

NGPS: Structure-Preserving Self-Supervised Denoising via Neighbor-Guided Patch Sampling

Jaehyun Cho¹  and YoungJoon Yoo^{1,2}  *

¹ Department of Artificial Intelligence, Chung-Ang University, 84 Heukseok-ro, Dongjak-gu, Seoul, Republic of Korea
{q7011805,yjyoo3312}@cau.ac.kr

² SNUAILAB, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Korea
yjyoo3312@snuailab.ai

Abstract. Neighboring-slice self-supervised denoising is attractive for volumetric medical imaging, yet inter-slice misalignment breaks anatomical correspondence and often yields ghosting and blurred margins when adjacent slices are used naïvely as targets. We propose Neighbor-Guided Patch Sampling (NGPS), a lightweight framework that constructs neighboring supervision under local inter-slice misalignment. To avoid learning from misleading targets, prior methods commonly mask discrepant regions, but this stabilizes training at the cost of leaving a non-trivial portion of neighboring evidence unexploited, particularly around high-frequency anatomical boundaries. NGPS addresses this by *decoupling* structure matching from signal retrieval: for each masked location, it searches a local neighborhood for structurally similar candidate patches using a simple guide image (*e.g.*, fast bilateral filtering), while retrieving the supervision signal directly from the *raw noisy* neighbor at the matched coordinates. By matching on a noise-attenuated guide while retrieving raw values from neighboring slices, NGPS constructs local pseudo targets without dense deformation-field estimation or spatial re-sampling. Across the evaluated CT and synthetic-Rician MRI settings, NGPS improves fidelity and structure-sensitive metrics. Code is available at <https://github.com/cv-cho/NGPS>.

Keywords: Self-supervised denoising · Medical image restoration · Low-dose CT · Structure preservation · Self-supervised training

1 Introduction

Medical imaging enables non-invasive diagnosis and longitudinal monitoring [28]. However, X-ray-based modalities such as CT and fluoroscopy expose patients and clinicians to ionizing radiation, motivating low-dose acquisition [19]. Dose reduction induces photon starvation, yielding quantum noise and reconstruction artifacts that obscure fine anatomical structures, especially in high-attenuation regions [1, 6, 24]. This motivates denoising methods that restore structural and signal fidelity while preserving clinically meaningful content [8, 13, 20].

* Corresponding author.

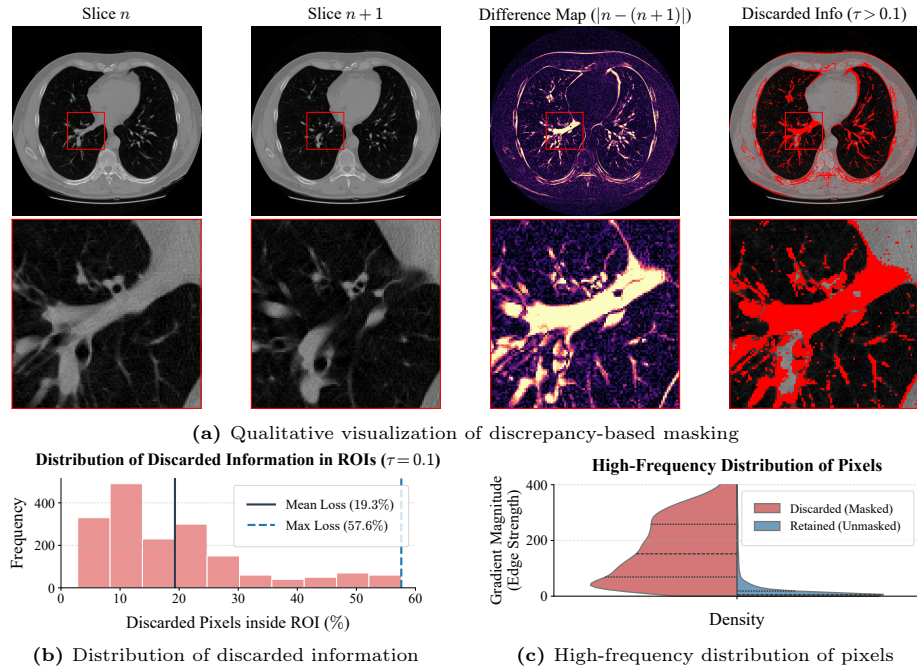


Fig. 1: Illustration of inter-slice anatomical misalignment and the resulting critical information loss from conventional masking strategies ($\tau = 0.1$) in 2.5 mm LIDC-IDRI data [2]. (a) The absolute difference map ($|n - (n + 1)|$) and the masked pixels (highlighted in red). (b) The histogram revealing the severity of this spatial information loss within critical Regions of Interest (ROIs) across the validation set. (c) The split violin plot showing that the discarded (masked) pixels are heavily concentrated in the high-frequency spectrum (high gradient magnitude) compared to the retained pixels.

Prior approaches include classical priors (*e.g.*, NLM, BM3D, TV) [5, 7, 10, 22] and supervised deep networks [27, 39], but paired clean targets are often infeasible and motion or anatomical variability further complicates alignment [17, 24, 34]. Self-supervised learning (SSL) alleviates this constraint by learning directly from noisy observations [3, 16, 18, 35]. For volumetric data, neighboring-slice SSL uses adjacent slices as supervision [25, 36, 41], yet inter-slice misalignment breaks the same-coordinate correspondence, so relevant anatomical evidence is often *displaced* rather than missing. Fig. 1 highlights that discrepancy-based masking can exclude a substantial portion of pixels within clinically relevant ROIs, removing *on average about 20%* of the available neighboring supervision. Moreover, the excluded regions are concentrated around high-gradient anatomical boundaries. This creates a persistent dilemma: masking-based remedies stabilize training by avoiding misleading targets, but systematically withhold supervision in structurally challenging regions. Registration-based alternatives explicitly estimate correspondences and resample neighboring slices. Under severe noise or larger

inter-slice gaps, the estimated alignment can become less reliable, while spatial resampling may smooth edges and add alignment overhead. These trade-offs motivate local target retrieval without dense warping. Therefore, a key challenge is to retrieve displaced anatomical evidence from a local neighborhood rather than treating it as unavailable.

Motivated by this, we propose *Neighbor-Guided Patch Sampling (NGPS)*, a misalignment-aware neighboring-slice SSL framework that is designed to (i) recover displaced supervision within a local neighborhood without dense deformation-field estimation or spatial warping, (ii) remain compatible with Noise2Noise-style [18] training, and (iii) preserve high-frequency anatomical margins with minimal overhead. NGPS reframes misalignment handling as a *training-time supervision construction* problem, rather than an inference-time non-local denoising prior. Specifically, NGPS constructs a neighboring pixel bank by matching pre-filtered guide patches across adjacent slices and retrieving target values from the corresponding raw noisy neighbors. The primary contributions are:

- **Misalignment as displaced supervision:** We identify a structural supervision gap in neighboring-slice SSL caused by displacement, clarifying the masking versus registration trade-off.
- **Neighbor-Guided Patch Sampling (NGPS):** We introduce a lightweight patch-based supervision recovery mechanism using guide-feature search and raw-target retrieval for misalignment-aware SSL.
- **Comprehensive validation:** We demonstrate consistent gains across low-dose CT and MRI benchmarks (AAPM-Mayo [24], LIDC-IDRI [2], IXI [4]), with improved preservation of fine anatomical margins.

2 Related Work

Self-supervised Learning for Denoising. To overcome the difficulty of obtaining paired noisy-clean images [24, 34], self-supervised learning (SSL) has become a practical alternative. Early image-domain methods, such as Blind-Spot Networks [3, 16], construct supervision without directly observing the target pixel, typically through masking or blind-spot/downsampling designs. While effective in broad settings, these schemes restrict the directly available context and can introduce smoothing or checkerboard artifacts [33]. Recent image-domain SSL methods broaden target construction through masking, downsampling, noise injection, or generative pseudo targets. DiffDenoise [11], for example, uses a conditional-diffusion pipeline to generate pseudo-clean supervision. These methods differ in how supervision is synthesized or withheld. In the neighboring-slice setting studied here, Fig. 1 shows that discrepancy masking removes supervision concentrated near high-gradient anatomical boundaries.

Volumetric SSL and Misalignment Handling. Volumetric medical SSL can exploit redundancy along different axes. Patch2Self [12] exploits diffusion-MRI q-space redundancy through held-out-volume regression. Our setting instead uses spatial redundancy across adjacent anatomical slices, which provide alternative

noisy observations under the standard slice-independent-noise assumption. This neighboring-slice formulation implicitly assumes anatomical correspondence; in practice, inter-slice displacement can make same-coordinate supervision blur shifted structures. Existing remedies largely follow registration- or masking-based strategies. Registration-based methods such as MSR2AU-Net [15] and Deformed2Self [36] estimate correspondences and warp neighboring slices. Such alignment can become less reliable under severe degradation or larger inter-slice gaps, and spatial resampling may alter local signal and noise statistics. Masking-based methods such as Noise2Sim [25] and NS-N2N [41] instead exclude discrepant pixels from the reconstruction loss. This avoids misleading same-coordinate targets but can reduce supervision near high-gradient boundaries, as quantified in Fig. 1. NGPS addresses this displaced-supervision regime through local retrieval rather than dense warping or exclusion.

Patch-based Sampling and Neighbor-Aware Processing. Patch-based self-similarity provides a complementary direction for constructing supervision without paired targets. Pixel2Pixel (P2P) [21] constructs a “PixelBank” by searching for similar patches within a single noisy image and sampling replacement targets from non-local neighbors. Extending this idea to volumetric SSL is non-trivial: exhaustive image-wide search can be costly, and same-image target selection can couple the retrieval process to the same noisy realization. Neighboring slices provide an alternative target source under the standard slice-independent-noise assumption, provided displaced correspondence can be resolved. At a broader conceptual level, Graph Flow Matching [29] introduces neighbor-aware aggregation in a flow-based image-generation setting rather than for denoising-target construction. NGPS differs from both settings by applying local guide-based patch search at locations flagged by inter-slice discrepancy and retrieving raw adjacent-slice values as training targets.

3 Methodology

We propose Neighbor-Guided Patch Sampling (NGPS), a misalignment-aware neighboring-slice SSL framework that constructs pseudo targets for misaligned regions via three steps, as illustrated in Fig. 2: (i) fast guide generation with conservative masking, (ii) decoupled structural matching on the guide, and (iii) raw target retrieval with a Top- K ensemble. We then train a 2D slice denoiser f_θ with a hybrid objective that uses standard same-coordinate supervision in static regions (blue in Fig. 2) and switches to NGPS-recovered targets in misaligned regions (purple). Furthermore, we incorporate an output-level regional consistency regularizer to reduce slice-to-slice flicker in non-masked areas.

3.1 Problem Setup and Motivation

Let $x, y \in \mathbb{R}^{H \times W \times D}$ denote a clean volumetric image and its noisy observation, modeled as $y_z(p) = x_z(p) + n_z(p)$, where $z \in \{1, \dots, D\}$ is the slice index and p is

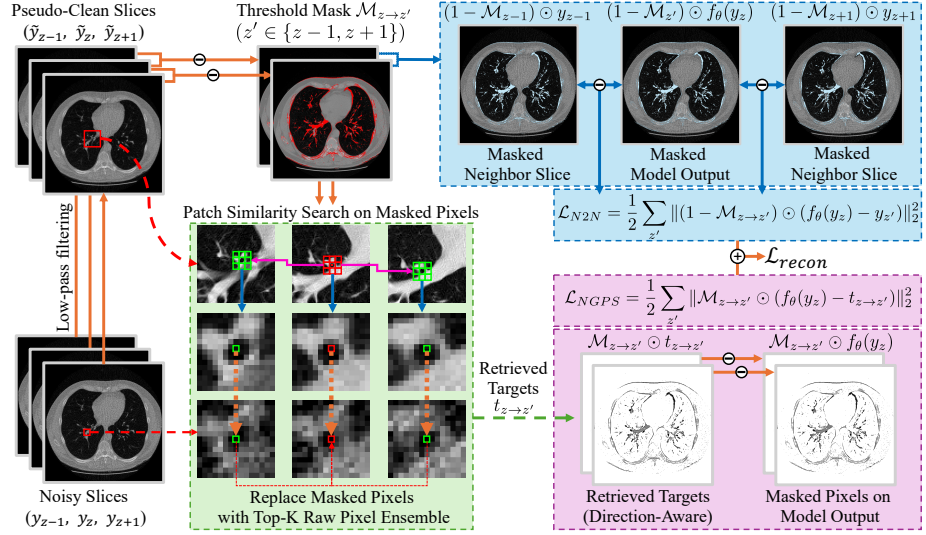


Fig. 2: Overall NGPS pipeline. A noisy slice triplet is low-pass filtered to produce noise-attenuated guide (“Pseudo-Clean”) slices, whose pairwise differences yield direction-aware threshold masks with reduced sensitivity to raw noise. For each masked location, NGPS searches a local window in the adjacent slice for the Top- K similar patches (green); the matched coordinates are then used to retrieve and average the corresponding center-pixel values from the original noisy adjacent slice, forming direction-aware retrieved targets. Unmasked static pixels use same-coordinate neighboring supervision to compute $\mathcal{L}_{N2N}$ (blue), whereas masked pixels use the retrieved targets to compute $\mathcal{L}_{NGPS}$ (purple). The two reconstruction terms are summed as $\mathcal{L}_{recon}$; regional consistency is applied separately to static regions as described in Sec. 3.4.

the spatial coordinate. Assuming zero-mean ($\mathbb{E}[n_z(p)] = 0$) and slice-independent noise, neighboring-slice SSL learns a denoiser using adjacent slices as pseudo targets [18]. Here, slice independence concerns noise across the input and target slices; it does not require pixel-wise i.i.d. noise within each slice. However, inter-slice misalignment breaks this same-coordinate correspondence. Under an inter-slice displacement δ , a local anatomical structure in the clean signal $x_z(p)$ appears shifted in the adjacent slice as $x_{z'}(p) \approx x_z(p - \delta)$. Naive same-coordinate supervision can encourage averaging between these shifted structures, yielding a superposition $\frac{1}{2}(x_z(p) + x_z(p - \delta))$ that can produce ghosting and blurred anatomical boundaries.

3.2 From Misalignment to Design Principles

To overcome this dilemma, NGPS reframes misalignment as *local displacement* rather than absence. Instead of discarding misaligned regions, we actively retrieve the corresponding anatomical evidence from a nearby coordinate q in the adjacent slice. This perspective establishes three core design requirements for

our framework: (i) **Recover displaced supervision** via explicit local retrieval rather than passive exclusion, (ii) **Retain raw neighboring values as training targets** while using filtered guides only for correspondence search, and (iii) **Remain lightweight** without dense deformation-field estimation or spatial re-sampling.

We use *selection bias* to denote noise-driven coordinate selection that induces correlation between input noise and the residual noise of a retrieved target. NGPS performs discrete local correspondence search. Unlike dense registration, it estimates no dense deformation field and performs no spatial warping or interpolation of neighboring slices.

3.3 Neighbor-Guided Patch Sampling (NGPS)

For each pixel flagged as misaligned, NGPS constructs a pseudo target $t_{z \rightarrow z'}(p)$ via (i) fast guide generation with conservative masking, (ii) decoupled patch matching on the guide, and (iii) raw-value retrieval with Top- K aggregation.

Guide generation and misalignment mask. We generate an edge-preserving guide volume $\tilde{y}$ by sequentially applying a 2D bilateral filter (BF) and a median filter (MF) to suppress quantum noise while preserving structural boundaries:

$$\tilde{y}_z(p) = \text{MF}\left(\text{BF}(y_z(p); \sigma_s, \sigma_r); \kappa\right), \quad (1)$$

where σ_s and σ_r are the spatial and range parameters of the BF, and κ is the kernel size of the MF. Let $\mathcal{N}(z) = \{z-1, z+1\}$ denote the set of available neighboring slice indices (applying reflection padding only at the volume ends). Instead of a single global mask, we define a direction-aware misalignment mask for each neighbor $z' \in \mathcal{N}(z)$ based on the same-coordinate guide disagreement:

$$\mathcal{M}_{z \rightarrow z'}(p) = \mathbb{K}\left(|\tilde{y}_z(p) - \tilde{y}_{z'}(p)| > \tau\right), \quad (2)$$

where τ is used to flag locations whose same-coordinate guide discrepancy exceeds the selected threshold. For pixels not flagged by $\mathcal{M}_{z \rightarrow z'}$, we use standard same-coordinate neighboring supervision. For flagged pixels, we avoid same-coordinate targets and invoke local NGPS retrieval.

Decoupled matching and raw retrieval. For each pixel $p \in \mathbb{Z}^2$ flagged as misaligned with respect to a specific neighboring slice $z' \in \mathcal{N}(z)$, we search for structurally similar candidate patches within a local spatial window Ω_p (e.g., 15×15) in slice z' . While the similarity search is performed at the patch level to ensure structural context, the supervision signal is retrieved purely as a single scalar value from the center pixel of the matched location. Let $\mathcal{P}_k(I, p) \in \mathbb{R}^{k \times k}$ extract a $k \times k$ patch centered at p from image I . We compute the guide-based sum of squared differences (SSD) costs:

$$\mathcal{D}(p, q; z') = \|\mathcal{P}_k(\tilde{y}_z, p) - \mathcal{P}_k(\tilde{y}_{z'}, q)\|_2^2, \quad q \in \Omega_p, \quad (3)$$

and subsequently retrieve the candidate target scalar from the *raw* center pixel, $y_{z'}(q) \in \mathbb{R}$. This separates correspondence search from regression-target retrieval, avoiding direct raw-patch matching while keeping the target value unfiltered.

The guide is computed from noisy observations and therefore does not make the selected coordinates strictly independent of noise. We accordingly treat guide-based matching as an empirical design choice intended to reduce noise-driven target selection, while raw-value retrieval keeps the regression target unfiltered.

Top- K ensemble target. To reduce the variance inherent to a single best match, we aggregate the Top- K candidate pixels for each neighboring direction z' . For a given neighbor z' , we rank the candidate coordinates in Ω_p and select the set $\{q^{(1)}, \dots, q^{(K)}\}$ yielding the smallest $\mathcal{D}(p, q; z')$. We define the direction-aware target $t_{z \rightarrow z'}(p)$ as the average of these K retrieved center pixels:

$$t_{z \rightarrow z'}(p) = \frac{1}{K} \sum_{k=1}^K y_{z'}(q^{(k)}). \quad (4)$$

This is inspired by the PixelBank philosophy [21] but specialized to volumetric SSL; the sampling is applied directionally (only for misaligned pixels) and the scalar targets are drawn exclusively from the adjacent slice z' to construct $t_{z \rightarrow z'}$.

3.4 Training objective

We optimize f_θ using a hybrid reconstruction loss and regional consistency term:

$$\mathcal{L}_{total} = \mathcal{L}_{recon} + \lambda \mathcal{L}_{RC}, \quad (5)$$

where λ is a hyperparameter that controls the strength of the regularization.

Hybrid reconstruction loss Guided by the directional misalignment mask $\mathcal{M}_{z \rightarrow z'}$, we adaptively switch the supervision signal. Specifically, we apply standard same-coordinate neighboring targets for static regions, and provide the retrieved NGPS targets $t_{z \rightarrow z'}$ for misaligned regions. The hybrid reconstruction loss is thus formulated as:

$$\mathcal{L}_{recon} = \mathcal{L}_{N2N} + \mathcal{L}_{NGPS}, \quad (6)$$

$$\mathcal{L}_{N2N} = \frac{1}{|\mathcal{N}(z)|} \sum_{z' \in \mathcal{N}(z)} \frac{\sum_p (1 - \mathcal{M}_{z \rightarrow z'}(p)) (f_\theta(y_z)(p) - y_{z'}(p))^2}{\sum_p (1 - \mathcal{M}_{z \rightarrow z'}(p)) + \epsilon}, \quad (7)$$

$$\mathcal{L}_{NGPS} = \frac{1}{|\mathcal{N}(z)|} \sum_{z' \in \mathcal{N}(z)} \frac{\sum_p \mathcal{M}_{z \rightarrow z'}(p) (f_\theta(y_z)(p) - t_{z \rightarrow z'}(p))^2}{\sum_p \mathcal{M}_{z \rightarrow z'}(p) + \epsilon}, \quad (8)$$

where $\mathcal{M}_{z \rightarrow z'}$ and $t_{z \rightarrow z'}$ are spatial maps matching the dimensions of the prediction $f_\theta(y_z)$, and $\epsilon > 0$ is a small numerical constant preventing division by zero. Each directional loss is normalized by the number of pixels assigned to its corresponding supervision region.

Table 1: Summary of datasets and noise simulation protocols used in our experiments.

Dataset	Modality	Anatomy	Spacing	Noise Simulation Type	Train/Test
AAPM-Mayo [24]	CT	Abdomen	1.0 mm	Realistic Quarter-Dose (Poisson injection in sinogram)	8 / 2
LIDC-IDRI [2]	CT	Thorax	1.25, 2.5 mm	Simulated Ultra-Low-Dose (Radon Transform, $P = 12.5K$)	66 / 6
IXI-T1 [4]	MRI	Brain	1.2 mm	Synthetic Rician ($\sigma \in \{5\%, 7\%, 9\%\}$)	52 / 5

Regional consistency regularization Equation 7 optimizes each slice independently and does not explicitly enforce volumetric coherence. In regions identified as static, adjacent clean signals are expected to be locally similar rather than identical. Following volumetric SSL practice [41], we therefore penalize prediction differences only on $(1 - \mathcal{M}_{z \rightarrow z'})$, encouraging inter-slice consistency without constraining regions flagged as displaced:

$$\mathcal{L}_{RC} = \frac{1}{|\mathcal{N}(z)|} \sum_{z' \in \mathcal{N}(z)} \|(1 - \mathcal{M}_{z \rightarrow z'}) \odot (f_{\theta}(y_z) - f_{\theta}(y_{z'}))\|_2^2. \quad (9)$$

4 Experiments

4.1 Experimental Setup

Datasets and Noise Simulation. To evaluate the proposed NGPS, we use three public medical volumetric datasets. The detailed configurations for each dataset are summarized in Table 1.

AAPM-Mayo (CT) [24]: As a realistic benchmark for low-dose CT (LDCT) denoising, we use abdominal CT scans from the AAPM Low-Dose CT Grand Challenge. Unlike simple additive noise, the Quarter-Dose (QD) images were generated by injecting Poisson noise directly into the *projection data* (sinograms) of Normal-Dose scans to simulate 25% of the full dose. This dataset serves as a gold standard for evaluating robustness against realistic, spatially correlated CT noise textures. For our experiments, we utilize the 1.0 mm slice reconstructions.

LIDC-IDRI (CT) [2]: To evaluate robustness under *ultra-low-dose* conditions with severe streak artifacts, we specifically curated a subset of thoracic CT scans acquired with GE Medical Systems scanners. The selected data consists of high-resolution and standard scans with slice thicknesses of exactly 1.25 mm and 2.5 mm. We simulate ULD projections via the Radon Transform with a photon count of 12,500, followed by Filtered Back Projection (FBP). This generates realistic streak artifacts and non-stationary noise distributions that pose significant challenges to conventional SSL methods. Unless otherwise noted, aggregate LIDC-IDRI results pool the 1.25-mm and 2.5-mm test subsets; spacing-specific results are reported separately.

IXI-T1 (MRI) [4]: For a controlled cross-modality evaluation, we use brain scans from the IXI dataset with a slice thickness of 1.2 mm. We add synthetic Rician noise at three intensity levels ($\sigma \in \{5\%, 7\%, 9\%\}$) to the clean volumetric data.

Table 2: Quantitative evaluation on the AAPM-Mayo [24] (Quarter-Dose) and LIDC-IDRI [2] (Ultra-Low-Dose) datasets. The best results are highlighted in **bold**, and the second-best are underlined. $\uparrow$ indicates higher is better, while $\downarrow$ indicates lower is better.

Method	AAPM-Mayo (Quarter-Dose CT)					LIDC-IDRI (Simulated ULD CT)				
	PSNR $\uparrow$	SSIM $\uparrow$	FSIM $\uparrow$	HFEN $\downarrow$	GMSD $\downarrow$	PSNR $\uparrow$	SSIM $\uparrow$	FSIM $\uparrow$	HFEN $\downarrow$	GMSD $\downarrow$
Baseline (Noisy)	30.30	0.7222	0.8772	0.3363	0.0874	22.14	0.4213	0.5920	0.8429	0.1959
BM3D [10]	34.67	0.7764	0.8547	0.3258	0.0900	25.07	0.5825	0.7768	0.6010	0.1449
DIP [31]	35.36	0.7895	0.8804	0.2766	0.0426	26.77	0.6235	0.8074	0.5910	0.1337
NAC [37]	34.61	0.8171	0.9062	0.2344	0.0664	24.83	0.5690	0.7500	0.6096	0.1439
ZS-N2N [23]	33.68	0.8129	0.9107	0.2911	0.0631	26.63	0.5979	0.8080	0.5545	0.1259
Pixel2Pixel [21]	33.71	0.8357	0.9117	0.3183	0.0739	25.58	0.5078	0.7277	0.6196	0.1237
Noise2Void [16]	32.57	0.7786	0.8977	0.3286	0.0739	26.10	0.5995	0.8295	0.4948	0.1078
NB2NB [14]	33.21	0.7922	0.9037	0.3187	0.0715	27.23	0.6063	0.8321	0.4687	0.1003
Filter2Noise [30]	35.23	0.8334	0.9265	0.2895	0.0625	28.81	0.7181	0.8716	0.5188	0.1036
Deformed2Self [36]	35.85	<u>0.8662</u>	0.9144	<u>0.2299</u>	<u>0.0377</u>	28.94	0.6703	0.8170	0.5134	0.0808
Noise2Sim [25]	35.49	0.8639	<u>0.9420</u>	0.2406	0.0390	28.81	0.7817	0.8749	0.5003	0.1040
NS-N2N [41]	<u>35.91</u>	0.8584	0.9235	0.2325	0.0396	<u>30.62</u>	<u>0.8080</u>	<u>0.8944</u>	<u>0.4406</u>	0.0777
Ours	36.68	0.8986	0.9470	0.2056	0.0362	31.03	0.8102	0.9168	0.4161	<u>0.0788</u>

Evaluation Metrics. We employ five quantitative metrics to assess different aspects of image quality. Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM) [32] are used to measure general signal fidelity. To evaluate perceptual quality and low-level feature preservation, we utilize Feature Similarity (FSIM) [40]. Furthermore, given the importance of preserving fine anatomical details in medical imaging, we employ High-Frequency Error Norm (HFEN) [26] and Gradient Magnitude Similarity Deviation (GMSD) [38] to specifically measure the restoration quality of edges and textures (lower values indicate better performance for HFEN and GMSD).

Implementation Details. Our method is implemented in PyTorch and evaluated on a single CPU/GPU workstation (AMD Ryzen 9 9950X, NVIDIA RTX 5090). Baselines with public implementations follow released/recommended settings; NS-N2N [41] is reimplemented from the paper using the same NAFNet [9] backbone as NGPS. NGPS uses fixed default hyperparameters across datasets (patch size $p=7$, window size $W=15$, Top- K $K=4$, masking threshold $\tau=0.05$), and we train with AdamW (2×10^{-4} learning rate, 10^{-5} weight decay) for 10 epochs with batch size 4, using $\lambda=0.5$. Additional NGPS architecture and training details are provided in Supplementary Sec. S1.

4.2 Quantitative Comparison

Results on Low-Dose and Ultra-Low-Dose CT Datasets. Table 2 compares quarter-dose AAPM-Mayo [24] and simulated-ULD LIDC-IDRI [2]. NGPS leads all five AAPM metrics. On LIDC-IDRI, it improves PSNR, FSIM, and HFEN over NS-N2N [41] by 0.41 dB, 0.0224, and 0.0245, respectively, whereas the SSIM margin is small (0.0022) and NS-N2N has a slightly lower GMSD

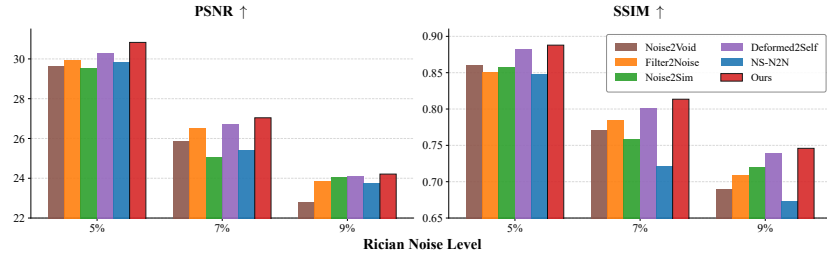


Fig. 3: Quantitative comparison on IXI dataset [4] (Simulated Rician noise). We add different levels of Rician noise (5%, 7%, and 9%) for corruption.

(0.0777 vs. 0.0788). Volume-level paired 95% confidence intervals exclude zero for PSNR and FSIM, but include zero for SSIM, HFEN, and GMSD (Supplementary Table S3). We therefore interpret the LIDC-IDRI result as improved fidelity and boundary-sensitive restoration rather than uniform dominance across all metrics. In the evaluated ULD and thicker-slice settings, Deformed2Self [36] shows larger degradation, while masking-based methods omit discrepant regions from their reconstruction losses. NGPS instead retrieves local supervision in these regions without dense warping.

Results on the IXI MRI Dataset with Rician Noise. We evaluate robustness on the IXI dataset [4] under increasing synthetic Rician noise (5%, 7%, and 9%), summarizing performance trends in Fig. 3 (full values in Supplementary Table S1). All methods degrade as the corruption level increases, but NGPS maintains the highest PSNR and SSIM across the tested levels. Among the baselines, Deformed2Self [36] ranks second overall, which is consistent with the usefulness of inter-slice context in this setting. The relative gaps among Deformed2Self, Noise2Sim [25], NS-N2N [41], and the single-image baselines vary across corruption levels. Nevertheless, NGPS remains highest in both metrics throughout the tested range, demonstrating consistent performance under the evaluated synthetic Rician corruption.

4.3 Qualitative Comparison

Fig. 4 compares quarter-dose AAPM-Mayo [24] and ULD LIDC-IDRI [2] CTs. The baselines exhibit clear visual trade-offs in the highlighted ROIs: single-image methods such as BM3D [10], Noisy-As-Clean [37], Filter2Noise [30], Pixel2Pixel [21], and ZS-N2N [23] either retain more residual noise or produce softer anatomical details. The warping-based baseline Deformed2Self [36] shows local artifacts in some regions, consistent with the difficulty of estimating correspondence under strong degradation. Masking-based SSL methods such as Noise2Sim [25] and NS-N2N [41] produce smoother boundaries in the displayed ROIs, which is consistent with omitting discrepant pixels from the reconstruction loss. In these

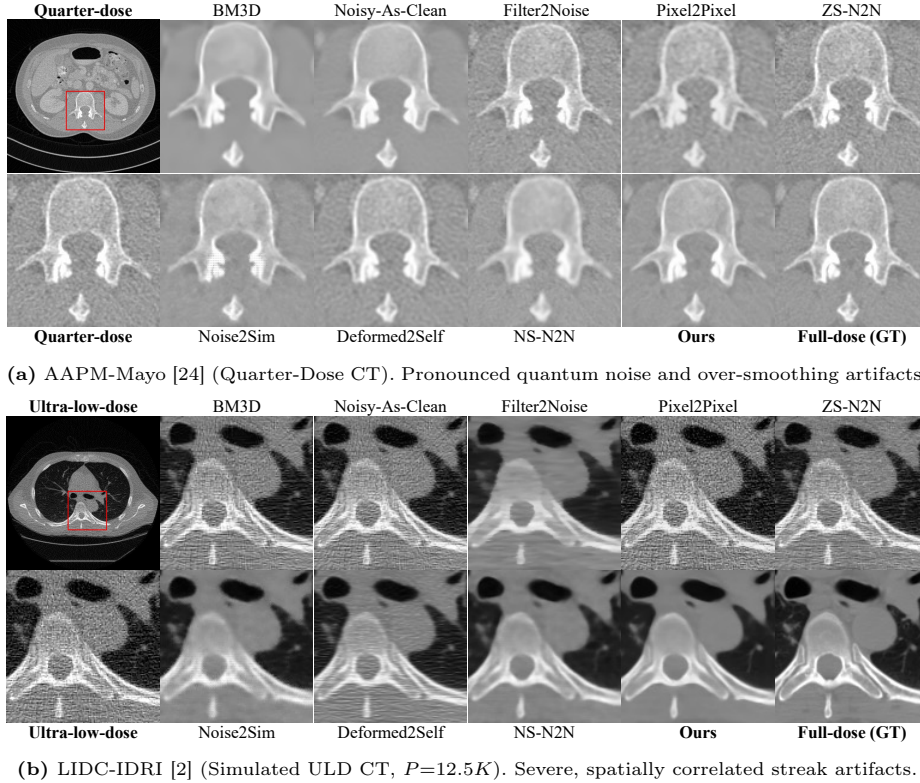


Fig. 4: Qualitative comparisons on quarter-dose and ultra-low-dose CT. The top-left shows the noisy input with a red ROI, while the rest display magnified ROIs.

examples, NGPS reduces residual corruption while retaining sharper anatomical margins through local raw-target retrieval.

Under 9% synthetic Rician noise on IXI MRI [4] (Fig. 5), similar tendencies are visible. Single-image methods retain more residual corruption or over-smoothing, the warping-based baseline shows local degradation under the tested corruption, and masking-based outputs show softer cortical boundaries in the highlighted area. In the highlighted example, NGPS retains sharper local structures while reducing residual corruption, consistent with the quantitative trends.

4.4 Discussion

Target-construction and Objective Ablations.

PixelBank-style counterfactual. Table 3 compares NGPS with two simplified PixelBank-style [21] target-construction variants. The same-slice variant retrieves targets from the center slice, following the intra-image PixelBank intuition. This

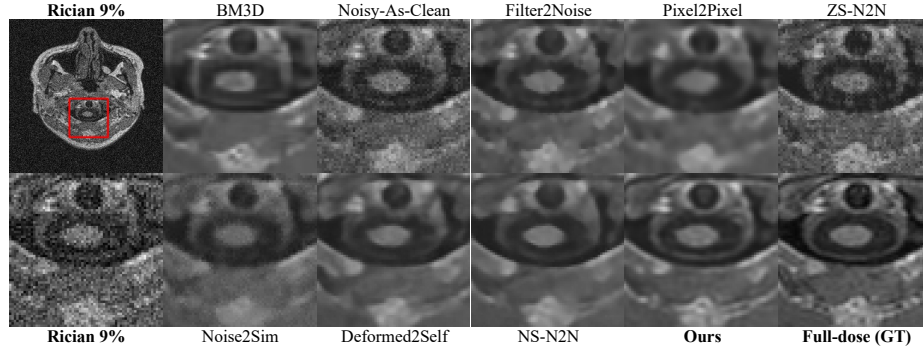


Fig. 5: Qualitative results on IXI [4] under 9% synthetic Rician noise. In the highlighted region, several baselines show more residual corruption or softer cortical boundaries. The warping- and masking-based baselines exhibit local structure loss in this example, whereas NGPS retains sharper anatomical details.

Table 3: PixelBank-style counterfactual training results. Entries are PSNR / SSIM.

Method	AAPM	LIDC 1.25 mm	LIDC 2.5 mm
Same-slice PixelBank	35.16 / 0.842	30.23 / 0.786	29.94 / 0.733
Adjacent-slice raw-patch bank	36.09 / 0.860	30.67 / 0.792	30.18 / 0.760
NGPS	36.68 / 0.899	31.08 / 0.810	30.92 / 0.799

tests whether intra-image patch retrieval alone can provide a sufficient self-supervised target. The adjacent-slice raw-patch bank instead moves the target source to a neighboring slice, but performs patch matching directly on raw noisy patches rather than on the noise-attenuated guide. This tests whether a direct cross-slice extension of raw PixelBank-style search is sufficient.

NGPS outperforms the adjacent-slice raw-patch bank by 0.41–0.74 dB and also outperforms the same-slice variant on all three settings. These comparisons indicate that neither intra-image retrieval nor direct raw cross-slice patch search reproduces the performance of the complete NGPS target-construction pipeline.

Loss-component ablation. Table 4 evaluates the contribution of reconstruction, regional consistency (RC), and inter-slice continuity (IC) on AAPM-Mayo [24]. In NS-N2N [41], discrepant pixels are removed from the masked reconstruction loss; therefore, IC can act as an auxiliary constraint by encouraging local inter-slice continuity in regions weakly supervised by the masked reconstruction term. With reconstruction alone, NS-N2N obtains 35.43 dB, while adding both RC and IC improves it to 35.91 dB.

RC improves NGPS from 36.50 to 36.68 dB, while adding IC to the RC-enabled NGPS objective decreases PSNR from 36.68 to 36.58 dB. In contrast, NS-N2N obtains its best result when both RC and IC are used. The detailed definition and motivation of $\mathcal{L}_{IC}$ are provided in Supplementary Sec. S4.1.

Table 4: Loss-component ablation on AAPM-Mayo [24]. We compare reconstruction, regional consistency (RC), and inter-slice continuity (IC).

Method	Recon.	RC ($\lambda = 0.5$)	IC ($\lambda = 1.0$)	PSNR (dB) $\uparrow$	SSIM $\uparrow$
NS-N2N [41]	✓			35.43	0.8481
	✓	✓		35.88	0.8578
	✓		✓	35.44	0.8473
	✓	✓	✓	35.91	0.8584
NGPS	✓			36.50	0.8894
	✓	✓		36.68	0.8986
	✓		✓	36.39	0.8938
	✓	✓	✓	36.58	0.8981

Table 5: Sensitivity to slice thickness on LIDC-IDRI [2]. PSNR drop is defined as $\text{PSNR}_{1.25\text{ mm}} - \text{PSNR}_{2.5\text{ mm}}$; lower is better, and the best results are highlighted in bold.

Method	1.25 mm (Thin Slice)		2.5 mm (Thick Slice)		PSNR drop (dB) $\downarrow$
	PSNR $\uparrow$	SSIM $\uparrow$	PSNR $\uparrow$	SSIM $\uparrow$	
Noise2Sim [25]	30.53	0.7963	27.70	0.7738	2.83
Deformed2Self [36]	30.52	0.7553	28.64	0.6550	1.88
NS-N2N [41]	30.83	0.8092	30.20	0.7877	0.63
Ours (NGPS)	31.08	0.8103	30.92	0.7994	0.16

Robustness to Slice Thickness and Masking Threshold. Table 5 isolates spacing changes using independently trained 1.25 mm and 2.5 mm models. NGPS decreases by 0.16 dB, compared with 0.63 dB for NS-N2N [41], 1.88 dB for Deformed2Self [36], and 2.83 dB for Noise2Sim [25]. The threshold sweep in Fig. 6a further shows that NGPS is comparatively stable across τ , whereas NS-N2N degrades under both strict and loose masking. These results support robustness within the evaluated spacing and threshold ranges; controlled larger-gap failures and confidence-based rejection are analyzed in Supplementary Sec. S4.4.

Multi-geometry Search-window Selection. We estimate match-offset CDFs using a 41×41 reference search across AAPM [24] 1.0 mm, IXI [4] 1.2 mm, and LIDC-IDRI [2] 1.25/2.5 mm, with $p=7$ and $K=4$. A 15×15 window covers over 80% of the selected offsets for every geometry and 98% on LIDC-IDRI 1.25 mm (Fig. 6b). Because larger windows add quadratic candidate cost with limited coverage gain, we use $W=15$ as a common default. Supplementary Table S5 shows that $W=15$ is within 0.02 dB of the best 1.25 mm result and performs best at 2.5 mm; we therefore use it as a practical common default rather than a universal optimum.

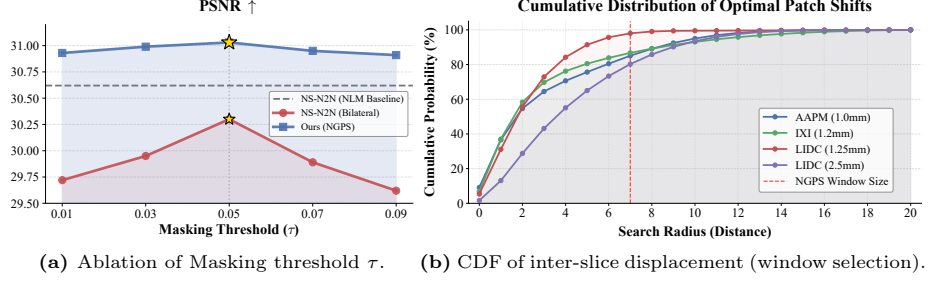


Fig. 6: Ablations for mask sensitivity and search-window design. Left: PSNR trends under varying τ . Right: displacement CDF used to set the NGPS search radius.

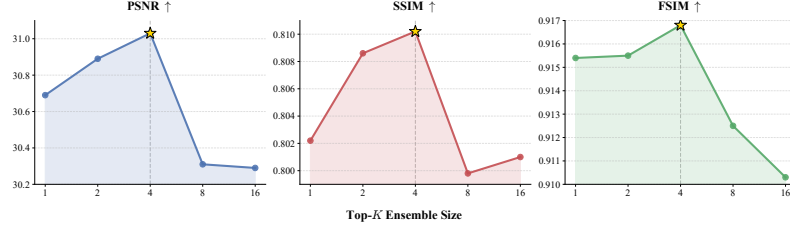


Fig. 7: Ablation study on the Top- K ensemble size in the NGPS module.

Table 6: Comparison of preprocessing time overhead for misalignment handling. Time is measured for a standard medical volume consisting of 100 slices (512×512 resolution).

Method	Target Preparation Strategy	Compute Device	Time (per 100 slices) ↓
Pixel2Pixel [21]	Patch-based Pixel Bank Creation	GPU	~5.38 s
Noise2Sim [25]	Mean Filter Masking	GPU	~0.005 s
NS-N2N [41]	NLM Denoising + Median Masking	CPU + GPU	~13.36 s
Ours (NGPS)	LPF Masking + Vectorized Patch Search	CPU + GPU	~0.72 s

Effect of Top- K Ensemble Size. In NGPS, K controls the number of retrieved candidate patches used to form the Top- K ensemble target. Fig. 7 shows a clear trade-off on the LIDC dataset [2]: performance improves from small ensembles to a moderate K , then declines when K becomes too large. The best results are achieved at $K = 4$ (PSNR 31.03 dB, SSIM 0.8102). With $K = 1$ or 2, the ensemble is too small to sufficiently reduce residual noise. With larger K (8 or 16), lower-quality or misaligned matches are increasingly included, which blurs fine structures. Thus, $K = 4$ provides the best tested noise-detail balance.

Computational Efficiency. Table 6 compares target-preparation cost for a 100-slice volume. NGPS requires approximately 0.72 s, about $19\times$ faster than Pixel2Pixel [21] and $7.5\times$ faster than Noise2Sim [25]; Noise2Sim [25] is faster but attains lower restoration metrics in Table 2. The efficiency of NGPS comes from restricting vectorized patch search to masked pixels and using the lightweight bilateral-plus-median guide.

5 Conclusion

We presented NGPS, a lightweight framework for constructing neighboring-slice supervision under inter-slice misalignment. NGPS forms noise-attenuated guides and direction-specific discrepancy masks, performs local Top- K patch search at flagged locations, and retrieves the corresponding raw adjacent-slice values as training targets. Its hybrid objective combines same-coordinate neighboring supervision at unflagged locations with NGPS-retrieved targets at flagged locations, together with regional consistency, without dense warping or a learnable alignment module. Across a realistic quarter-dose CT benchmark, simulated ultra-low-dose CT, and synthetic-Rician MRI, NGPS consistently improves fidelity and structure-sensitive restoration: it leads all AAPM [24] metrics, improves PSNR, FSIM, and HFEN over NS-N2N [41] on simulated ULD CT, and achieves the highest PSNR and SSIM across the evaluated MRI noise levels. These results support local supervision retrieval as an effective and practical strategy for structure-preserving volumetric self-supervised denoising.

Limitations and Future Work. NGPS assumes that a locally corresponding structure with sufficiently consistent appearance exists within the search window. This assumption can fail under large through-plane gaps, abrupt anatomical changes, staining variation, signal dropout, or other non-geometric inter-slice appearance changes. The optional match-cost gate mitigates degradation in the tested large-gap setting, but does not remove this assumption. The Noise2Noise [18] formulation does not require pixel-wise i.i.d. noise within each slice, but assumes zero-mean noise and negligible cross-slice correlation between input- and target-slice noise; cross-slice correlated artifacts remain outside the current validation scope. The fixed p , W , K , and τ values are practical defaults rather than universal optima. Finally, the IXI [4] experiments provide a controlled synthetic-noise evaluation; future work should prioritize acquisition-realistic low-field MRI, noisy 3D microscopy, and downstream clinical validation.

Acknowledgements

This work was supported by the SNUAILAB, the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) [RS-2021-II211341, Artificial Intelligence Graduate School Program (Chung-Ang University) and RS-2022-II220124, Development of Artificial Intelligence Technology for Self-Improving Competency-Aware Learning Capabilities]. Also, this research was supported by the "Regional Innovation System & Education (RISE)" through the Seoul RISE Center, funded by the Ministry of Education (MOE) and the Seoul Metropolitan Government. (2026-RISE-01-024-04)

References

1. Andreozzi, E., Pirozzi, M.A., Fratini, A., Cesarelli, G., Cesarelli, M., Bifulco, P.: A novel image quality assessment index for edgeaware noise reduction in low-dose fluoroscopy: Preliminary results. In: *Int. Conf. e-Health Bioeng. (EHB)*. pp. 1–5 (2020)
2. Armato III, S.G., McLennan, G., Bidaut, L., McNitt-Gray, M.F., Meyer, C.R., Reeves, A.P., Clarke, L.P.: The lung image database consortium (lidc) and image database resource initiative (idri): A completed reference database of lung nodules on ct scans. *Medical Physics* **38**(2), 915–931 (2011)
3. Batson, J., Royer, L.: Noise2self: Blind denoising by self-supervision. In: *ICML*. pp. 524–533 (2019)
4. Biomedical Image Analysis Group: IXI dataset. <https://brain-development.org/ixi-dataset/> (2018), accessed 25 Jun 2026
5. Buades, A., Coll, B., Morel, J.M.: A non-local algorithm for image denoising. In: *CVPR*. vol. 2, pp. 60–65 (2005)
6. Cesarelli, M., Bifulco, P., Cerciello, T., Romano, M., Paura, L.: X-ray fluoroscopy noise modeling for filter design. *Int. J. Comput. Assist. Radiol. Surg.* **8**(2), 269–278 (2013)
7. Chambolle, A.: An algorithm for total variation minimization and applications. *J. Math. Imaging Vis.* **20**(1), 89–97 (2004)
8. Chen, H., Zhang, Y., Kalra, M.K., Lin, F., Chen, Y., Liao, P., Wang, G.: Low-dose ct with a residual encoder-decoder convolutional neural network. *IEEE Trans. Med. Imaging* **36**(12), 2524–2535 (2017)
9. Chen, L., Chu, X., Zhang, X., Sun, J.: Simple baselines for image restoration. In: *European Conference on Computer Vision*. pp. 17–33. Springer (2022)
10. Dabov, K., Foi, A., Katkovnik, V., Egiazarian, K.: Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE TIP* **16**(8), 2080–2095 (2007)
11. Demir, B., Liu, Y., Chen, X., Chen, E.Z., Zhao, L., Mailhe, B., Sun, S.: Diffdenoise: self-supervised medical image denoising with conditional diffusion models. *arXiv preprint arXiv:2504.00264* (2025)
12. Fadnavis, S., Batson, J., Garyfallidis, E.: Patch2self: Denoising diffusion mri with self-supervised learning. *Advances in neural information processing systems* **33**, 16293–16303 (2020)
13. Fourati, W., Kammoun, F., Bouhlef, M.S.: Medical image denoising using wavelet thresholding. *J. Test. Eval.* **33**(5), 364–369 (2005)
14. Huang, T., Li, S., Jia, X., Lu, H., Liu, J.: Neighbor2neighbor: Self-supervised denoising from single noisy images. In: *CVPR*. pp. 14781–14790 (2021)
15. Jeon, S.Y., Wang, S., Wang, A.S., Gold, G.E., Choi, J.H.: Unsupervised training of a dynamic context-aware deep denoising framework for low-dose fluoroscopic imaging. *IEEE Trans. Instrum. Meas.* (2025)
16. Krull, A., Buchholz, T.O., Jug, F.: Noise2void-learning denoising from single noisy images. In: *CVPR*. pp. 2129–2137 (2019)
17. Lee, M.S., Park, S.W., Lee, S.Y., Kang, M.G.: Motion-adaptive 3d nonlocal means filter based on stochastic distance for low-dose x-ray fluoroscopy. *Biomedical Signal Processing and Control* **38**, 74–85 (2017)
18. Lehtinen, J., Munkberg, J., Hasselgren, J., Laine, S., Karras, T., Aittala, M., Aila, T.: Noise2Noise: Learning image restoration without clean data. In: *ICML*. pp. 2971–2980 (2018)

19. Li, M., Hsu, W., Xie, X., Cong, J., Gao, W.: Sacnn: Self-attention convolutional neural network for low-dose ct denoising with self-supervised perceptual loss network. *IEEE Trans. Med. Imaging* **39**(7), 2289–2301 (2020)
20. Luo, Y., Majoe, S., Kui, J., Qi, H., Pushparajah, K., Rhode, K.: Ultra-dense denoising network: application to cardiac catheter-based x-ray procedures. *IEEE Trans. Biomed. Eng.* **68**(9), 2626–2636 (2020)
21. Ma, Q., Jiang, J., Zhou, X., Liang, P., Liu, X., Ma, J.: Pixel2Pixel: A pixelwise approach for zero-shot single image denoising. *IEEE TPAMI* **47**(6), 4614–4629 (2025)
22. Manjon, J.V., Carbonell-Caballero, J., Lull, J.J., Garcia-Marti, G., Marti-Bonmati, L., Robles, M.: Mri denoising using non-local means. *Med. Image Anal.* **12**(4), 514–523 (2008)
23. Mansour, Y., Heckel, R.: Zero-Shot Noise2Noise: Efficient image denoising without any data. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. pp. 14018–14027 (2023)
24. McCollough, C.H., Bartley, A.C., Carter, R.E., Chen, B., Drees, T.A., Edwards, P., Fletcher, J.G.: Low-dose ct for the detection and classification of metastatic liver lesions: Results of the 2016 low dose ct grand challenge. *Medical Physics* **44**(10), e339–e352 (2017)
25. Niu, C., Li, M., Fan, F., Wu, W., Guo, X., Lyu, Q., Wang, G.: Noise suppression with similarity-based self-supervised deep learning. *IEEE Trans. Med. Imaging* **42**(6), 1590–1602 (2023)
26. Ravishankar, S., Bresler, Y.: Mr image reconstruction from highly undersampled k-space data by dictionary learning. *IEEE Trans. Med. Imaging* **30**(5), 1028–1041 (2011)
27. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: *MICCAI*. pp. 234–241 (2015)
28. Sagheer, S.V.M., George, S.N.: A review on medical image denoising algorithms. *Biomedical Signal Processing and Control* **61**, 102036 (2020)
29. Siddiqui, M.S.R., Eliasof, M., Haber, E.: Graph flow matching: Enhancing image generation with neighbor-aware flow fields. In: *Proceedings of the AAAI Conference on Artificial Intelligence*. vol. 40, pp. 25463–25471 (March 2026)
30. Sun, Y., Schneider, L.S., Mei, S., Wang, J., Hu, G., Gu, M., Ye, C., Wagner, F., Song, L., Bayer, S., Maier, A.: Filter2Noise: Interpretable self-supervised single-image denoising for low-dose ct with attention-guided bilateral filtering. *arXiv preprint arXiv:2504.13519* (2025)
31. Ulyanov, D., Vedaldi, A., Lempitsky, V.: Deep image prior. In: *CVPR*. pp. 9446–9454 (2018)
32. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: From error visibility to structural similarity. *IEEE TIP* **13**(4), 600–612 (2004)
33. Wang, Z., Liu, J., Li, G., Han, H.: Blind2unblind: Self-supervised image denoising with visible blind spots. In: *CVPR*. pp. 2027–2036 (2022)
34. Weigert, M., Schmidt, U., Boothe, T., Müller, A., Dibrov, A., Jain, A., Wilhelm, B., Schmidt, D., Broaddus, C., Culley, S., Rocha-Martins, M., Segovia-Miranda, F., Norden, C., Henriques, R., Zerial, M., Solimena, M., Rink, J., Tomancak, P., Royer, L., Jug, F., Myers, E.W.: Content-aware image restoration: pushing the limits of fluorescence microscopy. *Nature Methods* **15**(12), 1090–1097 (2018)
35. Xie, Y., Wang, Z., Ji, S.: Noise2same: Optimizing a self-supervised bound for image denoising. In: *Advances in Neural Information Processing Systems (NeurIPS)*. vol. 33, pp. 20320–20330 (2020)

36. Xu, J., Adalsteinsson, E.: Deformed2Self: Self-supervised denoising for dynamic medical imaging. In: MICCAI. pp. 25–35 (2021)
37. Xu, J., Huang, Y., Cheng, M.M., Liu, L., Zhu, F., Xu, Z., Shao, L.: Noisy-as-clean: Learning self-supervised denoising from corrupted image. *IEEE TIP* **29**, 9316–9329 (2020)
38. Xue, W., Zhang, L., Mou, X., Bovik, A.C.: Gradient magnitude similarity deviation: A highly efficient perceptual image quality index. *IEEE TIP* **23**(2), 684–695 (2014)
39. Zhang, K., Zuo, W., Chen, Y., Meng, D., Zhang, L.: Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE TIP* **26**(7), 3142–3155 (2017)
40. Zhang, L., Zhang, L., Mou, X., Zhang, D.: FSIM: A feature similarity index for image quality assessment. *IEEE TIP* **20**(8), 2378–2386 (2011)
41. Zhou, L., Zhou, Z., Huang, X., Wang, H., Zhang, X., Li, G.: Neighboring Slice Noise2Noise: Self-supervised medical image denoising from single noisy image volume. *arXiv preprint arXiv:2411.10831* (2024)