Supplementary Material
Learning Correction Filter via Degradation-Adaptive Regression for Blind Single Image Super-Resolution

Hongyang Zhou[1], Xiaobin Zhu[1], Jianqing Zhu[2], Zheng Han[1], Shi-Xue Zhang[1]
Jingyan Qin[1], Xu-Cheng Yin[1]
[1]University Of Science and Technology Beijing, [2]College of Engineering, Huaqiao University

1. Preliminaries

The SISR degradation model can be mathematically formulated as:

$$x = (y \otimes k) \downarrow_s, \tag{1}$$

where y denotes the HR image; x denotes the LR image; k is the blur kernel; $\otimes$ denotes convolution operator; $\downarrow_s$ denotes sub-sampling operator with stride of s. Note that Equation 1 can be written as:

$$x = Dy. \tag{2}$$

where D represents degradation operator. As assumed by Hussein *et al* [7], the signal y can be perfectly recovered from its samples x by the operator $U(DU)^{-1}$ as:

$$y = U(DU)^{-1}x = U(DU)^{-1}Dy, \tag{3}$$

where U is an upsampling operator which is the adjoint operator of D. Equation 3 indicates that the image y resides in the linear subspace can be spanned by the known degradation. Therefore, it can be perfectly recovered from the observations x_k by applying the pseudo inverse of U_k as:

$$y' = U_k(D_kU_k)^{-1}D_ky = U_k(D_kU_k)^{-1}x_k, \tag{4}$$

where D_k denotes known degradation operation; x_k denotes known degraded image; U_k denotes a DNN super-resolver trained under the assumption of known degradation and it can handle the x_k quite well with a learned prior.

According to the generalization of classical Whittaker-Shannon sampling theorem [8, 16, 17, 25], some works [2, 4, 23] consider that a signal sampled by a certain basis can be reconstructed by a different basis. By combining a digital correction filter with a reconstruction kernel, a sampled signal can be reconstructed. The correction filter is responsible for transforming the sampling coefficients associated with the sampling kernel to coefficients that fit the reconstruction kernel. Given a degraded LR image x_u that are obtained by a unknown degradation D_u, $x_u = D_uy$. According to above generalized sampling theory, the y' can be reconstructed as:

$$y' = U_kCx_u = U_kCD_uy, \tag{5}$$

where C is a correction operator for x_u. Combining equation 4 and equation 5, we can find the expression of C as follows:

$$\begin{aligned} U_kCD_uy &= U_k(D_kU_k)^{-1}D_ky \\ CD_u &= U_k(D_kU_k)^{-1} \\ CD_u &= U_k^{-1}D_k^{-1}D_k \\ CD_u &= U_k^{-1} \\ CD_uD_u^{-1} &= U_k^{-1}D_u^{-1} \\ C &= (D_uU_k)^{-1}. \end{aligned} \tag{6}$$

Then, we change the equation 6 as follows:

$$C = (D_uU_k)^{-1} \Rightarrow CD_u = U_k^{-1} \tag{7}$$

2. The Dictionary D

In our proposed degradation-adaptive regression module (DARM), the dictionary $\mathbb{D}$ we used is composed of 72 Gaussian and difference of Gaussians (DoG) filters [10]. As mentioned in [27, 20, 19, 10], Gaussians have strong representation ability. The mathematical formulation of Gaussian elliptic function is elaborated as:

$$G(x - x'; \Sigma) = \frac{1}{2\pi|\Sigma|^{\frac{1}{2}}} e^{-\frac{1}{2}(x-x')^T\Sigma^{-1}(x-x')} \tag{8}$$

where x and x' denote coordinates of neighboring pixels and central pixel; Σ denotes the covariance matrix. Based on the anisotropic Gaussian filters, the mathematical formulation for generating DoG filters can be elaborated as:

$$DoG(x-x'; \Sigma_1, \Sigma_2) = G(x-x'; \Sigma_1) - G(x-x'; \Sigma_2) \tag{9}$$

All 72 filters are normalized to sum total to 1. The detailed filters used in the experiment are shown in the figure 1.

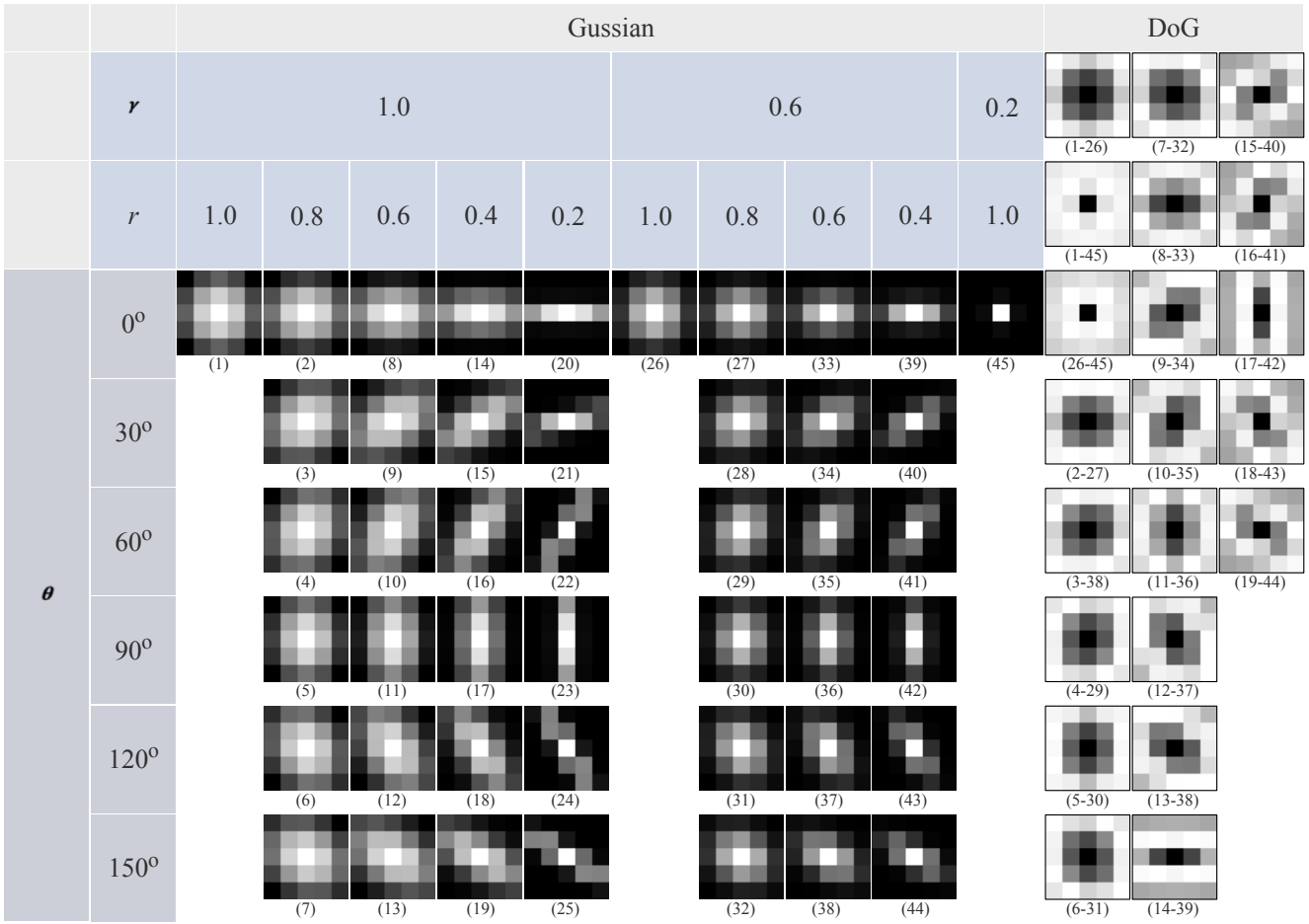

Figure 1. Visualization of the filters in the used dictionary which proposed by [10]. The symbols γ, θ, and r denote scaling, rotation, and elongation ratio respectively. The Gaussian is denoted by (a), and DoG (a-b) means the difference of Gaussian (a) and Gaussian (b).

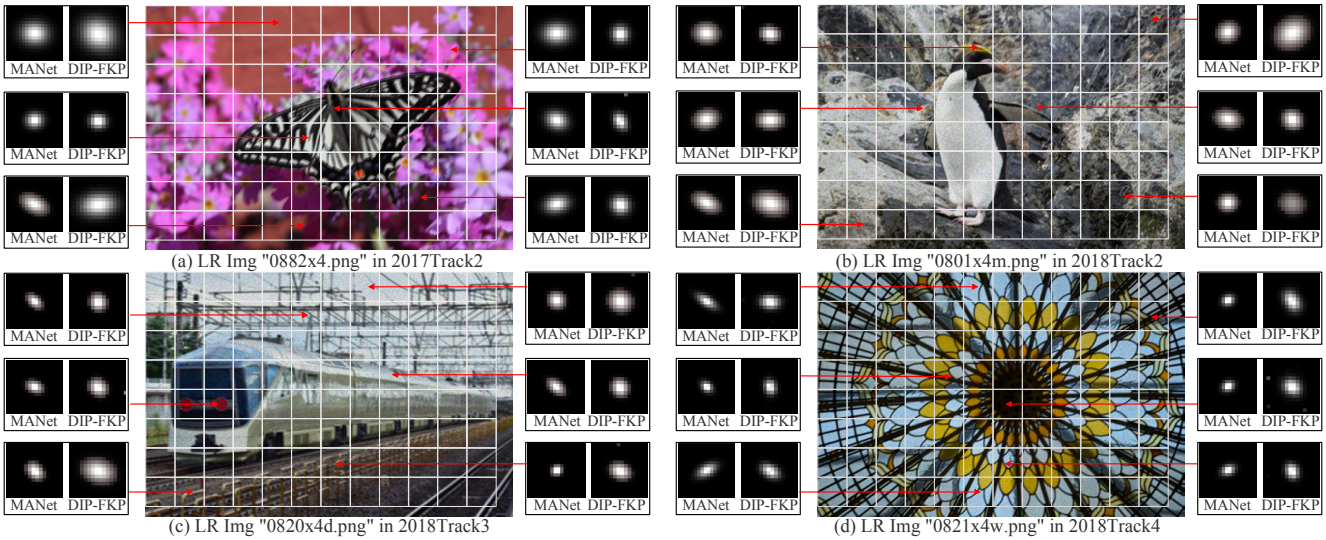

Figure 2. Illustration of degradation estimation for MANet [12] and DIP-FKP [13] on NTIRE2017 and NTIRE2018 testing datasets.

3. Spatial-Variant Degradation in LR Image

We conduct experiments on track2 of NTIRE2017 [21] and track2, track3, and track4 of NTIRE2018 [22]. We first divide the LR image into patches of size 40×40. Then we use MANet [12] and DIP-FKP [13] to estimate the degradation of different patches in LR image. As shown in fig-

ure 2, we can clearly see that the degradation of the four LR images is spatially variant. As mentioned by Liang *et al.* [12], due to different environmental factors like object motion, depth difference, out-of-focus, and camera shake, degradation at different locations of the image tend to be different. Although the official uses same degradation operators within an image, the given HR images are spatially variant degraded from the real HR ones due to camera acquisition in complex scenes, which make the LR images are spatially variant.

4. Training and Testing

We train our networks (DSEM, D, G, DARM) for input LR image to learn correction filters in an unsupervised manner. Then, we use learned correction filters to correct the input LR image to the pre-defined degradation. Finally, a trained SR model trained by pre-defined degradation is used for upsampling the degradation adapted LR image for SR. Note that when we input a new LR image, our networks (DSEM, D, G, DARM) will be reinitialized before training. The details training and testing process are presented in Algorithm 1.

Algorithm 1 Training and Testing

Input:
LR image: I^{LR}

Output:
Corrected LR image: I_c^{LR}
SR image: I^{SR}

Networks:
DSEM: Degradation Score Estimation Module
G: degradation Generator
D: Discriminator
DARM: Degradation-Adaptive Regression Module

Training:

1: Initialize parameters of DSEM, G, D, and DARM
2: Generate training data $X = \{x^i\}$, $Y = \{y^i\}$, and $X_b = \{x_b^i\}$ from I^{LR}
3: Use $X = \{x^i\}$ and $X_b = \{x_b^i\}$ to train DSEM
4: Use $X = \{x^i\}$, $X_b = \{x_b^i\}$, $Y = \{y^i\}$, trained DSEM, and pre-trained SR model to train D, G, and DARM
5: **return** trained DARM

Testing:

1: Use trained DARM to correct I^{LR} for getting I_c^{LR}
2: Use pre-trained SR model to upsample I_c^{LR} for getting I^{SR}
3: **return** I_c^{LR}, I^{SR}

5. Adaptability of Existing Super-Resolvers

We conduct experiments on track2 of NTIRE2018 [22]. To verify the effectiveness of our method, we compare our model with SRGAN [9], ESRGAN [24], SRFBN-s [11], EDSR [14], and RCAN [26]. The detailed experimental results are listed in Table 1. According to Table 1, our method can improve the performance of all methods in terms of PSNR/SSIM. We have noticed that our method fail to improve the performance of EDSR, RCAN and SRFBN-s in terms of LPIPS. But our method can improve the performance of SRGAN and ESRGAN on all indicators. As mentioned by Cheng *et al.* [15], EDSR, RCAN, and SRFBN-s are PSRN-oriented methods, which focus on higher PSNR and tend to generate images with too smooth and lost details. Although our method can effectively recover LR images, the characteristics of PSRN-oriented methods still lead to LPIPS degradation. While SRGAN and ESRGAN are Perceptual-Driven methods, which have inherent advantages in LPIPS. So our method can improve the performance of all indicators of the Perceptual-Driven methods. In general, the experimental results verify that our method can consistently improve the performance of trained super-resolvers.

Table 1. Quantitative Results on NTIRE2018 Track2 [22] testing dataset for scaling factor $\times 4$. $\uparrow$ denotes the larger the better. $\downarrow$ denotes the smaller the better. Red color indicates the best performance, and blue color indicates the second best performance.

Methods	PSNR↑	SSIM↑	LPIPS↓
SRGAN [9]	18.99	0.4582	0.6127
ESRGAN [24]	19.23	0.5699	0.5716
EDSR [14]	18.85	0.4873	0.6134
RCAN [26]	19.01	0.4891	0.6138
SRFBN-s [11]	18.87	0.4886	0.6085
SRGAN [9] + Correction Filter [7]	18.73	0.5555	0.6141
ESRGAN [24] + Correction Filter [7]	18.97	0.5845	0.5723
EDSR [14] + Correction Filter [7]	18.72	0.5628	0.6246
RCAN [26] + Correction Filter [7]	18.93	0.5789	0.6271
SRFBN-s [11] + Correction Filter [7]	18.74	0.5631	0.6255
Ours + SRGAN [9]	20.24	0.5741	0.6088
Ours + ESRGAN [24]	20.54	0.5927	0.5363
Ours + EDSR [14]	20.29	0.5763	0.6246
Ours + RCAN [26]	20.42	0.5801	0.6134
Ours + SRFBN-s [11]	20.32	0.5787	0.6225

6. Comparing with Correction_Filter

As shown in Tab. 1, Correction_Filter failed to improve the performance of SRGAN, ESRGAN, EDSR, RCAN, and SRFBN-s. The performance of Correction_Filter greatly relies on the accuracy of degradation estimation. Correction_Filter uses iterative approximation strategies to estimation the degradation kernel of LR image by initializing with bicubic kernel. However, it is difficult to solve complex degradation through this method. While our method coherently combine degradation esti-

mation with correction filter learning into a unified network, greatly improving the accuracy and adaptiveness. Besides, Correction_Filter only considers the spatially invariant degradation, failing to tackle spatially ones. In our method, we estimate degradations in a spatially variant way. The experimental results show that our method is more effective than Correction_Filter.

Table 2. Comparing the consumption of memory and time on a 510× 339 LR image for scale ×4 with Correction_Filter [7]. We only compare the process from LR input to correction.

Methods	Memory (GB)	Runtime (min)
Correction Filter [7]	23.72	7.36
Ours	7.50	2.47

We also compare the consumption of memory and runtime with Correction_Filter in Tab. 2. The memory and runtime is calculated on a 510× 339 LR image for scale ×4. Correction_Filter uses DBPN [6] in their iterative approximation strategies, which requires huge memory consumption (23.72 GB). While our method only need 7.50 GB. Besides, Correction_Filter needs multiple iterations to estimate the degenerate kernel, so it spends a lot of time. In contrast, our method can achieve better performance and spend less time.

7. Limitations

The first limitation of our method is that we need LR image with large enough size to learn correction filters. If the size of LR image is too small, the generated training data will be too small, which will affect the performance. In addition, if the size of LR image is smaller than the patch size in training settings, our method will not work unless the settings are modified. This is also a limitation for most existing methods [1, 18, 5].

The second limitation is that the current version of our method does not consider the corruptions introduced by noise. Although we introduced additional traditional method CBM3D [3] to solve this problem, there are still some problems with this method, which will also influence the performance of SR. In future work, we may add an extra learnable denoise module in our method.

References

[1] Sefi Bell-Kligler, Assaf Shocher, and Michal Irani. Blind super-resolution kernel estimation using an internal-gan. *Advances in Neural Information Processing Systems*, 32, 2019. 4

[2] Paul Leo Butzer. A survey of the whittaker-shannon sampling theorem and some of its extensions. In *Journal of Mathematical Research and Exposition*, volume 3, 1983. 1

[3] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. Color image denoising via sparse 3d collaborative filtering with grouping constraint in luminance-chrominance space. In *IEEE International Conference on Image Processing*, volume 1, pages I–313. IEEE, 2007. 4

[4] Yonina C Eldar. *Sampling theory: Beyond bandlimited systems*. Cambridge University Press, 2015. 1

[5] Mohammad Emad, Maurice Peemen, and Henk Corporaal. Dualsr: Zero-shot dual learning for real-world super-resolution. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 1630–1639, 2021. 4

[6] Muhammad Haris, Gregory Shakhnarovich, and Norimichi Ukita. Deep back-projection networks for super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1664–1673, 2018. 4

[7] Shady Abu Hussein, Tom Tirer, and Raja Giryes. Correction filter for single image super-resolution: Robustifying off-the-shelf deep super-resolvers. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1428–1437, 2020. 1, 3, 4

[8] Vladimir Aleksandrovich Kotelnikov. On the transmission capacity of the 'ether'and of cables in electrical communications. In *Proceedings of the first All-Union Conference on the technological reconstruction of the communications sector and the development of low-current engineering. Moscow*, 1933. 1

[9] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4681–4690, 2017. 3

[10] Wenbo Li, Kun Zhou, Lu Qi, Nianjuan Jiang, Jiangbo Lu, and Jiaya Jia. Lapar: Linearly-assembled pixel-adaptive regression network for single image super-resolution and beyond. *Advances in Neural Information Processing Systems*, 33:20343–20355, 2020. 1, 2

[11] Zhen Li, Jinglei Yang, Zheng Liu, Xiaomin Yang, Gwanggil Jeon, and Wei Wu. Feedback network for image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3867–3876, 2019. 3

[12] Jingyun Liang, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Mutual affine network for spatially variant kernel estimation in blind image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4096–4105, 2021. 2, 3

[13] Jingyun Liang, Kai Zhang, Shuhang Gu, Luc Van Gool, and Radu Timofte. Flow-based kernel prior with application to blind super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 10601–10610, 2021. 2

[14] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 136–144, 2017. 3

[15] Cheng Ma, Yongming Rao, Yean Cheng, Ce Chen, Jiwen Lu, and Jie Zhou. Structure-preserving super resolution with gradient guidance. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 7769–7778, 2020. 3

[16] Harry Nyquist. Certain topics in telegraph transmission theory. *Transactions of the American Institute of Electrical Engineers*, 47(2):617–644, 1928. 1

[17] Claude E Shannon. Communication in the presence of noise. *Proceedings of the IRE*, 37(1):10–21, 1949. 1

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

[19] Domen Tabernik, Matej Kristan, and Aleš Leonardis. Spatially-adaptive filter units for deep neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 9388–9396, 2018. 1

[20] Domen Tabernik, Matej Kristan, Jeremy L Wyatt, and Aleš Leonardis. Towards deep compositional networks. In *International Conference on Pattern Recognition*, pages 3470–3475. IEEE, 2016. 1

[21] Radu Timofte, Eirikur Agustsson, Luc Van Gool, Ming-Hsuan Yang, and Lei Zhang. Ntire 2017 challenge on single image super-resolution: Methods and results. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 114–125, 2017. 2

[22] Radu Timofte, Shuhang Gu, Jiqing Wu, and Luc Van Gool. Ntire 2018 challenge on single image super-resolution: Methods and results. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 852–863, 2018. 2, 3

[23] Michael Unser. Sampling-50 years after shannon. *Proceedings of the IEEE*, 88(4):569–587, 2000. 1

[24] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European conference on computer vision (ECCV) workshops*, pages 0–0, 2018. 3

[25] ET Whitaker. On the functions which are represented by the expansion of interpolating theory. In *Proc. Roy. Soc. Edinburgh*, volume 35, pages 181–194, 1915. 1

[26] 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*, pages 286–301, 2018. 3

[27] Daniel Zoran and Yair Weiss. From learning models of natural image patches to whole image restoration. In *International Conference on Computer Vision*, pages 479–486. IEEE, 2011. 1