



UNIVERSITÀ DEGLI STUDI DI PALERMO
DIPARTIMENTO DI INGEGNERIA

CORSO DI LAUREA MAGISTRALE IN INGEGNERIA INFORMATICA

Ricostruzione volumetrica della cavità orale da radiografia panoramica

TESI DI LAUREA DI
ANGELA CAFA'

RELATORE
Prof. ORAZIO GAMBINO

CONTRORELATORE
Prof. MARCO MORANA

CORRELATORE
Prof. SALVATORE CONTINO

ANNO ACCADEMICO 2024 – 2025

MAGISTRALE



Contents

1	Introduzione	3
2	Revisione della letteratura	6
2.1	CBCT e imaging dentale	6
2.1.1	Cone Beam Computed Tomography (CBCT)	7
2.1.2	Radiografia Panoramica	9
2.2	Tecniche di proiezione	11
2.3	Proiezione panoramica da CBCT	12
2.4	Ricostruzione 3D a partire da immagini panoramiche	13
2.4.1	Approcci basati su Deep Learning per la ricostruzione 3D da PX . .	13
2.4.2	Altri approcci 3D da PX	15
2.5	Modellazione della curva mandibolare	15
2.6	Ray tracing per proiezione medica	16
2.7	Adattamento di dominio tramite CycleGAN	17
2.8	Reti neurali per la ricostruzione 3D da immagini 2D	20
3	Metodologia	22
3.1	Overview della pipeline proposta	22
3.1.1	Addestramento e generazione	23
3.2	Simulazione dell'assorbimento dei raggi X nella generazione di immagini panoramiche sintetiche	24
3.2.1	Simulazione della proiezione panoramica tramite ray tracing	24
3.2.2	Generazione automatica di SimPX per il training	25
3.3	Architettura della rete cGAN	26
3.3.1	Schema dell'Architettura Generatore	27
3.3.2	Architettura del Discriminatore (PatchGAN)	28
3.3.3	Funzioni di perdita del cGAN	29
3.3.4	Strategie di Addestramento	32
3.3.5	Metriche di valutazione durante l'addestramento	33
3.3.6	Adattamento di dominio tramite CycleGAN	34
3.4	Ottimizzazione congiunta della curva e dei parametri di proiezione	36

4 Results	39
4.1 Introduzione	39
4.2 Ricostruzione CBCT da PX tramite cGAN	39
4.2.1 Esperimenti	39
4.3 Ottimizzazione con Trasformazioni sulla Curva e Ray Tracing	47
4.3.1 Metrica di costo	48
4.3.2 Algoritmo di ottimizzazione	48
5 Discussione	57
6 Conclusione	59

Chapter 1

Introduzione

Contesto

Negli ultimi decenni, l’evoluzione delle tecnologie di imaging ha profondamente trasformato la pratica clinica odontoiatrica. Tra le modalità diagnostiche più diffuse si collocano la radiografia panoramica (Panoramic X-ray) e la tomografia computerizzata a fascio conico (Cone Beam Computed Tomography, CBCT), due approcci complementari che differiscono per risoluzione spaziale, contenuto informativo, dose radiologica ed accessibilità.

La radiografia panoramica consente una visione complessiva dell’arcata dentale e delle strutture circostanti con una singola esposizione a bassa dose, rappresentando lo strumento di screening più comune nella pratica odontoiatrica. Tuttavia, in quanto immagine bidimensionale, la PX soffre di limitazioni legate alla perdita di informazioni tridimensionali, distorsioni geometriche e sovrapposizioni anatomiche.

Al contrario, la CBCT fornisce una ricostruzione tridimensionale dettagliata del distretto oro-maxillo-facciale, permettendo una valutazione accurata della morfologia ossea e dentale da qualsiasi angolazione. Nonostante la sua potenza diagnostica, l’impiego della CBCT comporta un’esposizione maggiore alle radiazioni ionizzanti e costi sensibilmente più elevati, fattori che ne limitano l’utilizzo a casi selezionati.

Questa compromesso tra qualità diagnostica e sicurezza del paziente ha stimolato la ricerca verso soluzioni ibride che combinino i vantaggi delle due tecniche. In particolare, risulta di grande interesse la possibilità di ricostruire un volume tridimensionale a partire da una singola immagine panoramica, riducendo così la necessità di eseguire una CBCT completa. Tale approccio, se realizzato in modo robusto e affidabile, potrebbe ampliare significativamente l’accesso a tecnologie 3D avanzate, specialmente in contesti a bassa disponibilità economica o in pazienti per cui l’esposizione radiologica deve essere minimizzata.

Motivazione

La possibilità di ottenere informazioni tridimensionali a partire da un’unica immagine panoramica rappresenta una sfida affascinante ma complessa nel campo dell’imaging odon-

toiatrico. La panoramica tradizionale, pur offrendo una visione complessiva dell'arcata dentale, è soggetta a distorsioni e non contiene informazioni dirette sulla profondità. Tuttavia, la geometria con cui viene acquisita suggerisce che parte della struttura tridimensionale è implicitamente proiettata lungo una traiettoria anatomica ben definita.

In questo contesto, la curva mandibolare assume un ruolo centrale: rappresenta il riferimento geometrico lungo cui si distribuiscono le principali strutture anatomiche della mandibola. Se stimata con precisione, questa traiettoria consente di ricostruire le informazioni tridimensionali implicite nella panoramica. In particolare, permette di generare sezioni planari orientate (note come MPR, Multiplanar Reformations), da cui è possibile ottenere un volume tridimensionale coerente con l'anatomia reale del paziente. Al contrario, una curva imprecisa compromette l'intera ricostruzione.

Per questo motivo, la presente tesi non si limita alla sola generazione volumetrica da immagini panoramiche, ma affronta anche il problema dell'ottimizzazione della curva mandibolare. Tale processo avviene confrontando una panoramica sintetizzata — ottenuta tramite proiezione dal volume ricostruito — con la panoramica reale del paziente. La traiettoria viene progressivamente affinata attraverso una funzione di errore che misura la coerenza tra le due immagini, utilizzando criteri di similarità strutturale.

In parallelo, vengono ottimizzati anche i parametri geometrici del sistema di proiezione, come la posizione e l'orientamento della sorgente, al fine di migliorare ulteriormente la qualità della ricostruzione.

Questo approccio permette non solo di migliorare la qualità della ricostruzione 3D, ma anche di adattarla in modo paziente-specifico, offrendo una potenziale alternativa non invasiva e a bassa dose alla CBCT convenzionale.

Obiettivo

L'obiettivo principale è sviluppare una pipeline per la ricostruzione tridimensionale del distretto dentale a partire da una singola radiografia panoramica, minimizzando la dipendenza da esami CBCT tradizionali. In particolare, viene proposto un framework che consente di:

- Predire una curva mandibolare paziente-specifica a partire da una PX reale.
- Generare immagini MPR (Multiplanar Reformation) lungo la curva stimata.
- Interpolare un volume tridimensionale coerente con la morfologia dentale del paziente.
- Eseguire il ray tracing sul volume interpolato per ottenere una proiezione panoramica sintetica.
- Confrontare la proiezione sintetica con la PX reale per ottimizzare iterativamente la curva mandibolare e i parametri del sistema di proiezione.

Questo approccio mira a ottenere un'alternativa a bassa dose alla CBCT, che sia personalizzabile, automatizzabile e integrabile nei flussi clinici esistenti.

Strategia

Per raggiungere gli obiettivi descritti, è stata progettata una pipeline strutturata in più fasi, ciascuna delle quali affronta uno specifico sottoproblema della ricostruzione tridimensionale. Il flusso proposto parte da una panoramica reale (PX) e procede attraverso le seguenti fasi:

1. **Stima iniziale della curva mandibolare:** mediante regressione statistica su traiettorie mandibolari acquisite automaticamente dal dataset, vengono stimati i punti di controllo di una curva spline cubica che approssima l'andamento medio dell'arcata dentale.
2. **Generazione delle immagini MPR:** la panoramica reale viene utilizzata come input per una rete generativa (cGAN), addestrata a produrre immagini MPR (Multiplanar Reformations) coerenti con l'anatomia dentale del paziente.
3. **Interpolazione del volume 3D:** le immagini MPR generate dalla rete, insieme alla curva spline stimata, vengono utilizzate per ricostruire un volume tridimensionale coerente con la morfologia del paziente.
4. **Simulazione della panoramica sintetica:** mediante ray tracing, viene proiettata la panoramica a partire dal volume ricostruito.
5. **Ottimizzazione:** la curva mandibolare iniziale e i parametri geometrici del sistema di proiezione vengono ottimizzati iterativamente, confrontando la panoramica oggetto di ottimizzazione con quella reale tramite metriche di similarità strutturale.

La rete generativa avversaria, addestrata separatamente su un dataset composto da coppie correlate di immagini panoramiche 2D e volumi MPR derivati da CBCT, costituisce il nucleo del processo di generazione volumetrica e garantisce la coerenza anatomica lungo l'intera pipeline.

Ogni fase è stata progettata in modo modulare, così da consentire la sostituzione o l'ottimizzazione dei singoli componenti, con l'obiettivo di massimizzare l'aderenza della ricostruzione sintetica all'anatomia reale del paziente.

Chapter 2

Revisione della letteratura

2.1 CBCT e imaging dentale

Nel contesto dell’odontoiatria moderna, le tecniche di imaging rappresentano strumenti indispensabili per una diagnosi accurata e una pianificazione terapeutica efficace. Tra le tecnologie più utilizzate si annoverano la Cone Beam Computed Tomography (CBCT) e la radiografia panoramica a raggi X (PX), ciascuna caratterizzata da specifici vantaggi e limitazioni.

La CBCT consente di acquisire immagini tridimensionali ad alta risoluzione [23], offrendo una visualizzazione dettagliata delle strutture anatomiche dentali e ossee [21]. È particolarmente indicata per la pianificazione implantare, la diagnosi di patologie complesse e la valutazione dell’articolazione temporo-mandibolare (ATM). Tuttavia, l’impiego della CBCT comporta un’esposizione più elevata alle radiazioni ionizzanti rispetto alla PX e presenta costi maggiori, fattori che ne limitano l’uso sistematico.[9]

La Radiografia Panoramica, al contrario, fornisce una vista bidimensionale dell’intera arcata dentale e delle strutture mandibolari, risultando più adatta a esami di screening generale, ortodonzia e monitoraggio dello sviluppo dentale. I suoi principali vantaggi sono la bassa dose di radiazioni e il costo contenuto. Tuttavia, soffre di limitazioni intrinsecche, come sovrapposizioni anatomiche, distorsioni geometriche e assenza di informazioni tridimensionali, che ne riducono l’efficacia nei casi diagnostici più complessi.

La letteratura recente evidenzia l’importanza di valutare criticamente l’impiego di ciascuna tecnica in funzione dell’applicazione clinica, con crescente interesse verso approcci ibridi. Tra questi, si colloca il framework Oral-3D[26], che si propone di combinare i vantaggi delle tecnologie PX e CBCT attraverso l’impiego di tecniche di deep learning per ricostruire modelli 3D a partire da immagini panoramiche 2D. L’obiettivo è migliorare la qualità diagnostica pur riducendo l’esposizione radiologica e contenendo i costi, rispondendo così a un’esigenza sempre più sentita in ambito clinico ed etico.

2.1.1 Cone Beam Computed Tomography (CBCT)

La *Cone Beam Computed Tomography* (CBCT) è una tecnica di imaging tridimensionale che utilizza un fascio conico di raggi X per acquisire dati volumetrici ad alta risoluzione delle strutture anatomiche cranio-facciali. Ampiamente adottata in odontoiatria e chirurgia maxillo-facciale, la CBCT consente una visualizzazione dettagliata di denti, radici, ossa mascellari e articolazioni temporo-mandibolari (ATM), risultando particolarmente utile per diagnosi complesse e pianificazioni chirurgiche.

Principio di funzionamento e procedura di acquisizione

Durante l'esame, il paziente viene posizionato in modo stabile all'interno del macchinario CBCT, che esegue una rotazione completa di 360° attorno alla testa. Durante questa rotazione, vengono acquisite centinaia di proiezioni bidimensionali da diverse angolazioni.



Figure 2.1: Esempio di installazione di un dispositivo CBCT, utilizzato per acquisire immagini del distretto odontomaxillofacciale e delle strutture ENT.

Le immagini ottenute possono essere esplorate nei tre piani ortogonali principali:

- **Assiale (Z):** sezione orizzontale, utile per valutare la disposizione radicolare.
- **Sagittale (X):** vista laterale, ideale per analizzare la posizione dei denti e delle radici.
- **Coronale (Y):** vista frontale, utile per osservare simultaneamente le arcate dentali e le ATM.

I dati acquisiti durante la scansione vengono elaborati da appositi software di ricostruzione per generare un volume tridimensionale del distretto anatomico esaminato. Tale volume può essere ulteriormente visualizzato tramite rendering 3D, migliorando la comprensione spaziale rispetto alle immagini bidimensionali tradizionali.[3]

La configurazione del fascio conico e la rotazione attorno a un centro fisso sono illustrate nella Figura 2.3. In questa rappresentazione schematica, il tubo radiogeno e il rilevatore ruotano simultaneamente attorno alla testa del paziente, acquisendo un volume cilindrico di dati proiettivi.

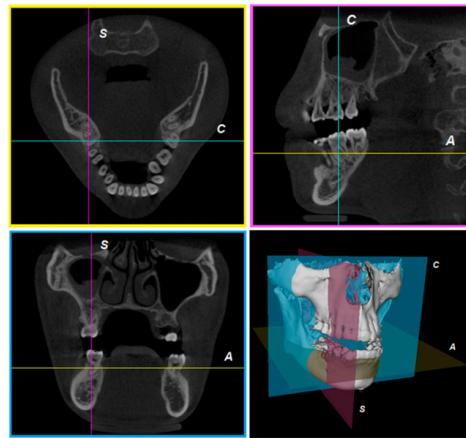


Figure 2.2: Riformattazione multiplanare. A, C e S indicano rispettivamente i piani assiale, coronale e sagittale. È inclusa anche una ricostruzione tridimensionale (3D).

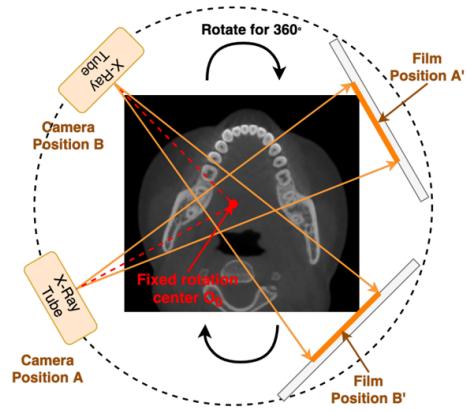


Figure 2.3: Processo di proiezione nella CT generale. Il tubo radiogeno e il rilevatore ruotano attorno al paziente per acquisire dati volumetrici da diverse angolazioni.

Vantaggi clinici La CBCT offre numerosi vantaggi rispetto alla radiografia tradizionale:

- **Alta risoluzione spaziale:** consente di distinguere dettagli fini delle strutture anatomiche.
- **Accuratezza diagnostica:** fondamentale nella pianificazione implantare, nell’endodonzia avanzata e nella chirurgia ortognatica.
- **Sezioni multiplanari:** migliorano la precisione dell’analisi anatomica e la localizzazione di lesioni.
- **Visualizzazione tridimensionale:** fornisce un modello completo del distretto cranio-facciale.

Limitazioni Nonostante i suoi numerosi vantaggi, l’impiego della CBCT presenta alcune criticità:

- **Esposizione maggiore alle radiazioni** rispetto a tecniche bidimensionali, seppur inferiore rispetto alla tomografia assiale convenzionale (TAC).
- **Costi elevati**, sia per l’acquisto dell’apparecchiatura che per l’esecuzione dell’esame.

- **Accessibilità limitata**, specialmente in contesti a bassa disponibilità economica o in pazienti pediatrici e vulnerabili.

Linee guida e prospettive future Secondo le linee guida dell'*ICRP* (International Commission on Radiological Protection), *ADA* (American Dental Association) e *SIRM* (Società Italiana di Radiologia Medica), l'uso della CBCT deve essere giustificato da un chiaro vantaggio clinico, rispettando sempre il principio *ALARA* (As Low As Reasonably Achievable) per la dose radiologica[2].

Per superare queste limitazioni, sono in fase di sviluppo approcci alternativi basati su *intelligenza artificiale* e *modelli generativi*, che mirano a ricostruire informazioni tridimensionali a partire da dati 2D, come le immagini panoramiche. In questa direzione si colloca il framework *Oral-3D*, che propone una pipeline innovativa per la ricostruzione 3D a partire da una semplice radiografia panoramica, con l'obiettivo di ridurre la dose radiologica, mantenendo un'elevata accuratezza diagnostica.

2.1.2 Radiografia Panoramica

La *Radiografia Panoramica* (PX), anche nota come ortopantomografia (OPT), è una tecnica di imaging extraorale che consente di acquisire un'immagine bidimensionale (2D) dell'intera bocca del paziente in un'unica vista panoramica. È ampiamente impiegata in ambito odontoiatrico per la diagnosi e la pianificazione dei trattamenti, grazie alla sua capacità di visualizzare simultaneamente le arcate dentali, la mandibola, il mascellare superiore e le articolazioni temporo-mandibolari (ATM).



Figure 2.4: Esempio di Radiografia Panoramica

Meccanismo di acquisizione

Durante l'esame, il paziente viene posizionato in modo stabile mediante supporti per mento e fronte; talvolta viene utilizzato un blocco di morso per mantenere la bocca aperta. Il sistema radiografico esegue una rotazione semicircolare attorno alla testa del paziente, proiettando un fascio di raggi X su un rilevatore posto sul lato opposto. Il movimento sincrono del tubo radiogeno e del rilevatore segue una traiettoria curva progettata per adattarsi alla forma dell'arco mandibolare.

Questa traiettoria curva definisce il *piano focale*: una regione geometrica dove le strutture anatomiche ricevono più proiezioni, risultando più nitide nell'immagine finale. Le strutture

al di fuori di questo piano, al contrario, vengono proiettate meno frequentemente, producendo immagini meno definite o distorte. [27]

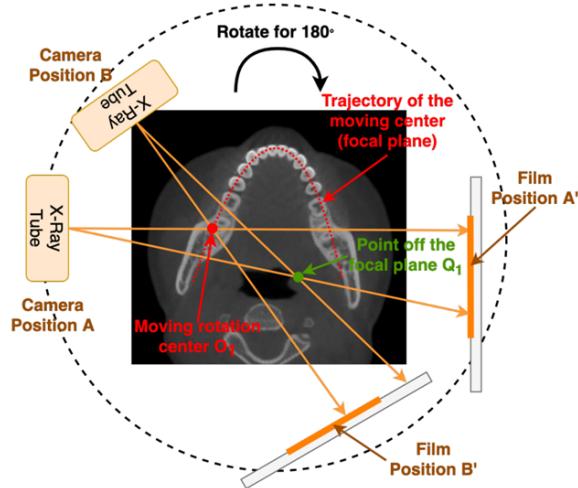


Figure 2.5: Schema del processo di acquisizione nella radiografia panoramica. I raggi X ruotano attorno al paziente seguendo una traiettoria curva (piano focale).

Vantaggi clinici La radiografia panoramica presenta numerosi vantaggi:

- **Bassa dose di radiazioni** ($5,5\text{--}22 \mu\text{Sv}$), nettamente inferiore rispetto a tecniche volumetriche.
- **Costo contenuto**, in media circa €49,29, che la rende accessibile per lo screening di massa.
- **Rapidità di acquisizione** e facilità d'uso in ambito ambulatoriale.

Limiti Tuttavia, la PX presenta anche delle limitazioni rilevanti:

- **Assenza di informazione tridimensionale**, che limita la valutazione di volumi, spessori ossei e profondità.
- **Distorsioni geometriche**, introdotte dalla particolare geometria di acquisizione possono alterare le proporzioni reali delle strutture, compromettendo l'accuratezza diagnostica.
- **Sovraposizioni anatomiche**, che possono mascherare patologie o ostacolare l'identificazione di strutture specifiche.

Utilizzi clinici La radiografia panoramica (PX) è spesso utilizzata come esame di primo livello per fornire una visione generale della bocca. Permette di individuare rapidamente eventuali denti inclusi o in sovrannumero, valutare la posizione dei denti del giudizio, rilevare anomalie come carie, cisti o lesioni ossee, e monitorare lo sviluppo dentale nei pazienti in età pediatrica. [17]

Prospettive e integrazione con nuove tecnologie L'integrazione tra imaging tradizionale e intelligenza artificiale rappresenta una direzione promettente per superare i limiti informativi della PX. In particolare, il framework *Oral-3D* adotta un'architettura generativa condizionata (cGAN) in grado di ricostruire un volume CBCT simulato a partire da una singola radiografia panoramica. Il processo si basa sull'estrazione di una curva mandibolare individuale, lungo cui vengono campionate immagini multiplanari (MPR) da cui si genera il volume 3D. Questo approccio consente una ricostruzione tridimensionale personalizzata, potenzialmente utile in contesti a bassa disponibilità di risorse o nei pazienti per i quali l'esposizione alla CBCT non è giustificabile.

2.2 Tecniche di proiezione

L'acquisizione di un volume CBCT consente di generare una vasta gamma di immagini derivate tramite riformattazione multiplanare e tecniche di proiezione, ognuna delle quali è adatta a specifiche esigenze diagnostiche. Tra queste, le più rilevanti nel contesto odontoiatrico sono la **riformattazione multiplanare (MPR)**, la **proiezione di massima intensità (MIP)** e la **proiezione panoramica** ottenuta tramite tecniche di ray tracing lungo percorsi curvi definiti sull'anatomia mandibolare.

Riformattazione Multiplanare (MPR)

La riformattazione multiplanare (MPR, *Multiplanar Reconstruction*) consente di generare immagini bidimensionali orientate arbitrariamente a partire da un volume tridimensionale. Nella pratica odontoiatrica, questa tecnica è comunemente utilizzata per ottenere viste assiali, sagittali o coronali del distretto maxillo-facciale, risultando particolarmente utile per l'analisi dettagliata delle strutture dentali e ossee.

Nel presente lavoro, la riformattazione MPR riveste un ruolo centrale in due fasi distinte. In primo luogo, rappresenta l'output tridimensionale generato direttamente da una rete neurale, che apprende a ricostruire un volume coerente con l'anatomia mandibolare del paziente a partire da una singola immagine panoramica bidimensionale. In secondo luogo, l'MPR viene interpolata lungo una curva spline che segue l'arcata mandibolare del paziente, producendo un volume tridimensionale coerente con l'anatomia dentale.

Questo volume è successivamente proiettato mediante ray tracing per simulare una panoramica sintetica. Tale proiezione consente di confrontare visivamente e geometricamente l'immagine sintetica con quella reale, facilitando la valutazione della qualità ricostruttiva e permettendo, nel ciclo di ottimizzazione, l'aggiornamento iterativo della curva spline e dei parametri geometrici, al fine di massimizzare l'allineamento tra le due immagini.

Questo approccio consente non solo di preservare la coerenza geometrica dell'anatomia mandibolare lungo l'arco dentale, ma anche di valutare in modo controllato e ripetibile l'efficacia del processo ricostruttivo in termini di fedeltà all'immagine panoramica di riferimento.

Proiezione di Massima Intensità (MIP)

La tecnica MIP (*Maximum Intensity Projection*) consiste nel proiettare un volume tridimensionale (come una CBCT) lungo una direzione ortogonale, selezionando per ciascun punto del piano il valore massimo di attenuazione tra i voxel presenti lungo quella direzione. Ad esempio, nella proiezione assiale, si considera ogni colonna verticale di voxel (lungo l'asse Z) e si seleziona il valore più elevato, ottenendo un'immagine bidimensionale che enfatizza le strutture ad alta densità, come i denti e i tessuti calcificati.

Sebbene in letteratura si parli talvolta di “raggi” che attraversano il volume, ciò rappresenta una semplificazione concettuale: non vengono tracciati raggi geometrici espliciti, bensì si esegue un’operazione di massimo lungo un asse specifico (ad esempio Z, per la vista assiale), senza modellare fisicamente il percorso del fascio radiante.

Nel contesto del presente studio, la MIP viene utilizzata per generare una vista assiale del volume CBCT. Questa proiezione fornisce una rappresentazione planare semplificata ma informativa dell'anatomia mandibolare, da cui viene automaticamente stimata una traiettoria curva che segue l'arcata dentale. Tale curva, ottenuta tramite segmentazione e interpolazione spline sulla MIP assiale, definisce il percorso lungo il quale vengono selezionate le immagini MPR dal volume CBCT.

L'impiego della MIP in questa fase garantisce un'elevata visibilità delle strutture ossee rilevanti e consente una segmentazione automatica dell'arcata dentale più robusta rispetto ad approcci più sensibili al rumore o alla variabilità anatomica.

Proiezione Panoramica Sintetica

La proiezione panoramica mira a riprodurre un'immagine analoga alla radiografia panoramica tradizionale (PX) a partire dal volume CBCT. In questo approccio, viene definita una traiettoria curva che segue l'arco mandibolare del paziente. Lungo tale traiettoria vengono generati raggi proiettivi tramite ray tracing: ciascun raggio attraversa il volume tridimensionale e campiona i dati lungo il proprio percorso. I valori di attenuazione raccolti vengono aggregati per generare un'immagine panoramica sintetica.

Questa tecnica permette di simulare fedelmente il processo di acquisizione panoramica mantenendo un controllo preciso sulla geometria di proiezione. Rispetto alla PX reale, consente di ridurre distorsioni e sovrapposizioni anatomiche, oltre a fornire una base coerente per valutare la qualità della ricostruzione volumetrica o per ottimizzare iterativamente la curva mandibolare nel processo di registrazione.

2.3 Proiezione panoramica da CBCT

La possibilità di sintetizzare immagini panoramiche a partire da volumi CBCT rappresenta un aspetto chiave in numerosi studi recenti in ambito odontoiatrico e radiologico.

Negli ultimi anni, sono stati proposti diversi approcci per simulare immagini panoramiche

a partire da dati volumetrici CBCT, con l’obiettivo di ottenere rappresentazioni bidimensionali comparabili alla radiografia panoramica convenzionale (PX). Tali metodi si basano su modelli fisici e geometrici del processo radiografico, e spesso integrano tecniche di riformattazione planare, interpolazione lungo traiettorie curve e simulazione del percorso dei raggi X attraverso il volume.

Impiego nelle pipeline di addestramento e ottimizzazione

Nel presente lavoro, la proiezione di immagini panoramiche sintetizzate da CBCT riveste un ruolo centrale in due fasi distinte ma correlate:

- **Creazione del dataset sintetico:** per l’addestramento della rete neurale, è stato necessario costruire una base di dati composta da coppie immagine panoramica – volume CBCT coerenti. A tal fine, si parte da un volume CBCT reale, si estrae la curva mandibolare e si applica un processo di ray tracing per ottenere l’immagine panoramica sintetica da associare al volume.
- **Fase di valutazione e ottimizzazione:** una volta ottenuto un volume ricostruito dalla rete a partire da una PX reale, viene generata una nuova panoramica simulata tramite ray tracing utilizzando una curva di regressione media. Questa immagine viene confrontata con la PX reale di partenza nell’ambito di un ciclo iterativo di ottimizzazione geometrica basata sull’errore proiettivo, al fine di aggiornare la traiettoria della curva mandibolare e migliorare la coerenza geometrica.

Tali passaggi sono fondamentali per garantire la coerenza 2D–3D all’interno della pipeline proposta, assicurando al tempo stesso un adattamento personalizzato dell’anatomia ricostruita alla radiografia panoramica reale del paziente.

2.4 Ricostruzione 3D a partire da immagini panoramiche

Negli ultimi anni, la ricostruzione tridimensionale dell’anatomia orale a partire da immagini bidimensionali ha suscitato un crescente interesse nella comunità scientifica, in particolare per le implicazioni cliniche e radiologiche in ambito odontoiatrico. L’obiettivo di ridurre l’esposizione a radiazioni mantenendo una rappresentazione anatomica accurata ha motivato numerosi studi che cercano di ottenere una volumetria 3D coerente a partire da radiografie panoramiche (PX).

2.4.1 Approcci basati su Deep Learning per la ricostruzione 3D da PX

Il framework *Oral-3D* [26] proposto da Song et al. (2021) affronta la sfida della ricostruzione 3D da immagini bidimensionali combinando la ricostruzione da vista singola con una trasformazione cross-modale, con l’obiettivo di recuperare sia la forma che la densità dell’oggetto target a partire da una singola immagine panoramica (PX).

Il contributo principale risiede nell'esplorazione del trasferimento cross-modale tra immagini di dimensioni differenti nel contesto dell'imaging dentale, mediante l'utilizzo di tecniche di *deep learning*. Questo approccio consente non solo di ricostruire la forma tridimensionale e la superficie della struttura ossea, ma anche di recuperare simultaneamente le informazioni di densità, fondamentali per una diagnosi odontoiatrica accurata.

Un aspetto particolarmente innovativo del framework è il disaccoppiamento tra il processo di ricostruzione della densità e quello della forma. Il metodo si articola in due fasi distinte:

1. **Back-projection:** una rete generativa viene addestrata per apprendere la trasformazione cross-dimensionale, proiettando l'immagine panoramica 2D nello spazio 3D e apprendendo automaticamente le informazioni di profondità relative a denti e mandibola, a partire da coppie di immagini 2D–3D.
2. **Deformazione morfologica:** la struttura 3D generata, inizialmente appiattita, viene adattata a una geometria curva mediante registrazione lungo l'arco dentale. Questo passaggio consente di ripristinare la forma anatomica originale, vincolando la posizione e la curvatura della mandibola sulla base di conoscenze a priori.

I risultati sperimentali evidenziano come *Oral-3D* sia in grado di ricostruire strutture orali 3D di alta qualità a partire da una singola immagine panoramica, preservando coerentemente anche le informazioni di densità. Questo rappresenta un avanzamento significativo nel campo dell'imaging dentale, apre nuove prospettive per la diagnosi e la pianificazione dei trattamenti odontoiatrici.

Questo framework ha costituito un punto di riferimento fondamentale per il presente lavoro, che ne adotta la struttura generale ma introduce modifiche sostanziali. In particolare, mentre in *Oral-3D* la deformazione del volume generato si basa sull'arco dentale ottenuto da una fotografia intraorale del paziente — operazione che richiede l'intervento manuale di un operatore per la segmentazione o la tracciatura della curva — nel nostro approccio tale informazione viene completamente sostituita da una curva spline parametrica. È previsto infatti che la curva iniziale — ottenuta, ad esempio, tramite regressione da un set di traiettorie medie — possa essere coinvolta in un ciclo di ottimizzazione iterativa, con l'obiettivo di aggiornare progressivamente i punti di controllo della spline. Tale processo consente di allineare la proiezione panoramica simulata con quella reale, adattando in maniera automatica la geometria mandibolare al singolo paziente, senza la necessità di input manuali o immagini supplementari.

Questa scelta progettuale elimina la dipendenza da dati aggiuntivi, come la fotografia intraorale richiesta in *Oral-3D*, riduce l'intervento dell'operatore e migliora la generalizzabilità del metodo. In questo modo, si apre la strada a una stima completamente automatica e personalizzata della traiettoria mandibolare, basata unicamente sull'immagine panoramica iniziale, rendendo il sistema più adatto a scenari clinici reali e facilmente integrabile in flussi di lavoro automatizzati.

2.4.2 Altri approcci 3D da PX

Approcci recenti come Oral-3Dv2 (2023)[27] e NeBLa (2023)[20] hanno esplorato nuove metodologie per la ricostruzione tridimensionale a partire da immagini panoramiche (PX), mirando a superare alcune limitazioni degli approcci basati su dataset accoppiati.

In particolare, Oral-3Dv2 adotta una rappresentazione implicita del volume attraverso una rete MLP (Multi-Layer Perceptron), apprendendo la mappatura tra coordinate spaziali e valori di attenuazione senza necessità di volumi CBCT direttamente appaiati alla PX. L'architettura sfrutta proiezioni sintetiche derivate da una geometria simulata, ma la traiettoria curva mandibolare resta fissata e non adattata al singolo paziente.

D'altro canto, NeBLa introduce una pipeline differenziabile basata sulla legge di Beer-Lambert, modellando in modo esplicito il processo di proiezione radiografica. La rete è suddivisa in moduli per la traduzione di stile, la ricostruzione del volume e il raffinamento finale. Anche in questo caso, l'attenzione è rivolta principalmente alla qualità volumetrica, mentre la definizione della traiettoria di proiezione resta semplificata.

Contributo del presente lavoro

A differenza degli approcci esistenti focalizzati principalmente sulla qualità della ricostruzione volumetrica, il presente lavoro pone al centro la stima e l'ottimizzazione della curva mandibolare. La traiettoria curva rappresenta un elemento chiave nella generazione delle immagini panoramiche sintetiche, e la sua corretta definizione consente non solo una proiezione più fedele all'anatomia reale, ma anche un miglioramento indiretto della qualità del volume CBCT ricostruito. In tal senso, il lavoro si distingue per l'introduzione di una pipeline in cui la curva mandibolare non è predefinita, ma appresa, ottimizzata e adattata al singolo paziente attraverso un processo iterativo basato su confronto proiettivo.

2.5 Modellazione della curva mandibolare

La corretta modellazione della curva mandibolare riveste un ruolo centrale nei metodi di sintesi panoramica da CBCT, poiché la traiettoria curva guida sia l'estrazione delle sezioni MPR sia il processo di proiezione panoramica finale. In letteratura sono stati proposti approcci differenti per definire tale traiettoria, che si possono distinguere in due grandi categorie: quelli basati su curve definite (es. parabole, ellissi) e quelli basati su spline interpolanti.

Gli approcci tradizionali utilizzano curve fisse o semi-fisse, ad esempio parabole o ellissi centrate geometricamente sul volume, per definire una traiettoria generica lungo l'arcata dentale. Tuttavia, questi metodi risultano spesso poco adattabili alla morfologia specifica del paziente.

Per superare tali limitazioni, diversi studi hanno introdotto l'uso di spline cubiche o curve parametrizzabili, che consentono una modellazione più flessibile e precisa della curva mandibolare. Le spline, grazie alla loro natura interpolante e alla possibilità di essere controllate

tramite un numero limitato di punti di controllo, permettono di adattare la traiettoria curva all'anatomia individuale del paziente, migliorando l'allineamento tra la struttura ossea reale e la proiezione sintetica generata.

Questa maggiore flessibilità geometrica è stata sfruttata in vari contesti dell'imaging medicale, come la riformattazione planare curvilinea (Curved Planar Reformation, CPR) e la ricostruzione di strutture anatomiche complesse, come vasi sanguigni o arcate dentali, a partire da dati volumetrici. Ad esempio, lavori come quelli di Vrtovec et al. (2006)[29] hanno dimostrato l'efficacia delle spline nell'approssimazione dell'arcata mandibolare per scopi ortodontici, mentre studi più generali sull'imaging medicale hanno impiegato spline per seguire traiettorie anatomiche nei volumi CT e MR.[4]

Nel presente lavoro, la curva mandibolare utilizzata per la generazione del dataset sintetico viene inizialmente estratta attraverso operazioni morfologiche applicate a proiezioni MIP assiali del volume CBCT. Tali operazioni permettono di isolare la struttura dentale e stimare una traiettoria approssimativa lungo l'arcata mandibolare. Successivamente, i punti ottenuti vengono ricondotti a una rappresentazione spline interpolante, che consente di ottenere una curva continua, regolare e facilmente modificabile, adatta all'utilizzo nel processo di ray tracing e nella fase successiva di ottimizzazione.

Questa rappresentazione ibrida — basata su una prima stima morfologica ottenuta tramite operazioni di sogliatura e filtraggio sull'immagine MIP assiale, seguita da una interpolazione mediante spline — consente di costruire una curva sufficientemente accurata per la generazione iniziale della proiezione panoramica. Al contempo, mantiene la flessibilità necessaria per l'adattamento nei passaggi successivi della pipeline. I punti di controllo individuati sulla curva spline costituiscono infatti dei parametri liberi che possono essere ottimizzati iterativamente, al fine di adattare la traiettoria proiettiva alla specifica morfologia anatomica del paziente e massimizzare la coerenza tra la panoramica sintetica e quella reale.

2.6 Ray tracing per proiezione medica

Il ray tracing è una tecnica computazionale originariamente sviluppata per la computer grafica, ma da tempo applicata anche all'ambito medico per simulare immagini di proiezione, come le radiografie digitali sintetiche (DRR, Digitally Reconstructed Radiographs). In campo radiologico, questa tecnica consente di modellare il percorso dei raggi X all'interno di un volume, calcolando l'attenuazione cumulativa lungo ciascun raggio.

In letteratura, il ray tracing è stato ampiamente impiegato per la simulazione di immagini proiettive a partire da dati tomografici, sia in ambito toracico che dentale. L'approccio classico prevede il campionamento del volume lungo traiettorie rettilinee e la somma delle intensità voxel lungo ciascun raggio, rappresentando così l'intensità rilevata sul piano del detector. Alcuni lavori più recenti hanno esteso questa formulazione rendendola *differenziabile*, ovvero compatibile con il calcolo automatico del gradiente, permettendone l'integrazione diretta all'interno di pipeline di apprendimento profondo end-to-end.

Tuttavia, il ray tracing non differenziabile trova ancora ampio utilizzo in contesti di ottimizzazione che non coinvolgono il deep learning. In questi casi, si possono utilizzare metodi di ottimizzazione tradizionali, come algoritmi basati su gradienti numerici o approcci iterativi, per affinare i parametri del modello. Questo approccio consente di ottenere proiezioni accurate e di ottimizzare il risultato finale senza la necessità di un apprendimento automatico. In questo modo, il ray tracing non differenziabile rimane uno strumento versatile e potente in molte applicazioni mediche.

Nel presente lavoro, il ray tracing viene utilizzato in forma non differenziabile per simulare l’immagine panoramica a partire da un volume CBCT. I raggi vengono tracciati lungo una traiettoria curva, definita da una *spline*, precedentemente stimata a partire dalla proiezione MIP assiale. L’intensità sul piano di proiezione viene ottenuta sommando i valori del volume lungo ciascun raggio. Questa rappresentazione sintetica è poi utilizzata sia nella fase di costruzione del dataset (per generare coppie CBCT–panoramica coerenti) sia nella fase di ottimizzazione, per confrontare la proiezione simulata con una panoramica reale.

Tale strategia permette di simulare fedelmente il processo panoramico, mantenendo un controllo esplicito sulla geometria della proiezione e sulla traiettoria dei raggi, senza la necessità di integrare l’intero modello all’interno del grafo computazionale del training.

Ad esempio, il framework DiffIR [31] implementa un ray tracing differenziabile per la generazione di proiezioni DRR a partire da volumi medicali. Tuttavia, nel presente lavoro si è scelto di adottare una pipeline non differenziabile, per mantenere una maggiore flessibilità sul controllo geometrico e sul tracciamento esplicito dei raggi.

2.7 Adattamento di dominio tramite CycleGAN

Uno degli aspetti affrontati in questo lavoro riguarda la discrepanza tra le immagini panoramiche reali, acquisite da dispositivi radiografici, e quelle sintetiche ottenute da volumi CBCT mediante proiezione.

Poiché il modello generativo è stato addestrato su immagini panoramiche sintetiche proiettate da CBCT, l’utilizzo diretto di immagini reali — caratterizzate da distribuzioni di intensità, contrasto e texture differenti — introdurrebbe un disallineamento semantico tra input e target (*domain shift*), riducendo l’efficacia della generazione. Per ovviare a questo problema, è necessaria una fase di armonizzazione tra i due domini.

A tale scopo, è stata adottata una *CycleGAN* [33], un’architettura avversaria non supervisionata per la traduzione di immagini tra domini non appaiati. La rete è stata addestrata per trasformare le immagini panoramiche reali nel dominio sintetico, preservando la struttura anatomica ma adattando l’aspetto visivo a quello delle proiezioni proiettate da CBCT. L’output della CycleGAN viene quindi utilizzato come input coerente per la cGAN di ricostruzione, migliorando la compatibilità tra l’immagine 2D iniziale e il volume 3D target.

Architettura e Funzionamento

Le CycleGAN (Cycle-Consistent Generative Adversarial Networks) sono una variante delle Generative Adversarial Network (GAN) progettata per la traduzione di immagini non supervisionata, particolarmente utile in assenza di coppie di immagini corrispondenti tra due domini distinti.

In questo contesto, i due domini coinvolti sono:

- **Dominio X:** immagini panoramiche reali (PX), acquisite tramite dispositivi radiografici;
- **Dominio Y:** immagini sintetiche ($SntPX$) ottenute tramite proiezione da volumi CBCT.

L'architettura della CycleGAN comprende:

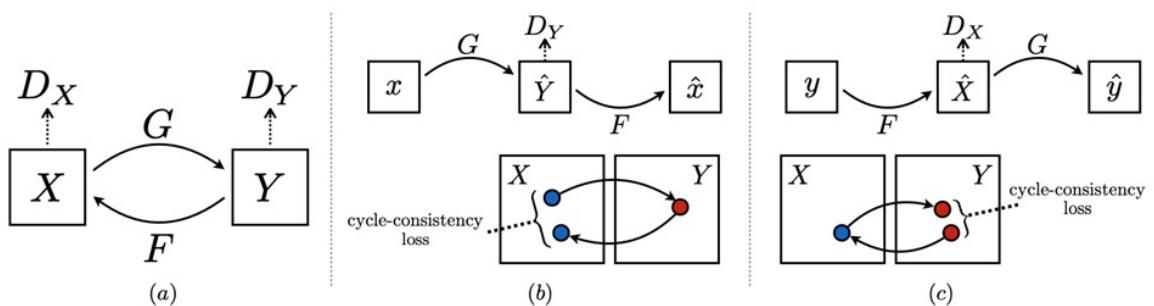


Figure 2.6: Architettura della CycleGAN utilizzata per la traduzione $PX \rightarrow SntPX$.

- **Due generatori:**

- $G : X \rightarrow Y$ mappa un'immagine PX reale nel dominio delle immagini sintetiche;
- $F : Y \rightarrow X$ esegue la trasformazione inversa, garantendo la coerenza ciclica.

- **Due discriminatori:**

- D_Y valuta se un'immagine appartiene realmente al dominio Y ;
- D_X valuta se un'immagine appartiene realmente al dominio X .

La rete è regolata da due perdite di consistenza ciclica:

- **Forward cycle:** $x \rightarrow G(x) \rightarrow F(G(x)) \approx x$
- **Backward cycle:** $y \rightarrow F(y) \rightarrow G(F(y)) \approx y$

Obiettivo della Formulazione

L'obiettivo della CycleGAN è apprendere due funzioni di mappatura:

- $G : X \rightarrow Y$
- $F : Y \rightarrow X$

in modo tale che le immagini trasformate siano realistiche e coerenti con la struttura originaria, anche in assenza di corrispondenze esplicite tra X e Y .

Adversarial Loss

Per ogni direzione di mappatura, viene impiegato un discriminatore che valuta il realismo dell'immagine generata. Le funzioni di perdita avversaria sono:

$$\mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) = \mathbb{E}_{y \sim p_{\text{data}}(y)} [\log D_Y(y)] + \mathbb{E}_{x \sim p_{\text{data}}(x)} [\log(1 - D_Y(G(x)))] \quad (2.1)$$

$$\mathcal{L}_{\text{GAN}}(F, D_X, Y, X) = \mathbb{E}_{x \sim p_{\text{data}}(x)} [\log D_X(x)] + \mathbb{E}_{y \sim p_{\text{data}}(y)} [\log(1 - D_X(F(y)))] \quad (2.2)$$

Cycle Consistency Loss

La consistenza ciclica garantisce che una trasformazione seguita dalla sua inversa riconduca all'immagine originale:

$$\mathcal{L}_{\text{cyc}}(G, F) = \mathbb{E}_{x \sim p_{\text{data}}(x)} [\|F(G(x)) - x\|_1] + \mathbb{E}_{y \sim p_{\text{data}}(y)} [\|G(F(y)) - y\|_1] \quad (2.3)$$

Obiettivo Complessivo

La funzione obiettivo globale è definita come:

$$\mathcal{L}(G, F, D_X, D_Y) = \mathcal{L}_{\text{GAN}}(G, D_Y, X, Y) + \mathcal{L}_{\text{GAN}}(F, D_X, Y, X) + \lambda \mathcal{L}_{\text{cyc}}(G, F) \quad (2.4)$$

con ottimizzazione formulata come:

$$G^*, F^* = \arg \min_{G, F} \max_{D_X, D_Y} \mathcal{L}(G, F, D_X, D_Y) \quad (2.5)$$

Identity Loss

Una loss addizionale utile è l'identity loss, impiegata per impedire modifiche non necessarie alle immagini già appartenenti al dominio target:

$$\mathcal{L}_{\text{idt}}(G, F) = \mathbb{E}_{y \sim p_{\text{data}}(y)} [\|G(y) - y\|_1] + \mathbb{E}_{x \sim p_{\text{data}}(x)} [\|F(x) - x\|_1] \quad (2.6)$$

2.8 Reti neurali per la ricostruzione 3D da immagini 2D

La ricostruzione tridimensionale a partire da immagini bidimensionali è un problema complesso e fortemente ambiguo, poiché molteplici strutture 3D possono proiettarsi sulla stessa immagine 2D [1]. Negli ultimi anni, reti neurali profonde sono state impiegate con successo per affrontare questo problema, sfruttando grandi dataset e introducendo conoscenze statistiche dell'anatomia (ad esempio modelli di forma) per inferire le informazioni mancanti [26].

Architetture principali. Diversi approcci sono stati proposti in letteratura:

- **Encoder–Decoder** puri, in cui un'immagine 2D viene codificata in uno spazio latente e poi decodificata in un volume 3D. Ad esempio, la rete 3D-R2N2 ha dimostrato questa strategia ricostruendo modelli 3D a partire da singole immagini 2D. Tali approcci volumetrici hanno mostrato efficacia ma soffrono di limitazioni in risoluzione e pesi computazionali elevati [5].
- **Variational Autoencoder (VAE)** o **GAN**, che aggiungono rispettivamente una regolarizzazione probabilistica o una perdita avversaria per aumentare la plausibilità della ricostruzione. I VAE affrontano l'ambiguità intrinseca generando molteplici forme 3D possibili per lo stesso input tramite campionamento nello spazio latente. Inoltre, combinando un VAE con un GAN (VAE-GAN) si uniscono la capacità di ricostruzione propria dei VAE con quella avversaria dei GAN, migliorando la qualità e realismo delle forme 3D generate.
- **Reti condizionate (cGAN)**, in cui sia il generatore sia il discriminatore sono condizionati da informazioni aggiuntive (ad esempio etichette o immagini). In questo schema, la rete apprende una mappatura diretta da input 2D a output 3D sfruttando un segnale condizionante (come un'immagine panoramica) per guidare la generazione [8].

Approccio seguito. Nel presente lavoro è stata adottata una *Conditional GAN (cGAN)*, in cui il generatore apprende a mappare un'immagine panoramica sintetica in un volume 3D (MPR) corrispondente. In pratica, si effettua una “back-projection” dell'immagine panoramica in un volume 3D mediante una rete generativa [26]. Il discriminatore valuta la coerenza tra la proiezione 2D di input e il volume generato, fungendo da supervisore avversario che guida il generatore a produrre volumi morfologicamente credibili e coerenti.

La scelta della cGAN è motivata dalla sua capacità di:

- catturare dettagli locali della morfologia mandibolare;
- mantenere una coerenza strutturale globale;
- apprendere la variabilità anatomica dai dati reali.

Vantaggi architetturali. Il generatore adottato nella cGAN implementata nel presente lavoro segue una struttura tipo *autoencoder*, con una fase di encoding che comprime l’informazione spaziale dell’immagine panoramica, seguita da una fase di decoding che ricostruisce il volume tridimensionale. Questa scelta architettonica è particolarmente adatta nei contesti in cui si desidera catturare una forte eterogeneità morfologica, come nel caso dell’anatomia mandibolare, caratterizzata da ampia variabilità inter-soggetto.

Diversi studi in letteratura hanno evidenziato come l’integrazione di un autoencoder all’interno del generatore di una GAN migliori la qualità e la coerenza delle ricostruzioni volumetriche, permettendo una modellazione più flessibile delle forme anatomiche [16]. Rispetto ad architetture puramente convolutive o a modelli implicitamente condizionati, il nostro schema autoencoder-guidato consente una separazione più netta tra rappresentazione latente e generazione, facilitando l’apprendimento di strutture ricorrenti ma morfologicamente variabili.

Inoltre, l’utilizzo di un discriminatore condizionato consente di vincolare la generazione del volume alla coerenza con l’immagine panoramica di input, guidando la rete verso una ricostruzione tridimensionale plausibile e morfologicamente compatibile. Questa combinazione rende il nostro modello particolarmente adatto all’integrazione in pipeline ibride basate su confronto confronto tra immagini simulate a partire dal volume generato e immagini 2D reali, come descritto nelle sezioni successive.

In letteratura è stato osservato che integrare una struttura di tipo *autoencoder* nel generatore di una cGAN può migliorare significativamente la qualità delle ricostruzioni 3D [24]. Tale configurazione consente una separazione più netta tra codifica e generazione, facilitando la modellazione di strutture complesse e variabili, come l’anatomia mandibolare.

Infine, l’architettura si presta naturalmente all’integrazione in pipeline ibride basate su approcci di tipo *analysis-by-synthesis*, in cui il confronto tra proiezioni simulate (ottenute dal volume generato) e immagini reali consente un raffinamento progressivo della coerenza 2D–3D, come avviene nella fase di ottimizzazione della curva mandibolare.

Chapter 3

Metodologia

3.1 Overview della pipeline proposta

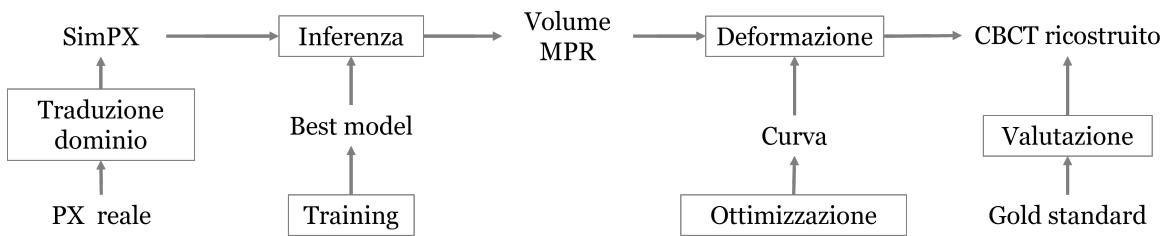


Figure 3.1: Struttura Generale della Pipeline

Il presente lavoro propone una pipeline ibrida per la ricostruzione tridimensionale di volumi dentali a partire da un'unica immagine panoramica (*Panoramic X-ray*, PX). L'approccio combina tecniche di apprendimento profondo con modelli geometrici esplicativi, con l'obiettivo di ottenere una ricostruzione volumetrica coerente con la morfologia anatomica del paziente, pur in assenza di acquisizioni tomografiche CBCT.

Il metodo adottato consiste di un modulo di *back-projection*, di natura generativa, che ricostruisce un volume 3D “appiattito” a partire dalla panoramica 2D, apprendendo informazioni di profondità latenti, e di un modulo di deformazione. La deformazione viene ottenuta mediante un processo di ottimizzazione iterativa che aggiorna i parametri di una curva spline, ovvero i suoi punti di controllo, e della geometria del sistema di proiezione (ray tracing).

Tale ottimizzazione è guidata dal confronto tra la panoramica reale e quella simulata ottenuta dal volume ricostruito, con l'obiettivo di massimizzare la coerenza tra la panoramica ricostruita e quella reale attraverso un adattamento progressivo della struttura generata alla morfologia specifica del paziente.

In sintesi, la pipeline si compone di due moduli principali:

- un modulo di *back-projection* generativo, una rete *conditional GAN* (cGAN) apprende la mappatura tra un'immagine panoramica sintetica e un volume MPR tridimensionale, rappresentando una prima stima della struttura 3D coerente con l'immagine di input;

- un modulo di deformazione geometrica, a partire dal volume MPR generato, viene ottimizzata la curva spline mandibolare con l’obiettivo di guidare in modo più accurato la deformazione geometrica del volume stesso. Questa curva, infatti, definisce la traiettoria lungo cui viene interpolato il volume per ricostruire la CBCT. La curva viene aggiornata iterativamente per minimizzare l’errore tra la panoramica proiettata (via ray tracing sul volume deformato) e quella reale.

L’intero processo di ottimizzazione è guidato dal confronto tra la proiezione simulata e quella reale, con l’obiettivo di adattare progressivamente la geometria mandibolare stimata alla morfologia effettiva del paziente. Il risultato finale è una ricostruzione tridimensionale che risulta coerente con l’immagine panoramica di partenza e il più possibile fedele al volume CBCT reale, assunto come *gold standard*.

3.1.1 Addestramento e generazione

Dataset e Preprocessing

Per l’addestramento del modello generativo è stato utilizzato il dataset pubblicato da Li et al. [15], intitolato *3D multimodal dental dataset based on CBCT and oral scan*, disponibile su Figshare¹. Il dataset desensibilizzato è stato raccolto in Cina coinvolgendo 300 pazienti di età compresa tra i 19 e i 28 anni, con l’approvazione di un comitato etico locale. Le scansioni intraorali incluse nel dataset originale non sono state impiegate in questo studio.

Sono stati selezionati 100 volumi CBCT, successivamente suddivisi in tre insiemi: 60 volumi per l’addestramento, 20 per la validazione e 20 per il test finale. Tale suddivisione è stata effettuata in maniera casuale. La scelta di dedicare l’80% dei dati all’addestramento è motivata dalla volontà di massimizzare la capacità di generalizzazione della rete, a fronte di una leggera riduzione dell’insieme di test. Quest’ultimo è stato utilizzato per una valutazione qualitativa e quantitativa delle prestazioni, ritenuta sufficiente in quanto i dati provengono da un campione omogeneo e pre-processato in modo standardizzato.

Ciascun volume CBCT presenta una risoluzione spaziale anisotropa, con dimensioni tipiche di $604 \times 604 \times 413$ voxel e spaziatura variabile lungo i tre assi. Al fine di uniformare i dati per l’addestramento del modello, è stato effettuato un preprocessing iniziale comprendente:

- **Isotropizzazione:** nei dati medici lo spacing tra voxel può variare lungo i tre assi ($x \neq y \neq z$), rendendo le misurazioni geometriche incoerenti. Per garantire proporzioni corrette e confrontabilità tra pazienti, i volumi sono stati riscalati a voxel isotropici ($x = y = z$) con spaziatura uniforme di 1 mm tramite interpolazione trilineare.
- **Normalizzazione dell’intensità:** i valori di intensità sono stati mappati in un intervallo fisso per favorire la stabilità numerica durante l’apprendimento;

¹<https://doi.org/10.6084/m9.figshare.26965903.v3>

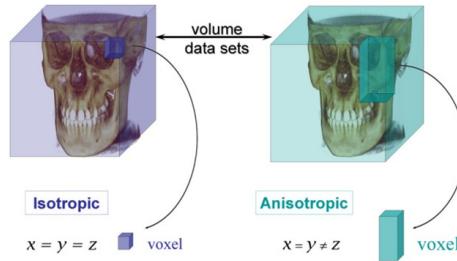


Figure 3.2: Esempio di volumi isotropici e anisotropici

- **Cropping e resize:** i volumi sono stati ridimensionati e ritagliati manualmente tramite un’interfaccia appositamente creata per adattarsi alla dimensione target del modello, pari a $288 \times 256 \times 160$ voxel;



Figure 3.3: Processo di crop

3.2 Simulazione dell’assorbimento dei raggi X nella generazione di immagini panoramiche sintetiche

Per simulare il processo proiettivo delle immagini panoramiche sintetiche (SntPX) a partire da volumi CBCT, è stato adottato un modello fisico basato sulla legge di Beer-Lambert, che descrive l’attenuazione dell’intensità dei raggi X lungo il cammino attraverso i tessuti.

L’attenuazione è modellata tramite il coefficiente di attenuazione lineare μ , il quale è proporzionalmente correlato ai valori delle unità Hounsfield (HU) presenti nel volume CBCT. L’immagine panoramica simulata viene ottenuta integrando questi coefficienti lungo ciascun raggio, riproducendo il comportamento fisico dell’acquisizione radiografica.

3.2.1 Simulazione della proiezione panoramica tramite ray tracing

La generazione dell’immagine panoramica sintetica (SimPX) a partire da un volume CBCT si basa su un modello fisico che simula l’attenuazione dei raggi X attraverso i tessuti, secondo la legge di Beer-Lambert. I dati volumetrici, rappresentati come una matrice tridimensionale in unità Hounsfield (HU), vengono convertiti in coefficienti di attenuazione lineare μ mediante la seguente relazione:

$$\mu = \mu_{\text{acqua}} + \frac{\text{HU}}{1000} \cdot (\mu_{\text{acqua}} - \mu_{\text{aria}}) \quad (3.1)$$

dove μ_{acqua} e μ_{aria} sono i coefficienti di attenuazione rispettivamente dell’acqua e dell’aria.

L'attenuazione lungo il cammino di un raggio L è descritta da:

$$I = I_0 \cdot \exp \left(- \int_0^L \mu(x) dx \right) \quad (3.2)$$

che in ambiente computazionale viene discretizzata come:

$$I = I_0 \cdot \exp \left(- \sum_i \mu_i \cdot \Delta s_i \right) \quad (3.3)$$

dove μ_i è il valore del voxel attraversato e Δs_i è il passo di campionamento lungo il raggio.

3.2.2 Generazione automatica di SimPX per il training

Poiché il framework di ricostruzione 3D si basa su apprendimento supervisionato, è necessario disporre di un dataset costituito da coppie di immagini panoramiche (PX) e volumi CBCT corrispondenti. In mancanza di acquisizioni perfettamente allineate, è stata implementata una pipeline per la generazione automatica di SimPX da volumi CBCT reali. Il processo si articola nei seguenti passaggi:

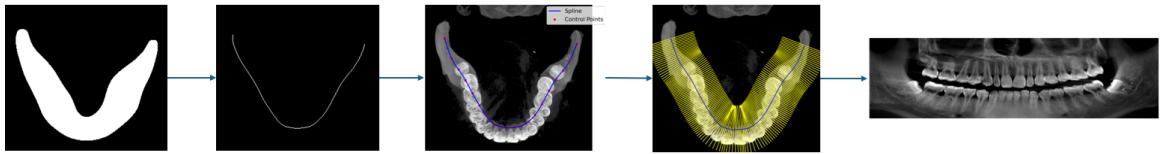


Figure 3.4: Simulazione della proiezione panoramica tramite ray tracing.

- **Definizione della traiettoria panoramica:** viene stimata una curva mandibolare anatomica direttamente dal volume CBCT. Una proiezione MIP assiale consente di visualizzare l'arco dentale, su cui vengono selezionati 20 punti di controllo utilizzati per costruire una spline cubica. Tale approccio è coerente con lavori recenti e consente un adattamento anatomico fedele e ottimizzabile.[4]
- **Simulazione dei raggi proiettivi e delle sezioni MPR:** i raggi X vengono tracciati lungo la traiettoria spline e, per ciascun raggio, viene generata una sezione MPR (Multiplanar Reconstruction). Le immagini MPR vengono infine assemblate e combinate tramite una trasformazione logaritmica:

$$I_{\text{SimPX}} = S \cdot \ln \left(\sum e^{\frac{\text{MPR}(x)}{S}} \right) \quad (3.4)$$

dove S è un parametro empirico che regola la sensibilità ai contrasti di densità, enfatizzando le strutture ossee.

Output strutturato per il training Tutti i dati generati — volume CBCT normalizzato, SntPX, traiettoria spline, immagini MPR — vengono salvati in un file .mat, strutturato per l’addestramento supervisionato del modello generativo.

3.3 Architettura della rete cGAN

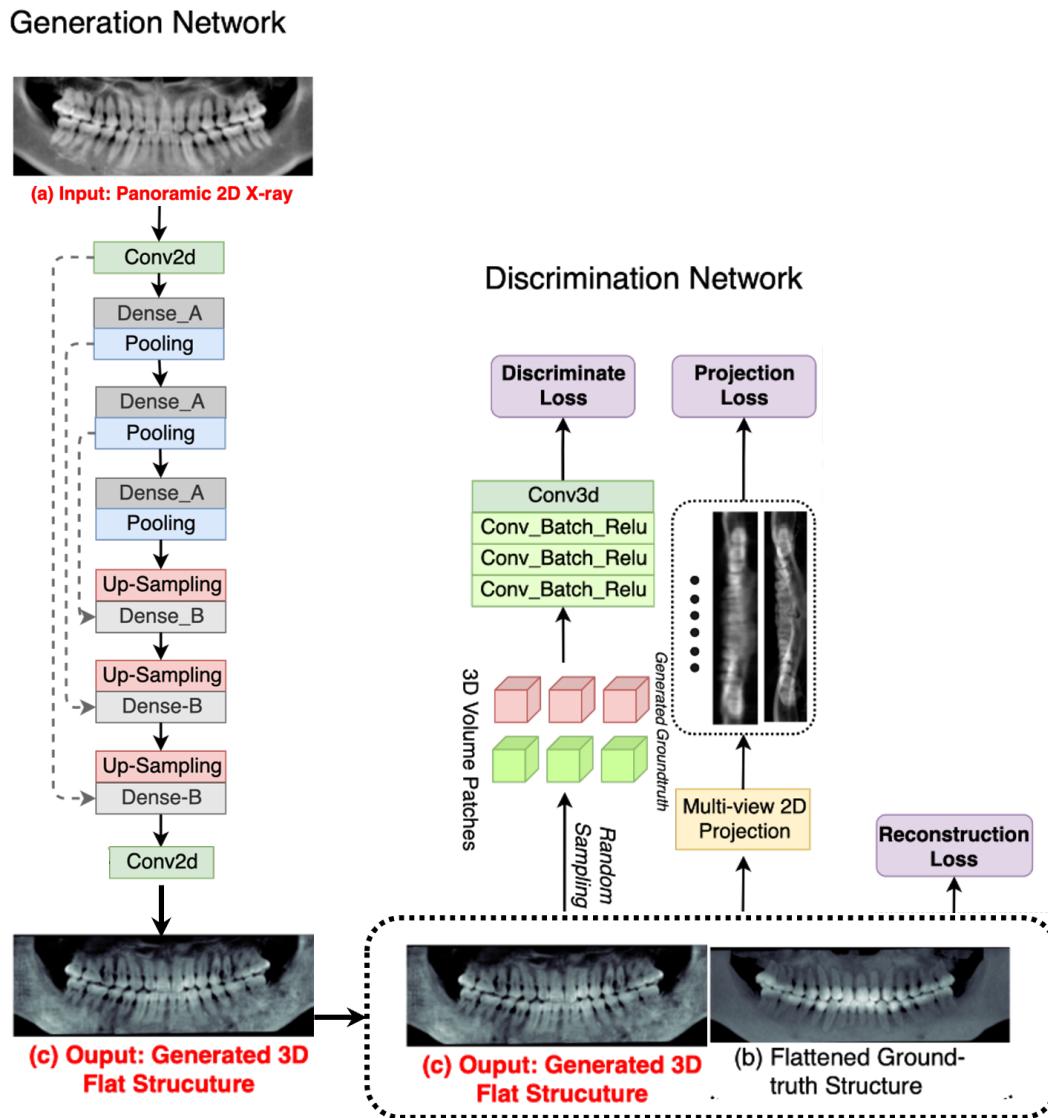


Figure 3.5: Architettura della rete cGAN

La rete generativa utilizzata in questo lavoro è una *Conditional GAN* (cGAN), progettata per apprendere una mappatura da un’immagine panoramica sintetica (SntPX) a un volume tridimensionale (MPR o CBCT). L’architettura è composta da due moduli principali:

- **Generatore (G):** riceve in input una SntPX e produce un volume tridimensionale coerente (MPR), che rappresenta una prima stima della struttura anatomica del paziente.
- **Discriminatore (D):** valuta la coerenza tra il volume di input e il volume generato, fungendo da supervisore avversario che guida il generatore a produrre volumi morfologicamente plausibili.

Il generatore adotta una struttura *encoder-decoder* con convoluzioni 2D, *Dense Blocks* e operazioni di upsampling, in cui la profondità del volume viene codificata nei canali del tensore di output. Il discriminatore sfrutta convoluzioni tridimensionali (3D) condizionate sull’immagine panoramica, secondo il paradigma *PatchGAN*.

La funzione di perdita complessiva combina una *loss* avversaria e una *loss* di ricostruzione supervisionata (tipicamente L1), con l’obiettivo di garantire sia il realismo morfologico del volume generato che l’accuratezza quantitativa rispetto al volume di riferimento.

3.3.1 Schema dell’Architettura Generatore

Il generatore implementa la trasformazione da immagine 2D a volume 3D apprendendo le informazioni di profondità latenti nell’input. La sua architettura è di tipo encoder-decoder: una sequenza di strati convolutivi 2D nella fase di encoding riduce progressivamente la risoluzione spaziale dell’immagine aumentando il numero di mappe di attivazione (feature maps), seguita da uno stadio di decoding che ripristina la risoluzione originale. Le informazioni tridimensionali vengono incorporate proprio nei canali del tensore di output, che codificano la profondità del volume 3D.

Durante la fase di generazione, il modello segue una sequenza di operazioni strutturate secondo un’architettura di tipo encoder-decoder. I principali passaggi sono:

1. **Normalizzazione dell’input:** l’immagine panoramica in ingresso viene normalizzata tramite divisione per 256.
2. **Encoding:** una serie di blocchi convoluzionali, inclusi Dense Blocks e operazioni di pooling, riducono progressivamente la risoluzione spaziale aumentando la profondità delle feature.
3. **Decoding:** tramite upsampling e skip connections, si ricostruisce la risoluzione originale, mantenendo la coerenza spaziale e arricchendo la rappresentazione volumetrica.
4. **Attivazione finale:** un layer di attivazione tanh vincola i valori voxel dell’output in un intervallo compreso tra $[-1, 1]$.

Per facilitare l’apprendimento della dimensione di profondità (non presente direttamente nella panoramica), il generatore utilizza *Dense Blocks* con collegamenti interni densi (*dense connections*), ispirati alla rete DenseNet. In particolare, si distinguono due tipi di blocchi: il **blocco A** (fase di encoding), che espande il numero di canali per estrarre le informazioni di profondità, e il **blocco B** (fase di decoding), che fonde le feature provenienti dallo strato di upsampling con quelle delle *skip connections*, mantenendo costante il numero di canali.

Grazie a questa architettura con *skip connections* e fusioni dense, il numero di canali del tensore in uscita dal generatore corrisponde al numero di slice (profondità D) del volume 3D ricostruito. L’ultimo layer applica una funzione di attivazione tanh per vincolare i valori dei voxel in un intervallo limitato (tipicamente $[-1, 1]$).

Table 3.1: Architettura del generatore (encoder-decoder) adottata in *Oral-3D*. Le dimensioni sono espresse come [canali, altezza, larghezza].

Strato	Dimensione Output
Input	[1, 160, 576]
Conv + ReLU	[10, 160, 576]
DenseBlock + MaxPool2D	[20, 80, 288]
DenseBlock + MaxPool2D	[40, 40, 144]
DenseBlock + MaxPool2D	[80, 20, 72]
UpSample + Skip Connection	[80, 40, 144]
UpSample + Skip Connection	[80, 80, 288]
UpSample + Skip Connection	[80, 160, 576]
Conv2D + tanh	[80, 160, 576]
Output	[80, 160, 576]

3.3.2 Architettura del Discriminatore (PatchGAN)

Il discriminatore è una rete convolutiva progettata per valutare il realismo locale del volume 3D generato, confrontandolo con volumi CBCT reali. *Oral-3D* adotta un discriminatore a patch (*PatchGAN*), analogo a quello proposto da Isola et al. [12] per il framework *pix2pix*, ma esteso al dominio tridimensionale.

Il discriminatore scorre piccoli sotto-volumi (patch 70^3 voxel) del volume generato e del corrispondente volume reale, producendo per ciascuna patch una classificazione binaria (real/fake). L’architettura interna riprende la classica struttura convoluzionale (Conv – Batch-Norm–LeakyReLU), ma impiega convoluzioni 3D. L’output finale è una mappa tridimensionale di probabilità (funzione sigmoide), che indica per ogni patch la stima di appartenenza a un volume reale. Questa formulazione locale contribuisce a migliorare la qualità delle strutture ad alta frequenza (come i bordi dentali), integrando la valutazione globale con una penalizzazione più fine.

Table 3.2: Architettura del discriminatore PatchGAN 3D. Le dimensioni sono relative a ciascuna patch di input di dimensione [1, 70, 70, 70] (canali, profondità, altezza, larghezza).

Strato	Dimensione Output
Input Patch	[1, 70, 70, 70]
Conv3D ($k=4$, $s=2$, $p=0$) + BN + LeakyReLU	[16, 34, 34, 34]
Conv3D ($k=4$, $s=2$, $p=0$) + BN + LeakyReLU	[32, 16, 16, 16]
Conv3D ($k=4$, $s=2$, $p=0$) + BN + LeakyReLU	[64, 7, 7, 7]
Conv3D ($k=4$, $s=1$, $p=0$) + BN + LeakyReLU	[128, 4, 4, 4]
Conv3D finale ($k=4$) + Sigmoid	[1, 1, 1, 1]
Output (flatten)	[1]

Nel metodo `discriminate()`, il discriminatore valuta separatamente le patch reali e quelle generate, minimizzando la distanza tra le predizioni e i target (0 per sintetico, 1 per reale) tramite LSGAN. Il metodo `inference()` viene invece usato durante l’addestramento del generatore, che cerca di ingannare il discriminatore facendo in modo che le patch sintetiche vengano classificate come reali.

Questo approccio patch-wise consente di guidare il generatore verso la produzione di dettagli locali realistici, mantenendo al contempo un'efficienza computazionale accettabile grazie al campionamento casuale di un numero ridotto di patch per volume.

3.3.3 Funzioni di perdita del cGAN

L'addestramento della rete cGAN è guidato da una combinazione di tre termini di perdita principali: la perdita di ricostruzione voxel-wise, la perdita di proiezione ortogonale e la perdita avversaria. Le prime due componenti supervisionate garantiscono coerenza anatomica e fedeltà volumetrica, mentre la terza promuove il realismo morfologico attraverso l'interazione competitiva tra generatore e discriminatore.

Perdita di proiezione

Un ulteriore termine di perdita è rappresentato dalla **perdita di proiezione** L_P , che confronta le proiezioni ortogonali (assiale, coronale e sagittale) dei volumi reale e generato. Tale perdita guida il generatore a distribuire coerentemente la densità volumetrica nei diversi piani, rafforzando la coerenza tridimensionale.

Nel presente lavoro, L_P è calcolata come media degli errori MSE tra le proiezioni medie (mean intensity projections) del volume generato e del volume reale, lungo ciascun asse spaziale:

$$L_P = \frac{1}{3} \sum_{d=1}^3 \mathbb{E} [\|P(y) - P(G(x))\|^2] \quad (3.5)$$

dove $P(\cdot)$ indica l'operatore di proiezione media lungo ciascuna direzione.

Questa strategia consente di penalizzare disallineamenti nella distribuzione globale della densità ossea, anche in presenza di ricostruzioni voxel-wise localmente accurate, migliorando la struttura complessiva del volume generato.

Perdita di ricostruzione voxel-wise

La **perdita di ricostruzione** \mathcal{L}_R è definita come l'errore quadratico medio (MSE, Mean Squared Error) tra il volume 3D reale y e il volume generato $G(x)$, valutato voxel per voxel. Questo termine guida l'apprendimento verso una ricostruzione fedele della distribuzione spaziale della densità ossea:

$$\mathcal{L}_R = \mathbb{E}_{x,y} [\|y - G(x)\|^2] \quad (3.6)$$

Dove:

- x è l'immagine panoramica sintetica (SntPX);
- y è il volume CBCT reale;

- $G(x)$ è il volume 3D predetto dal generatore.

Questo termine penalizza le deviazioni voxel-wise, incoraggiando il modello a riprodurre accuratamente l'intensità volumetrica associata alla morfologia anatomica reale. È particolarmente importante nella prima fase dell'addestramento, quando il discriminatore non è ancora attivo, poiché fornisce un segnale diretto e supervisionato che guida la rete verso una mappatura consistente tra input e output.

La sua presenza consente di preservare le strutture globali e locali dell'anatomia, costituendo un vincolo forte sulla fedeltà della ricostruzione rispetto ai dati acquisiti.

Sebbene la loss avversaria sia cruciale per assicurare il realismo percepito, è la perdita di ricostruzione voxel-wise a garantire l'accuratezza quantitativa della stima volumetrica. In altre parole, questo termine assicura che il volume generato non sia solo “realistico”, ma anche coerente con la struttura anatomica effettiva osservata nel CBCT di riferimento.

Adversarial Loss

Dopo 100 epoche, il generatore non è più addestrato solo a “ricostruire bene”, ma anche a ingannare il discriminatore, cercando di produrre volumi sempre più realistici. Il primo termine è la **perdita avversaria** (*adversarial loss*), che spinge il generatore G a produrre volumi 3D realistici secondo il giudizio del discriminatore D .

Nel presente lavoro, per stabilizzare l'addestramento del modello generativo, è stata adottata la formulazione *Least Squares GAN* (LSGAN) [18].

La funzione di perdita del discriminatore D e del generatore G è definita come segue:

$$\mathcal{L}_D = \mathbb{E}_y [(D(y) - 1)^2] + \mathbb{E}_x [D(G(x))^2] \quad (3.7)$$

$$\mathcal{L}_G = \mathbb{E}_x [(D(G(x)) - 1)^2] \quad (3.8)$$

dove:

- x è l'immagine panoramica in input (PX)
- y è il volume 3D reale (flattened CBCT)
- $G(x)$ è il volume 3D generato
- $D(\cdot)$ è la predizione del discriminatore

Questa formulazione penalizza quadraticamente gli errori del discriminatore, incentivando l'output del generatore a collocarsi nella regione desiderata (output ≈ 1 per immagini realistiche), e favorisce una convergenza più stabile e un gradiente meno rumoroso rispetto al classico GAN con loss a cross-entropia.

Interpretazione della loss Nel modello LSGAN, l’obiettivo del discriminatore è quello di:

- assegnare un valore vicino a 1 quando riceve un volume reale y , ovvero $D(y) \rightarrow 1$;
- assegnare un valore vicino a 0 quando riceve un volume generato, ovvero $D(G(x)) \rightarrow 0$.

La sua loss \mathcal{L}_D sarà quindi bassa se riesce a distinguere correttamente reale da sintetico.

Al contrario, il generatore G mira a *ingannare* il discriminatore, cercando di far apparire i volumi generati come reali. Per farlo, cerca di spingere il discriminatore a produrre $D(G(x)) \rightarrow 1$. La sua loss \mathcal{L}_G sarà quindi bassa se i dati sintetici risultano indistinguibili da quelli reali.

Equilibrio dinamico È importante sottolineare che non è auspicabile che la loss del discriminatore sia costantemente bassa. Un discriminatore troppo efficace impedisce al generatore di apprendere, producendo gradienti trascurabili. Viceversa, un discriminatore troppo debole non fornisce al generatore un segnale utile per migliorare. È quindi fondamentale mantenere un equilibrio dinamico tra i due moduli durante l’addestramento. Nel presente lavoro, per facilitare questo bilanciamento, il discriminatore è stato introdotto solo dopo le prime 100 epoche, lasciando inizialmente al generatore la possibilità di apprendere una rappresentazione grezza utile alla convergenza.

Confronto con la loss a cross-entropia Nella formulazione originale dei GAN [11], la funzione di perdita si basa sulla cross-entropia binaria, interpretando il compito del discriminatore come un problema di classificazione. In tal caso, le loss sono definite come:

$$\mathcal{L}_D^{(\text{BCE})} = -\mathbb{E}_y[\log D(y)] - \mathbb{E}_x[\log(1 - D(G(x)))] \quad (3.9)$$

$$\mathcal{L}_G^{(\text{BCE})} = -\mathbb{E}_x[\log D(G(x))] \quad (3.10)$$

Tuttavia, questa formulazione può generare gradienti deboli nelle fasi iniziali dell’addestramento, specialmente quando il discriminatore diventa troppo accurato. In tali condizioni, il generatore riceve segnali poco informativi, rendendo l’ottimizzazione instabile o lenta a convergere.

Per questo motivo, nel presente lavoro è stata adottata la formulazione LSGAN per la loss avversariale, in quanto consente una migliore stabilità numerica e un gradiente più utile nelle prime fasi in cui il generatore è ancora poco allenato. Inoltre, l’introduzione ritardata del discriminatore, prevista a partire dalla 100^a epoca, evita che quest’ultimo domini eccessivamente l’ottimizzazione iniziale, lasciando spazio al generatore per apprendere una mappatura di base guidata dalla sola loss ricostruttiva.

Dal punto di vista computazionale, il gradiente della funzione di perdita \mathcal{L}_G viene calcolato rispetto ai parametri del generatore θ_G , attraverso la seguente derivazione composta:

$$\frac{\partial \mathcal{L}_G}{\partial \theta_G} = \frac{\partial \mathcal{L}_G}{\partial D(G(x))} \cdot \frac{\partial D(G(x))}{\partial G(x)} \cdot \frac{\partial G(x)}{\partial \theta_G} \quad (3.11)$$

Ciò implica che, anche se il discriminatore D non viene aggiornato durante l'ottimizzazione del generatore G , esso resta parte integrante del percorso di retropropagazione, influenzando direttamente il segnale di apprendimento ricevuto dal generatore.

Nel caso della cross-entropia, quando $D(G(x)) \rightarrow 0$ (ossia quando il discriminatore è molto sicuro che il dato generato è falso), la derivata del logaritmo tende a zero, causando un fenomeno noto come *vanishing gradient*. Al contrario, la penalizzazione quadratica della LSGAN mantiene un gradiente proporzionale all'errore residuo ($D(G(x)) - 1$), anche quando $D(G(x))$ è lontano dal target desiderato. Questo consente di mantenere un flusso di apprendimento più costante ed efficace, in particolare nelle fasi iniziali dell'addestramento.

Funzione di perdita complessiva

La funzione totale di perdita utilizzata per addestrare il generatore è quindi definita come:

$$L_G^{\text{tot}} = \lambda_1 L_G + \lambda_2 L_R + \lambda_3 L_P \quad (3.12)$$

dove:

L_G è la perdita avversaria,

L_R è la perdita di ricostruzione voxel-wise,

L_P è la perdita sulle proiezioni ortogonali.

I pesi λ_i sono stati determinati empiricamente per bilanciare i contributi delle diverse componenti della funzione obiettivo. In particolare, il termine avversario è associato a un peso ridotto ($\lambda_{\text{adv}} = 0.05$), così da favorire la stabilità dell'addestramento, soprattutto nelle fasi iniziali. Al contrario, i termini di ricostruzione ($\lambda_R = 1.0$) e proiezione ($\lambda_P = 1.0$) rivestono un ruolo principale nel guidare il modello verso soluzioni fedeli alla struttura anatomica e coerenti nei diversi piani.

3.3.4 Strategie di Addestramento

Il training avviene in modo alternato: si aggiornano prima i pesi del discriminatore D minimizzando L_D e successivamente quelli del generatore G minimizzando L_G^{tot} . Per facilitare la convergenza, entrambe le reti impiegano tecniche di normalizzazione delle attivazioni (Batch Normalization) e funzioni di attivazione come *ReLU* o *LeakyReLU*.

Si utilizza l'ottimizzatore Adam [14], con learning rate iniziale pari a 10^{-3} , che viene ridotto di un fattore 10 ogni 50 epoche. Il batch size è molto ridotto (anche 1), per contenere i costi computazionali dovuti ai volumi 3D e aumentare la variabilità percepita dal discriminatore.

L'addestramento è prolungato (fino a 300 epoche nel caso *Oral-3D*) e monitorato tramite un set di validazione per applicare tecniche di *early stopping*. Grazie a queste strategie

e alla combinazione di perdite strutturate, il modello riesce ad apprendere una mappatura robusta 2D→3D, generando ricostruzioni volumetriche realistiche a partire da immagini panoramiche.

Ottimizzazione dei parametri di rete

Per l'ottimizzazione dei pesi del generatore è stato utilizzato l'algoritmo *Adam* [14], scelto per la sua efficienza nella gestione di gradienti rumorosi e per la sua capacità di adattarsi dinamicamente alla curvatura della funzione di perdita. Adam è particolarmente adatto al generatore, poiché consente una rapida convergenza anche in presenza di una funzione obiettivo in continua evoluzione, indotta dalla natura avversaria del training.

Per il discriminatore è stato invece adottato l'ottimizzatore *SGD* (Stochastic Gradient Descent), più semplice e meno aggressivo. Questa scelta riflette un compromesso: mentre il generatore deve reagire velocemente ai feedback del discriminatore, quest'ultimo dovrebbe evolvere più lentamente, per non diventare troppo dominante e impedire al generatore di ricevere gradienti significativi. L'uso di SGD introduce una dinamica più regolare nel discriminatore, contribuendo a mantenere un equilibrio stabile durante l'addestramento.

Nel presente lavoro, entrambi i moduli condividono lo stesso tasso di apprendimento, pur impiegando algoritmi di ottimizzazione differenti. Questa configurazione ha garantito una dinamica di addestramento stabile, senza fenomeni di predominanza da parte di uno dei due moduli.

3.3.5 Metriche di valutazione durante l'addestramento

Per monitorare la qualità della ricostruzione tridimensionale ottenuta dalla rete generativa cGAN, sono state adottate tre metriche quantitative fondamentali:

PSNR (Peak Signal-to-Noise Ratio) Il PSNR misura la differenza voxel-wise tra il volume generato $G(x)$ e il volume CBCT reale y . È calcolata secondo la formula:

$$PSNR = 20 \cdot \log_{10} \left(\frac{SCALE}{\sqrt{MSE}} \right) \quad (3.13)$$

dove MSE è l'errore quadratico medio tra i volumi e $SCALE$ è il range massimo dei valori normalizzati (es. 2.0).

SSIM (Structural Similarity Index) Lo SSIM valuta la somiglianza strutturale tra volumi, considerando luminanza, contrasto e struttura. L'implementazione è tridimensionale e si basa su convoluzioni locali con kernel gaussiani. I valori finali sono espressi in percentuale.

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \quad (3.14)$$

dove μ , σ , e σ_{xy} sono le statistiche locali.

DICE coefficient : misura l'overlap tra il volume generato binarizzato ($G(x) > -0.8$) e il volume MPR interpolato ($MPR > -0.5$). Il coefficiente di Dice è definito come:

$$Dice = \frac{2|A \cap B|}{|A| + |B|} \cdot 100 \quad (3.15)$$

dove A e B sono le maschere binarie confrontate del volume reale e generato.

Queste metriche vengono calcolate su ogni batch del set di validazione durante l'addestramento. I valori medi e le deviazioni standard vengono tracciati per monitorare la convergenza e selezionare il miglior modello. In particolare, le metriche sono combinate in uno *score complessivo* utilizzato per monitorare le prestazioni del modello e selezionare il checkpoint migliore.

3.3.6 Adattamento di dominio tramite CycleGAN

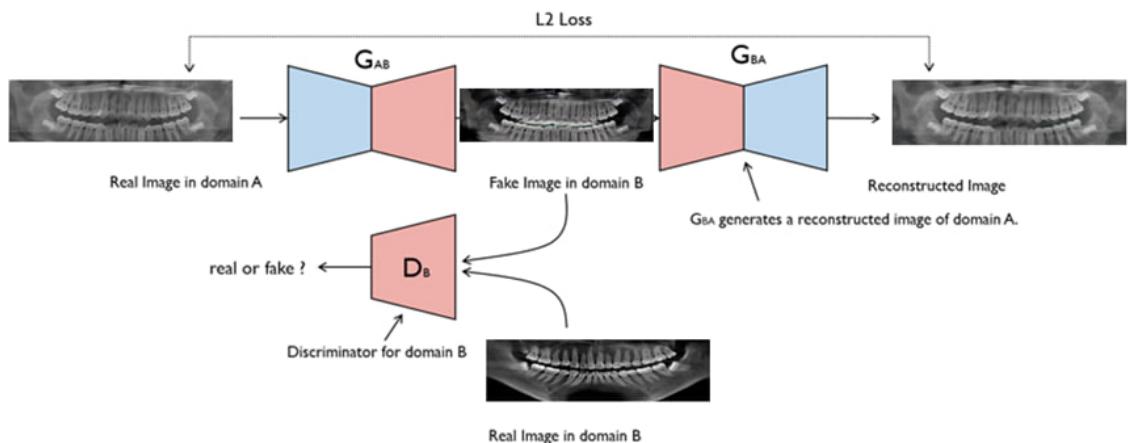


Figure 3.6: Configurazione della CycleGAN per la traduzione dal dominio reale a quello sintetico.

Per affrontare la discrepanza di dominio tra le immagini panoramiche reali (PX) e quelle sintetiche derivate da CBCT (SntPX), è stata implementata una CycleGAN. Questo approccio di traduzione di dominio non supervisionato consente di trasformare le immagini PX reali in immagini che rispecchiano lo stile delle SntPX, mantenendo inalterate le caratteristiche anatomiche.

La CycleGAN è composta da due coppie di reti neurali: due generatori e due discriminatori. I generatori trasformano le immagini da un dominio all'altro e viceversa, mentre i discriminatori valutano quanto le immagini generate siano realistiche. Durante l'addestramento, la CycleGAN massimizza la coerenza ciclica, assicurando che un'immagine convertita in un dominio possa essere riconvertita nella sua forma originale.

In questo lavoro, la CycleGAN svolge un ruolo ausiliario e non rappresenta il fulcro della pipeline. La sua funzione è quella di ridurre il domain shift tra immagini PX reali e SimPX,

migliorando la compatibilità tra dominio di addestramento e dominio reale. L’obiettivo principale resta infatti la ricostruzione tridimensionale a partire da immagini panoramiche, per la quale la CycleGAN fornisce un input adattato al modello generativo.

Dataset delle immagini panoramiche reali Le immagini panoramiche reali utilizzate per l’addestramento della CycleGAN provengono dal dataset della *Dentex Challenge 2023*, pubblicato su Kaggle². Il dataset contiene un’ampia collezione di radiografie panoramiche acquisite clinicamente, utilizzate per apprendere la mappatura dal dominio reale a quello sintetico. Poiché non è disponibile una corrispondenza esplicita con volumi CBCT, la traduzione di dominio avviene in maniera non supervisionata.

Prima dell’addestramento, le immagini PX reali sono state preprocessate mediante normalizzazione dell’intensità e ridimensionamento spaziale, al fine di renderle compatibili con le dimensioni delle SimPX derivate da CBCT.

Addestramento e Monitoraggio

Durante l’addestramento della CycleGAN, sono stati monitorati diversi parametri a ogni epoca, tra cui le loss dei discriminatori (D_X, D_Y), le loss dei generatori (G, F), la cycle consistency loss e la identity loss. Inoltre, è stato tracciato l’andamento del learning rate, inizialmente costante e successivamente decrescente in modo lineare.

Per garantire la stabilità del training, è fondamentale mantenere un *equilibrio* tra le varie componenti della funzione obiettivo: la adversarial loss deve essere bilanciata tra generatori e discriminatori per evitare il *mode collapse*, la cycle consistency loss deve restare contenuta per preservare la struttura delle immagini, e la identity loss deve risultare bassa, segno che non avvengono trasformazioni superflue.

Nel complesso, questo monitoraggio continuo ha permesso di mantenere una dinamica di apprendimento stabile ed efficace, favorendo una traduzione di dominio coerente e contribuendo al miglioramento della generalizzazione del modello anche su immagini PX reali, senza richiedere modifiche all’architettura del generatore principale.

Table 3.3: Valori delle perdite della CycleGAN all’epoca 200

Componente	Loss A	Loss B
Discriminatore (D)	0.151	0.194
Generatore (G)	0.476	0.406
Cycle consistency	0.266	0.309
Identità (Idt)	0.079	0.070

La Tabella 3.3 riporta le perdite principali osservate all’epoca 200 durante il training della rete CycleGAN. I valori ottenuti evidenziano una buona stabilità del modello, con un comportamento coerente sia nel ciclo diretto che inverso. In particolare, si osservano

²<https://www.kaggle.com/datasets/truthisneverlinear/dentex-challenge-2023>

errori di identità contenuti, una perdita di consistenza ciclica sotto la soglia critica $\approx 0,3$, e discriminatori ben bilanciati, il che indica un apprendimento efficace e regolare.

Una volta completato l’addestramento, la rete CycleGAN è stata utilizzata in modalità unidirezionale, sfruttando solo il generatore $G_{A \rightarrow B}$ per convertire le immagini panoramiche reali in immagini sintetiche nello stile *SntPX*, da utilizzare come input per la rete di ricostruzione 3D.

3.4 Ottimizzazione congiunta della curva e dei parametri di proiezione

Per perfezionare la ricostruzione tridimensionale a partire da una panoramica reale, viene introdotto un modulo di ottimizzazione iterativa che agisce su due livelli complementari: (1) la geometria della curva spline mandibolare e (2) i parametri del sistema di proiezione (come traslazione, rotazione e distanza fuoco-detettore). Questo processo consente di adattare dinamicamente la traiettoria di proiezione alle caratteristiche anatomiche del singolo paziente.

L’ottimizzazione è guidata da una funzione obiettivo che misura la similarità tra la panoramica reale e quella sintetizzata.

Il procedimento prevede la generazione iterativa di una CBCT ricostruita, ottenuta tramite interpolazione volumetrica lungo la curva corrente, e la relativa proiezione panoramica tramite ray tracing. La funzione di costo calcolata tra la *SntPX* e la *PX* reale guida l’aggiornamento simultaneo dei punti di controllo della curva spline e dei parametri geometrici della proiezione, utilizzando algoritmi di ottimizzazione numerica (es. *Levenberg-Marquardt, least squares*). Tale ottimizzazione congiunta consente di ottenere una ricostruzione tridimensionale anatomicamente plausibile e visivamente compatibile con l’immagine panoramica di input.

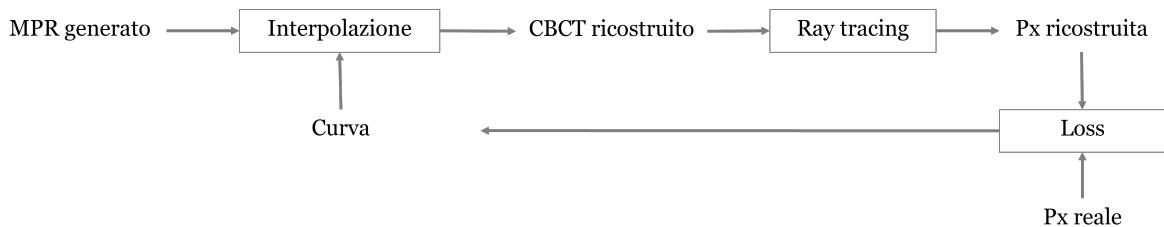


Figure 3.7: Schema a blocchi del ciclo di ottimizzazione iterativa della curva mandibolare. A partire da una curva iniziale, viene generata una CBCT tramite interpolazione, seguita da una proiezione panoramica mediante ray tracing. La *SntPX* risultante è confrontata con l’immagine reale per calcolare un errore che guida l’aggiornamento iterativo della curva fino alla convergenza.

Funzione d’errore: importanza della regolarità L’accuratezza della funzione d’errore è cruciale: deve fornire una discesa regolare e non soggetta a discontinuità, per guidare efficacemente l’ottimizzazione. Una funzione sufficientemente continua e robusta consente di evitare

oscillazioni o convergenze instabili, facilitando il raggiungimento di una soluzione anatomicamente plausibile. Questa necessità è ben documentata nella letteratura sull'ottimizzazione numerica, dove si sottolinea l'importanza di criteri obiettivo lisci e differenziabili per garantire stabilità e convergenza nei metodi iterativi [19].

Nel contesto della registrazione medica, metriche come la Mutual Information, la Normalized Cross-Correlation e la Structural Similarity Index Measure sono ampiamente adottate proprio per la loro capacità di offrire superfici di costo regolari e ottimizzabili [22, 28].

Viewpoint e Side Resizing Durante l'ottimizzazione, vengono considerati i parametri geometrici che determinano la configurazione del sistema di proiezione. In particolare:

- Il **viewpoint** rappresenta la posizione e l'orientamento del centro di proiezione, e determina l'angolazione da cui il fascio RX attraversa il volume (side view). La sua ottimizzazione consente di simulare una proiezione panoramica osservata dalla stessa prospettiva di acquisizione dell'immagine reale.
- Il **side resizing** corrisponde al fattore di scala applicato lateralmente alla proiezione, legato allo zoom e alla distanza focale. Questo parametro regola la larghezza apparente delle strutture nella panoramica simulata e permette di adattarne le proporzioni a quelle reali.

La combinazione di questi due fattori consente di generare immagini proiettive che riproducono fedelmente non solo la morfologia, ma anche la configurazione geometrica della scena originale.

Strategia di incremento dei punti di controllo

Per rendere l'ottimizzazione più stabile e computazionalmente sostenibile, la curva spline parametrica viene inizialmente definita da un numero ridotto di punti di controllo. Questo permette al sistema di agire su un numero limitato di parametri, semplificando la fase iniziale di convergenza. Una volta che l'errore si è ridotto e la forma della curva è sufficientemente vicina alla configurazione ottimale, è possibile aumentare la densità dei punti inserendone di nuovi tra quelli esistenti. Tale approccio progressivo, noto come *refinement adattivo*, permette di incrementare la risoluzione della curva solo laddove necessario, migliorando l'aderenza locale senza compromettere la stabilità dell'ottimizzazione globale. Inoltre, consente di assegnare più punti nei tratti ad alta curvatura, dove è richiesta una maggiore flessibilità, e meno dove la traiettoria è rettilinea.

Calcolo della curva media dei punti di controllo Per ottenere una curva di riferimento comune, è possibile calcolare la *curva media* a partire da un insieme di curve spline definite per ciascun paziente del dataset. Ogni curva è rappresentata da un insieme di M punti di controllo bidimensionali, cioè $\{(x_{i1}, y_{i1}), (x_{i2}, y_{i2}), \dots, (x_{iM}, y_{iM})\}$ per il paziente i -esimo, con $i = 1, \dots, N$.

La media dei punti di controllo viene calcolata punto per punto lungo la direzione dei pazienti. Per ciascun indice $j = 1, \dots, M$, si ottiene:

$$\bar{x}_j = \frac{1}{N} \sum_{i=1}^N x_{ij}, \quad \bar{y}_j = \frac{1}{N} \sum_{i=1}^N y_{ij} \quad (3.16)$$

Il risultato è un insieme di M punti medi $\{(\bar{x}_1, \bar{y}_1), (\bar{x}_2, \bar{y}_2), \dots, (\bar{x}_M, \bar{y}_M)\}$, che definisce la traiettoria media dell'arcata mandibolare all'interno del dataset.

Successivamente, su tali punti viene interpolata una curva spline (es. spline cubica naturale) per generare una rappresentazione continua e densa della curva, utile per applicazioni successive come la ricostruzione panoramica e il ray tracing simulato.

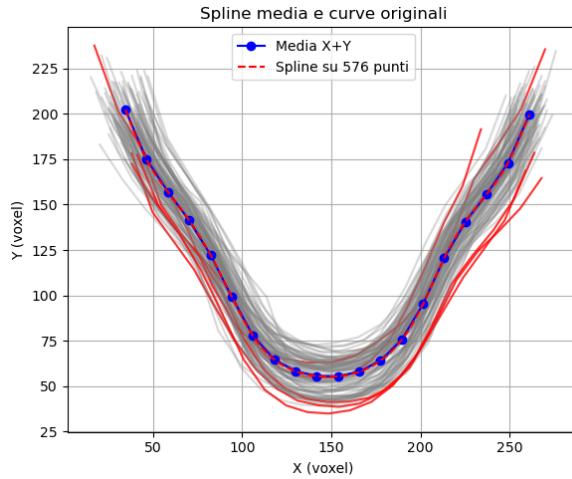


Figure 3.8: Curva di regressione media.

Chapter 4

Results

4.1 Introduzione

In questo capitolo vengono presentati i risultati sperimentali ottenuti durante lo sviluppo e la validazione del framework proposto per la ricostruzione tridimensionale a partire da un’immagine panoramica reale (*PX_realstyle*). Le valutazioni riguardano le singole fasi della pipeline, dalle generazioni iniziali prodotte dal modello generativo condizionato (cGAN), fino alla successiva fase di ottimizzazione della curva spline mandibolare, utilizzata per affinare la traiettoria di proiezione.

Vengono analizzati sia gli aspetti qualitativi (visualizzazione volumetrica, confronti tra proiezioni simulate e reali, allineamento delle strutture anatomiche) sia quelli quantitativi, tramite metriche numeriche (PSNR, SSIM, Dice) e curve di andamento delle funzioni di perdita durante l’addestramento e l’ottimizzazione.

Particolare attenzione è dedicata alla fase di deformazione della curva mandibolare, in cui una strategia di ottimizzazione iterativa (basata su confronto dei bordi o metrica di similarità) consente di adattare la proiezione simulata alla panoramica reale, migliorando l’allineamento geometrico e la coerenza anatomica.

4.2 Ricostruzione CBCT da PX tramite cGAN

Il primo stadio della pipeline prevede l’addestramento di una rete generativa condizionata (cGAN) con l’obiettivo di ricostruire un volume tridimensionale (*CBCT simulato*) a partire da una panoramica in ingresso. La rete è stata addestrata utilizzando coppie di immagini panoramiche sintetiche (*SimPX*) e volumi CBCT reali, in modo da apprendere la corrispondenza tra caratteristiche bidimensionali e strutture tridimensionali latenti.

4.2.1 Esperimenti

Per valutare le capacità della rete cGAN nella ricostruzione di volumi CBCT a partire da immagini panoramiche, Durante la fase di addestramento sono stati testati diversi approcci

per valutare l'impatto delle scelte progettuali: è stata condotta un'ampia serie di esperimenti. Le analisi hanno riguardato sia gli aspetti quantitativi che qualitativi, includendo:

- il monitoraggio delle curve di perdita durante l'addestramento;
- la valutazione delle prestazioni su un insieme di test, tramite metriche oggettive come PSNR, SSIM e Dice. Inoltre, per aggregare le metriche in un'unica misura sintetica, è stato definito un *Overall Score*, calcolato come media pesata normalizzata delle tre metriche:

$$\text{Overall Score} = \frac{1}{3} \left(\frac{\text{PSNR}}{20} + \frac{\text{SSIM}}{100} + \frac{\text{Dice}}{100} \right) \cdot 100$$

Tale formulazione assegna pari peso relativo alle tre metriche, considerando le rispettive scale di riferimento.

- il confronto visivo tra volumi generati e ground truth, per verificare la coerenza strutturale;
- l'analisi dell'impatto di diverse configurazioni di training, comprese tecniche di early stopping e strategie di adattamento del learning rate.

Le configurazioni finali sono state selezionate in base alle prestazioni ottenute sul set di validazione, privilegiando la stabilità della convergenza e la qualità della ricostruzione 3D.

Setup Sperimentale

Per tutti gli esperimenti presentati, è stato utilizzato un dataset composto da 100 pazienti, suddiviso in 60% per l'addestramento, 20% per la validazione e 20% per il test. Il batch size è stato fissato a 1, per contenere il carico computazionale dovuto alla natura tridimensionale dei volumi. L'addestramento è stato eseguito per 300 epoche, con validazione ogni 20 epoche.

Il modello è stato ottimizzato con l'algoritmo *Adam*, utilizzando un learning rate iniziale di 0,001. In alcuni esperimenti, sono stati testati scheduler del learning rate come *ReduceLROnPlateau* e *Cosine Annealing*, nonché strategie di *early stopping*, per migliorare la convergenza e prevenire l'overfitting.

Le sessioni di training sono state eseguite su GPU NVIDIA RTX 3090 (24 GB), in ambiente *PyTorch*, fissando un seed per garantire la riproducibilità dei risultati.

Scelta degli iperparametri

Durante la fase sperimentale sono stati eseguiti numerosi test per identificare la configurazione di iperparametri più efficace per l'addestramento del modello generativo. In particolare, sono stati esplorati i seguenti aspetti.

Learning rate

Come si osserva dalla Tabella, un learning rate iniziale più elevato (0,001) consente al modello di raggiungere prestazioni superiori in termini di PSNR, SSIM e Dice. Questo miglioramento può essere attribuito alla maggiore velocità di apprendimento nella fase iniziale, che risulta particolarmente efficace quando il generatore è pre-addestrato con una loss supervisionata, prima dell'introduzione del discriminatore. La scelta di un learning rate più basso (0,0002), sebbene più conservativa e in linea con molte architetture GAN presenti in letteratura [11, 18], comporta una convergenza più lenta e tende a produrre volumi meno coerenti, soprattutto in assenza di una forte componente avversaria nelle prime epoche. Questi risultati confermano che la combinazione di un pre-training supervisionato e un learning rate più alto favorisce un apprendimento più rapido ed efficace della struttura volumetrica, migliorando le prestazioni complessive del framework.

Configurazione	PSNR ↑	SSIM ↑	Dice ↑	Overall Score ↑
LR = 0.0002	19.81 ± 0.37	0.748 ± 1.70	0.5575 ± 7.33	76.53
LR = 0.001	19.64 ± 0.34	0.7495 ± 0.0162	0.5603 ± 0.0719	76.39

Table 4.1: Confronto tra diversi learning rate e relative metriche di valutazione sul test set.

Come mostrato nella Tabella 4.1, sono stati confrontati due diversi valori di learning rate iniziale: 0,0002 e 0,001. Entrambe le configurazioni producono risultati comparabili in termini di SSIM e PSNR, tuttavia il valore di Dice, che rappresenta la metrica di similarità volumetrica più sensibile alla sovrapposizione strutturale, è risultato più elevato per la configurazione con $LR = 0,001$ (0.5603 contro 0.5575).

Per tale motivo, il criterio di selezione è stato basato sul valore massimo della metrica Dice, ritenuta più indicativa della qualità volumetrica della ricostruzione. Di conseguenza, la configurazione con learning rate iniziale pari a 0,001 è stata adottata come impostazione definitiva per le successive fasi di addestramento e ottimizzazione.

Strategie di adattamento del learning rate

Durante la fase di addestramento, sono state sperimentate diverse strategie di modulazione del learning rate (LR), con l'obiettivo di migliorare la stabilità e la convergenza del modello generativo. Le principali tecniche analizzate includono:

1. Decadimento manuale. È stata adottata come configurazione di riferimento una strategia di *learning rate decay* manuale, con riduzioni a intervalli regolari. Il piano di schedulazione, riportato nella Tabella 4.2, prevede un decadimento ogni 50 epoche, fino a raggiungere un valore minimo di 10^{-6} :

$$LR = LR_{\text{iniziale}} \cdot 0,1^{\lfloor \frac{\text{epoca}}{50} \rfloor}$$

Intervallo di epoche	Formula	Learning Rate (lr)
0 – 49	$0.001 \cdot (0.1^0)$	0.001
50 – 99	$0.001 \cdot (0.1^1)$	0.0001
100 – 300	$0.001 \cdot (0.1^2)$	0.00001

Table 4.2: Schedulazione manuale del learning rate durante l’addestramento.

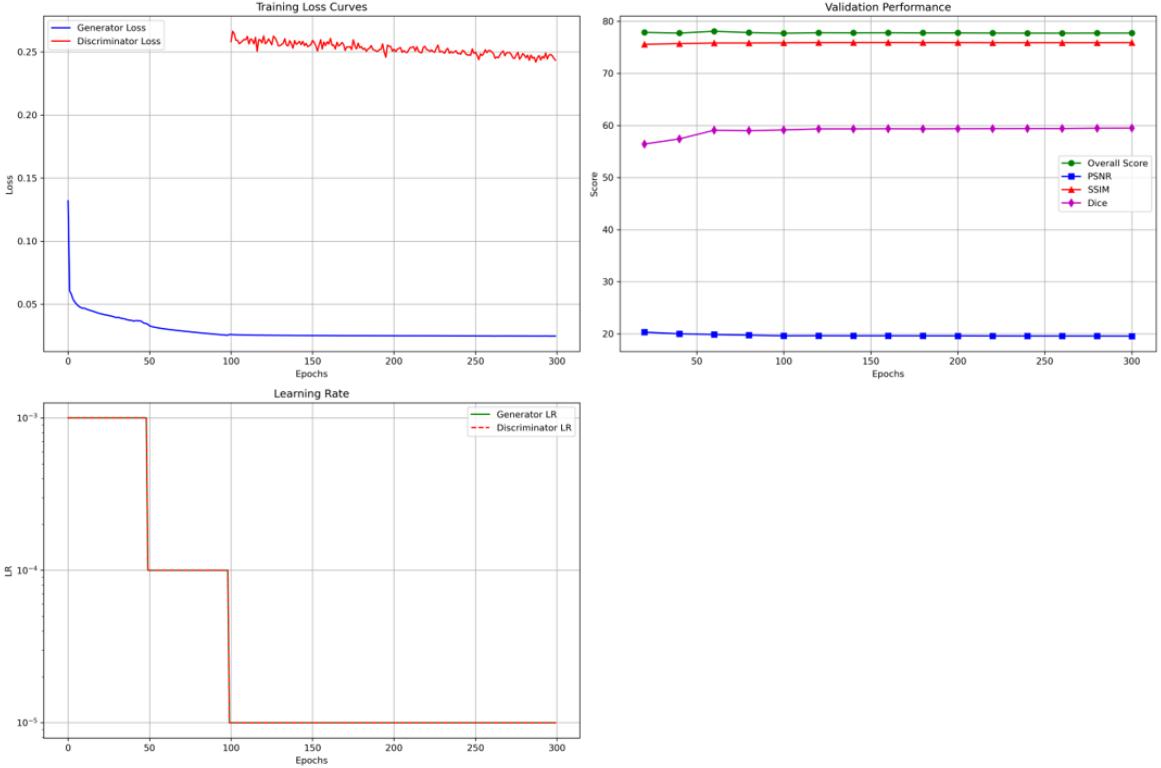


Figure 4.1: Perdite, metrice e andamento lr modalità manuale.

Questa modalità, sebbene non adattiva, consente un controllo esplicito sull’andamento del learning rate ed è risultata particolarmente stabile in combinazione con il pretraining supervisionato del generatore. Questa impostazione ha permesso di iniziare l’addestramento con un apprendimento più aggressivo, per poi raffinarlo progressivamente, riducendo il tasso di aggiornamento dei pesi man mano che il modello convergeva.

Scheduler automatici Oltre alla schedulazione manuale, sono stati sperimentati anche due scheduler automatici forniti da *PyTorch*, al fine di modulare dinamicamente il tasso di apprendimento durante l’addestramento:

2. ReduceLROnPlateau. È stato inoltre testato lo scheduler *ReduceLROnPlateau* di PyTorch, che adatta dinamicamente il learning rate in base al mancato miglioramento di una metrice di validazione. In questo caso, la metrice monitorata era il *Dice coefficient*, considerato particolarmente rilevante per valutare la coerenza morfologica tra il volume ricostruito e quello reale. I parametri utilizzati sono i seguenti:

- **Fattore di riduzione:** 0,5;

- **Pazienza:** 2 validazioni consecutive senza miglioramento;
- **Soglia di miglioramento:** 0,01;
- **LR minimo:** 10^{-6} .

Questa strategia ha dimostrato una buona capacità di adattamento, in particolare durante le fasi di plateau, contribuendo a stabilizzare il processo di apprendimento. Tuttavia, nel contesto specifico di questo studio, non ha portato a un miglioramento significativo delle metriche finali rispetto alla schedulazione manuale.

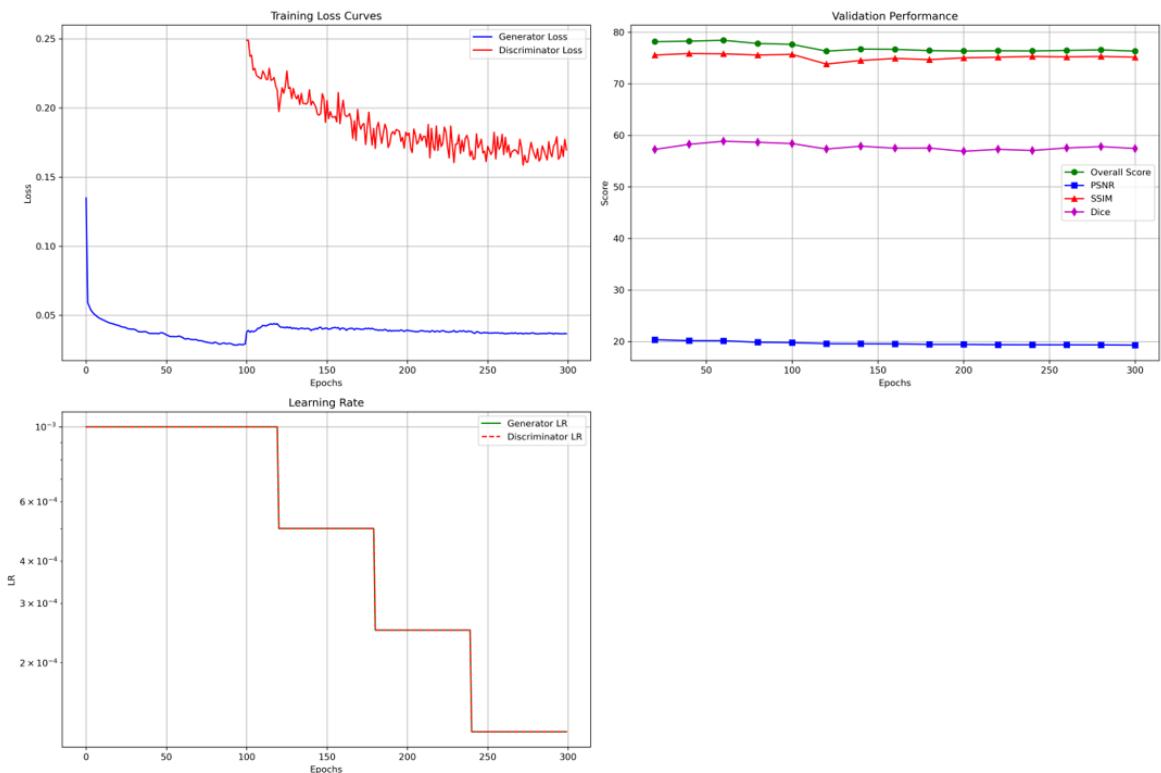


Figure 4.2: Perdite, metrice e andamento lr modalità plateau.

3. CosineAnnealingWarmRestarts. Infine, è stata valutata una strategia avanzata basata su decadimento cosenoide con riavvii. Questo scheduler introduce una dinamica oscillatoria del learning rate secondo una curva cosinusoidale, con cicli di "decadimento" e successiva "ripartenza" (*warm restart*), per incentivare la ricerca di configurazioni migliori nello spazio dei parametri ed evitare minimi locali. Ogni ciclo successivo ha durata doppia rispetto al precedente. La durata crescente dei cicli consente una progressiva focalizzazione sui minimi locali più promettenti, mitigando il rischio di convergenza prematura. I parametri adottati sono stati:

- **T₀:** 50 epoche (durata del primo ciclo);
- **T_{mult}:** 2 (ogni ciclo raddoppia in durata);

- **LR minimo:** 1×10^{-6} .

L’andamento ciclico del learning rate ha favorito una maggiore variabilità durante l’addestramento, ma non ha prodotto miglioramenti quantitativi significativi rispetto alla schedulazione manuale.

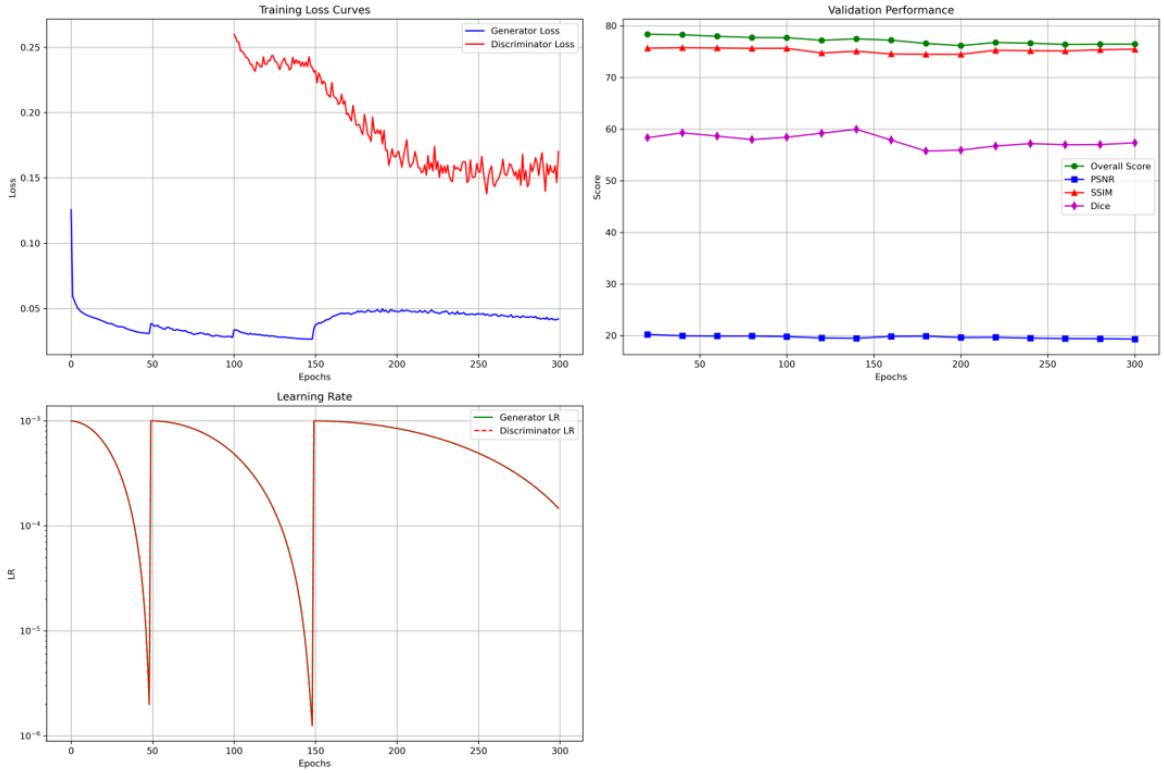


Figure 4.3: Perdite, metrice e andamento lr modalità cosine.

Conclusioni. La schedulazione manuale si è dimostrata particolarmente efficace e ben allineata al setup sperimentale adottato, garantendo risultati stabili, interpretabili e di buona qualità.

Sebbene gli scheduler automatici — come *ReduceLROnPlateau* e *CosineAnnealingWarmRestarts* — offrano una maggiore adattività e potenziale capacità di esplorazione dello spazio dei parametri, non sono emerse differenze significative in termini di metriche quantitative finali (PSNR, SSIM, Dice). Per questo motivo, la strategia manuale è stata mantenuta come impostazione predefinita nelle successive fasi di addestramento.

Configurazione	PSNR ↑	SSIM ↑	Dice ↑	Overall Score ↑
Plateau	19.38 ± 0.33	0.7422 ± 1.66	0.5436 ± 6.76	75.16
Cosine	19.44 ± 0.37	0.7467 ± 1.63	0.5427 ± 7.11	75.38
Manuale	19.64 ± 0.34	0.7495 ± 0.0162	0.5603 ± 0.0719	76.39

Table 4.3: Confronto delle prestazioni sul test set in funzione della strategia di scheduling del learning rate adottata

Nel complesso, l’efficacia di ciascun approccio dipende dalla natura del modello, dalla complessità del compito e dalla fase del processo di apprendimento. Per tale ragione, nel

presente lavoro sono state condotte sperimentazioni comparative, che hanno confermato la maggiore coerenza della schedulazione manuale nel contesto specifico della ricostruzione volumetrica da immagini panoramiche.

Si osservano comportamenti differenti nella modulazione del tasso di apprendimento:

Dal punto di vista delle prestazioni (metriche), non emergono differenze significative tra i tre approcci, ma la schedulazione manuale si è rivelata più stabile e coerente nel contesto specifico di questo studio.

Early Stopping

Durante l’addestramento è stato adottato un meccanismo di *early stopping*, con l’obiettivo di interrompere automaticamente il training qualora le prestazioni sul *validation set* non mostrino miglioramenti significativi per un certo numero di epoch consecutive. In particolare, è stata impostata una soglia di miglioramento minimo (*min_delta*) pari a 0,01, e una pazienza (*patience*) corrispondente a 2 validazioni consecutive senza miglioramento, equivalenti a 40 epoch.

La metrica monitorata per l’attivazione dell’early stopping è il Dice coefficient, calcolato tra il volume generato dal modello e il volume CBCT reale. Tale metrica è stata selezionata in quanto riflette la coerenza strutturale tridimensionale tra le due distribuzioni, risultando particolarmente indicativa della qualità della ricostruzione volumetrica. Un miglioramento viene riconosciuto solo se il valore di Dice supera di almeno 0,01 il massimo registrato fino a quel momento.

L’impiego di questa strategia ha consentito di ridurre i tempi complessivi di addestramento e di mitigare il rischio di *overfitting*, interrompendo l’ottimizzazione non appena le metriche di valutazione hanno mostrato una fase di stabilizzazione. L’early stopping non è stato attivato

Table 4.4: Riassunto dell’early stopping durante l’addestramento del modello

Parametro	Valore
Epoca di arresto (<i>early stopping</i>)	180
Soglia di miglioramento (<i>min_delta</i>)	0.01
Pazienza (<i>patience</i>)	2 validazioni consecutive
Valore massimo Dice	0.561
Valore massimo SSIM (opzionale)	0.749
Tempo di addestramento (hh:mm)	00:53

sin dall’inizio del training, ma solo dopo l’introduzione del discriminatore (epoca 100), per evitare di interferire con la fase di pre-addestramento supervisionato del generatore. La soglia di miglioramento minimo è stata fissata a 0.01 sulla metrica Dice, con valutazione ogni 20 epoch. Il training si è concluso automaticamente all’epoca 180.

Questo esperimento ha dimostrato l’efficacia dell’impiego dell’early stopping. Sebbene l’intero processo di addestramento del modello generativo possa richiedere in media circa 1 ora e 30 minuti, il criterio di arresto anticipato ha permesso di ottenere risultati comparabili in circa 53 minuti, riducendo sensibilmente i tempi computazionali. Questo conferma che,

in presenza di una metrica di validazione stabile, l'early stopping consente di risparmiare risorse senza compromettere la qualità della ricostruzione.

Valutazione Qualitativa

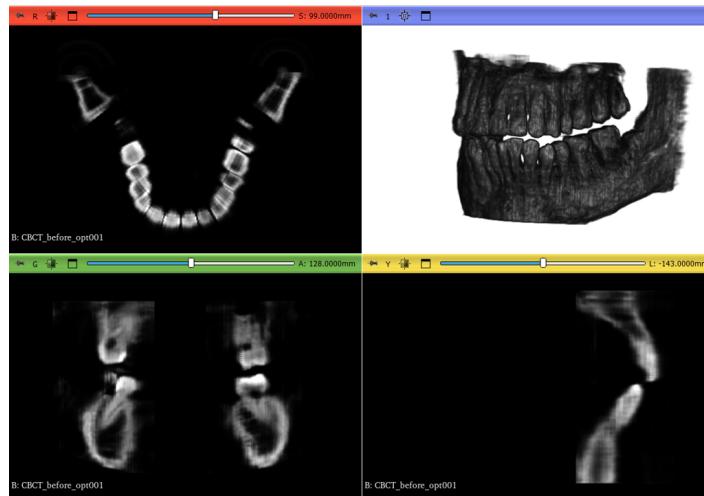


Figure 4.4: Visualizzazione volumetrica del CBCT simulato prima dell'ottimizzazione. Sono mostrati i tre piani ortogonali (assiale in alto a sinistra, coronale in basso a sinistra, sagittale in basso a destra) e la ricostruzione 3D globale (in alto a destra).



Figure 4.5: Visualizzazione volumetrica del CBCT simulato prima dell'ottimizzazione. Sono mostrati i tre piani ortogonali (assiale in alto a sinistra, coronale in basso a sinistra, sagittale in basso a destra) e la ricostruzione 3D globale (in alto a destra).

La Figura 4.4 mostra la visualizzazione del volume CBCT simulato generato dal modello prima della fase di ottimizzazione della curva mandibolare. Sono visibili i tre piani ortogonali (assiale, coronale, sagittale), che consentono di analizzare nel dettaglio la distribuzione della densità volumetrica e la morfologia dentale ricostruita. La vista tridimensionale (in alto a destra) fornisce un'indicazione globale della qualità del volume generato, utile per confronti visivi con la ground truth. A tale proposito, la Figura mostra un confronto diretto tra il volume CBCT reale e quello ricostruito generato dal modello cGAN. Si osserva una buona coerenza

strutturale, in particolare nei piani assiale e coronale, dove le arcate dentarie risultano ben delineate. La vista 3D evidenzia il mantenimento della morfologia mandibolare complessiva. Tuttavia, alcune discrepanze locali possono ancora emergere in regioni ad alta variabilità anatomica, indicando margini di miglioramento soprattutto nelle zone posteriori o nelle aree a bassa intensità. Le ricostruzioni evidenziano la capacità del modello di preservare le strutture ossee principali, con un buon grado di dettaglio anatomico e continuità volumetrica.

4.3 Ottimizzazione con Trasformazioni sulla Curva e Ray Tracing

Il processo di ottimizzazione implementato si articola attraverso una sequenza modulare in cui agiscono due distinti blocchi di trasformazione: uno applicato alla curva spline parametrica e uno ai parametri geometrici del ray tracing. Il diagramma a blocchi in Figura 4.6 schematizza l'intera pipeline.

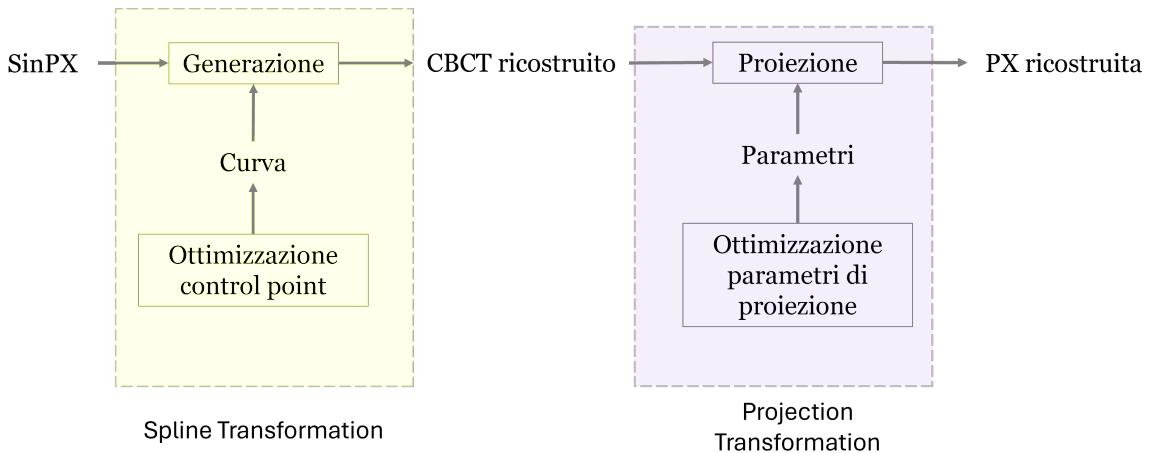


Figure 4.6: Pipeline ottimizzazione.

Si parte da una panoramica reale ($PX_realstyle$), che rappresenta l'immagine di riferimento. In parallelo, una curva spline iniziale viene elaborata da un modulo $\mathcal{T}_{\text{spline}}(\cdot)$ che modifica i punti di controllo della curva. La curva risultante viene utilizzata da un modulo \mathcal{R} per interpolare un volume tridimensionale simulato ($CBCT_final$) a partire dal volume MPR generato dalla rete.

Successivamente, il volume $CBCT_final$ viene elaborato dal modulo di ray tracing (RT), che genera una proiezione panoramica simulata (PX_sim). Prima della generazione della proiezione, vengono applicate trasformazioni geometriche affini, indicate come $\mathcal{T}_{\text{ray}}(\cdot)$, che modificano parametri quali traslazione, rotazione, scala e distanza di proiezione.

La proiezione simulata PX_sim viene quindi confrontata con $PX_realstyle$ tramite una funzione di dissimilarità (ad esempio DSSIM), il cui valore guida l'ottimizzazione.

In questo framework, l'ottimizzazione è congiunta: i parametri della curva spline e quelli del ray tracing vengono ottimizzati simultaneamente all'interno di un unico ciclo iterativo,

con l’obiettivo comune di minimizzare la distanza tra l’immagine simulata e quella reale.

Pur trattandosi di un’ottimizzazione congiunta, è mantenuta una distinzione concettuale tra le due famiglie di trasformazioni. Le modifiche alla curva spline modellano la traiettoria di campionamento anatomica, mentre le trasformazioni geometriche del ray tracing regolano il sistema di proiezione. Questa distinzione è fondamentale per garantire stabilità numerica, interpretabilità del processo e per evitare compensazioni non desiderate tra i due domini, che potrebbero ostacolare la convergenza.

4.3.1 Metrica di costo

La **Structural Dissimilarity Index** (DSSIM) è una metrica utilizzata per quantificare la differenza strutturale tra due immagini, ed è derivata dalla metrica SSIM (Structural Similarity Index Measure)[30]. Mentre la SSIM misura la somiglianza tra due immagini x e y , producendo un valore compreso tra -1 e 1 (dove 1 indica immagini identiche), la DSSIM rappresenta invece una funzione di *costo*, da minimizzare nei problemi di ottimizzazione.

La DSSIM è definita come:

$$\text{DSSIM}(x, y) = \frac{1 - \text{SSIM}(x, y)}{2}$$

Questa trasformazione consente di interpretare la dissimilarità come una distanza, ottenendo un valore compreso tra 0 (massima somiglianza) e 1 (massima dissimilarità).

La metrica SSIM si basa sul confronto tra tre aspetti principali delle immagini: la luminanza (cioè la media dei livelli di intensità), il contrasto (la deviazione standard) e la struttura (la correlazione tra i segnali). Combinando queste informazioni, la SSIM – e di conseguenza la DSSIM – riesce a cogliere differenze strutturali che sono percettivamente più rilevanti rispetto a semplici metriche basate su differenze di pixel, come MSE o L1. Questo rende la DSSIM particolarmente adatta nel contesto dell’imaging medico, dove è essenziale preservare i dettagli anatomici anche durante l’ottimizzazione.

Nel contesto dell’imaging medico, la metrica DSSIM viene spesso preferita alla SSIM diretta nei processi di ottimizzazione, in quanto rappresenta una funzione di costo da minimizzare [10]. La sua maggiore sensibilità a variazioni strutturali consente una guida più efficace durante l’adattamento dei parametri.

Nel presente lavoro, la DSSIM è stata utilizzata per guidare l’adattamento dei punti di controllo della curva spline e dei parametri di proiezione, con l’obiettivo di allineare la panoramica simulata alla struttura della panoramica reale.

4.3.2 Algoritmo di ottimizzazione

L’algoritmo di ottimizzazione implementato è di tipo iterativo e si basa sul principio della discesa del gradiente, con l’obiettivo di minimizzare la dissimilarità strutturale (DSSIM) tra la PX proiettata da CBCT ricostruita e quella reale. A ogni iterazione, i parametri liberi del

sistema (punti di controllo della curva spline e parametri geometrici del ray tracing) vengono aggiornati per ridurre il valore della funzione obiettivo.

In particolare, è stato adottato l'algoritmo *Trust Region Reflective (TRF)*, un metodo robusto per l'ottimizzazione non lineare vincolata. Il TRF, introdotto da Coleman e Li [6], si basa su una strategia a regioni di fiducia e integra trasformazioni riflesse per garantire il rispetto dei vincoli sui parametri. Questo approccio consente all'algoritmo di esplorare lo spazio delle soluzioni in modo controllato, mantenendo le variabili all'interno di intervalli fisicamente plausibili. Ciò lo rende particolarmente adatto a scenari come quello affrontato nel presente lavoro, in cui sia i punti di controllo della curva sia i parametri del ray tracing sono soggetti a vincoli.

L'algoritmo opera secondo i seguenti passi:

1. A partire dai punti di controllo iniziali della curva spline (*AvgShape*) e dai parametri iniziali del ray tracing (traslazione, rotazione, scala, distanza), si genera un volume CBCT simulato tramite interpolazione del volume MPR lungo la traiettoria spline.
2. Si calcola la proiezione panoramica (*SntPX*) mediante il modulo di ray tracing, che applica le trasformazioni geometriche correnti al sistema di proiezione.
3. Si confronta l'immagine *SntPX* con la panoramica reale (*PX_realstyle*), utilizzando come metrica di dissimilarità la DSSIM.
4. I parametri vengono aggiornati iterativamente mediante un approccio basato su regioni di fiducia (*Trust Region Reflective*), che stima numericamente il gradiente della funzione di costo e ne guida la discesa, garantendo il rispetto dei vincoli imposti.
5. Il processo si ripete fino al soddisfacimento di una delle condizioni di arresto: (i) una variazione trascurabile della funzione obiettivo tra iterazioni consecutive (*ftol1*), (ii) un gradiente sufficientemente piccolo (*gtol*), oppure (iii) il raggiungimento del numero massimo di valutazioni della funzione (*max_nfev*). In questo lavoro, tali soglie sono state impostate rispettivamente a 10^{-10} , 10^{-10} e 800.

Durante l'ottimizzazione, è stato introdotto un meccanismo di logging e visualizzazione che consente di monitorare l'andamento della curva spline, dell'immagine sintetizzata e del valore della funzione di costo a ogni step. Tale tracciamento è utile sia per l'analisi qualitativa, sia per diagnosticare eventuali stagnazioni o instabilità del processo di ottimizzazione.

Separazione dei parametri. Nel presente lavoro, tutti i parametri vengono ottimizzati congiuntamente all'interno di un unico processo. Tuttavia, si mantiene una distinzione concettuale tra due famiglie di variabili: da un lato i parametri associati alla curva spline (posizione dei punti di controllo), dall'altro quelli geometrici del ray tracing (traslazione, rotazione, scala, distanza di proiezione). Ciascun gruppo ha un ruolo funzionale distinto nella pipeline: la curva determina la morfologia del volume ricostruito, mentre la geometria di proiezione ne regola la modalità di visualizzazione.

Tale distinzione è utile per analizzare in modo separato l'effetto di ciascun blocco sulla funzione di costo, e consente un controllo più mirato del comportamento dell'ottimizzazione. Inoltre, permette di evitare fenomeni indesiderati come compensazioni tra trasformazioni (es. una traslazione della curva annullata da una traslazione inversa della proiezione), che potrebbero ostacolare la convergenza o mascherare la reale origine dell'errore.

Sebbene non sia stata adottata una separazione computazionale esplicita, questa suddivisione concettuale potrà essere sfruttata in futuro per progettare strategie di ottimizzazione modulare, basate su aggiornamenti alternati o vincoli specifici per ciascun sottoinsieme di parametri.

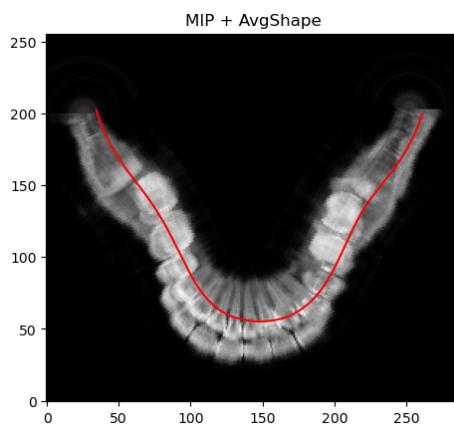


Figure 4.7: Proiezione MIP del volume CBCT ricostruito, con sovrapposta la curva mandibolare iniziale (*AvgShape*) usata come punto di partenza per l'ottimizzazione.

Parametri della curva spline Una volta generato il volume MPR tramite rete generativa, l'ottimizzazione mira a modificare la curva iniziale per adattarla alla morfologia effettiva del soggetto. L'obiettivo è ottenere una proiezione panoramica (SntPX) che risulti quanto più possibile coerente con l'immagine reale di riferimento.

La traiettoria lungo cui viene generata la panoramica è definita da una curva spline bidimensionale che approssima la forma dell'arco mandibolare. Tale curva è rappresentata da un insieme di punti di controllo $\mathbf{c} = \{(x_i, y_i)\}_{i=1}^N$, distribuiti su un piano assiale del volume CBCT.

Questi punti di controllo costituiscono i parametri geometrici della curva e vengono trattati come variabili ottimizzabili. L'ottimizzazione agisce sulle coordinate (x_i, y_i) di ciascun punto, con l'obiettivo di deformare la curva in modo da migliorare la coerenza tra la PX soggetta a trasformazioni e la PX reale.

Dal punto di vista formale, la stima della curva ottimale corrisponde alla minimizzazione della funzione di dissimilarità:

$$\min_{\mathbf{c}} \text{DSSIM}(\text{PX}_{\text{real}}, \text{PX}_{\text{sim}}(\mathbf{c})) \quad (4.1)$$

Ogni iterazione del processo di ottimizzazione modifica la posizione dei punti di controllo e genera una nuova curva spline, che viene utilizzata per estrarre un nuovo volume interpolato

e produrre la corrispondente proiezione panoramica.

Parametri geometrici del ray tracing Nel processo di proiezione panoramica, la geometria della scena radiografica viene definita da un insieme di parametri che controllano la posizione relativa tra la sorgente dei raggi X, il volume CBCT e il piano di proiezione. Questi parametri sono cruciali per ottenere una proiezione coerente con l'immagine reale e sono quindi inclusi nel processo di ottimizzazione.

I principali parametri geometrici ottimizzati sono:

- **Traslazione (t_x, t_y):** rappresenta uno spostamento rigido del volume CBCT nel piano ortogonale alla direzione di proiezione. Serve a centrare meglio l'area di interesse, in particolare l'arco mandibolare, rispetto alla traiettoria dei raggi. Viene espressa in voxel o millimetri e consente di correggere disallineamenti spaziali.
- **Rotazione (θ):** definisce una rotazione del volume CBCT rispetto all'asse ortogonale al piano panoramico. È utile per compensare inclinazioni anatomiche o errori di allineamento tra il volume ricostruito e la posizione attesa della mandibola nella proiezione reale. La rotazione è espressa in radianti o gradi.
- **Scala (zoom):** rappresenta un fattore di ingrandimento isotropo applicato al volume. Agisce sulla dimensione apparente della mandibola nella proiezione e consente di adattare la scala spaziale della simulazione a quella dell'immagine reale. È un parametro particolarmente utile in presenza di variazioni anatomiche o errori di calibrazione.
- **Distanza di proiezione (SDD):** indica la distanza tra la sorgente dei raggi e il piano del rivelatore panoramico. Questo parametro influisce sulla divergenza del fascio di raggi e sull'ingrandimento dell'immagine proiettata. Una distanza maggiore tende a produrre proiezioni più “parallele”, mentre una distanza minore accentua gli effetti prospettici.

Tutte le trasformazioni geometriche (t_x, t_y, θ, s, d) non agiscono direttamente sul volume CBCT, che rimane fisso nello spazio, ma modificano la geometria del sistema di proiezione, ovvero la posizione e l'orientamento della sorgente e della traiettoria dei raggi. In questo modo, si simulano rotazioni, traslazioni e variazioni di scala della sorgente RX rispetto al volume, riproducendo il comportamento di un sistema di acquisizione reale.

Tutti questi parametri sono trattati come variabili continue e ottimizzate congiuntamente ai punti di controllo della curva spline, all'interno della funzione di costo basata sulla dissimilarità tra la proiezione da CBCT ricostruita e quella reale. La stima dei parametri geometrici è formalizzata come un problema di ottimizzazione continua, in cui si cerca la configurazione ottimale (t_x, t_y, θ, s, d) che minimizza la dissimilarità strutturale tra le due PX:

$$\min_{t_x, t_y, \vartheta, s, d} \text{DSSIM}(\text{PX}_{\text{real}}, \text{PX}_{\text{sim}}(t_x, t_y, \vartheta, s, d)) \quad (4.2)$$

dove t_x, t_y rappresentano le traslazioni nel piano, ϑ la rotazione, s il fattore di scala e d la distanza di proiezione (SDD). La funzione $\text{PX}_{\text{sim}}(\cdot)$ indica la panoramica ottenuta tramite ray tracing, generata a partire dal volume ricostruito.

Vincoli sui parametri di proiezione. Per garantire stabilità numerica e coerenza fisica, sono stati imposti vincoli esplicativi sui parametri geometrici del ray tracing. In particolare, le traslazioni (t_x, t_y) sono limitate all'interno del campo visivo del volume CBCT, la rotazione ϑ è vincolata entro un intervallo simmetrico attorno allo zero, mentre la scala s e la distanza di proiezione d devono rimanere positive e comprese in un range fisiologicamente plausibile. Questi limiti impediscono che il sistema di proiezione assuma configurazioni non realistiche e vengono implementati direttamente all'interno del solver `least_squares` della libreria SciPy, mediante l'uso dei parametri `bounds`.

Vincoli morfologici sulla curva spline. Per evitare deformazioni innaturali della mandibola, è stata introdotta una regolarizzazione spaziale sui punti di controllo della curva spline. In particolare, il loro spostamento è limitato all'interno di una regione morfologicamente plausibile, coerente con la forma attesa dell'arco dentale. Questa regolarizzazione agisce come una barriera geometrica che impedisce alla curva di assumere configurazioni irrealistiche, migliorando al contempo la stabilità del processo di ottimizzazione.

Inizializzazione. La curva spline viene inizializzata a partire da una forma media (*AvgShape*), ottenuta mediante regressione statistica su un dataset di training. Per incoraggiare l'esplorazione dello spazio dei parametri ed evitare la convergenza verso minimi locali subottimali, nelle fasi iniziali dell'ottimizzazione è stato aggiunto un rumore gaussiano controllato ai punti di controllo iniziali. Inoltre, anche i parametri di proiezione (traslazione, rotazione e scala) sono stati inizializzati con valori leggermente perturbati rispetto a una configurazione neutra, al fine di simulare realistiche condizioni di disallineamento. Questo duplice disturbo iniziale consente al processo di esplorare varianti della forma media e della geometria proiettiva in modo più efficace.

Valutazione delle metriche di costo e risultati quantitativi. Sebbene la metrifica principale utilizzata per guidare l'ottimizzazione sia la Structural Dissimilarity Index (DSSIM), sono stati condotti anche esperimenti comparativi utilizzando la Normalized Cross Correlation (NCC) e la Mutual Information (MI). Per adattarle a un contesto di minimizzazione, entrambe sono state convertite in metriche di dissimilarità mediante l'opposto numerico (-NCC, -MI), come già proposto in precedenti studi di image registration (Mattes et al., 2001; Klein et al., 2009). Tuttavia, rispetto a NCC e MI, la DSSIM ha mostrato una maggiore sensibilità alle variazioni morfologiche nella regione mandibolare e una discesa del gradiente più stabile,

rendendola più efficace nel guidare la convergenza verso configurazioni visivamente coerenti. Al contrario, NCC e MI hanno evidenziato oscillazioni più marcate e una minore capacità di guidare l'ottimizzazione in modo robusto.

Per quantificare l'efficacia dell'ottimizzazione congiunta della curva spline e dei parametri geometrici di proiezione, sono state calcolate le metriche SSIM tra (i) la panoramica reale e quella ricostruita (S_{nPX}), e (ii) il volume CBCT reale e quello ottenuto dalla curva spline, sia prima che dopo l'ottimizzazione. I risultati, riportati in Tabella 4.5, evidenziano un miglioramento significativo nella proiezione panoramica, che rappresenta il target principale dell'ottimizzazione. Il miglioramento nella somiglianza volumetrica è più contenuto, ma risulta coerente con il fatto che il volume CBCT viene ricostruito indirettamente a partire da immagini MPR e dalla traiettoria ottimizzata.

Oltre all'ottimizzazione congiunta di tutti i parametri – includendo sia i punti di controllo della curva spline che i parametri geometrici del ray tracing – è stata condotta un'analisi comparativa mirata, volta a esplorare le prestazioni di diverse funzioni obiettivo e algoritmi di ottimizzazione. In particolare, sono state testate varie combinazioni di metriche di similarità e ottimizzatori, effettuando una valutazione quantitativa in ciascun caso.

Questa analisi ha permesso di identificare la combinazione più efficace in termini di coerenza visiva e metrica rispetto all'immagine reale. Nello specifico, la coppia SSIM + least squares ha prodotto i risultati più stabili e soddisfacenti, rappresentando quindi un riferimento per le successive fasi sperimentali.

Table 4.5: Confronto SSIM panoramica prima e dopo l'ottimizzazione, per ciascuna metrica.

Metrica di costo	SSIM Iniziale	SSIM Finale
DSSIM	0.5344	0.792
NCC	0.5344	0.6123
MI	0.5344	0.5841

Per garantire una comparabilità equa tra i volumi pre e post ottimizzazione, è stata applicata una normalizzazione lineare dei voxel, rimappandoli nell'intervallo [0, 1]. Questo assicura che i confronti visivi e le metriche come la SSIM non siano influenzati da differenze di range di intensità.

Evoluzione della funzione obiettivo. In Figura 4.8 è mostrato l'andamento della funzione di costo DSSIM durante le iterazioni dell'ottimizzazione. Si osserva una riduzione a gradini del valore medio di dissimilarità, con fasi di plateau seguite da improvvisi miglioramenti, alternate a oscillazioni locali. La curva smussata (in arancione) rappresenta una media mobile a 5 punti, utile per evidenziare la tendenza generale.

Tali oscillazioni sono attribuibili alla complessità della funzione obiettivo e all'interazione non lineare tra i parametri di controllo (punti della curva spline) e quelli geometrici del ray tracing. Nonostante ciò, l'andamento globale mostra una chiara convergenza verso una configurazione ottimizzata che massimizza la similarità strutturale con la panoramica reale.

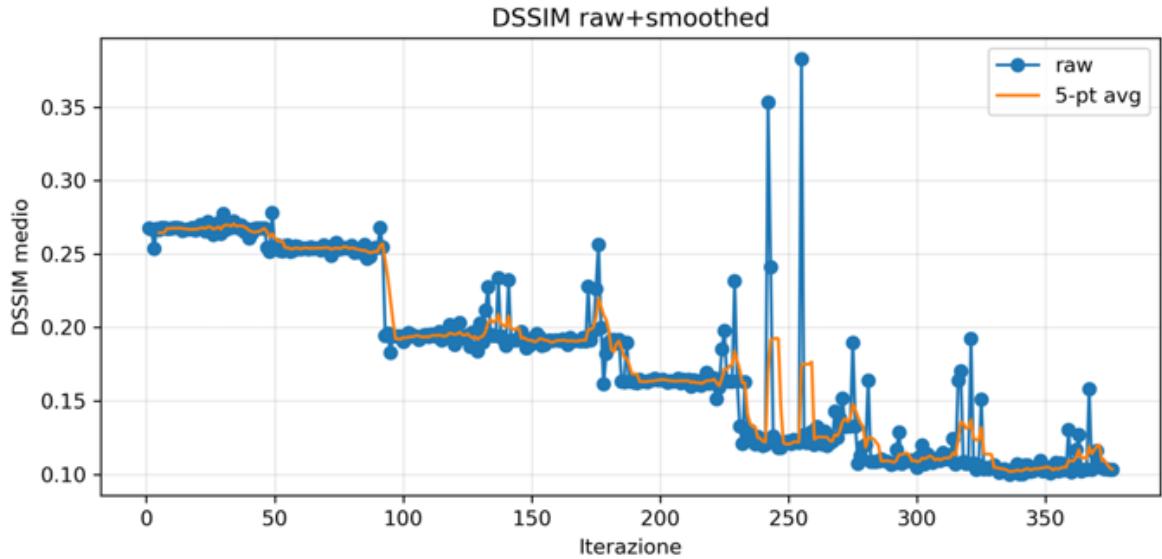


Figure 4.8: Andamento della funzione di costo DSSIM durante le iterazioni dell’ottimizzazione.

Confronto tra curve spline. La Figura 4.9 mostra il confronto tra tre configurazioni della curva spline mandibolare: la curva media (*AvgShape*), ottenuta tramite regressione sui dati di training; la curva iniziale perturbata, ottenuta aggiungendo rumore gaussiano casuale alla curva media ($\sigma = 3.0$ px); e la curva finale ottimizzata, risultante dal processo iterativo di minimizzazione della dissimilarità tra la panoramica non ottimizzata e quella reale.

L’introduzione di una perturbazione iniziale consente di testare la robustezza dell’algoritmo rispetto a condizioni iniziali non ideali. Si osserva come, a partire da una curva disturbata, l’ottimizzazione riesca a riconvergere verso una configurazione più coerente con l’anatomia mandibolare reale, migliorando la qualità della proiezione ottimizzata.

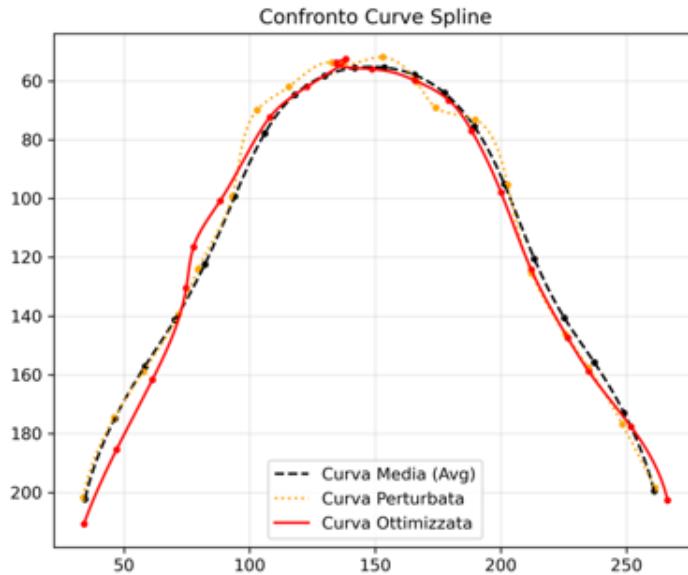


Figure 4.9: Confronto tra le curve spline mandibolari durante l’ottimizzazione.

Confronto qualitativo delle proiezioni panoramiche. In Figura 4.10 è mostrato un confronto tra tre immagini: la panoramica reale di riferimento, la proiezione iniziale ottenuta a partire da una curva spline leggermente modificata rispetto alla media, e la proiezione finale risultante dal processo di ottimizzazione. Si osserva un miglioramento qualitativo evidente, in particolare nella coerenza delle strutture mandibolari e nella riduzione di artefatti morfologici. L'ottimizzazione consente una migliore sovrapposizione tra le regioni di interesse, portando a un incremento del valore SSIM da 0,453 a 0,792. Questo risultato conferma l'efficacia dell'ottimizzazione congiunta dei punti di controllo della curva spline e dei parametri geometrici del ray tracing.

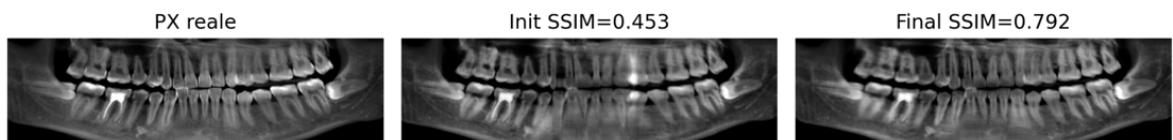


Figure 4.10: Confronto qualitativo tra panoramica reale (sinistra), proiezione con parametri iniziali rumorosi e proiezione finale dopo l'ottimizzazione. Si nota un miglioramento visibile nella coerenza morfologica e nella qualità dell'immagine, con un incremento del valore SSIM da 0,453 a 0,792.

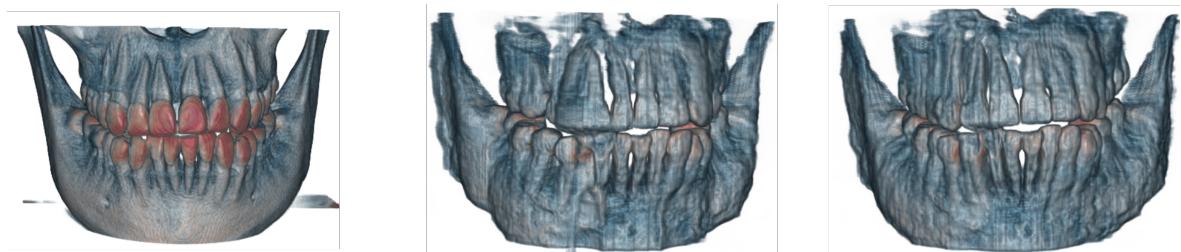


Figure 4.11: Confronto tra: (sinistra) volume CBCT reale, (centro) volume ricostruito prima dell'ottimizzazione, (destra) volume ricostruito dopo l'ottimizzazione.

Confronto delle ricostruzioni CBCT prima e dopo l'ottimizzazione. Per valutare l'efficacia del processo di ottimizzazione, è stato effettuato un confronto tra le ricostruzioni volumetriche CBCT ottenute con parametri iniziali, non ottimizzati o perturbati, e quelle ottenute dopo l'ottimizzazione. In particolare, la SSIM volumetrica è passata da 0,69 a 0,71, dimostrando un miglioramento nella qualità della ricostruzione.

L'immagine iniziale, ottenuta con parametri non ottimizzati, presenta alcune discrepanze rispetto alla morfologia reale, mentre la ricostruzione finale, ottenuta dopo l'ottimizzazione, mostra un'aderenza molto più elevata ai contorni anatomici reali. Questo confronto visivo evidenzia in modo chiaro l'impatto positivo dell'ottimizzazione sulla qualità del volume ricostruito.

Valutazione sul test set. L'ottimizzazione dei parametri della curva spline e del sistema di proiezione è stata applicata al test set, con l'obiettivo di valutare le prestazioni del framework

in uno scenario realistico, svincolato dai dati di addestramento. Tale scelta consente di misurare in modo oggettivo il miglioramento apportato dall'ottimizzazione, sia a livello di proiezione panoramica che di volume tridimensionale.

Per simulare una condizione iniziale sub-ottimale ma plausibile, i parametri iniziali sono stati perturbati: ai punti di controllo della curva spline è stato aggiunto rumore gaussiano con deviazione standard pari a 5 pixel, mentre i parametri geometrici del sistema di proiezione (rotazione, traslazione, distanza fuoco-detettore) sono stati alterati casualmente entro un intervallo del ± 10

In particolare, la metrica SSIM è stata calcolata prima e dopo l'ottimizzazione, sia tra le immagini panoramiche (SntPX vs PX reale) sia tra i volumi CBCT (volume ricostruito vs volume reale), al fine di quantificare l'efficacia del processo. In particolare, la metrica SSIM è stata calcolata prima e dopo l'ottimizzazione, sia tra le immagini panoramiche (SntPX vs PX reale) sia tra i volumi CBCT (volume ricostruito vs volume reale), al fine di quantificare l'efficacia del processo.

Poiché il generatore della rete neurale utilizza una funzione di attivazione \tanh , l'output volumetrico è inizialmente compreso nell'intervallo $[-1, 1]$. Per garantire un confronto coerente con i volumi CBCT reali, tale output è stato riscalato nell'intervallo $[0, 1]$ prima di essere utilizzato per l'interpolazione e la valutazione della metrica SSIM.

I risultati ottenuti sono riportati in Tabella 4.6 e confermano un incremento significativo sia nella qualità della proiezione che nella ricostruzione volumetrica, in linea con l'obiettivo principale del framework: ottenere un volume 3D coerente a partire da una panoramica dentale.

I valori riportati in tabella rappresentano la media e la deviazione standard calcolate su un test set composto da 20 pazienti, selezionati in modo da garantire variabilità anatomica e robustezza della valutazione. L'incremento osservato nella SSIM panoramica conferma

Table 4.6: Risultati medi su test set prima e dopo l'ottimizzazione

Metrica	Pre-Ottimizzazione	Post-Ottimizzazione
SSIM panoramica	0.48 ± 0.04	0.82 ± 0.03
SSIM volumetrica	0.68 ± 0.05	0.70 ± 0.04

l'efficacia del processo di ottimizzazione nel migliorare la coerenza tra la proiezione simulata e quella reale. Questo risultato è coerente con l'obiettivo del framework, che mira a ottenere una ricostruzione tridimensionale strutturalmente fedele a partire da una singola immagine panoramica.

Chapter 5

Discussione

Riflessioni sull'ottimizzazione congiunta

Una delle considerazioni emerse durante la progettazione del sistema riguarda la distinzione concettuale tra due categorie di trasformazioni: da un lato, le trasformazioni rigide globali (traslazione, rotazione, zoom), che agiscono sulla geometria del ray tracing modificando il centro di proiezione e l'orientamento del fascio di raggi; dall'altro, le deformazioni locali della traiettoria mandibolare, modellate tramite la modifica dei punti di controllo di una curva spline.

Sebbene inizialmente si fosse ipotizzata una separazione dei due livelli di ottimizzazione per semplificare l'analisi, l'implementazione finale ha adottato un approccio congiunto, in cui tutti i parametri — locali e globali — vengono aggiornati simultaneamente. Questa scelta ha dimostrato una maggiore efficacia, poiché consente di sfruttare in modo sinergico l'interazione tra deformazioni locali e disallineamenti globali. Ad esempio, piccole imprecisioni della curva possono essere compensate da micro-aggiustamenti del centro di proiezione, e viceversa.

In sintesi, l'ottimizzazione congiunta ha favorito una convergenza più rapida e una migliore aderenza tra la proiezione sintetizzata e quella reale, confermando la validità dell'approccio integrato rispetto a una strategia a fasi separate.

Nonostante l'ottimizzazione porti a un evidente miglioramento nella qualità della proiezione panoramica, l'incremento nella somiglianza volumetrica è più contenuto. In particolare, la SSIM volumetrica passa da 0,700 a 0,703 dopo l'ottimizzazione, indicando un incremento marginale. Questo risultato è tuttavia atteso, in quanto il volume CBCT viene ricostruito indirettamente a partire da immagini MPR e dalla traiettoria ottimizzata, e non è oggetto diretto dell'ottimizzazione. Inoltre, la qualità volumetrica risente della natura delle MPR generate, che può introdurre una perdita di dettaglio anatomico rispetto a una CBCT acquisita direttamente.

Prospettive future e possibili estensioni

Nonostante i risultati ottenuti dimostrino la validità dell'approccio proposto, vi sono margini di miglioramento che possono essere esplorati in lavori futuri. In primo luogo, l'ottimizzazione dei parametri è stata affrontata utilizzando metodi deterministici basati sul gradiente (ad esempio, Levenberg–Marquardt). Tuttavia, in presenza di superfici di errore altamente non lineari e con molteplici minimi locali, potrebbe risultare vantaggioso sperimentare algoritmi alternativi. Tra questi si annoverano approcci evolutivi come gli algoritmi genetici, metodi ispirati alla swarm intelligence come Particle Swarm Optimization [13], o strategie di discesa stocastica [14], in grado di esplorare meglio lo spazio dei parametri ed evitare la convergenza prematura.

Parallelamente, si potrebbe valutare l'impiego di funzionali di errore alternativi o di combinazioni pesate di metriche, costruendo una funzione di costo composita più robusta. Ad esempio, oltre alla DSSIM, si potrebbero integrare metriche basate sulla similarità strutturale locale, sull'accordo dei bordi (Edge Agreement), sulla correlazione dei gradienti (Gradient Correlation) [25], o metriche percettive ispirate al sistema visivo umano [32]. Queste opzioni potrebbero contribuire a una maggiore coerenza visiva nella panoramica e a una guida più efficace per l'ottimizzazione della curva mandibolare e dei parametri geometrici.

Algoritmi alternativi di ottimizzazione L'ottimizzazione congiunta dei parametri di proiezione e della curva spline può essere affetta dalla presenza di minimi locali, specialmente quando si utilizzano funzioni obiettivo non lineari o non derivabili come la DSSIM. Per affrontare tali situazioni, potrebbe risultare vantaggioso sperimentare algoritmi alternativi, capaci di esplorare lo spazio dei parametri in modo più ampio.

Tra questi, si annoverano approcci evolutivi come gli algoritmi genetici [7], che simulano i meccanismi della selezione naturale e della mutazione per esplorare soluzioni diverse in parallelo. Altri metodi promettenti sono quelli ispirati alla swarm intelligence, come il Particle Swarm Optimization (PSO) [13], in cui una popolazione di “particelle” (soluzioni candidate) aggiorna iterativamente la propria posizione nello spazio dei parametri sulla base sia della propria esperienza che di quella del gruppo.

Tali strategie non richiedono il calcolo del gradiente e sono particolarmente utili quando la funzione obiettivo è non continua, non differenziabile o presenta plateau che rendono difficile la convergenza dei metodi classici basati sulla discesa del gradiente.

Chapter 6

Conclusione

Questo lavoro ha affrontato il problema della ricostruzione tridimensionale del volume mandibolare a partire da una singola radiografia panoramica, proponendo un approccio ibrido che combina apprendimento profondo e modellazione geometrica esplicita. L'obiettivo principale era stimare una curva spline mandibolare personalizzata che, se usata come traiettoria nel processo di deformazione, restituisse una CBCT ricostruita coerente con l'anatomia del paziente.

I risultati ottenuti dimostrano un miglioramento significativo nella similarità tra la proiezione iniziale non ottimizzata e quella reale, specialmente utilizzando la metrica DSSIM come guida per l'ottimizzazione. È stata inoltre osservata una convergenza efficace della funzione obiettivo, con andamenti stabili e sensibili alle variazioni locali della curva spline. La metodologia proposta mostra dunque il potenziale per generare ricostruzioni 3D più fedeli alla morfologia del paziente rispetto a tecniche standard basate su approcci morfologici fissi o regressivi.

L'originalità del lavoro risiede nella combinazione tra una rete generativa per la ricostruzione iniziale del volume e una fase successiva di ottimizzazione geometrica esplicita basata su ray tracing. Questo schema modulare si presta a future integrazioni in pipeline differenziabili end-to-end o in scenari di imaging a bassa dose, risultando quindi promettente anche in ambito clinico.

Tuttavia, il metodo presenta alcune limitazioni, legate principalmente alla complessità computazionale del processo iterativo e alla dipendenza da una stima iniziale accurata della curva spline. Inoltre, la valutazione è stata condotta su un numero limitato di casi, il che richiede ulteriori validazioni su larga scala.

Come sviluppi futuri, si prevede di integrare l'ottimizzazione in un framework completamente differenziabile, migliorare l'efficienza computazionale e validare il metodo in scenari clinici reali, eventualmente includendo soggetti con variabilità anatomica marcata o condizioni patologiche.

Bibliography

- [1] Ahmed J. Afifi, Jannes Magnusson, Toufique A. Soomro, and Olaf Hellwich. Pixel2point: 3d object reconstruction from a single image using cnn and initial sphere. *IEEE Access*, 9:110–121, 2021.
- [2] American Dental Association Council on Scientific Affairs. The use of cone-beam computed tomography in dentistry: an advisory statement from the american dental association council on scientific affairs. *Journal of the American Dental Association*, 143(8):899–902, 2012.
- [3] S. Baccher, I. M. Gowdar, Y. Guruprasad, R. N. Solanki, R. Medhi, M. J. Shah, and D. N. Mehta. Cbct: A comprehensive overview of its applications and clinical significance in dentistry. *Journal of Pharmacy & Bioallied Sciences*, 16(Suppl 3):S1923–S1925, 2024.
- [4] Myungsoo Bae, Jae-Woo Park, and Namkug Kim. Semi-automatic and robust determination of dental arch form in dental cone-beam ct with b-spline approximation. *Computer Methods and Programs in Biomedicine*, 172:95–101, 2019.
- [5] Christopher B. Choy, Danfei Xu, JunYoung Gwak, Kevin Chen, and Silvio Savarese. 3d-r2n2: A unified approach for single and multi-view 3d object reconstruction, 2016.
- [6] Thomas F. Coleman and Yuying Li. On the convergence of interior-reflective newton methods for nonlinear minimization subject to bounds. *Mathematical Programming*, 67:189–224, 1994.
- [7] Yong Fan, Tianzi Jiang, and David J. Evans. Medical image registration using parallel genetic algorithms. In Stefano Cagnoni, Jens Gottlieb, Emma Hart, Martin Middendorf, and Günther R. Raidl, editors, *Applications of Evolutionary Computing*, pages 304–314, Berlin, Heidelberg, 2002. Springer Berlin Heidelberg.
- [8] Rocco Furferi. Deep learning approaches for 3d model generation from 2d artworks to aid blind people with tactile exploration. *Heritage*, 8(1), 2025.
- [9] NL Gerlach, H Ghaeminia, EM Bronkhorst, SJ Bergé, GJ Meijer, and TJJ Maal. Accuracy of assessing the mandibular canal on cone-beam computed tomography: a validation study. *Journal of Oral and Maxillofacial Surgery*, 72(4):666–671, 2014.

- [10] Vahid Ghodrati, Jun Shao, Mark Bydder, Zhaoyang Zhou, Wenli Yin, Kim-Lien Nguyen, Yonggang Yang, and Peng Hu. Mr image reconstruction using deep learning: evaluation of network structure and loss functions. *Quantitative Imaging in Medicine and Surgery*, 9(9):1516–1527, 2019.
- [11] Ian J. Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial networks, 2014.
- [12] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A. Efros. Image-to-image translation with conditional adversarial networks. *CoRR*, abs/1611.07004, 2016.
- [13] J. Kennedy and R. Eberhart. Particle swarm optimization. In *Proceedings of ICNN'95 - International Conference on Neural Networks*, volume 4, pages 1942–1948 vol.4, 1995.
- [14] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2017.
- [15] Xiang Li. 3d multimodal dental dataset based on cbct and oral scan. <https://doi.org/10.6084/m9.figshare.26965903.v3>, 2024. Dataset.
- [16] Yuhui Ma, Jiang Liu, Yonghuai Liu, Huazhu Fu, Yan Hu, Jun Cheng, Hong Qi, Yufei Wu, Jiong Zhang, and Yitian Zhao. Structure and illumination constrained gan for medical image enhancement. *IEEE Transactions on Medical Imaging*, 40(12):3955–3967, 2021.
- [17] Sreekanth Kumar Mallineni, Robert Prashanth Anthonappa, Jayakumar Jayaraman, and Nigel Martyn King. Radiographic localization of supernumerary teeth: a narrative review. *Frontiers in Dental Medicine*, 6:1495025, 2025.
- [18] Xudong Mao, Qing Li, Haoran Xie, Raymond Y. K. Lau, and Zhen Wang. Multi-class generative adversarial networks with the L2 loss function. *CoRR*, abs/1611.04076, 2016.
- [19] Jorge Nocedal and Stephen J. Wright. *Numerical optimization*. Springer series in operations research and financial engineering. Springer, New York, NY, 2. ed. edition, 2006.
- [20] Sihwa Park, Seongjun Kim, Doeyoung Kwon, Yohan Jang, In-Seok Song, and Seung Jun Baek. Nebla: Neural beer-lambert for 3d reconstruction of oral structures from panoramic radiographs, 2024.
- [21] Shanon Patel, C. Durack, F. Abella, H. Shemesh, M. Roig, and K. Lemberg. Cone beam computed tomography in endodontics - a review. *International Endodontic Journal*, 48(1):3–15, January 2015.

- [22] J. P. Pluim, J. B. A. Maintz, and M. A. Viergever. Mutual-information-based registration of medical images: a survey. *IEEE Transactions on Medical Imaging*, 22(8):986–1004, 2003.
- [23] William C Scarfe, Allan G Farman, and Predrag Sukovic. Clinical applications of cone-beam computed tomography in dental practice. *Journal of the Canadian Dental Association*, 72(1):75–80, 2006.
- [24] Ashish Shrivastava, Tomas Pfister, Oncel Tuzel, Josh Susskind, Wenda Wang, and Russell Webb. Learning from simulated and unsupervised images through adversarial training. *CoRR*, abs/1612.07828, 2016.
- [25] Mikael Sjödahl. Gradient correlation functions in digital image correlation. *Applied Sciences*, 9(10), 2019.
- [26] Weinan Song, Yuan Liang, Jiawei Yang, Kun Wang, and Lei He. Oral-3d: Reconstructing the 3d bone structure of oral cavity from 2d panoramic x-ray, 2021.
- [27] Weinan Song, Haoxin Zheng, Dezhao Tu, Chengwen Liang, and Lei He. Oral-3dv2: 3d oral reconstruction from panoramic x-ray imaging with implicit neural representation, 2023.
- [28] Aristeidis Sotiras, Christos Davatzikos, and Nikos Paragios. Deformable medical image registration: a survey. *IEEE Transactions on Medical Imaging*, 32(7):1153–1190, 2013.
- [29] Tomaz Vrtovec, Sébastien Ourselin, Lavier Gomes, Boštjan Likar, and Franjo Pernuš. Generation of curved planar reformations from magnetic resonance images of the spine. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2006*, volume 4191 of *Lecture Notes in Computer Science*, pages 135–143. Springer, 2006.
- [30] Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.
- [31] Bin Xia, Yulun Zhang, Shiyin Wang, Yitong Wang, Xinglong Wu, Yapeng Tian, Wenming Yang, and Luc Van Gool. Diffir: Efficient diffusion model for image restoration, 2023.
- [32] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. *CoRR*, abs/1801.03924, 2018.
- [33] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A. Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks, 2020.