

ALMA MATER STUDIORUM · UNIVERSITÀ DI BOLOGNA

SCUOLA DI SCIENZE

Corso di Laurea in Informatica

**RICOSTRUZIONE DI IMMAGINI DI
TOMOSINTESI MAMMARIA DIGITALE
TRAMITE RETI NEURALI
CONVOLUZIONALI**

Relatrice:
Chiar.ma Prof.
Elena Loli Piccolomini

Presentata da:
Carolina Bonafè

Sessione II
Anno Accademico 2023/2024

Indice

Introduzione	I
1 Principi della Tomosintesi mammaria digitale	1
1.1 Cenni su Imaging medico e Tomografia Computerizzata	1
1.1.1 Assorbimento dei raggi X: la Legge di Lambert-Beer	3
1.1.2 Ricostruzione delle immagini	5
1.2 La Tomosintesi mammaria digitale: procedure e vantaggi	9
2 I modelli di ricostruzione delle immagini	11
2.1 Le reti neurali convoluzionali	13
2.2 Encoder e decoder	17
2.2.1 U-net	18
2.2.2 Residual U-net	20
2.3 Il dataset	22
3 Risultati	27
3.1 Gli iperparamteri	28
3.2 Le metriche di valutazione	30
3.3 Evoluzione delle funzioni di loss durante l'addestramento	32
3.4 Valutazione delle immagini ricostruite	34
3.5 Analisi comaparativa dei modelli	37
Conclusioni	47

Elenco delle figure

1.1	Funzionamento generale della TC	3
1.2	Illustrazione della legge di Lambert-Beer	4
1.3	Esempio di sinogramma	6
1.4	Funzionamento della Tomosintesi Mammaria Digitale (DBT)	10
2.1	Esempio di una rete neurale convoluzionale	13
2.2	Illustrazione degli elementi di un neurone appartenente a una rete neurale convoluzionale	14
2.3	Applicazione del kernel nei livelli convoluzionali	15
2.4	Esempio di un'operazione effettuata da uno strato di MaxPooling	16
2.5	Architettura generica della U-Net	19
2.6	Blocco residuale della Residual U-Net	20
2.7	Schema della ricostruzione iniziata dal metodo iterativo e conclusa dalle reti neurali	24
2.8	Esempio di coppie input-target: sezione 22 della paziente 56550	25
3.1	esempio del processo di aggiustamento dei valori di grigio	27
3.2	Grafici delle funzioni di attivazione ReLu e sigmoide	30
3.3	Confronto tra l'evoluzione delle funzioni di loss L1 e MSE durante il training	32
3.4	Slice 17, con indicazioni su alcune microcalcificazioni presenti	34
3.5	Slice 21: confronto tra le ricostruzioni effettuate dai modelli con input e target.	36
3.6	Slice 34: Confronto tra le ricostruzioni effettuate da Residual U-Net e U-Net, combinate con le funzioni di loss L1 e MSE.	40

3.7	Slice 17 : Confronto tra le ricostruzioni effettuate da Residual U-Net e U-Net, combinate con le funzioni di loss L1 e MSE.	43
3.8	Ingrandimento di una microcalcificazione nella Slice 17: Residual Unet .	44
3.9	Ingrandimento di una microcalcificazione nella Slice 17 : U-Net	44
3.10	Ingrandimento di una microcalcificazione nella Slice 21 : Residual U-Net	45
3.11	Ingrandimento di una microcalcificazione nella Slice 21 : U-Net	45
3.12	Slice 12: Confronto tra le ricostruzioni effettuate da Residual U-Net e U-Net, combinate con le funzioni di loss L1 e MSE.	46

Elenco delle tabelle

3.1	Confronto tra Immagini di Input e Target tramite metriche di valutazione	37
3.2	U-Net: Confronto tra MSE e L1 utilizzando metriche di valutazione . . .	38
3.3	Residual U-Net: Confronto tra MSE e L1 utilizzando metriche di valuta- zione	38
3.4	Analisi comparativa delle differenze tra immagini ricostruite con diversi modelli, combinati con diverse funzioni di loss	41

Introduzione

Nel panorama della medicina moderna, l'intelligenza artificiale è emersa come una nuova forza rivoluzionaria, con impatti evidenti sui metodi di diagnosi e trattamento. In questo contesto, la sua applicazione alle tecniche di imaging medico rappresenta una frontiera promettente, offrendo nuove opportunità per migliorare la qualità delle immagini e ottimizzare i risultati clinici. L'introduzione di metodi basati sul deep learning nelle tecniche di imaging si allinea perfettamente con gli obiettivi fondamentali di queste ultime, ovvero ottenere ricostruzioni di elevata qualità, riducendo in modo significativo l'esposizione totale alle radiazioni per il paziente e garantendo nel contempo precisione e affidabilità nei risultati. Per ottenere esiti ottimali, i modelli di deep learning necessitano di un allenamento supervisionato su set di dati specifici per apprendere correttamente sia gli effetti di degradazione da rimuovere sia i dettagli anatomici da preservare, che caratterizzano ciascuna tipologia di imaging medico [1].

In questa tesi verrà analizzata l'applicazione di metodi di deep learning per la ricostruzione di immagini tomografiche. In particolare, sono stati implementati due modelli di reti neurali profonde, ovvero la U-Net e la Residual U-Net, impiegati per migliorare immagini bidimensionali pre-elaborate, ottenute da un processo di Tomosintesi mammaria digitale (DBT). Questa metodologia consente di ottenere informazioni dettagliate sulla struttura anatomica interna del seno, combinando proiezioni bidimensionali per generare una rappresentazione tridimensionale.

Le immagini fornite al modello sono state precedentemente ricostruite tramite un determinato algoritmo iterativo basato su modelli. Sebbene questi algoritmi siano un approccio consolidato nella ricostruzione tomografica, presentano un elevato costo computazionale, richiedendo numerose iterazioni per risultati di alta qualità. Per rispettare i requisiti clinici in termini di tempo, è possibile eseguire solo un numero limitato di

iterazioni, producendo immagini con dettagli significativi ma non ottimali. L’impiego delle reti neurali consente di completare la ricostruzione delle immagini in tempi significativamente più brevi, sostituendo le fasi finali dell’algoritmo iterativo. Questo approccio migliora ulteriormente la qualità delle immagini ricostruite, riducendo i tempi di elaborazione senza compromettere la risoluzione o i dettagli anatomici.

Verranno valutate le differenze di prestazioni dei modelli in analisi, combinati con le funzioni di loss MSE e L1, sia tramite confronto visivo delle immagini ottenute sia attraverso metriche di valutazione. Infatti, sebbene i modelli abbiano migliorato le immagini di input, le ricostruzioni ottenute da diversi modelli presentano caratteristiche differenti, che saranno valutate in funzione della loro rilevanza per la diagnosi e la prognosi.

Nel primo capitolo, in particolare nella sezione 1.1, verranno definiti i concetti di imaging medico e Tomografia Computerizzata, con un focus particolare sulle strategie per ridurre l’esposizione alle radiazioni. Questa sezione sarà seguita dalla presentazione della legge di Lambert-Beer, che descrive l’assorbimento dei raggi X da parte dei tessuti, e da una panoramica dei metodi di ricostruzione delle immagini tomografiche più comuni. Il primo capitolo si concluderà con la sezione 1.2, nella quale verrà illustrata la tecnica di imaging di riferimento di questa tesi, ovvero la Tomosintesi mammaria digitale.

Nel capitolo 2, verrà analizzato più nel dettaglio l’uso dei metodi di deep learning nel contesto delle ricostruzioni di immagini prodotte da processi di tomografia e tomosintesi. In particolare, nella sezione 2.1 verranno discusse le caratteristiche delle reti convoluzionali. A seguire, verranno presentati i due modelli utilizzati negli esperimenti di questa tesi, ovvero U-Net e Residual U-Net, con particolare attenzione alla loro architettura encoder-decoder. Infine, il dataset verrà introdotto nella sezione 2.3, che si concentrerà sulla descrizione dell’algoritmo iterativo basato su modelli utilizzato per generare le immagini contenute in esso, noto come Scaled Gradient Projection.

Infine, il terzo e ultimo capitolo, composto da cinque sezioni, contiene la valutazione dei risultati ottenuti dalle ricostruzioni. Nello specifico, nelle sezioni 3.1 e 3.2 verranno definiti gli iperparametri utilizzati nell’addestramento dei modelli e le metriche di valutazione utilizzate per stabilire le prestazioni dei modelli. Successivamente, sarà discusso l’andamento delle due funzioni di loss utilizzate nei modelli. Infine, saranno valutate e confrontate le prestazioni dei modelli e le immagini ricostruite, evidenziando le caratteristiche principali che queste devono possedere in un contesto medico.

Capitolo 1

Principi della Tomosintesi mammaria digitale

1.1 Cenni su Imaging medico e Tomografia Computerizzata

Esistono diverse opzioni di imaging disponibili, ognuna caratterizzata da specifiche proprietà che la rendono più o meno adatta per determinati tipi di diagnosi o strutture analizzate.

Il campo dell'imaging medico ha avuto origine con la scoperta dei raggi X da parte di Wilhelm Röntgen nel 1895. I raggi X sono una forma di radiazione elettromagnetica caratterizzata da un'energia elevata che consente loro di attraversare la maggior parte degli oggetti. Questa capacità ha favorito il loro utilizzo in campo medico per investigare le parti anatomiche degli esseri viventi, in particolare quelle dell'uomo. Infatti le radiazioni vengono assorbite in quantità variabili dai diversi tessuti, a seconda della loro densità radiologica, ovvero della loro capacità di assorbire o bloccare i raggi X. Un bersaglio, come ad esempio una pellicola nel caso radiografia analogica convenzionale, rileva la radiazione residua e la traduce in un'immagine, che riflette la densità e lo spessore delle strutture interne analizzate [2].

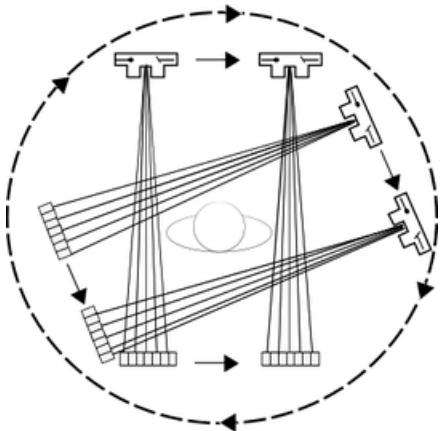
Tuttavia il metodo tradizionale non solo comporta l'uso di radiazioni ionizzanti, che possono essere dannose per i tessuti e risultano altamente invasive, ma è anche spesso fonte di risultati imprecisi. Infatti le radiografie ottenute con questo metodo sono immagini planari, ovvero forniscono una visione bidimensionale del contenuto della struttura. poiché rappresentano corpi tridimensionali in un singolo piano, queste immagini riescono a fornire solo una media delle informazioni spaziali reali, rendendo difficile la distinzione di strutture sovrapposte[2, 3].

Questa limitazione è stata parzialmente superata con l'introduzione dell'uso del computer in campo medico all'inizio degli anni Settanta, che ha portato allo sviluppo della Tomografia Computerizzata (TC), una delle principali modalità di imaging diagnostico non invasivo. Questa tecnica consente di visualizzare l'intera regione corporea tramite la realizzazione di immagini tridimensionali [2]. Durante una procedura di tomografia computerizzata, un fascio stretto di raggi X viene diretto verso il paziente e ruotato rapidamente attorno al corpo, con il fine di acquisire molte proiezioni da angolazioni diverse. I dati contenenti le diverse proiezioni ottenute vengono elaborati dal computer tramite degli algoritmi per generare un'immagine degli strati trasversali o assiali del corpo. Per questo motivo la TC è comunemente chiamata TAC, ovvero Tomografia Assiale Computerizzata. Il processo viene ripetuto per produrre molteplici immagini rappresentanti diverse sezioni delle strutture in analisi, visualizzabili singolarmente o impilate insieme per generare un'immagine 3D. Le rappresentazioni così ottenute riproducono informazioni ed evidenziano più dettagli rispetto alle radiografie tradizionali, il che rende questo metodo un utile strumento di screening e diagnostico .

L'obiettivo di ridurre l'esposizione totale alle radiazioni ionizzanti dei raggi X, motivato dai suoi effetti potenzialmente dannosi nei tessuti viventi, ha stimolato la ricerca e lo sviluppo di varie strategie che abbracciano approcci tecnici e tecniche di esame specifiche applicate nella pratica clinica. La prima è chiamata TC a basso dosaggio e si basa sulla riduzione della quantità di radiazione emessa durante ciascuna scansione. Un'altra soluzione invece consiste nel diminuire il numero di proiezioni a raggi X effettuate. Questa scelta consente di eseguire esami molto più rapidi, ma comporta l'ottenimento di dati tomografici incompleti. La ricostruzione delle immagini ottenute partendo dalle informazioni parziali costituisce l'obiettivo della tomografia sparsa.[1]

Nella prossima sezione verrà esamitata l'integrazione di questi due metodi di riduzione

Figura 1.1: Funzionamento generale della TC



Source: <https://www.rgmedicali.it/2019/02/18/tac-tomografia-computerizzata-cos-%C3%A8-e-a-cosa-s serve/>

A8-e-a-cosa-s serve/

delle radiazioni per lo sviluppo di una nuova tecnica di imaging, specifica per l'analisi del seno, nota come Tomosintesi Mammaria Digitale. Prima di procedere, è importante esaminare i principi generali di assorbimento dei raggi X, definiti tramite la legge di Lambert-Beer, e della ricostruzione generale dell'immagine, entrambi applicabili sia alla Tomografia Computerizzata che alla Tomosintesi Mammaria Digitale.

1.1.1 Assorbimento dei raggi X: la Legge di Lambert-Beer

In quanto radiazioni elettromagnetiche, i raggi X trasportano energia attraverso lo spazio tramite onde e fotoni. La loro capacità di penetrare la materia, con il conseguente assorbimento dei fotoni da parte dei tessuti e delle strutture, permettono di estrarre i dati di proiezione necessari per l'analisi. L'immagine dell'oggetto scansionato rappresenta la mappa dei coefficienti di attenuazione, descritta dal modello fisico dell'assorbimento dei fotoni espresso dalla legge di Lambert-Beer.

Quando un fascio di raggi X attraversa un corpo, un rilevatore registra la radiazione non assorbita e, a partire da essa, misura l'attenuazione dell'intensità, ossia la riduzione del numero di fotoni dovuta all'attraversamento del soggetto, variabile a seconda del materiale di quest'ultimo. Questa attenuazione è descritta da un coefficiente di attenuazione $\mu(w) \geq 0$, che dipende dalle proprietà della materia specifiche del punto attraversato w .

La legge di Lambert-Beer calcola l'attenuazione totale di un fascio di raggi X, assu-

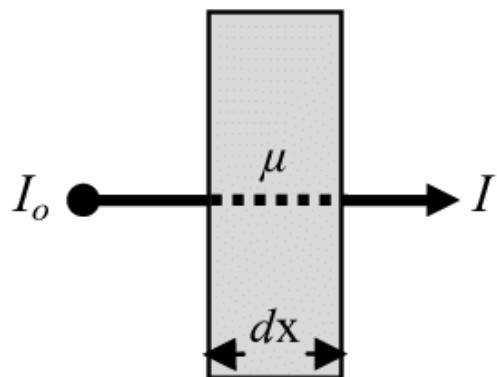
mendo che l'oggetto sia omogeneo, ovvero che il coefficiente di attenuazione sia uniforme in tutto il volume dell'oggetto e che tutte le intensità misurate siano positive. La formula è espressa come:

$$I = I_0 e^{-\mu w} \quad (1.1)$$

dove I_0 è il conteggio di fotoni emessi conosciuto, ossia l'intensità non attenuata dei raggi X alla sorgente, μ è il coefficiente di attenuazione lineare omogeneo e w è lo spessore dell'oggetto.

Per centrare sia l'immagine che la didascalia (inclusa la fonte) all'interno della figura, puoi racchiudere l'immagine e il testo della fonte all'interno di un ambiente center. Ecco come puoi farlo:

Figura 1.2: Illustrazione della legge di Lambert-Beer



Source: Amin, Mohamad and Sudiana, Dodi and Gunawan, Dadag. (2011). The Algebraic Reconstruction Technique of Lambert-Beer's attenuation approximation for parallel rays transmission projection

Realmente il coefficiente non è un valore costante ma varia con l'attenuazione del fascio, poiché gli spettri di energia dei fotoni cambiano man mano che i raggi X attraversano un materiale. Quando il fascio passa attraverso oggetti eterogenei, il coefficiente varia lungo il percorso e dipende dalle proprietà della materia nella posizione raggiunta. Il fattore $\mu(x)$ sarà quindi il coefficiente di attenuazione lineare lungo il percorso del raggio.

La seguente espressione rappresenta l'attenuazione complessiva del fascio in funzione del coefficiente di attenuazione lineare, che varia da punto a punto in un oggetto di

materiale vario. [4]

$$-\ln\left(\frac{I}{I_0}\right) = \int \mu(x) dx \quad (1.2)$$

L'immagine nella Tomografia Computerizzata rappresenta una mappa dei coefficienti di attenuazione, calcolati a partire dal set di dati di proiezione misurati. L'immagine digitale infatti può essere vista come una matrice di pixel, nella quale ognuno di questi corrisponde a un elemento tridimensionale di volume elementare in cui viene scomposto il corpo in esame, noto come voxel.

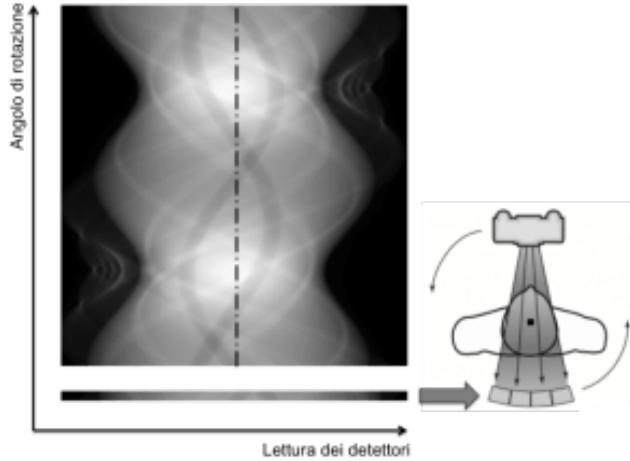
Durante l'elaborazione dell'immagine, è necessario determinare il valore del coefficiente di attenuazione per ogni voxel, al fine di valutare la capacità di assorbimento delle diverse parti del corpo e ricostruire le relative mappe di attenuazione corrispondenti [5]

1.1.2 Ricostruzione delle immagini

Fin'ora abbiamo discusso gli sforzi per ridurre la dose di radiazioni nella Tomografia Computerizzata mediante approcci pratici per ottimizzare le tecniche di acquisizione. Tuttavia questi metodi comportano la degradazione della qualità dell'immagine dovuto all'aumento del rumore e alla suscettibilità di artefatti che possono compromettere i risultati. Di recente, i progressi nella potenza di calcolo hanno consentito lo sviluppo di metodi basati su software per la ricostruzione delle immagini nella TC. [6]

Come descritto precedentemente, nelle scansioni tomografiche i raggi X, emessi da angolazioni diverse, passano attraverso il corpo e vengono assorbiti in modi differenti a seconda dei tessuti traversati. Ogni raggio fornisce informazioni su come il tessuto ha assorbito i raggi X, creando delle proiezioni bidimensionali che vengono poi successivamente ordinate per formare il sinogramma. Quest'ultimo costituisce la base per gli algoritmi di ricostruzione delle immagini, che utilizzano i dati radiografici raccolti per ricostruire una rappresentazione tridimensionale dei tessuti interni.

Figura 1.3: Esempio di sinogramma



Source: <https://www.aitertc.it/web>

Gli algoritmi modellano la ricostruzione come un problema inverso di misurazione lineare, poiché si mira ad ottenere una quantità sconosciuta, in questo caso l'immagine dell'interno del corpo umano, da misure indirette che contengono rumore. Nel caso della Tomografia Computerizzata, i problemi partono dalla definizione di un modello di misurazione diretto:

$$y = M(x) \quad (1.3)$$

dove M è un'operatore che effettua le misurazioni e genera i dati y a partire da un modello applicato ai parametri x .

Da questa definizione, si può costruire un modello inverso a quello di misurazione, ovvero M^{-1} , che ha l'obiettivo di ricostruire l'input a partire dall'output. Si deve dimostrare che :

$$M^{-1}(y) = x \quad (1.4)$$

Nel nostro caso i dati noti sono quelli di proiezione acquisiti durante la scansione, ovvero il sinogramma e si cerca di risalire all' immagine originale che può avere generato tali dati[7].

Nel contesto di tomografia con pochi punti di vista, il problema di ricostruzione è un problema mal posto, ovvero esistono infinite immagini coerenti con gli stessi dati di misurazione. Ciò si può dimostrare riscrivendo la formula 1.3 in termini di un sistema lineare:

$$Ax = b \quad (1.5)$$

dove il vettore $x \in \mathbb{R}^{n^2}$ rappresenta le immagini da ricostruire, b è un vettore m -dimensionale che contiene le misure di proiezione rumorose, ovvero i sinogrammi, e la matrice $A \in \mathbb{R}^{m \times n^2}$ è processo fisico di proiezione discreto. Infatti, si può intuire che con pochi punti di vista $m < n^2$, poiché il numero di misurazioni è inferiore al numero di pixel. Per questo il sistema è sotto-determinato e di conseguenza ammette infinite soluzioni in \mathbb{R}^{n^2}

Esistono vari tipi di algoritmi di ricostruzione, ognuno con i propri punti di forza e di debolezza. Tradizionalmente le immagini tomografiche sono state prodotte utilizzando gli algoritmi analitici, basati sulla retroproiezione nell'immagine finale, in seguito all'elaborazione preliminare dei dati acquisiti. A questo proposito, il metodo della retroproiezione filtrata (FBP) è l'algoritmo analitico più utilizzato e rappresenta lo standard di riferimento per la ricostruzione dei dati delle immagini tomografiche. Questo approccio si basa sulla retroproiezione dei dati, in cui ogni raggio viene tracciato all'indietro lungo la sua traiettoria originale. In questo modo ogni dato di proiezione presente nel sinogramma viene riportato nel dominio dell'immagine lungo la direzione del raggio originale che ha causato l'assorbimento, permettendo così di ricostruire una rappresentazione dettagliata dei tessuti interni. La retroproiezione è caratterizzata da una complessità computazionale relativamente bassa; tuttavia, in presenza di dati mancanti o di incompletezze nelle proiezioni, come nel caso della TC sparsa, dove ogni proiezione contribuisce solo parzialmente alla formazione dell'immagine finale, questo metodo può introdurre artefatti e rumore. Per attenuare tali problematiche, l'algoritmo FBP applica un filtro sui dati di proiezione prima della retroproiezione, capace di regolare caratteristiche fondamentali dell'immagine, come la risoluzione spaziale o il rumore, migliorando così la qualità e la nitidezza della ricostruzione finale. Nonostante la velocità caratterizzante il procedimento, Le principali limitazioni di FBP sono gli alti livelli di rumore e l'elevata suscettibilità agli artefatti L'FBP compromettendo quindi la qualità dell'immagine e la dose di radiazioni a favore di velocità di ricostruzione elevate[8].

Un'alternativa ampiamente utilizzata è rappresentata dalla ricostruzione iterativa basata su modelli. Gli algoritmi iterativi affrontano la ricostruzione delle immagini

corrispondenti ai dati di proiezione acquisiti come un problema di ottimizzazione, in cui i dati dell'immagine sono la soluzione ottimale sconosciuta al problema. Essi sono caratterizzati dalla presenza di una prima proiezione in avanti dallo spazio dell'immagine allo spazio di proiezione, ovvero una stima iniziale grezza per inizializzare il processo iterativo. A partire dall' immagine iniziale stimata, il sistema simula come dovrebbero essere i dati di proiezione se l'immagine fosse corretta; in seguito si confrontano i dati simulati con il sinogramma reale. Se ci sono differenze, la stima corrente dell'immagine viene aggiornata attraverso termini di regolarizzazione, per ridurre l'errore e ottenere una ricostruzione più accurata. Questo procedimento viene eseguito iterativamente, fino a quando la stima non corrisponde il più possibile ai dati acquisiti.

La ricostruzione iterativa basata su modelli si contraddistingue perchè richiede una modellazione del processo di imaging CT per la proiezione in avanti, al fine di consentire la convergenza graduale verso una rappresentazione più accurata. Questo modello incorporato nell'algoritmo che tiene conto delle proprietà tecniche del sistema di acquisizione e di ciascuno dei suoi componenti, come ad esempio la forma del fascio di raggi X, gli elementi del rivelatore e i voxel di ricostruzione. [6, 9]

Gli algoritmi iterativi basati su modelli, nonostante la loro capacità di fornire immagini di alta qualità mantenedno basse le dosi di radiazioni, presentano dei punti deboli legati alla velocità di ricostruzione dell'immagine, che risulta inferiore a quella dell'FBP. Tale velocità è influenzata dall'implementazione dell'algoritmo e tende a diminuire all'aumentare della complessità del modello utilizzato per descrivere il processo di imaging TC. [10]

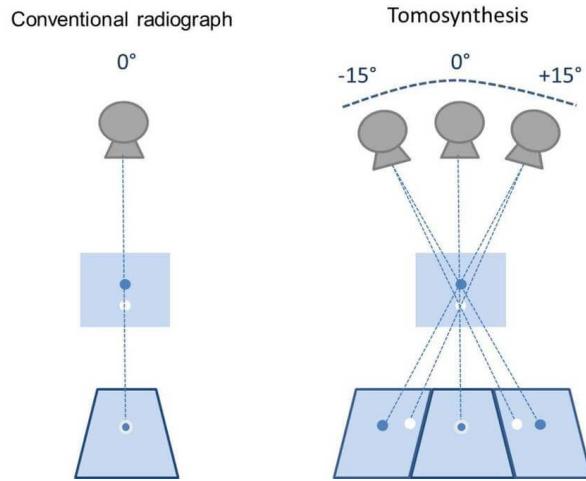
Recentemente, i metodi basati sul Deep Learning sono emersi per la ricostruzione tomografica con pochi punti di vista. In particolare, gli approcci di deep learning hanno raggiunto risultati eccellenti nel risolvere problemi inversi mal posti, come quello della CT a pochi punti di vista[11]. L'uso del Deep learning più popolare nella Tomografica Computerizzata è l'approccio basato su modelli appresi, dove una rete neurale viene utilizzata nei metodi iterativi per aggiornare al meglio l'immagine in ogni iterazione. L'intelligenza artificiale, nel contesto della tomografia, può essere applicata anche alla post-elaborazione appresa (LPP). In questo caso una rete neurale, dopo un'allenamento supervisionato, riesce a sopprimere il rumore e gli artefatti nelle immagini ricostruite con metodi tradizionali come l'FBP.

1.2 La Tomosintesi mammaria digitale: procedure e vantaggi

Un metodo affine alla Tomografia Computerizzata, che sfrutta sia un basso dosaggio di radiazioni sia un numero limitato di proiezioni, è la Tomosintesi Mammaria Digitale (DBT). Questa tecnica di imaging mammografico permette di ricostruire immagini tridimensionali della mammella dall'acquisizione di un numero ristretto di proiezioni bidimensionali ottenute con diverse angolazioni, mentre la sorgente si muove con un arco limitato, che varia da 15 a 60 gradi [12]. La tomosintesi infatti si differenzia dalla tomografia tradizionale in quanto è progettata per i casi nei quali l'anatomia umana non consente una rotazione completa della sorgente di raggi X, rendendo necessario l'uso di un angolo ridotto. Nel caso della DBT infatti, i raggi si focalizzano unicamente sulla parte anteriore del busto delle pazienti. Allo stesso tempo, poiché la sorgente utilizzata nella tomosintesi emette fasci da poche angolazioni lungo la traiettoria, le radiazioni sono limitate a una dose ridotta. Questo approccio consente una migliore identificazione dei tessuti adiposi e fibro-glandolari che compongono le mammelle, poiché questi hanno una bassa attenuazione dei raggi X. Come nella tomografia, le proiezioni acquisite vengono elaborate da un algoritmo di ricostruzione che calcola la loro posizione verticale per stimare la distribuzione dei tessuti nell'immagine tridimensionale.

La tomosintesi comporta numerosi vantaggi rispetto alla mammografia standard, che nonostante l'impiego di dosi relativamente basse di radiazioni ionizzanti, presenta ancora la limitazione di ottenere immagini bidimensionali come risultato delle radiografie, con la conseguente sovrapposizione dei tessuti. Questa situazione complica la valutazione della densità mammaria complessiva e della dose di radiazioni specifica per la paziente. Pertanto, la DBT può essere considerata più efficace della mammografia nel rilevare e visualizzare le masse mammarie, in quanto riesce a fornire anche informazioni sulla profondità dei tessuti[13]. Per questo motivo, essa sta rapidamente sostituendo la mammografia digitale come tecnica radiologica primaria per l'imaging del seno [8, 14].

Figura 1.4: Funzionamento della Tomosintesi Mammaria Digitale (DBT)



Meltzer, Carin and Båth, Magnus and Johnsson, Ase. (2014). A phantom study of tomosynthesis; do in-plane artifacts affect the estimated size of nodules?. [10.1594/ecr2014/C-1093](https://doi.org/10.1594/ecr2014/C-1093).

Ciò che rende possibile l'ottenimento di immagini tridimensionali, rispetto alla mammografia, è la capacità della sorgente di raggi X di ruotare attorno a un punto fisso, come accade nella tomografia. Questo permette di ottenere diverse proiezioni planari a bassa dose che verranno utilizzate per ricostruire il volume 3D della struttura in analisi. Nella tomosintesi, inoltre, i rilevatori sono stati ulteriormente migliorati, consentendo ad esempio una riduzione del tempo complessivo di acquisizione delle proiezioni.

La ricostruzione a strati della mammella nella direzione convenzionale è resa possibile dall'utilizzo di algoritmi di ricostruzione simili a quelli usati nella tomografia, che utilizzano le diverse posizioni nelle proiezioni degli stessi tessuti per calcolare la loro posizione verticale. Questi ultimi consentono di ottenere immagini coerenti con la vera struttura anatomica interna del corpo, compensando le limitazioni e gli effetti delle informazioni incomplete [2].

Capitolo 2

I modelli di ricostruzione delle immagini

Nel contesto della tomografia sparsa e della tomosintesi, gli algoritmi iterativi e analitici sono stati a lungo le tecniche predominanti per la ricostruzione delle immagini, fornendo le informazioni interne di un oggetto per la diagnosi e riducendo allo stesso tempo la dose di radiazioni. Tuttavia, a causa della scarsità di dati di proiezione, le immagini TC a poche viste ottenute con questi algoritmi di ricostruzione soffrono di distorsione strutturale e artefatti. Infatti, sebbene gli algoritmi di ricostruzione iterativa abbiano mostrato prestazioni migliori rispetto all'FBP, è ancora difficile soddisfare i requisiti diagnostici delle ricostruzioni, in termini di degradazione della qualità, tempo di ricostruzione e flessibilità, in particolare in quelle con poche viste [15]. Queste carenze hanno portato allo sviluppo di nuovi metodi alternativi per la ricostruzione, che mirano ad ottenere la combinazione ideale di alta qualità delle immagini, basse dosi e tempi di elaborazione rapidi. Questo aspetto distingue le soluzioni più recenti da quelle tradizionali, che invece tendono a privilegiare uno di questi aspetti a discapito degli altri [8, 16].

Recentemente, i metodi basati sul Deep Learning (DL) sono emersi rispetto agli approcci completamente convenzionali per risolvere problemi inversi dell'imaging medico, in particolare della tomografia e della tomosintesi, rappresentando una nuova generazione di tecniche che migliorano la qualità delle immagini, l'accuratezza e l'efficienza delle ricostruzioni [8]. I sistemi di deep learning coinvolgono reti neurali per trasformare le immagini in input in output utili, apprendendo le caratteristiche nel percorso verso la

produzione degli output.

I metodi di ricostruzione basati su deep learning si possono distinguere in due principali categorie: diretti o indiretti. I primi ricostruiscono un’immagine di alta qualità direttamente dal sinogramma, senza l’utilizzo di tecniche come retroproiezione filtrata (FBP) o ricostruzione iterativa (IR), mirando quindi ad imparare direttamente il processo di inversione.

Al contrario, i metodi indiretti prevedono una prima ricostruzione del sinogramma o di un’immagine ottenuta tramite FBP o IR, prima di migliorare ulteriormente la qualità dell’immagine con la rete [17]. Questi metodi mirano ad ottenere ricostruzioni con rumore ridotto, senza artefatti, applicando un processo di post-elaborazione appresa [18].

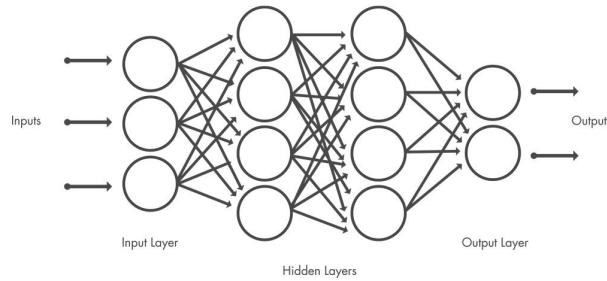
Tutti gli algoritmi DLR indiretti impiegano metodi di addestramento analoghi, fornendo alla rete un insieme di immagini abbinate: una a bassa dose e di scarsa qualità, utilizzata come input, e l’altra di qualità molto elevata, che rappresenta invece la ground truth. Questo utilizzo di dati abbinati rende il processo di addestramento supervisionato. In questi metodi non viene ricostruita l’immagine direttamente, ma è il sistema che, attraverso una serie di aggiustamenti avvenuti durante il training, impara da solo ad eseguire una regressione che perfeziona una ricostruzione iniziale grezza fino ad ottenere quella finale desiderata di alta qualità [11, 17]. Nello specifico, la maggior parte dei metodi di deep learning esistenti si basa su reti neurali convoluzionali, il cui utilizzo consolidato per l’analisi di immagini, insieme al rapido sviluppo [18], ha accelerato la loro applicazione nei metodi indiretti nel campo dell’imaging medico [**DBLP /corr/ZhangLQWYLH16**, 19], in particolare nella tomografia a bassa dose e sparsa [20, 21].

L’obiettivo del progetto descritto in questa tesi è analizzare il comportamento di alcune reti neurali convoluzionali applicate a immagini di tomosintesi precedentemente ricostruite iterativamente. Lo scopo è comprendere come, dopo l’allenamento, tali reti possano migliorare la qualità delle immagini di partenza. In questo capitolo verranno descritte le caratteristiche delle reti convoluzionali utilizzate in questo progetto e del dataset impiegato.

2.1 Le reti neurali convoluzionali

Le reti neurali convoluzionali (CNN) sono tra i modelli più rappresentativi nel campo del deep learning, grazie alla loro capacità di estrarre caratteristiche visive in modo efficiente attraverso un processo di propagazione delle informazioni. Il principale ambito di applicazione delle CNN è l'elaborazione di dati strutturati a griglia, come le immagini, dalle quali riescono ad apprendere in modo automatico e adattivo gerarchie spaziali di feature.

Figura 2.1: Esempio di una rete neurale convoluzionale



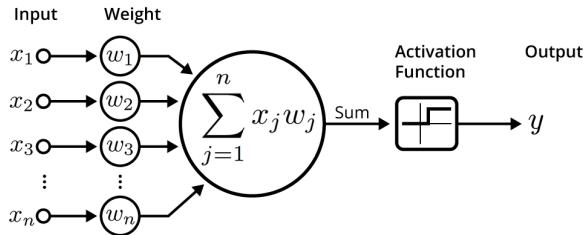
Source: <https://it.mathworks.com/discovery/convolutional-neural-network.html>

Le CNN, così come le altre reti neurali artificiali, si basano su neuroni organizzati in strati, che riescono ad apprendere rappresentazioni gerarchiche. Nelle CNN, i neuroni non sono completamente connessi come in altre reti tradizionali. Un neurone si connette esclusivamente ad alcuni neuroni dello strato precedente, definendo così il proprio campo ricettivo. L'input di un neurone i a livello l , scritto come a_i^l , viene solitamente costruito come una combinazione lineare dei valori di input in arrivo, corrispondenti al output dei neuroni del livello precedente, ovvero z_j^{l-1} :

$$a_i^j = \sum_j w_{ij}^l \cdot z_j^{l-1} + b_i^l \quad (2.1)$$

dove w_{ij}^l è il peso della connessione che va al neurone j appartemente allo strato $l-1$ al neurone i e b_i^l è il parametro bias.

Figura 2.2: Illustrazione degli elementi di un neurone appartenente a una rete neurale convoluzionale



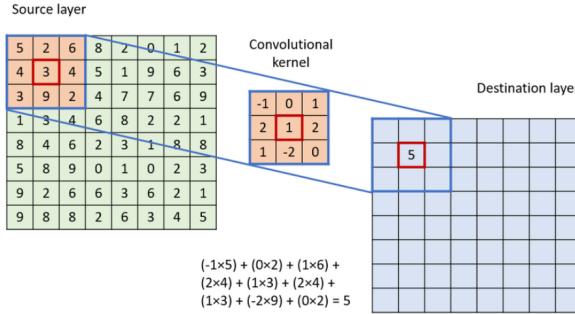
Source: <https://becominghuman.ai/>

Il funzionamento di una CNN può essere considerato come l'imitazione della corteccia animale, dove analogamente numerosi stimuli visivi a diverse scale vengono percepiti nel campo visivo. Così come i neuroni biologici rispondono a specifici segnali visivi nel campo visivo, i neuroni artificiali di una CNN rispondono a caratteristiche rilevanti nell'immagine. Nel caso di queste reti neurali, i diversi recettori che rispondono alle varie caratteristiche sono rappresentati da determinati filtri, noti come kernel. Tra i livelli collocati tra lo strato di input e quello di output, sono presenti strati convoluzionali che rivestono un ruolo fondamentale. Essi infatti applicano i kernel durante le operazioni di convoluzione, permettendo di rilevare pattern specifici all'interno delle immagini [21].

Uno strato di convoluzione è un componente fondamentale dell'architettura CNN per l'estrazione delle feature, che applica un'operazione di convoluzione. Quest'ultima consiste nello scorrimento del kernel sull'input, rappresentato da un tensore. Per ogni posizione di quest'ultimo, un neurone calcola il prodotto elemento per elemento con il kernel, considerato come un array multidimensionale di parametri o pesi. Dopo che il filtro ha analizzato una porzione dell'immagine, si sposta di un numero definito di posizioni, noto come passo o stride, e ripete il procedimento fino a coprire l'intera immagine. I prodotti ottenuti vengono sommati, includendo anche un bias, per ottenere il valore nella posizione corrispondente della feature map, ovvero il tensore di output. Questa operazione viene ripetuta con più kernel, il cui numero determina la profondità delle feature map di output, le quali verranno poi concatenate insieme per formare un output multidimensionale che rappresenta varie caratteristiche dell'immagine. In questa operazione i neuroni possono essere considerati i convolutori, perché i loro pesi, appresi durante

l’addestramento, corrispondono ai parametri del kernel, il quale agisce come connettore tra i neuroni di diversi strati. Infatti tutti i neuroni all’interno di uno stesso strato condividono i pesi e il bias, che sono associati al filtro che viene applicato sulle varie sezioni dell’immagine. Ciò significa che un kernel è in grado di rilevare lo stesso tipo di pattern, anche se appare in posizioni diverse[22]. Per modificare la dimensione dell’output di uno

Figura 2.3: Applicazione del kernel nei livelli convoluzionali



Source: <https://medium.com/>

strato convoluzionale, ad esempio al fine di mantenere la stessa dimensione dell’input, si può utilizzare la tecnica del padding. Essa consiste nell’aggiungere dati extra ai bordi dell’input in modo da manipolare la larghezza, l’altezza o la profondità. Senza i padding, i filtri non riescono a elaborare i bordi dell’input, portando ad una conseguente riduzione della dimensione dell’output ed eliminazione di informazioni nei lati dopo ogni convoluzione. La tecnica più usata è il padding zero, ovvero l’utilizzo di zeri come valori di padding. [23]

L’operazione di convoluzione consente al kernel di individuare pattern ricorrenti, come bordi o texture, che possono apparire in più punti dell’immagine. Ognuno dei kernel è progettato per estrarre caratteristiche diverse dell’input, formando così varie feature map che catturano informazioni differenti. Attraverso l’applicazione di filtri diversi sulle feature map ottenute nei livelli successivi, la rete riesce ad estrarre caratteristiche sempre più astratte e complesse dell’immagine.

L’insieme di più layer convoluzionali crea un vasto numero di feature map e, per ridurre la dimensionalità in piano di queste proprietà, solitamente viene aggiunto uno strato aggiuntivo, noto come pooling layer. Questo aiuta a minimizzare la spesa computazionale necessaria, preservando le informazioni discriminanti e scartando contemporaneamente i

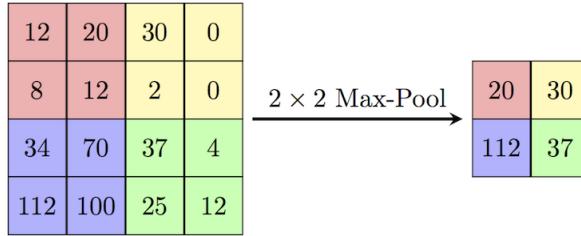
dettagli irrilevanti dell'immagine, riducendone la dimensione. L'architetture della CNN descrivono la combinazione degli strati convoluzionali con quelli di pooling.

Nei modelli presenti in questa tesi, viene utilizzato il metodo maxpooling, nel quale la feature map risultante dallo strato convoluzionale precedente viene divisa in regioni, chiamate finestre di pooling, e viene preso in considerazione solo il valore massimo all'interno di esse. Il pooling massimo può essere rappresentato matematicamente come:

$$a_{ijk} = \max_{(p,q) \in \mathcal{R}_{ij}} X_{kpq} \quad (2.2)$$

L'output dell'operazione di max pooling nella posizione (i, j) della k-esima feature map, ovvero a_{ijk} è il massimo tra tutti i valori nella k-esima feature map che si trovano all'interno della finestra di pooling R_{ij} [24]. Dopo l'operazione di convoluzione, i dati

Figura 2.4: Esempio di un'operazione effettuata da uno strato di MaxPooling



Source: <https://computersciencewiki.org/>

passano attraverso una funzione di attivazione, che aggiunge non linearità al modello per poter apprendere rappresentazioni complesse [25]. Infatti l'output di un neurone deriva dall'applicazione del kernel, quindi la feature map z_i^l , nonché input dei neuroni dello strato successivo è calcolabile tramite funzioni di attivazioni, generalmente scelte come fisse [26]:

$$f(\cdot) : z_i^l = f(a_i^l) \quad (2.3)$$

Le funzioni di attivazione determinano quali informazioni devono essere trasmesse al neurone successivo, simulando la funzione per cui solo i segnali elettrici neurali che superano una certa soglia possono essere trasmessi al neurone successivo. Una delle funzioni di attivazione più comuni è la a Rectified Linear Unit (ReLU) definita come $ReLU(a) = \max(0, a)$ [26].

Durante l’allenamento, l’obiettivo della rete è apprendere i parametri dei kernel θ , ovvero pesi e bias, con il fine di massimizzare la somiglianza di ogni output di training dati i corrispondenti input e parametri CNN per tutti gli n esempi di addestramento [27].

$$R_{\text{learn}} = \arg \max_{\theta \in \Theta} \prod_{n=1}^N P(y_n | x_n, \theta)$$

La complessità e le prestazioni generali di un’architettura CNN dipendono inoltre dagli iperparametri , che definiscono tra gli altri il numero e le caratteristiche di livelli nascosti, operazioni di pooling, tecniche di regolarizzazione o funzioni di costo. Di conseguenza, esiste un’ampia gamma di opzioni per implementare una CNN verso il caso d’uso specifico, nonché architetture predefinite e stabilite.

2.2 Encoder e decoder

Le reti neurali profonde utilizzate in questa tesi sono addestrate per migliorare la qualità delle immagini prodotte attraverso un processo di tomosintesi, ricostruite con un algoritmo iterativo basato su modelli interrotti, in modo da completare tale ricostruzione. Saranno utilizzati modelli addestrati in modo supervisionato e caratterizzati dalla divisione logica dell’architettura in due parti, ovvero encoder e decoder.

L’encoder, la sottorete che va dal livello di input a quello intermedio, segue il principio delle reti convoluzionali tradizionali, in cui l’output si ottiene applicando una funzione di attivazione a una combinazione lineare degli input, come illustrato nella formula 2.1. Durante questa fase avviene l’operazione di downsampling, ovvero la riduzione della risoluzione spaziale dell’input, che consente di apprendere una rappresentazione semplificata dei dati, catturando le caratteristiche principali. Questo blocco contiene gli strati delle reti convoluzionali tradizionali visti precedentemente, ovvero il livello di convoluzione, la funzione di attivazione e il livello di pooling. La rappresentazione compressa ottenuta in output dall’encoder è nota come codifica o rappresentazione latente.

Il decoder invece è la sottorete che va dal livello intermedio al livello di output. Il suo obiettivo è quello di riportare la codifica nello spazio originale, ovvero provare, a partire da essa, a ricostruire l’immagine di input. È infatti importante sottolineare che

nel decoder ogni livello opera in modo top-down, poiché l’obiettivo è ricostruire il segnale di input cercando di mappare la rappresentazione codificata all’immagine originale, mirando a ricostruire una rappresentazione il più fedele possibile [28]. In questo blocco viene eseguito l’upsampling, ovvero l’aumento progressivo della risoluzione spaziale, per ripristinare la rappresentazione originaria. Anche in questa fase si applicano funzioni non lineari per ricostruire l’output finale [29, 30].

Le reti utilizzate in questa tesi sono spesso impiegate per la segmentazione e la riduzione del rumore nelle immagini. Esistono altri modelli caratterizzati da un’architettura encoder-decoder, chiamati Autoencoder, i quali sono progettati per l’allenamento non supervisionato e hanno l’obiettivo di ricostruire il segnale in input.

Nelle sottosezioni seguenti, analizzeremo nel dettaglio le reti utilizzate e analizzate in questa tesi.

2.2.1 U-net

U-Net è un modello CNN sviluppato principalmente per l’attività di segmentazione delle immagini, basato su un’architettura di autoencoder composta da 23 strati convoluzionali . Questo modello estrae le caratteristiche a bassa risoluzione eseguendo ripetutamente operazioni convoluzionali su immagini ad alta risoluzione e riducendo le immagini convolute [31].

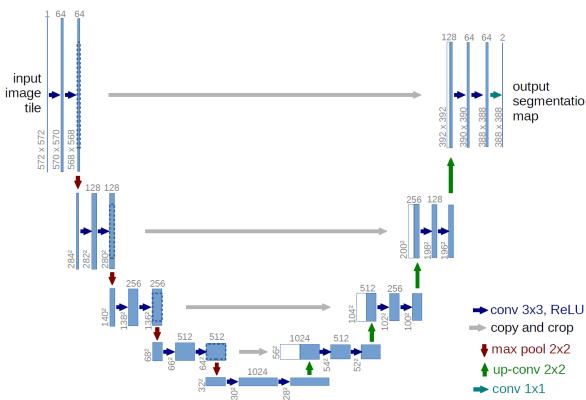
U-Net adotta una struttura simmetrica a forma di U per estrarre completamente le informazioni sulle caratteristiche nei campioni di immagini. Nel modello U-Net la parte di encoder implica un processo di contrazione che segue la tipica formazione di una rete neurale convoluzionale. Nella rete implementata e utilizzata in questa tesi, l’encoder è composto una sequenza di blocchi convoluzionali, che comprendono due strati di convoluzione con filtri di dimensione 3×3 , seguiti da una normalizzazione batch e da una funzione di attivazione ReLU. In particolare l’encoder inizia con un blocco convoluzionale e continua con una sequenza di altri 4 blocchi convoluzionali, intervallati da livelli di max pooling per ridurre la dimensione spaziale, con una dimensione di pooling 2×2 con stride 2. Dopo ogni down-sampling, il numero di filtri viene raddoppiato.

D’altra parte, il decoder esegue l’up-sampling delle feature map tramite blocchi specifici per ripristinare la dimensione originale. Essi sono composti da uno strato di upsample,

che aumenta la dimensione spaziale delle feature map di un fattore 2, raddoppiando ogni riga e ogni colonna. Viene utilizzata un’interpolazione bilineare, che considera i valori dei pixel circostanti per calcolare i valori dei nuovi pixel aggiunti. Questa operazione è fondamentale per produrre feature map di dimensione maggiore rispetto a quelle che riceve in input, al fine di soddisfare la dimensione dell’immagine originale [32]. Infatti l’upsample con interpolazione bilineare consente di spalmare i valori presenti nell’input su una posizione più ampia, espandendo la sua rappresentazione. Nei blocchi di Upsampling, lo strato di upsample è seguito da una convoluzione 3×3 , da una normalizzazione batch, utilizzata per stabilizzare il training, e da una funzione l’attivazione ReLU. Nel decoder, i blocchi di Upsampling sono intervallati da blocchi convoluzionali e similmente all’encoder , questa sequenza viene ripetuta quattro volte, dimezzando il numero di filtri in ogni fase. Nell’ultima fase, viene applicata una convoluzione aggiuntiva 1×1 per ridurre la feature map al numero richiesto di canali, generando così l’immagine segmentata.[33, 34, 35, 36, 37]

Ulteriori elementi presenti nell’architettura U-Net sono le skip connections, che collegano direttamente gli strati dell’encoder agli strati corrispondenti del decoder. Queste connessioni consentono di trasferire le informazioni dettagliate, estratte dall’encoder, direttamente al decoder, migliorando così la qualità della ricostruzione finale.

Figura 2.5: Architettura generica della U-Net



Source: <https://lmb.informatik.uni-freiburg.de/people/ronneber/u-net/>

2.2.2 Residual U-net

la Residual-U-Net (Res-U-Net) è una variante del modello U-Net che utilizza blocchi residui, ognuno dei quali è costituito da strati sovrapposti come la normalizzazione batch (BN), l'attivazione ReLU o lo strato di convoluzione. I blocchi residui riescono ad estrarre le caratteristiche e rafforzare la loro propagazione durante l'addestramento e il test, aiutando a riutilizzare e perfezionare le funzionalità tra i livelli [38, 39].

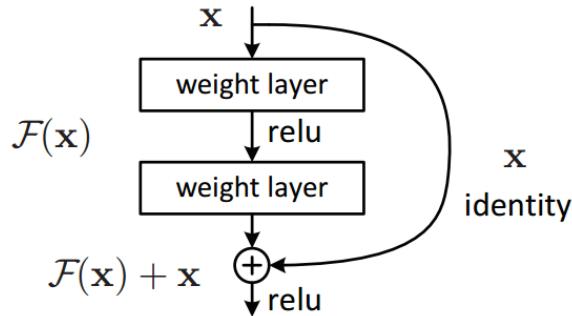
In questi blocchi non viene appresa una mappatura diretta, ma piuttosto una funzione residua. Sia $M(x)$ la distribuzione reale sottostante che un modello tradizionale, come la U-Net, cerca di apprendere, ovvero la mappatura dall'input x all'output y . Ciò è indicato nella seguente equazione:

$$y = M(x) \quad (2.4)$$

Nel caso dei blocchi residui, questi si concentrano sull'apprendere la funzione residua $F(x)$, ovvero la differenza tra l'output del modello applicato a x $M(x)$ e l'input originale x , formalizzata nella seguente formula:

$$F(x) = M(x) - x \quad (2.5)$$

Figura 2.6: Blocco residuale della Residual U-Net



Source: <https://medium.com/>

Apprendendo la funzione residua, il modello vuole capire quanto deve cambiare l'input per produrre l'output corretto.

Possiamo esprimere le operazioni di ogni blocco convoluzionale,in particolare ciò che esso vuole apprendere, tramite questa formula:

$$y_l = M(x) = F(x_l) + h(x_l) \quad (2.6)$$

$$y_{l+1} = f(y_l) \quad (2.7)$$

Dove x_l e y_l sono rispettivamente l'input e l'output dell' l -esimo blocco residuale, F è la funzione residua, $h(x_l) = x$ è la funzione d'identità e f è la funzione di attivazione Relu. Si può notare che la rete utilizza le informazioni originali, propagate tramite un percorso diretto. In questa equazione, l'utilizzo dell'input attraverso la funzione identità rappresenta una skip connection all'interno dello strato convoluzionale, nota come skip connection residua. Essa è utilizzata per migliorare le informazioni all'interno della rete e consentendo l'addestramento di reti più profonde. Le skip connections residuali rappresentano uno degli aspetti più innovativi dell'architettura della Residual U-Net, consentendo di preservare l'intera risoluzione spaziale. Quando le feature map generate da un livello vengono sommate all'input originale, si stabilisce un collegamento diretto tra queste informazioni. Questo approccio permette alla rete di recuperare dettagli e informazioni che potrebbero andare perduti durante il processo di compressione [37, 38, 40].

Grazie all'integrazione delle informazioni originali, il modello è in grado di concentrarsi sulle modifiche necessarie da applicare, distinguendo chiaramente tra le caratteristiche che devono essere modificate e quelle che dovrebbero rimanere invariate. Le skip connections residuali consentono, infatti, di trasferire le feature map e garantiscono che le informazioni critiche non vengano compromesse durante l'elaborazione. Questo meccanismo è ottimo per ridurre il rischio di scomparsa del gradiente e facilita l'apprendimento di rappresentazioni più complesse.

Nella Residual U-Net implementata, è presente un primo blocco residuale, il quale include una sequenza di un livello di convoluzione 3×3 con stride 1, una normalizzazione batch e una funzione di attivazione ReLU, ripetuta per due volte. Infine il risultato ottenuto viene concatenato all'input, rappresentando la skip connection, e il tutto viene passato ad una convoluzione finale per ridurre il numero di canali. Nella sottorete rappresentante l'encoder sono presenti quattro blocchi residui intervallati da operazioni di max pooling. Nella parte di decoder, per ripristinare le dimensioni originali e integrare le informazioni apprese dalla fase di downsampling, è presente una sequenza di un blocco di upsampling, identico a quelli utilizzati nella U-Net, e uno residuale, anch'esso ripetuta quattro volte.

La motivazione alla base dello sviluppo della residual U-Net è stata quella di superare la difficoltà nell’addestramento di reti neurali altamente profonde. È noto che le reti neurali sono in grado di convergere più rapidamente verso una soluzione quando sono presenti più livelli. Tuttavia, i risultati sperimentali hanno dimostrato che l’aumento del numero di livelli può causare un degrado delle prestazioni. L’Introduzione di skip connection residuali può aiutare a mitigare tali problemi [1, 31].

2.3 Il dataset

Il dataset utilizzato per l’allenamento e per il test dei modelli è composto da coppie di immagini bidimensionali ricostruite iterativamente.

Le immagini utilizzate rappresentano sezioni bidimensionali del seno di alcune pazienti, ottenute attraverso l’elaborazione dei dati di proiezioni a poche viste tramite un algoritmo iterativo.

Infatti sappiamo che la tomosintesi, come la tomografia a poche viste, è caratterizzata dall’utilizzo di misure di proiezioni limitate. Di conseguenza il problema di ricostruzione di immagini di questo tipo, come ricavato dalla formula 1.5 , è un problema mal posto e ciò comporta la presenza di artefatti e di rumore nelle immagini ricostruite.

Per affrontare questo tipo di problema vengono utilizzati i metodi iterativi basati su modelli, poiché prevedono l’integrazione di informazioni a priori, ovvero le proprietà del sistema di acquisizione, che possono aiutare a comprendere il comportamento atteso della ricostruzione date le misure di proiezioni. In questo contesto, il problema di ricostruzione può essere espresso come problema di ottimizzazione, nel quale si mira a minimizzare la funzione obiettivo $f(x)$ della forma

$$\min_{x>0} f(x) = J(x) + \lambda \cdot R(x) \quad (2.8)$$

dove

- $J(x)$ è la funzione di adattamento dai dati, che misura quanto le proiezioni calcolate da un’immagine f si discostano dai dati osservati, ovvero i sinogrammi. Essa dipende dal tipo dominante di rumore di uno specifico sistema CT fisico;
- $R(x)$ è la funzione di regolarizzazione, utilizzata per migliorare la qualità dell’immagine. Poiché molte immagini mediche sono quasi uniformi all’interno degli organi,

la funzione di regolarizzazione più ampiamente utilizzata per i problemi TC è la funzione di variazione totale (TV), nota per preservare i bordi all'interno delle immagini;

- λ è il parametro di bilaciamento

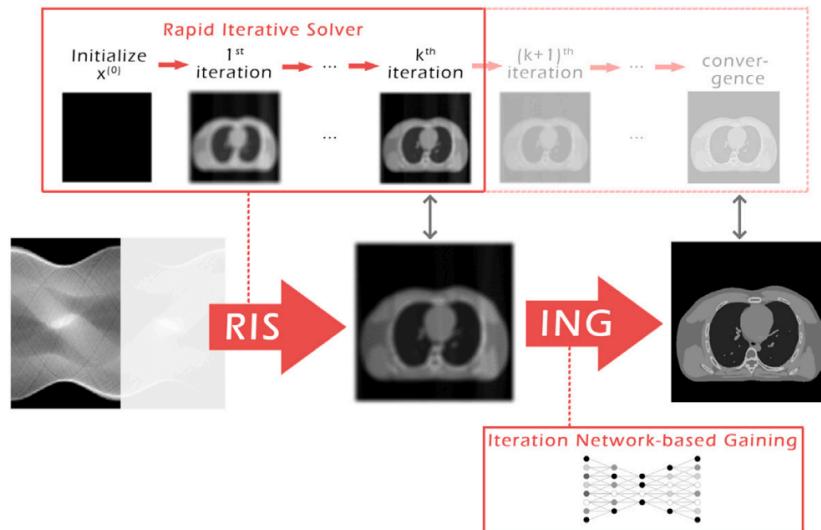
Per risolvere questo problema di ottimizzazione, e quindi per ottenere le immagini utilizzate in questa tesi, è stato utilizzato come risolutore un algoritmo iterativo noto come Scaled Gradient Projection (SGP). Nell'impostazione del problema inverso, risulta particolarmente complesso ottenere le immagini di ground truth, poiché non esiste un oracolo capace di indicare la soluzione esatta. Le immagini ideali, che rappresentano fedelmente la realtà, sono infatti sconosciute e inaccessibili, poiché devono anch'esse essere ricostruite a partire dai dati incompleti e imperfetti ottenuti dai dispositivi di acquisizione. Per questo motivo, l'algoritmo iterativo SGP è stato impiegato non solo per la generazione delle immagini di input, ma anche per quelle di target, ottenute dopo un numero maggiore di iterazioni [27].

L'algoritmo Scaled Gradient Projection è stato proposto come una soluzione alternativa agli approcci diretti, che risultano poco pratici nel contesto di problemi mal posti, poiché possono generare soluzioni instabili [1]. L'SGP è un algoritmo adatto per risolvere problemi di ottimizzazione, poiché segue un approccio di proiezione del gradiente accelerato, simile a quello di discesa. Infatti, come quest'ultimo, esso modifica i pesi lungo la direzione del gradiente per raggiungere il punto di minimo della funzione, ma attraverso un operatore di proiezione i risultati vengono proiettati su un insieme di dati appartenenti a un dominio considerato consentito [41].

Esso sceglie la lunghezza del passo introducendo una particolare matrice diagonale appropriata, nota come matrice di scaling, che migliora il condizionamento del problema.

È stata dimostrata la convergenza dello Scaled Gradient Projection al minimo globale dell'equazione 2.8, tuttavia, per generare le immagini del dataset non è stato necessario eseguire tutte le iterazioni. Come accennato nell'introduzione e nella sezione 1.2, gli algoritmi iterativi basati su modelli richiedono un numero elevato di iterazioni per produrre risultati ottimali. Tuttavia, poiché le immagini ottenute da un numero ridotto di iterazioni contengono comunque molti dettagli significativi della struttura esaminata, le reti neurali possono essere impiegate per completare la ricostruzione iniziata da un algoritmo iterativo interrotto. Questo approccio consente di ridurre il tempo di esecuzione e, di conseguenza, il costo computazionale [42, 43].

Figura 2.7: Schema della ricostruzione iniziata dal metodo iterativo e conclusa dalle reti neurali



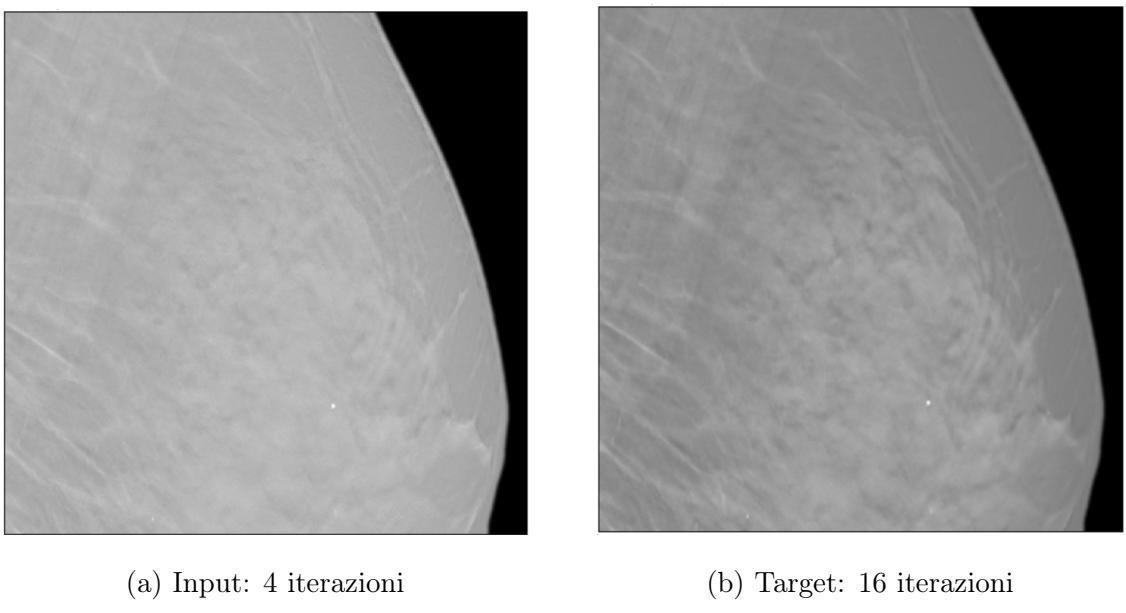
Source: <https://www.sciencedirect.com/science/article/pii/S0895611122001264>

Il dataset, fornito dall'azienda IMS, contiene immagini bidimensionali in scala di grigi, che rappresentano sezioni trasversali della mammella sinistra di pazienti reali. Queste immagini sono state generate dal metodo iterativo SGP partendo da 16 proiezioni. La sorgente utilizzata nel processo tomografico, da cui sono state ottenute queste proiezioni, è caratterizzata da un movimento ad angolo limitato, compreso tra -15 e 15 gradi.

Come già sottolineato in precedenza, esso è composto da coppie di immagini corrispondenti, una di input e una di target. Il dataset è organizzato per pazienti, ognuno

identificato da un codice numerico univoco. Ogni paziente presenta un set di 60 immagini che rappresentano le sezioni trasversali del seno. Per il training sono state impiegate le immagini di cinque pazienti, mentre quelle di una singola paziente sono state riservate per la fase di test. Le immagini di input sono il risultato di 4 iterazioni dell'algoritmo SGP, mentre le immagini di target sono state ottenute dopo 16 iterazioni. Le immagini sono state ridimensionate a 512x512 pixel per essere utilizzate nei modelli, garantendo la conservazione e la presenza delle aree significative.

Figura 2.8: Esempio di coppie input-target: sezione 22 della paziente 56550



L’obiettivo delle reti neurali è completare e affinare il processo di ricostruzione iniziato dall’algoritmo iterativo, migliorando l’immagine di input affinché assomigli il più possibile a quella di target corrispondente. Infatti, nelle immagini di input, proprio perché generate da un processo che non esegue tutte le iterazioni necessarie per raggiungere la soluzione ideale, le strutture anatomiche saranno più rumorose o meno visibili. È compito delle reti neurali apprendere la trasformazione tra le immagini di input e quelle di target, caratterizzate da una qualità migliore poiché ottenute in seguito a un numero maggiore di iterazioni.

Questa combinazione consente di accelerare il processo di ricostruzione delle immagini senza richiedere che l’SGP esegua tutte le iterazioni necessarie per raggiungere la soluzione ottimale, sfruttando la capacità delle reti neurali di affinare e completare le immagini a partire da una soluzione iniziale già promettente.

Capitolo 3

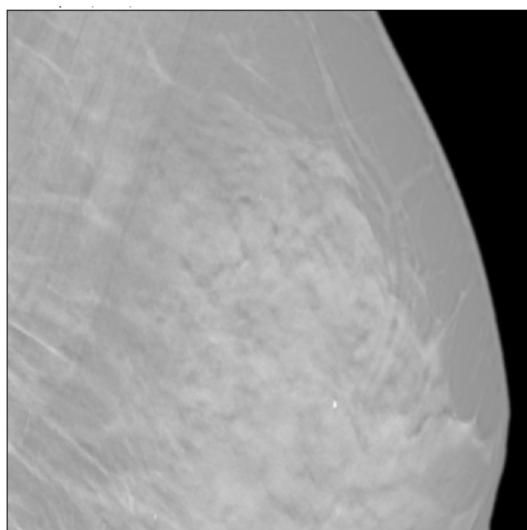
Risultati

In questo capitolo vengono presentate le immagini finali ottenute dall'uso della U-Net e della Residual U-Net per la ricostruzione delle immagini di input. Verranno evidenziate le differenze tra le immagini ricostruite e quelle target, sia attraverso la comparazione visiva diretta delle tre immagini, sia mediante l'analisi di metriche specifiche.

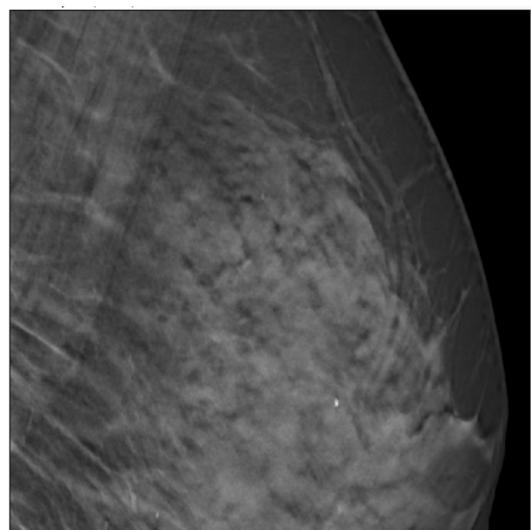
Inoltre verranno analizzate le differenze tra i risultati prodotti dai due modelli, considerando diverse combinazioni di iperparametri testati per ciascuna rete. Questo approccio ha lo scopo di valutare l'impatto delle variazioni parametriche sulle prestazioni della ricostruzione, permettendo di identificare le configurazioni ottimali per ciascun modello.

Figura 3.1: esempio del processo di aggiustamento dei valori di grigio

(a) Immagine originale



(b) Immagine post-elaborata



È importante sottolineare che, al fine di evidenziare le differenze nei risultati ottenuti

dai due modelli, le immagini presentate nelle prossime sezioni sono state sottoposte a un processo di post-elaborazione, in particolare per l’aggiustamento dei livelli di grigio. Ciò consente di enfatizzare le caratteristiche delle immagini, soprattutto in termini di nitidezza e di identificazione degli oggetti di interesse, come si può dedurre dall’esempio nella figura 3.1.

Prima di procedere con l’analisi dei risultati, è fondamentale fornire una descrizione degli altri iperparametri utilizzati, che influenzano direttamente il processo di apprendimento e, di conseguenza, le prestazioni finali dei modelli.

3.1 Gli iperparamteri

Ciò che distingue i diversi modelli e ne determina il successo nelle varie applicazioni sono gli iperparametri, ovvero impostazioni specifiche regolate dall’utente prima dell’allenamento dei modelli di deep learning. Questi parametri svolgono un ruolo cruciale nel governare il processo di apprendimento, influenzando direttamente le prestazioni e l’affidabilità di tali metodi nella pratica [44]. La loro regolazione consente di adattare il comportamento del modello per massimizzare le prestazioni e migliorare l’affidabilità nel contesto in cui vengono utilizzati.

Possiamo identificare alcuni iperparametri principali che possono contraddistinguere i diversi esperimenti di training. I primi esempi possono essere il numero di epoche, che definisce quante volte l’algoritmo lavora attraverso l’intero dataset, e la dimensione del batch, che indica invece il numero di campioni di dati che vengono elaborati dal modello prima dell’aggiornamento dei pesi.

In particolare per le reti analizzate , ovvero la U-Net e la Residual-U-Net, sono stati eseguiti diversi esperimenti di allenamento, variando il numero di epoche tra 60 e 100. Per motivi di utilizzo della memoria, la dimensione del batch è sempre stata mantenuta costante a 6. [45]

Un altro iperparametro importante è la funzione di loss, che nelle attività di apprendimento profondo valuta l’accuratezza dei modelli di previsione, misurando la differenza, e di conseguenza l’errore, tra il valore previsto e la verità di base. Se il valore della funzione è basso, ciò indica che i dati confrontati, in questo caso le immagini, sono simili tra loro. Negli esperimenti di training sono state utilizzate due funzioni di loss specifiche,

la Mean Square Error e la Mean Absolute Error. La Mean Squared error (MSE) calcola le differenza media quadratica tra le predizioni del modello e la ground truth. Essa viene definita dalla seguente formula:

$$MSE = \frac{1}{N} \sum_{t=1}^N (\hat{y}_t - y_t)^2 \quad (3.1)$$

La Mean Absolute Error (MAE) calcola le differenza media assoluta tra le predizioni del modello e la ground truth. Nelle librerie di Python questa funzione viene comunemente chiamata L1, nome che verrà utilizzato nelle sezioni successive. Essa viene definita dalla seguente formula:

$$L1 = \frac{1}{N} \sum_{t=1}^N |\hat{y}_t - y_t| \quad (3.2)$$

In entrambe le equazioni, N rappresenta il numero di campioni per ogni batch, \hat{y}_t sono i valori predetti dal modello e y_t rappresentano la ground truth. Nel caso di immagini, come in questo contesto, sia la MSE che la MAE quantificano la differenza tra i corrispondenti pixel delle immagini predette e quelle reali.

Oltre alla funzione di loss, viene anche impostata anche una funzione di attivazione, che fornisce la non linearità necessaria al modello per poter apprendere rappresentazioni complesse. La funzione utilizzata nei singoli strati convoluzionali è la ReLu, mentre come funzione di attivazione finale del modello è stata impiegata la funzione sigmoide

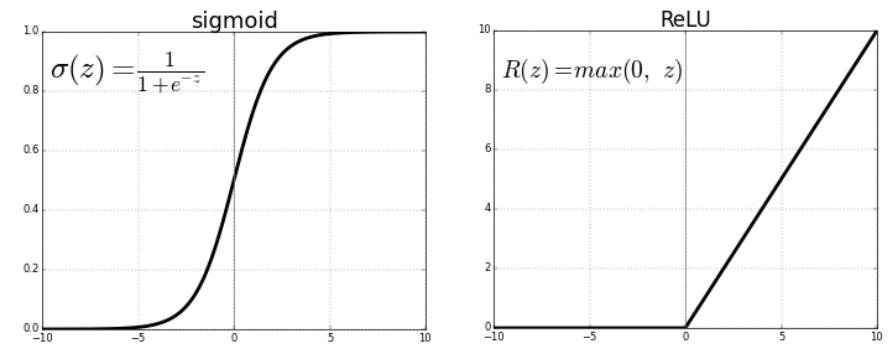
$$\sigma(x) = \frac{1}{1 + e^{-x}} \quad (3.3)$$

Questa funzione garantisce che l'output, in cui è mappato l'input, sia compreso tra 0 e 1 [46].

Nelle reti neurali inoltre vengono utilizzati degli algoritmi di ottimizzazione, che hanno l'obiettivo di trovare valori appropriati per ottimizzare i pesi e ridurre la funzione di loss. Uno degli algoritmi più utilizzati, e quello impiegato in questa tesi, è ADAM, che combina il concetto di discesa del gradiente con il momentum, memorizzando i gradienti passati per minimizzare la funzione di perdita e quindi trovare il valore minimo della funzione stessa.

Infine, è importante nominare un'ultimo iperparametro, ovvero il learning rate, che controlla la velocità con cui un modello apprende dai dati. Nel caso dei metodi che adottano l'approccio iterativo della discesa del gradiente, esso determina il cambiamento dei pesi. Nelle varie prove, il learning rate è stato mantenuto costante a 10^{-4}

Figura 3.2: Grafici delle funzioni di attivazione ReLu e sigmoide



Source: <https://towardsdatascience.com/>

3.2 Le metriche di valutazione

In questa sezione verranno illustrate le metriche utilizzate per valutare la qualità delle immagini in scala di grigi ricostruite nel contesto del miglioramento della qualità delle immagini di tomosintesi. Queste misure sono state calcolate confrontando l'immagine target con quella ricostruita dalla rete e con quella di input corrispondente.

Come metriva per calcolare l'errore, ovvero la differenza tra immagini, è stato utilizzato lo scarto quadratico medio (RMSE), definito come la radice quadrata della Mean Square Error. Esso valuta, nel caso specifico di immagini, le differenze tra pixel corrispondenti. Lo scarto quadratico medio è definito dalla seguente formula:

$$RMSE = \sqrt{\frac{1}{N} \sum_{i=1}^N (\hat{y}_i - y_i)^2} \quad (3.4)$$

dove \hat{y} sono le immagini target e y sono le immagini con cui esse vengono confrontate, che possono essere quelle di input oppure quelle predette dal modello. Piccole variazioni di pixel corrispondenti nelle due immagini possono influire notevolmente sul valore. Se si vuole che due immagini siano il più simili possibile, come nel caso di immagini ricostruite e di target corrispondente, il valore dell'RMSE deve essere il più basso possibile. Tuttavia lo scarto quadratico medio non tiene in considerazione la percezione visiva nelle immagini nella loro interezza, la quale è più sensibile a cambiamenti di struttura e contrasto piuttosto che alle differenze di pixel isolati.

Per questa ragione, è stato utile utilizzare un'altra metrica di valutazione, ovvero l'indice di similarità strutturale (SSIM). Questo indice è comune nell'ambiente medico per valutare la similarità delle strutture interne, quindi le forme, presenti nelle immagini rispetto a quelle di riferimento. Infatti esso si concentra sulla somiglianza visiva tra due immagini f e g in termini di luminanza l , contrasto c e struttura s .

$$SSIM(f, g) = l(f, g) * c(f, g) * s(f, g) \quad (3.5)$$

Il massimo valore ottenibile da l, c e g è 1 e di conseguenza il valore totale dell' SSIM varia tra 1, indicante immagini identiche in tutti e tre gli aspetti, e -1.

Infine è stato utilizzato come metrica di valutazione il Peak Signal-To-Noise Ratio (PSNR), misurato in decibel, equivalente al rapporto tra la potenza del rumore e la potenza massima del segnale. Il PSNR è particolarmente utile per evidenziare le differenze di qualità tra immagini rumorose, poiché altamente sensibile ai livelli di rumore. Il PSNR misura le differenze numeriche senza alcuna considerazione della conservazione della struttura.

$$PSNR = 10 \cdot \log_{10} \left(\frac{L^2}{\frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2} \right) \quad (3.6)$$

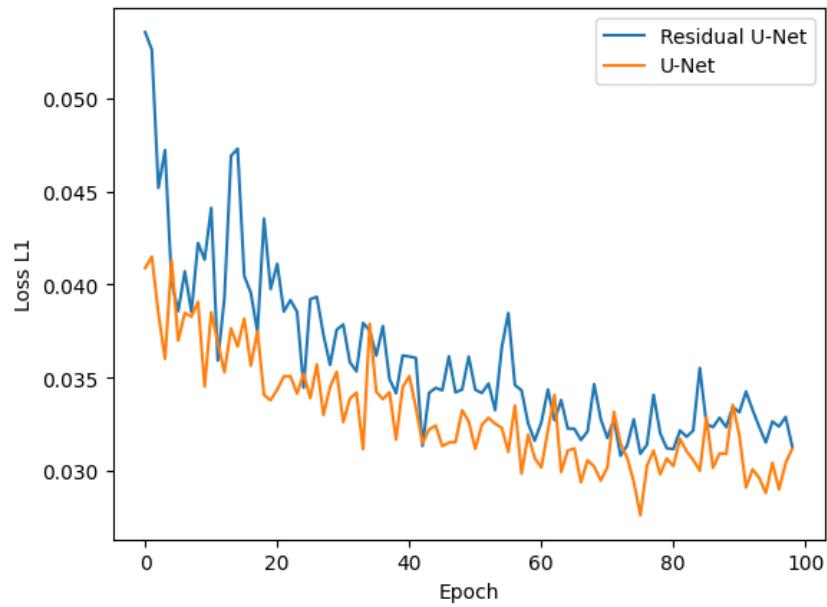
In questa formula, L corrisponde al rumore del segnale e il denominatore equivale al Mean Square Error calcolato tra l'immagine target e quella di input o ricostruita, che esprime la potenza massima del segnale. Un valore di PSNR più elevato indica una qualità migliore dell'immagine estratta in termini di riduzione del rumore. I valori tipici del PSNR variano da 20dB a 40dB e immagini di buona qualità hanno normalmente un valore superiore a 30 dB.

Nell'imaging medico la SSIM può essere considerata la metrica più importante nella ricostruzione, poiché essa si allinea maggiormente con la percezione umana nel valutare la qualità delle immagini e somiglianza strutturale, oltre alla percezione di dettagli anatomici critici. Questa caratteristica è fondamentale nei contesti di diagnosi, dove la fedeltà strutturale e la distinzione degli elementi di interesse è spesso più rilevante rispetto alla semplice differenza tra pixel distinti. Nella sezione 3.4 si approfondirà il concetto di qualità delle immagini nel contesto tomografico e diagnostico, esaminando come i modelli di ricostruzione possano ottenere tale qualità nelle immagini generate. Prima però vediamo un'analisi dell'andamento delle funzioni di loss durante l'allenamento dei modelli [47, 48, 49, 50, 51].

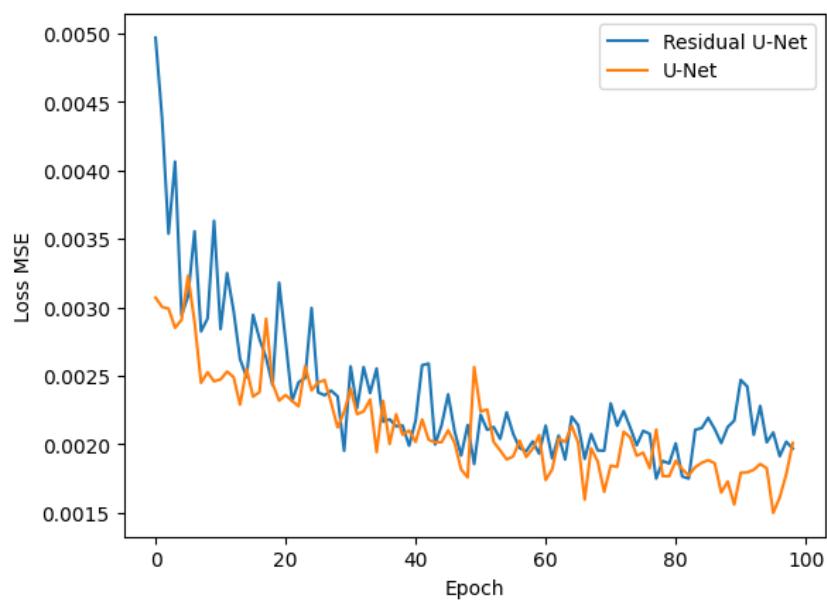
3.3 Evoluzione delle funzioni di loss durante l'addestramento

Figura 3.3: Confronto tra l'evoluzione delle funzioni di loss L1 e MSE durante il training

(a) Evoluzione della funzione di loss L1



(b) Evoluzione della funzione di loss MSE



Nella seguente sezione, sono presentate alcune considerazioni relative alla decrescita delle due funzioni di loss L1 e MSE applicate ai modelli considerati in questa tesi. In particolare i grafici della figura 3.2, fanno riferimento ad un allenamento di 100 epoche con una dimensione del batch uguale a 6. Dai grafici rappresentati nella figura 3.2, si può subito notare che la MSE presenta complessivamente valori più piccoli e una più evidente decrescita rispetto alla L1, soprattutto considerando la Residual U-Net. Infatti considerano le formule precedentemente definite di queste funzioni di loss, sappiamo che la L1, che calcola la media dei valori assoluti, considera ogni errore in maniera lineare, mentre la MSE, che invece calcola la media dei valori al quadrato, tende ad amplificare le differenze tra pixel maggiori, che influiscono sul risultato in modo proporzionale. Di conseguenza, valori minori nel grafico relativo alla MSE implicano un'ottima riduzione di errori significativi, indicando che gli algoritmi tendono a focalizzarsi maggiormente su questi per ridurre il loro impatto. Nonostante un'influenza lineare delle differenze nel risultato, si osservano oscillazioni significative in L1, implicando una minore stabilità della MSE in questo contesto.

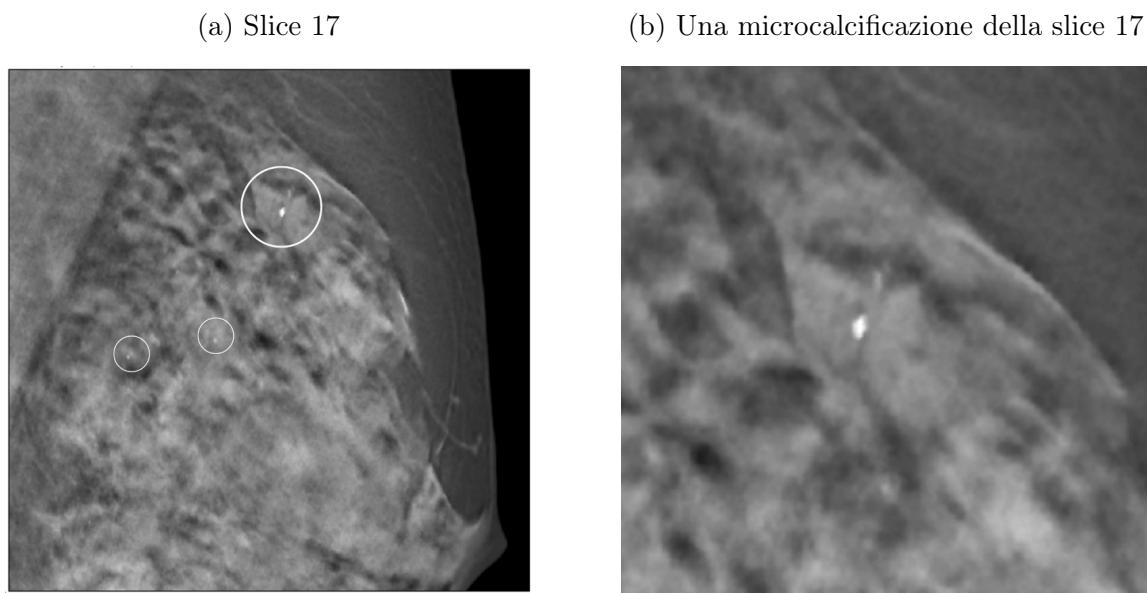
Analizzando nello specifico il grafico relativo alla funzione di loss L1, entrambi i modelli hanno inizialmente un errore molto alto, soprattutto la Residual U-Net con una loss iniziale corrispondente a 0.069. La funzione L1 ha un andamento decrescente nel corso delle epoche in entrambi i modelli. La Residual U-Net tuttavia presenta maggiori fluttazioni di valori, soprattutto nelle epoche iniziali, e valori della loss indicativamente maggiori rispetto alla U-Net. Da ciò si può dedurre che, nonostante la convergenza nella centesima epoca a valori simili, la U-net presenta prestazioni leggermente migliori.

Questa considerazione è applicabile anche nel caso della funzione di loss MSE, nonostante, come si può vedere nel secondo grafico rappresentato, la differenza sia meno netta. Come nel caso precedente, la Residual U-net ha un valore di loss iniziale molto più elevato della U-net, che invece presenta un valore relativamente basso fin da subito. Entrambe le curve dimostrano che l'andamento della funzione nei due modelli è decrescente, soprattutto nel caso della Residual U-net, la quale, nonostante partisse da un valore di loss significativamente più alto, già dalla ventesima epoca ha valori simili a quelli della U-net. Tuttavia la residual U-Net è caratterizzata da maggiori fluttazioni di valori come nel caso della funzione di loss L1. Analizzando le ultime 20 epoche, possiamo concludere che anche in questo caso la U-Net dimostra migliori prestazioni.

3.4 Valutazione delle immagini ricostruite

La qualità delle immagini prodotte in processi tomografici è cruciale per garantire diagnosi accurate e affidabili, poiché esse indirizzano direttamente le capacità del medico di identificare e analizzare elementi di interesse. Nel caso specifico della Tomosintesi Mammaria Digitale, una delle sfide principali della diagnosi è infatti l'identificazione corretta di elementi critici nel tessuto mammario, come le microcalcificazioni, che possono essere indicatori di tumori maligni o cancro al seno. Le microcalcificazioni sono piccoli depositi di calcio nei tessuti mammari, e la loro identificazione precoce può fare la differenza nella prognosi del paziente. Se le immagini non sono abbastanza nitide, c'è il rischio di non riconoscere queste strutture o di scambiare elementi non interessanti per microcalcificazioni. Nelle radiografie, queste ultime possono essere osservate sia nei casi di tumori benigni che nei casi di quelli maligni, distinte in base alla distribuzione spaziale, morfologia e cambiamento nel tempo. Quelle visualizzate maggiormente nelle immagini del dataset utilizzato, come quella mostrata nella figura 3.3, si presentano come strutture rotonde e compatte, con un centro chiaro e contorni regolari.[52]

Figura 3.4: Slice 17, con indicazioni su alcune microcalcificazioni presenti



La misura in cui le microcalcificazione sono visibili è direttamente legata alla qualità dell'immagine in termini di risoluzione, livelli di rumore e distorsione. Di conseguenza, è

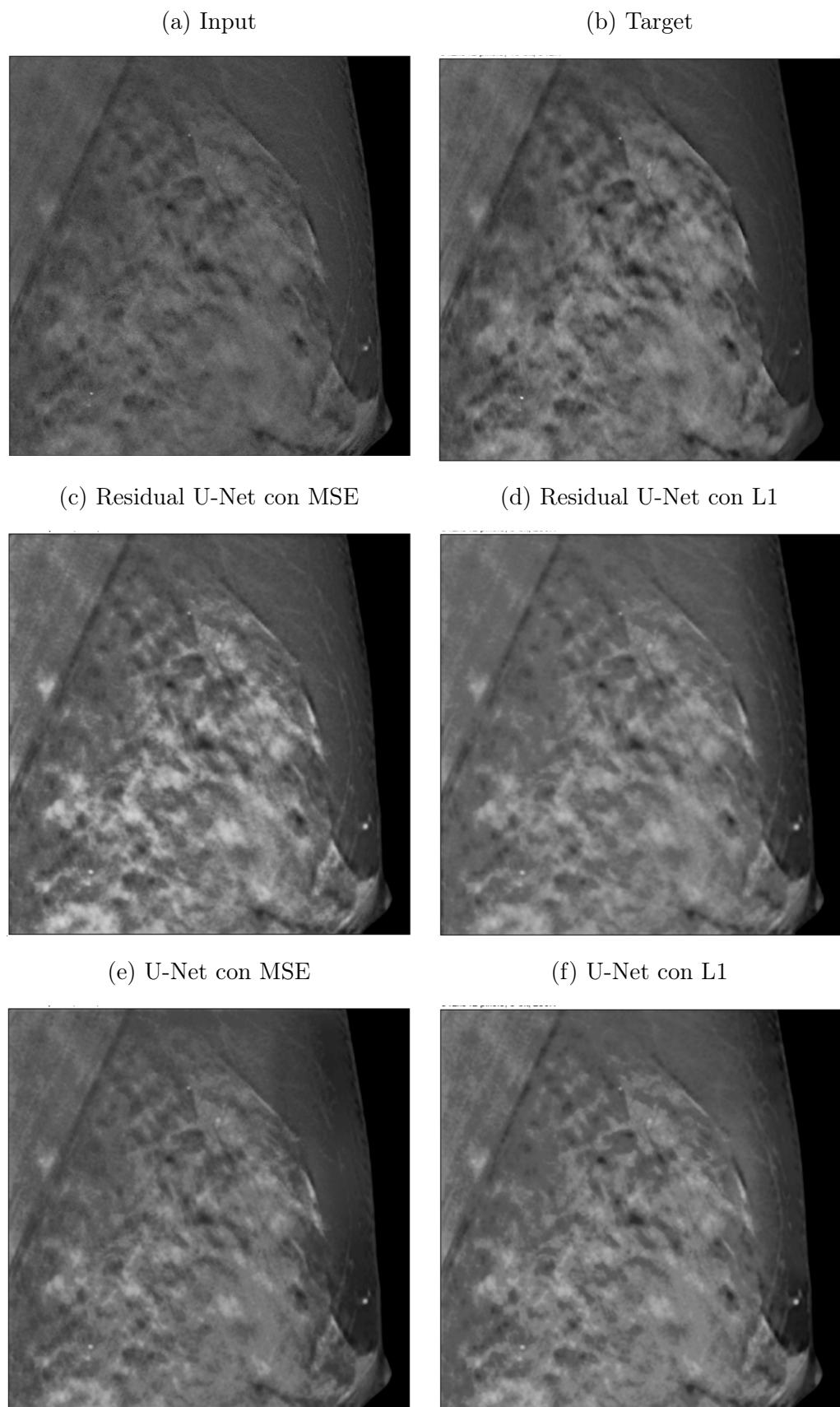
fondamentale esaminare i modelli di ricostruzione in relazione alla qualità e, soprattutto alla visibilità delle microcalcificazioni nelle immagini da essi prodotte. [53]

L'obiettivo degli strumenti di ricostruzione è garantire l'ottenimento di immagini nitide, nelle quali microcalcificazioni e altri elementi di interesse per la diagnosi siano chiaramente visibili e facilmente distinguibili da artefatti o da altre strutture circostanti. Oltre alla visibilità, è importante anche preservare la posizione, ma soprattutto la dimensione effettiva di queste microcalcificazioni. Quando un modello di ricostruzione, come le reti neurali, non riesce a migliorare adeguatamente la qualità delle immagini, non solo la visibilità delle microcalcificazioni può risultare compromessa, ma è anche possibile che vengano generati nuovi artefatti non presenti nell'immagine originale. È fondamentale, quindi, evitare l'introduzione di elementi finti e altre alterazioni, poiché ciò potrebbe portare a diagnosi errate.

Al fine di valutare questi aspetti delle immagini ricostruite, è necessario effettuare un confronto tra queste ultime con le immagini di input e di target. Tutte le ricostruzioni sono state prodotte da modelli allenati per 100 epoche e con dimensione del batch equivalente a 6. Inoltre tutte le immagini proposte in seguito si riferiscono a quelle della paziente di test, che sono state ricostruite dai modelli. Nella figura 3.4, è stato proposto un confronto tra immagini di input, target e ricostruite.

È evidente da questo esempio che le immagini ricostruite mostrano una qualità superiore rispetto a quelle di input, in particolare per quanto riguarda il contrasto e la nitidezza. È fondamentale, per le considerazioni elaborate precedentemente, sottolineare anche l'integrità delle strutture reali: gli elementi effettivi, in particolare le macrocalcificazioni, sono stati preservati senza alcuna aggiunta o rimozione. Tuttavia, nonostante i modelli combinati con le due funzioni di loss abbiano generato immagini più nitide rispetto agli input originali, le ricostruzioni differiscono tra loro nella qualità, soprattutto nel contrasto e nella resa delle microcalcificazioni. Il confronto di tali modelli e delle immagini prodotte da essi viene presentato nella sezione successiva.

Figura 3.5: Slice 21: confronto tra le ricostruzioni effettuate dai modelli con input e target.



3.5 Analisi comparativa dei modelli

Questa sezione si concentra sul confronto delle prestazioni tra i due modelli in esame, U-Net e Residual U-Net, utilizzando le funzioni di perdita MSE e L1. Esso si concretizza nell’analisi sia delle differenze visive tra le immagini ricostruite, sia tramite l’utilizzo di metriche di valutazione, approfondite nella sezione 3.2. Le metriche e le immagini qui riportate si riferiscono a un allenamento dei modelli condotto con un numero di epoch pari a 100 e una dimensione del batch di 6.

Iniziamo analizzando le prestazioni dei modelli U-Net e ResU-Net, evidenziando le differenze comportate dall’uso delle due funzioni di loss L1 e MSE attraverso i valori ottenuti nelle metriche di valutazione. E’ stata calcolata la media dei valori di ciascuna metrica per ogni confronto di immagini che verrà effettuato in seguito.

Dalla tabella 3.1 si può dedurre dal valore della metrica SSIM che le differenze strutturali tra le immagini di input e di target sono relativamente basse, ovvero i dettagli sono ben preservati. Inoltre il valore corrispondente all’RMSE è relativamente basso, indicando un errore medio quadratico ridotto tra le immagini di input e target. Tuttavia, il PSNR, pari a 24.6102, indica che la qualità dell’immagine in termini di rumore in input rispetto a quella di target è buona, ma non eccellente.

Tabella 3.1: Confronto tra Immagini di Input e Target tramite metriche di valutazione

Metric	value
PSNR	24.6102
SSIM	0.9246
RMSE	0.0637

Considerando invece il confronto delle due funzioni di loss, nel caso della U-Net, si può dedurre dalla tabella 3.2 che l’utilizzo della funzione L1 risulta più performante in questo modello, con valori di tutte e tre le metriche migliori rispetto a quelli ottenuti con MSE.

Questo può suggerire che le immagini ottenute dal modello U-Net con L1 come funzione di loss sono qualitativamente migliori rispetto a quelle ottenute da MSE in ter-

Tabella 3.2: U-Net: Confronto tra MSE e L1 utilizzando metriche di valutazione

Metric	MSE	L1
PSNR	27.2992	27.6718
SSIM	0.9386	0.9460
RMSE P	0.0570	0.0554

mini di riduzione del rumore, precisione tra i pixel e conservazione delle caratteristiche strutturali.

Considerando invece la Residual U-Net, si osserva che la qualità delle immagini prodotte dal modello utilizzando le due funzioni di loss non presenta differenze significative, indicando che entrambi i metodi offrono prestazioni comparabili.

I risultati suggeriscono che l'utilizzo della funzione di perdita L1 nella Residual U-Net offre una leggera superiorità in termini di PSNR e RMSE, migliorando la qualità complessiva dell'immagine e riducendo l'errore rispetto a MSE. Tuttavia, esaminando la metrica SSIM, considerata quella più rilevante nel contesto della ricostruzione, si osserva che, nonostante entrambe le funzioni di loss generino immagini con una buona somiglianza strutturale, la MSE presenta un valore della metrica leggermente superiore. Ciò suggerisce una migliore preservazione delle caratteristiche strutturali con l'utilizzo della MSE, anche se questa differenza non è significativa.

Nel complesso, l'utilizzo della funzione di loss L1 nella Residual U-Net suggerisce una qualità delle immagini leggermente superiore, meno influenzate dal rumore. Tuttavia bisogna sottolineare che è altrettanto importante considerare le caratteristiche strutturali, meglio preservate nelle immagini ricostruite da U-Net con MSE come funzione di Loss.

Tabella 3.3: Residual U-Net: Confronto tra MSE e L1 utilizzando metriche di valutazione

Metric	MSE	L1
PSNR	28.0337	28.8369
SSIM	0.9495	0.9417
RMSE	0.0518	0.0512

Sebbene la differenza non sia particolarmente decisiva, dall'analisi effettuata si può dedurre che la prestazione della funzione di loss L1 applicata alla ricstruzione sia

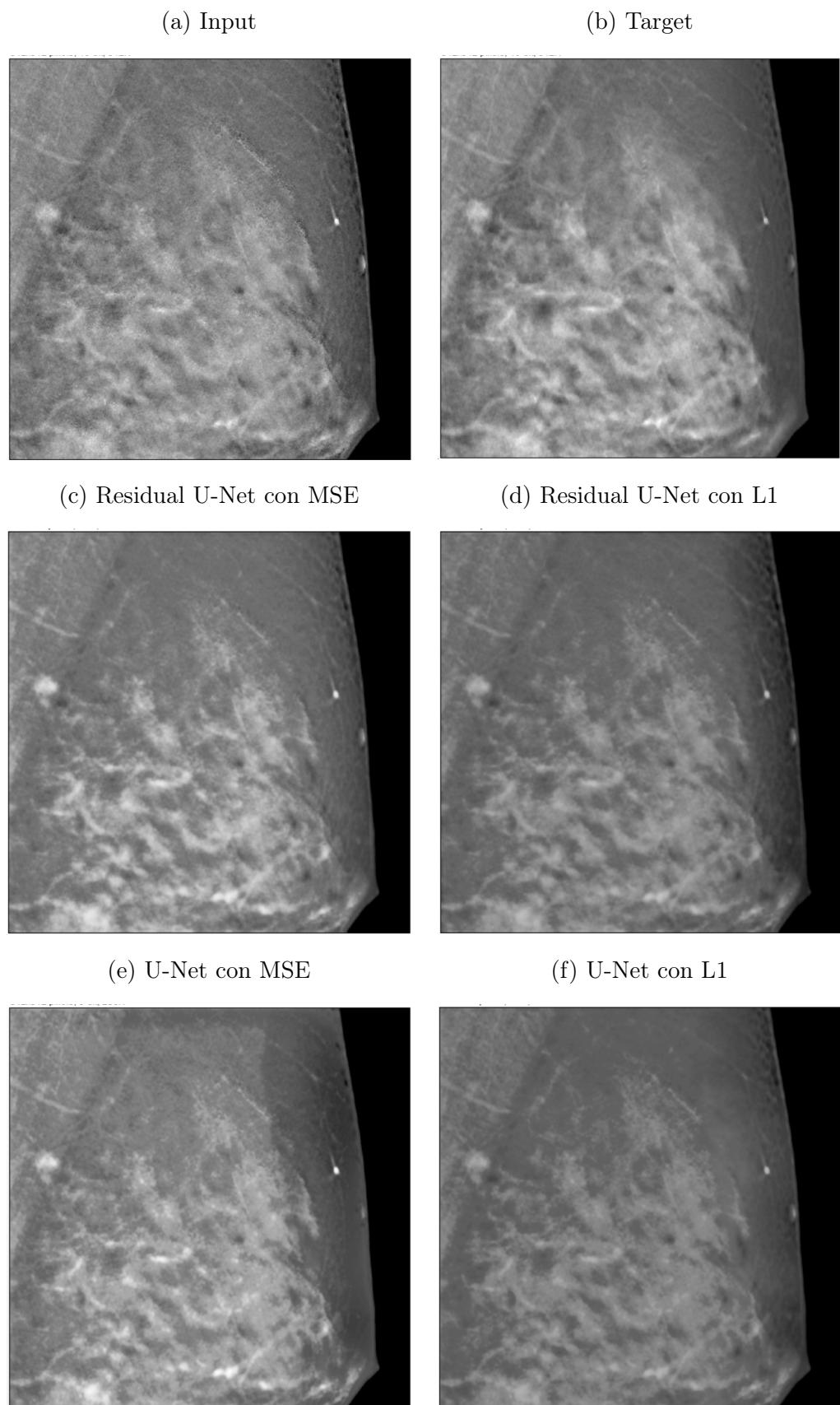
migliore rispetto alla MSE in entrambi i modelli.

Tuttavia, questa differenza in termini di prestazioni non è rispecchiata nel confronto percepito ad occhio nudo delle immagini prodotte. Ad esempio consideriamo la figura 3.5, che illustra immagini relative alla sezione 34, caratterizzate da evidenti microcalcificazioni, identificabili già nell'input.

Possiamo vedere dalle immagini che quelle prodotte utilizzando la MSE risultano più contrastate e nitide, mettendo maggiormente in evidenza sia la struttura dei tessuti che le microcalcificazioni, che appaiono più definite e chiare. Al contrario, nelle immagini generate con la L1, tali elementi non sono risaltati allo stesso modo, risultando confusi e meno distinguibili. Tale considerazione può essere applicata ad entrambi i modelli, che risultano più performanti con la funzione di loss MSE.

Queste variazioni facilmente percepibili nelle immagini prodotte, anche se piccole, indicano che la MSE produce in realtà immagini migliori in termini di rumore e nitidezza, caratteristiche che mettono in risalto anche la visibilità degli elementi di interesse. Infine, possiamo osservare che le immagini prodotte con la MSE mostrano una maggiore somiglianza con il target, in quanto anch'esso caratterizzato da un contrasto più marcato rispetto a quelle generate con la L1.

Figura 3.6: Slice 34: Confronto tra le ricostruzioni effettuate da Residual U-Net e U-Net, combinate con le funzioni di loss L1 e MSE.



Nelle considerazioni effettuate sui grafici rappresentanti le funzioni di loss, è stato evidenziato il fatto che la MSE è più sensibile ad errori significativi tra i pixel. Ciò può comportare una migliore conservazione dei dettagli fini, poiché il modello è incentivato a minimizzare non solo le differenze complessive, ma anche a preservare le piccole variazioni che possono costituire dettagli critici. Le immagini prodotte utilizzando la funzione di loss L1 al contrario, possono apparire meno dettagliate poiché il modello si focalizza sulla minimizzazione della media delle differenze. con questa funzione di loss gli errori significativi in alcune aree dell'immagine non influiscono in modo sostanziale sul risultato complessivo, portando a una ricostruzione che può trascurare alcuni dettagli.

la figura 3.5 è ottima anche per mettere in risalto non solo la superiorità dei risultati ottenuti dalla residual U-Net utilizzando la funzione di loss MSE, ma anche la qualità inferiore delle immagini generate dalla U-Net con la funzione di loss L1 rispetto alle altre.

Per approfondire questo punto, riorganizziamo ora i dati delle tabelle precedenti per mettere a confronto diretto le prestazioni dei due modelli.

Tabella 3.4: Analisi comparativa delle differenze tra immagini ricostruite con diversi modelli, combinati con diverse funzioni di loss

Metric	MSE		L1	
	U-Net	Res U-Net	U-Net	Res U-Net
PSNR	27.2992	<u>28.0337</u>	27.6718	<u>28.8369</u>
SSIM	0.9386	<u>0.9495</u>	<u>0.9460</u>	0.9417
RMSE	0.0570	<u>0.0518</u>	0.0554	<u>0.0512</u>

Nella tabella 3.4, emerge che la Residual U-Net presenta prestazioni migliori rispetto alla U-Net, indipendentemente dalla funzione di loss utilizzata. In particolare, con la funzione di loss MSE, le immagini sono migliori soprattutto in termini di PSNR e SSIM, che indicano una qualità visiva superiore delle immagini ricostruite. Anche l'RMSE risulta migliore nella Residual U-Net, sebbene la differenza rispetto alla U-Net non sia altrettanto marcata.

Ciò è in contrasto con quanto osservato nei grafici nella sezione 3.3, dove la una funzione di loss della U-Net appariva più stabile e leggermente inferiore. Tuttavia, è plausibile che la Residual U-Net, grazie alla presenza di skip connection residue che

facilitano una stretta connessione tra l’immagine di input e quella generata, sia in grado maggiormente di preservare e riprodurre con maggiore fedeltà i dettagli anatomici presenti nelle immagini originali, riflettendoli correttamente in quelle ricostruite [54].

Con l’utilizzo della funzione di loss L1, la Residual U-Net continua a mostrare prestazioni superiori rispetto alla U-Net, sebbene le differenze siano meno pronunciate rispetto all’uso della MSE. In particolare, l’RMSE ma soprattutto il PSNR evidenziano una qualità di ricostruzione migliore nella Residual U-Net. Tuttavia, l’SSIM indica che la U-Net riesce a conservare in modo più efficace le caratteristiche strutturali delle immagini.

Nella figura 3.6, è proposto un esempio di immagini prodotte tra la Residual U-Net e U-Net, per vedere se le loro differenze rispecchiano quelle emerse dal confronto dei valori delle metriche.

Si può notare da questa figura che effettivamente la Residual U-Net produce immagini più nitide rispetto alla U-Net. Come nell’esempio precedente, tra le immagini ricostruite, quella che si distingue per qualità superiore è quella generata dalla Residual U-Net utilizzando la MSE, mentre l’immagine meno convincente risulta essere quella prodotta dalla U-Net con la funzione di loss L1. Quest’ultima infatti appare poco chiara, a causa del basso contrasto che rende difficile distinguere gli elementi di interesse, come le microcalcificazioni. Nella sezione precedente è stata evidenziata proprio l’importanza della visibilità di questi elementi, poiché le immagini saranno poi impiegate come strumento di diagnosi. Di conseguenza questo fattore è un criterio fondamentale per la valutazione dei modelli e delle funzioni di loss. Nelle figure 3.7 e 3.8, è stato proposto un confronto delle rappresentazioni della microcalcificazione più visibile nell’esempio precedente.

Figura 3.7: Slice 17 : Confronto tra le ricostruzioni effettuate da Residual U-Net e U-Net, combinate con le funzioni di loss L1 e MSE.

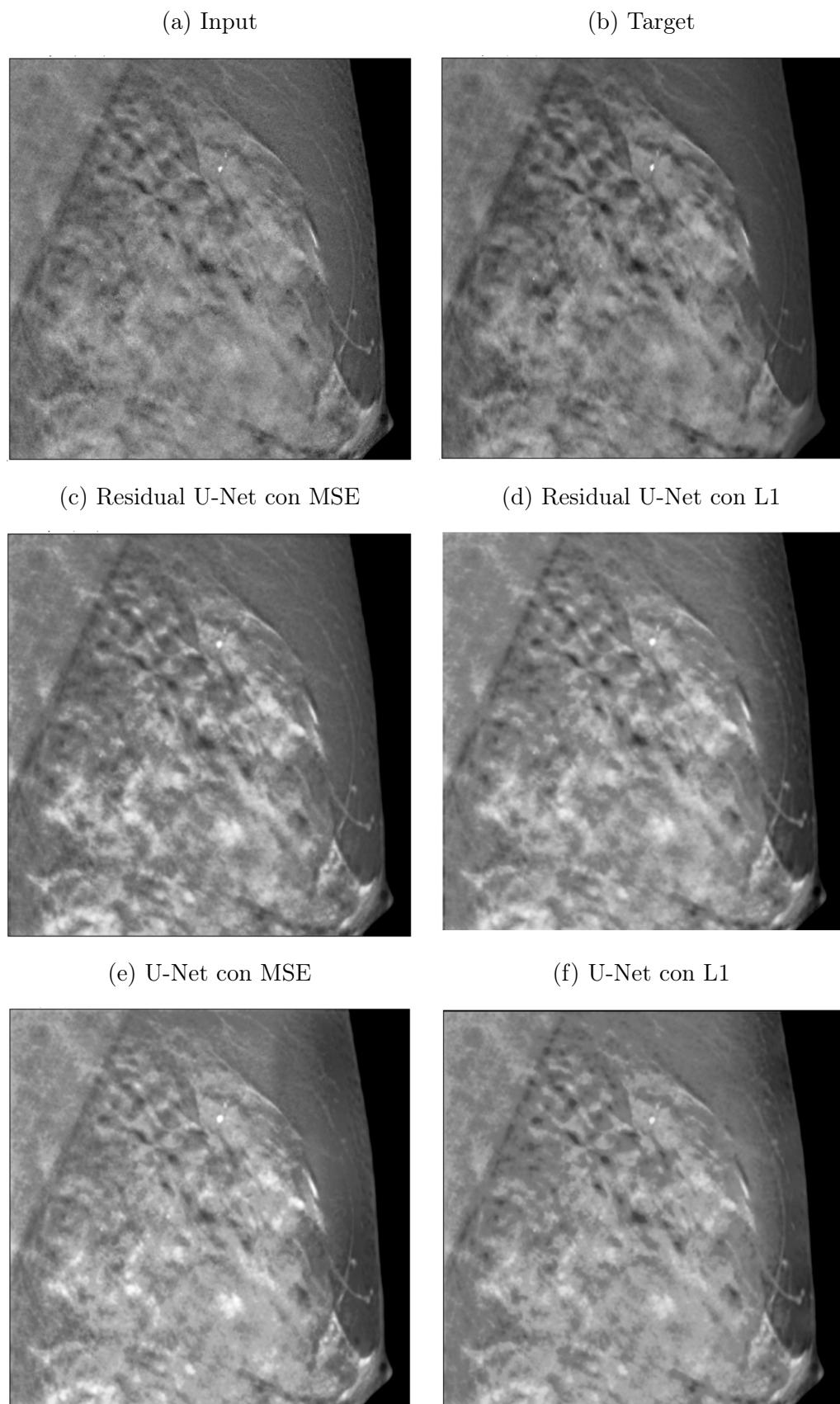
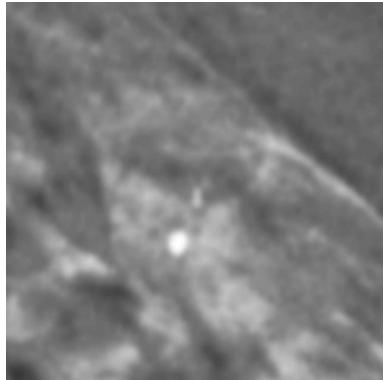


Figura 3.8: Ingrandimento di una microcalcificazione nella Slice 17: Residual Unet

(a) Residual Unet con MSE



(b) Residual Unet con L1

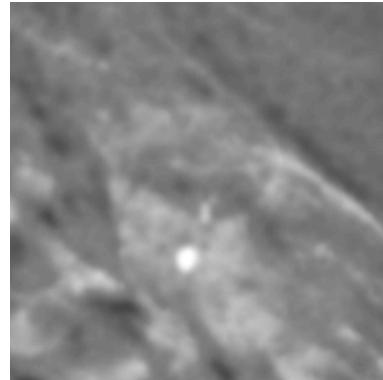
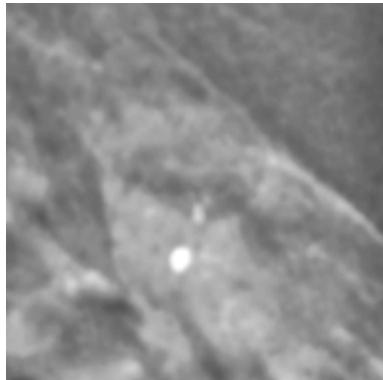
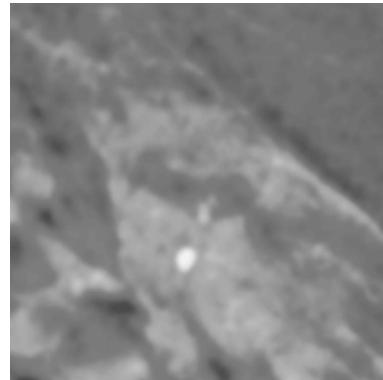


Figura 3.9: Ingrandimento di una microcalcificazione nella Slice 17 : U-Net

(a) Unet con MSE



(b) Unet con L1



E' evidente che con l'utilizzo della Residual U-Net utilizzata in combinazione con la MSE, la microcalcificazione risulta più visibile, anche grazie al contrasto di essa con i tessuti cicostanti. Le immagini prodotte dalla U-Net con MSE e dalla Residual U-Net con L1 ingrandite presentano una qualità approssimativamente simile, sebbene nella seconda il contrasto sia più marcato. Al contrario, la microcalcificazione appare meno visibile nell'immagine generata dalla U-Net con L1, dove i contorni risultano meno definiti e il ridotto contrasto con l'area circostante ne rende difficile l'individuazione.

Bisogna tenere in considerazione che le prestazioni dei modelli applicati per ricostruire l'immagine dipendono dalle caratteristiche dell'immagine di input. Infatti ci sono casi in cui le differenze di prestazioni dei modelli sono meno marcate. Tuttavia possiamo ricono-

scere nella Residual U-net una migliore abilità complessiva nel ricostruire immagini che riescono ad evidenziare maggiamente i dettagli interessanti

Si propongono infine ulteriori ingrandimenti di microcalcificazioni, in particolare della slice 21, rappresentati nelle figure 3.9 e 3.10. E' possibile notare che i dettagli della microcalcificazione presente nell'immagine prodotta dalla Residual U-Net utilizzando la funzione di loss MSE, sono accentuati rispetto alle altre, le quali invece non differiscono particolarmente le une dalle altre.

Figura 3.10: Ingrandimento di una microcalcificazione nella Slice 21 : Residual U-Net

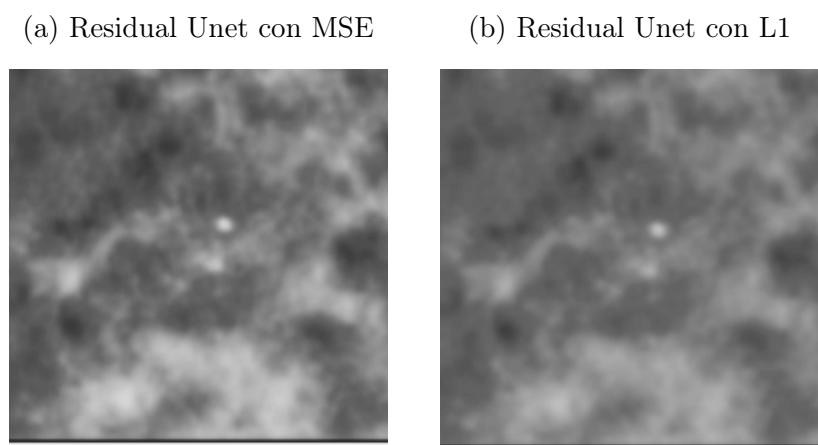


Figura 3.11: Ingrandimento di una microcalcificazione nella Slice 21 : U-Net

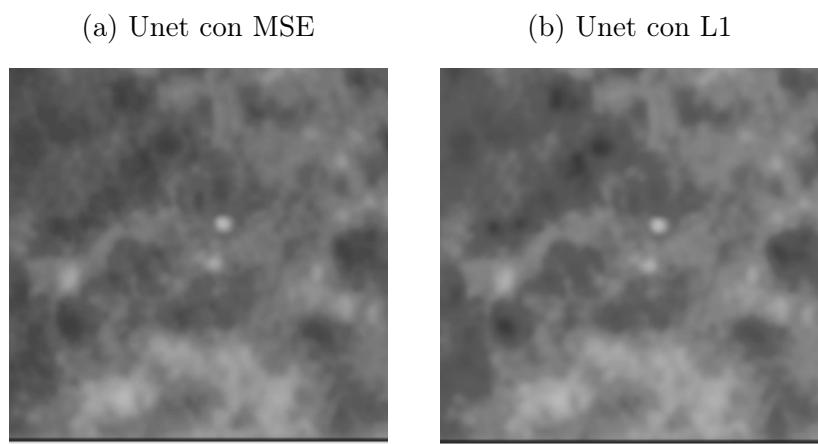
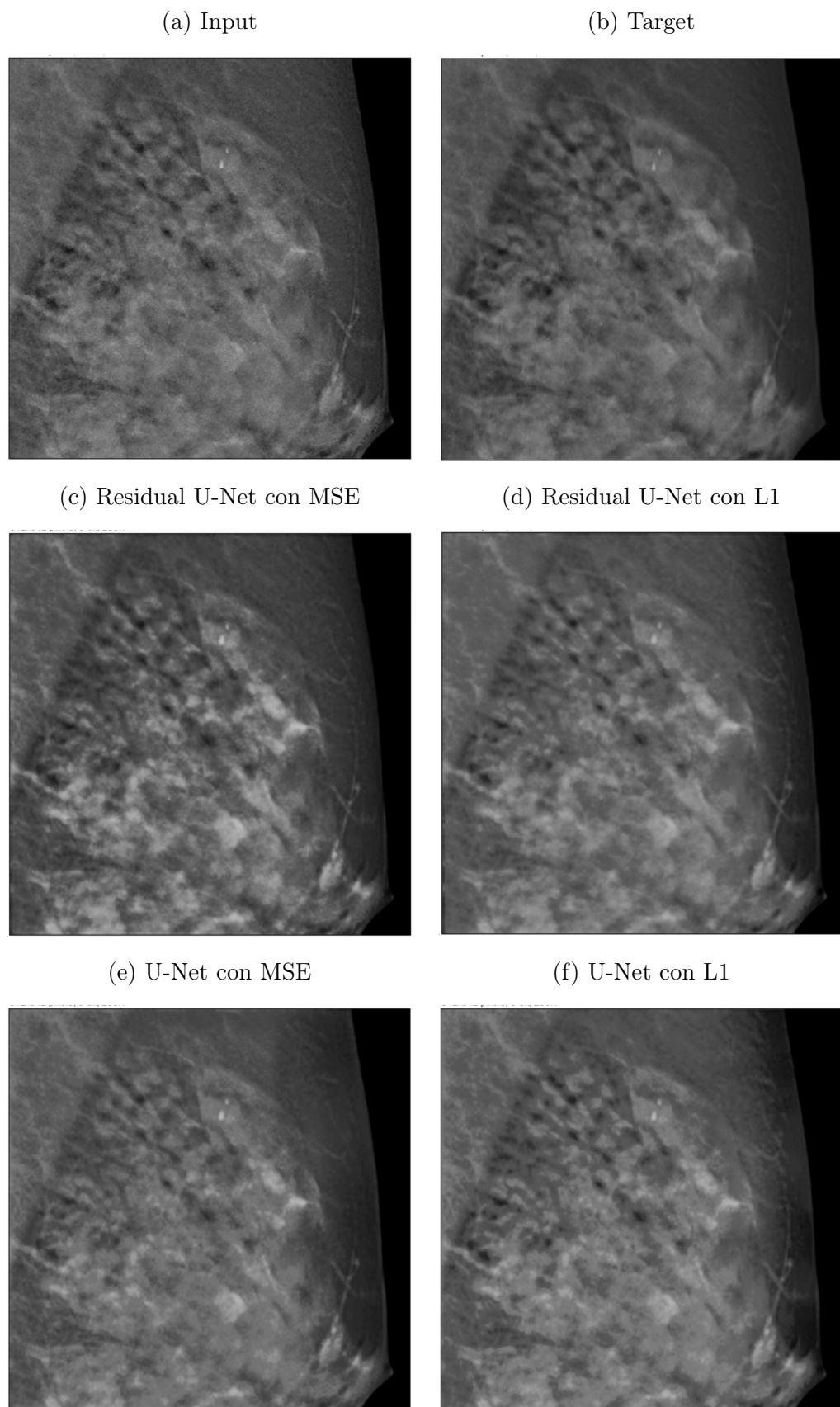


Figura 3.12: Slice 12: Confronto tra le ricostruzioni effettuate da Residual U-Net e U-Net, combinate con le funzioni di loss L1 e MSE.



Conclusioni

I modelli U-Net e Residual U-Net, utilizzati per la ricostruzione di immagini nel contesto dell’imaging medico, hanno dimostrato di migliorare la qualità delle immagini di input, sebbene con risultati differenti. L’adozione di diverse funzioni di perdita ha avuto un impatto significativo sulla qualità delle immagini finali, evidenziando una leggera superiorità della MSE.

È evidente dalle analisi che la U-Net ha ottenuto risultati inferiori rispetto alla sua versione residuale, in particolare quando è stata utilizzata insieme alla funzione di perdita L1. La visibilità ridotta delle microcalcificazioni presenti nelle immagini ricostruite da questa combinazione, pur essendo chiaramente distinguibili già in quelle di input, potrebbe suggerire che in situazioni in cui tali calcificazioni siano meno evidenti, il modello potrebbe non essere altrettanto efficace nel garantirne l’identificazione.

È importante sottolineare che, sebbene le metriche siano stati strumenti fondamentali per valutare le differenze tra le immagini, in alcuni casi non sono riuscite a definire con precisione quale fosse la ricostruzione migliore in relazione all’immagine target corrispondente. Ad esempio, nel caso della U-Net, sebbene le immagini prodotte utilizzando la Mean Squared Error (MSE) fossero qualitativamente superiori rispetto a quelle ottenute con la L1, i valori di tutte e tre le metriche hanno indicato i risultati opposti. Questo solleva interrogativi sull’adeguatezza delle metriche tradizionali nel catturare la complessità delle immagini mediche. Poiché la valutazione visiva delle immagini è fondamentale, considerando che esse fungono da strumenti diagnostici per i professionisti del settore, è impossibile condurre un’analisi delle immagini mediche esclusivamente attraverso metriche quantitative.

Infine è importante sottolineare che gli esperimenti descritti in questa tesi sono stati concepiti come un primo passo preliminare, propedeutico alla sua estensione per ri-

costruzioni tridimensionali. Infatti, dopo aver validato l'efficacia dei modelli e delle funzioni di loss, è possibile procedere all'applicazione degli stessi principi a immagini tridimensionali, effettivamente utilizzate nel contesto dell'imaging medico.

Bibliografia

- [1] Davide Evangelista, Elena Morotti e Elena Loli Piccolomini. «RISING: A new framework for model-based few-view CT image reconstruction with deep learning». In: *Computerized Medical Imaging and Graphics* 103 (2023), p. 102156. ISSN: 0895-6111. DOI: <https://doi.org/10.1016/j.compmedimag.2022.102156>. URL: <https://www.sciencedirect.com/science/article/pii/S0895611122001264>.
- [2] Marco Salvatore e Arturo Brunetti. «La grande scienza. Imaging diagnostico computerizzato». In: *Storia della Scienza* (2003).
- [3] Elena Morotti e Elena Loli Piccolomini. *Sparse Regularized CT Reconstruction: An Optimization Perspective*. Liverpool: Ke Chen, 2022. DOI: [10.1007/978-3-030-03009-4_123-1](https://doi.org/10.1007/978-3-030-03009-4_123-1).
- [4] Martin Berger et al. «X-ray Imaging». In: *Medical Imaging Systems: An Introductory Guide*. A cura di Andreas Maier et al. Springer, 2018. Cap. 7.
- [5] R C Chen et al. «Measurement of the linear attenuation coefficients of breast tissues by synchrotron radiation computed tomography». In: *Physics in Medicine and Biology* 55.17 (ago. 2010), p. 4993. DOI: [10.1088/0031-9155/55/17/008](https://doi.org/10.1088/0031-9155/55/17/008). URL: <https://dx.doi.org/10.1088/0031-9155/55/17/008>.
- [6] Wolfram Stiller. «Basics of iterative reconstruction methods in computed tomography: A vendor-independent overview». In: *European Journal of Radiology* 109 (2018), pp. 147–154. ISSN: 0720-048X. DOI: <https://doi.org/10.1016/j.ejrad.2018.10.025>. URL: <https://www.sciencedirect.com/science/article/pii/S0720048X18303747>.

- [7] Zhou Yu et al. «Fast Model-Based X-Ray CT Reconstruction Using Spatially Nonhomogeneous ICD Optimization». In: *IEEE Transactions on Image Processing* 20.1 (2011), pp. 161–175. DOI: [10.1109/TIP.2010.2058811](https://doi.org/10.1109/TIP.2010.2058811).
- [8] C.M. McLeavy et al. «The future of CT: deep learning reconstruction». In: *Clinical Radiology* 76.6 (2021), pp. 407–415. ISSN: 0009-9260. DOI: <https://doi.org/10.1016/j.crad.2021.01.010>. URL: <https://www.sciencedirect.com/science/article/pii/S0009926021000672>.
- [9] Ke Li, Jie Tang e Guang-Hong Chen. «Statistical Model Based Iterative Reconstruction (MBIR) in Clinical CT Systems: Experimental Assessment of Noise Performance». In: *Medical Physics* 41.4 (2014).
- [10] Emil Y. Sidky et al. «Do CNNs Solve the CT Inverse Problem?» In: *IEEE Trans. Biomed. Eng.* 68.6 (2021), pp. 1799–1810. DOI: [10.1109/TBME.2020.3020741](https://doi.org/10.1109/TBME.2020.3020741). URL: <https://doi.org/10.1109/TBME.2020.3020741>.
- [11] Harshit Gupta et al. «CNN-Based Projected Gradient Descent for Consistent CT Image Reconstruction». In: *IEEE Transactions on Medical Imaging* 37.6 (2018), pp. 1440–1453. DOI: [10.1109/TMI.2018.2832656](https://doi.org/10.1109/TMI.2018.2832656).
- [12] Stefania Naccarato et al. «Tomosintesi: caratterizzazione dosimetrica». In: set. 2009.
- [13] Ming Fan et al. «Computer-aided detection of mass in digital breast tomosynthesis using a faster region-based convolutional neural network». In: *Methods* 166 (2019). Deep Learning in Bioinformatics, pp. 103–111. ISSN: 1046-2023. DOI: [10.1016/j.ymeth.2019.02.010](https://doi.org/10.1016/j.ymeth.2019.02.010). URL: <https://www.sciencedirect.com/science/article/pii/S1046202318303608>.
- [14] Zewen Li et al. «A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects». In: *IEEE Transactions on Neural Networks and Learning Systems* 33.12 (2022), pp. 6999–7019. DOI: [10.1109/TNNLS.2021.3084827](https://doi.org/10.1109/TNNLS.2021.3084827).
- [15] Dobin Yim et al. «Deep learning-based image reconstruction for few-view computed tomography». In: *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment* 1011 (2021), p. 165594. ISSN: 0168-9002. DOI: <https://doi.org/10.1016/j.nima.2021.165594>.

2021.165594. URL: <https://www.sciencedirect.com/science/article/pii/S0168900221005799>.

- [16] Timothy P. Szczykutowicz et al. «A Review of Deep Learning CT Reconstruction: Concepts, Limitations, and Promise in Clinical Practice». In: *Current Radiology Reports* 10 (2022).
- [17] Lennart R. Koetzier et al. «Deep Learning Image Reconstruction for CT: Technical Principles and Clinical Prospects». In: *Radiology* 306.3 (2023). PMID: 36719287, e221257. DOI: 10.1148/radiol.221257. eprint: <https://doi.org/10.1148/radiol.221257>. URL: <https://doi.org/10.1148/radiol.221257>.
- [18] Young June Kim et al. «Deep Learning-Based Denoising Algorithm Compared to Iterative Reconstruction and Filtered Back Projection: A 12-Reader Phantom Study». In: *European Radiology* 31 (2021). URL: <https://doi.org/10.1007/s00330-021-07810-3>.
- [19] Qing Li et al. «Low-dose Computed Tomography Image Reconstruction via a Multi-stage Convolutional Neural Network with Autoencoder Perceptual Loss Network». In: *Quantitative Imaging in Medicine and Surgery* 12.3 (2022).
- [20] Jonas Teuwen et al. «Deep learning reconstruction of digital breast tomosynthesis images for accurate breast density and patient-specific radiation dose estimation». In: *Medical Image Analysis* 71 (2021), p. 102061. ISSN: 1361-8415. DOI: <https://doi.org/10.1016/j.media.2021.102061>. URL: <https://www.sciencedirect.com/science/article/pii/S1361841521001079>.
- [21] Teja Kattenborn et al. «Review on Convolutional Neural Networks (CNN) in vegetation remote sensing». In: *ISPRS Journal of Photogrammetry and Remote Sensing* 173 (2021), pp. 24–49. ISSN: 0924-2716. DOI: <https://doi.org/10.1016/j.isprsjprs.2020.12.010>. URL: <https://www.sciencedirect.com/science/article/pii/S0924271620303488>.
- [22] Tianyi Liu et al. «Implementation of Training Convolutional Neural Networks». In: *CoRR* abs/1506.01195 (2015). arXiv: 1506.01195. URL: <http://arxiv.org/abs/1506.01195>.

- [23] Fahad Alrasheedi, Xin Zhong e Pei-Chi Huang. «Padding Module: Learning the Padding in Deep Neural Networks». In: *IEEE Access* 11 (2023), pp. 7348–7357. DOI: [10.1109/ACCESS.2023.3238315](https://doi.org/10.1109/ACCESS.2023.3238315).
- [24] Nadeem Akhtar e U. Ragavendran. «Interpretation of intelligence in CNN-pooling processes: a methodological survey». In: *Neural Comput. Appl.* 32.3 (2020), pp. 879–898. DOI: [10.1007/S00521-019-04296-5](https://doi.org/10.1007/S00521-019-04296-5). URL: <https://doi.org/10.1007/S00521-019-04296-5>.
- [25] Andrinandrasana David Rasamoelina, Fouzia Adjailia e Peter Sinčák. «A Review of Activation Function for Artificial Neural Network». In: *2020 IEEE 18th World Symposium on Applied Machine Intelligence and Informatics (SAMI)*. 2020, pp. 281–286. DOI: [10.1109/SAMI48414.2020.9108717](https://doi.org/10.1109/SAMI48414.2020.9108717).
- [26] Andrea Apicella et al. «A survey on modern trainable activation functions». In: *Neural Networks* 138 (2021), pp. 14–32. ISSN: 0893-6080. DOI: [10.1016/j.neunet.2021.01.026](https://doi.org/10.1016/j.neunet.2021.01.026). URL: <https://www.sciencedirect.com/science/article/pii/S0893608021000344>.
- [27] Michael T. McCann, Kyong Hwan Jin e Michael Unser. «Convolutional Neural Networks for Inverse Problems in Imaging: A Review». In: *IEEE Signal Processing Magazine* 34.6 (2017), pp. 85–95. DOI: [10.1109/MSP.2017.2739299](https://doi.org/10.1109/MSP.2017.2739299).
- [28] Matthew D. Zeiler et al. «Deconvolutional networks». In: *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*. 2010, pp. 2528–2535. DOI: [10.1109/CVPR.2010.5539957](https://doi.org/10.1109/CVPR.2010.5539957).
- [29] Jin Zheng e Lihui Peng. «An Autoencoder-Based Image Reconstruction for Electrical Capacitance Tomography». In: *IEEE Sensors Journal* 18.13 (2018), pp. 5464–5474. DOI: [10.1109/JSEN.2018.2836337](https://doi.org/10.1109/JSEN.2018.2836337).
- [30] Pengzhi Li, Yan Pei e Jianqiang Li. «A comprehensive survey on design and application of autoencoder in deep learning». In: *Applied Soft Computing* 138 (2023), p. 110176. ISSN: 1568-4946. DOI: [10.1016/j.asoc.2023.110176](https://doi.org/10.1016/j.asoc.2023.110176). URL: <https://www.sciencedirect.com/science/article/pii/S1568494623001941>.

- [31] Anita Khanna et al. «A deep Residual U-Net convolutional neural network for automated lung segmentation in computed tomography images». In: *Biocybernetics and Biomedical Engineering* 40.3 (2020), pp. 1314–1327. ISSN: 0208-5216. DOI: 10.1016/j.bbe.2020.07.007. URL: <https://www.sciencedirect.com/science/article/pii/S0208521620300887>.
- [32] Hongyang Gao et al. «Pixel Transposed Convolutional Networks». In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* 42.5 (2020), pp. 1218–1227. DOI: 10.1109/TPAMI.2019.2893965.
- [33] Ning Han et al. «Multi-level U-net network for image super-resolution reconstruction». In: *Displays* 73 (2022), p. 102192. ISSN: 0141-9382. DOI: 10.1016/j.displa.2022.102192. URL: <https://www.sciencedirect.com/science/article/pii/S0141938222000361>.
- [34] Nahian Siddique et al. «U-Net and Its Variants for Medical Image Segmentation: A Review of Theory and Applications». In: *IEEE Access* 9 (2021), pp. 82031–82057. DOI: 10.1109/ACCESS.2021.3086020.
- [35] Norel Ya Qine Abderrahim, Saadane Abderrahim e Azmi Rida. «Road Segmentation using U-Net architecture». In: *2020 IEEE International conference of Moroccan Geomatics (Morgeo)*. 2020, pp. 1–4. DOI: 10.1109/Morgeo49228.2020.9121887.
- [36] Allard A. Hendriksen et al. «Deep denoising for multi-dimensional synchrotron X-ray tomography without high-quality reference data». In: *Scientific Reports* 11 (2021), p. 11895. DOI: 10.1038/s41598-021-91084-8.
- [37] Nabil Ibtehaz e M. Sohel Rahman. «MultiResUNet : Rethinking the U-Net architecture for multimodal biomedical image segmentation». In: *Neural Networks* 121 (2020), pp. 74–87. ISSN: 0893-6080. DOI: <https://doi.org/10.1016/j.neunet.2019.08.025>. URL: <https://www.sciencedirect.com/science/article/pii/S0893608019302503>.
- [38] Lina Huang et al. «Segmenting Medical Images: From UNet to Res-UNet and nnUNet». In: (lug. 2024). DOI: 10.48550/arXiv.2407.04353.

- [39] Hao Zhang et al. «Infrared Image Segmentation for Photovoltaic Panels Based on Res-UNet». In: ott. 2019, pp. 611–622. ISBN: 978-3-030-31653-2. DOI: 10.1007/978-3-030-31654-9_52.
- [40] Michal Drozdzal et al. «The Importance of Skip Connections in Biomedical Image Segmentation». In: *Deep Learning and Data Labeling for Medical Applications - First International Workshop, LABELS 2016, and Second International Workshop, DLMIA 2016, Held in Conjunction with MICCAI 2016, Athens, Greece, October 21, 2016, Proceedings*. A cura di Gustavo Carneiro et al. Vol. 10008. Lecture Notes in Computer Science. 2016, pp. 179–187. DOI: 10.1007/978-3-319-46976-8_19. URL: https://doi.org/10.1007/978-3-319-46976-8%5C_19.
- [41] Tadashi Wadayama e Satoshi Takabe. «Deep Learning-Aided Trainable Projected Gradient Decoding for LDPC Codes». In: *2019 IEEE International Symposium on Information Theory (ISIT)*. 2019, pp. 2444–2448. DOI: 10.1109/ISIT.2019.8849215.
- [42] S Bonettini, R Zanella e L Zanni. «A scaled gradient projection method for constrained image deblurring». In: *Inverse Problems* 25.1 (nov. 2008), p. 015002. DOI: 10.1088/0266-5611/25/1/015002. URL: <https://dx.doi.org/10.1088/0266-5611/25/1/015002>.
- [43] E. Loli Piccolomini et al. «Reconstruction of 3D X-ray CT images from reduced sampling by a scaled gradient projection algorithm». In: *Comput. Optim. Appl.* 71.1 (2018), pp. 171–191. DOI: 10.1007/S10589-017-9961-2. URL: <https://doi.org/10.1007/s10589-017-9961-2>.
- [44] Prasanna Balaprakash et al. «DeepHyper: Asynchronous Hyperparameter Search for Deep Neural Networks». In: *2018 IEEE 25th International Conference on High Performance Computing (HiPC)*. 2018, pp. 42–51. DOI: 10.1109/HiPC.2018.00014.
- [45] Aditya Devarakonda, Maxim Naumov e Michael Garland. «AdaBatch: Adaptive Batch Sizes for Training Deep Neural Networks». In: *ArXiv* abs/1712.02029 (2017). URL: <https://api.semanticscholar.org/CorpusID:3302972>.

- [46] Won-Kee Hong. «Understanding artificial neural networks: analogy to the biological neuron model». In: gen. 2023, pp. 7–13. ISBN: 9780443152528. DOI: 10.1016/B978-0-443-15252-8.00003-0.
- [47] Yiming Lei et al. «CT Image Denoising and Deblurring With Deep Learning: Current Status and Perspectives». In: *IEEE Transactions on Radiation and Plasma Medical Sciences* 8.2 (2024), pp. 153–172. DOI: 10.1109/TRPMS.2023.3341903.
- [48] Mónica Chillarón et al. «Computed tomography medical image reconstruction on affordable equipment by using Out-Of-Core techniques». In: *Computer Methods and Programs in Biomedicine* 193 (2020), p. 105488. ISSN: 0169-2607. DOI: <https://doi.org/10.1016/j.cmpb.2020.105488>. URL: <https://www.sciencedirect.com/science/article/pii/S0169260719316190>.
- [49] Alexander Tanchenko. «Visual-PSNR measure of image quality». In: *Journal of Visual Communication and Image Representation* 25.5 (2014), pp. 874–878. ISSN: 1047-3203. DOI: <https://doi.org/10.1016/j.jvcir.2014.01.008>. URL: <https://www.sciencedirect.com/science/article/pii/S1047320314000091>.
- [50] De Rosal Igantius Moses Setiadi. «PSNR vs SSIM: imperceptibility quality assessment for image steganography». In: *Multimedia Tools and Applications* 80 (2021). Access provided by Università degli Studi di Bologna - Area Biblioteche e Servizi allo Studio, pp. 8423–8444. DOI: 10.1007/s11042-020-10035-z. URL: <https://doi.org/10.1007/s11042-020-10035-z>.
- [51] Dimple Sethi, Sourabh Bharti e Chandra Prakash. «A comprehensive survey on gait analysis: History, parameters, approaches, pose estimation, and future work». In: *Artificial Intelligence in Medicine* 129 (2022), p. 102314. ISSN: 0933-3657. DOI: <https://doi.org/10.1016/j.artmed.2022.102314>. URL: <https://www.sciencedirect.com/science/article/pii/S0933365722000793>.
- [52] Azadeh Elmi et al. «Calcifications on DBT and Synthetic Views: Update and Management Strategies». In: *Current Radiology Reports* 8 (mag. 2020). DOI: 10.1007/s40134-020-00352-4.
- [53] Nur Athiqah Harron et al. «Microcalcification Detection for Digital Breast Tomosynthesis Images Using Faster-RCNN». In: *2024 IEEE 14th International Conference on Image Processing Theory, Tools and Applications (IPTA)* (2024).

rence on Control System, Computing and Engineering (ICCSCE). 2024, pp. 150–155. DOI: [10.1109/ICCSCE61582.2024.10696812](https://doi.org/10.1109/ICCSCE61582.2024.10696812).

- [54] Hu Chen et al. «Low-Dose CT With a Residual Encoder-Decoder Convolutional Neural Network». In: *IEEE Transactions on Medical Imaging* 36.12 (2017), pp. 2524–2535. DOI: [10.1109/TMI.2017.2715284](https://doi.org/10.1109/TMI.2017.2715284).

Ringraziamenti

Qui possiamo ringraziare il mondo intero!!!!!!!!!

Ovviamente solo se uno vuole, non è obbligatorio.