Zero-Shot Image Denoising via Hybrid Prior-Guided Pseudo Sample Generation

Supplementary Material

Algorithm 1 Local Prior-Guided Down-Sampling ($\mathcal{S}_d$)

Input: Noisy image $\mathbf{x} \in \mathbb{R}^{h\times w\times c}$, window size k
Output: Downsampled samples $(\mathbf{x}_\ell, \mathbf{y}_\ell)$, $(\mathbf{x}_\hbar, \mathbf{y}_\hbar)$
1: Initialize $\mathbf{x}_\ell$, $\mathbf{y}_\ell$, $\mathbf{x}_\hbar$, $\mathbf{y}_\hbar \in \mathbb{R}^{h/2\times w/2\times c}$ as zero tensors
2: Obtain Sobel gradients of $\mathbf{x}$ using Sobel operators $\mathbf{S}_{\text{hor}}$ and $\mathbf{S}_{\text{ver}}$: $\mathbf{G}_{\text{hor}}(\mathbf{x}) = \mathbf{x} * \mathbf{S}_{\text{hor}}$, $\mathbf{G}_{\text{ver}}(\mathbf{x}) = \mathbf{x} * \mathbf{S}_{\text{ver}}$
3: Compute the aggregated gradients using $k \times k$ average pooling, and average them over c channels: $\bar{\mathbf{G}}_{\text{hor}}(\mathbf{x}) = \mathbf{G}_{\text{hor}}(\mathbf{x}) * \mathbb{I}_{k\times k}/k^2$, $\bar{\mathbf{G}}_{\text{ver}}(\mathbf{x}) = \mathbf{G}_{\text{ver}}(\mathbf{x}) * \mathbb{I}_{k\times k}/k^2$.
4: Compute the magnitude map of aggregated gradients: $\mathbf{G}(\mathbf{x}) = \sqrt{\bar{\mathbf{G}}^2_{\text{hor}}(\mathbf{x}) + \bar{\mathbf{G}}^2_{\text{ver}}(\mathbf{x})}$
5: /* Sobel gradient-guided down-sampling */
6: **for** each non-overlapping 2×2 patch in $\mathbf{G}(\mathbf{x})$ **do**
7: Extract 4 elements within the patch: g_1, g_2, g_3, g_4
8: Record the coordinates of them as: $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3, \mathbf{p}_4$
9: Sort in ascending order, e.g., $g_3 \le g_2 \le g_4 \le g_1$
10: /* Create low-low / high-high instance pairs */
11: $\mathbf{x}_\ell \leftarrow \mathbf{x}(\mathbf{p}_3)$, $\mathbf{y}_\ell \leftarrow \mathbf{x}(\mathbf{p}_2)$
12: $\mathbf{x}_\hbar \leftarrow \mathbf{x}(\mathbf{p}_4)$, $\mathbf{y}_\hbar \leftarrow \mathbf{x}(\mathbf{p}_1)$
13: **end for**
14: **return** $(\mathbf{x}_\ell, \mathbf{y}_\ell)$, $(\mathbf{x}_\hbar, \mathbf{y}_\hbar)$

Algorithm 2 Generation of Pixel Bank $\mathbf{B}_p$ for $\mathcal{S}_r$

Input: Noisy image $\mathbf{x} \in \mathbb{R}^{h\times w\times c}$, neighborhood size l, Top-K, number of candidate pixels M, and σ_G
Output: Pixel Bank $\mathbf{B}_p \in \mathbb{R}^{h\times w\times c\times K}$
1: Extract the luminance (Y channel) of the YCbCr color space of $\mathbf{x}$, denoted as $\mathbf{x}_Y \in \mathbb{R}^{h\times w\times 1}$
2: **for** each coordinate $\mathbf{p}$ in $\mathbf{x}_Y$ **do**
3: Extract the $l \times l$ patch centered at $\mathbf{p}$: $\mathbf{N}_l(\mathbf{p})$
4: /* Global random sampling s.t. a Gaussian */
5: Sample M candidate coordinates conditioned on a Gaussian: $\mathcal{C}(\mathbf{p}) = \{\mathbf{p}_1, \ldots, \mathbf{p}_M\}$, where $p(\mathbf{p}_i) \sim \mathcal{N}(\mathbf{p}, \sigma_G^2\mathbf{I})$ and $1 \le i \le M$
6: **for** $1 \le i \le M$ **do**
7: Extract the neighborhood $\mathbf{N}_l(\mathbf{p}_i)$ centered at $\mathbf{p}_i$
8: Compute the distance $D(\mathbf{p}, \mathbf{p}_i)$ using L_1 norm
9: **end for**
10: /* Top-K candidate selection */
11: Find the coordinates $\{\tilde{\mathbf{p}}_1, \ldots, \tilde{\mathbf{p}}_K\}$ corresponding to the K smallest values in $\{D(\mathbf{p}, \mathbf{p}_i)\}_{i=1}^M$
12: Extract the K pixels of $\mathbf{x}$ located at $\tilde{\mathbf{p}}_1, \ldots, \tilde{\mathbf{p}}_K$ and form a tensor $\mathbf{v} \in \mathbb{R}^{1\times 1\times c\times K}$ for the pixel $\mathbf{x}(\mathbf{p})$
13: Store these pixels in the pixel bank: $\mathbf{B}_p[\mathbf{p}] \leftarrow \mathbf{v}$
14: **end for**
15: **return** $\mathbf{B}_p$

1. More Details about the Samplers

1.1. Local Prior-Guided Down-Sampling

We provide more details about our $\mathcal{S}_d$ in Alg. 1. As can be seen, it makes use of image locality ($\mathcal{P}_1$) by aggregating the Sobel gradients within a local window of size $k \times k$. Then the down-scaled instances are intuitively generated through pairing image pixels with low-low and high-high gradient magnitudes, as shown in lines 6 $\sim$ 12 in Alg. 1. The reason for gradient pairing lies in that the potential signals of image pixels with high gradients may have fundamentally different statistics from those with low gradients, which is probably more in line with the assumption of consistent underlying signal in Noise2Noise [4].

1.2. Construction of the Pixel Bank

The core component of our $\mathcal{S}_r$ is the pixel bank $\mathbf{B}_p$, which is built following the procedure described in Alg. 2. Unlike Pixel2Pixel [6], the pixels in our $\mathbf{B}_p$ are not sampled from a local window centered around the target pixel, but from the entire image restrained by a Gaussian distribution. This simple improvement yields two substantial benefits:

- Our method makes better use of non-local self-similarity ($\mathcal{P}_2$) within natural images, thus producing higher-quality candidate pixels for pseudo sample generation;
- By directly performing random sampling conditioned on a Gaussian instead of exhaustively searching over a local window, it significantly improves model efficiency.

2. Rationale

We provide additional experimental results and analyses in this supplementary material to further verify the design and effectiveness of our ZS-HPD model. The primary content of this part consists of the following:

- **Window Size**: We analyze the impact of varying the window size for gradient aggregation in our local sampling strategy ($\mathcal{S}_d$).
- **Patch Size**: We study how different patch sizes (i.e., l of $\mathbf{N}_l(u_0, v_0)$) affect the performance of our global sampler ($\mathcal{S}_r$) in measuring patch similarity.
- **Radius for Spectrum Division**: The sensitivity analysis about the radius r for spectrum division is performed to verify our choice.
- **Gaussian**: The standard deviation of the Gaussian affects the range of global sampling, so we explore the effect of the parameter σ on the performance of our $\mathcal{S}_r$.
- **Top-K**: This parameter decides the number of candidate

Algorithm 3 Hybrid Prior-Guided Zero-Shot Denoising

Input: Noisy image $\mathbf{x}$, down-sampler $\mathcal{S}_d$, random sampler $\mathcal{S}_r$, denoising network $\mathcal{F}_\theta$, total iterations T
Output: Trained denoising network $\mathcal{F}_\theta^*$

/* **Step 1**: Down-scaled sample generation with $\mathcal{S}_d$ */
$(\mathbf{x}_\ell, \mathbf{y}_\ell), (\mathbf{x}_\hbar, \mathbf{y}_\hbar) \leftarrow \mathcal{S}_d(\mathbf{x})$
/* Augment with inverted pairs */
$\mathbf{B}_d = \{(\mathbf{x}_\ell, \mathbf{y}_\ell), (\mathbf{x}_\hbar, \mathbf{y}_\hbar), (\mathbf{y}_\ell, \mathbf{x}_\ell), (\mathbf{y}_\hbar, \mathbf{x}_\hbar)\}$
/* **Step 2**: Build the pixel band $\mathbf{B}_p$ with $\mathcal{S}_r$ */
$\mathbf{B}_p \leftarrow \mathcal{S}_r(\mathbf{x})$
/* **Step 3**: Denoising network training */
for $t = 1$ to T **do**
    Randomly select a training pair $(\mathbf{x}_d, \mathbf{y}_d)$ from $\mathbf{B}_d$
    Randomly select a training pair $(\mathbf{x}_p, \mathbf{y}_p)$ from $\mathbf{B}_p$
    $\hat{\mathbf{y}}_d = \mathcal{F}_\theta(\mathbf{x}_d), \quad \hat{\mathbf{y}}_p = \mathcal{F}_\theta(\mathbf{x}_p)$
    $\mathcal{L}_d = \mathrm{SWL}(\mathbf{y}_d, \hat{\mathbf{y}}_d), \quad \mathcal{L}_p = \mathrm{SWL}(\mathbf{y}_p, \hat{\mathbf{y}}_p)$
    Total loss $\mathcal{L} = \mathcal{L}_d + \mathcal{L}_p$
    Minimize $\mathcal{L}$ with AdamW optimizer
end for
return Trained denoising network $\mathcal{F}_\theta^*$

Table 1. The performance of our ZS-HPD on Kodak24 [3] when window size for gradient aggregation changes (Gaussian noise).

Sampler	Window Size	$\sigma = 10$	$\sigma = 25$	$\sigma = 50$
$\mathcal{S}_d$	3×3	**33.57**	29.76	26.33
	5×5	**33.57**	**29.88**	**26.36**
	7×7	33.45	29.55	26.28
	9×9	33.43	29.54	26.26

Table 2. Denoising results of our ZS-HPD on Kodak24 [3] when patch size for similarity measurement changes (Poisson noise).

Sampler	Patch Size	$\lambda = 10$	$\lambda = 25$	$\lambda = 50$
$\mathcal{S}_r$	3×3	26.16	28.13	30.05
	5×5	26.49	28.76	**30.27**
	7×7	**26.85**	**28.79**	30.12
	9×9	26.66	28.76	29.98

pixels in our pixel bank $\mathbf{B}_p$, and we investigate its effect on denoising performance here.

Besides, we also present extra visual results to further verify the effectiveness of our ZS-HPD. The detailed studies are exhibited in the following. Except for the hyperparameters to be analyzed, the remaining settings of the framework are consistent with those described in the manuscript.

3. Supplementary Results

3.1. Window Size for Gradient Aggregation

To study the impact of the window size utilized for gradient aggregation, we set it to be 3×3, 5×5, 7×7, 9×9, and assess the performance of the proposed ZS-HPD on Kodak24 [3] with Gaussian noise ($\sigma = 10, 25, 50$). The results are shown in Table 1, and it can be seen that a larger window size does not necessarily lead to better results. According to the prior of spatial locality $\mathcal{P}_1$, this is probably because that increased window size results in weakened spatial coherence between the pixels within it, thereby affecting the guidance provided by aggregated gradients for subsequent downsampling. On the other hand, larger window sizes typically signify greater computational requirements. So, in our settings, we opt for a 5×5 size as our final choice.

Table 3. Impact of the radius r for frequency separation on McMaster18 [9] with different Gaussian noise.

Component	Noise Level (σ)	Radius (r) 10%	20%	30%	50%	70%
SWL	10	34.49	34.34	**34.50**	34.34	34.34
	25	29.60	**29.85**	29.34	29.24	29.40
	50	25.08	**25.10**	25.06	25.04	25.00

Table 4. Impact of different standard deviation (σ_G) for gaussian sampling on McMaster18 [9] with different Poisson noise.

Sampler	Noise Level (λ)	Standard Deviation (σ_G) 10	15	20	25	30
$\mathcal{S}_r$	10	27.36	**27.93**	27.87	27.87	27.73
	25	30.09	**30.20**	30.13	30.01	29.91
	50	**31.75**	31.64	31.47	31.36	31.31

3.2. Patch Size for Similarity Measurement

We experiment with patch sizes of 3×3, 5×5, 7×7, 9×9 to investigate the optimal patch size for measuring non-local self-similarity. This study is conducted on the Kodak24 [3] dataset with Poisson noise levels of λ = 10, 25, and 50. As shown in Table 2, it can be seen that the model performs overall better when the patch size $l = 7$. In fact, the patch size l can behave to control the impact of noise on similarity measurement to some extent. On the one hand, it is difficult for small image patches to overcome noise interference. On the other hand, if l is too large, it becomes hard to effectively exploit the non-local self-similarity ($\mathcal{P}_2$). Our observations are in line with Pixel2Pixel [6], even though it is verified with real-world noise. Therefore, in our experiments, we simply choose 7×7 for all cases.

3.3. The Radius for Frequency Division

In our spectral weighted loss (SWL), a hyperparameter that may affect the performance of our approach exists, i.e., the radius r for frequency division. To this end, we conduct a sensitivity analysis on this hyperparameter. We test a range of values for the radius r: 0.1, 0.2, 0.3, 0.5, 0.7 (normalized by image size) on McMaster18 [9] with Gaussian noise (σ = 10, 25 and 50), and the results are shown in Table 3. It can be seen that when the value of r is between 0.2 and 0.3, the model performance is better than in other cases. The value of r is intuitively determined to be 0.2 in our experiments.

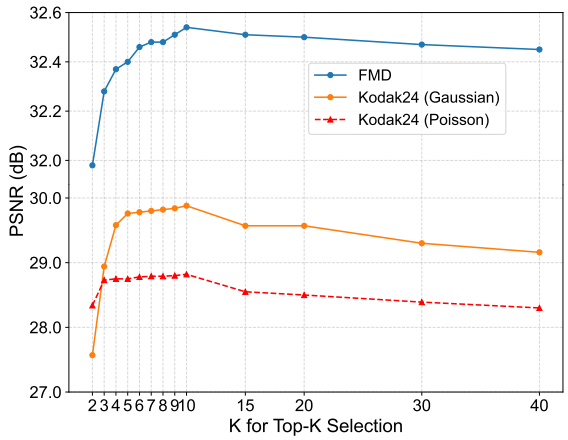

Figure 1. The impact of K in Top-K candidate selection on model performance. The results are collected on FMD [10] and Kodak24 [3]. The noise levels for Gaussian noise and Poisson noise are set to $\sigma = 25$ and $\lambda = 25$, respectively.

3.4. Standard Deviation for Gaussian Sampling

To investigate the effect of the sampling range of our $\mathcal{S}_r$, we dive to study how the standard deviation σ_G of the Gaussian affect the results of our ZS-HPD. We conduct experiments with McMaster18 [9] by setting σ_G to 10, 15, 20, 25, and 30. The results are illustrated in Table 4. It can be seen that when σ_G = 15, the model reaches the best results for λ = 10 and λ = 25, but the best result for λ = 50 corresponds to σ_G = 10. So although we are seeking to effectively explore the non-local self-similarity $\mathcal{P}_2$, the restraint of spatial locality $\mathcal{P}_1$ is equally important. In our implementation, we choose σ_G = 15 for all cases.

3.5. Top-K Candidates for Pixel Bank

We conduct a comprehensive study to explore how the value of K in top-K candidate selection affects the performance of our ZS-HPD. We evaluate the denoising performance for Gaussian noise ($\sigma = 25$) and Poisson noise ($\lambda = 25$) with varying K values on Kodak24 [3], and for real-world noise on FMD [10]. The quantitative comparisons are visualized in Fig. 1. It can be observed that the performance initially improves as K increases, since a larger pool of high-quality candidates introduces necessary sample diversity for model training. However, continuously increasing K results in the degradation of the performance, which indicates an inherent trade-off between the benefit of improved sample diversity and the detriment resulting from low-correlation candidate pixels. In our experiments, we adopt $K = 10$ for all trials.

4. More Visual Results

We provide more visual results of our ZS-HPD compared to other advanced zero-shot denoising methods, including DIP [2], BM3D [1], ZS-N2N [7], ZS-N2M [5], Pixel2Pixel [6] etc., as shown in Fig. 2 $\sim$ Fig. 5. In most cases, our ZS-HPD can restore the underlying structures in these test images more reliably than other methods, generating visual effects that are closer to the ground truth. The quantitative PSNR and SSIM [8] values below each small image patch also confirm this observation. These visual results prove the superiority of the proposed ZS-HPD over other advanced methods, and to some extent, also illustrate the significant role of adequately exploring the local and non-local priors of natural images in zero-shot image denoising.

References

[1] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. *IEEE Transactions on Image Processing*, 16:2080–2095, 2007. 3

[2] Ulyanov Dmitry, Vedaldi Andrea, and Lempitsky Victor. Deep image prior. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9446–9454, 2018. 3

[3] Richard Franzen. Kodak lossless true color image suite. https://r0k.us/graphics/kodak/, 2010. Accessed: 2024-05-27. 2, 3

[4] 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*, pages 4620–4631, 2018. 1

[5] Duo Liu, Yiqi Shi, Guoyin Zhang, Sizhao Li, and Liguo Zhang. Zero-shot noise2mean: Gap minimization for efficient denoising from a single noisy image. In *Proceedings of the AAAI Conference on Artificial Intelligence*, pages 5406–5414, 2025. 3

[6] Qing Ma, Jiang Junjun, Zhou Xiong, Liang Pengwei, Liu Xianming, and Ma Jiayi. Pixel2pixel: A pixelwise approach for zero-shot single image denoising. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 47(6):4614–4629, 2025. 1, 2, 3

[7] Youssef Mansour and Reinhard Heckel. Zero-shot noise2noise: Efficient image denoising without any data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14018–14027, 2023. 3

[8] Zhou Wang, Alan Conrad Bovik, Hamid R. Sheikh, and Eero P. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004. 3

[9] Lei Zhang, Xiaolin Wu, Antoni Buades, and Xin Li. Color demosaicking by local directional interpolation and nonlocal adaptive thresholding. *Journal of Electronic Imaging*, 20: 023016, 2011. 2, 3

[10] Yide Zhang, Yinhao Zhu, Evan Nichols, Qingfei Wang, Siyuan Zhang, and Cody Smith. A poisson-gaussian denoising dataset with real fluorescence microscopy images. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 11710–11718, 2019. 3

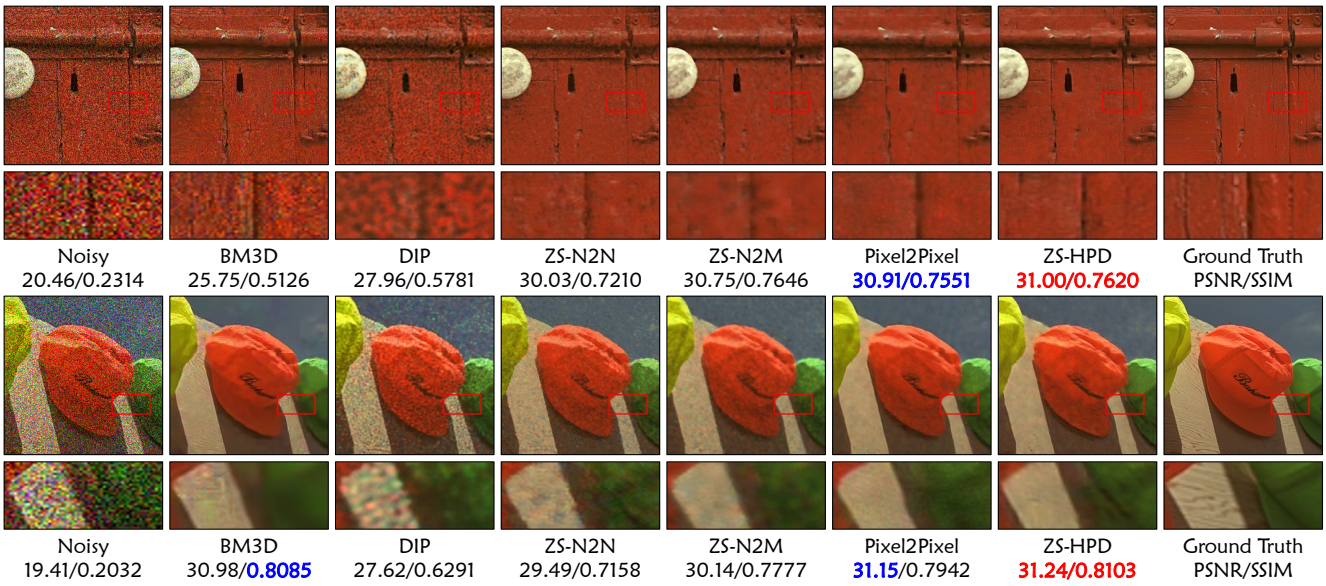

Figure 2. Visual results of several compared methods on two testing images from Kodak24 with Poisson ($\lambda = 10$) noise. The best and second-best PSNR and SSIM results are highlighted in **red** and **blue**, respectively.

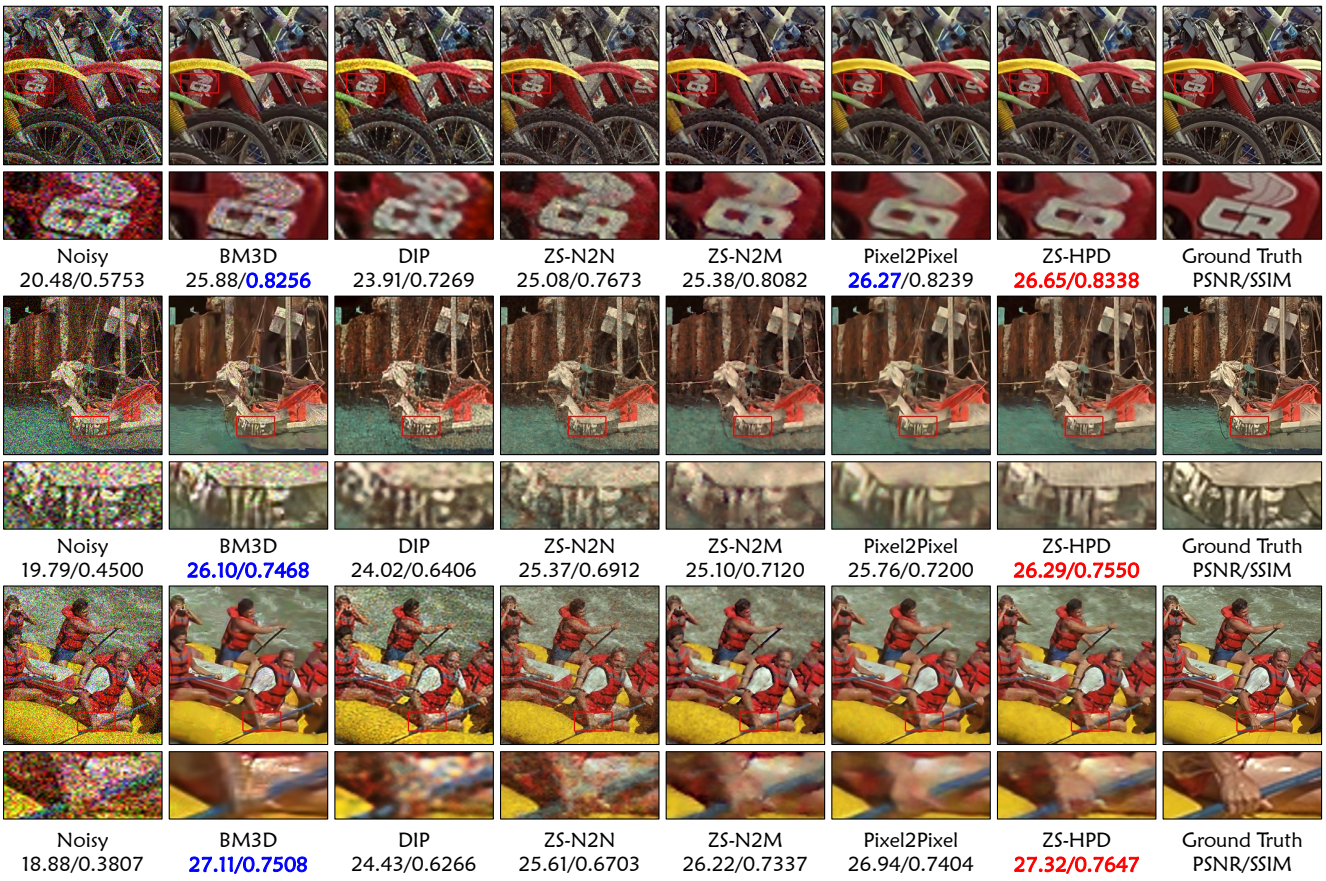

Figure 3. Visual results of several compared methods on three testing images from Kodak24 with Poisson ($\lambda = 10$) noise. The best and second-best PSNR and SSIM results are highlighted in **red** and **blue**, respectively.

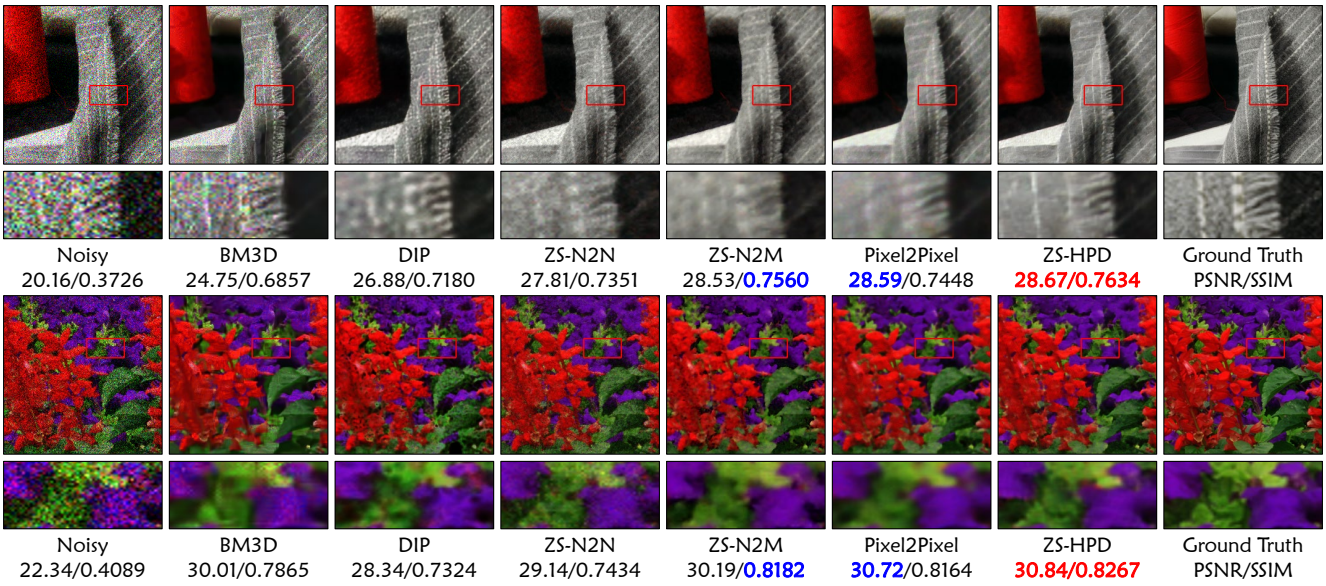

Figure 4. Visual results of several compared methods on two testing images from McMaster18 with Poisson ($\lambda = 10$) noise. The best and second-best PSNR and SSIM results are highlighted in **red** and **blue**, respectively.

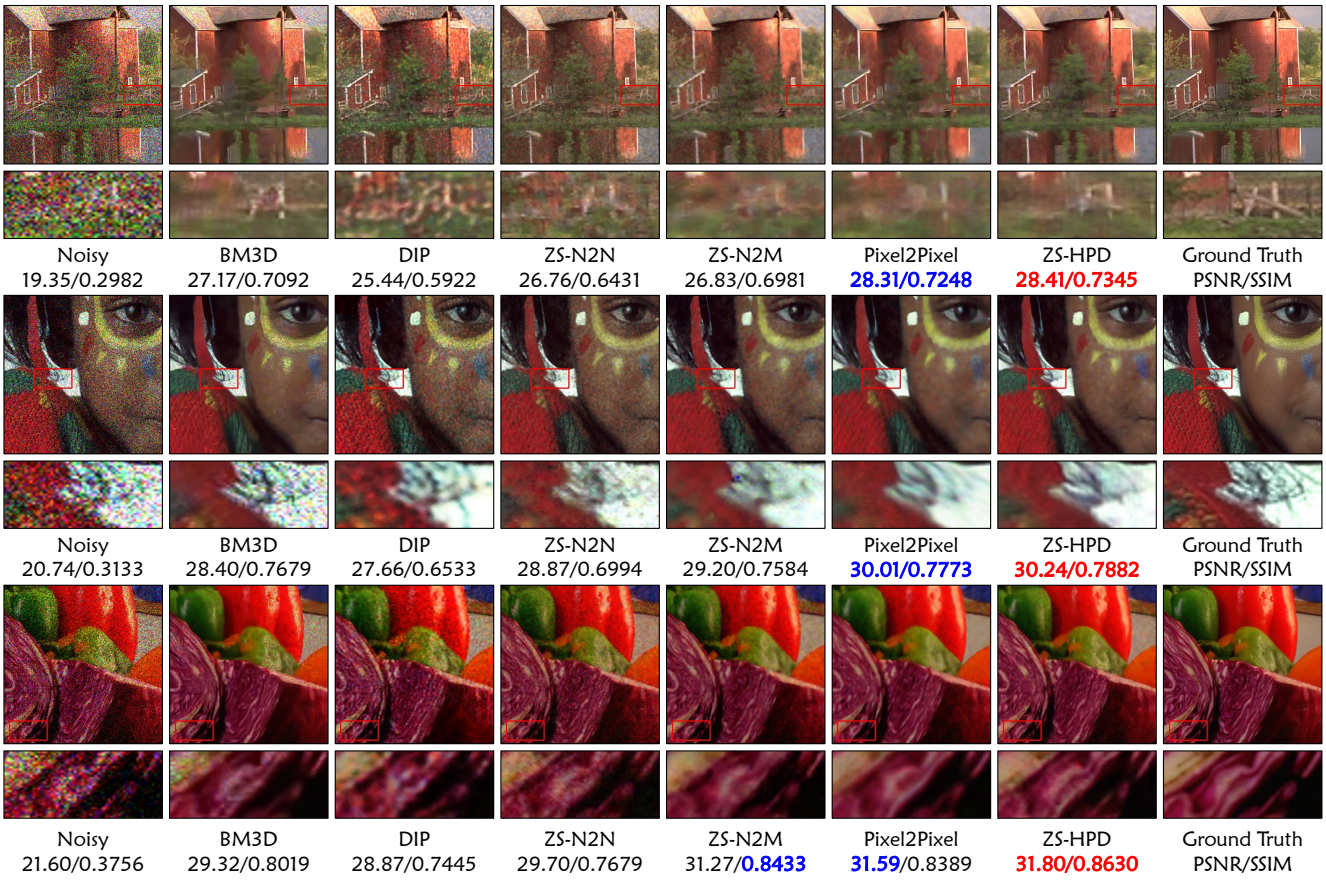

Figure 5. Visual results of several compared methods on three testing images from Kodak24 and McMaster18 with Poisson ($\lambda = 10$) noise. The best and second-best PSNR and SSIM results are highlighted in **red** and **blue**, respectively.