# Consistency is Key to Image Denoising

Oindrila Haldar*, Saptarshi Mandal*, Chandra Sekhar Seelamantula
*Department of Electrical Engineering*
*Indian Institute of Science*
Bangalore, India
{oindrilah, saptarshiman}@alum.iisc.ac.in, css@iisc.ac.in

***Abstract*—Without noise, image processing and pattern recognition would be a fait accompli!**
**Image denoising remains a fundamental challenge in practical image processing applications. Deep neural networks (DNNs) have set new performance benchmarks over classical techniques by using supervised learning to predict either the clean image or the noise exclusively. This paper brings both paradigms on the same footing and examines them from the minimum mean-squared error (MMSE) estimation perspective, which requires that their outputs must satisfy a linear constraint, referred to as the consistency criterion. We show that jointly optimizing the image prediction and noise prediction models by enforcing the consistency criterion bridges the gap between the two paradigms and improves denoising performance across multiple settings. Experiments show up to 0.49 dB Peak Signal-to-Noise Ratio (PSNR) gain on CBSD68, Kodak24, McMaster, Urban100 datasets, and 0.46 dB on SIDD dataset, across diverse architectures like DnCNN, SwinIR, CycleISP, and Restormer, under both synthetic and real-world noise.**

***Index Terms*—Image Denoising, Loss functions, Estimation**

## I. Introduction

Noise is ubiquitous and inevitable in any imaging modality, from pinhole cameras to digital sensors. Its presence not only degrades image quality but also hampers downstream computer vision tasks such as classification, segmentation among others. Denoising techniques aim to recover the underlying clean image from a noisy observation, thereby enhancing the perceptual quality of images and videos for an enriched experience as well as accurate downstream processing.

Classical image denoising techniques estimate the underlying clean image by imposing assumptions on the image prior. In contrast, deep learning models can be trained to predict either the clean image or the noise. The distinction between image and noise prediction models is more fundamental than a superficial matter of choosing different training targets. Clean images possess rich semantic structure whereas noise is largely random and has little spatial structure. From the manifold perspective in high dimensions, natural images allow for limited pixel combinations and are therefore embedded in a low-dimensional manifold [1], whereas noise is distributed broadly across the space, as shown in Figure 1. Consequently, the manifold of noisy images can be regarded as a perturbation of the manifold of clean images. The distinction in the manifold structure of image versus noise is what makes denoising possible. From the function learning perspective, although the same loss function may be applied, the resulting loss landscapes differ drastically. Our work exploits the strengths of learning in both landscapes to achieve improved performance.

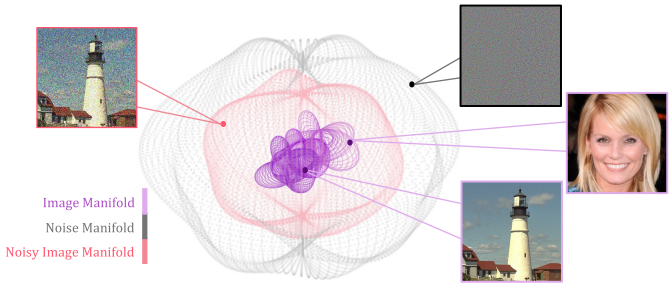


Fig. 1: *Illustration of clean, noisy, and noise manifolds.*

Although both image and noise prediction strategies have been studied fairly extensively, they are usually treated in isolation. As a result, the connection between the two has remained underexplored, particularly with respect to their loss landscapes and manifold structures. This article aims to bridge this gap and provide a unified framework.

### A. Our Contributions

In this work, we propose a *unified framework* that leverages the strengths of optimizing in both the image and noise landscapes to achieve improved denoising performance. While image prediction models aim to directly estimate the clean image from its noisy counterpart, noise prediction models learn to predict the noise component, which is then subtracted from the noisy input. To bridge these two paradigms, we introduce a novel *consistency loss*, which jointly constrains the image and noise prediction models, further improving the peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM). Moreover, given the two fine-tuned models, we introduce a simple method to optimally combine their outputs, thereby producing superior denoising estimates compared to using either model in isolation. Extensive experiments across several datasets and architectures demonstrate that the proposed formulation gives an edge over the image prediction or noise prediction models alone.

## II. Related Works

Extensive research has been conducted to enhance deep learning techniques for image denoising, focusing on novel *network architectures*, improved *loss functions* and alternative

*Equal Contribution

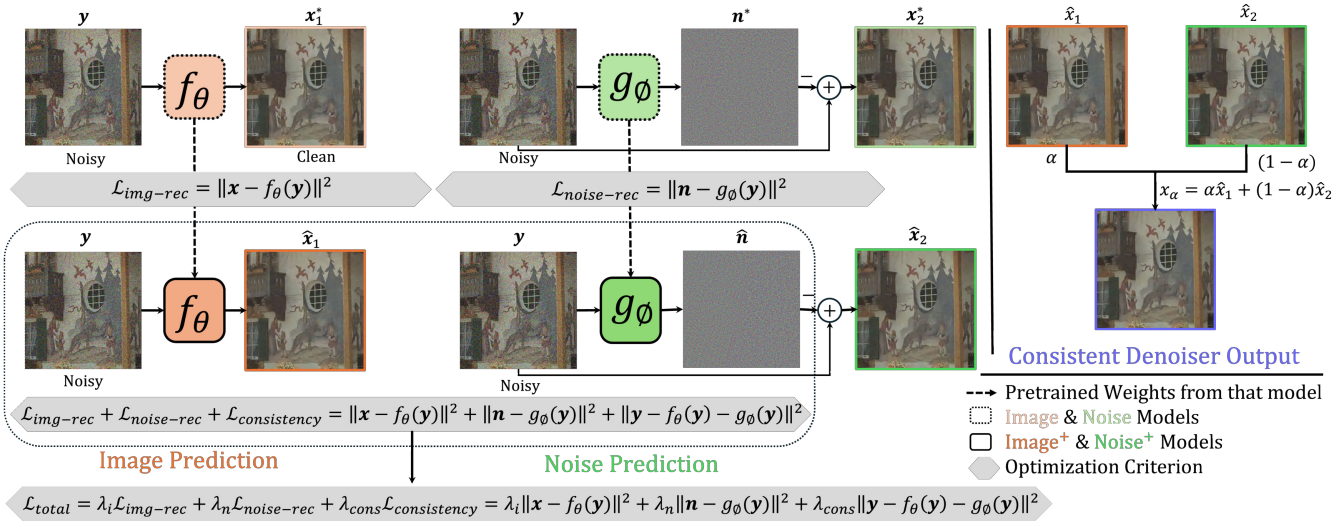


Fig. 2: An overview of the complete consistency-driven denoising pipeline, which consists of the image and noise prediction models, followed by fine-tuning of both models with the consistency criterion. The final Consistent Denoiser output is obtained through an optimal convex combination of their predictions using the learned parameter $\alpha^*$.

*learning paradigms* such as directly predicting the clean image or estimating the noise for subsequent removal.

### A. Architectures for Image Denoising

A wide range of network architectures have been designed to improve image denoising performance, evolving from convolutional models to diffusion-based frameworks.

Early deep learning denoising approaches demonstrated that neural networks could achieve performance comparable to traditional methods such as BM3D. With the rise of Convolutional Neural Networks (CNNs), models such as *DnCNN* [2] achieved SOTA performance by learning residual noise for blind Gaussian denoising, while Mao et al. [3] introduced a fully convolutional encoder-decoder architecture with skip connections.

The success of transformers in NLP and vision, motivated their use in image restoration. *Restormer* [4] leveraged long-range pixel dependencies for image denoising, deblurring, and defocusing, while *Uformer* [5], UNet-styled transformer utilized non-overlapping window-based self attention and multi-scale modulation.

More recently, diffusion models have achieved SOTA performance in image generation by formulating denoising as a progressive refinement process. Since the introduction of *Denoising Diffusion Probabilistic Models (DDPM)* [6], variants such as *DDRM* [7] have extended diffusion to linear inverse problems, with further refinements proposed in [8], [9].

### B. Loss Functions for Image Denoising

Loss functions play a critical role in training deep learning-based denoising models, by shaping the loss landscape and influencing optimization stability and performance. A well-designed loss function smoothens the loss landscape, making it locally convex improving training stability and efficiency.

One of the most widely used training objectives is the *Mean Squared Error (MSE)* [2] which ensures pixel-wise fidelity, but often falls short in capturing perceptual quality. To address this limitation, *Perceptual Loss* [10] compares high-level feature representations extracted from a pre-trained network (typically VGG-16), while *SSIM loss* [11] helps preserve structural details beyond pixel fidelity. Additionally, other specialized objectives such *Total Variation (TV) loss* [12], *Edge-aware losses* [13], *Contextual Loss* [14] help in improving texture consistency and preserving fine details. In this work, we focus more on the *Mean Squared Error (MSE)* loss, which can be used to jointly learn both image and noise.

### C. Image Prediction vs. Noise Prediction

Another design choice in image denoising is the prediction target, i.e., whether the model directly estimates the clean image (**image prediction**) from its noisy counterpart or predicts the noise component (**noise prediction**), which is then subtracted from the noisy input. Classical techniques estimate the underlying clean signal using statistical priors [15] whereas, deep learning based methods have the flexibility to predict either the clean image or the noise, depending on the chosen loss function for training. This approach is commonly employed in residual learning architectures [2]. Recent diffusion models have also leveraged noise prediction for iterative denoising, achieving state-of-the-art performance in both synthetic and real-world noise settings [7], [16]. Unlike methods that rely either on image or noise prediction, our formulation leverages both paradigms, distinguishing it from existing residual-learning, dual-branch, and self-consistency-based denoising approaches [2], [17].

## III. The Consistency Paradigm

Let $\mathbf{x} \in \mathbb{R}^{m \times n}$ denote a clean image and $\mathbf{y} = \mathbf{x} + \mathbf{n}$ its noisy observation, where $\mathbf{n}$ represents additive noise with arbitrary

| Architecture | Method | CBSD68 | Kodak24 | McMaster | Urban100 |
|---|---|---|---|---|---|
| DnCNN [2] | Noise | 28.42 | 29.52 | 28.84 | 27.94 |
| | Image | 28.37 | 29.41 | 28.39 | 27.95 |
| | Noise$^+$ | 28.47 | 29.52 | 28.93 | 28.06 |
| | Image$^+$ | 28.60 | 29.65 | 28.87 | 28.24 |
| | Consistent ($\alpha^* = 0.19$) | **28.62** | **29.69** | **28.98** | **28.26** |
| MIRNet [18] | Noise | 29.69 | 31.20 | 30.87 | 30.40 |
| | Image | 29.66 | 31.18 | 30.74 | 30.27 |
| | Noise$^+$ | 29.72 | 31.25 | 31.00 | 30.62 |
| | Image$^+$ | 29.73 | 31.26 | 30.99 | 30.59 |
| | Consistent ($\alpha^* = 0.54$) | **29.76** | **31.30** | **31.06** | **30.67** |
| CycleISP [19] | Noise | 29.50 | 31.00 | 30.52 | 29.92 |
| | Image | 29.37 | 30.83 | 30.36 | 29.74 |
| | Noise$^+$ | 29.50 | 31.00 | 30.52 | 29.92 |
| | Image$^+$ | 29.52 | 31.01 | 30.55 | 29.95 |
| | Consistent ($\alpha^* = 0.43$) | **29.55** | **31.05** | **30.61** | **30.00** |
| Restormer [4] | Noise | 29.60 | 31.25 | 30.91 | 30.78 |
| | Image | 29.62 | 31.14 | 30.75 | 30.49 |
| | Noise$^+$ | 29.59 | 31.26 | 30.93 | 30.82 |
| | Image$^+$ | 29.66 | 31.29 | 30.88 | 30.77 |
| | Consistent ($\alpha^* = 0.71$) | **29.67** | **31.32** | **30.99** | **30.88** |
| SwinIR [20] | Noise | 29.23 | 30.64 | 30.16 | 29.18 |
| | Image | 29.20 | 30.58 | 30.15 | 29.21 |
| | Noise$^+$ | 29.21 | 30.59 | 30.18 | 29.21 |
| | Image$^+$ | 29.16 | 30.54 | 30.09 | 29.13 |
| | Consistent ($\alpha^* = 0.64$) | **29.25** | **30.66** | **30.25** | **29.26** |

TABLE I: *Comparison of denoising performance (PSNR, dB) under additive white Gaussian noise ($\sigma = 50$) across various methods.* **Bold** indicates the best performance across methods for each architecture.

| Architecture | Method | CBSD68 | Kodak24 | McMaster | Urban100 |
|---|---|---|---|---|---|
| DnCNN [2] | Noise | 30.31 | 28.55 | **25.93** | **25.65** |
| | Image | 29.91 | 28.09 | 24.55 | 25.09 |
| | Noise$^+$ | 30.40 | 28.59 | 25.84 | 25.62 |
| | Image$^+$ | 30.21 | 28.46 | 25.04 | 25.44 |
| | Consistent ($\alpha^* = 0.49$) | **30.41** | **28.60** | 25.61 | 25.60 |
| MIRNet [18] | Noise | 31.18 | **29.40** | 27.00 | 27.03 |
| | Image | 31.10 | 29.33 | 27.00 | 26.94 |
| | Noise$^+$ | 31.19 | 29.34 | **27.08** | 27.13 |
| | Image$^+$ | 31.17 | 29.35 | 26.91 | 27.07 |
| | Consistent ($\alpha^* = 0.57$) | **31.22** | 29.38 | 27.07 | **27.15** |
| CycleISP [19] | Noise | 31.00 | 29.27 | 26.85 | 26.78 |
| | Image | 30.90 | 29.18 | 26.69 | 26.62 |
| | Noise$^+$ | 31.00 | 29.27 | 26.85 | 26.78 |
| | Image$^+$ | 31.02 | 29.27 | 26.87 | 26.77 |
| | Consistent ($\alpha^* = 0.46$) | **31.05** | **29.31** | **26.93** | **26.83** |
| Restormer [4] | Noise | 31.05 | **29.21** | 26.95 | 26.97 |
| | Image | 30.97 | 29.10 | 26.82 | 26.63 |
| | Noise$^+$ | 31.07 | 29.18 | 26.97 | 26.97 |
| | Image$^+$ | 31.03 | 29.08 | 26.99 | 26.76 |
| | Consistent ($\alpha^* = 0.74$) | **31.10** | 29.18 | **27.02** | **26.97** |
| SwinIR [20] | Noise | 30.89 | 29.24 | 26.86 | 26.93 |
| | Image | 30.90 | 29.21 | 26.87 | 26.88 |
| | Noise$^+$ | 30.88 | 29.25 | 26.87 | 26.94 |
| | Image$^+$ | 30.83 | 29.17 | 26.77 | 26.83 |
| | Consistent ($\alpha^* = 0.65$) | **30.93** | **29.28** | **26.92** | **26.97** |

TABLE II: *Comparison of denoising performance (PSNR, dB) under additive white Gaussian noise ($\sigma = 25$) across various methods.* ***Bold*** *indicates the best performance across methods for each architecture.*

distribution.

In standard image prediction, a denoiser $f_\theta(\mathbf{y})$ estimates $\hat{\mathbf{x}}$ by minimizing the loss, $\mathcal{L}(\cdot,\cdot)$ between prediction and ground truth, $\hat{\mathbf{x}} = \arg\min_\theta \mathcal{L}(f_\theta(\mathbf{y}), \mathbf{x})$, leading to the empirical risk minimization given in Eq. (1).

$$\theta^\star = \arg\min_\theta \frac{1}{N}\sum_{i=1}^{N} \mathcal{L}(f_\theta(\mathbf{y}_i), \mathbf{x}_i), \tag{1}$$

for dataset $\mathcal{D} = \{(\mathbf{x}_i, \mathbf{y}_i)\}_{i=1}^N$, where $N$ is the total number of training exemplars.

In the case of noise prediction, the training is geared towards having the model, $g_\phi(\mathbf{y})$ predict the noise $\mathbf{n}_i = \mathbf{y}_i - \mathbf{x}_i$. Following analogously as in the image prediction scenario, the corresponding risk minimization problem is given in Eq. (2).

$$\phi^\star = \arg\min_\phi \frac{1}{N}\sum_{i=1}^{N} \mathcal{L}(g_\phi(\mathbf{y}_i), \mathbf{n}_i). \tag{2}$$

The clean estimate then follows as $\hat{\mathbf{x}} = \mathbf{y} - g_{\phi^\star}(\mathbf{y})$.

We leverage both image- and noise-prediction models by first training them independently and then jointly fine-tuning with a **Consistency Loss** to achieve more accurate estimates of the clean image.

Given the observation model $\mathbf{y} = \mathbf{x} + \mathbf{n}$, the MMSE estimators of $\mathbf{x}$ and $\mathbf{n}$ are given by conditional expectations $\mathbb{E}[\mathbf{x}|\mathbf{y}]$ and $\mathbb{E}[\mathbf{n}|\mathbf{y}]$, respectively, which must satisfy the following *consistency* criterion:

$$\mathbb{E}[\mathbf{y}|\mathbf{y}] = \mathbb{E}[\mathbf{x}|\mathbf{y}] + \mathbb{E}[\mathbf{n}|\mathbf{y}] \Rightarrow \mathbf{y} = \mathbb{E}[\mathbf{x}|\mathbf{y}] + \mathbb{E}[\mathbf{n}|\mathbf{y}].$$

Effectively, this is a *partitioning* of the conditional expectations. When trained separately with the MSE loss, the image and the noise model capture their respective conditional expectations, $\hat{\mathbf{x}} = f_{\theta^\star}(\mathbf{y}) \approx \mathbb{E}[\mathbf{x}|\mathbf{y}]$ and $\hat{\mathbf{n}} = g_{\phi^\star}(\mathbf{y}) \approx \mathbb{E}[\mathbf{n}|\mathbf{y}]$, resulting in

$$\mathbf{y} = \hat{\mathbf{x}} + \hat{\mathbf{n}}. \tag{3}$$

When the image and noise prediction models are trained independently, the consistency criterion Eq. (3) is not guaranteed! Consistency is enforced explicitly by means of a new loss (given in Eq. (4)

$$\mathcal{L}_{\text{consistency}} = \mathbb{E}[\|\mathbf{y} - f_\theta(\mathbf{y}) - g_\phi(\mathbf{y})\|^2], \tag{4}$$

which, in practice, will be approximated by its sample average over the dataset.

The proposed strategy is to train the image and noise prediction models separately and subsequently fine-tune both of them using the renewed loss given in Eq. (5)

$$\begin{aligned}\mathcal{L}_{\text{total}} = \lambda_i \mathcal{L}(\mathbf{x}, f_\theta(\mathbf{y})) + \lambda_n \mathcal{L}(\mathbf{y} - \mathbf{x}, g_\phi(\mathbf{y})) \\ + \lambda_{\text{cons}} \mathcal{L}_{\text{consistency}}(\mathbf{y}, f_\theta(\mathbf{y}) + g_\phi(\mathbf{y})).\end{aligned} \tag{5}$$

Fine-tuning with the consistency constraint ensures that the models jointly maintain the valid decomposition of the noisy observation into its corresponding clean image and noise constituents as given in Eq. (3).

Upon fine-tuning, we obtain two estimates of the clean image, from the image prediction model, $\hat{\mathbf{x}}_1 = f_\theta(\mathbf{y})$ and noise prediction model, $\hat{\mathbf{x}}_2 = \mathbf{y} - g_\phi(\mathbf{y})$. To further obtain a single estimate, we combine the two predictions through

a convex combination rather than a simple average, $\hat{\mathbf{x}}_\alpha = \alpha\hat{\mathbf{x}}_1 + (1-\alpha)\hat{\mathbf{x}}_2, 0 \leq \alpha \leq 1$. To determine the optimal weight factor $\alpha$, the MSE over the dataset $\{\mathbf{x}_i, i = 1, 2, ..., N\}$ is minimized and the minimization results in the following closed-form expression for the optimal value of $\alpha$ given by $\alpha^*$ in Eq. (6).

$$L(\alpha) = \frac{1}{N}\sum_{i=1}^{N} \|\mathbf{x}^i - \hat{\mathbf{x}}_\alpha^i\|^2$$
$$\alpha^* = \frac{\sum_{i=1}^{N}(\hat{\mathbf{x}}_1^i - \hat{\mathbf{x}}_2^i)^\top(\mathbf{x}^i - \hat{\mathbf{x}}_2^i)}{\sum_{i=1}^{N}\|\hat{\mathbf{x}}_1^i - \hat{\mathbf{x}}_2^i\|^2}. \quad (6)$$

Figure 2 provides an overview of the proposed framework.

## IV. Experimental Results

The experiments were carried out under two scenarios: **Gaussian color image denoising** where, image patches of size $128\times128$ were extracted from the CBSD68 [22], McMaster [23], Kodak24 [24], and Urban100 [25] datasets, with additive Gaussian noise applied at noise levels $\sigma \in \{15, 25, 50\}$ and **Real image denoising**, where paired clean and noisy patches were sampled from the SIDD dataset [21].

The models were trained on 1-8 NVIDIA A6000 (48 GB) GPUs, depending on model size. Both the image and noise prediction models were optimized using AdamW ($\beta_1 = 0.9$, $\beta_2 = 0.999$) with $\ell_2$ loss for 150K iterations. Training employed a cosine annealing schedule, with the learning rate decaying from $10^{-4}$ to $10^{-6}$. A batch size of 8 was used, along with data augmentation via rotations (90°, 180°, 270° counterclockwise) and horizontal/vertical flipping. Subsequently, both models were fine-tuned for an additional 150K iterations using the total loss defined in equation 5, with the same training configuration. The loss weights were set to $\lambda_i = \lambda_n = \lambda_{cons} = \frac{1}{3}$. The fine-tuned models are referred to as Image$^+$ and Noise$^+$, respectively.

Table I, Table II and Table III demonstrates that the novel *consistency* loss improves performance across architectures, with Image prediction achieving gains of up to **+0.49 dB** at $\sigma = 25$ and **+0.48 dB** at $\sigma = 50$. Noise prediction also benefits with improvements up to **+0.10 dB** and **+0.22 dB**, respectively. The convex combination with $\alpha$ further enhances performance in most cases with an average improvement of **0.15 dB**. On real-world datasets such as SIDD [21], improvements of upto **+0.34 dB** for Image Prediction and upto **+0.46 dB** for Noise prediction are observed. Though the gains are achieved at the additional computational cost of training two models, the denoising performance at higher noise levels justify the added cost. Additionally, during inference no extra cost is incurred if only the best fine-tuned model is deployed.

Figure 3 provides a visual comparison across methods. These results show that the gains from the proposed framework diminishes as $\sigma$ decreases, an effect observed across all datasets and architectures. This trend aligns with the manifold hypothesis, which suggests that at lower noise levels, noisy images lie closer to the clean image manifold, limiting the potential for improvement.

| Method | CycleISP | | SwinIR | | Restormer | | MIRNet | |
|---|---|---|---|---|---|---|---|---|
| | PSNR↑ | SSIM↑ | PSNR↑ | SSIM↑ | PSNR↑ | SSIM↑ | PSNR↑ | SSIM↑ |
| Noise | 38.70 | 0.9107 | 37.92 | 0.9053 | 39.38 | 0.9169 | 39.07 | 0.9155 |
| Image | 38.54 | 0.9100 | 38.04 | 0.9061 | 39.27 | 0.9158 | 39.25 | 0.9165 |
| Noise$^+$ | 38.84 | 0.9128 | 38.38 | 0.9088 | 39.50 | 0.9179 | 39.36 | 0.9174 |
| Image$^+$ | 38.78 | 0.9123 | 38.38 | 0.9089 | 39.43 | 0.9175 | 39.32 | 0.9171 |
| Consistent | **38.85** | **0.9130** | **38.43** | **0.9095** | **39.50** | **0.9179** | **39.37** | **0.9174** |

TABLE III: *Denoising performance on the SIDD-Validation dataset (PSNR in dB and SSIM).* **Bold** numbers denote the best performance for each architecture across all methods.

## V. Conclusion and Future Work

In this paper, we revisited image denoising from an MMSE estimation perspective, treating image and noise prediction on equal footing. Comparing their performance using an array of neural network denoisers and several datasets, we observed that the noise prediction network almost always outperforms the image prediction network. From the MMSE estimation perspective, the image and noise predictions must satisfy an additional *consistency* criterion, which we explicitly incorporated into the loss function. Our experiments demonstrate that finetuned models trained with this consistency-based formulation consistently outperform standard image and noise prediction models, highlighting the effectiveness of enforcing consistency for improved generalization. Furthermore, since our formalism does not rely on any specific assumption about the probability distribution of noise, it generalizes seamlessly to real-world scenarios with diverse noise characteristics. An extensive validation in several practical imaging modalities (MRI, CT, low-light photography, etc.) would be an interesting future direction. The role of consistency could also be explored in other linear inverse problems such as deblurring, compressed image recovery, super-resolution, etc. Moreover, its application to downstream tasks such as image generation using diffusion models such as DDPM [6] presents another promising avenue for future work.

## References


[1] Y. Bengio, A. Courville, and P. Vincent, "Representation learning: A review and new perspectives," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 8, pp. 1798–1828, 2013.

[2] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE Transactions on Image Processing*, vol. 26, no. 7, pp. 3142–3155, 2017.

[3] X. Mao, C. Shen, and Y.-B. Yang, "Image restoration using very deep convolutional encoder-decoder networks with symmetric skip connections," *Advances in Neural Information Processing Systems*, vol. 29, 2016.

[4] S. W. Zamir, A. Arora, S. Khan, M. Hayat, F. S. Khan, and M. Yang, "Restormer: Efficient transformer for high-resolution image restoration," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022.

[5] Z. Wang, X. Cun, J. Bao, W. Zhou, J. Liu, and H. Li, "Uformer: A general u-shaped transformer for image restoration," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 17683–17693.

[6] J. Ho, A. Jain, and P. Abbeel, "Denoising diffusion probabilistic models," *Advances in Neural Information Processing Systems*, vol. 33, pp. 6840–6851, 2020.

[7] B. Kawar, M. Elad, S. Ermon, and J. Song, "Denoising diffusion restoration models," *Advances in Neural Information Processing Systems*, vol. 35, pp. 23593–23606, 2022.

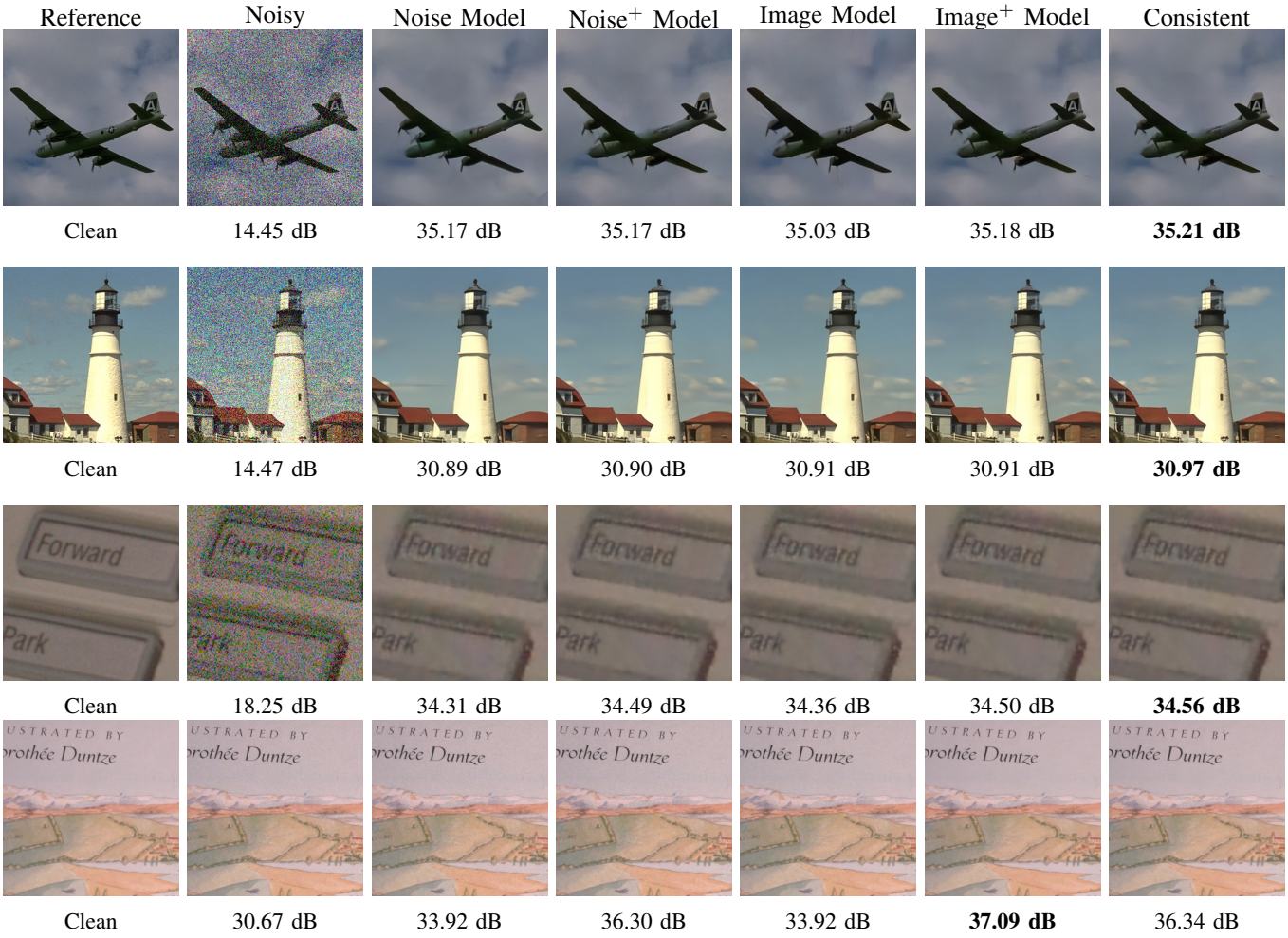


Fig. 3: *Visual comparison of denoising performance across architectures and methods. Rows 1 & 2: CycleISP [19] & Restormer [2] on CBSD68 at $\sigma = 50$ respectively. Rows 3 & 4: SwinIR [20] & Restormer [4] on SIDD [21], respectively.*


[8] M. T. Shahriar and H. Li, "A study of image pre-processing for faster object recognition," *arXiv preprint arXiv:2011.06928*, 2020.

[9] Z. Kadkhodaie and E. Simoncelli, "Stochastic solutions for linear inverse problems using the prior implicit in a denoiser," *Advances in Neural Information Processing Systems*, vol. 34, pp. 13242–13254, 2021.

[10] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *European Conference on Computer Vision*, 2016, pp. 694–711.

[11] Z. Wang et al., "Image quality assessment: from error visibility to structural similarity," *IEEE Transactions on Image Processing*, vol. 13, no. 4, pp. 600–612, 2004.

[12] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: Nonlinear Phenomena*, vol. 60, no. 1-4, pp. 259–268, 1992.

[13] K. Nazeri, E. Ng, T. Joseph, F. Qureshi, and M. Ebrahimi, "Edge-connect: Structure guided image inpainting using edge prediction," in *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, 2019.

[14] R. Mechrez, I. Talmi, and L. Zelnik-Manor, "The contextual loss for image transformation with non-aligned data," in *European Conference on Computer Vision*, 2018, pp. 768–783.

[15] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, "Color image denoising via sparse 3d collaborative filtering with grouping constraint in luminance-chrominance space," in *IEEE International Conference on Image Processing*, 2007, vol. 1, pp. I–313.

[16] H. Chung, B. Sim, and J. C. Ye, "Come-closer-diffuse-faster: Accelerating conditional diffusion models for inverse problems through stochastic contraction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2022, pp. 12413–12422.

[17] et al. S. Liu, "Lg-dbnet: Local and global dual-branch network for sar image denoising," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 62, pp. 1–15, 2024.

[18] S. W. Zamir et al., "Learning enriched features for fast image restoration and enhancement," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 45, no. 2, pp. 1934–1948, 2022.

[19] S. W. Zamir et al., "Cycleisp: Real image restoration via improved data synthesis," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2020, pp. 2696–2705.

[20] J. Liang et al., "Swinir: Image restoration using swin transformer," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021, pp. 1833–1844.

[21] A. Abdelhamed et al., "A high-quality denoising dataset for smartphone cameras," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1692–1700.

[22] D. Martin et al., "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proceedings of the IEEE International Conference on Computer Vision*, 2001.

[23] L. Zhang et al., "Color demosaicking by local directional interpolation and nonlocal adaptive thresholding," *Journal of Electronic Imaging*, vol. 20, no. 2, pp. 023016, 2011.

[24] R. Franzen, "True color kodak images," Online, 1999.

[25] J.-B. Huang et al., "Single image super-resolution from transformed self-exemplars," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 5197–5206.