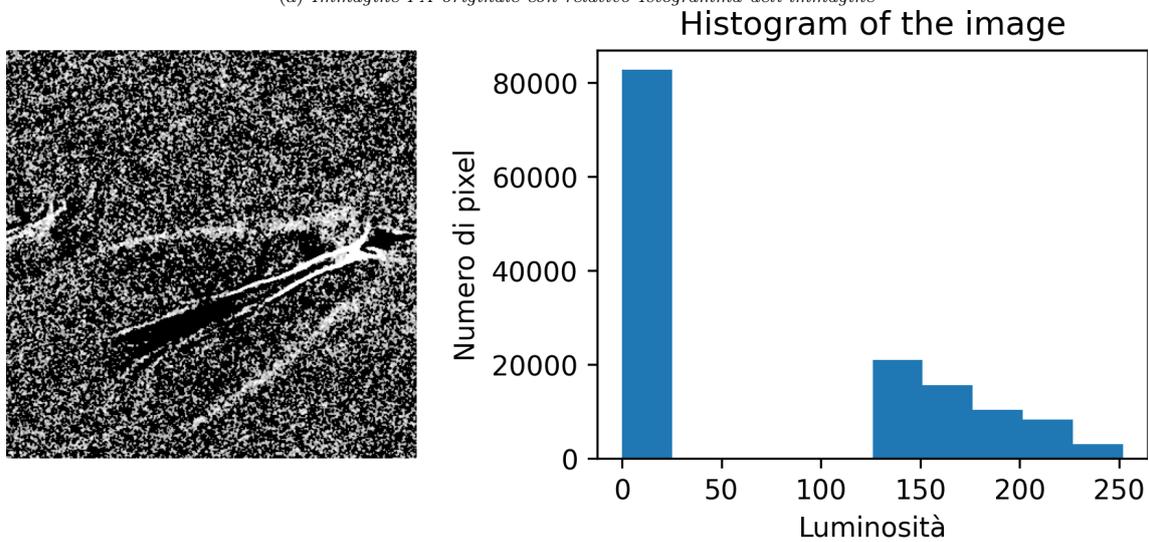
Figura 2.13: Iistogramma di riferimento medio calcolato su due volumi OCT

L'algoritmo di Histogram Matching è quello implementato da Nyul et al.[16] in cui sono stati imposti i parametri:

- histogram_levels = 256;
- match_points = 95;
- "ThresholdAtMeanIntensity = "True" per escludere i pixel del background e ottimizzare l'algoritmo.[17]



(a) Immagine PA originale con relativo Istogramma dell'immagine



(b) Immagine PA dopo Histogram Matching con corrispondente istogramma

Figura 2.14

2.2.3 Costruzione del Dataset 2D

La rete neurale da addestrare sarà di tipo bidimensionale. Occorre perciò estrarre anche in questo caso da ciascun volume pre-processato singole patch. I volumi .nrrd pre-processati e le rispettive segmentazioni manuali vengono caricate e si itera sulla prima dimensione estraendo le singole slice. Vengono visualizzati di seguito i volumi a disposizione:

Tabella 2.3: Volumi del Dataset PAT

Nome_Volume	Dimensioni
NeedleInGel_01s	950x288x288
NeedleInGel_12s	950x288x288
NeedleInGel_21s	950x288x288
NeedleInGel_42s	650x288x288
NeedleInGel_43s37	950x288x288
NeedleInGel_43s42	650x288x288
NeedleInBeef_10s	950x284x288
NeedleInBeef_33s	950x284x288
NeedleInMaxsPhantom_20s	950x284x288
NeedleInBeef_26s54	950x284x288
NeedleInBeef_26s44	950x284x288
NeedleInUSgel_22s	950x284x288
NeedleInMaxsPhantom_17s	950x284x288

Il totale delle immagini a disposizione è pari a 11750. Si è scelto di utilizzare 6 volumi per il training (5100 immagini 2D), 3 per il validation set (2850 immagini 2D) e 4 per il test set (3800 immagini 2D). Poichè le immagini contenenti l'ago risultavano inferiori rispetto alla totalità delle immagini, si è scelto di scartare una parte di immagini del volume non contenente l'ago evitando così l'overfit sulla classe 0 relativa allo sfondo. Questa selezione ha riguardato il training set, mentre per il test set si sono considerati i volumi nella loro interezza per valutare anche le performance sulle immagini che non contenessero la feature di interesse. Pertanto, in definitiva il dataset sarà diviso in:

- TRAINING SET – 2802 elementi
- VALIDATION SET- 830 elementi. In questo caso la dimensione del validation test è circa il 25 % del training set.
- TEST SET – 3800 elementi.

Anche in questo caso le immagini sono state preventivamente salvate in formato .png nelle rispettive cartelle.

Capitolo 3

Metodi

In questo capitolo vengono illustrati i metodi mediante i quali si è giunti alla soluzione del problema posto in esame. In particolare viene presentata una descrizione generale del modello della rete utilizzata per il task, la sua architettura e principio di funzionamento. Successivamente si specificano le feature implementate per ciascun modello sviluppato. La pipeline di inference comprendente l'applicazione del modello, la pipeline di riduzione dei tempi computazionali e l'applicazione di tecniche di post-processing. Infine nella sezione finale si descrivono le metriche utilizzate per la valutazione delle performance e la funzione di loss scelta per il training della rete.

3.1 Deep Learning Neural Network

Il Deep Learning è una tecnica di apprendimento automatico basata su reti neurali con un numero elevato di layer nascosti di elaborazione non lineare al fine di estrarre caratteristiche dei dati e trasformarli in diversi livelli di astrazione.

Si tratta di una tecnica di apprendimento che consente di elaborare grandi quantità di dati ed è in grado di eseguire automaticamente l'estrazione delle caratteristiche dai dati grezzi.

La rete neurale Convoluzionale è una rete di Deep Learning progettata per l'elaborazione di array di dati strutturati come le immagini e costituiscono lo stato dell'arte per molte applicazioni tra cui la classificazione e la segmentazione di immagini.

L'architettura di una rete neurale convoluzionale è di tipo feed-forward multistrato, realizzata impilando diversi layer nascosti in sequenza, al fine di garantire un apprendimento gerarchico.

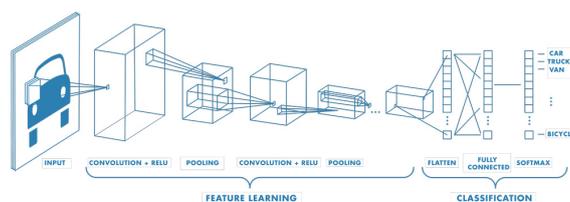


Figura 3.1: Architettura rete CNN che mostra i layer di convoluzione + ReLu (attivazione)+ layer di Pooling + Layer Fully Conncted

I layer nascosti sono livelli convoluzionali, seguiti dai livelli di attivazione a loro volta seguiti, nella maggioranza dei casi, dai layer di pooling.

- **Layer Convoluzionale:** è l'elemento chiave di una rete neurale. Si tratta di un'operazione di filtraggio tra un kernel e l'immagine di input lungo la quale viene fatto scorrere al fine di estrarre le features dell'immagine. Per la definizione di tale layer è necessario impostare la dimensione del kernel, lo stride (ovvero di quanto scorre il kernel lungo l'immagine), zero-padding ovvero si sceglie se aggiungere o meno una cornice di zeri lungo le dimensioni dell'immagine.

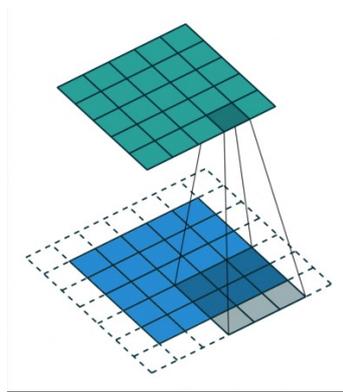


Figura 3.2: Layer Convoluzionale:

- **Layer di Attivazione:** successivamente, l'immagine passa attraverso la funzione di attivazione. Tale funzione ha l'effetto di aggiungere non linearità alla rete neurale. Le funzioni di attivazione più utilizzate sono la Sigmoide, per problemi di classificazione binaria; la Softmax, ovvero un'estensione della funzione sigmoide applicabile a problemi multiclasse: queste due funzioni associano ad ogni elemento dell'output la probabilità compresa tra 0 ed 1, di appartenere ad una classe; la funzione ReLU, che lascia invariati i valori positivi della matrice e porta a zero i valori negativi.

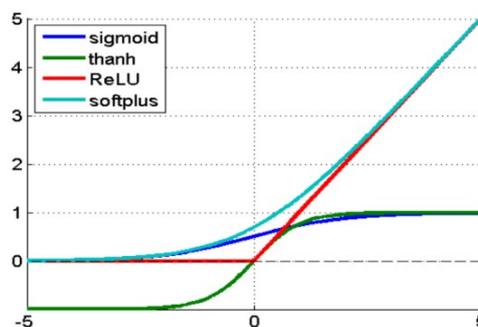


Figura 3.3: Esempi di funzioni di attivazione

- Layer di Pooling: Vengono utilizzati per ridurre la dimensione della mappa di caratteristiche, mantenendo le informazioni più rilevanti. Pertanto esegue un'operazione di feature selection.

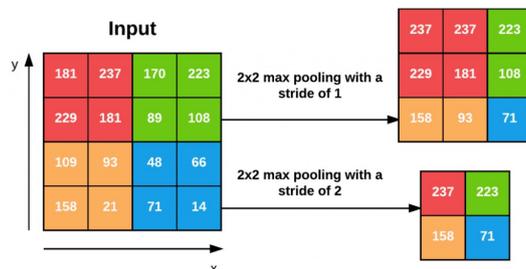


Figura 3.4: Esempio: a sinistra l'input è un volume 4×4 . A destra l'effetto del max pooling con stride pari a 1 e pari a 2. Questa applicazione riduce le dimensioni spaziali dell'input.

- Layer Fully Connected (FC): è costituito dai pesi e dai bias lungo i neuroni e viene utilizzato per collegarli tra due livelli diversi. Questi strati sono solitamente collocati prima dello strato di uscita e costituiscono gli ultimi strati di un'architettura CNN. In questo strato, l'immagine di ingresso proveniente dagli strati precedenti viene appiattita e inviata allo strato FC. Il vettore appiattito viene poi sottoposto a pochi altri strati FC, dove di solito si svolgono le operazioni matematiche. In questa fase inizia il processo di classificazione. Il motivo per cui due strati sono connessi è che due strati completamente connessi funzionano meglio di un singolo strato connesso. [18]

Nel seguente lavoro, l'algoritmo di segmentazione è stato implementato mediante il training di una Fully Convolutional Neural Network (FCN) utilizzata spesso per la segmentazione di immagini mediche, la rete U-Net con encoder ResNet_34. [19]. Una rete fully convolutional (FCN) è un'estensione delle CNN sviluppata per l'elaborazione di immagini e la segmentazione semantica. A differenza delle reti neurali convoluzionali tradizionali, che richiedono input di dimensioni fisse, le FCN sono progettate per accettare input di dimensioni arbitrarie e produrre output di dimensioni corrispondenti. Si tratta di un'architettura generale che utilizza convoluzioni trasposte come metodo di upsampling. [20]

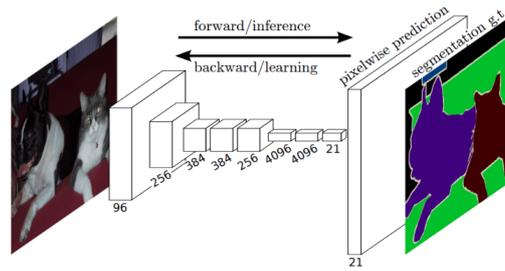


Figura 3.5: FCN Dopo un upsampling per recuperare la risoluzione completa dell'immagine all'ultimo livello, ogni pixel viene classificato utilizzando la softmax per generare le segmentazioni [21].

L'architettura U-Net, utilizzata nel presente lavoro, è un'architettura specifica di FCN sviluppata per la segmentazione di immagini biomediche, con il vantaggio di introdurre lunghe 'skip-connection' tra il percorso di contrazione e il percorso di espansione, permettendo l'integrazione di informazione a diverse scale e localizzare la segmentazione.

La rete è così chiamata per la sua forma ad 'U' che la rende simmetrica e la divide in due parti: un percorso di contrazione e un percorso di espansione.

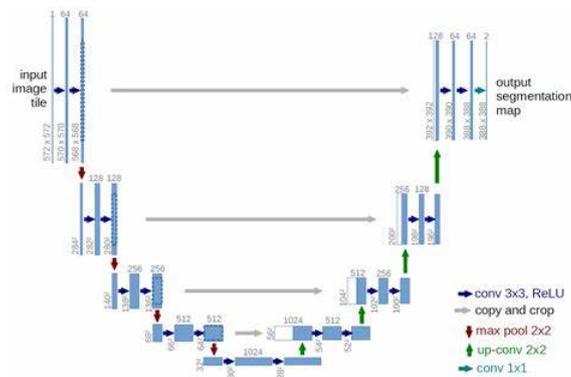


Figura 3.6: Architettura di una rete U- Net

Il percorso di contrazione (encoder), a sinistra, comprende il processo convoluzionale, tramite il quale l'immagine di input viene ridotta in termini di dimensioni. Consiste nell'applicazione ripetuta di due convoluzioni con kernel 3x3, ciascuno seguito da un layer ReLU e un'operazione di max pooling 2x2 con stride pari a 2 per il downsampling.

La funzione ReLU, che restituisce un valore pari a zero per ogni ingresso negativo, mantenendo come mantenere gli input positivi, viene utilizzata per introdurre la non linearità, rendendo il modello in grado di apprendere tipi di dati più complessi. Gli strati di max pooling, invece, sono utilizzati per campionare le mappe di caratteristiche prodotte dagli strati convoluzionali. Ad ogni fase di downsampling vengono raddoppiati il numero di canali di features mentre le dimensioni spaziali si dimezzano.

Il percorso di espansione (decoder), a destra, prevede invece layer convoluzionali 2x2, di attivazione e upsampling. Il decoder sovracampiona la feature map fino ad ottenere una matrice delle stesse dimensioni dell'immagine di in-

put. L'ultimo layer è costituito da una funzione di attivazione che permette di ottenere tante mappe quanti sono i canali di output, in cui ogni valore rappresenta la probabilità del pixel (o voxel) di appartenere ad una classe. Nei task di segmentazione multi-classe, viene spesso utilizzata la funzione Softmax; in questo caso, l'ultimo layer prende il nome di Softmax layer. [19]

L'intera pipeline, dal pre-processing al training e calcolo delle metriche è stata sviluppata in MONAI che fornisce dei framework basati su PyTorch ottimizzati per le immagini mediche, ed eseguita in Google Colab con GPU Tesla T4.

3.1.1 ResNet-34

La ResNet34 è un modello di rete neurale convoluzionale a 34 layer , preaddestrato sul set di dati ImageNet. La ResNet, a differenza delle reti neurali tradizionali, prende i residui di ogni strato e li utilizza nei successivi layer connessi. [22].

Questo tipo di rete risolve infatti il problema del 'vanishing gradient': quando la rete è troppo profonda, i gradienti da cui viene calcolata la funzione di perdita si riducono velocemente a zero, bloccando l'apprendimento poichè il valore dei pesi non viene aggiornato. La ResNet permette di superare questo problema con l'introduzione di 'connessioni residue', ovvero introducendo delle scorciatoie dirette che consentono il passaggio del segnale originale non modificato attraverso gli strati successivi.

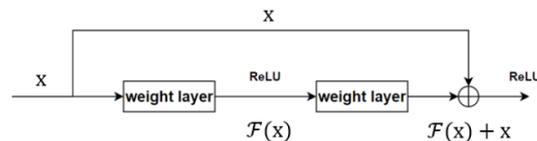


Figura 3.7: Esempio del blocco di connessione residua

In pratica, una connessione residua aggiunge l'input originale all'output dell'elaborazione dello strato corrente. Questo viene realizzato tramite un'operazione di somma element-wise tra l'input originale e l'output dello strato. In altre parole, l'output dello strato corrente rappresenta la "differenza residua" rispetto all'input originale.

L'aggiunta delle connessioni residue consente al segnale originale di propagarsi direttamente attraverso la rete senza alcuna trasformazione. Le connessioni residue consentono alla rete di imparare gli errori residui, ovvero le differenze tra l'output desiderato e l'output attuale, semplificando l'ottimizzazione dei pesi degli strati successivi. [23]

L'architettura della ResNet-34 è mostrata in Figura

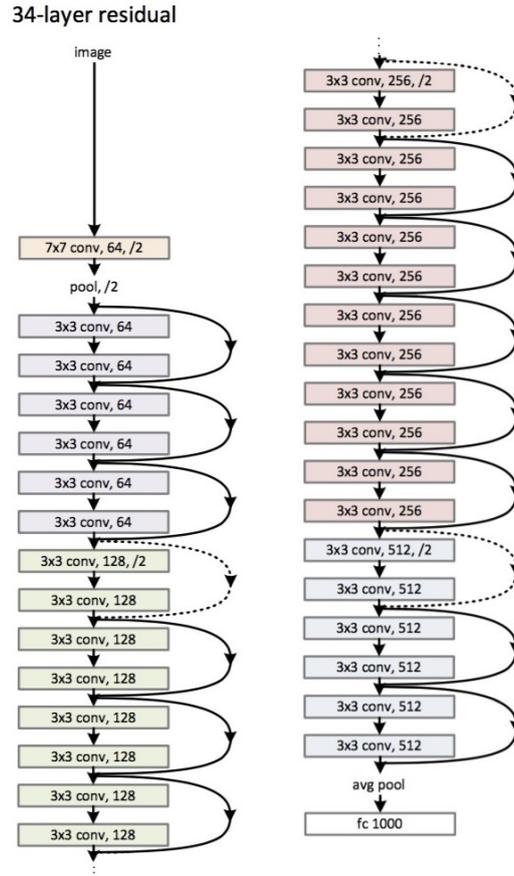


Figura 3.8: Architettura ResNet-34 con connessioni residue (la linea tratteggiata indica l'aumentare delle dimensioni con il padding). Il Downsampling è effettuato con layer convoluzionali con passo pari a 2.

Essa viene importata dalla libreria `segmentation_models_pytorch`, che fornisce le architetture preallenate sui più importanti benchmark. Al fine di rendere confrontabili i risultati finali, si è scelto di mantenere la stessa architettura della rete per i due modelli. In particolare essa presenta 4 layer con 16, 32, 64, 128 canali rispettivamente. L'encoder_depth è settata a 4, pertanto i tensori generati ad ogni stadio dell'encoder avranno risoluzione

$$[(H, W), (H/2, W/2), (H/4, W/4), (H/8, W/8), (H/16, W/16)]$$

I canali di input sono settati a 1, mentre per quanto riguarda i canali di output sono stati settati a 2.

Questa scelta dipende dal fatto che le maschere manuali che si stanno utilizzando sono a due canali per permettere l'utilizzo della funzione di attivazione Softmax.

Difatti, a seguito di alcuni addestramenti con funzione di attivazione sigmoide e maschere single-channel, la rete non produceva risultati degni di nota. Questo probabilmente dipende dal fatto che l'utilizzo di maschere su 2 canali con una funzione di attivazione Softmax può consentire alla rete di catturare informazioni più complesse e contestuali per la segmentazione binaria. Con due canali, la rete può produrre una mappa di segmentazione per ciascuna

classe e considerare il contesto delle due classi contemporaneamente. Questo può migliorare la capacità della rete di discriminare tra le classi e gestire meglio situazioni di ambiguità. Questa scelta è stata mantenuta per tutte e tre i modelli considerati.

Le maschere a 2 canali sono state ottenute in modo che sul primo canale avrò i pixel appartenenti all'ago settati a 1 mentre quelli del background a 0, mentre sul secondo canale si ha la versione complementare (pixel ago a 0 e sfondo a 1).

Viene riportata di seguito una maschera a due canali a titolo di esempio estratta da un volume OCT. Lo stesso procedimento è stato applicato anche per le maschere PA.

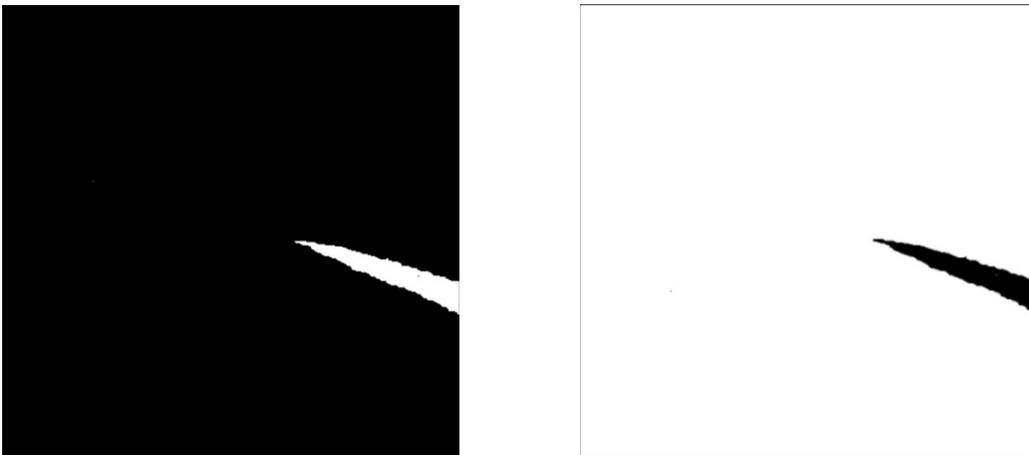


Figura 3.9: A sinistra CH0: pixel ago a 1 e sfondo 0. A destra CH1: pixel ago a 0 e sfondo a 1

3.1.2 OCT model

La rete importata da Monai lavora sui dizionari. Vengono quindi importate le immagini di training e validation e le rispettive maschere manuali, e caricate nel dizionario con chiavi 'image' e 'label'. Successivamente è stato aggiunto un canale lungo la prima dimensione al fine di rendere le immagini di tipo 'channel-first'.

Per migliorare le performance della rete, sul training set è stato eseguito il Data Augmentation mediante alcune trasformazioni Random presenti in Monai. La Data Augmentation è un insieme di trasformazioni applicabili ai dati del training set, il cui scopo è quello di aumentarne la variabilità, così da rendere la rete più robusta a determinate variazioni.

In particolare:

- RandFlipd: Esegue il flip lungo tutti e 3 gli assi, con una probabilità di 0.1 (ovvero c'è la probabilità del 10% di ottenere l'immagine flipata lungo l'asse specificato). Questo introduce ancora più casualità all'interno delle trasformazioni.
- RandRotate90d: Rotazioni random di 90° con una probabilità del 10 % (valore di default);
- RandShiftIntensityd : shift dell'intensità dell'immagine (non applicata alla label) con un offset di 0.10 e una probabilità del 50 %

Applicare queste tecniche di Data Augmentation permette sia di ottenere training set di dimensioni maggiori, sia di allenare la rete in modo tale che apprenda caratteristiche invarianti alle trasformazioni applicate. Quindi fare augmentation dei dati in ingresso, aiuta a ridurre l'overfitting.

Mentre per quanto riguarda il validation, non sono state applicate le trasformazioni random.

Al fine di ottenere un training ripetibile è stato impostato il set_determinism con SEED pari a 3. Il training della rete è stato condotto per 15 epoche, con batch size pari a 8. La funzione di loss utilizzata è la Dice Loss, con funzione di attivazione Softmax. L'ottimizzatore scelto è Adam con learning rate pari a 10^{-6} .

Durante l'allenamento, il Validation Set è utilizzato per evitare che ci sia overfitting nel modello. È stato scelto di valutarlo dopo ogni epoca per ottenere risultati più precisi. Per farlo, è stata eseguita la sliding window inference con batch pari a 8, che applica il modello, di volta in volta, ad una finestra dell'immagine di partenza, per poi scorrere su di essa e ripetere l'operazione. La grandezza della finestra è impostata pari a 144x144.

Una volta ottenuto l'output, vengono applicate le seguenti trasformazioni:

- Funzione di attivazione Softmax;
- Threshold imposta a 0.5 per assegnare la classe 0 ai pixel con valore inferiore alla soglia, classe 1 per valori superiori.

Viene calcolato il Mean Dice come metrica di validazione: se il valore risultante supera quello dell'epoca precedente, si salva il modello. Il risultato migliore

è stato ottenuto alla 14esima epoca.

Il tempo necessario ad effettuare il training è stato di 1 ora e 28 minuti.

3.1.3 PA model

Per quanto riguarda il modello addestrato su immagini PAT, vengono eseguite le stesse operazioni di trasformazione sul training e sul validation applicate nel modello OCT. Ciò che è stato modificato sono gli iperparametri.

In particolare, in questo caso il batch size è stato impostato a 8 sia nel training sia nel validation, con un numero di epoche pari a 30. Per quanto riguarda l'algoritmo di ottimizzazione è stato utilizzato Adam con learning rate pari a $10e^{-5}$.

Il risultato migliore è stato ottenuto alla 26esima epoca, valore oltre il quale la Dice Metric comincia a decrementare.

Il tempo necessario ad effettuare il training è stato di circa 40 minuti.

3.2 Inference

La procedura di inference consiste nel testare i modelli su nuovi dati, ovvero sui dati che compongono il Test set, al fine di valutare la capacità di generalizzazione.

La pipeline di inference prevede che l'input sia il volume da segmentare; da ciascun volume, vengono estratte di volta in volta le patch bidimensionali e applicato il pre processing definito in fase di training che comprende normalizzazione tra il 1° e il 99° percentile, il reshape e l'applicazione del kernel gaussiano. Viene poi aggiunto un canale lungo la prima dimensione al fine di ottenere la rappresentazione channel-first.

Su ciascuna di essa viene poi applicato il modello e, dopo l'inference, all'output viene applicata la funzione di attivazione Softmax e l'assegnazione di ogni pixel ad una classe mediante la soglia imposta a 0.5. Vengono poi ripristinate le dimensioni originali.

Tutte le patch estratte vengono poi raggruppate nuovamente per riformare il volume di segmentazione automatica, che viene poi salvato nel Drive con estensione .nrrd.

3.2.1 Pipeline di riduzione dei tempi computazionali

La riduzione dei tempi computazionali nella procedura di needle tracking è uno degli obiettivi principali insito nello sviluppo stesso di questa procedura. Difatti, in alcune applicazioni come la biopsia guidata da immagini, è fondamentale che questo avvenga in tempo reale.

Ridurre i tempi computazionali consente infatti di ottenere aggiornamenti immediati sulla collocazione dell'ago, fondamentale per guidare l'ago in modo preciso. Questo aspetto consente, di conseguenza, di accelerare il flusso di lavoro complessivo.

Diverse strategie possono essere adoperate al fine di ottenere un miglioramento dal punto di vista della tempistica in fase di predizione: quello che si vorrebbe ottenere è un algoritmo che consenta di processare più frame nel minor tempo possibile.

Alcune tecniche mediante le quali è possibile ridurre i tempi computazionali prevedono:

- La riduzione dell'immagine, mediante tecniche di ridimensionamento o campionamento: questo consentirà di ottenere immagini più piccole, ovvero un numero minore di pixel da elaborare.
- Estrazione delle Regioni di Interesse: se l'ago è infatti presente solo in una specifica regione dell'immagine, è possibile estrarre solo quella regione di interesse, al fine di ridurre l'area di calcolo.

- Parallelizzazione: si può sfruttare la capacità del calcolo in parallelo, mediante per esempio l'utilizzo della GPU, accelerando così il processo rispetto alla CPU tradizionale.
- Estrazione delle features del primo ordine.

Diversi accorgimenti tra quelli elencati sono stati messi in campo al fine di minimizzare il tempo necessario ad effettuare la predizione su ciascun volume. Per quanto riguarda i volumi OCT, la pipeline di pre-processing presuppone il ridimensionamento dell'immagine: questo step consente di ridurre l'area da processare passando da 512x896 pixel a 512x512. Il ridimensionamento in fase di inference, piuttosto che un crop, consente di preservare tutte le caratteristiche dell'immagine.

Questo è utile ai fini di generalizzazione dell'algoritmo in fase di testing: non si conosce la posizione precisa dell'ago all'interno dell'immagine, pertanto il crop avrebbe potuto rimuovere pixel dell'oggetto da riconoscere e segmentare. In secondo luogo, tutto l'algoritmo è stato implementato in Google Colab, una piattaforma in cloud offerta da Google, al fine di poter allocare la memoria GPU e sfruttarne l'accelerazione. Google Colab offre due modelli di GPU: Nvidia Tesla T4 [1.59GHz memory clock, 16Gbytes memory, 8.1 TFLOPS performance] e Nvidia K80 [0.82GHz memory clock, 12Gbytes memory, 4.1 TFLOPS performance], a seconda della disponibilità, che tuttavia risulta limitata nell'applicazione nella sua versione gratuita.

In particolare l'addestramento del modello e la fase di predizione sono stato svolti utilizzando il framework CUDA per l'utilizzo delle GPU, sfruttando l'architettura in parallelo.

Un' ultima strategia utilizzata in fase di inference al fine di diminuire il tempo necessario per la predizione, si basa sull'estrazione delle caratteristiche.

In particolare, le features del primo ordine rappresentano le statistiche di base calcolate direttamente dai valori dei pixel dell'immagine, senza considerarne le relazioni spaziali e strutturali più complesse.

Le features del primo ordine che possono essere estratte dalle immagini sono:

1. La media, il cui valore indica una misura dell'intensità media dell'immagine.
2. La deviazione standard (o la varianza) che indica la variazione o la dispersione dei valori di intensità nell'immagine.
3. La mediana, che rappresenta il valore centrale dell'intervallo di intensità dell'immagine.
4. I percentili, i quali indicano il valore sotto il quale cade una determinata percentuale dei valori di intensità.

- Intensità dei pixel, ovvero il valore numerico che indica la luminosità o il livello di colore di un pixel nell'immagine. La distribuzione viene rappresentata mediante l'istogramma.

Per quanto riguarda i volumi OCT, la pipeline di riduzione dei tempi di inferenza è riportata nella Figura 3.1:

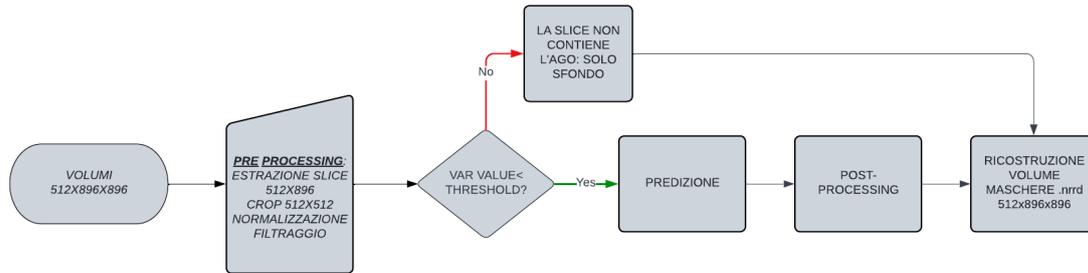


Figura 3.10: Pipeline di Inferenza sui volumi OCT

Essendo ciascun volume composto da 896 slice di cui solo una ridotta porzione contenente l'ago, ovvero la feature di interesse, si vuole limitare la predizione ad un volume di dimensioni minori così da diminuire il tempo necessario per la ricostruzione totale dello stesso.

Per fare ciò si sono osservati gli istogrammi delle luminosità delle due occorrenze:

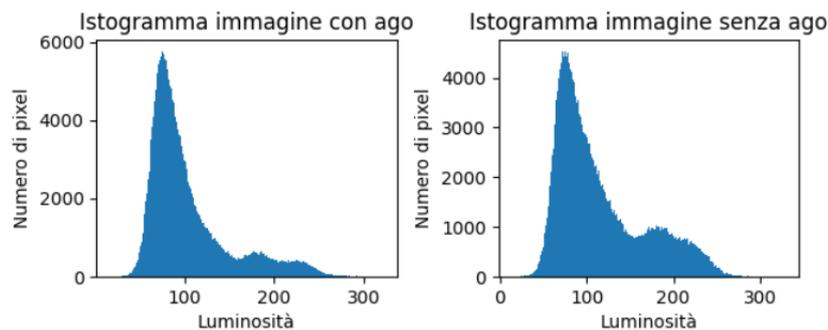


Figura 3.11: A sinistra: istogramma calcolato su immagine contenente ago. A destra: istogramma dell'immagine che non contiene l'ago

Poiché non si riscontrano evidenti differenze, tali da poter discriminare le due casistiche, si sono individuati elementi caratterizzanti nelle features del primo ordine, ovvero nel calcolo di parametri statistici: in particolare calcolando il valor medio e la varianza di ciascuna slice all'interno del volume. Calcolando i due parametri si è poi scelto di porre attenzione sulla varianza, in quanto il range di discriminazione, rispetto a quello che si crea usando il valor medio,

risulta maggiore.

Al fine di visualizzare e comprendere al meglio la strategia implementata, vengono riportate di seguito, a titolo di esempio, due slice estratte dal volume OCT N2P2, la prima riporta la feature di interesse, la seconda invece è solo background.

Immagine con ago

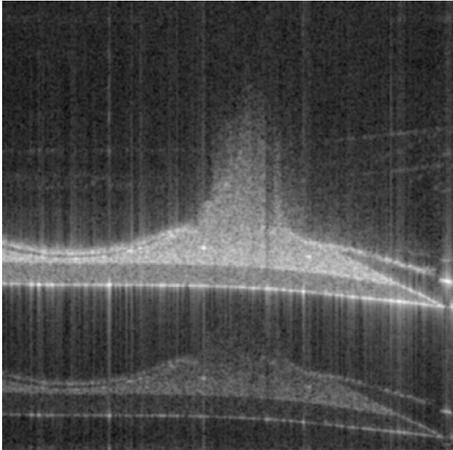


Immagine senza ago

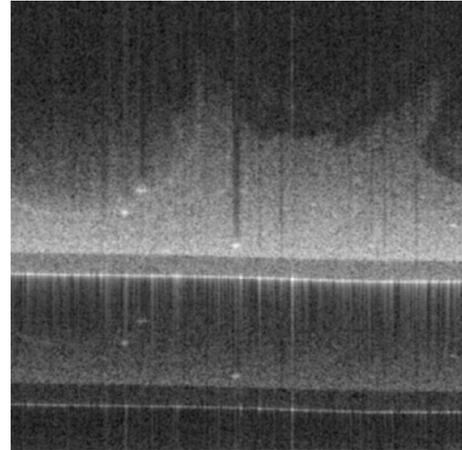


Figura 3.12: A sinistra: immagine estratta dal volume OCT con ago. A destra: immagine estratta dallo stesso volume solo sfondo

Calcolando la varianza delle due immagini risulta che:

- La prima immagine ha una varianza pari a circa 2000.
- La seconda immagine ha una varianza pari a circa 2570.

Questo trend di discesa della varianza nelle immagini in cui l'ago è presente viene confermato se si visualizza la Figura seguente che riporta l'andamento della varianza calcolato per tutto il volume OCT.

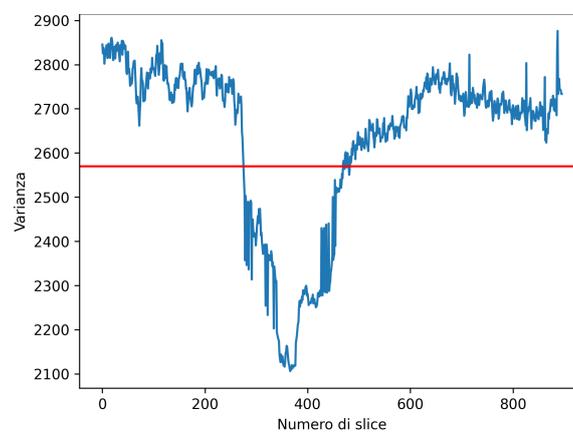


Figura 3.13: Andamento della varianza calcolato per ogni frame del volume OCT

Il picco verso il basso che si riscontra nel grafico rappresenta, appunto, la porzione di volume a varianza più bassa contenente l'ago. Questo cambiamento a livello di valore di varianza si può ricercare nel fatto che l'ago, inserendosi nel fantoccio crea una zona di ombra che ne aumenta la varianza. Infatti queste immagini risultano complessivamente più scure rispetto al caso in cui l'ago è assente.

Con questo approccio quindi, applicando una soglia di un opportuno valore, si potrebbe discriminare tra le due tipologie di immagini.

Il valore di soglia è stato ricavato effettuando una media della varianza calcolata sugli 896 frame contenuti in ogni volume. I 9 valori ricavati da ciascun volume vengono poi mediati sul totale degli stessi.

Tramite questo procedimento il valore di soglia è risultato circa pari a 2600, considerando anche un certo margine di errore.

Durante la fase di predizione, le sole slice che rispettano tale criterio verranno selezionate e ad esse verrà applicato il modello. In tal modo, i tempi di predizione, risultano notevolmente ridotti.

Viene riportato nella tabella sottostante un confronto in termini di tempo necessario per processare ciascun volume, per intero senza pre processing, quando viene effettuato solo il pre processing e quando viene utilizzata la strategia della varianza.

Tabella 3.1: Tempi di predizione Volumi OCT

Nome_Volume	No Pre Processing	Pre Processing	Pre Processing e Var
N2P1	134.02s	45.48s	30.70s
N2P2	130.44s	45.29s	30.57s
N2P3	137.64s	45.36s	29.49s
N6P2	132.64s	44.83s	32.07s
N6P3	130.51s	45.32s	32.8s
N6P4	131.02s	53.48s	33.4s
N7P1	131.16s	37.63s	20.53s
N7P2	131.79s	41.95s	24.21s
N7P3	132.33s	34.57s	22.92s

Osservando la tabella si nota come il crop sull'immagine permetta di effettuare una riduzione notevole dei tempi computazionali, che diminuisce ulteriormente quando si applica il criterio della varianza.

Per quanto riguarda invece i volumi PAT, la strategia implementata al fine di ridurre i tempi computazionali, concentrando la predizione su un sottoinsieme del volume complessivo, si basa sulla ricerca del contorno.

In particolare viene riportato il flow-chart riassuntivo dei passaggi svolti durante l'inferenza.

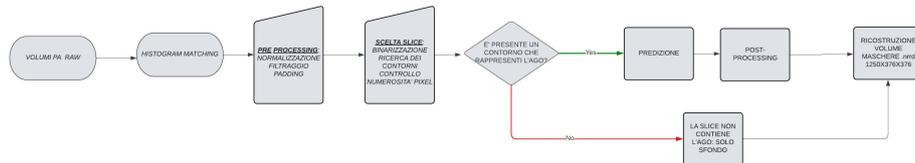


Figura 3.14: Pipeline di inferenza sui volumi PAT

Come è possibile osservare dalla figura, il volume sul quale avviene la predizione è quello ottenuto dopo aver eseguito l'Histogram Matching.

Dal volume vengono estratte le singole slice e applicato il pre processing definito in fase di training, al fine di ridurre il rumore dell'immagine.

L'algorithmo di scelta delle slice si compone di diversi step:

- L'immagine viene binarizzata con soglia automatica di OTSU.

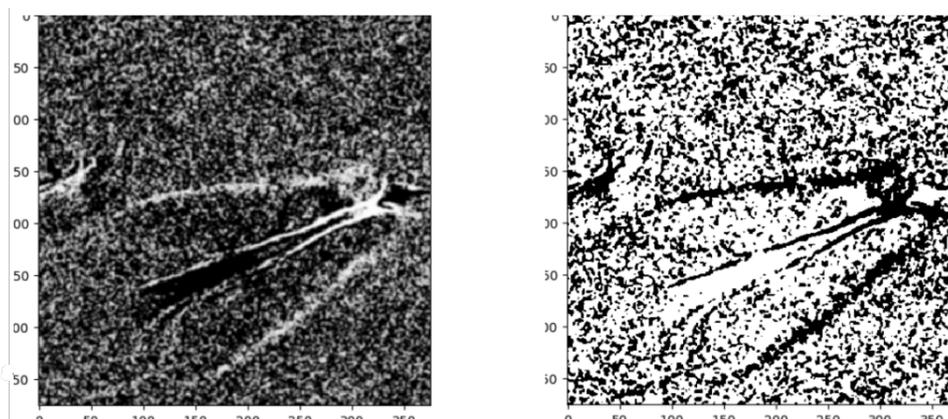


Figura 3.15: A sinistra: immagine estratta dal volume PAT dopo HM. A destra: immagine binarizzata

- Con la funzione Contour, della libreria skimage, vengono ricercati ed evidenziati nell'immagine binarizzata tutti i contorni in essa presenti.

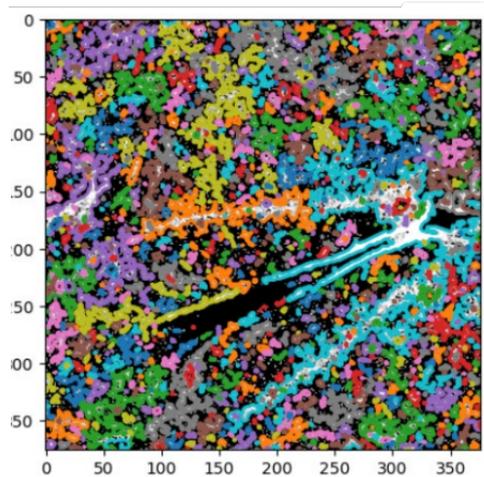


Figura 3.16: Ricerca dei contorni all'interno dell'immagine

- Al fine di ottenere e cercare il solo contorno dell'ago al fine di localizzarlo, si filtrano i contorni trovati in base alla numerosità dei pixel appartenenti ad essi mediante la funzione `ContourArea` della libreria `cv2`. Questa operazione permette di scartare i cluster meno numerosi.

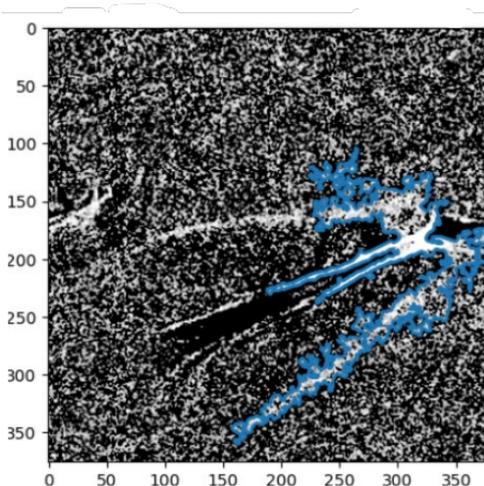


Figura 3.17: Contorno dell'ago trovato in base alla numerosità dei pixel

Vengono riportati di seguito i risultati in termini di tempo necessario per eseguire la predizione. In questo caso, non viene eseguito il crop dell'immagine in quanto si tratta di patch di dimensioni 288x288.

Tabella 3.2: Tempi di predizione Volumi PAT

Nome_Volume	No Pre Processing	Pre Processing e Contour
NeedleInGel_01s	11.02s	3.87s
NeedleInGel_12s	8.8s	7.56s
NeedleInGel_21s	7.9s	4.98s
NeedleInGel_42s	10.09s	3.62s
NeedleInGel_43s37	9.99s	5.04s
NeedleInGel_43s42	7.09s	3.77s
NeedleInBeef_10s	10.72s	1.048s
NeedleInBeef_33s	8.34s	4.95s
NeedleInMaxsPhantom_20s	10.18s	5.69s
NeedleInBeef_26s54	15.17s	6.33s
NeedleInBeef_24s44	9.59s	6.27s
NeedleInUSgel_22s	9.86s	4.21s
NeedleInMaxsPhantom_17s	9.73s	7.01s

Si nota, osservando i risultati, che in alcuni casi i tempi necessari ad effettuare la predizione sull'intero volume si sono dimezzati.

3.3 Post Processing

Ciascuna maschera di segmentazione automatica ottenuta subisce poi un processo di post-processing al fine di migliorare il risultato complessivo.

Per entrambi i modelli si è scelto di implementare gli stessi step di post processing, assegnando valori diversi ai parametri nel caso delle due tipologie di volume, ed è stato eseguito dopo la predizione, prima che le slice vengano inserite nel volume ricostruito.

Pertanto, è stato eseguito slice per slice ed implementato con le funzioni di `skimage.morphology`.

Il post-processing eseguito consiste nell'eliminazione di piccoli oggetti e nel riempimento dei buchi.

Per quanto riguarda i volumi OCT, in primo luogo, è stato effettuato lo step di rimozione degli oggetti di piccole dimensioni in base alla loro grandezza: si è scelto di eliminare, in base ad alcune prove effettuate e trovando il migliore compromesso, quelli con una dimensione inferiore a 500 pixel, dove i pixel sono considerati appartenenti allo stesso oggetto in base alla connettività che ne definisce il vicinato. Tale tecnica risulta essere molto efficace nella rimozione di discontinuità della segmentazione.

Successivamente è stata applicata la tecnica di post processing di chiusura, al fine di rifinire il profilo dell'ago. Tale tecnica è un operatore morfologico usato spesso nel processing dell'immagine per rimuovere piccoli buchi o spazi nelle immagini binarie e si compone di una dilatazione, seguita da un'erosione. La dilatazione ingrandisce l'immagine mentre l'erosione la rimpicciolisce, a seconda dell'elemento strutturale che verrà sovrapposto ad ogni pixel dell'immagine. In questo caso si è impostata la chiusura dei gaps con area inferiore a 40 pixel.

Viene riportata di seguito un slice estratta dal volume N6P4 a titolo di esempio, al fine di mostrare gli effetti del post-processing.

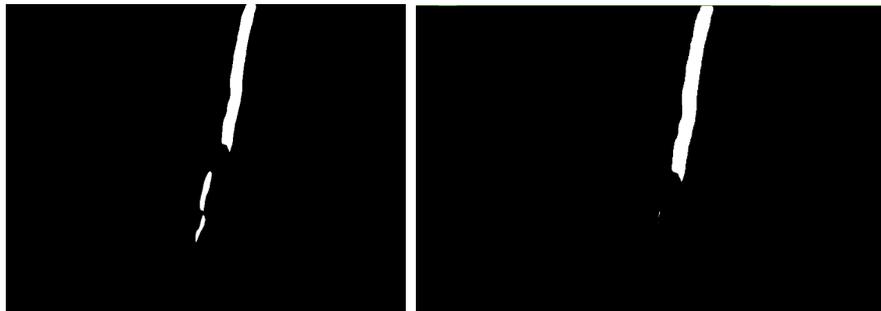


Figura 3.18: A destra: Slice estratta dal volume OCT N6P4 senza applicazione di post processing. A sinistra: Slice estratta dal volume OCT a cui sono state applicate le tecniche di post processing.

Osservando le figure a confronto è possibile notare come il post processing implementato abbia rimosso le discontinuità e i piccoli cluster di pixel isolati riconosciuti erroneamente come oggetto.

Nella tabella seguente vengono riportati i risultati ottenuti a seguito di applicazione delle tecniche post processing. In generale, la metrica Dice risulta aumentata, migliorando sensibilmente i risultati ottenuti.

OCT VOLUMES	DICE SENZA POST PROCESSING	DICE CON POST PROCESSING
N2P1	0.803	0.819
N2P2	0.821	0.850
N2P3	0.793	0.871
N6P2	0.722	0.726
N6P3	0.720	0.853
N6P4	0.689	0.955
N7P1	0.710	0.710
N7P2	0.614	0.633
N7P3	0.598	0.611

Tabella 3.3: Risultato dell'applicazione del post processing sui volumi OCT

Per quanto riguarda i volumi PAT, anche in questo caso è stata effettuata la rimozione degli oggetti più piccoli secondo cui vengono rimossi i gaps che presentano una dimensione inferiore a 800 pixel, mentre la chiusura è impostata per dimensioni inferiori ai 50 pixel.

Anche in questo caso viene riportata una slice estratta dal volume PAT, prima e dopo il post processing.



Figura 3.19: A sinistra: Slice estratta senza applicazione di Post Processing. A destra: Slice con applicazione di tecniche di Post Processing

Di seguito, invece, viene riportata la tabella che mostra i risultati complessivi:

PAT VOLUMES	DICE SENZA POST PROCESSING	DICE CON POST PROCESSING
NeedleInGel_01	0.84	0.86
NeedleInGel_12	0.85	0.89
NeedleInGel_21	0.88	0.91
NeedleInGel_42	0.79	0.81
NeedleInGel_43s37	0.89	0.92
NeedleInGel_43s42	0.90	0.93
NeedleInBeef_10	0.48	0.52
NeedleInBeef_33	0.61	0.65
NeedleInMaxsPhantom_20	0.47	0.51
NeedleInBeef_26s54	0.60	0.62
NeedleInBeef_26s44	0.44	0.51
NeedleInMaxsPhantom_17	0.50	0.61
NeedleInUSgel_22	0.55	0.60

Tabella 3.4: Risultato dell'applicazione del post processing sui volumi PAT

Anche in questo caso, la metrica risulta notevolmente aumentata, salvo qualche eccezione su alcuni volumi il cui coefficiente Dice presentava valori elevati anche senza applicazione del post processing.

3.4 Metriche di valutazione e loss

Gli algoritmi di Deep Learning utilizzano un approccio di discesa del gradiente stocastica per ottimizzare ed apprendere l'obiettivo.

Le funzioni di Loss vengono utilizzate per determinare l'errore tra l'output dell'algoritmo e il valore target specificato, pertanto misurano quanto bene la rete neurale modella i dati di addestramento. Durante l'allenamento, si mira a ridurre al minimo questa perdita tra gli output previsti e quelli target. [24]. Una volta definita la funzione di loss, l'algoritmo di ottimizzazione minimizza il suo valore e questo avviene mediante il 'backpropagation' dell'errore della rete, aggiornando i pesi moltiplicando la derivate della funzione loss rispetto alla derivata di ogni peso per il learning rate e sottraendo questo valore dai pesi calcolati allo step precedente.

Questa operazione è ripetuta per un certo numero di operazioni fino a che non si incontra un criterio di stop.

Pertanto, funzioni di perdita appropriate possono rafforzare il processo di apprendimento e ottenere buoni risultati di segmentazione[25]. La funzione di perdita prevalentemente utilizzata per la segmentazione delle immagini mediche è la Dice Loss, di tipo 'region based' per risolvere i problemi di sbilanciamento delle classi nella segmentazione medica. Essa è stata formulata appositamente per la segmentazione degli oggetti in base al coefficiente Dice (DSC).

Nelle immagini mediche, è comune che la regione di interesse occupi solo un'area molto piccola della scansione. Ciò può far sì che il processo di apprendimento rimanga intrappolato in minimi locali della funzione di perdita e quindi produrre una rete le cui previsioni sono fortemente orientate verso lo sfondo e le regioni in primo piano vengono rilevate solo parzialmente o addirittura mancate[26]. Pertanto, la DiceLoss, che è una quantità compresa tra 0 e 1, è stata proposta.

La DSC tra due volumi binari può essere descritta come:

$$DiceLoss = 1 - DSC = 1 - (2 * |P \cap G|) / (|P| + |G|)$$

Scopo di questo progetto è la segmentazione dell'ago, pertanto è opportuno valutare la classificazione di ciascun pixel come oggetto o sfondo nelle segmentazioni automatiche ottenute.

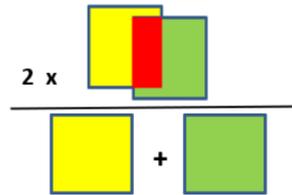
Questo step è fondamentale al fine di validare la scelta degli iperparametri e l'architettura stessa della rete per l'obiettivo preposto. Le performance della rete sono state valutate con diverse metriche: il Mean Dice Coefficient, la Precision, la Recall per la segmentazione, mentre l'Housdorff Distance per i contorni.

Il *Mean Dice Coefficient* è un indice di sovrapposizione tra la predizione del modello e il Ground Truth. Si calcola come il numero degli elementi in comune (veri positivi) moltiplicato per due e diviso per il numero di elementi di

ogni set :

$$DSC = (2 * |P \cap G|) / (|P| + |G|)$$

Graficamente rappresentato nella Figura:



Il suo valore può essere tra 0 e 1, dove 1 indica la perfetta sovrapposizione tra la segmentazione automatica e quella manuale, mentre 0 la mancanza di sovrapposizione.

In questo lavoro sia la DiceLoss sia il Mean Dice durante la fase di training sono stati calcolati mediante le librerie di MONAI.

La *Precision* è il numero di pixel veri positivi (TP), ovvero i pixel appartenenti alla classe ago e predetti come tali, diviso il numero totale di veri positivi (TP) e falsi positivi (FP), ovvero pixel appartenenti allo sfondo ma predetti come oggetto.

Quindi misura la proporzione dei pixel identificati correttamente rispetto a tutti i pixel che sono stati identificati come oggetto.

$$Precision = TP / (TP + FP)$$

La *Recall* è invece il numero di pixel veri positivi TP divisi per il numero totale di veri positivi e falsi negativi FN, ovvero pixel appartenenti all'ago ma classificati come sfondo. Misura quanto bene il modello identifica le istanze dell'oggetto di interesse.

$$Recall = TP / (TP + FN)$$

Infine, la *Hausdorff Distance*, è una metrica che quantifica la differenza tra due insiemi di punti nello spazio. Viene definita come la massima distanza tra un punto in un insieme e il punto più vicino nell'altro insieme.

Tale metrica è utilizzata al fine di confrontare la bontà della segmentazione confrontando i contorni estratti dall'immagine segmentata con i contorni della maschera di riferimento.[27]

Per i calcolo di queste metriche è stata utilizzata la repository `surface_distance`

¹ per il calcolo della Hausdorff Distance e la libreria seg-metrics di PyPi per le metriche Dice, Precision, Recall [28] calcolate per il testing del modello. Le metriche sono state calcolate sui volumi ricostruiti.

¹<https://github.com/deepmind/surface-distance>

Capitolo 4

Risultati

Nel capitolo seguente vengono illustrati i risultati ottenuti in questo lavoro di tesi al fine di presentare una valutazione complessiva delle prestazioni dell'architettura di rete precedentemente descritta e dell'elaborazione dei dati effettuata.

In particolare, verranno confrontati i due modelli addestrati, quello per OCT e quello PAT, in base alla funzione di Loss e alle metriche ottenute.

Verranno in primis mostrati i risultati ottenuti sul modello per dati OCT, considerando il set di dati descritto nella sezione 2.1. I 9 volumi sono stati divisi in training set, validation set e test set. Nello specifico, si è scelto di utilizzare i volumi N7P1, N7P2, N7P3 per il test. I restanti volumi sono quindi stati divisi in 80% training e 20% validation. Pertanto, sono state estratte complessivamente 8064 immagini, di cui 4303 per il training set, 1075 immagini per il validation set e 2688 immagini per il test set.

Si è scelto di avere un test set più ampio poichè, data la ripetitività delle caratteristiche delle immagini, non sarebbe stato necessario un training set di dimensioni maggiori, a fronte di una maggiore di variabilità del test set che consenti di capire quanto il modello sia in grado di generalizzare i dati non visti.

Di fatti, l'addestramento ha previsto anche l'applicazione di Data Augmentation sul training set, consentendo di aumentare artificialmente le dimensioni del dataset mediante l'applicazione di trasformazioni, quali rotazioni e flip casuali, alle immagini originali.

Questo è stato possibile perchè non influiscono sulla morfologia della struttura dell'ago ma, anzi, consentono di esplorare una possibilità maggiore di modalità di inserzione, al fine di migliorare la robustezza del modello e prevenire l'overfitting.

Seguono i risultati ottenuti sul modello per dati PAT, considerando il set di dati descritto nella sezione 2.2. In questo caso il dataset comprende 13 volumi, per un totale di 11750 immagini. Di questi, 6 volumi sono stati utilizzati per il training set, 3 per il validation e 4 per il test. In questo caso, la porzione interessata dall'inserzione dell'ago è pari a circa 1/3 del volume totale. Pertanto, in fase di addestramento, per evitare l'overfitting sulla classe 0, alcuni volumi sono stati dimezzati, rimuovendo le immagini che presentassero solo sfondo. Per il test set, invece si mantengono i volumi nella loro interezza al fine di verificare il corretto funzionamento dell'algoritmo. Si sono ottenute

2802 immagini per il training set, 830 immagini per il validation set e 3800 immagini per il test set, seguendo le stesse considerazioni fatte per il dataset OCT.

In fase di inference, le strategie implementate e illustrate nella sezione al fine di ridurre il volume su cui avviene la predizione si sono dimostrate efficaci nell'obiettivo di diminuire il tempo computazionale e rendere più rapida la predizione sull'intero volume, nell'ottica di un utilizzo in real time.

In modo particolare, la segmentazione dei volumi OCT richiedeva tempistiche elevate a fronte della dimensione delle immagini e della complessità delle stesse, pertanto il ridimensionamento e la selezione delle sole slice utili ha permesso di diminuire il tempo necessario ad eseguire il processo, passando da una media di 130 secondi per ciascun volume a circa 30 secondi.

Le prestazioni dei modelli sono state poi valutate con le metriche Dice, Precision, Recall e Hausdorff Distance.

Gli esperimenti sono stati condotti utilizzando diversi iperparametri e funzioni di loss al fine di trovare la configurazione ottimale per ciascun modello.

In particolare, vengono mostrati i risultati ottenuti con la funzione di perdita Dice Loss descritta nella Sezione 3.4 nella sua configurazione più performante. Per l'addestramento della ResNet-34 sui dati OCT il numero massimo di epoche è stato impostato a 15. Il tasso di apprendimento è stato impostato a 10^{-6} con l'ottimizzatore Adam, che lo modifica durante l'addestramento, essendo meno sensibile al valore iniziale. La batch size è stata impostata a 8, sia sul training set sia sul validation set. Nella Figura 4.1 sono mostrate la curva Dice Loss sul training set e la curva del coefficiente Mean Dice calcolato sul validation set: l'algoritmo salva i parametri ottenuti nell'epoca su cui si riscontra il valore migliore di Dice. Mentre le Tabelle 4.1 e 4.2 mostrano le metriche di valutazione descritte nella Sezione 3.4 sulla base dei risultati ottenuti.

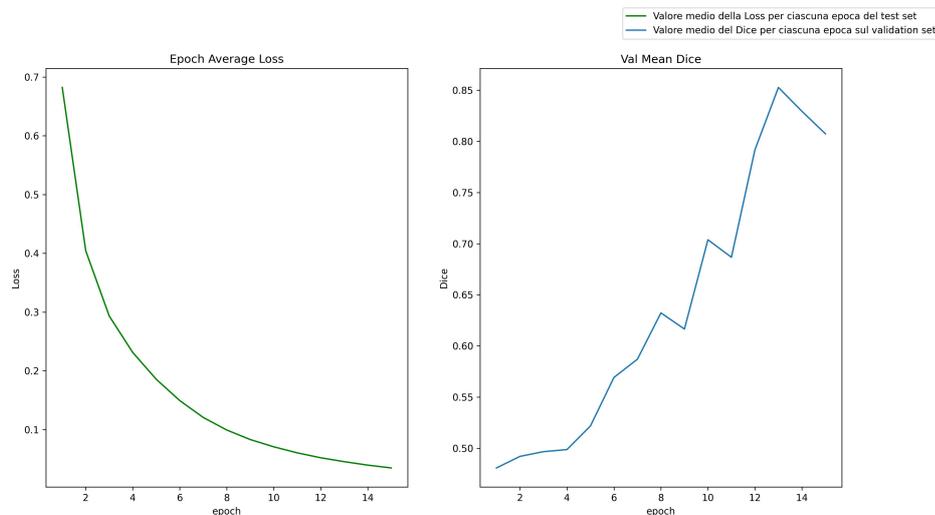


Figura 4.1: Prestazioni della rete OCT. A destra: valore medio della loss calcolata per ciascuna epoca. A sinistra: valore medio del Dice calcolato per ogni epoca sul set di validazione.

Il risultato migliore è stato ottenuto alla 14esima epoca, ed è stato salvato. Il training oltre quell'epoca decrementa la loss ma le prestazioni sul validation peggiorano.

Il valore medio del Dice in fase di validazione è in linea con il coefficiente Dice calcolato per ciascun volume e riportato nella tabella seguente. Le metriche mostrate sono calcolate sulla classe '1', ovvero sull'oggetto ago, sia per il set di training sia per quello di test.

TRAINING SET	DICE SCORE	PRECISION	RECALL	HD
N2P3	0.87	0.87	0.95	24.59
N6P2	0.73	0.82	0.92	567.99
N6P3	0.85	0.89	0.73	54.27
N6P4	0.95	0.89	0.62	75.69

Tabella 4.1: Prestazioni calcolate sul training set del dataset OCT

VALIDATION SET	DICE SCORE	PRECISION	RECALL	HD
N2P1	0.82	0.91	0.89	224.47
N2P2	0.85	0.94	0.87	22.29

Tabella 4.2: Prestazioni calcolate sul validation set del dataset OCT

TEST SET	DICE SCORE	PRECISION	RECALL	HD
N7P1	0.71	0.85	0.84	14.07
N7P2	0.63	0.77	0.62	518.60
N7P3	0.61	0.80	0.56	261.49

Tabella 4.3: Prestazioni calcolate sul test set del dataset OCT

Per quanto riguarda il modello addestrato sui dati PAT, il numero massimo di epoche è stato impostato a 30, con learning rate pari a 10^{-5} . Il risultato migliore è stato ottenuto all'epoca 26. Il valore della loss non decrementava ulteriormente oltre l'epoca 20, difatti anche le variazioni sul dice risultano minime.

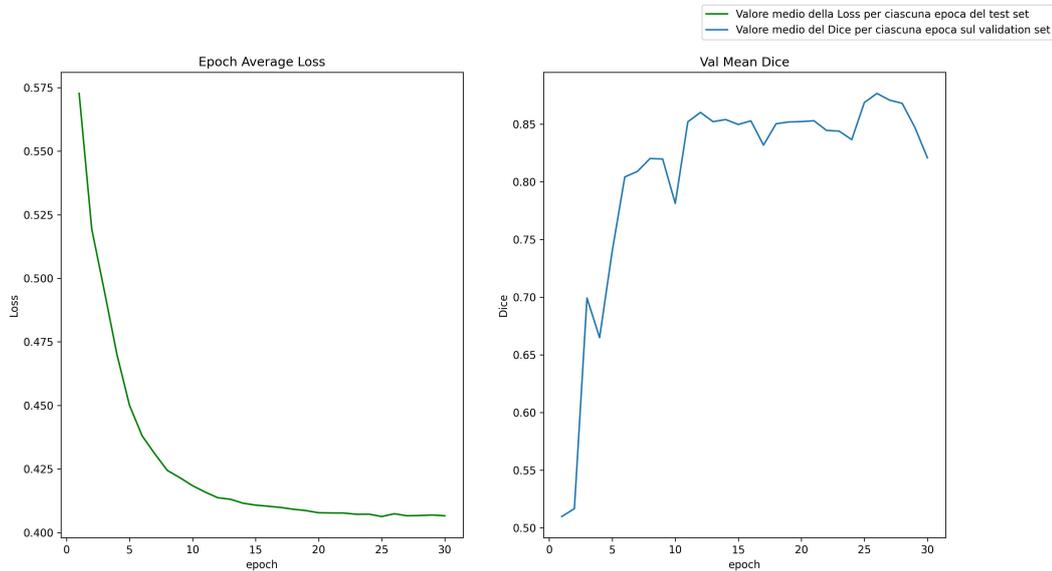


Figura 4.2: Prestazioni della rete PAT

Vengono riportate di seguito le metriche calcolate sul training set, sul validation set e sul test set del dataset PAT su ciascun volume; anche in questo le valutazioni fanno riferimento alla classe '1' ago.

TRAINING SET	DICE SCORE	PRECISION	RECALL	HD
NeedleInGel_01	0.86	0.76	0.97	666.07
NeedleInGel_12	0.89	0.85	0.92	715.85
NeedleInGel_21	0.91	0.87	0.95	597.99
NeedleInGel_43s37	0.92	0.88	0.95	656.26
NeedleInGel_43s42	0.93	0.96	0.90	381.43
NeedleInBeef_10	0.52	0.78	0.41	218.03

Tabella 4.4: Performance calcolate sul training set del dataset PAT

VALIDATION SET	DICE SCORE	PRECISION	RECALL	HD
NeedleInGel_42	0.81	0.71	0.96	388.367
NeedleInBeef_33	0.65	0.87	0.52	33.73
NeedleInMaxsPhantom_20	0.51	0.57	0.45	95.93

Tabella 4.5: Performance calcolate sul validation set del dataset PAT

TEST SET	DICE SCORE	PRECISION	RECALL	HD
NeedleInBeef_26s54	0.62	0.85	0.48	46.08
NeedleInBeef_26s44	0.51	0.84	0.32	62.78
NeedleInMaxsPhantom_17	0.61	0.67	0.43	80.84
NeedleInUSgel_22	0.60	0.74	0.49	61.67

Tabella 4.6: Performance calcolate sul test set del dataset PA

Le Tabelle 4.7, 4.8 e 4.9 riportano invece il valore medio e la deviazione standard calcolati per ciascuna metrica e per ciascun set al fine di poter confrontare i due modelli sui dataset.

Si nota come, sia per i volumi OCT sia per i volumi PAT utilizzati in fase di addestramento, le metriche calcolate siano elevate, superando l'80% in più della metà dei casi. Questo indica che il modello è in grado di adattarsi e di apprendere i pattern presenti nei dati di addestramento. Il dato della Hausdorff Distance presenta tuttavia una deviazione standard elevata, indice di una alta variabilità. Questo si deve alla presenza di alcuni volumi con valori outliers. Questo potrebbe essere dovuto al fatto che la distanza di Hausdorff tiene conto non solo della sovrapposizione tra maschera automatica e manuale, ma anche della loro forma e geometria, che potrebbe influire sul valore della distanza di Hausdorff.

Le performance calcolate sul validation set si dimostrano coerenti con quanto detto per il training set. Questi valori permettono di fornire una stima sulla capacità di generalizzazione del modello e sull'efficacia del modello su dati che non sono stati utilizzati durante l'addestramento.

Per quanto riguarda le performance sul test set, le metriche di sovrapposizione si attestano mediamente sul 60% per il coefficiente Dice, 80% per la Precision, circa 55% per il Recall. Ciò che mostra una maggiore discrepanza è il valore di Hausdorff Distance tra i due dataset, difatti i dati OCT mostrano una maggiore variabilità mentre i dati PAT riportano un valore minore, indice di una maggiore confrontabilità tra i contorni delle segmentazioni delle maschere automatiche e delle maschere manuali.

Dati i valori delle metriche ottenute sul validation set di entrambi i dati, un decremento delle performance sul test set potrebbe ricercarsi in cause che non dipendano dall'overfitting, quanto piuttosto dalle caratteristiche delle immagini stesse. In particolare, i volumi OCT del test set differiscono dai volumi utilizzati nel training set e validation set per via della posizione in cui è stato inserito l'ago: difatti, mentre nei volumi utilizzati per l'addestramento e la validazione l'inserzione avveniva lateralmente, localizzata quindi in un'estremità dell'immagine, nei volumi del test set l'ago viene inserito centralmente, sviluppandosi per tutta la grandezza dell'immagine.

Per quanto riguarda il dataset PAT, invece, i volumi utilizzati nel test set

risultano più rumorosi rispetto a quelli utilizzati nell'addestramento della rete. Le tecniche di pre processing implementate non sono risultati sufficienti a migliorare le performance una volta applicato il modello.

Tabella 4.7: Valore medio e deviazione standard delle performance calcolate sul Training set di ciascun dataset

Training Set	Dice	Precision	Recall	Hausdorff Distance
OCT	0.85 ± 0.09	0.87 ± 0.03	0.81 ± 0.15	180.63 ± 259.08
PAT	0.84 ± 0.15	0.85 ± 0.07	0.85 ± 0.21	539.27 ± 196.21

Tabella 4.8: Valore medio e deviazione standard delle performance calcolate sul Validation set di ciascun dataset

Validation Set	Dice	Precision	Recall	Hausdorff Distance
OCT	0.83 ± 0.02	0.92 ± 0.02	0.88 ± 0.02	123.38 ± 142.35
PAT	0.67 ± 0.15	0.72 ± 0.15	0.65 ± 0.27	172.67 ± 189.36

Tabella 4.9: Valore medio e deviazione standard delle performance calcolate sui test set di ciascun dataset

Test Set	Dice	Precision	Recall	Hausdorff Distance
OCT	0.65 ± 0.05	0.80 ± 0.04	0.67 ± 0.14	264.72 ± 252.24
PAT	0.58 ± 0.05	0.77 ± 0.08	0.43 ± 0.07	62.84 ± 14.21

Le Figure di seguito riportano alcuni esempi di predizione dei due modelli applicati su slice estratte dai 3 volumi di test del dataset OCT (Figura 4.3) e sui 4 volumi di test del dataset PAT (Figura 4.4).

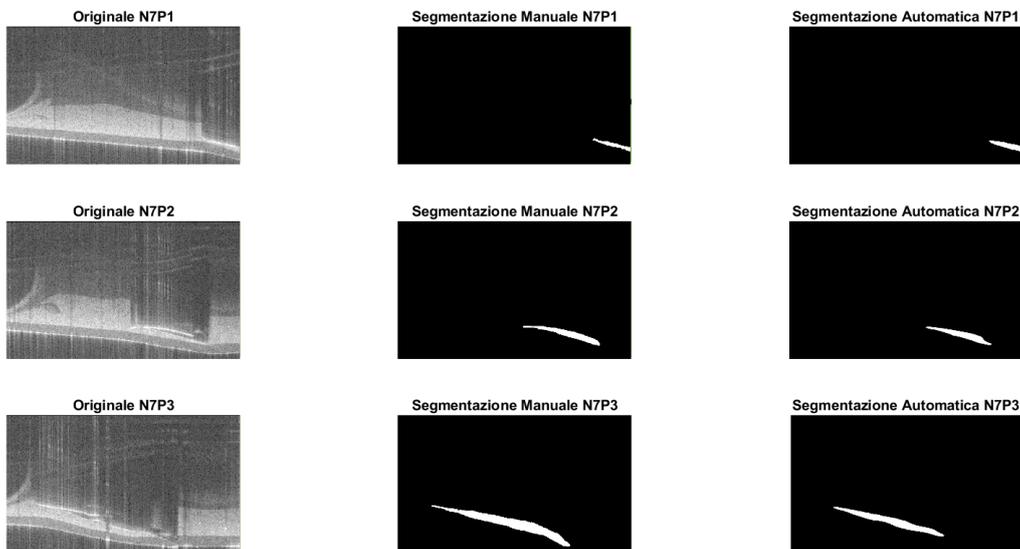


Figura 4.3: TEST OCT. Da sinistra verso destra: Immagine grayscale estratta dal volume OCT, Segmentazione Manuale ottenuta con Slicer3D, Segmentazione Automatica ottenuta con la rete neurale.

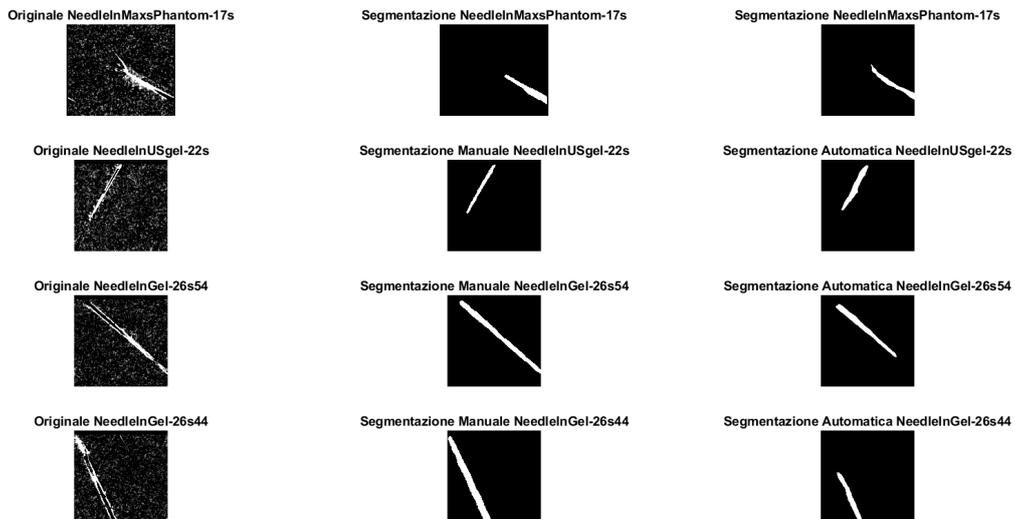


Figura 4.4: TEST PAT Da sinistra verso destra: Immagine grayscale estratta dal volume PAT, Segmentazione Manuale ottenuta con Slicer3D, Segmentazione Automatica ottenuta con la rete neurale.

Infine, volendo ampliare l'utilizzo di un singolo modello su diverse modalità di imaging, il modello OCT è stato testato anche con le MIP dei volumi combinati needle+vessels PAT illustrati nella Sezione 2.2.

Il calcolo della Maximum Intensity Projection lungo l'asse di profondità di tali volumi ha permesso di ottenere una visualizzazione più chiara dell'immagine risultante, in quanto si discriminano efficacemente le strutture di interesse (ovvero ago e vasi). Tali immagini risultano essere promettenti per l'utilizzo dell'algoritmo, come confermano le metriche ricavate sulle immagini a disposizione, coerenti con quelle ottenute sul test set dei volumi OCT.

Tabella 4.10: Valore medio e deviazione standard delle performance calcolate sulle MIP

	Dice	Precision	Recall	Hausdorff Distance
MIP Images	0.63 ± 0.08	0.67 ± 0.15	0.61 ± 0.12	54.86 ± 35.08

Questo probabilmente si deve al fatto che, osservando l'istogramma dell'immagine MIP del volume 'merged', esso risulti morfologicamente simile all'istogramma delle immagini estratte dai volumi OCT in fase di training della rete, come mostrato di seguito.

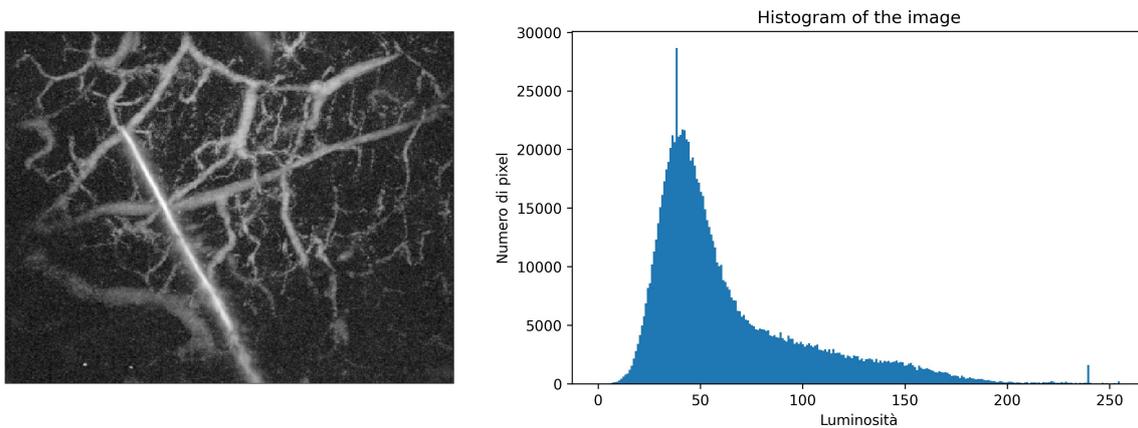


Figura 4.5: MIP calcolata su volume PA needle+vessels e relativo istogramma

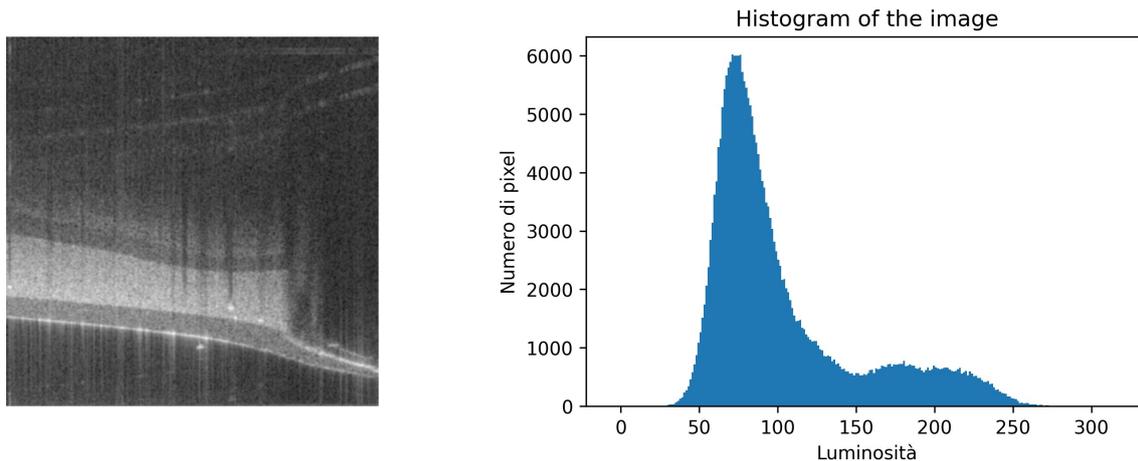


Figura 4.6: Slice estratta da volume OCT utilizzata per il training e relativo istogramma

Di contro, le immagini dei volumi PAT 'needle', utilizzate nel training del secondo modello non risultavano performanti in quanto il loro istogramma delle immagini ha una morfologia completamente non adattabile, anche a fronte dell'applicazione di tecniche di modifica dello stesso quale per esempio l'Histogram Matching.

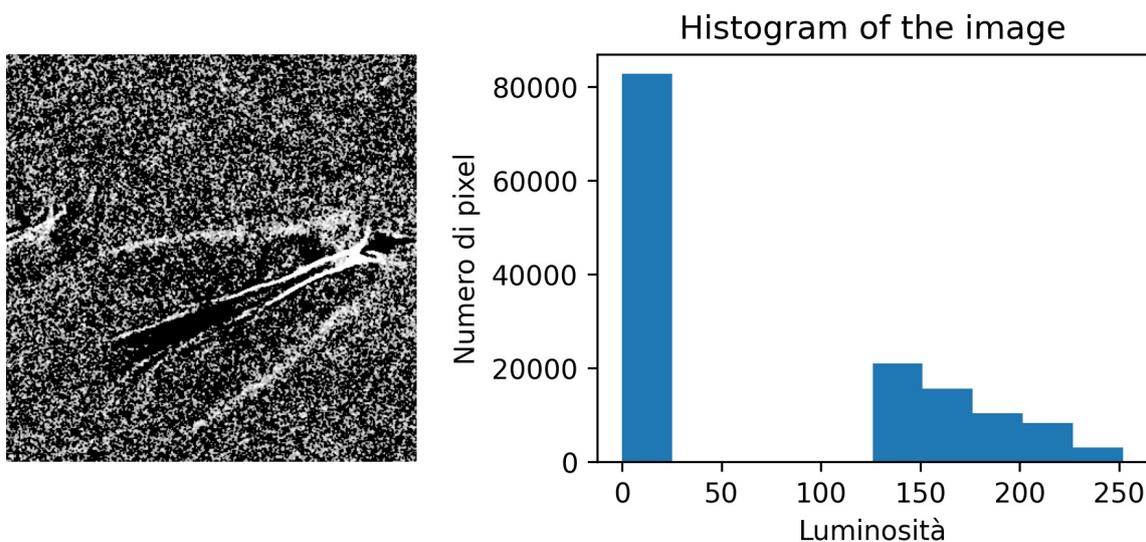


Figura 4.7: Immagine estratta dal volume PA 'needle' DOPO Histogram Matching

Questo non consente all'algoritmo di riconoscimento di individuare, nella maggioranza dei casi, l'ago nell'immagine. Di seguito vengono riportate le predizioni effettuate con modello OCT sulle MIP dei volumi PAT.

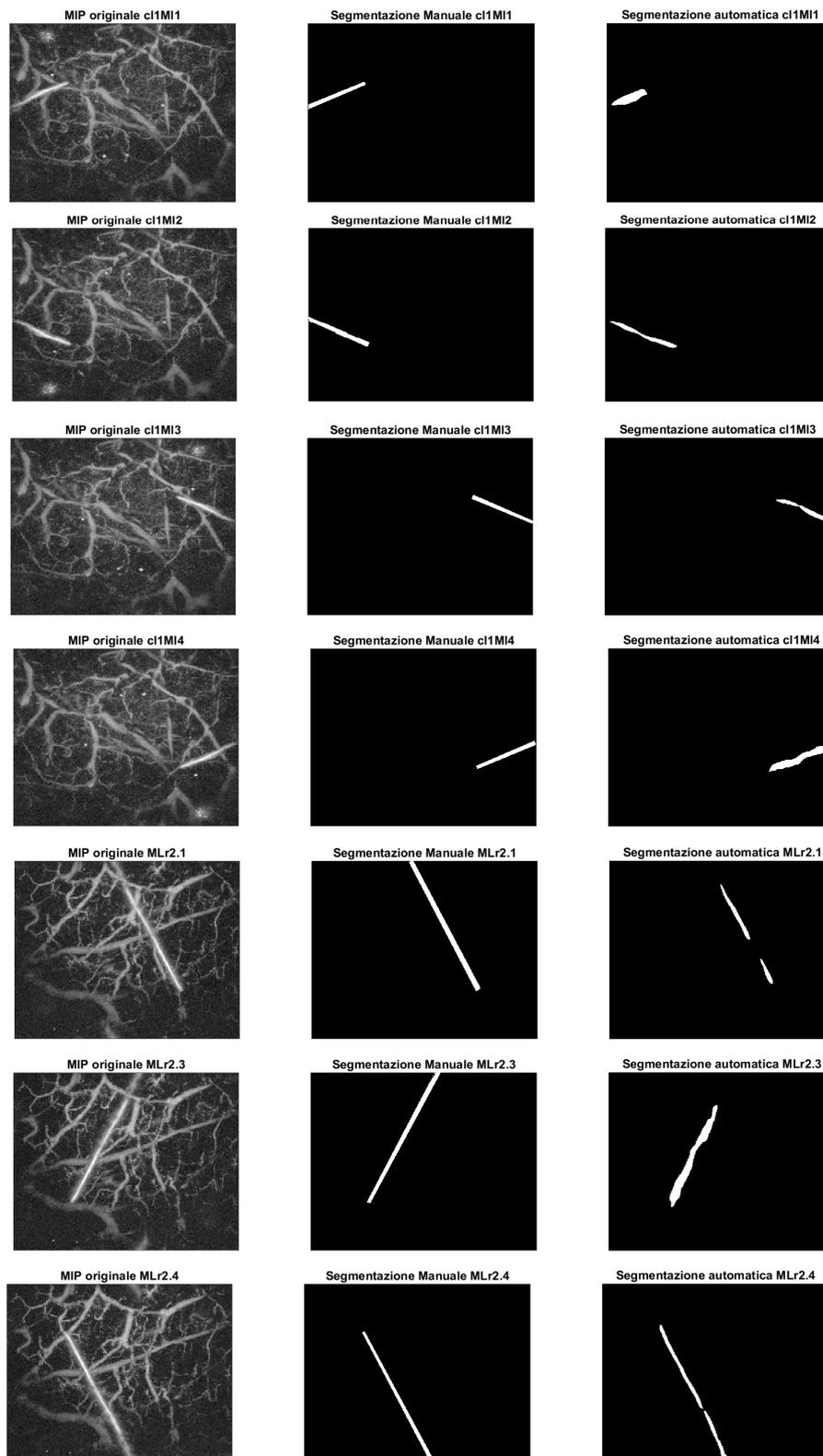


Figura 4.8: Da sinistra verso destra: Immagine MIP calcolata dal volume PAT needle+vessels, Segmentazione Manuale ottenuta con Slicer3D, Segmentazione Automatica ottenuta con la rete neurale

Capitolo 5

Conclusione

L'identificazione accurata dell'ago è una sfida chiave in interventi percutanei, biopsie, chirurgia fetale e medicina cardiovascolare.

In queste procedure, la guida mediante immagini è utilizzata di routine nella pratica clinica come mezzo di supporto visivo per le procedure percutanee basate su aghi, come la somministrazione di anestesia locale, biopsie tumorali e accesso venoso centrale in cardiologia e medicina fetale. Per tale motivo, risulta cruciale l'identificazione efficiente e accurata dell'ago nelle immagini, in modo particolare in situazioni complesse, quando la punta dell'ago è a grandi profondità, quando l'angolo di inserimento è ripido e quando il diametro dell'ago è piccolo, in quanto può comportare il danneggiamento di una struttura critica.

D'altra parte, il tracciamento dell'ago contribuisce a migliorare la sicurezza del paziente, riducendone i rischi.

In questo lavoro di tesi, è sviluppato un sistema di tracciamento dell'ago basato su un algoritmo di Deep Learning con l'obiettivo di fornire segmentazioni accurate, e quindi il rilevamento, dell'ago basato su imaging ottico.

In modo particolare, i dataset utilizzati comprendono volumi di Tomografie a Coerenza Ottica e volumi di Tomografie Fotoacustiche, da cui sono state estratte le immagini bidimensionali opportunamente elaborate con tecniche di *Image Enhance*, al fine di utilizzarle per l'addestramento della rete neurale Resnet-34.

Nonostante si tratti di un problema di segmentazione binaria, i risultati mostrano che l'architettura risulta performante efficacemente quando si utilizza la funzione di attivazione Softmax e il ground truth formato da due canali speculari. Questo permette di aumentare l'apprendimento, a differenza dell'utilizzo della funzione di attivazione sigmoide, la quale non produceva miglioramenti in termini di prestazioni durante il processo di addestramento. Essendo i volumi di dimensioni elevate, al fine di ridurre i tempi necessari a produrre la predizione su ciascuno di essi, sono state adottate diverse tecniche.

In primis, l'algoritmo è stato implementato sfruttando la GPU Nvidia Tesla T4 offerta da Google Colab al fine di parallelizzare le operazioni e velocizzarle. Successivamente, in fase di inference le immagini sono state ridimensionate così da ridurre il numero di pixel da elaborare e infine sono state proposte due

tecniche, per ciascun tipo di dataset, per selezionare in modo più puntuale la porzione delle slice che presentano la feature di interesse. Difatti le immagini all'interno del volume interessate dalla presenza dell'ago si presentano in numero ridotto (circa $1/3$) rispetto a quelle che presentano solo sfondo. L'applicazione del modello è quindi avvenuta solo su un sottoinsieme selezionato di slice, scelte in base al calcolo della varianza per quanto riguarda i volumi OCT e alla ricerca dei contorni per quanto riguarda i volumi PAT. Tali accorgimenti sono risultati efficienti al fine di diminuire i tempi computazionali, che nella maggior parte dei volumi risultano dimezzati.

Per quanto riguarda le metriche di valutazione, i due modelli raggiungono risultati simili in termini di valore di Dice, Precisione e Accuratezza sia sul set di testing sia su quello di training, a fronte invece dell'indice Hausdorff Distance, che misura la massima distanza tra due punti, uno nell'immagine predetta dalla rete e l'altro nell'immagine di riferimento, che risulta essere inferiore quando calcolato sui volumi PAT di test.

I valori elevati ottenuti sugli altri dati sono riconducibili alla dipendenza di tale indice dalla geometria e dalla forma delle regioni di segmentazione per cui anche una piccola discrepanza tra di esse può portare a valori elevati della distanza di Hausdorff, anche se le metriche come l'indice Dice e la precisione indicano una buona sovrapposizione generale.

Infine, il modello sviluppato su OCT è stato testato anche su immagini MIP del dataset PAT, con l'obiettivo di estendere l'applicazione di esso su modalità di imaging differenti.

Le prestazioni calcolate sulle immagini a disposizione risultano essere promettenti in quanto producono risultati confrontabili con quelli ottenuti per i dati OCT. Tuttavia, condizione fondamentale affinché questo sia possibile è che le immagini abbiano caratteristiche simili a quelle utilizzate per il training, per tale motivo non è stato possibile testare il modello OCT sul dataset PAT che presenta invece solo l'ago.

5.1 Limiti e Sviluppi Futuri

La rete ha mostrato ottime prestazioni nonostante le dimensioni ridotte del dataset, in modo particolare di quello OCT composto da soli 9 volumi.

Inoltre è opportuno evidenziare che le maschere di riferimento non risultano estremamente accurate in quanto elaborate manualmente utilizzando Slicer3D. Questo potrebbe contribuire al peggioramento delle metriche di valutazione.

La velocità dell'algoritmo, nonostante le strategie utilizzate, potrebbe non essere sufficiente per applicazioni in tempo reale. In modo particolare, il kernel CUDA ha permesso di ottimizzare alcune fasi di elaborazione dell'algoritmo, tuttavia la GPU offerta da Google Colab presenta dei tempi di utilizzo limitati. Anche la dimensione della RAM ha compromesso le velocità dell'algoritmo

in quanto spesso nella ricostruzione volumetrica raggiungeva il massimo della sua capacità e di conseguenza il riavvio del runtime.

Risulta quindi necessario testare l'algoritmo su un runtime locale e con una GPU locale. Questo permetterebbe anche di evitare il caricamento, volta per volta, dei dati da Google Drive a Google Colab, che di certo ne rallenta le prestazioni.

Infine, sviluppi futuri potrebbero riguardare l'inserimento di un dataset di lavoro più ampio al fine di migliorare le performance e soprattutto di raggiungere l'obiettivo di un riconoscimento su più modalità di imaging.

Di fatti, l'applicazione del modello OCT sulle MIP dei volumi PAT, ha messo in luce la possibilità di creare un modello Multi-Modale con prestazioni notevoli nel riconoscere l'ago tra le strutture vascolari.

Tuttavia le 8 immagini a disposizione non sono sufficienti per generalizzare il risultato.

Bibliografia

- [1] Mingyue Ding e Aaron Fenster. *"A real-time biopsy needle segmentation technique using Hough Transform"* (cit. a p. 1).
- [2] Carsten Stevenson Majid Maybody e Stephen B. Solomon. *"Overview of Navigation Systems in Image-Guided Interventions"* (cit. a p. 1).
- [3] BRENTON KELLER MARK DRAELOS GAO TANG SINA FARSIU ANTHONY N. KUO KRIS HAUSER e JOSEPH A. IZATT. *"Real-time corneal segmentation and 3D needle tracking in intrasurgical OCT"*. 2018 (cit. a p. 3).
- [4] Joanna Czajkowska Bartłomiej Pyciński Jan Juszczak Ewa Pietka. *"Biopsy needle tracking technique in US images"*. Computerized Medical Imaging e Graphics Volume 65 Pages 93-101, 2018 (cit. a p. 4).
- [5] Nassir Navab Jakob Weiss Nicola Rieke Mohammad Ali Nasser Mathias Maier Abouzar Eslami. *"Fast 5DOF needle tracking in iOCT"* (cit. a p. 4).
- [6] Aumann Donner S Fischer J Müller F. *"Tomografia a coerenza ottica (OCT): principio e realizzazione tecnica."* 2019 (cit. a p. 4).
- [7] Yu Chen et al. «"Review of advanced imaging techniques"». In: *Journal of pathology informatics* 3 (mag. 2012), p. 22. DOI: 10.4103/2153-3539.96751 (cit. a p. 4).
- [8] McLeod. *"Novel approaches to needle tracking and visualisation"* (cit. a p. 5).
- [9] Beard. *"Biomedical photoacoustic imaging."* (Cit. a p. 6).
- [10] Robert Prevedel akub Czuchnowski. *"Fotoacustica: vedere attraverso il suono"* (cit. a p. 6).
- [11] K. M. Meiburger. *"Quantitative Ultrasound and Photoacoustic Imaging for the Assessment of Vascular Parameters. Springer"* (cit. alle pp. 6, 7).
- [12] K. J. Draper C. C. Blake D. B. Gowman L. nad Downey e A. Fenster. *"An Algorithm for Automatic Needle Localization in Ultrasound guided Breast Biopsies."* (Cit. a p. 8).
- [13] Martin Barva Jan Kybic Jean-Martial Mari Christian Cachard. *"Localizing Metal Electrode from 3D Ultrasound Data Using Ransac and Intensity Priors."* 2005 (cit. a p. 10).
- [14] Joshua Shubert Muyinatu A. Lediju Bell. *"Photoacoustic Based Visual Servoing of Needle Tips to Improve Biopsy on Obese Patients"* (cit. a p. 11).
- [15] Yuuki Watanabe e Toshiki Itagaki. *"Real-time display on Fourier domain optical coherence tomography system using a graphics processing unit"* (cit. a p. 16).

- [16] Nyul LG Udupa JK e Zhang X. "*New variants of a method of MRI scale standardization* " (cit. a p. 25).
- [17] "*ITK. Insight Toolkit*". URL: https://itk.org/Doxygen/html/classitk_1_1HistogramMatchingImageFilter.html (cit. a p. 25).
- [18] MK Gurucharan. "*Basic CNN Architecture: Explaining 5 Layers of Convolutional Neural Network*". URL: <https://www.upgrad.com/blog/basic-cnn-architecture/> (cit. a p. 29).
- [19] Olaf Ronneberger Philipp Fischer e Thomas Brox. "*U-net: Convolutional networks for biomedical image segmentation*" (cit. alle pp. 29, 31).
- [20] Evan Shelhamer, Jonathan Long e Trevor Darrell. «Fully Convolutional Networks for Semantic Segmentation». In: *CoRR* abs/1605.06211 (2016). arXiv: 1605.06211. URL: <http://arxiv.org/abs/1605.06211> (cit. a p. 29).
- [21] E. Shelhamer J. Long e T. Darrell. «"Fully convolutional networks for semantic segmentation"». In: () (cit. a p. 30).
- [22] "<https://roboflow.com/model/resnet-34>" (cit. a p. 31).
- [23] URL: <https://towardsdatascience.com/understanding-and-visualizing-%20resnets-442284831be8> (cit. a p. 31).
- [24] URL: <https://deeptai.org/machine-learning-glossary-and-terms/loss-function> (cit. a p. 47).
- [25] BERNAL J KUSHIBAR K ASFAW D S et al. "*Deep convolutional neural networks for brain image analysis on magnetic resonance imaging: A review*" (cit. a p. 47).
- [26] DROZDZAL M VORONTSOV E CHARTRAND G et al. "*The importance of skip connections in biomedical image segmentation*" (cit. a p. 47).
- [27] Huttenlocher D.P., Klanderma G.A. e W.J. Rucklidge. «Comparing images using the Hausdorff distance». In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 15.9 (1993), pp. 850–863. DOI: 10.1109/34.232073 (cit. a p. 48).
- [28] Jingnan Jia. *A package to compute segmentation metrics: seg-metrics*. 2020. DOI: 10.5281/zenodo.3995075. URL: https://github.com/Ordgod/segmentation_metrics (cit. a p. 49).

Ringraziamenti

Infine, vorrei dedicare una sezione di questo lavoro per esprimere i miei ringraziamenti a tutte le persone che hanno sostenuto e contribuito alla realizzazione di questo lavoro di tesi.

Innanzitutto, vorrei esprimere la mia gratitudine alla Prof.ssa Kristen M. Meiburger per aver accettato di essere la relatrice della mia tesi di laurea, avendomi dato la possibilità di approfondire un ambito per il quale ho nutrito grande interesse durante i miei studi.

Ringrazio il mio correlatore, l'Ing. Francesco Marzola, per il suo contributo e supporto durante lo sviluppo di questa tesi di laurea. La sua competenza, i suoi suggerimenti e la sua disponibilità nel condividere le sue conoscenze sono stati fondamentali per il successo di questo lavoro di ricerca.

Desidero esprimere la mia sentita riconoscenza alla mia famiglia, che ha giocato un ruolo fondamentale nel mio percorso accademico. Il loro continuo supporto, affetto e incoraggiamento sono stati inestimabili e mi hanno dato la forza necessaria per raggiungere questo obiettivo.