Advanced Deep Learning

Anatomy-Aware De-noising Framework for Low-Dose CT at Variable Dosages

GROUP 9

Rohan Bansal
Ajay
Dipanshu Thakur
Dipan Sen

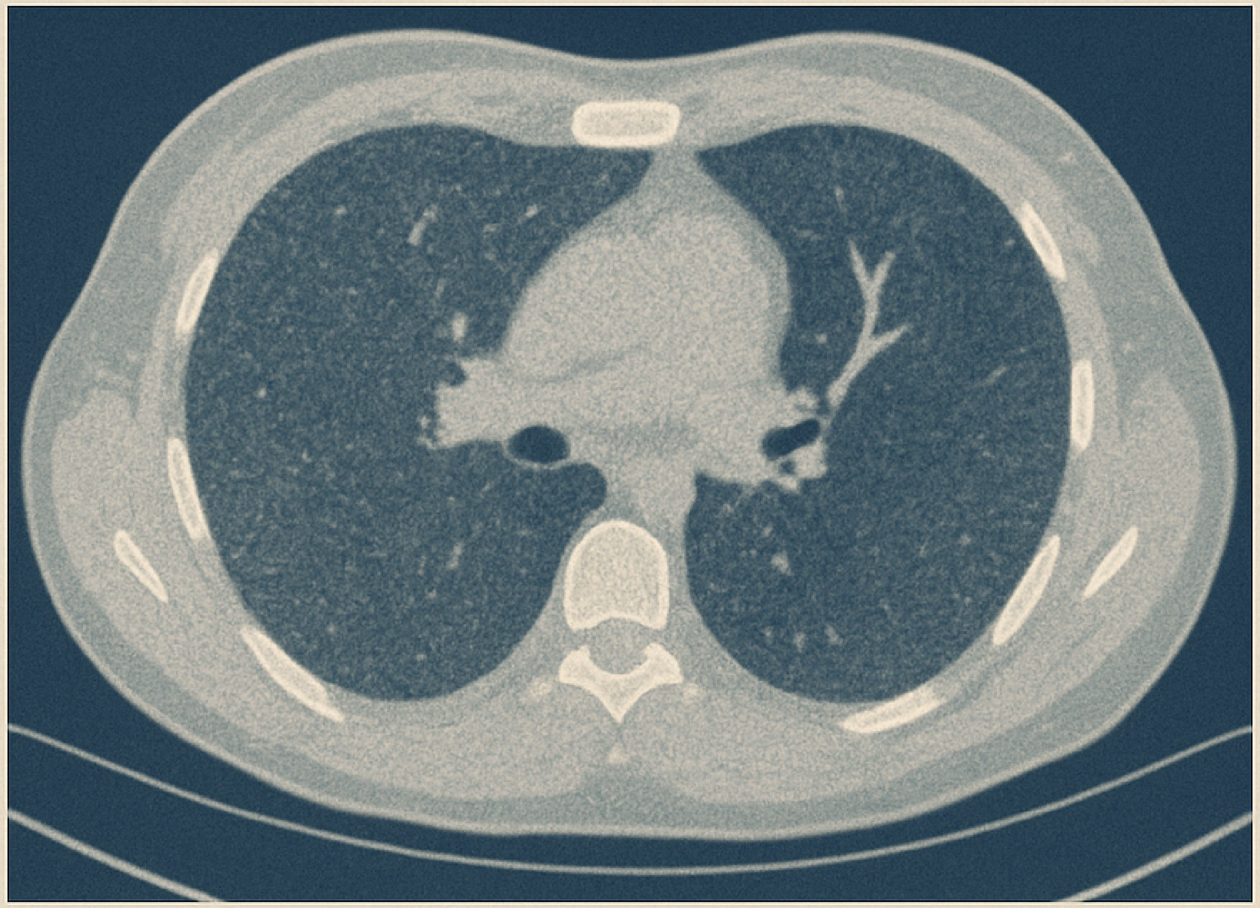
CHAPTER 1

Problem Statement

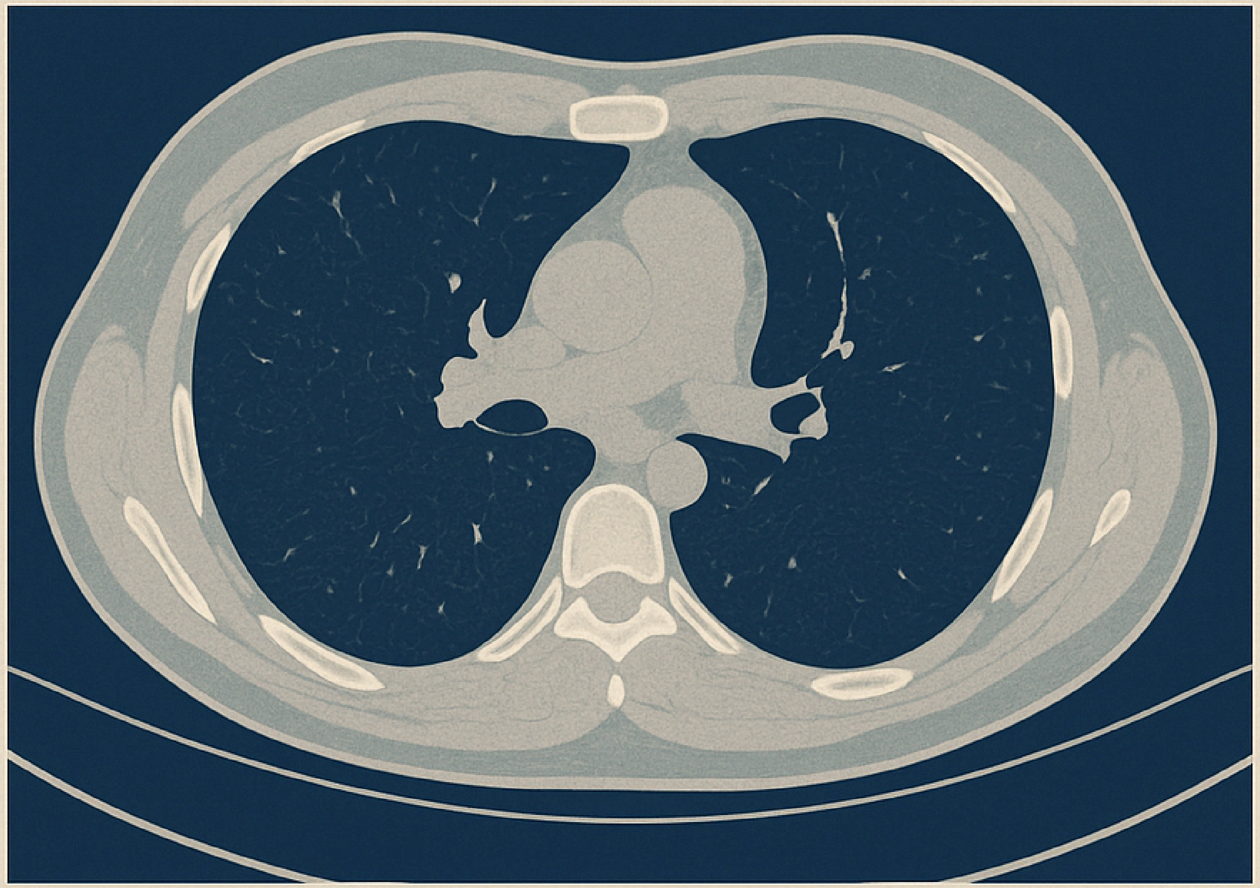
Problem Statement and Introduction

- **Radiation vs. fidelity trade-off:** LDCT reduces patient dose but substantially increases noise and artifacts, degrading the visibility of fine anatomical structures.
- **Project aim:** design a denoiser that restores NDCT-level quality across multiple dose levels (10-75%) while preserving subtle anatomy.
- **Safety-driven validation:** ensure clinical reliability using both image metrics and anatomy-aware checks like radiomic stability.

LDCT



NDCT



Dataset Preparation

Goal: Create realistic low-dose CT (LDCT) data by adding physics-based noise in projection (counts) space, then reconstruct.

LoDoPaB (parallel-beam)

- Convert image to projections .
- Pick dose (photons per ray); add Poisson shot noise (optionally electronic).
- Convert noisy counts to log, and reconstruct with FBP.

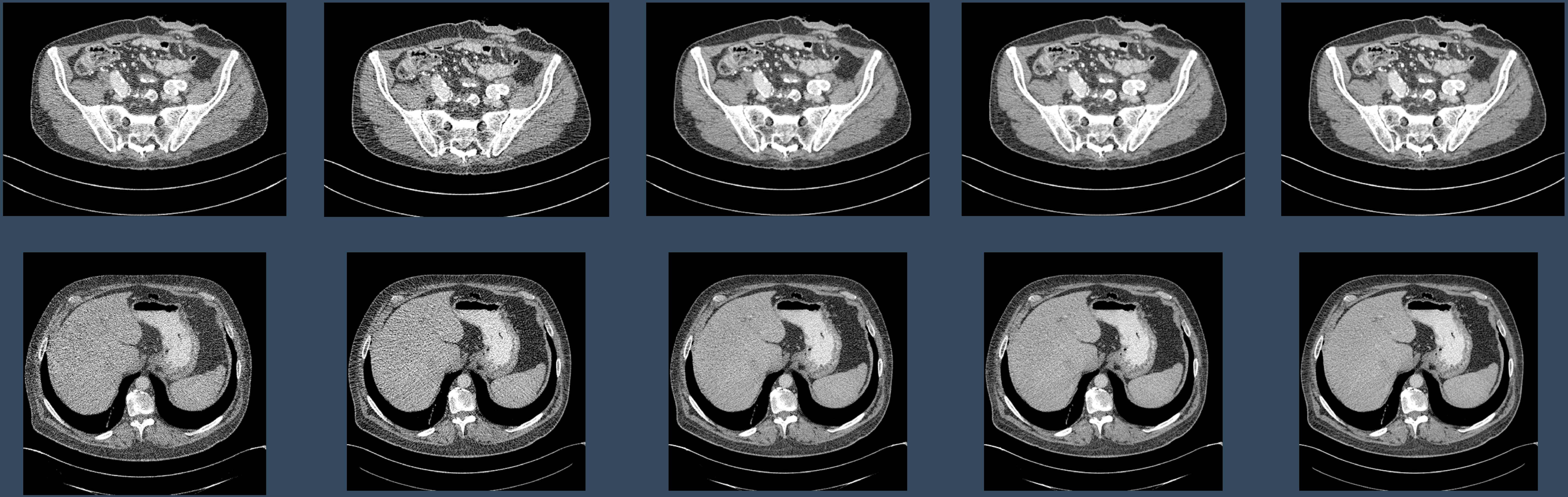
AAPM (fan-beam)

- Start from projections.
- Scale counts by dose.
- Add Poisson shot noise (optionally electronic); convert to log; reconstruct to image.
-

Output: NDCT–LDCT pairs at 10%, 25%, 50%, 70% dose

Zeng, D. et al., "A Simple Low-dose X-ray CT Simulation from High-dose CT Scans"

Samples from generated Dataset



10% 25% 50% 70% NDCT

CHAPTER 2

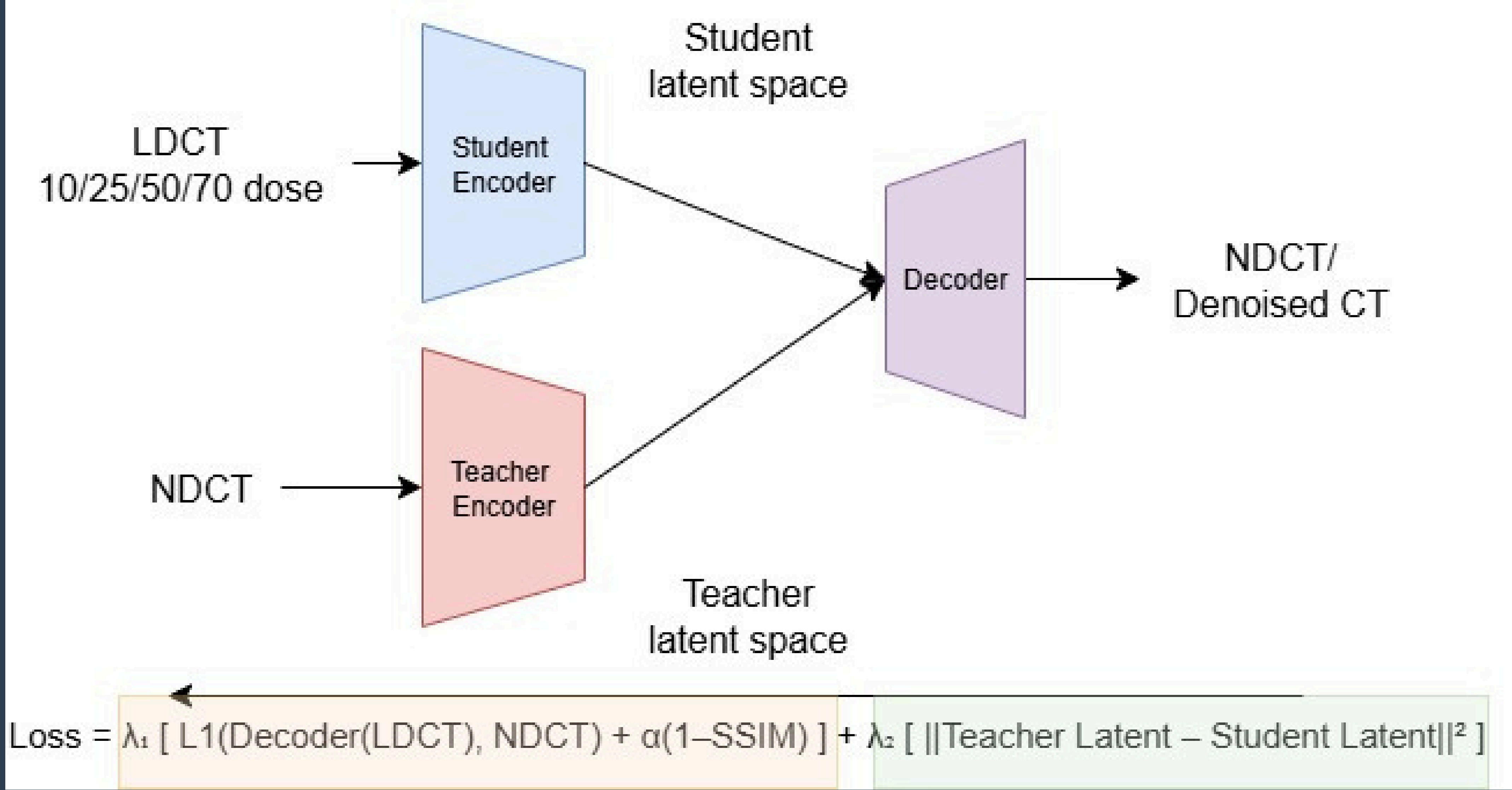
Proposed Solution

Baseline Architecture

Goal: Reconstruct high-quality Normal-Dose CT (NDCT) images from Low-Dose CT (LDCT)

Architecture Overview

- **Teacher Network**
  - Autoencoder trained only on NDCT images.
  - Learns a clean latent representation of anatomical structures.
- **Student Network**
  - Separate LDCT encoder trained to mimic the teacher's latent space.
  - Uses the shared decoder (from the teacher) to reconstruct NDCT-like outputs.



Proposed Improvements

- **Teacher–student latent distillation:** Student enc. LDCT → learns → Teacher NDCT latents.
- **Structural & perceptual constraints:** Multi-scale perceptual feature matching and gradient-consistency losses maintain edges and texture.
- **Feature map distillation:** Student intermediate feature maps are aligned to the teacher's (layer-wise L2) to transfer structural representation quality.

$$L\_base = \lambda_1 [\ L_1(\text{Denoised}, \text{NDCT} + \alpha(1 - \text{SSIM})\ ] + \lambda_2 \|\ z\_t - z\_s\ \|^2$$

$$L\_edge = |\nabla(\text{Denoised}) - \nabla(\text{NDCT})|_1 \text{ (Sobel gradient consistency)}$$

$$L\_msssim = 1 - \text{MS-SSIM}(\text{Denoised}, \text{NDCT})$$

$$L\_perc = ||\varphi(\text{Denoised}) - \varphi(\text{NDCT})||^2$$

$$L\_feat\_distillation = \delta\|F\text{\^{}}s - F\text{\^{}}t\|^2$$

CHAPTER 3

Implementation

Key Aspects of Implementation

- **Teacher–student encoder:** NDCT → Teacher encoder → Z_t and LDCT (+ dose conditioning MLP) → Student encoder → Z_s → shared decoder → reconstruction.

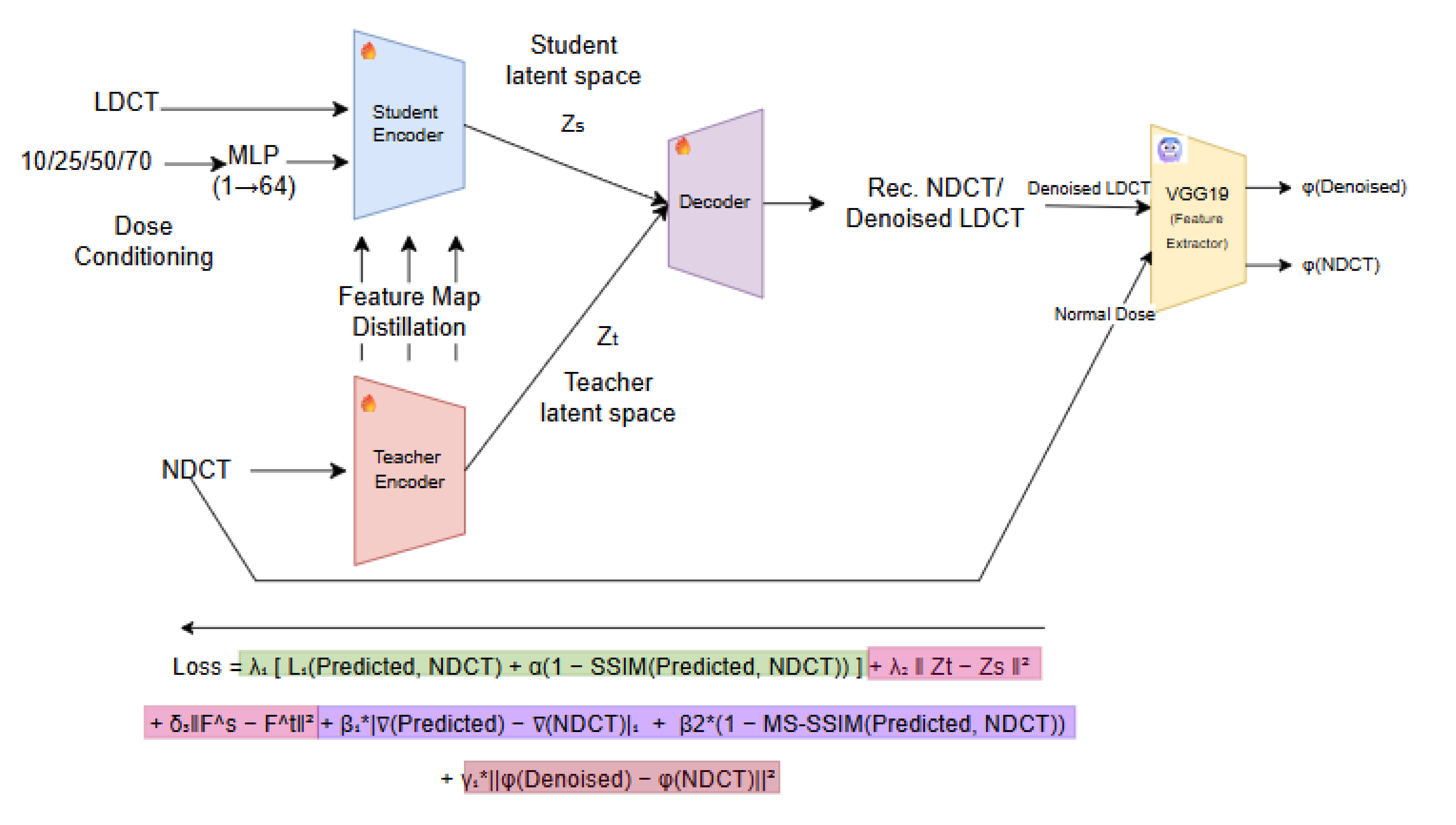
- **Dose conditioning:** scalar dose (10/25/50/70) passed through a small MLP (1→64) and injected into the student encoder to handle dose-dependent noise.

- **Feature-map + latent distillation:** layer-wise feature alignment and latent L2 loss ($\|Z_t - Z_s\|^2$) — transfers structural representation, not just output.

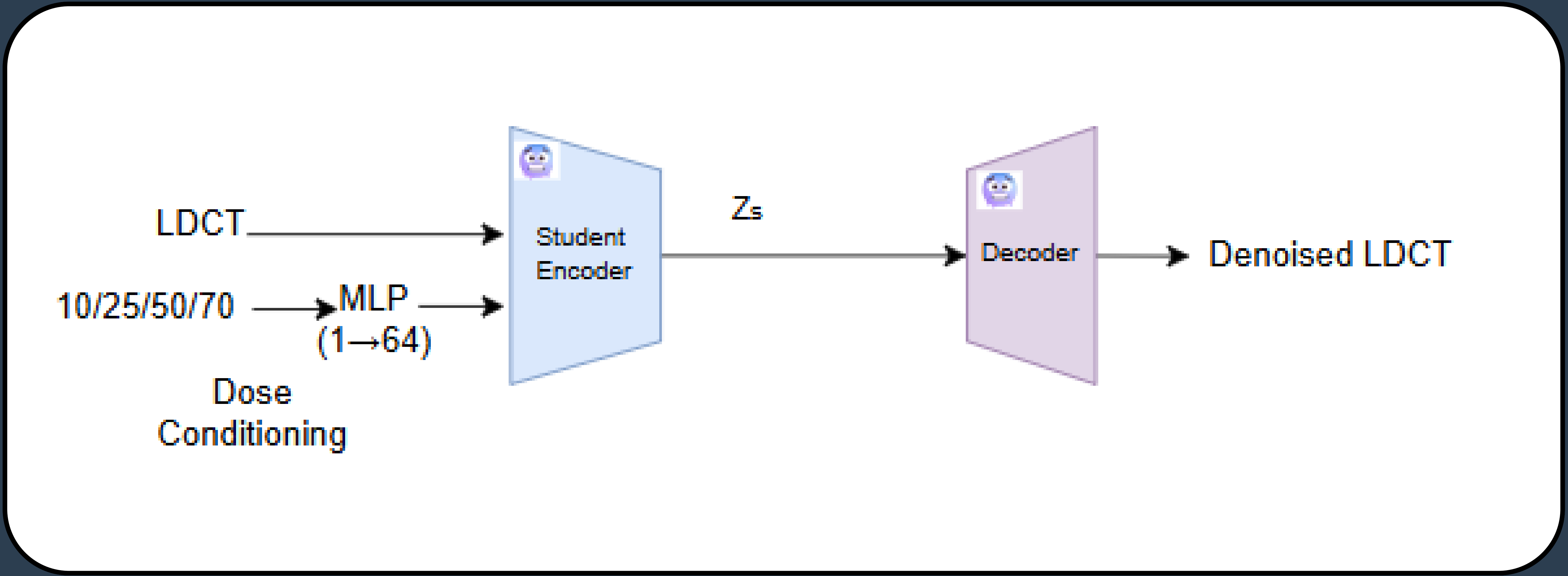
- **Compact loss summary (callout):** Recon = pixel (L1) + $\alpha(1-\text{SSIM})$; + $\lambda\_\text{latent}\|Z_t - Z_s\|^2$; + feature-map L2 + gradient consistency (Sobel) + MS-SSIM + VGG perceptual.

- **Training recipe highlight:** train teacher then freeze teacher and train student with frozen teacher to mimic NDCT feature maps and latent.

Training Pipeline



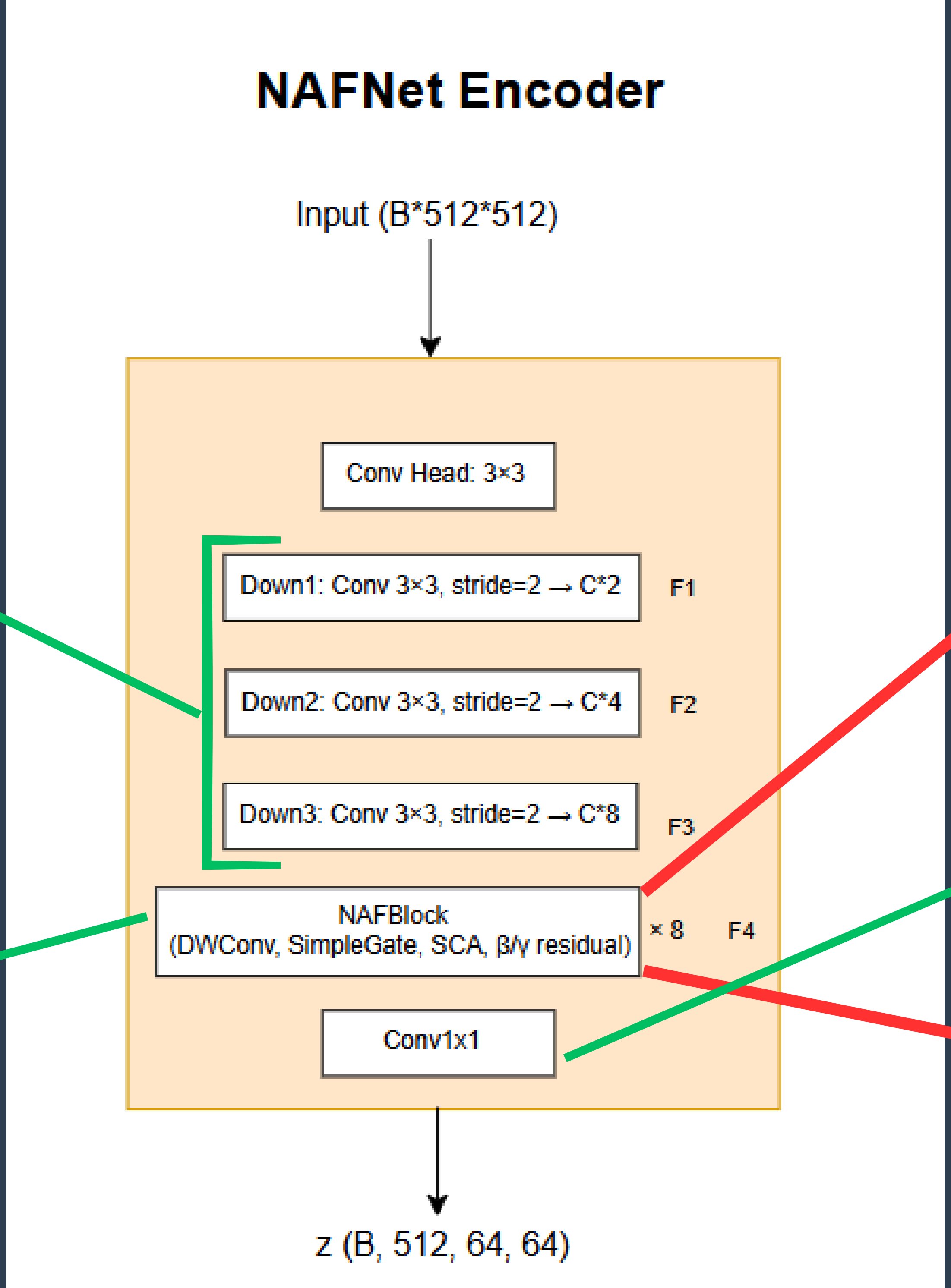
Inference Pipeline



- **Single-pass denoising:** LDCT is fed into the trained student encoder, producing the latent representation Zs used by the decoder to generate the final denoised CT.
- **Dose-aware inference:** The input dose level (10/25/50/70%) is mapped through a lightweight MLP conditioning vector, enabling dose-adaptive noise suppression.

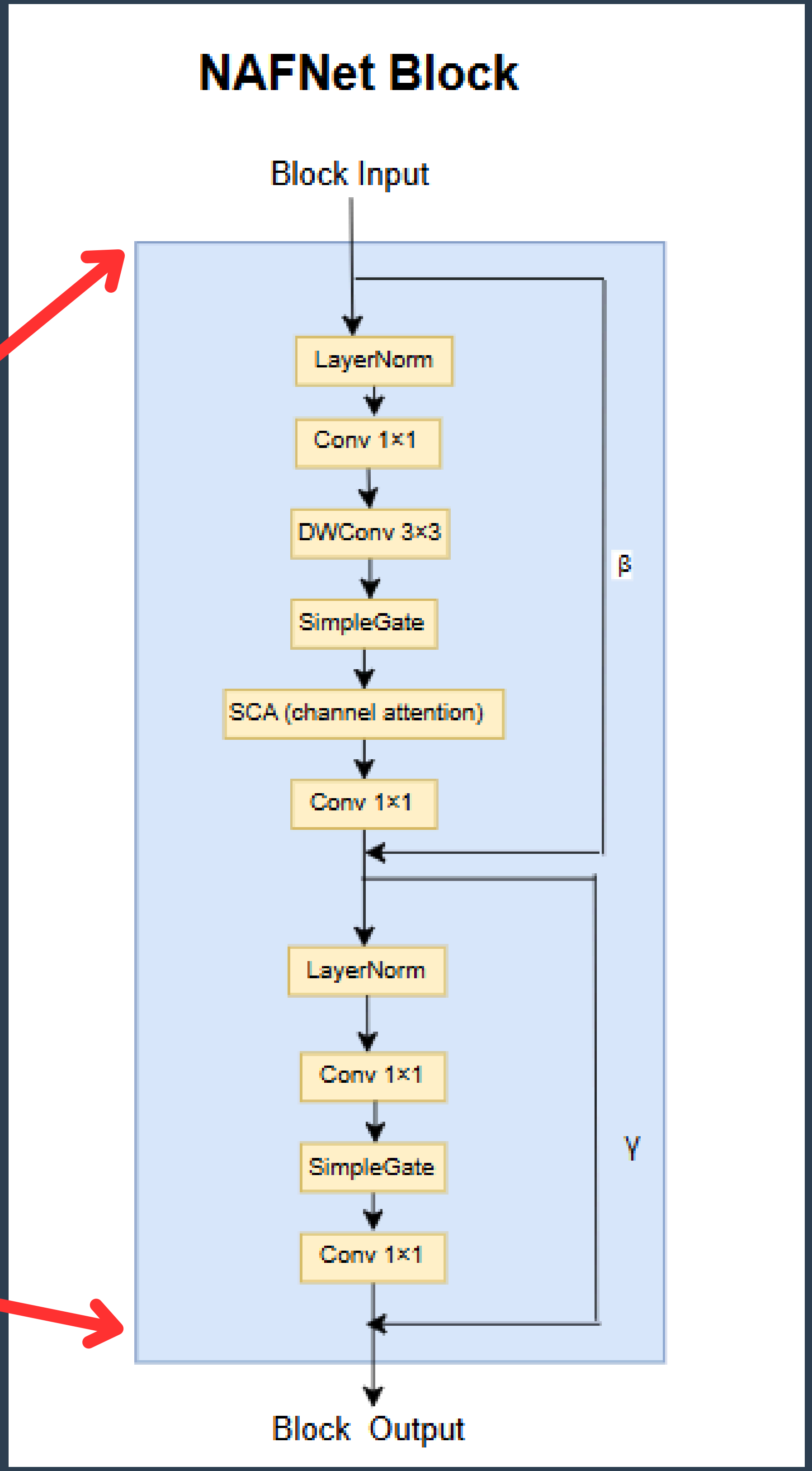
Encoder Architecture

Progressive stride-2 downsampling builds a multi-scale anatomical feature hierarchy.



Gated depthwise conv + channel attention with residual scaling → efficient, stable feature extraction.

Latent projection compresses features into a distillation-friendly representation (Z).



Link: https://www.researchgate.net/figure/Illustration-of-a-Simplified-Channel-Attention-in-NAFNet-40-b-Channel-Attention-in_fig8_379493321

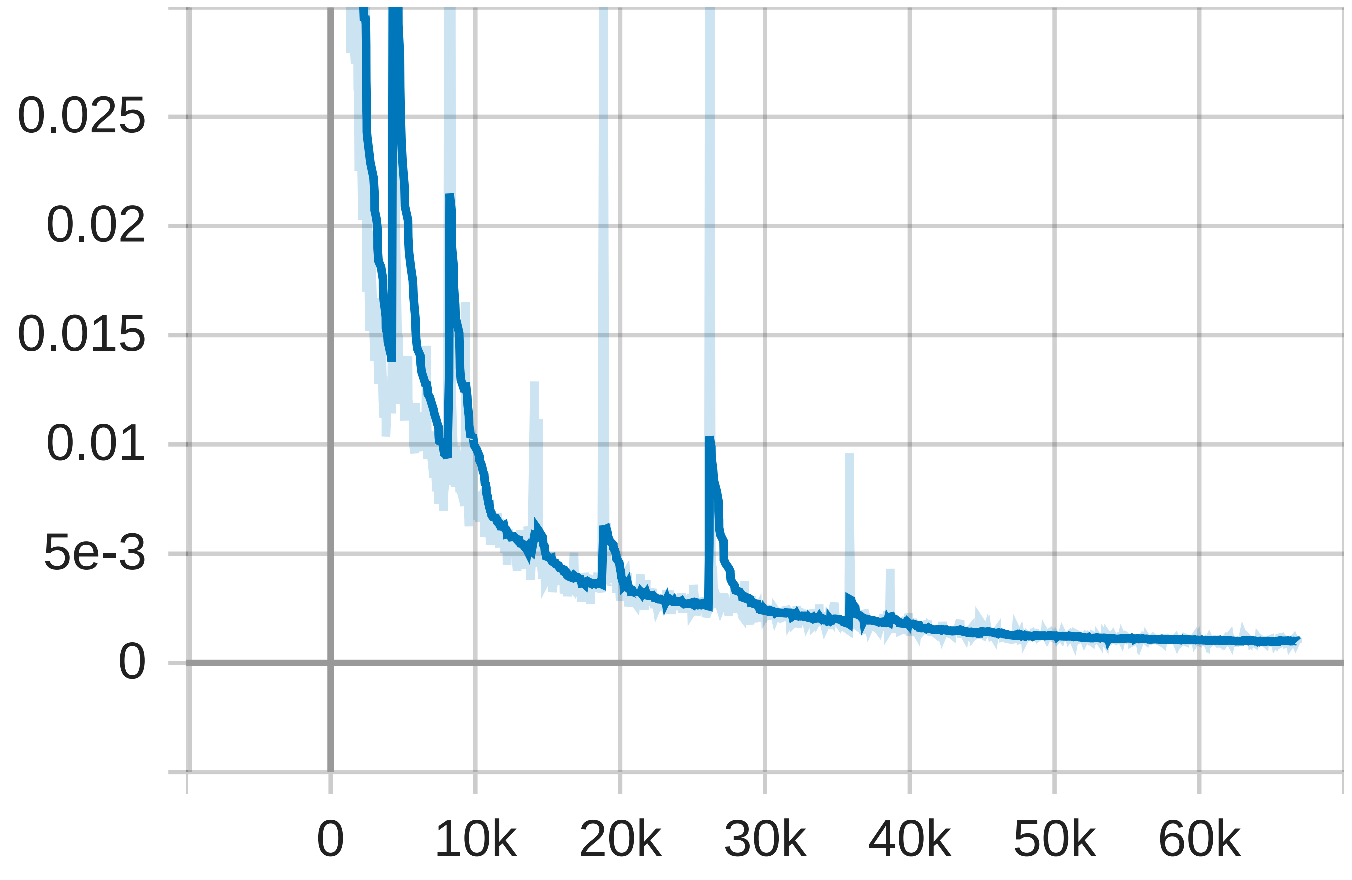
CHAPTER 4

Results and Analysis

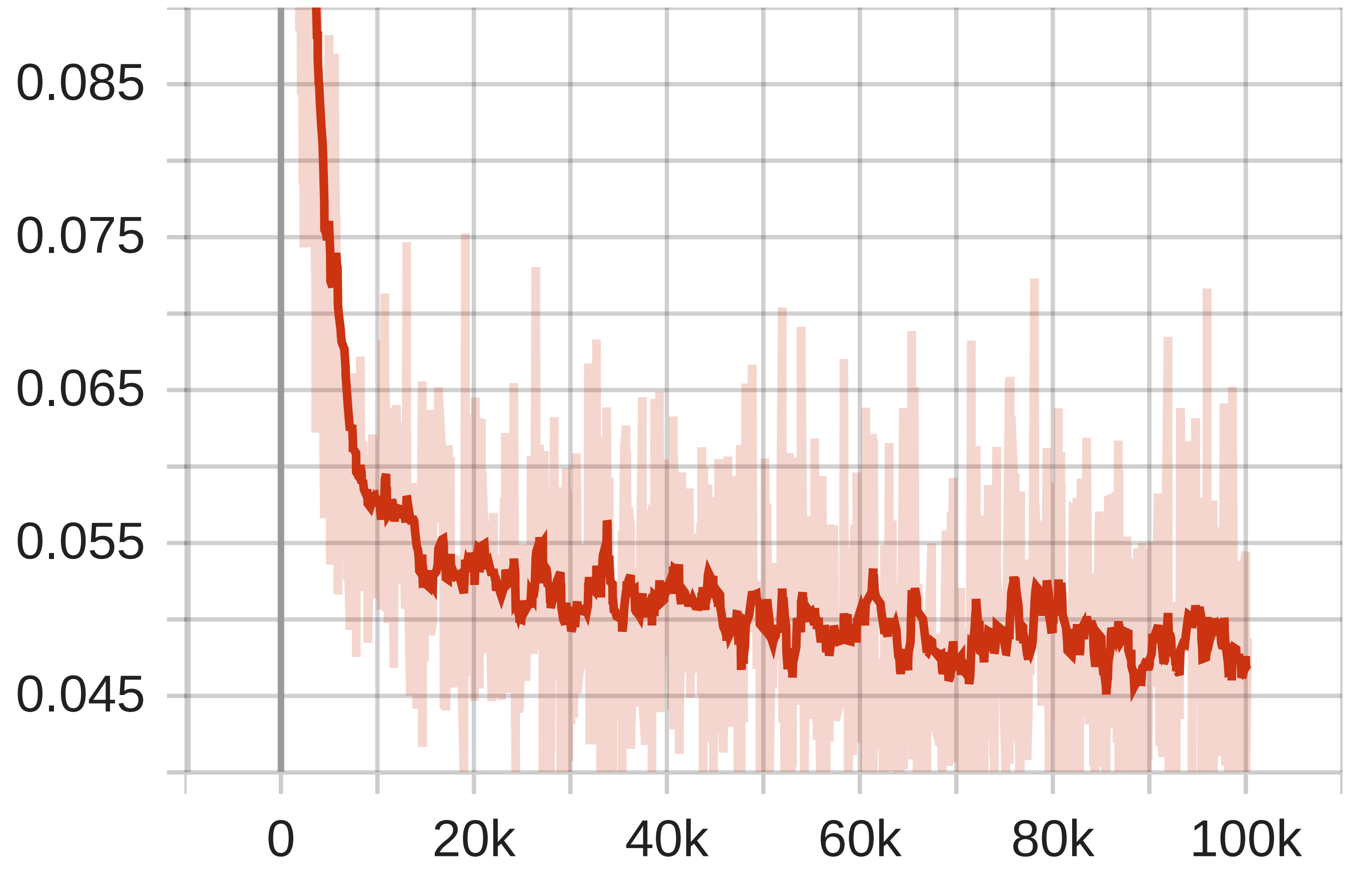
Quantitative Analysis

Metric	10% GT	10% Baseline	10% Current	25% GT	25% Baseline	25% Current	50% GT	50% Baseline	50% Current	70% GT	70% Baseline	70% Current
PSNR	29.34	37.69	**40.81**	34.92	37.42	**39.99**	33.28	38.14	**43.42**	33.83	38.19	**44.01**
SSIM	0.598	0.903	**0.949**	0.808	0.899	**0.941**	0.814	0.914	**0.969**	0.841	0.915	**0.973**
RMSE	0.0360	0.0132	**0.00927**	0.0184	0.0136	**0.0102**	0.0236	0.0125	**0.00688**	0.0224	0.0124	**0.00643**

Table 1: Comparison across dose levels (Current = NAFNet)



Teacher Loss (Total)

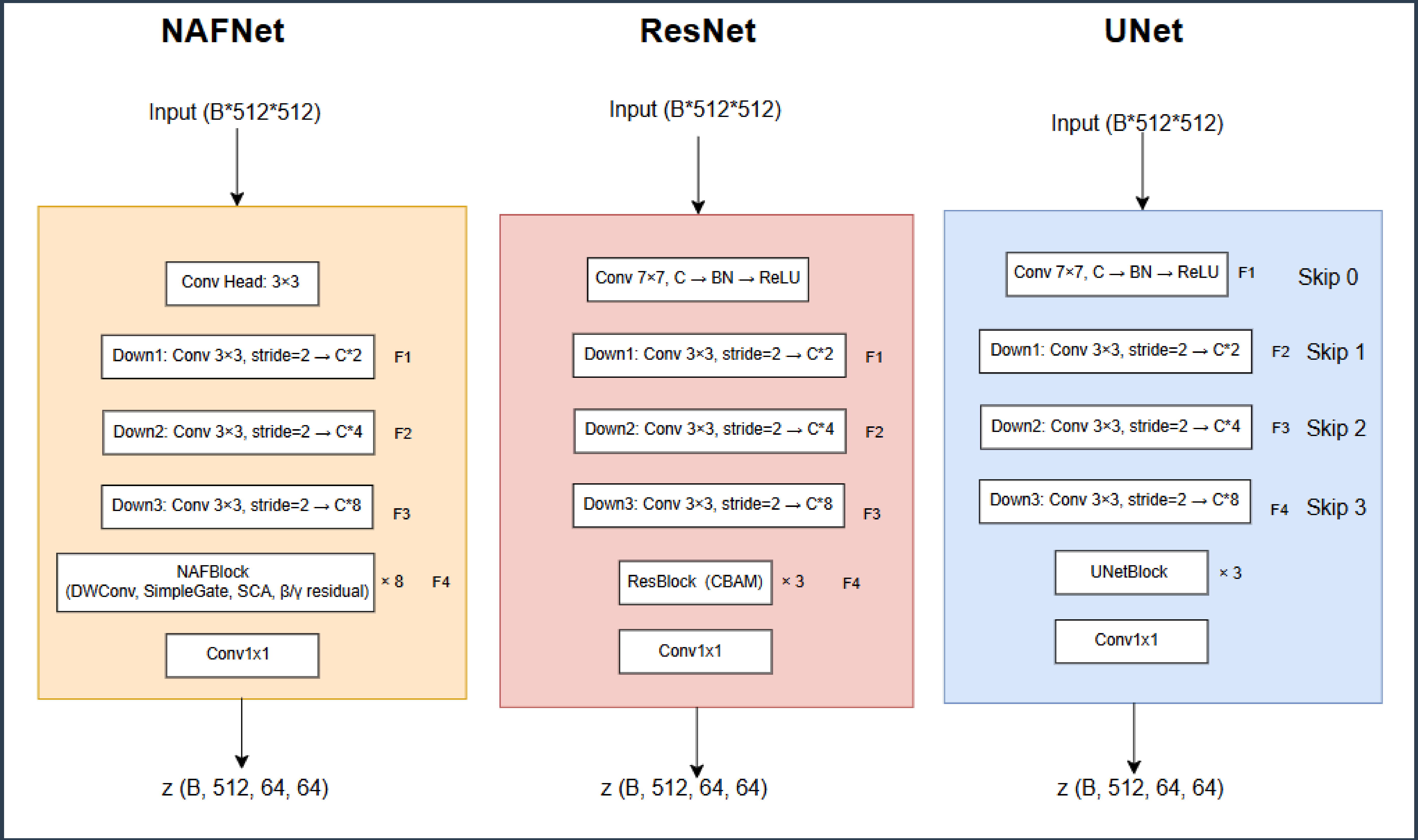


Student Loss (Total)

Key Insights

- **Consistent improvement across all dose levels:** The proposed model achieves higher PSNR/SSIM and lower RMSE than the baseline at 10%, 25%, 50%, and 70% doses **— showing strong dose-robust generalization.**
- **Largest gains:** PSNR jumps by ~3–4 dB over baseline, and SSIM reaches 0.94+, demonstrating superior recovery of fine anatomical details under extreme noise.
- **High-dose performance remains strong:** Even at 70% dose, the model surpasses both Ground truth and baseline, proving stable enhancement rather than overfitting to low-dose **— no major artifacts.**
- **Teacher training stability:** Teacher loss rapidly decreases and flattens smoothly, indicating a well-formed NDCT latent space that the student can reliably learn from.
- **Student convergence trend:** Student loss steadily declines toward a stable band - successful distillation and reconstruction learning despite higher initial noise.

Ablation Studies

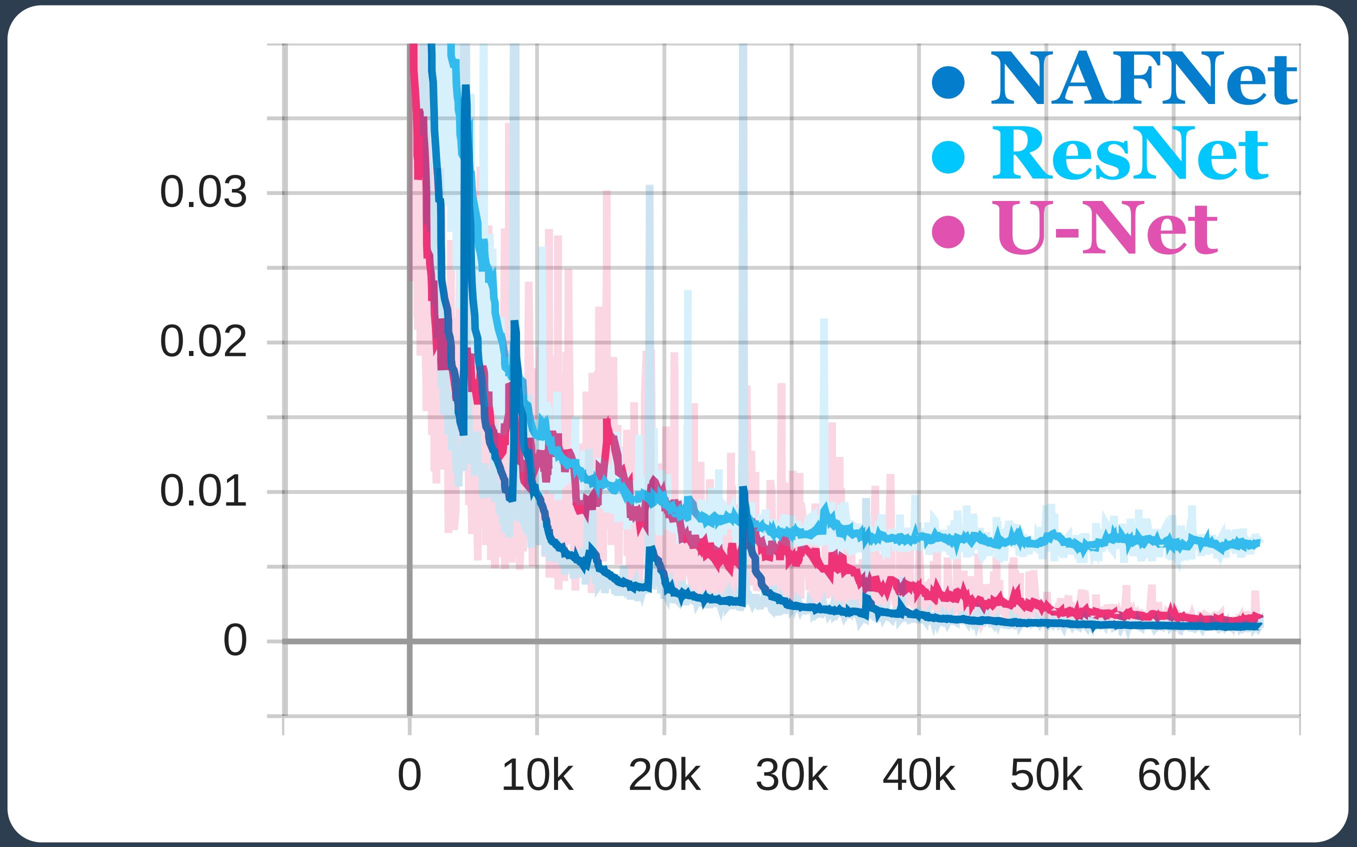


Link: https://arxiv.org/abs/1512.03385

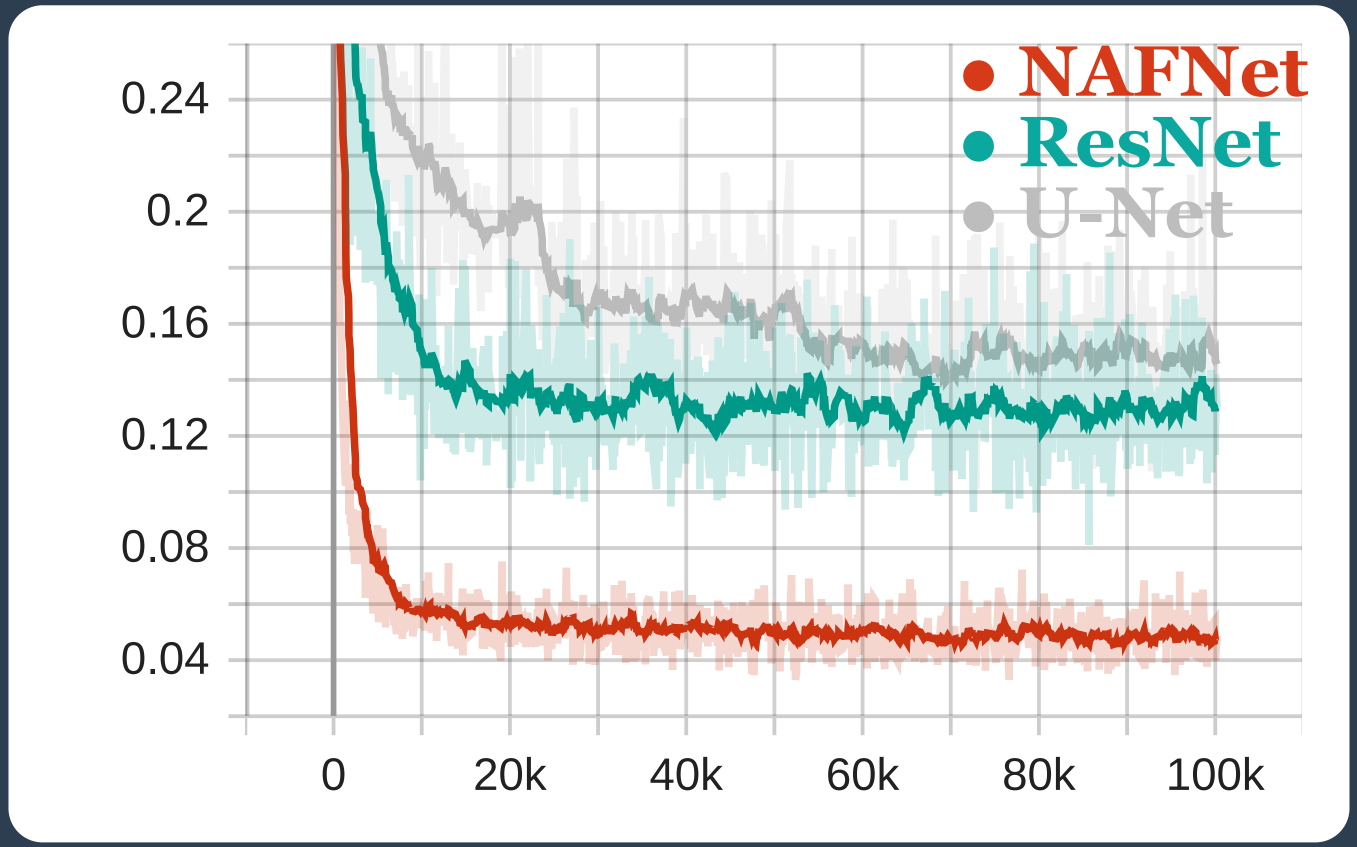
Link: https://www.researchgate.net/publication/345261676_U-Net_and_its_variants_for_medical_image_segmentation_theory_and_applications

Results from Ablation

Metric	10% PSNR	10% SSIM	10% RMSE	25% PSNR	25% SSIM	25% RMSE	50% PSNR	50% SSIM	50% RMSE	70% PSNR	70% SSIM	70% RMSE
ResNet	39.1803	0.9197	0.01110	38.5782	0.9132	0.01193	40.9355	0.9413	0.00905	41.2465	0.9448	0.00873
UNet	36.9165	0.8995	0.01477	37.7752	0.8949	0.01314	38.2611	0.9291	0.01278	38.3739	0.9314	0.01263
NAFNet (RadImageNet)	40.2783	0.9443	0.00985	39.4457	0.9360	0.01085	42.8251	0.9673	0.00735	43.3560	0.9713	0.00691
NAFNet (No Dose, VGG)	40.4855	0.9470	0.00961	39.6989	0.9398	0.01054	43.0054	0.9683	0.00720	43.5575	0.9722	0.00676
NAFNet (VGG)	**40.8063**	**0.9487**	**0.00927**	**39.9871**	**0.9413**	**0.01020**	**43.4179**	**0.9692**	**0.00688**	**44.0137**	**0.9731**	**0.00643**



Teacher Loss (Total)



Student Loss (Total)

https://github.com/BMEII-AI/RadImageNet

Key Insights

- Per-model metric rows show consistent numeric gaps across doses, indicating the architectures produce reproducible, dose-dependent behavior.
- The “NAFNet (No Dose)” entry shows values very close to the dose-conditioned variant — suggesting that explicit dose-conditioning produces only a small change for the dataset.
- All teacher loss traces exhibit a rapid early drop followed by a long tail, indicating fast initial learning of NDCT reconstruction and then fine-tuning of latent structure.
- Student losses show a sharp initial decrease and then settle into distinct steady bands with varying levels of short-term variance.
- **The NAFNet** trace reaches a lower, tighter steady band with small variance, while the other traces show higher steady-state variance and a slower long-tail — indicating differences in how quickly and stably each student/architecture internalizes the teacher latents.

Best and Worst Performing Variants

Table 1: Parameter Summary of Teacher and Student Encoders

Architecture	Teacher Encoder Params	Student Encoder Params	Decoder Params
NAFNet	4,257,024	4,527,552	4,256,513
ResNet	4,088,550	4,192,134	4,087,527
U-Net	4,088,550	4,292,518	4,476,167

1. **NAFNet (BEST):** efficient local filtering (DWConv) + SimpleGate preserves high-frequency anatomical edges while SCA focuses channels, so denoising keeps structure.

2. **ResNet:** lacks the gating + attention interplay, so it smooths small features more.

3. **U-Net (WORST):** skip pathways help structure but can reintroduce noise and lack the efficient gated micro-ops that recover tiny details.

4. **No-Dose NAFNet:** strong learned priors + robust training let model generalize across dose levels; conditioning helps but is not strictly necessary for core reconstruction.

Qualitative Analysis - Demonstration

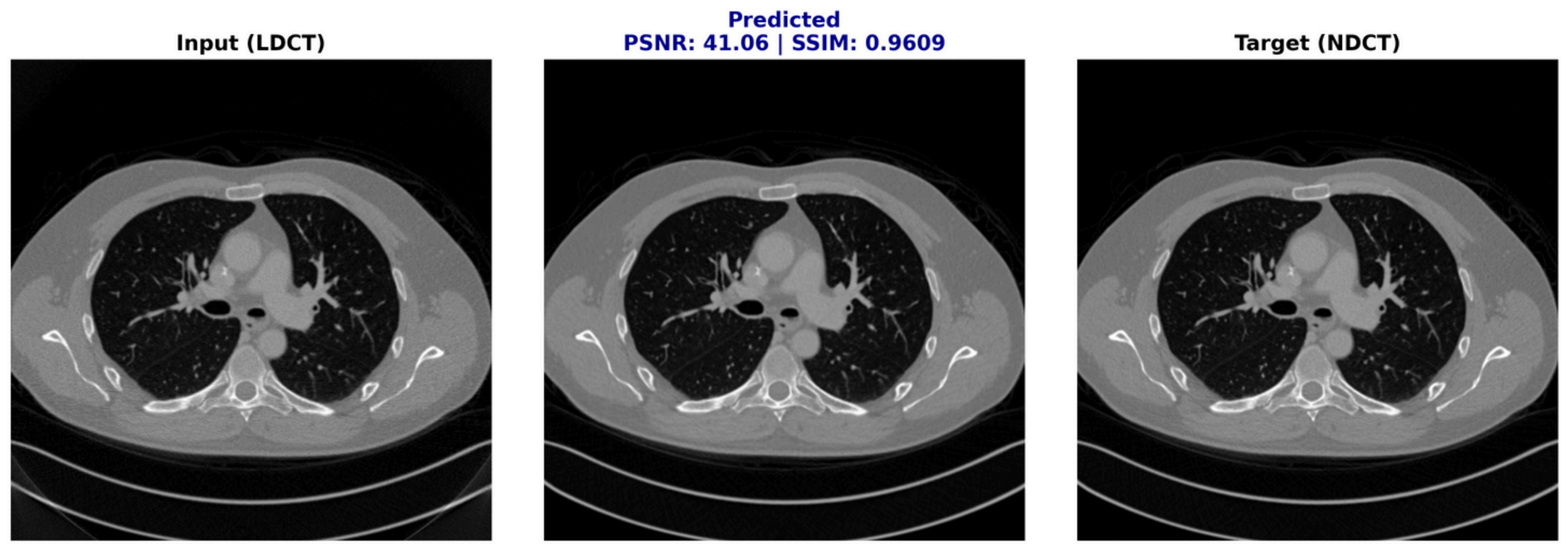


Fig 1. 10% Dose, (PSNR, SSIM)=(41.06,0.9609)

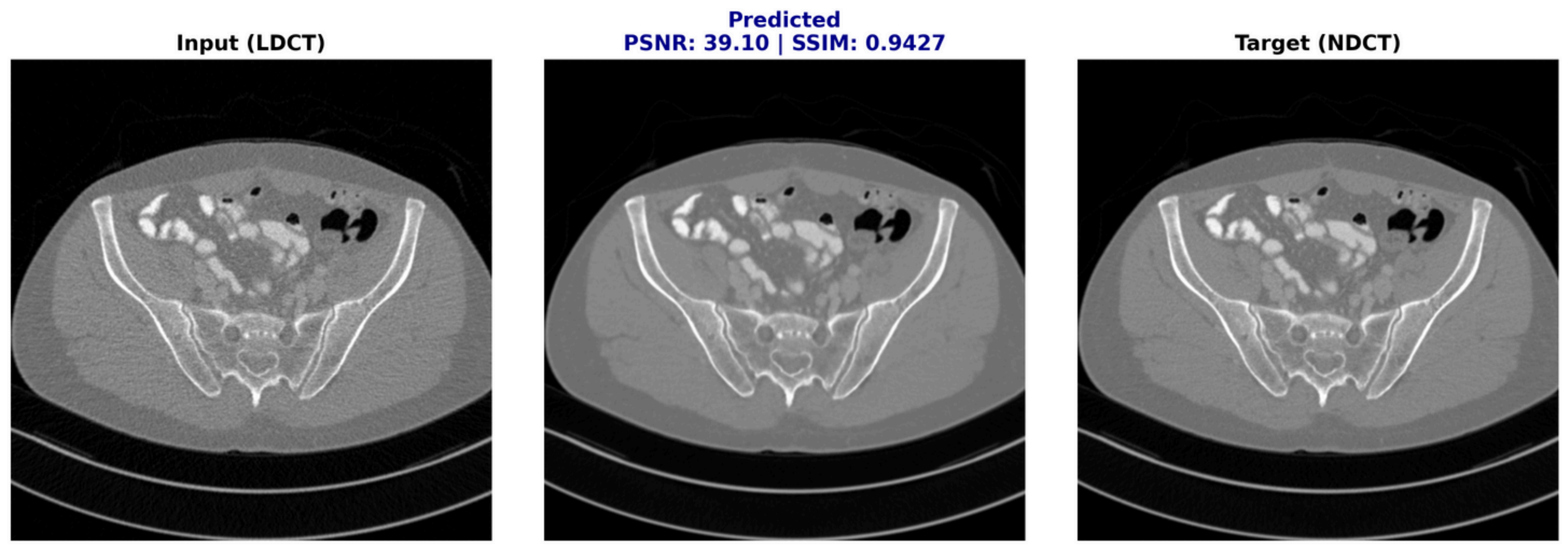


Fig 2. 25% Dose, (PSNR, SSIM)=(39.10, 0.9427)

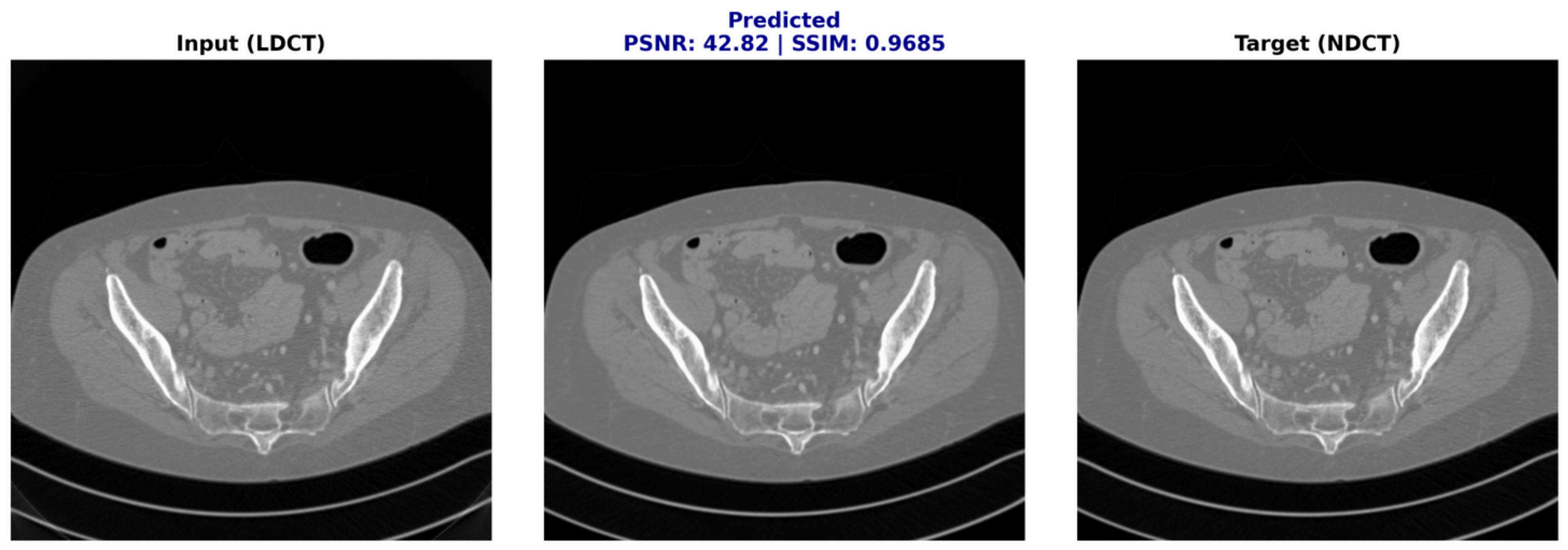


Fig 3. 50% Dose, (PSNR, SSIM)=(42.82, 0.9685)

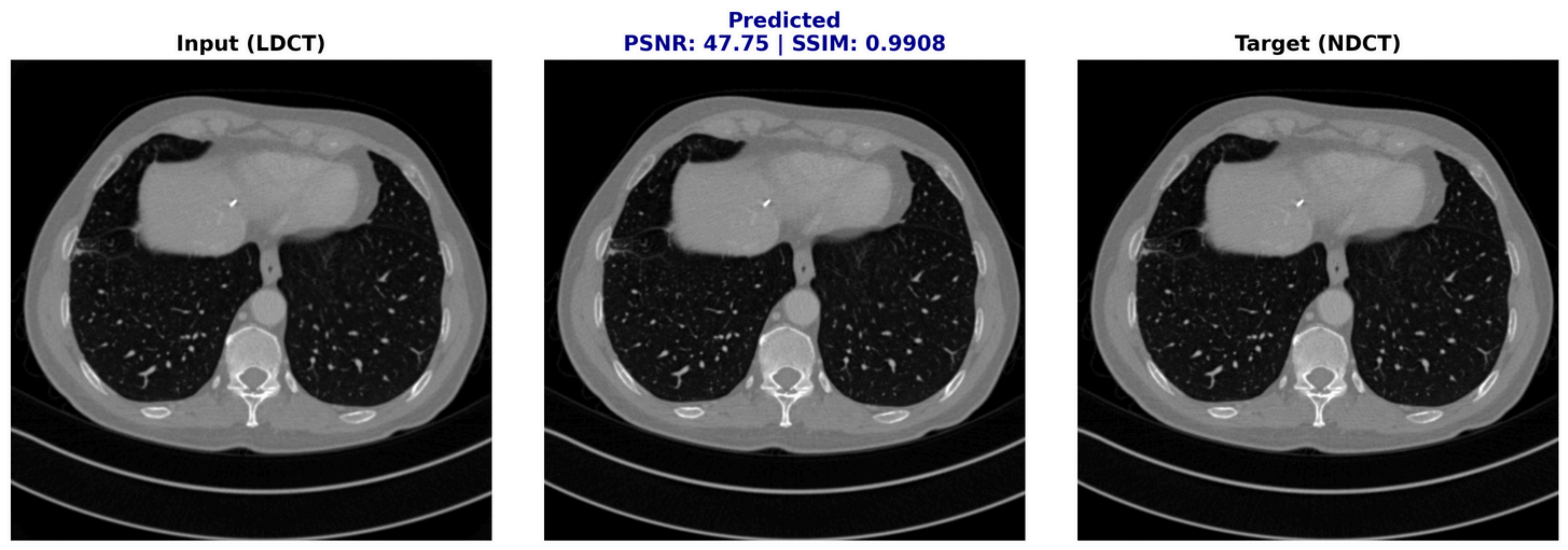


Fig 4. 70% Dose, (PSNR, SSIM)=(47.75,0.9998)

Key Insights

- **10% dose —** Predicted image strongly reduces noise but shows noticeable over-smoothing of fine soft-tissue texture and tiny vessels.
- **25% dose** — Predicted image delivers effective denoising with well-preserved edges and small structures, minimal smoothing.
- **50% dose —** Predicted image is nearly indistinguishable from NDCT, preserving anatomy and texture with high fidelity.
- **70% dose —** Predicted image maintains excellent detail and contrast with negligible residual noise and no obvious artifacts.

CHAPTER 5

Conclusion and Future Work

Conclusion

1. Dose-Adaptive Generalization with Teacher-Student Distillation
Successfully developed a NAFNet-based framework that achieves NDCT-quality reconstruction across 10-70% dose levels through anatomy-aware latent distillation, multi-scale perceptual constraints, and dose-conditioning, outperforming ResNet and U-Net baselines.

2. Robust Clinical Performance with Structural Preservation
Demonstrated superior quantitative results (PSNR improvement of 3-4 dB, SSIM reaching 0.94+) and qualitative anatomical fidelity, maintaining critical edge details and texture even at extreme 10% dose with minimal over-smoothing or artifacts.

3. Efficient Architecture Through Gated Feature Learning
NAFNet's depthwise convolution with SimpleGate and SCA attention mechanisms enabled stable training convergence and efficient feature extraction, proving that gating operations preserve high-frequency anatomical details better than residual-only or skip-connection-heavy architectures.

Future Works Possible...

1. Mamba-Integrated Hybrid Architecture for Global-Local Context
Develop a Mamba-based model integrating selective state-space mechanisms with attention and CNN modules to capture global dependencies and local texture simultaneously, reducing computational complexity from while improving contextual awareness across CT slices.

2. Cross-Anatomy and Multi-Modal Generalization
Extend the framework to further-more anatomical regions and integrate multi-modal medical imaging, e.g, MRI, PET-CT fusion, using unified latent representations, enabling transfer learning across different imaging protocols and dose distributions.

3. Real-Time Clinical Deployment with Uncertainty Quantification
Optimize the model for real-time inference on clinical CT scanners using model compression (pruning, quantization) and develop Bayesian uncertainty estimation modules to flag low-confidence reconstructions, ensuring radiologist trust and safety-critical validation.

THANK YOU !!!

GROUP 9