A More Focus on Multi-degradation Method for Single Image Super-Resolution

Ngoc-Khanh Nguyen[1,2], Thanh-Danh Nguyen[1,2], and Vinh-Tiep Nguyen[1,2]

[1]University of Information Technology, VNU-HCM, Vietnam
[2]Vietnam National University, Ho Chi Minh City, Vietnam
18520901@gm.uit.edu.vn, {danhnt, tiepnv}@uit.edu.vn

***Abstract*—Single Image Super-resolution (SISR) aims at reconstructing a High-Resolution (HR) image from a Low-Resolution (LR) one. Recent works, especially in the deep learning-based approach, mainly define and resolve the problems of LR images degraded by a fixed degradation kernel, typically bicubic interpolation. However, this assumption can hardly be practical since an input image may suffer from many other deteriorations (e.g. blur or noise). Previous works tackle such multi-degradations by proposing new models, targeting at lessening the restrictions of learning-based method and taking advantages of CNN architecture. Unfortunately, they ignore the existing state-of-the-art CNN-based SISR models that are trained on a fixed degradation kernel. In this work, we introduce a context-extending module that generates on-the-fly more realistic types of degradation. We also come up with a comprehensive cross-degradation loss function enabling the model to better adapt real-world conditions. With this proposal, we can generalize arbitrary end-to-end learning-based networks. Evaluating by Peak Signal-to-Noise Ratio (PSNR) metric, our proposed method outperforms the EDSR baseline a significant amount of 34.5% (from 20.14dB to 27.09dB) on noisy images meanwhile sustaining the comparable results on the bicubic downsampling factor.**

***Index Terms*—single image super-resolution, image reconstruction, multi-degradation**

Fig. 1. Exemplary input and output images of our proposed framework under the condition of noisy degradation.

I. Introduction

Single Image Super-resolution (SISR) in computer vision aims to enhance the resolution of an arbitrary low-resolution image. Therefore, the applications of SISR vary in many aspects that have high demand in super-resolution image such as: artistic photography, medical image understanding or urban observation. SISR has gained special attention from research community for decades. In this task, the relationship between an LR image and its ground truth HR image can be various. As to the fact that SISR is an ill-posed problem, an LR image can construct many versions of HR images and vice versa. To be specific, an LR image is considered to be generated by applying transformation methods on an HR image, particularly bicubic interpolation which downsamples an HR image. Based on the assumption that bicubic is the unique degradation of LR images, several work are published to super-resolve bicubic LR images. Besides, other factors degrading HR images such as noise or blur may contribute to a more complex LR image. Fig. 1 shows exemplary samples on enhancing noisy images.

Technically, SISR is categorized into three different approaches: interpolation-based methods [1], [2], model-based methods [3]–[6] and learning-based methods [7]–[9]. Such interpolation-based methods like bicubic interpolation or nearest-neighbor are fast in computation but low in accuracy. Even though model-based methods can flexibly handle to estimate realistic SR image, they are time-consuming because of the optimization procedure. Furthermore, they are not end-to-end SISR manner and require a complicated handcraft hyper-parameters design. Consequently, the learning-based approach, particularly CNN-based method, attracts most attention of research community due to its efficiency and the high-speed computing. CNN-based methods exploit the hierarchical features of LR image to reconstruct high-quality SR image.

The approach of learning-based methods outperforms other strategies in tackling bicubic degradation. However, these methods are not practical in real-world conditions because their performance sinks when the input degradation differs from bicubic. On the one hand, the quality of real-world images is affected by a combination of degradations such as noise and blur. On the other hand, the majority of SISR models employing CNN architecture is trained with synthetic paired dataset which mainly focuses on bicubic downsampling kernels. These reasons result in a huge challenge for learning-based approach to deal with real-world LR images. Thus, recently there are several proposals [10]–[16] aiming at addressing multi-degradations via a single SISR network.

In this paper, we introduce an end-to-end framework targeting at improving the performance of existing bicubic degradation oriented CNN-based models in real-world scenarios, especially with noisy images, resulting in a more applicable SISR model. We propose a pipeline including **an online image augmenting component**, helping the model learn two types of degradation at the same time. Furthermore, by introducing **a new loss function**, our model do not only well study bicubic degradation, but also boosts the results when taking in noisy images, making the model is more applicable when practising with real LR images.

Our contributions are listed in two main points:

- Propose a general end-to-end framework solving multi-degradation problem by using an online data augmentation component.
- Propose a novel Comprehensive Cross-Degradation loss function focusing on reconstructing super-resolution images

The rest of this paper is organized as follows. Section II reviews previous work on the three aforementioned approaches. Section III presents our proposed methods to deal with multi-degradation factor. Section IV shows how effective our method yields over baselines. Finally, Section V concludes our work.

II. Related work

A. CNN-based Super-Resolution

In the past few years, many efforts have been made in SISR by employing CNN architecture to super-resolve an LR image to an HR one. A work proposed by Dong *et al.* [17] with a three-layer CNN architecture for SISR is the pioneer of this approach. However, this architecture is not capable of representing high-level features which badly affects its performance. After that, several work [18]–[21] have extended the network in depth and further using residual connection to avoid overfitting [7], [8]. Kim *et al.* [18] introduces a 20-layer CNN-based network to achieve better result. Lim *et al.* [7] proposes a large architecture with 256 filters at each layer and additionally used residual connection based on ResNet [22]. Lai *et al.* [23] adopts Laplacian pyramid to predict the sub-band residuals of the LR input image. At this time, state-of-the-art result on Set5 dataset belongs to [24]. This architecture employs a resampler module to estimate weights and offsets for the downsampling kernel. GAN-based methods [9] are also utilized to predict more natural SR images. However, these networks only work well under a fixed degradation and suffers from a severe drop when dealing with images in real-world settings, making it not applicable.

B. Multi-Degradation Super-Resolution

Recently, many strategies have been proposed to tackle multi-degradations via a single SISR model. The so-called *non-blind* super-resolution approach assumes the degradations at testing time are available in training phase, meanwhile *blind* one can deal with unseen degradations in the input LR images. The blind super-resolution branch recently receives considerable attention from research community [25]–[29]. It is safely said that non-blind method is an intermediate step to solve blind one. The first effort to resolve LR images under multi-degradations is introduced by Zhang *et al.* [10]. Its core idea is taking not only an LR image as input but also a corresponding blur kernel reduced by PCA algorithm. Nonetheless, this proposal is limited to Gaussian blur kernel and due to its lack of skip connection, the deeper layers are less affected by the degradation maps. Thus, Gu *et al.* [15] improves the work of [10] by applying correction onto the input degradation maps and increasing its influence. Particularly, with each input, it takes a number of iterations to do exactly these things: predict SR image, correct and update the input degradation maps based on the feedback from the estimated SR image.

Deriving from [10], Xu *et al.* [12] employes dynamic convolution to achieve higher performance on tackling with blur and noise kernel. To better deal with non-available degradations in testing phase, Schocher *et al.* [16] proposes a network which completely ignores training phase. It is trained from scratch using the LR input image and its downsampled version to find the optimal parameters. These parameters are then used to predict the original LR image, the model thus can easily adapt to various degradations. Nevertheless, this architecture converges slowly and is time-consuming at testing time. Soh *et al.* [30] adopts meta learning and transfer learning techniques to improve the work of [16]. The model is first trained on prior degradation to achieve parameters θ_t, which helps it converge more quickly when transferring to other degradations. Zhang *et al.* proposes a deep unfolding network [11] that takes advantages of both CNN-based and model-based methods to super-resolve multi-degradation images. Indeed, such model can be trained end-to-end and is able to deal with flexible multi-degradations. Besides, there are works that aim to plug an additional module to super-resolve prior factor [31], [32]. Unfortunately, they suffer from high computational cost. To the best of our knowledge, there is a work [14] aiming at correcting the input LR image. This method can work with existing learning-based methods and requires no additional information. They use a correction filter to modify the downsampling kernel of the input LR image, forcing it to mimic the ideal bicubic degradation, from which the SR image can be easily reconstructed.

III. Methodology

A. Problem Formulation

The degradation model of an LR image can be represented as follows:

$$I^{LR} = (I^{HR} \otimes k) \downarrow_s + n, \tag{1}$$

where k is a blur kernel, $\downarrow_s$ is a downsampling operation with scale factor s and n are usually Additive White Gaussian Noise (AWGN) with noise level σ. For convenience, most of the learning-based SISR models assume that only downsampling interpolation exists, leading to a critical drop in performance

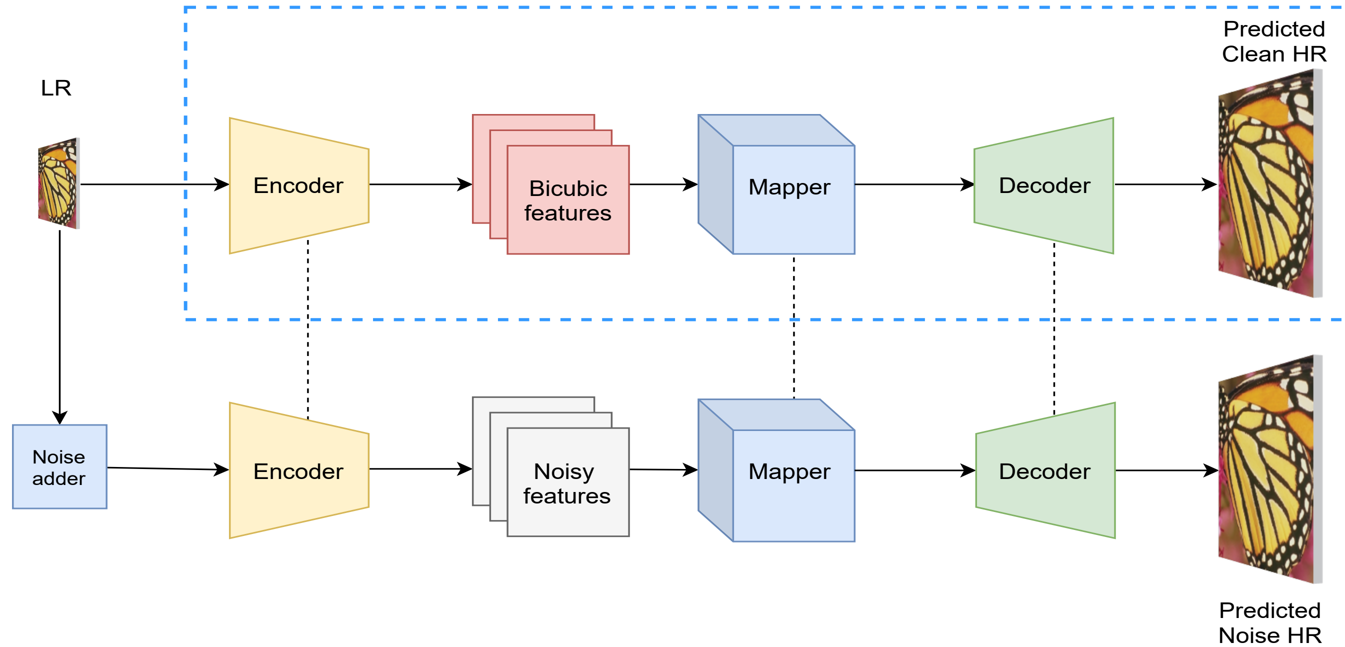

Fig. 2. The proposed framework that solves two factors of degradation by online data augmentation. We only take in bicubic degradation and generate on-the-fly noisy image. The blue dashed rectangle indicates the straightforward path for testing without adding noise to the input, while the black dash implies shared parameters.

when testing with real-world images. In this paper, we propose a framework that helps existing CNN-based SISR models have the ability to study complex degradations, particularly Gaussian noise at level 30 following bicubic kernel.

B. Our proposed method

1) Online Augmentation: As in Equation 1, the most simple observation is bicubic downsampling if we do not consider k and n, and coincidentally $\downarrow_s$ is bicubic interpolation. The formula now becomes:

$$I^{LR} = I^{HR} \downarrow_s, \tag{2}$$

For this convenience, bicubic degradation has been widely used as a benchmark for evaluating SISR models. Furthermore, we notice that the complex degradations can be derived from this simple bicubic. Based on this assumption, we propose a method that performs online augmentation. In detail, we take only an input, i.e., bicubic LR image and the bicubic-to-noise transformation is applied afterward, making our framework parallel process both simple and complex downsampling kernels. We divide bicubic oriented CNN-based architecture into three different components: encoder, LR-HR mapper and decoder. Usually, the encoder is the very first convolution block in the architecture. For example, as in our EDSR [7] baseline, the encoder is the convolution operation right before the sequence of residual blocks, which is the LR-HR mapper. Finally, the decoder is the upsampling operator that upscales the LR feature size into HR space. Specifically, the decoder of EDSR is the PixelShuffle [33] module.

In our framework, the three components of the encoder, LR-HR mapper and decoder share weights between the two processes. As aforementioned, several SISR models poorly perform when the input degradation deviates from bicubic downsampling kernel. Our framework address this problem by providing an additional branch in pipeline. This branch is supposed to bring about the information of the complex degradation that we want our model to learn, along with the bicubic degradation. Thanks to this pipeline, we use exactly the same amount of bicubic data but the model is still able to capture a wide range of deterioration in LR images. Additionally, although the number of parameters stays unchanged compared to the original model, the performance in noisy images is boosted. As a result, **our method performs well on both bicubic and noisy images**.

2) Comprehensive Cross-Degradation Loss Function: Besides the online augmentation, our model learns to present both bicubic and noisy degradation simultaneously in training phase. The LR-HR mapper in our framework is in charge of representing both simple and complex degradation features. Actually, since noisy degradation is more complex, a model tends to bias the easier task, which results in better performance when reconstructing bicubic LR images. Consequently, we may witness a downtrend in recovering LR noisy images and an uptrend on one another. However, we expect to keep the result in bicubic images as high as possible and together improve the performance in generating SR image from a noisy LR one. To ensure this expectation, we introduce a multi-component loss function as below:

$$\mathcal{L}_{SR} = \alpha\mathcal{L}_{ft} + \beta\mathcal{L}_n + \gamma\mathcal{L}_b, \tag{3}$$

where, $\mathcal{L}_{ft}$ follows the formula in Eq. (4) and $\mathcal{L}_n, \mathcal{L}_b$ follow the formula in Eq. (5). The perceptual loss $\mathcal{L}_{ft}$ measures the

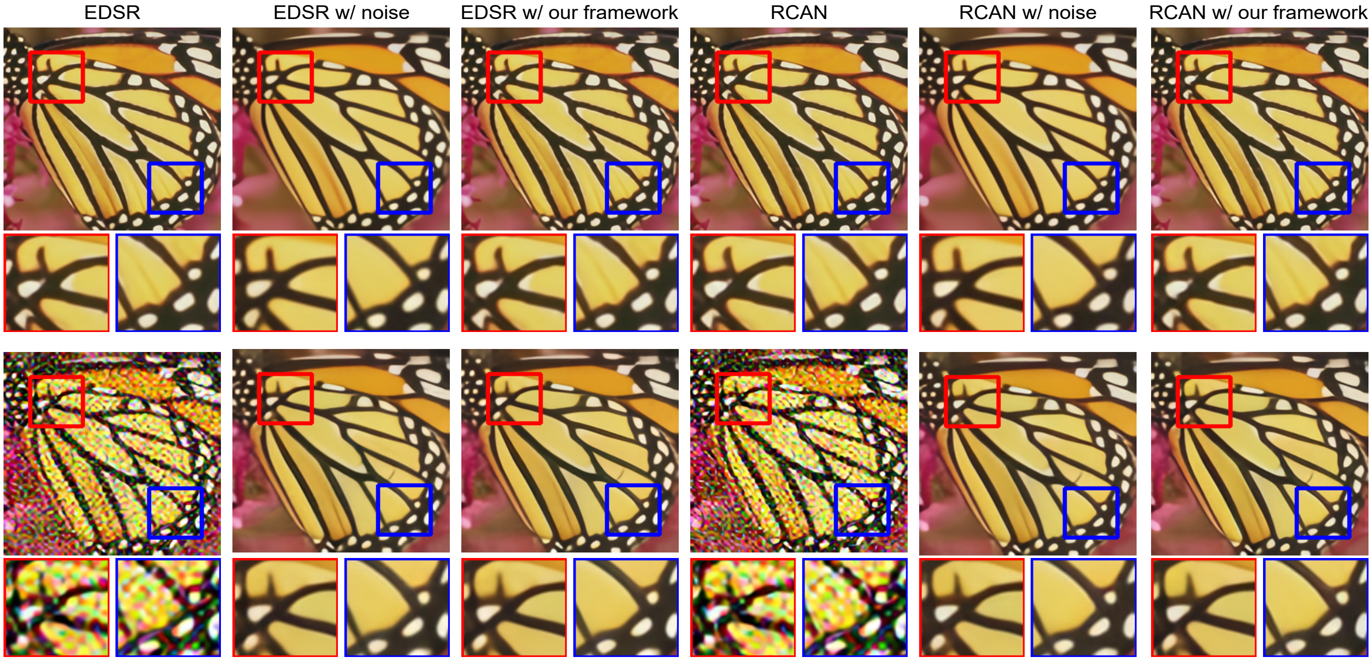

Fig. 3. Comparison between the results of our framework and the original baselines. Left to right: EDSR, EDSR trained with noisy dataset, EDSR with our framework, RCAN, RCAN trained with noisy dataset and RCAN with our framework; Top: results on bicubic images; Bottom: results on noisy images.

TABLE I
SISR PERFORMANCE COMPARISON ON SET5 BENCHMARK WITH SCALE FACTOR 4

Model	PSNR ⇑			SSIM ⇑		
	Bicubic	Noise	*F_1-score*	Bicubic	Noise	*F_1-score*
EDSR [7]	**32.35**	20.14	24.82	0.84	0.35	0.50
EDSR w/ noise	28.14	27.05	27.58	0.80	0.78	0.79
EDSR w/ our framework	32.30	**27.09**	**29.45**	**0.90**	**0.78**	**0.83**
RCAN [34]	**32.32**	19.81	24.56	**0.90**	0.35	0.50
RCAN w/ noise	27.83	**27.05**	27.43	0.79	**0.78**	0.79
RCAN w/ our framework	32.17	27.00	**29.36**	0.89	0.78	**0.83**

difference between feature map extracted from a bicubic image and a noisy image after the encoder module, as shown below:

$$\mathcal{L}_{ft} = \mathcal{L}_1(ft_n, ft_b), \tag{4}$$

where ft_n and ft_b represent noisy feature and bicubic feature, respectively. Here, we want to narrow the representation of these two features. Our assumption is based on the fact that existing CNN-based SISR models have already achieved remarkable performance on bicubic LR images. Therefore, if the noisy image is represented close enough to the bicubic one, we can predict a better SR image because the model now can treat the noisy feature maps exactly like the bicubic ones. However, this assumption could lead to mode collapse because given two representations, the closest distance between them is 0 ($ft_b = ft_n = 0$). Thus, to easily satisfy any distance equation, the model marks the representations as 0, leading the feature maps to contain no information. In order to overcome this drawback, we propose adding two terms to our Comprehensive Cross-Degradation Loss Function, which follows this equation:

$$\mathcal{L}_x = \mathcal{L}_1(HR, SR_x), \tag{5}$$

with $x = n$ if we measure the difference between SR image reconstructed from noisy feature and $x = b$ in case of bicubic feature. These terms help force ft_n and ft_b to estimate SR images under the condition that estimated images must be close to the ground truth HR. By doing so, we give a constraint that our extracted features do not fall into mode collapse.

To leverage our proposed loss function in Eq. (3), there must be different tracks between training and testing phase. At training time, our network generates two inputs (as in Fig. 2) to learn both simple and complex degradation simultaneously. When testing, our pipeline is considered as the bicubic oriented learning-based method without adding noise to the taken LR image.

IV. EXPERIMENT

A. Dataset

In this work, we only consider Gaussian noise level 30 following bicubic interpolation as complex degradation. We use 800 HR training images with resolution of $2K$ in DIV2K dataset [35] and adopt augmentations such as flip, rotate and

TABLE II
FURTHER STUDY ON CNN-BASED METHODS FOR SISR

Model	PSNR			SSIM		
	Bicubic	Noise	F_1-score	Bicubic	Noise	F_1-score
D_DBPN [20]	**32.08**	20.03	24.66	**0.89**	0.35	0.50
D_DBPN w/ noise	12.97	12.93	12.95	0.43	**0.43**	0.43
D_DBPN w/ our framework	31.04	**20.96**	**25.02**	0.88	0.40	**0.55**
RDN [8]	**32.33**	20.05	24.75	**0.9**	0.35	0.5
RDN w/ noise	7.76	7.76	7.76	0.24	0.24	0.24
RDN w/ our framework	31.04	**20.72**	**24.85**	0.88	**0.39**	**0.54**

shift to increase ten times the number of images. So, our total training phase uses 8,000 samples of images. The LR images are generated by bicubically downsampling HR images. As aforementioned in Section III, the noisy counterpart is generated on-the-fly from the bicubic input image. At each iteration, 32 LR images with patch size 48×48 are fed into the network. Next, 32 noisy images are constructed by adding noise to the input images. Each type of degradations walks through every component of the model to reconstruct its SR counterpart.

B. Our Super-Resolution Results

In our experiments, we use Adam optimization with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. We use EDSR [7] and RCAN [34] as our baselines. However, due to the limitation of computational resources, we reduce the number of residual blocks of RCAN architecture to 5. The initial learning rate was set to $1e^{-4}$ and decreased to half once every 200 epochs. The three components $\mathcal{L}_{ft}$, $\mathcal{L}_x$ and $\mathcal{L}_n$ are $\mathcal{L}_1$ pixel-wise loss function. The hyper-parameter α, β and γ are all set equal to 1.

To fairly compare with the original baselines [7], [34], we also evaluate our performance by *PSNR* and *SSIM* metrics. The values are measured on channel *Y* of *YCrCb* color space. However, our ultimate target is not only to keep high performance on bicubic dataset, but also increase the reconstruction result on noisy images. Therefore, we use the *f_1-score* metric to compare our work with traditional baselines. This metric raises high value only if the model performs well on both target degradations. As a result, *f_1-score* on bicubic and noisy images is best fit to measure our aforementioned goal. The visualization of output images is shown in Fig. 3.

The results on EDSR and RCAN baseline are shown in Table I. The bicubic oriented CNN-based baselines are originally trained with bicubic dataset. These models are only capable of well studying seen domain, leading to a critical drop on the other. An alternative approach helping the above methods deal with both bicubic and noisy images is to train it with noisy dataset. Even though the result is boosted on noisy domain, the models witnesses a sharp drop in the performance of bicubic images. Therefore, our framework outperforms above approaches because it reconstructs high-quality SR images regardless of the taken degradation. In a word, the *f_1-score* metric presents our approach get the highest performance over baselines.

TABLE III
ABLATION STUDY ON OUR PROPOSED LOSS FUNCTION

Model	PSNR ⇑			SSIM ⇑		
	Bicubic	Noise	F_1-score	Bicubic	Noise	F_1-score
EDSR w/o bicubic loss	27.74	26.92	27.32	0.79	0.78	0.78
EDSR w/o noisy loss	29.17	21.89	25.01	0.83	0.45	0.58
RCAN w/o bicubic loss	26.95	25.57	26.24	0.77	0.73	0.75
RCAN w/o noisy loss	31.74	20.72	25.07	0.89	0.38	0.53

C. Ablation Study

In order to demonstrate the effectiveness of each component in our proposed Comprehensive Cross-Degradation loss function, we setup experiments to train our framework with the combination of a single perceptual and whether bicubic or noisy loss function. The experimental results of our ablation are shown in Table III. In case of removing noisy loss, the cooperation of perceptual loss and bicubic loss function shows a critical drop in reconstructing noisy LR images measured by PSNR and SSIM metric. As aforementioned in Section III, since recovering bicubic images is easier, the model tends to choose to learn mapping bicubic features to HR images. As a result, the process leads to a fail in attempting to predict noisy inputs. In the other case which the bicubic loss is deprecated, the model focuses on super-resolving noisy images without attention on bicubic features. Therefore, even though the results of reconstructing images improve, the performance in predicting bicubic images is lower than its baselines.

To further investigate the performance of our proposed framework, we establish experiments on RDN [8] and D_DBPN [20] model. The results are presented in Table II. From the reported results, we notice that our framework fails in enhancing the applicable ability of the two taken models. This poor performance is caused because the given models are unable to super-resolve noisy images. Since these architectures are initially designed to cope with bicubic degradation, their latter high-level features directly take in nearly every low-level features beforehand by using dense connection, which leads to information saturation. This mechanism works well on bicubic degradation because the inverse function is more simple than noise (as in Equation 1), making the high-level features still be capable of mapping LR and HR features. When dealing with more complex degradation, these models fail because the saturated high-level information are unable to reconstruct such a complicated function.

V. Conclusion

In this work, we propose a general end-to-end framework for the problem of single image super-resolution. Our proposed method focus on resolving multi-degradation by using an online data augmentation module. Besides, we also introduce a novel Comprehensive Cross-Degradation loss function that supports better reconstructing the super-resolution images. To this end, arbitrary models are able to reach high performance on various types of degradation thanks to applying our framework. In the future, we aim to improve our framework by making it able to satisfy arbitrary degradation. By doing so, we can help bicubic oriented learning-based methods more flexible in real-world conditions.

Acknowledgments. This research is funded by University of Information Technology - Vietnam National University Ho Chi Minh City under grant number D1-2021-18. Besides, we want to give special thanks to Phuoc-Hieu Le and Anh-Khoa Nguyen Vu for their valuable supports.

References

[1] R. Keys, "Cubic convolution interpolation for digital image processing," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 29, no. 6, pp. 1153–1160, 1981.

[2] C. E. Duchon, "Lanczos filtering in one and two dimensions," *Journal of Applied Meteorology and Climatology*, vol. 18, no. 8, pp. 1016–1022, 1979.

[3] S. Dai, M. Han, W. Xu, Y. Wu, Y. Gong, and A. K. Katsaggelos, "Softcuts: a soft edge smoothness prior for color image super-resolution," *IEEE Transactions on Image Processing*, vol. 18, no. 5, pp. 969–981, 2009.

[4] J. Sun, Z. Xu, and H.-Y. Shum, "Image super-resolution using gradient profile prior," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2008, pp. 1–8.

[5] Q. Yan, Y. Xu, X. Yang, and T. Q. Nguyen, "Single image superresolution based on gradient profile sharpness," *IEEE Transactions on Image Processing*, vol. 24, no. 10, pp. 3187–3202, 2015.

[6] A. Marquina and S. J. Osher, "Image super-resolution by tv-regularization and bregman iteration," *Journal of Scientific Computing*, vol. 37, no. 3, pp. 367–382, 2008.

[7] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, "Enhanced deep residual networks for single image super-resolution," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2017, pp. 136–144.

[8] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 2472–2481.

[9] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2017, pp. 105–114.

[10] K. Zhang, W. Zuo, and L. Zhang, "Learning a single convolutional super-resolution network for multiple degradations," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 3262–3271.

[11] K. Zhang, L. V. Gool, and R. Timofte, "Deep unfolding network for image super-resolution," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 3217–3226.

[12] Y.-S. Xu, S.-Y. R. Tseng, Y. Tseng, H.-K. Kuo, and Y.-M. Tsai, "Unified dynamic convolutional network for super-resolution with variational degradations," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 12 496–12 505.

[13] J. W. Soh, S. Cho, and N. I. Cho, "Meta-transfer learning for zero-shot super-resolution," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 3516–3525.

[14] S. A. Hussein, T. Tirer, and R. 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*, 2020, pp. 1428–1437.

[15] J. Gu, H. Lu, W. Zuo, and C. Dong, "Blind super-resolution with iterative kernel correction," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 1604–1613.

[16] A. Shocher, N. Cohen, and M. Irani, ""zero-shot" super-resolution using deep internal learning," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 3118–3126.

[17] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Proceedings of the European Conference on Computer Vision*. Springer, 2014, pp. 184–199.

[18] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1646–1654.

[19] ——, "Deeply-recursive convolutional network for image super-resolution," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1637–1645.

[20] M. Haris, G. Shakhnarovich, and N. Ukita, "Deep back-projection networks for super-resolution," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 1664–1673.

[21] T. Tong, G. Li, X. Liu, and Q. Gao, "Image super-resolution using dense skip connections," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 4809–4817.

[22] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2016, pp. 770–778.

[23] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Deep laplacian pyramid networks for fast and accurate super-resolution," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2017, pp. 624–632.

[24] W. Sun and Z. Chen, "Learned image downscaling for upscaling using content adaptive resampler," *IEEE Transactions on Image Processing*, vol. 29, pp. 4027–4040, 2020.

[25] Y. Chen, Y. Tai, X. Liu, C. Shen, and J. Yang, "Fsrnet: End-to-end learning face super-resolution with facial priors," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 2492–2501.

[26] A. Lugmayr, M. Danelljan, and R. Timofte, "Unsupervised learning for real-world super-resolution," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2019, pp. 3408–3416.

[27] D. Ren, K. Zhang, Q. Wang, Q. Hu, and W. Zuo, "Neural blind deconvolution using deep priors," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 3341–3350.

[28] Z. Shen, W.-S. Lai, T. Xu, J. Kautz, and M.-H. Yang, "Deep semantic face deblurring," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018, pp. 8260–8269.

[29] R. Yasarla, F. Perazzi, and V. M. Patel, "Deblurring face images using uncertainty guided multi-stream semantic networks," *IEEE Transactions on Image Processing*, vol. 29, pp. 6251–6263, 2020.

[30] J. W. Soh, S. Cho, and N. I. Cho, "Meta-transfer learning for zero-shot super-resolution," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2020, pp. 3516–3525.

[31] K. Zhang, W. Zuo, S. Gu, and L. Zhang, "Learning deep cnn denoiser prior for image restoration," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2017, pp. 3929–3938.

[32] K. Zhang, W. Zuo, and L. Zhang, "Deep plug-and-play super-resolution for arbitrary blur kernels," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2019, pp. 1671–1681.

[33] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1874–1883.

[34] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proceedings of the European Conference on Computer Vision*, 2018, pp. 286–301.

[35] E. Agustsson and R. Timofte, "Ntire 2017 challenge on single image super-resolution: Dataset and study," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2017, pp. 126–135.