Image Restoration Learning via Noisy Supervision in Fourier Domain

Haosen Liu, Jiahao Liu, Shan Tan, *Member, IEEE*, and Edmund Y. Lam, *Fellow, IEEE*

***Abstract*—Noisy supervision refers to supervising network learning with targets corrupted by noise, encompassing both weakly supervised learning with noisy targets and fully unsupervised denoising using unpaired noisy images. It alleviates the data collection burden and enhances the practical applicability of deep learning techniques. Existing methods face two main limitations: they are ineffective at handling noise with long-range correlations, commonly found in real-world scenarios such as low-light imaging and remote sensing, and rely on pixel-wise loss functions that offer limited supervision for image deblurring and super-resolution. This work addresses these challenges by leveraging the Fourier domain, where spatially correlated noise exhibits sparsity and independence, and Fourier coefficients capture global information that enables stronger supervision. We prove that Fourier coefficients of a wide range of noise converge in distribution to the Gaussian distribution and establish a statistical equivalence between learning with clean and noisy targets in the Fourier domain. Based on these insights, we develop a weakly supervised framework for image restoration learning with noisy targets, and construct a fully unsupervised denoising method tailored to stripe-wise noise. Extensive experiments show that our approaches achieve superior performance in both quantitative metrics and perceptual quality.**

***Index Terms*—Noisy supervision, weakly supervised learning, unsupervised denoising, image restoration, Fourier domain.**

I. Introduction

IMAGE restoration aims to provide an accurate estimation $\hat{z}$ of a high-quality latent image z from its degraded low-quality measurement x. Traditional image restoration methods rely on hand-designed image prior models. Some representative techniques include the variational methods [1], sparse coding [2], collaborative filtering [3], and the low-rank methods [4], [5]. These methods often have limited performance as their models struggle to effectively capture the complex properties of real-world images. In recent years, deep learning methods have gained increasing attention in image restoration tasks due to their larger model capacity. They have significantly outperformed traditional model-based methods in areas such as image denoising [6], [7], super-resolution [8], [9], and deblurring [10]. However, the outstanding performance of deep learning methods is typically based on supervised learning, which requires access to a set of high-quality images as training targets. Unfortunately, obtaining such images in practice is often challenging or even impossible, limiting the practical applicability of these methods.

In imaging systems, a fundamental trade-off exists between signal-to-noise ratio (SNR) and spatiotemporal resolution. When images are captured at high spatial resolutions or ultra-high speeds, the limited amount of incident light received at each pixel leads to low SNR. This issue becomes even more pronounced in low-light conditions. Consequently, when collecting data, one is faced with a choice between obtaining high-resolution but noisy images or clean but low-resolution images. In many situations, it is generally impractical to collect high-resolution and clean images as training targets. To address this challenge, some approaches propose post-processing the captured high-resolution but noisy images to generate pseudo-clean targets. These methods either estimate the latent clean image from multiple (*e.g.*, 150) noisy measurements of the same scene [11] or apply denoising techniques such as BM3D [3] to remove noise in each noisy observation [12]. However, these post-processing strategies have significant drawbacks. The former approach requires the scene to be static so that the repetitive measurements are feasible, not applicable to dynamic scenes. The latter approach introduces unwanted artifacts and may inadvertently smooth out image details. These limitations call for a paradigm that supervises the network learning directly with noisy images as training targets, thereby avoiding the need for either repeated acquisition or artifact-prone processing.

This noisy supervision paradigm encompasses both weakly supervised learning with noisy targets and unsupervised denoising using unpaired noisy images. A pioneering work in this field is the weakly supervised method Noise2Noise (N2N) [13], which establishes a fundamental statistical equivalence, *i.e.*, if the noise n in the noisy target y is zero-mean and independent of the input x, then

$$\mathbb{E}\left[\|f_\theta(x) - y\|_2^2\right] = \|f_\theta(x) - z\|_2^2 + const, \tag{1}$$

where $\|\cdot\|_2$ denotes the ℓ_2 norm, $f_\theta(\cdot)$ denotes the network with parameters θ, z is the latent clean image, $x = g(z)$

Received 28 August 2025; revised 19 March 2026 and 8 May 2026; accepted 28 May 2026. Date of publication 11 June 2026; date of current version 16 June 2026. This work was supported in part by the Research Grants Council of Hong Kong under Grant GRF 17201822 and in part by the Theme-Based Research Scheme under Grant T45-701/22-R. The associate editor coordinating the review of this article and approving it for publication was Dr. Charles Kervrann. *(Corresponding author: Edmund Y. Lam.)*

Haosen Liu and Edmund Y. Lam are with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong, SAR, China (e-mail: hsliu@eee.hku.hk; elam@eee.hku.hk).

Jiahao Liu and Shan Tan are with the Key Laboratory of Image Processing and Intelligent Control, Ministry of Education, School of Artificial Intelligence and Automation, Huazhong University of Science and Technology, Wuhan 430074, China (e-mail: jiahaoliu_aia@hust.edu.cn; shantan@hust.edu.cn).

This article has supplementary downloadable material available at https://doi.org/10.1109/TIP.2026.3700926, provided by the authors.

Digital Object Identifier 10.1109/TIP.2026.3700926

is the image degraded by the degradation model $g(\cdot)$, and $y = z + n$ is the image corrupted by the zero-mean noise n. Note that the independence between n and x can naturally hold if the degraded image x and the noisy image y are two independent observations. Inspired by N2N, many unsupervised denoising techniques have been proposed by using this equivalence along with additional noise assumptions such as the spatial independence of noise (Noise2Void [14], Noise2Self [15], Self2Self [16], Neighbor2Neighbor [17]) and the prior knowledge of the noise model (Noiser2Noise [18], Recorrupted 2Recorrupted [19]).

However, these existing methods formulate noisy supervision mainly in the spatial domain, which leads to two practical limitations. The first limitation lies in the noise structure. Noise in many real-world imaging systems can exhibit strong spatial correlation [20], [21]. Such correlation can degrade the performance of N2N [13] and undermine the effectiveness of N2N-inspired unsupervised methods, because spatially correlated noise can violate the spatial noise-independence assumptions adopted by many of these methods and make accurate noise modeling more challenging. The second limitation lies in the supervision form. The spatial-domain loss treats each local pixel individually, failing to provide global supervision that can enhance image restoration performance [22]. Addressing both issues requires a noisy-supervision strategy that is robust to structured noise and, at the same time, capable of providing global guidance.

The Fourier domain provides a natural way to meet these requirements. On the noise side, spatially correlated noise often has a sparse and approximately independent representation in the Fourier domain. The most representative example is stripe-wise noise, which is commonly observed in low-light vision [23], remote sensing [24], infrared imaging [25], and microscopy imaging [26]. Although such noise exhibits long-range spatial correlation, in its Fourier-domain representation, the noise energy is typically concentrated on a small subset of approximately independent coefficients, allowing the supervision to rely more on less-contaminated coefficients while suppressing the influence of noise-corrupted ones. On the supervision side, each Fourier coefficient is computed from the entire image and therefore encodes global image information. A loss defined over Fourier coefficients can enhance the recovery of global frequency structures, leading to better image super-resolution [27] and deblurring [28] results. These two properties make the Fourier domain a natural choice for constructing noisy supervision that is robust to structured noise and effective for restoration tasks requiring global supervision.

To formalize this intuition, we first analyze the statistical behavior of noise in the Fourier domain. Similar to using the central limit theorem (CLT) in the analysis of discrete cosine transform coefficients in image compression [29], this work conducts a comprehensive analysis of noise properties in the Fourier domain, showing that the Fourier coefficients of many types of noise converge in distribution to the Gaussian distribution. This enables a unified treatment of different noise models under a common framework. We then investigate the impact of noise on a Fourier-based loss and establish the statistical equivalence between using noisy targets and clean targets for such a loss function. Leveraging this equivalence, we develop an efficient weakly supervised learning framework, termed image restoration learning via noisy supervision in the Fourier domain (IR-NSF). To validate its effectiveness, we apply it to multiple image restoration tasks, diverse network architectures, and various noise models. Overall, IR-NSF achieves superior performance in super-resolution, deblurring, and stripe-wise noise removal.

In addition to enabling weakly supervised learning with noisy targets, the proposed Fourier-based statistical equivalence also facilitates the development of unsupervised techniques. This is demonstrated through a Fourier-based unsupervised stripe removal (USR). The USR generates pseudo training pairs by swapping roughly estimated stripe noise between unpaired noisy images. These pairs are then used to train the network, with the loss computed solely on Fourier coefficients affected by stripe artifacts. This design allows the network to effectively eliminate structured stripe noise while preserving the underlying image content. It can be used independently as a stripe removal or combined with other unsupervised methods for fully unsupervised removal of both stripe-wise and pixel-wise noise.

In summary, our main contributions are:

- *Noise Statistics in the Fourier Domain*: We prove that the Fourier coefficients of a wide range of noise types converge in distribution to the Gaussian distribution, and demonstrate that spatially correlated noise can exhibit sparsity and independence in the Fourier domain.
- *Fourier-based Statistical Equivalence*: We establish a Fourier-based statistical equivalence between learning with noisy targets and clean targets under a mild assumption that noise in the noisy target is zero-mean, independent of the input in the training pair, and has Fourier coefficients that follow Gaussian distributions.
- *Noisy Supervision Methods:* Building on the Fourier-based statistical equivalence, we develop a weakly supervised framework applicable to diverse image restoration tasks, along with an unsupervised method tailored to stripe-wise noise. These methods achieve state-of-the-art results on both synthetic and real-world benchmarks.

The remainder of this paper is organized as follows: Section II reviews related work. Section III analyzes the statistical properties of noise in the Fourier domain, establishes a Fourier-based statistical equivalence between training with noisy and clean targets, and introduces the IR-NSF framework for weakly supervised image restoration. Section IV presents extensive experiments demonstrating the effectiveness of IR-NSF across diverse restoration tasks and benchmarks. Section V applies the proposed equivalence to the unsupervised setting, introducing a novel Fourier-based unsupervised method, USR, for removing stripe-wise noise using unpaired noisy images. Finally, Section VI concludes this manuscript.

II. Related Work

A. Noise Model

The most widely adopted noise model is the additive, independent, and homogeneous Gaussian distribution [30].

However, it is insufficient for effectively modeling the signal-dependent noise such as the Poisson-Gaussian noise. To address this limitation, Foi *et al*. proposed a heteroscedastic Gaussian model [31]. While the heteroscedastic Gaussian model is more realistic than the homogeneous one, it still fails to account for structured noise commonly observed in various applications. For example, recent studies have demonstrated the presence of periodic noise and row noise in images captured under extremely low-light conditions [23]. Similarly, periodic stripes have been observed in whisk-broom imaging systems, while non-periodic stripes occur in push-broom imaging for remote sensing [24]. Notably, [32] emphasizes the importance of reducing stripe noise for the CMOS image sensor, as stripe-wise noise tends to be significantly more visible than pixel-wise noise. Moreover, noise affecting projections in X-ray medical imaging is commonly assumed to be stationary and correlated [33]. To describe these types of noise, Fehrenbach *et al*. [20] and Mäkinen *et al*. [21] proposed a simple yet effective model, *i.e*.,

$$n = h * \eta, \tag{2}$$

where h denotes a correlation kernel, η denotes zero-mean i.i.d. noise, and * denotes the convolution operator. By setting different h and η, this model can represent both the uncorrelated and correlated noise [20], [21].

In this work, we will show that although different types of noise mentioned above follow various distributions in the spatial domain, their Fourier coefficients all approximately follow the Gaussian distribution. This enables a unified analysis of diverse noise under a common framework. Additionally, we will demonstrate that spatially correlated noise can exhibit sparsity and independence in the Fourier domain, highlighting the potential advantages of utilizing it for developing noisy supervision methods.

B. Noisy Supervision for Image Restoration

Existing methods in this field are all developed in the spatial domain, with a primary focus on the denoising task. Some representatives are the N2N and its variants. Most of them are established based on the statistical equivalence given by (1), while differing in the way of constructing input-target pairs (x, y). The N2N collects training pairs from two independent measurements of the same scene. Its variants aim at unsupervised denoising and generate input-target pairs from unpaired noisy images. For instance, Noise2Void [14], Noise2Self [15], Self2Self [16], and Neighbor2Neighbor [17] assume spatial independence of the noise and propose dividing each noisy image into two parts, where one serves as the input and the other as the target. On the other hand, variants such as Noiser2Noise [18] and Recorrupted2Recorrupted [19] assume the given noise model and its parameters, generating input-target pairs by further adding noise to the captured image. However, due to the limitations of their assumptions, these variants cannot effectively handle spatially correlated noise.

Additionally, there are other types of methods explored in this field. For instance, SDA-SURE [34] develops noisy supervision based on Stein's unbiased risk estimator but requires noise to be i.i.d. Gaussian. The AmbientGAN [35] and noise-robust GAN (NR-GAN) [36] make the learning process of a generative adversarial network (GAN) robust to corruption. The SS-GMM [37] learns the image prior with the Gaussian mixture model (GMM) from only a single noisy image. However, these methods rely on extra assumptions, such as prior knowledge of noise model [35], rotation-invariance [36], and i.i.d. Gaussian [37], which can limit their practicality in real-world applications. Another approach, the deep image prior (DIP) [38], utilizes the early-stopping strategy to regularize the generator training. But how to determine the stopping point remains unclear, undermining its practical performance.

In this work, we propose that the Fourier domain is better suited for developing noisy supervision. We establish a Fourier-based statistical equivalence that not only enables a novel weakly supervised learning framework applicable to diverse image restoration tasks and noise models but also opens new avenues for advancing unsupervised denoising techniques.

C. Fourier-Based Supervision

The Fourier transform has long been a fundamental tool in image processing and has recently been integrated into deep learning designs. Chen *et al*. [39] improved the generalization behavior of convolutional neural networks (CNN) for image classification by utilizing the amplitude-phase recombination strategy. This approach was based on the observation that humans rely heavily on phase information for recognition. In contrast, Jiang *et al*. emphasized the necessity of both amplitude and phase for image reconstruction [40]. They proposed a novel method that measures the frequency distance between real and fake images to supervise learning for image reconstruction and synthesis. Similarly, Fuoli *et al*. [27] demonstrated that developing supervision using Fourier amplitude and phase can yield superior perceptual results for image super-resolution (SR). Li et al. [41] validated that combining wavelet-domain Transformer with Fourier supervision can expand the range of pixel utilization for lightweight SR. Furthermore, Fourier-based supervision has also shown promising results in the image deblurring task [28]. Beyond Fourier supervision, some works, such as FourierSR [42], have developed novel network modules using the Fourier transform. However, these approaches typically require clean targets, posing challenges in practical scenarios.

In this work, our goal is to develop efficient Fourier-based supervision directly from noisy targets, which are more readily available. By addressing this challenge, we aim to bridge the gap between the advancements made by these existing methods and their practical applicability.

III. Fourier-Based Statistical Equivalence and Weakly Supervised Image Restoration Learning

Following the introduction to the notation and definitions used in our theoretical analysis, this section proves that the Fourier coefficients of diverse noise converge in distribution to Gaussian distributions and are mutually independent. Building on this, we establish the statistical equivalence between learning with noisy targets and clean targets in the Fourier domain.

Finally, we construct a weakly supervised image restoration framework that leverages noisy targets for training.

A. Notation and Definition

In this work, the latent clean image and the noisy target image are denoted by z and y, respectively. The noise corrupting the target image is defined as $n = y - z$. They are denoted by 2-dimensional variables of the size $U \times V$. We use $[u, v]$ to denote the position indices in the spatial domain, *e.g.*, the noise n at the position $[u, v]$ is denoted by $n[u, v]$. The Fourier transform adopted in this work refers to the 2-dimensional discrete Fourier transform (2D-DFT). For a 2-dimensional variable such as the noise n, its Fourier transform is denoted by $\mathscr{F}(n)$ and is calculated as

$$\mathscr{F}(n)[k,l] = \frac{1}{UV}\sum_{u=0}^{U-1}\sum_{v=0}^{V-1} n[u,v]e^{-j2\pi\left(\frac{ku}{U}+\frac{lv}{V}\right)}, \tag{3}$$

where $[k, l]$ denote the position indices in the Fourier domain. The real and imaginary coefficients of $\mathscr{F}(n)$ are denoted by $a(n)$ and $b(n)$, respectively. They are calculated as

$$a(n)[k,l] = \frac{1}{UV}\sum_{u=0}^{U-1}\sum_{v=0}^{V-1} n[u,v]\cos\left(2\pi\left(\frac{ku}{U}+\frac{lv}{V}\right)\right), \tag{4}$$

and

$$b(n)[k,l] = \frac{-1}{UV}\sum_{u=0}^{U-1}\sum_{v=0}^{V-1} n[u,v]\sin\left(2\pi\left(\frac{ku}{U}+\frac{lv}{V}\right)\right). \tag{5}$$

These equations demonstrate that the Fourier coefficients of a variable are either the average or a weighted sum of its values across all positions, thereby containing the global information of this variable.

B. Noise Analysis in the Fourier Domain

1) Spatially-Independent Noise: Our analysis begins by considering the Poisson-Gaussian noise. For a Poisson distribution $\mathcal{P}(\lambda)$, it can be approximated by a Gaussian distribution, with the approximation error less than 1% when $\lambda \geq 5$ [43]. For a captured noisy image, its parameter of Poisson noise is jointly determined by the clean image z and sensor-specific factors such as the quantum efficiency and the analog gain [31]. Therefore, the Poisson-Gaussian noise can be approximately described by the heteroscedastic Gaussian model [31], *i.e.*,

$$n[u,v] \sim \mathcal{N}\big(0, \gamma z[u,v] + \nu\big), \tag{6}$$

where γ and ν are noise parameters determined by sensor-specific factors. For noise following such a model, it is relatively intuitive that each of its Fourier coefficients follows a Gaussian distribution since the linear combination of independent Gaussian variables is still a Gaussian variable.

Moreover, we observe that such a property is not only possessed by the Poisson-Gaussian noise that can be approximated by the heteroscedastic Gaussian model, but by any type of i.i.d. noise whose distribution might look very different from the Gaussian distribution. More specifically, we have the following theorem.

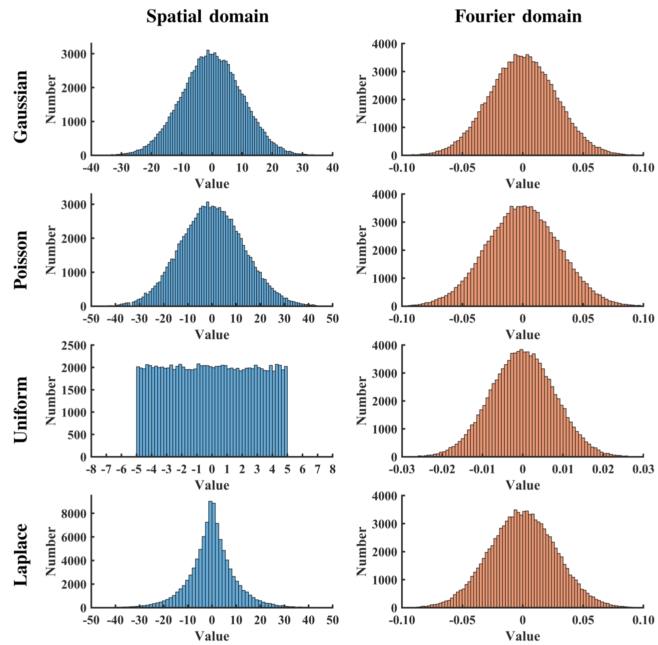


Fig. 1. Histograms of noise in the spatial domain and its coefficient in Fourier domain. For noise following different distributions in the spatial domain, their Fourier coefficients consistently approximate the Gaussian distribution.

Theorem 1: Suppose n is i.i.d. noise with bounded cumulants. As UV approaches infinity, each Fourier coefficient of n will (i) converge in distribution to a Gaussian distribution and (ii) be independent of other coefficients.

Proof: The proof is provided in the Appendix A. □

Remark 1 (Gaussian Approximation): Theorem 1 indicates that, when the image size (*i.e.*, UV) is sufficiently large, each Fourier coefficient of i.i.d. noise n approximately follows a Gaussian distribution. In practice, UV is generally much greater than 10^3, which is large enough for this approximation.

Remark 2 (Mean and Variance): A Gaussian distribution is characterized by its mean and variance. Since the Fourier transform is linear, the mean value of $\mathscr{F}(n)$ can be obtained by $\mathbb{E}[\mathscr{F}(n)] = \mathscr{F}(\mathbb{E}[n])$. Therefore, if the noise n has a zero mean, its Fourier coefficients will also have a zero mean. Regarding the variance of $\mathscr{F}(n)$, it is scaled by the normalization factor in (3). For the i.i.d. noise with a variance of σ_n^2, most of its Fourier coefficients will hold the variance of $\sigma_n^2/2UV$. But it is worth noting that this scaling affects both signal and noise equally, and thus the signal-to-noise ratio in the Fourier domain remains the same as in the spatial domain.

To illustrate our analysis, a statistical experiment is conducted on several noise distributions, and the results are visualized in Fig. 1. In this experiment, we at first repeatedly generate noise instances n of size 255×255 from a candidate distribution. Then, we calculate $\mathscr{F}(n)$ for each noise instance. Finally, we select a fixed pixel and a Fourier coefficient to plot the histograms. In the case of Poisson noise, it was generated based on a selected image, and the mean value was subtracted to maintain a zero mean. As Fig. 1 shows, the histograms of the Fourier coefficients approximate Gaussian distributions for noise following different distributions in the spatial domain. This empirical evidence supports our analysis regarding the noise distribution in the Fourier domain.

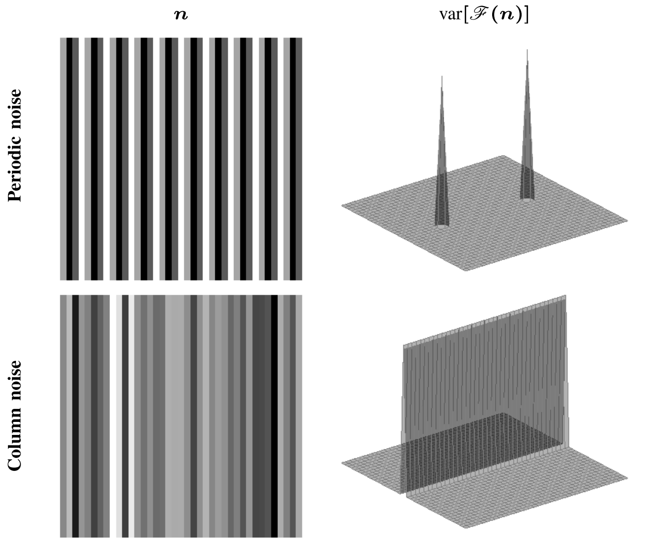


Fig. 2. Visualization of periodic noise, column noise, and their corresponding var[$\mathscr{F}(n)$]. The column noise and periodic noise are generated in the same way as [23] and [44], respectively. Their var[$\mathscr{F}(n)$] are zero at most positions, only accumulating at a few positions.

2) Spatially-Correlated Noise: For the spatially-correlated noise, we focus our discussions on the stationary noise model shown in (2), *i.e.*, $n = h*\eta$, where η is assumed to be zero-mean i.i.d. noise. Despite its simplicity in mathematical form, this model effectively describes various real-world noise sources such as stripe noise, band-limited noise, and pink noise [20], [21]. For such noise, we have the following theorem.

Theorem 2: Suppose the spatially-correlated noise n can be modeled by (2). As UV approaches infinity, each Fourier coefficient of n will (i) converge in distribution to the Gaussian distribution and (ii) be independent of other coefficients.

Proof: The proof is provided in the Appendix B. □

Remark 3 (Mean and Variance): For the mean and variance of $\mathscr{F}(n)$,

$$\mathbb{E}[\mathscr{F}(n)] = \mathscr{F}(h) \cdot \mathscr{F}(\mathbb{E}[\eta]) = 0, \tag{7}$$

and

$$\text{var}[\mathscr{F}(n)] = |\mathscr{F}(h)|^2 \cdot \text{var}[\mathscr{F}(\eta)]. \tag{8}$$

An important observation from [21] is that $|\mathscr{F}(h)|$ for diverse correlation kernels h exhibits sparsity. To illustrate this, Fig. 2 visualizes the periodic noise, column noise, and their corresponding variance distributions in the Fourier domain. For such noise, its energy is concentrated in only a small fraction of Fourier coefficients, while the vast majority remain nearly unaffected. This leads to a significant disparity in signal-to-noise ratio (SNR) across coefficients: the affected ones exhibit very low SNR, whereas the remaining ones retain high SNR. This observation suggests that different Fourier coefficients may provide supervisory signals of varying reliability. Therefore, enhancing supervision from noisy targets requires effectively exploiting the largely reliable coefficients while mitigating the impact of the few severely corrupted ones.

In summary, we have proved that the Fourier coefficients of a wide range of noise approximately follow the Gaussian distribution if the image size UV is large enough. This allows for a unified treatment of different types of noise under a common framework. Additionally, we have pointed out that noise exhibiting spatial correlation can maintain both the independence and the sparsity in the Fourier domain. These findings emphasize the advantages of utilizing the Fourier domain for noisy supervision and provide insights into the development of effective noise mitigation strategies.

C. Statistical Equivalence in the Fourier Domain

This subsection aims to demonstrate the efficacy of utilizing noisy targets in the Fourier domain to supervise image restoration learning, regarding which we have the following theorem.

Theorem 3: Let $\hat{z} = f_\theta(x)$ be the network output and $y = z+n$ the noisy target, where n is independent of x. Define the loss function in the Fourier domain as

$$L_\varphi\big(\hat{z}, y\big) = \sum_{k=0}^{U-1}\sum_{l=0}^{V-1} \ell_\varphi\Big(\mathscr{F}(\hat{z})[k,l] - \mathscr{F}(y)[k,l]\Big), \tag{9}$$

where $\ell_\varphi : \mathbb{C} \to \mathbb{R}$ is given by $\ell_\varphi(a + jb) = \varphi(a, b)$ and $\varphi : \mathbb{R}^2 \to \mathbb{R}$ is a penalty defined on the two-dimensional real plane corresponding to the complex coefficient. Suppose the real and imaginary Fourier coefficients of n at each position (approximately) follow independent zero-mean Gaussian distributions of the joint probability $p(a, b) = p_a(a)p_b(b)$, the following two statements hold.

(i) Utilizing the penalty φ with the noisy target y is statistically equivalent to utilizing its blurred version $\phi = \varphi * p$ with the clean target, *i.e.*,

$$\mathbb{E}[L_\varphi\,(\hat{z}, y)] = \sum_{k=0}^{U-1}\sum_{l=0}^{V-1} \ell_\phi\Big(\mathscr{F}(\hat{z})[k,l] - \mathscr{F}(z)[k,l]\Big), \tag{10}$$

where $\ell_\phi(a + jb) = \phi(a, b)$.

(ii) If $\varphi(a, b)$ is strictly increasing in the absolute value of each argument, the optimal point for (10) is $\hat{z}^\star = z$.

Proof: The proof is provided in the Appendix C. □

Remark 4 (Statistical Equivalence): This theorem indicates that supervising the image restoration learning with noisy targets y is statistically equivalent to the one with clean targets z, sharing the same optimal points.

Remark 5 (Noise Assumption): The noise assumption involved in this theorem is very mild. When the input image x and the target image y are captured from two independent observations, the independence between the noise n and the input x will be satisfied. Furthermore, based on our analysis above, the Fourier coefficients of a wide range of noise models will (approximately) follow independent Gaussian distributions when the image size is large enough. Consequently, this theorem is broadly applicable to diverse noise commonly observed in practical applications.

Remark 6 (Penalty Function): This theorem requires the penalty function $\varphi(a, b)$ to be strictly increasing in the absolute value of each argument. Representative penalty functions satisfying this requirement include the ℓ_2 norm $\varphi(a, b) = \sqrt{a^2 + b^2}$, the ℓ_1 norm $\varphi(a, b) = |a|+|b|$, and the separable Huber function

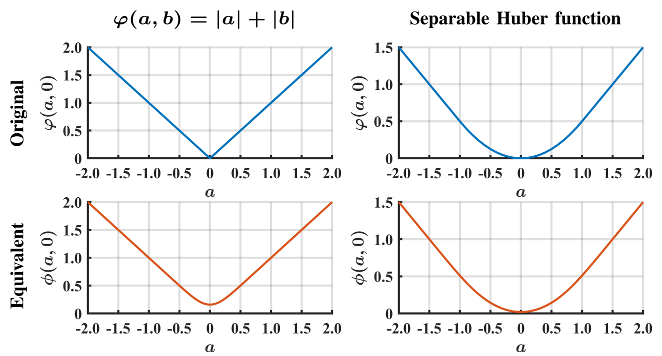


Fig. 3. Statistical equivalence. Utilizing $\varphi(a,b)$ with noisy targets is statistically equivalent to utilizing $\phi(a,b) = \varphi(a,b) * p(a,b)$ with clean targets.

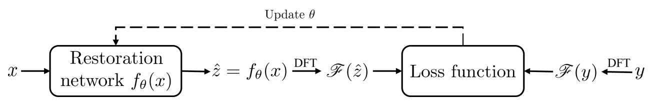


Fig. 4. The IR-NSF framework. It supervises image restoration learning with noisy targets y in the Fourier domain.

$\varphi(a,b) = H(a) + H(b)$, where $H(t)$ is the Huber loss defined as

$$H(t) = \begin{cases} \frac{1}{2}t^2, & |t| \leq \delta, \\ \delta\left(|t| - \frac{1}{2}\delta\right), & |t| > \delta. \end{cases} \tag{11}$$

The threshold δ controls the transition between the quadratic and linear regions. We empirically set it as 1. Our ablation studies in the supplementary material demonstrate that both the ℓ_2 and ℓ_1 norms perform well for image super-resolution and deblurring tasks, whereas the separable Huber function is more effective for image denoising. Accordingly, in the experimental section, we adopt the separable Huber function for denoising and the ℓ_1 norm for super-resolution and deblurring. To illustrate the statistical equivalence in this theorem, Fig. 3 visualizes the original penalty function $\varphi(a,b)$ and its blurred version $\phi(a,b) = \varphi(a,b) * p(a,b)$ along a one-dimensional cross-section $b = 0$. Clearly, φ and ϕ differ only slightly in a small region near the origin, implying that they will yield similar learning results.

D. Weakly Supervised Learning With Noisy Targets

Building on our developed Fourier-based statistical equivalence shown in (10), we construct a weakly supervised learning framework termed image restoration via noisy supervision in the Fourier domain (IR-NSF). As illustrated in Fig. 4, IR-NSF supervises image restoration learning with noisy targets in the Fourier domain. Its loss function is

$$\sum_{s=1}^{S} L_\varphi\big(f_\theta(x_s), y_s\big), \tag{12}$$

where S denotes the total number of training pairs, x_s and y_s denotes the s-th training image pair. Although this loss function does not explicitly involve expectation over the noise n as (10) does, we empirically observe that the statistical equivalence in (10) still holds when the number of training images is large enough. The justification of omitting expectation over the noise n in (12) lies in the inherent self-similarity of natural images: within each image and across the training set, there are typically numerous similar patches. Consequently, aggregating the loss over these similar image patches implicitly approximates the expectation with respect to the noise n. Notably, since IR-NSF differs from N2N only in the loss evaluation, the additional computation comes from the 2D-DFT used in the Fourier-domain loss. This cost is substantially smaller than the forward and backward passes of the restoration network, so the overall training time of IR-NSF is nearly identical to that of N2N. Moreover, IR-NSF does not alter the network architecture and, therefore, introduces no additional computational overhead at inference time. As an architecture-agnostic method, its overall computational cost is primarily determined by the backbone network employed.

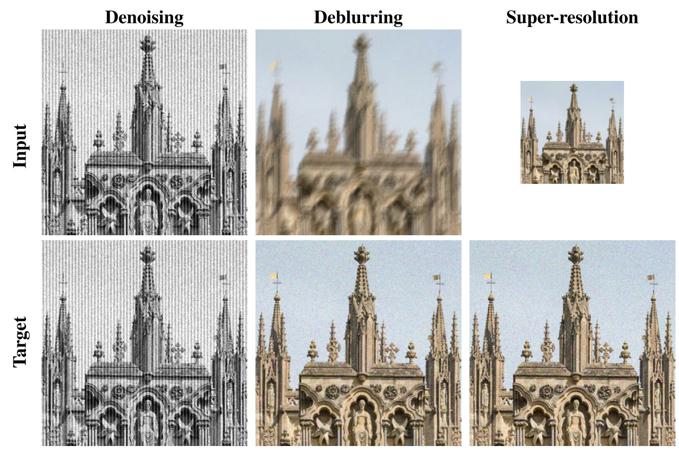


Fig. 5. Illustration of training pairs for DN, SR, and DB. These images are generated from 0149 in DIV2K.

IV. Experimental Evaluation of IR-NSF

This section provides a comprehensive evaluation of our proposed weakly supervised learning framework, IR-NSF, covering convergence analysis, image restoration performance, and data collection burden. The assessment is conducted through both synthetic and real-world experiments, with subsections A-D dedicated to synthetic experiments and subsection E focusing on real-world experiments.

A. Experimental Setting

Image Restoration Tasks. In synthetic experiments, IR-NSF is evaluated on three distinct image restoration tasks: grayscale image denoising (DN), color image super-resolution (SR), and color image deblurring (DB). For DN, we consider a mixture of i.i.d. Gaussian noise of $\sigma = 5$, Poisson noise, and the periodic noise shown in Fig. 2. For SR, we simulate the low-resolution images by downsampling from higher resolution ones. For DB, as suggested by [45], we generate blurry images with eight motion blur kernels from [46] and further add i.i.d. Gaussian noise with a standard deviation of $\sigma = 2$.

Training and Testing Data. The degraded input images and noisy target images are synthesized with source images from publicly available datasets. During the training phase, we employ 800 4K-resolution training images from the DIV2K dataset [47]. To accelerate the data loading process, these images are pre-processed by cropping them into 22988 image patches of size 480×480. During the testing phase, we utilize images from the Set5 [48], Set14 [49], Kodak24,[1] McMaster [50], Urban100 [51], and Manga109 [52] datasets. For DN, the source images are converted to grayscale, while for SR and DB, the color information is retained. The degraded input images are synthesized using the aforementioned degradation models, and the noisy target images are generated by adding synthetic noise of the specified model to the latent clean image. For DN, the noise model is identical for both the input and target images. For SR and DB, Gaussian noise with a standard deviation of $\sigma = 10$ is utilized for generating noisy targets. Without explicit specification, the noise generation is carried out on the fly during the training process. In Fig. 5, training pairs for different tasks are illustrated.

Comparison Methods. Our proposed IR-NSF framework is designed for image restoration learning using noisy targets. To ensure fairness, we primarily compare IR-NSF with Noise2Noise (N2N) [13], which adopts the same setting as ours. To compare the statistical equivalences established by N2N and IR-NSF, we also apply both frameworks to clean targets. These alternative cases are referred to as N2C and IR-CSF, respectively.

Evaluation Metrics. We employ the Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) as quantitative evaluation metrics for the restored images in all tasks. However, the calculation of these metrics varies depending on the specific image restoration task. For DN and DB, the metrics are calculated using all color channels of the restored images. For SR, the calculation follows the approach in [53], where the PSNR and SSIM are measured on the Y channel of the YCbCr color space, and an equal number of border pixels is ignored corresponding to the scaling factor.

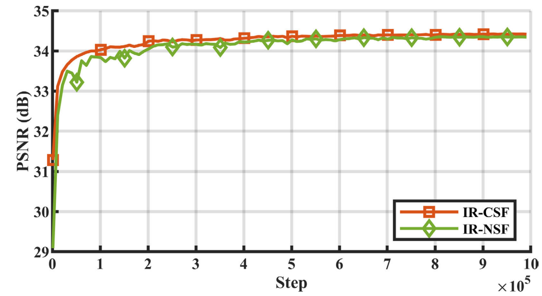
Network Architectures. To validate that our proposed learning framework can be applied to diverse network architectures, we utilize two different networks for each image restoration task. Specifically, we employ ResNet (for DN), RDN [45] (for DN), EDSR [53] (for SR), RCAN [54] (for SR), MIMO [28] (for DB), and MIMO+ [28] (for DB). In the case of ResNet, it is a modified version of EDSR in which the up-sample block is removed and a long skip connection is added between the input and output. The remaining networks align with the descriptions provided in their respective papers.

Training Settings. The training of these networks is conducted based on the source codes provided by their respective authors. The default settings are maintained, except for the modifications introduced here. For MIMO and MIMO+, the number of training epochs is adjusted to 300, with the learning rate halved every 50 epochs. All experiments are performed on the AMD Ryzen Threadripper PRO 3955WX CPU and the NVIDIA GeForce RTX 4090 GPU.

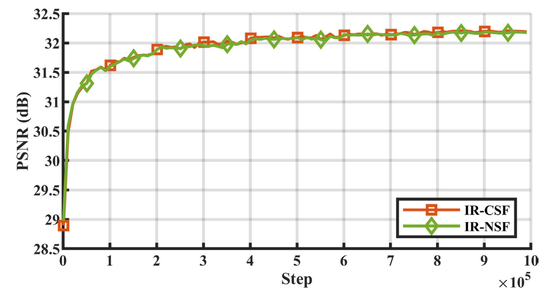
B. Convergence Analysis

In Fig. 6, the convergence curves of using clean targets (IR-CSF) and noisy targets (IR-NSF) are compared. The

[1] http://r0k.us/graphics/kodak/



(a) DN experiment with ResNet



(b) SR experiment with EDSR

Fig. 6. PSNR (dB) on Urban100 as a function of training step. For (a) the denoising task, noise type is the mixture of Poisson-Gaussian and periodic noise. For (b) the super-resolution task, the noisy target used by IR-NSF is corrupted by Gaussian noise. Noisy images are generated on the fly with the source images from DIV2K. For both tasks, training with noisy targets (*i.e.*, IR-NSF) or clean targets (*i.e.*, IR-CSF) converges to almost the same results.

convergence curve is plotted as PSNR on Urban100 against the training step. Clearly, for the SR experiment shown in Fig. 6(b), the convergence curves for IR-NSF and IR-CSF are almost consistent everywhere. As for the DN experiment shown in Fig. 6(a), there are obvious differences between the two curves at the early training stage. This is because noisy targets for DN consist of periodic noise, which is more difficult to tackle than the i.i.d. Gaussian noise in the target for the SR experiment. However, the extra difficulty posed by the periodic noise can be successfully overcome in the late training stage. As Fig. 6(a) shows, the curves of IR-CSF and IR-NSF finally converge to close points. Overall, experiments shown in Fig. 6 demonstrate that training with noisy targets (IR-NSF) has an equivalent effect to training with clean targets (IR-CSF).

C. Image Restoration Results

The PSNR and SSIM results of image denoising, super-resolution, and deblurring experiments are reported in Tables I-III, respectively. To support perceptual evaluation, several visual results are provided in Figs. 7–9. Due to space limitations, results with RDN, RCAN, and MIMO+ are reported in the supplementary material.

1) Denoising: From Fig. 7 and Table I, it can be observed that IR-CSF and IR-NSF achieve similar performances. In contrast, there is a noticeable performance gap between N2C and N2N. On Urban100, the PSNR of N2N is over 0.3 dB worse than those of the other three methods. Also, N2N results in the loss of more image details, as illustrated in the zoomed regions in Fig. 7. These results demonstrate that, compared to

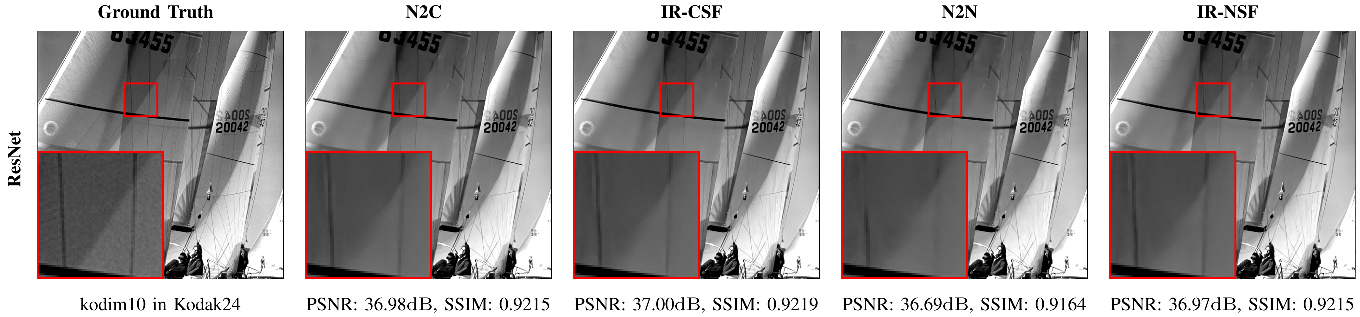


Fig. 7. Visual results of denoising experiments. The image contrast is adjusted for better comparison.

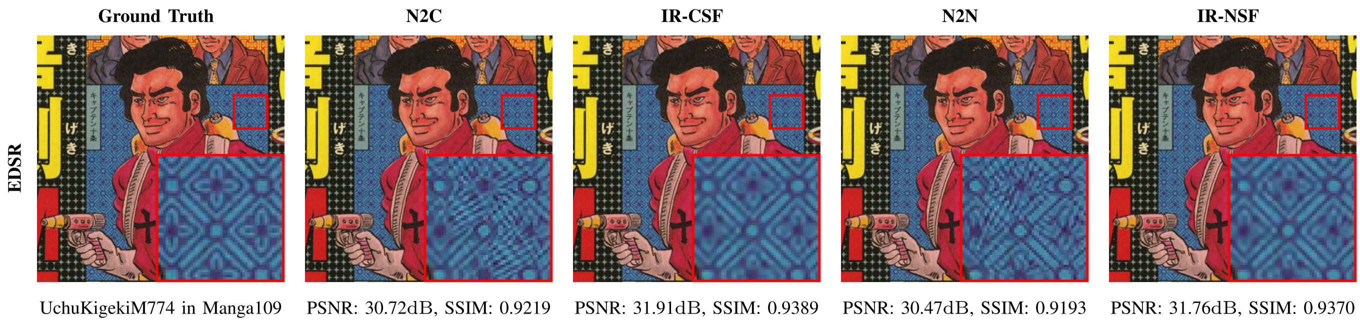


Fig. 8. Visual results of super-resolution experiments.

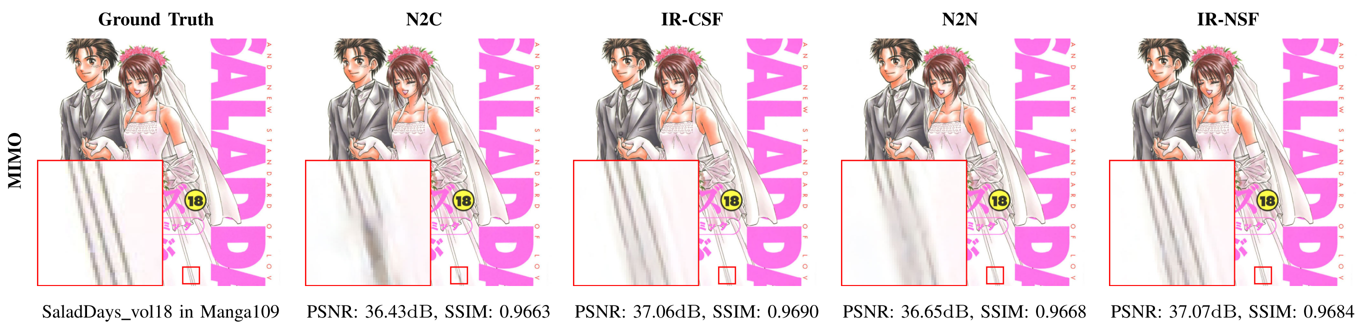


Fig. 9. Visual results of deblurring experiments.

TABLE I

AVERAGE PSNR (dB) AND SSIM RESULTS OF DENOISING EXPERIMENTS. THE RESULTS OF IR-NSF ARE HIGHLIGHTED IN GRAY

Network	Target	Method	Set5		Set14		Kodak24		McMaster		Urban100		Manga109	
			PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
ResNet	Clean	N2C	35.07	0.9315	33.77	0.9147	34.45	0.9176	35.92	0.9398	34.39	0.9477	36.16	0.9508
		IR-CSF	35.07	0.9317	33.77	0.9149	34.45	0.9176	35.92	0.9399	34.41	0.9478	36.16	0.9508
	Noisy	N2N	34.94	0.9295	33.62	0.9125	34.32	0.9155	35.77	0.9380	34.04	0.9450	35.97	0.9492
		IR-NSF	35.06	0.9312	33.75	0.9146	34.43	0.9175	35.90	0.9396	34.34	0.9474	36.14	0.9505

training in the spatial domain, training in the Fourier domain can reduce the gap between using clean and noisy targets. This improvement can be attributed to the fact that the periodic noise affects all spatial pixels but only influences a small portion of the Fourier coefficients.

2) Super-Resolution: From Fig. 8 and Table II, it can be observed that IR-NSF achieves a higher PSNR (up to 0.2 dB higher) and better visual results than N2N, despite there being almost no performance gap between training with noisy targets and clean targets in both the spatial domain and the Fourier domain for this experimental setting. This result validates that establishing supervision in the Fourier domain is more effective than in the spatial domain. This could be attributed to Fourier coefficients capturing global information, whereas spatial pixels capture only limited local information, allowing the former to provide more meaningful supervision.

3) Deblurring: From Table III, it can be observed that IR-NSF outperforms N2N over 0.3 dB. Moreover, as Fig. 9 shows, edges in images recovered by IR-NSF are clearer and sharper compared to those obtained with N2N. Similar

TABLE II

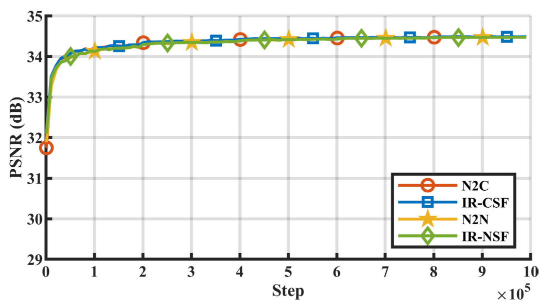
AVERAGE PSNR (dB) AND SSIM RESULTS OF SUPER-RESOLUTION EXPERIMENTS. THE RESULTS OF IR-NSF ARE HIGHLIGHTED IN GRAY

Network	Target	Method	Set5		Set14		Kodak24		McMaster		Urban100		Manga109	
			PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
EDSR	Clean	N2C	37.95	0.9602	33.54	0.9175	34.01	0.9229	38.32	0.9636	32.05	0.9277	38.37	0.9763
		IR-CSF	38.07	0.9607	33.68	0.9181	34.08	0.9237	38.45	0.9642	32.20	0.9286	38.75	0.9772
	Noisy	N2N	37.94	0.9603	33.54	0.9173	33.99	0.9227	38.29	0.9635	32.04	0.9277	38.27	0.9762
		IR-NSF	38.04	0.9606	33.65	0.9183	34.06	0.9234	38.42	0.9642	32.20	0.9288	38.55	0.9769

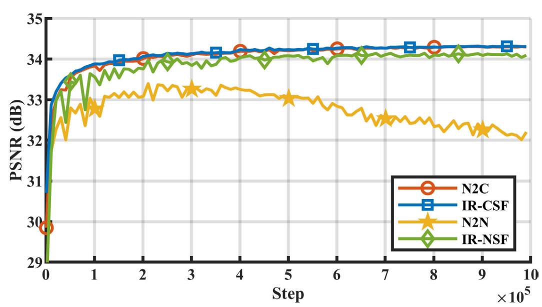
TABLE III

AVERAGE PSNR (dB) AND SSIM RESULTS OF DEBLURRING EXPERIMENTS. THE RESULTS OF IR-NSF ARE HIGHLIGHTED IN GRAY

Network	Target	Method	Set5		Set14		Kodak24		McMaster		Urban100		Manga109	
			PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
MIMO	Clean	N2C	32.75	0.9162	31.56	0.8954	33.09	0.9186	34.56	0.9325	30.75	0.9218	34.61	0.9431
		IR-CSF	32.95	0.9179	32.05	0.9004	33.64	0.9224	34.74	0.9344	31.30	0.9274	34.98	0.9455
	Noisy	N2N	32.79	0.9157	31.75	0.8967	33.12	0.9189	34.49	0.9322	30.84	0.9221	34.65	0.9430
		IR-NSF	32.96	0.9179	32.05	0.9000	33.59	0.9223	34.73	0.9344	31.31	0.9274	34.88	0.9454

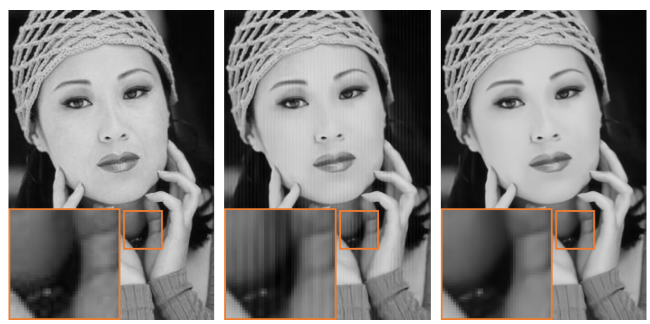


(a) Poisson-Gaussian noise



(b) Mixed noise

Fig. 10. PSNR (dB) on Urban100 as a function of training step for limited noise realizations. For (a), the noise type is the Poisson-Gaussian noise, and the training set is the DIV2K. For (b), the noise type is the mixture of Poisson-Gaussian and periodic noise. The training set is the combination of DIV2K and Flickr2K.



(a) Ground Truth (b) N2N (c) IR-NSF

Fig. 11. Visual results of denoising experiments under limited training data. The statistical equivalence for N2N becomes invalid in this case. Consequently, the periodic noise in the training target appears in its recovered image. (a) Ground Truth. (b) N2N (PSNR=33.27 dB, SSIM= 0.8984). (c) IR-NSF (PSNR=34.68 dB, SSIM= 0.9507).

to the SR experiments, there is almost no performance gap between IR-NSF/N2N and IR-CSF/N2C, demonstrating that the advantage of IR-NSF over N2N, in this case, is due to the Fourier-based supervision being more effective for the image deblurring task compared to spatial domain supervision.

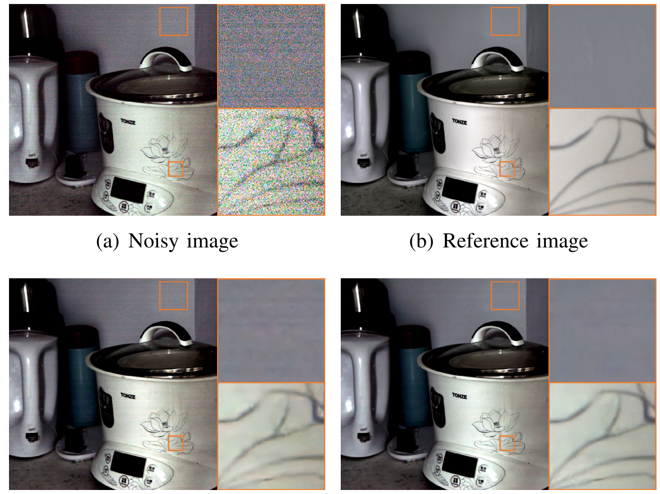
D. Data Collection Burden

The IR-NSF and N2N are based on the statistical equivalences between training on noisy and clean targets. The term '*statistical*' implies that such an equivalence could be invalid when the number of training data is limited. In this subsection, we compare N2N and IR-NSF from the perspective of the required amount of training data. The task used for this comparison is the image denoising task. Two types of noise have been considered. The first noise type is the i.i.d. Poisson-Gaussian noise. In this case, we employ 800 images from DIV2K as the source images. For each source image, two noisy images are independently generated to serve as training pairs. As such, there are 800 training pairs for this case in total. For the two noisy images in each pair, we do not specify which is the input and which is the target. Instead, their roles are randomly and dynamically determined at each training iteration to maximize their use. The second noise type is a mixture of Poisson-Gaussian and periodic noise. In this case, 2650 source images from Flickr2K [55] are additionally selected, yielding 3450 training pairs in total. In Fig. 10, the convergence curves of N2N and IR-NSF are plotted. The plots of N2C and IR-CSF are also provided for reference. In Fig. 10(a), we observe that all of the convergence curves are almost consistent everywhere. This indicates that when the target image is corrupted by i.i.d. Poisson-Gaussian noise, the statistical equivalences used by N2N and IR-NSF will be effective on a dataset containing an equal or larger number of images than DIV2K. However, things become different when the target image is further corrupted by periodic noise. In Fig 10(b), the convergence curve of N2N shows a noticeable deviation from that of N2C, indicating that the statistical equivalence for N2N becomes invalid in this case, even though more training data have been adopted. Consequently, as shown in Fig. 11, the periodic noise in the corrupted target will also appear in the image denoised by N2N. By contrast, there is only a minor performance gap between IR-NSF and IR-CSF. Also, the noise pattern is suppressed in the image processed by IR-NSF. These results demonstrate that the statistical equivalence for IR-NSF remains valid in this case. Therefore, we claim that, compared to N2N, IR-NSF has a lower requirement for the amount of training data when the target is corrupted by spatially correlated noise, such as periodic noise.

E. Real-World Experiments

Besides synthetic experiments, the proposed method is validated through real-world experiments on two publicly available datasets. One is the See-in-the-Dark (SID) [56], and the other is the real local motion blur dataset (ReLoBlur) [57].

1) Real-World Denoising Experiment on SID: The real-world denoising experiment is conducted on the SID Sony set [56], which contains sequences of short-exposure low-light RAW images alongside their corresponding long-exposure RAW images that serve as ground-truth References. These RAW images are of the resolution of 4240×2832. In our experiments, 279 short-exposure image pairs from the training set are used for network training. Evaluation is conducted on 36 short-exposure and long-exposure pairs from the validation set and 90 pairs from the test set. The short-exposure RAW images are typically contaminated by both pixel-wise and stripe-wise noise. To perform denoising, these noisy RAW images are packed into 4-channel tensors for the network to process, and the denoised RAW image is obtained by unpacking the network output. The network architecture and training procedure are similar to those used in the synthetic denoising experiment. Due to the flickering effect of light sources powered by alternating current, pixel intensities in regions around these light sources could exhibit significant discrepancies between paired short-exposure images. These pixels are excluded from loss calculation and backpropagation to prevent training instability. The trained networks are evaluated with image pairs from the SID Sony validation and test sets. To compensate for inaccuracies and fluctuations in camera setups (*e.g.*, shutter speed, analog gain, and black level), two scalar parameters (a scaling factor and a bias) are computed to post-process the denoised short-exposure image to minimize the mean squared error against the reference long-exposure image. In Table IV and Fig. 12, the evaluation metrics and the visual results are provided for quantitative and perceptual comparisons. As Table IV shows, IR-NSF outperforms N2N in both PSNR and SSIM. Additionally, Fig. 12 indicates that IR-NSF is better at suppressing the stripe-wise noise and preserving image textures. Moreover, we provide an additional visual comparison on a challenging outdoor HDR scene from the SID dataset in the supplementary material (Fig. S4), where IR-NSF again demonstrates superior texture preservation over N2N under extreme lighting conditions.



(a) Noisy image (b) Reference image

(c) N2N (d) IR-NSF

Fig. 12. Visual results of denoising experiments on SID. These RGB images are obtained by applying the white balance and demosaicking to the RAW images. The image contrast is adjusted for better comparison. (a) Noisy image. (b) Reference image. (c) N2N [13] (PSNR=45.40 dB, SSIM=0.9908). (d) IR-NSF (PSNR=46.03 dB, SSIM=0.9917).

TABLE IV
QUANTITATIVE RESULTS ON REAL-WORLD DENOISING DATASET SID [56]. THE RESULTS OF IR-NSF ARE HIGHLIGHTED IN GRAY

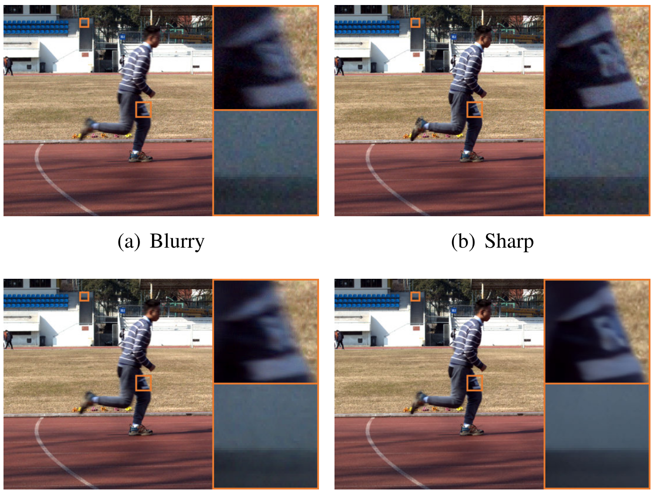
Method	SID Validation		SID Test	
	PSNR	SSIM	PSNR	SSIM
N2N	36.85	0.8466	40.10	0.9124
IR-NSF	37.07	0.8476	40.32	0.9140

2) Real-World Deblurring Experiment on ReLoBlur: The ReLoBlur [57] comprises 2,405 paired images at a resolution of 2152×1436, with 2,010 pairs allocated for training and the remaining 395 pairs for testing. It is captured by utilizing a beam splitter to divert incident light from the target scene to two synchronized cameras with different exposure times. Each scene consists of static background regions and locally moving objects. The images captured by paired cameras are blurry and sharp, respectively. To enhance the content consistency of these paired images, they are post-processed by color correction, photometrical alignment, and geometrical alignment. It is worth noting that the noise reduction operator is not involved in these post-processing steps, and thus, the noise can be observed in both the blurry and sharp images. In other words, training targets provided by ReLoBlur are generally noisy, especially for scenes under relatively low lighting conditions. Utilizing this dataset, we conduct a comparison between N2N and the proposed IR-NSF. In this case, we adopt the same network architecture and training settings as [57]. For the metrics, in addition to PSNR and SSIM, we also follow [57] and adopt the weighted PSNR ($PSNR_w$) and weighted SSIM ($SSIM_w$), which are calculated only on locally blurry regions.

TABLE V

QUANTITATIVE RESULTS ON REAL-WORLD DEBLURRING DATASET RELOBLUR [57]. PSNR_W AND SSIM_W DENOTE WEIGHTED PSNR AND WEIGHTED SSIM THAT ARE CALCULATED ON THE LOCALLY BLURRY REGION. THE RESULTS OF IR-NSF ARE HIGHLIGHTED IN GRAY

Method	PSNR	SSIM	PSNR_w	SSIM_w
N2N	34.19	0.9186	27.18	0.8505
IR-NSF	34.43	0.9250	27.39	0.8630



(a) Blurry (b) Sharp (c) N2N (d) IR-NSF

Fig. 13. Deblurring results on ReLoBlur. The image contrast is adjusted for better comparison. In each image, the zoomed region on the top right corner is from the moving object, and the one on the bottom right is from the static background. Noise can be observed in the zoomed background regions of both the (a) blurry image and the (b) sharp image. (c) N2N [13] (PSNR=33.79 dB, SSIM= 0.8904). (d) IR-NSF (PSNR=34.10 dB, SSIM=0.9052).

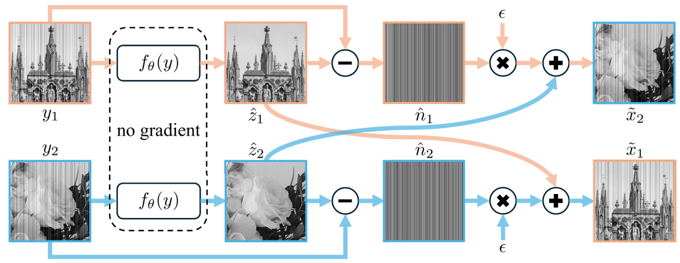
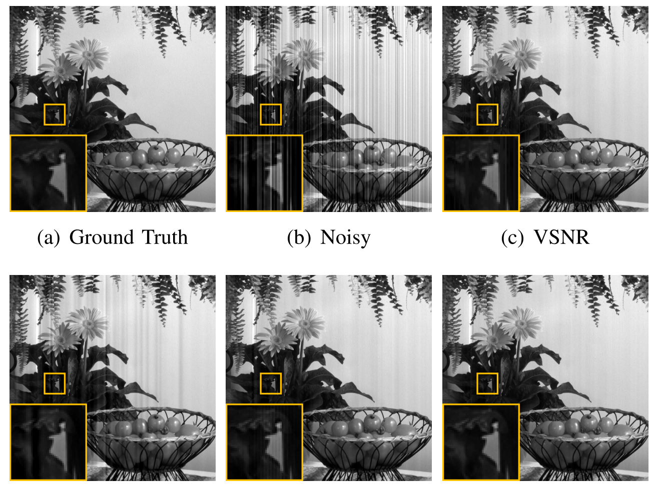


Fig. 14. Data generation procedure of our proposed USR.



(a) Ground Truth (b) Noisy (c) VSNR (d) SEID (e) SelfBSR (f) USR

Fig. 15. Visual comparison of destriping methods. (a) Ground Truth. (b) Noisy images. (c) VSNR (PSNR=36.93 dB, SSIM=0.9746). (d) SEID (PSNR=33.47 dB, SSIM=0.9481). (e) Self-BSR (PSNR=37.30 dB, SSIM=0.9777). (f) USR (PSNR=39.92 dB, SSIM=0.9915). The image contrast is adjusted for better comparison.

Although the sharp image is somewhat noisy, these metrics can still evaluate the deblurring performance since the blurring effect generally has a stronger influence on these metrics than the noise. From Table V, it can be observed that the proposed IR-NSF outperforms N2N in both PSNR and SSIM on the ReLoBlur dataset. In Fig. 13, a visual comparison is provided. As the zoomed regions show, IR-NSF yields clearer, sharper results than N2N, demonstrating its superiority on the real-world deblurring task.

V. UNSUPERVISED STRIPE REMOVAL

Besides enabling weakly supervised learning with noisy targets, the Fourier-based statistical equivalence can also facilitate the development of unsupervised denoising methods using single noisy images. To demonstrate this, this section develops a Fourier-based unsupervised method that achieves state-of-the-art performance in removing stripe-wise noise.

Compared to spatially independent pixel-wise noise, stripe-wise noise poses a greater challenge for unsupervised denoising methods. Existing state-of-the-art unsupervised approaches typically assume that noise is spatially independent, which limits their effectiveness when dealing with stripe-wise noise that exhibits strong spatial correlations along specific lines or across the entire image. Although several techniques (e.g., [59], [60], [61]) have been proposed to relax the assumption of noise independence, they are generally effective only for locally correlated noise and fail to address noise with global spatial dependencies. This limitation hinders their applicability in scenarios such as extremely low-light imaging and remote sensing, where stripe noise is a common occurrence. To address this issue, we propose an unsupervised stripe removal (USR) based on the statistical equivalence in the Fourier domain.

Given a set of unpaired images corrupted by stripe noise, USR constructs training pairs via a noise-swapping strategy, as illustrated in Fig. 14. At each training iteration, the noisy image $y_i = z_i + n_i$ is first processed by the network under training to produce an estimate $\hat{z}_i$ of the latent clean image. To stabilize training, gradient backpropagation is stopped during this step. The stripe noise is then estimated by $\hat{n}_i = y_i - \hat{z}_i$. Subsequently, a synthetic training input is generated as $\tilde{x}_i = \hat{z}_i + \epsilon \cdot \hat{n}_j$, where $\hat{n}_j$ is the stripe noise estimated from another randomly sampled noisy image in the dataset. The amplification factor $\epsilon > 1$ is introduced to compensate for the potential underestimation of the stripe noise, which often occurs during training. The resulting pair $(\tilde{x}_i, y_i)$ is then used as the training input and target, respectively. As shown in Fig. 2, stripe noise only affects a subset of Fourier

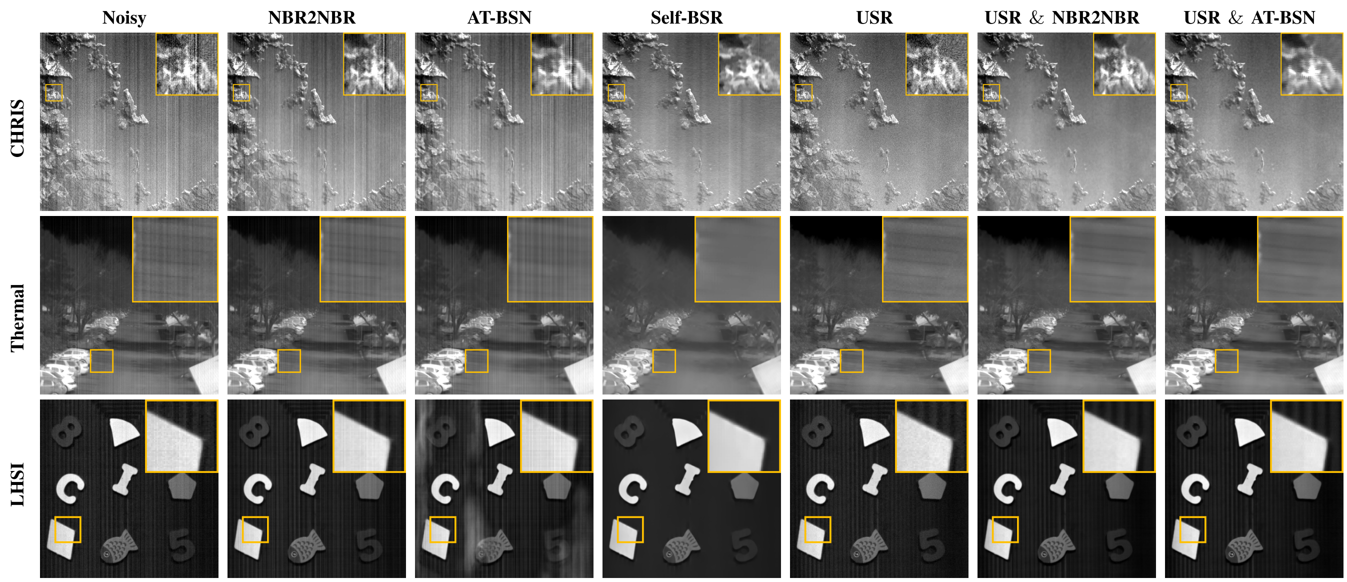


Fig. 16. Visual comparison of unsupervised techniques on real-world images corrupted by mixed pixel-wise and stripe-wise noise.

TABLE VI

AVERAGE PSNR (dB) AND SSIM RESULTS OF VSNR [20], SEID [44], SELF-BSR [58] AND USR. THESE METHODS ARE ADOPTED TO REMOVE THE STRIPE NOISE. THE GAUSSIAN NOISE IS POST-PROCESSED BY BM3D [3]. THE RESULTS OF USR ARE HIGHLIGHTED IN GRAY. THE BEST RESULTS ARE HIGHLIGHTED IN BOLD

Noise	Method	Set5		Kodak24		McMaster	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Row	VSNR	38.52	**0.9915**	36.16	0.9845	38.08	0.9841
	SEID	35.90	0.9474	35.69	0.9818	35.05	0.9565
	Self-BSR	36.04	0.9774	35.41	0.9858	36.27	0.9737
	USR	**39.99**	0.9905	**38.51**	**0.9953**	**39.84**	**0.9888**
Row & Gaussian	VSNR	34.40	0.9375	33.30	0.9481	34.81	0.9418
	SEID	34.81	0.9134	34.92	0.9533	36.22	0.9499
	Self-BSR	34.75	0.9424	33.58	0.9459	34.41	0.9408
	USR	**35.89**	**0.9470**	**35.08**	**0.9533**	**36.23**	**0.9524**

coefficients, implying that the remaining coefficients can be left unprocessed. Accordingly, the loss function for USR is computed solely over these corrupted coefficients, *i.e.*,

$$L_{\varphi}^{USR} = \sum_{(k,l)\in\Omega} \ell_{\varphi}\Big(\mathscr{F}(f_{\theta}(\tilde{x}))[k,l] - \mathscr{F}(y)[k,l]\Big), \tag{13}$$

where Ω denotes the set of positions where Fourier coefficients are corrupted by stripe noise. For vertical stripes, $\Omega = \{(k,l) : k = 0\}$. For stripes with mixtures of vertical and horizontal orientations, $\Omega = \{(k,l) : k = 0 \text{ or } l = 0\}$. The network thus learns to recover only the coefficients at positions in Ω. To obtain the destriped image, we replace the Fourier coefficients of the original image with those from the network output only at these positions, leaving all other coefficients unchanged. This enables USR to preserve both image structure and pixel-wise noise while effectively removing stripes. Consequently, USR can be applied independently for stripe removal or combined with other unsupervised denoisers for fully unsupervised removal of both stripe-wise and pixel-wise noise.

The performance of USR is validated on both the synthetic and real-world experiments. In synthetic experiments, we evaluate its destriping performance under two scenarios: with and without the interference of Gaussian noise. The training data is synthesized from source images in the DIV2K, while the test sets are Set5, Kodak24, and McMaster. The comparison methods include: VSNR [20], an effective algorithm for removing stationary noise, SEID [44], an EM-based stripe estimation framework that can integrate various image denoising methods, and Self-BSR [58], an unsupervised deep learning technique designed for simultaneous stripe estimation and image recovery. For both synthetic scenarios, these methods and USR are used exclusively for destriping. When Gaussian noise is present, BM3D is subsequently applied to post-process the destriped images. As shown in Table VI, the proposed USR achieves the highest PSNR and SSIM in almost all cases. Furthermore, Fig. 15 demonstrates that USR better suppresses stripe noise compared to VSNR, SEID, and Self-BSR, with noticeably fewer remaining artifacts. In addition, we compare the computational complexity of these destriping algorithms. As reported in Table VII, USR requires only 17.8% of the GFLOPs of Self-BSR, and its running time is substantially lower than that of the other methods. Notably, USR is agnostic to the choice of backbone network. While we adopt ResNet for USR in this work, its computational complexity can be further reduced by utilizing a lighter network.

The effectiveness of USR is further validated on three real-world datasets: the Compact High Resolution Imaging Spectrometer (CHRIS)[2] remote sensing dataset, the thermal imaging (TI) dataset [62], and the low-light hyperspectral imaging (LHSI) indoor dataset [63]. In all cases, the captured images are corrupted by both pixel-wise and stripe-wise noise. The stripes in CHRIS images are vertical, while those in TI and LHSI exhibit a mixture of vertical and horizontal orientations. We compare our method against several unsupervised

[2]https://www.brockmann-consult.de/beam/data/products/

TABLE VII

COMPARISON OF COMPUTATIONAL COMPLEXITY OF DESTRIPING METHODS. THE RUNNING TIME IS MEASURED AS THE AVERAGE PROCESSING TIME PER IMAGE ON MCMASTER. VSNR [20] AND SEID [44] ARE EVALUATED ON THE INTEL(R) CORE(TM) I7-12700 CPU, WHILE SELF-BSR [58] AND OUR PROPOSED USR ARE RUN ON THE NVIDIA GEFORCE RTX 4090 GPU

Method	VSNR	SEID	Self-BSR	USR
GFLOPs	-	-	447.10	79.80
Time (s)	0.74	446.21	0.09	0.02

denoising baselines, including NBR2NBR [17], AT-BSN [61], and Self-BSR [58]. As shown in Fig. 16, both NBR2NBR and AT-BSN effectively remove pixel-wise noise but fail to eliminate stripe artifacts. Self-BSR, which is designed to address both noise types, either leaves residual stripes or over-smoothes image textures. In contrast, our proposed USR effectively removes stripe noise while preserving detailed image structures. Since stripe noise typically produces more visually prominent artifacts than pixel-wise noise, USR alone yields visually more appealing results than existing methods. Furthermore, we observe that for these mixed-noise scenarios, combining USR with unsupervised denoisers such as NBR2NBR or AT-BSN provides a more effective solution than using Self-BSR alone. This two-stage strategy leverages the complementary strengths of USR in stripe removal and existing denoisers in handling pixel-wise noise, achieving superior overall restoration. Beyond performance, this strategy also offers advantages in computational complexity. In this work, both USR and NBR2NBR adopt ResNet as their backbone, resulting in comparable GFLOPs and runtime. Consequently, the computational cost of their combination would be approximately twice that of USR alone. According to Table VII, even this combined cost remains lower than that of Self-BSR.

VI. CONCLUSION

Motivated by the key insights that Fourier coefficients contain global information and that spatially correlated noise, such as periodic and column noise, exhibits sparsity and independence in the Fourier domain, this work proposes a novel concept of establishing noisy supervision in the Fourier domain. To support this concept, we conduct a comprehensive statistical analysis demonstrating that the Fourier coefficients of a wide range of noise converge in distribution to the Gaussian distribution, enabling a unified treatment of diverse noise models. Building on this analysis, we establish a Fourier-based statistical equivalence between learning with noisy targets and clean targets under mild assumptions. This equivalence forms the theoretical basis for IR-NSF, an effective weakly supervised learning framework that operates with noisy targets in the Fourier domain. Extensive experiments across multiple restoration tasks and network architectures demonstrate that IR-NSF delivers superior performance in super-resolution, deblurring, and stripe-wise noise removal. Beyond the weakly supervised setting, the newly developed statistical equivalence also inspires the design of a Fourier-based unsupervised stripe removal, USR, which operates on unpaired noisy images and demonstrates superior results on both synthetic and real-world benchmarks. Collectively, these findings highlight the versatility and strength of Fourier-based statistical equivalence, offering new tools and perspectives for advancing image restoration learning under noisy supervision.

ACKNOWLEDGMENT

The authors would like to thank Dr. Tao Liu, Chutian Wang, Chao Qu, and Dr. Shuo Zhu for their helpful discussions. The source code and trained models are publicly available at https://github.com/haosennn/IR-NSF

REFERENCES

[1] L. I. Rudin, S. Osher, and E. Fatemi, “Nonlinear total variation based noise removal algorithms,” *Phys. D, Nonlinear Phenomena*, vol. 60, nos. 1–4, pp. 259–268, Nov. 1992.

[2] M. Aharon, M. Elad, and A. Bruckstein, “*rmK*-SVD: An algorithm for designing overcomplete dictionaries for sparse representation,” *IEEE Trans. Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[3] K. Dabov, A. Foi, V. Katkovnik, and K. Egiazarian, “Image denoising by sparse 3-D transform-domain collaborative filtering,” *IEEE Trans. Image Process.*, vol. 16, no. 8, pp. 2080–2095, Aug. 2007.

[4] W. Dong, G. Shi, and X. Li, “Nonlocal image restoration with bilateral variance estimation: A low-rank approach,” *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 700–711, Feb. 2013.

[5] S. Gu, L. Zhang, W. Zuo, and X. Feng, “Weighted nuclear norm minimization with application to image denoising,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 2862–2869.

[6] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, “Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising,” *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3142–3155, Jul. 2017.

[7] X.-J. Mao, C. Shen, and Y.-B. Yang, “Image restoration using very deep convolutional encoder–decoder networks with symmetric skip connections,” in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 29, 2016, pp. 2810–2818.

[8] C. Dong, C. C. Loy, K. He, and X. Tang, “Learning a deep convolutional network for image super-resolution,” in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 184–199.

[9] P. Zhang and E. Y. Lam, “From local to global: Efficient dual attention mechanism for single image super-resolution,” *IEEE Access*, vol. 9, pp. 114957–114964, 2021.

[10] S. Nah, T. H. Kim, and K. M. Lee, “Deep multi-scale convolutional neural network for dynamic scene deblurring,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3883–3891.

[11] A. Abdelhamed, S. Lin, and M. S. Brown, “A high-quality denoising dataset for smartphone cameras,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 1692–1700.

[12] J. Rim, H. Lee, J. Won, and S. Cho, “Real-world blur dataset for learning and benchmarking deblurring algorithms,” in *Proc. 16th Eur. Conf. Comput. Vision (ECCV)*, Aug. 2020, pp. 184–201.

[13] J. Lehtinen et al., “Noise2Noise: Learning image restoration without clean data,” in *Proc. Int. Conf. Mach. Learn.*, 2018, pp. 2965–2974.

[14] A. Krull, T.-O. Buchholz, and F. Jug, “Noise2void-learning denoising from single noisy images,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 2129–2137.

[15] J. Batson and L. Royer, “Noise2Self: Blind denoising by self-supervision,” in *Proc. Int. Conf. Mach. Learn.*, 2019, pp. 524–533.

[16] Y. Quan, M. Chen, T. Pang, and H. Ji, “Self2self with dropout: Learning self-supervised denoising from single image,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2020, pp. 1890–1898.

[17] T. Huang, S. Li, X. Jia, H. Lu, and J. Liu, “Neighbor2neighbor: Self-supervised denoising from single noisy images,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2021, pp. 14781–14790.

[18] N. Moran, D. Schmidt, Y. Zhong, and P. Coady, “Noisier2noise: Learning to denoise from unpaired noisy data,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2020, pp. 12064–12072.

[19] T. Pang, H. Zheng, Y. Quan, and H. Ji, “Recorrupted-to-recorrupted: Unsupervised deep learning for image denoising,” in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2021, pp. 2043–2052.

[20] J. Fehrenbach, P. Weiss, and C. Lorenzo, "Variational algorithms to remove stationary noise: Applications to microscopy imaging," *IEEE Trans. Image Process.*, vol. 21, no. 10, pp. 4420–4430, Oct. 2012.

[21] Y. Mäkinen, L. Azzari, and A. Foi, "Collaborative filtering of correlated noise: Exact transform-domain variance for improved shrinkage and patch matching," *IEEE Trans. Image Process.*, vol. 29, pp. 8339–8354, 2020.

[22] T. Liu, J. Cheng, and S. Tan, "Spectral Bayesian uncertainty for image super-resolution," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2023, pp. 18166–18175.

[23] K. Monakhova, S. R. Richter, L. Waller, and V. Koltun, "Dancing under the stars: Video denoising in starlight," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 16241–16251. [Online]. Available: https://openaccess.thecvf.com/content/CVPR2022/html/Monakhova_Dancing_Under_the_Stars_Video_Denoising_in_Starlight_CVPR_2022_paper.html

[24] Y. Chang, M. Chen, L. Yan, X.-L. Zhao, Y. Li, and S. Zhong, "Toward universal stripe removal via wavelet-based deep convolutional neural network," *IEEE Trans. Geosci. Remote Sens.*, vol. 58, no. 4, pp. 2880–2897, Apr. 2020.

[25] Z. He, Y. Cao, Y. Dong, J. Yang, Y. Cao, and C.-L. Tisse, "Single-image-based nonuniformity correction of uncooled long-wave infrared detectors: A deep-learning approach," *Appl. Opt.*, vol. 57, no. 18, p. D155, 2018.

[26] J. Huisken, J. Swoger, F. Del Bene, J. Wittbrodt, and E. H. K. Stelzer, "Optical sectioning deep inside live embryos by selective plane illumination microscopy," *Science*, vol. 305, no. 5686, pp. 1007–1009, Aug. 2004.

[27] D. Fuoli, L. Van Gool, and R. Timofte, "Fourier space losses for efficient perceptual image super-resolution," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 2360–2369.

[28] S.-J. Cho, S.-W. Ji, J.-P. Hong, S.-W. Jung, and S.-J. Ko, "Rethinking coarse-to-fine approach in single image deblurring," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 4641–4650.

[29] E. Y. Lam and J. W. Goodman, "A mathematical analysis of the DCT coefficient distributions for images," *IEEE Trans. Image Process.*, vol. 9, no. 10, pp. 1661–1666, Oct. 2000.

[30] G. Chen, F. Zhu, and P. A. Heng, "An efficient statistical method for image noise level estimation," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 477–485.

[31] A. Foi, M. Trimeche, V. Katkovnik, and K. Egiazarian, "Practical poissonian-Gaussian noise modeling and fitting for single-image raw-data," *IEEE Trans. Image Process.*, vol. 17, no. 10, pp. 1737–1754, Oct. 2008.

[32] J. Nakamura, *Image Sensors and Signal Processing for Digital Still Cameras*. Boca Raton, FL, USA: CRC Press, 2006.

[33] J. Zheng, J. A. Fessler, and H.-P. Chan, "Detector blur and correlated noise modeling for digital breast tomosynthesis reconstruction," *IEEE Trans. Med. Imag.*, vol. 37, no. 1, pp. 116–127, Jan. 2018.

[34] S. Soltanayev and S. Y. Chun, "Training deep learning based denoisers without ground truth data," in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 3257–3267.

[35] A. Bora, E. Price, and A. G. Dimakis, "AmbientGAN: Generative models from lossy measurements," in *Proc. Int. Conf. Learn. Representations*, 2018.

[36] T. Kaneko and T. Harada, "Noise robust generative adversarial networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2020, pp. 8401–8411.

[37] H. Liu, X. Liu, J. Lu, and S. Tan, "Self-supervised image prior learning with GMM from a single noisy image," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 2825–2834.

[38] V. Lempitsky, A. Vedaldi, and D. Ulyanov, "Deep image prior," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 9446–9454.

[39] G. Chen, P. Peng, L. Ma, J. Li, L. Du, and Y. Tian, "Amplitude-phase recombination: Rethinking robustness of convolutional neural networks in frequency domain," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 458–467.

[40] L. Jiang, B. Dai, W. Wu, and C. C. Loy, "Focal frequency loss for image reconstruction and synthesis," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV)*, Oct. 2021, pp. 13919–13929.

[41] W. Li, H. Guo, Y. Hou, G. Gao, and Z. Ma, "Dual-domain modulation network for lightweight image super-resolution," *IEEE Trans. Multimedia*, early access, Feb. 16, 2026, doi: 10.1109/TMM.2026.3660149.

[42] W. Li, H. Guo, Y. Hou, and Z. Ma, "FourierSR: A Fourier token-based plugin for efficient image super-resolution," *IEEE Trans. Image Process.*, vol. 35, pp. 732–742, 2026.

[43] B. Mandracchia, X. Hua, C. Guo, J. Son, T. Urner, and S. Jia, "Fast and accurate sCMOS noise correction for fluorescence microscopy," *Nature Commun.*, vol. 11, no. 1, p. 94, Jan. 2020.

[44] L. Song and H. Huang, "Simultaneous destriping and image denoising using a nonparametric model with the EM algorithm," *IEEE Trans. Image Process.*, vol. 32, pp. 1065–1077, 2023.

[45] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image restoration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 7, pp. 2480–2495, Jul. 2021.

[46] A. Levin, Y. Weiss, F. Durand, and W. T. Freeman, "Understanding and evaluating blind deconvolution algorithms," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 1964–1971.

[47] E. Agustsson and R. Timofte, "NTIRE 2017 challenge on single image super-resolution: Dataset and study," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 126–135.

[48] M. Bevilacqua, A. Roumy, C. Guillemot, and M.-L.-A. Morel, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," in *Proc. Procedings Brit. Mach. Vis. Conf.*, 2012, pp. 135.1–135.10.

[49] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *Proc. Int. Conf. curves Surf.*, 2010, pp. 711–730.

[50] X. Wu, "Color demosaicking by local directional interpolation and nonlocal adaptive thresholding," *J. Electron. Imag.*, vol. 20, no. 2, Apr. 2011, Art. no. 023016.

[51] J.-B. Huang, A. Singh, and N. Ahuja, "Single image super-resolution from transformed self-exemplars," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 5197–5206.

[52] Y. Matsui et al., "Sketch-based Manga retrieval using Manga109 dataset," *Multimedia Tools Appl.*, vol. 76, no. 20, pp. 21811–21838, Oct. 2017.

[53] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced deep residual networks for single image super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jul. 2017, pp. 136–144.

[54] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 286–301.

[55] R. Timofte et al., "NTIRE 2017 challenge on single image super-resolution: Methods and results," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jul. 2017, pp. 114–125.

[56] C. Chen, Q. Chen, J. Xu, and V. Koltun, "Learning to see in the dark," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 3291–3300.

[57] H. Li, Z. Zhang, T. Jiang, P. Luo, H. Feng, and Z. Xu, "Real-world deep local motion deblurring," in *Proc. AAAI Conf. Artif. Intell.*, 2023, vol. 37, no. 1, pp. 1314–1322.

[58] C. Qu, Z. Chen, J. Zhang, X. Chen, and J. Han, "Self-BSR: Self-supervised image denoising and destriping based on blind-spot regularization," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 35, no. 9, pp. 8666–8678, Sep. 2025.

[59] W. Lee, S. Son, and K. M. Lee, "AP-BSN: Self-supervised denoising for real-world images via asymmetric PD and blind-spot network," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2022, pp. 17725–17734.

[60] J. Li et al., "Spatially adaptive self-supervised learning for real-world image denoising," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2023, pp. 9914–9924.

[61] S. Chen, J. Zhang, Z. Yu, and T. Huang, "Exploring efficient asymmetric blind-spots for self-supervised denoising in real-world scenarios," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2024, pp. 2814–2823.

[62] V. Saragadam, A. Dave, A. Veeraraghavan, and R. G. Baraniuk, "Thermal image processing via physics-inspired deep networks," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshops (ICCVW)*, Oct. 2021, pp. 4040–4048.

[63] X. Li, G. Li, and B. Zhao, "Low-light hyperspectral image enhancement," *IEEE Trans. Geosci. Remote Sens.*, vol. 60, Aug. 2022, Art. no. 5537013, doi: 10.1109/TGRS.2022.3201206.