



**UNIVERSITÀ
DEGLI STUDI
DI PADOVA**



**DIPARTIMENTO
DI INGEGNERIA
DELL'INFORMAZIONE**

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE

CORSO DI LAUREA IN INGEGNERIA BIOMEDICA

**“Applicazione di reti neurali per la segmentazione di organi molli in
immagini biomediche”**

Relatore: Dott.ssa ILARIA TONIOLO

Laureando/a: LORENZO BOTTIN

ANNO ACCADEMICO 2022 – 2023

Data di laurea: 21 novembre 2023

Indice

ABSTRACT.....	3
Capitolo 1: IMMAGINI BIOMEDICHE	5
1.1 Tecniche di acquisizione.....	5
1.1.1 Tomografia computerizzata	5
1.1.2 Risonanza magnetica	7
1.2 Analisi delle bioimmagini.....	9
Capitolo 2: RETI NEURALI CONVOLUZIONALI	11
2.1 Storia.....	11
2.2 Struttura.....	12
2.2.1 Modello McCulloch-Pitts	13
2.2.2 Neocognitrone	15
2.3 Applicazioni alla medicina.....	17
2.3.1 Segmentazione.....	17
2.3.2 Algoritmi di segmentazione automatica	18
Capitolo 3: CNN APPLICATI ALL'ANALISI DI TESSUTI MOLLI.....	19
3.1 Analisi comparativa	19
3.1.1 Sistema di acquisizione.....	19
3.1.2 Preelaborazione.....	20
3.1.3 Rete neurale convoluzionale	23
3.2 Tecnica di analisi	27
3.3 Segmentazione multiorgano.....	30
Capitolo 4: DISCUSSIONE	33
Capitolo 5: CONCLUSIONI	37
BIBLIOGRAFIA.....	39

ABSTRACT

Le immagini biomediche sono uno strumento di grande importanza nella diagnosi. Negli ultimi decenni la ricerca e la tecnologia hanno apportato importanti innovazioni nella strumentazione clinica, sia nel caso di esami strumentali di risonanze magnetiche (MRI), sia nel caso di tomografie computerizzate (CT), consentendo un miglioramento di risoluzione delle immagini biomediche e implementando nuove sequenze di acquisizione *ad hoc* per il quesito clinico. Tali migliorie hanno portato ad avere immagini più pesanti, formate da *slices* (o “fette”) più fitte, da cui la necessità di sviluppare algoritmi per automatizzare l’analisi di una tale mode di dati e guidare la diagnosi. A tale scopo sembra promettente l’utilizzo di algoritmi di intelligenza artificiale di reti neurali, con particolare riferimento alle reti convoluzionali (CNN).

In questa tesi verranno presentate tecniche applicative delle CNN per l’analisi di organi molli in MRI e CT. In particolare, i tipi di analisi studiati sono la segmentazione degli organi e la determinazione della linea centrale negli organi tubulari (intestino tenue e vasi sanguigni).

Capitolo 1

IMMAGINI BIOMEDICHE

In questo capitolo saranno presentate le tecniche di acquisizione delle immagini biomediche e come vengono impiegate in diagnosi e per la progettazione di protesi.

1.1 Tecniche di acquisizione

Le immagini biomediche sono uno strumento di grande importanza medica in quanto permettono di osservare l'organismo umano al suo interno senza necessità di eseguire operazioni chirurgiche rischiose per il paziente. Grazie alle immagini i medici riescono a diagnosticare malattie, pianificare ed eseguire interventi chirurgici guidati. È inoltre possibile tramite tecniche di mappatura tridimensionale e utilizzo di appositi software creare protesi su misura per il paziente.

L'avvento delle bioimmagini risale alla fine del '800 quando Wilhelm C. Röntgen scoprì i Raggi X, per mezzo dei quali ottenne una proiezione radiografica di un tessuto umano [1]. Negli anni a venire molti studi vennero condotti a riguardo trasformando le radiografie in una pratica medica di uso quotidiano.

I Raggi X sono un intervallo di frequenze delle onde elettromagnetiche. La tecnica di radioproiezione sfrutta il fenomeno di assorbimento delle onde da parte della materia. Si avrà ~~Avremo~~ quindi delle onde emesse da una sorgente con una certa intensità, le quali superato un corpo materiale (es. tessuto umano) avranno una diversa intensità inferiore in funzione del tipo di materia attraversato [2].

1.1.1 Tomografia computerizzata (TC)

Negli anni '70 venne inventata una strumentazione medica che, sfruttando i raggi X, riproduce immagini in sezione e tridimensionali dell'anatomia: la tomografia computerizzata. Il dispositivo è formato da un lettino sul quale si distende il paziente, da un emettitore (sorgenti di raggi X) e da dei sensori (rilevatori) disposti attorno al lettino. L'emettitore è un tubo radiogeno (figura 1.1) costituito da un contenitore in vetro al boro in cui è stato fatto un vuoto. All'interno sono presenti due filamenti metallici, ai quali è applicata una differenza di potenziale dell'ordine dei kilo Volt, che servono per accelerare gli elettroni. Durante il loro moto gli elettroni impattano sul materiale dell'anodo al quale trasferiscono quasi tutta l'energia in forma cinetica, ma una piccola parte (meno dell'1%) viene convertita in raggi X.

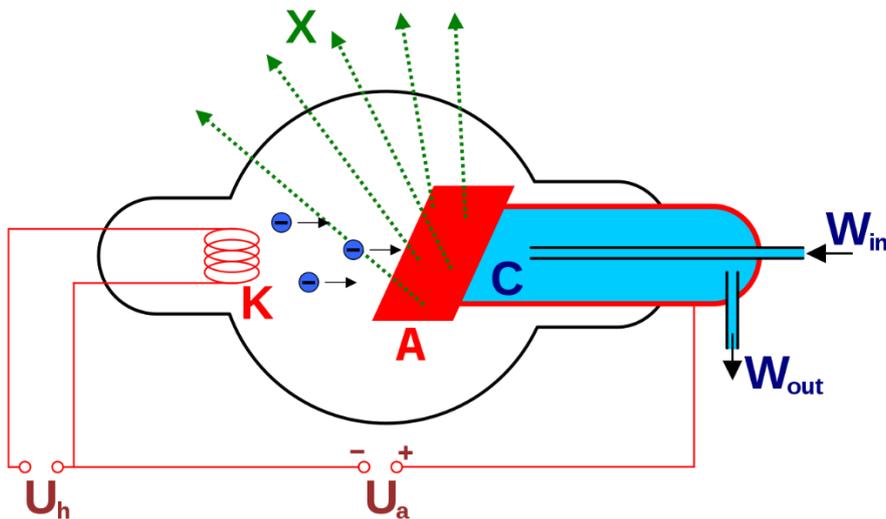


Figura 1.1: schema di un tubo radiogeno con K catodo, A anodo e X raggi di fotoni. Url fonte: https://it.wikipedia.org/wiki/Tubo_radiogeno#/media/File:Roentgen-Roehre.svg.

Esistono diverse generazioni di tomografi che si differenziano per disposizione e traiettoria degli emettitori e dei sensori, oltre che per il tempo di acquisizione delle immagini [2]. Nella prima generazione il sistema è costituito dall'emettitore e da un singolo rivelatore che traslano parallelamente all'asse del lettino e ruotano attorno ad esso. La necessità di due movimenti implica un tempo di acquisizione lungo (dell'ordine della decina di minuti). Nelle successive generazioni è stato aumentato il numero di rilevatori per velocizzare il tempo di acquisizione ed è stata modificata la traiettoria dell'emettitore. Lo stato dell'arte delle tomografie computerizzate prevede la CT elicoidale nella quale l'emettitore ruota attorno al paziente che trasla assialmente, così facendo la sorgente segue una traiettoria elicoidale rispetto al paziente [3]. Il tempo di acquisizione della CT elicoidale è inferiore al minuto, tempo durante il quale il paziente rimane in apnea riducendo gli artefatti dovuti al movimento degli organi interni. Una volta acquisiti i dati vengono elaborati da un calcolatore che li converte in un modello tridimensionale dell'organismo. Tale modello è formato da pixel volumetrici aventi una gradazione di grigio funzione della capacità di assorbimento dei Raggi X di quel volume. Dalla visualizzazione computerizzata il medico vede la posizione e la forma degli organi permettendogli di individuare eventuali anomalie dovute a malattie. Un esempio di macchinario per CT è il Somatom definition AS (figura 1.2) dell'azienda Siemens Medical Solution USA Inc. Ha una risoluzione temporale di 150 ms con un tempo di rotazione minimo dell'emettitore di 0.3 s. Opera con fasci di fotoni a diverse energie generati dal tubo radiogeno con differenza di potenziale di 70, 80, 100, 120 o 140 kV. Altre aziende che sviluppano macchinari CT sono Canon medical sysetm USA, Toshiba IT & Control System Corporation,



Figura 1.2: immagine del Somaton definition AS, uno scanner a tomografia computerizzata dell'azienda Siemens Medical Solution USA Inc. Fonte [4].

Nell'utilizzo della CT è importante considerare gli effetti dannosi dei Raggi X. Essendo raggi ionizzanti hanno effetti collaterali sull'organismo, causano infatti il rischio di sviluppare tumori. Esiste una dose limite di raggi ai quali una persona può essere esposta per non sviluppare effetti dannosi. Quindi esiste un numero limitato di CT che una persona può effettuare da cui la necessità di altri metodi di imaging medico.

1.1.2 Risonanza magnetica (MRI)

Nel 1946 Bloch e Purcell scoprirono il fenomeno della risonanza magnetica nucleare [2]. Questo fenomeno per essere descritto richiede il formalismo della meccanica quantistica, sfrutta infatti le proprietà di spin degli atomi e il modo in cui interagiscono con campi magnetici variabili. Tramite opportuni macchinari di risonanza magnetica è possibile determinare la composizione di un campione senza effettuare test fisico-chimici che lo distruggerebbero. Per questo motivo fin da subito la risonanza magnetica è stata utilizzata nell'ambito della chimica analitica con la spettroscopia MR. In seguito, negli anni '70, si inizia a pensare a come utilizzare il fenomeno per l'imaging medico e negli anni '80 inizia la diffusione in ambito clinico della nuova tecnologia conosciuta come MRI (Magnetic Resonance Imaging). Il suo successo è dovuto alla capacità di discriminazione dei tessuti molli e all'utilizzo di radiazioni elettromagnetiche non ionizzanti sicure per i pazienti.

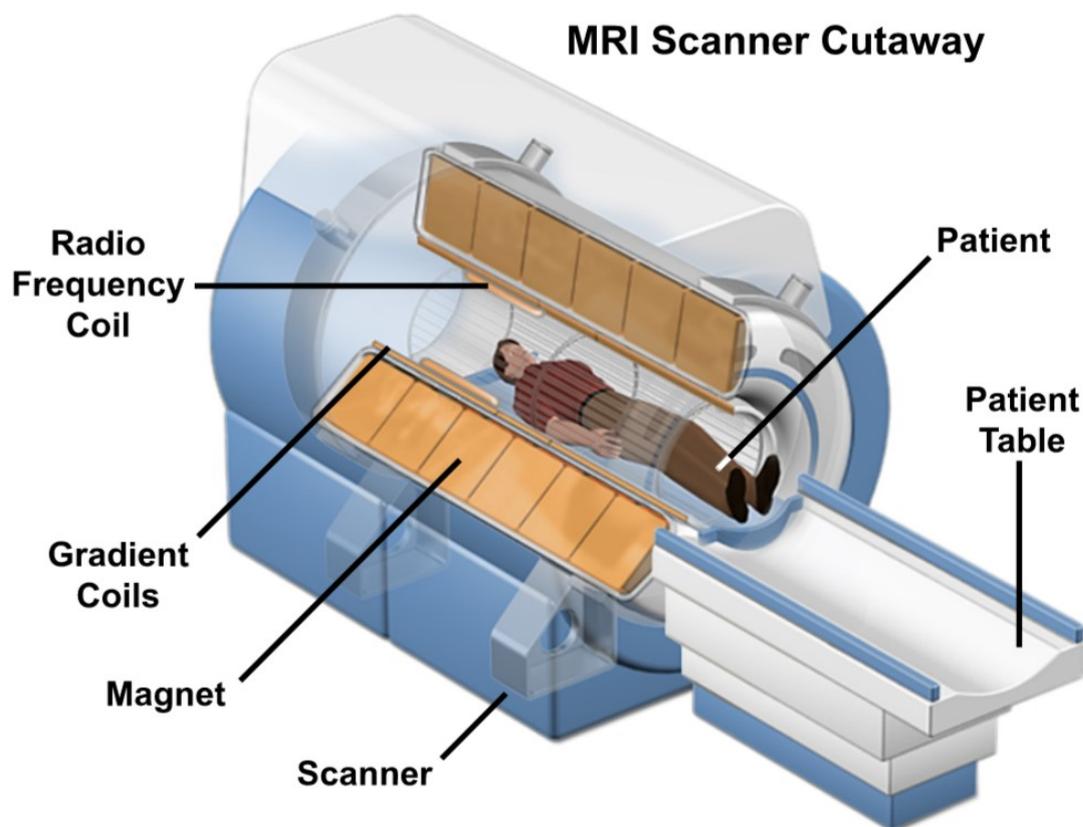


Figura 1.3: schema rappresentativo di un macchinario per MRI. Url fonte: <https://nationalmaglab.org/magnet-academy/read-science-stories/science-simplified/mri-a-guided-tour/>.

Il macchinario per MRI (figura 1.3) è formato da una struttura cilindrica cava nella quale sono disposte delle bobine che generano campi magnetici variabili dell'ordine di decine di milliTesla (mT) e i rilevatori che misurano come interagiscono con i tessuti del paziente. Quest'ultimo si trova disteso su un lettino all'interno del macchinario. È inoltre presente un magnete che genera un campo magnetico statico dell'ordine di qualche Tesla (T). Durante la risonanza al paziente viene chiesto di restare fermo per la durata della procedura che è tipicamente di qualche minuto. La risoluzione temporale non è molto elevata (qualche immagine al secondo), quella spaziale è di alcuni millimetri.

Un esempio di macchinario per MRI è il Magnetom Skyra (figura 1.4) dell'azienda Siemens Medical Solution USA Inc. Ha un campo magnetico statico di 3T. La stessa azienda produce macchinari a 1.5T e 7T. Questi sono i valori tipici dei campi magnetici statici per i macchinari attualmente in commercio anche per le altre aziende produttrici (Philips S.p.A., Canon medical sysetm USA, Fujifilm, ...). Ad oggi non sono stati riscontrati effetti nocivi del campo magnetico statico a nessuna potenza, né per i pazienti né per i medici. Per questo motivo l'MRI è preferita alla TC in casi particolari come le donne incinte per le quali i raggi X potrebbero creare danni al feto.



Figura 1.4: immagine del Magnetom Skyra, un macchinario per risonanza magnetica dell'azienda Siemens Medical Solution USA Inc. Fonte [4].

1.2 Analisi delle bioimmagini

Dopo l'acquisizione dei dati, indipendentemente dalla tecnica utilizzata, è necessaria un'elaborazione computerizzata per rendere le immagini fruibili per l'operatore sanitario. Si inizia integrando le varie misure per ricostruire l'immagine in una scala di grigi dipendente dal parametro misurato. Si passa poi ad un filtraggio utile a migliorare la qualità dell'immagine tramite riduzione del rumore. Una tecnica consiste nell'analizzare lo spettro dell'immagine nel dominio delle frequenze (tramite trasformata di Fourier o altre) ed eliminare o attenuare le frequenze caratteristiche del rumore, in questo modo ritornando al dominio iniziale il rumore risulta attenuato e i dati d'interesse sono meglio visibili.

Dall'immagine filtrata il medico individua elementi utili a diagnosi e operazioni. Ad esempio, nelle cine-MRI (risonanze magnetiche che acquisiscono più immagini in un arco di tempo) i medici misurano manualmente il diametro dell'intestino per valutarne la velocità di contrazione [5]. Dai calcoli ottenuti è possibile individuare eventuali patologie intestinali. Dal punto di vista tecnico queste procedure sono svolte con software dove vengono evidenziati con puntatore i punti di interesse e il computer ne determina la distanza [6]. Tale lavoro deve essere svolto per ogni istante di acquisizione delle immagini e per ogni sezione del modello tridimensionale, il che richiede un elevato dispendio di tempo da parte del medico.

Dalle bioimmagini è inoltre possibile mappare un'area anatomica e sulla base della specifica geometria del paziente costruire una protesi su misura [7]. Questo processo prevede la segmentazione dell'area d'interesse, la creazione di un modello 3D digitale della protesi e la stampa 3D della stessa che verrà poi impiantata nel paziente tramite operazione chirurgica. L'applicazione di questo metodo si estende a varie regioni anatomiche come la creazione di protesi mandibolari, protesi d'anca o ricostruzione dell'arco aortico.

Capitolo 2

RETI NEURALI CONVOLUZIONALI

In questo capitolo verranno presentate le reti neurali convoluzionali (*Convolutional Neural Network*, CNN), come nascono, come sono strutturate e le loro applicazioni nell'analisi delle immagini biomediche.

2.1 Storia

Le reti neurali sono una tipologia di algoritmi utilizzati nell'analisi di immagini, audio, linguaggi e molti altri tipi di dati. Vengono impiegati in svariati ambiti dalla medicina [8] alla finanza, dai videogiochi alle scienze atmosferiche [9]. Negli ultimi anni le loro potenzialità e utilizzo sono ampiamente cresciute, diventando di interesse per molte attività produttive e di ricerca.

Come si evince dal nome, le reti neurali derivano dai neuroni, le cellule responsabili delle capacità percettive e intellettive umane. Nel '900 grazie a studi in ambito anatomico e fisiologico del sistema nervoso si era a conoscenza della struttura e il funzionamento dei neuroni come elementi singoli. Rimaneva però ignoto il funzionamento delle reti del sistema nervoso e di come queste fossero capaci, ad esempio, di elaborare semplici percezioni visive e riconoscere gli oggetti che vediamo. Per questo desiderio di conoscenza sono stati condotti studi con l'obiettivo di modellizzare matematicamente i neuroni e le loro interazioni. Primo fra questi è il modello matematico delle reti neurali fatto da McCulloch e Pitts nella pubblicazione "A logical calculus of the ideas immanent in nervous activity" (1943) [10], nel quale i neuroni vengono paragonati a degli elaboratori con logica booleana aventi una soglia di attivazione e un'uscita binaria. Nel 1958 Rosenblatt pubblica "The Perceptron: A Probabilistic Model for Information Storage and Organization in the Brain" [11], articolo nel quale viene sviluppato il modello McCulloch-Pitts creando un layer di "neuroni" dotato di capacità di apprendimento grazie alla presenza di pesi per gli ingressi modificabili in apprendimento. Grazie a questa rete riuscì ad insegnare a un computer come distinguere delle carte contrassegnate sulla sinistra da altre contrassegnate sulla destra. Queste due ricerche sono alla base di tutti i tipi di reti neurali. Nello specifico la tipologia di rete convoluzionale ha avuto sviluppo grazie al lavoro di Fukushima che nel 1980 pubblica "Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position" [12]. In questo articolo viene presentata la rete neurale Neocognitron considerata il predecessore delle reti

convoluzionali. Nel 1989 LeCun et al. nel loro lavoro creano una rete capace di riconoscere caratteri scritti a mano (LeNet) [13] e usano per la prima volta il termine convoluzione applicato alle reti neurali. Questa rete viene ulteriormente sviluppata negli anni fino alla creazione di LeNet-5 considerato il modello base delle reti convoluzionali [14]. Negli anni a seguire sono state sviluppate svariate reti convoluzionali ognuna con le proprie caratteristiche utili a svolgere specifiche attività, ma ciò che più di tutto ha permesso il progredire delle reti è stato il miglioramento tecnologico degli hardware che ha portato un enorme capacità di calcolo necessaria al funzionamento delle reti.

2.2 Struttura

Esistono diversi tipi di reti neurali, ma si basano tutti su un modello comune [16]. Questo modello prevede un elemento base, il *nodo* (Figura 2.1), che costituisce l'unità di elaborazione. Il nodo ha una struttura simile ad un neurone: un corpo centrale a cui afferiscono un numero variabile di sinapsi che generano un impulso trasmesso tramite l'assone, questi elementi nelle reti neurali si traducono in una *funzione di nodo* con dei parametri generati da nodi precedenti e che fornisce un valore d'uscita a nodi successivi, in formula:

$$Output = y = f(x_1, x_2, \dots, x_n) \quad (2.1)$$

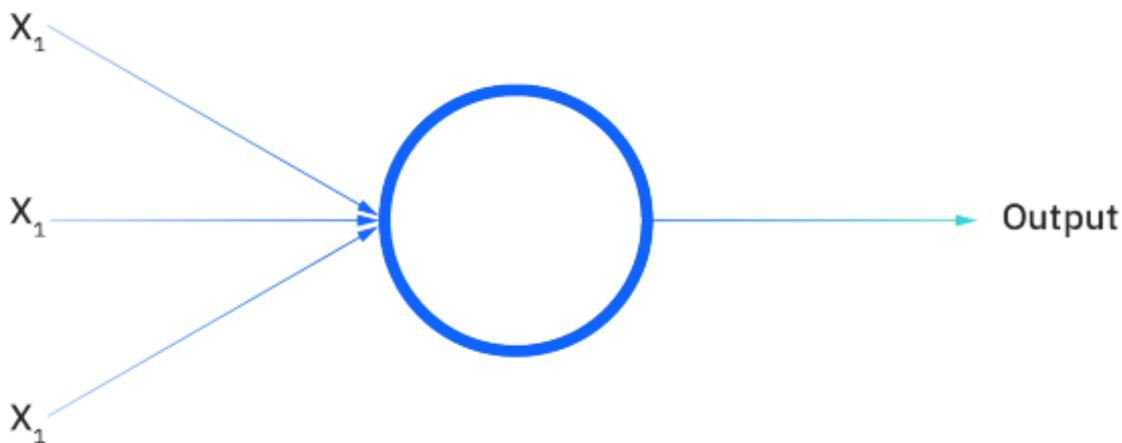


Figura 2.1: immagine che schematizza l'elemento nodo di una generica rete neurale. A sinistra si vedono tre generici ingressi collegati tramite sinapsi (freccette) al nodo che genera l'uscita (Output) secondo una qualche funzione di nodo. Immagine tratta da [15].

I nodi sono collegati tra loro a formare delle reti. Queste reti sono classificate in base a eventuali percorsi chiusi dei nodi, ovvero partendo da un nodo e proseguendo verso l'uscita è possibile ritornare al nodo iniziale. Tali reti sono definite cicliche e sono di ordine n , dove n è il numero di nodi che costituiscono il circolo massimo. Le reti convoluzionali di nostro interesse sono di

ordine zero ovvero non hanno cicli. Pertanto, di seguito è spiegata solo la struttura di quest'ultime.

Le reti di ordine zero hanno i nodi organizzati in livelli come in Figura 2.2: uno iniziale che costituisce l'ingresso dei dati (input layer), un numero variabile di livelli centrali nascosti (multiple hidden layers) e un livello finale che costituisce l'uscita della rete (output layer).

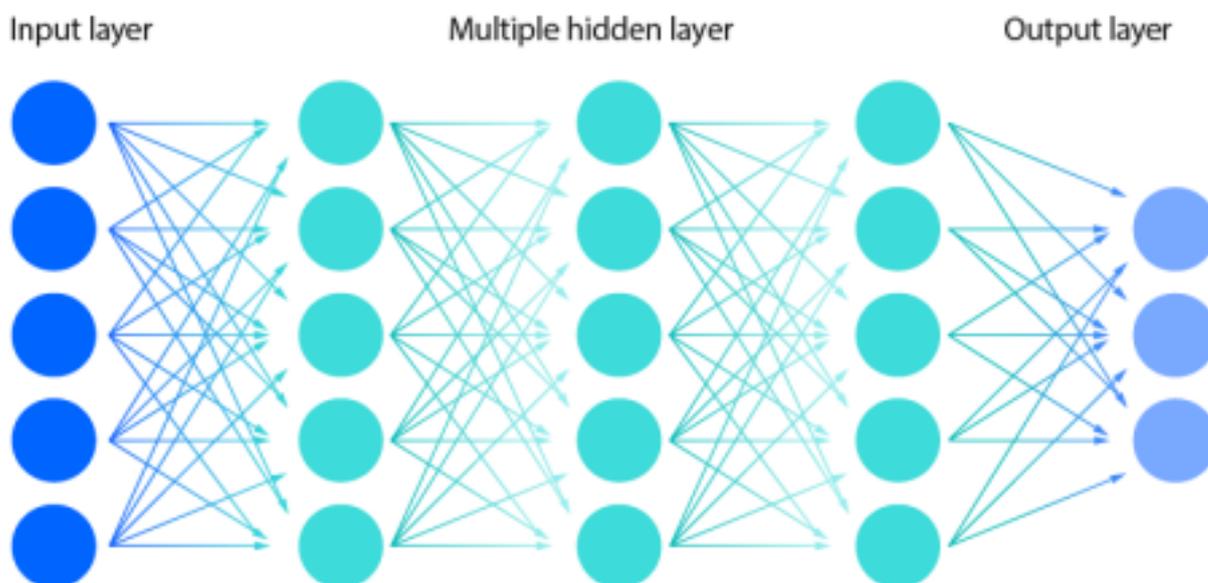


Figura 2.2: Immagine che rappresenta un'ipotetica rete neurale con in livello di ingresso (input layer), una serie di livelli nascosti (multiple hidden layers) e un livello di uscita (output layer). Immagine tratta da [15].

Al fine di esemplificare la struttura vengono di seguito presentati i modelli storici di McCulloch-Pitts e Fukushima.

2.2.1 Modello McCulloch-Pitts [10]

Questo modello è il più semplice e da questo si parte per comprendere il funzionamento delle reti neurali. Ci sono cinque ipotesi alla base:

1. L'attività dei neuroni è del tipo tutto o niente.
2. Un determinato numero di sinapsi deve essere attivo entro un certo periodo di tempo per poter iniziare un impulso nel neurone, e questo numero è indipendente dalla precedente attività e posizione del neurone.
3. L'unica differenza temporale significativa è quella delle sinapsi.
4. L'attività inibitoria previene l'eccitazione del neurone in ogni momento.
5. La struttura della rete non cambia nel tempo.

In queste ipotesi il neurone rappresenta l'elemento *nodo* della struttura precedentemente definita. Da queste assunzioni possiamo affermare che la logica dei nodi è di tipo booleano, ovvero i parametri e l'uscita della *funzione di nodo* hanno dominio $\{0,1\}$. Inoltre, affinché un nodo abbia uscita pari a 1 la somma degli ingressi attivi deve superare un valore soglia e questo valore non dipende da eventi passati: il nodo non ha memoria. L'unico modo in cui una rete così definita può "ricordare" è attraverso la presenza di circoli in cui i nodi si attivano mutuamente per un tempo indefinito. Infine, se è presente un collegamento inibitorio attivo l'uscita del nodo è 0.

Si va ora a presentare la *funzione di nodo* per farlo definiamo alcuni parametri. Sia n il nodo considerato con soglia di attivazione θ e siano $N_{e1}, N_{e2}, \dots, N_{ep}$ le sinapsi eccitatorie e $N_{i1}, N_{i2}, \dots, N_{iq}$ le sinapsi inibitorie agenti su n con attività $N(t)$ all'istante t . Sia k l'insieme dei sottoinsiemi di $\{N_{e1}, N_{e2}, \dots, N_{ep}\}$ tale per cui la loro somma supera θ . La *funzione di nodo* risulta:

$$y = \prod_{m=1}^q \sim N_{im}(t) \wedge \sum_{a \in k} \prod_{s \in a} N_{es}(t) \quad (2.2)$$

Dove il simbolo di produttoria indica una congiunzione e la sommatoria una disgiunzione. Di seguito il suo equivalente in codice informatico:

```

IF <sum of Ni> > 0
    y = 0
ELSE IF <sum of Ne> > θ
    y = 1
ELSE
    y = 0

```

Si nota che per reti di ordine 0 se sostituiamo ricorsivamente nella funzione i valori delle attività dei nodi agenti con le rispettive funzioni di nodo allora la funzione dipende unicamente dai valori dei nodi di ingresso della rete nell'istante considerato.

A titolo di esempio viene riportata la figura 2.3 appartenente alla pubblicazione di McCulloch e Pitts dove possiamo vedere quattro schemi di reti neurali rappresentanti le espressioni logiche:

- a. Identità
- b. $N_1 \vee N_2$
- c. $N_1 \wedge N_2$
- d. $N_1 \wedge \sim N_2$

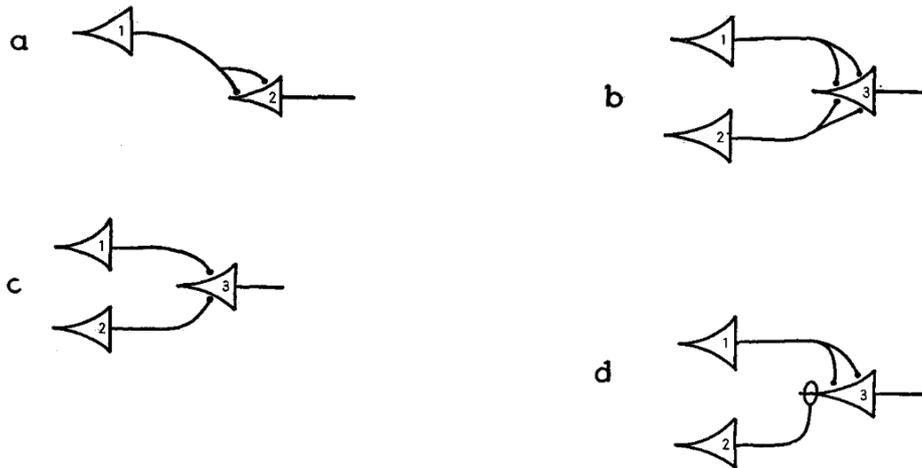


Figura 2.3: Immagine che raffigura quattro esempi di rete neurale creati basandosi sul modello McCulloch e Pitts. Immagine tratta da [10].

McCulloch e Pitts enunciano una serie di teoremi che definiscono le proprietà del loro modello. In particolare, i primi due affermano che:

- Ogni rete esprime una proposizione booleana;
- Ogni proposizione booleana può essere espressa da una rete.

Questi due teoremi definiscono l'ambito di funzionamento del modello che, pur essendo semplice, è in grado di rappresentare molte operazioni elementari. A queste reti neurali manca la capacità di apprendere, presente invece nel modello del perceptrone di Roseblatt che introduce la somma pesata e non più semplice delle sinapsi eccitatorie con i valori dei pesi che variano durante la fase di apprendimento. Inoltre, l'ipotesi di un dominio booleano per l'uscita dei nodi ne limita le potenzialità, infatti, già nel perceptrone il dominio viene notevolmente espanso.

2.2.2 Neocognitrone [12]

Il neocognitrone ha una struttura più complessa del modello McCulloch-Pitts che si avvicina agli algoritmi moderni. Come mostrato in figura 2.4, la rete è organizzata con un livello input U_0 e una sequenza variabile di moduli ognuno dei quali formato da due livelli ($U_{S1} - U_{C1}$). Ogni livello è a sua volta costituito da un numero variabile K_1 di piani contenenti i nodi organizzati in due dimensioni. Grazie a questa caratteristica bidimensionale la rete è in grado di analizzare dati come immagini tenendo in considerazione la loro struttura planare. Ogni modulo ha le stesse caratteristiche: delle sinapsi modificabili in ingresso al livello U_S , grazie alle quali la rete è in grado di apprendere, e delle sinapsi statiche che collegano U_S e U_C .

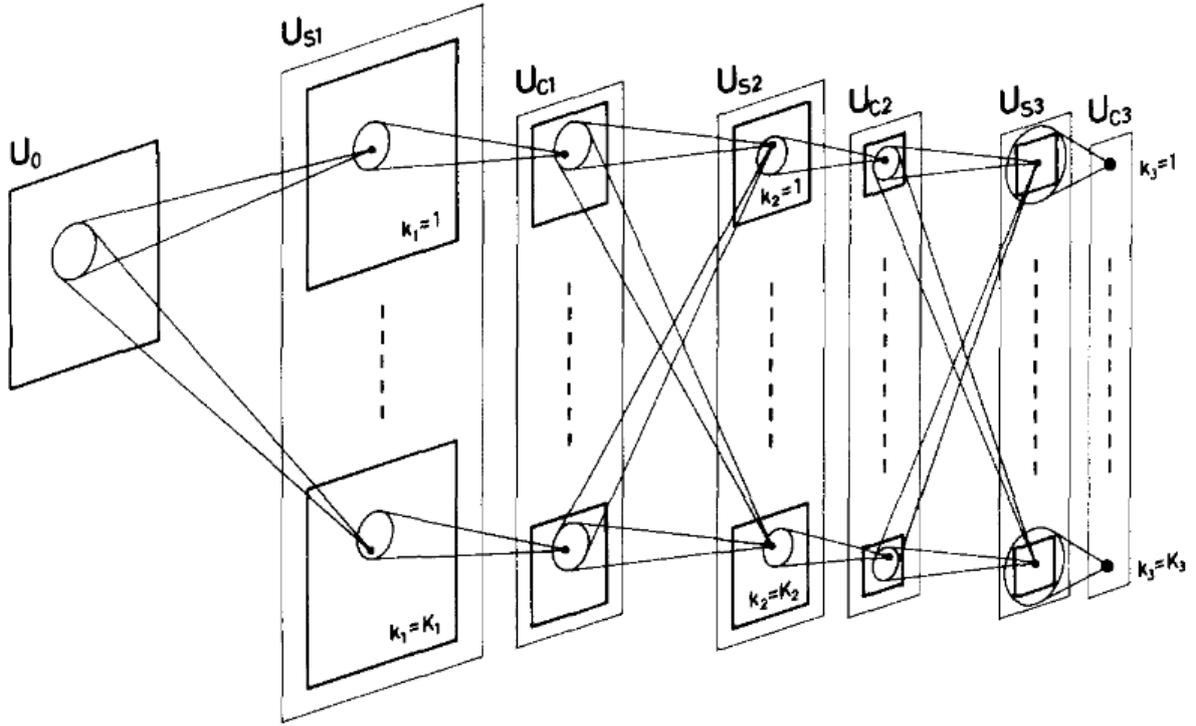


Figura 2.4: Immagine che raffigura lo schema logico del neocognitrone. I rettangoli disegnati con linea sottile rappresentano i livelli e i quadrati con linea spessa i piani. Immagine tratta da [12].

Di seguito vengono presentate le funzioni di nodo e le funzioni di apprendimento della rete, senza eseguirne un'analisi approfondita, ma concentrandosi su quegli aspetti che rendono la rete il primo esempio di rete neurale convoluzionale.

Definiamo $u_{Sl}(k_l, n)$ e $u_{Cl}(k_l, n)$ le uscite dei nodi eccitatori rispettivamente del livello S e C del l -esimo modulo appartenenti al piano k_l in posizione $n = (n_x, n_y)$, e siano $v_{Sl}(n)$ e $v_{Cl}(n)$ i nodi inibitori. La funzione di un nodo appartenente al livello S ha formula:

$$u_{Sl}(k_l, n) = r_l \varphi \left[\frac{1 + \sum_{k_{l-1}=1}^{K_{l-1}} \sum_{v \in S_l} a_l(k_{l-1}, \mathbf{v}, k_l) \cdot u_{Cl-1}(k_{l-1}, n + \mathbf{v})}{1 + \frac{2r_l}{1+r_l} b(k_l) \cdot v_{Cl-1}(n)} - 1 \right] \quad (2.3)$$

dove $a_l(k_{l-1}, \mathbf{v}, k_l)$ e $b(k_l)$ sono i pesi applicati alle sinapsi che vengono modificati durante l'apprendimento e r_l un parametro da cui dipende l'efficacia delle sinapsi inibitorie. φ è una funzione molto presente nelle reti neurali perché introduce la non linearità, essenziale per analizzare caratteristiche complesse dei dati in ingresso. Questa funzione è conosciuta come ReLU (Rectified Linear Unit) ed è così definita:

$$\varphi[x] = \begin{cases} x & \text{se } x > 0 \\ 0 & \text{altrimenti} \end{cases}$$

Possiamo notare che i valori di uscita dei nodi sono di tipo analogico ovvero valori maggiori o uguali a 0. In (2.3) la sommatoria su S_l è ciò che caratterizza questa rete come convoluzionale. Infatti, il filtro bidimensionale definito dai pesi $a_l(k_{l-1}, \mathbf{v}, k_l)$ viene applicato al livello precedente in un intorno ad n definito da S_l , questo è fatto per ogni n . Tale procedimento ha come scopo la ricerca di un determinato schema che per le immagini è ad esempio la differenza di intensità dei colori per l'individuazione dei bordi di un oggetto.

Di seguito è riportata la funzione di un nodo appartenente al livello C:

$$u_{Cl}(k_l, n) = \psi \left[\frac{1 + \sum_{\mathbf{v} \in D_l} d_l(\mathbf{v}) \cdot u_{Sl}(k_l, n + \mathbf{v})}{1 + v_{Sl}(n)} - 1 \right]$$

ha una struttura molto simile alla controparte del livello S. Cambiano i pesi delle sinapsi eccitatorie $d_l(\mathbf{v})$ che sono dei valori predefiniti che non variano durante l'apprendimento. Cambia anche la funzione non lineare che è:

$$\psi[x] = \varphi \left[\frac{x}{\alpha + x} \right]$$

dove α è un ulteriore parametro predefinito.

Con questa rete neurale Fukushima riuscito ad addestrare un elaboratore a distinguere i numeri e le lettere ottenuti in ingresso come immagine con gradazione di grigi. La rete è in grado di riconoscere i caratteri anche in caso di distorsioni dell'immagine e presenza di rumore.

2.3 Applicazioni alla medicina

Algoritmi come le reti neurali stanno ottenendo un vasto impiego nelle tecniche di analisi computazionale per le immagini biomediche: computer aided detection (CADx) e computer aided diagnosis (CAD). Queste tecniche prevedono come risultato finale la segmentazione, classificazione e individuazione di anomalie [8]. Per segmentazione si intende la suddivisione dell'immagine in aree logiche per il riconoscimento degli oggetti in essa presenti. La classificazione è la capacità di distinguere gli oggetti, ovvero se nell'immagine è presente il cuore piuttosto che l'intestino. L'individuazione di anomalie vuol dire individuare difetti anatomici o funzionali come la presenza di tumori. Di seguito approfondiremo la segmentazione.

2.3.1 Segmentazione

Lo scopo della segmentazione precedentemente descritto è quello di dividere l'immagine in base agli oggetti in essa presenti individuando delle regioni d'interesse (ROIs). È uno strumento

utile al medico, il quale ha necessità di studiare la struttura o l'organo d'interesse nella sua forma tridimensionale che non è visibile nelle immagini CT e MRI perché sono bidimensionali. La segmentazione manuale richiede un tempo oneroso che può essere evitato usando algoritmi di segmentazione automatica. Allo stato attuale algoritmi di questo tipo hanno precisione e affidabilità sufficiente solo per la segmentazione di alcune parti anatomiche come le ossa che hanno un elevato contrasto in immagini da CT e MRI. Per altri organi come vasi sanguigni o intestino, in generale tutti i tessuti molli, il contrasto non è sufficiente per segmentarli con facilità. Questo problema viene parzialmente risolto nell'intestino e nei vasi sanguigni tramite l'iniezione di un mezzo di contrasto che è capace di interagire con i raggi X della CT o con i campi magnetici dell'MRI rendendo evidente la ROI.

2.3.2 Algoritmi di segmentazione automatica

Nella segmentazione di regioni ad elevato contrasto vengono impiegati algoritmi classici come la proiezione a massima intensità (maximum intensity projection MIP) o la proiezione a minima intensità (minimum intensity projection MinIP) ottenendo ottimi risultati [17]. Per i tessuti molli invece gli algoritmi classici non sono efficaci, si stanno quindi svolgendo promettenti ricerche sull'impiego delle reti neurali convoluzionali (CNN). I problemi di questi algoritmi sono il lungo tempo di addestramento, l'elevata differenza anatomica tra pazienti diversi e la necessità di avere le immagini già segmentate per poterle usare come verità sulla base della quale l'algoritmo apprende. Per rendere l'algoritmo sufficientemente preciso è necessario sottoporli una grande mole di dati da cui apprendere [8]. In molti casi l'apprendimento è semi-assistito, ovvero un operatore umano deve comunicare alla rete l'uscita corretta, così che possa sapere come migliorarsi. Quest'operazione consiste nella segmentazione manuale dei dati di addestramento e tali dati non sono facilmente reperibili.

Capitolo 3

CNN APPLICATI ALL'ANALISI DI TESSUTI MOLLI

In questo capitolo è presentato lo stato dell'arte dell'uso delle reti neurali convoluzionali per la segmentazione e diagnostica di organi molli. Vengono presi in considerazione articoli relativi alla segmentazione di diverse aree anatomiche: intestino tenue ([18], [19]), vasi sanguigni ([20], [21], [22], [23]), nervi ([24]) e sistema muscoloscheletrico ([25]). Sono tutti articoli che si propongono di sviluppare metodi di segmentazione o diagnostica eccetto per [25] che è un'indagine comparativa condotta su 101 articoli sull'argomento.

3.1 Analisi comparativa

I metodi presentati negli articoli sono accumulati tutti dall'uso di reti convoluzionali ma differiscono per molti aspetti: strumenti di acquisizione, struttura della rete neurale, scopo dell'applicazione, e altri. Di seguito vengono analizzate le differenze.

3.1.1 Sistema di acquisizione

Come presentato nel Capitolo 1, esistono diverse strumentazioni e diversi principi fisici che ci permettono di acquisire immagini mediche. In tabella 3.1 vediamo le tecniche usate negli studi analizzati. Gli scanner ottici sono *TopCon TRV-50* e *Canon CR5 nonmydriatic 3CCD* in [22], e *Heidelberg Retina Tomograph* in [24]. Entrambi questi articoli si occupano della regione oculare, il primo analizza i vasi sanguigni della retina e il secondo i nervi della cornea. La particolarità di quest'area anatomica è la possibilità di osservare con metodi ottici non invasivi strutture anatomiche che normalmente sono visibili solo con operazioni. Inoltre, la strumentazione usata ha costi economici di molto inferiori a quelli di MRI e CT. Per questi motivi tale regione è ampiamente studiata per la diagnosi di patologie.

Tecnica di acquisizione	Articoli
MRI	[18] [19]
CT	[20] [21]
Ultrasuoni	[23]
Scanner ottico	[22] [24]

Tabella 3.1: Tecniche di acquisizione sulla sinistra e relativi articoli che le utilizzano sulla destra.

Al fine di questa tesi sono stati raccolti gli articoli in modo da mostrare come le reti neurali siano applicabili a tutti i tipi di immagini mediche, sono stati presi uno/due studi per tecnica di acquisizione. Nello studio [25], invece, sono state comparate numerose ricerche e in tabella 3.2

si vede per ogni tipologia di immagine quante ricerche la hanno utilizzata. Per la suddivisione in regioni anatomiche e i collegamenti agli articoli si rimanda alla tabella 1 dell'articolo [25]. La tabella 3.2 è rappresentativa dello stato dell'arte della ricerca nella segmentazione del sistema muscoloscheletrico. Si nota un maggiore interesse nei confronti delle tecniche di MRI e CT rispetto ai Raggi X e agli ultrasuoni. Nonostante [25] si limiti ad analizzare una specifica regione anatomica l'interesse per MRI e CT è generalizzabile a tutti gli organi interni dell'organismo eccezion fatta per l'occhio che attraverso la pupilla rende visibile il suo interno anche tramite scanner ottici.

Tecnica di acquisizione	Numero di studi
MRI	41
CT	43
Raggi X	10
Ultrasuoni	10

Tabella 3.2: Tecniche di acquisizione sulla sinistra e relativo numero di studi che le utilizzano sulla destra. I dati sono relativi all'articolo "Deep Learning-Based Medical Images Segmentation of Musculoskeletal Anatomical Structures: A Survey of Bottlenecks and Strategies" [25].

Dalla tabella 3.2 notiamo che sono maggiormente utilizzate come tecniche di acquisizione MRI e CT. Questa preferenza è riscontrabile anche per gli altri organi molli come intestino e sistema pelvico. Il motivo è che tali tecniche hanno una migliore risoluzione e creano un buon contrasto per gli organi molli permettendone una migliore analisi. La ricerca non è comunque concentrata unicamente su quei metodi perché esistono diversi vantaggi nell'utilizzo delle altre tecniche: i costi economici di utilizzo sono inferiori, lo spazio richiesto dai macchinari è ridotto, le macchine per MRI e CT richiedono una stanza a loro dedicata invece, il macchinario per ultrasuoni è sufficientemente piccolo da poter essere impiegato anche in piccoli ambulatori e non solo nei grandi ospedali.

3.1.2 Preelaborazione

Successivamente all'acquisizione dei dati avviene la parte di elaborazione, che ha come prima fase la preelaborazione. Tale fase non viene sempre applicata e si passa direttamente all'uso della rete neurale, come possiamo notare dalla tabella 3.3.

Con preelaborazione	Senza preelaborazione
[19] [21] [23]	[18] [19] [21] [23]

Tabella 3.3. Ricerche divise in base alla presenza o assenza di tecniche di preelaborazione.

Negli studi analizzati, gli algoritmi di preelaborazione prevedono analisi in base ai colori o alle frequenze.

Nell'articolo "Automatic coronary artery segmentation of CCTA images using UNet with a local contextual transformer" [20] viene applicato un filtro che seleziona l'ampiezza di unità Hounsfield (HU) relativa alle arterie. HU è un'unità di misura relativa alla radio densità. Con questo metodo vengono risaltate le arterie e viene eliminato del rumore dell'immagine (figura 3.1).

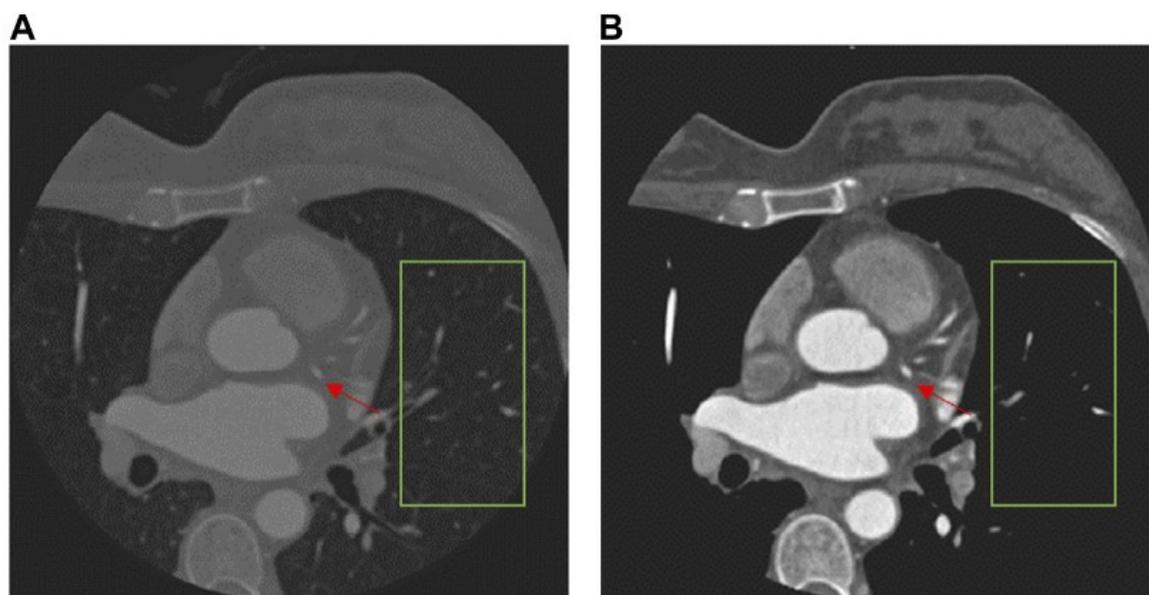


Figura 3.1: A sinistra l'immagine prima dell'applicazione del filtro e a destra dopo il filtro. L'immagine è tratta dalla figura 1 dell'articolo "Automatic coronary artery segmentation of CCTA images using UNet with a local contextual transformer" [20].

Nell'articolo "Retinal Vessel Segmentation Algorithm Based on Residual Convolutional Neural Network" [22] viene applicato un altro tipo di preelaborazione. Inizialmente dei tre canali RGB viene selezionato il canale verde (green) che ha un maggiore contrasto per le regioni d'interesse (figura 3.2) e tale canale viene trasformato in un'immagine a scala di grigi. Successivamente viene applicato un filtro Gaussiano per eliminare il rumore. Infine, viene applicato il metodo di equalizzazione dell'istogramma per migliorare il contrasto dell'immagine (figura 3.3).

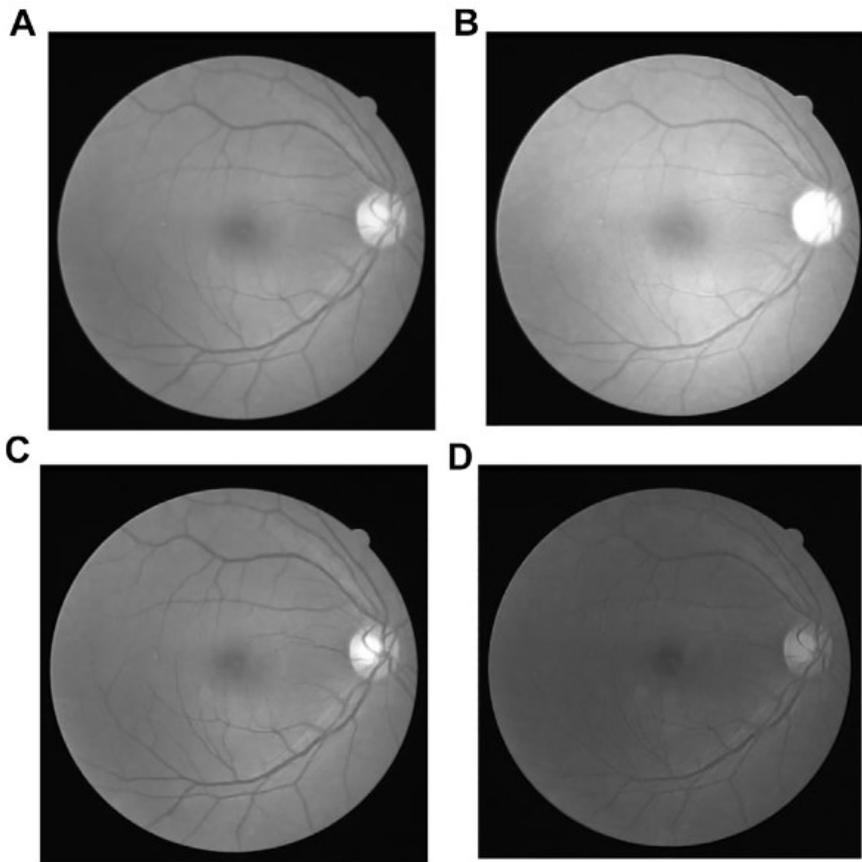


Figura 3.2: l'immagine originale in scala di grigi (A) e la sua divisione in canali, rosso (B), verde (C) e blu (D). L'immagine è tratta dalla figura 6 dell'articolo [22].

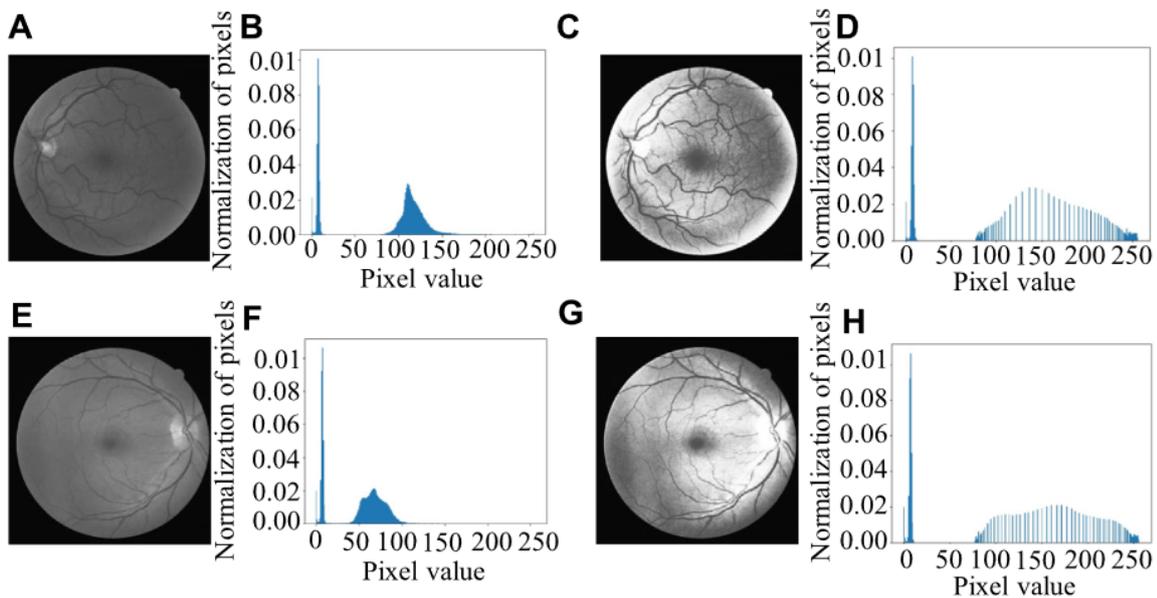


Figura 3.3: immagine a scala di grigi del canale verde (A, E) e relativi istogrammi dei valori dei pixel (B, F), immagine dopo l'equalizzazione degli istogrammi (C, G) e relativi istogrammi (D, H). L'immagine è tratta dalla figura 8 dell'articolo [22].

Infine, nell'articolo "Segmentation of Corneal Nerves Using a U-Net-Based Convolutional Neural Network" come preelaborazione viene prima applicata una finestra dell'immagine,

viene cioè scartato il bordo con uno spessore di dieci pixel perché a causa della strumentazione utilizzate tale area risulta poco illuminata e talvolta sfocata. Dopo l'immagine viene ridimensionata di un fattore 0,7 per ottenere una dimensione di 256 x 256 pixel.

Un altro tipo di preelaborazione non considerata nella tabella 3.3 è la *data augmentation* che viene utilizzata in quasi tutti gli articoli. Il suo scopo è quello di aumentare il dataset disponibile per l'addestramento della rete neurale. L'elaborazione prevede di applicare delle modifiche alle immagini disponibili ruotandole, traslandole o aggiungendo del rumore. Questa tecnica è essenziale in questo tipo di studi perché le immagini MRI e CT rese disponibili per ricerca sono limitate e questa carenza rende difficile giungere a risultati significativi.

3.1.3 Rete neurale convoluzionale

La maggior parte degli articoli utilizza la rete neurale U-Net o un suo sviluppo. Questa rete è stata presentata per la prima volta nel 2015 da Ronneberger, Fischer e Brox nell'articolo "U-Net: Convolutional Networks for Biomedical Image Segmentation" [26]. U-Net (figura 3.4) è una rete convoluzionale organizzata a gradini. Ad ogni gradino corrisponde un livello di astrazione, che analizza le caratteristiche dell'immagine, e una riduzione della dimensione della matrice. I gradini sono organizzati in una scala discendente a cui segue una scala ascendente di egual numero di gradini, formando così una U, da cui il nome della rete. Nella parte discendente ogni gradino è chiamato encoder ed è formato da due livelli convoluzionali e si collega al gradino successivo tramite un livello di max pool 2x2 che riduce le dimensioni della matrice. Arrivati in fondo alla scala si inizia a risalire tramite i decoder formati da un livello di up-conv 2x2 che aumenta le dimensioni della matrice al quale viene aggiunta la matrice del corrispondente encoder, seguono due livelli convoluzionali. In tutti i livelli convoluzionali è applicata una convoluzione 3x3 alla quale viene applicata la funzione di attivazione ReLU.

Nelle ricerche qui presentate due utilizzano U-Net ([23] [24]), una il suo sviluppo per analisi di immagini 3D (3D U-Net, [18]), un'altra usa U-Net con l'aggiunta di alcuni moduli di elaborazione ([20]). Le altre ricerche sfruttano altri tipi di CNN, una rete chiamata residual CNN ([22]) e un'altra una rete molto simile a U-Net che presenta la sequenza encoder-decoder ([19]).

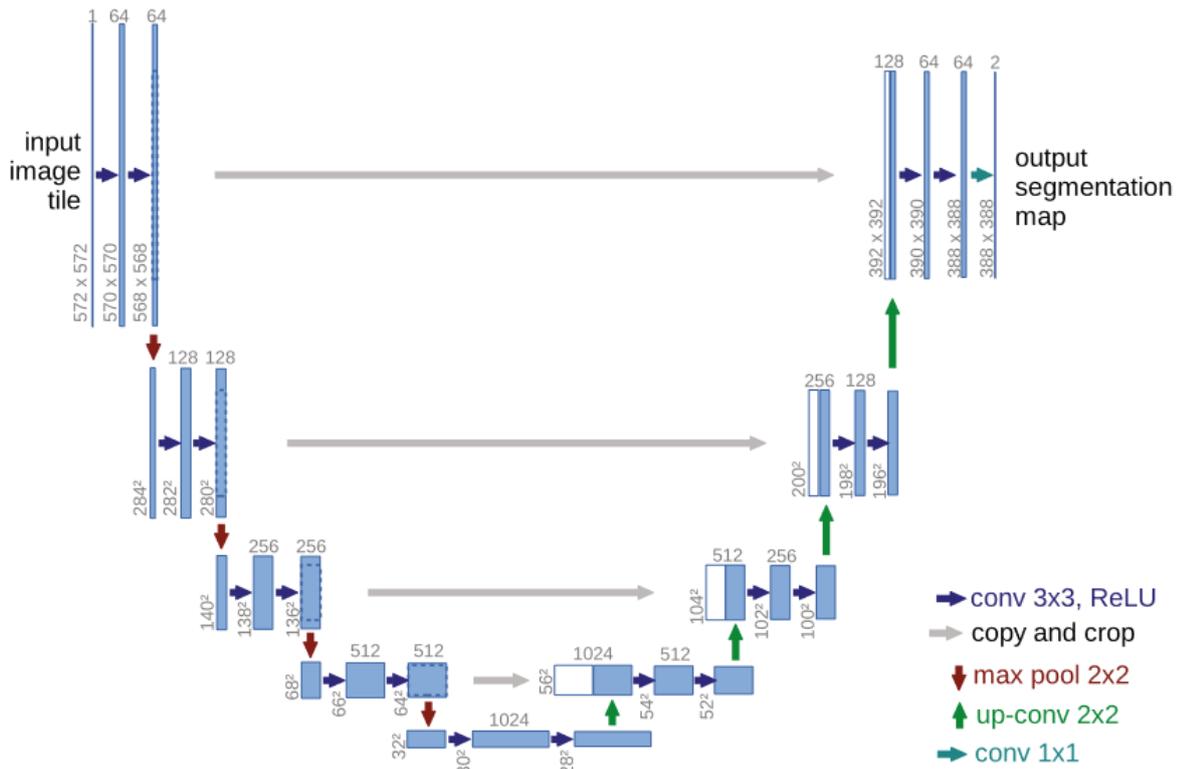


Figura 3.4: schema della rete neurale convoluzionale U-Net. L'immagine è tratta dall'articolo di Ronneberger et al. [26].

La rete di maggiore interesse è quella presentata in “Automatic coronary artery segmentation of CCTA images using UNet with a local contextual transformer” [20]. Questa rete, che è uno sviluppo di U-Net, presenta risultati significativamente migliori dell'uso di U-Net. Nello specifico la rete si chiama DR-LCT-UNet con DR e LCT che sono i nomi di due moduli, rispettivamente dense residual module e local contextual transformer, aggiunti nello schema di U-Net come mostrato in figura 3.5. In tabella 3.4 vediamo i risultati dei test condotti con i diversi moduli dove SA indica il self-attention module, un modulo presente in alcune reti neurali che ha lo scopo di guidare l'apprendimento della rete.

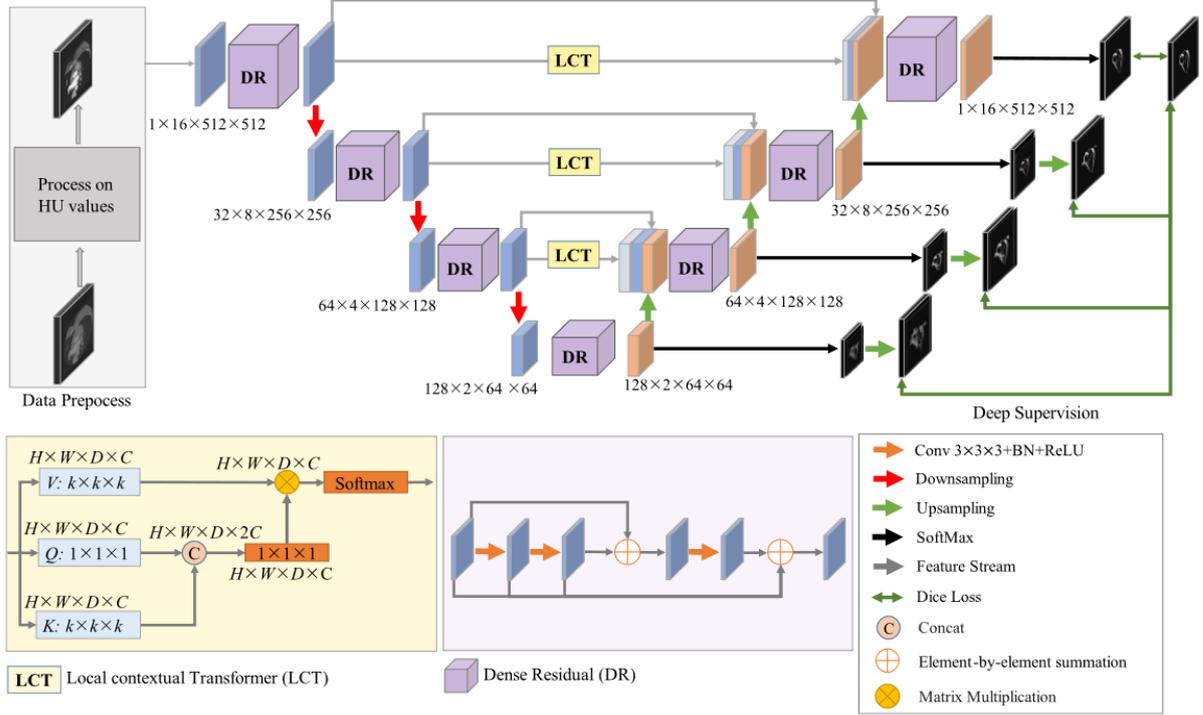


Figura 3.5: schema della rete neurale convoluzionale DR-LCT-UNet. L'immagine è presa dall'articolo di Q. Wang et al. [20].

Method	SA	LCT	DR	R	DSC↑	Recall↑	Precision↑
3D-UNet					0.837	0.844	0.837
SA-UNet	✓				0.843	0.835	0.857
LCT-UNet		✓			0.852	0.846	0.869
R-UNet				✓	0.841	0.832	0.882
DR-UNet			✓		0.852	0.863	0.847
DR-LCT-UNet		✓	✓		0.858	0.863	0.858

Tabella 3.4: risultati degli esperimenti condotti con varie reti neurali. Legenda: self-attention module (SA), local contextual transformer (LCT), dense residual module (DR), residual block (R). Le spunte indicano quali moduli vengono utilizzati. L'immagine è tratta dall'articolo di Q. Wang et al. [20].

Il confronto è eseguito su tre parametri statistici: precisione (precision), recupero (recall) e Dice similarity coefficient (DSC). Questi parametri hanno formule:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (3.1)$$

$$\text{Recall} = \frac{TP}{TP + FN} \quad (3.2)$$

$$\text{DSC} = \frac{2TP}{2TP + FN + FP} \quad (3.3)$$

Dove TP indica i veri positivi (true positive), FP i falsi positivi (false positive) e FN i falsi negativi (false negative). Positivo e negativo indica se la rete neurale ha determinato un pixel come appartenente o meno all'area da segmentare. Vero e falso significa che corrisponde o no alla segmentazione fatta dall'operatore umano, considerata la verità. Il recupero è indice della capacità della rete di riconoscere le zone d'interesse mentre la precisione ci dice con quanta sicurezza vengono riconosciute. Il DSC è indice dell'efficacia della rete. Più questi indici sono vicini al valore 1, tanto migliore è la rete. Dalla tabella 3.4 vediamo che la rete DR-LRC-UNet rispetto alla classica 3D-UNet ha un miglioramento del 2,1% in precisione, 1,9% in recupero e 2,1% in DSC. La rete DR-LRC-UNet dimostra un'importante evoluzione delle capacità anche rispetto agli altri tipi di rete eccetto per la precisione che è maggiore nella rete R-UNet che è una rete U-Net con un modulo "residuale" aggiunto. Questa rete ha però un peggioramento nell'indice recupero.

I livelli convoluzionali delle reti prese in esame hanno tipicamente un nucleo 3x3, ovvero gli elementi che vengono analizzati in ogni passo della convoluzione sono disposti in una griglia di 9 elementi. L'articolo "The assessment of small bowel motility with attentive deformable neural network" [19] ha la particolarità di usare un nucleo 3x3 che però non è rigido ma può variare forma, come mostrato in figura 3.6. Questa tecnica aumenta le possibili caratteristiche individuabili dal livello convoluzionale a scapito delle prestazioni computazionali della rete, perché devono essere registrate in Runtime le posizioni degli elementi del nucleo che sono variabili e non più rigide. Il modello presentato nell'articolo ha come scopo il calcolo del calibro dell'intestino tenue in una sequenza di cine-MRI. In tabella 3.5 sono presenti i risultati dei test condotti paragonati ad un altro tipo di rete e all'analisi manuale eseguita da un medico esperto. Il paragone è fatto in base alla media e alla relativa deviazione standard e successivamente vengono comparati gli errori quadratici medi (mean square error MSE) rispetto all'analisi manuale. Si vede che il nucleo variabile riduce l'MSE di circa 1,0 che rappresenta un miglioramento rispetto al metodo con nucleo rigido.

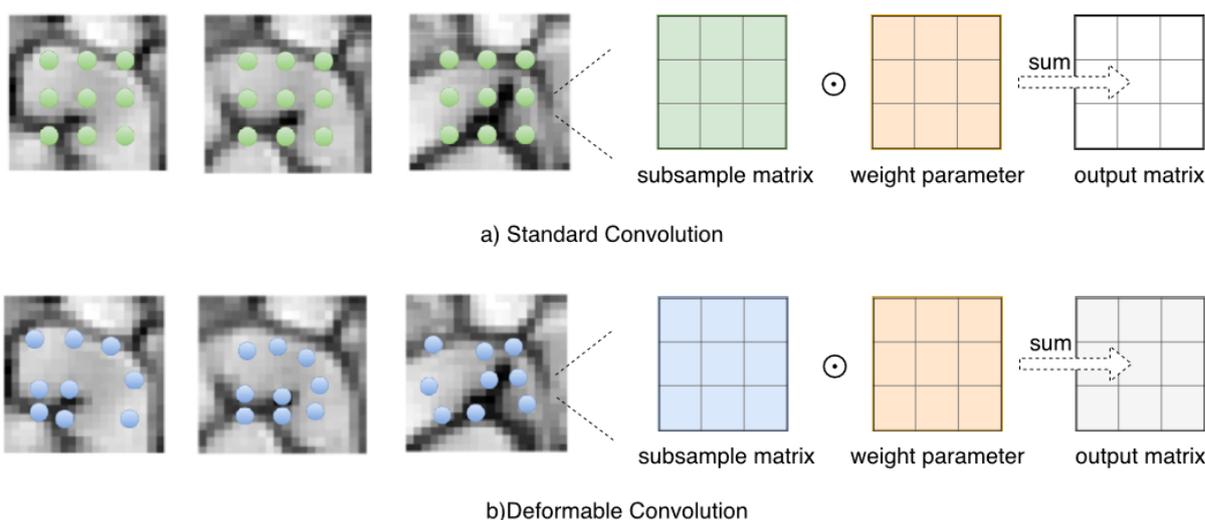


Figura 3.6: sopra la convoluzione standard con nucleo rigido e sotto la convoluzione con nucleo variabile. L'immagine è tratta dall'articolo di X. Wu et al. [19].

	Method	mean \pm std	MSE
Previous	Manual	10.4 \pm 4.53 mm	
	FCN-LSTM	11.23 \pm 2.03 mm	4.472
Proposed	CNN + Encoder-Decoder	8.87 \pm 2.63 mm	4.18
	CNN + Attentive	8.86 \pm 2.25 mm	4.20
	Deformable+Encoder-Decoder	8.86 \pm 2.84 mm	3.34
	Deformable + Attentive	9.13 \pm 2.62 mm	3.27

Tabella 3.5: vengono comparati i metodi usati nei test paragonati ad un modello di un articolo precedente FCN-LSTM e all'analisi manuale svolta da un medico esperto. La comparazione è fatta in termini di media e deviazione standard e errore quadratico medio (mean square error MSE). L'immagine è tratta dall'articolo di X. Wu et al. [19].

3.2 Tecnica di analisi

Gli approcci alla segmentazione possono essere diversi. Si veda la differenza tra i metodi degli articoli “Automatic coronary artery segmentation of CCTA images using UNet with a local contextual transformer” [20] e “Coronary artery centerline extraction in cardiac CT angiography using a CNN-based orientation classifier” [21]. Entrambi hanno come scopo la segmentazione delle arterie coronarie ma mentre nel primo si segmenta direttamente l'immagine, nel secondo c'è un passaggio intermedio che è il calcolo della linea centrale dei vasi sulla base della quale viene effettuata la segmentazione delle aree circostanti. Nello specifico nel primo articolo viene dato come ingresso alla CNN una matrice tridimensionale della scala di grigi ottenuta tramite CT e la rete da come uscita una maschera tridimensionale che rappresenta le aree contenenti le arterie coronarie. In figura 3.7 si vedono le mappe delle

arterie coronarie create dalla segmentazione manuale, da due reti convoluzionali di riferimento, tra cui 3D-UNet, e dalla rete sviluppata dagli autori (DR-LCT-UNet). Il secondo metodo ha sempre come ingresso una matrice tridimensionale ottenuta da CT ma per l'analisi parte da un punto all'interno di una arteria coronaria determinato tramite una rete neurale e, a partire da quello, tramite un'altra rete, si muove passo dopo passo all'interno dell'arteria coronaria tracciando la linea centrale del vaso e ogni volta calcola il diametro stimato dello stesso. Da questi dati viene ricostruita una mappa volumetrica delle coronarie come mostrato in figura 3.8.

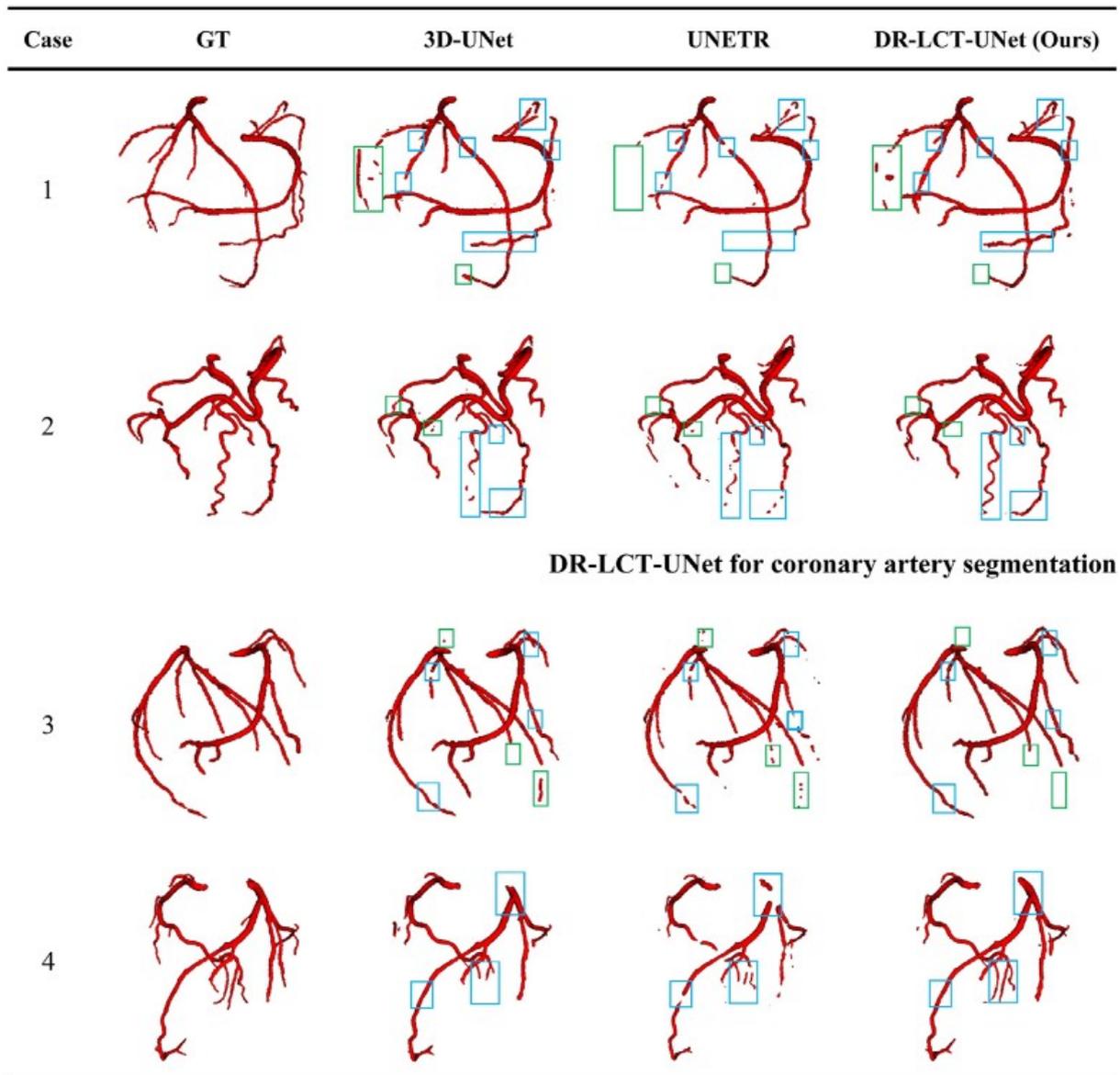


Figura 3.7: immagini volumetriche delle arterie coronarie. Da sinistra le colonne rappresentano le segmentazioni: manuale, da rete 3D-UNet, da rete UNETR e da rete DR-LCT-UNet. L'immagine è tratta dall'articolo di Q. Wang et al. [20].

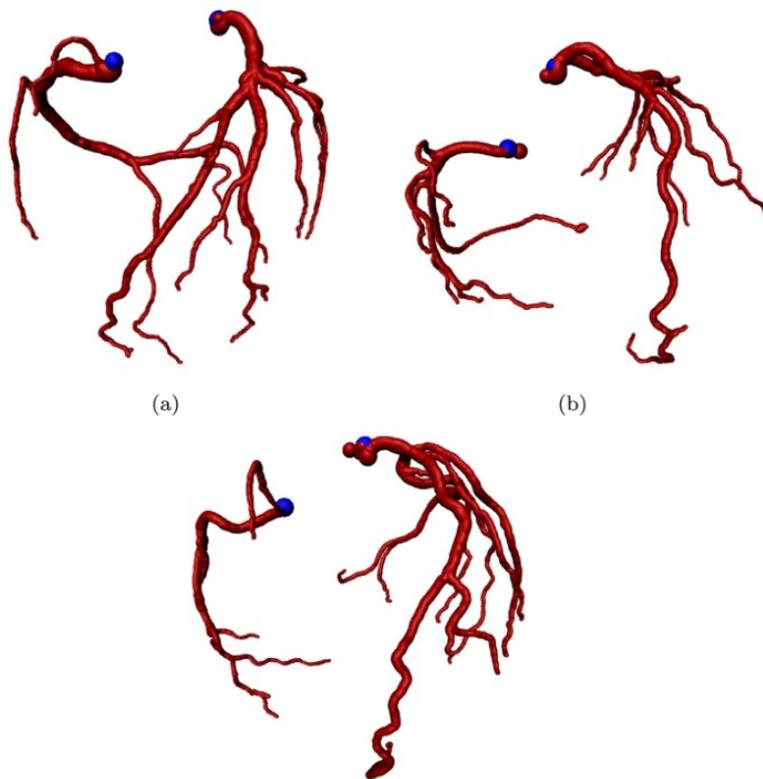


Figura 3.8: l'immagine rappresenta tre ricostruzioni tridimensionali delle arterie coronarie ottenute tramite metodo presentato nell'articolo di J. M. Wolterink et al. [21].

La tecnica del calcolo della linea centrale presentata da J. M. Wolterink et al. [21] ha ispirato il lavoro di L. D. van Harten et al. [18] che hanno usato un metodo simile per la segmentazione dell'intestino tenue. Nel loro lavoro hanno creato due reti convoluzionali a cascata. La prima calcola la linea centrale dell'intestino come per il metodo di Wolterink e la seconda analizza un intorno di tale linea e segmenta la regione che contiene l'intestino tenue. In generale il metodo della linea centrale è applicabile a qualsiasi parte anatomica con struttura tubulare. Un esempio di risultato della prima parte è visibile in figura 3.9 dove la linea arancione è tracciata dalla rete convoluzionale mentre quella bianca dal medico.

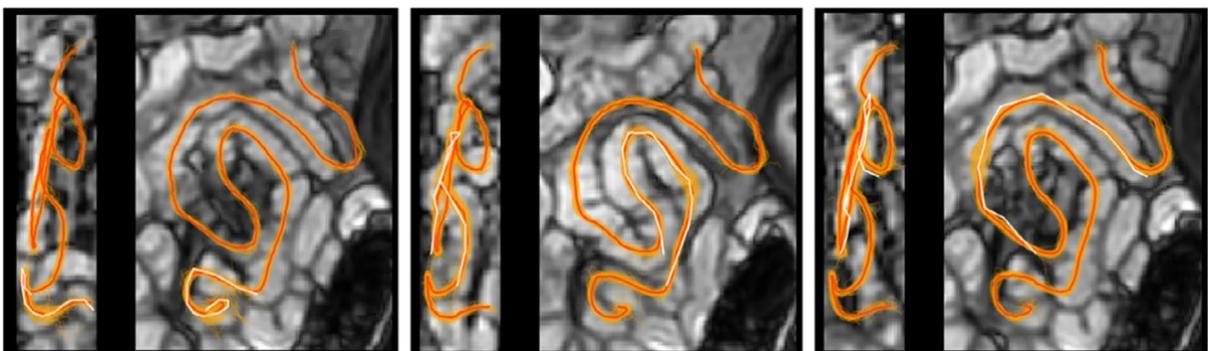


Figura 3.9: esempi di calcolo della linea centrale dell'intestino tenue tramite rete neurale (arancione) e manualmente da un medico esperto (bianco). L'immagine è tratta dall'articolo di L. D. van Harten et al. [18].

La segmentazione ha come scopo di creare un'informazione utile alla diagnosi del medico o di algoritmi di diagnosi automatica. Per questo motivo alcuni approcci di analisi delle immagini mediche integrano nelle reti neurali il processo di diagnosi evitando l'intermedio della segmentazione. Un esempio è presentato nell'articolo "The assessment of small bowel motility with attentive deformable neural network" [19], nel quale gli autori sviluppano un metodo per calcolare il diametro dell'intestino al variare del tempo da una sequenza di cine-MRI. Da quest'informazione è possibile diagnosticare disturbi come la malattia di Crohn. In altri casi invece viene usata la segmentazione come intermedio, lo vediamo nell'articolo "Segmentation of Corneal Nerves Using a U-Net-Based Convolutional Neural Network" [24]. In esso gli autori prima segmentano i nervi della cornea e dall'immagine ottenuta con un altro algoritmo automatico determinano se il paziente soffre di neuropatia diabetica. I risultati sono accurati al 100% se vengono analizzate almeno cinque immagini per paziente.

3.3 Segmentazione multiorgano

Fino ad ora sono stati mostrati metodi che permettono la segmentazione o analisi di singole aree anatomiche. Esistono anche algoritmi che segmentano più organi contemporaneamente. Un esempio è presentato nell'articolo "Automatic multi-organ segmentation in computed tomography images using hierarchical convolutional neural network" [27]. In esso gli autori analizzano due diverse aree anatomiche: la zona pelvica maschile e la regione della testa e del collo. Le immagini sono ottenute tramite CT. La tecnica prevede due fasi: la prima di segmentazione grossolana e la seconda di rifinitura come mostrato in figura 3.10. La prima fase identifica le regioni di interesse dei vari organi in modo approssimato permettendo la frammentazione dell'immagine originale in più parti che vengono poi sottoposte alla segmentazione finale.

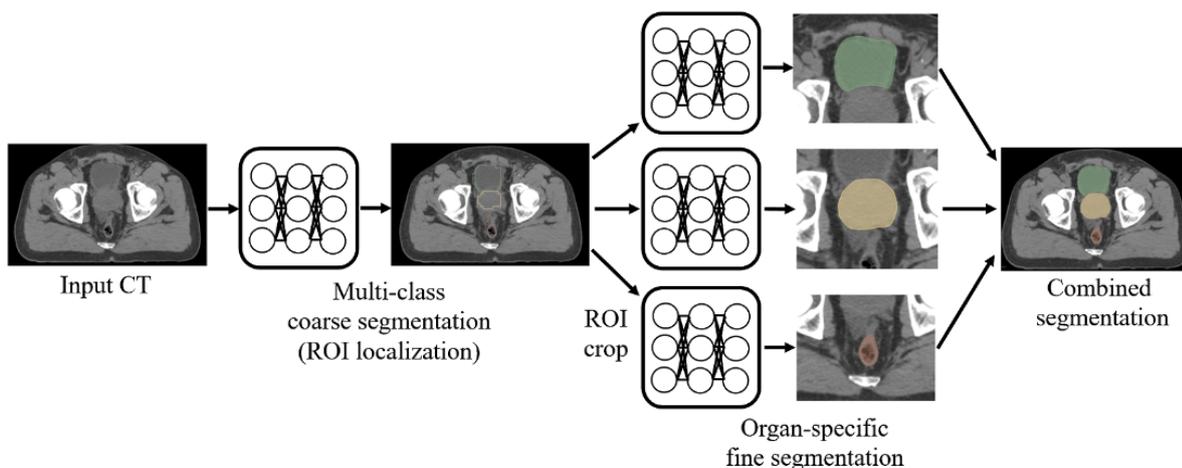


Figura 3.10: schema della segmentazione multiorgano. L'immagine è tratta dall'articolo di S. Sultana et al. [27].

La prima fase usa la rete neurale convoluzionale 3D-UNet con gli encoder formati da due livelli convoluzionali con nucleo 3x3x3 e un livello di pooling 2x2x2, i decoder sono uguali eccetto per il livello di pooling che non diminuisce la dimensione della matrice ma la aumenta. La seconda fase ha una struttura più complessa chiamata GAN (figura 3.11), è formata da un generatore e un discriminatore che competono per generare una segmentazione accurata. Il generatore è una rete U-Net simile a quella della prima fase.

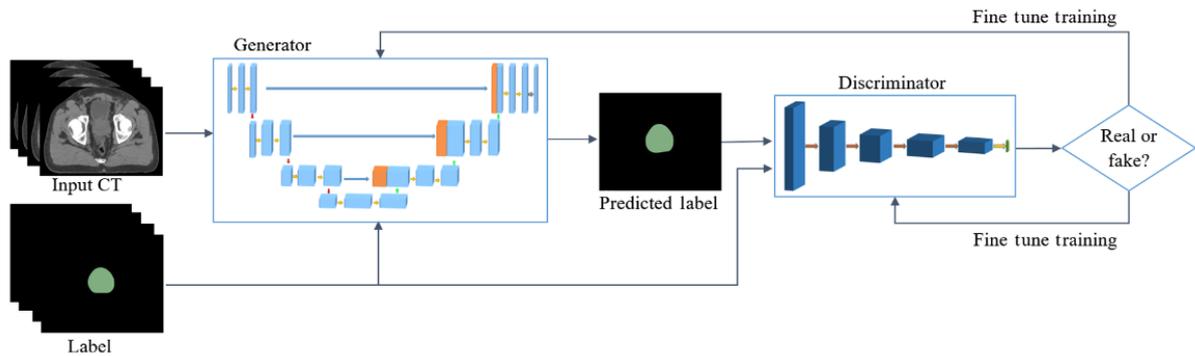


Figura 3.11: schema della rete GAN. L'immagine è tratta dall'articolo di S. Sultana et al. [27].

I risultati dei test condotti sul modello sono rappresentati nelle tabelle 3.6 e 3.7. Il Dice similarity coefficient è superiore al 90% per tutti e tre gli organi della zona pelvica e di poco inferiore per la zona della testa e del collo. Da questi risultati è evidente che le prestazioni della segmentazione multiorgano sono paragonabili a quella di organi specifici.

Metrics	Method	Prostate	Bladder	Rectum
DSC	Hierarchical UNet-GAN	0.91 ± 0.05	0.95 ± 0.06	0.90 ± 0.09

Tabella 3.6: risultati degli esperimenti condotti sulla rete neurale UNet-GAN per la zona anatomica pelvica. La tabella è tratta dall'articolo di S. Sultana et al. [27].

Metrics	Method	Left PG	Right PG	Left SMG	Right SMG
DSC	Hierarchical UNet-GAN	0.87 ± 0.04	0.86 ± 0.03	0.87 ± 0.04	0.85 ± 0.05

Tabella 3.7: risultati degli esperimenti condotti sulla rete neurale UNet-GAN per la zona anatomica della testa e del collo. La tabella è tratta dall'articolo di S. Sultana et al. [27].

Capitolo 4

DISCUSSIONE

Le reti neurali convoluzionali permettono un'analisi rapida e dettagliata delle bioimmagini e questo sta portando miglioramenti nelle attività mediche di diagnosi, chirurgia e creazione di protesi su misura a partire da strutture segmentate. Negli ultimi anni, molti studi stanno sviluppando metodi di segmentazione ottenendo una precisione sempre più simile a quella di un medico esperto. Tuttavia, i risultati non sono ancora giunti ad uno sviluppo utile alla pratica clinica come strumento a supporto del medico.

Per rendere le CNN funzionanti è necessario addestrarle con un ampio dataset di immagini pre-segmentate ed etichettate, che servono alla rete come verità dalla quale migliorarsi. Si è visto che non sono disponibili dataset così numerosi, e le ragioni sono diverse. Anzitutto c'è scarsa disponibilità di immagini, perché le tecniche di acquisizione di maggior interesse, MRI e CT, sono costose e le scansioni richiedono un lungo tempo di esecuzione. Per questi motivi non è possibile eseguire un numero elevato di scansioni altrimenti la ricerca avrebbe un costo economico elevato. Inoltre, CT e Raggi X sono dannosi per l'organismo quindi si tende a non eseguirle senza motivo su soggetti sani. Un altro problema è che le immagini devono essere pre-segmentate da medici esperti. Quest'attività richiede un tempo oneroso, quindi non può essere eseguita un numero elevato di volte. La soluzione alla scarsità di dataset è la creazione di un database pubblico di immagini mediche, dal quale i ricercatori possano liberamente attingere per addestrare e validare i propri modelli. Purtroppo, esistono solo alcuni piccoli database pubblici con immagini pre-segmentate e quindi gli studi vengono effettuati con dataset di qualche decina di scansioni.

La scarsa disponibilità di immagini mediche rende difficile l'addestramento e la validazione delle reti neurali. Il motivo è la differenza anatomica che gli organi molli hanno nei pazienti. Ad esempio, il modo con cui l'intestino tenue è avvolto varia tra una persona e l'altra. Le reti devono essere in grado di riconoscere le ROIs indipendentemente dal paziente, ma avendo a disposizione poche scansioni per addestrarsi sviluppano una limitata adattabilità alle diverse anatomie.

Un'eccezione al problema è la tecnica di acquisizione con scanner ottici, perché ha un costo economico contenuto e un breve tempo di esecuzione. La tecnica è stata usata negli studi [22] e [24], nei quali sono state acquisite immagini da 60 ([22]) e 559 ([24]) pazienti. Il dataset in [24] è molto più ampio rispetto a quelli ottenuti da MRI e CT, per questo motivo la rete alla

fine dell'addestramento ha ottenuto una sensibilità di 0.972 ([24]). Nonostante l'ottimo risultato questo metodo ha un'applicabilità limitata al solo sistema oculare.

Nell'analisi di organi molli, molti degli studi sulle reti neurali precedenti a quelli presentati si limitavano a creare le segmentazioni delle immagini bidimensionali ottenute tramite CT e MRI, non creavano una struttura tridimensionale degli organi. Questo approccio aveva problematiche soprattutto per gli algoritmi di diagnosi che sfruttavano le immagini segmentate. Ad esempio, calcolare il diametro dell'intestino tenue da singole immagini causa un'errata stima del valore reale, perché l'organo ha struttura tubolare e quindi per avere il reale diametro il piano dell'immagine analizzata dovrebbe essere perpendicolare all'asse del condotto intestinale. Inoltre, se si vuole studiare come varia nel tempo il diametro di un tratto dell'intestino bisogna considerare che, essendo un organo molle, l'intestino è soggetto ad ogni minima perturbazione meccanica (battito cardiaco, respirazione, contrazione dell'intestino stesso, ...). Questo comporta uno spostamento nello spazio e quindi in una stessa immagine al variare del tempo non è presente lo stesso tratto di intestino. Di conseguenza le ricerche degli ultimi anni si stanno concentrando sull'analisi tridimensionale, usano infatti CNN tipo 3D-UNet che è un'evoluzione di U-Net progettata per dati tridimensionali.

L'approccio alla segmentazione non è unico. In questa tesi sono stati presentati due metodi. Il primo, più basilare, prevede di dare in ingresso ad una singola CNN l'immagine e ottenere in uscita una maschera dell'organo segmentato ([20], [22], [23] e [24]). L'altro metodo parte dall'identificazione grezza della ROI a cui segue la segmentazione precisa dell'area circostante ([18], [21] e [27]). Il secondo metodo è potenzialmente migliore perché alla CNN per la segmentazione vengono forniti dei dati con poca variabilità delle caratteristiche. Ad esempio, una rete che deve segmentare l'intestino tenue, non dovrà imparare ad escludere dalla ROI il tessuto cardiaco o polmonare, perché non lo riceverà in ingresso. Tuttavia, allo stato dell'arte, gli algoritmi che individuano le porzioni di immagini da segmentare non sono abbastanza precisi. In [18] e [21] viene prima calcolata la linea centrale, rispettivamente dell'intestino e delle arterie, che viene usata come linea guida per la segmentazione. Gli algoritmi che svolgono quest'operazione a volte tracciano la linea al di fuori dell'organo. Il problema viene risolto aumentando la sensibilità della parte dell'algoritmo che decide quando interrompe la creazione della linea. Tuttavia, questa modifica ha come conseguenza che parte dell'organo non viene considerato parte della ROI e quindi nella maschera della segmentazione sono presenti dei vuoti.

Studio	Tecnica di acquisizione	ROI	Sensibilità	Precisione	DICE
[18]	MRI	Intestino	0,77	0,85	0,79
[20]	CT	Arterie coronarie	0,863	0,858	0,858
[22]	Scanner ottico	Vasi sanguigni della retina	0,83	0,979	
[23]	Ultrasuoni	Carotide e giugulare	0,85	0,92	0,88
[24]	Scanner ottico	Nervi della cornea	0,972		
[27]	CT	Zona pelvica			0,91 (prostata) 0,95 (vescica) 0,90 (retto)
[27]	CT	Testa e collo			0,87 (ghiandole parotidi) 0,86 (ghiandole submandibolari)

Tabella 5.1: confronto tra i metodi in [18], [20], [22], [23], [24] e [27] di tecnica di acquisizione, ROI, Sensibilità, Precisione e DICE. Non per tutti i metodi sono presenti i tre indici statistici.

In tabella 5.1 sono riportati, se presenti, sensibilità, precisione e DICE delle CNN analizzate in [18], [20], [22], [23], [24] e [27]. La sensibilità media delle CNN è 0,857 e la precisione media è 0,902. L'eccezione in positivo è la CNN in [24] che ha una sensibilità di 0,972, la ragione come visto è almeno in parte riconducibile ad un dataset più ampio. Lo studio [27] esegue una segmentazione multiorgano della zona pelvica e di testa e collo. Per ogni DICE tra parentesi è riportata la zona anatomica a cui si riferisce.

Capitolo 5

CONCLUSIONI

La segmentazione e analisi di immagini mediche è un lavoro che attualmente è svolto dai medici e richiede loro un grande dispendio di tempo. La soluzione a questo problema è la creazione di algoritmi computerizzati per l'automatizzazione del processo, ed in particolare sono promettenti le reti neurali convoluzionali (CNN), un particolare tipo di algoritmi di intelligenza artificiale con proprietà utili all'analisi di immagini. Allo stato dell'arte molte ricerche sono state effettuate, su varie regioni anatomiche (intestino, vasi sanguigni, nervi, tessuto muscolare), usando CNN per la segmentazione di immagini da MRI, CT, ultrasuoni e scanner ottici, e hanno ottenuto risultati migliori anno dopo anno. Tuttavia, le CNN sono nate recentemente, alla fine degli anni '90, e l'interesse per l'applicazione all'ambito medico è degli ultimi anni. Per questo motivo le reti non sono ancora state validate per essere utilizzate in ambito clinico. Per raggiungere quest'obbiettivo è essenziale avere a disposizione più dataset pubblici da poter usare per addestrare e validare le CNN, oltre a migliori hardware per velocizzare la fase di apprendimento che allo stato attuale richiede diverse ore. Nonostante la loro giovane età ed i vari problemi, è verosimile che le CNN saranno implementate nell'arco di qualche anno come strumento a supporto dei medici.

BIBLIOGRAFIA

- [1] J. H. Scatliff e P. J. Morris, “From Röntgen to Magnetic Resonance Imaging: The History of Medical Imaging”, 2014.
- [2] G. Coppini, S. Diciotti e G. Valli, *Bioimmagini, Collana di Ingegneria Biomedica*, III Edizione, vol. n°8, Bologna, Pàtron Editore, 2012.
- [3] J. Hsieh e T. Flohr, “Computed tomography recent history and future perspectives”, 2021.
- [4] <https://www.siemens-healthineers.com>.
- [5] C. S.de Jonge, A. J. P. M. Smout, A. J. Nederveen, J. Stoker, “Evaluation of gastrointestinal motility with MRI: Advances, challenges and opportunities”, 2017.
- [6] F. C. F. Dionisio, L. S. Oliveira, M. de A. Hernandez, E. E. Engel, P. M. de Azevedo-Marques e M. H. Nogueira-Barbosa, “Manual versus semiautomatic segmentation of soft-tissue sarcomas on magnetic resonance imaging: evaluation of similarity and comparison of segmentation times”, 2021.
- [7] A. Aimar, A. Palermo e B. Innocenti, “The Role of 3D Printing in Medical Applications: A State of the Art”, 2019.
- [8] S. M. Anwar, M. Majid, A. Qayyum, M. Awais, M. Ainowami e M. K. Khan, “Medical Image Analysis using Convolutional Neural Networks: A Review”, 2018.
- [9] M. W. Gardner e S. R. Dorling, “Artificial neural network (the multilayer perceptron) – A review of applications in the atmospheric sciences”, 1998.
- [10] W. S. McCulloch e W. Pitts, “A logical calculus of the ideas immanent in nervous activity”, 1943.
- [11] F. Rosenblatt, “The perceptron: A probabilistic model for information storage and organization in the brain”, 1958.
- [12] K. Fukushima, “Neocognitron: A Self-organizing Neural Network Model for a Mechanism of Pattern Recognition Unaffected by Shift in Position”, 1980.
- [13] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard e L. D. Jackel, “Backpropagation Applied to Handwritten Zip Code Recognition”, 1989.

- [14] Z. Li, F. Liu, W. Yang, S. Peng e J. Zhou, “A Survey of Convolutional Neural Networks: Analysis, Application, and Prospects”, 2022.
- [15] <https://www.ibm.com/it-it/topics/neural-networks>.
- [16] J. Gu, Z. Wang, J. Kuen, L. Ma, A. Shahroudy, B. Shuai, T. Liu, X. Wang, G. Wang, J. Cai e T. Chen, “Recent advances in convolutional neural networks”, 2017.
- [17] F. Schick, “Tissue segmentation: a crucial tool for quantitative MRI and visualization of anatomical structures”, 2016.
- [18] L D. van Harten, C. S. de Jonge, K. J. Beek, J. Stoker e I. Išgum, “Untangling and segmenting the small intestine in 3D cine-MRI using deep learning”, 2022.
- [19] X. Wu, M. Zhong, Y. Guo e H. Fujita, “The assessment of small bowel motility with attentive deformable neural network”, 2019.
- [20] Q. Wang, L. Xu, L. Wang, X. Yang, Y. Sun, B. Yang e S. E. Greenwald, “Automatic coronary artery segmentation of CCTA images using UNet with a local contextual transformer”, 2023.
- [21] J. M. Wolterink, R. W. van Hamersvelt, M. A. Viergever, T. Leiner e I. Išgum, “Coronary artery centerline extraction in cardiac CT angiography using a CNN-based orientation classifier”, 2018.
- [22] S. Xu, Z. Chen, W. Cao, F. Zhang e B. Tao, “Retinal Vessel Segmentation Algorithm Based on Residual Convolutional Neural Network”, 2021.
- [23] L. A. Groves, B. VanBerlo, N. Veinberg, A. Alboog, T. M. Peters e E. C. S. Chen, “Automatic segmentation of the carotid artery and jugular vein from 2D ultrasound images for 3D vascular reconstruction”, 2020.
- [24] A. Colonna, F. Scarpa e A. Ruggeri, “Segmentation of Corneal Nerves Using a U-Net-Based Convolutional Neural Network”, 2018.
- [25] L. Bonaldi, A. Pretto, C. Pirri, F. Uccheddu, C. G. Fontanella e C. Stecco, “Deep Learning-Based Medical Images Segmentation of Musculoskeletal Anatomical Structures: A Survey of Bottlenecks and Strategies”, 2023.
- [26] O. Ronneberger, P. Fischer e T. Brox, “U-Net: Convolutional Networks for Biomedical Image Segmentation”, 2015.

- [27] S. Sultana, A. Robinson, D. Y. Song e J. Lee, “Automatic multi-organ segmentation in computed tomography images using hierarchical convolutional neural network”, 2020.