ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

SCUOLA DI SCIENZE

Corso di Laurea in Informatica

Strategie Plug-and-Play per la ricostruzione tomografica bidimensionale

Relatore:
Chiar.ma Prof.
Elena Loli Piccolomini

Presentata da:
Chiara Bertocchi

Sessione IV
Anno Accademico 2024/2025

Introduzione

Il presente lavoro di tesi si propone di esplorare le potenzialità e i limiti tecnici dell'approccio **Plug-and-Play (PnP)** applicato nella ricostruzione di immagini nella **Tomografia Computerizzata (CT)**. Negli ultimi anni, la ricerca nel campo dell'imaging computazionale è stata caratterizzata da una dicotomia tra i metodi variazionali classici, solidi matematicamente ma limitati da regolarizzatori analitici troppo semplici, e gli approcci puramente data-driven basati sul Deep Learning, capaci di prestazioni straordinarie ma spesso criticati per la loro natura di "scatola nera" e la scarsa interpretabilità fisica.

Il paradigma Plug-and-Play si colloca in questo scenario come una soluzione di compromesso d'avanguardia: esso permette di integrare la potenza espressiva di reti neurali pre-addestrate (utilizzate come denoiser) all'interno di schemi di ottimizzazione iterativa tradizionali. Questo garantisce che la ricostruzione finale rimanga coerente con la fisica del sistema di acquisizione, beneficiando al contempo della capacità delle reti profonde di apprendere prior complessi dai dati. L'obiettivo della tesi è validare tale approccio attraverso una rigorosa campagna sperimentale, analizzandone la robustezza in scenari critici di campionamento ridotto.

La struttura della tesi è organizzata come segue:

- Il Capitolo 1 definisce i fondamenti teorici della Tomografia Computerizzata analizzando l'operatore matematico della *Trasformata di Radon*, il *Fourier Slice Theorem* e le principali geometrie di acquisizione. Viene, inoltre, formalizzato il concetto di *problema inverso*, discutendone le criticità legate al mal-condizionamento e alla necessità di regolarizzazione per la ricostruzione di immagini di alta qualità.

- Il Capitolo 2 fornisce una panoramica del framework *DeepInverse*, libreria open-source che costituisce il motore computazionale dell'intero progetto. In questa sezione vengono descritti i moduli fondamentali utilizzati per la modellazione fisica, la gestione dei dataset e l'implementazione degli algoritmi di

ottimizzazione. Viene, inoltre, evidenziata la modularità del framework e la facilità di standardizzazione della ricerca su problemi inversi.

- Il Capitolo 3 entra nel dettaglio delle scelte implementative e del setup sperimentale. Vengono descritte le fasi di pre-processing dei dati, le strategie di *data augmentation* e le procedure di *fine-tuning* adottate per i denoiser DnCNN e DRUNet. Viene, inoltre, presentata la configurazione dei parametri, come il *continuation schedule*, essenziale per garantire la convergenza degli algoritmi.
- Il Capitolo 4 è dedicato all'esposizione e all'analisi dei risultati ottenuti dalla sperimentazione. La valutazione si articola in: analisi del comportamento del modello in regime di *Low-Dose CT* (riduzione drastica degli angoli di proiezione), confronto tra denoiser di tipo blind e non-blind, e infine una comparazione diretta tra il paradigma PnP e l'approccio *Unfolded* (Learned Primal-Dual). Attraverso l'utilizzo di metriche quantitative (MSE, PSNR, SSIM) e analisi delle distribuzioni statistiche tramite boxplot, vengono tratte le conclusioni sulla stabilità e l'efficacia clinica dei modelli proposti.

Indice

1 Tomografia computerizzata **1**
- 1.1 Trasformata di Radon e Fourier slice Theorem . . . 2
- 1.2 Geometrie di acquisizione . . . 4
  - 1.2.1 Parallela . . . 4
  - 1.2.2 Fan-Beam . . . 5
- 1.3 Il problema inverso nella Ricostruzione Tomografica . . . 6

2 Strumenti e metodi **7**
- 2.1 Il framework DeepInverse . . . 7
  - 2.1.1 Metodi di ricostruzione . . . 8
  - 2.1.2 Addestramento delle reti usate per la ricostruzione di immagini 9
  - 2.1.3 Modellazione fisica . . . 10
- 2.2 Confronto tra approcci Variazionali e Deep Learning . . . 11
  - 2.2.1 Approccio Variazionale (model-based) . . . 11
  - 2.2.2 Approccio Deep Learning (data-driven) . . . 12
- 2.3 Plug-and-Play . . . 12
  - 2.3.1 Definizione matematica del metodo . . . 13
  - 2.3.2 Proximal Gradient Descent (PGD) . . . 13
  - 2.3.3 Denoiser come Prior . . . 14
  - 2.3.4 Convergenza e proprietà del punto fisso . . . 16
- 2.4 Approccio Unfolded . . . 17
  - 2.4.1 Learned Primal-Dual . . . 17

3 Caso di Studio **19**
- 3.1 Dataset utilizzato . . . 19
  - 3.1.1 Scelta implementativa 1: creazione e suddivisione del dataset . 20
  - 3.1.2 Scelta implementativa 2: data augmentation . . . 21
- 3.2 Operatore fisico di acquisizione tomografica . . . 21
  - 3.2.1 Geometria e configurazione . . . 22
  - 3.2.2 Scelta implementativa 1: calcolo di una stepsize stabile . . . 22

3.3 Training dei denoiser 22
3.3.1 Strategia di addestramento per DRUNet 23
3.3.2 Strategia di addestramento per DnCNN 24
3.4 Configurazione dell'algoritmo PnP e Unfolded 24
3.4.1 Continuation Schedule 25
3.4.2 Implementazione Unfolded 26

4 Risultati e analisi **27**
4.1 Metriche di confronto 27
4.1.1 Mean Squared Error (MSE) 28
4.1.2 Peak Signal-to-Noise Ratio (PSNR) 29
4.1.3 Indice di similarità strutturale (SSIM) 30
4.2 Sparse-view CT: Analisi della riduzione delle proiezioni 32
4.2.1 Baseline sperimentale: 180 proiezioni 33
4.2.2 Riduzione della dose a 90 proiezioni 36
4.2.3 Riduzione della dose a 60 proiezioni 39
4.3 Analisi comparativa tra i metodi di ricostruzione 43
4.3.1 Confronto tra modellazione analitica e approcci Data-driven . 43
4.3.2 Confronto tra PnP-DRUNet e PnP-DnCNN 45
4.4 Analisi comparativa tra approcci Plug-and-Play e metodi Unfolded . 46

Conclusioni **49**

Bibliografia **51**

Ringraziamenti **53**

Elenco delle tabelle

4.1 Metriche medie ottenute con 180 proiezioni angolari. 33
4.2 Metriche medie ottenute con 90 proiezioni angolari. 37
4.3 Metriche medie ottenute con 60 proiezioni angolari. 40
4.4 Confronto delle metriche medie ottenute attraverso PnP-DRUNet e Unfolded Primal-Dual (180 proiezioni). 46

Elenco delle figure

1.1 Schema dell'acquisizione dell'oggetto attraverso Computed Tomography 2
1.2 Schema dell'acquisizione dell'oggetto tramite geometria parallela. . . 4
1.3 Schema dell'acquisizione dell'oggetto tramite geometria divergente. . . 5

2.1 Schema della modularità del framework DeepInverse 8
2.2 Architettura del denoiser DnCNN 15
2.3 Architettura del denoiser DRUNet 16

3.1 Esempi di campioni presenti all'interno del dataset Koule 20
3.2 Esempio di applicazione sequenziale di un flip verticale e una rotazione di 90° 21
3.3 Schema dell'iterato inziale con FBP contro semplice back-projection . 25

4.1 Confronto tra la fedeltà di immagini modificate con tipi diversi di distorsioni. (a) groud truth, (b) Aumento del contrasto, (c) cambio della luminanza, (d) contaminazione con rumore gaussiano, (e) contaminazione con rumore a impulsi, (f) compressione JPEG, (g) blurring, (h) spacial scaling. 29
4.2 Confronto tra la fedeltà di immagini modificate con tipi diversi di distorsioni. (a) groud truth, (b) aumento del contrasto, (c) degradazione con rumore bianco uniforme, (d) blurring. 30
4.3 Confronto di diverse distorsioni strutturali applicate alla medesima immagine: si noti come distorsioni visivamente molto diverse possano produrre lo stesso valore di MSE, evidenziando la necessità di metriche strutturali come l'SSIM. 32
4.4 Confronto visivo delle ricostruzioni di campioni del test set con 180 angoli: Sinogramma, Ground Truth, GBP, TV, PnP-DRUNet, PnP-DnCNN 35
4.5 Distribuzione statistica delle metriche PSNR, SSIM e MSE per i diversi modelli di ricostruzione. 36

4.6 Confronto visivo delle ricostruzioni di campioni del test set con 90 angoli: Sinogramma, Ground Truth, GBP, TV, PnP-DRUNet, PnP-DnCNN 38
4.7 Distribuzione statistica delle metriche PSNR, SSIM e MSE per i diversi modelli di ricostruzione. 39
4.8 Confronto visivo delle ricostruzioni di campioni del test set con 60 angoli: Sinogramma, Ground Truth, GBP, TV, PnP-DRUNet, PnP-DnCNN 41
4.9 Distribuzione statistica delle metriche PSNR, SSIM e MSE per i diversi modelli di ricostruzione. 42
4.10 Confronto prestazionale tra metodi Deep e tradizionali al diminuire del numero di viste. Si noti come i modelli PnP mantengano un distacco netto (circa 7-8 dB) rispetto alla TV anche nello scenario di sottocampionamento estremo. 43
4.11 Campioni di ricostruzione tramite Total Variation: si noti l'effetto di livellamento sulle strutture minori. 44
4.12 Confronto diretto tra la regolarizzazione TV e l'approccio Plug-and-Play. 44
4.13 Confronto visivo delle ricostruzione eseguite attraverso approccio Unfolded e PnP 47
4.14 Distribuzione statistica delle metriche PSNR, SSIM e MSE per il confronto tra gli approcci PnP e Unfolded. 48

Capitolo 1

Tomografia computerizzata

La **Tomografia Computerizzata (CT)** è stata introdotta per superare i limiti intrinsechi della radiografia tradizionale a raggi X. Quest'ultima, infatti, produce proiezioni bidimensionali di volumi tridimensionali, causando la sovrapposizione di strutture anatomiche diverse o il mascheramento di dettagli critici, limitando sensibilmente l'accuratezza diagnostica[2].

La CT si basa sull'acquisizione di proiezioni dell'oggetto da molteplici angolazioni. Matematicamente, questo processo può essere descritto introducendo un nuovo sistema di coordinate (r, s) ruotato di un angolo ϕ rispetto al sistema di riferimento dell'oggetto (x, y). La trasformazione di coordinate è definita dalla **matrice di rotazione**:

$$\begin{pmatrix} r \\ s \end{pmatrix} = \begin{pmatrix} \cos\phi & \sin\phi \\ -\sin\phi & \cos\phi \end{pmatrix} \begin{pmatrix} x \\ y \end{pmatrix} \tag{1.1}$$

dove

- $f(x, y)$ rappresenta l'oggetto o la sezione anatomica sotto studio.
- $p(r, \phi)$ rappresenta la proiezione acquisita per un determinato angolo ϕ.
- ϕ è l'angolo di incidenza del fascio di raggi X.
- s rappresenta la coordinata lungo la direzione del raggio X, ovvero il percorso di integrazione dell'attenuazione.

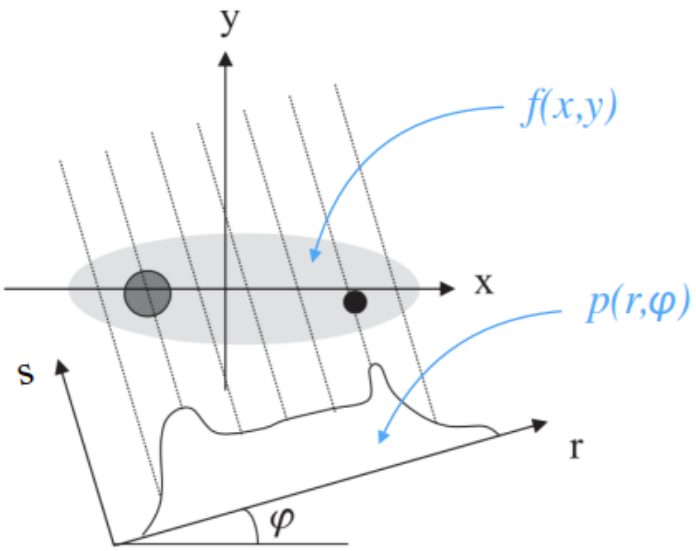

Figura 1.1: Schema dell'acquisizione dell'oggetto attraverso Computed Tomography

L'insieme di tutte le proiezioni $p(r, \phi)$ acquisite al variare dell'angolo ϕ viene organizzato in una struttura dati denominata **sinogramma**. Attraverso algoritmi di ricostruzione, è possibile invertire questo processo per risalire alla distribuzione spaziale $f(x, y)$, ottenendo così sezioni trasversali nitide e prive di sovrapposizioni.

1.1 Trasformata di Radon e Fourier slice Theorem

Alla base delle ricostruzioni per la Tomografia Computerizzata si trova la **Trasformata di Radon**, introdotta dal matematico Johann Radon nel 1917[1]. Essa rappresenta l'operatore matematico che modella il processo di acquisizione, trasformando una funzione bidimensionale $f(x, y)$ (che rappresenta la distribuzione spaziale dell'oggetto) nel corrispondente sinogramma $p(r, \phi)$ attraverso il calcolo di integrali di linea. La trasformata di Radon $\mathcal{R}$ descrive l'attenuazione subita dal fascio di raggi X lungo la retta di integrazione s. In un sistema di coordinate ruotato, la formula è definita come:

$$\mathcal{R}f(r, \phi) = \int_{-\infty}^{+\infty} f(r\cos\phi - s\sin\phi, r\sin\phi + s\cos\phi), ds \tag{1.2}$$

Per una rappresentazione più generale, è possibile utilizzare la funzione *delta di Dirac* δ per selezionare tutti i punti (x, y) appartenenti alla retta di proiezione identificata

dall'equazione $r = x\cos\phi + y\sin\phi$:

$$\mathcal{R}f(r,\phi) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} f(x,y)\delta(r - x\cos\phi - y\sin\phi), dx, dy \tag{1.3}$$

In questa formulazione, r indica la distanza perpendicolare della retta dall'origine, mentre ϕ rappresenta l'angolo di inclinazione del raggio rispetto all'asse delle ascisse.

La traformata di Radon permette, quindi, di rappresentare matematicamente il passaggio dall'oggetto al relativo sinogramma, ovvero la generazione dei dati acquisiti. Tuttavia, per definire il passaggio inverso necessario alla ricostruzione, si utilizza il **Fourier Slice Theorem**[2] che mette in relazione diretta la Trasformata di Radon con la Trasformata di Fourier. Il teorema afferma che la trasformata di Fourier unidimensionale di una proiezione (ottenuta tramite Radon ad un angolo θ) è equivalente a una "fetta" della trasformata di Fourier bidimensionale dell'oggetto intero, presa allo stesso angolo θ.

Matematicamente, se indichiamo con $\hat{g}_\theta(\omega)$ la trasformata della proiezione e con $\hat{f}$ quella dell'immagine, si ha:

$$\hat{g}_\theta(\omega) = \hat{f}(\omega\cos\theta, \omega\sin\theta) \tag{1.4}$$

Questo risultato è di cruciale importanza: esso implica che, accumulando proiezioni da un numero sufficiente di angolazioni tra 0 e π, è possibile mappare l'intero spazio delle frequenze dell'oggetto. Idealmente, la ricostruzione finale si otterrebbe applicando una trasformata di Fourier inversa ai dati così raccolti. Tuttavia, nella pratica clinica e numerica, questo approccio diretto presenta due criticità principali:

- **Campionamento Radiale**: Le proiezioni forniscono dati su una griglia polare (a raggiera). Poiché i sistemi di calcolo operano su griglie cartesiane quadrate, è necessaria un'interpolazione che può introdurre errori, specialmente alle alte frequenze dove i dati sono più radi.
- **Effetto di sfocatura (Retroproiezione)**: Il tentativo di riportare i dati dal sinogramma allo spazio immagine tramite la *retroproiezione* (spalmando i dati lungo le direzioni di acquisizione) introduce una sfocatura intrinseca proporzionale a $1/|\underline{x}|$.

Per ovviare a tali problemi, si ricorre all'algoritmo di **Filtered Backprojection (FBP)**, che applica un filtro correttivo (filtro a rampa) alle proiezioni prima della retroproiezione, garantendo una ricostruzione nitida e fedele alla struttura originale[3].

1.2 Geometrie di acquisizione

Il processo di acquisizione dei dati in CT è fortemente influenzato dalla configurazione spaziale della **sorgente** di radiazioni e dei **rilevatori**. Tale configurazione, definita **geometria di acquisizione**, determina la modalità in cui i raggi X attraversano l'oggetto e, di conseguenza, la struttura matematica dei dati raccolti nel sinogramma. Vengono di seguito descritte le principali metodologie di scansione.

1.2.1 Parallela

La geometria **Parallela** rappresenta la configurazione più lineare e intuitiva per l'acquisizione tomografica. Essa si basa sull'utilizzo di raggi paralleli ed quidistanti che non presentano alcun tipo di divergenza o intersezione, indipendentemente dalla distanza percorsa. Nello specifico, il sistema, composto da sorgente e rilevatore, trasla lungo una linea retta per acquisire una serie di proiezioni parallele; una volta completata la traslazione, l'intero apparato ruota di un piccolo incremento per ripetere il processo fino a coprire l'arco di scansione necessario. La geometria è caratterizzata principalmente dalla coordinata radiale (r), che indica la distanza del raggio dall'origine, e dall'angolo di rotazione (ϕ), che determina l'orientamento del fascio rispetto all'oggetto.

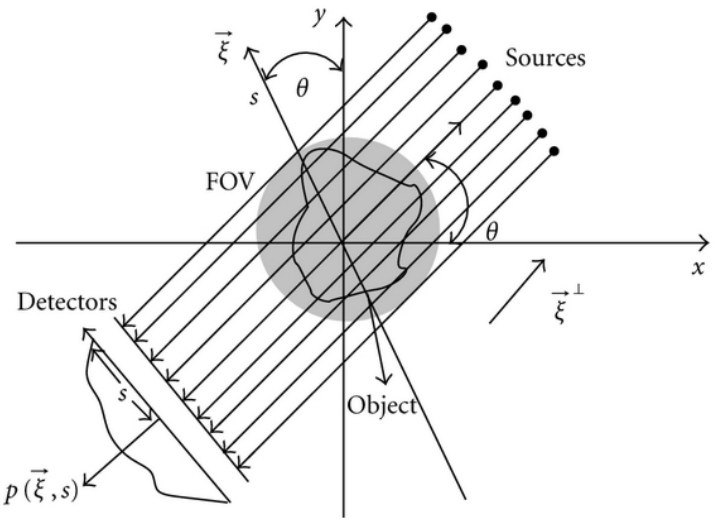

Figura 1.2: Schema dell'acquisizione dell'oggetto tramite geometria parallela.

Sebbene questa geometria sia stata fondamentale per i primi prototipi di scanner CT, oggi viene utilizzata raramente in ambito clinico a causa dei lunghi tempi di

acquisizione richiesti dal movimento di traslazione.

1.2.2 Fan-Beam

La geometria **Fan-Beam** descrive un setup in cui la sorgente genera un fascio di raggi a forma di ventaglio divergente (si veda la Figura 1.3), in grado di coprire l'intera sezione bidimensionale dell'oggetto catturata dal rilevatore. In questa configurazione, la sorgente e il detector ruotano attorno all'isocentro del macchinario per acquisire le proiezioni necessarie alla ricostruzione. Questa geometria è caratterizzata da un insieme di parametri geometrici, tra i quali assumono maggiore importanza la distanza sorgente-isocentro, l'angolo di vista (β) e l'angolo del singolo raggio all'interno del ventaglio (γ).

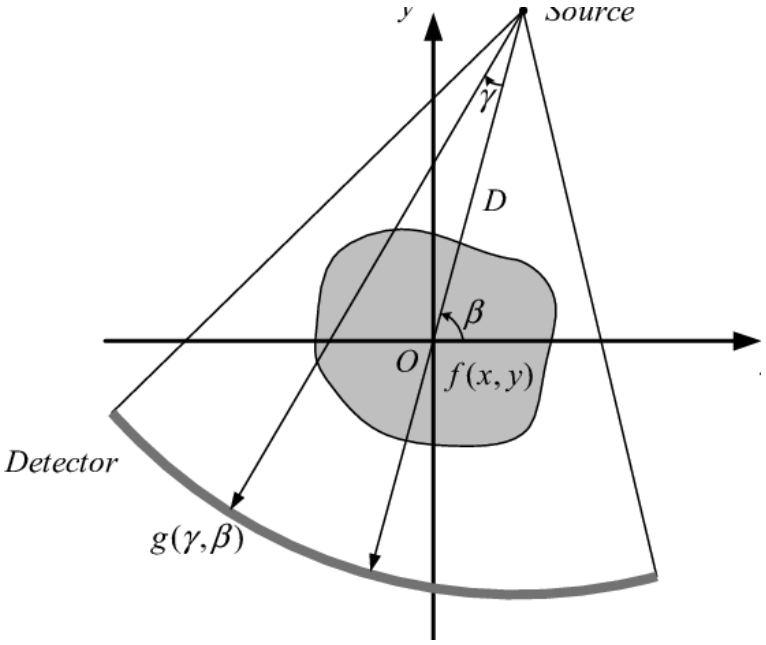

Figura 1.3: Schema dell'acquisizione dell'oggetto tramite geometria divergente.

Dal punto di vista analitico, la relazione che lega l'angolo di un raggio parallelo (ϕ) alle coordinate della geometria Fan-Beam è espressa dalla formula:

$$\phi = \beta + \gamma \tag{1.5}$$

dove β rappresenta l'angolo di rotazione della sorgente e γ l'angolazione del raggio rispetto all'asse centrale del ventaglio.

Rispetto ai sistemi a fasci paralleli, la geometria Fan-Beam offre una maggiore sensibilità fotonica e, a parità di dimensioni del rilevatore, permette di ottenere un campo

di vista (*Field of View*) notevolmente più ampio. Di conseguenza, nelle applicazioni cliniche e industriali si predilige l'utilizzo di geometrie divergenti. Tuttavia, la geometria a fasci paralleli rimane il modello teorico di riferimento per lo studio analitico della trasformata di Radon.

1.3 Il problema inverso nella Ricostruzione Tomografica

La ricostruzione di un'immagine CT può essere descritta formalmente come la risoluzione di un **problema inverso**[5]. Quest'ultimo consiste nel recuperare la distribuzione spaziale interna dell'oggetto $f(x, y)$ partendo dalle proiezioni acquisite. Il processo di acquisizione viene modellato attraverso un sistema di equazioni lineari:

$$y = Ax + \epsilon \tag{1.6}$$

dove:

- $x \in \mathbb{R}^N$ rappresenta l'immagine originale.
- A è l'operatore di proiezione (matrice di sistema) che approssima la trasformata di Radon in base alla geometria scelta (Parallela o Fan-Beam).
- $y \in \mathbb{R}^M$ è il sinogramma misurato.
- ϵ rappresenta il rumore statistico e le incertezze di misura intrinseche al sistema.

Nelle applicazioni reali, ottenere x risulta notevolmente difficoltoso poiché il problema è spesso **mal posto**. Questo accade specialmente in scenari di **Low-Dose CT**, dove il numero di proiezioni M è significativamente inferiore al numero di incognite N. Di conseguenza, per superare questi limiti, vengono utilizzate tecniche di ottimizzazione iterativa che cercano di minimizzare una funzione di costo composta da un termine di fedeltà ai dati e un termine di regolarizzazione (prior):

$$\min_x \frac{1}{2}|Ax - y|^2 + \lambda \mathcal{R}(x) \tag{1.7}$$

Questo approccio costituisce la base logica per i modelli analizzati in questa tesi, come l'Unfolded Primal-Dual[5] e il PnP-DRUNet[6], i quali sostituiscono o integrano il termine di regolarizzazione $\mathcal{R}(x)$ con reti neurali profonde addestrate per apprendere le caratteristiche strutturali delle immagini tomografiche.

Capitolo 2

Strumenti e metodi

In questo capitolo vengono illustrate le principali caratteristiche di **DeepInverse**[4], la libreria open-source utilizzata per lo sviluppo del progetto di tesi. Successivamente, viene presentata la classificazione dei metodi di ricostruzione tomografica, distinguendo tra approcci variazionali e metodi basati sul Deep Learning. Infine, viene descritto in dettaglio l'approccio **Plug-and-Play** e l'approccio **Unfolded**, approfondendone il funzionamento e l'integrazione all'interno del framework considerato.

2.1 Il framework DeepInverse

DeepInverse è una libreria open-source sviluppata su ambiente **Pytorch**, specificamente progettata per la risoluzione di problemi inversi nel campo dell'imaging. Il framework si distingue per la sua elevata modularità, permettendo di gestire l'intero workflow di ricostruzione: dalla definizione di operatori di forward, alla risoluzione di problemi variazionali complessi.

Attualmente, le reti neurali profonde rappresentano la scelta principale per la risoluzione di molti problemi inversi. Tuttavia, nonostante il crescente interesse della ricerca, la maggior parte degli algoritmi basati su deep learning viene sviluppata "ad hoc" per applicazioni specifiche; ciò rende difficile la generalizzazione del codice al di fuori dell'ambito di training e complica la riproducibilità dei risultati sperimentali. DeepInverse mira a superare tali limitazioni fornendo un framework unificato e modulare, che favorisca la standardizzazione e il riutilizzo del codice tra diversi domini applicativi (Figura 2.1).

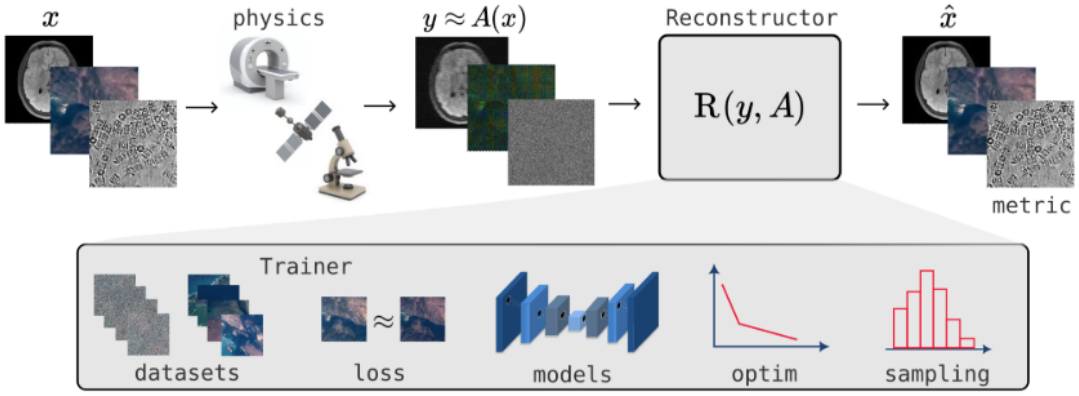

Figura 2.1: Schema della modularità del framework DeepInverse

2.1.1 Metodi di ricostruzione

DeepInverse unifica i principali algoritmi di ricostruzione presenti in letteratura attraverso la seguente formula generale:

$$\hat{x} = R_\theta(y, A_\xi, \sigma) \tag{2.1}$$

dove R_θ rappresenta l'algoritmo di ricostruzione caratterizzato da un insieme di parametri addestrabili θ, mentre $\hat{x}$ è l'immagine ricostruita. In termini di implementazione, la libreria permette di calcolare la stima attraverso una sintassi intuitiva, tipicamente espressa come:

```
x_hat = model(y, physics)
```

Gli algoritmi disponibili nella libreria sono classificati in tre macro-categorie, a seconda dell'integrazione tra modello fisico e componenti di apprendimento:

- **Optimization-based methods**: come formalizzato da **Chambolle e Pock (2016)** [8], l'operatore di ricostruzione R_θ viene definito risolvendo un problema di minimizzazione della forma:

  $$\hat{x} = R_\theta(y, A_\xi, \sigma) \in \arg\min_x f_\sigma(y, A_\xi(x)) + \lambda g(x) \tag{2.2}$$

  In questa formulazione, $f_\sigma(y, A_\xi(x))$ rappresenta il termine di **data fidelity**, che misura la coerenza tra la stima e i dati osservati y attraverso il modello fisico A_ξ, mentre $g(x)$ è il termine di **regolarizzazione** (prior) che incorpora

informazioni a priori sulla struttura dell'immagine per rendere il problema *ben posto*.

Il modulo `optim` di DeepInverse offre diverse strategie per modellare il prior $g(x)$: si spazia dai metodi classici basati su operatori espliciti (come la norma L_1), a soluzioni che impiegano **regolarizzatori appresi** direttamente dai dati. Una menzione particolare va all'approccio **Plug-and-Play (PnP)**, tecnica che permette di sostituire l'operatore di prossimità del regolarizzatore con un denoiser pre-addestrato[6] (ad esempio una rete U-Net), garantendo notevole flessibilità nella modellazione del rumore.

- **Sampling-based methods**: vengono definiti da equazioni differenziali stocastiche e campionamento probabilistico. In questo paradigma, l'obiettivo è generare una sequenza di campioni $\{x_t\}$ che converga verso la distribuzione a posteriori $p(x|y)$:

$$x_{t+1} \sim p(x_{t+1}|x_t, y, D_\sigma, A_\xi, \sigma) \quad \text{per } t = 0, \ldots, T-1 \tag{2.3}$$

  In questo modo, x_T rappresenta un campione approssimato della distribuzione corretta. Il vantaggio principale di questo approccio risiede nella possibilità di effettuare campionamenti multipli, permettendo la **quantificazione dell'incertezza** della ricostruzione. A tal fine, il modulo `sampling` implementa framework che spaziano dai **modelli di diffusione** (che sfruttano la capacità generativa di modelli pre-addestrati) agli algoritmi di tipo **Langevin**, i quali utilizzano tecniche di *Markov Chain Monte Carlo* per esplorare lo spazio delle soluzioni.

- **Non-iterative methods**: eliminano la necessità di cicli di ottimizzazione complessi per gli scenari in cui l'efficienza computazionale è prioritaria. Il modulo `models` mette a disposizione approcci come l' **artifact removal**, in cui i dati y vengono riportati nel dominio dell'immagine tramite un operatore di *backprojection* ($A_\xi^\top$), per poi essere processati da un denoiser D_σ in un singolo passaggio. Inoltre, la libreria supporta **reti generative condizionate**, che introducono una variabile latente z per mappare i dati in una varietà di possibili ricostruzioni $R_\theta(y, z)$, e i recenti **Foundation Models**, addestrati *end-to-end* per generalizzare su una vasta gamma di operatori di forward.

2.1.2 Addestramento delle reti usate per la ricostruzione di immagini

L'**addestramento** di una rete neurale è un processo iterativo volto a ottimizzare i pesi della rete stessa per minimizzare una funzione obiettivo. Questo processo si basa su un insieme di dati denominato **training set**. Una volta conclusa la fase di

addestramento, le prestazioni del modello devono essere verificate su un insieme di dati indipendenti, definito **test set**, per garantire la capacità di generalizzazione.

Per implementare correttamente questo workflow, DeepInverse mette a disposizione i seguenti strumenti:

- **Funzione di perdita** (loss function): una funzione matematica utilizzata per misurare la discrepanza tra il valore stimato dal modello e il valore reale (**ground truth**). Nel contesto dei problemi inversi, l'obiettivo dell'algoritmo di ricostruzione consiste, quindi, nel minimizzare tale funzione, riducendo l'errore tra l'immagine ricostruita e quella originale. Il modulo `loss` del framework fornisce diverse implementazioni standard per questo scopo.
- **Dataset**: una collezione di immagini appartenenti ad uno specifico dominio applicativo. Il modulo `datasets` non solo include alcuni dataset di riferimento pronti all'uso, ma fornisce anche strumenti per importare e manipolare dataset personalizzati, permettendo una gestione flessibile del caricamento dei dati (*data loading*).
- **Valutazione**: per quantificare in modo oggettivo le prestazioni del modello sul test set, vengono utilizzate le **metriche** di qualità dell'immagine (come MSE, PSNR o SSIM). Questi strumenti sono raccolti nel modulo `metric` e permettono di confrontare l'efficacia di diversi approcci di ricostruzione.

Per semplificare la fase di sviluppo sperimentale, il framework introduce il modulo `trainer`. Questa componente astrae la complessità del loop di training, fornendo strumenti predefiniti per il monitoraggio delle metriche in tempo reale e per la gestione della riproducibilità, uno dei problemi critici identificati nella ricerca attuale sui problemi inversi.

2.1.3 Modellazione fisica

Un elemento distintivo di DeepInverse è la gestione esplicita della fisica del problema tramite il modulo `physics`. Questo modulo consente la definizione esplicita dell'operatore di forward A come un oggetto lineare all'interno del grafo di calcolo di PyTorch, garantendo l'accesso immediato all'operatore aggiunto $A^\top$ e al calcolo dei gradienti. Tale astrazione permette di testare lo stesso algoritmo di ricostruzione su problemi fisici differenti (es. inpainting, denoising o tomografia) semplicemente sostituendo l'oggetto physics.

2.2 Confronto tra approcci Variazionali e Deep Learning

Per comprendere a pieno il potere della ricostruzione tomografica, oggetto di studio della tesi, è necessario analizzare e confrontare metodi classici e quelli basati sull'apprendimento automatico.

2.2.1 Approccio Variazionale (model-based)

Nei **metodi variazionali**, la ricostruzione è trattata come un problema di ottimizzazione deterministico. Nello specifico il problema inverso viene riformulato nella seguente equazione:

$$\hat{x} = \arg\min_x \frac{1}{2}|A_\xi x - y|_2^2 + \lambda \mathcal{R}(x) \tag{2.4}$$

Dove il primo termine $\frac{1}{2}\|A_\xi x - y\|_2^2$, è la data fidelity e il secondo termine, $\mathcal{R}(x)$, è il **regolarizzatore**, pesato dal parametro di regolarizzazione $\lambda > 0$. Quest'ultimo ha il compito di bilanciare la ricostruzione verso soluzioni che rispettino determinate proprietà fisiche o geometriche.

Uno dei regolarizzatori più utilizzato è la **Total Variation (TV)**, definita come la norma L_1 del gradiente dell'immagine, ovvero:

$$\mathcal{R}_{TV}(x) = |\nabla x|_1 = \int |\nabla x|, d\Omega \tag{2.5}$$

Il vantaggio principale dei metodi variazionali risiede nella loro natura di modelli "trasparenti". A differenza delle reti neurali profonde, spesso criticate per la loro capacità decisionale, ogni iterazione di un algoritmo variazionale ha un'interpretazione fisica o geometrica diretta. Inoltre, l'assenza di una fase di addestramento elimina il rischio di **bias indotto dai dati**; il risultato dipende esclusivamente dalla fisica del sistema e dal modello teorico scelto, rendendo il metodo robusto.

Matematicamente, se il funzionale di costo $J(x) = f(y, Ax) + \lambda \mathcal{R}(x)$ è **strettamente convesso**, è garantita l'esistenza di un unico minimo globale $\hat{x}$. Ciò comporta che algoritmi che verificano questa condizione convergano sempre verso la medesima soluzione, indipendentemente dal punto di partenza x_0.

Nonostante il rigore teorico, i metodi variazionali sono limitati dall'eccessiva semplicità dei prior definiti analiticamente. Di conseguenza questo spesso risulta in un fallimento di modellazione della complessità statistica di immagini mediche reali. Questa limitazione rappresenta la principale motivazione tecnologica dietro il passaggio verso i metodi basati su Deep Learning, capaci di apprendere prior molto più

ricchi e flessibili rispetto a quanto sia possibile descrivere con una semplice funzione analitica.

2.2.2 Approccio Deep Learning (data-driven)

I metodi basati su **Deep Learning** mirano a risolvere il problema inverso apprendendo una mappatura parametrica dei dati, spostando il fulcro della ricostruzione dalla modellazione matematica manuale all'estrazione di caratteristiche statistiche dai dataset. Questo permette di superare i limiti geometrici dei metodi variazionali. Inoltre, una volta addestrata, la rete può produrre ricostruzioni di alta qualità in frazioni di secondo, un vantaggio competitivo enorme rispetto alle centinaia di iterazioni richieste dai metodi variazionali.

Tuttavia, la principale criticità risiede nella generalizzazione. Una rete addestrata su una specifica anatomia o geometria (es. torace) potrebbe fallire o produrre artefatti se applicata ad un'altra differente (es. cranio). In più, esiste il rischio di **allucinazioni numeriche**, dove la rete genera dettagli anatomicamente plausibili ma non presenti nel paziente reale, a causa di bias nel training set [6].

2.3 Plug-and-Play

I metodi **Plug-and-Play** (PnP) rappresentano una classe di algoritmi iterativi il cui obiettivo è combinare termini di fedeltà ai dati (*data-fidelity*) con denoisers profondi, sfruttando schemi di ottimizzazione classica come **PGD** (Proximal Gradient Descent, noto anche come ISTA) o **ADMM** [6]. Sebbene approcci completamente data-driven abbiano mostrato risultati sensibilmente superiori rispetto all'utilizzo di regolarizzatori esplici (come la Total Variation), i loro output spesso non corrispondono alla soluzione di un problema di minimizzazione ben definito. Questa mancanza di interpretabilità è particolarmente limitante in applicazioni critiche come l'imaging medico, dove la coerenza fisica del risultato è un requisito fondamentale [5;6]. Il paradigma Plug-and-Play concretizza questa unione sostituendo i passaggi legati all'operatore prossimale del regolarizzatore con uno o più denoiser pre-addestrati (D_σ):

$$Prox_{\tau\lambda g} \longrightarrow D_\sigma \tag{2.6}$$

Questo approccio permette di disaccoppiare la gestione della fisica del problema dalla modellazione del prior. Invece di definire una funzione di regolarizzazione $g(x)$ in forma chiusa, si utilizza l'**implicito prior** appreso da un denoiser come DnCNN o DRUNet. Sotto specifiche condizioni di non-espansività del denoiser e convessità della fedeltà ai dati, è possibile garantire la **convergenza al punto fisso** (fixed point convergence) [6] dell'algoritmo.

2.3.1 Definizione matematica del metodo

Il framework Plug-and-Play è stato introdotto per la prima volta da **Venkatakrishnan et al. nel 2013**[6] per la ricostruzione di immagini model-based. Il primo algoritmo di ottimizzazione considerato in questo contesto è stato l'**Alternating Direction Method of Multipliers** (ADMM), un classico schema di proximal splitting utilizzato per minimizzare funzioni composte. Nel caso della ricostruzione di immagini, il modello può essere formulato matematicamente partendo da una misurazione degradata y e da un dato incognito x. Definendo $p(x|y)$ come la likelihood condizionale e $p(x)$ come il prior della variabile x, la stima della **Massima Probabilità a Posteriori** (MAP), indicata con $\hat{x}$ si ottiene come segue:

$$\hat{x} = \arg\max_x p(y|x)p(x) = \arg\min_x f(x,y) + g(x) \tag{2.7}$$

In questa espressione, f rappresenta il termine di fedeltà ai dati, mentre g rappresenta il termine di regolarizzazione o prior[2]. Un esempio classico di questa formulazione è la regolarizzazione Total Variation (TV) applicata in presenza di rumore gaussiano additivo. In questo scenario, il termine di data fidelity è definito come $f(x,y) = \frac{1}{2\sigma^2}\|Ax - y\|_2^2$, mentre il termine di prior è $g(x) = \lambda\|\nabla x\|_1$.

Per risolvere il problema di minimizzazione con funzione f e g convesse, gli algoritmi di proximal splitting come ADMM prevedono l'applicazione alternata dei singoli **operatori prossimali** ($prox_f, prox_g$) o dei subgradienti ($\partial f, \partial g$). L'intuizione fondamentale del metodo PnP risiede nell'osservazione che il passaggio di regolarizzazione del prior può essere interpretato e sostituito da un'operazione di denoising più generica e potenzialmente molto più complessa di un semplice operatore analitico.

2.3.2 Proximal Gradient Descent (PGD)

L'algoritmo **Proximal Gradient Descent**, noto anche come *composite gradient descent* o *generalized gradient descent* è progettato per risolvere problemi di ottimizzazione in cui la funzione obiettivo ha una struttura composta[8]:

$$\min_x f(x) = g(x) + h(x) \tag{2.8}$$

Si assume che g sia una funzione convessa e differenziabile, mentre h sia convessa ma non necessariamente differenziabile. In questo contesto il classico algoritmo di discesa del gradiente non può essere applicato direttamente. L'idea alla base del PGD consiste nel trattare i due termini in modo distinto: si effettua un'approssimazione quadratica solo sulla parte differenziabile (g), lasciando invariata la funzione h. Di conseguenza, l'aggiornamento per trovare il nuovo punto x_{k+1} diventa la soluzione del seguente problema di minimizzazione locale:

$$x_{k+1} = \arg\min_x \frac{1}{2\eta}|x - (x_k - \eta\nabla g(x_k))|_2^2 + h(x) \tag{1.9}$$

In questa formulazione, il primo termine agisce come una penalità di prossimità, imponendo che il punto x_{k+1} rimanga vicino all'aggiornamento fornito dal gradiente di g. Il secondo termine, invece, assicura che il valore della funzione h rimanga contenuto. Questa operazione viene formalizzata attraverso il **mapping prossimale** (prox_h), definito come l'operatore che minimizza la somma tra una penalità di distanza quadratica e la funzione h:

$$\text{prox}h(x) = \arg\min z \, \frac{1}{2}|z - x|_2^2 + h(z) \tag{1.10}$$

Dato un passo $\eta > 0$, l'aggiornamento iterativo del sistema si riduce a due fasi combinate in un unico passaggio:

- **Gradient Step**: si compie un passo nella direzione del gradiente di g.
- **Proximal Step**: si applica l'operatore prossimale al risultato ottenuto.

$$x_{k+1} = \text{prox}_{\eta h}(x_k - \eta \nabla g(x_k)) \tag{1.11}$$

Algorithm 1: Proximal Gradient Descent (PGD)

Input: Punto iniziale $x^{(0)}$, funzione differenziabile g, funzione convessa h, step size $\{t_k\}_{k=1,2,\ldots}$
Output: Minimo stimato $\hat{x}$

repeat
  Calcola il gradiente della parte liscia: $\nabla g(x^{(k-1)})$
  Esegui il passo di discesa nel gradiente:
  $z^{(k)} \leftarrow x^{(k-1)} - t_k \nabla g(x^{(k-1)})$
  Applica l'operatore prossimale relativo a h:
  $x^{(k)} \leftarrow \text{prox}_{h,t_k}(z^{(k)})$ `// dove` $\texttt{prox}_{h,t}(z) = \arg\min_u \frac{1}{2t}\|z - u\|_2^2 + h(u)$
  $k \leftarrow k + 1$
until *criterio di arresto soddisfatto*
return $x^{(k)}$

2.3.3 Denoiser come Prior

Come analizzato nelle sezioni precedenti, il framework PnP si basa sulla sostituzione dell'operatore prossimale con un prior implicito definito da un denoiser preaddestrato. La scelta del modello di denoising è cruciale, poiché ne determina la capacità di generalizzazione e la qualità della ricostruzione finale. All'interno di

DeepInverse, vengono principalmente considerati due approcci allo stato dell'arte proposti da **Zhang et al.**[6]:

- **DnCNN** (Zhang et al., 2017): rappresenta una pietra miliare nel denoising basato su Deep Learning. A differenza degli approcci classici che tentano di mappare direttamente l'immagine corrotta nella versione pulita, la DnCNN adotta una strategia di **residual learning**. La rete è progettata per stimare esclusivamente il rumore residuo $R(y)$, definendo l'immagine pulita come:

$$D(y) = y - R(y) \tag{2.9}$$

  dove y è l'input corrotto. L'architettura standard è composta da un numero variabile di layer (solitamente 17 o 20) con convoluzioni 3 x 3, Batch Normalization e ReLU. Una caratteristica distintiva della DnCNN è la sua capacità di gestire il **blind denoising**: la rete può essere addestrata su un ampio range di livelli di rumore σ, eliminando la necessità di specificare il rumore esatto in fase di inferenza.

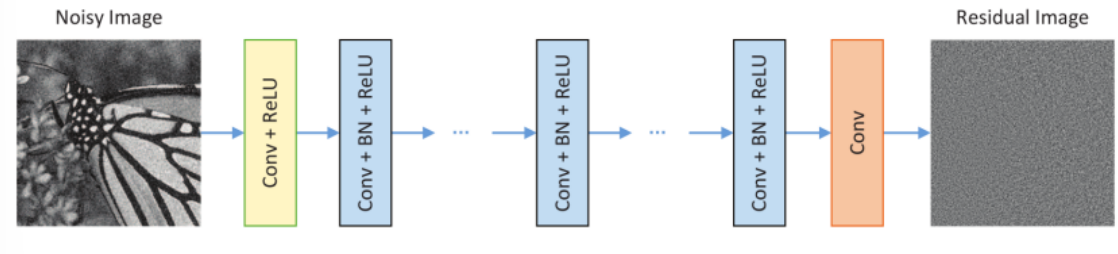

Figura 2.2: Architettura del denoiser DnCNN

- **DRUNet** (Zhang et al., 2021): è una rete più complessa che combina i vantaggi delle *U-Net* con i blocchi residuali, specificamente ottimizzata per essere inserita in algoritmi iterativi PnP. L'architettura si sviluppa su 4 livelli di scala con un numero di canali crescente (64, 128, 256, 512). A differenza della DnCNN, la DRUNet è un denoiser **non-blind**, ovvero richiede come input aggiuntivo una mappa del livello di rumore (*noise level map*). Zhang sottolinea che questa caratteristica è fondamentale per il PnP: durante le iterazioni dell'algoritmo (come PGD), il parametro di regolarizzazione agisce con un "livello di rumore efficace" che cambia nel tempo. La DRUNet permette di modulare la forza del denoising ad ogni passo, offrendo un controllo superiore alla convergenza. Inoltre, l'impiego di convoluzioni *bias-free* ne migliora la stabilità quando applicata iterativamente, riducendo il rischio di accumulo di errori sistematici (bias) nelle fasi finali della ricostruzione.

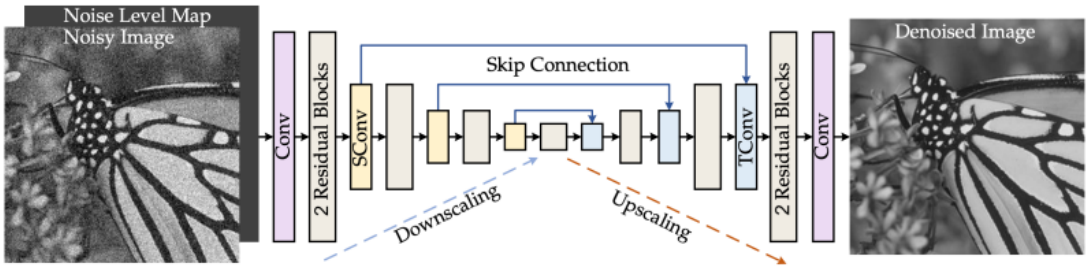

Figura 2.3: Architettura del denoiser DRUNet

2.3.4 Convergenza e proprietà del punto fisso

L'analisi della convergenza risulta essere uno step fondamentale per poter garantire che l'algoritmo utilizzato non solo produca immagini visivamente coerenti, ma che sia numericamente stabile e affidabile. In questo contesto, la stabilità del PnP può essere dimostrata attraverso la **teoria dei punti fissi**. Se consideriamo il problema come la somma di una funzione "liscia" g (data-fidelity) e una "non liscia" h (prior), il successo del processo dipende dalla lunghezza del passo η scelto per ogni iterazione.

Seguendo l'analisi della discesa del gradiente prossimale, se il passo η è abbastanza piccolo (minore di $1/L$, dove L è la costante di Lipschitz del gradiente), l'errore diminuisce costantemente. Matematicamente, l'algoritmo converge con un tasso di $O(1/k)$[8]:

$$f(x^{(k)}) - f^* \leq \frac{|x^{(0)} - x^*|_2^2}{2\eta k} \tag{2.10}$$

Questo significa che la velocità con cui l'algoritmo raggiunge la soluzione ottimale è la stessa dei metodi matematici più robusti. Per assicurare che questa sequenza di immagini non oscilli all'infinito ma si fermi su un risultato stabile (convergenza), il denoiser deve essere "ben condizionato" (tecnicamente si dice che deve essere **non-espansivo**), ovvero non deve amplificare eccessivamente le variazioni tra un'iterazione e l'altra.

In conclusione, l'algoritmo PnP riesce a coniugare la stabilità matematica dei metodi variazionali con la potenza espressiva del Deep Learning. Questo paradigma permette, quindi, di fondere i vantaggi di entrambi i mondi, ottenendo un punto di equilibrio tra il rigore della modellazione fisica e la complessità della distribuzione statistica dei dati reali.

2.4 Approccio Unfolded

Si va ora ad introdurre un approccio basato sulle reti **unrolled** (o **unfolded**). A differenza dei metodi Plug-and-Play che utilizzano reti neurali pre-addestrate come moduli esterni, gli approcci unrolled progettano la rete stessa ricalcando la struttura di un algoritmo di ottimizzazione iterativo. Tale metodologia permette di superara la natura di "black-box" tipica delle reti end-to-end, garantendo che ogni passaggio della rete mantenga un preciso significato fisico e matematico.

Nello specifico un layer di una rete unrolled corrisponde a una singola iterazione dell'algoritmo che si intende "srotolare". Per esempio, se si sta modellando la discesa del gradiente, ogni layer esegue lo stesso aggiornamento di passo che il gradient descent effettuerebbe. Ogni neurone applica le medesime operazioni matematiche di uno step del processo iterativo, trasformando l'input e passandolo al layer successivo. La rete così ottenuta può, quindi, essere addestrata e gli iperparametri possono essere appresi, mentre le informazioni sul modello diretto sono integrate nelle matrici dei pesi e la conoscenza a priori viene incorporata nella funzione di attivazione.

2.4.1 Learned Primal-Dual

Per ottenere un ulteriore aspetto di confronto su cui poter valutare le prestazioni del modello PnP, viene preso in considerazione l'algoritmo **unfolded primal-dual** di Adler and Öktem[5], in cui sia la data fidelity che il prior sono moduli learned, distinti per ogni iterazione. Si considera prima la versione iterativa dell'algoritmo per trasformarla successivamente nella sua versione unfolded. Il metodo classico primal-dual per i programmi lineari e i problemi di ottimizzazione si basa sull'esistenza di soluzioni ammissibili per il problema primale e quello duale. Dato il programma lineare:

$$\min c^T x \quad \text{soggetto a: } Ax \geq b, x \geq 0 \tag{2.11}$$

il suo duale è definito come:

$$\max b^T y \quad \text{soggetto a: } A^T y \leq c, y \geq 0 \tag{2.12}$$

dove $A \in \mathbb{Q}^{m \times n}$ e T denota la trasposta. Nel metodo primale-duale, si assume di avere una soluzione duale ammissibile y; inizialmente si può impostare $y = 0$. Il metodo cerca quindi di trovare una soluzione primale x che rispetti le condizioni di slackness complementare rispetto a y. Nello specifico, esistono due tipi di condizioni di slackness complementare:

- Primali: $x_j > 0 \Rightarrow A_j y = c_j$.
- Duali: $y_i > 0 \Rightarrow A_i x = b_i$.

Se queste condizioni sono soddisfatte, sia x che y sono ottimali. In caso contrario, l'algoritmo aggiorna la soluzione duale per incrementare il valore della funzione obiettivo. Nella versione Learned, quindi "srotolando l'algoritmo", l'architettura della rete sostituisce gli operatori prossimali classici con blocchi convoluzionali addestrabili[5].

Capitolo 3

Caso di Studio

In questo capitolo viene esposta nel dettaglio l'architettura del progetto sperimentale, il cui obiettivo risiede nella ricostruzione di immagini bidimensionali in scala di grigi a partire da proiezioni tomografiche corrotte da rumore. Le immagini prese in considerazione sono definite *sintetiche* in quanto, non derivando da pazienti reali, rappresentano dati generati artificialmente. Questo permette di operare in un ambiente controllato grazie alla disponibilità di una ground truth nota, facilitando così una valutazione oggettiva e quantitativa delle prestazioni degli algoritmi di ricostruzione utilizzati.

Vengono inoltre analizzate le scelte implementative adottate per la generazione dei risultati, con particolare attenzione agli operatori e agli strumenti forniti dal framework DeepInverse. Nello specifico, verrà illustrato come tali componenti siano stati integrati e configurati per modellare l'intero workflow di imaging, seguendo i paradigmi introdotti nel capitolo precedente.

3.1 Dataset utilizzato

Per lo svolgimento degli esperimenti è stato utilizzato il Dataset **COULE**, reperibile sul repository pubblico *kaggle.com*. In linea con quanto anticipato, si tratta di un dataset generato artificialmente che contiene immagini in scala di grigi composte da ellissi e linee uniformi sovrapposte. Tali strutture geometriche sono progettate per simulare in modo semplificato l'anatomia umana, rendendo il dataset uno strumento valido per l'addestramento e la valutazione di reti neurali in applicazioni di Computed Tomography (CT), quali la *Sparse-view CT* e il *Denoising* [2]. Inoltre, il dataset contiene immagini non corrotte, ovvero ciò che viene definito *ground truth*. Tale caratteristica risulta cruciale per la fase di training, in quanto permette di utilizzare

tali immagini per il calcolo della funzione di perdita, guidando la rete verso una ricostruzione che sia il più fedele possibile al dato reale.

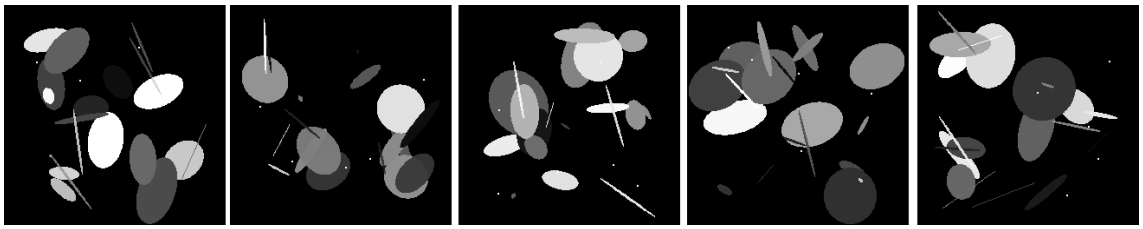

Figura 3.1: Esempi di campioni presenti all'interno del dataset Koule

Il dataset presenta una struttura predefinita composta da un numero contenuto di campioni: 400 immagini destinate alla fase di addestramento (training set) e 30 immagini riservate alla fase di valutazione (test set). Sebbene questa suddivisione sia quella suggerita dal repository originale, il framework DeepInverse permette di rimodulare tali proporzioni in base alle necessità sperimentali. Dal punto di vista tecnico, le immagini hanno una risoluzione di 256 x 256 pixel, con un'intensità luminosa codificata in un range di valori interi [0, 255]. Una caratteristica rilevante del dataset è la presenza di elementi ad alto contrasto (valori molto vicini a 255, corrispondenti al bianco puro), che rappresentano le componenti più critiche da ricostruire correttamente, specialmente in scenari di Sparse Tomography, dove l'incompletezza dei dati tende a generare artefatti in corrispondenza delle variazioni brusche di intensità.

3.1.1 Scelta implementativa 1: creazione e suddivisione del dataset

Il framework DeepInverse mette a disposizione il modulo `Datasets`, progettato per definire nuovi set di dati o generare varianti a partire da collezioni esistenti. Tale modulo integra diverse trasformazioni applicabili alle immagini per ottimizzare la fase di caricamento.

In questo caso specifico, data la limitata numerosità campionaria del test set originale, si è scelto di ignorare la suddivisione suggerita dal repository per incrementare le immagini destinate alla valutazione del modello. La procedura ha previsto il download locale dell'intero dataset tramite l'utility `kagglehub`. Successivamente, è stato applicato uno shuffle stocastico dei dati utilizzando un seed fisso, garantendo così la piena riproducibilità dell'esperimento. È stata, infine, definita una nuova partizione 80-20: l'80 % dei dati andrà a caratterizzare il nuovo training set, mentre il restante 20% verrà utilizzato per la validazione e il test del modello.

3.1.2 Scelta implementativa 2: data augmentation

Sebbene la nuova ripartizione garantisca un test set più robusto, la conseguente riduzione delle immagini nel training set potrebbe limitare la capacità di generalizzazione dei denoiser. Per risolvere tale criticità, è stata introdotta una strategia di **data augmentation**, che permette di generare stocasticamente nuovi campioni applicando trasformazioni geometriche sulle immagini esistenti. Nello specifico, ad ogni iterazione di caricamento vengono applicate le seguenti operazioni:

- Flip orizzontale o verticale, ciascuno con una probabilità del 50%;
- Rotazione casuale scelta tra gli angoli dell'insieme $\{0°, 90°, 180°, 270°\}$

Queste trasformazione posso essere applicate anche in sequenza, permettendo non solo di aumentare virtualmente la dimensione del dataset, ma anche di mitigare i possibili **bias strutturali** presenti nei dati originali (come, ad esempio, una prevalenza di ellissi con un orientamento preferenziale), rendendo il modello finale più robusto rispetto alle variazioni geometriche dell'anatomia simulata.

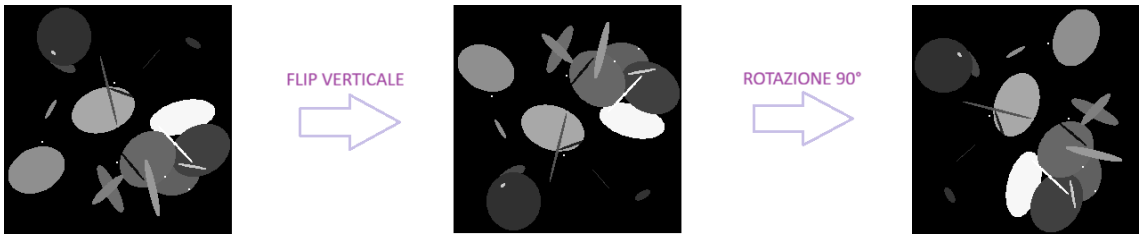

Figura 3.2: Esempio di applicazione sequenziale di un flip verticale e una rotazione di 90°

3.2 Operatore fisico di acquisizione tomografica

Poiché il dataset fornisce esclusivamente le immagini originali, è necessario simulare il processo di acquisizione per ottenere i relativi **sinogrammi**. Tale operazione viene eseguita attraverso il modulo `physics` di DeepInverse. Per riprodurre le caratteristiche dell'oggetto di studio, il framework offre due opzioni principali: `TomographyWithAstra` e `Tomography`. Entrambe implementano la trasformata di Radon discreta, definendo l'operatore di proiezione A (*forward projection*) e il suo operatore aggiunto $A^\top$ (*back projection*). Tuttavia, la classe `TomographyWithAstra` risulta essere sensibilmente più efficiente, in quanto sfrutta il backend della libreria **ASTRA Toolbox** per il calcolo accelerato su GPU[4]. Di conseguenza, è stata scelta quest'ultima per modellare il sistema fisico.

3.2.1 Geometria e configurazione

Per stabilire un ambiente di test bilanciato tra complessità e stabilità numerica, è stata impostata una geometria di acquisizione con **180 angoli di proiezione** distribuiti uniformemente nell'intervallo $[0°, 180°]$. Al fine di simulare condizioni realistiche, il sinogramma risultante viene corrotto da un rumore Gaussiano additivo con deviazione standard $\sigma = 0.008$. La scelta è ricaduta su una geometria di tipo **fan-beam** (a fascio divergente), la cui configurazione richiede i seguenti parametri:

- **Dimensioni del detector**, ovvero il numero di pixel del sensore. Quest'ultimo è stato impostato per essere superiore alla diagonale dell'immagine (altezza $\times \sqrt{2}$). Una corretta campionatura del detector insieme ai parametri geometrici, è essenziale per evitare artefatti di troncamento o distorsione nella ricostruzione tramite *Filtered Back Projection* (FBP).
- **Distanze sorgente-oggetto-detector**. Sono stati definiti il raggio della sorgente (distanza tra sorgente e centro di rotazione) e il raggio del detector (distanza tra detector e centro di rotazione). Questi valori determinano il grado di divergenza del fascio e influenzano direttamente il fattore di ingrandimento dell'immagine acquisita.

3.2.2 Scelta implementativa 1: calcolo di una stepsize stabile

Per garantire la convergenza dell'algoritmo PGD utilizzato nel paradigma PnP, lo step-size η (passo di aggiornamento) deve essere scelto in funzione delle proprietà spettrali dell'operatore fisico. Nello specifico, la teoria dell'ottimizzazione richiede che $\eta \leq 1/L$, dove L è la norma spettrale (o costante di Lipschitz del gradiente) dell'operatore $A^\top A$[8]. Per ottenere una stima accurata di L, è stato utilizzato il **metodo delle potenze**, che calcola il modulo del più grande autovalore dell'operatore[8]. Lo step-size finale è stato quindi calcolato come:

$$\eta = \frac{1}{L} \times \alpha \tag{3.1}$$

dove α è un fattore di sicurezza (0.1) introdotto per prevenire instabilità numeriche, oscillazioni o divergenze che potrebbero verificarsi a causa di errori di approssimazione durante il calcolo iterativo.

3.3 Training dei denoiser

Come discusso nel capitolo precedente, l'efficacia dell'algoritmo PnP risiede nella qualità del prior implicito fornito dal denoiser. Per questo studio, si è scelto di confrontare le prestazioni di un modello blind (*DnCNN*) e di un modello non-blind

(*DRUNet*). Entrambe le reti sono state sottoposte a una fase di **fine tuning**, una tecnica di *transfer learning* che non parte da un'inizializzazione casuale dei pesi, ma ottimizza i parametri pre-addestrati messi a disposizione da DeepInverse per adattarli alle caratteristiche specifiche del dataset utilizzato. Sebbene le due architetture richiedano procedure di ottimizzazione differenti, entrambi i processi di training condividono l'utilizzo di una **funzione di loss composita**. L'impiego della sola loss L_1, pur essendo una scelta standard in presenza di rumore Gaussiano, tende a produrre risultati eccessivamente levigati e poveri di dettagli ad alta frequenza. Per ovviare ciò, la funzione obiettivo è stata definita come combinazione pesata delle seguenti componenti:

- **L1 Loss**: agisce come termine di fedeltà globale, minimizzando la differenza assoluta pixel per pixel tra l'immagine ricostruita e la ground truth.
- **SSIM Loss**: basata sull'indice di similarità strutturale, assicura che le relazioni locali tra pixel siano preservate, mantenedo il contrasto e la struttura dell'immagine.
- **Gradient Loss**: penalizza la discrepanza tra i gradienti delle immagini, costringendo la rete a preservare i bordi netti e prevenendo l'effetto "sfocatura".

3.3.1 Strategia di addestramento per DRUNet

Il fine-tuning della DRUNet è stato articolato in due fasi distinte per riflettere la natura iterativa del PnP:

1. *Fase di apprendimento strutturale*: 30 epoche dedicate all'apprendimento della geometria e delle strutture fondamentali delle immagini. Il livello di rumore σ è stato campionato casualmente nell'intervallo $[0.01, 0.1]$. La funzione di perdita è stata configurata come: $0.5 \cdot L_1 + 0.2 \cdot (1.0 - SSIM) + 0.3 \cdot \text{GradLoss}$. Il learning rate è stato dimezzato dopo 15 epoche per stabilizzare la convergenza.
2. *Fase di rifinitura*: 15 epoche supplementari per conferire al denoiser una capacità di intervento più "chirurgica", utile a rimuovere residui di rumore molto deboli. Questa fase modella le iterazioni finali del PnP, dove l'immagine è già parzialmente pulita. L'intervallo di rumore è stato ridotto a $[0.001, 0.02]$ e la perdita è stata sbilanciata a favore del dettaglio: $0.3 \cdot L_1 + 0.2 \cdot (1.0 - SSIM) + 0.5 \cdot \text{GradLoss}$.

In questa seconda fase è stato introdotto il meccanismo di **Batch Identity**, che consiste nell'includere condizionalmente durante il training alcuni batch privi di rumore ($\sigma = 0$). Questa tecnica è fondamentale per evitare l'iper-correzione dell'immagine da parte della rete e per garantire una maggiore stabilità del punto fisso durante la ricostruzione iterativa[6].

3.3.2 Strategia di addestramento per DnCNN

L'addestramento della rete DnCNN ha seguito una procedura simile alla prima fase di DRUNet per quanto concerne il numero di epoche e i pesi della funzione di perdita. Tuttavia, sono state adottate due variazioni per adattarsi alla natura blind della rete: il learning rate viene dimezzato ogni 10 epoche e l'intervallo del livello di rumore è stato esteso a $[0.001, 0.1]$ per coprire l'intera gamma di degradazioni possibili in un singolo ciclo di addestramento.

3.4 Configurazione dell'algoritmo PnP e Unfolded

Per concretizzare il confronto metodologico introdotto nei capitoli precedenti, sono state implementate due pipeline di ricostruzione distinte: una basata sul denoiser DnCNN e l'altra sulla rete DRUNet. Entrambi gli approcci utilizzano l'algoritmo Proximal Gradient Descent (PGD) come schema di ottimizzazione sottostante e condividono la medesima funzione di data-fidelity. Nello specifico, è stata adottata la **Mean Squared Error (L_2)** che rappresenta la scelta statisticamente ottimale quando il rumore presente nel sinogramma è di natura gaussiana[8;2]. Un elemento chiave della configurazione è l'utilizzo della **filtered back projection** (FBP) come iterato iniziale (x_0) per le ricostruzioni. Questa scelta è preferibile rispetto all'impiego di un iterato nullo (costante a zero) per le seguenti ragioni tecniche:

- Recupero delle alte frequenze. Il filtro a rampa applicato nelle FBP compensa la sfocatura naturale della semplice back-projection, amplificando le alte frequenze. Sebbene ciò incrementi la visibilità del rumore, fornisce una base strutturale dell'immagine molto più definita.
- Accelerazione della convergenza. Partire da una stima FBP permette all'algoritmo PnP di iniziare il processo di raffinamento da un'immagine già anatomicamente plausibile. Il denoiser può, quindi, concentrarsi fin da subito sulla rimozione degli artefatti e del rumore, portando a una convergenza più rapida verso il punto fisso rispetto a un'inizializzazione priva di informazioni strutturali.

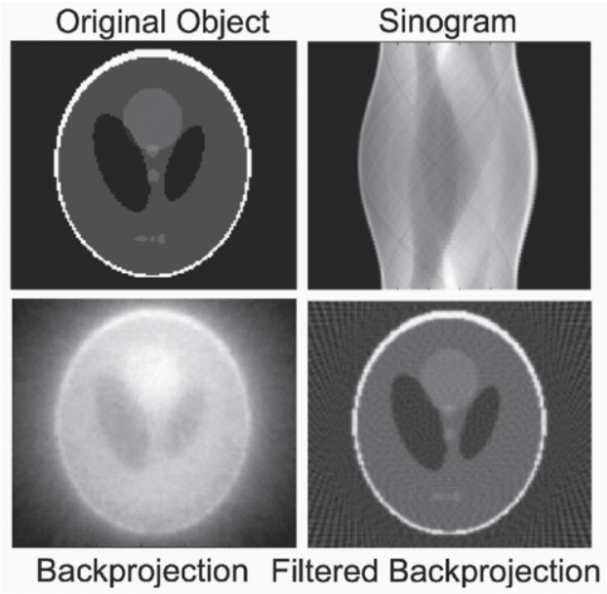

Figura 3.3: Schema dell'iterato inziale con FBP contro semplice back-projection

Inoltre, lo step-size è stato mantenuto costante secondo la stima della norma spettrale calcolata nella sezione 2.3.2 per garantire le condizioni di convergenza e stabilità introdotte precedentemente.

3.4.1 Continuation Schedule

A differenza della ricostruzione basata su DnCNN che, operando in modalità blind, non richiede parametri di rumore in input, per la DRUNet è stato adottato un approccio di **Continuation Schedule**. Questa strategia consiste nel suddividere il processo di ottimizzazione in segmenti sequenziali, ciascuno caratterizzato da parametri differenti[6].

Nello specifico, le variabili interessate dal continuation schedule sono il numero di iterazioni e il livello di rumore σ fornito in input al denoiser. La configurazione ottimale è stata determinata seguendo un approccio euristico, analizzando sia l'andamento quantitativo delle metriche sia la qualità visiva degli iterati intermedi (ad esempio, osservando l'iterato iniziale si può notare una forte dominanza del rumore. Di conseguenza, per le prime iterazioni viene impostato un valore di σ elevato). Per determinare il numero esatto di iterazioni per ogni segmento, è stato utilizzato come supporto un grafico dell'andamento del PSNR. Si ricerca il punto di massimo incremento prima che si stabilizzi oppure si verifichi un calo, sintomo di un eccessivo livellamento dei dettagli introdotto dal denoiser. Seguendo questa metodologia, la configurazione che ha prodotto i risultati migliori è la seguente:

- Livelli di rumore(σ_{sch}): $[0.1, 0.01, 0.001]$
- Iterazioni per segmento($iter_{sch}$): $[3, 5, 20]$

Tale approccio non è stato applicato alla pipeline con DnCNN poiché, grazie alla sua natura blind, la rete adatta autonomamente la forza del filtraggio. In questo secondo caso, è stato sufficiente determinare il numero totale di iterazioni (fissato a 6) monitorando la curva di convergenza complessiva.

3.4.2 Implementazione Unfolded

L'architettura Learned Primal-Dual è stata realizzata srotolando l'algoritmo iterativo in una sequenza di blocchi funzionali addestrabili. In questo schema, ogni "strato" della rete neurale non è una generica operazione matematica, ma rappresenta una specifica iterazione del processo di ricostruzione che integra la fisica del problema[5]. Il cuore del modello è costituito dalla classe `PDNetIteration`, che definisce una singola iterazione dell'algoritmo attraverso due passaggi fondamentali:

- **Dual Step (Data Fidelity)**: implementato tramite la classe `fStepPDNet`, opera nello "spazio dei dati" (il sinogramma). Riceve l'attuale stima dell'immagine e la confronta con i dati acquisiti per garantire che la ricostruzione sia coerente con le misure fisiche effettuate.
- **Primal Step (Prior)**: implementato dalla classe `gStepPDNet`, agisce invece nello "spazio immagine". Questo modulo utilizza una rete neurale (`PDNetPrior`) che impara a riconoscere le forme geometriche (le ellissi) e a regolarizzare la ricostruzione, eliminando artefatti e rumore senza perdere i dettagli strutturali.

L'algoritmo srotolato è stato configurato per un totale di 8 iterazioni. Un aspetto fondamentale di questa implementazione è che ogni iterazione possiede i propri parametri addestrabili indipendenti; ciò permette al modello di "imparare" una strategia differenziata: ad esempio, le prime iterazioni possono concentrarsi sulla rimozione degli artefatti grossolani, mentre le ultime rifiniscono i dettagli più sottili.

Per aumentare la capacità del modello di elaborare informazioni complesse, l'intero sistema è stato addestrato per 5 epoche totali, utilizzando una funzione di perdita basata sull'errore quadratico medio (MSE).

Capitolo 4

Risultati e analisi

In questo capitolo vengono presentati e analizzati i risultati ottenuti dalle sperimentazioni numeriche. L'obiettivo principale è valutare l'efficacia del paradigma Plug-and-Play nella ricostruzione tomografica, confrontando le prestazioni delle reti addestrate con i metodi classici della letteratura. L'analisi si articola su tre livelli fondamentali:

1. **Analisi Quantitativa**: l'utilizzo di metriche standard permette di stabilire dei benchmark oggettivi per misurare la fedeltà della ricostruzione rispetto alla ground truth.
2. **Analisi della Convergenza**: attraverso lo studio dell'andamento iterativo delle metriche, viene verificata la stabilità numerica degli algoritmi e la capacità dei denoiser di agire come regolarizzatori efficaci all'interno del ciclo di ottimizzazione.
3. **Analisi della robustezza**: l'impiego di rappresentazioni grafiche basate su *boxplot* consente di osservare la distribuzione dell'errore e la deviazione standard delle performance sull'intero test set, garantendo che i risultati non siano limitati a singoli casi favorevoli ma siano generalizzabili.

4.1 Metriche di confronto

La valutazione della qualità delle immagini ricostruite rappresenta uno dei principali strumenti di controllo impiegato nell'imaging computazionale. Per calibrare correttamente il modello e convalidare i risultati della fase di test, è necessario introdurre dei parametri numerici che quantifichino la distanza tra la stima prodotta dall'algoritmo e l'immagine originale.

4.1.1 Mean Squared Error (MSE)

La **mean squared error (MSE)** rappresenta una delle modalità più diffuse per quantificare la differenza tra valori stimati e i valori reali dell'oggetto ricostruito. In statistica, l'MSE è definita come una *funzione di rischio* e corrisponde al valore atteso della **squared error loss** o **quadrartic loss**[7].

Dal punto di vista probabilistico, l'MSE misura la media dei quadrati degli "errori", intesi come la discrepanza tra la stima e l'oggetto reale. Esiste un legame profondo tra questa metrica e il concetto di **cross-entropy**: nel caso di dati affetti da rumore con distribuzione Gaussiana, minimizzare l'MSE equivale a minimizzare la cross-entropy tra la distribuzione dei dati originali e quella delle predizioni del modello. In questo senso, l'MSE funge da indicatore sintetico che comprime l'intero dataset di training in un singolo valore scalare, quantificando la capacità del modello di riprodurre la realtà osservata. Tuttavia, tale compressione può risultare limitante: essendo un valore puramente numerico, l'MSE non fornisce indicazioni su quali aspetti dell'immagine (bordi, contrasto) siano stati ricostruiti fedelmente e quali invece presentino artefatti. Per questo motivo, sarebbe riduttivo limitarsi ad utilizzare una singola metrica.

Matematicamente, si considerino due immagini x e y di dimensione finita, dove x rappresenta la *ground truth* e y la ricostruzione. Sia N il numero totale di pixel dell'immagine, l'MSE tra le due immagini è definito come:

$$\mathrm{MSE}(x, y) = \frac{1}{N} \sum_{i=1}^{N} (x_i - y_i)^2 \tag{4.1}$$

dove x_i e y_i rappresentano il valore dell'i-esimo pixel rispettivamente in x e in y. Definito l'errore locale come $e_i = x_i - y_i$, l'MSE non è altro che la norma L_2 al quadrato dell'errore, mediata sul numero di campioni.

$$\mathrm{MSE}(x, y) = \frac{1}{N} \|\mathbf{e}\|_2^2 \tag{4.2}$$

L'impiego dell'MSE nell'ottimizzazione offre vantaggi significativi:

- Semplicità computazionali: il calcolo delle metriche richiede operazioni elementari ed estremamente efficienti in termini di memoria, poiché ogni pixel può essere processato indipendentemente.
- Proprietà matematiche: la funzione è convessa, simmetrica e differenziabile, caratteristiche che la rendono ideale per gli algoritmi basati sulla discesa del gradiente.

Nonostante queste proprietà, l'MSE soffre di limiti nella percezione visiva umana: un'immagine con un leggero spostamento geometrico potrebbe avere un MSE molto alto pur essendo visivamente perfetta, mentre un'immagine eccessivamente levigata potrebbe avere un MSE basso pur avendo perso dettagli fondamentali. Per tale motivo, nella letteratura dell'imaging, l'MSE viene spesso utilizzato come base per il calcolo del **Peak Signal-to-Noise Ratio (PSNR)**, o integrato da metriche differenti.

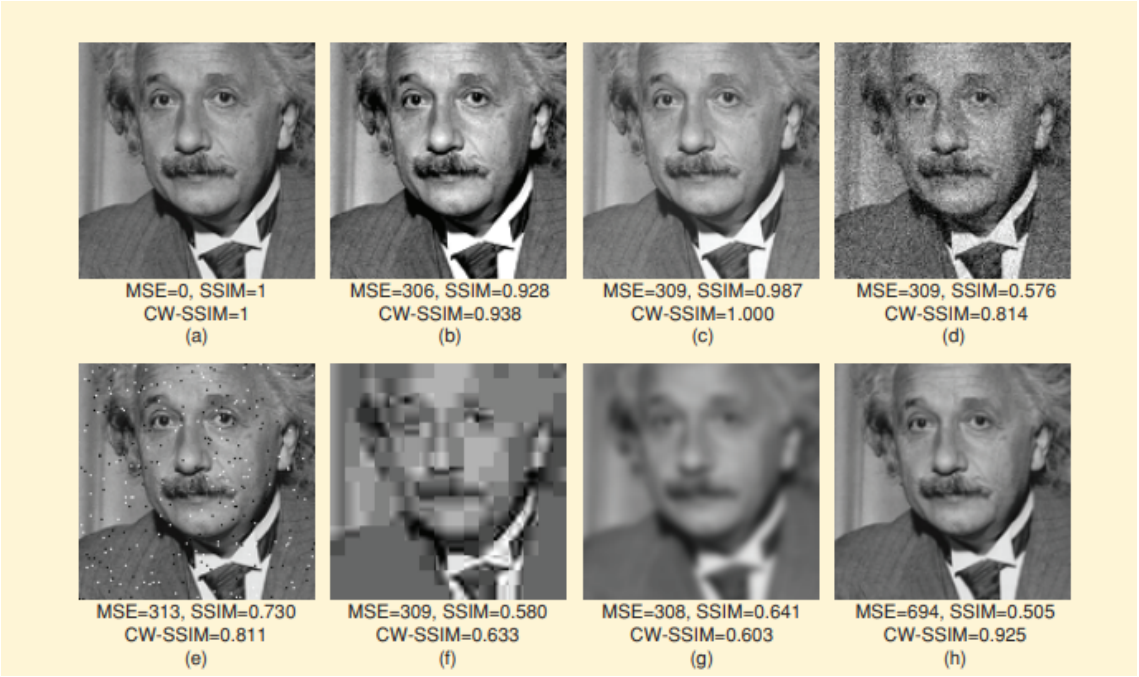

Figura 4.1: Confronto tra la fedeltà di immagini modificate con tipi diversi di distorsioni. (a) groud truth, (b) Aumento del contrasto, (c) cambio della luminanza, (d) contaminazione con rumore gaussiano, (e) contaminazione con rumore a impulsi, (f) compressione JPEG, (g) blurring, (h) spacial scaling.

4.1.2 Peak Signal-to-Noise Ratio (PSNR)

Il **Peak Signal-to-Noise Ratio** è una metrica definita come il rapporto tra la massima potenza possibile di un segnale e la potenza del rumore che ne corrompe la fedeltà della rappresentazione. Poiché molti segnali hanno un intervallo dinamico molto ampio, il PSNR viene solitamente espresso in scala logaritmica utilizzando i decibel (dB)[2]. Dato un segnale x (ground truth) e la sua ricostruzione y, il PSNR è calcolato a partire dall'MSE come segue:

$$\mathrm{PSNR}(x, y) = 10 \cdot \log_{10} \left(\frac{R^2}{\mathrm{MSE}(x, y)} \right) \tag{4.3}$$

Dove R rappresenta il massimo valore possibile dei pixel nell'immagine (il "picco" del segnale). Nel caso specifico di questo studio, poiché le immagini vengono normalizzate nell'intervallo [0, 1], il valore di R è pari a 1.

Dalla formulazione matematica si evince che il valore del PSNR aumenta proporzionalmente alla diminuzioen dell'MSE. Per questo motivo, un valore di PSNR più elevato è indice di una fedeltà superiore della ricostruzione. Al contratio, valori bassi suggeriscono una significativa discrepanza numerica tra le immagini confrontate.

Sebbene il PSNR sia lo standard de facto per la valutazione della qualità nelle applicazioni di compressione e ricostruzione di immagini, esso condivide alcuni dei limiti elencati precedentemente dell'MSE. Basandosi esclusivamente su una differenza quadratica calcolata pixel per pixel, la metrica non tiene conto delle complesse proprietà biologiche del sistema visivo umano (*Human Visual System, HVS*), il quale percepisce le distorsioni in modo differenziato a seconda della loro struttura. Di conseguenza, due immagini caratterizzate dal medesimo valore di PSNR possono presentare una qualità percepita profondamente diversa: ad esempio, l'occhio umano tende a tollerare maggiormente un rumore bianco distribuito uniformemente rispetto a distorsioni strutturate, artefatti di 'blurring' o sbavature dei bordi. Sebbene tali degradazioni producano lo stesso calo quantitativo nel rapporto segnale-rumore, l'impatto sulla leggibilità diagnostica dell'immagine risulta essere molto differente.

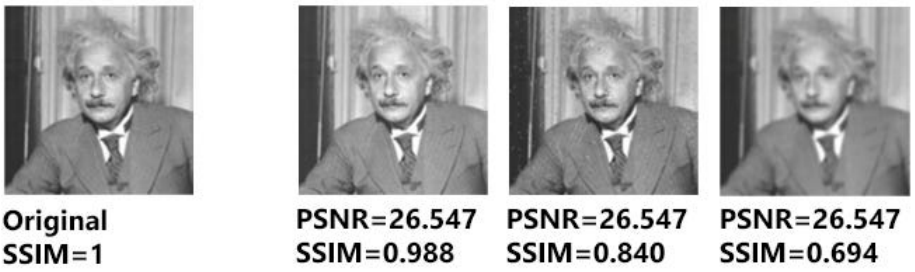

Figura 4.2: Confronto tra la fedeltà di immagini modificate con tipi diversi di distorsioni. (a) groud truth, (b) aumento del contrasto, (c) degradazione con rumore bianco uniforme, (d) blurring.

4.1.3 Indice di similarità strutturale (SSIM)

L'**indice di similarità strutturale (SSIM)** è una metrica che quantifica la differenza strutturale tra due immagini, riflettendo la differenza percepita dal sistema visivo umano. A differenza del PSNR, l'SSIM si fonda sul presupposto che le imma-

gini naturali possiedano una struttura intrinseca, caratterizzata da forti correlazioni spaziali tra pixel[7].

Poiché il sistema visivo umano si è evoluto per estrarre informazioni strutturali dalle scene, una misura di fedeltà realmente efficace deve dare priorità alla conservazione di tali configurazioni. A tal fine, l'SSIM opera una distinzione fondamentale tra due tipologie di alterazioni:

- **Distorsioni non strutturali**: variazioni legate alla luminanza, contrasto o shift spaziali. Tali modifiche non alterano la configurazione degli oggetti nell'immagine e l'occhio umano è in grado di compensarle senza percepire una perdita di qualità significativa.
- **Distorsioni strutturali**: rumore Gaussiano additivo, sfocatura(blur) o compressione con perdita (lossy). Questi fenomeni degradano significativamente l'informazione strutturale e vengono percepiti come un calo drastico della qualità.

Formalmente, siano $x = \{x_i \mid i = 1, \ldots, N\}$ e $y = \{y_i \mid i = 1, \ldots, N\}$ rispettivamente il segnale originale e quello ricostruito. L'indice SSIM viene calcolato combinando tre fattori indipendenti:

$$\mathrm{SSIM}(x, y) = \frac{(2\bar{x}\bar{y} + c_1)(2\sigma_{xy} + c_2)}{(\bar{x}^2 + \bar{y}^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{4.4}$$

Questa equazione può essere scomposta nel prodotto di tre componenti indipendenti:

$$\mathrm{SSIM}(x, y) = \frac{\sigma_{xy}}{\sigma_x \sigma_y} \cdot \frac{2\bar{x}\bar{y}}{\bar{x}^2 + \bar{y}^2} \cdot \frac{2\sigma_x \sigma_y}{\sigma_x^2 + \sigma_y^2} \tag{4.5}$$

Le tre componenti misurano rispettivamente:

1. Correlazione: rappresenta il coefficiente di correlazione tra x e y. Misura il grado di somiglianza strutturale e ha un range dinamico di $[-1, 1]$. Il valore massimo 1 si ottiene quando y è legato a x da una relazione lineare positiva.
2. Luminanza: confronta la media della luminanza tra le due immagini. Assume valore 1 se e solo se $\bar{x} = \bar{y}$.
3. Contrasto: misura quanto le varianze delle due immagini siano simili. Assume valore 1 se e solo se $\sigma_x = \sigma_y$.

L'SSIM assume valori compresi nell'intervallo $[0, 1]$ per segnali non negativi. Un valore pari a 1 indica che le due immagini sono identiche. Rispetto alle metriche basate sull'MSE, l'SSIM fornisce una valutazione molto più coerente con il giudizio soggettivo umano, rendendola uno strumento indispensabile per validare algoritmi

di ricostruzione in ambito medicale, dove la conservazione dei bordi e delle texture è fondamentale per la diagnosi.

Figura 4.3: Confronto di diverse distorsioni strutturali applicate alla medesima immagine: si noti come distorsioni visivamente molto diverse possano produrre lo stesso valore di MSE, evidenziando la necessità di metriche strutturali come l'SSIM.

4.2 Sparse-view CT: Analisi della riduzione delle proiezioni

La prima fase di valutazione sperimentale si focalizza sullo scenario della **Sparse-view CT**, spesso identificata nel contesto clinico di **Low-Dose CT**. L'obiettivo primario di questo approccio è la riduzione della dose di radiazioni somministrate al paziente, necessario per minimizzare i rischi radiologici pur mantenendo un contenuto informativo sufficiente per una diagnosi accurata.

Dal punto di vista computazionale, la riduzione della dose si traduce in una diminuzione del numero di proiezioni angolari acquisite. Questo trasforma il problema

inverso in un sistema fortemente sottodeterminato, dove l'operatore fisico A perde ulteriori informazioni, rendendo la ricostruzione estremamente suscettibile alla generazione di artefatti (tipicamente striature note come *streaking artifacts*) [2;3]. Di conseguenza, per testare la robustezza e la capacità di generalizzazione del modello proposto, l'analisi è stata condotta partendo da una configurazione di riferimento stabile con **180 proiezioni** (considerata *full-view* per questo dataset), riducendo poi progressivamente il campionamento a **90** e a **60 proiezioni**.

L'incremento dell'indeterminatezza del problema richiede un adeguamento dei parametri del paradigma Plug-and-Play in quanto, al diminuire delle informazioni catturate, il contributo del prior fornito dai denoiser diventa più cruciale. Diventa, quindi, necessario valutare una rimodulazione del numero di iterazioni dell'algoritmo e una ricalibrazione del livello di rumore σ nel continuation schedule, al fine di bilanciare correttamente la fedeltà ai dati acquisiti con la regolarizzazione strutturale imposta dalle reti neurali.

4.2.1 Baseline sperimentale: 180 proiezioni

Questa prima analisi si riferisce alla configurazione standard descritta nel Capitolo 2, in cui il sistema di acquisizione opera con 180 proiezioni angolari. In tale scenario, il problema inverso è ben condizionato e il campionamento è sufficiente per permettere ai metodi classici di produrre ricostruzioni di buona qualità. I risultati ottenuti in questa fase fungono da **riferimento (baseline)** per valutare il degrado prestazionale nei casi di sottocampionamento successivi. Di seguito vengono riportati i risultati medi ottenuti sul test set e i relativi confronti visivi:

Tabella 4.1: Metriche medie ottenute con 180 proiezioni angolari.

Metrica	**Linear (FBP)**	**TV**	**PnP-DRUNet**	**PnP-DnCNN**
PSNR (dB) ↑	16.63	21.71	**30.56**	28.89
SSIM ↑	0.098	0.429	**0.935**	0.906
MSE ↓	$2.18 \cdot 10^{-2}$	$6.90 \cdot 10^{-3}$	$9.00 \cdot 10^{-4}$	$1.32 \cdot 10^{-3}$

Una prima osservazione che si può dedurre dalla tabella è il netto distacco prestazionale tra i metodi basati su Deep Learning e gli approcci tradizionali. Tra i due metodi deep, la DRUNet si distingue maggiormente, beneficiando della sua natura *non-blind* che permette una regolarizzazione più precisa. Inoltre, l'analisi del boxplot rafforza queste osservazioni, fornendo indicazioni cruciali sulla robustezza dei modelli:

- *Affidabilità delle prestazioni*: si nota come la discrepanza tra i valori massimi e minimi ottenuti per **PnP-DRUNet** e **PnP-DnCNN** sia estremamente ridotta.
- *Superiorità strutturale*: il valore di SSIM prossimo a 0.93 per DRUNet conferma che la rete è in grado di recuperare quasi perfettamente le strutture geometriche presenti nelle immagini del test set, laddove la semplice FBP fallisce drasticamente (SSIM ≈ 0.10) a causa del rumore residuo che maschera i dettagli.

Come evidenziato precedentemente, le metriche quantitative non sempre riflettono fedelmente la qualità visiva di una ricostruzione. Per questo motivo vengono esaminati cinque campioni del test set che rappresentano diverse sfide geometriche, permettendo di analizzare nel dettaglio il comportamento dei modelli.

Dall'osservazione delle ricostruzioni (Figure 4.6 a ed e) si evince che i singoli impulsi luminosi ad alto contrasto, considerati una delle sfide critiche della ricostruzione tomografica, vengono preservati correttamente. A differenza di quando potrebbe accadere con filtri di denoising standard, tali elementi non vengono erroneamente interpretati come rumore e rimossi, confermando l'accuratezza del prior appreso dalle reti. Ulteriori osservazioni includono:

- **Gestione delle sovrapposizioni**: nei casi b) e c), la sovrapposizione di elissi con intensità luminosa simile viene gestita con precisione. Le forme risultano corrette e prive di sbavature lungo i bordi, indicando un'ottima capacità di segmentazione strutturale.
- **Recupero di geometrie sottili**: le ellissi lunghe e strette, che nella ricostruzione FBP tendono a confondersi con il rumore di fondo, vengono ricostruite nitidamente dai modelli PnP.
- *Confronto tra modelli*: mentre la ricostruzione TV tende a produrre immagini con regioni uniformi ma bordi talvolta eccessivamente semplificati, le soluzioni PnP-DRUNet e PnP-DnCNN rimangono più fedeli alla ground truth, eliminando quasi totalmente gli artefatti granulari introdotti dalla FBP senza sacrificare la risoluzione spaziale.

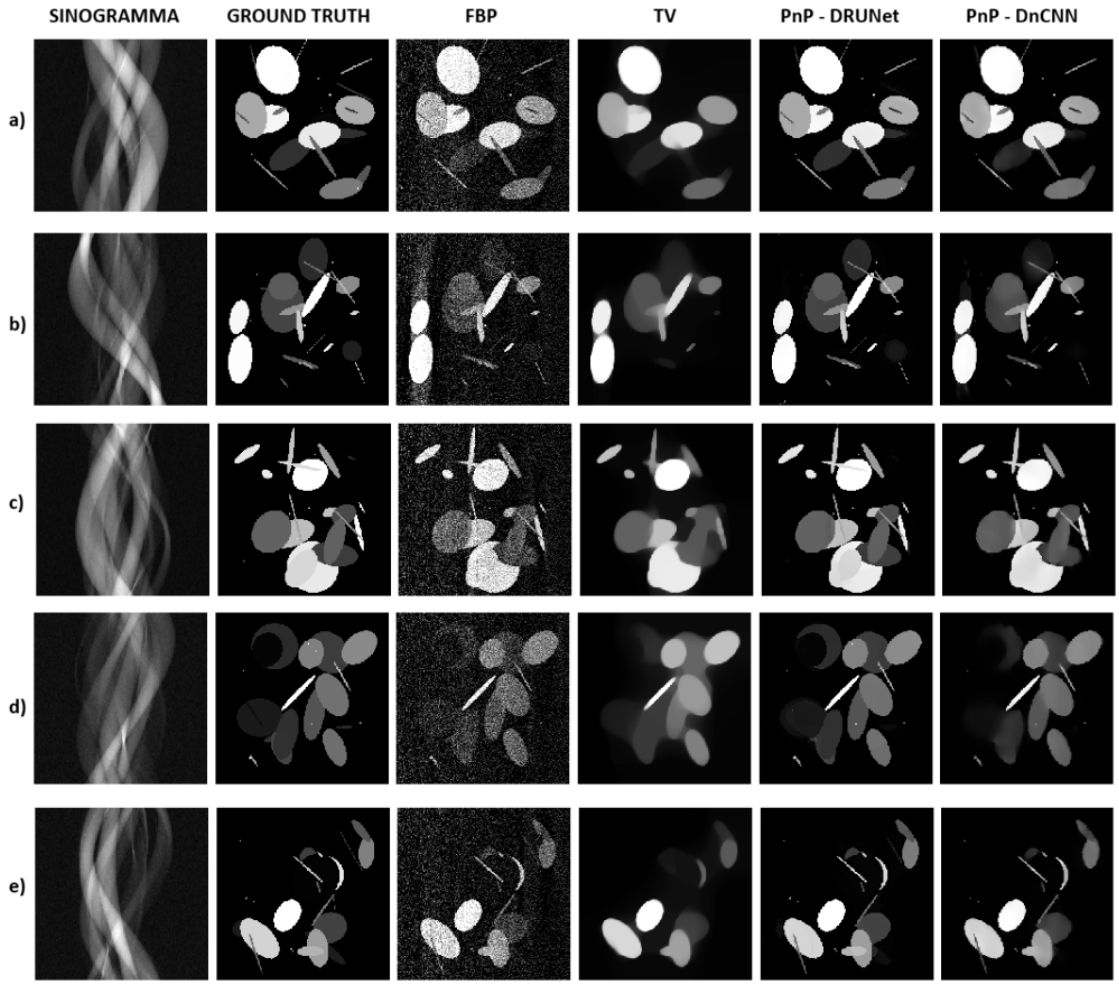

Figura 4.4: Confronto visivo delle ricostruzioni di campioni del test set con 180 angoli: Sinogramma, Ground Truth, GBP, TV, PnP-DRUNet, PnP-DnCNN

Si introduce un ulteriore osservazione relativa alla robustezza dei modelli attraverso l'analisi dei **grafici boxplot**. Questi grafici permettono di valutare la distribuzione delle metriche sull'intero test set, evidenziando la dispersione dei risultati rispetto ai valori medi.

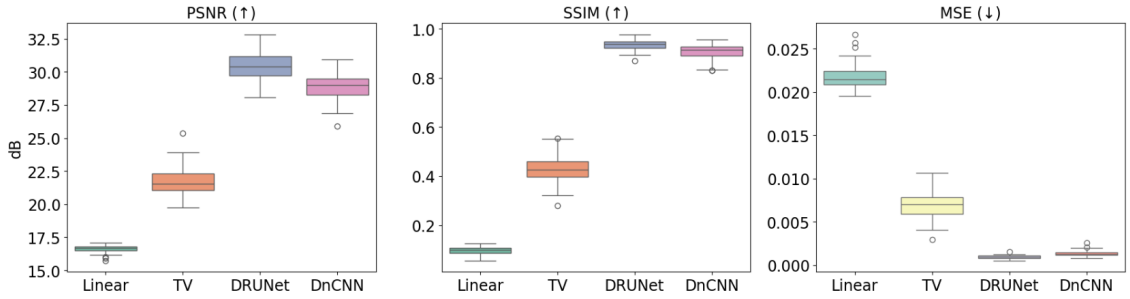

Figura 4.5: Distribuzione statistica delle metriche PSNR, SSIM e MSE per i diversi modelli di ricostruzione.

L'analisi dei boxplot (Figura 4.7) conferma la superiorità e la stabilità dei metodi Plug-and-Play. Infatti, minore è l'altezza della "box", che rappresenta l'intervallo interquartile, minore risulta essere la variabilità delle prestazioni tra le diverse immagini del dataset. Si noti come i modelli PnP-DRUNet e PnP-DnCNN presentino box estremamente contenute, specialmente nella metrica SSIM, indicando che la qualità della ricostruzione è costante e non dipende dalla specifica geometria delle ellissi presenti nel campione. La presenza di un numero molto ridotto di punti isolati (*outliers*) al di fuori del boxplot per i modelli PnP sottolinea la capacità di quest'ultimi di generalizzare correttametne anche su casi più complessi, garantendo una robustezza operativa verificata su tutto il test set.

In conclusione questi risultati validano l'affidabilità dell'approccio proposto: il modello non solo raggiunge picchi di qualità elevata, ma lo fa in modo predicibile e stabile su tutta la casistica analizzata.

4.2.2 Riduzione della dose a 90 proiezioni

In questa fase della sperimentazione si procede a dimezzare il numero di proiezioni acquisite, riducendo gli angoli a 90. Tale scenario rappresenta un livello di difficoltà intermedio, in cui l'incompletezza dei dati inizia a produrre artefatti visibili nelle ricostruzioni lineari. Per individuare la configurazione ottimale dei parametri dell'algoritmo Plug-and-Play, è stata condotta una ricerca empirica monitorando l'andamento delle metriche al variare del numero di iterazioni. L'analisi di questi trial ha evidenziato che la combinazione dei parametri utilizzata nella configurazione *baseline* (180 proiezioni) garantisce prestazioni ancora eccellenti, in quanto tentativi di ricalibrazione del continuation schedule o del numero di iterazioni non hanno prodotto miglioramenti significativi. Di conseguenza, si è scelto di mantenere invariati i parametri per garantire la coerenza nel confronto tra i diversi scenari di campionamento.

Seguendo la metodologia introdotta nella sezione precedente, l'analisi delle prestazioni verrà articolata attraverso la presentazione delle metriche medie, il confronto visivo delle ricostruzioni e lo studio della robustezza mediante i grafici boxplot.

Tabella 4.2: Metriche medie ottenute con 90 proiezioni angolari.

Metrica	**Linear (FBP)**	**TV**	**PnP-DRUNet**	**PnP-DnCNN**
PSNR (dB) ↑	16.34	21.67	**29.96**	28.60
SSIM ↑	0.095	0.426	**0.927**	0.896
MSE ↓	$2.33 \cdot 10^{-2}$	$6.96 \cdot 10^{-3}$	$1.03 \cdot 10^{-3}$	$1.41 \cdot 10^{-3}$

Come previsto, la riduzione del numero di proiezioni acquisite comporta una minore densità di informazioni nel sinogramma, riflettendosi in un lieve decremento delle metriche medie rispetto allo scenario a 180 angoli. Tuttavia, lo scarto prestazionale risulta estremamente contenuto: il PnP-DRUNet subisce un calo di soli 0.33 dB relativamente alla metriche PSNR e mantiene un valore di SIMM superiore a 0.92. Questo comportamento è un chiaro indicatore della robustezza dei modelli deep: nonostante il problema inverso diventi più sottodeterminato, i prior appresi dalle reti neurali riescono a compensare efficacemente l'assenza di dati misurati, producendo risultati che restano qualitativamente eccellenti e decisamente superiori ai metodi classici.

Al fine di valutare l'impatto diretto della riduzione del numero di proiezioni, per il confronto visivo sono stati selezionati alcuni dei medesimi campioni analizzati nello scenario precedente. Come ipotizzato, la ricostruzione tramite FBP mostra un incremento del rumore di fondo e degli artefatti a strisce dovuto alla riduzione degli angoli di acquisizione. Nonostante ciò, i modelli Plug-and-Play mantengono una qualità visiva elevata, non mostrando un calo prestazionale evidente nelle strutture principali. Al contrario, i limiti dei metodi tradizionali emergono con maggiore chiarezza: la regolarizzazione TV, pur eliminando il rumore, produce immagini eccessivamente levigate che peccano di precisione e resa dei dettagli.

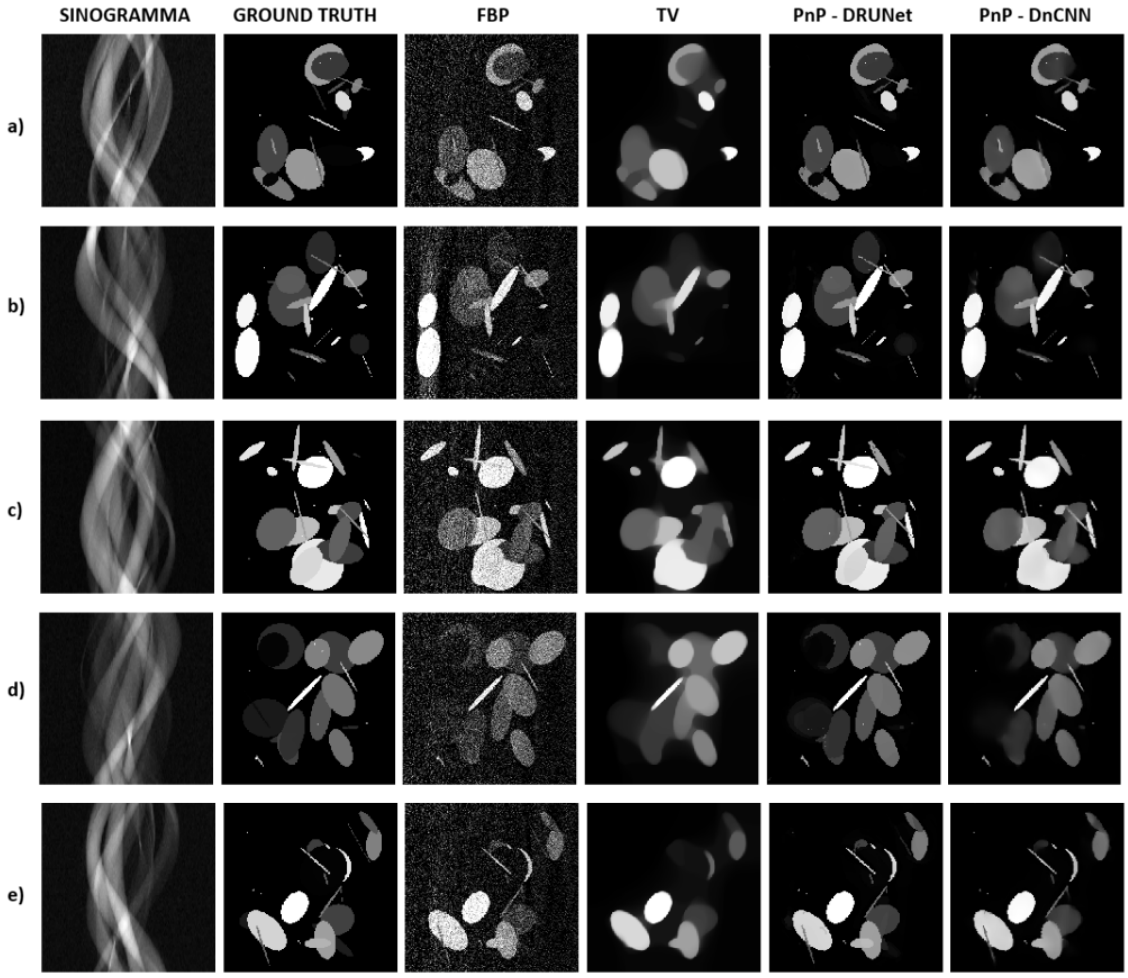

Figura 4.6: Confronto visivo delle ricostruzioni di campioni del test set con 90 angoli: Sinogramma, Ground Truth, GBP, TV, PnP-DRUNet, PnP-DnCNN

Sebbene i risultati complessivi risultino ottimi, un'analisi più minuziosa rivela i primi ostacoli introdotti dal sottocampionamento: nelle ricostruzioni relative alle immagini a) e c) si osserva che alcuni elementi di dimensioni ridotte non vengono perfettamente recuperati dai metodi PnP. Sebbene tale perdita risulti trascurabile per la visione d'insieme, essa rappresenta il limite intrinseco della riduzione delle viste. Infatti, con sole 90 proiezioni, il solo prior della rete, per quanto avanzato, non riesce a ricostruire tutte le frequenze spaziali con assoluta fedeltà, poiché parte dell'informazione originale non è più presente nel dato acquisito.

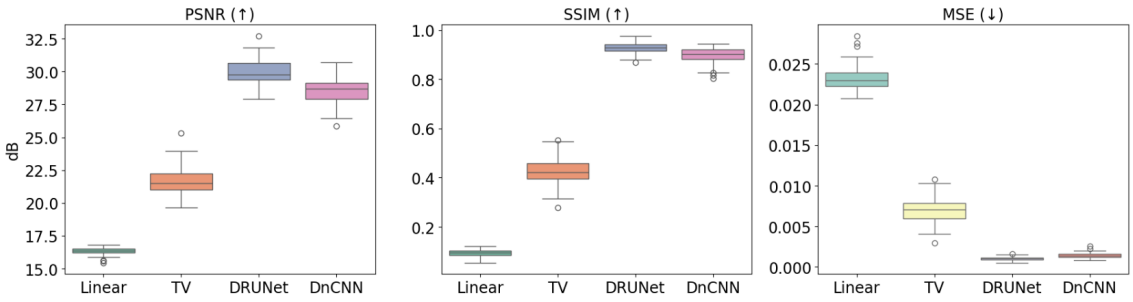

Figura 4.7: Distribuzione statistica delle metriche PSNR, SSIM e MSE per i diversi modelli di ricostruzione.

Poiché i boxplot riflettono la distribuzione delle metriche e i valori medi di questo scenario non si discostano significativamente dalla baseline sperimentale, i grafici risultano molto simili a quelli discussi precedentemente. Di conseguenza, le osservazioni relative alla robustezza del modello rimangono sostanzialmente invariate. L'unica differenza degna di nota è un lieve incremento del numero di punti isolati verso il basso nello scenario a 90 angoli. Questo fenomeno suggerisce che, per configurazioni geometriche particolarmente complesse, la carenza di dati angolari inizi a rappresentare una sfida critica anche per i prior profondi, sebbene l'effetto resti limitato a pochi casi specifici.

In sintesi, i modelli Plug-and-Play dimostrano una capacità di adattamento superiore, garantendo ricostruzioni di alta qualità anche in scenari di Low-Dose CT. La diminuzione degli angoli, infatti, non compromette la stabilità dei risultati, che si mantengono coerenti con quanto ottenuto nel caso a 180 proiezioni.

4.2.3 Riduzione della dose a 60 proiezioni

Questo test rappresenta la sfida computazionale più complessa introdotta finora: il numero di proiezioni acquisite viene ridotto a un terzo rispetto alla baseline, per un totale di soli 60 angoli. Tale configurazione comporta un forte sottodeterminamento del problema inverso con una conseguente e drastica perdita di informazioni all'interno del sinogramma. Questo rende la ricostruzione molto più complicata e si ha la necessità di riconfigurare i parametri dei modelli, al fine di massimizzare il recupero delle strutture geometriche e del dettaglio fine. Nello specifico:

- PnP-DnCNN: trattandosi di un denoiser di tipo blind, l'unica variabile di controllo esterna è il numero di iterazioni dell'algoritmo PnP. Sulla base dell'analisi dei grafici di convergenza, che mostrano una stabilizzazione delle metriche dopo il decimo passaggio, il numero di iterazioni è stato incrementato a 10.

- PnP-DRUNet: per il modello non-blind è stato invece ricalibrato il continuation schedule. Mentre i livelli di rumore (σ) sono stati mantenuti inalterati per preservare la risposta della rete, è stato aumentato il numero di iterazioni per ogni stadio della sequenza. La nuova configurazione adottata è $iteration\_schedule = [4, 10, 30]$.

Le modifiche apportate ai parametri hanno permesso di ottenere le metriche medie riportate nella seguente tabella:

Tabella 4.3: Metriche medie ottenute con 60 proiezioni angolari.

Metrica	**Linear (FBP)**	**TV**	**PnP-DRUNet**	**PnP-DnCNN**
PSNR (dB) ↑	15.96	21.60	**28.90**	28.12
SSIM ↑	0.092	0.423	0.923	**0.926**
MSE ↓	$2.54 \cdot 10^{-2}$	$7.08 \cdot 10^{-3}$	$1.33 \cdot 10^{-3}$	$1.58 \cdot 10^{-3}$

Come previsto, si osserva un calo generalizzato delle metriche dovuto alla forte riduzione del numero di proiezioni. Tuttavia, grazie alla riconfigurazione dei parametri, non si riscontrano degradazioni drastiche delle prestazioni. I risultati rimangono competitivi, confermando la capacità dei modelli di estrarre informazioni strutturali anche in condizioni di campionamento estremo. Un dato estremamente interessante riguarda il valore dell'SSIM medio per il modello PnP-DRUNet, che risulta essere superiore rispetto al caso di test precedente. Questo fenomeno, apparentemente controintuitivo, può essere ricondotto al contributo del denoiser che, lavorando più a lungo, tende a produrre superfici molto uniformi e bordi netti che la metrica in questione premia fortemente.

Dalle ricostruzioni (Figura 4.8) si possono osservare alcuni piccoli artefatti morfologici dovuti alla scarsità dei dati:

- **Incompletezza strutturale**: nel campione a), alcune ellissi non appaiono completamente "piene" o presentano un'intensità non uniforme al loro interno.
- **Distorsioni geometriche**: nel campione d), si notano strutture che, pur mantenendo bordi definiti, presentano forme irregolari che si discostano dal profilo ellittico standard.

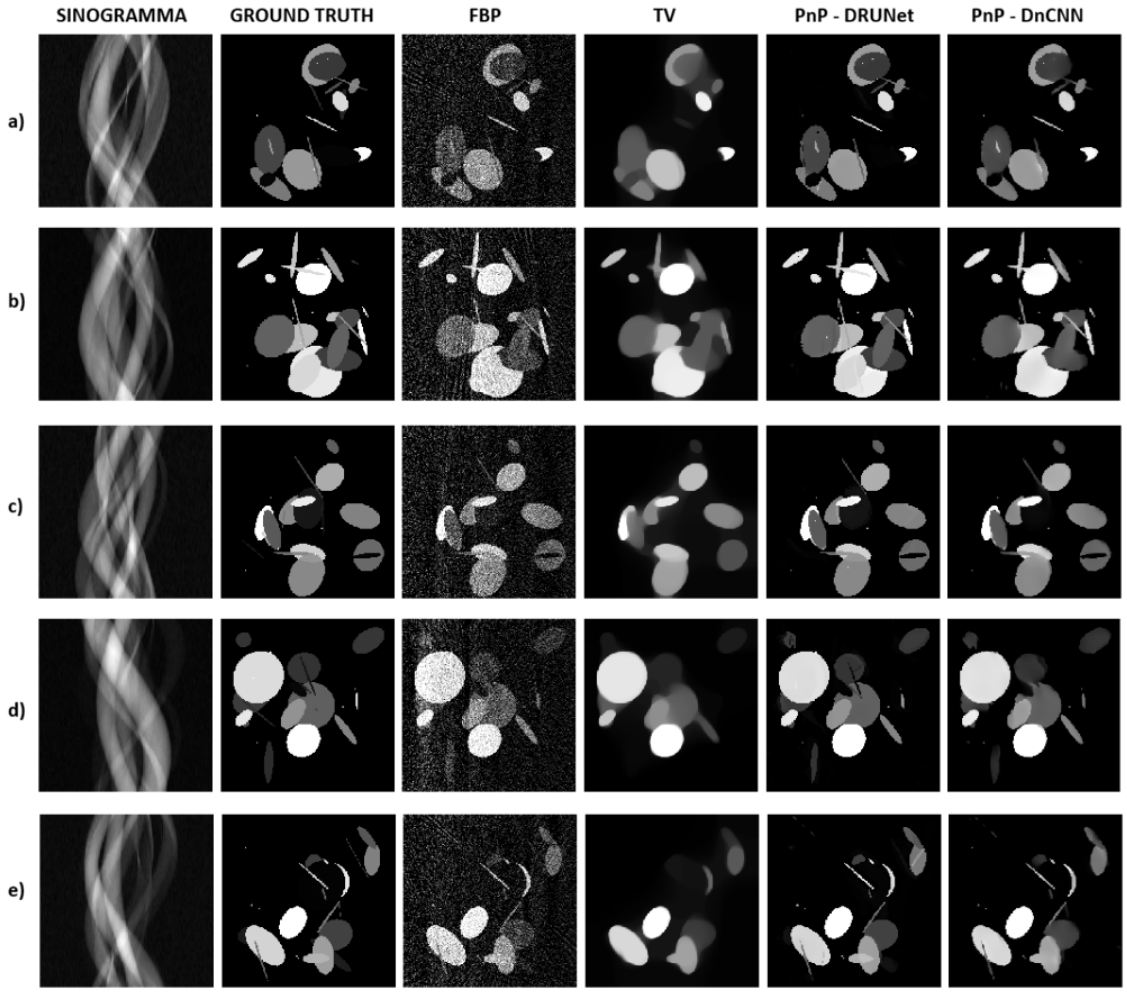

Figura 4.8: Confronto visivo delle ricostruzioni di campioni del test set con 60 angoli: Sinogramma, Ground Truth, GBP, TV, PnP-DRUNet, PnP-DnCNN

Queste imprecisioni sono la conseguenza diretta di un numero di informazioni estremamente contenuto nel signogramma; i modelli operano al massimo delle loro capacità di inferenza, ma non possono garantire una ricostruzione perfetta laddove la base di partenza è fisicamente incompleta. Tuttavia i risultati rimangono visivamente soddisfacenti: nei campione c) ed e), i modelli PnP riescono a recuperare correttamente persino i singoli impulsi luminosi ad alto contrasto, dimostrando una capacità di generalizzazione che i metodi tradizionali non riescono a raggiungere.

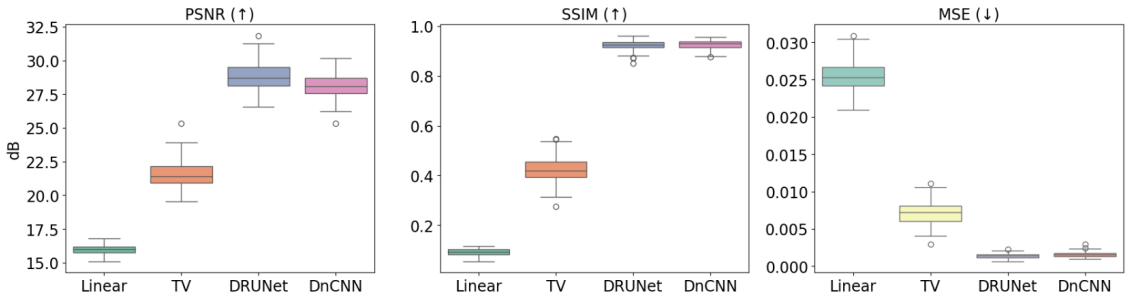

Figura 4.9: Distribuzione statistica delle metriche PSNR, SSIM e MSE per i diversi modelli di ricostruzione.

Attraverso l'analisi dei boxplot si nota un allungamento verticale delle box per i modelli PnP-DRUNet e PnP-DnCNN, in particolare per quanto concerne la metrica PSNR. Tale fenomeno è sintomo del fatto che, in regime di forte sottocampionamento, le prestazioni iniziano a dipendere maggiormente dalla complessità geometrica che caratterizza il campione specifico. Si osserva, inoltre, una presenza più marcata di punti isolati verso il basso. Nonostante l'aumento della varianza del PSNR, le box dell'SSIM rimangono compatte e posizionate su valori molto alti. Come discusso in precedenza, questo conferma che, sebbene l'accuratezza puntuale (PSNR) oscilli di più a causa della carenza di dati, il denoiser riesce a mantenere una coerenza strutturale costante, "forzando" la ricostruzione verso forme pulite e definite.

In conclusione, il confronto tra i tre stadi di testing appena esposti dimostra che i modelli Plug-and-Play non solo offrono prestazioni superiori, ma garantiscono una robustezza operativa che degrada in modo "dolce" (*graceful degradation*). Infatti, anche nello scenario più critico a 60 angoli, le prestazioni dei modelli PnP rimangono complessivamente superiori a quelle ottenute dal metodo TV a 180 angoli. Per visualizzare chiaramente questo comportamento, viene riportato di seguito l'andamento del valore medio della metrica PSNR in funzione del numero di proiezioni acquisite.

Tale risultato valida l'efficacia dell'approccio proposto per applicazione di Low-Dose CT, in quanto la riduzione della dose di radiazioni non pregiudica la qualità strutturale e la nitidezza dei borsi, elementi indispensabili per garantire l'accuratezza della diagnosi.

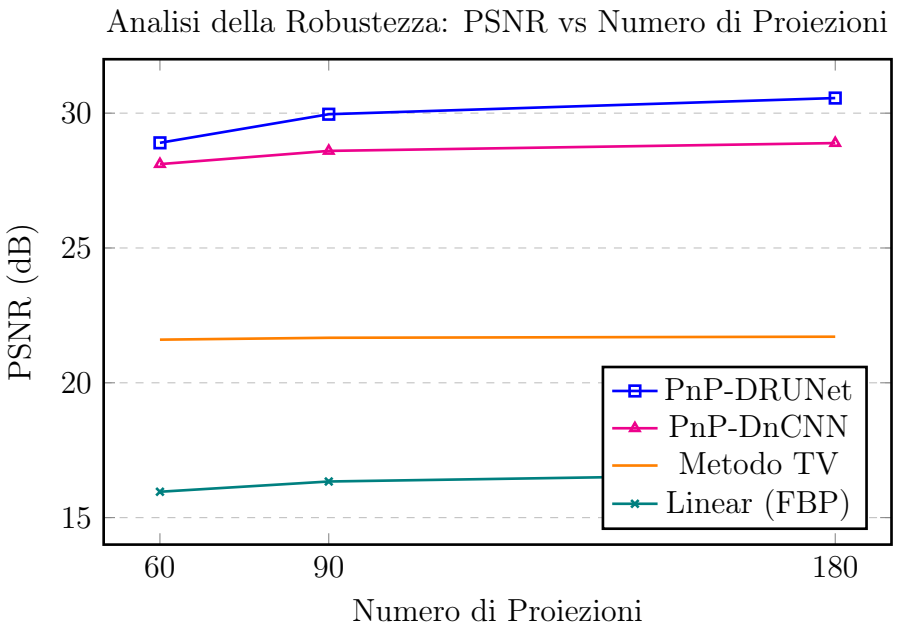

Figura 4.10: Confronto prestazionale tra metodi Deep e tradizionali al diminuire del numero di viste. Si noti come i modelli PnP mantengano un distacco netto (circa 7-8 dB) rispetto alla TV anche nello scenario di sottocampionamento estremo.

4.3 Analisi comparativa tra i metodi di ricostruzione

In questa sezione vengono esaminate nel dettaglio le discrepanze prestazionali tra i metodi di ricostruzione utilizzati, prendendo come riferimento la baseline sperimentale a 180 proiezioni. L'obiettivo è isolare fattori metodologici che determinano la superiorità di un modello rispetto all'altro.

4.3.1 Confronto tra modellazione analitica e approcci Data-driven

Il marcato distacco prestazionale che si è potuto osservare nella sezione precedente è riconducibile alla natura intrinseca dei due paradigmi. I metodi tradizionali si fondano su un approccio puramente analitico che modella la fisica del problema attraverso vincoli matematici, mentre i modelli Plug-and-Play sfruttano la potenza dei dati. Se da un lato l'approccio classico mantiene una validità universale indipendentemente dal target, dall'altro i modelli Deep Learning, seppur addestrati specificamente su un dominio di dati, dimostrano una capacità superiore di estrazio-

ne e ricostruzione dell'informazione latente. Nello specifico, la differenza principale emerge nell'operatore di regolarizzazione:

- **Total Variation (TV)**: si basa su un'assunzione a priori deterministica che interpreta l'immagine come un insieme di regioni costanti a tratti (*piecewise constant*)[8]. Sebbene questa proprietà sia utile per sopprimere il rumore, essa impone un vincolo geometrico semplificativo che tende a livellare le strutture più fini e a smussare i bordi a basso contrasto. Questo fenomeno può essere osservato nel dettaglio nella Figura 4.11, che riporta quattro campioni di ricostruzione tramite TV in cui emergono chiaramente le criticità legate all'eccessiva semplificazione dei dettagli.

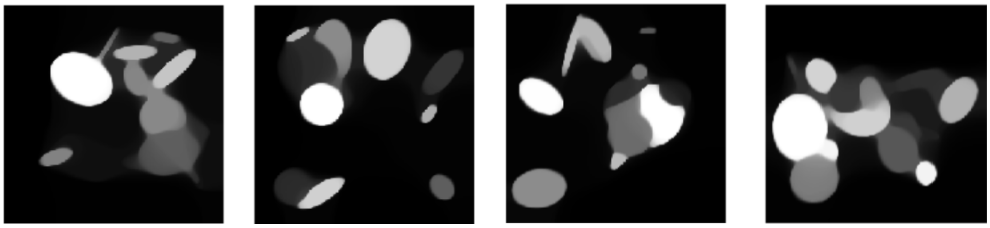

Figura 4.11: Campioni di ricostruzione tramite Total Variation: si noti l'effetto di livellamento sulle strutture minori.

- **Paradigma PnP**: in opposizione al precedente, i denoiser profondi non impongono una forma geometrica fissa. Essi sostituiscono la regolarizzazione analitica con un *prior* appreso, capace di riconoscere e preservare la complessità strutturale delle ellissi senza sacrificare la nitidezza dei contorni o la fedeltà delle texture. Questo vantaggio qualitativo può essere apprezzato nella Figura 4.12, che propone un confronto diretto tra le metodologie (per facilitare l'analisi visiva e concentrare l'attenzione sulla resa delle strutture sono stati omessi il sinogramma e la ricostruzione FBP).

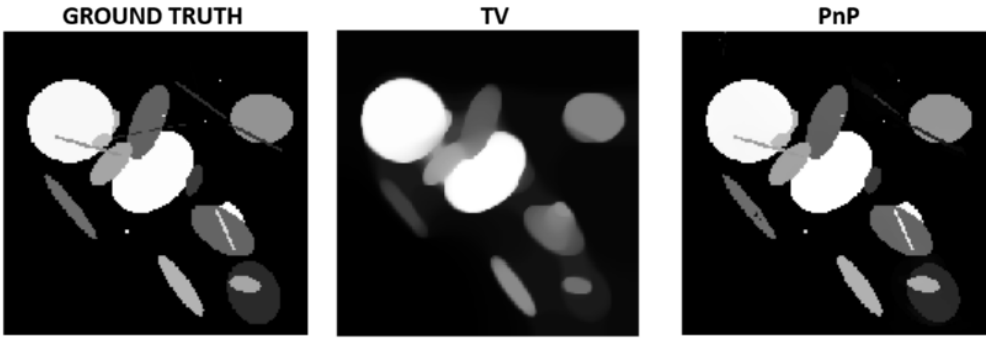

Figura 4.12: Confronto diretto tra la regolarizzazione TV e l'approccio Plug-and-Play.

4.3.2 Confronto tra PnP-DRUNet e PnP-DnCNN

Si procede ad analizzare il confronto diretto tra i due algoritmi Plug-and-Play, i quali si differenziano sostanzialmente per la tipologia di denoiser integrato come *prior*. Di conseguenza, la ragione delle differenti prestazioni risiede nelle caratteristiche intrinseche dei modelli di denoising piuttosto che nella struttura dell'algoritmo PnP stesso. Infatti, sebbene i parametri operativi, quali il numero di iterazioni e la gestione del livello di rumore, differiscano leggermente, essi sono stati ottimizzati seguendo la medesima procedura empirica e non costituiscono, pertanto, la causa primaria del divario numerico osservato precedentemente. In generale, i risultati dei test mostrano che il modello PnP-DRUNet ottiene prestazioni sistematicamente superiori. Le ragioni di tale discrepanza possono essere ricondotte a due fattori principali:

- **Architettura della rete**: a differenza della DnCNN, che mantiene una risoluzione costante lungo tutta la profondità della rete, la DRUNet sfrutta una struttura a "U" (*U-Net*) con connessioni residue. Questo design permette di estrarre feature a diverse scale spaziali, riuscendo a catturare efficacemente sia il contesto globale dell'immagine sia i dettagli più fini dei bordi e delle texture.
- **Flessibilità del livello di rumore**: si ricorda, inoltre, che, mentre la DnCNN opera in modalità blind, la DRUNet riceve in input una mappa di rumore σ. All'interno del paradigma PnP, questa caratteristica è fondamentale: permette all'algoritmo di modulare dinamicamente l'aggressività della regolarizzazione nei confronti dell'immagine degradata, riducendo progressivamente la forza del denoiser all'aumentare delle iterazioni secondo un *continuation schedule* prestabilito.

In definitiva, l'analisi comparativa evidenzia come la superiorità dei modelli Plug-and-Play non risieda semplicemente in una maggiore potenza computazionale, ma nella capacità di integrare una conoscenza strutturale complessa (*learned prior*) all'interno del processo iterativo. Laddove la modellazione analitica della Total Variation fallisce nel distinguere tra rumore e dettagli fini a causa della sua natura rigida e deterministica, i denoiser profondi dimostrano una flessibilità adattiva superiore. In particolare, l'architettura multiscala e la natura *non-blind* della DRUNet si confermano le caratteristiche chiave per massimizzare la fedeltà della ricostruzione, permettendo una gestione dinamica della regolarizzazione che risulta preclusa sia ai metodi classici che ai modelli *blind* come la DnCNN.

4.4 Analisi comparativa tra approcci Plug-and-Play e metodi Unfolded

In questa sezione viene affrontato il confronto diretto tra le due metodologie ibride analizzate in questo lavoro di tesi: il paradigma **Plug-and-Play** e l'approccio **Unfolded**. Il confronto è effettuato sulla casistica *baseline* a 180 proiezioni per valutare il potenziale massimo di entrambi i metodi. Per quanto riguarda il paradigma PnP, viene preso in cosiderazione esclusivamente il modello che utilizza il denoiser DRUNet come prior, in quanto ha dimostrato nelle sezioni precedenti prestazioni superiori.

Tabella 4.4: Confronto delle metriche medie ottenute attraverso PnP-DRUNet e Unfolded Primal-Dual (180 proiezioni).

Metrica	**Linear (FBP)**	**PnP-DRUNet**	**Unfolded**
PSNR (dB) ↑	16.63	30.56	**30.73**
SSIM ↑	0.098	**0.935**	0.870
MSE ↓	$2.20 \cdot 10^{-2}$	$9.00 \cdot 10^{-4}$	$8.00 \cdot 10^{-4}$

Dall'analisi dei dati in Tabella 4.4, si osserva che la ricostruzione tramite Unfolded Primal-Dual ottiene risultati lievemente migliori nelle metriche PSNR e MSE rispetto a PnP. Tuttavia, il PnP-DRUNet mantiene un vantaggio significativo nella metrica SSIM, con un distacco di circa 0.06. Questi risultati sperimentali possono esssere spiegati dalla differente natura dei due approcci: mentre gli unfolded ottimizzano l'intera architettura per minimizzare l'errore quadratico (MSE) portando così a valori di PSNR superiori, i metodi PnP beneficiano di un denoiser pre-addestrato estremamente efficace nel preservare la coerenza strutturale e la nitidezza dei bordi, aspetti che l'indice SSIM premia maggiormente. Queste differenze possono essere osservate visivamente nelle ricostruzioni eseguite.

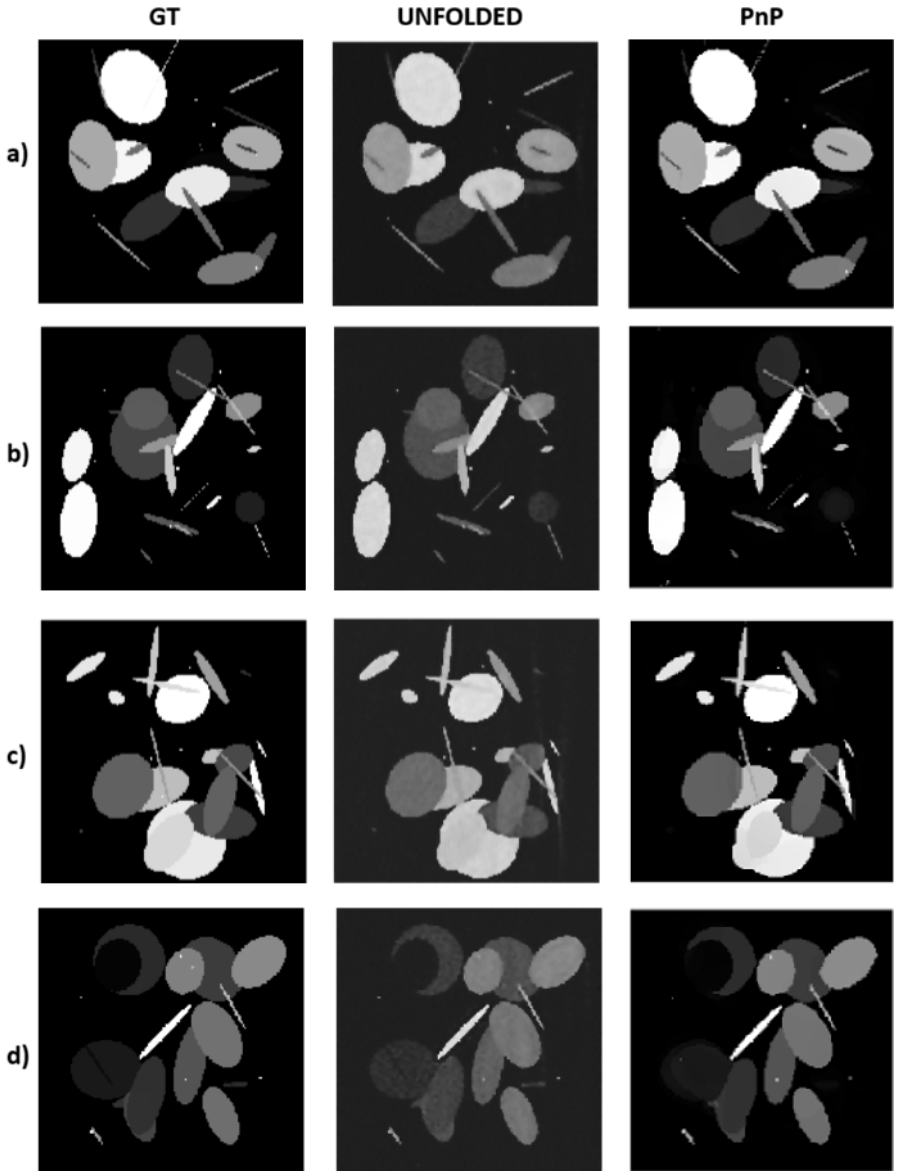

Figura 4.13: Confronto visivo delle ricostruzione eseguite attraverso approccio Unfolded e PnP

Sebbene l'approccio Unfolded presenti valori di PSNR e MSE mediamente superiori, le ricostruzioni mostrano un rumore residuo di fondo più evidente e una minore definizione dei bordi rispetto al paradigma PnP-DRUNet. Osservando i campioni c) e d) nella Figura 4.13, si può notare come il modello PnP riesca a separare meglio le ellissi sovrapposte, mantenendo un contrasto netto. D'altra parte, il modello Unfolded, pur recuperando correttemente la geometria generale delle immagini, presenta superfici meno uniformi. Ciò è probabilmente causato dal fatto che l'addestramento

della rete tende a privilegiare la fedeltà statistica ai valori dei pixel (minimizzazione MSE) piuttosto che la pulizia formale del prior, caratteristica che invece emerge nei metodi PnP.

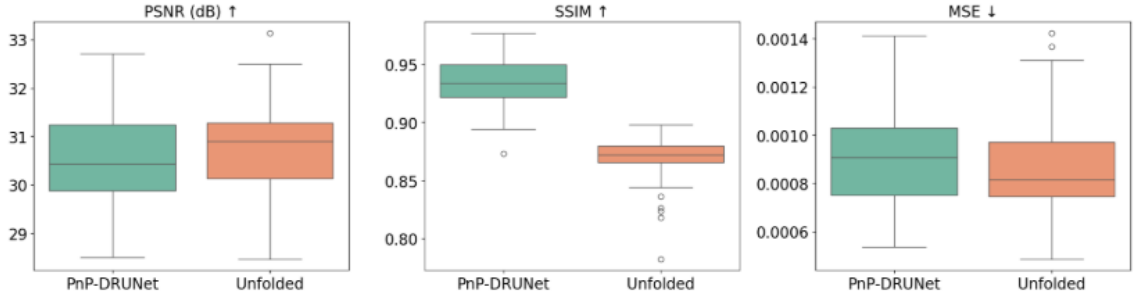

Figura 4.14: Distribuzione statistica delle metriche PSNR, SSIM e MSE per il confronto tra gli approcci PnP e Unfolded.

Attraverso l'analisi dei boxplot (Figura 4.14), che permettono un confronto diretto tra le distribuzioni delle metriche sul test set, è possibile trarre conclusioni rilevanti sulla stabilità dei due approcci. Nello specifico si può notare che le "box" relative all'approccio Unfolded risultano essere lievemente più compatte e posizionate su valori mediani migliori rispetto al PnP. Questo suggerisce una precisione numerica superiore nella maggior parte dei casi. Tuttavia, l'Unfolded presenta una significativa presenza di outliers varso il basso nel SSIM e verso l'alto nel MSE. Questo fenomeno indica che, sebbene il modello sia generalmente molto efficace, tende a produrre errori più marcati su campioni caratterizzati da geometrie particolarmente complesse o atipiche.

In sintesi, mentre l'Unfolded eccelle nell'approssimazione puntale, l'approccio PnP si dimostra più affidabile nel preservare la coerenza globale dell'immagine, non presentando cadute prestazionali drastiche. Di conseguenza, il paradigma Plug-and-Play si conferma la scelta preferibile per applicazioni in cui la nitidezza dei contorni, la separazione delle strutture e la stabilità visiva sono requisiti fondamentali.

Conclusioni

Il presente lavoro di tesi ha indagato l'efficacia del paradigma **Plug-and-Play** nella ricostruzione di immagini per la tomografia computerizzata (CT). L'obiettivo principale risiedeva nel superamento dei limiti intrinsechi dei metodi di ricostruzione tradizionali, integrando il rigore fisico del modello di acquisizione con la potenza espressiva dei prior appresi tramite modelli di Deep Learning. Tali limiti sono stati messi alla prova in scenari di *Low-Dose CT* per analizzare il comportamento del modello al diminuire delle proiezioni angolari e, di conseguenza, della densità delle informazioni presenti all'interno del sinogramma di partenza.

La sperimentazione, condotta a partire da una baseline a 180 proiezioni e ridotta progressivamente a 90 e 60 viste, ha permesso di simulare contesti di *Low-Dose CT*, dove la riduzione della dose di radiazioni somministrata al paziente è di primario interesse clinico. L'analisi ha evidenziato che, mentre la regolarizzazione classica *Total Variation* (TV) mostra evidenti limiti nel preservare i dettagli fini a causa del suo comportamento piecewise constant, i modelli PnP garantiscono una ricostruzione fedele dei bordi e delle geometrie complesse. In particolare, il modello PnP-DRUNet si è distinto come il più performante, raggiungendo valori di SSIM superiori a 0.92 anche nello scenario di sottocampionamento più estremo, confermando l'efficacia dell'architettura multiscala e della gestione dinamica del rumore (continuation schedule).

Un contributo significativo di questa tesi è stato il confronto diretto tra il paradigma PnP e gli approcci Unfolded (nello specifico, il modello Learned Primal-Dual). L'analisi comparativa ha rivelato un importante trade-off metodologico: i metodi Unfolded hanno dimostrato una precisione numerica superiore nelle metriche puntuali (PSNR e MSE), mentre il paradigma Plug-and-Play, pur ottenendo valori di PSNR leggermente inferiori, ha mostrato una superiorità netta nella conservazione della struttura globale e nella pulizia visiva (SSIM elevato), grazie all'utilizzo di denoiser specializzati capaci di agire come regolarizzatori più robusti e meno soggetti a rumore residuo di fondo.

Un aspetto fondamentale emerso dallo studio è la robustezza statistica del modello PnP. L'analisi tramite boxplot ha evidenziato come la distribuzione delle metriche rimanga compatta, con una presenza estremamente ridotta di *outliers* rispetto all'approccio Unfolded. Questo dimostra una capacità di generalizzazione essenziale per le applicazioni pratiche, provando che le prestazioni non dipendono esclusivamente dalla specifica configurazione strutturale delle immagini degradate.

Nonostante i risultati promettenti, il lavoro ha fatto emergere alcune sfide, legate principalmente alla necessità di una calibrazione accurata dei parametri iterativi e al carico computazionale richiesto dalle architetture profonde durante le fasi di ottimizzazione. Tuttavia, il vantaggio principale del paradigma Plug-and-Play risiede nella sua natura ibrida, che coniuga i benefici degli approcci iterativi con quelli dei modelli profondi: una volta eseguito il training dei denoiser, la loro conoscenza del dataset può essere sfruttata direttamente, permettendo di contenere il numero di iterazioni finali. Inoltre, lo schema dell'algoritmo PnP garantisce una trasparenza superiore rispetto alle reti end-to-end, che operano spesso come *black-box* di difficile interpretazione.

In conclusione, i risultati ottenuti pongono solide basi per l'adozione di algorimti di ricostruzione basati su regolarizzazione profonda nei sistemi CT di nuova generazione, delineando un percorso verso uno screening medico più sicuro, efficiente e caratterizzato da immagini ad alta qualità.

Le prospettive future prevedono l'estensione del paradigma a dataset clinici reali e lo studio di funzioni di perdita ibride, al fine di affinare ulteriormente la rimozione degli artefatti in contesti anatomici complessi.

Riferimenti bibliografici

[1] J. Radon, *On the Determination of Functions from Their Integral Values Along Certain Manifolds* (translated from German), Berichte über die Verhandlungen der Königlich Sächsischen Gesellschaft der Wissenschaften zu Leipzig, vol. 69, pp. 262–277, 1917.

[2] A. C. Kak and M. Slaney, *Principles of Computerized Tomographic Imaging*, SIAM: Society for Industrial and Applied Mathematics, 2001.

[3] T. M. Buzug, *Computed Tomography: From Photon Statistics to Modern Cone-Beam CT*, Springer Science & Business Media, 2008.

[4] J. Tachella, S. Chun, and D. Dong, *DeepInverse: A Pytorch Library for Solving Imaging Inverse Problems with Deep Learning*, arXiv preprint arXiv:2307.05740, 2023.

[5] J. Adler and O. Öktem, *Learned Primal-Dual Reconstruction*, IEEE Transactions on Medical Imaging, vol. 37, no. 6, pp. 1322–1332, 2018.

[6] K. Zhang, Y. Li, W. Zuo, E. Schuster, and L. Zhang, *Plug-and-Play Image Restoration with Deep Denoiser Prior*, IEEE Transactions on Pattern Analysis and Machine Intelligence, 2021.

[7] Z. 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, vol. 13, no. 4, pp. 600–612, 2004.

[8] A. Chambolle and T. Pock, *An Introduction to Continuous Optimization for Imaging*, Acta Numerica, vol. 25, pp. 161–319, 2016.

Ringraziamenti

Innanzitutto, desidero ringraziare profondamente la professoressa Piccolomini: mi ha seguita con pazienza ed è sempre stata disponibile al confronto, aiutandomi in ogni momento di necessità. Non avrei potuto fare scelta migliore.

Ringrazio la mia famiglia che, nonostante io possa essere una persona difficile e riservata tra le mura di casa, mi è sempre stata accanto, supportandomi emotivamente ed economicamente. Il traguardo che raggiungo oggi è anche merito vostro, dei vostri sacrifici e della fiducia incrollabile che avete riposto in me, lasciandomi sempre la libertà di inseguire le mie ambizioni. Grazie per aver sopportato i miei silenzi, le mie ansie e i miei sbalzi d'umore, accogliendoli sempre con un gesto silenzioso o una parola di conforto quando ne avevo più bisogno.

Una menzione speciale va sicuramente ai miei animali domestici, indipendentemente dal fatto che siano ancora qui con me o meno: mi hanno mostrato un amore senza limiti e occuperanno sempre un posto speciale nel mio cuore.

Ora, come potrei non ringraziare una persona che è diventata famiglia non per nascita, ma per scelta? Grazie alla mia migliore amica *Eli* (so che "Elisabetta" non ti sarebbe piaciuto) per esserci sempre stata, per capirmi come nessun altro e per accettare anche le parti più strane di me. In questi anni siamo cambiate e cresciute insieme, ma la certezza di averti al mio fianco è rimasta l'unico punto fermo in mezzo a tutto questo caos. Ti voglio un bene immenso e sono sicura che questo non cambierà mai.

Un ringraziamento va ai colleghi che hanno alleggerito questi quattro anni di lavoro con le chiacchiere a bordo vasca e i "gossipponi". Ringrazio anche tutti i bimbi a cui ho insegnato: vedere la loro crescita mi ha regalato un'immensa soddisfazione e spero che i momenti passati insieme siano stati altrettanto gioiosi per loro.

Ringrazio le mie compagne di squadra che mi hanno fatta sentire parte integrante della *Vis* e mi hanno permesso di staccare la mente. Anche se dovessi smettere con il basket per investire nel mio futuro professionale, il legame con voi non si spezzerebbe affatto.

Un grazie va a tutti gli amici del mio "paesino sperduto" e a quelli di Bologna e dintorni. Anche se la mia voglia di uscire non è sempre ai massimi livelli, ogni volta che sto con voi mi diverto sinceramente, soprattutto se c'è della Coca-Cola. Vi voglio bene.

In conclusione, ringrazio me stessa per aver superato i momenti difficili, per aver perseverato e per essere diventata la persona che sono ora. Mi ringrazio per aver creduto nelle mie capacità e, senza voler peccare di presunzione, sono fiera di aver concluso questo percorso senza mai essere bocciata a un esame. Questo traguardo è la dimostrazione che la determinazione e il tempo dedicato ripagano sempre.

Per fortuna non ho scritto questi ringraziamenti a mano, altrimenti la pagina sarebbe piena di macchie di lacrime. Grazie di cuore a tutti e buona fortuna a me per il futuro.