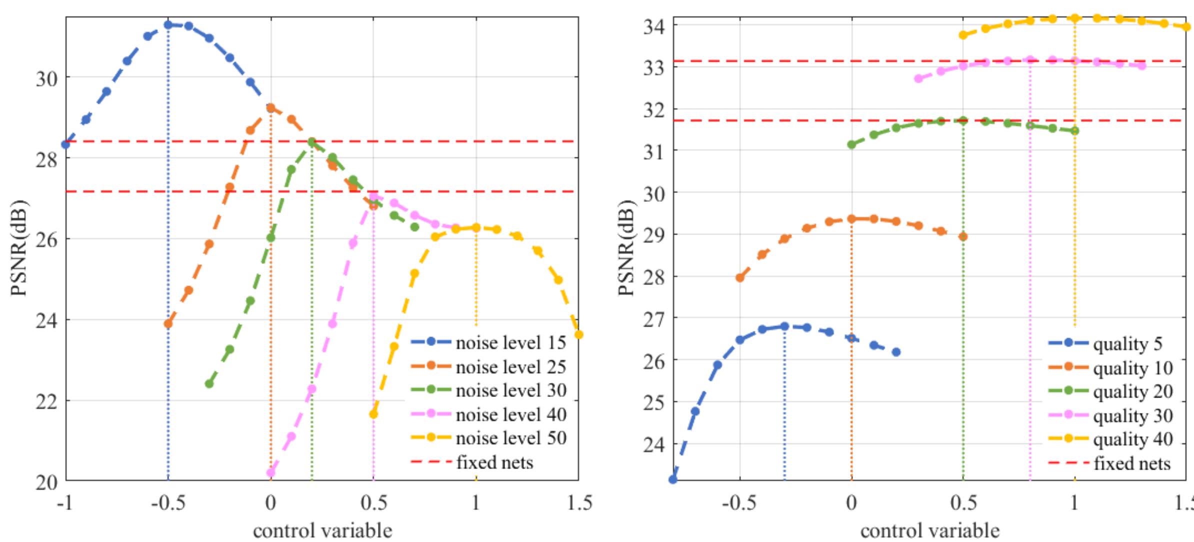
CFSNet: Toward a Controllable Feature Space for Image Restoration
Supplementary Material

We first provide more model analyses about the proposed coupling module in Sec. 2. Then we extend our model to two different image restoration tasks in Sec. 1.1 and Sec. 2.1, *i.e.*, color image denoising and general image super-resolution. Furthermore, additional results can be found in Sec. 2.2.

1. More Examples and Analyses



(a) image denoising (b) JPEG deblocking

Figure 1. PSNR-α_{in} curve for five noise levels (*i.e.*, 15, 25, 30, 40, and 50) on the (gray) BSD68 dataset and PSNR-α_{in} curve for five quality factors (*i.e.*, 5, 10, 20, 30, and 40) on the LIVE1 dataset. Note that our CFSNet is only trained on two endpoints (*e.g.*, noise level 25 and 50), but by adjusting the control variable α_{in}, our CFSNet can still achieve similar performance compared with the fixed networks (red dotted line) at two intermediate working points (*e.g.*, noise level 30 and 40).

Control mechanism analyses In the ablation study of the main paper, we exhibit the reconstruction results of our CFSNet with different input control variables. Here, we provide more details on handling tasks with different degradation levels. To be more specific, we change control variable α_{in} with an interval of 0.1 for each degradation level. Besides, we use the main branch of our CFSNet as the fixed network, which is trained for each specific intermediate degradation level with corresponding data (red dotted line in the Fig.1). As shown in Fig.1, the observations can be summarized as: 1) Whether it is an image denoising task or a JPEG image deblocking task, we can always find the optimal control variable to obtain the result of minimal reconstruction distortion. On the other hand, the optimal α_{in} of different degradation levels is continuous. Taking image denoising as an example, the larger the noise level, the larger the optimal control variable. The noise level 15 even corresponds to a negative optimal control variable. This is because when we train our framework, we limit the two endpoint noise levels (noise level 25 and 50) to two endpoint control variables (0 and 1), respectively. Therefore, our framework establishes an implicit correspondence between control variables and degradation levels. 2) Compared with the fixed networks (red dotted line), our CFSNet can achieve similar performance at two invisible intermediate working points

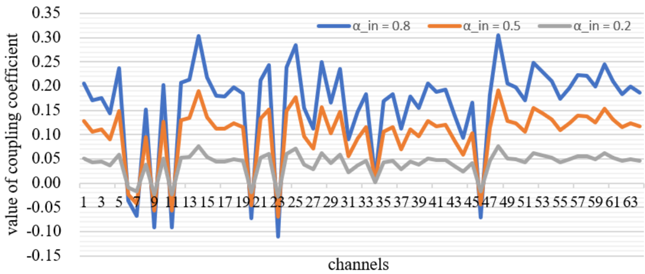


Figure 2. The coupling coefficient curve of the second module in image denoising.

methods	Denoising	DeJPEG
M10T5-top	27.07($\alpha_{in} = 0.6$)	33.10($\alpha_{in} = 0.8$)
M10T5-last	26.85($\alpha_{in} = 0.5$)	33.10($\alpha_{in} = 0.9$)
M10T10	27.12($\alpha_{in} = 0.5$)	33.17($\alpha_{in} = 0.8$)
M20T20	27.19($\alpha_{in} = 0.5$)	33.24($\alpha_{in} = 0.8$)
M30T30	27.28($\alpha_{in} = 0.5$)	33.26($\alpha_{in} = 0.8$)

Table 1. Model analysis. The average PSNR(dB) for unseen noise level $\sigma 40$ on (gray) BSD68 and unseen quality factor $q30$ on LIVE1. M and T denote the number of main blocks and the number of tuning blocks, respectively.

of both restoration tasks, which indicates that our CFSNet can preferably emulate unknown working points. In addition, we can empirically provide a reference value of the control variable for each degradation level to make our framework more user-friendly.

The number of basic modules Here, we further evaluate the effect of the number of main blocks and tuning blocks. From the experimental results (Tab. 1), we observe that: 1) The more coupling modules, the better the reconstruction results. 2) Adding coupling modules only in the first 5 (M10T5-top) or last 5 (M10T5-last) layers is inferior to dense stacking mode (M10T10) in image denoising and JPEG image deblocking.

The coupling coefficients learned First, since the manifold flatness of each layer of the network is different, the coefficients vary by channels and modules. This verifies that the coupling coefficients of CFSNet are learned adaptively from the training process. In addition, as shown in Fig. 2, the variation of coupling coefficients with different α_{in} is consistent, because each α_{in} corresponds to an intermediate latent representation.

1.1. Color Image Denoising

In the main paper, we show the superiority of our CFSNet in gray image denoising task. Here, we further extend our model to color image denoising. We generate noisy color images by adding AWGN noise to clean RGB images with different noise levels $\sigma = 15, 25, 30, 40,$ and 50. We evaluate our CFSNet using the Kodak24 [3] dataset. For the specific implementation, we use the same model settings as the gray image denoising task.

We show the color image denoising visual results in Fig. 3. We can see that: 1) Our CFSNet can play a role in controlling the trade-off between noise reduction and detail preservation. In particular, compared with DnCNN-B [7] and FFDNet [9], our CFSNet can even achieve similar visual effects when dealing with the noise level $\sigma = 15$ which beyond the preset range ($[25, 50]$). This further verifies the rationality of our theoretical analysis. 2) The optimal visual quality is image specific or scene specific. For example, "flower" (3rd row) enjoys more realistic texture when $\alpha_{in} = 0.4$, while "leaf" (4th row) enjoys smoother artifact-free result when $\alpha_{in} = 1.0$. However, the fixed method (*e.g.*, DnCNN-B [7]) can not adequately meet specific needs. Therefore, it is necessary to adjust the reconstruction results according to specific goals.

2. Model Analyses

2.1. General Image Super-resolution

In the main paper, we show the perception-distortion trade-off of image super-resolution modeled by a simple bicubic downsampling operation. Here, in order to further demonstrate the flexibility of our CFSNet, we extend our framework to a more challenging degradation model. To be specific, when we train the main branch in Step 1, we still adopt bicubic downsampling as the degradation setting. However, when we train the tuning branch in Step 2, we first blur the HR image by Gaussian kernel of size 17×17 with standard deviation 2.6, then we bicubic downsample it with scale factor 3 to produce an LR image. For testing, we use three standard benchmark datasets: Set5 [1], B100 [6], Urban100 [4], and we follow [8] to use a popular 7×7 Gaussian kernel with width 1.6 which never appear in training process. Besides, both branches of our framework are trained based on the same MAE loss. In more details, we adopt a small model that contains 10 main blocks and 10 tuning blocks (i.e., $M = 10$) for a fair comparison.

Fig. 4 shows the visual results of the proposed CFSNet. Tab. 2 shows the average PSNR results for SRCNN [2], VDSR [5], IRCNN [8] and our CFSNet. Several observations can be summarized as follows: 1) Our CFSNet can make a nice compromise between blur removal and detail sharpening. Specifically, the large control variable α_{in} leads to over sharpening artifacts. In contrast, incomplete blur elimination results can be observed using a small control variable α_{in}.

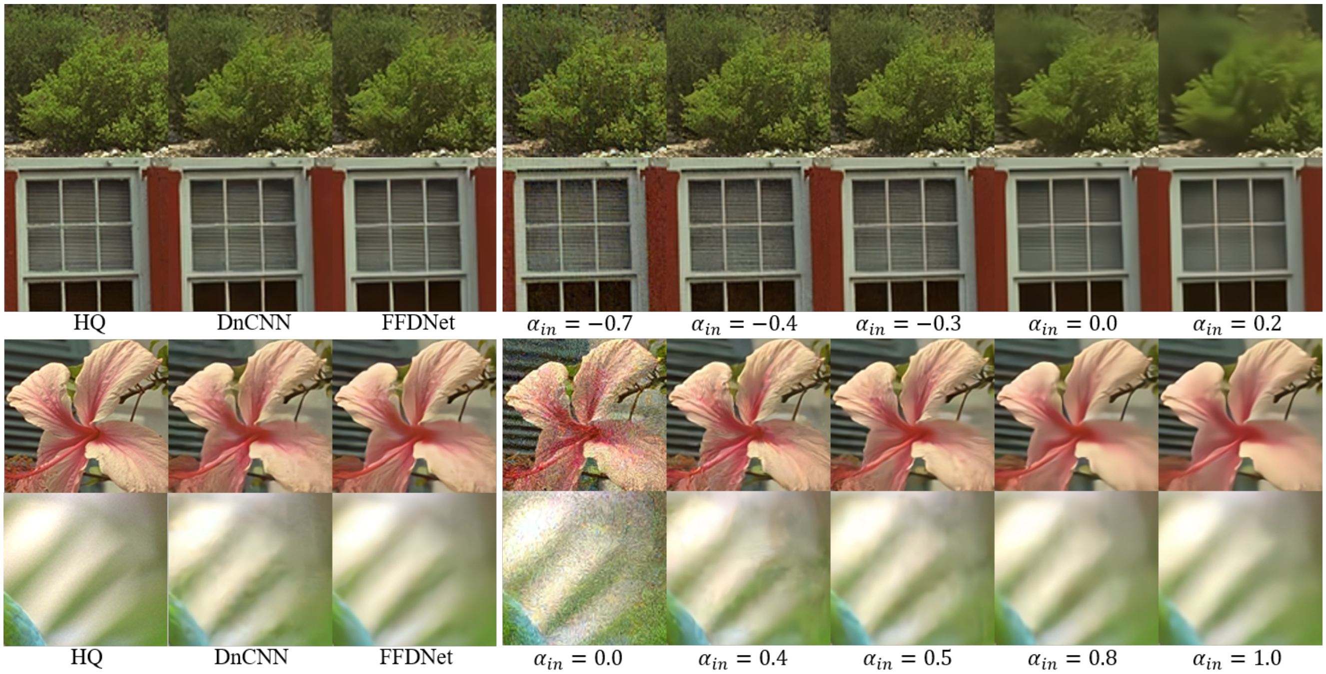


Figure 3. Color image denoising results with unknown noise level $\sigma = 15$ (first two rows) and $\sigma = 40$ (last two rows). $\alpha_{in} = 0.5$ and $\alpha_{in} = -0.3$ correspond to the highest PSNR results, respectively.

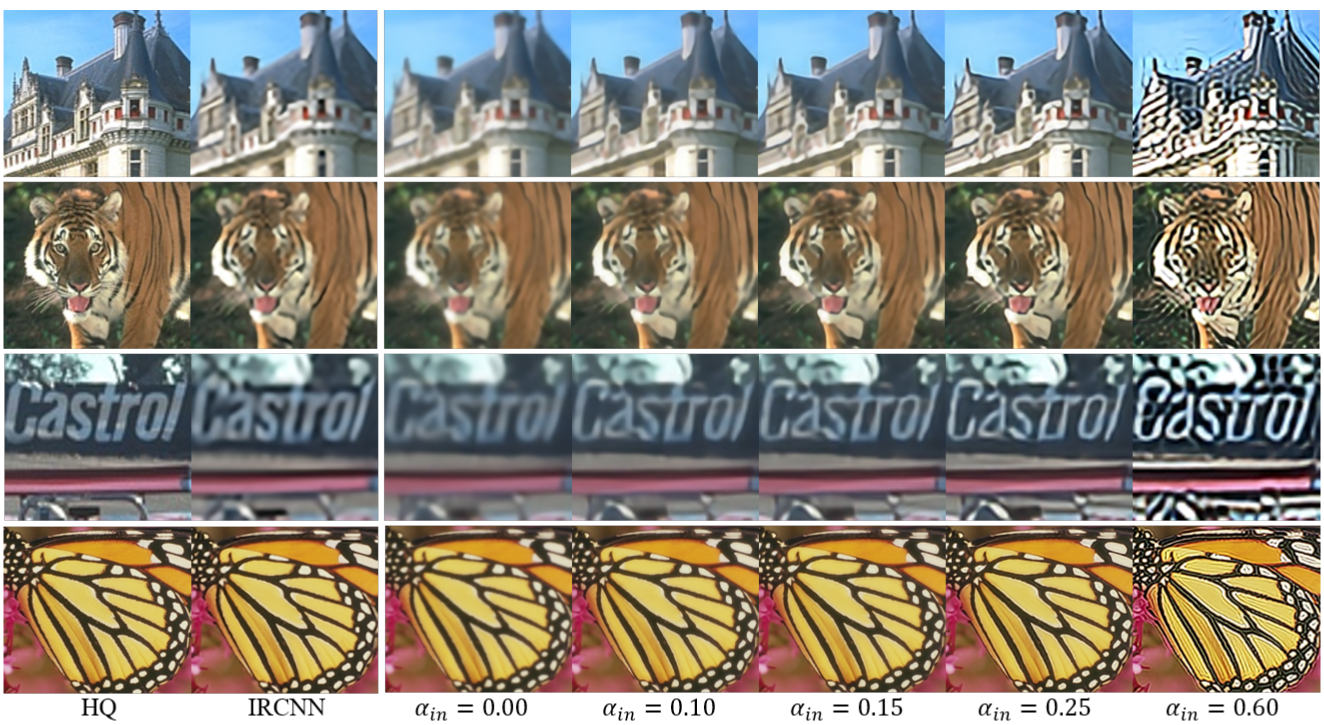


Figure 4. Visual results of single image super-resolution with unseen degradation (the blur kernel is 7×7 Gaussian kernel with standard deviation 1.6, the scale factor is 3). $\alpha_{in} = 0.15$ corresponds to the highest PSNR results.

2) Our CFSNet achieves the best PSNR results on all test datasets. 3) the lowest distortion result ($\alpha_{in} = 0.15$) does not necessarily mean the best visual effects. For example, the result of "tiger" is more visually acceptable when $\alpha_{in} = 0.25$.

Figure 5. Perceptual and distortion balance for $4\times$ image super-resolution.

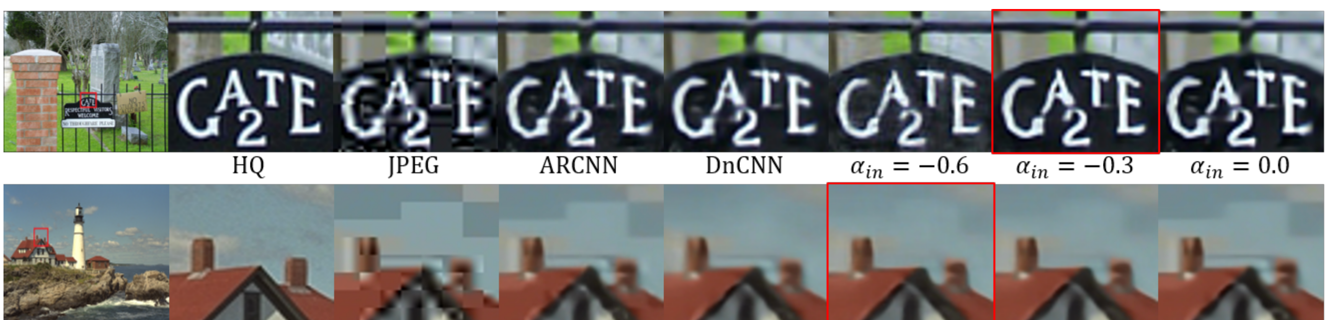


Figure 6. JPEG image artifacts removal results of "cemetry" and "lighthouse2" (LIVE1) with unknown quality factor 5. $\alpha_{in} = -0.3$ corresponds to the highest PSNR results, and the best visual results are marked with red boxes.

Table 2. The average general image super-resolution results of PSNR (dB) on three benchmark datasets. Note that the degradation settings of all the testing images are not included in the training stages of our CFSNet.

Dataset	SRCNN	VDSR	IRCNN	CFSNet
Set5	32.05	33.25	33.38	33.50
B100	28.13	28.57	28.65	28.79
Urban100	25.70	26.61	26.77	27.33

2.2. Additional Results

We present additional perceptual and distortion balance results for $4\times$ image super-resolution in Fig. 5, and additional JPEG image artifacts removal results in Fig. 6.

References

[1] Marco Bevilacqua, Aline Roumy, Christine Guillemot, and Marie Line Alberi-Morel. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. 2012.

[2] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307, 2016.

[3] Rich Franzen. Kodak lossless true color image suite. *source: http://r0k. us/graphics/kodak*, 4, 1999.

[4] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5197–5206, 2015.

[5] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1646–1654, 2016.

[6] David Martin, Charless Fowlkes, Doron Tal, and Jitendra Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *null*, page 416. IEEE, 2001.

[7] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Transactions on Image Processing*, 26(7):3142–3155, 2017.

[8] Kai Zhang, Wangmeng Zuo, Shuhang Gu, and Lei Zhang. Learning deep cnn denoiser prior for image restoration. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3929–3938, 2017.

[9] Kai Zhang, Wangmeng Zuo, and Lei Zhang. Ffdnet: Toward a fast and flexible solution for cnn-based image denoising. *IEEE Transactions on Image Processing*, 27(9):4608–4622, 2018.