arXiv:2508.02967v1 [cs.CV] 5 Aug 2025

# Towards Robust Image Denoising with Scale Equivariance

**Dawei Zhang, Xiaojie Guo***

Tianjin University

## Abstract

Despite notable advances in image denoising, existing models often struggle to generalize beyond in-distribution noise patterns, particularly when confronted with out-of-distribution (OOD) conditions characterized by spatially variant noise. This generalization gap remains a fundamental yet underexplored challenge. In this work, we investigate *scale equivariance* as a core inductive bias for improving OOD robustness. We argue that incorporating scale-equivariant structures enables models to better adapt from training on spatially uniform noise to inference on spatially non-uniform degradations. Building on this insight, we propose a robust blind denoising framework equipped with two key components: a Heterogeneous Normalization Module (HNM) and an Interactive Gating Module (IGM). HNM stabilizes feature distributions and dynamically corrects features under varying noise intensities, while IGM facilitates effective information modulation via gated interactions between signal and feature paths. Extensive evaluations demonstrate that our model consistently outperforms state-of-the-art methods on both synthetic and real-world benchmarks, especially under spatially heterogeneous noise. Code will be made publicly available.

## Introduction

The field of image denoising has progressed significantly, evolving from traditional prior-based methods to deep learning frameworks that deliver state-of-the-art performance. However, these advances are largely confined to a supervised learning paradigm, which exposes a fundamental vulnerability, *i.e.*, models trained on paired data with in-distribution (ID) noise often fail to generalize when confronted with out-of-distribution (OOD) noise patterns that are more complex and spatially varying. As illustrated in Fig. 1, performance degrades sharply under such mismatched conditions. This generalization gap is commonly attributed to overfitting, *i.e.*, networks tend to entangle signal representations with the noise characteristics seen during training. Although recent self-supervised and zero-shot methods offer improved flexibility, they typically underperform in terms of denoising quality and remain fragile under true OOD settings. As such, the field faces a critical dilemma: *how to attain the high fidelity of supervised denoising without inheriting its brittleness*.

*Corresponding author.

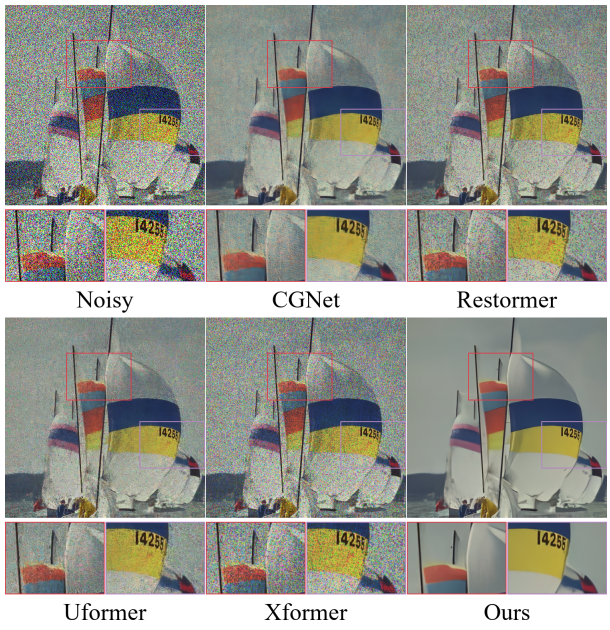



Figure 1: Visual comparison on a sample with OOD noise.

To tackle this issue, existing approaches have explored several directions. One straightforward strategy is to diversify the training noise distribution, using synthesized complex noise such as mixtures (Guo et al. 2019; Zhang et al. 2023b) or adversarial perturbations (Chen et al. 2023). Another line of work focuses on reformulating the learning objective, *e.g.*, by encouraging distortion-invariant representations through MaskedDenoising (Chen et al. 2023) or CLIP-based alignment (Cheng, Liang, and Tan 2024). A third direction leverages noise adaptation, where pretrained models are fine-tuned for specific OOD conditions (Kim, Kim, and Baik 2024; Ha et al. 2024; Chihaoui and Favaro 2025). While effective to some extent, these methods primarily focus on data, objectives, or adaptation protocols, leaving the network architecture itself largely unexamined. In contrast, we argue that OOD robustness should be viewed as an inductive property of the network design rather than merely a byproduct of training strategies.

The core challenge in generalizing from uniform to non-uniform noise lies in the entanglement of features with spatially varying noise level maps. Ideally, a denoising model

should "factor out" the influence of the noise level. However, most existing architectures (see Fig. 1) are ill-equipped to do so, lacking an explicit mechanism to disentangle this interaction. Motivated by this, we revisit a fundamental design principle, namely *scale equivariance*, which is a property that ensures consistent behavior under input scaling and is closely related to first-order homogeneity. We find that enforcing scale equivariance within the network architecture can suppress the impact of spatially varying noise levels, leading to improved generalization and train-test consistency. To this end, we design operations and modules that promote scale-equivariant behavior.

Modern denoising networks often adopt normalization layers to stabilize training. However, conventional normalization schemes (*e.g.*, BatchNorm, LayerNorm) are scale-independent and can inadvertently amplify the entanglement between feature maps and noise levels. We address this by proposing a simple Constant Scaling (CS) operation, which normalizes features by the number of channels, reducing variance without introducing scale distortion. Moreover, pixel-wise OOD noise introduces inconsistent distortions across feature dimensions. To mitigate this issue, we introduce the Normalized Self-Modulator (NSM), which transforms pixel features into a normalized space and modulates them using their original values to maintain scale equivariance. Together, CS and NSM constitute our Heterogeneous Normalization Module (HNM). Another critical component is the activation function. Standard nonlinearities, especially those involving exponential terms, can disrupt scale properties. Apart from common activations, the star operation/gating mechanism (Ma et al. 2024) achieves high-dimensional nonlinear mapping in a low-dimensional space but is typically second-order homogeneous. To retain first-order scale equivariance, we modify and extend the gating design via introducing the Interactive Gating Module (IGM). This module uses a dual-signal scaling strategy, allowing information flow to adapt jointly to both feature and gating signals, thereby enhancing selective filtering under non-uniform noise conditions.

Our major contributions are summarized as follows:

- We identify *scale equivariance* as a key role for mitigating the train-test inconsistency caused by non-uniform noise levels. Guided by this insight, we design a robust blind denoising network tailored for OOD scenarios.
- We propose the Heterogeneous Normalization Module, which integrates feature scaling and normalized self-modulation to strike a stable feature representation and flexible feature correction. Moreover, the Interactive Gating Module is proposed to provide powerful nonlinear representations and improve the feature selection.
- Extensive experiments reveal the superior denoising robustness of our network against OOD noises, especially in high-level noise conditions. Our model consistently outperforms existing approaches with advanced perceptual quality across diverse degradation patterns.

## Related Work

This part will briefly review representative works related to deep learning-based and robust image denoising.

**Deep learning-based image denoising.** Deep learning has become the dominant paradigm in image denoising, with early successes driven by Convolutional Neural Networks (CNNs)(Zamir et al. 2020; Chang et al. 2020; Zhang et al. 2021; Shen et al. 2023; Ghasemabadi et al. 2024), and more recent advances powered by Vision Transformer (ViT) architectures (Wang et al. 2022; Zamir et al. 2022; Li et al. 2023; Zhang et al. 2024; Chen et al. 2024), which leverage sophisticated self-attention mechanisms and normalization strategies. While these designs excel on ID data, their supervised nature makes them prone to overfitting, thus limiting their OOD robustness. To address this, alternative paradigms such as self-supervised learning (Zhang et al. 2023a; Liao et al. 2024; Li, Zhang, and Zuo 2025; Liu et al. 2025) and zero-shot denoising (Quan et al. 2020; Mansour and Heckel 2023; Ma et al. 2025) have emerged. These approaches enable model training directly on unpaired or noisy data without requiring clean targets, for offering greater flexibility. However, they often fall short in denoising quality and still face OOD limitations, as their training remains implicitly tied to the noise distribution in the available data. Thus, a fundamental challenge remains, saying how to retain the high performance of supervised frameworks while overcoming their inherent lack of robustness under OOD noise. This question continues to motivate much of the current research in robust image denoising.

**Robust image denoising.** Robust image denoising refers to a model's ability to maintain high performance when exposed to unseen or OOD noise patterns. Existing efforts to enhance such robustness can be broadly categorized as follows. One common approach is to enrich the training data by simulating complex noise, such as via mixture noise modeling (Guo et al. 2019; Zhang et al. 2023b) or adversarial perturbations (Chen et al. 2023). Although these strategies help improve diversity during training, they also risk contaminating the evaluation setup by inadvertently exposing the model to OOD-like noise during training, thus complicating generalization assessment. A second direction involves learning degradation-invariant representations, encouraging models to focus on clean image content rather than the noise distribution. Representative methods include MaskedDenoising and CLIPDenoising (Cheng, Liang, and Tan 2024). However, these methods often come at the cost of reduced image quality, particularly under high noise levels. A third class of methods emphasizes noise adaptation. For instance, LAN (Kim, Kim, and Baik 2024) and noise translation techniques (Ha et al. 2024) attempt to map OOD noise into a known or ID domain. However, such methods may suffer from scalability limitations or require prior exposure to real-world noise distributions. MID (Chihaoui and Favaro 2025) presents another strategy by fine-tuning a denoiser trained solely on simple synthetic noise using self-supervised objectives. While effective, this approach relies on ensemble inference, making it computationally intensive.

In contrast to the mentioned schemes that primarily focus on data augmentation, objective design, or adaptation

techniques, this work takes a fundamentally different perspective, *i.e.*, how *network architecture itself* can serve as an inductive bias for robustness.

## Problem Analysis

The canonical model for denoising (Liang et al. 2021; Li et al. 2023) [1] models an observed noisy image $Y$ as the sum of its underlying clean signal $X$ and an additive white Gaussian noise component $\lambda \cdot N$:

$$Y = X + \lambda \cdot N, N_{ij} \sim \mathcal{N}(0, 1), \tag{1}$$

where $\lambda$ is a scalar representing a uniform noise level. While this model is widely adopted for training deep denoisers, it introduces a critical mismatch with real-world noise characteristics, which are rarely spatially uniform. In practice, due to sensor response variations and environmental influences (*e.g.*, lighting, exposure, ISO), the noise level varies across spatial locations. Consequently, the scalar $\lambda$ should be replaced by a spatially varying map $\Lambda$, leading to a distributional shift between training and inference. This train-test discrepancy from uniform to non-uniform noise lies at the heart of poor OOD generalization. We identify a core reason for this degradation as feature entanglement. Conventional denoising networks tend to learn features that are tightly coupled to the magnitude of the noise. When the noise level becomes spatially varying, this coupling breaks down, leading to destabilized representations and degraded performance. To achieve robustness under such conditions, the key is to decouple feature representations from the noise level map. We argue that this can be systematically addressed by enforcing scale equivariance in the network design. Also known as first-order homogeneity, scale equivariance ensures that the network's output scales linearly with the input, formally expressed as follows:

**Definition 1** (Scale Equivariance)**.** *A function $\mathcal{H}$ obeying the $\mathcal{H}(0) = 0$ is said to be first-order homogeneous if, for input $I$ and a scalar $k > 0$, it satisfies the following condition:*

$$\mathcal{H}(k \cdot I) = k \cdot \mathcal{H}(I).$$

For a denoising network $\mathcal{F}$ endowed with the above property, its behavior during training under a uniform noise level $\lambda$ can be described as follows:

$$\mathcal{F}(X + \lambda \cdot N) = \lambda \cdot \mathcal{F}(\frac{X}{\lambda} + N) \quad \xrightarrow{\text{target}} \quad X. \tag{2}$$

As shown in Eq. (2), the network no longer learns a direct mapping from a noisy image to its clean counterpart. Instead, the effective learning task becomes denoising the input $X/\lambda + N$ to produce $X/\lambda$, thereby removing only the additive noise $N$. This formulation decouples the denoising mechanism from the global noise magnitude, embedding a scale-equivariant inductive bias directly into the network architecture. This decoupling extends naturally to the more challenging OOD setting with spatially variant noise $\Lambda$:

$$\mathcal{F}(X + \Lambda \odot \tilde{N}) = \Lambda \odot \mathcal{F}(\frac{X}{\Lambda} + \tilde{N}) \xrightarrow{\text{target}} X, \tag{3}$$

where $\odot$ denotes Hadamard product and $\tilde{N}$ is a base noise term that may originate from a different distribution. Remarkably, the network's internal processing remains structurally consistent: regardless of whether the noise level is uniform or spatially varying, the effective input is normalized to unit noise variance across the image. This normalization aligns with recent observations (Chen et al. 2023; Cheng, Liang, and Tan 2024) that denoisers inherently prioritize learning the noise pattern over image content. By enforcing a consistent computational pattern across domains, scale equivariance provides a principled and architecture-level path to achieving robustness.

The importance of strict first-order homogeneity is best understood by examining common violations, which we term architectural pathologies for this task: *1) Zero-Order Homogeneity (Scale Independence).* Consider a network with this property, *i.e.*, $\mathcal{F}(X + \lambda \cdot N) = \mathcal{F}(X/\lambda + N)$, its learning objective becomes mapping $X/\lambda + N$ to $X$. This imposes a flawed dual task: the network must simultaneously denoise the input and learn the global scaling $\lambda$, which forces an entanglement between features and the noise level seen during training. This entanglement is precisely what fails when $\lambda$ becomes a spatially-varying map; *2) Higher-Order Homogeneity.* A network with second-order homogeneity, *i.e.*, $\mathcal{F}(X+\lambda\cdot N) = \lambda^2\cdot\mathcal{F}(X/\lambda+N)$ involves similar entanglement phenomena. The objective becomes mapping $X/\lambda + N$ to $X/\lambda^2$, creating a distorted target. Moreover, the second-order property involves a quadratic growth in feature variance, increasing risk of learning collapse; *3) Exponential Activations.* Many exponential-based activation functions like GELU break homogeneity. They introduce input-dependent scaling errors that compound throughout the network, rendering OOD performance fragile.

**A First-Principles Validation.** With scale equivariance equipped, even a simple network should generalize well under distributional shifts. To test this in blind denoising, we build a baseline U-Net only from demonstrably scale-equivariant components: standard convolutions and ReLU activation. Despite its simplicity and lack of advanced modules, this network serves as a clean testbed to observe the effect of scale equivariance without confounding factors.

**Lemma 1** (Homogeneity of Basic Operators)**.** *Both convolution and ReLU are first-order homogeneous.*

*Proof.* Given an input $I$ and a scalar $k > 0$, the convolution output at position $(i, j)$ is:

$$\text{Conv}(k \cdot I)_{i,j} = k \cdot \sum_{m,n} I_{i-m,j-n} \cdot W_{m,n} = k \cdot \text{Conv}(I)_{i,j}.$$

Similarly, the relationship $\text{ReLU}(k \cdot I)_{i,j} = k \cdot \text{ReLU}(I)_{i,j}$ also holds. Thus, the claim is established. □

Building on this, the entire residual block, including the skip connection, preserves this property as well.

**Lemma 2** (Homogeneity of Residual Block)**.** *The 'Conv-ReLU-Conv' residual block is first-order homogeneous.*

*Proof.* Given Lemma 1, we have:

$$\begin{aligned}\text{RB}(k \cdot I) &= \text{Conv}(\text{ReLU}(\text{Conv}(k \cdot I))) + k \cdot I \\ &= k \cdot (\text{Conv}(\text{ReLU}(\text{Conv}(I))) + I) = k \cdot \text{RB}(I).\end{aligned}$$

[1] Without loss of generality, we adopt it through this work.

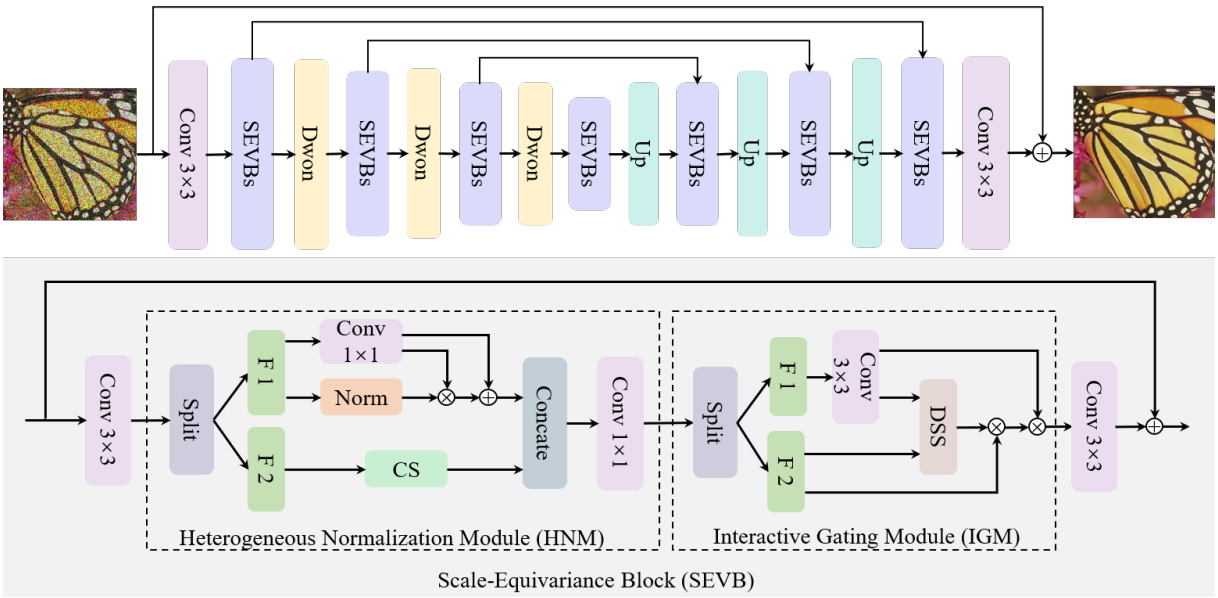



Figure 2: Overview of the proposed SEVNet with scale-equivariant design. CS represents Constant Scaling. DSS involves the dual-signal scaling term. The split operation divides the feature map into equally shaped sub-maps along the channel dimension.

| Method | In-Distribution | Out-Of-Distribution | | |
|---|---|---|---|---|
| | Gaussian $\sigma = 20$ | Speckle $\sigma = 90$ | Poisson $\alpha = 6$ | Mixture |
| Uformer | 32.89/0.9450 | 23.47/0.7510 | 22.92/0.7253 | 20.07/0.5812 |
| NAFNet | 32.28/0.9402 | 22.69/0.7047 | 22.30/0.6901 | 20.15/0.5995 |
| GRL | 32.77/0.9419 | 24.15/0.8021 | 23.65/0.7868 | 21.19/0.6990 |
| CGNet | 32.85/0.9454 | 22.77/0.6999 | 22.32/0.6829 | 20.02/0.5850 |
| Xformer | **32.99/0.9456** | 21.80/0.6386 | 21.15/0.6096 | 17.89/0.4682 |
| MHNet | 32.87/0.9452 | 21.70/0.6474 | 21.03/0.6150 | 17.56/0.4552 |
| AKDT | 32.54/0.9417 | 22.58/0.6774 | 22.08/0.6568 | 19.52/0.5468 |
| CRWKV | 32.39/0.9377 | 20.17/0.5845 | 19.60/0.5594 | 16.57/0.4299 |
| Baseline | 32.85/0.9447 | **24.49/0.8286** | **24.04/0.8183** | **21.67/0.7554** |

Table 1: Quantitative comparisons in PSNR/SSIM on the Urban100 dataset. The mixture noise is composed of Speckle and Poisson. The best results are highlighted in **bold**, while the second-best ones are underlined.

The property holds for the entire block. □

As reported in Tab. 1, this simple baseline outperforms those sophisticated but non-equivariant models on OOD tasks while remaining competitive on ID cases. The comparison corroborates our theoretical analysis.

## Designing a Scale-Equivariant Network

Our first-principles validation demonstrated that even a simple equivariant network can outperform sophisticated yet non-compliant ones. To further advance performance, we now introduce several novel modules designed to solve key challenges in constructing a high-performance, scale-equivariant denoiser. These components collectively define our SEVNet architecture, as illustrated in Fig. 2. Due to the page limit, formal proofs of equivariance for all modules are provided in the appendix.

**Constant Scaling.** Modern networks often rely on normalization operations, *e.g.*, Layer Norm, to stabilize training. While effective, these operations are scale-independent and thus conflict with our equivariant objective. To resolve this, we propose Constant Scaling (CS) as an equivariant stabilizer applied to the feature map $F \in \mathbb{R}^{c\times h\times w}$:

$$\mathrm{CS}(F) = \frac{F}{\sqrt{c}} \odot \eta, \tag{4}$$

where $\eta \in \mathbb{R}^c$ is a learnable, per-channel parameter initialized to 1. This operation normalizes the initial feature magnitude (preventing uncontrolled variance growth), while allowing $\eta$ to adapt channel-wise amplitudes during training.

**Normalized Self-Modulator.** While CS ensures global stability, it does not address local feature distortions caused by noise with spatial amplitude variations. To counteract this, we propose the Normalized Self-Modulator (NSM), a dynamic pixel-wise module designed to provide localized feature correction while preserving scale equivariance. The core idea is to perform a temporary local normalization, followed by an adaptive rescaling that restores critical magnitude in an equivariant manner. NSM performs self-modulation using affine parameters derived from the token's original pre-normalized state. This can be formulated as:

$$\begin{aligned} [\gamma, \beta] &\leftarrow \mathrm{Conv}_{1\times1}(F), \\ \mathrm{NSM}(F) &= \gamma \odot \frac{X - \mu(F)}{\sqrt{\sigma^2(F)}} + \beta, \end{aligned} \tag{5}$$

where $\mu$ and $\sigma^2$ are the per-token mean and variance, respectively. Since the affine parameters $\gamma \in \mathbb{R}^{c\times h\times w}$ and $\beta \in \mathbb{R}^{c\times h\times w}$ are derived via a linear transformation of the input, they are first-order homogeneous. This ensures that the entire NSM remains strictly scale-equivariant, providing adaptive feature correction without violating our principle.

| Method | In-Distribution | Out-Of-Distribution | | | | | |
|---|---|---|---|---|---|---|---|
| | Gaussian | Speckle | Poisson | Mixture | Variant 1 | Variant 2 | Variant 3 |
| | | | | Kodak24 | | | |
| SwinIR (Liang et al. 2021) | 33.50/0.9029 | 27.55/0.7931 | 27.23/0.7820 | 22.79/0.5887 | 25.99/0.7457 | 26.31/0.7541 | 25.63/0.7313 |
| Uformer (Wang et al. 2022) | 33.62/0.9062 | 27.20/0.7656 | 26.86/0.7506 | 22.27/0.5102 | 25.42/0.6950 | 25.82/0.7102 | 25.07/0.6750 |
| Restormer (Zamir et al. 2022) | **33.70/0.9076** | 26.12/0.6978 | 25.74/0.6807 | 20.11/0.4279 | 23.19/0.6078 | 24.01/0.6280 | 22.87/0.5809 |
| NAFNet (Chen et al. 2022) | 33.42/0.9048 | 25.39/0.6681 | 25.15/0.6580 | 22.16/0.5422 | 23.49/0.6087 | 23.99/0.6227 | 23.33/0.5967 |
| GRL (Li et al. 2023) | 33.52/0.9025 | 27.58/0.7981 | 27.29/0.7904 | 23.66/0.6789 | 26.19/0.7643 | 26.47/0.7684 | 25.87/0.7560 |
| CGNet (Ghasemabadi et al. 2024) | 33.65/0.9075 | 25.43/0.6564 | 25.11/0.6416 | 21.74/0.4981 | 23.32/0.5953 | 23.88/0.6106 | 23.04/0.5714 |
| Xformer (Zhang et al. 2024) | 33.66/0.9069 | 25.29/0.6388 | 24.82/0.6165 | 18.91/0.3507 | 22.56/0.5433 | 23.15/0.5582 | 22.10/0.5143 |
| MHNet (Gao et al. 2025) | 33.64/0.9074 | 24.88/0.6390 | 24.43/0.6163 | 18.67/0.3462 | 21.91/0.5361 | 22.72/0.5573 | 21.64/0.5121 |
| AKDT (Brateanu et al. 2025) | 33.48/0.9050 | 25.57/0.6565 | 25.20/0.6391 | 20.94/0.4493 | 23.39/0.5887 | 23.94/0.6018 | 23.05/0.5637 |
| ESC (Lee, Yun, and Ro 2025) | 33.51/0.9028 | 26.52/0.7017 | 26.16/0.6843 | 21.58/0.4691 | 24.44/0.6261 | 24.94/0.6402 | 24.03/0.6021 |
| CRWKV (Chen et al. 2025) | 33.50/0.9038 | 23.89/0.5978 | 23.43/0.5763 | 17.68/0.3380 | 20.73/0.5026 | 21.48/0.5186 | 20.26/0.4726 |
| Ours | 33.68/0.9069 | **27.94/0.8164** | **27.66/0.8106** | **24.23/0.7368** | **26.64/0.7908** | **26.91/0.7932** | **26.34/0.7858** |
| | | | | Urban100 | | | |
| SwinIR (Liang et al. 2021) | 32.71/0.9415 | 25.53/0.8338 | 25.19/0.8238 | 20.78/0.6490 | 23.80/0.7858 | 26.09/0.8529 | 23.75/0.7831 |
| Uformer (Wang et al. 2022) | 32.89/0.9450 | 25.03/0.8055 | 24.71/0.7936 | 20.07/0.5812 | 23.11/0.7408 | 25.60/0.8331 | 23.10/0.7374 |
| Restormer (Zamir et al. 2022) | **33.12/0.9469** | 24.31/0.7547 | 24.00/0.7439 | 18.56/0.5142 | 21.65/0.6759 | 24.27/0.7797 | 21.64/0.6722 |
| NAFNet (Chen et al. 2022) | 32.28/0.9402 | 23.87/0.7430 | 23.64/0.7356 | 20.15/0.5995 | 22.05/0.6832 | 22.44/0.7000 | 22.04/0.6824 |
| GRL (Li et al. 2023) | 32.77/0.9419 | 25.53/0.8362 | 25.21/0.8279 | 21.19/0.6990 | 23.91/0.7973 | 26.08/0.8541 | 23.85/0.7952 |
| CGNet (Ghasemabadi et al. 2024) | 32.85/0.9454 | 24.13/0.7449 | 23.85/0.7361 | 20.02/0.5850 | 22.14/0.6835 | 22.56/0.7012 | 22.09/0.6790 |
| Xformer (Zhang et al. 2024) | 32.99/0.9456 | 23.74/0.7138 | 23.37/0.6996 | 17.89/0.4682 | 21.24/0.6326 | 21.69/0.6467 | 21.19/0.6276 |
| MHNet (Gao et al. 2025) | 32.87/0.9452 | 23.51/0.7224 | 23.14/0.7073 | 17.56/0.4552 | 20.81/0.6290 | 21.41/0.6505 | 20.84/0.6262 |
| AKDT (Brateanu et al. 2025) | 32.54/0.9417 | 24.05/0.7304 | 23.76/0.7197 | 19.52/0.5468 | 22.06/0.6697 | 22.46/0.6841 | 22.01/0.6657 |
| ESC (Lee, Yun, and Ro 2025) | 32.97/0.9439 | 24.38/0.7463 | 24.07/0.7351 | 19.42/0.5445 | 22.25/0.6793 | 22.64/0.6917 | 22.18/0.6747 |
| CRWKV (Chen et al. 2025) | 32.39/0.9377 | 22.00/0.6544 | 21.67/0.6410 | 16.57/0.4299 | 19.42/0.5743 | 19.94/0.5889 | 19.34/0.5675 |
| Ours | 32.94/0.9453 | **26.14/0.8644** | **25.84/0.8590** | **21.94/0.7698** | **24.59/0.8371** | **26.66/0.8771** | **24.54/0.8357** |

Table 2: Quantitative denoising comparisons (PSNR/SSIM) on the different datasets. The average results across 3 noise levels for Speckle and Poisson noise are reported. The best results are marked in **bold**, while the second-best ones are underlined.

**Heterogeneous Normalization Module.** While individual CS and NSM modules provide stabilization, they offer a trade-off between efficiency and expressive power. To get the best of both worlds, we combine them into a single Heterogeneous Normalization Module (HNM). The design is simple yet effective: we partition the feature channels and process one subset with the lightweight CS for baseline stability, while the other subset is processed by the more powerful NSM for dynamic, fine-grained correction. This heterogeneous strategy creates a robust, multi-level stabilization mechanism that achieves adaptive feature modulation.

**Interactive Gating Module.** Although ReLU satisfies the requirement of scale equivariance, its limited expressiveness may hinder performance. In contrast, more sophisticated activations such as GELU and Swish offer stronger nonlinear modeling capabilities but inherently violate scale equivariance, introducing a potential bottleneck in equivariant design. To bridge this gap, we propose the Interactive Gating Module (IGM), a novel activation function that is both highly expressive and strictly scale-equivariant by construction. Specifically, given the input feature map $F \in \mathbb{R}^{c\times h\times w}$, we first split it along the channel dimension into two equal parts $F_1 \in \mathbb{R}^{\frac{c}{2}\times h\times w}$ and $F_2 \in \mathbb{R}^{\frac{c}{2}\times h\times w}$, and then feed them into a gating unit. We define the IGM as follows:

$$[F_1, F_2] \leftarrow \text{Split}(F),\ F_v \leftarrow \text{Conv}_{3\times3}(F_1),\ F_g \leftarrow F_2,$$
$$\text{IG}(F_v, F_g) = \frac{F_v \odot F_g}{\sqrt{\sigma^2(F_v) + \sigma^2(F_g)}}, \tag{6}$$

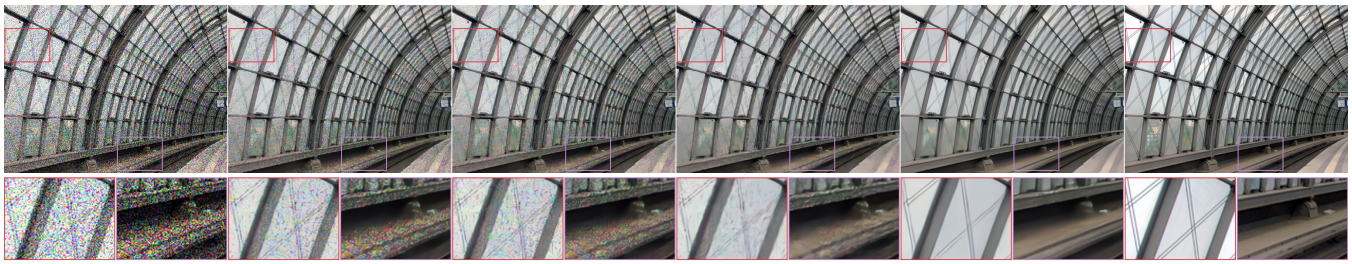
where $\sigma^2$ stands for the per-token variance. A brilliance of our IGM is the dual-signal scaling term, which not only ensures the entire module remains scale-equivariant but also enables the gating strength to be co-determined by both the feature $F_v$ and the gating signal $F_g$. This is achieved by transforming the gating signal's variance into a relative importance weight $\frac{\sigma^2(F_g)}{\sigma^2(F_v)+\sigma^2(F_g)}$. Beyond its mathematical compliance, the IGM is interactive. The denominator normalizes the output based on the combined energy (variance) of both the value and the gate. This prevents the output from exploding when both signals are strong, allowing the gate's influence to be modulated by its own magnitude relative to the value signal. This stable interaction enhances the network's feature selection capability beyond a simple ReLU, providing the desired equivariant non-linearity.

## Experiments

Our SEVNet is built on the principle of scale equivariance. Together, these components create a network that is theoretically sound for OOD generalization. We now turn to the experimental results to verify these claims.

**Datasets** We train models on the DIV2K (Timofte et al. 2017) dataset with Gaussian noise of various levels ($\sigma \in [5, 20]$) for synthetic OOD noise removal, and ISP noise (Guo et al. 2019) for handling real-world cases. We evaluate models on signal-dependent noise (Poisson, Speckle, and their mixtures). To enhance the non-uniformity of the

(a) Poisson Noise




(b) Variant 2 of Speckle Noise

(c) Variant 3 of Speckle Noise

Figure 3: Qualitative comparisons on OOD noises. Please refer to the appendix to see more visual results.

level map, variants (Variant 1-3) of Speckle noise are synthesized by spatial functions (Yue et al. 2024): sine-cosine, peaks, and Gaussian kernels. The evaluation is conducted on standard test sets (Kodak24, CBSD68, Urban100) and real-world benchmarks (CC, PolyU, HighISO) (Ha et al. 2024). Denoising quality is quantified by the Peak Signal-to-Noise Ratio (PSNR) and the Structural Similarity Index (SSIM).

| **Method** | **CC** PSNR/SSIM | **PolyU** PSNR/SSIM | **HighISO** PSNR/SSIM |
|---|---|---|---|
| SwinIR | 33.61/0.8583 | 36.02/0.9151 | 35.53/0.8797 |
| Uformer | 34.30/0.8766 | 36.18/0.9182 | 35.96/0.8891 |
| Restormer | 33.71/0.8576 | 36.02/0.9146 | 35.57/0.8788 |
| NAFNet | 34.88/0.8952 | 36.30/0.9230 | 36.38/0.9002 |
| GRL | 33.92/0.8638 | 36.14/0.9181 | 35.75/0.8842 |
| CGNet | 33.69/0.8580 | 36.05/0.9155 | 35.58/0.8802 |
| Xformer | 33.65/0.8567 | 36.01/0.9141 | 35.54/0.8786 |
| MHNet | 33.83/0.8634 | 36.05/0.9156 | 35.70/0.8826 |
| AKDT | 34.48/0.8838 | 36.17/0.9197 | 36.05/0.8921 |
| ESC | 34.08/0.8697 | 36.30/0.9208 | 35.94/0.8880 |
| CRWKV | 31.13/0.7507 | 31.98/0.7718 | 31.66/0.7571 |
| Ours+NT | **36.95/0.9486** | **37.60/0.9557** | **39.36/0.9675** |

Table 4: Quantitative comparisons on real-world noise.

**Quantitative Comparison.** As shown in Tab. 2, our scale-equivariant network exhibits superior OOD robustness without compromising ID performance. Despite incorporating advanced modules and popular components, these existing networks violate scale equivariance, leading to severe performance degradation in OOD scenarios. More results and efficiency comparisons can be found in the appendix.

**Qualitative Comparison.** As shown in Fig. 3, even in severely degraded conditions, our network effectively removes various OOD noises while preserving clear edges and some textures. In contrast, the denoising results of competing methods are poor and exhibit different degradation patterns, *e.g.*, residual noise and artifacts.

**Real-World Extension.** The applicability of our scale-equivariant network, which assumes independent noise, can be successfully extended to real-world data by integrating a lightweight, SEVB-based noise translator (Ha et al. 2024). The strong performance of this approach (Tab. 4), surpassing complex yet inefficient models, highlights a highly promising path for synthetic-to-real generalization that involves shifting the focus from building ever-more-complex denoisers to combining simple, principled networks, *i.e.*, those that

| Setting | Speckle $\sigma = 90$ PSNR/ SSIM | Poisson $\alpha = 60$ PSNR/SSIM | Mixture $(\sigma = 90, \alpha = 6)$ PSNR/SSIM | Variant 1 PSNR/SSIM | Variant 2 PSNR/SSIM | Variant 3 PSNR/SSIM |
|---|---|---|---|---|---|---|
| Heterogeneous Normalization Module (HNM) | | | | | | |
| (1a) Ours | **26.86/0.7933** | **26.42/0.7838** | **24.23/0.7368** | **26.64/0.7908** | **26.91/0.7932** | **26.34/0.7858** |
| (1b) HNM→ Full CS | 26.78/0.7888 | 26.35/0.7793 | 24.17/0.7317 | 26.57/0.7858 | 26.83/0.7887 | 26.27/0.7812 |
| (1c) w/ CS + w/o NSM | 26.74/0.7885 | 26.30/0.7791 | 24.16/0.7322 | 26.52/0.7855 | 26.79/0.7885 | 26.23/0.7810 |
| (1d) w/ HNM, IGM→ELU | 26.77/0.7894 | 26.33/0.7799 | 24.16/0.7325 | 26.55/0.7866 | 26.82/0.7893 | 26.25/0.7818 |
| (1e) w/ HNM, IGM→GELU | 26.74/0.7872 | 26.29/0.7775 | 24.13/0.7291 | 26.52/0.7844 | 26.79/0.7871 | 26.23/0.7794 |
| (1f) w/o HNM, IGM→ELU | 22.79/0.5859 | 21.51/0.5205 | 16.63/0.2782 | 21.56/0.5565 | 22.42/0.5896 | 20.84/0.5186 |
| (1g) w/o HNM, IGM→GELU | 25.73/0.7241 | 25.17/0.7007 | 22.41/0.5495 | 25.40/0.7119 | 25.74/0.7228 | 24.97/0.6938 |
| (1h) Scale-independent affine | 26.54/0.7778 | 26.09/0.7666 | 23.85/0.7101 | 26.32/0.7747 | 26.59/0.7778 | 26.01/0.7684 |
| Interactive Gating Module (IGM) | | | | | | |
| (2a) w/o IGM | 26.75/0.7887 | 26.31/0.7789 | 24.15/0.7317 | 26.53/0.7858 | 26.80/0.7886 | 26.24/0.7811 |
| (2b) IGM → ReLU | 26.80/0.7892 | 26.36/0.7797 | 24.17/0.7317 | 26.59/0.7865 | 26.85/0.7891 | 26.28/0.7814 |
| (2c) Dual→Single scaling | 26.83/0.7916 | 26.39/0.7821 | 24.20/0.7347 | 26.62/0.7891 | 26.88/0.7916 | 26.31/0.7840 |
| (2d) HNM→ LN, w/ IGM | 26.39/0.7684 | 25.90/0.7545 | 23.46/0.6744 | 26.13/0.7636 | 26.41/0.7676 | 25.83/0.7562 |
| (2e) HNM→LN, IGM→ReLU | 25.53/0.6979 | 24.82/0.6612 | 21.88/0.5057 | 25.15/0.6898 | 25.54/0.7034 | 24.68/0.6640 |
| (2f) HNM→LN, IGM→GELU | 26.24/0.7583 | 25.78/0.7426 | 23.07/0.5991 | 25.96/0.7485 | 26.28/0.7576 | 25.65/0.7388 |

Table 3: Ablation studies on the Kodak24 dataset.

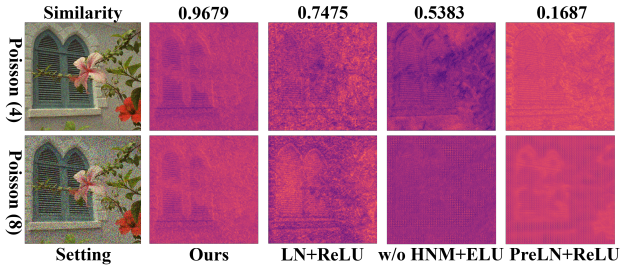



Figure 4: Comparison in visual and cosine similarities of feature maps under different noise level maps.

are scale-equivariant, with noise adaptation methods.

**Ablation Study.** In Tab. 3, we validate the effectiveness of the HNM and IGM, as well as the role of scale equivariance:

*1) Impact of HNM*. Applying CS to the entire feature mapping (1b) or bypassing NSM (1c) leads to a performance drop, confirming the necessity of combining both components. Interestingly, HNM allows the network to maintain comparable generalization even when using exponential activations like ELU (1d) and GELU (1e). Without HNM, however, these activations (1f-1g) severely impair OOD capability, likely because HNM shifts features to intervals where the exponential effect is less pronounced. For the NSM's design, generating its affine parameters from the normalized features (1h), also degrades performance, though the decline is moderate since the NSM operates on only a portion of the features. Collectively, these results (1e-1h) demonstrate that scale equivariance is critical for robust denoising and that our HNM can partially suppress the adverse effects of violating this principle.

*2) Impact of IGM*. Removing IGM (2a) or replacing it with ReLU (2b) both yield suboptimal results, demonstrating IGM's superior nonlinear capabilities. compared with the scaling term obtained from gating (2c), the dual-signal scaling term further enhances OOD robustness through its feature selection effect. Under the scale-independent Layer Norm (LN) setting, our scaling term (2d) enables the network to perceive the scale but does not give accurate equivariance, thereby retaining weak generalization. This resilience is not found in standard activations like ReLU (2e) and GELU (2f). More importantly, without the scaling term, IGM becomes a second-order homogeneous operation, and the resulting quadratic growth in feature variance causes training to collapse, making it impossible to report the corresponding ablation results. Ultimately, these results (2d-2f) demonstrate that a scale-equivariant design is key for OOD denoising and that our IGM partially mitigates the performance degradation caused by scale-invariant operations.

**Feature Analysis.** To further explain the advantage of scale equivariance, we visualize feature maps (Fig. 4) from the final decoder stage. Under vastly different corruption degrees, our scale-equivariant network produces remarkably consistent feature maps. Conversely, designs that violate this principle heighten the network's sensitivity to variations in the level map. For instance, a network with scale-independence (middle and preceding Layer Norm) or exponential activation (ELU) yields dissimilar feature pairs. In short, Scale equivariance indeed forces the network to learn robust representations that are decoupled from the noise level map.

## Conclusion

We have demonstrated that scale equivariance is a key architectural principle for OOD denoising, allowing networks to generalize from uniform to non-uniform noise level maps. Guided by this principle, we develop several network components. Specifically, HNM stabilizes and corrects feature distributions. IGM provides powerful nonlinear representations and refines the transmission of information flow. Our work not only delivers a superior OOD denoiser but also advocates for a new perspective in designing deep networks for robust image restoration. Future work could explore this principle for more spatially non-uniform restoration tasks.

## Acknowledgement

This work was supported by the National Natural Science Foundation of China under Grant nos.62372251 and 62072327.

## References

Brateanu, A.; Balmez, R.; Avram, A.; and Orhei, C. 2025. AKDT: Adaptive Kernel Dilation Transformer for Effective Image Denoising. In Bashford-Rogers, T.; Meneveaux, D.; Ammi, M.; Ziat, M.; Jänicke, S.; Purchase, H. C.; Radeva, P.; Furnari, A.; Bouatouch, K.; and de Sousa, A. A., eds., *Proceedings of the 20th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, VISIGRAPP 2025 - Volume 3: VISAPP, Porto, Portugal, February 26-28, 2025*, 418–425. SCITEPRESS.

Chang, M.; Li, Q.; Feng, H.; and Xu, Z. 2020. Spatial-Adaptive Network for Single Image Denoising. In Vedaldi, A.; Bischof, H.; Brox, T.; and Frahm, J., eds., *Computer Vision - ECCV 2020 - 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part XXX*, volume 12375 of *Lecture Notes in Computer Science*, 171–187. Springer.

Chen, B.; Chai, T.; Jiang, W.; Xu, Y.; Zhou, G.; and Wu, X. 2025. Multi-View Learning with Context-Guided Receptance for Image Denoising. *CoRR*, abs/2505.02705.

Chen, H.; Gu, J.; Liu, Y.; Magid, S. A.; Dong, C.; Wang, Q.; Pfister, H.; and Zhu, L. 2023. Masked Image Training for Generalizable Deep Image Denoising. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2023, Vancouver, BC, Canada, June 17-24, 2023*, 1692–1703. IEEE.

Chen, L.; Chu, X.; Zhang, X.; and Sun, J. 2022. Simple Baselines for Image Restoration. In Avidan, S.; Brostow, G. J.; Cissé, M.; Farinella, G. M.; and Hassner, T., eds., *Computer Vision - ECCV 2022 - 17th European Conference, Tel Aviv, Israel, October 23-27, 2022, Proceedings, Part VII*, volume 13667 of *Lecture Notes in Computer Science*, 17–33. Springer.

Chen, X.; Li, Z.; Pu, Y.; Liu, Y.; Zhou, J.; Qiao, Y.; and Dong, C. 2024. A Comparative Study of Image Restoration Networks for General Backbone Network Design. In Leonardis, A.; Ricci, E.; Roth, S.; Russakovsky, O.; Sattler, T.; and Varol, G., eds., *Computer Vision - ECCV 2024 - 18th European Conference, Milan, Italy, September 29-October 4, 2024, Proceedings, Part LXXI*, volume 15129 of *Lecture Notes in Computer Science*, 74–91. Springer.

Cheng, J.; Liang, D.; and Tan, S. 2024. Transfer CLIP for Generalizable Image Denoising. *CoRR*, abs/2403.15132.

Chihaoui, H.; and Favaro, P. 2025. Unsupervised Real-World Denoising: Sparsity is All You Need. *CoRR*, abs/2503.21377.

Gao, H.; Zhang, Y.; Yang, J.; and Dang, D. 2025. Mixed hierarchy network for image restoration. *Pattern Recognit.*, 161: 111313.

Ghasemabadi, A.; Janjua, M. K.; Salameh, M.; Zhou, C.; Sun, F.; and Niu, D. 2024. Cascadedgaze: Efficiency in global context extraction for image restoration. *arXiv preprint arXiv:2401.15235*.

Guo, S.; Yan, Z.; Zhang, K.; Zuo, W.; and Zhang, L. 2019. Toward Convolutional Blind Denoising of Real Photographs. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 1712–1722.

Ha, I.; Ryou, D.; Seo, S.; and Han, B. 2024. Learning to Translate Noise for Robust Image Denoising. *CoRR*, abs/2412.04727.

Kim, C.; Kim, T. H.; and Baik, S. 2024. LAN: Learning to Adapt Noise for Image Denoising. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2024, Seattle, WA, USA, June 16-22, 2024*, 25193–25202. IEEE.

Lee, D.; Yun, S.; and Ro, Y. 2025. Emulating Self-attention with Convolution for Efficient Image Super-Resolution. *CoRR*, abs/2503.06671.

Li, J.; Zhang, Z.; and Zuo, W. 2025. Rethinking Transformer-Based Blind-Spot Network for Self-Supervised Image Denoising. In Walsh, T.; Shah, J.; and Kolter, Z., eds., *AAAI-25, Sponsored by the Association for the Advancement of Artificial Intelligence, February 25 - March 4, 2025, Philadelphia, PA, USA*, 4788–4796. AAAI Press.

Li, Y.; Fan, Y.; Xiang, X.; Demandolx, D.; Ranjan, R.; Timofte, R.; and Gool, L. V. 2023. Efficient and Explicit Modelling of Image Hierarchies for Image Restoration. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2023, Vancouver, BC, Canada, June 17-24, 2023*, 18278–18289. IEEE.

Liang, J.; Cao, J.; Sun, G.; Zhang, K.; Gool, L. V.; and Timofte, R. 2021. SwinIR: Image Restoration Using Swin Transformer. In *IEEE/CVF International Conference on Computer Vision Workshops, ICCVW 2021, Montreal, BC, Canada, October 11-17, 2021*, 1833–1844. IEEE.

Liao, X.; Zheng, T.; Zhong, J.; Zhang, P.; and Ren, C. 2024. Asymmetric Mask Scheme for Self-supervised Real Image Denoising. In Leonardis, A.; Ricci, E.; Roth, S.; Russakovsky, O.; Sattler, T.; and Varol, G., eds., *Computer Vision - ECCV 2024 - 18th European Conference, Milan, Italy, September 29-October 4, 2024, Proceedings, Part XXV*, volume 15083 of *Lecture Notes in Computer Science*, 199–215. Springer.

Liu, H.; Fu, J.; Xie, Q.; and Meng, D. 2025. Rotation-Equivariant Self-Supervised Method in Image Denoising. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2025, Nashville, TN, USA, June 11-15, 2025*, 12720–12730. Computer Vision Foundation / IEEE.

Ma, Q.; Jiang, J.; Zhou, X.; Liang, P.; Liu, X.; and Ma, J. 2025. Pixel2Pixel: A Pixelwise Approach for Zero-Shot Single Image Denoising. *IEEE Trans. Pattern Anal. Mach. Intell.*, 47(6): 4614–4629.

Ma, X.; Dai, X.; Bai, Y.; Wang, Y.; and Fu, Y. 2024. Rewrite the Stars. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2024, Seattle, WA, USA, June 16-22, 2024*, 5694–5703. IEEE.

Mansour, Y.; and Heckel, R. 2023. Zero-Shot Noise2Noise: Efficient Image Denoising without any Data. In *IEEE/CVF*

*Conference on Computer Vision and Pattern Recognition, CVPR 2023, Vancouver, BC, Canada, June 17-24, 2023*, 14018–14027. IEEE.

Quan, Y.; Chen, M.; Pang, T.; and Ji, H. 2020. Self2Self With Dropout: Learning Self-Supervised Denoising From Single Image. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2020, Seattle, WA, USA, June 13-19, 2020*, 1887–1895. Computer Vision Foundation / IEEE.

Shen, H.; Zhao, Z.; Zhang, W.; and et. al. 2023. Adaptive Dynamic Filtering Network for Image Denoising. In Williams, B.; Chen, Y.; and Neville, J., eds., *Thirty-Seventh AAAI Conference on Artificial Intelligence, AAAI 2023, Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence, IAAI 2023, Thirteenth Symposium on Educational Advances in Artificial Intelligence, EAAI 2023, Washington, DC, USA, February 7-14, 2023*, 2227–2235. AAAI Press.

Timofte, R.; Agustsson, E.; Gool, L. V.; Yang, M.; and et. al. 2017. NTIRE 2017 Challenge on Single Image Super-Resolution: Methods and Results. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2017, Honolulu, HI, USA, July 21-26, 2017*, 1110–1121. IEEE Computer Society.

Wang, Z.; Cun, X.; Bao, J.; Zhou, W.; Liu, J.; and Li, H. 2022. Uformer: A General U-Shaped Transformer for Image Restoration. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2022, New Orleans, LA, USA, June 18-24, 2022*, 17662–17672. IEEE.

Yue, Z.; Yong, H.; Zhao, Q.; Zhang, L.; Meng, D.; and Wong, K. K. 2024. Deep Variational Network Toward Blind Image Restoration. *IEEE Trans. Pattern Anal. Mach. Intell.*, 46(11): 7011–7026.

Zamir, S. W.; Arora, A.; Khan, S.; Hayat, M.; Khan, F. S.; and Yang, M. 2022. Restormer: Efficient Transformer for High-Resolution Image Restoration. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 5718–5729.

Zamir, S. W.; Arora, A.; Khan, S. H.; Hayat, M.; Khan, F. S.; Yang, M.; and Shao, L. 2020. Learning Enriched Features for Real Image Restoration and Enhancement. In Vedaldi, A.; Bischof, H.; Brox, T.; and Frahm, J., eds., *Computer Vision - ECCV 2020 - 16th European Conference, Glasgow, UK, August 23-28, 2020, Proceedings, Part XXV*, volume 12370 of *Lecture Notes in Computer Science*, 492–511. Springer.

Zhang, D.; Zhou, F.; Jiang, Y.; and Fu, Z. 2023a. MM-BSN: Self-Supervised Image Denoising for Real-World with Multi-Mask based on Blind-Spot Network. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2023 - Workshops, Vancouver, BC, Canada, June 17-24, 2023*, 4189–4198. IEEE.

Zhang, J.; Zhang, Y.; Gu, J.; Dong, J.; Kong, L.; and Yang, X. 2024. Xformer: Hybrid X-Shaped Transformer for Image Denoising. In *The Twelfth International Conference on Learning Representations, ICLR 2024, Vienna, Austria, May 7-11, 2024*. OpenReview.net.

Zhang, K.; Li, Y.; Liang, J.; Cao, J.; Zhang, Y.; Tang, H.; Fan, D.; Timofte, R.; and Gool, L. V. 2023b. Practical Blind Image Denoising via Swin-Conv-UNet and Data Synthesis. *Mach. Intell. Res.*, 20(6): 822–836.

Zhang, Y.; Li, K.; Li, K.; Sun, G.; Kong, Y.; and Fu, Y. 2021. Accurate and Fast Image Denoising via Attention Guided Scaling. *IEEE Trans. Image Process.*, 30: 6255–6265.

# Appendix

## Theoretical Analysis

Following the design principle of scale equivariance, we propose the Constant Scaling (CS), Normalized Self-Modulator (NSM), and Interactive Gating Module (IGM) to construct a more powerful OOD denoiser. The proofs of equivariance for the main steps within these network components are provided below. To facilitate the proofs, we uniformly represent the input as vectors, as the key operations are performed on pixel feature vectors.

**Lemma 3** (Homogeneity of CS)**.** *The Constant Scaling operation is first-order homogeneous.*

*Proof.* Given an input $\mathbf{v} \in \mathbb{R}^c$ and a scalar $k > 0$, we need to show that $\text{CS}(k \cdot \mathbf{v}) = k \cdot \text{CS}(\mathbf{v})$.

Since the learnable parameter $\eta \in \mathbb{R}^c$ is independent of the input $\mathbf{v}$, it remains unchanged when the input is scaled. We proceed as follows:

$$\text{CS}(k \cdot \mathbf{v}) = \frac{k \cdot \mathbf{v}}{\sqrt{c}} \odot \eta = k \cdot \text{CS}(\mathbf{v}).$$

This completes the proof. □

**Lemma 4** (Homogeneity of NSM)**.** *The Normalized Self-Modulator (NSM) operation is first-order homogeneous.*

*Proof.* Given an input $\mathbf{v} \in \mathbb{R}^c$ and a scalar $k > 0$, we need to show that $\text{NSM}(k \cdot \mathbf{v}) = k \cdot \text{NSM}(\mathbf{v})$.

First, we analyze the components:

- **Affine Parameters:** The $1 \times 1$ convolution is a linear transformation and thus satisfies first-order homogeneity:

$$[k \cdot \beta, k \cdot \gamma] \leftarrow \text{Conv}_{1\times1}(k \cdot \mathbf{v}).$$

- **Mean and Standard Deviation:** The mean and variance of the scaled vector $k \cdot \mathbf{v}$ are:

$$\mu(k \cdot \mathbf{v}) = \frac{1}{c}\sum_{i=1}^{c}(k \cdot \mathbf{v}_i) = k \cdot \mu(\mathbf{v}).$$

$$\begin{aligned}\sigma^2(k \cdot \mathbf{v}) &= \frac{1}{c}\sum_{i=1}^{c}(k \cdot \mathbf{v}_i - k \cdot \mu(\mathbf{v}))^2 \\ &= \frac{k^2}{c}\sum_{i=1}^{c}(\mathbf{v}_i - \mu(\mathbf{v}))^2 = k^2 \cdot \sigma^2(\mathbf{v}).\end{aligned}$$

  Therefore, the standard deviation is $\sigma(k \cdot \mathbf{v}) = k \cdot \sigma(\mathbf{v})$.
- **Normalized Vector:** The normalized vector part of NSM is scale-independent (zero-order homogeneous):

$$\frac{k \cdot \mathbf{v} - \mu(k \cdot \mathbf{v})}{\sigma(k \cdot \mathbf{v})} = \frac{\mathbf{v} - \mu(\mathbf{v})}{\sigma(\mathbf{v})}.$$

Now, we combine these results to analyze the full NSM operation:

$$\begin{aligned}\text{NSM}(k \cdot \mathbf{v}) &= k \cdot \gamma \odot \frac{k \cdot \mathbf{v} - \mu(k \cdot \mathbf{v})}{\sigma(k \cdot \mathbf{v})} + k \cdot \beta \\ &= k \cdot \text{NSM}(\mathbf{v}).\end{aligned}$$

This completes the proof. □

**Lemma 5** (Homogeneity of IGM)**.** *The Interactive Gating Module (IGM) is a first-order homogeneous system.*

*Proof.* Given an input $\mathbf{v} \in \mathbb{R}^c$ and a scalar $k > 0$, we need to show that $\text{IGM}(k \cdot \mathbf{v}) = k \cdot \text{IGM}(\mathbf{v})$.

We will analyze the homogeneity of the numerator and the denominator of the IGM formula separately. Here, the convolution and split operations are omitted

- **Numerator Analysis:** The numerator is a Hadamard product. Its homogeneity is determined as follows:

$$(k \cdot \mathbf{v}_v) \odot (k \cdot \mathbf{v}_g) = k^2 \cdot (\mathbf{v}_v \odot \mathbf{v}_g).$$

  Thus, the numerator is second-order homogeneous.
- **Denominator Analysis:** The denominator is the square root of the sum of variances. The variance of a scaled vector $\sigma^2(k \cdot \mathbf{v})$ is second-order homogeneous. Applying this property to the denominator, we get:

$$\sigma^2(k \cdot \mathbf{v}_v) + \sigma^2(k \cdot \mathbf{v}_g) = k^2 \cdot \left(\sigma^2(\mathbf{v}_v) + \sigma^2(\mathbf{v}_g)\right).$$

Combining the results for the numerator and the denominator, we analyze the full IGM operation:

$$\begin{aligned}\text{IGM}(k \cdot \mathbf{v}_v, k \cdot \mathbf{v}_g) &= \frac{(k \cdot \mathbf{v}_v) \odot (k \cdot \mathbf{v}_g)}{\sqrt{\sigma^2(k \cdot \mathbf{v}_v) + \sigma^2(k \cdot \mathbf{v}_g)}} \\ &= k \cdot \text{IGM}(\mathbf{v}_v, \mathbf{v}_g).\end{aligned}$$

This completes the proof. □

## Experiments

**Implementation details**. We extract $128 \times 128$ size of patches for training. The data augmentation is conducted on input patches with random horizontal flips and rotations. The batch size is set to 4. For synthetic OOD noises, we adopt the Adam solver with $\beta_1 = 0.9$ and $\beta_2 = 0.999$ for 1000 epochs. The learning rate starts from $2 \times 10^{-4}$ and decays by a factor of 0.5 every 200 epochs. For real-world cases, we first freeze the pre-trained blind denoiser. Then, we set the epoch and initial learning rate to 250 and $1 \times 10^{-4}$, respectively, to train the lightweight noise translator based on our SEVB. The corresponding training process and loss function refer to this noise adaptive method (Ha et al. 2024). All the experiments are conducted with PyTorch and NVIDIA GeForce RTX 3090 GPUs. Our network and others are trained under the blind denoising task. To facilitate image processing, data centralization and its corresponding inverse operation are applied at the start and end of all networks, respectively.

**Synthetic OOD Noise.** For the OOD scenarios, we synthesize several pixel-dependent noises, and the corresponding noise level maps are spatially non-uniform. *1) Poisson noise:* $Y = X + \alpha \cdot N_p$, where $Y$, $X$, and $N_p$ respectively represent the observed image, the clear image, and the Poisson noise component. The synthesis of $N_p$ depends on the pixels of $X$. $\alpha \in \{4, 5, 6\}$ represents the scaling factor of the noise level. *2) Speckle noise:* $Y = X + \sigma \cdot \sqrt{X} \odot N$, where $N$ is derived from the standard Gaussian distribution. The noise level at each position is affected by the corresponding pixel, while $\sigma \in \{60, 80, 90\}$ controls the global damage intensity. *3)*

| Method | Speckle Noise | | | Poisson Noise | | |
|---|---|---|---|---|---|---|
| | $\sigma = 60$ | $\sigma = 80$ | $\sigma = 90$ | $\alpha = 4$ | $\alpha = 5$ | $\alpha = 6$ |
| Kodak24 | | | | | | |
| SADNet (Chang et al. 2020) | 28.58/0.8236 | 26.76/0.7703 | 25.97/0.7409 | 28.20/0.8138 | 26.77/0.7711 | 25.53/0.7227 |
| SwinIR (Liang et al. 2021) | 29.11/0.8368 | 27.22/0.7866 | 26.33/0.7560 | 28.71/0.8272 | 27.21/0.7859 | 25.78/0.7328 |
| Uformer (Wang et al. 2022) | 28.97/0.8326 | 26.83/0.7567 | 25.81/0.7075 | 28.53/0.8202 | 26.82/0.7563 | 25.22/0.6752 |
| Restormer (Zamir et al. 2022) | 28.21/0.7837 | 25.71/0.6839 | 24.45/0.6257 | 27.74/0.7670 | 25.75/0.6847 | 23.74/0.5904 |
| NAFNet (Chen et al. 2022) | 26.75/0.7182 | 25.06/0.6563 | 24.36/0.6297 | 26.38/0.7040 | 25.07/0.6557 | 23.99/0.6144 |
| GRL (Li et al. 2023) | 29.08/0.8360 | 27.23/0.7900 | 26.42/0.7682 | 28.68/0.8265 | 27.22/0.7897 | 25.96/0.7549 |
| CGNet (Ghasemabadi et al. 2024) | 27.08/0.7213 | 25.02/0.6411 | 24.19/0.6068 | 26.61/0.7017 | 25.01/0.6383 | 23.72/0.5848 |
| Xformer (Zhang et al. 2024) | 27.84/0.7554 | 24.74/0.6145 | 23.28/0.5466 | 27.25/0.7298 | 24.74/0.6132 | 22.46/0.5065 |
| MHNet (Gao et al. 2025) | 27.11/0.7399 | 24.43/0.6212 | 23.10/0.5559 | 26.56/0.7157 | 24.42/0.6183 | 22.32/0.5149 |
| AKDT (Brateanu et al. 2025) | 27.49/0.7389 | 25.12/0.6382 | 24.09/0.5923 | 26.99/0.7175 | 25.11/0.6362 | 23.51/0.5637 |
| ESC (Lee, Yun, and Ro 2025) | 28.46/0.7851 | 26.08/0.6850 | 25.01/0.6351 | 27.98/0.7656 | 26.10/0.6834 | 24.41/0.6040 |
| CRWKV (Chen et al. 2025) | 26.50/0.7053 | 23.29/0.5742 | 21.87/0.5139 | 25.86/0.6785 | 23.33/0.5721 | 21.11/0.4782 |
| Ours | **29.35/0.8463** | **27.61/0.8096** | **26.86/0.7933** | **28.97/0.8385** | **27.60/0.8094** | **26.42/0.7838** |
| Urban100 | | | | | | |
| SADNet (Chang et al. 2020) | 26.06/0.8468 | 23.95/0.7910 | 23.08/0.7617 | 25.68/0.8380 | 24.03/0.7932 | 22.67/0.7458 |
| SwinIR (Liang et al. 2021) | 27.37/0.8794 | 25.10/0.8264 | 24.11/0.7955 | 26.90/0.8699 | 25.11/0.8266 | 23.56/0.7750 |
| Uformer (Wang et al. 2022) | 27.05/0.8699 | 24.56/0.7956 | 23.47/0.7510 | 26.58/0.8583 | 24.62/0.7971 | 22.92/0.7253 |
| Restormer (Zamir et al. 2022) | 26.54/0.8294 | 23.82/0.7419 | 22.56/0.6927 | 26.08/0.8172 | 23.94/0.7470 | 21.98/0.6675 |
| NAFNet (Chen et al. 2022) | 25.43/0.7915 | 23.49/0.7328 | 22.69/0.7047 | 25.07/0.7817 | 23.55/0.7351 | 22.30/0.6901 |
| GRL (Li et al. 2023) | 27.34/0.8787 | 25.09/0.8277 | 24.15/0.8021 | 26.87/0.8690 | 25.11/0.8279 | 23.65/0.7868 |
| CGNet (Ghasemabadi et al. 2024) | 25.92/0.8019 | 23.69/0.7330 | 22.77/0.6999 | 25.49/0.7901 | 23.75/0.7352 | 22.32/0.6829 |
| Xformer (Zhang et al. 2024) | 26.26/0.8078 | 23.16/0.6950 | 21.80/0.6386 | 25.71/0.7906 | 23.26/0.6987 | 21.15/0.6096 |
| MHNet (Gao et al. 2025) | 25.81/0.8121 | 23.01/0.7076 | 21.70/0.6474 | 25.30/0.7961 | 23.09/0.7108 | 21.03/0.6150 |
| AKDT (Brateanu et al. 2025) | 25.98/0.7973 | 23.59/0.7164 | 22.58/0.6774 | 25.53/0.7835 | 23.66/0.7187 | 22.08/0.6568 |
| ESC (Lee, Yun, and Ro 2025) | 26.53/0.8206 | 23.86/0.7307 | 22.74/0.6876 | 26.05/0.8059 | 23.96/0.7338 | 22.21/0.6655 |
| CRWKV (Chen et al. 2025) | 24.41/0.7439 | 21.42/0.6347 | 20.17/0.5845 | 23.88/0.7253 | 21.54/0.6383 | 19.60/0.5594 |
| Ours | **27.83/0.8944** | **25.74/0.8584** | **24.84/0.8405** | **27.40/0.8877** | **25.76/0.8589** | **24.36/0.8305** |
| CBSD68 | | | | | | |
| SADNet (Chang et al. 2020) | 27.02/0.8109 | 25.24/0.7532 | 24.50/0.7243 | 26.70/0.8011 | 25.31/0.7556 | 24.15/0.7093 |
| SwinIR (Liang et al. 2021) | 28.05/0.8367 | 26.21/0.7864 | 25.36/0.7580 | 27.67/0.8272 | 26.21/0.7861 | 24.86/0.7376 |
| Uformer (Wang et al. 2022) | 27.74/0.8274 | 25.67/0.7559 | 24.72/0.7124 | 27.33/0.8145 | 25.69/0.7552 | 24.19/0.6843 |
| Restormer (Zamir et al. 2022) | 26.74/0.7652 | 24.32/0.6738 | 23.19/0.6258 | 26.28/0.7488 | 24.38/0.6757 | 22.59/0.5984 |
| NAFNet (Chen et al. 2022) | 25.04/0.6973 | 23.41/0.6347 | 22.77/0.6084 | 24.66/0.6823 | 23.40/0.6330 | 22.41/0.5917 |
| GRL (Li et al. 2023) | 28.00/0.8344 | 26.18/0.7848 | 25.40/0.7614 | 27.62/0.8245 | 26.19/0.7849 | 24.97/0.7476 |
| CGNet (Ghasemabadi et al. 2024) | 25.81/0.7191 | 23.83/0.6454 | 23.03/0.6132 | 25.39/0.7032 | 23.85/0.6450 | 22.61/0.5943 |
| Xformer (Zhang et al. 2024) | 26.52/0.7474 | 23.66/0.6260 | 22.37/0.5674 | 25.96/0.7251 | 23.68/0.6262 | 21.66/0.5336 |
| MHNet (Gao et al. 2025) | 25.72/0.7292 | 23.20/0.6241 | 23.20/0.6241 | 25.22/0.7085 | 23.22/0.6234 | 21.29/0.5317 |
| AKDT (Brateanu et al. 2025) | 26.09/0.7295 | 23.82/0.6379 | 22.87/0.5962 | 25.62/0.7108 | 23.84/0.6377 | 22.36/0.5721 |
| ESC (Lee, Yun, and Ro 2025) | 27.13/0.7766 | 24.93/0.6908 | 23.96/0.6488 | 26.70/0.7598 | 24.96/0.6902 | 23.43/0.6229 |
| CRWKV (Chen et al. 2025) | 24.17/0.6607 | 21.14/0.5363 | 19.89/0.4813 | 23.49/0.6323 | 21.11/0.5322 | 19.17/0.4461 |
| Ours | **28.17/0.8407** | **26.44/0.7978** | **25.70/0.7786** | **27.81/0.8320** | **26.46/0.7980** | **25.31/0.7679** |

Table 5: Quantitative denoising comparisons (PSNR/SSIM) on the different datasets. The best results are marked in **bold**, while the second-best ones are underlined.

*Mixture Noise:* $Y = X + \sigma \cdot \sqrt{X} \odot N + \alpha \cdot N_p$, where $\sigma = 90$ and $\alpha = 6$. Its noise level map and base noise are more complex. *4) Variants of Speckle Noise.* To further enhance the level map non-uniformity of Speckle noise, we adopt spatial functions (sine-cosine, peaks, and Gaussian kernels) to construct its variants (Variant 1-3). The form of the variant is $Y_{ij} = X_{ij} + \phi_{ij} \cdot \sqrt{X_{ij}} \cdot N_{ij}$, where $\phi_{ij}$ represents the local scaling factor generated by the spatial position $(i, j)$. The overall value of $\phi$ is normalized within the range of $[60, 120]$. *5) Extension of base noise.* We replace the $N$ in the Speckle noise with the $N_l$ sampled from the Laplace distribution, expanding the evaluation of OOD robustness.

**Feature Analysis.** We provide additional examples of feature comparison. As shown in Fig. 5, our scale-equivariant network consistently captures highly similar features, even when faced with significant differences in global corruption intensity and inconsistent spatial variation patterns in the level map. Conversely, networks that violate this design principle are overly sensitive to such variations. For instance, networks incorporating scale-independent Layer Norm or exponential activations (w/o HNM+ELU) produce markedly dissimilar feature maps. Furthermore, while these non-equivariant designs may yield usable features at low noise levels, they severely distort the network's representa-

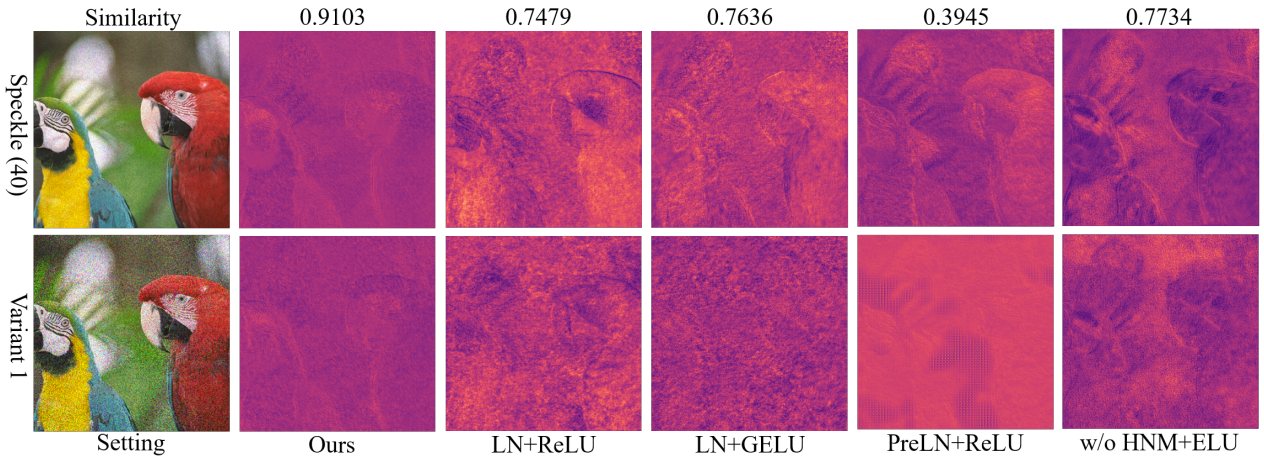
| Method | Speckle Noise* | | | Speckle Noise variant* | | |
|---|---|---|---|---|---|---|
| | $\sigma = 60$ | $\sigma = 80$ | $\sigma = 90$ | Variant 1 | Variant 2 | Variant 3 |
| Kodak24 | | | | | | |
| SADNet (Chang et al. 2020) | 26.83/0.7717 | 25.15/0.7064 | 24.41/0.6701 | 24.23/0.6663 | 24.45/0.6730 | 23.92/0.6471 |
| SwinIR (Liang et al. 2021) | 27.36/0.7871 | 25.40/0.7147 | 24.49/0.6725 | 24.15/0.6625 | 24.51/0.6752 | 23.83/0.6441 |
| Uformer (Wang et al. 2022) | 26.92/0.7501 | 24.74/0.6414 | 23.81/0.5901 | 23.51/0.5845 | 23.86/0.5965 | 23.24/0.5654 |
| Restormer (Zamir et al. 2022) | 25.77/0.6786 | 23.21/0.5588 | 22.12/0.5065 | 21.40/0.5024 | 21.97/0.5160 | 21.08/0.4791 |
| NAFNet (Chen et al. 2022) | 25.30/0.6619 | 23.92/0.6122 | 23.35/0.5919 | 22.86/0.5785 | 23.21/0.5906 | 22.66/0.5665 |
| GRL (Li et al. 2023) | 27.29/0.7845 | 25.52/0.7358 | 24.78/0.7138 | 24.59/0.7091 | 24.82/0.7139 | 24.32/0.7014 |
| CGNet (Ghasemabadi et al. 2024) | 25.13/0.6382 | 23.55/0.5727 | 22.94/0.5467 | 22.44/0.5422 | 22.82/0.5539 | 22.17/0.5219 |
| Xformer (Zhang et al. 2024) | 24.95/0.6188 | 21.96/0.4765 | 20.79/0.4214 | 20.42/0.4198 | 20.80/0.4277 | 20.05/0.4031 |
| MHNet (Gao et al. 2025) | 24.41/0.6095 | 21.75/0.4741 | 20.63/0.4181 | 20.00/0.4132 | 20.53/0.4248 | 19.70/0.3960 |
| AKDT (Brateanu et al. 2025) | 25.23/0.6368 | 23.16/0.5437 | 22.34/0.5057 | 21.95/0.5040 | 22.33/0.5125 | 21.66/0.4875 |
| ESC (Lee, Yun, and Ro 2025) | 25.95/0.6752 | 23.77/0.5722 | 22.90/0.5282 | 22.54/0.5227 | 22.91/0.5323 | 22.24/0.5075 |
| CRWKV (Chen et al. 2025) | 21.95/0.4964 | 19.36/0.3847 | 18.37/0.3416 | 17.90/0.3385 | 18.30/0.3429 | 17.58/0.3261 |
| Ours | **27.69/0.8088** | **26.02/0.7730** | **25.31/0.7579** | **25.13/0.7557** | **25.35/0.7576** | **24.87/0.7509** |
| Urban100 | | | | | | |
| SADNet (Chang et al. 2020) | 24.03/0.7909 | 22.30/0.7274 | 21.61/0.6954 | 21.44/0.6904 | 21.64/0.6978 | 21.40/0.6859 |
| SwinIR (Liang et al. 2021) | 25.32/0.8264 | 23.19/0.7531 | 22.31/0.7150 | 22.09/0.7076 | 22.32/0.7174 | 22.05/0.7053 |
| Uformer (Wang et al. 2022) | 24.75/0.7940 | 22.48/0.6964 | 21.58/0.6519 | 21.35/0.6467 | 21.59/0.6569 | 21.34/0.6440 |
| Restormer (Zamir et al. 2022) | 23.89/0.7386 | 21.35/0.6349 | 20.32/0.5882 | 19.84/0.5822 | 20.21/0.5942 | 19.81/0.5783 |
| NAFNet (Chen et al. 2022) | 23.68/0.7357 | 22.06/0.6778 | 21.41/0.6519 | 21.07/0.6383 | 21.32/0.6524 | 21.04/0.6371 |
| GRL (Li et al. 2023) | 25.21/0.8224 | 23.24/0.7648 | 22.46/0.7385 | 22.29/0.7342 | 22.49/0.7389 | 22.24/0.7322 |
| CGNet (Ghasemabadi et al. 2024) | 23.83/0.7333 | 22.01/0.6660 | 21.31/0.6372 | 20.97/0.6275 | 21.24/0.6414 | 20.90/0.6232 |
| Xformer (Zhang et al. 2024) | 23.40/0.7013 | 20.63/0.5836 | 19.58/0.5366 | 19.31/0.5342 | 19.58/0.5425 | 19.25/0.5305 |
| MHNet (Gao et al. 2025) | 23.01/0.7000 | 20.37/0.5776 | 19.32/0.5270 | 18.86/0.5211 | 19.24/0.5327 | 18.86/0.5181 |
| AKDT (Brateanu et al. 2025) | 23.74/0.7181 | 21.70/0.6373 | 20.90/0.6032 | 20.63/0.5996 | 20.88/0.6086 | 20.57/0.5955 |
| ESC (Lee, Yun, and Ro 2025) | 23.81/0.7271 | 21.58/0.6371 | 20.73/0.5991 | 20.48/0.5948 | 20.72/0.6025 | 20.41/0.5905 |
| CRWKV (Chen et al. 2025) | 21.45/0.6315 | 19.00/0.5289 | 18.08/0.4880 | 17.70/0.4843 | 18.02/0.4921 | 17.61/0.4783 |
| Ours | **25.91/0.8593** | **24.00/0.8193** | **23.23/0.8005** | **23.06/0.7975** | **23.27/0.8005** | **23.01/0.7959** |
| CBSD68 | | | | | | |
| SADNet (Chang et al. 2020) | 25.31/0.7543 | 23.80/0.6912 | 23.16/0.6590 | 23.02/0.6554 | 23.19/0.6624 | 22.94/0.6510 |
| SwinIR (Liang et al. 2021) | 26.36/0.7855 | 24.50/0.7176 | 23.66/0.6805 | 23.44/0.6732 | 23.64/0.6826 | 23.37/0.6695 |
| Uformer (Wang et al. 2022) | 25.79/0.7500 | 23.77/0.6535 | 22.91/0.6082 | 22.69/0.6031 | 22.91/0.6136 | 22.63/0.5998 |
| Restormer (Zamir et al. 2022) | 24.43/0.6703 | 22.13/0.5682 | 21.15/0.5223 | 20.63/0.5166 | 21.00/0.5294 | 20.49/0.5077 |
| NAFNet (Chen et al. 2022) | 23.52/0.6344 | 22.26/0.5809 | 21.77/0.5593 | 21.43/0.5500 | 21.64/0.5603 | 21.31/0.5434 |
| GRL (Li et al. 2023) | 26.26/0.7783 | 24.56/0.7260 | 23.85/0.7024 | 23.72/0.6985 | 23.86/0.7023 | 23.64/0.6976 |
| CGNet (Ghasemabadi et al. 2024) | 23.97/0.6446 | 22.42/0.5796 | 21.82/0.5528 | 21.49/0.5485 | 21.71/0.5589 | 21.32/0.5359 |
| Xformer (Zhang et al. 2024) | 23.90/0.6302 | 21.23/0.5053 | 20.18/0.4546 | 19.90/0.4511 | 20.16/0.4603 | 19.81/0.4493 |
| MHNet (Gao et al. 2025) | 23.21/0.6153 | 20.74/0.4933 | 19.70/0.4416 | 19.21/0.4357 | 19.56/0.4467 | 19.14/0.4339 |
| AKDT (Brateanu et al. 2025) | 23.97/0.6375 | 22.04/0.5514 | 21.28/0.5163 | 21.01/0.5140 | 21.24/0.5217 | 20.88/0.5091 |
| ESC (Lee, Yun, and Ro 2025) | 24.87/0.6843 | 22.88/0.5950 | 22.08/0.5558 | 21.84/0.5504 | 22.07/0.5604 | 21.67/0.5440 |
| CRWKV (Chen et al. 2025) | 21.04/0.5211 | 18.57/0.4077 | 17.63/0.3637 | 17.29/0.3604 | 17.56/0.3674 | 17.18/0.3586 |
| Ours | **26.55/0.7974** | **24.93/0.7549** | **24.25/0.7365** | **24.13/0.7337** | **24.27/0.7358** | **24.05/0.7345** |

Table 6: Quantitative denoising comparisons (PSNR/SSIM) on the different datasets. The best results are marked in **bold**, while the second-best ones are underlined. "*" indicates that the base noise follows the standard Laplacian distribution.

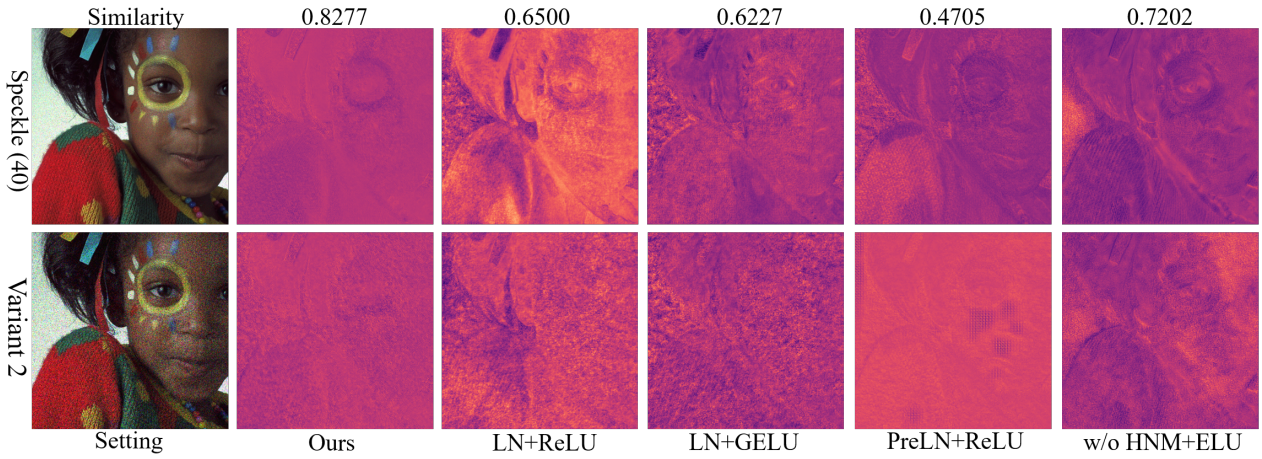
tions at high levels. particularly, placing Layer Norm at the front of a block (PreLN+ReLU) results in more severe feature distortion under high noise compared to putting it in the middle (LN+ReLU). In short, scale equivariance enables the network to generate aligned feature maps that are decoupled from the noise level map, thereby laying the foundation for robust OOD denoising.

**Quantitative Comparison.** We provide more detailed quantitative comparisons for Speckle and Poisson noise. As shown in Tab. 5, our scale-equivariant network consistently achieves superior OOD denoising performance across various corruption levels. Conversely, existing methods, which violate the principle of scale equivariance, suffer from a lack of robustness, even when equipped with advanced modules. Furthermore, when the random distribution used to synthesize Speckle noise is replaced from Gaussian to Laplacian, our scale-equivariant network still demonstrates OOD denoising advantage, as shown in Tab. 6. This indicates its robustness to changes in the base noise distribution.

**Qualitative Comparison.** As shown in Fig. 6 to Fig. 13, our method produces clean denoised images, restoring clear edges and partial texture information, even at high levels of corruption. In contrast, affected by the non-uniform noise level map, the denoising capabilities of existing methods are




(a) Speckle Noise of $\sigma = 40$ and its variant 1




(b) Speckle Noise of $\sigma = 40$ and its variant 2

Figure 5: Comparison in visual and cosine similarities of feature maps under different noise level maps.

| Method | Params (M) | MACs (G) | Times (ms) | PSNR/SSIM |
|---|---|---|---|---|
| SADNet | 4.23 | 18.97 | 0.8 | 25.70/0.7238 |
| SwinIR | 3.14 | 204.99 | 53.1 | 25.92/0.7325 |
| Uformer | 20.63 | 41.55 | 5.1 | 25.44/0.6844 |
| Restormer | 26.13 | 141.24 | 4.8 | 23.67/0.6038 |
| NAFNet | 29.16 | 16.25 | 3.4 | 23.92/0.6161 |
| GRL | 5.11 | 316.57 | 10.4 | 26.18/0.7593 |
| CGNet | 185.38 | 115.08 | 5.0 | 23.75/0.5956 |
| Xformer | 25.22 | 143.12 | 33.0 | 24.36/0.5898 |
| MHNet | 67.54 | 127.40 | 6.7 | 23.98/0.5878 |
| AKDT | 11.43 | 56.89 | 0.7 | 25.08/0.6292 |
| ESC | 0.94 | 51.59 | 0.5 | 24.61/0.6206 |
| CRWKV | 21.08 | 113.55 | 31.2 | 23.00/0.5585 |
| Ours | 31.32 | 158.69 | 1.4 | 26.62/0.7889 |
| Ours(lw) | 16.28 | 76.50 | 0.7 | 26.56/0.7857 |

Table 7: Comparison of different networks in terms of complexity, per-image inference time, and average OOD evaluation on the Kodak24 dataset.

spatially inconsistent and exhibit different patterns of degradation, *e.g.*, coarse results, residual noise, and artifacts.

**Real-World Extension.** The principle of scale equivariance assumes independently distributed noise, a condition not met by spatially correlated real-world camera noise. To bridge this gap, we construct the SEVB-based noise translator trained on synthetic ISP noise. This translator maps correlated noise to a distribution that our pretrained denoiser can process. As shown in Fig. 14 and Fig. 15, denoising networks trained directly on ISP noise fail to generalize, whereas our approach demonstrates outstanding robustness. This suggests that our scale-equivariant network offers a powerful denoising prior, and combining it with noise adaptation provides an effective path for generalizing from synthetic to real data.

**Complexity Comparison** According to Tab. 7, our model offers fast inference, though its complexity is not the lowest. We observe that OOD robustness does not correlate with model complexity, as both more and less complex networks perform poorly. To confirm this, we build a lightweight version of our network that retains comparable robustness with only half the complexity. This result confirms that OOD robustness is primarily determined by the scale-equivariant design, not model capacity.

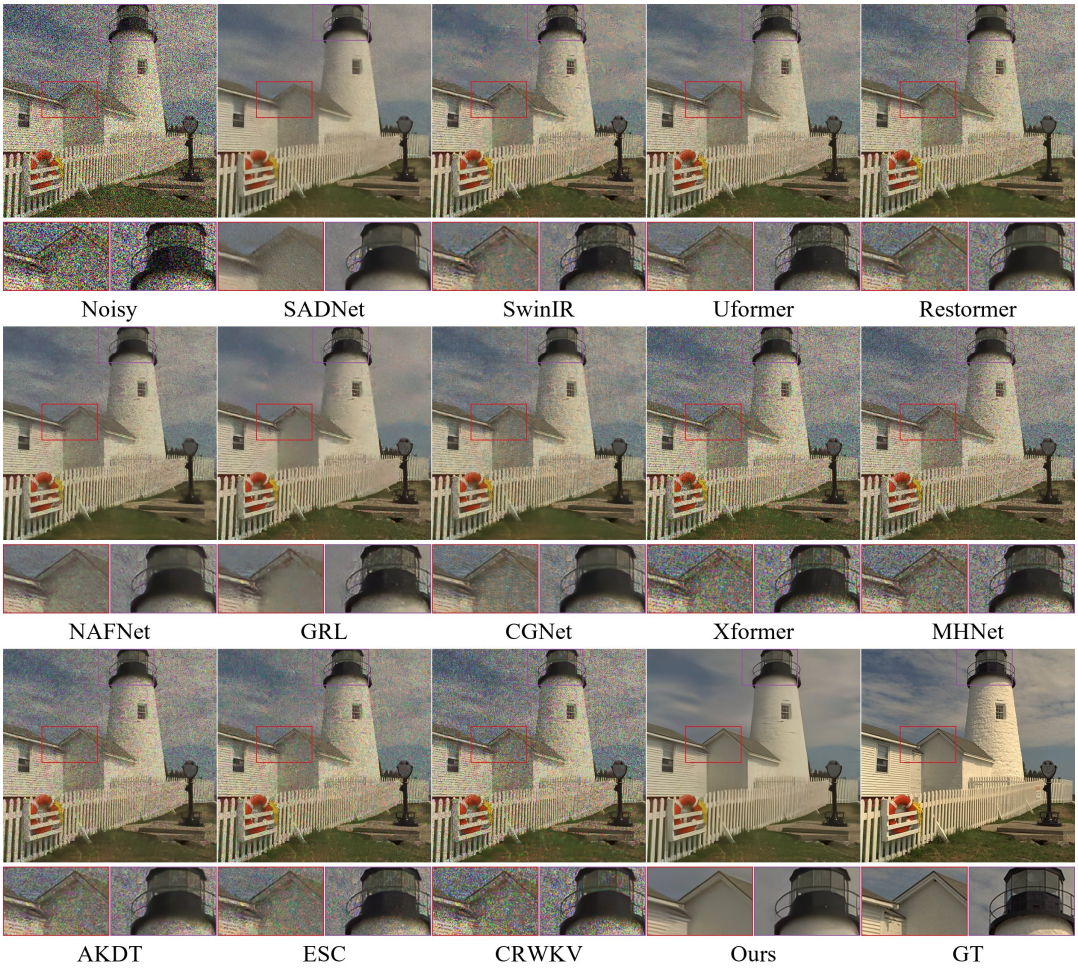



Figure 6: Qualitative comparisons on mixture noise of $\sigma = 90$ and $\alpha = 6$.

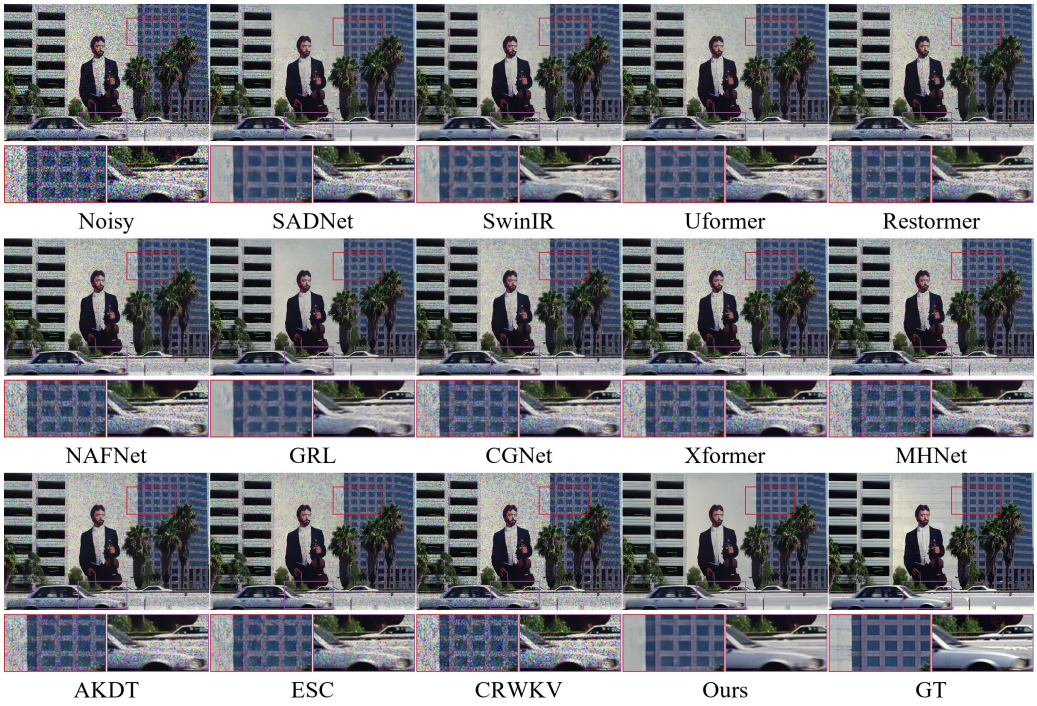



Figure 7: Qualitative comparisons on Poisson noise of $\alpha = 6$.

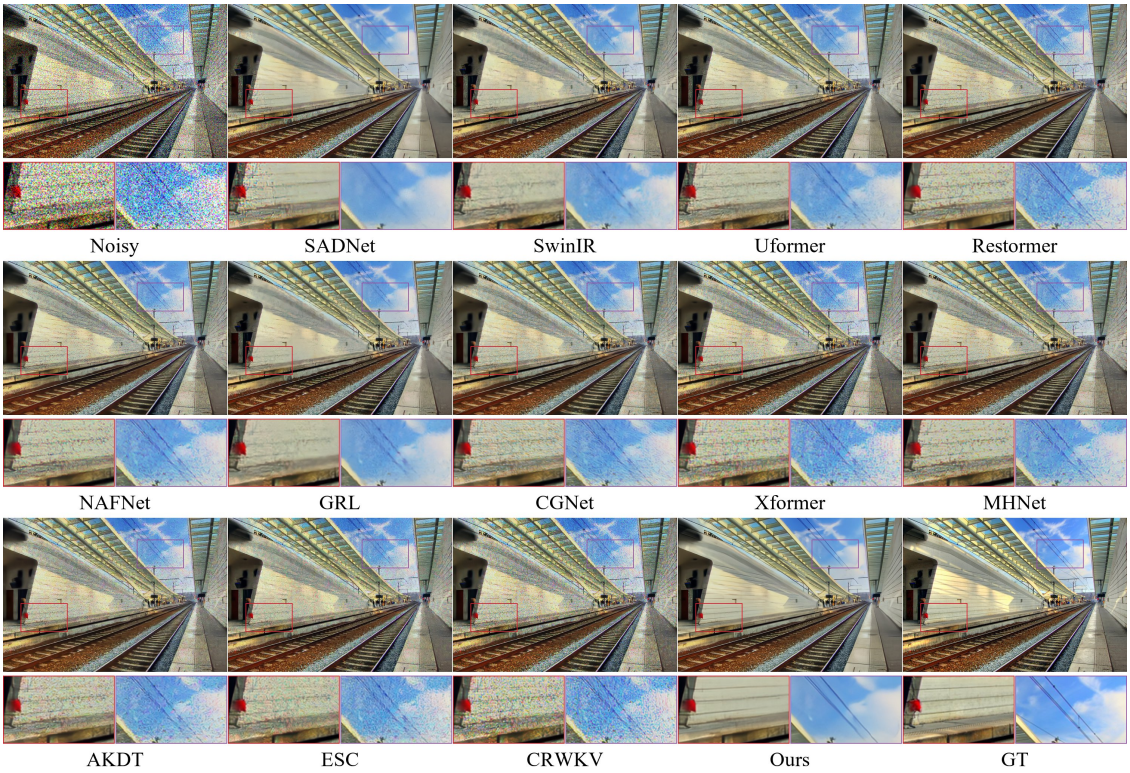



Figure 8: Qualitative comparisons on Poisson noise of $\alpha = 5$.

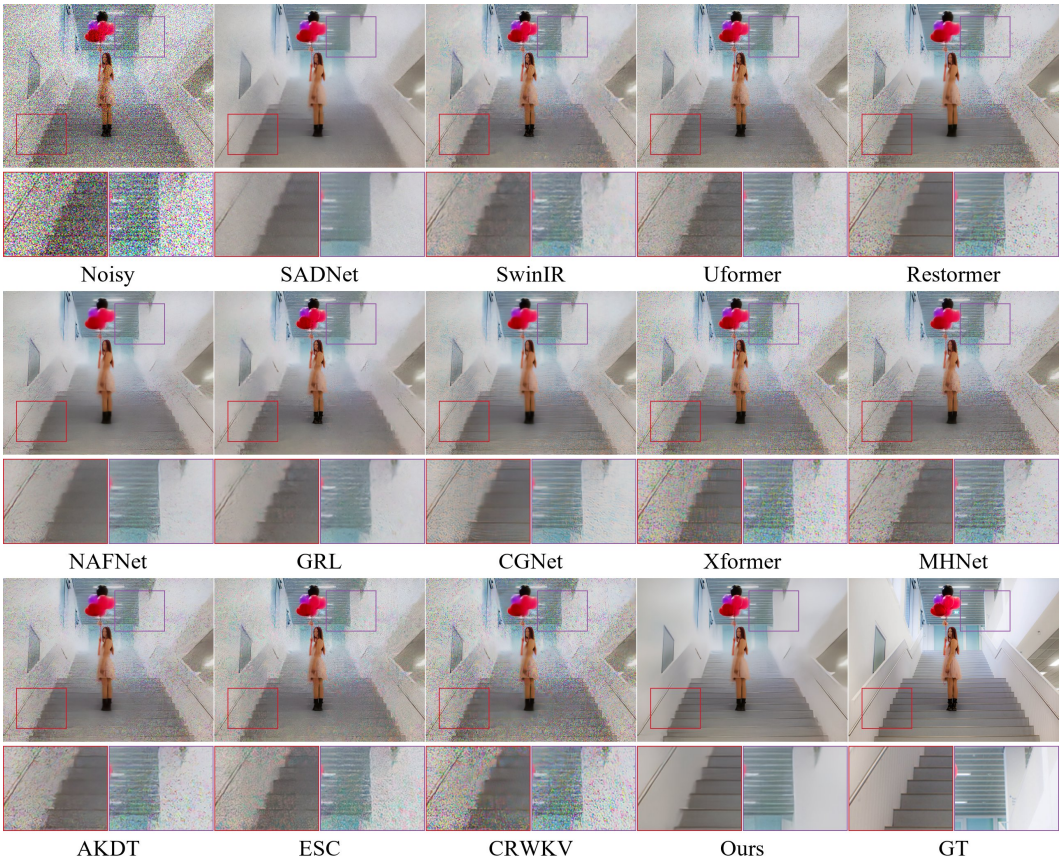



Figure 9: Qualitative comparisons on variant 1 of Speckle noise.

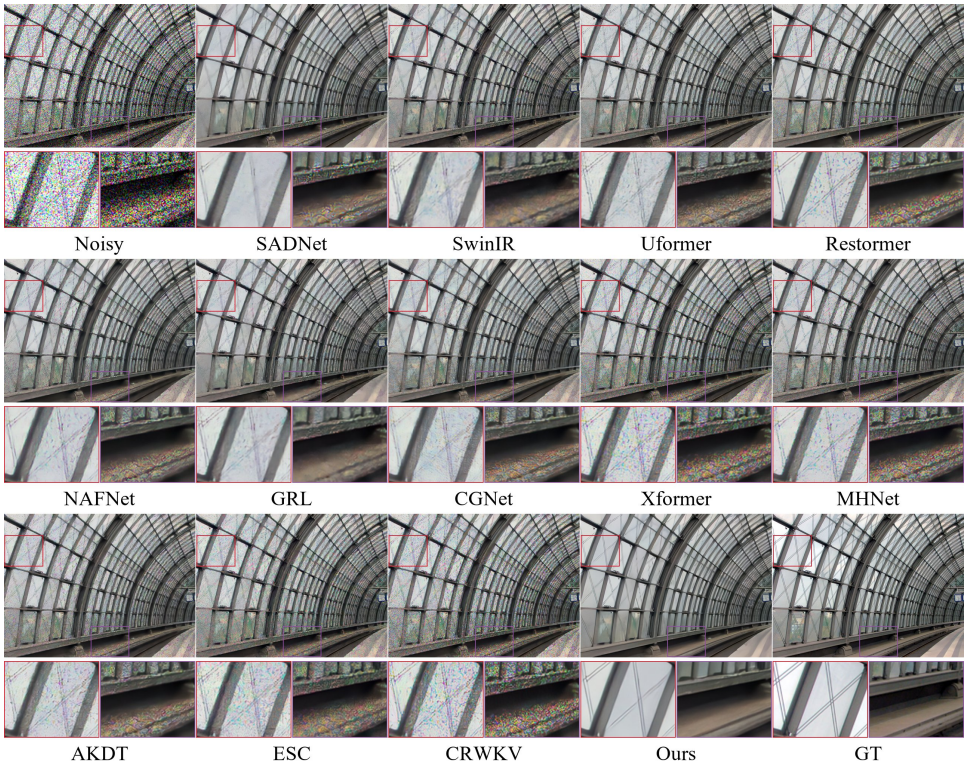



Figure 10: Qualitative comparisons on variant 2 of Speckle noise.

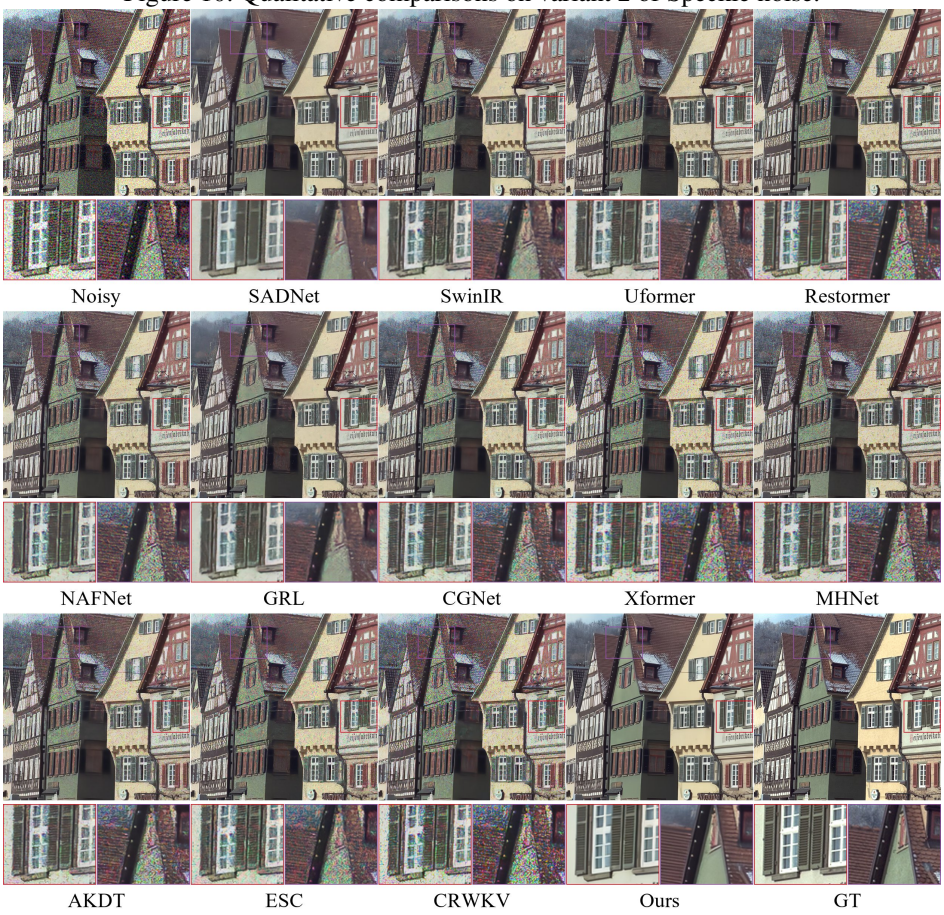



Figure 11: Qualitative comparisons on variant 3 of Speckle noise.

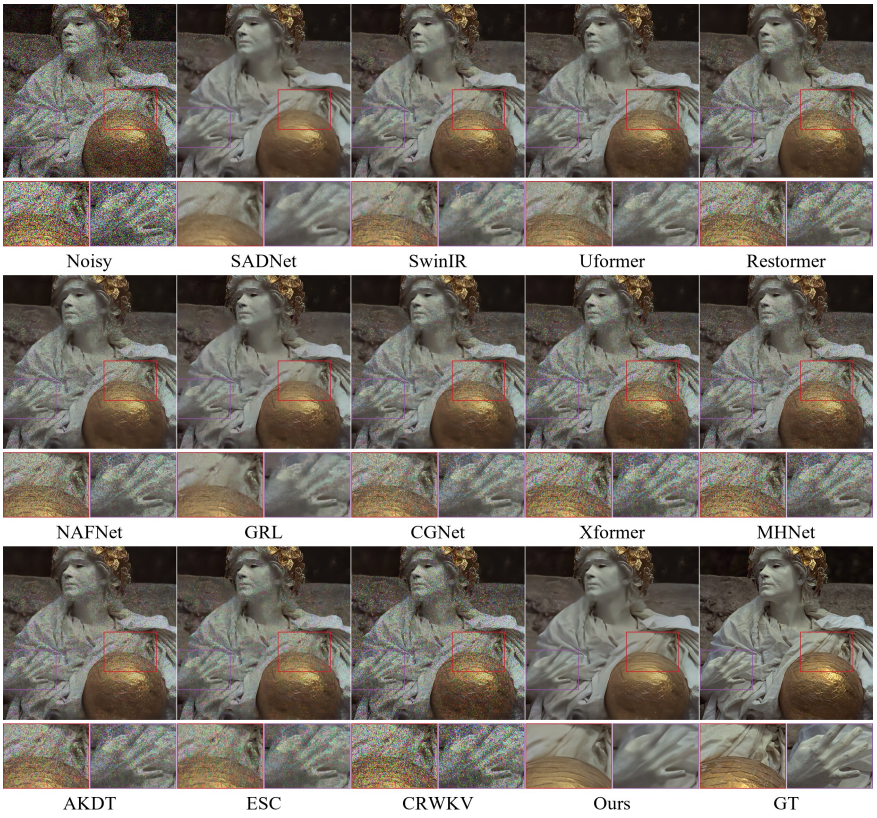



Figure 12: Qualitative comparisons on variant 1 of Speckle noise*.

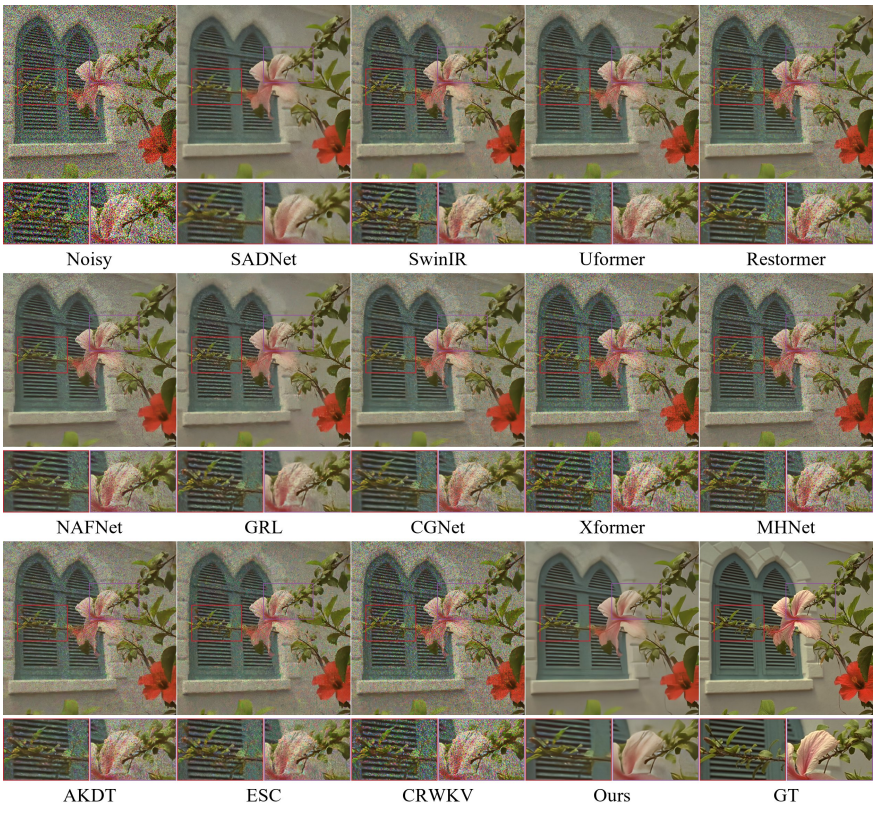



Figure 13: Qualitative comparisons on variant 3 of Speckle noise*.

Figure 14: Qualitative comparisons on CC benchmark.

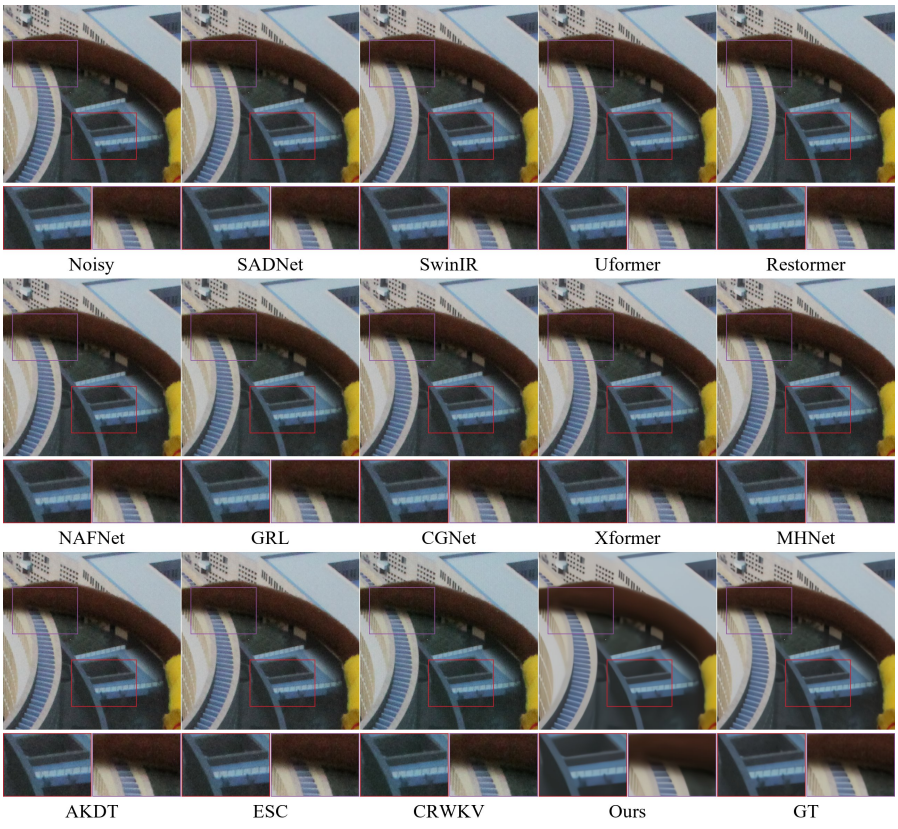



Figure 15: Qualitative comparisons on HighISO benchmark.