Supplementary Material of "Blind Image Super-resolution with Elaborate Degradation Modeling on Noise and Kernel"

Zongsheng Yue[1], Qian Zhao[2], Jianwen Xie[3], Lei Zhang[4], Deyu Meng[2,5], Kwan-Yee K. Wong[1]

[1]The University of Hong Kong, Hong Kong, China [2]Xi'an Jiaotong University, Xi'an, China
[3]Cognitive Computing Lab, Baidu Research, Bellevue, USA
[4]The Hong Kong Polytechnic University, Hong Kong, China
[5]Peng Cheng Laboratory, Shenzhen, China

Abstract

In this supplementary material, we provide more details on the calculation of the E-Step and the network architectures of the image generator G. Besides, we also present more additional experiments.

1. Calculation Details on the E-Step

Given current model parameters $\{\boldsymbol{\alpha}_{\text{old}}, \boldsymbol{L}_{\text{old}}, \boldsymbol{\lambda}_{\text{old}}\}$, we denote the posterior of $\boldsymbol{z}$ under them as $p_{\text{old}}(\boldsymbol{z}|\boldsymbol{y})$. In E-Step, our goal is to sample $\boldsymbol{z}$ from $p_{\text{old}}(\boldsymbol{z}|\boldsymbol{y})$ using Langevin dynamics [12]:

$$
\begin{aligned}
\boldsymbol{z}^{(\tau+1)} &= \boldsymbol{z}^{(\tau)} + \frac{\delta^2}{2}\left[\frac{\partial}{\partial \boldsymbol{z}}\log p_{\text{old}}(\boldsymbol{z}|\boldsymbol{y})\right]\Bigg|_{\boldsymbol{z}=\boldsymbol{z}^{(\tau)}} + \delta\boldsymbol{\zeta}^{(\tau)} \\
&= \boldsymbol{z}^{(\tau)} + \frac{\delta^2}{2}\left[\frac{\partial}{\partial \boldsymbol{z}}\log p_{\text{old}}(\boldsymbol{z},\boldsymbol{y})\right]\Bigg|_{\boldsymbol{z}=\boldsymbol{z}^{(\tau)}} + \delta\boldsymbol{\zeta}^{(\tau)} \\
&= \boldsymbol{z}^{(\tau)} - \frac{\delta^2}{2}\left[\frac{\partial}{\partial \boldsymbol{z}} g(\boldsymbol{z})\right]\Bigg|_{\boldsymbol{z}=\boldsymbol{z}^{(\tau)}} + \delta\boldsymbol{\zeta}^{(\tau)},
\end{aligned} \tag{1}
$$

where

$$
\begin{aligned}
g(\boldsymbol{z}) = &\frac{1}{2}\left\|\frac{1}{\boldsymbol{\lambda}_{\text{old}}}\odot\left\{\boldsymbol{y} - \left[G(\boldsymbol{z};\boldsymbol{\alpha}_{\text{old}}) * h(\boldsymbol{L}_{\text{old}})\right]\downarrow_s^d\right\}\right\|_2^2 \\
&+ \rho\sum_{k=1}^{2}|f_k * G(\boldsymbol{z};\boldsymbol{\alpha}_{\text{old}})|^\gamma + \frac{1}{2}\|\boldsymbol{z}\|_2^2,
\end{aligned} \tag{2}
$$

τ indexs the time step for Langevin dynamics, δ denotes the step size, $\boldsymbol{\zeta}$ is the Gaussian white noise used to prevent trapping into local modes, $\odot$ represents the Hadamard product. As for the derivation of $g(\boldsymbol{z})$, we firstly factorize $p_{\text{old}}(\boldsymbol{z},\boldsymbol{y})$ as follows:

$$
p_{\text{old}}(\boldsymbol{z},\boldsymbol{y}) = p(\boldsymbol{y}|\boldsymbol{\alpha}_{\text{old}}, \boldsymbol{L}_{\text{old}}, \boldsymbol{\lambda}_{\text{old}}, \boldsymbol{z})p(\boldsymbol{\alpha}_{\text{old}}|\boldsymbol{z})p(\boldsymbol{z}), \tag{3}
$$

where $p(\boldsymbol{y}|\boldsymbol{\alpha}_{\text{old}}, \boldsymbol{L}_{\text{old}}, \boldsymbol{\lambda}_{\text{old}}, \boldsymbol{z})$, $p(\boldsymbol{\alpha}_{\text{old}}|\boldsymbol{z})$, and $p(\boldsymbol{z})$ are defined in Eq. (5), Eq. (11), and Eq. (12) of the maintext, respectively. By substituting these three terms into Eq. (3), we can easily obtain the formulation in Eq. (2) after simple derivation.

Table 1. Performances of the proposed BSRDM with different settings of ρ on Set14. The PSNR/SSIM/LPIPS results are all averaged on different degradations combined with camera sensor noise and six different blur kernels (see Fig. 2 of the maintext) under scale factor 2.

ρ	Metrics		
	PSNR↑	SSIM↑	LPIPS↓
0	27.20	0.725	0.379
0.01	27.37	0.737	0.378
0.10	27.84	0.762	0.366
0.20	28.01	0.771	0.360
0.30	28.06	0.774	0.356
0.40	28.09	0.774	0.355
0.50	28.06	0.772	0.355
1.00	27.56	0.744	0.383

Table 2. Performances of the proposed BSRDM with different settings of γ on Set14. The PSNR/SSIM/LPIPS results are all averaged on different degradations combined with camera sensor noise and six different blur kernels (see Fig. 2 of the maintext) under scale factor 2.

γ	Metrics		
	PSNR↑	SSIM↑	LPIPS↓
0.67	28.01	0.771	0.360
1.00	27.83	0.760	0.367
2.00	27.27	0.738	0.375

2. Network Architecture

As for the generator G in the maintext, we follow the "hourglass" archtechture in DIP [10]. However, we used a very tiny version that contains much fewer parameters as shown in Sec. 5.4 of the maintext. The detailed network architecture is shown in Fig. 1. Note that, as for the upsampling operation, the nearest interpolation is employed.

3. Experimental Results

3.1. Hyper-parameter Analysis

As shown in Sec. 3.2 of the maintext, our proposed BSRDM mainly involves two hyper-parameters, i.e., ρ and γ. Next, we empirically analyse the sensitiveness of

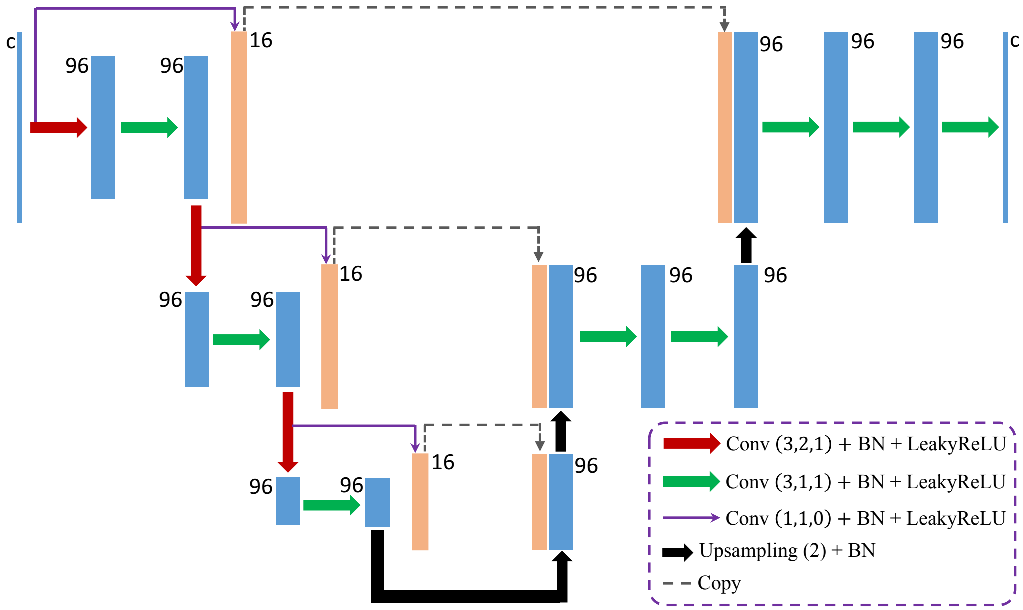
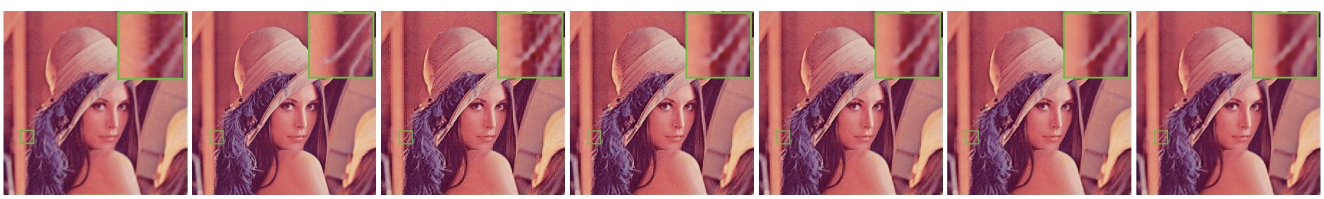
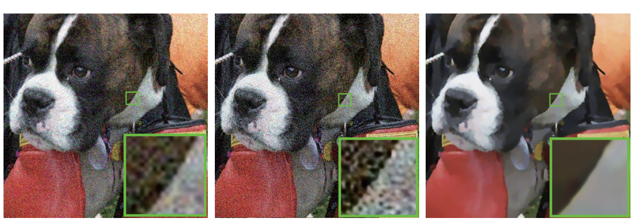


Figure 1. The detailed network architecture of the generator G. "Conv (k,p,s)" represents the 2-D convolution operator with kernel size k, stride s and reflection padding size p, "BN" represents the Batch Normlization layer, "LeakyReLU" represents the LeakyReLU activation function with negative slope 0.25, and "Upsampling (s)" represents the nearest interpolation operator with scale factor s. The blue or orange rectangles denote the feature maps of the intermediate layers, and the numbers along them are the corresponding number of channels.



(a) Zoomed LR (x2) (b) HR Image (c) ρ = 0 (d) ρ = 0.10 (e) ρ = 0.20 (f) ρ = 0.40 (g) ρ = 1.00

Figure 2. One typical example of the proposed method under different settings of ρ for the degradation with camera sensor noise on Set14. From left to right: (a) the zoomed LR image, (b) the HR image, (c)-(d) the super-resolved results of BSRDM under different ρ values.



(a) Zoomed LR (x4) (b) DIPFKP (c) BSRDM

Figure 3. Visual super-resolution results of the "dog" example in RealSRSet [14]. From left to right: (a) the zoomed LR image, (b)-(c) the recovered HR images of DIPFKP [5] and the proposed BSRDM, respectively.

BSRDM to them.

Hyper-parameter ρ: Intuitively, the hyper-paramter ρ controls the relative importance of the hyper-Laplacian prior in our method. Table 1 lists the PSNR/SSIM performance of our proposed BSRDM under different ρ values on Set14 [13], and one corresponding visual results are shown in Fig. 2. It can be easily seen that BSRDM performs very stably and well in range of [0.2, 0.5], but larger ρ value tends to produce more smooth results. Therefore, taking both of the quantitative and qualitative results into consideration, we set ρ to be 0.20 in our experiments.

Hyper-parameter γ: The hyper-paramter γ reflects the strength of the sparsity constraint on the image gradients. The Eq. (11) of the maintext degenerates into the traditional Laplacian or Gaussian distribution when γ equals 1 or 2. Dilip Krishnan and Rob Fergus [4] pointed out that the hyper-Laplacian with $\gamma = {}^2\!/_3$ is a better model of image gradients than a Laplacian or a Gaussian. Here, we list the quantitative performance of our BSRDM under different settings of γ in Table 2. It can be easily observed that BSRDM achieves the best results when γ equals to ${}^2\!/_3$, which is in accordance with the conclusion of Dilip Krishnan and Rob Fergus [4].

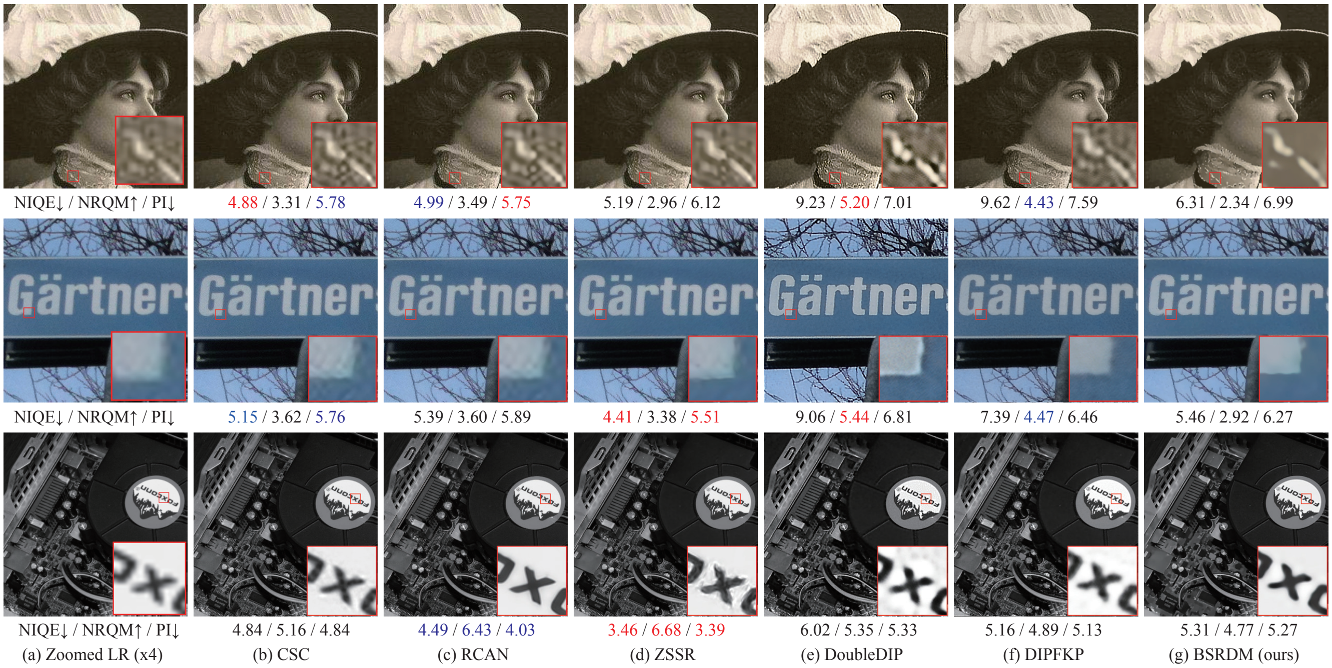


Figure 4. Three typical visual results on the RealSRSet [14] with scale factor 4. The best and second best non-reference metrics are highlighted in red and blue. Please zoom in for best view.

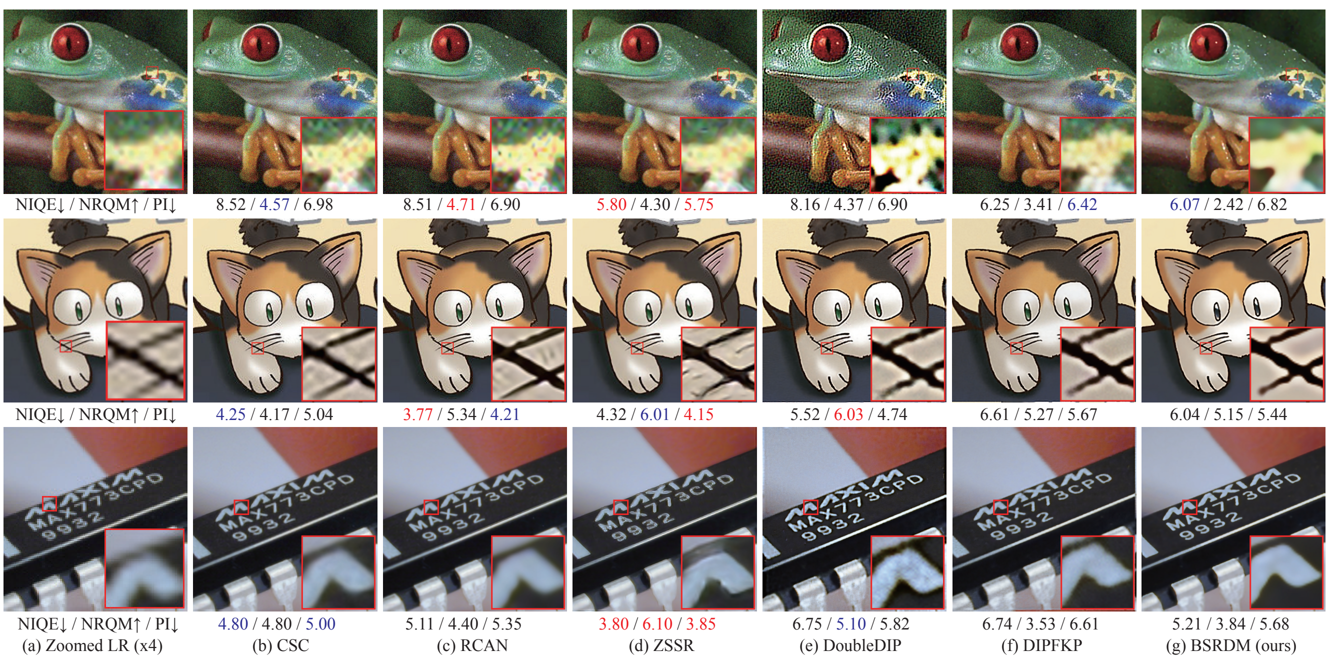


Figure 5. Three typical visual results on the RealSRSet [14] with scale factor 4. The best and second best non-reference metrics are highlighted in red and blue. Note that this figure is the same with the Fig. 5 of the maintext, but the non-reference metrics (i.e., NIQE, NRQM and PI) are additionally marked for each image. Please zoom in for best view.

3.2. Limitations

Figure 3 displays a real super-resolution example, in which the LR image is heavily corrupted by camera sensor noise. It can be easily observed that the current SotA method DIPFKP cannot handle such case with complicated real noise, its recovered result contains obvious artifacts. On the contrary, the proposed BSRDM is able to remove most of the noise and obtains clean super-resolved HR image. Even though achieving superior performance, BSRDM still has two major limitations. Firstly, the recovered image

Table 3. The non-reference NIQE, NRQM and PI comparison results of different methods on the RealSRSet data set. The best and second best results are highlighted in red and blue.

Metrics	Methods					
	CSC [3]	RCAN [15]	ZSSR [9]	DoubleDIP [8]	DIPFKP [5]	BSRDM (ours)
NIQE↓	5.87	5.61	4.73	7.29	7.04	6.23
NRQM↑	4.16	4.58	5.36	5.22	4.45	3.99
PI↓	5.85	5.51	4.69	6.04	6.29	6.12

Table 4. PSNR/SSIM/LPIPS results of different comparison methods on DIV2K100. All the results are averaged on six differen degradations with blur kernels as shown in Fig. 2 of the maintext. The best results are highlighted in **bold**. The gray results indicate unfair comparisons due to the mismatched degradations.

Noise types	Scale	Metrics	Methods					
			RCAN [15]	ZSSR-B [9]	ZSSR-NB [9]	DoubleDIP [8]	DIPFKP [5]	BSRDM (ours)
Case 1	×2	PSNR↑	25.92	26.00	**30.52**	25.17	27.38	29.07
		SSIM↑	0.720	0.734	**0.855**	0.689	0.749	0.800
		LPIPS↓	0.343	0.322	**0.284**	0.448	0.398	0.337
	×3	PSNR↑	22.99	23.13	27.18	22.05	26.68	**28.22**
		SSIM↑	0.598	0.616	0.766	0.579	0.718	**0.769**
		LPIPS↓	0.407	0.397	0.376	0.517	0.452	**0.373**
	×4	PSNR↑	21.16	21.43	26.85	20.17	25.89	**27.20**
		SSIM↑	0.526	0.548	**0.736**	0.514	0.696	0.732
		LPIPS↓	0.467	0.462	0.423	0.546	0.474	**0.414**
r Case 2	×2	PSNR↑	25.49	25.69	27.72	24.88	27.21	**28.14**
		SSIM↑	0.689	0.708	0.761	0.685	0.748	**0.779**
		LPIPS↓	0.415	0.397	0.397	0.460	0.415	**0.385**
	×3	PSNR↑	22.77	22.91	25.71	21.69	26.16	**26.84**
		SSIM↑	0.580	0.599	0.702	0.566	0.698	**0.730**
		LPIPS↓	0.497	0.480	0.470	0.541	0.492	**0.401**
	×4	PSNR↑	21.16	21.24	25.10	20.06	25.10	**25.71**
		SSIM↑	0.519	0.538	0.672	0.503	0.660	**0.685**
		LPIPS↓	0.551	0.540	0.517	0.582	0.535	**0.509**

of BSRDM is smooth, since the L_2 loss function (see Eq. (17) of the maintext) and the hyper-Laplacian prior on image gradients in it both favor smoothing the generated HR image. Secondly, BSRDM cannot hallucinate more image textures that not exists in the observed LR image, e.g., hairs of the dog in Fig. 3, and is thus inferior to the GAN-based methods [11, 11] on this point. In the future, it might be expected to develop more powerful image priors specifically to overcome these limitations.

3.3. Experiments on the Real Data

3.3.1 More Visual Results

Figure 4 displays three more visual super-resolution results on RealSRSet [14] with scale factor 4. In the first (top row) and second (middle row) examples, the LR image is with obvious camera sensor noise. The comparison methods cannot deal with such degradation with complicated real noises, while our BSRDM is able to remove most of the noises, indicating the effectiveness of the proposed non-i.i.d. noise modeling method. In the third example (bottom row), it can be easily seen that the recovered HR image by BSRDM is with sharper clearer details.

3.3.2 Disscussion on the Non-reference Metrics

Since the ground-truth for the RealSRSet [14] is not available, three non-reference metrics (i.e., NIQE [7], NRQM [6] and PI [2]) are considered as quantitative evaluation. As shown in Table 3, BSRDM and the current SotA method DIPFKP [5] both fail to achieve promising results. However, in Fig. 4 and Fig. 5, we can easily observed that the recovered results by BSRDM is evidently better than other comparison methods. We argue that these non-reference metrics are not consistent with our perceptual visual system. In the future work, we will make our best effort to develop more rational non-reference metric to match with and facilitate current researches on SISR.

3.4. Experiments on the Synthetic Data

In Table 4, we list the performance comparisons of different methods on the dataset DIV2K100 [1]. Note that, due to the computer memory limitation, we cannot give the results of CSC [3] in Table 4. It can be easily observed that the proposed BSRDM illustrates obvious superiorities than the comparison methods, which is consistent with that on Set14 in Table 1 of the maintext, Furthermore, we display more visual results of different methods on the synthetic data sets in Fig. 6 (Gaussian noise) and Fig. 7 (camera sensor noise).

References

[1] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE/CF conference on computer vision and*

pattern recognition workshops (CVPRW), pages 126–135, 2017. 4

[2] Yochai Blau, Roey Mechrez, Radu Timofte, Tomer Michaeli, and Lihi Zelnik-Manor. The 2018 pirm challenge on perceptual image super-resolution. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, pages 0–0, 2018. 4

[3] Shuhang Gu, Wangmeng Zuo, Qi Xie, Deyu Meng, Xiangchu Feng, and Lei Zhang. Convolutional sparse coding for image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 1823–1831, 2015. 4

[4] Dilip Krishnan and Rob Fergus. Fast image deconvolution using hyper-laplacian priors. In *Advances in Neural Information Processing Systems*, volume 22, pages 1033–1041, 2009. 2

[5] Jingyun Liang, Kai Zhang, Shuhang Gu, Luc Van Gool, and Radu Timofte. Flow-based kernel prior with application to blind super-resolution. pages 10601–10610, 2021. 2, 4

[6] Chao Ma, Chih Yuan Yang, Xiaokang Yang, and Ming Hsuan Yang. Learning a no-reference quality metric for single-image super-resolution. *Computer Vision and Image Understanding*, 158:1–16, 2017. 4

[7] Anish Mittal, Rajiv Soundararajan, and Alan C Bovik. Making a "completely blind" image quality analyzer. *IEEE Signal processing letters*, 20(3):209–212, 2012. 4

[8] Dongwei Ren, Kai Zhang, Qilong Wang, Qinghua Hu, and Wangmeng Zuo. Neural blind deconvolution using deep priors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3341–3350, 2020. 4

[9] Assaf Shocher, Nadav Cohen, and Michal Irani. "zero-shot" super-resolution using deep internal learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3118–3126, 2018. 4

[10] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 9446–9454, 2018. 1

[11] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 1905–1914, 2021. 4

[12] Max Welling and Yee W Teh. Bayesian learning via stochastic gradient langevin dynamics. In *Proceedings of the International Conference on Machine Learning (ICML)*, pages 681–688. Citeseer, 2011. 1

[13] Roman Zeyde, Michael Elad, and Matan Protter. On single image scale-up using sparse-representations. In *Proceedings on International Conference on Curves and Surfaces*, pages 711–730. Springer, 2010. 2

[14] Kai Zhang, Jingyun Liang, Luc Van Gool, and Radu Timofte. Designing a practical degradation model for deep blind image super-resolution. *arXiv preprint arXiv:2103.14006*, 2021. 2, 3, 4

[15] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European conference on computer vision (ECCV)*, pages 286–301, 2018. 4

Figure 6. Visual super-resolution results of different methods for the degradation with Gaussian noise under scale factor 3. The blur kernel is shown on the upper-right conner of the zoomed LR image.

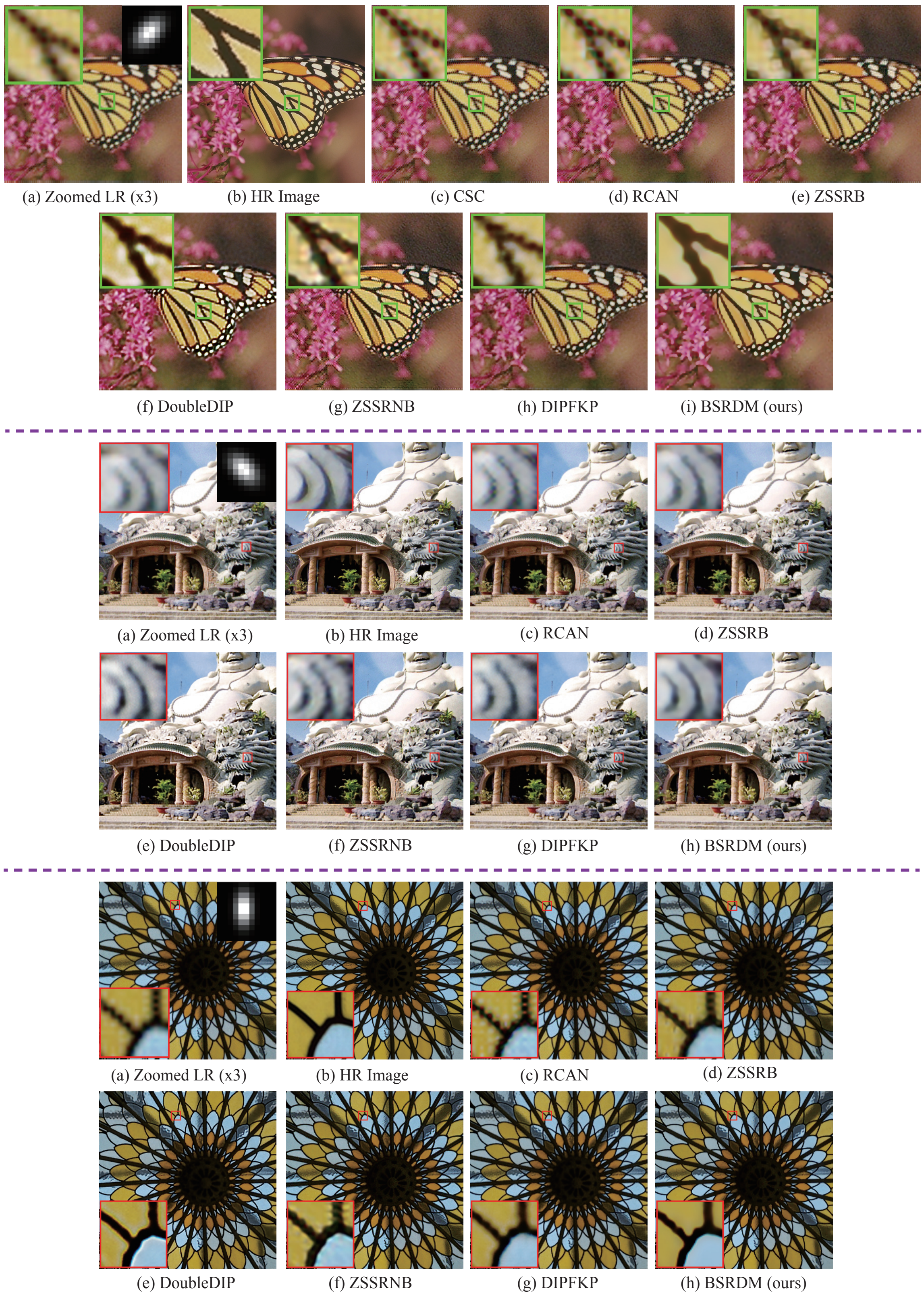


Figure 7. Visual super-resolution results of different methods for the degradation with camera sensor noise under scale factor 3. The blur kernel is shown on the upper-right conner of the zoomed LR image. Note that due to the computer memory limitation, we cannot provide the super-resolution result of the method CSC for the second and third example in DIV2K100.