Alma Mater Studiorum · Università di Bologna

SCUOLA DI SCIENZE
Corso di Laurea in Informatica

RETI NEURALI PER RICOSTRUZIONE DI IMMAGINI TOMOGRAFICHE DENTALI A BASSA DOSE

Relatrice:
Chiar.ma Prof.
Elena Loli Piccolomini
Correlatore:
Dott.
Davide Evangelista

Presentata da:
Filippo Speziali

III Sessione
Anno Accademico 2023/2024

Abstract

In questa tesi viene studiato l'utilizzo di reti neurali convoluzionali per la ricostruzione di immagini tomografiche a bassa dose. In particolare, verranno confrontate le prestazioni di una rete Unet con due sue varianti che implementano connessioni residuali per input bidimensionali, insieme a una variante adatta per input tridimensionali. Sono stati condotti esperimenti su due tipi di dataset di immagini tomografiche ricostruite con l'algoritmo Feldkamp-Davis-Kress (FDK). Il primo dataset è composto da 3 volumi acquisiti su dei fantocci reali in cui si abbassa il numero di proiezioni. In particolare è composto sia da volumi con metà che da un quarto delle proiezioni. Un secondo dataset è composto da 3 volumi acquisiti sui fantocci dopo averli esposti a 100 volte la dose di radiazioni standard. Infine la rete tridimensionale viene testata su un paziente reale, sia per ricostruire un immagine a metà viste sia simulando una dose elevata.

Indice

Introduzione **1**

1 Tomografia a bassa dose **1**
1.1 L'imaging e la tomografia computerizzata 1
1.1.1 Proiezioni e retroproiezioni 2
1.1.2 Limitare la dose 5
1.2 Dataset 6
1.2.1 Dataset ad alta dose 6
1.2.2 Dataset sottocampionato 8
1.2.3 Paziente reale 9

2 Modelli **11**
2.1 Reti neurali convoluzionali 11
2.1.1 Rete encoder-decoder 15
2.2 Reti studiate 16
2.2.1 Unet 17
2.2.2 Residual Unet 18
2.2.3 Residual Unet 3d 20

3 Risultati **23**
3.1 Iperparametri 23
3.2 Overfitting 24
3.3 Confronto numerico sui modelli 27
3.3.1 Dataset ad alta dose 27
3.3.2 Dataset con metà delle proiezioni 28
3.3.3 Dataset con un quarto delle proiezioni 29
3.4 Confronto sulle immagini ricostruite 30
3.4.1 Dataset ad alta dose 30
3.4.2 Dataset con metà delle proiezioni 32
3.4.3 Dataset con un quarto delle proiezioni 36

3.5 Risultati sui pazienti reali 38
3.5.1 Conversione ad alta dose 38
3.5.2 Conversione con metà delle proiezioni 40
3.6 Conclusioni 41

Introduzione

Nell'ambito dell'imaging medico, l'integrazione dell'intelligenza artificiale sta rivoluzionando le pratiche diagnostiche e terapeutiche, aprendo nuove frontiere nella ricostruzione e nell'analisi delle immagini. In questa tesi si approfondirà l'utilizzo di questa tecnologia nell'ambito dell'imaging tomografico, che fornisce informazioni dettagliate e tridimensionali delle strutture anatomiche interne. In particolare verranno utilizzate delle reti neurali profonde per la ricostruzione delle immagini tomografiche dentali a bassa dose. L'obiettivo di queste ricostruzioni è ottenere delle immagini di qualità, utilizzabili per la diagnostica, riducendo il piú possibile la quantità di radiazioni elettromagnetiche che impattano il paziente. Le reti sono state addestrate utilizzando un metodo supervisionato, ovvero sono stati forniti i dati delle immagini da ricostruire in input insieme alle relative ricostruzioni desiderate in output. I dati su cui le reti sono state allenate sono stati gentilmente forniti dall' azienda SeeThrough S.r.l [1]. La ricostruzione delle immagini a bassa dose avviene mediante due approcci. Il primo approccio è la ricostruzione di un immagine acquisita abbassando il numero di proiezioni. In questo approccio le immagini presentano numerosi artefatti che la rete proverà a rimuovere. La seconda strategia consiste nell'usare una dose di radiazioni più elevata del normale per simulare un effetto ad alta dose. In questo caso le immagini a dose standard sono meno rumorose delle immagini ad alta dose, la rete deve imparare a rimuovere questo rumore. Infine SeeThrough ha fornito anche il volume prodotto da una scansione su un paziente reale per verificare se l'apprendimento delle ricostruzioni sui fantocci possa applicarsi anche nella realtà. Le reti prese in analisi sono delle varianti del modello Unet con l'aggiunta di connessioni residuali. In particolare vengono utilizzate e confrontate 3 reti che lavorano su due dimensioni e infine viene presentata una rete che che lavora sul 3d. Nel primo capitolo saranno trattati i temi dell'imaging e della tomografia computerizzata seguiti dall'illustrazione dei metodi volti a ridurre l'esposizione alle radiazioni nella sezione 1. In seguito, sarà presentato il dataset utilizzato per condurre gli esperimenti di questa tesi nella sezione 1.2. Nel secondo capitolo, verranno esaminati i metodi di intelligenza artificiale noti per risolvere questo genere di problematiche nella sezione 2.1. Successivamente, nella sezione 2.2, saranno illustrate le

reti neurali impiegate sul dataset a disposizione. All'inizio del capitolo 3, verranno esposti gli esperimenti che sono stati condotti. Successivamente nella sezione 3.3, saranno mostrati i risultati delle ricostruzioni, mentre nella sezione 3.4 verranno analizzate le immagini prodotte dalle reti. Infine nella sezione 3.5 sarà mostrato l'utilizzo della rete su un paziente reale.

Capitolo 1

Tomografia a bassa dose

1.1 L'imaging e la tomografia computerizzata

L'imaging medico è un insieme di procedure grazie alle quali è possibile conoscere ed esaminare una precisa area del corpo umano, non visibile dall'esterno, attraverso delle immagini. Questo processo nasce dopo la scoperta dei raggi X nel 1895 da parte di Wilhelm Conrad Roentgen. I raggi X sono onde elettromagnetiche in grado di penetrare i tessuti del corpo umano, che vengono poi assorbiti in modo differente da tessuti con diverse densità. La loro scoperta è alla base della nascita della radiografia, che può essere considerata la prima pratica di imaging medico. In questa pratica una sorgente di raggi X viene fatta passare attraverso il corpo e finisce su un rilevatore. In base alla densità dei tessuti attraversati verrà trattenuta una certa quantità di raggi X. I tessuti densi, come le ossa, assorbono più raggi X rispetto ai tessuti molli. L'immagine risultante è una rappresentazione bidimensionale delle varie strutture interne, con aree più chiare che indicano una maggiore trasmissione dei raggi X e aree più scure che indicano una maggiore assorbanza.
La radiografia tradizionale presenta dei limiti. Le proiezioni prodotte generano immagini bidimensionali di un corpo tridimensionale. Alcune informazioni possono essere nascoste o imprecise. Questo effetto dipende dal fatto che il fascio incidente subisce l'attenuazione di tutti i tessuti che incontra lungo il suo tragitto, con la conseguenza che quando raggiunge il rilevatore porta con sé questa informazione. Per superare queste limitazioni è nata la tomografia computerizzata che ha rivoluzionato la diagnosi medica ma ha anche avuto enormi effetti su altri campi [5].
La tomografia computerizzata, detta TC (in inglese Computed Tomography o CT), è una tecnica di imaging medico che utilizza raggi X per creare immagini dettagliate del corpo umano. A differenza della radiografia, che produce immagini bidimensionali di una singola proiezione, la tomografia genera immagini tridimensionali del

corpo. Il suo principio è l'utilizzo di proiezioni provenienti da più angoli, che permettono di ricostruire l'immagine dell'oggetto. Il primo utilizzo pratico è stato nel 1971, anno in cui è stato eseguito il primo esame TAC. Da allora sono emerse altre modalità di imaging tomografico, come la risonanza magnetica (MRI) o la tomografia ad ultrasuoni, che hanno reso la TC una parte indispensabile della medicina.
Nelle prossime sezioni verrà approfondito il funzionamento della TC introducendo i concetti teorici fisici e matematici che descrivono il processo di acquisizione dei dati e in seguito verrà presentato il dataset su cui sono stati svolti gli studi di questa tesi.

1.1.1 Proiezioni e retroproiezioni

Quando i raggi X entrano in contatto con la materia si verificano diversi fenomeni fisici tra la radiazione e i tessuti. La loro lunghezza d'onda è molto breve, compresa tra i 0.01 e i 10 nanometri, molto più breve della luce visibile (400 - 700 nanometri). L'energia che viene prodotta da una radiazione elettromagnetica si può calcolare dalla relazione di Plank-Einstein:

$$E = \frac{hc}{\lambda} \tag{1.1}$$

E è l'energia del fotone, h la costante di Plank, c la velocità della luce e λ la lunghezza d'onda. Si può notare, dunque, come ad una lunghezza d'onda più corta corrisponde un energia più elevata. Grazie a queste proprietà i raggi X sono in grado di penetrare la materia. Attraversando la materia alcuni fotoni verranno assorbiti, altri deviati a seconda del materiale con cui vanno a contatto. Quando un fascio di raggi X di intensità I_0 passa attraverso un oggetto con densità uniforme la sua attenuazione I si può calcolare attraverso la legge di Lambert Beer's [9]

$$I = I_0 e^{-\mu d} \tag{1.2}$$

dove d è lo spessore dell'oggetto e μ è il coefficiente di attenuazione lineare dell'oggetto.

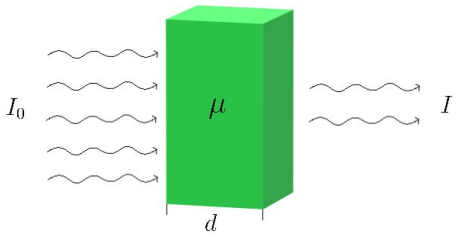


Figura 1.1: Diagramma schematizzato dei raggi X che attraversano una sostanza omogenea

Nell'equazione (1.2), I, I_0 e d possono tutti essere misurati. Il coefficiente di attenuazione lineare può essere ottenuto come $\mu = ln(\frac{I_0}{I})/d$. Il valore di μ riflette le proprietà fisiche di diverse sostanze ed è correlato al valore del pixel dell'immagine tomografica. Quando un fascio di raggi X passa attraverso n oggetti con densità uniforme I_i e diversi coefficienti di attenuazione μ_i l'equazione (2) diventa:

$$I = I_0 e^{-(\mu_1 d_1 + ... + \mu_n d_n)} \tag{1.3}$$

A questo punto si può trovare un'equazione a più variabili da cui è possibile calcolare i coefficienti di attenuazione.

$$\sum_{i=0}^{n} \mu_i d_i = ln\left(\frac{I_0}{I}\right) \tag{1.4}$$

Nella pratica, però, le sostanze di un oggetto non sono equamente distribuite. Per calcolare la proiezione dei raggi X lungo il percorso L, si utilizza l'integrale dei coeffiecienti di attenuazione dei materiali lungo L:

$$p = \int_L \mu dl = ln\left(\frac{I_0}{I}\right) \tag{1.5}$$

A questo punto nella TC moderna si procede facendo variare l'angolo con cui vengono emessi i fasci di raggi X ottenendo più proiezioni dello stesso oggetto tridimensionale.

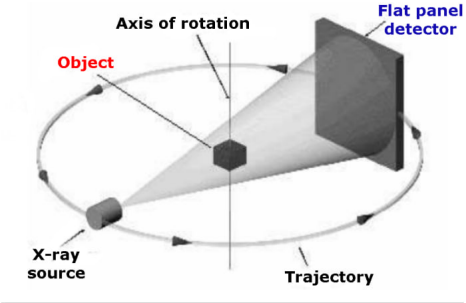


Figura 1.2: Esempio di scanner TC, i raggi X vengono emessi in un fascio conico

Producendo molteplici proiezioni da angolazioni diverse, nel caso bidimensionale si genera una rappresentazione grafica dei dati nota come sinogramma [7]. Quest'ultimo è una visualizzazione grafica in cui gli angoli di proiezione sono rappresentati su un asse, mentre sull'altro asse è indicata la posizione lungo il rilevatore. I valori contenuti nei pixel del sinogramma indicano la quantità di radiazione rilevata per ciascuna proiezione. In caso di tomografia volumetrica, invece, si ottiene una mappa tridimensionale.

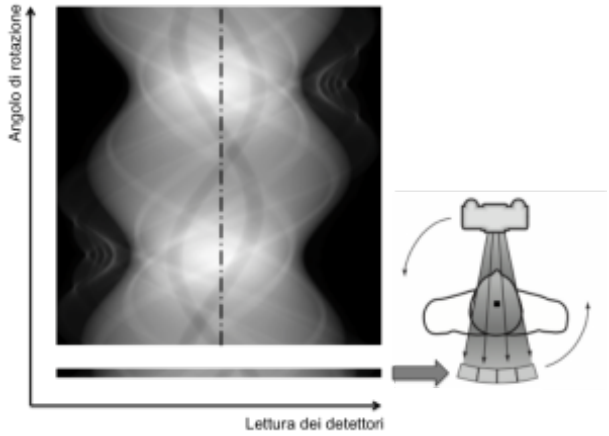


Figura 1.3: Esempio di sinogramma

A questo punto per arrivare ad avere un immagine tomografica bisogna compiere una retroproiezione. L'obiettivo è quello di trovare i diversi valori di attenuazione, ciascuno dei quali contribusice al valore di un certo pixel nell'immagine. La somma dei contributi di ciascuna retroproiezione su un'unica sezione permette di ottenere la ricostruzione tomografica. Più proiezioni si hanno, maggiore sarà la qualità dei risultati ottenuti. Uno degli alogoritmi piu' utilizzati per fare retroproiezione si chiama Filtered back projection (FBP). Questo algoritmo applica un filtro per rimuovere

il rumore dovuto all'alta frequenza e successivamente effettua una retroproiezione. Nel 3d ci si riferisce all'FBP come l'algoritmo di Feldkamp-Davis-Kress (FDK) [2] che è stato utilizzato per ricostruire le immagini del dataset di questa tesi. Gli algoritmi analitici come l'FDK producono molto rumore, sono quindi preferibili alcuni algoritmi iterativi ma richiedono più tempo di calcolo per ricostruire l'immagine.

1.1.2 Limitare la dose

La tomografia è un'innovativa tecnica di imaging che ha rivoluzionato la diagnosi e il monitoraggio di numerose condizioni mediche. Tuttavia, nonostante i suoi evidenti vantaggi, la tomografia può comportare un'esposizione del paziente alle radiazioni ionizzanti, che costituiscono un rischio per la salute a lungo termine, aumentando la probabilità di sviluppare patologie come il cancro. Per mitigare questo rischio, e limitare l'influenza dei raggi X sul paziente, si possono adottare diverse strategie, tra cui l'impiego di tecniche come la tomografia a viste sparse o l'applicazione di dosi ridotte di radiazioni.

L'utilizzo di viste sparse consiste nell' abbassare il numero di proiezioni, consentendo di ridurre l'esposizione complessiva alle radiazioni e acquisendo immagini in tempi più rapidi. Il problema che sorge utilizzando questa tecnica è che, diminuendo le proiezioni, si perdono informazioni sull'immagine. Applicando l'algoritmo di ricostruzione FBP sui dati incompleti si ottiene una qualità dell'immagine molto degradata. In particolare si ottengono diversi artefatti, come striature nell'immagine in cui mancano informazioni che potrebbero essere preziose per una diagnosi corretta.

Anche la semplice riduzione del numero di fotoni dei raggi X, controllando la corrente o la tensione del tubo, consentono di abbassare la dose di radiazione. In questo caso l'algoritmo di ricostruzione FBP produrrà immagini rumorose.

In entrambi i casi vengono prodotte delle immagini che prima di essere utilizzate dai medici devono essere migliorate e ricostruite. Uno degli approcci possibili è tramite l'apprendimento automatico utilizzando delle reti neurali convolutive. L'apprendimento automatico supervisionato è un ramo dell'intelligenza artificiale in cui si addestrano modelli computazionali su un insieme di esempi di input e output desiderato. Durante il processo di addestramento il modello cerca di imparare una mappatura tra gli input e gli output. Il termine "supervisionato" deriva dal fatto che durante l'addestramento il modello conosce l'output ideale che dovrebbe produrre, chiamato ground truth. Ovviamente meno dati si hanno a disposizione più questo compito risulta difficile. Avere a disposizione un grande numero di dati non è facile, sia per via della privacy dei pazienti sia perchè essendo la tomografia una pratica in cui ci si espone ai raggi X non è possibile ripetere tante volte questi esami. Un

approccio possibile è utilizzare dei fantocci, cioè degli oggetti progettati per essere scansionati o addirittura un vero teschio umano. Nel caso di questa tesi vengono utilizzati i dati provenienti dalle scansioni di 3 fantocci umani. Questo ha permesso di ottenere immagini utilizzando una dose elevata, che sui pazienti reali non sarebbe possibile utilizzare, che viene usata come ground truth. Si cerca quindi di ottenere un effetto ad alta dose partendo da una dose standard. Inoltre sono stati prodotti due dataset in cui varia il numero delle proiezioni. Rispettivamente si è provato a generare immagini con metà e con un quarto delle proiezioni. Nel prossimo paragrafo verrà illustrato nel dettaglio il dataset utilizzato.

1.2 Dataset

I dati a disposizione sono dei volumi tomografici dentali, ricostruiti tramite FDK, di 3 fantocci umani. La rete è stata allenata su due di questi e un terzo è stato tenuto come test. Sebbene sui fantocci si possano effettuare tutte le scansioni che si vogliono presentano delle differenze importanti con i pazienti reali. In primo luogo i fantocci non generano artefatti da movimento. Questo perchè restano immobili per tutta la scansione a differenza di un paziente reale. Inoltre i tessuti sono simulati e non presentano saliva. Per questo oltre alle scansioni sui tre fantocci viene considerato un paziente reale per testare le prestazioni della rete. Le immagini dei dataset presi in considerazione hanno tutte la stessa dimensione. Sono immagini 836 x 836 pixel e in particolare per ogni volume ci sono 920 sezioni.

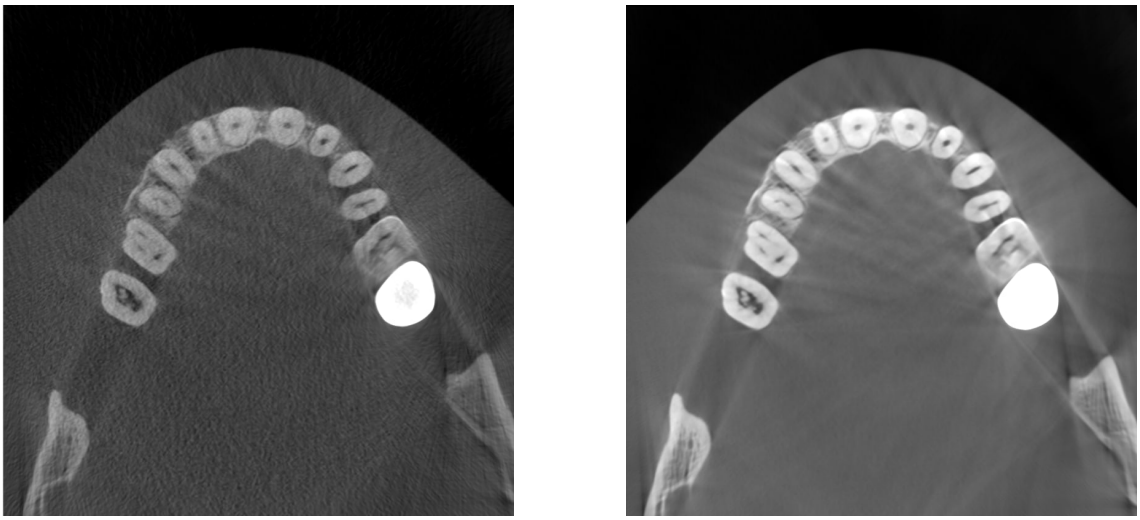
1.2.1 Dataset ad alta dose

Il primo tipo di scansione che è stata fatta è una scansione ad "alta dose". Più precisamente viene utilizzata 100 volte la dose standard. Con una dose così elevata si ottengono delle immagini precisissime. L'obiettivo della conversione della rete è quello di migliorare la qualità delle immagini simulando una scansione ad alta dose. Due fantocci sono stati usati per il train della rete 1.4 e uno per il test.

(a) Dose standard **(b)** Alta dose

Figura 1.4: Un esempio di input e output del dataset ad alta dose

Uno dei fantocci usato per il train possiede degli impianti dentali 1.5. I metalli producono delgli artefatti che la rete tenterà di rimuovere.



(a) Dose standard **(b)** Alta dose

Figura 1.5: In figura è mostrato una sezione del volume di un paziente in cui è presente del metallo

1.2.2 Dataset sottocampionato

Il dataset sottocampionato è formato da una serie di acquisizioni in cui si è ridotto il numero di proiezioni. In particolare sono stati provati 2 approcci diversi. Il primo è stato quello di acquisire un dataset con metà delle proiezioni, più precisamente sono stati considerati 240 angoli di proiezione su 480. Il secondo approccio è stato quello di prendere un quarto delle proiezioni, quindi acquisizioni con 120 angoli.

Diminuendo le proiezioni si trovano artefatti e si perdono informazioni dell'immagine. Gli artefatti si presentano come delle striature sull'immagine in cui manca informazione. Nella figura 1.6b è rappresentata l'immagine a 480 proiezioni, ovvero l'obiettivo della rete. Nella figura 1.6a è rappresentata un acquisizione a 240 proiezioni.

(a) 240 proiezioni **(b)** 480 proiezioni

Figura 1.6: Un esempio di input e output del dataset con metà delle proiezioni

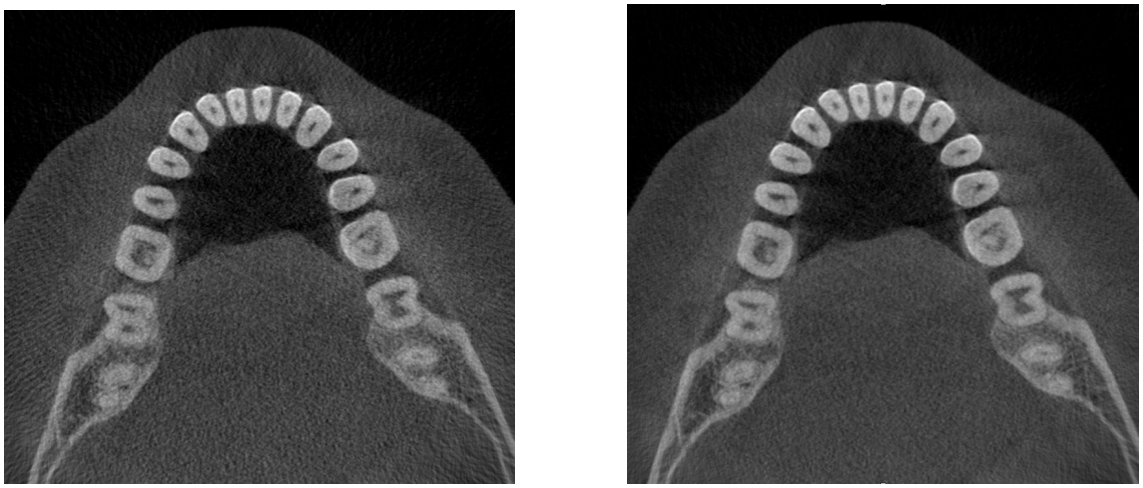
Acquisendo solo un quarto delle proiezioni gli artefatti sono molto più marcati del dataset con metà delle proiezioni. La rete dovrà ricostruire molta più informazione.

(a) 120 proiezioni **(b)** 480 proiezioni

Figura 1.7: Un esempio di input e output del dataset con un quarto delle proiezioni

1.2.3 Paziente reale

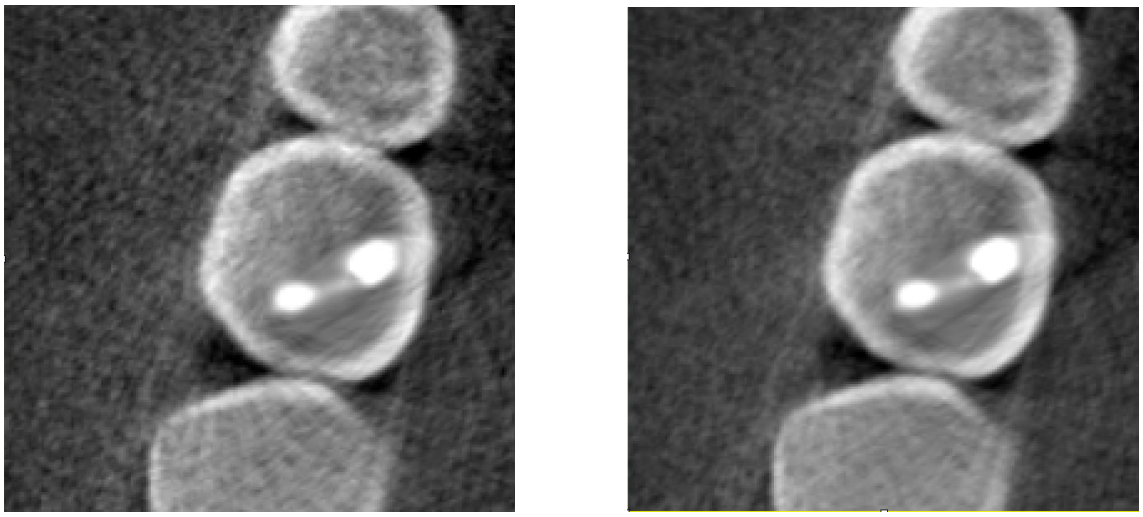
Come anticipato i pazienti reali presentano differenze sostanziali rispetto ai fantocci utilizzati per il train. Di conseguenza è stato utilizzato un paziente reale per testare sia la conversione ad alta dose sia la conversione diminuendo le proiezioni. Per queste conversioni le reti sono state allenate su tutti e 3 i fantocci e il paziente reale è stato tenuto come test. Nel caso della conversione ad alta dose ovviamento non si ha a disposizione una ground truth in quanto non è possibile eseguire una scansione con 100 volte la dose standard su un paziente reale. Invece Si ha a disposizione solo l'immagine ricostruita usando FDK con 480 proiezioni 1.9b e l'immagine con metà proiezioni 1.9a.



(a) 240 proiezioni **(b)** 480 proiezioni

Figura 1.8: Un esempio di input e output del paziente reale

Il paziente reale presenta anche degli impianti dentali per verificare se la rete riesce anche a ricostruire questa feature. Si mostra un esempio nella figura ingrandita 1.9.



(a) 240 proiezioni **(b)** 480 proiezioni

Figura 1.9: Immagini ingrandite su un impianto dentale nel paziente reale

Capitolo 2

Modelli

Nel processo di imaging TC, due metodi comuni per abbassare la dose sono ridurre il flusso di raggi X verso il rilevatore o la diminuzione del numero di proiezioni. Questi metodi riducono il tempo e la quantità di raggi X a cui si è esposti ma amplifiano anche il rumore e gli artefatti che potrebbero causare il deterioramento dell'immagine ricostruita. La ricostruzione dell'immagine TC consiste nel risolvere un sistema lineare di equazioni sottodeterminato. Uno dei metodi più usati per risolvere questo problema è l'utilizzo di algoritmi iterativi. I metodi iterativi per risolvere sistemi lineari sottodeterminati sono utilizzati quando si desidera trovare una soluzione approssimata o una soluzione che soddisfi determinati criteri di convergenza. Questi metodi non cercano di trovare una soluzione esatta, ma iterano su una sequenza di approssimazioni che si avvicinano gradualmente alla soluzione desiderata. Questi algoritmi hanno dei limiti e negli anni sono state sperimentati altri tipi di soluzioni, ad esempio l'uso del machine learning. Le tecniche di intelligenza artificiale/apprendimento automatico hanno ottenuto notevoli successi nell'ambito della visione artificiale, nell'analisi delle immagini e in molte altre aree. Una caratteristica principale dietro a questi successi è l'uso di reti neurali artificiali profonde addestrate con grandi quantità di dati. In particolare, lavorando con le immagini, vengono utilizzate le reti neurali convoluzionali che risultano adatte all'estrazione e l'analisi delle caratteristiche visive. In questo capitolo verranno introdotte le idee e i concetti chiave per capire il funzionamento di queste reti 2.1. Successivamente verranno presentati i modelli utilizzati per gli esperimenti di questa tesi 2.2.

2.1 Reti neurali convoluzionali

Le reti neurali convoluzionali, note anche come reti convolutive, sono diventate strumenti fondamentali nel campo del computer vision per diversi motivi. Sono

progettate appositamente per gestire dati strutturati come le immagini. Le immagini possono essere rappresentate come una griglia di pixel. Ogni cella della griglia contiene un valore a cui corrisponde quanto luminoso è quello specifico pixel. Le immagini del dataset utilizzato in questa tesi sono immagini in scala di grigio.

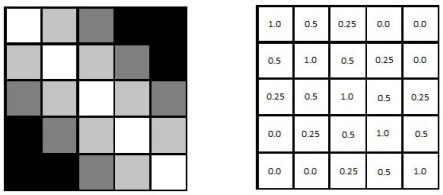


Figura 2.1: Esempio di rappresentazione di un immagine in una matrice

Le reti convoluzionali riescono a sfruttare la struttura spaziale delle immagini, considerando le relazioni tra i pixel e riducendo la complessità computazionale rispetto ad altri approcci. Riescono a farlo tramite diversi strati convolutivi. Ogni strato convolutivo applica un filtro (kernel) sull'immagine che ne estrae diverse feature. Un esempio si trova in figura 2.2. Questo filtro effettua una convoluzione sull'immagine. Questa operazione viene effettuata scorrendo il filtro sull'immagine e calcolando il prodotto scalare tra i valori del filtro e i valori dei pixel corrispondenti. Questo processo viene ripetuto per ogni posizione possibile dell'immagine di input. Il risultato è una nuova matrice chiamata feature map. Per fare in modo che questa matrice abbia le stesse dimensioni dell'input della convoluzione si aggiunge uno strato di padding, ovvero aggiungendo un bordo con valore 0.

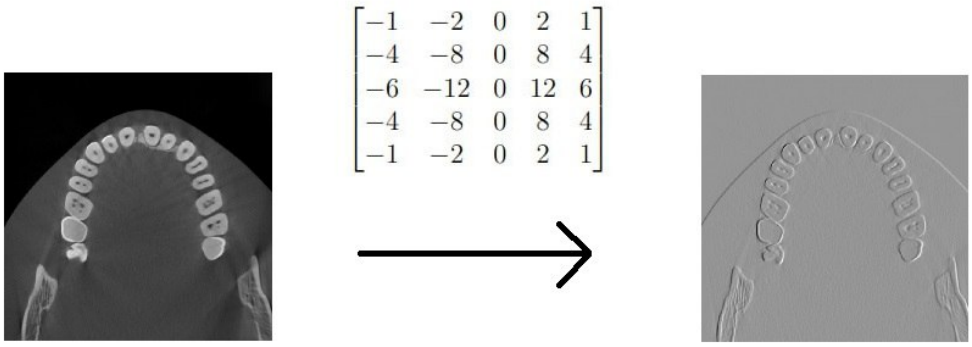


Figura 2.2: Filtro di sobel, kernel che estrae gli edge verticali dell'immagine

Lo scopo della rete è imparare i pesi di questi filtri convolutivi. Dopo ogni filtro convolutivo viene posta una funzione di attivazione per introdurre non linearità nell'output del neurone. Una delle funzioni di attivazione più utilizzata, e che verrà anche usata nei modelli di questa tesi, è la Rectified Linear Unit (ReLU). Ogni neurone ha una certo campo recettivo, cioè l'area dell'immagine da cui è influenzato. Inizialmente, nel caso delle convoluzioni, il campo recettivo Δ corrisponde alla grandezza del filtro K. Posizionando più livelli convoluzionali in successione si aumenta il campo recettivo, poiché il filtro di un certo livello è condizionato dai filtri utilizzati nei livelli precedenti (figura 2.3). Per calcolare quanto aumenta il receptive field applicando un filtro f con stride σ dopo il filtro g possiamo utilizzare la seguente formula:

$$rf = \sigma_f \times (\Delta_g - 1) + \Delta_f \tag{2.1}$$

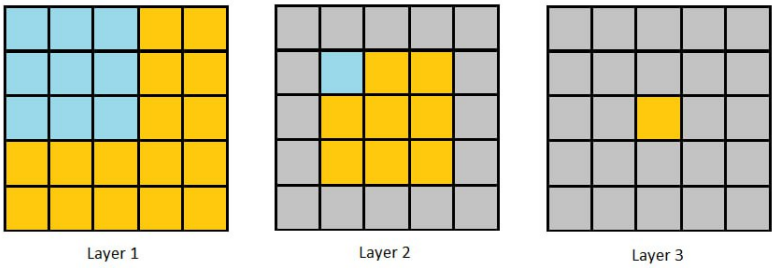


Figura 2.3: Receptive field con kernerl 3x3

Un altro modo per aumentare il campo recettivo che viene utilizzato nelle reti convoluzionali è introdurre uno strato di pooling. Il pooling consiste nel ridurre le dimensioni spaziali dell'input. Esistono vari metodi per effettuare questa operazione. In questa tesi viene utilizzato il maxpooling. Nel maxpooling si utilizza un filtro di dimensione K che prende il massimo dei valori dell'immagine su cui viene applicato (figura 2.4). Prendendo una activation map di dimensioni $W \times W$ e applicando il kernel K con uno stride S la dimensione in output equivarrà a:

$$W' = \frac{W - K}{S} + 1 \tag{2.2}$$

Anche per il maxpooling il receptive field Δ è uguale a K.

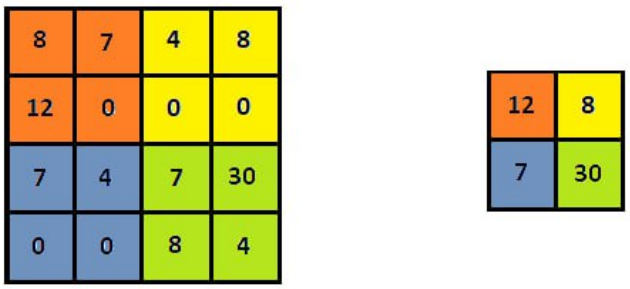


Figura 2.4: Esempio di operazione maxpooling con stride 2 e kernel 2x2

2.1.1 Rete encoder-decoder

Una rete encoder-decoder, o autoencoder, è un tipo di architettura di rete neurale artificiale che segue due percorsi: Un perscordo di compressione e uno di decompressione. L'encoder prende l' input e lo trasforma in una rappresentazione compressa. Questa rappresentazione compressa è chiamata spazio latente o codifica. Il ruolo dell'encoder è catturare le caratteristiche più importanti presenti nei dati in input mentre scarta dettagli non necessari. Il decoder invece prende la rappresentazione compressa prodotta dall'encoder e tenta di ricostruire i dati originali da essa. Gli autoencoder hanno molteplici applicazioni nel campo della computer vision tra cui la riduzione del rumore e la ricostruzione di immagini. Attraverso un addestramento supervisionato la rete può imparare a rimuovere il rumore dalle immagini. L'encoder impara a catturare le caratteristiche essenziali dell'input, mentre il decoder, partendo dalla codifica fatta dall'encoder cerca di produrre una versione dell'immagine ripulita dal rumore (Figura 2.5).

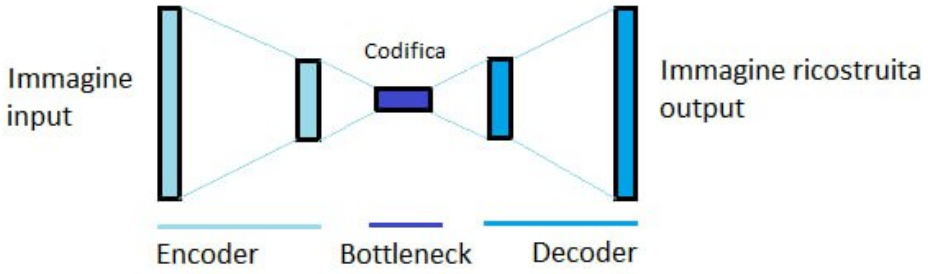


Figura 2.5: Struttura schematica di un autoencoder

Per la compressione dell'immagine si utilizzando delle operazioni di downsampling, come ad esempio l'operazione maxpool descritta in precedenza. Per la decompressione si utilizza invece una operazione di upsampling. Uno dei metodi per effettuare l'upsampling è ad esempio una convoluzione trasposta. In una convoluzione trasposta l'input viene ingrandito inserendo nella matrice dell'imagine, tra un pixel e un altro, dei nuovi valori inizializzati a 0 e successivamente effettuando una convoluzione con un kernel di dimensione K. Scelto un kernel, uno stride S e un padding P, si calcolano quanti zeri inserire con la seguente formula

$$z = S - 1 \tag{2.3}$$

Inoltre si calcola un nuovo padding P' da aggiungere

$$P' = K - P - 1 \tag{2.4}$$

Infine si applica la convoluzione sempre con stride unitario. Per determinare quanto viene ingrandito un input di dimensione $W \times W$ si può utilizzare la formula

$$W' = (W - 1) \times S + K - 2P \tag{2.5}$$

Infine definiamo il campo recettivo della convoluzione trasposta $\Delta = K$

2.2 Reti studiate

In questa sezione saranno illustrate le reti che sono state soggetto di analisi per questa tesi. Come modello di partenza è stata scelta la Unet2.2.1. Questo perchè

in letteratura è noto il suo utilizzo in compiti come denoising ed è già stata impiegata per ricostruzione di immagini tomografiche[6] [3]. Vengono poi proposte due variazioni al modello classico della Unet che utilizzano delle connessioni residuali 2.2.2. Infine viene proposta una rete sempre basata sulla struttura della Unet ma che lavora sul 3d. 2.2.3

2.2.1 Unet

La U-Net è una rete neurale convoluzionale sviluppata presso il Dipartimento di Informatica dell'Università di Freiburg per la segmentazione di immagini biomediche.[8]

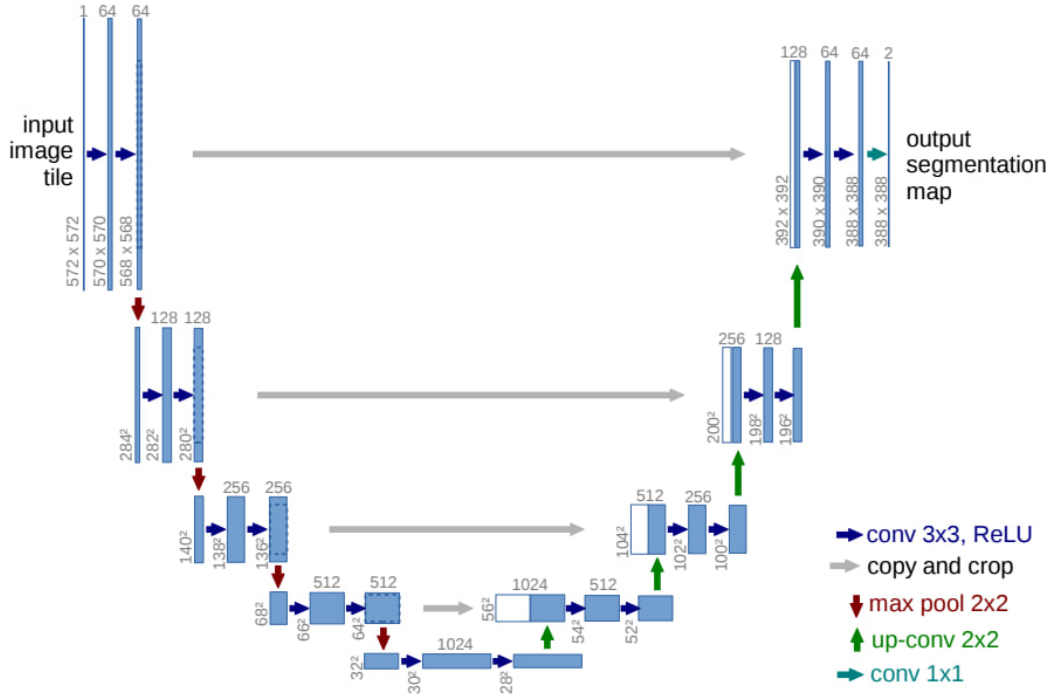


Figura 2.6: Struttura originale della Unet

L'architettura consiste in un percorso di contrazione per catturare il contesto e un percorso simmetrico di espansione per garantire una precisa localizzazione. Prima di ogni operazione di contrazione / espansione si applicano 2 convoluzioni con kernel 3x3 ognuna seguita dalla funzione di attivazione Relu. L'operazione di contrazione consiste in un 2x2 maxpooling con stride 2 e per ogni livello vengono raddoppiate le dimensioni delle features. Per il percorso di espansione viene applicata un operazione di upsampling. Prima di applicare la coppia di convoluzioni 3x3 nel percorso

di espansione viene concatenata la mappa di feature del livello di contrazione corrispondente. Queste concatenazioni prendono il nome di skip connection. Nella struttura originale riportata in figura le convoluzioni sono senza padding ed è quindi necessario fare un crop dell'immagine per rispettare le dimensioni.
Ciò che ha reso questa struttura molto efficace per la segmentazione semantica è la sua capacità di riuscire a catturare il contesto e allo stesso tempo essere molto precisa nella sua localizzazione. Queste caratteristiche sono state sfruttate anche per la risoluzione di altri problemi come la risoluzione di immagini e il deblurring. L'architettura consente di acquisire caratteristiche sia di basso che di alto livello dell'immagine di input, il che può essere utile per le attività di deblurring.
In questa tesi verrà analizzato come performa una semplice Unet per la ricostruzione delle immagini tomografiche dentali. La rete originale è stata riadattata togliendo l'operazione di crop e togliendo un livello di profondità per questioni di risorse computazionali.

2.2.2 Residual Unet

Una residual Unet è una Unet in cui vengono aggiunte delle connessioni residuali. Una connessione residua permette ad alcuni dati di raggiungere le ultime parti della rete neurale saltando alcuni livelli. Un esempio di blocco residuale è il seguente:

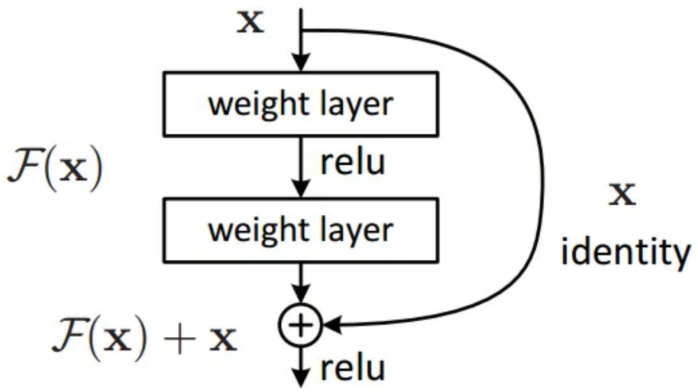


Figura 2.7: Blocco residuale [4]

Le connessioni residuali consentono alla rete di imparare facilmente a identificare le caratteristiche dell'immagine originale che dovrebbero essere conservate dopo il denoise. In pratica, una connessione residuale consente alla rete di imparare a differenziare tra ciò che deve essere modificato (gli artefatti) e ciò che dovrebbe rimanere invariato (i dettagli dell'immagine). La complessità del problema viene ridotta in

quanto la rete deve imparare solo le differenze rispetto all'input originale anziché ricostruire completamente l'immagine. Esistono diversi modi per implementare delle connessioni residuali utilizzando una Unet. In questa tesi verranno utilizzati 2 modelli differenti che utilizzano queste connessioni. Il primo modello è una semplice Unet in cui l'input viene sommato all'output della rete come mostrato in figura.2.8. Ci si riferirà questa rete con il nome di ResUnet.

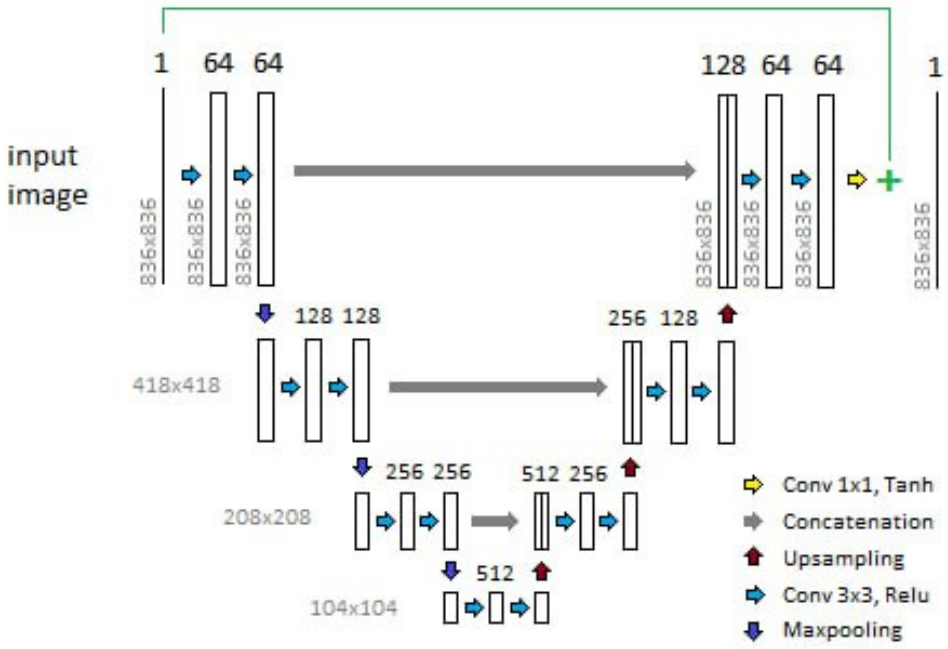


Figura 2.8: Struttura della ResUnet

Il secondo modello che implementa le connessioni residuali è una ResUnet in cui viene utilizzato un blocco convolutivo per ogni livello di profondità sia nella parte di compressione che di decompressione della rete, come mostrato in figura 2.9. Per riferirisi a questa rete verrà usato il nome FullResUnet.

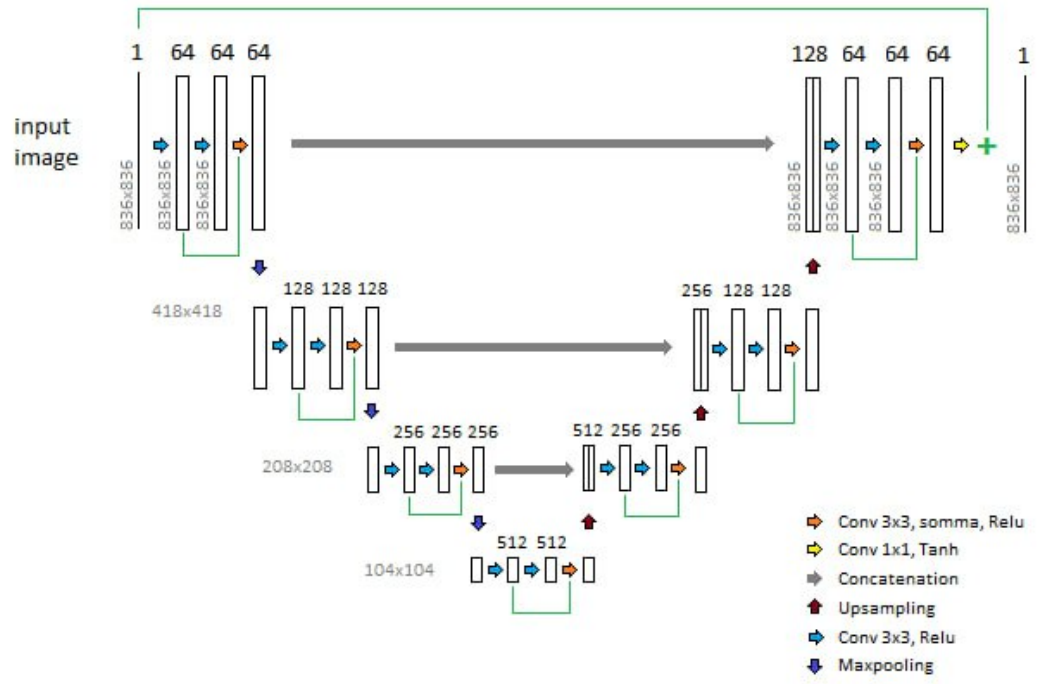


Figura 2.9: Struttura della FullResUnet

2.2.3 Residual Unet 3d

Le reti descritte nelle precedenti sezioni lavorano solo con immagini bidimensionali. Un elemento importante nella ricostruzione dell'intero volume è mantenere la coerenza delle sezioni bidimensionali una volta che vengono messe assieme. Una rete che lavora solo sul 2d non è allenata a mantenere questa coerenza. In figura 2.10 è mostrata un architettura di rete 3d proposta. Questa rete prende un blocco di 4 immagini alla volta e quindi viene anche allenata sulla profondità del volume. Per effettuare la ricostruzione di un intero volume la rete procede a ricostruire un blocco alla volta con stride 1. Successivamente viene fatta una media delle sezioni che si vanno a sovrapporre per ottenere un volume con la stessa profondità dell'originale.

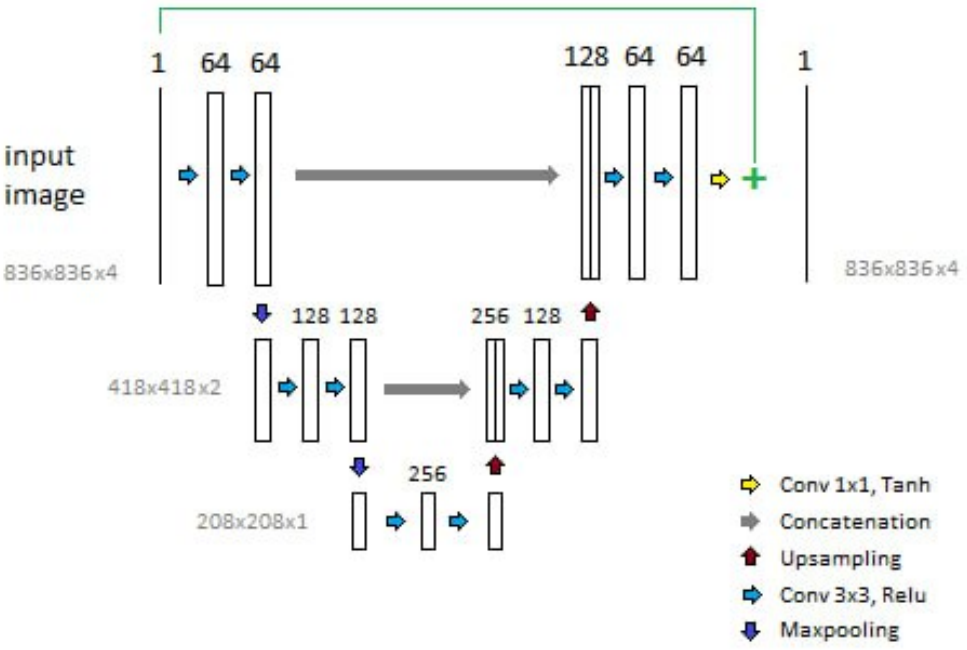


Figura 2.10: Struttura della ResNet3d

Capitolo 3

Risultati

In questo capitolo verranno esposti i risultati dei modelli presentati nel capitolo precedente. Gli iperparametri che descrivono gli esperimenti sono riportati nella sezione 3.1. Gli esperimenti che verranno mostrati sono i seguenti:
per ogni dataset sono state allenate le reti discusse nel capitolo 1, su 20 epoche e utilizzando MSE come loss. Inizialmente si parlerà del fenomeno dell'overfitting discutendo se i modelli presi in esame presentino questo problema 3.2. Successivamente si confronteranno le prestazioni dei modelli 2d sui dataset con un quarto e metà delle proiezioni e sul dataset ad alta dose. Verranno confrontati, dividendo i risultati per dataset, i valori della loss di train e di test nella sezione confronto numerico sui modelli 3.3. Successivamente, sempre dividendo per dataset, nella sezione 3.4 verranno confrontate le immagini ricostruite dalle diverse reti. In questa sezione verranno anche mostrate le differenze nella ricostruzione apportate dalla rete 3d. Infine verranno mostrati i risultati delle ricostruzioni della rete 3d su dei pazienti reali. 3.5

3.1 Iperparametri

Gli iperparametri nel deep learning sono parametri esterni al modello, che ne influenzano il processo di apprendimento, ma non vengono appresi direttamente durante il processo di addestramento.

Loss function

Per tutti gli esperimenti si è utlizzata la loss MSE, che è una misura della media dei quadrati delle differenze tra i valori predetti dal modello e i valori reali degli esempi nel set di dati. Sia I l'immagine originale e K l'immagine ricostruita, entrambe di

dimensione $M \times N$, si definisce l'MSE come

$$MSE = \frac{1}{M \times N} \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} (I(i,j) - K(i,j))^2 \tag{3.1}$$

- $I(i,j)$ è il valore del pixel i,j dell'immagine I
- $K(i,j)$ è il valore del pixel i,j dell'immagine K

Epoche e batch size

Le epoche rappresentano il numero di volte che un modello passa attraverso l'intero set di dati di addestramento durante il processo di apprendimento. Negli esperimenti verranno usate 20 epoche poichè si sono rivelate sufficienti per raggiungere dei buoni risultati. La batch size invece rappresenta il numero di campioni che verranno propagati nella rete dopo i quali verranno aggiornati i pesi. Come batch size sono state scelte 4 immagini. Le immagini sono molto grandi e, per limitazione delle risorse di calcolo, non si è potuto incrementare questo iperparametro. Nella rete 3d, che prende già un blocco di 4 immagini, la batch size è 1.

Learning rate e funzione di ottimizzazione

Il learning rate o tasso di apprendimento è la dimensione del passo che ad ogni iterazione porta verso il minimo di una funzione di perdita. Come learning rate è stato scelto $1 - e04$. La funzione di ottimizzazione utilizzata è stata Adam, abbreviazione di Adaptive Moment Estimation. È un algoritmo di ottimizzazione per aggiornare i pesi della rete in modo iterativo. Adam è considerato un'estensione della discesa del gradiente stocastico ed è noto per la sua efficacia nella gestione dei gradienti sparsi.

3.2 Overfitting

L'overfitting è un problema comune nell'apprendimento automatico. Si verifica quando un modello è troppo complesso rispetto alla complessità dei dati su cui è addestrato. In altre parole, il modello si adatta così strettamente ai dati di addestramento che perde la capacità di generalizzare, producendo previsioni o risultati meno accurati quando viene applicato ai dati di test. Un modo per monitorare la presenza di overfitting è quello di utilizzare una percentuale del dataset (solitamente tra il 10% e il 20%) come dataset di validazione. Alla fine di ogni epoca si calcola la loss sul dataset di validazione, senza aggiornare i parametri della rete. Successivamente si confronta l'andamento della loss sui dati di train con l'andamento della loss sul dataset di validazione. Se la loss sul dataset di validazione tende ad aumentare

rispetto alla loss sui dati di train allora si è in presenza di overfitting.

Sono stati condotti diversi test per verficare se i modelli utilizzati presentassero il fenomeno dell'overfitting. Questi sono alcuni risultati ottenuti in base al modello e al dataset. Il 10% del dataset è stato usato per crare il validation set. La loss utilizzata è il Mean Squared Error riportato come MSE.

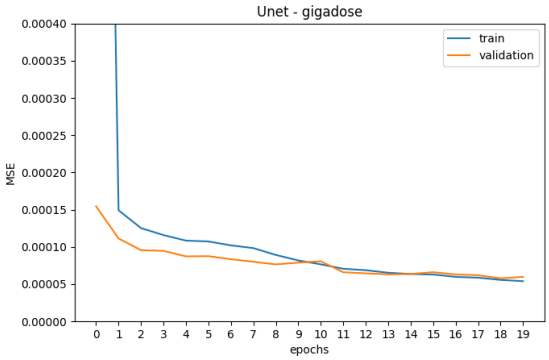


Figura 3.1: Andamento della loss per la rete Unet sul dataset ad alta dose

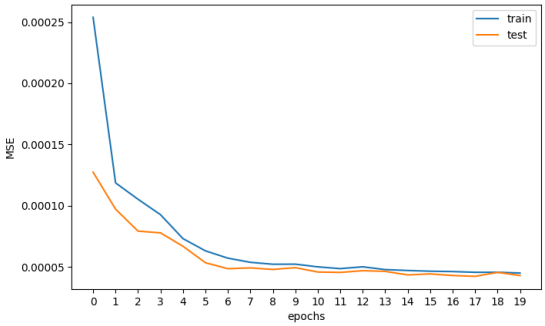


Figura 3.2: Andamento della loss per la rete ResUnet sul dataset ad alta dose

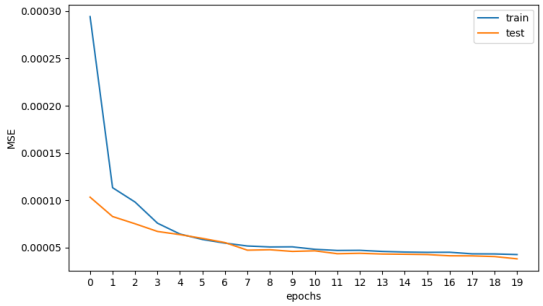


Figura 3.3: Andamento della loss per la rete FullResUnet sul dataset ad alta dose

Si può osservare come in nessun caso si presenti overfitting. Questo perchè i dati sono immagini di grandi dimensioni e le reti non sono abbastanza complesse da generare overfitting in così poche epoche. Si trovano risultati analoghi sui dataset sottocampionati.

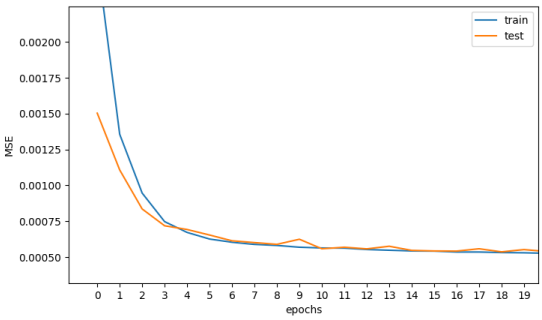


Figura 3.4: Andamento della loss per la rete ResUnet sul dataset con metà delle proiezioni

Infine si mostra un risultato ottenuto su 70 epoche. In questo grafico 3.5 si può osservare un leggero overfitting dopo 70 epoche. Dopo 20 epoche la rete performa già molto bene, il fenomeno dell'overfitting viene quindi evitato.

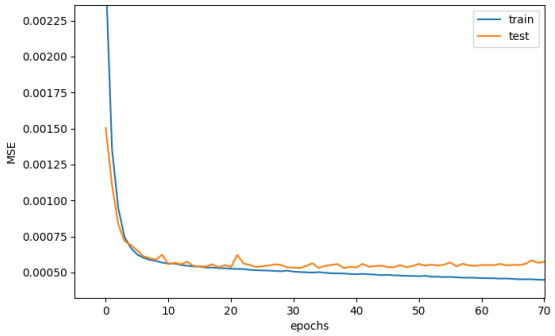


Figura 3.5: andamento della loss per la rete ResUnet sul dataset con metà delle proiezioni e 70 epoche

3.3 Confronto numerico sui modelli

In questo paragrafo vengono mostrate e confrontate le performance delle diverse reti prese in esame, analizzando esclusivamente il valore della loss. Sono stati effettuati e ripetuti diversi test e verranno riportati solo alcuni dei risultati più significativi.

3.3.1 Dataset ad alta dose

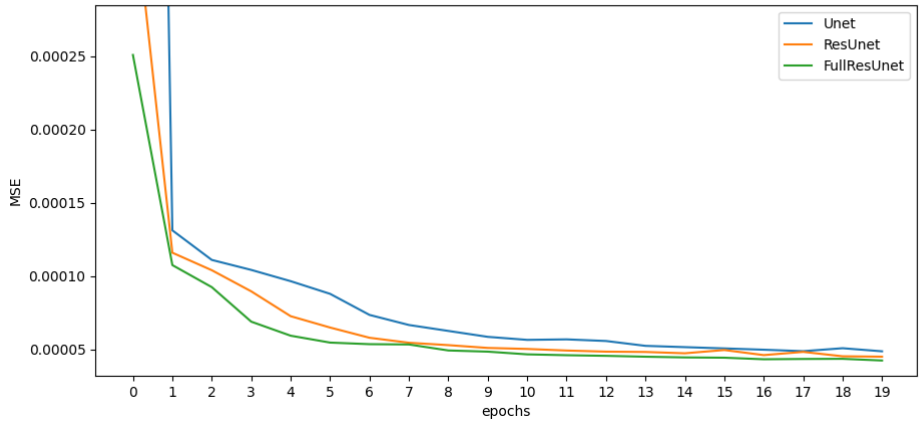


Figura 3.6: Andamento della loss del train delle diverse reti sul dataset Gigadose

Epoca	17	18	19
Unet	4.8843e-05	5.0847e-05	4.8744e-05
ResUnet	4.8347e-05	4.5340e-05	4.5081e-05
FullResUnet	4.3512e-05	4.3613e-05	4.2467e-05

Tabella 3.1: Confronto numerico dei valori dell'MSE per le ultime 3 epoche

Dai precedenti risultati si possono trarre le seguenti valutazioni. La Unet semplice è il modello che performa peggio. Come è stato osservato nella descrizione del modello questa rete, non presentando la connessione residuale che somma l'input all'output, cerca di ricostruire l'intera immagine, problema che risulta più complesso. Le altre 2 reti cercano invece di rimuovere il rumore preservando la struttura dell'immagine. La rete che performa meglio è la FullResUnet.

3.3.2 Dataset con metà delle proiezioni

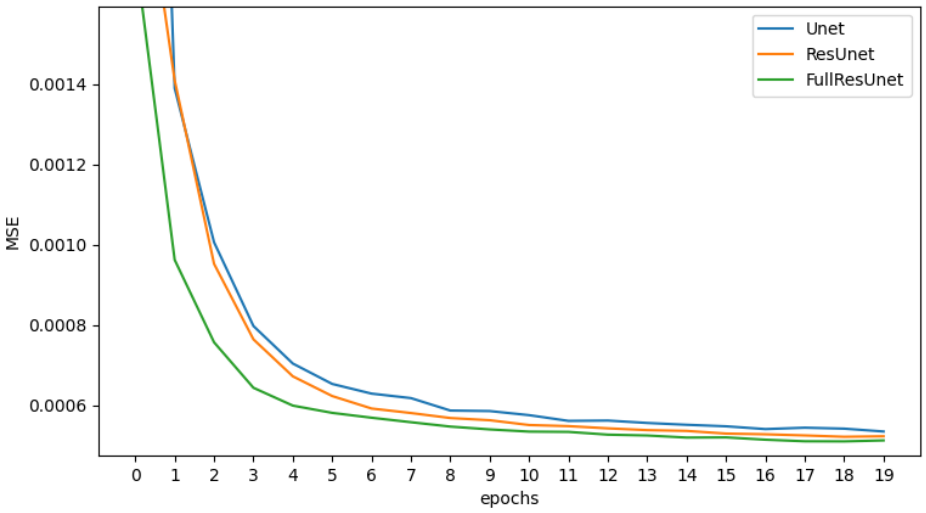


Figura 3.7: andamento della loss del train delle diverse reti sul dataset sottocampionato con metà delle proiezioni

Epoca	17	18	19
Unet	5.4468e-04	5.4228e-04	5.3516e-04
ResUnet	5.2539e-04	5.2207e-04	5.2331e-04
FullResUnet	5.1082e-04	5.1056e-04	5.1296e-04

Tabella 3.2: Confronto numerico dei valori dell'MSE per le ultime 3 epoche

Si può notare nuovamente come la Unet sia la rete che performa peggio seguita dalla ResUnet e dalla FullResUnet. Infine sono riportati i valori dell' MSE ottenuti sul dataset di test.

	MSE
Unet	6.021-04
ResUnet	5.592e-04
FullResUnet	5.501e-04

Tabella 3.3: Confronto numerico dei valori dell'MSE sul dataset di test

3.3.3 Dataset con un quarto delle proiezioni

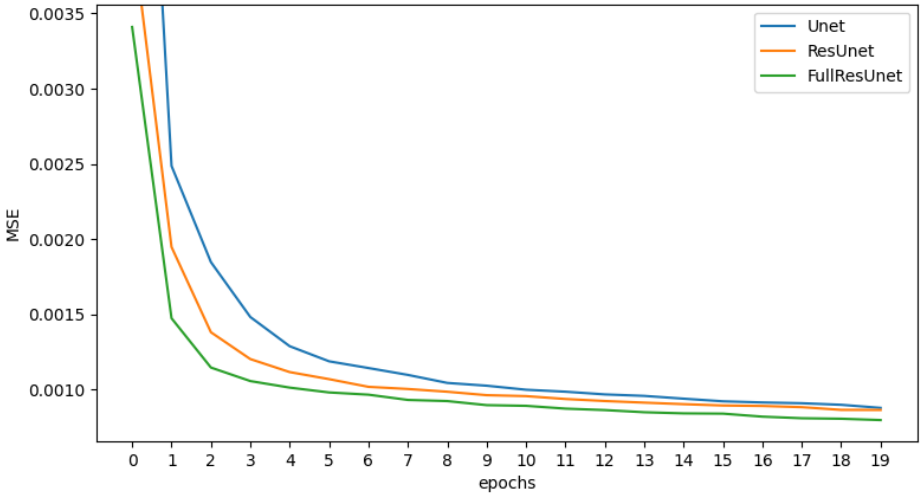


Figura 3.8: Andamento della loss del train delle diverse reti sul dataset sottocampionato con un quarto delle proiezioni

Epoca	17	18	19
Unet	9.0854e-04	8.9804e-04	8.7757e-04
ResUnet	8.8236e-04	8.6435e-04	8.638e-04
FullResUnet	8.087e-04	8.0535e-04	7.965e-04

Tabella 3.4: Confronto numerico dei valori dell'MSE per le ultime 3 epoche

Da questi risultati si nota nuovamente come la performance delle 3 reti sia sempre nello stesso ordine. La FullResUnet sembra discostarsi di più dagli altri due modelli.

	MSE
Unet	9.821-04
ResUnet	9.092e-04
FullResUnet	8.551e-04

Tabella 3.5: Confronto numerico dei valori dell'MSE sul dataset di test

3.4 Confronto sulle immagini ricostruite

In questa sezione verranno mostrate e confrontate alcune delle ricostruzioni fatte sul paziente di test, seguendo l'ordine dei dataset utilizzati. Per ciascun dataset, ogni rete ha generato un volume completo del paziente di test. Successivamente, questi volumi sono stati confrontati tra di loro e con la ground truth, sia manualmente sia utilizzando l'indice di similarità strutturale (SSIM). Il SSIM è una metrica di qualità dell'immagine utilizzata per valutare quanto due immagini digitali siano simili in termini di struttura. Per ciascun volume prodotto, è stato calcolato l'indice di similarità strutturale (SSIM) rispetto alla ground truth. Un valore SSIM più vicino a 1 indica una maggiore similarità strutturale tra le due immagini, mentre un valore più vicino a 0 indica una maggiore discrepanza strutturale.

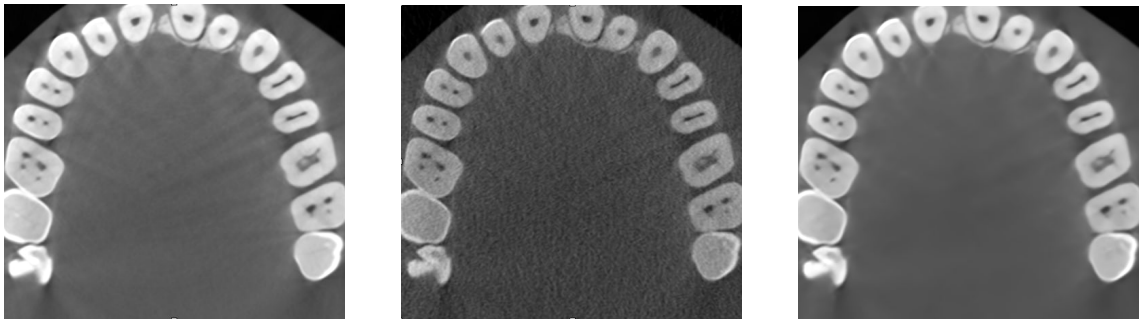
3.4.1 Dataset ad alta dose

La qualità della ricostruzione è molto alta per tutte le reti. Come si può notare dalla tabella 3.6 la FullResUnet sui dati di test raggiunge il valore medio dell'SSIM più alto. La differenza rispetto alle ricostruzioni delle altre reti non è una differenza significativa.

Unet	0.954
ResUnet	0.959
FullResUnet	0.967

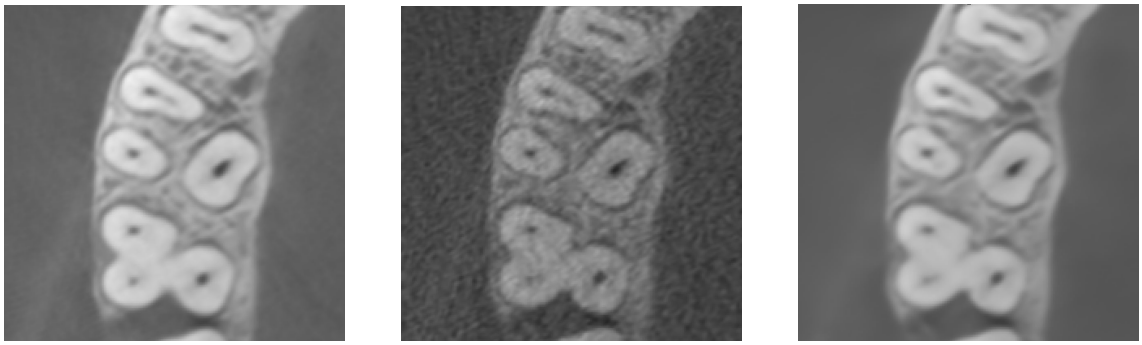
Tabella 3.6: Valore medio SSIM sul dataset di test

In figura 3.9 viene presentata la ricostruzione della rete FullResUnet su una arcata inferiore. Si può notare come la rete, oltre ad aver tolto il rumore presente nella immagine a dose standard, elimini alcuni artefatti presenti anche nella ground truth. Questo miglioramento non porta con sé problematiche sui punti dell'immagine con più dettagli. Infatti si può notare dalla figura 3.10 che anche ingrandendo un punto con molti particolari sui denti, la rete non produce distorsioni o modifiche dell'immagine che potrebbero portare a una diagnosi scorretta.



(a) 100 volte la dose **(b)** Dose standard **(c)** Ricostruzione FullResUnet

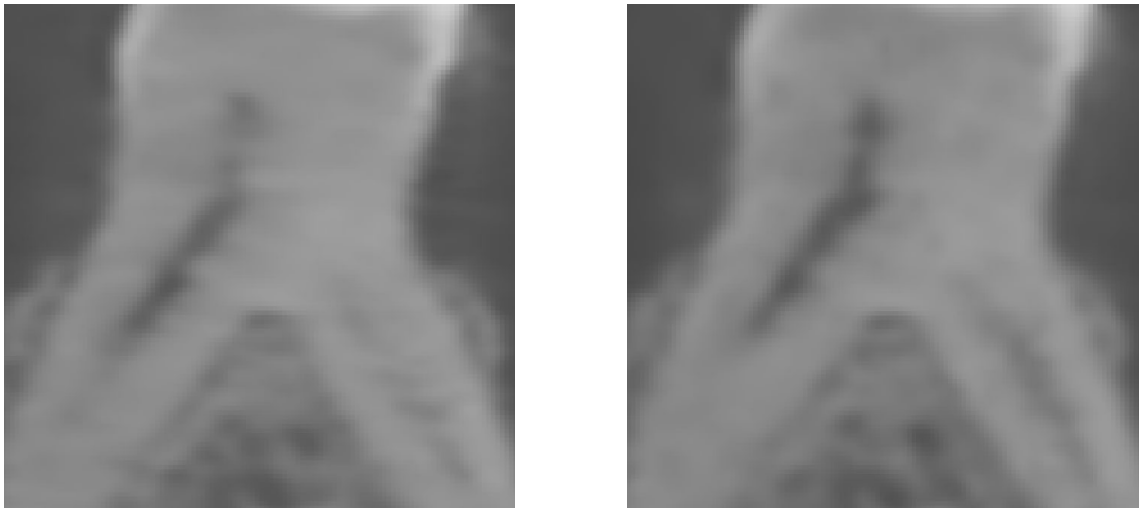
Figura 3.9: A sinistra la ground truth, al centro l'immagine con metà proiezioni e a destra la ricostruzione della rete



(a) 100 volte la dose **(b)** Dose standard **(c)** Ricostruzione FullResUnet

Figura 3.10: Da sinistra ground truth, immagine dose standard e ricostruzione della rete

Una volta prodotto un volume completo del paziente di test si può effettuare un reslice e visualizzare le immagini su un altro asse. Uno degli obiettivi importanti che si vogliono raggiungere è quello di ottenere coerenza tra ogni slice. Le reti che lavorano sul 2d ricostruiscono un immagine alla volta e quindi faticano a ricostruire fedelmente l'immagine sull'asse z. La rete 3d analizzata in questa tesi riesce su questo dataset a produrre delle immagini che mantengono la coerenza anche sulla profondità. La figura 3.11 mostra un esempio di dettaglio che la rete 2d non riesce ricostruire in modo adeguato.

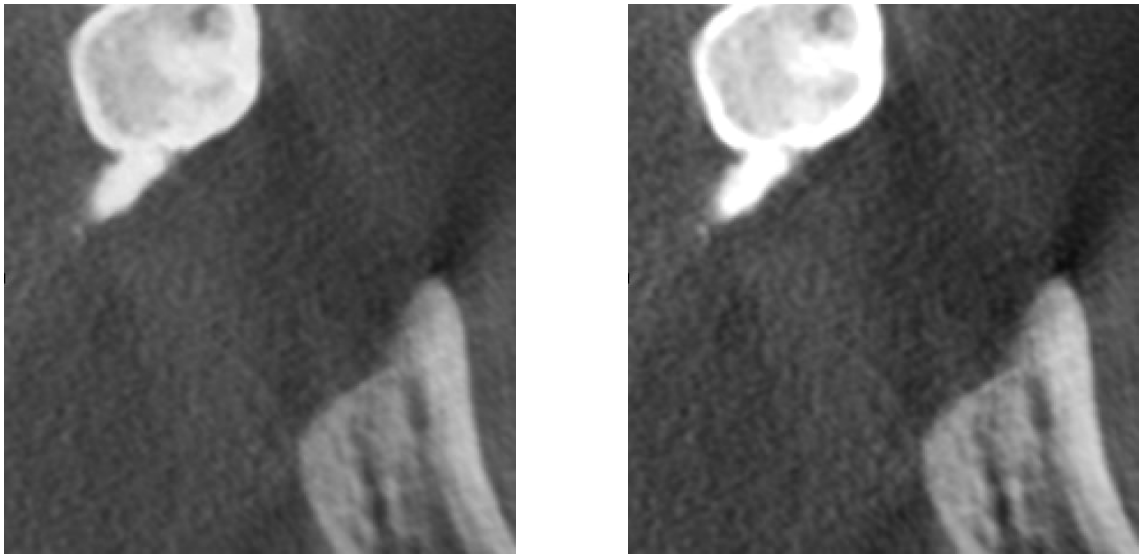


(a) ResUnet2d **(b)** ResUnet3d

Figura 3.11: Differenza tra una ricostruzione 2d e 3d. La ricostruzione 2d non riproduce un dettaglio importante all'interno del dente.

3.4.2 Dataset con metà delle proiezioni

La qualità della ricostruzione delle diverse reti è abbastanza simile. La ResUnet e la FullResUnet sono praticamente indistinguibili. Come mostrato in figura 3.12 la Unet sembra presenti un po' più di rumore. L'immagine è leggermente più sgranata e mantiene delle ombre più marcate rispetto alla ricostruzione dell ResUnet.



(a) ResUnet **(b)** Unet

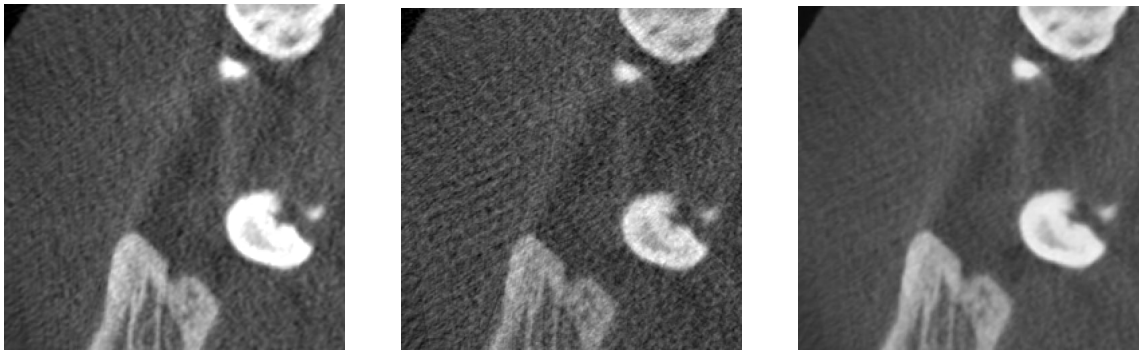
Figura 3.12: Differenza tra una ricostruzione della ResUnet e una ricostruzione della Unet

Di seguito sono riportati i risultati medi della Structural similarity index measure sul dataset di test.

Unet	0.752
ResUnet	0.798
FullResUnet	0.799

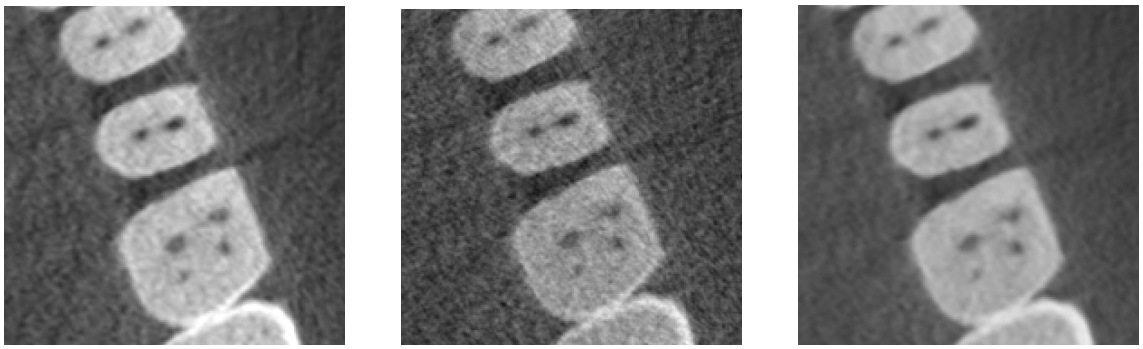
Tabella 3.7: Valore medio SSIM sul dataset di test

Confrontando invece la ricostruzione con l'output della ground truth si trovano i seguenti risultati. Nella figura 3.13 si può notare come la ricostruzione della ResUnet sia molto buona e fedele all'originale. Addirittura sembra rendere più liscia l'immagine rispetto alla ground truth. Anche sui denti 3.14 presenta questo vantaggio. Invece se si prende un punto con dei dettagli, come le radici dei denti, si può osservare come la rete non riesca a riprodurli fedelmente e tenda ad essere un po' troppo sfuocata. In figura 3.15 viene mostrato un esempio di punto in cui la rete tende a sfuocare negativamente l'immagine. In questo caso la radice di un dente viene poco definita e nella ricostruzione sembra fondersi con l'osso.



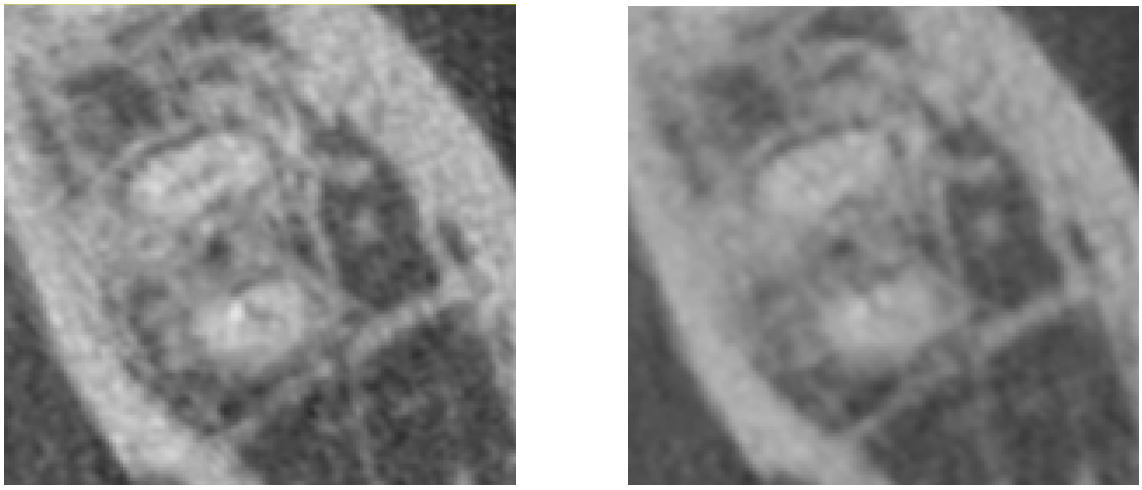
(a) 480 proiezioni **(b)** 240 proiezioni **(c)** Ricostruzione ResUnet

Figura 3.13: A sinistra la ground truth, al centro l'immagine con metà proiezioni e a destra la ricostruzione della rete



(a) 480 proiezioni **(b)** 240 proiezioni **(c)** Ricostruzione ResUnet

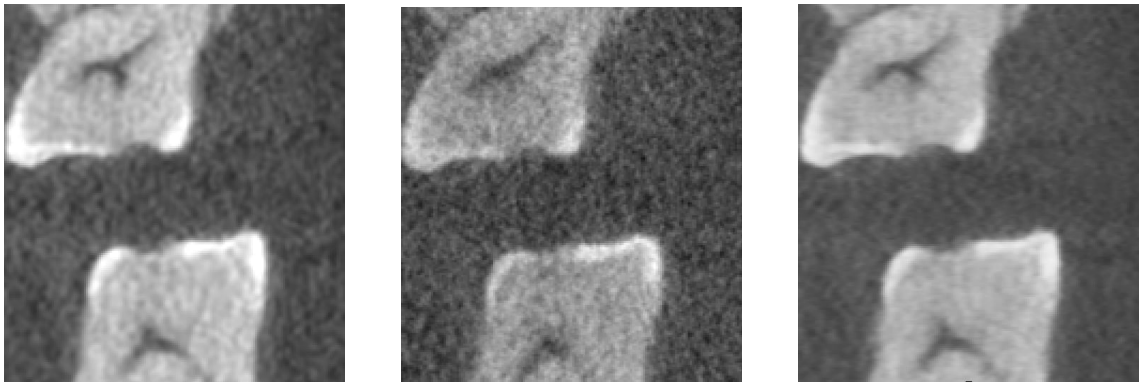
Figura 3.14: Da sinistra ground truth, acquizione con metà proiezioni e ricostruzione della rete sui denti



(a) 480 proiezioni **(b)** ResUnet

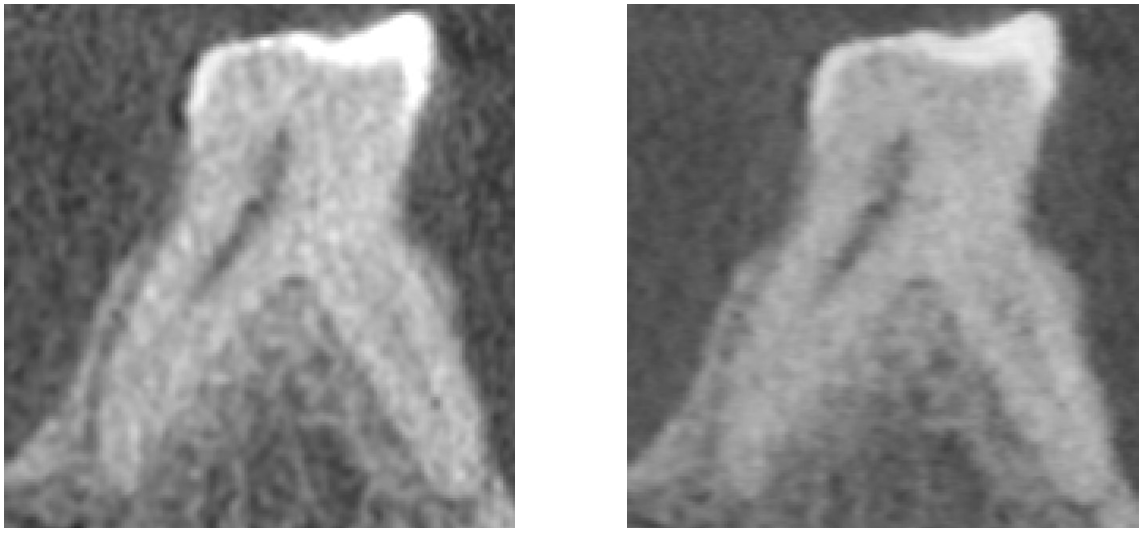
Figura 3.15: A sinistra la ground truth, a destra la ricostruzione della rete.

Dopo aver testato la rete su tutte le immagini del paziente di test si può ricostruire un intero volume. Successivamente si possono prendere le immagini su un asse diverso facendo un reslice del volume 3d. La rete, lavorando in 2d, non ha imparato nulla riguardo alla profondità dell'immagine. La ricostruzione risulta comunque buona 3.16 sebbene non perfetta. Infatti, come mostrato nell'immagine (3.17) si ha comunque il problema che la rete tende troppo a sfuocare i dettagli con l'aggiunta che fatica a ricreare la coerenza sull'asse z.



(a) 480 proiezioni **(b)** 240 proiezioni **(c)** Ricostruzione ResUnet

Figura 3.16: Da sinistra ground truth, acquisizione con metà proiezioni e ricostruzione della rete sui denti



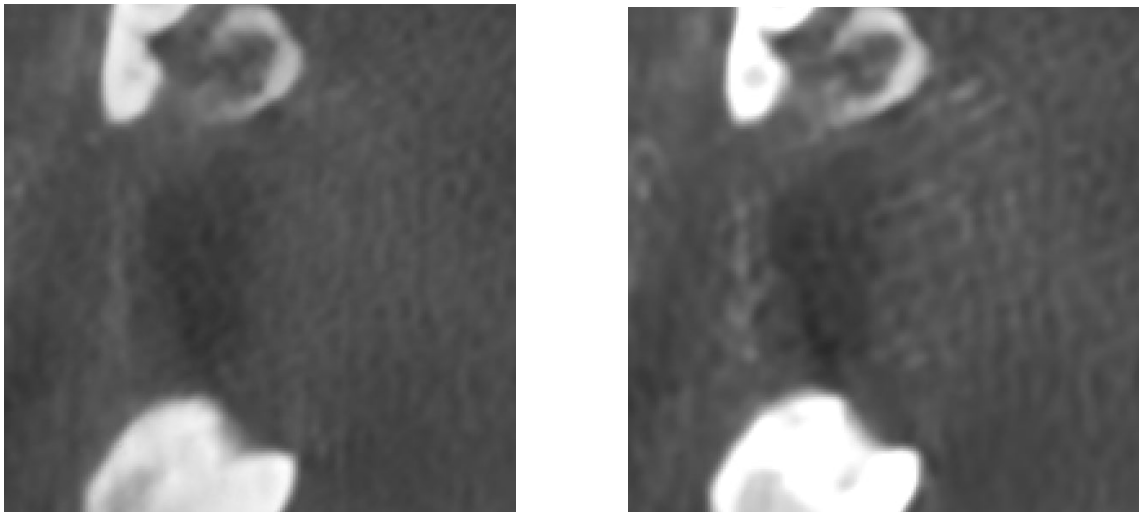
(a) 480 proiezioni **(b)** ResUnet

Figura 3.17: A sinistra la ground truth, a destra la ricostruzione della rete.

Purtroppo la rete 3d non riesce a fare meglio che la rete 2d. Questo perchè prende 4 slice di profondità e potrebbero non essere abbastanza per catturare il rumore del dataset con metà delle proiezioni.

3.4.3 Dataset con un quarto delle proiezioni

Il problema di ricostruire l'immagine partendo da solo un quarto delle proiezioni si è rivelato un problema molto difficile. Questo perchè gli artefatti sono molti e spesso manca troppa informazione nell'immagine sottocampionata per poter essere ricostruita fedelmente. Le reti hanno nuovamente un comportamento simile ma la classica Unet performa leggermente peggio nonostante i valori medi dello structural similarity index measure 3.8. La Unet, in alcuni punti in cui gli artefatti erano più marcati, non riesce ad eleminare del tutto il rumore lasciando delle striature come mostrato in figura 3.18



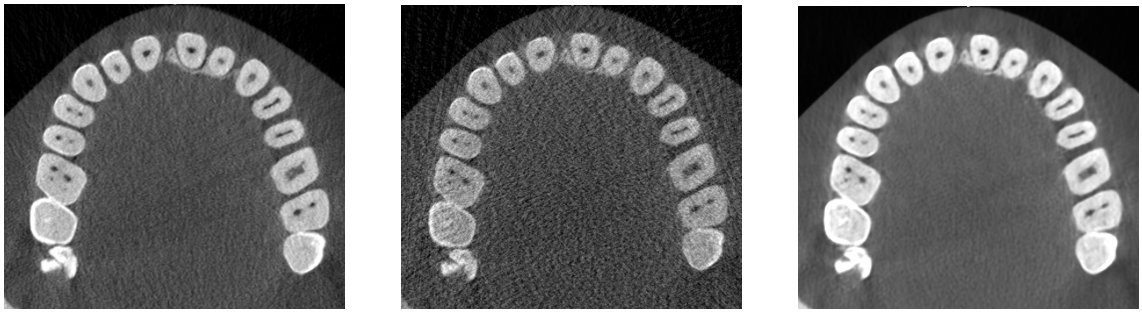
(a) FullResUnet **(b)** Unet

Figura 3.18: Differenza tra una ricostruzione della FullResUnet e una ricostruzione della Unet

Unet	0.654
ResUnet	0.662
FullResUnet	0.672

Tabella 3.8: Valore medio SSIM sul dataset di test. L'SSIM medio sui dati in input era di 0.244

La FullResUnet e la ResUnet ancora una volta sono praticamente indistinguibili. Il denoise è molto efficacie e si può vedere un esempio di come nuovamente l'immagine ricostruita dalla rete sembri meno rumorosa dell'originale 3.19.

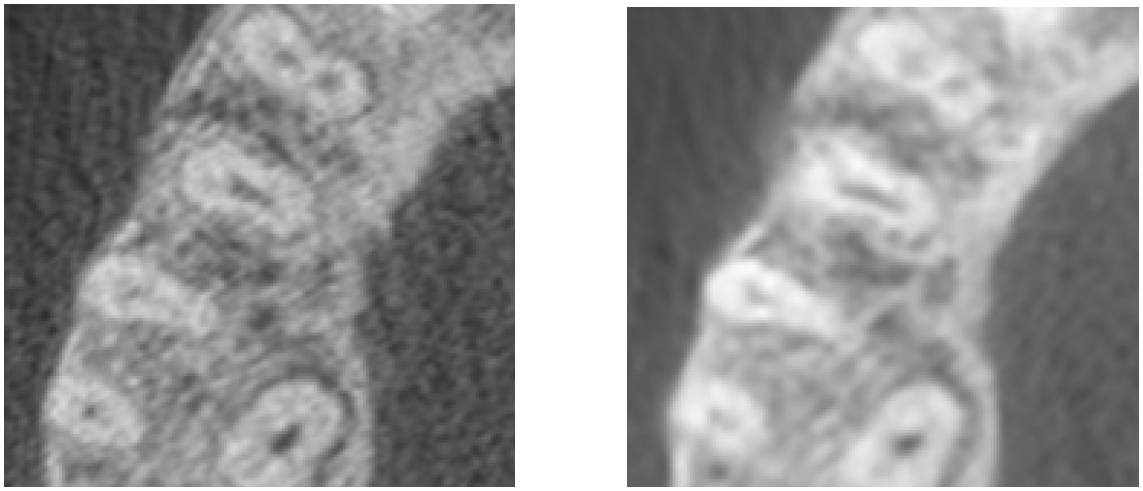


(a) 480 proiezioni **(b)** 120 proiezioni **(c)** Ricostruzione FullResUnet

Figura 3.19: Ground truth, un quarto delle proiezioni e ricostruzione della rete sui denti

Purtroppo bisogna guardare bene i dettagli e questa ricostruzione non riesce a riprodurre fedelmente certi punti. Questi errori rischiano di portare a delle diagnosi

scorrette. Si può notare ad esempio nella figura 3.20 come la sfuocatura dell'immagine della ricostruzione della rete modifichi i denti. Questi dettagli che la rete aggiunge rendono queste immagini inutilizzabili. La ricostruzione a un quarto delle dosi risulta troppo ottimista per essere realmente utilizzata.



(a) Ground truth **(b)** FullResUnet

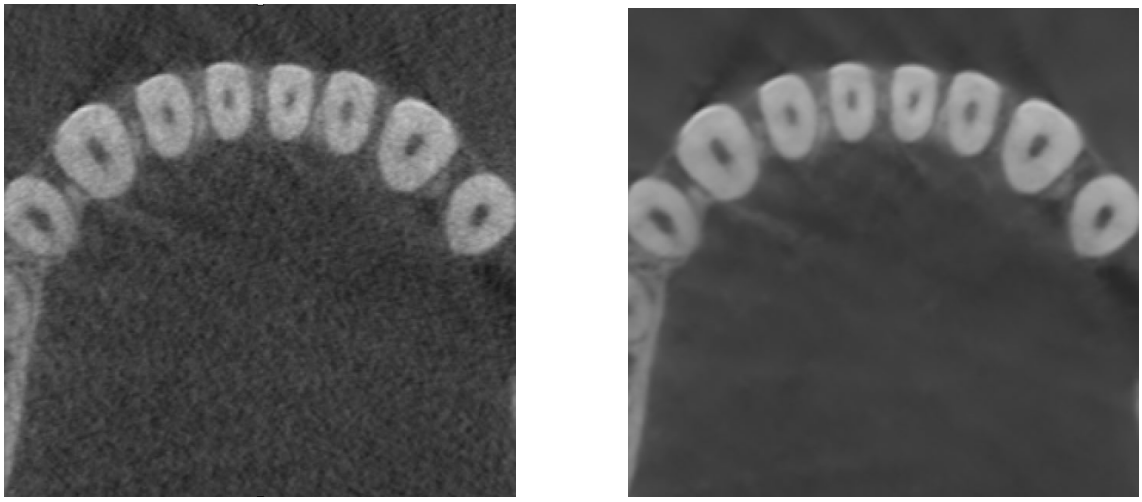
Figura 3.20: Differenza tra una ricostruzione della FullResUnet e una ricostruzione della Unet

3.5 Risultati sui pazienti reali

In questa sezione vengono mostrati i risultati della rete 3d, allenata solo sui fantocci, su un paziente reale. In particolare, sono stati esaminati i risultati della conversione con metà delle proiezioni e della conversione ad alta dose.

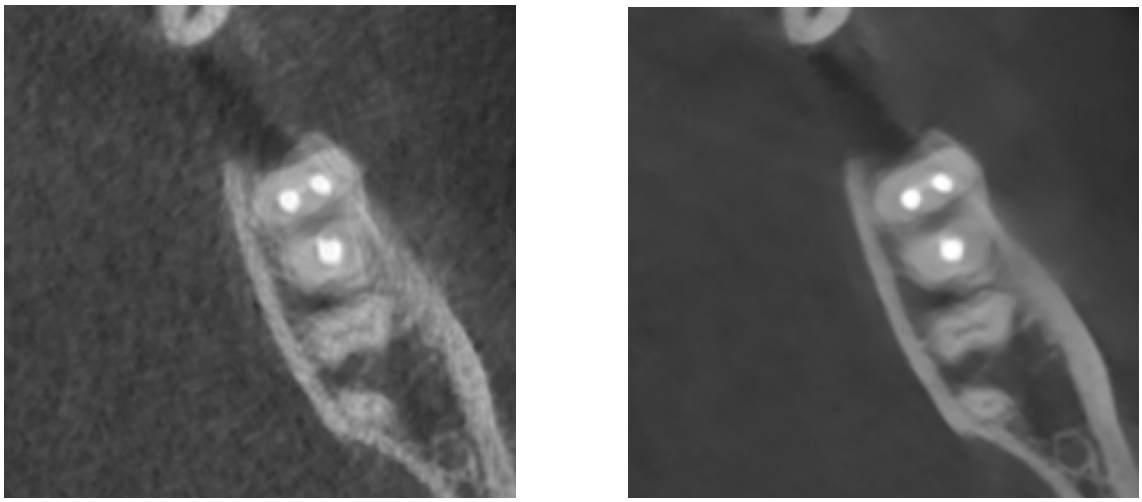
3.5.1 Conversione ad alta dose

La rete riesce a creare un effetto ad alta dose anche sul paziente reale. Nell'immagine 3.21 viene mostrata la ricostruzione dell'arcata inferiore. Nell'immagine 3.22, invece, viene mostrata una ricostruzione su un impianto dentale. Nel dataset di train, uno dei fantocci possedeva dei metalli e la rete è riuscita a riprodurre con successo questa feature.



(a) Paziente reale **(b)** Paziente reale con effetto ad alta dose

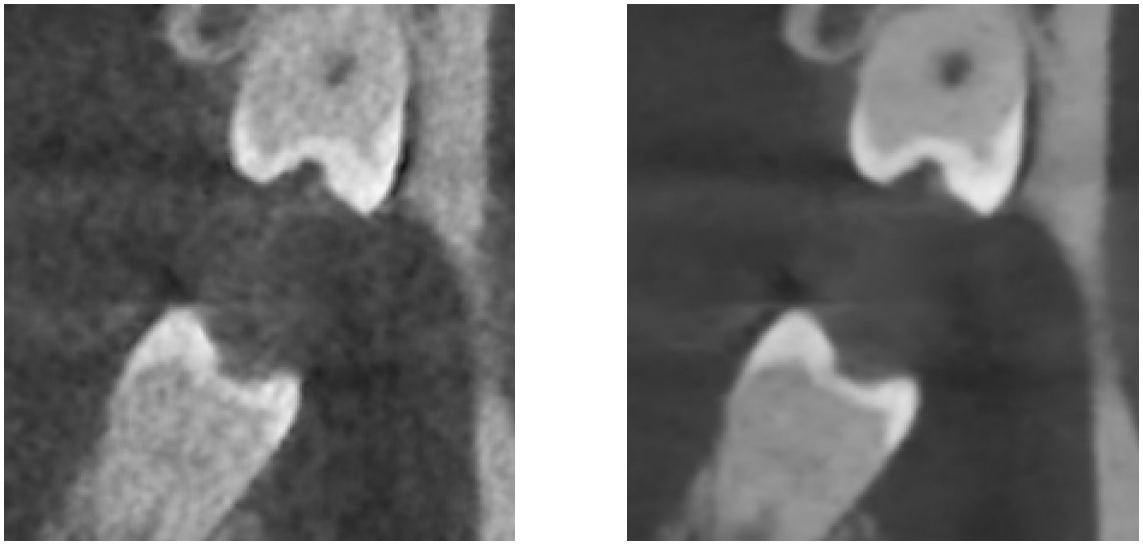
Figura 3.21: Effetto ad alta dose applicato a un paziente reale



(a) Paziente reale **(b)** Paziente reale con effetto ad alta dose

Figura 3.22: Effetto ad alta dose applicato a un impianto dentale del paziente reale

In generale la rete mostra degli ottimi risultati anche sull'asse Z, come mostrato dalla figura 3.23. Solamente in alcuni punti tende a sfuocare l'immagine.

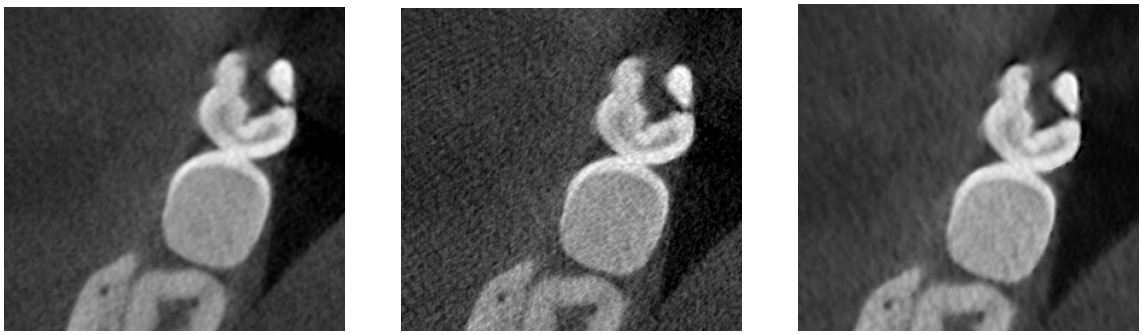


(a) Paziente reale **(b)** Paziente reale con effetto ad alta dose

Figura 3.23: Effetto ad alta dose su un paziente reale visualizzato sull'asse Z

3.5.2 Conversione con metà delle proiezioni

La conversione con metà delle proiezioni presenta diverse problematiche. In alcuni punti la rete si comporta molto bene, ad esempio in figura 3.24 si può vedere come la rete rimuova gli artefatti rimanendo fedele alla ground truth.

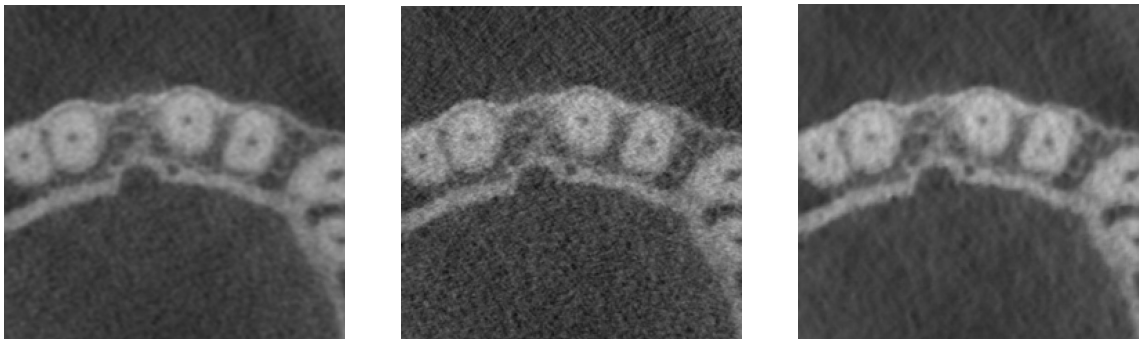


(a) 480 proiezioni **(b)** 240 proiezioni **(c)** ricostruzione rete 3d

Figura 3.24: ground truth, metà delle proiezioni e ricostruzione della rete

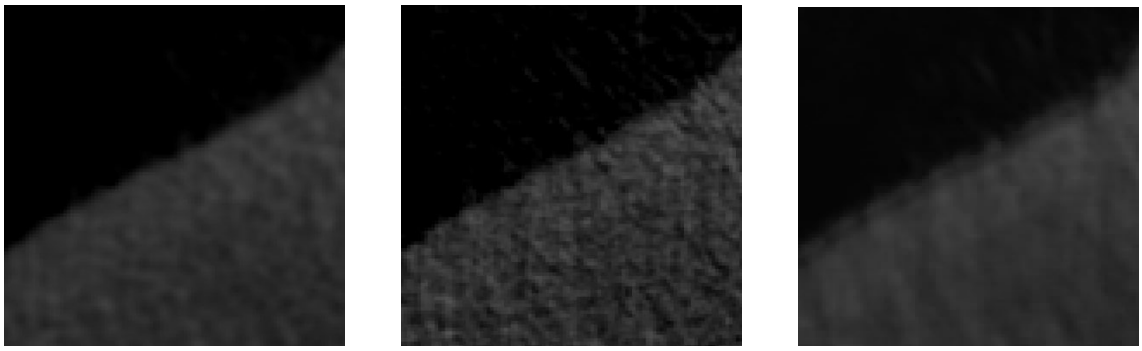
Come è stato mostrato con i risultati ottenuti solo sui fantocci, la rete tende a sfuocare alcuni edge che possono fondere parti che dovrebbero essere distinte tra loro. Viene mostrato un esempio in figura 3.25. Il problema più significativo però riguarda alcuni edge che vengono sfuocati a tal punto da creare nuovi artefatti 3.26.

Questo rende la conversione della rete molto problematica poichè può portare a diagnosi errate. Questi artefatti potrebbero essere dovuti al fatto che il paziente reale durante la tomografia non è immobile come invece lo sono i fantocci utilizzati per l'addestramento.



(a) 480 proiezioni **(b)** 240 proiezioni **(c)** ricostruzione rete 3d

Figura 3.25: ground truth, metà delle proiezioni e ricostruzione della rete sui denti



(a) 480 proiezioni **(b)** 240 proiezioni **(c)** ricostruzione rete 3d

Figura 3.26: ground truth, metà delle proiezioni e ricostruzione su un bordo

3.6 Conclusioni

La rete Unet non sempre produce risultati ottimali, tuttavia l'aggiunta di connessioni residue in tutte le conversioni migliora le prestazioni complessive. Tra le due varianti delle reti Unet residuali non ci sono sostanziali differenze. Nel dataset ad alta dose, tutte le reti mostrano risultati significativi, portando l'SSIM a un valore superiore a 0,95. Le reti che operano con input bidimensionali non manifestano una coerenza robusta sull'asse z, a differenza di quelle con input tridimensionali che migliorano notevolmente la transizione tra le sezioni. Nel dataset con metà delle

proiezioni invece la rete Unet ha dei risultati peggiori rispetto alle versioni che implementano le connessioni residuali. La conversione è molto buona sebbene alcuni dettagli risultano sfuocati e potrebbero essere problematici. Sull'asse z il problema peggiora e la rete tridimensionale non riesce a migliorare significativamente il problema in quanto prende solo 4 immagini di profondità. La ricostruzione con un quarto delle proiezioni invece risulta troppo sottocampionata per tutte le reti. Gli artefatti sembrano troppo marcati per garantire una ricostruzione sicura con così pochi volumi per il train. Infine sul paziente reale è molto notevole l'effetto ad alta dose. Sembra molto coerente con l'immagine originale e si potrebbe tentare anche una dose più bassa. La ricostruzione con metà proiezioni invece, sebbene in gran parte dell'immagine svolga un ottimo lavoro, presenta degli artefatti che andrebbero evitati per una corretta diagnosi. Servirebbero più dati per far imparare alle reti anche gli artefatti dovuti dal movimento del paziente. In conclusione, l'impiego delle reti neurali mostra un grande potenziale, poiché hanno dimostrato un'efficacia notevole anche con una quantità limitata di dati, applicando con successo la conversione imparata sui fantocci anche sui pazienti reali.

Riferimenti bibliografici

[1] See through. URL `https://www.seethrough.one/`.

[2] L. A. Feldkamp, L. C. Davis, and J. W. Kress. Practical cone-beam algorithm. *J. Opt. Soc. Am. A 1, 612-619*, 1984.

[3] Doga Gunduzalp, Batuhan Cengiz, Mehmet Ozan Unal, and Isa Yildirim. 3d u-netr: Low dose computed tomography reconstruction via deep learning and 3 dimensional convolutions. *arXiv:2105.14130 [cs.CV]*, 2021.

[4] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. *arXiv:1512.03385 [cs. CV]*, 2015.

[5] S. Hughes. Ct scanning in archaeology, computed tomography - special application. *InTech*, 2011.

[6] Satoru Mizusawa, Yuichi Sei, Ryohei Orihara, and Akihiko Ohsuga. Computed tomography image reconstruction using stacked u-net. *Computerized Medical Imaging and Graphics*, 90:101920, 2021. ISSN 0895-6111. doi: https://doi.org/10.1016/j.compmedimag.2021.101920. URL `https://www.sciencedirect.com/science/article/pii/S0895611121000690`.

[7] J. Radon. Uber die bestimmung von funktionen durch ihre integralwerte langs gewisser mannigfaltigkeiten. *Journal of Mathematical Physics, 69, 262-277*, 1917.

[8] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. *arXiv:1505.04597 [cs. CV]*, 2015.

[9] Ge wang, Yi Zhang, Xiaojing Ye, and Xuanqin Mou. *Machine Learning for Tomographic Imaging*. IOP Publishing, 2020.