A. Framework

For the backbone network of our methods, it can be a *hourglass* architecture like DIP [29], or a *decoder* architecture like DD [9]. We show the frameworks in Fig. 13.

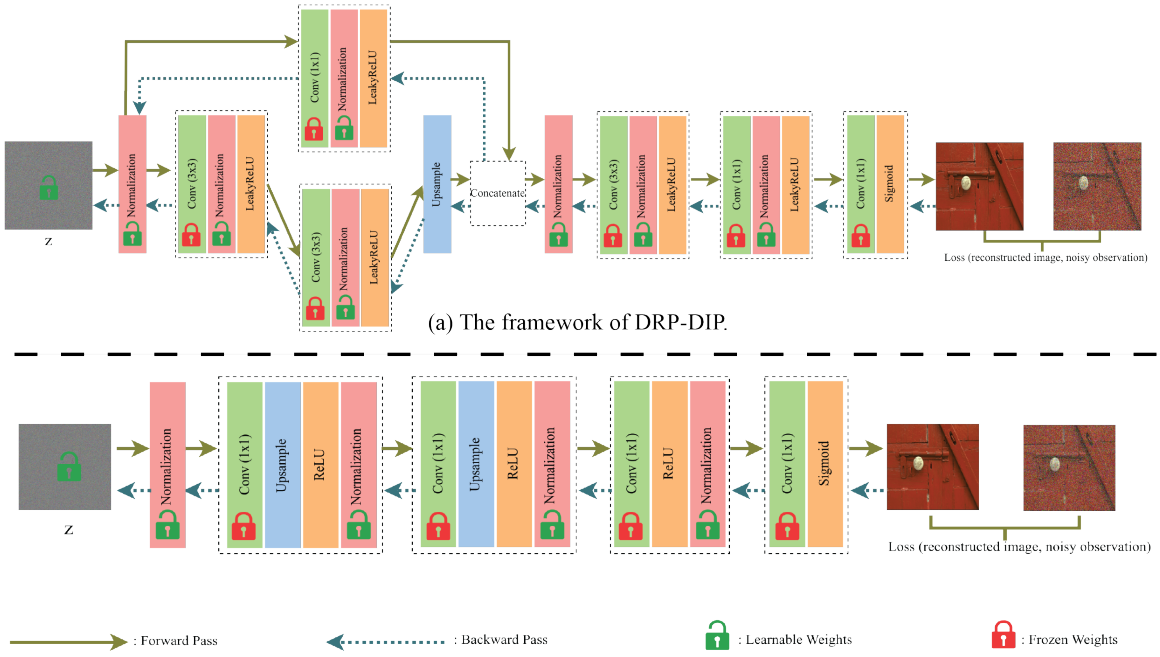

Figure 13. The framework of our method: (a) for DRP-DIP and (b) for DRP-DD.

In addition, Tab. 4 shows the details including the number of CNN layers and the number of learnable parameters of DIP [29].

Table 4. The details of different scale/depth of backbone network.

	No. of CNN layers	No. of parameters
Scale-1	6	82,979
Scale-2	11	201,143
Scale-3	16	319,307
Scale-4	21	437,471
Scale-5	26	555,635

B. OPT Time Measurement

Processing time is an essential concern in many real-world applications such as self-driving cars. Therefore, in addition to the restoration quality (e.g., measured by PSNR), in this paper, we also consider the optimization (OPT) time. However, ensuring a fair comparison yet having a large scale of experiments across different models is a nontrivial task as the OPT time can be easily affected by the execution environment and other concurrent tasks on the same hardware.

To address this issue, 1) for each model, we only measure its forward-pass and backward-pass time in each iteration (we do not take the I/O time of saving models/reconstructed images into account); 2) we estimate the correlation between iterations and OPT time. With this correlation, we obtain per-iteration OPT time denoted as $\mathrm{T}_{per-iter}$. After that, it is easy to estimate the total OPT time (T_{total}) given any number of iterations (M):

$$\mathrm{T}_{total} = f(M) = \mathrm{T}_{per-iter} \times M. \tag{9}$$

To estimate $\mathrm{T}_{per-iter}$, we run our experiments in a clean pure execution environment in which the computational resource is completely reserved for our experiments and no other concurrent tasks are run. The hardware we use is one NVIDIA Quadro RTX 6000 with 24 GB GDDR6 GPU Memory. We evaluate our methods (DRP-DIP and DRP-DD), DIP, and different variants of DD on the denoising task of removing low-level Gaussian noise. To save computational cost while still advocating statistical benefits, we set M to be $1,000$ for each model and record the time it needs for each iteration. We then show the histogram of these recorded optimization time in Fig. 14. As one can observe, the optimization time for each iteration is very stable and concentrated to its mean value across all models we considered. Therefore, it is plausible to use the mean value to serve as $\mathrm{T}_{per-iter}$.

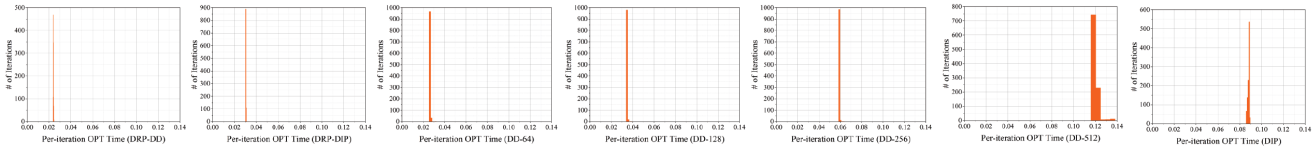

Figure 14. The histogram of per-iteration optimization time among 1,000 iterations ($M = 1000$) on removing low-level Gaussian noise across different models (from left to right): DRP-DD, DRP-DIP, DD-64, DD-128, DD-256, DD-512, and DIP. **DRP-DD** and **DRP-DIP** are our proposed methods.

C. Comparison with Shallow Deep Image Prior

"Would a simplified shallow and compact version of DIP itself with/without TV can achieve high restoration quality while requiring less optimization time?"

To answer this question, we compare DRP-DIP with the simplified shallow and compact version of DIP with or without TV regularization on image denoising (see details in Sec. 4.2). We report the comparisons in Tab. 5. From Tab. 5, one can observe that 1) S-DIP w/o TV does not perform well in restoring images (e.g., *low PSNR*) while adding TV helps (e.g., S-DIP w/ TV yields slightly higher PSNR)—this observation is consistent with the finding in [17]; 2) Both S-DIP (with/without TV) do not work as well as our proposed DRP-DIP in terms of restoration quality (e.g., our proposed DRP-DIP yields the highest PSNR in all settings compared to others); 3) Both S-DIP (with/without TV) do not show significant OPT time reduction compared to our proposed DRP-DIP.

Table 5. Comparison of DRP-DIP, S-DIP w/o TV, and S-DIP w/ TV on image denoising. We report the mean values of the PSNR (numbers outside parentheses) and the OPT time (numbers inside parentheses).

		DRP-DIP	S-DIP w/o TV	S-DIP w/ TV
Gaussian	Low	28.82 (5.22)	22.10 (393.03)	22.78 (1376.10)
	High	24.44 (2.47)	19.85 (157.01)	21.00 (588.18)
Impulse	Low	31.34 (26.86)	22.53 (1246.32)	22.54 (1266.50)
	High	25.32 (8.87)	17.30 (21.49)	18.75 (56.69)
Shot	Low	27.86 (3.95)	21.84 (384.14)	22.68 (1357.92)
	High	22.82 (2.19)	19.12 (89.97)	20.24 (579.70)
Speckle	Low	28.89 (4.50)	22.17 (401.30)	22.89 (1392.83)
	High	24.15 (2.16)	19.94 (179.02)	20.99 (535.16)

D. Performance on Clean Images

It is also intriguing to explore the scenario where the observations are clean and noise-free and to see whether our proposed RDP-DIP shows any advantages over the original DIP. We, therefore, reconstruct images from clean and noise-free observations (e.g., denoising a "clean" image). For simplicity and computational cost-saving, we here only compare our DRP-DIP with the original DIP, and for fairness comparisons, we further limit the optimization time cap to 300 seconds/5 minutes. Fig. 15 shows the quantitative comparisons. One can observe that our proposed DRP-DIP demonstrates significant advantages over the original DIP that it retains higher reconstruction quality while requiring far less optimization time, which is consistent with the findings we have in the setting of denoising noisy images.

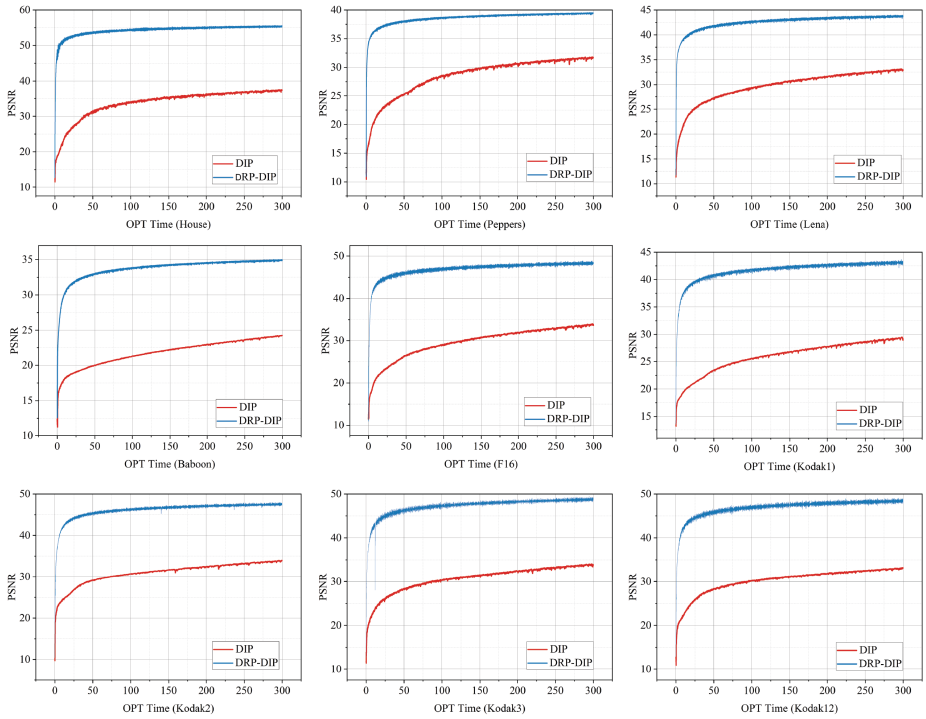

Figure 15. We reconstruct images when the observations are clean and noise-free using DRP-DIP and DIP.

E. Comparison with Total Variation

Here we compare our method DRP-DIP with TV without DIP (directly optimizing over pixels). Tab. 6 presents results for image denoising (Sec. 4.2). One can observe that DRP-DIP in general yields a higher PSNR. Besides the PSNR, we also want to emphasize that TV requires meticulous tuning on its hyperparameter on different noise types and noise levels (e.g., we set our search range for the weight of TV as $[10^{-2}, 10^{2}]$ and report the best PSNR) while the hyperparameters of our DRP-DIP are quite uniform across noise and noise levels. For impulse noise, both TV and our method adopt ℓ_1 loss. We further compare their performance for image super-resolution (Sec. 4.3) with the upsampling factor $t = 4$. We show the visualization result in Fig. 16. One can observe that TV introduces some artifacts into the restoration and makes it blurry; while the restoration by our DRP-DIP is much sharper and brighter.

Table 6. Comparison of DRP-DIP and TV in image denoising.

		TV	DRP-DIP
Gaussian	Low	28.14	28.82
	High	23.95	24.44
Impulse	Low	26.70	31.34
	High	15.15	25.32
Shot	Low	27.03	27.86
	High	22.71	22.82
Speckle	Low	28.20	28.89
	High	23.65	24.15

F. Ablation Studies

Here, we investigate the impacts of λ, the width of the backbone network, the size of the input seed, and different normalization methods, regarding the performance of our proposed methods. For simplicity and computational cost-saving, we

Figure 16. A visualization of image SR with 4× factors by TV and DRP-DIP.

only experiment with DRP-DIP on image denoising where we follow the exact same noise setting as that of Sec. 4.2. For quantitative measurement, we obtain its mean PSNR and OPT time among all 9 images that we test for each noise and noise level.

Different λ. We fix all other hyper-parameters and only vary the λ in the set $\{0.1, 0.2, 0.3, 0.4, 0.45, 0.5, 0.6, 0.7, 0.8, 0.9, 1.0\}$. We report the experimental results in Fig. 17. For restoration quality, as one can see, when λ is too small (e.g., $\lambda = 0.1$), the regularization power on the reconstructed image is very low, leading to low-quality reconstruction. It is similar to using our proposed method without TV, which has been discussed in Sec. 3.3. As we increase λ, the regularization power also keeps increasing and the restoration quality is becoming better. However, if we set λ to be extremely large (e.g., $\lambda = 1.0$), too much regularization power is given to the TV, and in turn, it harms the quality of restoration. For optimization time, the trend increases monotonically as the increase of λ, which is also expected as when λ is larger, it poses more regularization on the model, thus slows down the optimization process.

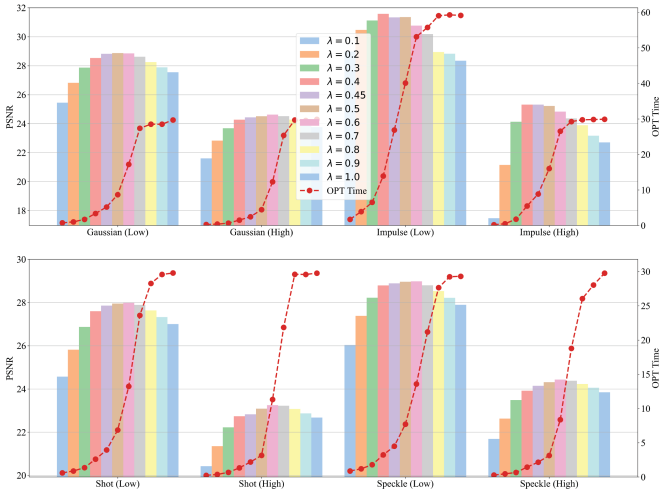

Figure 17. The PSNR and OPT time of DRP-DIP on image denoising when using different lambda.

Network width. Based on the observation in Fig. 17, we choose $\lambda = 0.45$ as default considering both the restoration quality and the optimization time. Similarly, we fix other hyperparameters and only vary the backbone network width in the set $\{64, 128, 256, 512\}$ (e.g., Width-64 denotes that the channel number for the convolution layer in the backbone network is 64). The experimental results are reported in Fig. 18. As one can observe, different widths of the backbone network have negligible impact on the restoration quality, although Width-64 yields the best restoration quality in most cases. For optimization time, although it is consistently decreasing as the network width increases, the variance is low. Therefore, these observations suggest that our proposed method is insensitive to the width of the backbone network.

Input size. Based on the observations in Fig. 17 and Fig. 18, we set $\lambda = 0.45$ and network width as 64, we then vary the input size as $512 \times 512 \times \{3, 8, 16, 32\}$ where 512×512 is equal to the spatial dimension of image height $\times$ width and

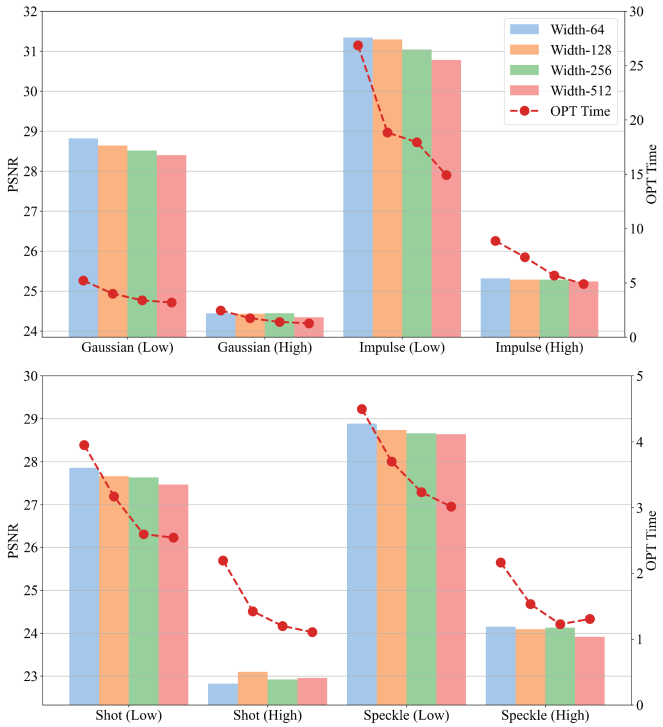

Figure 18. The PSNR and OPT time of DRP-DIP on image denoising when using backbone network with different network widths.

$\{3, 8, 16, 32\}$ denotes the different input channels. The experimental results are reported in Fig. 19. From Fig. 19, using a smaller size of input seed (e.g., $512 \times 512 \times 3$) yields a slightly better restoration quality while it requires a slightly longer optimization time. In general, the size of the input seed does not significantly change the performance of our proposed method.

Different normalization methods. Batch normalization (BN) [11] is widely used in deep learning models and can significantly smooth the optimization landscape, allowing for more stable and significantly faster training [25]. The original DIP also use BN. However, BN is not the only option for normalization. Alternatively, one can also choose instance normalization (IN) [28], layer normalization (LN) [1], and group normalization (GN) [31]. In addition, we only have one degraded observation; thus, the BN in our experiments is equal to instance normalization (IN). To investigate the influence of other normalization methods, we experiment our method with LN and GN on image denoising (Sec. 4.2) and report the performance in Tab. 7. One can observe that LN performs poorly, whereas GN marginally outperforms BN in PSNR, but takes longer to reach its peak.

Table 7. Different normalization methods when used in DRP-DIP. We report results in the form: mean PSNR (OPT time).

		DRP-DIP w/BN	DRP-DIP w/GN	DRP-DIP w/LN
Gaussian	Low	28.82 (5.22)	28.87 (9.88)	23.88 (26.67)
	High	24.44 (2.47)	24.60 (4.46)	18.83 (0.16)
Impulse	Low	31.34 (26.89)	31.54 (44.86)	19.43 (0.22)
	High	25.32 (8.87)	25.34 (13.43)	13.95 (0.27)
Shot	Low	27.86 (3.95)	28.04 (6.20)	22.78 (29.13)
	High	22.82 (2.19)	23.24 (6.26)	17.56 (0.15)
Speckle	Low	28.89 (4.50)	29.09 (8.00)	24.37 (0.17)
	High	24.15 (2.16)	24.40 (4.64)	18.97 (0.15)

G. Additional Experimental Results

We show additional results of the spectral analysis in the noise-free scenario in Fig. 20; image super-resolution in Fig. 21 and Fig. 25; image denoising in Fig. 22, Fig. 23 and Fig. 24; and image inpainting in Fig. 26 and Fig. 27.

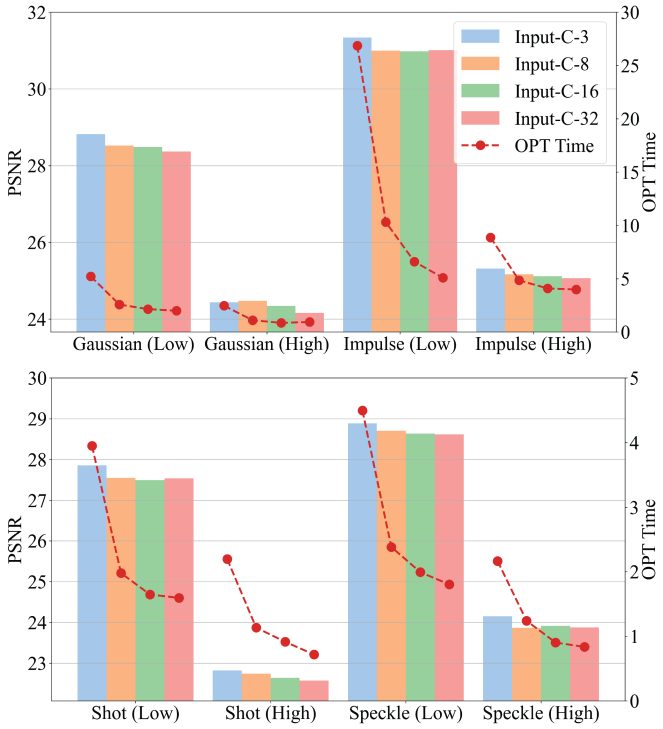

Figure 19. The PSNR and OPT time of DRP-DIP on image denoising when using different sizes for input seed.

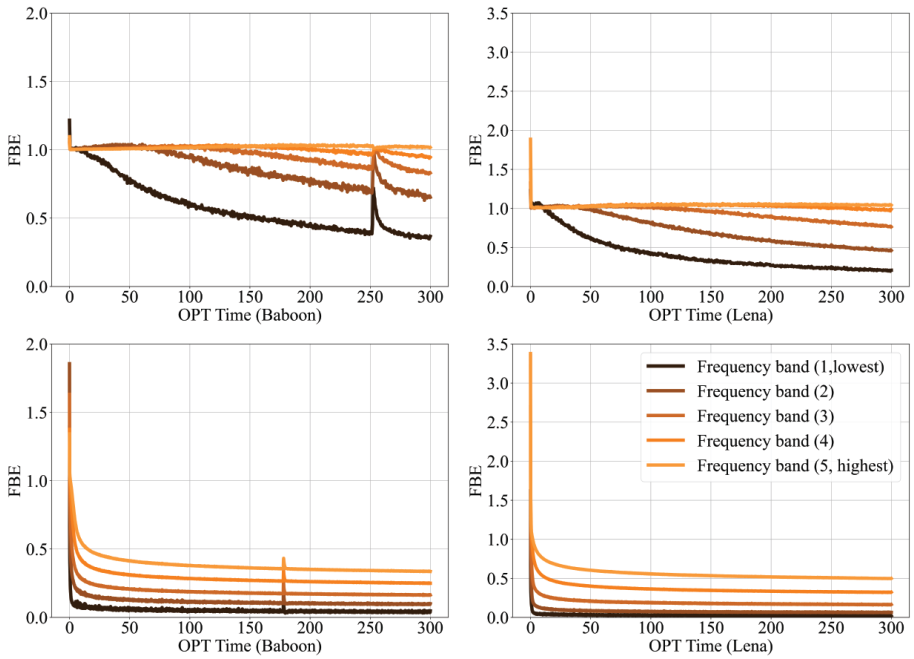

Figure 20. We compare the spectral bias of DIP (first row) and DRP-DIP (second row) under the noise-free scenario.

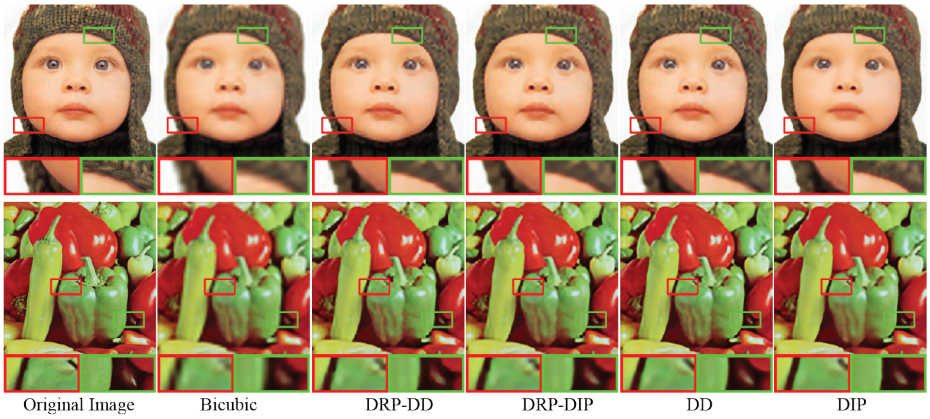

Figure 21. A visualization of image SR with $8\times$ factors. Our methods (DRP-DD and DRP-DIP) produce restorations that are visually superior to Bicubic and comparable to DD and DIP.

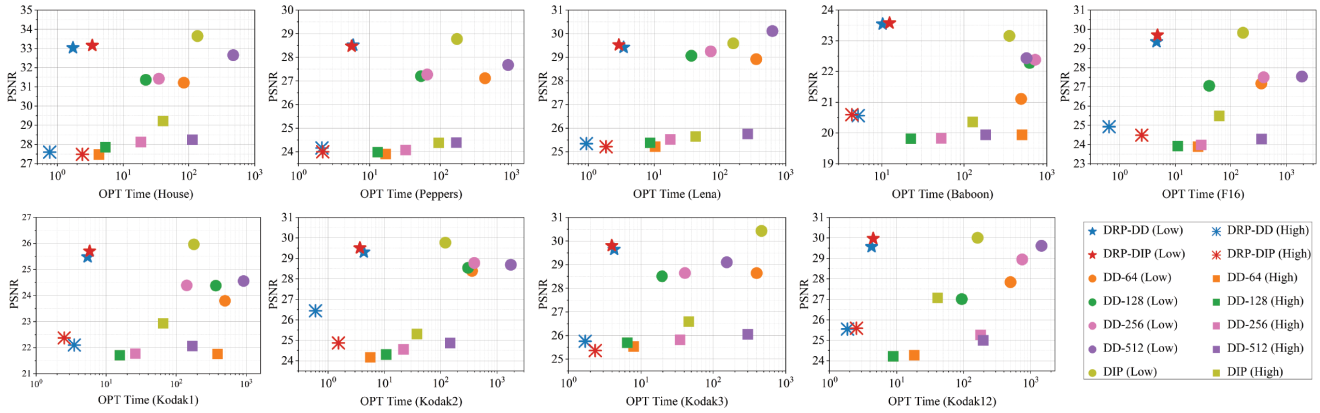

Figure 22. Quantitative comparison of image denoising on low- and high-level Gaussian noise. DRP-DD and DRP-DIP are our proposed methods.

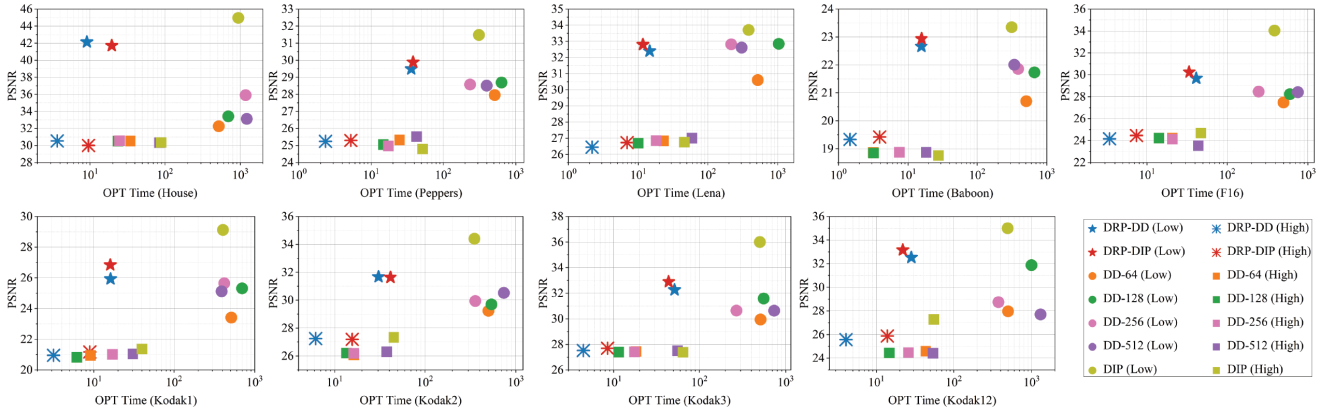

Figure 23. Quantitative comparison of image denoising on low- and high-level impulse noise. DRP-DD and DRP-DIP are our proposed methods.

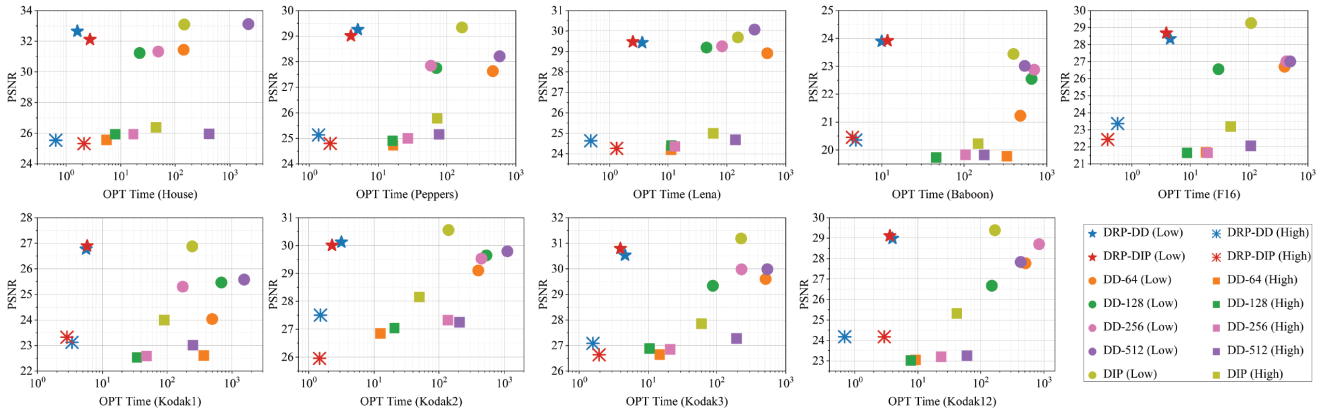

Figure 24. Quantitative comparison of image denoising on low- and high-level speckle noise. DRP-DD and DRP-DIP are our proposed methods.

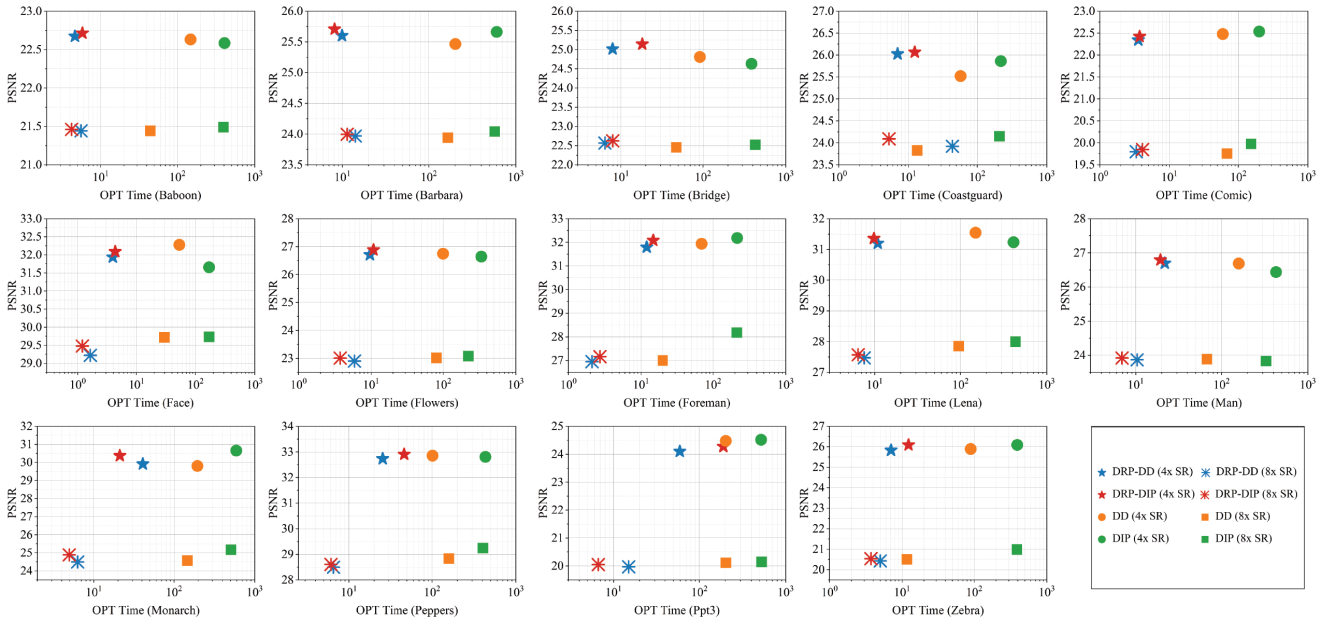

Figure 25. Quantitative comparison of image super-resolution on Set14. DRP-DD and DRP-DIP are our proposed methods.

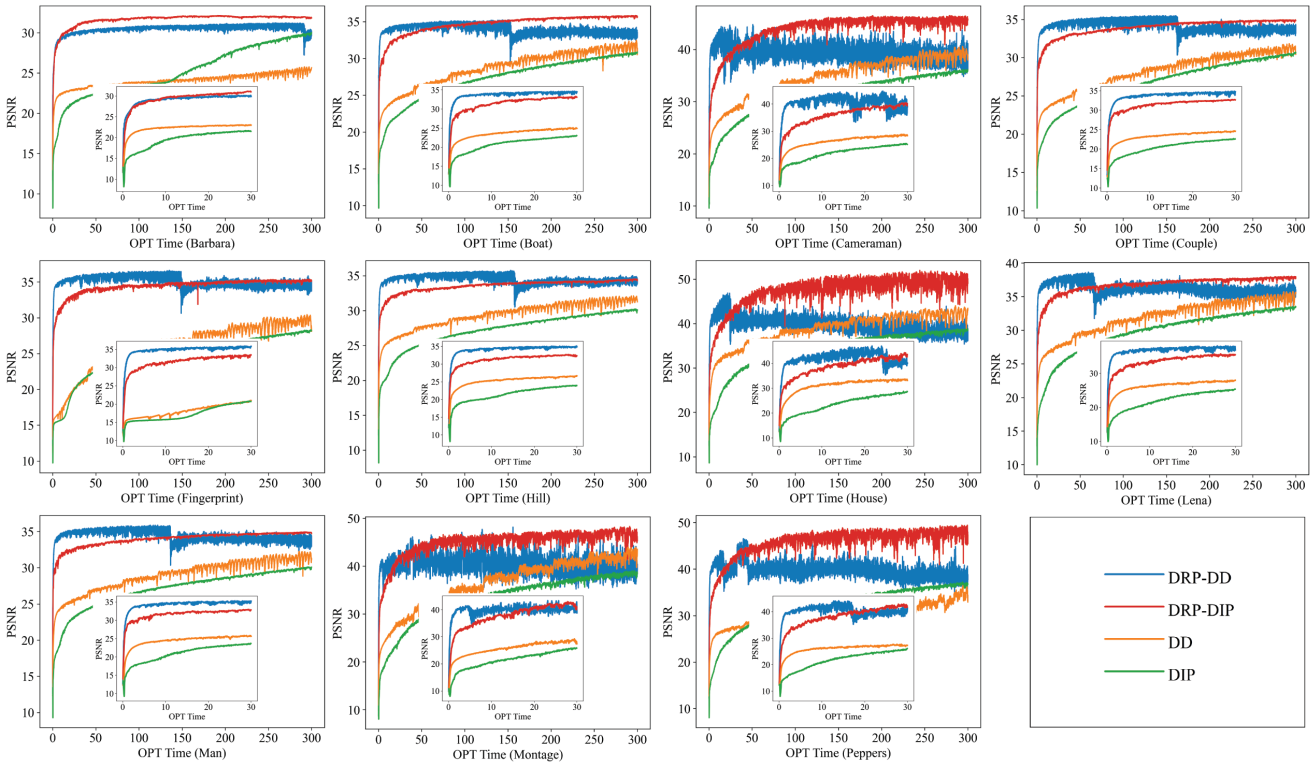

Figure 26. Quantitative comparison of image inpainting with 30% of missing pixels. The small subfigure shows the comparison in the first 30 second. DRP-DD and DRP-DIP are our proposed methods.

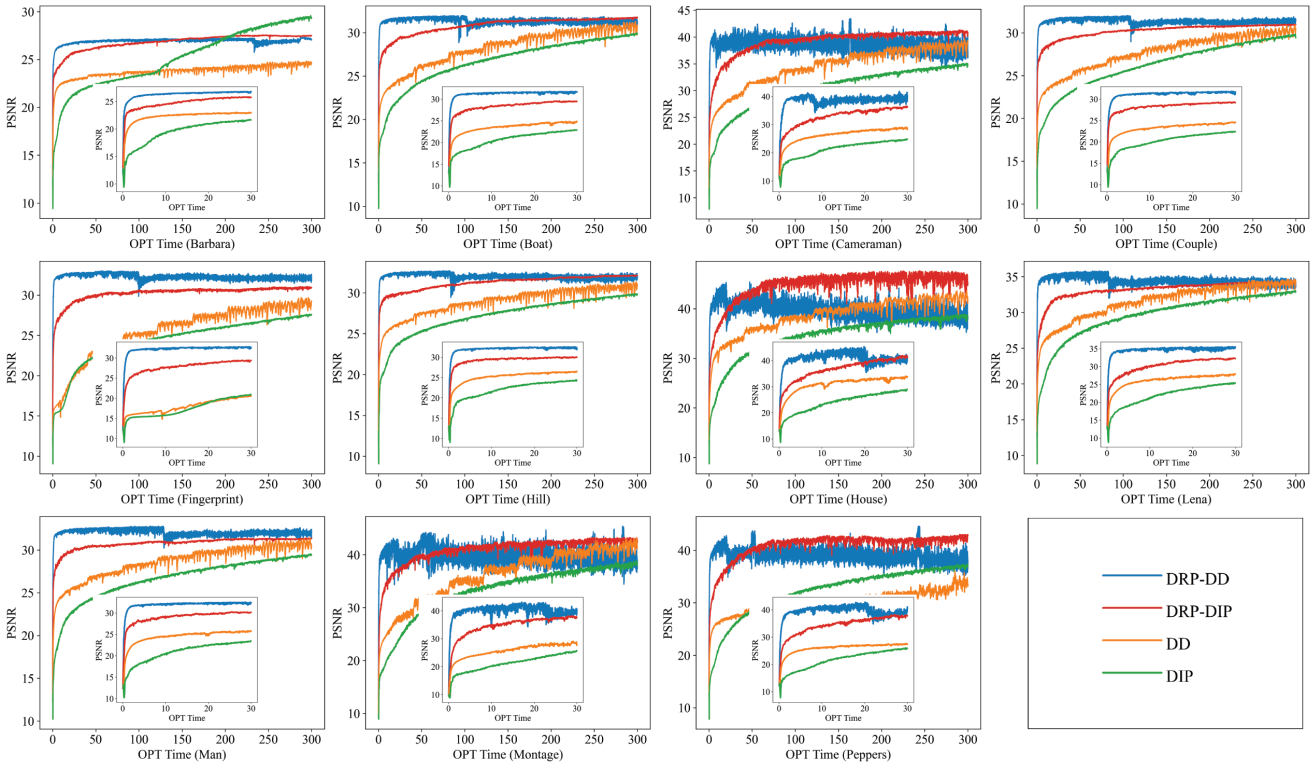

Figure 27. Quantitative comparison of image inpainting with 50% of missing pixels. The small subfigure shows the comparison in the first 30 second. DRP-DD and DRP-DIP are our proposed methods.