ENHANCED IMAGE RESTORATION VIA SUPERVISED TARGET FEATURE TRANSFER

Yuzhao Chen[⋆†] *Tao Dai*[⋆†] *Xi Xiao*[⋆†] *Jian Lu*[‡] *Shu-Tao Xia*[⋆†]

[⋆]Tsinghua Shenzhen International Graduate School, Tsinghua University, Shenzhen, China
[†] Peng Cheng Laboratory, Shenzhen, China
[‡] Shenzhen Key Laboratory of Advanced Machine Learning and Applications, Shenzhen University, China
chen-yz19@mails.tsinghua.edu.cn, daitao.edu@gmail.com, {xiaox,xiast}@sz.tsinghua.edu.cn

ABSTRACT

Deep learning has obtained remarkable success for image restoration. However, most existing deep image restoration models are trained by minimizing the pixel-level reconstruction error between restored images and target images (ground truth), while neglecting the rich information from the intermediate feature layers, thus hindering the representational power of networks. To address this problem, we propose a Supervised Target Feature Transfer (STFT) framework to enhance the power of feature expression of the deep image restoration models. Specifically, we introduce a self-supervised antoencoder-based *target feature extractor* to extract compact feature representation of target images, which serves as supervision signals to train the deep backbone models at the same time. With such feature-level supervised information, deep backbone model can be enhanced by transfer learning of such *target features*. Moreover, we theoretically analyze our STFT training strategies and demonstrate that it imposes learnable prior information on the backbone restoration model. Extensive experiments demonstrate the effectiveness of our proposed framework compared with the state-of-the-art image restoration models.

***Index Terms*—** Transfer Learning, Image Restoration, Target Features, Autoencoder

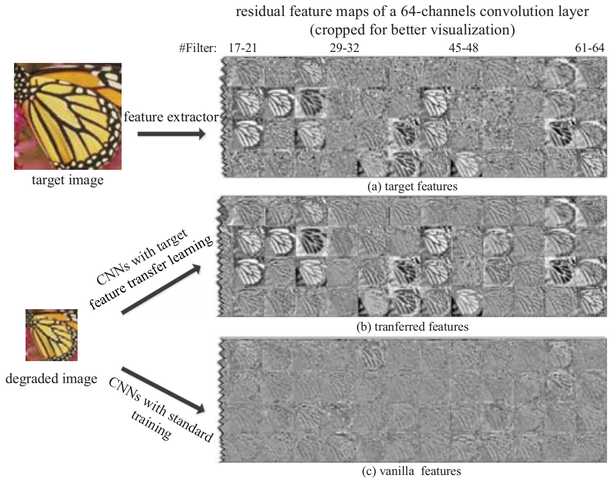

Fig. 1. Feature visualization in intermediate layers. (a): features from target image captured by autoencoder-based feature extractor. (b): features from degrade image with restoration model trained by our STFT method. (c): features from degraded image with restoration model trained by standard training.

1. INTRODUCTION

Image restoration is a seriously ill-posed inverse problem that attempts to reconstruct high-quality images from degraded images distorted by different types of degradations. Depending on the type of degradation, image restoration can be further divided into various tasks, such as image super-resolution, image denoising and JPEG image deblocking.

Recently, deep learning has achieved great success on various image restoration tasks. Most existing image restoration methods focus on the design of network architecture [1, 2, 3, 4], the construction of loss function [5, 6] and the way of training strategy [7, 8]. However, most existing deep restoration methods are trained by minimizing the pixel-level reconstruction error between restored images and target images, while neglecting the rich feature-level information in the intermediate layers from target images. As a result, such training strategy with pixel-level reconstruction error leads to the deficiency of low-frequency and high-frequency information in generated features ((see the bottom row in Fig.1). On the other hand, autoencoder based networks can provide more compact representations of target images (see the first row in Fig.1).

The above observations raise a question whether we can transfer the compact representations from target images into the feature space of existing deep restoration models. Perhaps it is natural to adopt feature distillation methods [9, 10, 11, 12, 13, 14] by transferring the knowledge from a complex teacher model to a lightweight student model. However, these efforts mainly focus on distilling the knowledge from soft labels produced by teacher models, and thus can not directly applicable for image regression (e.g., image restoration) without label information. Thus, how can we transfer the knowledge from a teacher model into existing deep restoration models is still an open problem.

To remedy the issue, we propose a Supervised Target Feature

Corresponding author: Tao Dai

This work is supported in part by the National Natural Science Foundation of Guangdong Province (2018A030313422), the National Natural Science Foundation of China (61771273, 61972219, 61773229), Guangdong Basic and Applied Basic Research Foundation (2019A151511034), the National Key Research and Development Program of China (2018YFB1800600, 2018YFB1800204), the China Postdoctoral Science Foundation (2019M660645), the R&D Program of Shenzhen (JCYJ20180508152204044, JCYJ20190813174403598, JCYJ20190813165003837) and the project "PCL Future Greater-Bay Area Network Facilities for Large-scale Experiments and Applications (LZC0019)".

Authorized licensed use limited to: Tsinghua University. Downloaded on April 20,2022 at 07:07:56 UTC from IEEE Xplore. Restrictions apply.

Transfer (STFT) framework to enhance the power of feature expression for deep image restoration models. Our STFT framework mainly consists of a deep image restoration backbone model and a self-supervised autoencoder based feature extractor. Specifically, we design an autoencoder network with the similar (or a relatively lightweight one) structure as the given image restoration backbone model to extract features from target images. To transfer the rich knowledge from autoencoder network to the image restoration backbone model, we train the network by minimizing feature-level reconstruction error in the intermediate feature layers. Since the target features from autoencoder have more compact representation for target images, such training learning actually would distill the diverse knowledge to the restoration model, and enhance the power of feature expression. As illustrated in Fig.1, the trained model with pixel-level reconstruction error contains less semantic information, while our STFT training with feature-level reconstruction error exhibits richer and more diverse textures.

In summary, the main contributions of our work can be summarized as follows:

- We propose a Supervised Target Feature Transfer (STFT) framework by enhancing the power of feature expression of deep image restoration models. To this end, we propose a self-supervise autoencoder-based feature extractor, which extracts compact representation as supervision signals.
- Our STFT enhances the current deep restoration models by minimizing the pixel-level reconstruction error in the output layer and feature-level reconstruction error in the intermediate layers. Experiments on different image restoration tasks demonstrate the effectiveness of our method compared with other state-of-the-art methods.

2. THE PROPOSED METHOD

2.1. Observations

To find the reason why the feature distillation methods fail in working on the image restoration task, we visualized the feature maps extracted by the trained classification network and image denoising network respectively and perform a comparison. The examples of experimental results are provided in the supplement materials. In short, the observation gives a little explanation for the phenomenon: these low-level image processing tasks require the networks to capture representations with highly fine textures and structures due to the property of regression problem, which makes them to be very sensitive to the specific value of the intermediate generated features. Existing distillation methods, however, are not capable of providing sufficient precise knowledge to the restoration model.

To this end, we believe that there are two key factors that will lead to the success of knowledge transfer learning for image restoration models: accurate target features that work as supervision signals, and proper features transfer strategy. These two concerns are involved in our proposed method described below.

2.2. Supervised Target Feature Transfer

The purpose of STFT can be formulated as follows. Suppose that a restoration model $\mathcal{M}_\theta$ parameterized by parameter θ learns the mapping that restores the distorted image I_{dis} to the target image (Ground Truth) I_{GT}:

$$f_\theta : I_{dis} \to I_{GT} \tag{1}$$

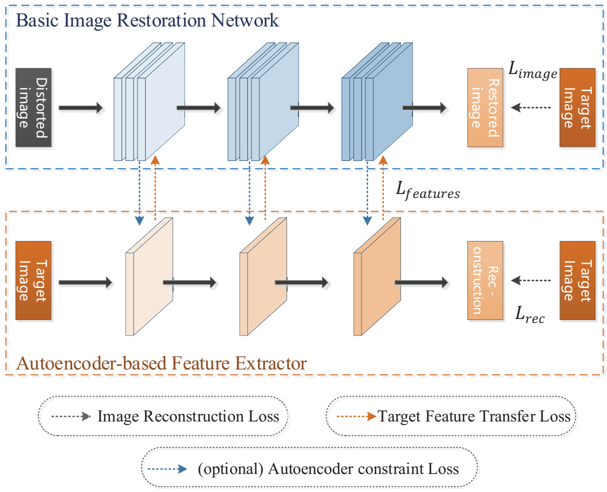

Fig. 2. The proposed STFT framework. The restoration network is not only trained for mapping distorted images to restored fimages, but also predict the target features in advance before reconstructing images. The blur dotted arrow indicates the loss that imposes a constraint on the autoencoder, as illustrated in section 2.4.

Our method further requires the deep network to model the mapping that extracts I_{dis} to the feature representation of its target image in the embedding space $\mathcal{F}_{GT}$:

$$g_{\theta_1} : I_{dis} \to \mathcal{F}_{GT} \tag{2}$$

where $\theta_1 \in \theta$, represents the parameters of the layers that learn to transfer target feature in the network. To supervise the learning of this additional mapping, we train a self-supervised feature extractor $\mathcal{M}_{ext}$ based on autoencoder. It has been show that autoencoder is powerful tool for extracting meaningful features as well as reconstructing the input data. From this perspective, a well-trained feature extractor is supposed to learn a bijection mapping:

$$h : I_{GT} \to \mathcal{F}_{GT} \text{ and } h^{-1} : \mathcal{F}_{GT} \to I_{GT} \tag{3}$$

which represents the encoding and decoding stages respectively. The $\mathcal{M}_{ext}$ is easy to train by optimizing the reconstruction loss:

$$L_{rec} = \|\mathcal{M}_{ext}(I_{GT}) - I_{GT}\|_1 \tag{4}$$

Intuitively, to prevent from information missing, the STFT method requires that the dimension of the target feature to match the corresponding one that receives supervision. According to our experiments, it would cause performance degradation of the model with any transformation of the features. Hence our feature extractor does not downsample the input images like the traditional autoencoder model [15, 16], but maintains the basically same architecture as the given image restoration network. Interestingly, we have empirically found that our STFT method is not sensitive to the size of the feature extractor (or the number of parameters). Therefore, a simply yet effective strategy to reduce the additional training cost is to apply a relatively lightweight autoencoder network by cutting off those repeatedly stacked layers in the given restoration networks while reserving the ones that selected to extract target features. As we will see, the target features captured by such autoencoder are qualified to work well as the ideal supervision signals of learning mapping g_{θ_1}.

An overall framework of our STFT is depicted in Fig.2. Different from traditional image restoration methods that only reduce the

Authorized licensed use limited to: Tsinghua University. Downloaded on April 20,2022 at 07:07:56 UTC from IEEE Xplore. Restrictions apply.

(a) Image super-resolution results. The best average PSNR(dB) on benchmark in 5×10^5 iterations.

	Set5		Set14		B100		Urban100	
	L_{pixel}	$+L_{feat}$	L_{pixel}	$+L_{feat}$	L_{pixel}	$+L_{feat}$	L_{pixel}	$+L_{feat}$
EDSR	27.10	**27.18**	25.31	**25.41**	25.09	**25.16**	22.85	**22.94**
RCAN	27.16	**27.27**	25.36	**25.47**	25.14	**25.19**	22.98	**23.05**

(b) Image dehazing results.

	NYU2		#params
	L_{pixel}	$+L_{feat}$	
AOD	20.088	**20.238**	$< 10K$
PPDN	29.707	**29.923**	$> 30M$

(c) Image denoising results. The best average PSNR(dB) on benchmark in 5×10^5 iterations.

DnCNN	Kodak24		BSD68		McMaster		FFDNet	Kodak24		BSD68		McMaster	
	L_{pixel}	$+L_{feat}$	L_{pixel}	$+L_{feat}$	L_{pixel}	$+L_{feat}$		L_{pixel}	$+L_{feat}$	L_{pixel}	$+L_{feat}$	L_{pixel}	$+L_{feat}$
$\sigma = 25$	33.50	**33.66**	32.32	**32.47**	33.83	**34.06**	$\sigma = 25$	33.32	**33.46**	32.07	**32.18**	33.53	**33.67**
$\sigma = 50$	29.35	**29.44**	28.07	**28.14**	28.64	**28.79**	$\sigma = 50$	29.72	**29.80**	28.37	**28.51**	28.99	**29.07**

Table 1. Performance comparison between standard training (L_{pixel}) and our target feature transfer learning method ($+L_{feat}$).

reconstruction error between I_{dis} and I_{GT} , we consider the feature difference between the generated features $\mathcal{F}_{gen}$ and $\mathcal{F}_{GT}$ to force the model to focus on what information is essential for the reconstruction. This multi-task can be learned by a total objective function:

$$L_{image}(\mathcal{M}_\theta(I_{dis}), I_{GT}) + \lambda_1 \sum_{d=1}^{D} L_{features}(\mathcal{F}^d_{gen}, \mathcal{F}^d_{GT}) \quad (5)$$

where D is the number of layers that directly supervised by the corresponding target features, λ_1 is a hyperparameter. Notice that $\mathcal{F}_{gen} = \mathcal{M}_{\theta_1}(I_{dis})$. As usual, we set the image loss as the pixel-wise loss to pursue as high PSNR as possible. After considering the observation in the above section, we believe that the normal pixel-wise loss is also the most suitable distance function to transfer the feature knowledge. We empirically investigate various distance metrics in section 3.3 to verify our choice.

2.3. Theoretical Analysis

The STFT actually works by providing a good prior assumption to the restoration model. From the perspective of Bayesian statistic, any other loss item except the pixel-wise image loss is the given prior to the restoration model. This can be seen by the following fact. Given model $\mathcal{M}$, we can find the MAP parameter θ of it:

$$P(\theta \mid x, y; \mathcal{M}) \propto P(x \mid y, \theta; \mathcal{M}) P(\theta \mid y; \mathcal{M}) \quad (6)$$

Recall that $y = Hx + v$, therefore $H^\dagger y = x + H^\dagger v$, where $H\dagger$ is the pseudo inverse matrix. Assume that $H^\dagger v$ is the gaussian white noise, and we use the model to fit the restoration mapping, then the likelihood of parameter θ is

$$P(x \mid y, \theta; \mathcal{M}) \propto exp(-\sum_i \|x_i - \mathcal{M}_\theta(y_i)\|_2) \quad (7)$$

Denote the prior $P(\theta|y; \mathcal{M}) = exp(-\lambda_1 H(\theta, y))$, then the eq.(6) can be transformed into:

$$P(\theta \mid x, y; \mathcal{M}) \propto exp(-C(\theta, x, y, \mathcal{M})) \quad (8)$$

where $C(\theta, x, y, \mathcal{M}) = \sum_i \|x_i - \mathcal{M}_\theta(y_i)\|_2 + \lambda_1 H(\theta, y)$, the former item is the regular pixel-wise image loss, and the latter one is the prior item. In the STFT method, the prior knowledge is learned by conducting an auxiliary task (learn the mapping of (2)) and thus more flexible than a fixed given prior assumption:

$$H(\theta, y) = \sum_i \|\mathcal{F}_i - \mathcal{M}_{\theta_1}(y_i)\|_p \quad (9)$$

We can expect it will provide correct inductive bias to the model. This is natural since the auxiliary task is a part of the main task (learn the mapping (1)):

$$f_\theta(I_{dis}) = h^{-1}(g_{\theta_1}(I_{dis})) \quad (10)$$

2.4. Training Strategies

We further propose an effective training strategies by joint optimizing the two parts of STFT framework. During the training of $\mathcal{M}_\theta$, we can train $\mathcal{M}_{ext}$ from scratch by optimizing

$$L_{rec} + \lambda_2 \sum_{d=1}^{D} L_{features}(\mathcal{F}^d_{GT}, \mathcal{F}^d_{gen}) \quad (11)$$

where λ_2 is a weight that controls the strength of restrain on the autoencoder, and empirically set to 0.1 for the balance of the restriction strength and the accuracy of the target features. Compare with the usual setting that a teacher model is well-trained in advance and fixed, we adopt the generated features of the image restoration model to work as supervision signals and use same feature loss to guide the training of the feature extractor and thus compress its search space of target representation, which ease the restoration model to draw knowledge from it. The evaluation that exemplify the effectiveness of this strategy is provided in the supplements. We notice that this training strategy is a bit like the mutual learning method [17], in which the two (or more) identical networks are trained for solving the same task based on the mutual teaching of each other. We further investigate these methods in the ablation study section 3.3.

3. EXPERIMENTAL RESULTS

3.1. Implement Details

For the task of image dehazing, we use 25000 pairs of images of the training set of NYU2 [18] as training data. As for the remaining tasks, 800 training images of DIV2K [19] plus 2500 training images of Flickr2K [20] are adopted as the training data. Data augmentations of random rotation are performed in each experiment. In each training batch, 16 patches with the size of 96×96 (24×24 for the task of $\times 4$ super-resolution) are randomly extracted as inputs. The basic models and feature extractors are trained by ADAM optimizer [21] with $\beta_1 = 0.9$, $\beta_2 = 0.999$ and $\epsilon = 10^{-8}$. The initial learning rates are set according the corresponding papers of the chosen baseline model, and decay by a factor of 2 every 1.5×10^5 iterations. For all of the experiments, we adopt L_1 loss as the image reconstruction loss, and L_2 loss as the feature transfer loss.

3.2. Evaluation study

Image Super-resolution. For image super-resolution task, we use the manually generated $\times 4$ downscale images with additional unknown noise and blur as degraded input data and four standard benchmark datasets [22, 23, 24] for testing. We adopt EDSR [1] and RCAN [2] as the basic model. In order to shorten the training time, we set the baseline EDSR model with 16 residual blocks with

Authorized licensed use limited to: Tsinghua University. Downloaded on April 20,2022 at 07:07:56 UTC from IEEE Xplore. Restrictions apply.

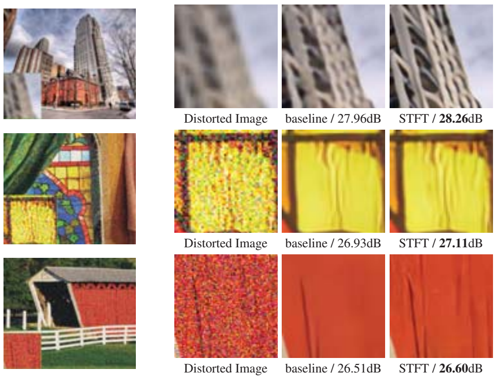

Fig. 3. Visual comparison between Distorted Image and Restored Image w/wo STFT. The first row gives an example of super-resolution and the last two rows give the results of denoising.

ARCNN	Live1		BSD500 val.	
	L_{pixel}	$+L_{feat}$	L_{pixel}	$+L_{feat}$
$q = 10$	29.32	**29.41**	28.95	**29.02**
$q = 30$	32.80	**32.91**	32.24	**32.32**

Table 2. JPEG artifacts removal results. The best average PSNR(dB) on Live1 and the validation set of BSD500 in 2×10^5 iterations.

64 filters in each convolutional layer, and the baseline RCAN model with 4 residual groups and 5 residual blocks for each group. The hyperparameters of λ_1 are set to 1 and 50, respectively. The Fig.3 displays the visual comparsion between our results and the baseline methods, and the Tab.1(a) gives the quantitative comparisons . We can see that the STFT method boost the baseline of about 0.1dB in PSNR, while at the same time achieving higher visual quality.

Image Denoising. The degraded images are created by using the target images with gaussian noise of level 25 and 50 respectively. We use the benchmark datasets of [25, 26] as test data. We adopt DnCNN [27] and FFDNet [28] as baseline model and follow their settings. The hyperparameters λ_1 are both set to 1. We also provide visual comparison in Fig.3, it can be observed that STFT method generate restored images with sharper edges and finer texture. The quantitative evaluation is demonstrated in Tab.1(c), which shows that the STFT significantly enhances the basic denoising model.

Image Dehazing. For image dehazing, we use the network of AOD [29] and PPDN [30] respectively. Apart from the training data, We use the remain 2256 pairs training set of NYU2 as testset. The hyperparameters of λ_1 are set to 100 and 1 for training AOD and PPDN respectively. Tab.1(b) gives the results. Notably, while the AOD is a extremely lightweight model (#parameters $<$ 10K) and the PPDN is a large network (#parameters $>$ 30M), the consistent performance gain indicates that our method is capable of adapting to arbitrary restoration model.

JPEG Artifacts Removal. For JPEG artifacts removal, we evaluate our method on the classic ARCNN [31]. Live1 [32] and validation set of BSD500 is used for testing. The compressed images are generated by the target images with compression quality of 10 and 30 respectivly. Comparison results are provided in Tab.2, which shows that the models that trained by STFT achieve consistent and considerable improvement.

3.3. Ablation study

Feature Transfer Loss Distance. Based on the discovery and analysis in the previous section, we consider the pixel-wise loss to be the most effective loss distance for target feature transfer learning. In

		Kodak24	BSD68	McMaster
Baseline ($DnCNN_{17}$)		33.50	32.32	33.83
Target Feature Transfer		**33.66**	**32.46**	**34.06**
Mutual Learning		33.48	32.31	33.84
Feature Distillation	$DnCNN_{29}$	33.68	32.45	34.07
	$DnCNN_{17}$	33.50	32.30	33.90

Table 3. Performance comparison between three kinds of feature transfer learning methods.

this section, we further explore several loss metrics that have been used in other vision tasks, such as the focal loss used in [33], feature similarity based on inner product and KL-divergency in [34], max mean discrepancy (MMD) with polynomial kernel in [35]. We investigate six loss distances as feature transfer loss in our STFT framework to perform super-resolution based on RCAN. The results are shown on Fig.4 We can see that the L_2 loss and L_1 loss obtain the best performance gain while those loss distances proposed in the existing distillation methods fail in dealing with the feature transfer learning in these low-level regression tasks. Besides, the Focal loss, which is also based on pixel-wise loss, can also achieve a weaker performance gain. These verify the rationality of using pixel-wise loss to transfer the restoration based target features.

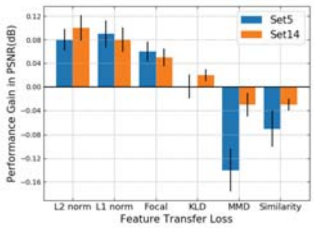

Fig. 4. Investigations of different feature transfer loss. We conduct three repeated experiment for evaluation of each loss.

Role of the Autoencoder. In the framework of STFT, we use an autoencoder to work as target feature extractor. There exists two other options: 1) use a restoration network with the same architecture as given baseline model, like the mutual learning [17] , 2) use a more complex and powerful teacher network. To further analyze the role of autoencoder, we conduct an ablation study and the results are shown in Tab.3. We set the normally trained denoising network DnCNN with 17 convolutional layers ($DnCNN_{17}$) as baseline. For the test of normal feature distillation, we first train a teacher model, DnCNN with 29 layers ($DnCNN_{29}$), which can achieve performance close to our approach, and then distill its features to the student $DnCNN_{17}$, the result is given in the last row. Based on the results, we can confirm that the autoencoder is the key factor to extract accurate target feature for boosting the restoration model, while the other two methods are invalid for providing precise knowledge to the image restoration model.

4. CONCLUSION

We present a novel training framework for enhancing the power of feature representation of existing image restoration models. The core idea of our method is to adopt an autoencoder-based feature extractor to extract target features, which contain sufficient accurate information to perfectly restore the degraded images. We further analyze the training strategies for transferring target features in the proposed STFT framework. Extensive experiments verify the effectiveness of our method and achieves remarkable performance gain on the restoration models.

Authorized licensed use limited to: Tsinghua University. Downloaded on April 20,2022 at 07:07:56 UTC from IEEE Xplore. Restrictions apply.

5. REFERENCES

[1] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee, "Enhanced deep residual networks for single image super-resolution," in *CVPR'w*, 2017.

[2] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu, "Image super-resolution using very deep residual channel attention networks," in *ECCV*, 2018.

[3] Tao Dai, Jianrui Cai, Yongbing Zhang, Shu-Tao Xia, and Lei Zhang, "Second-order attention network for single image super-resolution," in *CVPR*, 2019.

[4] Xiao-Jiao Mao, Chunhua Shen, and Yu-Bin Yang, "Image restoration using convolutional auto-encoders with symmetric skip connections," *arXiv preprint arXiv:1606.08921*, 2016.

[5] Roey Mechrez, Itamar Talmi, Firas Shama, and Lihi Zelnik-Manor, "Maintaining natural image statistics with the contextual loss," in *ACCV*, 2018.

[6] Justin Johnson, Alexandre Alahi, and Li Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *ECCV*, 2016.

[7] 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 *ECCV*, 2018.

[8] Jingwen Chen, Jiawei Chen, Hongyang Chao, and Ming Yang, "Image blind denoising with generative adversarial network based noise modeling," in *CVPR*, 2018.

[9] Adriana Romero, Nicolas Ballas, Samira Ebrahimi Kahou, Antoine Chassang, Carlo Gatta, and Yoshua Bengio, "Fitnets: Hints for thin deep nets," *arXiv preprint arXiv:1412.6550*, 2014.

[10] Sergey Zagoruyko and Nikos Komodakis, "Paying more attention to attention: Improving the performance of convolutional neural networks via attention transfer," *arXiv preprint arXiv:1612.03928*, 2016.

[11] Junho Yim, Donggyu Joo, Jihoon Bae, and Junmo Kim, "A gift from knowledge distillation: Fast optimization, network minimization and transfer learning," in *CVPR*, 2017.

[12] Jangho Kim, SeongUk Park, and Nojun Kwak, "Paraphrasing complex network: Network compression via factor transfer," in *NeurIPS*, 2018.

[13] Byeongho Heo, Minsik Lee, Sangdoo Yun, and Jin Young Choi, "Knowledge transfer via distillation of activation boundaries formed by hidden neurons," in *AAAI*, 2019.

[14] Byeongho Heo, Jeesoo Kim, Sangdoo Yun, Hyojin Park, Nojun Kwak, and Jin Young Choi, "A comprehensive overhaul of feature distillation," *arXiv preprint arXiv:1904.01866*, 2019.

[15] Hugo Larochelle, Dumitru Erhan, Aaron Courville, James Bergstra, and Yoshua Bengio, "An empirical evaluation of deep architectures on problems with many factors of variation," in *ICML*, 2007.

[16] Jonathan Masci, Ueli Meier, Dan Cireşan, and Jürgen Schmidhuber, "Stacked convolutional auto-encoders for hierarchical feature extraction," in *ICANN*, 2011.

[17] Ying Zhang, Tao Xiang, Timothy M Hospedales, and Huchuan Lu, "Deep mutual learning," in *CVPR*, 2018.

[18] Pushmeet Kohli Nathan Silberman, Derek Hoiem and Rob Fergus, "Indoor segmentation and support inference from rgbd images," in *ECCV*, 2012.

[19] Eirikur Agustsson and Radu Timofte, "Ntire 2017 challenge on single image super-resolution: Dataset and study," in *CVPR'w*, July 2017.

[20] 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 *CVPR'w*, 2017.

[21] Diederik P Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

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

[23] Roman Zeyde, Michael Elad, and Matan Protter, "On single image scale-up using sparse-representations," in *International conference on curves and surfaces*, 2010.

[24] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja, "Single image super-resolution from transformed self-exemplars," in *CVPR*, 2015.

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

[26] Lei Zhang, Xiaolin Wu, Antoni Buades, and Xin Li, "Color demosaicking by local directional interpolation and nonlocal adaptive thresholding," *Journal of Electronic imaging*, vol. 20, no. 2, pp. 023016, 2011.

[27] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE TIP*, vol. 26, no. 7, pp. 3142–3155, 2017.

[28] Kai Zhang, Wangmeng Zuo, and Lei Zhang, "Ffdnet: Toward a fast and flexible solution for cnn-based image denoising," *IEEE TIP*, vol. 27, no. 9, pp. 4608–4622, 2018.

[29] Boyi Li, Xiulian Peng, Zhangyang Wang, Jizheng Xu, and Dan Feng, "An all-in-one network for dehazing and beyond," *arXiv preprint arXiv:1707.06543*, 2017.

[30] He Zhang, Vishwanath Sindagi, and Vishal M Patel, "Multi-scale single image dehazing using perceptual pyramid deep network," in *CVPR'w*, 2018.

[31] Ke Yu, Chao Dong, Chen Change Loy, and Xiaoou Tang, "Deep convolution networks for compression artifacts reduction," *arXiv preprint arXiv:1608.02778*, 2016.

[32] HR Sheikh, "Live image quality assessment database release 2," *http://live. ece. utexas. edu/research/quality*, 2005.

[33] Weimin Tan, Bo Yan, and Bahetiyaer Bare, "Feature super-resolution: Make machine see more clearly," in *CVPR*, June 2018.

[34] Yifan Liu, Ke Chen, Chris Liu, Zengchang Qin, Zhenbo Luo, and Jingdong Wang, "Structured knowledge distillation for semantic segmentation," in *CVPR*, 2019.

[35] Zehao Huang and Naiyan Wang, "Like what you like: Knowledge distill via neuron selectivity transfer," *arXiv preprint arXiv:1707.01219*, 2017.

Authorized licensed use limited to: Tsinghua University. Downloaded on April 20,2022 at 07:07:56 UTC from IEEE Xplore. Restrictions apply.