

Enhancing Structural Quality in Medical X-ray Super-Resolution via Domain-Aware Loss

Aryan Tyagi

1aryantyagi@gmail.com

Guru Gobind Singh Indraprastha University

Neeraj Sharma

Indian Institute of Technology (BHU)

Prachi Pant

Guru Gobind Singh Indraprastha University

Mohini Manav

Lala Lajpat Rai Memorial Medical College

Article

Keywords: Super-resolution, Medical imaging, X-ray, Deep learning, Transformer, ESRGAN, Loss function

Posted Date: October 19th, 2026

DOI: <https://doi.org/10.21203/rs.3.rs-10532589/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Additional Declarations: No competing interests reported.

Abstract

Upscaling medical X-ray images is more challenging than enhancing natural images, because preserving accurate structural and boundary information is more important for diagnostic correctness than producing a realistic appearance. Super-resolution methods such as SRGAN, ESRGAN, and the Hybrid Attention Transformer (HAT) perform well on natural images, but when applied to medical images they can cause texture hallucination, obscure anatomical boundaries, and remove clinically important features. This study presents a comprehensive comparison of four approaches: standard SRGAN, standard ESRGAN with VGG-based perceptual loss, HAT with L1 loss, and the proposed optimised models ESRGAN_USELoss and HAT_USELoss, which incorporate a domain-specific loss function designed for X-ray imaging. The proposed USELoss (U-Net, Sobel and Edge Loss) replaces the computationally heavy VGG19 perceptual loss with a lightweight U-Net encoder that preserves multi-scale anatomical structures, and adds a Sobel-based edge loss to improve boundary accuracy, prioritising diagnostic fidelity over perceptual sharpness. On the PadChest X-ray dataset the quantitative results show large gains: standard ESRGAN reaches only 22.21 dB PSNR, whereas ESRGAN_USELoss reaches 34.04 dB, while HAT_USELoss achieves the best SSIM (0.9215) and LPIPS (0.1607). The proposed methods improve edge-preservation metrics by 5.5%, demonstrating that domain-specific loss functions are effective for clinically reliable medical image super-resolution.

<https://github.com/1aryantyagi/Segmentation-Paper> is the Github repository of the code.

Introduction

X-ray imaging is one of the most widely used modalities in clinical diagnosis because it is non-invasive, low cost, and fast [3]. Image quality is directly linked to diagnostic confidence and to the detection of subtle pathological changes [1, 27], yet practical limitations such as radiation dose, equipment constraints, and acquisition time often yield images of inadequate resolution [14]. Deep-learning super-resolution can reconstruct high-resolution images from low-dose inputs, supporting the ALARA principle while preserving diagnostic quality [20, 22, 25]. The problem is that super-resolution methods designed for natural images do not transfer reliably to medical images, where structural fidelity is paramount.

A wide range of approaches has been proposed to improve image super-resolution, as summarised in Table 1. Early convolutional methods such as SRCNN [7], VDSR [13], EDSR [17] and RCAN [32] established that learned reconstruction outperforms classical interpolation. Generative-adversarial methods including SRGAN [15] and ESRGAN [26] introduced perceptual and adversarial losses for photo-realistic textures, while transformer-based models such as SwinIR [16] and HAT [6] exploited self-attention to capture long-range dependencies. In parallel, the loss function itself has been studied extensively, from pixel and structural losses [33, 34] to VGG-based perceptual loss [12, 24] and adversarial loss [10]. However, most of this work targets natural images; when transferred to medical imaging, adversarial training can hallucinate anatomically implausible structures and ImageNet-based perceptual losses introduce a domain mismatch that degrades structural fidelity.

Table 1
Previous approaches to image super-resolution and loss-function design

Ref	First author surname	Description
7	Dong et al	Introduced SRCNN, the first CNN for image super-resolution, showing end-to-end learning outperforms classical interpolation.
13	Kim et al	Proposed VDSR, using a very deep network with residual learning to ease training of deeper models.
17	Lim et al	Developed EDSR with enhanced deep residual blocks for single-image super-resolution.
32	Zhang et al	Introduced RCAN, very deep residual channel-attention networks for super-resolution.
15	Ledig et al	Proposed SRGAN, combining adversarial training with VGG-based perceptual loss for photo-realistic textures.
26	Wang et al	Developed ESRGAN with Residual-in-Residual Dense Blocks and a relativistic discriminator, improving perceptual quality.
8	Dosovitskiy et al	Introduced the Vision Transformer (ViT), enabling long-range dependency modelling for vision tasks.
16	Liang et al	Proposed SwinIR, applying window-based self-attention to image restoration.
6	Chen et al	Developed HAT, a hybrid attention transformer combining window and channel self-attention for image restoration.
30	Zamir et al	Proposed Restormer, an efficient transformer for high-resolution image restoration.
31	Zhang et al	Introduced a Swin-Conv-UNet hybrid CNN-transformer for blind image denoising.
4	Cao et al	Applied Swin-Unet, a pure-transformer U-Net, to medical image segmentation.
5	Chen et al	Proposed TransUNet, combining transformers with U-Net for medical image segmentation.
11	Hatamizadeh et al	Developed UNETR, a transformer-based architecture for 3D medical image segmentation.
33	Zhao et al	Compared L1, L2, SSIM and MS-SSIM losses, showing L1 yields sharper results than L2.
12	Johnson et al	Introduced perceptual loss based on pre-trained VGG features for super-resolution and style transfer.
34	Wang et al	Proposed the Structural Similarity Index (SSIM) for structure-aware image quality assessment.

Ref	First author surname	Description
10	Goodfellow et al	Introduced generative adversarial networks, the basis of the adversarial loss.
9	Gonzalez et al	Described gradient-domain (Sobel/Laplacian) operators for edge-preservation losses.
24	Simonyan et al	Proposed the VGG networks (trained on ImageNet) widely used to compute perceptual loss.

While prior work has focused mainly on architectural advances, the role of loss-function design for domain-specific medical super-resolution remains underexplored. This study investigates the hypothesis that a domain-aware loss, USELoss (U-Net Sobel Edge Loss), improves structural and edge-preservation metrics (e.g. SSIM, PSNR, edge discrepancy) relative to standard VGG-based perceptual loss, across both GAN-based (ESRGAN) and transformer-based (HAT) architectures, without requiring architectural redesign.

Methodology

Dataset

Experiments used the PadChest chest X-ray dataset, a large collection of clinically annotated radiographs acquired at San Juan Hospital, Spain [2]. The dataset contains high-quality posterior-anterior chest X-ray images with expert radiological reports, making it well suited to evaluating medical image super-resolution. PadChest is a publicly available dataset and is acknowledged accordingly.

A total of 600 images were used, split at the image level into 500 for training and 100 for testing. The test set was never used for model selection or hyper-parameter tuning, and identical data splits were applied across all methods to guarantee fair comparison. No additional validation set was used; all reported metrics are computed on the held-out test set.

Model

Three representative architectures - SRGAN, ESRGAN, and the Hybrid Attention Transformer (HAT) - were evaluated, capturing the evolution from early GAN-based to advanced GAN and transformer-based super-resolution. A unified, domain-specific loss framework (USELoss) was applied across these architectures to ensure anatomically meaningful reconstruction and to enable a fair, consistent comparison in which performance differences are attributable to architecture rather than to the loss formulation.

Pre-Processing

The original 16-bit grayscale images were converted to 8-bit format with intensity values rescaled to [0, 255]. High-resolution (HR) images were resized to 224 x 224 pixels using Lanczos interpolation to preserve structure, and low-resolution (LR) images of 56 x 56 pixels were produced by bicubic downsampling (scale factor 4). Images were converted to RGB format for compatibility across all tested architectures. Data augmentation (random horizontal flipping and 90 deg, 180 deg, 270 deg rotations) was applied only within the training set, so that no examples were leaked between the training and test sets.

SRGAN (Super-Resolution GAN)

SRGAN is a GAN-based approach focused on perceptual quality rather than pixel accuracy. It uses a deep residual generator (SRResNet) with skip connections and pixel-shuffle upsampling to reconstruct HR images from LR inputs, while a discriminator is trained jointly to distinguish real from generated images. Training combines a VGG-based perceptual loss with an adversarial loss, producing sharper outputs than regression methods, though training instability and batch-normalisation artefacts remain limitations.

ESRGAN (Enhanced SRGAN)

ESRGAN improves SRGAN by redesigning both architecture and training. It replaces residual blocks with Residual-in-Residual Dense Blocks (RRDB) for stronger feature extraction, removes batch normalisation to avoid artefacts, and introduces a relativistic discriminator to better guide adversarial learning. These changes improve texture detail, training stability, and overall visual fidelity over SRGAN.

HAT (Hybrid Attention Transformer)

HAT shifts from CNN-GAN methods to transformer-based modelling. It employs a residual-in-residual structure with hybrid attention blocks that combine spatial self-attention and channel attention and uses overlapping cross-attention to enable interaction across image regions for improved global context. Unlike GAN-based methods, HAT relies only on a pixel-wise loss, achieving high-quality reconstruction with strong stability and long-range dependency learning.

Proposed Domain-Aware Loss (USELoss)

The main contribution of this work is USELoss (U-Net Sobel Edge Loss), a loss framework designed for the medical-imaging domain. Unlike perceptual losses derived from networks trained on natural images, USELoss incorporates anatomical-structure and boundary constraints that are better aligned with medical imaging. To ensure a fair comparison between architectures, identical loss components and weights are used for both ESRGAN and HAT variants; the adversarial term is included only for the GAN-based ESRGAN.

For ESRGAN

$$L_{\text{total}} = \lambda_1 L_{\text{pixel}} + \lambda_2 L_{\text{unet}} + \lambda_3 L_{\text{edge}} + \lambda_4 L_{\text{adv}}$$

For HAT

$$L_{\text{total}} = \lambda_1 L_{\text{pixel}} + \lambda_2 L_{\text{unet}} + \lambda_3 L_{\text{edge}}$$

The weighting scheme $(\lambda_1, \lambda_2, \lambda_3, \lambda_4) = (0.5, 1.0, 0.5, 0.005)$ is identical for all models, so any performance difference is due to architecture rather than the loss formulation.

Pixel Loss (L_{pixel}). Measures the difference between corresponding pixels in the reconstructed and target images. By reducing pixel-level inconsistency, the model preserves anatomical detail, maintains intensity consistency, and improves reconstruction accuracy.

U-Net-Based Perceptual Loss (L_{unet}). Standard perceptual loss in ESRGAN depends on VGG networks trained on natural images, which limits their usefulness for medical imaging. L_{unet} replaces VGG-based feature loss with a lightweight U-Net encoder whose multi-scale representations emphasise anatomical structure and boundary information, resolving the domain mismatch and keeping reconstructed structure faithful rather than texture-driven.

Edge Preservation Loss (L_{edge}). Accurate boundary representation is essential for interpreting medical images. L_{edge} uses a Sobel-based edge loss to enforce gradient consistency between super-resolved and reference images, preserving fine anatomical edges while reducing artificial sharpening. The Sobel kernels are:

$$\mathbf{G}_x = \begin{array}{|c|c|c|} \hline -1 & 0 & 1 \\ \hline -2 & 0 & 2 \\ \hline -1 & 0 & 1 \\ \hline \end{array}$$
$$\mathbf{G}_y = \begin{array}{|c|c|c|} \hline -1 & -2 & -1 \\ \hline 0 & 0 & 0 \\ \hline 1 & 2 & 1 \\ \hline \end{array}$$

Adversarial Loss (L_{adv}). Adversarial training can create structures that appear realistic but are anatomically implausible. To limit this risk, ESRGAN retains the adversarial loss at a much lower weight, improving perceived clarity without increasing the likelihood of hallucinated textures. HAT omits the adversarial loss, since its self-attention mechanism already captures long-range structural dependencies and global context.

Training

All models were trained on an NVIDIA L4 GPU (24 GB VRAM) using CUDA-enabled acceleration and the Adam optimiser ($\beta_1 = 0.9$, $\beta_2 = 0.99$) with an initial learning rate of 2×10^{-4} . ESRGAN used a two-stage strategy of pixel-based pre-training followed by full USELoss fine-tuning, whereas HAT was trained directly with the full USELoss formulation. ESRGAN-based models were trained for 200 epochs with the learning rate halved at scheduled milestones and a batch size of 32. HAT-based models were trained for multiple iterations with a cosine-annealing learning-rate schedule and a batch size of 8, owing to the higher memory demands of transformer architectures. HAT configuration parameters (window size, embedding dimension, attention heads, and depth) were kept identical to the original HAT formulation so that the effect of the proposed loss could be isolated. LR input patches of 64×64 pixels were upsampled to 256×256 HR patches, and all models were trained under identical pre-processing and evaluation conditions.

Evaluation and Statistical Analysis

Model performance was assessed with a broad set of quantitative metrics spanning pixel fidelity, perceptual quality, structural similarity, and edge preservation. Reconstruction quality was measured with Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), and Multi-Scale SSIM (MS-SSIM). Pixel-wise reconstruction error was measured with Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE).

Perceptual quality was assessed with Learned Perceptual Image Patch Similarity (LPIPS), Visual Information Fidelity (VIF), and Feature Similarity Index Measure (FSIM); LPIPS measures perceptual distance in a learned feature space (lower is better), while FSIM measures feature-level consistency. Additional global similarity measures included Normalised Cross-Correlation (NCC), Universal Quality Index (UQI), and Signal-to-Noise Ratio (SNR). Boundary preservation was evaluated using Sobel-based edge-difference metrics, namely the total edge-magnitude difference (edge sum) and the mean per-pixel edge discrepancy (edge mean). All metrics are reported as mean $\pm$ standard deviation across the held-out test set.

Reproducibility

The study uses the publicly available PadChest dataset [2], and the pre-processing pipeline, data splits, training schedules, loss weights, and evaluation metrics are specified in full above to support replication of the reported experiments.

Results

Table 2 reports the main image-quality metrics for all evaluated models. Standard ESRGAN shows a clear drop in performance, with a PSNR of only 22.21 dB and an SSIM of 0.5545, indicating that it does not generalise to medical images when trained with a standard VGG-based perceptual loss. In contrast, SRGAN, HAT, and all USELoss-based variants exceed 33 dB PSNR with consistently high structural similarity. Adding USELoss to ESRGAN raises PSNR from 22.21 dB to 34.04 dB and SSIM from 0.5545 to 0.9160 - an improvement of more than 50% in PSNR - underscoring the importance of domain-aware

perceptual supervision. HAT_USELoss achieves the highest SSIM (0.9215) among transformer-based methods, while baseline HAT achieves the highest MS-SSIM, indicating strong cross-scale consistency.

Table 2
Reconstruction quality across all models (PSNR, SSIM, MS-SSIM; mean $\pm$ standard deviation).

Model	PSNR (dB) $\uparrow$	SSIM $\uparrow$	MS-SSIM $\uparrow$
SRGAN	33.71 $\pm$ 2.34	0.9014 $\pm$ 0.0262	0.9831 $\pm$ 0.0055
ESRGAN	22.21 $\pm$ 0.75	0.5545 $\pm$ 0.0599	0.9147 $\pm$ 0.0113
ESRGAN_USELoss	34.04 $\pm$ 2.39	0.9160 $\pm$ 0.0238	0.9846 $\pm$ 0.0051
HAT	33.97 $\pm$ 2.86	0.9202 $\pm$ 0.0243	0.9864 $\pm$ 0.0046
HAT_USELoss	33.96 $\pm$ 2.81	0.9215 $\pm$ 0.0243	0.9857 $\pm$ 0.0049

Table 3 summarises pixel-wise reconstruction errors. Standard ESRGAN produces much larger errors (MSE > 396, RMSE 19.84) than any other method, confirming that adversarial learning based on a natural-image perceptual loss is unstable in the medical domain. HAT achieves the lowest MAE (3.03), indicating high average pixel-level accuracy, whereas ESRGAN_USELoss achieves the lowest MSE and RMSE, indicating better robustness to outliers and greater overall stability. These results show that USELoss effectively regularises both GAN-based and transformer-based models toward anatomically accurate reconstructions.

Table 3
Pixel-wise reconstruction errors (MAE, MSE, RMSE; mean $\pm$ standard deviation).

Model	MAE $\downarrow$	MSE $\downarrow$	RMSE $\downarrow$
SRGAN	3.40 $\pm$ 0.88	32.37 $\pm$ 21.66	5.46 $\pm$ 1.59
ESRGAN	14.36 $\pm$ 1.64	396.47 $\pm$ 71.08	19.84 $\pm$ 1.74
ESRGAN_USELoss	3.17 $\pm$ 0.85	30.16 $\pm$ 20.77	5.26 $\pm$ 1.56
HAT	3.03 $\pm$ 0.90	32.78 $\pm$ 25.77	5.40 $\pm$ 1.90
HAT_USELoss	3.08 $\pm$ 0.90	32.62 $\pm$ 25.22	5.40 $\pm$ 1.87

Table 4 reports perceptual-quality metrics. Standard ESRGAN shows substantial perceptual degradation, with the highest LPIPS (0.4102) and the lowest VIF and FSIM, consistent with the visual artefacts observed in its reconstructions. Both ESRGAN_USELoss and HAT_USELoss show large perceptual improvements: HAT_USELoss achieves the lowest LPIPS (0.1607) and the highest FSIM (0.9434). USELoss improves every perceptual metric, demonstrating that it is an effective, domain-appropriate replacement for natural-image perceptual supervision.

Table 4
Perceptual quality (LPIPS, VIF, FSIM; mean $\pm$ standard deviation).

Model	LPIPS $\downarrow$	VIF $\uparrow$	FSIM $\uparrow$
SRGAN	0.1830 $\pm$ 0.0306	0.5849 $\pm$ 0.0497	0.9297 $\pm$ 0.0145
ESRGAN	0.4102 $\pm$ 0.0400	0.2619 $\pm$ 0.0428	0.7620 $\pm$ 0.0287
ESRGAN_USELoss	0.1645 $\pm$ 0.0279	0.6270 $\pm$ 0.0486	0.9371 $\pm$ 0.0135
HAT	0.1658 $\pm$ 0.0278	0.6408 $\pm$ 0.0536	0.9373 $\pm$ 0.0158
HAT_USELoss	0.1607 $\pm$ 0.0270	0.6348 $\pm$ 0.0523	0.9434 $\pm$ 0.0140

Table 5 summarises additional global similarity measures. ESRGAN_USELoss achieves the highest NCC, UQI, and SNR, indicating strong global correspondence and signal fidelity, whereas standard ESRGAN drops clearly on all three, again confirming its unsuitability for medical image super-resolution without domain adaptation. All USELoss-based methods perform comparably, indicating that the proposed loss generalises well across architectures.

Table 5
Global similarity (NCC, UQI, SNR; mean $\pm$ standard deviation).

Model	NCC $\uparrow$	UQI $\uparrow$	SNR (dB) $\uparrow$
SRGAN	0.9954 $\pm$ 0.0012	0.9954 $\pm$ 0.0012	28.85 $\pm$ 2.34
ESRGAN	0.9653 $\pm$ 0.0136	0.9616 $\pm$ 0.0140	17.36 $\pm$ 1.52
ESRGAN_USELoss	0.9962 $\pm$ 0.0010	0.9962 $\pm$ 0.0010	29.19 $\pm$ 2.39
HAT	0.9959 $\pm$ 0.0014	0.9959 $\pm$ 0.0014	29.12 $\pm$ 2.86
HAT_USELoss	0.9960 $\pm$ 0.0013	0.9959 $\pm$ 0.0014	29.11 $\pm$ 2.81

Table 6 reports edge-preservation metrics. Standard ESRGAN produces the largest edge discrepancy, indicating excessive false edges and boundary distortion. Adding USELoss to ESRGAN cuts edge error by about 46%, demonstrating the effectiveness of Sobel-based edge supervision. HAT_USELoss attains the lowest edge sum and edge mean of all tested models, preserving fine anatomical boundaries best, and reduces edge error by about 5.5% relative to baseline HAT - showing that explicit edge-aware constraints also benefit transformer-based architectures.

Table 6
Edge preservation via Sobel-based edge discrepancy (mean $\pm$ standard deviation).

Model	Edge Sum ↓	Edge Mean ↓
SRGAN	790,426 $\pm$ 223,833	15.75 $\pm$ 4.46
ESRGAN	1,342,383 $\pm$ 259,756	26.75 $\pm$ 5.18
ESRGAN_USELoss	721,645 $\pm$ 205,396	14.38 $\pm$ 4.09
HAT	741,839 $\pm$ 219,860	14.78 $\pm$ 4.38
HAT_USELoss	700,866 $\pm$ 209,175	13.97 $\pm$ 4.17

Table 7 reports the best-performing model for each metric. ESRGAN_USELoss is best on most fidelity and error-based metrics, while HAT_USELoss is best on perceptual and edge-preservation metrics. This indicates that GAN-based and transformer-based methods derive distinct advantages from USELoss, yet both achieve balanced, clinically meaningful reconstructions.

Table 7
Best model per evaluation metric, showing trade-offs between reconstruction accuracy, perceptual quality, and boundary preservation.

Metric	Best Model	Value
PSNR	ESRGAN_USELoss	34.04 dB
SSIM	HAT_USELoss	0.9215
MS-SSIM	HAT	0.9864
LPIPS	HAT_USELoss	0.1607
MAE	HAT	3.03
MSE	ESRGAN_USELoss	30.16
RMSE	ESRGAN_USELoss	5.26
NCC	ESRGAN_USELoss	0.9962
UQI	ESRGAN_USELoss	0.9962
SNR	ESRGAN_USELoss	29.19 dB
VIF	HAT	0.6408
FSIM	HAT_USELoss	0.9434
Edge Sum	HAT_USELoss	700,866
Edge Mean	HAT_USELoss	13.97

Figure 1 compares PSNR, SSIM, and MS-SSIM across all models, clearly showing the failure of standard ESRGAN and the consistent gains of USELoss-based variants. Figure 2 presents a radar chart of normalised metrics, in which ESRGAN_USELoss and HAT_USELoss exhibit the most balanced performance profiles. The visual comparisons in Fig. 3 show that standard ESRGAN produces large artefacts and colour shifts, while SRGAN yields smoother but less detailed reconstructions; HAT preserves large structures but can over-smooth small boundaries, whereas HAT_USELoss produces sharper edges and better anatomical detail without adding false textures. The edge visualisations in Figs. 4 and 5 show that USELoss-based methods suppress false edge generation while maintaining boundary continuity, which is strongly linked to improved diagnostic interpretability of small anatomical structures.

Discussion

The performance degradation of standard ESRGAN (PSNR 22.21 dB versus > 34 dB for the alternatives) highlights a fundamental limitation of VGG-based perceptual loss in medical imaging. VGG19 features, optimised for natural images, prioritise texture and semantic content rather than anatomical structure. This mismatch leads to poor reconstruction fidelity in X-ray imaging, where structural continuity, edge precision, and density variation are critical; consequently ESRGAN exhibits severe structural distortion (SSIM 0.5545), indicating substantial loss of clinically relevant information.

The proposed USELoss framework addresses this limitation by replacing VGG supervision with a U-Net encoder whose inductive biases are better aligned with medical imaging. Its multi-scale feature extraction and skip connections preserve fine anatomical detail and global structure simultaneously. Despite being far lighter (~ 7.5M parameters versus ~ 143M for VGG19), the U-Net encoder provides more relevant structural representations, improving both efficiency and reconstruction quality.

Edge supervision further improves performance by explicitly enforcing boundary accuracy through a Sobel-based loss. This is particularly important in clinical scenarios such as fracture detection or tumour-boundary delineation, where minor edge deviations can affect diagnosis. The combination of L1 loss (intensity fidelity), U-Net perceptual loss (structural coherence), and edge loss (boundary precision) forms a complementary framework that addresses multiple failure modes of conventional loss functions.

The results confirm that loss design, rather than architecture alone, is the primary determinant of performance. ESRGAN improves dramatically when paired with USELoss (PSNR 34.04 dB, SSIM 0.9160), showing that its capacity was previously underutilised, while transformer-based HAT models benefit further on perceptual and edge metrics, achieving the highest SSIM (0.9215) and best LPIPS. This suggests that although advanced architectures contribute to performance, domain-specific supervision is essential for clinically meaningful outputs.

From a practical standpoint, USELoss adds negligible computational overhead, since the U-Net encoder is used only during training and edge computations are lightweight, giving inference efficiency

comparable to baseline models and supporting real-world deployment. Limitations remain, including evaluation on a single dataset (PadChest), reliance on bicubic degradation, and the absence of clinical validation; future work must address robustness across modalities and real-world artefacts.

Conclusion

This study demonstrates that domain-specific loss design is critical for effective medical image super-resolution. Conventional perceptual losses, particularly those based on VGG networks, fail to capture the structural characteristics essential for diagnostic imaging. The proposed USELoss framework integrates U-Net-based perceptual supervision, explicit edge constraints, and controlled adversarial learning to align optimisation objectives with clinical requirements, yielding large gains across reconstruction, perceptual, and edge-preservation metrics for both GAN-based and transformer-based architectures. The ability to reconstruct high-quality images from lower-resolution inputs has important clinical implications, including potential radiation-dose reduction and improved accessibility in resource-constrained settings, and the framework is computationally efficient and compatible with existing imaging pipelines. Future work will extend the approach to diverse datasets and modalities such as CT and MRI and conduct clinical validation through radiologist studies.

Declarations

Author Contribution

Aryan Tyagi: Conceptualization, Methodology, Software, Data Curation, Formal Analysis, Investigation, Validation, Visualization, Writing – Original Draft Preparation, Writing – Review & Editing. Aryan Tyagi developed the proposed USELoss framework, implemented and trained the super-resolution models, performed experiments and statistical analyses, generated figures and tables, interpreted the results, and prepared the initial manuscript draft. Neeraj Sharma: Supervision, Methodology, Validation, Writing – Review & Editing. Neeraj Sharma provided technical guidance on medical image analysis and deep learning methodology, contributed to study design and validation of experimental findings, supervised the research activities, and critically reviewed and revised the manuscript. Prachi Pant: Data Curation, Investigation, Formal Analysis, Visualization, Writing – Review & Editing. Prachi Pant assisted with dataset preparation, preprocessing, experimental evaluation, result interpretation, figure preparation, and manuscript review and editing. Mohini Manav: Clinical Validation, Resources, Supervision, Writing – Review & Editing. Mohini Manav provided clinical expertise regarding radiological relevance and diagnostic considerations, contributed to interpretation of medical imaging results, supervised the clinical aspects of the study, and reviewed and edited the manuscript. All authors read and approved the final manuscript and agree to be accountable for all aspects of the work.

Data Availability

The data that support the findings of this study are publicly available from the PadChest dataset repository: Bustos A. et al., PadChest: A large chest X-ray image dataset with multi-label annotated reports, Medical Image Analysis, 2020. The dataset can be accessed at: <https://bimcv.cipf.es/bimcv-projects/padchest/>. The code used to train and evaluate the models presented in this study is available at: <https://github.com/1aryantyagi/Segmentation-Paper>. No new datasets were generated during the current study.

References

1. Bastos, A. D. L. & Nogueira, M. D. S. Image quality in diagnostic radiology: A guide to methodologies for radiologists. *Radiol. Bras.* **58**, e20240088 (2025).
2. Bustos, A., Pertusa, A., Salinas, J. M., De La Iglesia-Vaya, M. & PadChest A large chest x-ray image dataset with multi-label annotated reports. *Med. Image Anal.* **66**, 101797 (2020).
3. Calli, E., Sogancioglu, E., Van Ginneken, B., Van Leeuwen, K. G. & Murphy, K. Deep learning for chest X-ray analysis: A survey. *Med. Image Anal.* **72**, 102125 (2021).
4. Cao, H. et al. Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation. arXiv; (2021).
5. Chen, J. et al. TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation. arXiv; (2021).
6. Chen, X. et al. HAT: Hybrid Attention Transformer for Image Restoration. arXiv; (2023).
7. Dong, C., Loy, C. C., He, K. & Tang, X. Learning a Deep Convolutional Network for Image Super-Resolution. In: Computer Vision - ECCV 2014. Vol. 8692. Cham: Springer; 184–199. (2014).
8. Dosovitskiy, A. et al. *An Image is Worth 16x16 Words* (Transformers for Image Recognition at Scale. arXiv, 2020).
9. Gonzalez, R. C., Woods, R. E. & Masters, B. R. Digital Image Processing, Third Edition. *J. Biomed. Opt.* **14** (2), 029901 (2009).
10. Goodfellow, I. *NIPS 2016 Tutorial: Generative Adversarial Networks* (arXiv, 2017).
11. Hatamizadeh, A. et al. *UNETR: Transformers for 3D Medical Image Segmentation* (arXiv, 2021).
12. Johnson, J., Alahi, A. & Fei-Fei, L. Perceptual Losses for Real-Time Style Transfer and Super-Resolution. In: Computer Vision - ECCV 2016. Vol. 9906. Cham: Springer; 694–711. (2016).
13. Kim, J., Lee, J. K. & Lee, K. M. *Accurate Image Super-Resolution Using Very Deep Convolutional Networks* (arXiv, 2015).
14. Krupinski, E. A. Current perspectives in medical image perception. *Atten. Percept. Psychophys.* **72** (5), 1205–1217 (2010).
15. Ledig, C. et al. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network. arXiv; (2016).
16. Liang, J. et al. SwinIR: Image Restoration Using Swin Transformer. arXiv; (2021).
17. Lim, B., Son, S., Kim, H., Nah, S. & Lee, K. M. Enhanced Deep Residual Networks for Single Image Super-Resolution. arXiv; (2017).

18. Mabotuwana, T., Bhandarkar, V. S., Hall, C. S. & Gunn, M. L. Detecting Technical Image Quality in Radiology Reports. *AMIA Annu Symp Proc.* ;2018:780–788. (2018).
19. Mahesh, M. Essential Role of a Medical Physicist in the Radiology Department. *RadioGraphics* **38** (6), 1665–1671 (2018).
20. Mendonca, R. P. D. et al. Principles of radiological protection and application of ALARA, ALADA, and ALADAIP: A critical review. *Braz Oral Res.* **39**, e14 (2025).
21. Oakley, P. A. & Harrison, D. E. Death of the ALARA Radiation Protection Principle as Used in the Medical Sector. *Dose Response.* **18** (2), 155932582092164 (2020).
22. Qiu, D., Cheng, Y. & Wang, X. Medical image super-resolution reconstruction algorithms based on deep learning: A survey. *Comput. Methods Programs Biomed.* **238**, 107590 (2023).
23. Shin, M., Seo, M., Lee, K. & Yoon, K. Super-resolution techniques for biomedical applications and challenges. *Biomed. Eng. Lett.* **14** (3), 465–496 (2024).
24. Simonyan, K. & Zisserman, A. *Very Deep Convolutional Networks for Large-Scale Image Recognition* (arXiv, 2014).
25. Strauss, K. J. & Kaste, S. C. The ALARA concept in pediatric interventional and fluoroscopic imaging: a white paper executive summary. *Pediatr. Radiol.* **36** (S2), 110 (2006).
26. Wang, X. et al. ESRGAN: Enhanced Super-Resolution Generative Adversarial Networks. In: *Computer Vision - ECCV 2018 Workshops*. Vol. 11133. Cham: Springer; 63–79. (2019).
27. Wilczynski, N. L. Quality of Reporting of Diagnostic Accuracy Studies: No Change Since STARD Statement Publication - Before-and-after Study. *Radiology* **248** (3), 817–823 (2008).
28. Xiao, H. et al. Deep learning for medical imaging super-resolution: A comprehensive review. *Neurocomputing* **630**, 129667 (2025).
29. Yu, M., Xu, Z. & Lukasiewicz, T. A general survey on medical image super-resolution via deep learning. *Comput. Biol. Med.* **193**, 110345 (2025).
30. Zamir, S. W. et al. Restormer: Efficient Transformer for High-Resolution Image Restoration. arXiv; (2021).
31. Zhang, K. et al. Practical Blind Image Denoising via Swin-Conv-UNet and Data Synthesis. arXiv; (2022).
32. Zhang, Y. et al. Image Super-Resolution Using Very Deep Residual Channel Attention Networks. In: *Computer Vision - ECCV 2018*. Vol. 11211. Cham: Springer; 294–310. (2018).
33. Zhao, H., Gallo, O., Frosio, I. & Kautz, J. Loss Functions for Image Restoration With Neural Networks. *IEEE Trans. Comput. Imaging.* **3** (1), 47–57 (2017).
34. Wang, Z., Bovik, A. C., Sheikh, H. R. & Simoncelli, E. P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **13** (4), 600–612 (2004).

Figures

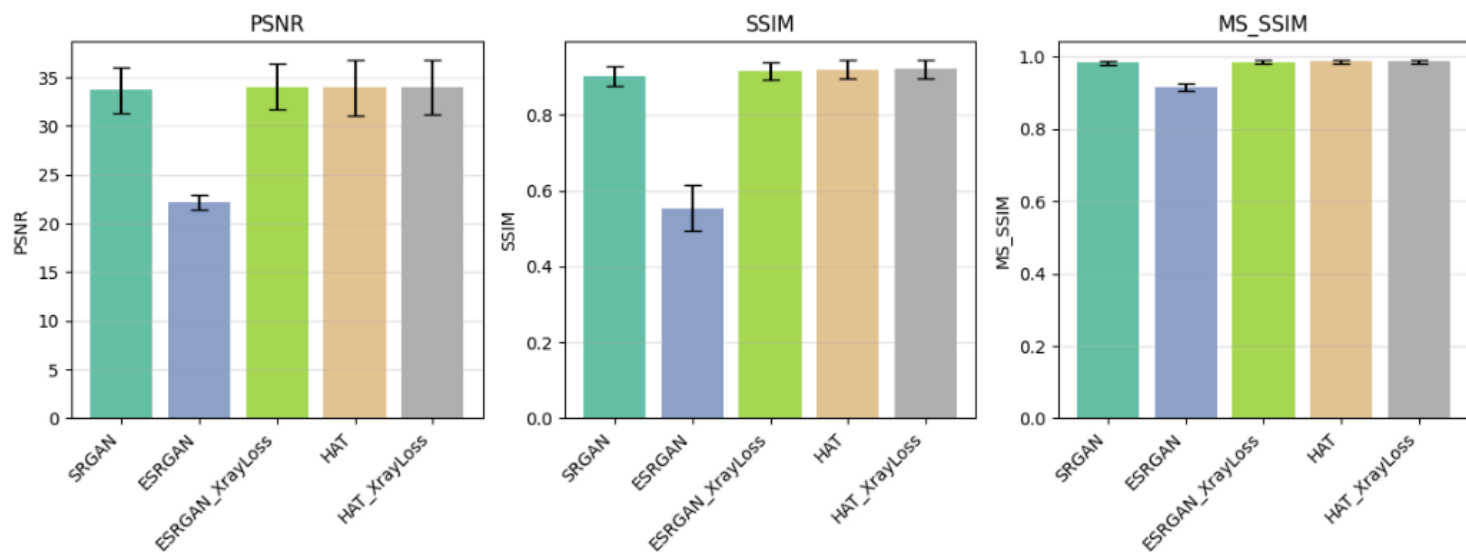


Figure 1

Bar chart comparing PSNR, SSIM, and MS-SSIM across all evaluated models.

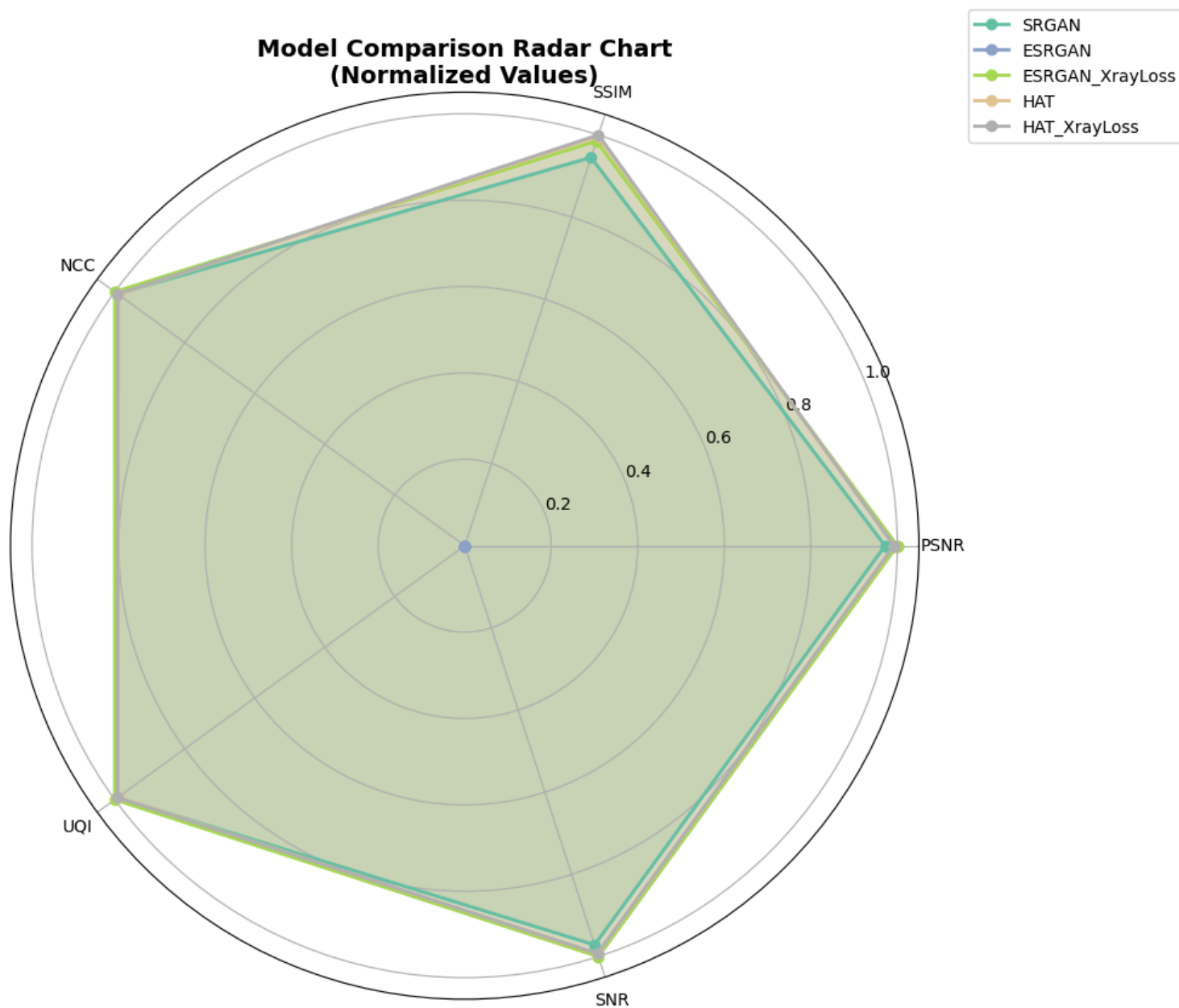


Figure 2

Radar chart comparing normalised performance metrics across all models.

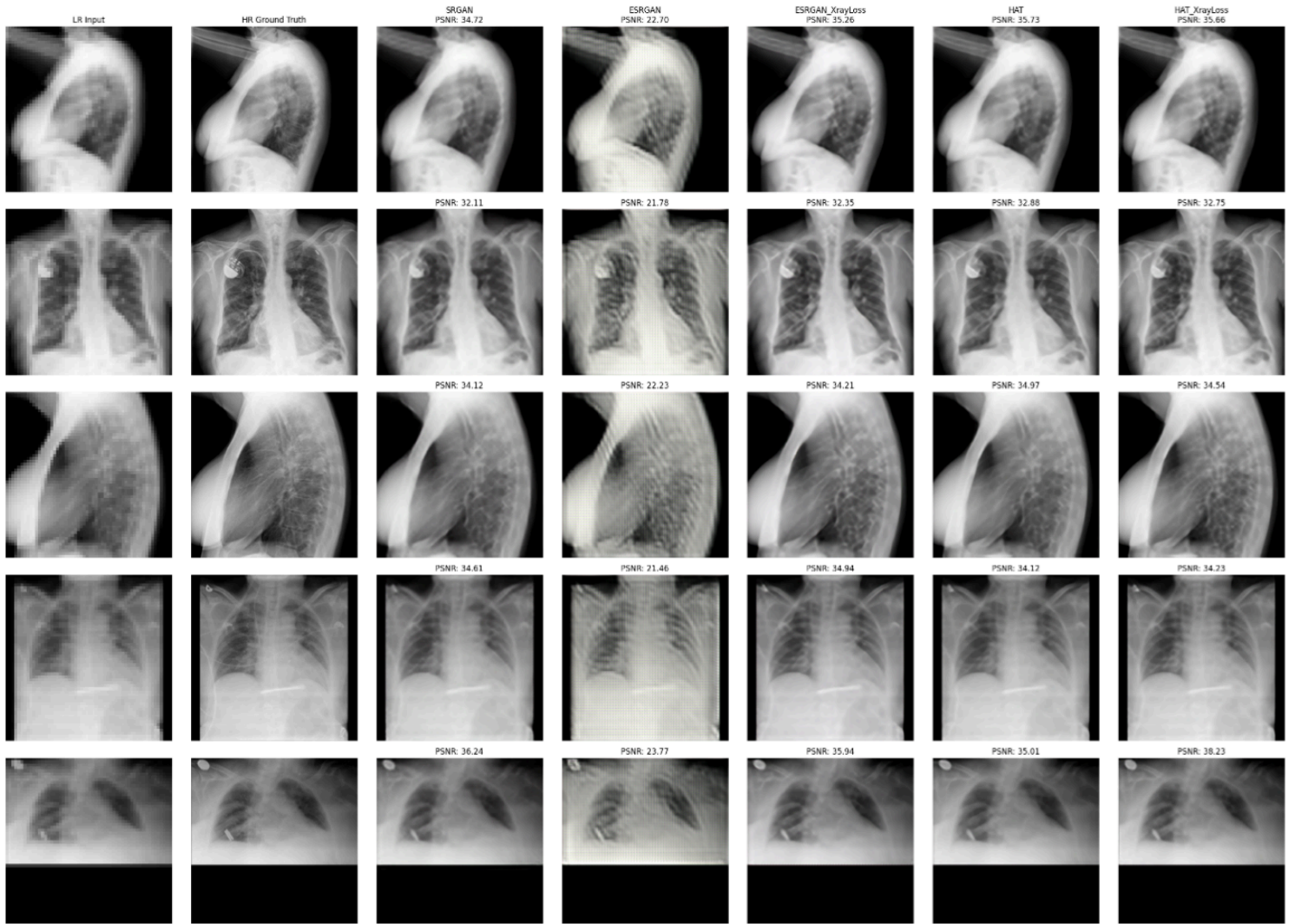


Figure 3

Visual quality comparison of super-resolution reconstructions across all models

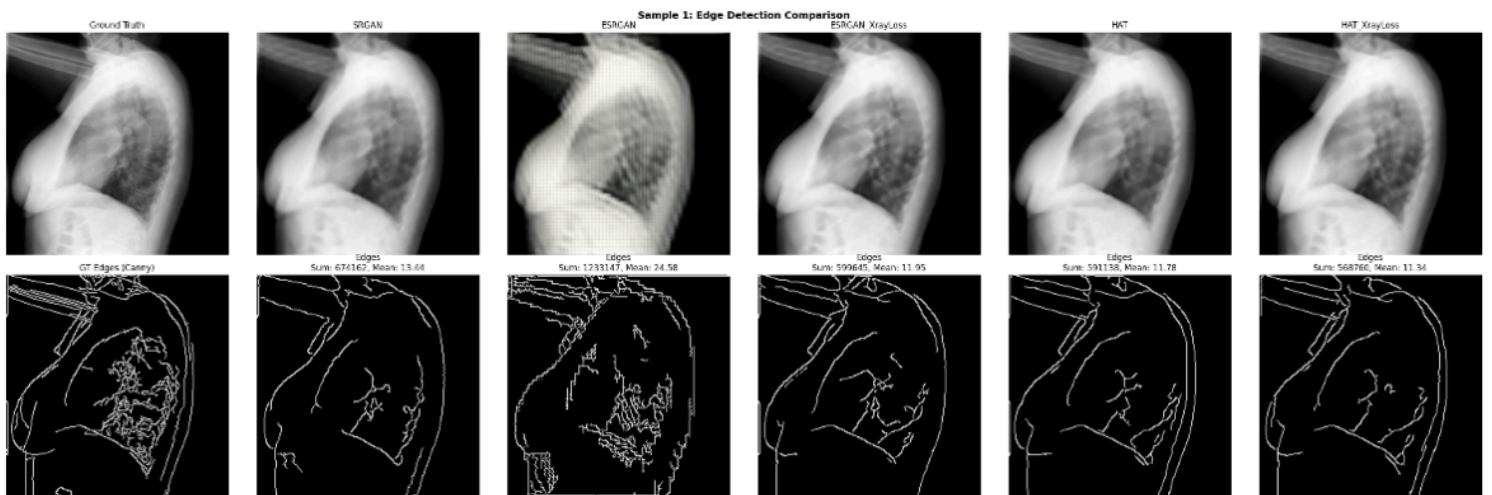


Figure 4

Edge-preservation analysis (Part 1).

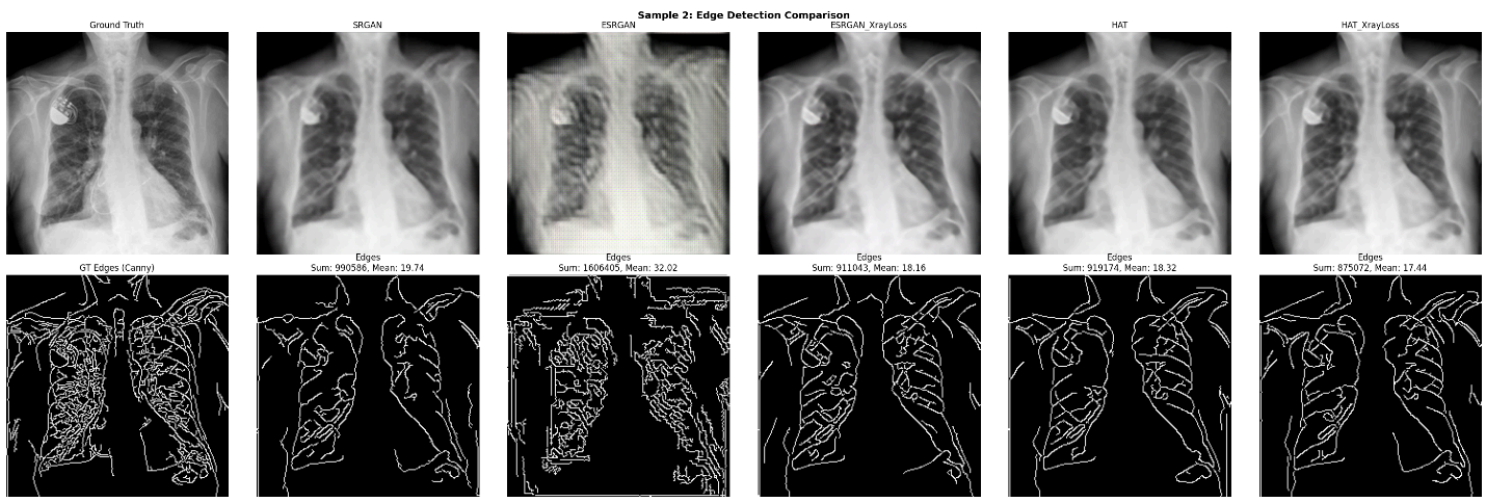


Figure 5

Edge-preservation analysis (Part 2).