



UNIVERSITÀ DI PARMA

DIPARTIMENTO DI INGEGNERIA E ARCHITETTURA

Corso di Laurea Triennale in Ingegneria dei Sistemi Informativi

Addestramento di una rete neurale encoder-decoder con dati limitati per la segmentazione del femore fetale da immagini ecografiche

*Training an encoder-decoder neural network with limited data
for fetal femur segmentation from echographic images*

CANDIDATO:
Dmitri Ollari Ischimji

RELATORE:
Prof. Claudio Ferrari

Dedicato a Diego

Indice

1	Introduzione	7
2	Lavori correlati	9
2.1	Segmentazione	9
2.2	Fully Convolutional Network	10
2.3	U-Net	11
2.3.1	U-Net: Convolutional Networks for Biomedical Image Segmentation	11
2.4	Segmentazione ossea	12
2.5	Segmentazione di vasi sanguigni	12
3	Metodi	15
3.1	Etichettatura	19
3.2	Modello	19
3.3	Architettura U-Net	20
3.3.1	Convoluzione	21
3.3.2	Max Pooling	21
3.3.3	Encoder	21
3.3.4	Decoder	22
3.3.5	Bridge	22
3.3.6	Output	23
3.4	Rimozione della Sliding Window	23
3.5	Modifica della Struttura Encoder-Decoder	23
3.6	Metriche	24
3.6.1	Dice BCE Loss	24
3.6.2	Intersection over Union(IoU)	26
3.7	Validazione del modello	26
3.8	Validazione del Modello	26

4	Risultati Sperimentali	29
4.1	Addestramento	29
4.2	Analisi con il Nuovo Ecografo	32
4.2.1	Confronto delle Immagini	33
4.2.2	Analisi della Distribuzione dei Valori	33
4.2.3	Risultati Quantitativi	34
4.2.4	Discussione	34
5	Conclusioni	37
	Bibliografia	39

Elenco delle figure

2.1	Segmentazione semantica	9
2.2	CNN	10
2.3	U-Net	11
3.1	Immagini utilizzate per l'addestramento	16
3.2	Immagini utilizzate nel test set	17
3.3	Immagini utilizzate nel validation set	17
3.4	Rappresentazione della Dice Loss	25
3.5	Rappresentazione grafica della Cross-validation	27
4.1	Processo di addestramento del modello	29
4.2	Errore e accuratezza della prima porzione di dati	30
4.3	Errore e accuratezza della seconda porzione di dati	30
4.4	Errore e accuratezza della terza porzione di dati	31
4.5	Errore e accuratezza della quarta porzione di dati	31
4.6	Errore e accuratezza della quinta porzione di dati	31
4.7	Immagine originale, segmentazione manuale e segmentazione del modello	32
4.8	Distribuzione Intensità dei pixel della segmentazione manuale, della segmentazione del modello e confronto tra le due	32
4.9	Immagine Originale, Maschera di segmentazione ottenuta dal modello e Risultato finale a seguito dell'estrazione del femore .	33
4.10	Immagine Originale, Maschera di segmentazione ottenuta dal modello e Risultato finale a seguito dell'estrazione del femore .	33
4.11	Immagine Originale, Maschera di segmentazione ottenuta dal modello e Risultato finale a seguito dell'estrazione del femore .	33
4.12	Distribuzione dei Valori	34
5.1	Correlazione tra luminosità e peso alla nascita	37

Elenco delle tabelle

3.1	Encoding e Decoding originali	24
3.2	Encoding e Decoding modificati	24
4.1	Risultati Sperimentali	34

Capitolo 1

Introduzione

La segmentazione semantica gioca un ruolo cruciale nell'analisi delle immagini mediche, consentendo di identificare e isolare strutture anatomiche di interesse. Questa tesi si concentra sull'applicazione di reti neurali convoluzionali (CNN), in particolare sull'utilizzo dell'architettura U-Net per la segmentazione binaria di immagini ecografiche fetali, allo scopo di estrarre e delineare i femori.

Le immagini ecografiche fetali rappresentano una sfida complessa nella segmentazione, richiedendo un'accurata identificazione delle strutture anatomiche. La segmentazione semantica binaria mira all'etichettatura specifica di pixel associati ai femori, fornendo una comprensione dettagliata delle strutture anatomiche esaminate.

L'approccio di questa tesi si basa sull'uso della rete neurale convoluzionale U-Net, nota per la sua efficacia nella segmentazione di immagini biomediche. La peculiarità di U-Net è la sua capacità di catturare dettagli locali pur mantenendo una visione globale dell'immagine, rendendola particolarmente adatta per la segmentazione dettagliata, come l'estrazione dei femori dalle ecografie fetali.

Attraverso l'analisi, l'implementazione e l'ottimizzazione di questa architettura, il lavoro mira a migliorare l'accuratezza e l'efficienza della segmentazione, fornendo uno strumento affidabile per l'identificazione automatica dei femori nelle immagini ecografiche fetali. L'obiettivo è di apportare un contributo significativo all'avanzamento delle tecnologie di estrazione delle informazioni dalle immagini ecografiche, automatizzando e facilitando una valutazione più precisa della densità minerale ossea fetale (BMD).

Capitolo 2

Lavori correlati

2.1 Segmentazione

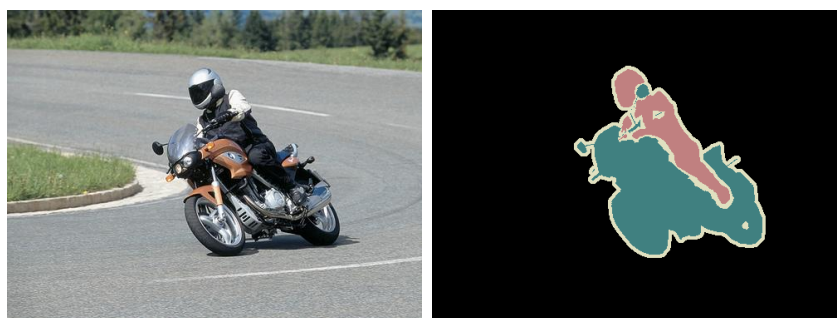


Figura 2.1: Segmentazione semantica

La segmentazione semantica rappresenta un campo di grande interesse e rilevanza nell'ambito dell'elaborazione delle immagini e della visione artificiale. Questa tecnica si distingue per la sua capacità d'interpretare il contenuto delle immagini a un livello semantico, andando oltre la semplice divisione dell'immagine in regioni omogenee basate su caratteristiche visive come il colore o la texture. Nello specifico, la segmentazione semantica si prefigge l'obiettivo di attribuire un'etichetta semantica a ogni singolo pixel dell'immagine, consentendo così d'identificare e categorizzare le diverse parti che compongono la scena. L'obiettivo principale della segmentazione semantica è quello di fornire una comprensione approfondita del contenuto visivo presente in un'immagine. Ciò si traduce nella capacità d'identificare e categorizzare oggetti e regioni, rendendo possibile un'analisi dettagliata e una migliore interpretazione dei dati visivi. Un esempio di applicazione

della segmentazione semantica la si può visionare nella figura 2.1, in questo caso l'obiettivo della segmentazione era quello di estrapolare le informazioni relative al motociclista in una classe e le informazioni relative al veicolo in un'altra classe, separando entrambe le classi dallo sfondo.

2.2 Fully Convolutional Network

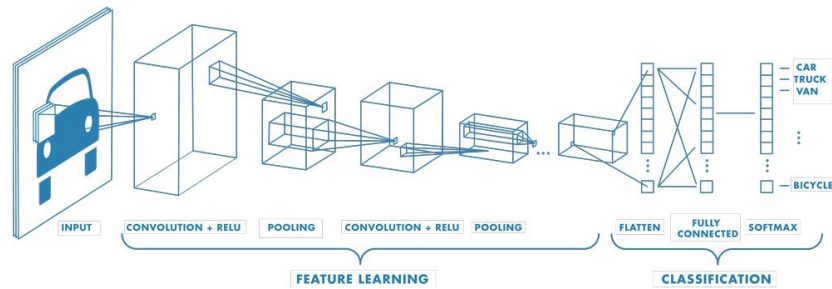


Figura 2.2: CNN

L'articolo *Fully Convolutional Networks for Semantic Segmentation* [Long et al., 2015] propone l'utilizzo di una tipologia di reti neurali convoluzionali (CNN) che permettono grazie all'assenza di layer completamente connessi di elaborare immagini di qualunque dimensione. Questa nuova tipologia di reti migliora notevolmente le capacità di apprendimento delle reti neurali permettendo di produrre mappe di segmentazione più precise grazie alla loro capacità di apprendimento d'informazioni spaziali.

Le motivazioni riguardanti l'ampio utilizzo nel settore della *computer vision* sono legate all'assenza di strati completamente connessi (lineari) che vincolano l'ingresso alla medesima grandezza per ogni singola immagine, permettendo di fornire in ingresso l'intera immagine e non frammenti della stessa così da aumentare l'apprendimento spaziale della rete.

Questa maggior flessibilità comporta un addestramento libero da limitazioni sull'ingresso comportando una maggiore tolleranza agli errori e al rumore rendendo questa tipologia di reti particolarmente adatte a contesti poveri di dati.

Le motivazioni per l'ampio utilizzo di queste reti nel settore della *computer vision* risiedono nell'assenza di strati completamente connessi, che solitamente vincolano l'ingresso a dimensioni fisse per ogni immagine. Questa caratteristica permette alle reti di elaborare l'intera immagine anziché solo frammenti, potenziando così l'apprendimento spaziale.

La maggiore flessibilità offerta da queste reti si traduce in un addestramento meno vincolato da limitazioni dimensionali dell'input, risultando in una maggiore tolleranza agli errori e al rumore. Di conseguenza, questa tipologia di reti si rivela particolarmente adatta in contesti caratterizzati da una scarsità di dati.

2.3 U-Net

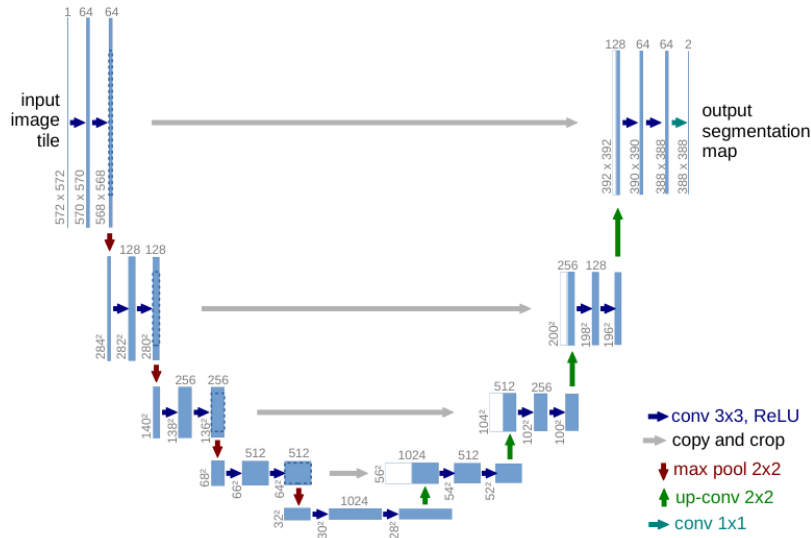


Figura 2.3: U-Net

2.3.1 U-Net: Convolutional Networks for Biomedical Image Segmentation

Nel 2015, Olaf Ronneberger, Philipp Fischer e Thomas Brox hanno introdotto il modello di rete neurale convoluzionale *U-Net* per la segmentazione semantica di immagini biomedicali [Ronneberger et al., 2015]. Progettata specificatamente per affrontare le sfide della segmentazione in ambito biomedico, come la necessità di segmentare con precisione strutture anatomiche con un numero limitato di immagini di addestramento, U-Net ha rappresentato un notevole avanzamento.

Il modello *U-Net* si distingue per una struttura simmetrica, dove la parte contrattiva (downsampling) cattura il contesto e quella espansiva (upsam-

pling) permette una localizzazione precisa. Questa configurazione consente alla rete di fondere informazioni di contesto con quelle locali, migliorando significativamente la precisione della segmentazione.

Una delle innovazioni chiave di *U-Net* è l'introduzione di collegamenti a salti tra le parti contrattive ed espansive, che trasferiscono caratteristiche spaziali ad alta risoluzione dalla parte contrattiva a quella espansiva, aumentando la precisione nella localizzazione delle strutture segmentate.

Grazie a queste caratteristiche, il modello *U-Net* ha ottenuto risultati eccellenti nella segmentazione di immagini biomedicali, anche con un numero limitato di immagini di addestramento, diventando così un punto di riferimento nella segmentazione semantica biomedica.

2.4 Segmentazione ossea

Il lavoro *Towards whole-body CT Bone Segmentation* [Klein et al., 2018] rappresenta un'importante analisi nello sviluppo di metodi e algoritmi avanzati per la segmentazione ossea in immagini ottenute tramite tomografia computerizzata (TC) di tutto il corpo. Questo studio pone l'accento sull'importanza della segmentazione ossea in campo medico, sia per la diagnosi di condizioni patologiche sia per analisi dettagliate del tessuto osseo.

Il contributo fondamentale dell'articolo risiede nell'esplorazione di approcci innovativi e nell'ottimizzazione di tecniche algoritmiche per identificare e isolare con precisione le strutture ossee nelle immagini TC. Viene sottolineata l'importanza dell'uso di metodologie avanzate nell'elaborazione delle immagini e dell'applicazione di algoritmi di visione artificiale e machine learning per ottenere una segmentazione accurata.

L'articolo assume un ruolo significativo nel campo dell'informatica medica, evidenziando l'impiego di soluzioni informatiche per un'analisi approfondita delle immagini mediche e riconoscendo l'importanza delle tecniche di segmentazione ossea per applicazioni cliniche e di ricerca biomedica.

2.5 Segmentazione di vasi sanguigni

L'articolo *Accurate Retinal Vessel Segmentation via Octave Convolution Neural Network* [Fan et al., 2020] introduce un approccio innovativo per la segmentazione precisa dei vasi sanguigni retinici utilizzando le reti neurali a convoluzione ottava. Questa tecnica gioca un ruolo cruciale nell'analisi delle immagini retiniche in ambito medico.

Il lavoro mette in luce i vantaggi offerti dalle reti neurali a convoluzione ottava, che utilizzano differenti frequenze spaziali per catturare dettagli a diverse scale. Questo approccio risulta particolarmente efficace nella segmentazione dei vasi sanguigni retinici, contribuendo significativamente alla comprensione e diagnosi delle patologie oculari.

L'articolo dimostra l'efficacia di queste reti nel rilevare e isolare i vasi sanguigni della retina, mostrando risultati più accurati rispetto ai metodi tradizionali. In definitiva, *Accurate Retinal Vessel Segmentation via Octave Convolution Neural Network* offre un contributo importante nel campo della segmentazione vascolare retinica, sottolineando l'efficacia e l'importanza delle reti neurali a convoluzione ottava nella diagnostica medica.

Capitolo 3

Metodi

Tutti i dati sono stati raccolti da analisi ecografiche effettuate presso l'**Azienda Ospedaliera Universitaria di Parma**, nel periodo tra **Aprile 2022** e **Gennaio 2023** da un team di **medici esperti**. Per standardizzare le immagini raccolte, sono stati applicati i seguenti parametri:

- Indice di massa corporea (BMI)
- Età
- Problematiche durante la gravidanza
- Problematiche dopo la gravidanza
- Femore centrato nell'inquadratura

Le immagini hanno subito un processo di *preprocessing* aggiuntivo per uniformare le dimensioni e la risoluzione. In particolare, sono state ridimensionate a immagini **1280px di larghezza** e **876px di altezza**.

Inoltre data la tipologia del problema, si è scelto di convertire le immagini da **RGB** a immagini in **scala di grigi** per ridurre la complessità del problema e per ridurre il quantitativo di dati necessari per l'addestramento della rete.

Sono state realizzate manualmente delle **maschere** di segmentazione per ogni immagine, in modo da avere un *ground truth* da confrontare con le predizioni del modello.

Data la scarsa quanti di dati a disposizione per l'addestramento della U-Net, si è scelto di utilizzare alcune tecniche di *data augmentation* per aumentare la quantità di dati a disposizione. In particolare si è scelto di utilizzare le seguenti tecniche applicate in modo casuale per ogni coppia **immagine-maschera**

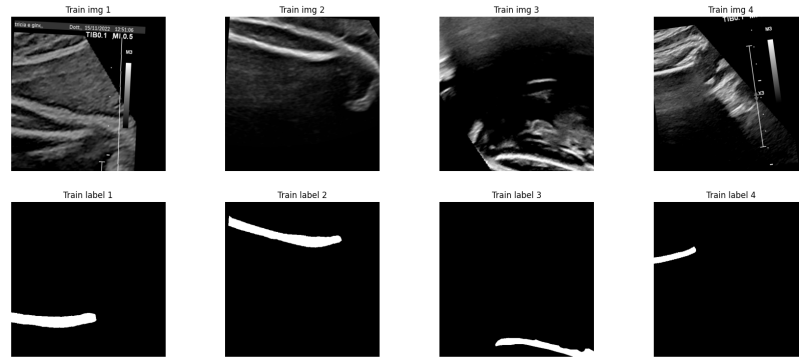


Figura 3.1: Immagini utilizzate per l'addestramento

- *Flip* orizzontale e verticale
- *Rotazioni* di 35°
- *Rumore* Gaussiano

Queste tecniche hanno migliorato notevolmente le segmentazioni ottenute tramite la rete U-Net, aumentando la robustezza della rete alle variazioni di luce e al rumore nelle immagini.

L'efficacia della *data augmentation* nella fase di addestramento del modello è evidente, migliorando l'adattabilità a contesti non controllati e la generalizzazione del modello. Come mostrato in Figure 3.1, la *data augmentation* è stata applicata in modo casuale per ogni coppia **immagine-maschera** solo durante l'addestramento, mantenendo inalterate le immagini usate per il controllo del modello.

Soltanto le immagini utilizzate per l'addestramento sono state sottoposte a *data augmentation*, le immagini contenute nel *test set* e nel *validation set* sono rimaste invariate.

È fondamentale che i set di test e validazione riflettano accuratamente le condizioni reali di utilizzo, per valutare efficacemente le prestazioni del modello. L'introduzione di variazioni artificiali tramite *data augmentation* in questi set potrebbe portare a una valutazione distorta delle capacità del modello.

Applicare la *data augmentation* ai set di test e validazione potrebbe risultare in una valutazione imprecisa delle prestazioni del modello, fornendo un'immagine ottimistica ma non realistica della sua capacità di generalizzare su dati non modificati.

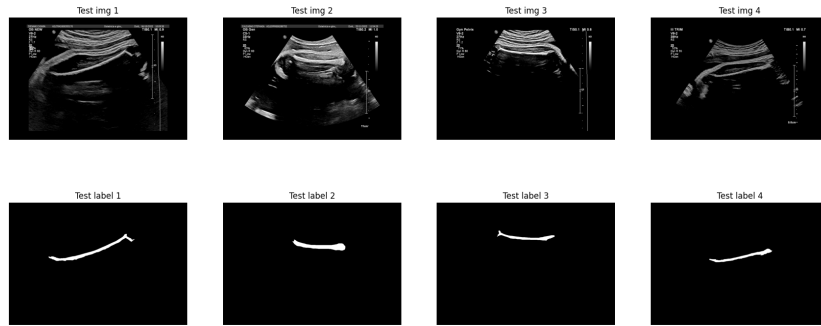


Figura 3.2: Immagini utilizzate nel test set

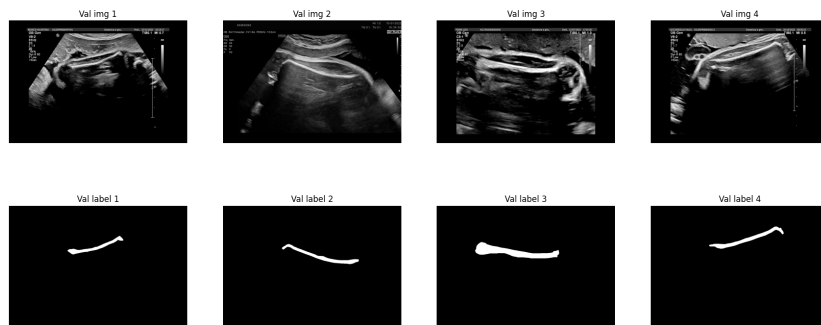


Figura 3.3: Immagini utilizzate nel validation set

Mentre la data augmentation durante l'addestramento espone la rete a una varietà maggiore di dati, è essenziale valutare la rete su dati non alterati per assicurarsi che abbia imparato a generalizzare correttamente.

3.1 Etichettatura

Le immagini utilizzate per l'addestramento della rete sono state etichettate manualmente utilizzando un software di etichettatura chiamato **LabelMe** [Wada, 2023].

Il processo di etichettatura comporta la selezione accurata e la delimitazione del perimetro del femore in ciascuna immagine. Sono applicati indicatori specifici attorno alla regione di interesse, in modo da isolare il femore dal resto dell'immagine. Questo approccio di etichettatura mirata è cruciale per fornire alla rete neurale esempi precisi da cui imparare, migliorando così la sua capacità di identificare correttamente le strutture anatomiche nelle immagini.

La corretta etichettatura è vitale non solo per l'addestramento efficace della rete, ma anche per garantire l'affidabilità e l'accuratezza delle segmentazioni future. Essa permette alla rete di riconoscere le variazioni sottili e le caratteristiche specifiche del femore, che sono essenziali per applicazioni mediche precise e affidabili.

3.2 Modello

Il modello scelto come punto di partenza per questo studio è il **U-Net** (Figure 2.3), la cui architettura originale è stata sviluppata da Olaf Ronneberger, Philipp Fischer e Thomas Brox nel 2015 [Ronneberger et al., 2015]. Questo studio propone un'architettura di rete neurale convoluzionale specificamente progettata per la segmentazione semantica di immagini biomediche, in cui ogni singolo pixel dell'immagine è classificato in una delle categorie pertinenti al problema analizzato.

Lo studio propone un'architettura di rete neurale convoluzionale per la segmentazione semantica di immagini biomediche, classificando ogni singolo pixel dell'immagine in una delle varie categorie del problema analizzato, la rete convoluzionale emersa da questa analisi rimane tutto' oggi una delle più utilizzate in ambito medico per la segmentazione semantica data la sue performance e la sua versatilità.

L'implementazione iniziale ricalca il modello realizzato da Olaf Ronneberger, Philipp Fischer e Thomas Brox, utilizzando il framework **PyTorch** [Team,] come base per la realizzazione della rete.

3.3 Architettura U-Net

L'architettura proposta da Olaf Ronneberger, Philipp Fischer e Thomas Brox è composta di 4 parti principali:

- **Encoder:** Visionabile graficamente come la parte discendente della U-Net
- **Bridge:** Visionabile graficamente come la linea di congiunzione fra la parte discendente e la parte ascendente della U-Net
- **Decoder:** Visionabile graficamente come la parte ascendente della U-Net
- **Output:** Visionabile graficamente come l'ultimo layer della U-Net

Le applicazioni che si appoggiano a modelli derivati dall'architettura U-net dominano settori come la medicina e la biologia, in particolare la segmentazione di immagini biomediche, come la segmentazione di immagini ecografiche, la segmentazione di immagini TC e la segmentazione di immagini RM.

I motivi principali del successo di U-Net includono:

- **Segmentazione dettagliata:** U-Net è capace di produrre segmentazioni dettagliate e precise grazie alle sue innovative "skip connections", che consentono al modello di catturare sia i dettagli di basso livello che il contesto di alto livello.
- **Architettura compatta:** Nonostante la sua elevata capacità di dettaglio, il modello rimane relativamente snello e può essere addestrato con successo anche con dataset di dimensioni moderate.
- **Adattabilità:** Originariamente concepita per applicazioni mediche, U-Net ha dimostrato un'elevata versatilità, adattandosi a una vasta gamma di applicazioni.

Tuttavia, come ogni modello, U-Net presenta anche alcuni svantaggi. Il principale riguarda la necessità di un dataset di addestramento ampio e accuratamente etichettato, che può essere un ostacolo in contesti con dati limitati. Inoltre, U-Net richiede una quantità significativa di memoria per memorizzare i pesi del modello, il che può rappresentare un problema nel caso di immagini ad alta risoluzione.

3.3.1 Convoluzione

La convoluzione è una delle operazioni fondamentali utilizzate nelle reti neurali convoluzionali (CNN) per estrarre le caratteristiche significative da un' input. Essa impiega un filtro (o kernel) che viene applicato sull'input.

La convoluzione è una delle operazioni fondamentali utilizzate nelle reti neurali convoluzionali per estrarre le caratteristiche significative da un' input, la convoluzione coinvolge un filtro (o kernel) e l'input su cui si applica.

Il processo di convoluzione consiste nel sommare ogni elemento di un'immagine al suo vicino, pesando ogni singola operazione mediante l'utilizzo del filtro(o kernel) il calcolo della feature map di uscita è calcolata come segue:

$$\left(\begin{bmatrix} a & b & c \\ d & e & f \\ g & h & i \end{bmatrix} * \begin{bmatrix} 1 & 2 & 3 \\ 4 & 5 & 6 \\ 7 & 8 & 9 \end{bmatrix} \right) [2, 2] = \quad (3.1)$$

$$= (i \cdot 1) + (h \cdot 2) + (g \cdot 3) + (f \cdot 4) + (e \cdot 5) + (d \cdot 6) + (c \cdot 7) + (b \cdot 8) + (a \cdot 9) \quad (3.2)$$

3.3.2 Max Pooling

Il *max pooling* è un'operazione chiave all'interno delle reti neurali convoluzionali (CNN), inclusa la rete U-Net.

Il *max pooling* è utilizzato per ridurre la dimensione delle feature map, consentendo di ridurre la complessità del problema da approssimare, comportando una maggior resistenza al *overfitting*, migliorando la capacità di generalizzazione del modello e di ottenere una rappresentazione più invariante rispetto alle piccole variazioni spaziali nell'input.

3.3.3 Encoder

La fase di *encoding* rappresenta la prima parte della rete U-Net, ed è composta da una serie di strati di convoluzione (vedi Sezione 3.3.1) e max pooling (vedi Sezione 3.3.2). Questi strati lavorano per ridurre progressivamente la dimensione spaziale dell'immagine, aumentando contemporaneamente il numero di canali delle *features*.

Nello specifico, la fase di *encoding* comprende tre componenti principali:

- **Strato iniziale:** Questo strato applica diverse operazioni di convoluzione ai dati di input per estrarre caratteristiche di basso livello, come bordi e texture.

- **Downsampling:** Successivamente, si utilizzano operazioni di max pooling o convoluzione con un passo (stride) superiore a 1 per ridurre la dimensione delle feature map.
- **Strati intermedi:** Questi strati applicano ulteriori operazioni di convoluzione per catturare caratteristiche di complessità crescente.

3.3.4 Decoder

La fase di *decoding* nella rete U-Net è incaricata di ricostruire l'immagine segmentata a partire dalle informazioni estratte durante la fase di *encoding*. Le fasi principali del decoder includono:

- **Upsampling:** La fase di decoding inizia con l'operazione di upsampling, che serve a ripristinare gradualmente la dimensione delle feature map ai livelli originali dell'immagine. Ciò viene fatto utilizzando operazioni come la trasposta della convoluzione (de convoluzione) o l'interpolazione bilineare. L'obiettivo è ottenere feature map di dimensioni compatibili con quelle dell'immagine di input.
- **Skip Connections:** Un aspetto distintivo della U-Net sono le skip connections, o connessioni di salto. Queste connessioni collegano le feature map estratte durante l'encoding alle corrispondenti feature map nella fase di decoding. Ciò consente di combinare informazioni multi-scala, in modo che il modello possa accedere sia a dettagli fini che a contesto di alto livello. Le skip connections sono fondamentali per migliorare la precisione della segmentazione.
- **Convoluzione nel Decoding:** Dopo l'upsampling e l'integrazione delle skip connections, vengono applicate operazioni di convoluzione per raffinare ulteriormente le feature map. Queste convoluzioni possono avere lo scopo di "mescolare" le informazioni o di catturare dettagli specifici a livelli più alti.

3.3.5 Bridge

Il *bridge* è una componente cruciale dell'architettura U-Net, che facilita il trasferimento di informazioni rilevanti tra l'encoder e il decoder attraverso le *skip connections*. Questa fase consente di considerare dettagli sia di basso che di alto livello durante il processo di segmentazione.

3.3.6 Output

La parte finale della rete U-Net è costituita da uno o più strati di convoluzione che riducono la profondità delle feature map alla dimensione richiesta per l'output finale, producendo così l'immagine segmentata.

3.4 Rimozione della Sliding Window

Con gli avanzamenti tecnologici nel campo dell'hardware e del software, si è osservato un notevole miglioramento nei tempi di elaborazione delle immagini e nella capacità di gestire immagini di grandi dimensioni. Pertanto, si è deciso di abbandonare l'approccio della sliding window, che è più conservativo in termini di utilizzo della memoria e tempo di elaborazione.

Al suo posto, si è adottato un metodo più moderno che sfrutta la potenza di calcolo delle GPU attuali e la loro memoria dedicata, significativamente più ampia rispetto alle generazioni precedenti. L'utilizzo di immagini complete, anziché porzioni separate tramite sliding window, ha prodotto un "effetto collaterale" positivo: una migliore comprensione spaziale delle immagini da parte della rete.

La visualizzazione dell'intera immagine consente alla rete di acquisire una migliore comprensione del contesto spaziale, migliorando così la precisione della segmentazione.

3.5 Modifica della Struttura Encoder-Decoder

Durante le fasi sperimentali, è emerso che aumentando il numero di iterazioni di convoluzione nei vari strati di encoding e decoding si ottengono risultati migliori in termini di segmentazione. Questo miglioramento, tuttavia, comporta un aumento del tempo di elaborazione e del consumo di memoria.

Di conseguenza, si è deciso di modificare le fasi di *encoding* e *decoding* come segue:

L'intento di questa modifica è sfruttare la capacità di estrazione di feature delle operazioni di convoluzione e di riduzione delle feature meno rilevanti tramite max pooling. Aggiungendo iterazioni di convoluzioni, la rete può estrarre più feature importanti per la classificazione dei pixel, mentre il max pooling aiuta a eliminare le feature meno significative, riducendo la dimensione delle feature map.

Layer	In channels	Out channels
Encoder	1	64
Encoder	64	128
Encoder	128	256
Encoder	256	512
Decoder	1024	512
Decoder	512	256
Decoder	256	128
Decoder	128	64

Tabella 3.1: Encoding e Decoding originali

Layer	In channels	Out channels
Encoder	1	16
Encoder	16	32
Encoder	32	64
Encoder	64	128
Encoder	128	256
Encoder	256	512
Decoder	1024	512
Decoder	512	256
Decoder	256	128
Decoder	128	64
Decoder	64	32
Decoder	32	16

Tabella 3.2: Encoding e Decoding modificati

3.6 Metriche

Per valutare le prestazioni del modello nel contesto del task di segmentazione, si è scelto di utilizzare la metrica *Dice BCE Loss* per la funzione di perdita e *Intersection over Union (IoU)* per misurare l'accuratezza.

3.6.1 Dice BCE Loss

La Dice Loss è una metrica di perdita basata sul coefficiente di Dice, che è una misura della somiglianza tra due campioni. È particolarmente utile per i dati non bilanciati, come è spesso il caso nelle immagini mediche dove la

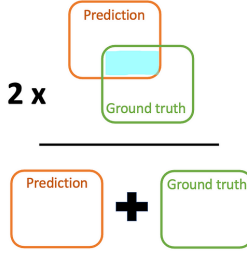
$$\text{Dice} = \frac{2 \times \text{Area of overlap}}{\text{Total area}} = \frac{2 \times \text{Area of overlap}}{\text{Prediction} + \text{Ground truth}}$$


Figura 3.4: Rappresentazione della Dice Loss

regione di interesse occupa una piccola parte dell'immagine. Il coefficiente di Dice è definito dall'equazione 3.4.

La *Binary Cross-Entropy (BCE) Loss* è ampiamente usata nei problemi di classificazione binaria, specialmente quando i dati di output sono probabilità che variano tra 0 e 1. Questa metrica valuta quanto le predizioni del modello si discostino dai valori reali (etichette).

$$L = L_{\text{Dice}} + L_{\text{BCE}} \quad (3.3)$$

$$L_{\text{Dice}} = 1 - \frac{2 \sum_i^N p_i g_i + \varepsilon}{\sum_i^N p_i^2 + \sum_i^N g_i^2 + \varepsilon} \quad (3.4)$$

$$L_{\text{BCE}} = -\frac{1}{N} \left[\sum_{i=1}^N g_i \log(p_i) + (1 - g_i) \log(1 - p_i) \right] \quad (3.5)$$

Dunque, la *loss* totale viene calcolata mediante:

$$L = 1 - \frac{2 \sum_i^N p_i g_i + \varepsilon}{\sum_i^N p_i^2 + \sum_i^N g_i^2 + \varepsilon} - \frac{1}{N} \left[\sum_{i=1}^N g_i \log(p_i) + (1 - g_i) \log(1 - p_i) \right] \quad (3.6)$$

Dove:

- p_i rappresenta la probabilità predetta dal modello per il pixel i .
- g_i è il valore di *ground truth* per il pixel i .

3.6.2 Intersection over Union(IoU)

Per la metrica dell'accuratezza della segmentazione, è stata utilizzata la metrica *Intersection over Union* (IoU), poiché è una metrica che permette di valutare la capacità di segmentazione del modello facendo il rapporto tra l'area di intersezione tra la maschera predetta e quella di *ground truth* e l'area di unione tra le due maschere, formalmente:

$$\text{IoU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} \quad (3.7)$$

Dove **TP** è il numero di *True Positive*, **FP** è il numero di *False Positive* e **FN** è il numero di *False Negative*.

3.7 Validazione del modello

La *cross-validation* (validazione incrociata) è una tecnica fondamentale nell'ambito del machine learning e dell'addestramento di modelli predittivi. Essenzialmente, la cross-validation è un metodo per valutare le prestazioni di un modello in modo robusto, valutandolo su più insiemi di dati per ottenere stime più affidabili delle sue capacità predittive. Questo processo aiuta a mitigare il rischio di overfitting e offre una migliore comprensione delle prestazioni del modello.

3.8 Validazione del Modello

La *cross-validation* (validazione incrociata) è un metodo essenziale nel machine learning per valutare le prestazioni di un modello in modo robusto. Questa tecnica comporta la valutazione del modello su più insiemi di dati per ottenere stime più affidabili delle sue capacità predittive, contribuendo a mitigare il rischio di overfitting.

Fornisce stime più affidabili delle prestazioni del modello, riducendo il rischio di ottenere stime di prestazioni spurie a causa di una singola divisione dei dati.



Figura 3.5: Rappresentazione grafica della Cross-validation

Capitolo 4

Risultati Sperimentali

4.1 Addestramento

Il modello è stato addestrato utilizzando la tecnica della *cross-validation* (vedi Figura 3.5), con una suddivisione dei dati in cinque parti uguali (cinque folds). In questo contesto di apprendimento supervisionato, al modello sono state fornite immagini originali insieme alle loro corrispondenti segmentazioni effettuate manualmente.

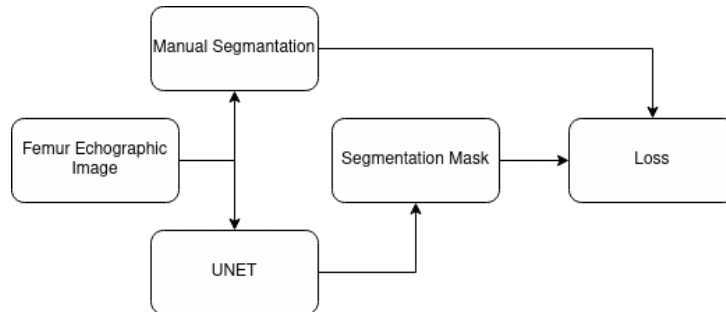


Figura 4.1: Processo di addestramento del modello

L'addestramento è stato eseguito per 5 iterazioni, utilizzando un fold diverso in ciascuna iterazione. L'errore finale è stato calcolato come la media degli errori ottenuti in ciascuna delle cinque iterazioni.

L'**errore** complessivo è stato calcolato come media degli errori ottenuti in ciascuna delle cinque iterazioni di addestramento, utilizzando la formula 3.6. Il risultato è un errore medio del 7.9%.

Parallelamente, l'**accuratezza** complessiva è stata calcolata come la media delle accuratezze ottenute in ciascuna iterazione, impiegando la formula 3.7. Ciò ha portato a un'accuratezza media del 92.1%.

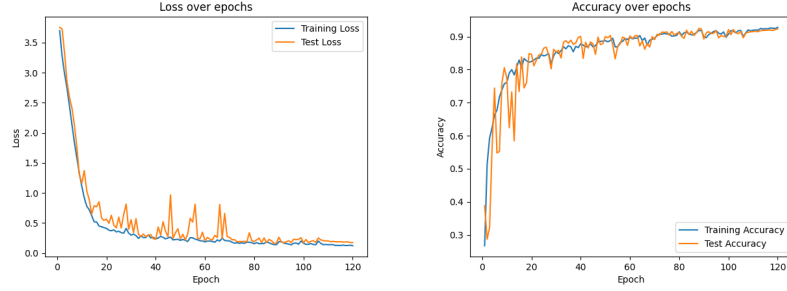


Figura 4.2: Errore e accuratezza della prima porzione di dati

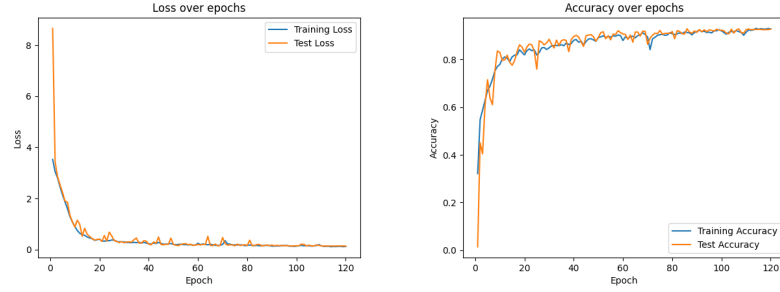


Figura 4.3: Errore e accuratezza della seconda porzione di dati

Oltre alle analisi quantitative, è stato fondamentale condurre analisi qualitative sulla segmentazione ottenuta dal modello, specialmente considerando il suo impiego in ambito medico per la segmentazione dei femori.

Nelle immagini seguenti viene riportato uno delle immagini prese in considerazione per l'addestramento del modello e vengono mostrate le segmentazione manuali, le segmentazioni ottenute dal modello e la differenze nella classificazione dei pixel tra le due segmentazioni.

Per confermare l'efficacia del modello nella segmentazione delle immagini, si è proceduto con l'analisi della distribuzione dei pixel. Questa analisi confronta direttamente la segmentazione manuale con quella ottenuta dal modello. L'obiettivo è dimostrare che la distribuzione dei pixel nelle segmentazioni del modello corrisponde strettamente a quella delle segmentazioni manuali, indicando così una segmentazione accurata.

I risultati mostrano che la distribuzione dei pixel nelle segmentazioni generate dal modello si allinea strettamente a quella delle segmentazioni manuali, suggerendo che il modello è in grado di segmentare le immagini con precisione.

Dall'analisi dei risultati sia qualitativi che quantitativi, emerge che il modello mostra una performance molto promettente. In particolare, si evidenzia

CAPITOLO 4. RISULTATI SPERIMENTALI

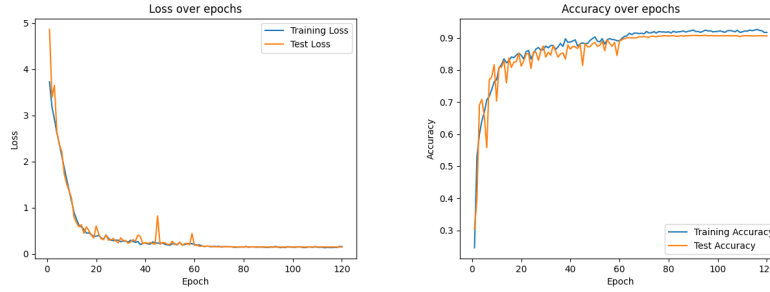


Figura 4.4: Errore e accuratezza della terza porzione di dati

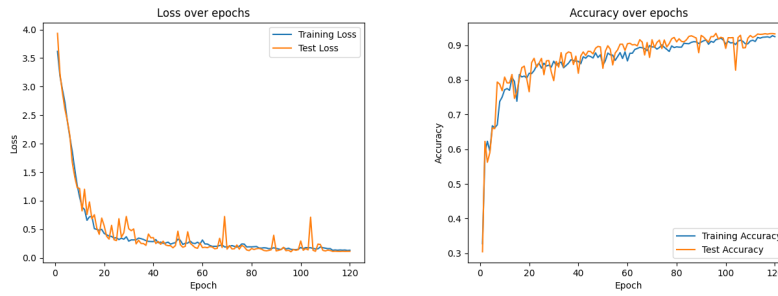


Figura 4.5: Errore e accuratezza della quarta porzione di dati

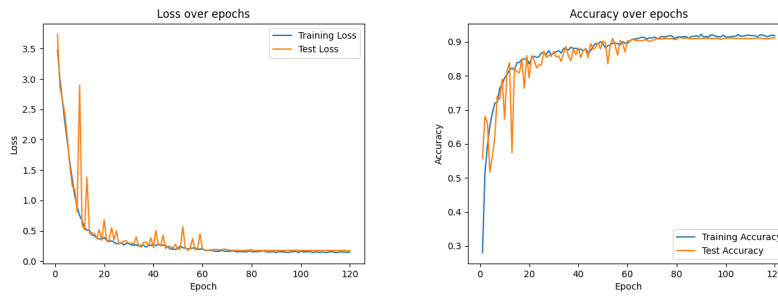


Figura 4.6: Errore e accuratezza della quinta porzione di dati



Figura 4.7: Immagine originale, segmentazione manuale e segmentazione del modello

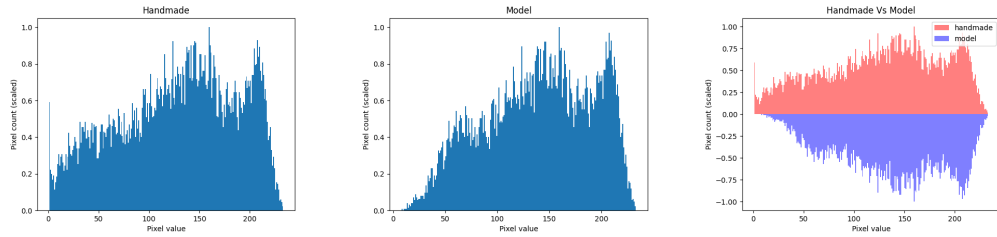


Figura 4.8: Distribuzione Intensità dei pixel della segmentazione manuale, della segmentazione del modello e confronto tra le due

che il modello supera abbondantemente un'accuratezza del 90%, offrendo un ampio margine per ulteriori miglioramenti attraverso l'incremento del numero di immagini disponibili per l'addestramento.

Degno di nota è la mancanza di ulteriori segmentazioni, non è necessario effettuare ulteriori segmentazioni manuali per espandere il set di addestramento. Le nuove immagini raccolte possono essere segmentate automaticamente utilizzando il modello e poi impiegate per rafforzare ulteriormente il processo di addestramento. Questo approccio non solo semplifica il processo di raccolta dati, ma accelera anche la possibilità di miglioramento e adattamento del modello a nuovi dati.

4.2 Analisi con il Nuovo Ecografo

In questa sezione, esploriamo l'impatto dell'introduzione di un nuovo ecografo sulle prestazioni del nostro modello di segmentazione. Questo nuovo strumento rappresenta una sfida significativa per il modello, in quanto introduce un set di dati con caratteristiche visive diverse da quelle su cui è stato originariamente addestrato.

4.2.1 Confronto delle Immagini

Sono state fornite al modello alcune immagini acquisite dal nuovo ecografo. Queste non hanno subito nessun tipo di elaborazione, ma sono state direttamente segmentate dal modello.



Figura 4.9: Immagine Originale, Maschera di segmentazione ottenuta dal modello e Risultato finale a seguito dell'estrazione del femore



Figura 4.10: Immagine Originale, Maschera di segmentazione ottenuta dal modello e Risultato finale a seguito dell'estrazione del femore

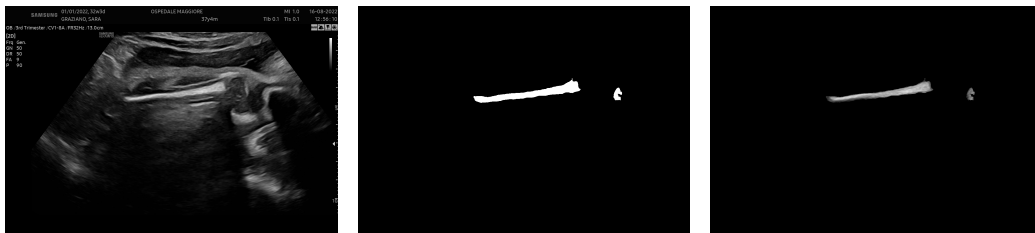


Figura 4.11: Immagine Originale, Maschera di segmentazione ottenuta dal modello e Risultato finale a seguito dell'estrazione del femore

4.2.2 Analisi della Distribuzione dei Valori

Abbiamo inoltre esaminato la distribuzione dei valori nelle immagini acquisite dal nuovo ecografo per comprendere meglio come il modello gestisce le nuove informazioni.

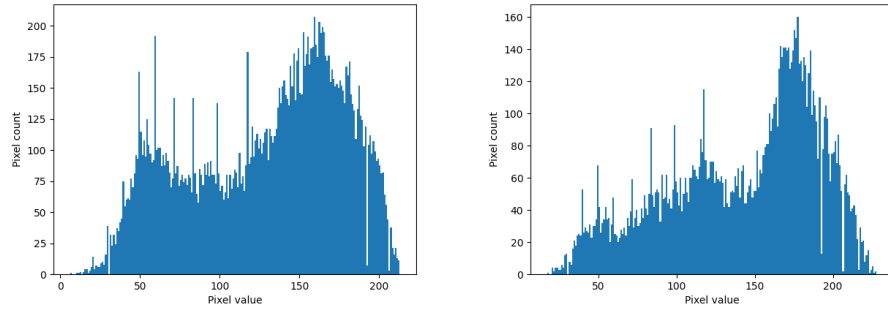


Figura 4.12: Distribuzione dei Valori

4.2.3 Risultati Quantitativi

Di seguito sono riportati i risultati quantitativi derivanti dall'analisi delle immagini del nuovo ecografo.

Immagine	Media	Deviazione Standard
0	128.225	47.961
1	119.319	46.750
2	175.710	41.832
3	141.379	47.968
4	129.142	50.137
5	91.6911	38.671
6	110.249	48.143
7	115.113	43.262
8	155.277	51.628
9	114.937	53.817
10	145.433	57.311
11	152.924	51.922
12	128.146	53.493

Tabella 4.1: Risultati Sperimentali

4.2.4 Discussione

I risultati emersi dall'analisi delle immagini ottenute dal nuovo ecografo indicano che, nonostante il modello non sia stato specificamente addestrato su questi dati, la sua performance qualitativa appare accettabile. Ciò suggerisce un potenziale significativo per miglioramenti attraverso un addestramento

mirato, che potrebbe affinare ulteriormente la capacità del modello di gestire con efficacia set di dati simili.

Capitolo 5

Conclusioni

Il modello sviluppato in questo studio è stato progettato per automatizzare la segmentazione di immagini di femori fetali [Dall'Asta, 2023b] [Dall'Asta, 2023a]. L'utilizzo del modello consente di automatizzare un compito che altrimenti richiederebbe tempo e impegno da parte di un professionista, liberando risorse per altre attività.

Specificamente, il modello è stato impiegato per automatizzare l'estrazione dei pixel del femore di un feto nelle settimane 35-37 di gestazione. I pixel estratti sono stati analizzati per determinare la luminosità, che nelle ecografie è correlata alla **densità minerale ossea (BMD)** del femore fetale, e per studiarne la correlazione con il peso di nascita.

Una correlazione debole è stata rilevata tra la luminosità e il peso alla nascita, come illustrato nella Figura 5.1.

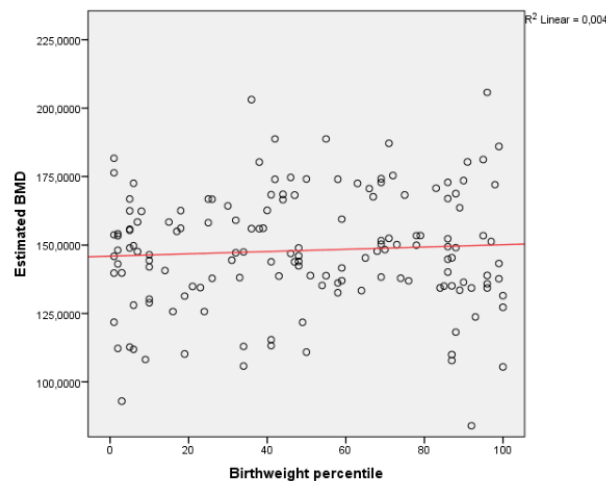


Figura 5.1: Correlazione tra luminosità e peso alla nascita

La rete neurale ha dimostrato di essere efficace nel riconoscere i pixel contenenti informazioni sull'area del femore. Inoltre, i risultati prodotti dalla rete suggeriscono la possibilità di utilizzare questo approccio per ulteriori analisi.

Bibliografia

- [Dall'Asta, 2023a] Dall'Asta, Ferrari, O. I. C. D. P. G. (2023a). Development of an artificial intelligence-based algorithm for the prenatal estimation/quantification of the fetal bone mineral density.
- [Dall'Asta, 2023b] Dall'Asta, Ferrari, O. I. C. D. P. G. (2023b). Prenatal estimation/quantification of the bone mineral density of the fetal femur by means of an artificial intelligence-based algorithm and relationship with birthweight: preliminary results from a pilot study.
- [Fan et al., 2020] Fan, Z., Mo, J., Qiu, B., Li, W., Zhu, G., Li, C., Hu, J., Rong, Y., and Chen, X. (2020). Accurate retinal vessel segmentation via octave convolution neural network.
- [Klein et al., 2018] Klein, A., Warszawski, J., Hillengaß, J., and Maier-Hein, K. H. (2018). Towards whole-body ct bone segmentation. In Maier, A., Deserno, T. M., Handels, H., Maier-Hein, K. H., Palm, C., and Tolxdorff, T., editors, *Bildverarbeitung für die Medizin 2018*, pages 204–209, Berlin, Heidelberg. Springer Berlin Heidelberg.
- [Long et al., 2015] Long, J., Shelhamer, E., and Darrell, T. (2015). Fully convolutional networks for semantic segmentation.
- [Ronneberger et al., 2015] Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation.
- [Team,] Team, P. Pytorch.
- [Wada, 2023] Wada, K. (Sep 25, 2011 – Oct 25, 2023). labelme.