RECORRUPTED-TO-NOISE: A UNIFYING FRAMEWORK FOR RECORRUPTION-BASED SELF-SUPERVISED DENOISING

Anonymous authors
Paper under double-blind review

ABSTRACT

Self-supervised learning has emerged as an effective paradigm for image denoising when clean ground-truth data are expensive or unavailable. Existing approaches often rely on restrictive assumptions regarding noise independence or distribution, limiting their applicability across imaging scenarios. In this paper, we introduce a *Recorrupted-to-Noise* (R2N) self-supervised denoising framework, which offers a unified theoretical interpretation of recorruption-based denoising methods. R2N recorrupts noisy observations and trains a network to predict the recorruption noise. The validity of R2N rests on conditional projective equivalence (CPE) between the observation noise and the recorruption noise, a weaker condition than those required by many existing self-supervised denoising methods. This condition supports a broad range of noise models, including general additive noise, natural exponential-family models, and certain multiplicative noise models. We also develop a test-time strategy for non-uniform noise with unknown parameters. We further analyze the effect of the recorruption coefficient on denoising error and establish a robustness bound under observation-model mismatch. Extensive experiments demonstrate that R2N outperforms existing self-supervised methods in challenging scenarios involving broader noise models and unknown noise parameters.

1 INTRODUCTION

Image denoising is a fundamental problem in image processing and underpins natural and scientific imaging applications (Parhad et al., 2024; Belthangady & Royer, 2019; Bendory et al., 2020; Zeng et al., 2024; Izadi et al., 2023). In many applications, acquiring clean images is costly, time-consuming, or even infeasible. This motivates the development of blind denoising methods, which aim to recover clean signals directly from noisy observations (Lehtinen et al., 2018; Monroy et al., 2025; Qu et al., 2024; Bepler et al., 2020).

One major line of work on blind denoising relies on paired noisy observations. Representative works include Noise2Noise (N2N) (Lehtinen et al., 2018) and its variants (Calvarons, 2021), which assume access to two conditionally independent noisy measurements of the same scene. When the noises have zero conditional mean and are independent given the underlying clean image, theoretical analysis shows that using one noisy image as input and the other as the target yields the same MSE-optimal predictor as training with clean targets (Lehtinen et al., 2018).

To reduce reliance on paired observations, recorruption-based methods construct training input–target pairs by further corrupting noisy observations through masking or additional noise. These methods can be broadly divided into two classes according to their mathematical foundations. Mask-based methods predict selected pixels or regions from their surrounding context, relying on pixel-wise noise independence and spatial correlations in the clean signal. Typical examples include Noise2Self (Batson & Royer, 2019), Noise2Void (Ji et al., 2025; Krull et al., 2019), and Neighbor2Neighbor (Huang et al., 2021), as well as their variants (Wang et al., 2022; Yu et al., 2023). Such context-based prediction can be implemented through blind-spot constraints (Krull et al., 2019) or input-preprocessing strategies (Huang et al., 2021; Batson & Royer, 2019).

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

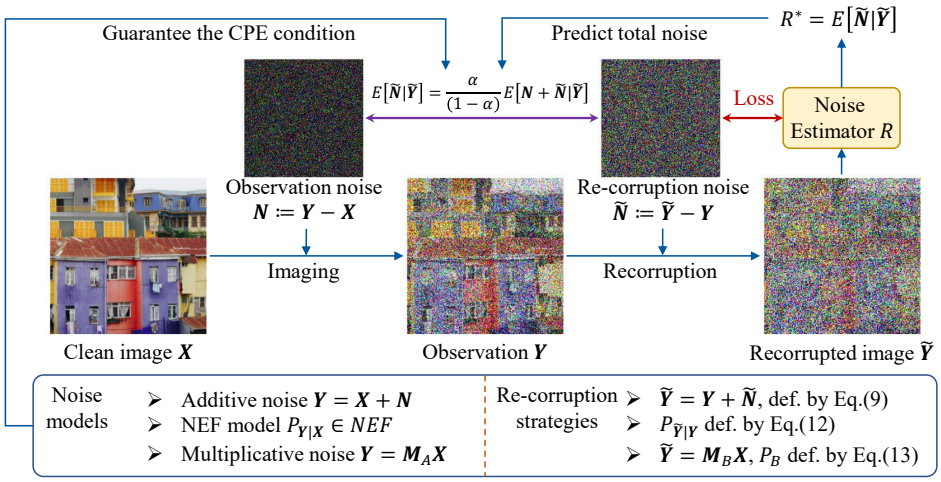

Figure 1: Overview of the R2N framework. R2N recorrupts noisy observations, and a neural network is trained to predict the recorruption noise. Under the CPE condition, the conditional expectations of the recorruption noise and the observation noise are proportional, which allows the predicted recorruption noise to be rescaled into an estimate of the total noise.

The second class is grounded in probabilistic and statistical estimation theory, including SURE-based approaches (Kim et al., 2020), Noise2Score (N2S) (Kim & Ye, 2021; Wei et al., 2025), Recorrupted-to-Recorrupted (R2R) (Pang et al., 2021), and its generalization GR2R (Krull et al., 2024; Monroy et al., 2025). These approaches provide theoretical guarantees for denoising under specific noise distributions. For instance, N2S uses Tweedie's formula to handle Gaussian noise exactly and Poisson or Gamma noise approximately (Kim & Ye, 2021). R2R and GR2R construct pseudo-independent pairs by recorrupting noisy inputs, extending their applicability to noise distributions in the natural exponential family (NEF) (Pang et al., 2021; Monroy et al., 2025). Despite these theoretical guarantees, their applicability depends on noise-models assumptions, limiting their use for more general observation models.

In this work, we propose a unifying framework for recorruption-based self-supervised denoising, *Recorrupted-to-Noise* (R2N). The core idea is to recorrupt noisy images and train a neural network to estimate the recorruption noise (Fig. 1). When the recorruption noise and observation noise satisfy *conditional projective equivalence* (CPE), the predicted noise can be rescaled and subtracted from the recorrupted observation to recover the posterior mean of the clean image. CPE is a weaker condition than those required by many existing self-supervised denoising methods. R2N therefore applies to a broad range of noise models, including general additive noise, NEF observation models, and certain multiplicative noise models. We also develop a test-time strategy for non-uniform noise with unknown parameters. Our analysis shows that the mainstream recorruption-based denoising methods can be viewed as special cases of the R2N framework. We further analyze the effect of the recorruption coefficient on denoising error and the error caused by observation-model mismatch. Experiments demonstrate that R2N effectively handles complex noise settings and achieves competitive performance across diverse conditions, with substantial gains in challenging noise settings.

2 Related Work

Supervised learning The supervised paradigm assumes access to paired samples consisting of a clean image X and a noisy observation Y. A standard training objective is the mean-squared error (MSE):

$$\mathcal{L}_{\mathrm{SUP}}(F) = \mathbb{E}\left[\|F(Y) - X\|^2\right], \tag{1}$$

where F denotes the denoising network. Assuming finite second moments, the minimizer of (1) is the conditional expectation of X given Y, i.e., $F^*_{\mathrm{SUP}} = \mathbb{E}[X|Y]$. Thus, even with perfect ground-truth supervision, the best achievable estimator in the MSE sense is the posterior mean, not the exact noise-free realization.

N2N The N2N framework (Lehtinen et al., 2018) trains from paired noisy observations: for each clean image X, there are two corrupted observations Y_i, $i = 1, 2$. The corresponding noises $N_i = Y_i - X$ for $i = 1, 2$ are assumed to have zero conditional mean and to be conditionally independent given X. This setting can arise when multiple frames of the same scene are available. N2N optimizes an MSE loss using one noisy view as input and the other as the target:

$$\mathcal{L}_{\mathrm{N2N}}(F) = \mathbb{E}\left[\|F(Y_1) - Y_2\|^2\right]. \tag{2}$$

Under these assumptions, the optimal predictor coincides with the supervised solution, i.e., $F^*_{\mathrm{N2N}} = \mathbb{E}[X|Y_1]$. Thus N2N recovers the posterior-mean estimator while avoiding the need for clean targets.

R2R-type methods R2R-type methods (Pang et al., 2021; Monroy et al., 2025) construct training pairs from a single noisy observation. Under certain noise models, R2R constructs a pair of recorrupted samples $\tilde{Y}_1, \tilde{Y}_2$ from the original noisy observation Y such that $\tilde{Y}_1$ and $\tilde{Y}_2$ have conditionally independent noises with zero conditional mean given the clean image, satisfying the N2N requirements. The training objective becomes

$$\mathcal{L}_{\mathrm{GR2R}}(F) = \mathbb{E}\left[\|F(\tilde{Y}_1) - \tilde{Y}_2\|^2\right] \tag{3}$$

which has the same form as $\mathcal{L}_{\mathrm{N2N}}$. R2R handles Gaussian corruption directly, while GR2R (Monroy et al., 2025) extends applicability to the NEF. GR2R also considers non-Gaussian additive noise, although its practical applicability is constrained by the complexity of pair construction and the restrictions associated with low-order denoising mappings.

Mask-based methods Mask-based methods, including Neighbor2Neighbor (Huang et al., 2021), Noise2Void (Krull et al., 2019), and Noise2Self (Batson & Royer, 2019), exploit spatial redundancy within a single noisy image. Their shared principle is to predict withheld pixels from the remaining observations. A representative masking objective is

$$\mathcal{L}_{\mathrm{mask}}(F) = \mathbb{E}_{Y,J}\left[\|F(Y_{J^c})_J - Y_J\|^2\right], \tag{4}$$

where J indexes withheld pixels and Y_{J^c} denotes the remaining observations. Under conditionally independent and zero-mean noise condition, the network learns to estimate clean pixel values from the surrounding observations. These approaches avoid specifying a parametric noise distribution but rely on spatial signal predictability and pixel-wise noise independence.

3 Main Contributions

In this section, we first introduce the R2N framework and the CPE condition, then describe CPE-admissible constructions for additive noise, NEF observation models, and multiplicative noise. Next, we establish connections to existing denoising methods and analyze the effect of the recorruption coefficient. Finally, we introduce a test-time coefficient estimation strategy to adapt R2N to unknown or non-uniform noise levels.

3.1 R2N Framework

R2N applies an additional corruption to a noisy observation and trains a network to predict the resulting recorruption noise. With a suitable observation model and recorruption strategy, this noise estimate can be converted into a denoised image.

Let $X \sim P_X(X)$ denote the unknown clean image, $Y \sim P_{Y|X}(Y|X)$ the noisy observation, and $N = Y - X$ the observation noise. Recorrupting the observation yields $\tilde{Y}$, where the additional recorruption noise is denoted by $\tilde{N} = \tilde{Y} - Y$. The CPE condition is defined as

$$\mathbb{E}[\tilde{N}|\tilde{Y}] = \frac{\alpha}{(1-\alpha)}\,\mathbb{E}[N|\tilde{Y}], \tag{5}$$

where $\alpha \in (0, 1)$ is a constant referred to as the *recorruption coefficient*. For a given observation model $P_{Y|X}$, we seek a recorruption kernel $P_{\tilde{Y}|Y}$ that satisfies CPE. We call $P_{Y|X}$ CPE-admissible when such a kernel exists.

The definition can be generalized to a matrix by viewing N and $\tilde{N}$ as vectors and replacing the scalar α with matrices. For clarity, we focus on the scalar case in this paper.

Given the CPE condition, R2N trains a neural network R to predict $\tilde{N}$ from $\tilde{Y}$ by minimizing the following objective:

$$\mathcal{L}_{\mathrm{R2N}}(R) = \mathbb{E}\left[\|R(\tilde{Y}) - \tilde{N}\|^2\right]. \tag{6}$$

The optimal predictor R^*_{R2N} is the minimum mean-squared error (MMSE) estimator of the recorruption noise:

$$R^*_{\mathrm{R2N}}(\tilde{Y}) = \mathbb{E}[\tilde{N}|\tilde{Y}]. \tag{7}$$

Under CPE, this estimator yields the posterior mean of the clean image:

$$\mathbb{E}[X|\tilde{Y}] = \tilde{Y} - \mathbb{E}[N + \tilde{N}|\tilde{Y}] = \tilde{Y} - \frac{1}{\alpha} R^*_{\mathrm{R2N}}(\tilde{Y}). \tag{8}$$

This is the same estimator obtained by supervised learning of X given $\tilde{Y}$ under the MSE loss.

3.2 Applicability of the CPE Condition

The R2N framework accommodates a range of observation models through suitable choices of the recorruption kernel.

R2N rests on the CPE condition, which requires a CPE-admissible observation model $P_{Y|X}$ and a compatible recorruption kernel $P_{\tilde{Y}|Y}$. Despite its concise form, Eq. 5 is an integral constraint whose full solution is difficult to characterize. We introduce three sets of additional assumptions that yield constructions for general additive noise, NEF observation models, and certain multiplicative noise models. The assumptions and main results are summarized below. The derivations and examples are provided in Appendix A, and Appendix B further analyzes the model mismatch problem.

Assumption A (corresponding to general additive noise): The observation noise N is independent of X, and the recorruption noise $\tilde{N}$ is independent of (X, N).

Suppose P_N is an n-divisible (Bose et al., 2002) distribution, i.e., N can be written as the sum of n independent and identically distributed (i.i.d.) random variables $N_i, i = 1, \cdots, n$. Let the recorruption noise be

$$\tilde{N} = \tilde{N}_1 + \cdots + \tilde{N}_k, \tag{9}$$

where $\tilde{N}_i$ are i.i.d. random variables with the same distribution as N_1. The CPE condition then holds with $\alpha = \frac{k}{n+k}$.

Every distribution is 1-divisible, so this construction can be applied to general additive noise with $k = n = 1$, where the recorruption is equivalent to adding an independent noise sample with the same distribution as the observation noise. In this case, the CPE condition can be easily verified by the linearity of conditional expectation.

Assumption B (corresponding to NEF): There exists an auxiliary random variable S such that $\mathbb{E}[S|X] = X$, and S is conditionally independent of $\tilde{Y}$ given X, with $Y = (1-\alpha)\tilde{Y} + \alpha S$.

This construction explicitly relates Y and $\tilde{Y}$. It directly implies CPE, but an implementable recorruption requires $X \to Y \to \tilde{Y}$ be a Markov chain. Specifically, $\tilde{Y}$ must be generated from Y without access to X. A sufficient condition is that

$$\begin{aligned} \log P_{\tilde{Y}|X}(\tilde{y}|x) &= (1-\alpha)\tilde{y}^T\phi(x) + \xi_1(\tilde{y}) + \eta_1(x), \\ \log P_{S|X}(s|x) &= \alpha s^T\phi(x) + \xi_2(s) + \eta_2(x), \end{aligned} \tag{10}$$

where $\phi, \xi_i, \eta_i, i = 1, 2$, are functions chosen so that $P_{\tilde{Y}|X}$ and $P_{S|X}$ are valid probability distributions. The induced observation model and recorruption kernel are

$$P_{Y|X}(y|x) = \exp\left(y^T\phi(x) + \eta_1(x) + \eta_2(x) + \log c_\alpha(y)\right), \tag{11}$$

$$P_{\tilde{Y}|Y}(\tilde{y}|y) \propto \exp\left(\xi_1(\tilde{y}) + \xi_2\left(\frac{y - (1-\alpha)\tilde{y}}{\alpha}\right)\right). \tag{12}$$

Here $c_\alpha(y)$ is a derived factor. Equation 11 indicates that the R2N framework is applicable when the observation model belongs to the NEF.

Assumption C (corresponding to multiplicative noise): Suppose $Y = M_A X, \tilde{Y} = M_B Y$, where M_A and M_B are random diagonal matrices whose i-th diagonal entries are $m_{A,i}, m_{B,i}$. $\{m_{A,i}\}$ and $\{m_{B,i}\}$ are each assumed to be i.i.d., mutually independent, and independent of the clean image.

In this case, let $\bar{m}_{A,k} = \mathbb{E}[m_{A,i}^k]$ and $\bar{m}_{B,k} = \mathbb{E}[m_{B,i}^k]$ be the k-th moments of $m_{A,i}$ and $m_{B,i}$, respectively. The CPE condition imposes the recursion

$$\bar{m}_{B,k+1} = \frac{\bar{m}_{A,k+1} - \alpha \bar{m}_{A,k}}{(1-\alpha)\bar{m}_{A,k+1}} \bar{m}_{B,k}, \quad \bar{m}_{B,0} = 1. \tag{13}$$

When the recursion defines a valid moment sequence, the multiplicative noise is CPE-admissible, and the sequence specifies the compatible recorruption multipliers.

The above constructions do not exhaust the scope of R2N, but they already cover a broad range of practical imaging models.

3.3 Unifying Mainstream Recorruption-based Denoising Methods

As introduced in Section 1, there are mainly two groups of recorruption-based self-supervised denoising methods: the mask-based methods and statistic-based methods. We relate these methods to R2N through objective transformations and a limiting interpretation of masking.

For the statistic-based methods, GR2R generalizes R2R, and SURE-based methods are limiting cases of it (Monroy et al., 2025). It therefore suffices to examine the relationship between GR2R and R2N. In GR2R, $\tilde{Y}_2 = \frac{1}{\alpha} Y - \frac{(1-\alpha)}{\alpha} \tilde{Y}_1$, so the loss in Eq. 3 can be rewritten as

$$\mathcal{L}_{\mathrm{GR2R}}(F) = \mathbb{E}\left[\| F(\tilde{Y}_1) - (\frac{1}{\alpha} Y - \frac{(1-\alpha)}{\alpha} \tilde{Y}_1) \|^2 \right] = \mathbb{E}\left[\| (F - \mathrm{Id})(\tilde{Y}_1) + \frac{1}{\alpha} \tilde{N}_1 \|^2 \right], \tag{14}$$

where $\tilde{N}_1 = \tilde{Y}_1 - Y$. This corresponds to R2N after identifying $\alpha(\mathrm{Id} - F)$ as R, with the losses related by the constant factor $1/\alpha^2$. Specifically, $\mathcal{L}_{\mathrm{GR2R}}(F) = \alpha^{-2} \mathcal{L}_{\mathrm{R2N}}(R)$.

Although the objectives are similar, GR2R constructs a conditionally independent pair $(\tilde{Y}_1, \tilde{Y}_2)$, which requires specific observation models, such as additive Gaussian noise or observation models in the NEF. R2N instead requires CPE, further admitting general additive noise and certain multiplicative noise models.

Mask-based methods predict withheld pixels from the remaining observations. One advantage is that the noise distribution need not be specified explicitly. These methods also admit a limiting interpretation within R2N. Under the additive-noise setting, consider different values of α at different pixels. Let $\alpha \to 1$ at masked pixels, corresponding formally to increasingly strong recorruption that suppresses their observed values. For visible pixels, let $\alpha = 0$ to retain the original observations. In practice, this loss of access to selected pixels is implemented by masking rather than by injecting unbounded noise, so no explicit noise distribution is required. This argument provides an intuitive connection between R2N and mask-based methods.

3.4 Influence of the Recorruption Coefficient

For some observation models, the recorruption coefficient α is not uniquely determined. The choice of α affects the overall denoising performance.

Define the total denoising error as the MSE between the reconstructed image and the clean image:

$$\ell_t(R;\alpha) \coloneqq \mathbb{E}\left[\| (\tilde{Y} - \frac{1}{\alpha} R(\tilde{Y})) - X \|^2 \right], \tag{15}$$

The posterior-mean orthogonality principle decomposes this error into two components: the irreducible error ℓ_i and the reducible error ℓ_r, defined as follows:

$$\ell_i(\alpha) \coloneqq \mathbb{E}[\| \mathbb{E}[X|\tilde{Y}] - X \|^2] \tag{16}$$

$$\ell_r(R;\alpha) \coloneqq \mathbb{E}\Big[\| (\tilde{Y} - \tfrac{1}{\alpha} R(\tilde{Y})) - \mathbb{E}[X|\tilde{Y}] \|^2 \Big] = \frac{1}{\alpha^2 (1-\alpha)^2} \mathbb{E}\Big[\| (1-\alpha) R(\tilde{Y}) - \alpha \mathbb{E}[N|\tilde{Y}] \|^2 \Big]. \tag{17}$$

For finite second moments, the decomposition satisfies $\ell_t(R;\alpha) = \ell_i(\alpha) + \ell_r(R;\alpha)$.

The irreducible error ℓ_i is intrinsic to denoising the recorrupted observation. As α increases, $\tilde{Y}$ typically contains more noise, making $\mathbb{E}[X|\tilde{Y}]$ less informative about X, and tending to increase ℓ_i.

The reducible error arises from the gap between the trained neural network and the theoretical optimum R^*_{R2N}, which can be reduced by improving the training process. Although the distribution of N is independent of α, the conditional mean $\mathbb{E}[N|\tilde{Y}]$ and the approximating error can be assumed to remain relatively stable across different α. The reducible error mainly scales with $\frac{1}{\alpha^2(1-\alpha)^2}$, suggesting that the reducible error follows a U-shaped trend as α varies in $(0, 1)$.

In summary, the selection of α reflects a trade-off between irreducible and reducible errors. Combining the above analysis, the irreducible error tends to increase with larger α, while the reducible error approximately exhibits a U-shaped trend. Consequently, the optimal α is expected to lie below 0.5. The analysis will be quantitatively verified in the experiments.

3.5 Test-Time Parameter Estimation

We propose an adaptive strategy that estimates the denoising coefficient from a test observation, allowing R2N to be used when the precise noise parameters are unknown.

For many noise types, particularly infinitely divisible distributions such as Gaussian and Cauchy (see Appendix A.1), R2N training does not require the full noise distribution. During training, the choice of α only affects the intensity parameters of the recorruption noise, such as σ for Gaussian or γ for Cauchy. Consequently, these noise parameters can be directly set without explicitly deriving them from the real noise and α. The exact value can be estimated at test time, for example, assuming approximate orthogonality between the predicted noise and the clean image, α can be estimated as

$$\frac{1}{\hat{\alpha}} = \frac{\langle R^*_{\mathrm{R2N}}(\tilde{Y}), \tilde{Y}\rangle}{\|R^*_{\mathrm{R2N}}(\tilde{Y})\|^2}, \tag{18}$$

where $\langle\cdot,\cdot\rangle$ denotes the inner product.

The same strategy can accommodate non-uniform noise parameters. For instance, in biological or biomedical images, colored noise can be modeled using approximately independent Gaussian Fourier coefficients with a frequency-dependent, bell-shaped power profile (Bendory et al., 2020). In such cases, Eq. (18) can be adapted for ring-wise estimation:

$$\frac{1}{\hat{\alpha}_r} = \frac{\mathrm{Re}\int_{\Omega_r} \overline{\widehat{R}(\omega)}\,\widehat{\tilde{Y}}(\omega)\,d\omega}{\int_{\Omega_r} |\widehat{R}(\omega)|^2\,d\omega}, \tag{19}$$

Here, $\widehat{R} = \mathcal{F}[R^*_{\mathrm{R2N}}(\tilde{Y})]$ and $\widehat{\tilde{Y}} = \mathcal{F}[\tilde{Y}]$ are Fourier transforms, and $\Omega_r = \{\omega : r - \delta \le \|\omega\| < r + \delta\}$ is a frequency annulus. The overline denotes complex conjugation.

This test-time calibration strategy simplifies the application of R2N to noise conditions with unknown and non-uniform parameters.

4 Experiments

To comprehensively evaluate the performance of the proposed R2N method, we compare it against both supervised and self-supervised denoising approaches, including PURE (Kim et al., 2020), Noise2Score (Kim & Ye, 2021), GR2R (Monroy et al., 2025), and Neighbor2Neighbor (Huang et al., 2021). The experiments cover a wide range of noise settings, including common distributions such as Gaussian, Poisson, and Gamma noise, as well as more challenging distributions, such as Cauchy and composite distributions. Evaluations are conducted on both grayscale and color images, with fastMRI (Knoll et al., 2020) images used for the grayscale experiments and DIV2K images for the color experiments. All methods employ a U-Net architecture to predict either the denoised image or the noise. We report the peak signal-to-noise ratio (PSNR) and structural similarity index measure (SSIM) as quantitative metrics.

Table 1: PSNR results on various noise distributions. R2N* denotes the R2N method with an estimated denoising coefficient.

Noise Distribution	PURE	N2S	GR2R	Nei2Nei	R2N* (ours)	R2N (ours)	SUP
Mul. Beta	-	-	-	15.79/0.77	31.82/0.90	**31.95/0.90**	31.16/0.91
Mul. Bernoulli	-	-	-	15.69/0.81	33.35/0.95	**33.57/0.95**	33.88/0.95
Composited	-	-	-	29.59/0.84	29.68/0.84	**30.08/0.87**	30.30/0.87
Cauchy	-	-	-	28.67/0.81	29.18/0.84	**29.26/0.85**	29.53/0.85
Gaussian	26.28/0.74	**27.53/0.80**	26.99/0.78	26.70/0.76	26.85/0.76	27.16/0.78	27.02/0.78
Poisson	21.70/0.46	23.78/0.56	27.10/0.79	26.84/0.78	26.85/0.77	**27.16/0.79**	27.17/0.80
Gamma	-	15.26/0.36	27.43/0.82	27.00/0.79	27.21/0.80	**27.46/0.82**	27.14/0.80

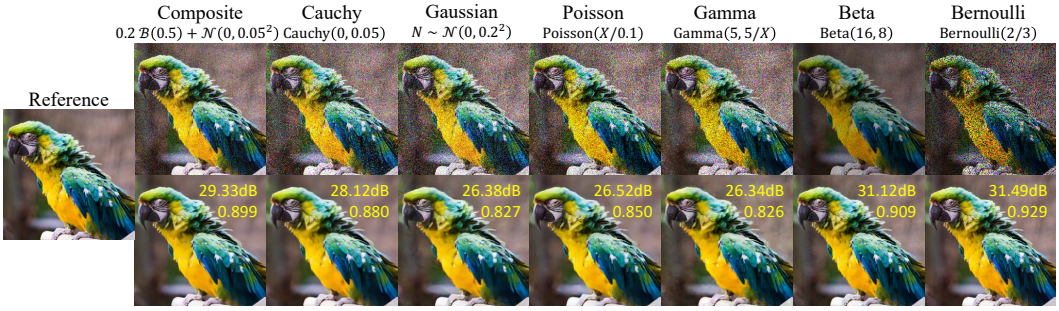

Figure 2: Denoising results of the R2N method under various noise distributions. The leftmost column shows the clean image and each remaining column represents a noise type. The top row shows the noisy observation, and the bottom row represents the denoising result produced by R2N. The PSNR and SSIM are reported in the top-right corner of each denoised image.

4.1 Evaluation under Diverse Noise Distributions

To assess the versatility of the proposed R2N framework, we conduct experiments across a wide range of noise distributions, including common noise models: Gaussian, Poisson, and Gamma, as well as some challenging scenarios: additive Cauchy noise, composite Bernoulli–Gaussian noise, and multiplicative Beta and Bernoulli noise.

The experiments are conducted on the DIV2K dataset. During training, the original images are randomly sampled and then corrupted by the simulated noise, whose distributions are annotated in Fig. 2. For each noise type, R2N is evaluated twice: once using fixed α and once using estimated α via Eq. (18). For NEF observation models, R2N employs the recorruption strategy introduced in Section 3.2 Assumption B with $\alpha = 0.2$. The explicit Poisson and Gamma kernels are given in Appendix A.2. For additive Cauchy noise, we adopt the strategy described in Section 3.2 Assumption A with $\frac{k}{n} = \frac{1}{2}$, using the Cauchy construction in Appendix A.1. The composite Bernoulli–Gaussian noise is treated as general additive noise, and the strategy in Section 3.2 Assumption A is applied by independently resampling the composite noise, as detailed in Appendix A.1. For the multiplicative settings, we consider Beta and Bernoulli noise. The corresponding recorruption multipliers are generated according to Eqs. (77) and (80) in Appendix A.3, respectively. In all GR2R experiments, α is fixed at 0.2.

Table 1 reports quantitative comparisons under different noise settings. For Gaussian noise, all approaches achieve strong performance comparable to supervised learning. For the other NEF models, R2N, GR2R, and the supervised baseline yield similar results, which is consistent with the theoretical analysis in Section 3.2 Assumption B. Beyond these standard noise models, we further consider additive Cauchy noise, composite Bernoulli–Gaussian noise, and multiplicative Beta and Bernoulli

noise. These challenging settings fall outside the assumptions of several model-specific baselines, including PURE, N2S, and GR2R, while Neighbor2Neighbor remains applicable as a general self-supervised baseline. Across all these settings, R2N consistently outperforms Neighbor2Neighbor and achieves performance comparable to supervised learning, with particularly substantial improvements under multiplicative noise. These results highlight the robustness and broad adaptability of R2N across diverse and complex noise scenarios.

Fig. 2 presents qualitative comparisons across different noise conditions. The results consistently demonstrate effective noise suppression, with corresponding PSNR and SSIM values annotated in the upper right corners.

4.2 Effect of the Recorruption Coefficient

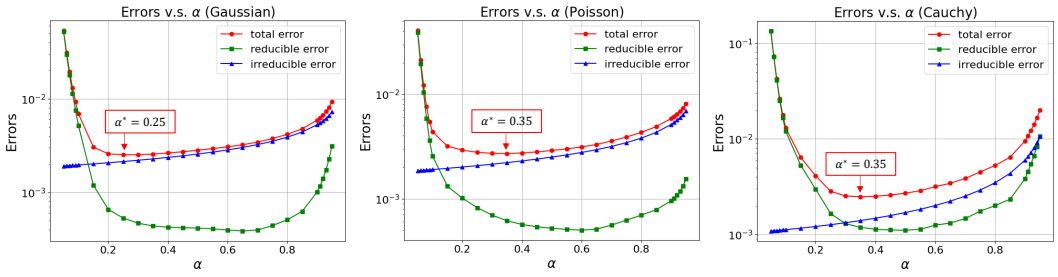

Figure 3: Trends of denoising error across different selections of the recorruption coefficient α. The red, green, and blue curves represent the total error, reducible error, and irreducible error, respectively. The best α corresponding to the smallest total error is annotated as α^* in the figure.

As discussed in Section 3.4, the recorruption coefficient α introduces a tradeoff between irreducible and reducible errors. To test the impact of α, we measure both errors across a range of α values under multiple noise settings.

We conduct this study on the DIV2K dataset. The original images are corrupted with Gaussian, Poisson, and additive Cauchy noise to simulate observations. The coefficient α is varied from 0.05 to 0.95, where a step size of 0.05 is used in the main region $(0.1, 0.9)$, while finer steps of 0.01 are applied for the boundary regions $(0.05, 0.1)$ and $(0.9, 0.95)$.

To compute the irreducible error as defined in Eq. (16), a supervised network is trained to predict $\mathbb{E}[X|\tilde{Y}]$. The same network is also used to calculate the reducible error defined in Eq. (17). The total error is then obtained by Eq. (15).

Fig. 3 illustrates the irreducible, reducible, and total errors across different values of α under the three noise settings. For all noise types, the irreducible error increases monotonically with α, whereas the reducible error exhibits a U-shaped behavior. The total error approximately follows the sum of the two components, dominated by the reducible error when α is close to zero and by the irreducible error in the central region. The α values corresponding to the lowest total error lie in the range $(0.2, 0.4)$, with the specific values indicated in Fig. 3. These trends align closely with the theoretical analysis presented in Section 3.4.

4.3 Unknown and Non-uniform Noise Levels

The final experiment evaluates the performance of the R2N framework under unknown and non-uniform noise levels. We use the fastMRI dataset (Knoll et al., 2020) for this study and simulate two types of non-uniform Gaussian noise. In the first scenario, the noise variance σ^2 varies across images within the range $(0.7\,\sigma_0, 1.3\,\sigma_0)$, while within each image the noise is i.i.d. across pixels. The second scenario considers colored Gaussian noise, whose Fourier coefficients are independent. Specifically, the noise variance follows $\sigma(r) = \sigma_0 \left(c_1 + c_2 \exp\left(-c_3 r^2\right)\right)$, where r denotes the frequency magnitude normalized by the Nyquist frequency. The constants are set as $c_1 = 0.7$, $c_2 = 0.6$, and $c_3 = 2$.

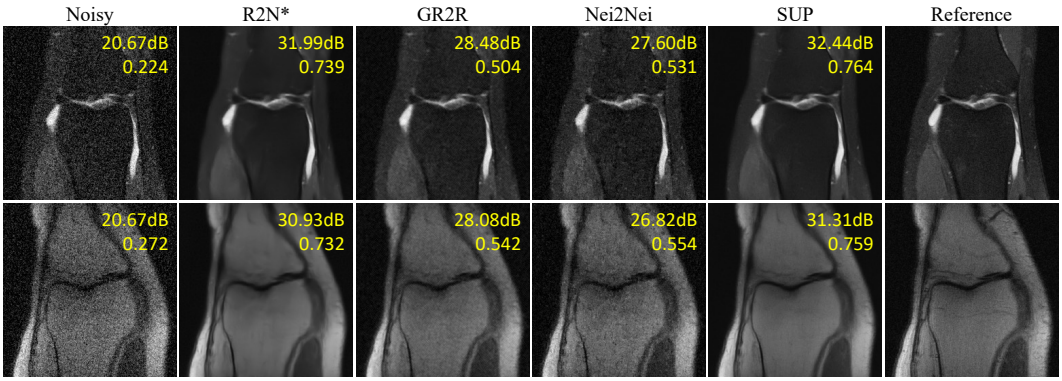

Figure 4: Denoising results under non-uniform noise settings. The first row shows results for i.i.d. Gaussian noise with unknown variance, and the second row shows the colored Gaussian case. The denoising performance metrics (PSNR and SSIM) are shown in the top-right corner of each image.

Table 2: PSNR results under various noise levels. R2N* denotes the R2N method with estimated noise levels.

Noise Level (σ_0)	GR2R	R2N* (ours)	R2N (ours)	SUP
0.1	28.79	**30.78**	31.08	31.32
0.2	24.54	**29.53**	29.49	29.76
0.3	21.44	**28.29**	28.19	29.00
0.5	17.66	**26.83**	27.05	27.82

Table 3: PSNR results under various levels for colored Gaussian noise. R2N* denotes the R2N method with estimated noise levels.

Noise Level (σ_0)	GR2R	Nei2Nei	R2N* (ours)	R2N (ours)	SUP
0.1	27.84	27.10	**30.26**	30.54	30.92
0.2	23.48	22.34	**28.81**	28.73	29.32
0.3	20.56	19.34	**27.44**	27.59	28.51
0.5	16.48	17.23	**25.77**	26.15	27.31

We assume that the average noise level $\bar{\sigma}$ over the dataset is available. During training, R2N adds i.i.d. Gaussian noise with variance $0.25\bar{\sigma}^2$ to the observations and trains the model to estimate the recorruption noise. At test time, $\hat{\alpha}$ is estimated using Eq. (18) and Eq. (19) for both scenarios. We also test R2N with the true α as a reference.

For comparison, we implement the GR2R method, using the average variance $\bar{\sigma}^2$ as the noise-level estimate. Supervised learning is also tested as a benchmark.

Figure 4 presents qualitative results, while Tables 2 and 3 report the corresponding PSNR values. R2N with estimated α achieves performance similar to R2N with true α, both approaching the supervised baseline across different noise settings, demonstrating the method's effectiveness. By contrast, GR2R performs poorly because it does not accommodate varying noise levels. Neighbor2Neighbor also yields inferior performance, as the colored Gaussian noise violates its pixel-wise independence assumption.

5 Conclusion

We presented R2N, a self-supervised denoising framework that trains on a recorrupted observation to predict the added noise. Under the CPE condition, this prediction can be rescaled to recover the posterior mean of the clean image. The condition is weaker the assumptions required in many existing self-supervised methods, supporting general additive and NEF observation models as well as certain multiplicative models. Our analysis shows how representative recorruption-based methods arise as special cases or limiting forms of R2N. We further characterized the trade-off induced by the recorruption coefficient and bounded the effect of observation-model mismatch. Experiments across diverse noise distributions, unknown noise levels, and non-uniform noise confirm the broad applicability of R2N and its consistent advantage in challenging settings. Future work will extend the CPE analysis to richer observation models and develop more flexible test-time adaptation strategies.

AI USE STATEMENT

Generative AI tools were used to assist with language polishing and code development. Language assistance included correcting grammar, refining wording, and improving clarity and readability. AI-assisted code development was conducted under human direction and supervision, and all AI-assisted code was manually reviewed by the authors. All AI-assisted text edits were reviewed to ensure that they preserved the intended meaning and technical accuracy. The authors take responsibility for the final content, analyses, and conclusions.

REFERENCES

Joshua Batson and Loic Royer. Noise2self: Blind denoising by self-supervision. In *International Conference on Machine Learning*, pp. 524–533. PMLR, 2019.

Chinmay Belthangady and Loïc A. Royer. Applications, promises, and pitfalls of deep learning for fluorescence image reconstruction. *Nature Methods*, 16(12):1215–1225, 2019.

Tamir Bendory, Alberto Bartesaghi, and Amit Singer. Single-particle cryo-electron microscopy: Mathematical theory, computational challenges, and opportunities. *IEEE Signal Processing Magazine*, 37(2):58–76, 2020.

Tristan Bepler, Kotaro Kelley, Alex J. Noble, and Bonnie Berger. Topaz-denoise: General deep denoising models for cryoem and cryoet. *Nature Communications*, 11(1):5208, 2020.

Arup Bose, Anirban Dasgupta, and Herman Rubin. A contemporary review and bibliography of infinitely divisible distributions and processes. *Sankhyā: The Indian Journal of Statistics, Series A*, pp. 763–819, 2002.

Adria Font Calvarons. Improved noise2noise denoising with limited data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 796–805, 2021.

Tao Huang, Songjiang Li, Xu Jia, Huchuan Lu, and Jianzhuang Liu. Neighbor2neighbor: Self-supervised denoising from single noisy images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14781–14790, 2021.

Saeed Izadi, Darren Sutton, and Ghassan Hamarneh. Image denoising in the deep learning era. *Artificial Intelligence Review*, 56(7):5929–5974, 2023.

Yongqiang Ji, Lan Xie, Gaozheng Xu, Zheng Zhu, Yu Gu, Jie Fei, Yuwei Li, Yao Chen, and Kuijie Cai. Sonar image denoising based on noise2void self-supervised learning. *Remote Sensing Letters*, 16(5):560–571, 2025.

Kwanyoung Kim and Jong Chul Ye. Noise2score: tweedie's approach to self-supervised image denoising without clean images. *Advances in Neural Information Processing Systems*, 34:864–874, 2021.

Kwanyoung Kim, Shakarim Soltanayev, and Se Young Chun. Unsupervised training of denoisers for low-dose ct reconstruction without full-dose ground truth. *IEEE Journal of Selected Topics in Signal Processing*, 14(6):1112–1125, 2020.

Florian Knoll, Jure Zbontar, Anuroop Sriram, Matthew J Muckley, Mary Bruno, Aaron Defazio, Marc Parente, Krzysztof J Geras, Joe Katsnelson, Hersh Chandarana, et al. fastmri: A publicly available raw k-space and dicom dataset of knee images for accelerated mr image reconstruction using machine learning. *Radiology: Artificial Intelligence*, 2(1):e190007, 2020.

Alexander Krull, Tim-Oliver Buchholz, and Florian Jug. Noise2void-learning denoising from single noisy images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2129–2137, 2019.

Alexander Krull, Hector Basevi, Benjamin Salmon, Andre Zeug, Franziska Müller, Samuel Tonks, Leela Muppala, and Aleš Leonardis. Image denoising and the generative accumulation of photons. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 1528–1537, 2024.

Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, and Timo Aila. Noise2Noise: Learning image restoration without clean data. In *Proceedings of the 35th International Conference on Machine Learning*, volume 80, pp. 2965–2974. PMLR, 2018.

Brayan Monroy, Jorge Bacca, and Julián Tachella. Generalized recorrupted-to-recorrupted: Self-supervised learning beyond gaussian noise. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 28155–28164, 2025.

Tongyao Pang, Huan Zheng, Yuhui Quan, and Hui Ji. Recorrupted-to-recorrupted: Unsupervised deep learning for image denoising. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2043–2052, 2021.

Saurabh Vijay Parhad, Krishna K Warhade, and Sanjay S Shitole. Speckle noise reduction in sar images using improved filtering and supervised classification. *Multimedia Tools and Applications*, 83(18):54615–54636, 2024.

Liying Qu, Shiqun Zhao, Yuanyuan Huang, Xianxin Ye, Kunhao Wang, Yuzhen Liu, Xianming Liu, Heng Mao, Guangwei Hu, Wei Chen, et al. Self-inspired learning for denoising live-cell super-resolution microscopy. *Nature Methods*, 21(10):1895–1908, 2024.

Zejin Wang, Jiazheng Liu, Guoqing Li, and Hua Han. Blind2unblind: Self-supervised image denoising with visible blind spots. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 2027–2036, 2022.

Xiangbin Wei, Yuanfeng Wang, Ao Xu, Lingyu Zhu, Dongyong Sun, Keren Li, Yang Li, and Qi Qin. Noise2score3d: Tweedie's approach for unsupervised point cloud denoising. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 25993–26003, 2025.

Xiaojun Yu, Chenkun Ge, Mingshuai Li, Miao Yuan, Linbo Liu, Jianhua Mo, Perry Ping Shum, and Jinna Chen. Self-supervised blind2unblind deep learning scheme for oct speckle reduction. *Biomedical Optics Express*, 14(6):2773–2795, 2023.

Xiangrui Zeng, Nibras Abdullah, and Putra Sumari. Self-supervised learning framework application for medical image analysis: A review and summary. *Biomedical Engineering Online*, 23(1):107, 2024.

APPENDICES

A PROOF OF THE CPE CONDITIONS

Let $X \sim P_X(X)$ denote the clean image and let $Y \sim P_{Y|X}(Y|X)$ denote the noisy observation. The recorrupted observation is denoted by $\tilde{Y}$. Define $N = Y - X$, $\tilde{N} = \tilde{Y} - Y$, and $N_T = N + \tilde{N}$.

We seek conditions under which the CPE condition in Eq. (5) holds for any distribution of X. Equivalently,

$$\mathbb{E}\left[\tilde{N} \mid \tilde{Y}\right] = \alpha \mathbb{E}\left[N + \tilde{N} \mid \tilde{Y}\right]. \tag{20}$$

Equivalently,

$$0 = \mathbb{E}\left[(1-\alpha)\tilde{N} - \alpha N \mid \tilde{Y}\right] = \mathbb{E}\left[\mathbb{E}\left[(1-\alpha)\tilde{N} - \alpha N \mid X, \tilde{Y}\right] \middle| \tilde{Y}\right]. \tag{21}$$

A sufficient condition is

$$\mathbb{E}\left[(1-\alpha)\tilde{N} - \alpha N \mid X, \tilde{Y}\right] \equiv 0. \tag{22}$$

Since $\tilde{N} = N_T - N$ and $N_T = N + \tilde{N} = \tilde{Y} - X$ is fixed when conditioning on $(X, \tilde{Y})$, Eq. (22) is equivalent to

$$\mathbb{E}\left[(1-\alpha)(N_T - N) - \alpha N \mid X, \tilde{Y}\right] = 0 \tag{23}$$

$$\Leftrightarrow \quad \mathbb{E}\left[N \mid X, \tilde{Y}\right] = (1-\alpha)N_T. \tag{24}$$

A.1 ASSUMPTION A: INDEPENDENT ADDITIVE NOISE ASSUMPTION.

For additive noise, Eq. (22) can first be written in a general conditional-density form. By Bayes' formula,

$$P_{N|X,\tilde{Y}}(n|X,\tilde{Y}) = \frac{P_{N|X}(n|X)P_{\tilde{Y}|X,N}(\tilde{Y}|X,n)}{P_{\tilde{Y}|X}(\tilde{Y}|X)}. \tag{25}$$

Thus, the equivalent equation becomes

$$\mathbb{E}\left[N \mid X, \tilde{Y}\right] = (1-\alpha)N_T \tag{26}$$

$$\Leftrightarrow \quad \int n \frac{P_{N|X}(n|X)P_{\tilde{Y}|X,N}(\tilde{Y}|X,n)}{P_{\tilde{Y}|X}(\tilde{Y}|X)} \mathrm{d}n = (1-\alpha)N_T \tag{27}$$

$$\Leftrightarrow \quad \int nP_{N|X}(n|X)P_{\tilde{Y}|X,N}(\tilde{Y}|X,n)\mathrm{d}n = (1-\alpha)N_T \int P_{N|X}(n|X)P_{\tilde{Y}|X,N}(\tilde{Y}|X,n)\mathrm{d}n \tag{28}$$

$$\Leftrightarrow \quad \int nP_{N|X}(n|X)P_{\tilde{N}|X,N}(N_T - n|X,n)\mathrm{d}n = (1-\alpha)N_T \int P_{N|X}(n|X)P_{\tilde{N}|X,N}(N_T - n|X,n)\mathrm{d}n. \tag{29}$$

Under the independent additive noise assumption, the original noise is independent of the clean image, $N \perp X$, and the recorruption noise is independently sampled, $\tilde{N} \perp (X, N)$. Therefore,

$$P_{N|X}(n|X) = P_N(n), \qquad P_{\tilde{N}|X,N}(\tilde{n}|X,n) = P_{\tilde{N}}(\tilde{n}). \tag{30}$$

The condition above then reduces to

$$\int nP_N(n)P_{\tilde{N}}(N_T-n)\mathrm{d}n$$
$$= (1-\alpha)N_T \int P_N(n)P_{\tilde{N}}(N_T-n)\mathrm{d}n \tag{31}$$
$$\Leftrightarrow \quad [(nP_N(n)) *_n (P_{\tilde{N}}(n))](N_T) = (1-\alpha)N_T\,[(P_N(n)) *_n (P_{\tilde{N}}(n))](N_T). \tag{32}$$

The equation holds almost surely for every value $N_T = n_T$, i.e.,

$$[(nP_N(n)) *_n (P_{\tilde{N}}(n))](n_T) = (1-\alpha)n_T\,[(P_N(n)) *_n (P_{\tilde{N}}(n))](n_T). \tag{33}$$

Apply Fourier transformation to both sides with respect to n_T, and denote $\hat{P}_\cdot$ as the Fourier transformation of $P_\cdot$. The left side equals

$$\mathcal{F}_n\left[nP_N(n)\right](\omega)\mathcal{F}_n\left[P_{\tilde{N}}(n)\right](\omega) = \frac{j}{2\pi}\hat{P}'_N(\omega)\hat{P}_{\tilde{N}}(\omega). \tag{34}$$

Similarly, the right-hand side equals

$$(1-\alpha)\frac{j}{2\pi}\frac{d}{d\omega}\left(\hat{P}_N(\omega)\hat{P}_{\tilde{N}}(\omega)\right) = (1-\alpha)\frac{j}{2\pi}\left[\hat{P}'_N(\omega)\hat{P}_{\tilde{N}}(\omega) + \hat{P}_N(\omega)\hat{P}'_{\tilde{N}}(\omega)\right]. \tag{35}$$

Assume that the characteristic functions are differentiable and nonzero on the frequency region under consideration. Hence, the condition becomes

$$\hat{P}'_N(\omega)\hat{P}_{\tilde{N}}(\omega) = (1-\alpha)\left[\hat{P}'_N(\omega)\hat{P}_{\tilde{N}}(\omega) + \hat{P}_N(\omega)\hat{P}'_{\tilde{N}}(\omega)\right] \tag{36}$$
$$\Leftrightarrow \quad \alpha\hat{P}'_N(\omega)\hat{P}_{\tilde{N}}(\omega) = (1-\alpha)\hat{P}_N(\omega)\hat{P}'_{\tilde{N}}(\omega) \tag{37}$$
$$\Leftrightarrow \quad \frac{\hat{P}'_{\tilde{N}}(\omega)}{\hat{P}_{\tilde{N}}(\omega)} = \frac{\alpha}{1-\alpha}\frac{\hat{P}'_N(\omega)}{\hat{P}_N(\omega)} \tag{38}$$
$$\Leftrightarrow \quad \left(\log\hat{P}_{\tilde{N}}(\omega)\right)' = \frac{\alpha}{1-\alpha}\left(\log\hat{P}_N(\omega)\right)' \tag{39}$$
$$\Leftrightarrow \quad \hat{P}_{\tilde{N}}(\omega) = \hat{P}_N(\omega)^{\frac{\alpha}{1-\alpha}}, \quad \text{a.e.} \tag{40}$$

where the integration constant is zero because both characteristic functions equal 1 at $\omega = 0$. The equality is understood on zero-free regions of the characteristic functions.

This is the required divisible additive-noise relation. It is admissible when the right-hand side is a valid characteristic function. If N is composed of n independent root components with characteristic function $\varphi(\omega)$, i.e.,

$$\hat{P}_N(\omega) = \varphi(\omega)^n, \tag{41}$$

then choosing $\alpha = \frac{k}{n+k}$ gives

$$\hat{P}_{\tilde{N}}(\omega) = \hat{P}_N(\omega)^{k/n} = \varphi(\omega)^k, \tag{42}$$

which means that the recorruption noise consists of k independent root components. The special case $k = n$ gives $\alpha = 0.5$ and recovers the same-distribution additive construction.

Concrete additive-noise examples. The same-distribution construction applies, for example, to composite Bernoulli–Gaussian noise. Let $N_i = bB_i + G_i$, where $B_i \sim \mathrm{Bernoulli}(p)$ and $G_i \sim \mathcal{N}(0,\sigma^2)$ are independent, and draw an independent copy $\tilde{N}_i$ for recorruption. Exchangeability of N and $\tilde{N}$ gives

$$\mathbb{E}[\tilde{N}|\tilde{Y}] = \mathbb{E}[N|\tilde{Y}], \qquad \alpha = \tfrac{1}{2}, \tag{43}$$
$$\mathbb{E}[X|\tilde{Y}] = \tilde{Y} - 2R^*_{\mathrm{R2N}}(\tilde{Y}). \tag{44}$$

This construction requires neither a zero noise mean nor divisibility beyond $n = 1$.

Table 4: Infinitely divisible noise distributions and their n-th convolution roots. The Gaussian distribution is parameterized by its variance; γ and c are scale parameters.

Distribution	Density and n-th convolution root
Gaussian	$f_G(z;\mu,\sigma^2) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(z-\mu)^2}{2\sigma^2}\right)$ $V_j \sim \mathcal{N}(\mu/n, \sigma^2/n)$
Cauchy	$f_C(z;x_0,\gamma) = \frac{1}{\pi\gamma}\left[1+\left(\frac{z-x_0}{\gamma}\right)^2\right]^{-1}$ $V_j \sim \text{Cauchy}(x_0/n, \gamma/n)$
Lévy	$f_L(z;\mu,c) = \sqrt{\frac{c}{2\pi}}\frac{\exp\left(-\frac{c}{2(z-\mu)}\right)}{(z-\mu)^{3/2}}, \quad z>\mu$ $V_j \sim \text{Lévy}(\mu/n, c/n^2)$

Infinitely divisible distributions allow more flexible choices. Table 4 lists the Gaussian, Cauchy, and Lévy examples and their convolution roots: if $V_1, \ldots, V_n$ are independent with the listed root distribution, then their sum has the original distribution.

Writing $\rho = \alpha/(1-\alpha)$, the corresponding independent recorruption laws are

$$N_i \sim \mathcal{N}(\mu,\sigma^2): \quad \tilde{N}_i \sim \mathcal{N}(\rho\mu, \rho\sigma^2), \tag{45}$$

$$N_i \sim \text{Cauchy}(x_0,\gamma): \quad \tilde{N}_i \sim \text{Cauchy}(\rho x_0, \rho\gamma), \tag{46}$$

$$N_i \sim \text{Lévy}(\mu,c): \quad \tilde{N}_i \sim \text{Lévy}(\rho\mu, \rho^2 c). \tag{47}$$

These laws are well defined for every $\rho > 0$, extending the integer-root construction $\rho = k/n$ to arbitrary $\alpha \in (0,1)$. For centered Gaussian noise, the recorruption variance is $\alpha\sigma^2/(1-\alpha)$; for centered Cauchy noise, its scale is $\alpha\gamma/(1-\alpha)$. In particular, the Cauchy setting $k/n = 1/2$ in the experiments gives $\alpha = 1/3$ and recorruption scale $\gamma/2$. For heavy-tailed laws such as Cauchy and Lévy, divisibility alone does not ensure a finite population MSE; conditional-mean identities and MSE optimality require their respective integrability conditions.

A.2 Assumption B: Construction assumption.

In the original setting, $\tilde{Y}$ is generated from Y, so $P_{\tilde{Y}|X,Y}(\tilde{y}|x,y)$ is independent of x. We introduce a constructive relation between $\tilde{Y}$ and Y and derive a sufficient condition for this independence. Assume that

$$Y = (1-\alpha)\tilde{Y} + \alpha S \tag{48}$$

where S is an auxiliary random variable conditionally independent of $\tilde{Y}$ given X, and the conditional distribution $P_{S|X}(s|x)$ satisfies

$$\mathbb{E}[S|X] = X. \tag{49}$$

Under this assumption, a sufficient conditional form of the CPE identity follows directly:

$$\mathbb{E}[\tilde{N}|X,\tilde{Y}] = \alpha\mathbb{E}[\tilde{Y}-S|X,\tilde{Y}] = \alpha(\tilde{Y}-X) = \alpha\mathbb{E}[N+\tilde{N}|X,\tilde{Y}]. \tag{50}$$

The remaining part is to find the conditions when the independence of $P_{\tilde{Y}|X,Y}$ holds. By Bayes' formula,

$$P_{\tilde{Y}|X,Y}(\tilde{y}|x,y) = \frac{P_{\tilde{Y}|X}(\tilde{y}|x)P_{Y|X,\tilde{Y}}(y|x,\tilde{y})}{P_{Y|X}(y|x)} \tag{51}$$

$$\propto \frac{P_{\tilde{Y}|X}(\tilde{y}|x)P_{S|X}\left(\frac{y-(1-\alpha)\tilde{y}}{\alpha}\middle|x\right)}{P_{Y|X}(y|x)}, \tag{52}$$

where the Jacobian factor depends only on α and is absorbed by normalization. The equation is the normalization of the numerator with respect to $\tilde{y}$.

A sufficient condition for $P_{\tilde{Y}|X,Y}(\tilde{y}|x,y)$ to be independent of x is that

$$P_{\tilde{Y}|X}(\tilde{y}|x)P_{S|X}\left(\frac{y-(1-\alpha)\tilde{y}}{\alpha}\middle|x\right) = a(x,y)b(y,\tilde{y}). \tag{53}$$

The reason is $a(x,y)$ will be canceled out in both the numerator and denominator. In the exponential perspective, again, there is a straightforward sufficient condition that

$$\log(P_{\tilde{Y}|X}(\tilde{y}|x)) = (1-\alpha)\tilde{y}^T\phi(x) + \xi_1(\tilde{y}) + \eta_1(x) \tag{54}$$

$$\log(P_{S|X}(s|x)) = \alpha s^T\phi(x) + \xi_2(s) + \eta_2(x) \tag{55}$$

then we have

$$\begin{aligned}&\log(P_{\tilde{Y}|X}(\tilde{y}|x)) + \log\left(P_{S|X}\left(\frac{y-(1-\alpha)\tilde{y}}{\alpha}\middle|x\right)\right)\\ &= y^T\phi(x) + \xi_1(\tilde{y}) + \eta_1(x) + \xi_2\left(\frac{y-(1-\alpha)\tilde{y}}{\alpha}\right) + \eta_2(x)\end{aligned} \tag{56}$$

which can be split as

$$\log a(x,y) = y^T\phi(x) + \eta_1(x) + \eta_2(x) \tag{57}$$

$$\log b(y,\tilde{y}) = \xi_1(\tilde{y}) + \xi_2\left(\frac{y-(1-\alpha)\tilde{y}}{\alpha}\right) \tag{58}$$

In this case, the corresponding distribution of Y given X is

$$P_{Y|X}(y|x) = \int a(x,y)b(y,\tilde{y})d\tilde{y} \tag{59}$$

$$= a(x,y)\int b(y,\tilde{y})d\tilde{y} \tag{60}$$

$$= a(x,y)c_\alpha(y) \tag{61}$$

$$= \exp(y^T\phi(x) + \eta_1(x) + \eta_2(x) + \log c_\alpha(y)) \tag{62}$$

where $c_\alpha(y) \coloneqq \int b(y,\tilde{y})d\tilde{y}$ absorbs the α-dependent constants. The result indicates that $P_{Y|X}$ is a natural exponential family distribution.

Poisson observations. For photon-counting observations $Y_i|X_i \sim \text{Poisson}(X_i)$, the kernel in Eq. (12) becomes binomial thinning:

$$U_i|Y_i \sim \text{Binomial}(Y_i, 1-\alpha), \qquad \tilde{Y}_i = \frac{U_i}{1-\alpha}. \tag{63}$$

Indeed, given X_i, Poisson splitting makes U_i and $V_i = Y_i - U_i$ independent Poisson variables with means $(1-\alpha)X_i$ and αX_i. Defining $S_i = V_i/\alpha$ gives $Y_i = (1-\alpha)\tilde{Y}_i + \alpha S_i$ and $\mathbb{E}[S_i|X_i] = X_i$, as required by Assumption B. For scaled counts $Y_i = \tau K_i$ with $K_i|X_i \sim \text{Poisson}(X_i/\tau)$ and known $\tau > 0$, the same construction gives

$$U_i|Y_i \sim \text{Binomial}(Y_i/\tau, 1-\alpha), \qquad \tilde{Y}_i = \frac{\tau U_i}{1-\alpha}. \tag{64}$$

Gamma observations. The Gamma setting in the experiments admits an analogous split. Let $Y_i|X_i \sim \text{Gamma}(L, L/X_i)$ for $X_i > 0$, using the shape–rate parameterization, so that $\mathbb{E}[Y_i|X_i] = X_i$. Draw B_i independently of (X_i, Y_i) and set

$$B_i \sim \text{Beta}((1-\alpha)L, \alpha L), \qquad \tilde{Y}_i = \frac{B_i Y_i}{1-\alpha}. \tag{65}$$

By the beta–gamma identity, $U_i = B_i Y_i$ and $V_i = (1-B_i)Y_i$ are conditionally independent given X_i, with

$$U_i|X_i \sim \text{Gamma}((1-\alpha)L, L/X_i), \qquad V_i|X_i \sim \text{Gamma}(\alpha L, L/X_i). \tag{66}$$

Thus $S_i = V_i/\alpha$ again satisfies Assumption B. Both the Poisson and Gamma kernels can be sampled using only Y_i and the noise parameters, without access to X_i. The Gaussian kernel is the independent additive construction above with zero mean and recorruption variance $\alpha\sigma^2/(1-\alpha)$.

A.3 Assumption C: Multiplicative noise assumption.

We next consider the coordinatewise multiplicative model used in the main text. For a fixed coordinate i, write

$$Y_i = m_{A,i}X_i, \qquad \tilde{Y}_i = m_{B,i}Y_i = m_{A,i}m_{B,i}X_i, \qquad t_i = m_{A,i}m_{B,i}, \tag{67}$$

where $m_{A,i}$ is the original multiplicative noise, $m_{B,i}$ is an independent recorruption multiplier, and t_i is the total multiplier. We assume that $m_{A,i}$ is independent of X_i, and that $m_{B,i}$ is independently sampled from both X_i and $m_{A,i}$. Assume that, for some $\alpha \in (0,1)$,

$$\mathbb{E}[m_{A,i}|t_i] = \alpha + (1-\alpha)t_i \tag{68}$$

almost surely. Since $\tilde{N}_i = \tilde{Y}_i - Y_i = (t_i - m_{A,i})X_i$ and $N_i + \tilde{N}_i = \tilde{Y}_i - X_i = (t_i - 1)X_i$, conditioning on $(X_i, \tilde{Y}_i)$ fixes $t_i = \tilde{Y}_i/X_i$ whenever $X_i \neq 0$. When $X_i = 0$, both sides below vanish. Therefore

$$\mathbb{E}[\tilde{N}_i|X_i, \tilde{Y}_i] = (t_i - \mathbb{E}[m_{A,i}|t_i])X_i = \alpha(t_i - 1)X_i = \alpha\mathbb{E}[N_i + \tilde{N}_i|X_i, \tilde{Y}_i]. \tag{69}$$

Applying the same argument coordinatewise and then taking conditional expectation again with respect to $\tilde{Y}$ yields the CPE identity

$$\mathbb{E}[\tilde{N}|\tilde{Y}] = \alpha\mathbb{E}[N + \tilde{N}|\tilde{Y}]. \tag{70}$$

Thus, Eq. (68) is a sufficient product-domain condition for multiplicative noise.

The same condition can be written as an integral equation. Suppose $m_{A,i}, m_{B,i} > 0$ have densities $P_{m_{A,i}}$ and $P_{m_{B,i}}$. The density of $t_i = m_{A,i}m_{B,i}$ is the Mellin convolution

$$P_{t_i}(t) = \int_0^\infty P_{m_{A,i}}(a)P_{m_{B,i}}(t/a)\frac{da}{a}. \tag{71}$$

Using the conditional density of $m_{A,i}$ given $t_i = t$, Eq. (68) is equivalent to

$$\int_0^\infty (a - \alpha - (1-\alpha)t)\, P_{m_{A,i}}(a)P_{m_{B,i}}(t/a)\frac{da}{a} = 0 \tag{72}$$

for all t in the support of t_i. Hence, after choosing the original multiplier distribution $P_{m_{A,i}}$, the recorruption distribution $P_{m_{B,i}}$ is determined by a Mellin-convolution equation.

A useful equivalent characterization is obtained from moments. Define $\bar{m}_{A,k} = \mathbb{E}[m_{A,i}^k]$ and $\bar{m}_{B,k} = \mathbb{E}[m_{B,i}^k]$, with $\bar{m}_{B,0} = 1$, and assume the required moments are finite. Multiplying Eq. (68) by t_i^k and taking expectation gives

$$\bar{m}_{A,k+1}\bar{m}_{B,k} = \alpha\bar{m}_{A,k}\bar{m}_{B,k} + (1-\alpha)\bar{m}_{A,k+1}\bar{m}_{B,k+1}. \tag{73}$$

Therefore, for a prescribed $m_{A,i}$ and coefficient α, the moments of the required recorruption multiplier must satisfy

$$\bar{m}_{B,k+1} = \frac{\bar{m}_{A,k+1} - \alpha\bar{m}_{A,k}}{(1-\alpha)\bar{m}_{A,k+1}}\,\bar{m}_{B,k}, \qquad k \geq 0, \tag{74}$$

or equivalently

$$\bar{m}_{B,k} = \prod_{\ell=0}^{k-1} \frac{\bar{m}_{A,\ell+1} - \alpha\bar{m}_{A,\ell}}{(1-\alpha)\bar{m}_{A,\ell+1}}. \tag{75}$$

The sequence $\{\bar{m}_{B,k}\}_{k\geq 0}$ must be a valid moment sequence. In particular, if $m_{B,i}$ is required to lie in $[0,1]$, then $\{\bar{m}_{B,k}\}$ must satisfy the Hausdorff moment condition $(-1)^r\Delta^r\bar{m}_{B,k} \geq 0$ for all $r, k \geq 0$, which also guarantees uniqueness on $[0,1]$. More generally, for $m_{B,i} \in \mathbb{R}_+$, one needs a valid Stieltjes moment sequence together with moment determinacy.

Beta attenuation. The beta multiplier family gives a closed-form continuous example. Take $r, s, u > 0$. Let $U_1 \sim \Gamma(r, 1)$, $U_2 \sim \Gamma(s, 1)$, and $U_3 \sim \Gamma(u, 1)$ be independent, and define

$$m_{A,i} = \frac{U_1 + U_2}{U_1 + U_2 + U_3}, \qquad m_{B,i} = \frac{U_1}{U_1 + U_2}. \tag{76}$$

Then $m_{A,i}$ and $m_{B,i}$ are independent, with

$$m_{A,i} \sim \mathrm{Beta}(r + s, u), \qquad m_{B,i} \sim \mathrm{Beta}(r, s). \tag{77}$$

Moreover, $t_i = m_{A,i} m_{B,i} = U_1/(U_1 + U_2 + U_3)$. Conditioning on t_i, the normalized remaining mass gives

$$m_{A,i} = t_i + (1 - t_i)V, \qquad V \sim \mathrm{Beta}(s, u), \tag{78}$$

and hence

$$\mathbb{E}[m_{A,i}|t_i] = t_i + (1 - t_i)\frac{s}{s+u} = \frac{s}{s+u} + \frac{u}{s+u} t_i. \tag{79}$$

Thus Eq. (68) holds with $\alpha = s/(s + u)$.

For a given observation multiplier $m_{A,i} \sim \mathrm{Beta}(a, b)$, this construction uses $s = \alpha b/(1 - \alpha)$, $r = a - s$, and $u = b$, requiring $0 < \alpha < a/(a+b)$. It provides the explicit recorruption distribution for continuous Beta attenuation.

Bernoulli masking. Bernoulli masking is a discrete special case. Let

$$m_{A,i} \sim \mathrm{Bernoulli}(p) \quad \text{and} \quad m_{B,i} \sim \mathrm{Bernoulli}(q) \tag{80}$$

be independent, with $0 < p < 1$ and $0 \le q < 1$. For $t_i = 1$, the condition is automatic because $m_{A,i} = 1$. For $t_i = 0$,

$$\mathbb{E}[m_{A,i}|t_i = 0] = \mathbb{P}(m_{A,i} = 1, m_{B,i} = 0 | m_{A,i} m_{B,i} = 0) = \frac{p(1-q)}{1-pq}. \tag{81}$$

Therefore the multiplicative condition holds with

$$\alpha = \frac{p(1-q)}{1-pq}. \tag{82}$$

Equivalently, for a prescribed original retention probability p and coefficient $0 < \alpha \le p$, the recorruption retention probability can be chosen as

$$q = \frac{p - \alpha}{p(1 - \alpha)}. \tag{83}$$

B Stability under Observation-Model Mismatch

This section analyzes how mismatch between the assumed observation model $P_{Y|X}$ and the true observation model $Q_{Y|X}$ affects the R2N estimator. We keep the notation of the main paper: X is the clean image, $Y = X + N$ is the noisy observation, $\tilde{Y} = Y + \tilde{N}$ is the recorrupted observation, and α is the recorruption coefficient.

Let P denote the joint distribution induced by the assumed observation model, the clean-image distribution, and the recorruption kernel. Let Q denote the corresponding true joint distribution. The clean-image distribution and the recorruption kernel are assumed to be shared by P and Q, while

$$Q_{Y|X} = P_{Y|X} + \Delta P_{Y|X}. \tag{84}$$

Equivalently, in density form,

$$q(y|x) = p(y|x) + \Delta p(y|x), \tag{85}$$

where $\int \Delta p(y|x)\, dy = 0$ for each x.

Under the true distribution Q, the conditional optimum of the R2N training objective is

$$R^*_{\mathrm{R2N}}(\tilde{Y}) = \mathbb{E}_Q[\tilde{N}|\tilde{Y}], \tag{86}$$

and the resulting denoising estimator is

$$\hat{X}_{\mathrm{R2N}}(\tilde{Y}) = \tilde{Y} - \frac{1}{\alpha}\mathbb{E}_Q[\tilde{N}|\tilde{Y}]. \tag{87}$$

The desired posterior mean under the true model satisfies

$$\mathbb{E}_Q[X|\tilde{Y}] = \tilde{Y} - \mathbb{E}_Q[N + \tilde{N}|\tilde{Y}]. \tag{88}$$

Therefore the denoising residual is

$$\begin{aligned} r(\tilde{Y}) &\coloneqq \hat{X}_{\mathrm{R2N}}(\tilde{Y}) - \mathbb{E}_Q[X|\tilde{Y}] \\ &= \mathbb{E}_Q\left[(N + \tilde{N}) - \alpha^{-1}\tilde{N}\middle|\tilde{Y}\right]. \end{aligned} \tag{89}$$

Define

$$G \coloneqq (N + \tilde{N}) - \alpha^{-1}\tilde{N}. \tag{90}$$

The ideal CPE identity under the assumed model P is

$$\mathbb{E}_P[\tilde{N}|\tilde{Y}] = \alpha\mathbb{E}_P[N + \tilde{N}|\tilde{Y}], \tag{91}$$

which is equivalently

$$\mathbb{E}_P[G|\tilde{Y}] = 0. \tag{92}$$

Thus the residual caused by model mismatch can be written as

$$r(\tilde{Y}) = \mathbb{E}_Q[G|\tilde{Y}]. \tag{93}$$

Conditional KL control. For a fixed value $\tilde{Y} = \tilde{y}$, denote by $P_{\tilde{y}}$ and $Q_{\tilde{y}}$ the conditional distributions of the remaining variables under P and Q, respectively. By Eq. (92), $\mathbb{E}_{P_{\tilde{y}}}[G] = 0$. For vector-valued images, we apply the scalar argument to every unit direction u and define $G_u = \langle u, G\rangle$. Donsker–Varadhan's variational inequality gives, for any $\lambda > 0$,

$$\mathbb{E}_{Q_{\tilde{y}}}[G_u] \le \frac{D_{\mathrm{KL}}(Q_{\tilde{y}}\|P_{\tilde{y}}) + \log \mathbb{E}_{P_{\tilde{y}}} \exp(\lambda G_u)}{\lambda}. \tag{94}$$

Applying the same inequality to $-G_u$ gives

$$-\mathbb{E}_{Q_{\tilde{y}}}[G_u] \le \frac{D_{\mathrm{KL}}(Q_{\tilde{y}}\|P_{\tilde{y}}) + \log \mathbb{E}_{P_{\tilde{y}}} \exp(-\lambda G_u)}{\lambda}. \tag{95}$$

Assume that $G|\tilde{Y} = \tilde{y}$ is directionally sub-Gaussian under the assumed model with proxy variance σ_G^2, uniformly over $\tilde{y}$; that is, for every unit vector u,

$$\log \mathbb{E}_{P_{\tilde{y}}} \exp(\lambda G_u) \le \frac{\sigma_G^2\lambda^2}{2} \tag{96}$$

for all $\lambda \in \mathbb{R}$. Let

$$D(\tilde{y}) \coloneqq D_{\mathrm{KL}}(Q_{\tilde{y}}\|P_{\tilde{y}}). \tag{97}$$

Combining Eqs. (94)–(96), we obtain

$$|\langle u, r(\tilde{y})\rangle| \le \inf_{\lambda>0}\left(\frac{D(\tilde{y})}{\lambda} + \frac{\sigma_G^2\lambda}{2}\right) = \sqrt{2\sigma_G^2 D(\tilde{y})}. \tag{98}$$

Taking the supremum over all unit vectors u gives the same pointwise bound for $\|r(\tilde{y})\|$. Squaring and integrating over the true marginal distribution of $\tilde{Y}$ gives

$$\mathbb{E}_Q\|r(\tilde{Y})\|^2 \le 2\sigma_G^2\mathbb{E}_Q D_{\mathrm{KL}}(Q_{\tilde{Y}}\|P_{\tilde{Y}}). \tag{99}$$

By the chain rule of KL divergence,

$$D_{\mathrm{KL}}(Q\|P) = D_{\mathrm{KL}}(Q(\tilde{Y})\|P(\tilde{Y})) + \mathbb{E}_Q D_{\mathrm{KL}}(Q_{\tilde{Y}}\|P_{\tilde{Y}}). \tag{100}$$

Since the first term is nonnegative,

$$\mathbb{E}_Q\|r(\tilde{Y})\|^2 \le 2\sigma_G^2 D_{\mathrm{KL}}(Q\|P). \tag{101}$$

Because P and Q share the clean-image distribution and the recorruption kernel, the joint KL reduces to the conditional KL of the observation model:

$$D_{\mathrm{KL}}(Q\|P) = \mathbb{E}_X D_{\mathrm{KL}}(Q_{Y|X}(\cdot|X)\|P_{Y|X}(\cdot|X)). \tag{102}$$

Combining Eqs. (99)–(102) yields

$$\mathbb{E}_Q\|r(\tilde{Y})\|^2 \le 2\sigma_G^2\mathbb{E}_X D_{\mathrm{KL}}(Q_{Y|X}(\cdot|X)\|P_{Y|X}(\cdot|X)). \tag{103}$$

Relation to ΔP. For small density perturbations $q(y|x) = p(y|x) + \Delta p(y|x)$, the conditional KL divergence admits the expansion

$$D_{\mathrm{KL}}(Q_{Y|X=x} \| P_{Y|X=x}) = \int (p(y|x) + \Delta p(y|x)) \log \left(1 + \frac{\Delta p(y|x)}{p(y|x)}\right) dy$$
$$= \frac{1}{2} \int \frac{\Delta p(y|x)^2}{p(y|x)} dy + O(\|\Delta p(\cdot|x)\|^3), \tag{104}$$

where the first-order term vanishes because $\int \Delta p(y|x)\, dy = 0$. Hence Eq. (103) implies

$$\mathbb{E}_Q \|r(\tilde{Y})\|^2 = O\left(\mathbb{E}_X \int \frac{\Delta p(y|X)^2}{p(y|X)} dy\right). \tag{105}$$

This shows that, under the conditional sub-Gaussian assumption, the mean-squared R2N bias caused by observation-model mismatch is second order in the natural weighted size of ΔP.

C Additional Empirical Results

This section reports additional experiments that support two assumptions used in the main paper and further evaluates R2N on a real microscopy denoising benchmark. The first group examines the empirical orthogonality between the clean image and the noise predicted by R2N, which is the key approximation behind the test-time coefficient estimator in Eq. (18) of the main paper. The second group studies the stability of the estimated recorruption coefficient $\hat{\alpha}$. The final group evaluates denoising performance on the FMDD microscopy dataset, where the real corruption is better described as mixed Poisson–Gaussian noise.

Cosine-similarity metric. For an image x and a noise-like signal z, we compute the signed cosine similarity

$$c(x, z) = \frac{\langle x, z \rangle}{\|x\|_2 \|z\|_2}. \tag{106}$$

The ideal orthogonality condition corresponds to $c(x, z) = 0$. In the simulated experiments, x is the clean DIV2K image and z is the noise predicted by the trained R2N network. In the FMDD experiment, we evaluate both the real noise, obtained from the noisy-clean image difference, and the R2N-predicted noise.

Table 5: Signed cosine similarity between clean DIV2K images and the R2N-predicted noise under different synthetic noise models.

Noise model	Mean cosine	Std. cosine
Gaussian–Bernoulli	0.0726	0.0212
Cauchy	0.0565	0.0304
Gaussian	0.0565	0.0304
Poisson	0.0443	0.0575
Gamma	0.0312	0.0487

Table 5 and Fig. 5 show that the predicted noise remains weakly correlated with the clean image across all tested noise models. The mean signed cosine similarity is below 0.08 in every case and is lowest for the Gamma and Poisson settings. These results are consistent with the approximation used by the test-time estimator: the predicted noise is dominated by corruption-dependent components rather than by image structure.

We further test whether the orthogonality behavior is stable when the Gaussian noise level changes. As shown in Table 6 and Fig. 6, the average cosine similarity stays within the narrow range $[0.0304, 0.0472]$ for all tested noise levels. This suggests that the approximation is not sensitive to a single preset Gaussian variance.

Table 7 and Fig. 7 extend the orthogonality check to real microscopy data. The real noise has mean cosine similarity close to zero but a large standard deviation, indicating substantial sample-to-sample variability in the real imaging noise. The R2N-predicted noise has a small positive mean

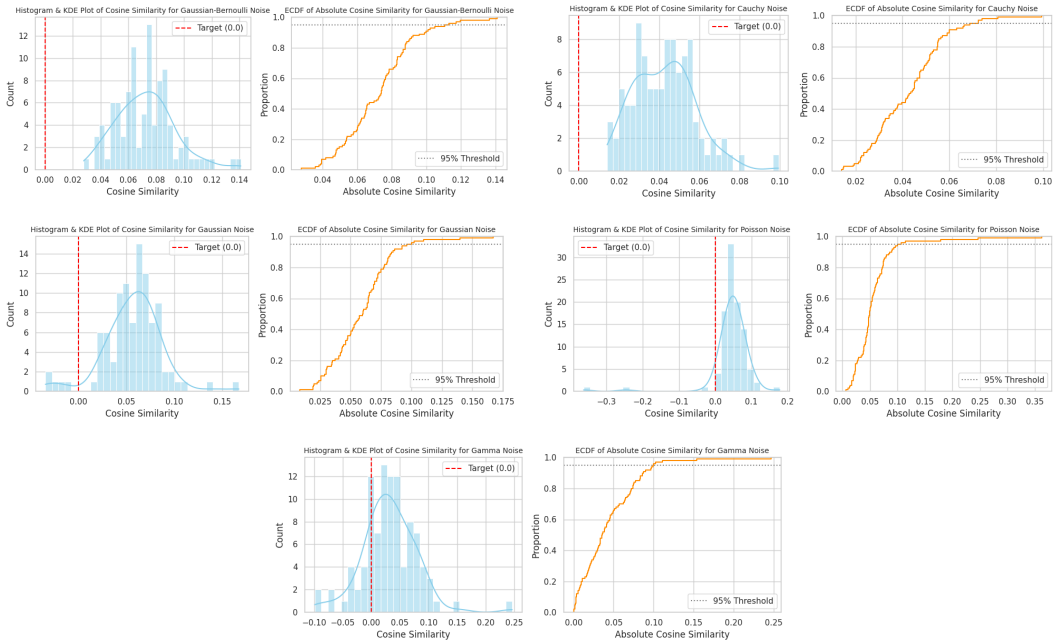

Figure 5: Distribution of the cosine similarity between clean DIV2K images and the R2N-predicted noise for five synthetic noise models. Each panel shows the signed cosine-similarity histogram and the empirical cumulative distribution of the absolute cosine similarity.

Table 6: Signed cosine similarity between clean DIV2K images and R2N-predicted noise for Gaussian noise at different noise levels.

Noise level	Mean cosine	Std. cosine
0.1	0.0472	0.0734
0.2	0.0304	0.0387
0.3	0.0436	0.0329
0.5	0.0360	0.0190

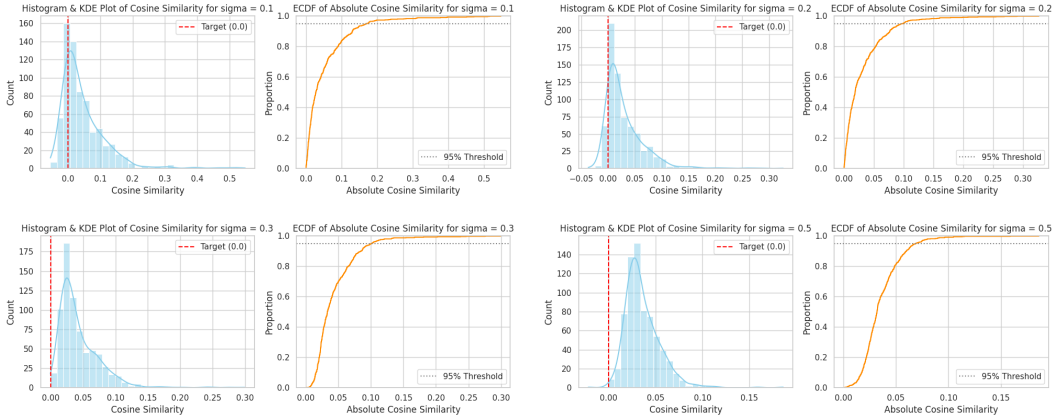

Figure 6: Distribution of the cosine similarity between clean DIV2K images and R2N-predicted noise under Gaussian corruption with noise levels 0.1, 0.2, 0.3, and 0.5.

Table 7: Signed cosine similarity on the FMDD microscopy dataset.

Noise signal	Mean cosine	Std. cosine
Real noise	0.0075	0.3223
R2N-predicted noise	0.1048	0.1037

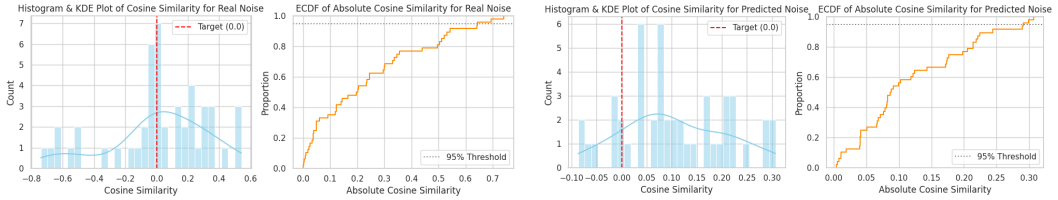

Figure 7: Cosine-similarity distributions on FMDD. Left: similarity between the clean image and the real noise. Right: similarity between the clean image and the R2N-predicted noise.

cosine similarity and a smaller standard deviation, which indicates that the learned noise estimate is not exactly orthogonal to the clean image but remains weakly aligned in aggregate. This residual correlation is consistent with the model-mismatch analysis in Section B, since FMDD noise is mixed Poisson–Gaussian while the tested self-supervised methods use Gaussian recorruption models.

Stability of α estimation. We also evaluate the stability of the test-time estimator for α across several synthetic noise distributions on DIV2K. For each distribution, we compare the estimated coefficient $\hat{\alpha}$ with the known ground-truth coefficient used by the simulation.

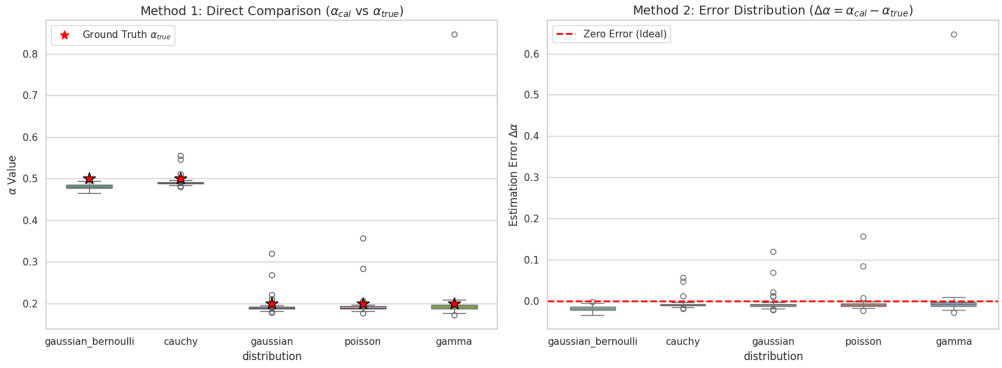

Figure 8: Stability of the test-time α estimator on DIV2K. The left panel compares the estimated coefficient with the ground-truth coefficient for each noise distribution. The right panel reports the estimation error $\Delta\alpha = \hat{\alpha} - \alpha_{\text{true}}$.

Fig. 8 shows that the estimated coefficients are concentrated near the true values for Gaussian–Bernoulli, Cauchy, Gaussian, and Poisson noise. Gamma noise exhibits one pronounced outlier, but the box statistics remain close to the target value. These results support the use of Eq. (18) in the main paper as a practical test-time calibration rule, while also showing that heavy-tailed or strongly mismatched cases may produce occasional outliers.

Real microscopy denoising on FMDD. Finally, we evaluate R2N on FMDD, a real microscopy dataset whose noise is empirically modeled as a mixture of Poisson and Gaussian components. To stress-test model mismatch, all compared methods are run with Gaussian-noise assumptions. R2N uses Gaussian recorruption with noise level 0.2 and $\alpha = 0.2$. GR2R uses Gaussian recorruption with noise level 0.05 and $\alpha = 0.2$; this noise level was selected after testing multiple candidates and gave the best GR2R result. N2S is run under the Gaussian setting, where noise levels 0.05, 0.1, and 0.2 lead to similar performance. PURE is run with Gaussian noise level 0.05.

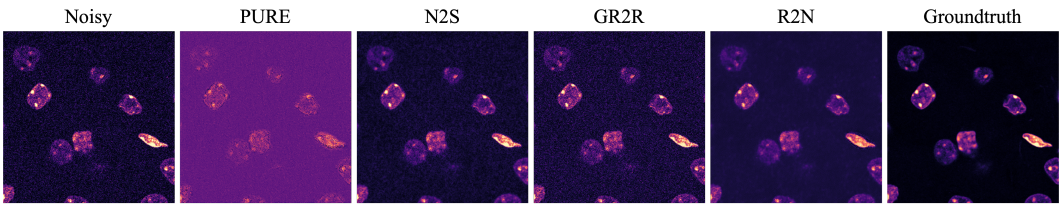

Figure 9: Qualitative FMDD denoising comparison. From left to right, the columns show the noisy input, PURE, N2S, GR2R, R2N, and the clean reference.

Table 8: Denoising performance on the FMDD microscopy dataset. All methods are evaluated under Gaussian-noise assumptions although the real noise is mixed Poisson–Gaussian.

Method	PSNR (dB)	SSIM
R2N	**31.15**	**0.821**
GR2R	26.89	0.510
N2S	27.82	0.542
PURE	19.79	0.170

Table 8 and Fig. 9 show that R2N achieves the best FMDD performance among the tested self-supervised baselines, improving over the strongest baseline by 3.33 dB PSNR and 0.279 SSIM. The qualitative comparison shows that R2N suppresses background noise while preserving cellular structures more faithfully than the competing methods. These results support the robustness claim in the main paper: even when the assumed Gaussian observation model does not exactly match the mixed Poisson–Gaussian microscopy noise, R2N remains effective because the induced bias is controlled by the degree of model mismatch.