Real-World Image Super-Resolution by Exclusionary Dual-Learning

Hao Li, Jinghui Qin, Zhijing Yang, Pengxu Wei, Jinshan Pan, Member, IEEE, Liang Lin, Senior Member, IEEE, and Yukai Shi

Abstract—Real-world image super-resolution is a practical image restoration problem that aims to obtain high-quality images from in-the-wild input, has recently received considerable attention with regard to its tremendous application potentials. Although deep-learning-based methods have achieved promising restoration quality on real-world image super-resolution datasets, they ignore the relationship between L1- and perceptual-minimization and roughly adopt auxiliary large-scale datasets for pre-training. In this paper, we discuss the image types within a corrupted image and the property of perceptual- and Euclidean-based evaluation protocols. Then we propose a method, Real-World Image Super-Resolution by Exclusionary Dual-Learning (RWSSR-EDL) to address the feature diversity in perceptual- and L1-based cooperative learning. Moreover, a noise-guidance data collection strategy is developed to address the training time consumption in multiple datasets optimization. When an auxiliary dataset is incorporated, RWSSR-EDL achieves promising results and relapses any training time increment by adopting the noise-guidance data collection strategy. Extensive experiments show that RWSSR-EDL achieves competitive performance over state-of-the-art methods on four in-the-wild image super-resolution datasets.

Index Terms—Deep representation learning, dual learning, efficient training, real-world image restoration.

I. INTRODUCTION

REAL-WORLD image super-resolution (RealSR) aims at restoring in-the-wild images collected from poor-quality sensors with unknown degraded kernels. Since RealSR realizes image restoration under real-world scenarios, it plays a remarkable role in many human-centric applications, such as mobile photo enhancement, automatic pilot, etc.

Manuscript received 9 June 2021; revised 21 January 2022; accepted 1 June 2022. Date of publication 10 June 2022; date of current version 30 October 2023. This work was supported in part by the National Natural Science Foundation of China (NSFC) under Grant 62002069, in part by the Science and Technology Project of Guangdong Province under Grants 2021A151011341 and 2022A151011835, in part by the China Postdoctoral Science Foundation funded project under Grant 2021M703687, and in part by Guangzhou Science and Technology Plan Project under Grants 20200203086 and 102020690. The Associate Editor coordinating the review of this manuscript and approving it for publication was Prof. Mea Wang. (Hao Li and Jinghui Qin contributed equally to this work. (Corresponding author: Yukai Shi.)

Hao Li, Zhijing Yang, and Yukai Shi are with the School of Information Technology, Guangdong University of Technology, Guangzhou 510006, China (e-mail: liha09605@gmail.com; yzhj@gdut.edu.cn; yksli@gdut.edu.cn).

Jinshan Pan is with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing 210094, China (e-mail: sdluran@gmail.com; span@njust.edu.cn).

Jinghui Qin, Pengxu Wei, and Liang Lin are with the School of Data and Computer Science, Sun Yat-sen University, Guangzhou 510006, China (e-mail: qingjing@mail.sysu.edu.cn; weipx3@mail.sysu.edu.cn; linliang@ieee.org).

Digital Object Identifier 10.1109/TMM.2022.3181457

1520-9210 © 2022 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission. See https://www.ieee.org/publications/rights/index.html for more information.

Traditional single image super-resolution (SISR) [1]–[4] obtains high-resolution images from low-quality ones with known and fixed degradation models (e.g., Gaussian blur followed by Bicubic downsampling). As the in-the-wild corrupted images own complicated degradation kernels, traditional SISR model exhibits limited capacity when applying to in-the-wild applications. To address this issue, RealSR constructs the real-world image pairs by incorporating poor- and high-quality sensors to take low- and high-quality images, respectively. Compared with manual downsample kernel, the degraded kernel in RealSR is inherently more complicated and closed to in-the-wild degradation. As the mobile platform has a limited optical sensor size and a large number of users, mobile photography enhancement is one of the most challenging applications for RealSR. Thus, RealSR prefers to entertain the human visual system. Typical evaluation protocols in traditional SR (e.g., PSNR and SSIM [5]) focus on pixel-level similarity and fail to reflect human perception well. Recently, various kinds of efforts [6]–[8] are proposed to reflect the human visual system in image quality assessment. LPIPS [7] argues that the widely used pixel-to-pixel measuring methods (e.g., L2/PSNR, SSIM, and FSIM) are contrary to human perception when estimating the perceptual similarity of images.

Recently, RealSR methods [9], [10] usually adopt LPIPS and PSNR as default evaluation metrics and achieve superior scores either on PSNR or LPIPS. As depicted in Fig. 1, high PSNR increment samples exhibit simple background, smooth structure, and relatively weak LPIPS increment. Besides, high LPIPS gain samples own complicated texture and less PSNR gain. On one hand, images with complex textures are easier to get high scores on LPIPS on account of psychophysical similarity measurements [7], while the adversarial training is adept at generating artificial texture. On the other hand, Euclidean-based measurements inherently prefer L1 minimization. Although the metrics based on adversarial loss and L1-norm-based loss are able to generate the results with favorable both LPIPS and PSNR values, the prior methods still utilize the weighted ratio to simply join them all.

Nevertheless, multiple datasets are another challenge for real-world image restoration frameworks. In RealSR task, many datasets [9]–[13] are proposed to address the real-world degradation and pixel displacement. Extensive training on multiple RealSR datasets is beneficial for generalization enhancement as well as performance improvement [13]. However, typical backbones [14]–[16] require several days to be converged on.
LI et al.: REAL-WORLD IMAGE SUPER-RESOLUTION BY EXCLUSIONARY DUAL-LEARNING

Fig. 1. We obtain high-profit results of DASR [18] from the perspective of PSNR and LPIPS index. Its discernable that the left high PSNR gain examples (>1.5 dB) obtain clear background, smooth structure and inconspicuous LPIPS promotion. However, the high LPIPS profits samples (>0.25) contain complicated texture and a few artifacts, which also obtain less PSNR gain. Though PSNR and LPIPS are both used for evaluating image restoration, they fail to reach an agreement.

one standard dataset [17] with a single GPU. The incremental number of RealSR datasets would raise more requirements toward efficiency. Specifically, the training efficiency would suffer a heavier burden when auxiliary datasets are incorporated for pre-training. This motivates us to investigate a dataset distillation strategy for multiple dataset collaborative training.

Previously, research communities pay considerable attention to model distillation as large-scale dataset plays a important role in training time consumption. For instance, DIV2K has 1,000 images with 2K resolution, NTIRE2020 and AIM2020 challenges provide 3,450 images with 2K resolution, respectively. Though having large-scale training, especially when encountering data augmentation, can significantly improve performance, the following training time cost certainly receives an obvious increment.

To tackle the above challenge, we observe that a full image contains many image types, which lead to a large variance in structure signal. Inspired by this, we introduce a Dual-Learning strategy based on an exclusionary attention mechanism to facilitate different types of images to exhibit diverse representation under multi-task learning paradigms. To address training time consumption issue, we also explore noise extraction and blending mechanism across multiple datasets, and proposed an efficient data collection strategy. With the proposed data collection strategy, we certainly relax the training time growth problem while incorporating an auxiliary large-scale dataset for mix-training. Our contributions can be summarized as:

- We propose an efficient training paradigm (i.e., NGDC algorithm) that extra time consumption is relaxed when incorporating auxiliary datasets for training.
- We proposed an exclusionary mask mechanism, namely RWSR-EDL, that brings significant improvement toward multi-loss training conditions in RealSR.
- We provide a comprehensive comparison on four challenging real-world SR benchmarks (e.g., AIM2019 [9], NTIRE2020 [10], CameraSR [11], RealSR [12]) to demonstrate that RWSR-EDL achieves a clear performance improvement on PSNR and LPIPS index both, and present a high-quality image restoration.

II. RELATED WORK

Single Image Super Resolution: Deep learning-based SR methods have achieved significant improvements over conventional SR methods on restoration quality. Among these methods, Dong et al. proposed the first CNN-based SR method called SR-CNN [19], which utilizes a three-layer CNN to learn a nonlinear mapping between LR and HR image pairs. Lim et al. proposed EDSR [14], using simplified residual blocks [20] for training SR model, which can achieve great improvement over restoration.
quality. Luo et al. [21] proposed a Feature Enhanced Pyramid Cascading and Deformable convolution (FEPCD) module to align multiple low-resolution burst images in the feature level. Zhang et al. proposed a residual in residual (RIR) [22] to address the difficulty of training a deep SR network with a channel attention mechanism to improve the representation ability of the SR network. Wang et al. proposed ESRGAN [23], which introduced a Residual-in-Residual Dense Block (RDB) into the generative network. Furthermore, FASRGAN [1] explored a novel pixel-level discriminator for a fine-grained generative adversarial learning and obtains interesting results. Hu et al. [24] address the simplicity of the downsampling kernel in image SR by adopting multiple degradation metrics, and achieves promising results on complex corrupted images. Recently, many dual-way neural network [2] are applied to image restoration, however, they only address the traditional image SR problem and neglect the complexity of real-world cases. Mei et al. [25] proposed the first Cross-Scale Non-Local (CS-NL) attention module and a powerful Self-Exemplar Mining (SEM) cell into deep neural networks. Jiang et al. [26] proposed a hierarchical dense connection network (HDN) for SR, achieved the improvements of both reconstruction performance and efficiency.

Although the above deep learning-based SR methods bring significant improvement over SISR, they cannot be generalized well on real-world images as their assumed the known and fixed degradation process does not hold for real-world images.

Real-World Super Resolution: Recently, Real-world super-resolution attracts considerable attention due to its distinguished practicality. Different from the traditional image SR that generally focuses on simple and uniform synthesized degradation, RealSR needs to handle complicated and rough degradation in real-world cases. To address the above challenges, Shocher et al. proposed zero-shot super-resolution (ZSSR) [27], they realize an unsupervised CNN-based SR method by training an image-specific SR network with internal data rather than employing external data. Bell-Kligler et al. proposed KernelGAN [28], to generate down-sampling kernel from label images by adopting kernel estimation GAN, which used in ZSSR for degraded kernel estimating. Fritsche et al. proposed the DSGAN model [29] to generate LR-HR pairs and then apply ESRGAN-FS on corresponding generated images. Pang et al. proposed FAN [30] to extract the different frequency components of the LR images, which can be used to recover the HR images by preserving more high-frequency details. Wei et al. proposed a DASR framework [18] by calculating the domain distance between LR images and real images with the domain-gap aware training and domain-distance weighted supervision. Ji et al. [31] proposed an RWSR model based on ESRGAN by using kernel estimation and noise injection.

However, the above methods only employ a single structure network to obtain the enhanced image, they ignore that the multi-branch network can learn more diverse image features with various prior knowledge. In contrast, we use a dual-learning refinement module and an exclusionary mask generator to further extract the diverse representation. With our exclusionary dual-learning strategy, we obtain the final enhanced image with rich high-frequency details, which indeed address the complicated degradation in real-world super-resolution.

Since the ground-truth label may absent in some image enhancement applications, some no-reference image quality assessment (IQA) metrics [32]–[34] are proposed. Spatial-Spectral Entropy-based Quality (SSEQ) [32] is capable of assessing the quality of a distorted image across multiple distortion categories. Naturalness Image Quality Evaluator (NIQE) [33] uses the multivariate Gaussian (MVG) model to fit the quality-aware features extracted from images. And Perception Index (PI) [34] combines the no-reference image quality measures of Ma et al. [35] and NIQE [33] to get the score.

Noise Modeling Based Denoising: Reducing the effect of noise is a critical issue in real-world image restoration. Recently, some noise modeling-based approaches have been proposed to address the real-world noise estimation and reduction problem. Lebrun et al. proposed Noise Clinic (NC) [36], a method to estimate the noise model dependent on signal and frequency followed by using non-local Bayes (NLB) [37]. Zhang et al. proposed FFDNet [38], a non-blind Gaussian denoising network that can obtain charming results on real-world noise cases, nevertheless, it requires manual intervention to select noise level.

However, real-world image noise is indeed distinct from Additive White Gaussian Noise (AWGN) and the network trained by the handcraft degradation gives a poor performance when applied to complex real-world noise. Guo et al. proposed CBDNet [39], which is composed of two sub-networks, that is noise estimation and non-blind denoising net. They use a noise modeling method that is able to generate realistic noise and also adopt real-world clean images for paired-wised training. Chen et al. proposed GCBD [40] to extract the smaller image with clear background in noisy images, and then incorporating GAN to generate more fake noise samples for training denoising CNN network. Anwarf et al. proposed RIDNet [41], a single-state denoising network with feature attention for real-world noise reduction.

Different from the above approaches, our RWSR-EDL is free from noise modeling, instead, we make use of inherent noise sampling strategy towards the real-world images to construct paired training data, which can eliminate the biases introduced by noise modeling. Also, we incorporate the NGDC strategy that effectively distills a large-scale auxiliary dataset to obtain target real-world noise type. With NGDC strategy, the performance is clearly improved without training time increment.

III. METHODOLOGY

A. Overview

Traditional single image super-resolution (SISR) aims at restoring a high-quality image \( I_{HR} \) from a low-quality image \( I_{LR} \). In traditional SISR, the \( I_{LR} \) is synthesized from \( I_{HR} \) with a down-scaling operation:

\[
I_{LR} = (I_{HR} \otimes K_{Gauss}) \downarrow_k
\]

(1)

where \( K_{Gauss} \) is a Gaussian blur kernel and \( \downarrow_k \) denotes the image degradation procedure with downscale factor \( k \). To obtain a low-quality \( I_{LR} \), \( \downarrow_k \) typically adopts a bicubic-based down-sampling algorithm. In contrast to traditional SISR, real-world super-resolution aims to address real-world image degeneration metrics by capturing \( I_{HR} \) and \( I_{LR} \) with different quality optical

Authorized licensed use limited to the terms of the applicable license agreement with IEEE. Restrictions apply.
sensors and resolution settings where the degraded metrics (e.g., \(\downarrow_k\)) and blur kernel are unknown. Therefore, \(\{I_{LR}, I_{HR}\}\) pairs inherently has different resolution properties in the real-world data collection procedure.

To improve the quality of real-world images and attack existed noises and artifacts in real-world images, we propose a novel single image super-resolution framework to take precedence over learned image features. As shown in Fig. 2, there are two key components in our RWSR-EDL: 1) noise-guidance data collection for an efficient training in multiple large-scale RealSR datasets; 2) exclusionary mask generator for relaxing multi-loss optimization in RealSR. In our method, we first enforce the LR images and ground-truth images yield similar noise distribution by embedding random noise into LR images. The intermediate HR images generated by the main feature extractor will be entered into the exclusionary dual-network. Finally, our framework learns diverse intermediate representations adaptively to pursue high PSNR and LPIPS scores by incorporating exclusionary mask and multi-loss optimization.

**B. Main Generator**

Since the goal of our work is to improve the perceptual quality of SR images under the real-world setting, we adopt an effective image enhancement backbone for feature extraction, which consists of 23 residual-in-residual dense blocks (RRDBs [23]) and incorporate paired-wised training. Given \(I_{LR}\), we generate intermediate HR image \(I_G\) with the main generator as follows:

\[
I_G = Generator(I_{LR}).
\]  

(2)

**C. Exclusionary Mask Generation**

To obtain a better super-resolving image by further improving the intermediate HR image \(I_G\), we employ a deep feature extraction, which consists of two branches with the same network design but various initialization metrics to fully explore diverse feature representation. The three branches both contain ResBlocks which has two 3 × 3 convolutional layers. The first convolutional layer with 3 input channels and 64 output channels is followed by a ReLU activation function while the second convolutional layer has 64 input channels and 3 output channels. To this end, each ResBlock in feature extraction phase can be formulated as follows:

\[
I^R_x = I_G + f_{RB}(I_G).
\]  

(3)

where \(x\) represents the \(x_{th}\) ResBlock in the three branch, \(f_{RB}\) indicates the ResBlock, \(I_G\) and \(I^R_x\) means the input and output of ResBlock, respectively. Similar to (3), we obtain \([I^R_x, I^H_x, I^m_x]\) w.r.t three parallel branches.

As the refined images \(I^R_x\) and \(I^H_x\) are obtained, we deploy a soft-mask generator to demonstrate the exclusionary dual-learning. More specifically, \(I^R_x\) and \(I^H_x\) are optimized with various loss functions to pursuing PSNR and LPIPS promotion both by applying exclusionary masks. Let \(I^m_x\) represent the output of the second branch, we apply a softmax operator on \(I^m_x\) to normalize all feature values into 0.0 to 1.0 as an adaptive mask. Specifically, a soft-mask for channel index can be generated as follows:

\[
M_{\alpha} = softmax(I^m_x) = \frac{\exp(z_n)}{\sum_{n=1}^{N} \exp(z_n)}
\]  

(4)

where \(n\) and \(z_n\) are channel number and specific value of \(I^m_x\) with channel \(n\) index, respectively. To this end, we obtain the \(M_{\alpha}\) for exclusionary dual-learning.

**D. Loss Function**

Since the exclusionary dual-learning is proposed relax the conflicts of perceptual- and L2-based optimizations, we first briefly introduce the the losses we used in this paper, including Pixel Loss, Perceptual Loss, and Adversarial Loss.

**Pixel Loss:** we use L1 loss, which is a widely used loss function for general image restoration, to train our generator to recover as much effective pixel as possible. The L1 loss is defined by the manhattan distance between the reconstructed image \(I_{SR}\) and the ground-truth image \(I_{HR}\) as follows:

\[
\mathcal{L}_{pix}(I_{SR}, I_{HR}) = \frac{1}{N} \sum_{i=1}^{N} \| I_{HR} - I_{SR} \|_1
\]  

(5)

where \(N\) is the samples of training set.

**Perceptual Loss:** To further enhance the high-frequency features (such as edges) in the SR image, we deploy a perceptual loss based on feature space. Specially, we extract the features of \(I_{SR}\) and \(I_{GT}\) with a pre-trained VGG-19, and compute their
loss as follows:
\[
\mathcal{L}_{\text{per}}(I_{SR}^i, I_{HR}^i) = \frac{1}{N} \sum_{i=1}^{N} \left\| VGG_{19}(I_{HR}^i) - VGG_{19}(I_{SR}^i) \right\|_1
\]

(6)

where \( N \) is the samples of training set and \( VGG_{19} \) denotes a pre-trained VGG-19 model [42].

Adversarial Loss: We also deploy adversarial loss to enhance the SR image’s texture to make it more realistic. The adversarial loss is defined as follows:
\[
\mathcal{L}_{\text{adv}}(I_{SR}^i, I_{HR}^i) = \frac{1}{N} \sum_{i=1}^{N} \left\{ -E[\log(1 - \sigma(D(I_{HR}^i) - \mathcal{L}_{\text{pix}}(I_{SR}^i, I_{HR}^i))) - E[\log(\sigma(D(I_{SR}^i)) - \mathcal{L}_{\text{pix}}(I_{SR}^i, I_{HR}^i)))] \right\}
\]

where \( N \) is the samples of the training set, \( \sigma \) represents a sigmoid function, and \( D(\cdot) \) is the discriminator. In adversarial learning, patch discriminator takes advantage of typical used VGG-128 with some aspects. First, patch discriminator is a fully convolutional network, which is free from image size restriction by getting rid of the fully-connected layer. Second, local feature representation is excavated with the limited receptive field. We then apply patch discriminator instead of VGG-128 as \( D(\cdot) \).

Exclusionary Dual-Learning: As illustrated in Fig. 1, a region with high LPIPS profits often bears low PSNR increment. However, a typical utilization of \( \{\mathcal{L}_{\text{adv}}, \mathcal{L}_{\text{per}}, \mathcal{L}_{\text{pix}}\} \) in imageSR [43] is employ them together with weighted average:
\[
\tilde{\mathcal{L}}_{\text{all}} = \alpha \mathcal{L}_{\text{adv}}(I_{SR}^i, I_{HR}^i) + \beta \mathcal{L}_{\text{per}}(I_{SR}^i, I_{HR}^i) + \gamma \mathcal{L}_{\text{pix}}(I_{SR}^i, I_{HR}^i),
\]

(7)

where \( [\alpha, \beta, \gamma] \) are empirical weight factors. \( \tilde{\mathcal{L}}_{\text{all}} \) incorporates all losses with fixed weight factors for optimization and ignores the diversity of image types in term of perceptual- and L1-norm-based metrics. Contrary to the (7), we adopt exclusionary masks in multiple losses training to avoid domain-conflicts and obtain more accurate feature representation:
\[
\mathcal{L}_{\text{all}} = \mathcal{L}_{\text{adv}}(M_\alpha \cdot I_R^i, I_{HR}^i) + \mathcal{L}_{\text{per}}(M_\alpha \cdot I_R^i + M_\beta \cdot I_R^i, I_{HR}^i) + \mathcal{L}_{\text{pix}}(M_\beta \cdot I_R^i, I_{HR}^i),
\]

(8)

where \( \cdot \) is a matrix dot operation. Meanwhile, we obtain \( M_\alpha \) with (4), and \( M_\beta \) is obtained from \( M_\alpha + M_\beta = 1 \) to promise the two masks have exclusionary property.

In (8), \( M_\alpha \) enforces partial region of \( I_R^i \) optimized with \( \mathcal{L}_{\text{adv}} \) and avoid the distortion of \( \mathcal{L}_{\text{pix}} \). Simultaneously, \( M_\beta \) performs spatial attention on \( I_R^i \) for \( \mathcal{L}_{\text{pix}} \) optimization and free from \( \mathcal{L}_{\text{adv}} \). The overall output \( [I_R^i + I_R^i] \) is further smooth by \( \mathcal{L}_{\text{per}} \) in (8). With the proposed exclusionary dual-learning mechanism, we can demonstrate a fine-grained multi-loss optimization and achieve high PSNR and LPIPS profits both.

Difference to Spatial Attention Mechanism: Compared with spatial attention mechanism [44], [45], which simply highlights the feature according to the high entropy information, the proposed exclusionary dual-learning mechanism enforces the branches capture diverse feature with exclusionary masks. More specific, the generated soft-mask utilizes a branch to obtain high-value feature representation as usual, then, another branch is enforced to capture extensive information. With this competitive mechanism, our model can learn complementary features at the same time and demonstrates promising results on complex real-world conditions. To this end, we address that different image types exhibit different signal properties with deeply learned representation and the adaptive mask is able to reconstruct better visual-quality SR images. In Ablation Study Section, we will show that using the adaptive mask is superior to using a plain dual-way neural network.

E. Noise-Guidance Data Collection

Data Preparation: As the real-world images inherently own a certain proportion of noise, which leads to the restored images contain spare artifacts, we apply a down-sampling to reduce the negative impact. Specifically, we incorporate a bicubic kernel \( K_{bic} \) on the source image \( I_{src} \) to implement noise remove and obtain the \( I_{HR} \):
\[
I_{HR} = (I_{src} \otimes K_{bic}) \downarrow_k.
\]

(9)

Similarly, we can obtain the corrupted image \( I_{LR} \) with:
\[
I_{LR} = (I_{HR} \otimes K_{bic}) \downarrow_k.
\]

(10)

Noise Sampling: The high-frequency signal among \( I_{src} \) has a certain amount of drawback as the (9) reduces the pattern information. The noise distribution on \( I_{LR} \), therefore, meets a significant change. In order to keep the \( I_{LR} \) and \( I_{src} \) has similar noise distribution, we directly sampling the noise in \( I_{src} \) with a grid strategy. Specifically, a sliding window with a size of \( s \times s \) and the stride is \( s \) as well is used to capture smaller images. The mean and variance value of the captured patch is computed for sample sifting.

Suppose variance and mean value of a fetched patches \( n_i \) is \( \sigma_i \) and \( m_i \). We obtain the CharAcTeristic Interval (CATI) (i.e., \( [\sigma_{\theta_1}, \sigma_{\theta_2}] \) and \( [m_{\theta_1}, m_{\theta_2}] \) by acquiring bottom 2% patches from the perspective of variance. When the \( \sigma_i \) and \( m_i \) pertain to CATI, we regard \( n_i \) as a noise patch, and joint it into the noise patch bank \( N \). Otherwise, we regard the rest patch as noiseless patch.

As shown in Fig. 4, the noiseless patch, which contains richer texture and detail, has larger \( \sigma \). We regard these variances exceeded patch as noiseless patch along with two aspects: 1) Complex structure and texture may cover up the noise. 2) Noise signal collection across multiple datasets benefit from the clean background. To this end, the patch with smaller \( \sigma \) and positive \( m \), which enjoys plain content and help us purely extract the inherent noise, is collected into noise bank.

Then, we randomly fetch \( n_i \) from the constructed \( N \) and infuse \( n_i \) into \( I_{LR} \) to generate training input \( I'_{LR} \) as:
\[
I'_{LR} = I_{LR} + n_i.
\]

(11)

As illustrated in (11), the training pairs become more complex and the noise resistance capability of our model is enhanced with the discriminative learning.

Noise-Guidance Data Collection: In inspired by the noise sampling phrase, we extend it to cross-dataset noise collection. Assuming the noise samples, which owns noise type similar to the target domain, is existed in an additional dataset. We aim at distilling auxiliary datasets and indexing similar noise domain.
image. More specifically, we collect a noisy bank $N_a$ from the target dataset $D_a$ by incorporating its CATI. Then, we apply CATI, which was obtained from $D_a$ already, to the auxiliary dataset $D_b$ as well. Notice here that $D_b$ is usually larger than $D_a$. For instance, suppose a source image $I_{LR}^a$ in $D_a$ pertains to the CATI, we generate the corresponding $\{I_{HR}^a, I_{LR}^b, N^b\}$ and merge them into $\{I_{HR}^a, I_{LR}^b, N^a\}$. Finally, the distilled training pairs $\{I_{LR}^{a+b}, I_{HR}^{a+b}, N^{a+b}\}$. We sketch the overall algorithm in Algorithm 1. With NGDC, we can significantly reduce the data volume of auxiliary dataset as well as achieve better performance.

**Difference to Impressionism:** As depicted in [31], Impressionism collects noise patches by the following rule: $\sigma(p) < v$, where $\sigma(\cdot)$ denotes the function to calculate variance in image patch $p$, and $v$ is value of a fixed threshold. However, $v$ is an empirical value and needs a manual search when using a different training set. To remedy this, as demonstrated in Algorithm 1, we obtain the patches from the bottom 2% variance by calculating the variance of all patches, and this rule can be applied to any training set. As shown in Fig. 3, NGDC realizes an efficient training paradigm that achieves a faster convergence by collecting valuable noise patches across multiple training sets.

IV. EXPERIMENTS

A. Dataset and Evaluation Protocols

We use following four real-world SR datasets for comprehensive comparisions to validate our RWSR-EDL:

- **RealSR** [12] consists of 595 LR-HR image pairs, which are collected under a real-world scenario with various optical resolution. To stay in step with other baseline methods [18], [29], we use 200 RealSR image pairs and 800 clear images from DIV2K for training. Then, we conduct specified 50 LR-HR pairs, which are collected by Canon camera, for testing. Due to the images from the RealSR dataset are

![Fig. 3. PSNR/LPIPS curves of different settings on RealSR [12] with training iteration index. With NGDC strategy, although we incorporate auxiliary large-scale dataset(i.e., NTIRE2020 [10]) for training, the training time receives no gain while the performance is clearly improved.]

![Algorithm 1: Algorithm of NGDC.](image)

| Requirement: Target dataset $D_a$, auxiliary dataset $D_b$, noise bank $N_a = \emptyset$ and $N_b = \emptyset$ |
| --- |
| 1. for $I_{LR}^a$ in $D_a$ do |
| 2. Compute $\sigma_i$ and $m_i$ by calculating $I_{HR}^a$ |
| 3. $I_{HR}^a = (I_{HR}^a \otimes K_bic) \downarrow_k$ |
| 4. $I_{LR}^b = (I_{HR}^b \otimes K_bic) \downarrow_k$ |
| 5. end for |
| 6. Construct $N_a$ from bottom 2% $\sigma$ value images and compute corresponding CATI $\in [\sigma, m]$ |
| 7. for $I_{LR}^b$ in $D_b$ do |
| 8. for patch $p_j^b$ in $I_{LR}^b$ do |
| 9. Compute $\sigma$ and $m$ for $p_j^b$ |
| 10. if $\{(\sigma, m)|p_j^b\} \in CATI$ then |
| 11. $N_b = N_b + p_j^b$ |
| 12. end if |
| 13. end for |
| 14. end for |
| 15. Construct $\{I_{LR}^{a+b}, I_{HR}^{a+b}, N^{a+b}\}$ |
| 16. for $I_{LR}^i$ in $D_b$ do |
| 17. for patch $p_j^i$ in $I_{LR}^i$ do |
| 18. Compute $\sigma$ and $m$ for $p_j^i$ |
| 19. if $\{(\sigma, m)|p_j^i\} \in CATI$ then |
| 20. $N_i = N_i + p_j^i$ |
| 21. $I_{HR}^i = (I_{HR}^i \otimes K_bic) \downarrow_k$ |
| 22. $I_{LR}^i = (I_{HR}^i \otimes K_bic) \downarrow_k$ |
| 23. end if |
| 24. end for |
| 25. end for |
| 26. Obtain $\{I_{LR}^{a+b}, I_{HR}^{a+b}, N^{a+b}\}$ via $D_b$ as same as $\{I_{LR}^{a+b}, I_{HR}^{a+b}, N^{a+b}\}$ |
| 27. Construct $\{I_{LR}^{a+b}, I_{HR}^{a+b}, N^{a+b}\}$ |

![Fig. 4. Noise and Noiseless patch.](image)
captured under different optical settings, they inherently have different resolutions, we adopt × 4 scale to evaluate our RWSR-EDL.

- **CameraSR** [11] is a real-world SR dataset, which consists of 200 LR-HR pairs collected by iPhoneX and Nikon camera, respectively. In our experiment, we used 80 real-world images, which are collected by iPhoneX (e.g., No.021-100) and 800 clean images from DIV2K for training. The rest 20 pairs of LR-HR pairs (No.001-020) are incorporated for testing.

- **NTIRE2020 Challenge** [10] contains 3550 images, which downscaled with unknown noise to simulating inherent optical sensor noise. In our experiment, we use 3450 images, which consists of 2650 images from Flickr2K and 800 images from DIV2K, for training. The testing data contains 100 images from the DIV2K validation set with the same degradation operation as the training image. We adopt the × 4 scale to evaluate our RWSR-EDL.

- **AIM2019 Challenge** [9]: The training and testing images provided by AIM2019 are the same as NTIRE2020. Nevertheless, different and undisclosed downgrade metrics were used to generate corrupted input. We adopt the × 4 scale to evaluate our RWSR-EDL as well.

In evaluation protocols, we adopted Peak Signal-to-Noise Ratio (PSNR), Structural Similarity (SSIM), Learned Perceptual Image Patch Similarity (LPIPS) [7], Naturalness Image Quality Evaluator (NIQE) [33] and Mean Opinion Score (MOS) to verify RWSR-EDL. Meanwhile, PSNR and SSIM are the most commonly used evaluation metrics in image restoration, as they focus on pixel fidelity rather than human perception. In contrast, LPIPS and NIQE pays more attention to human visual perception, the lower values of LPIPS and NIQE score indicate better perception quality in the SR observations. We collect the MOS results by recruiting 20 volunteers for subjective assessment on NTIRE2020 challenge data. Specifically, 100 pairs of patches from the NTIRE2020 testset are randomly partitioned and the volunteers were shown a side-by-side comparison of each method’s result and the referenced HR image. They were then asked to evaluate the quality of the SR image w.r.t. the reference image using the 6-level scale defined as: 0 - ‘Perfect’, 1 - ‘Almost Perfect’, 2 - ‘Slightly Worse’, 3 - ‘Worse’, 4 - ‘Much Worse’, 5 - ‘Terrible’.

### B. Implementation Details and Competing Methods

In our experiments, we use flip and random rotation with angles of 90°, 180° and 270° for data augmentation. The images are cropped into 128 × 128 are input and the batchsize is 16. We use Adam [46] as optimizer, where \( \beta_1 = 0.9 \) and \( \beta_2 = 0.999 \). The initial learning rate is \( 1 \times 10^{-4} \) and our RWSR-EDL is trained on an NVIDIA Tesla V100 server. Since the LR and HR images in the CameraSR dataset have the same resolution, we removed the corresponding downsampling and upsampling layers in CameraSR experiment. For the NGDC strategy, we adopt data of NTIRE2020 challenge [10] as an auxiliary dataset to present an efficient mix training in all experiments except the NTIRE2020 experiment itself.

### C. Quantitative and Qualitative Comparisons

**RealSR:** As depicted in Table I, we compare the state-of-the-art methods on RealSR dataset. Compared with DASR, our RWSR-EDL achieves clear improvement over three evaluation metrics. For instance, as DASR is the recently proposed method, RWSR-EDL obviously surpasses DASR with 1.57 dB and 0.045 on PSNR and SSIM, which justify the effectiveness of RWSR-EDL. Compared with the zero-shot learning method, RWSR-EDL still achieves 1.76 dB improvement. Similarly, as shown in Table II, the quantitative performance in terms of NIQE indicates considerable improvement using RWSR-EDL compared to other models. Also, we present the visual comparison in Fig. 5. It can be observed that other one-fold structure methods obtain blurry results on the RealSR dataset. By contrast, RWSR-EDL obtains clear structure and sharp details, which verify the effectiveness of the exclusionary dual-learning mechanism and NGDC strategy.
CameraSR: In Table III, we present the quantitative comparison on CameraSR dataset. Compared with DASR, our model achieve 1.05 dB improvement. Besides, RWSR-EDL still obtains 0.008 improvements over the LPIPS index, which justifies that our soft-mask mechanism makes a good balance between L1- and perceptual- minimization. As shown in Fig. 6, RWSR-EDL also presents a high-quality restoration with more details. For example, compared with DASR, our method recovers more clear lines of the bridge on the upper row of Fig. 6, which indeed show that the soft-mask strategy preserves sharp structure with adversarial optimization.

AIM2019 Challenge: As depicted in Table IV, Fig 7 and Fig.8, we compare RWSR-EDL with state-of-the-arts on data of the AIM2019 challenge. It notes that FSSR is the winning entry on the AIM2019 challenge. Compared with FSSR, our RWSR-EDL achieves 0.55 dB and 0.046 gain with PSNR and SSIM index. Also, RWSR-EDL consistently achieves first place among three evaluation metrics, which fully justifies that our exclusionary
dual-learning mechanism helps RWSR-EDL realizes effective spatial attention among the multi-task paradigm.

**NTIRE2020 Challenge:** In this experiment, we purely adopt data of NTIRE2020 and set the NGDC strategy down, to observe the effectiveness of our exclusionary dual-learning intuitively. As shown in Table V, RWSR-EDL achieves obvious improvement among three evaluation metrics. As Impressionism is the champion method in the NTIRE2020 challenge with high-quality enhancement, our model still performs superior results, which indeed verify the effectiveness of the proposed learning paradigm. Although EDSR enjoys much more parameter number and adopts L1 loss only, RWSR-EDL achieves 0.09 dB gain with PSNR index. Besides, EDSR exhibits a weak performance on perceptual-based evaluation protocol. Moreover, we confirmed the superior perceptual performance of RWSR-EDL by using MOS testing. The proposed model achieves the best result, with a 8.5% better MOS than Impressionism. The qualitative results depicted in Fig. 9 and Fig. 10 verify that RWSR-EDL obtains significant visual quality improvement over Impressionism with realistic texture, clear structure, and fewer artifacts.

**Efficiency Analysis:** Despite achieving superior results on quantitative comparisons, RWSR-EDL still presents a competitive running efficiency. The comparison in Table VI reveals that RWSR-EDL achieves competing efficiency. Compared with ESRGAN, RWSR-EDL exhibits faster running efficiency and significant restoration quality improvement. Compared with DASR, our model performs obvious quantitative promotion with similar time consumption.

**D. Ablation Study**

In this section, we make an ablation study on the proposed components, which are exclusionary dual-learning, patch-discriminator, and NGDC strategy.

**Exclusionary Dual-Learning:** We make a study on the exclusionary dual-learning mechanism to justify the effectiveness of the proposed exclusionary dual-learning strategy. As depicted in Table VII, ‘Single Branch’ shows a similar performance with ‘Dual-Learning w/o Mask,’ which justifies that a simply dual-way structure is useless to RealSR yet bring significant parameter increment. However, a combination of exclusionary

---

**TABLE IV**

| Methods  | PSNR↑ | SSIM↑ | LPIPS↓ |
|----------|-------|-------|--------|
| ZSSR [27] | 22.327 | 0.6022 | 0.630 |
| ESRGAN [23] | 21.382 | 0.5478 | 0.543 |
| CinCGAN [47] | 21.602 | 0.6129 | 0.461 |
| FSSR [29] | 20.820 | 0.5103 | 0.390 |
| Impressionism [31] | 21.021 | 0.5978 | 0.376 |
| DASR [18] | 21.780 | 0.5725 | 0.346 |
| RWSR-EDL | 22.335 | 0.6187 | 0.342 |

**TABLE V**

| Methods  | PSNR↑ | SSIM↑ | LPIPS↓ | MOS↓ |
|----------|-------|-------|--------|------|
| EDSR [14] | 25.31 | 0.6383 | 0.5784 | 2.875 |
| ESRGAN [23] | 19.06 | 0.2423 | 0.7552 | 3.250 |
| ZSSR [27] | 25.13 | 0.6268 | 0.6160 | 2.905 |
| KernelGAN [28] | 18.46 | 0.3826 | 0.7307 | 3.155 |
| FASRGAN [1] | 21.86 | 0.6214 | 0.5499 | 2.740 |
| Impressionism [31] | 24.82 | 0.6619 | 0.2270 | 2.430 |
| RWSR-EDL | 25.40 | 0.6819 | 0.2222 | 2.225 |

---

**TABLE VI**

| Method | ESRGAN | FSSR | DASR | RWSR-EDL |
|--------|--------|------|------|----------|
| Time (frame/s) | 0.7971 | 0.7918 | 0.7465 | 0.7632 |
| Parameter | 16,697,987 | 16,697,987 | 16,697,987 | 16,729,694 |

**TABLE VII**

| Test Set | RealSR [12] | Metric | PSNR↑ | LPIPS↓ | Parameter |
|----------|-------------|--------|-------|--------|-----------|
| Single Branch | 27.691 | 0.255 | 16,708,556 |
| Dual-Learning w/o Mask | 27.688 | 0.254 | 16,719,125 |
| Dual-Learning w/ Mask | 27.775 | 0.250 | 16,729,694 |
soft-mask and dual-learning structure exhibits an obvious improvement over PSNR and LPIPS metrics both. Without obvious parameter increment, ‘Dual-Learning w/ Mask’ surpasses ‘Dual-Learning w/o Mask’ with 0.1 dB improvement and 0.005 LPIPS gain, which clearly justify that the proposed exclusionary soft-mask mechanism is adequate to the dual-way learning.

**NGDC Strategy:** To fully justify the effectiveness of NGDC strategy, we plot the training curves with PSNR and LPIPS index in Fig. 3. Specifically, we conduct the evaluation on the testset of RealSR during the training phrase to plot the curves. As depicted in Fig. 3, ‘Ours(RealSR + NTIRE2020) w/ NGDC’ is converged while the ‘Ours(RealSR + NTIRE2020) w/o NGDC’ still stays at a climbing state, which fully justifies that the proposed NGDC strategy can extract useful noise from the auxiliary dataset and utilize noise stacking to bring superior performance.

In Table IX, we also present a quantitative study on NGDC strategy. As noise sampling strategy give a competitive performance within a single dataset, they fail to demonstrate similar promotion when incorporating auxiliary large-scale dataset as the training data is too miscellaneous. For example, ‘Ours(RealSR + NTIRE2020) w/o NGDC’ shows a lower score on all evaluation metrics than ‘Our(RealSR) w/ Data Augmentation’. To fair comparison, ‘Ours(RealSR + NTIRE2020) w/o NGDC’ and ‘Ours(RealSR + NTIRE2020) w/ NGDC’ both adopt data augmentation strategy as same as ‘Ours(RealSR) w/ Data Augmentation’. To our surprise, our NGDC strategy can significant improve the results when incorporating auxiliary large-scale dataset and noise sampling strategy, which well verify that auxiliary dataset exists a certain portion of negative samples and noise-guidance data collection is necessary.

**Patch Discriminator:** We make an experiment on the discriminator of the adversarial training. As shown in Table VIII, ‘Patch-D’ achieves 27.75 dB and the ‘VGG-128’ obtains 26.67 dB, which justify the effectiveness of patch discriminator in

---

**TABLE VIII**

ABLATION STUDY ON DISCRIMINATOR

| Test Set       | RealSR [12] |
|----------------|-------------|
| Metric         | PSNR↑ SSIM↑  |
| VGG-128        | 27.673 0.8058 0.254 |
| Patch-D        | 27.775 0.8095 0.250 |

**TABLE IX**

ABLATION STUDY ON NOISE-GUIDANCE DATA COLLECTION (NGDC) STRATEGY. WITH AN AUXILIARY DATASET, NGDC CAN CONSISTENTLY PROMOTE PERFORMANCE IN ALL EVALUATION METRICS WITHOUT ANY TRAINING TIME INCREMENT

| Test Set                                           | RealSR [12] |
|----------------------------------------------------|-------------|
| Metric                                             | PSNR↑ SSIM↑ LPIPS↓ |
| Ours(RealSR) w/o Noise Sampling                    | 27.715 0.8078 0.252 |
| Ours(RealSR) w/ Data Augmentation                  | 27.775 0.8095 0.250 |
| Ours(RealSR + NTIRE2020) w/o NGDC                 | 27.557 0.8022 0.264 |
| Ours(RealSR + NTIRE2020) w/ NGDC                  | 27.802 0.8110 0.247 |

---

"0810" image from NTIRE20 test set

"0845" image from NTIRE20 test set

Fig. 9. Super-resolution results on the NTIRE2020 [10] challenge data.

Fig. 10. Super-resolution results on the NTIRE2020 [10] challenge data.
real-world image enhancement. Moreover, patch discriminator also brings a significant improvement on SSIM and LPIPS. To this end, we replace VGG-128 with patch discriminator as default discriminator in the adversarial optimization.

**Loss Function:** We compare different settings of loss on the RealSR dataset. As depicted in Table X, our model achieves promising results on PSNR by incorporating $L_{\text{per}}$ and $L_{\text{pix}}$. Compared with the typical weighted sum loss $L_{\text{all}}$ in (7), our method achieves 1.68 dB and 0.025 gain with PSNR and LPIPS index. However, the proposed model shows a weaker PSNR score when incorporating the adversarial loss function. To our surprise, our full model achieves further improvement as the adversarial loss is incorporated, which justify that $L_{\text{per}}$ is important in RealSR issue. Also, our full model achieves the best score among Table X verify the effectiveness of the proposed exclusionary dual-learning in L1- and perceptual- based cooperative optimization.

**V. CONCLUSION AND FUTURE WORK**

In this work, we revival exclusionary dual-learning to facilitate deep representation that exhibits more diversity under L1- and perceptual- minimization. With the novel exclusionary dual-learning mechanism, our RWSR-EDL demonstrates a charming result on real-world image super-resolution. Moreover, we present a noise-guidance data collection strategy, which yields an efficient training paradigm in multiple datasets learning. Experimental results show that RWSR-EDL surpasses state-of-the-art real-world image super-resolution methods with a clear margin and closed parameter number on perceptual- and euclidean- based evaluation protocols.

We aim at extending our work by following directions. First, we would like to extend the proposed NGDC strategy to a real-world image denoising task, aim at collecting more noise samples with a given image and demonstrate efficient denoising training. Second, we are considering involving explainable deep learning into our framework to further improve the exclusionary dual-learning mechanism.

**REFERENCES**

[1] Y. Yan et al., “Fine-grained attention and feature-sharing generative adversarial networks for single image super-resolution,” *IEEE Trans. Multimedia*, vol. 24, pp. 1473–1487, 2022.

[2] K. Jiang et al., “Dual-path deep fusion network for face image hallucination,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 33, no. 1, pp. 378–391, Jan. 2022.

[3] C. Tian et al., “Coarse-to-fine CNN for image super-resolution,” *IEEE Trans. Multimedia*, vol. 23, pp. 1489–1502, 2021.

[4] M. Li, Z. Zhang, J. Yu, and C. W. Chen, “Learning face image super-resolution through facial semantic attribute transformation and self-attentive structure enhancement,” *IEEE Trans. Multimedia*, vol. 23, pp. 468–483, 2021.

[5] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, “Image quality assessment: From error visibility to structural similarity,” *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[6] M. Lucic, K. Kurach, M. Michalski, S. Gelly, and O. Bousquet, “Are GANs created equal? A large-scale study,” in *Proc. Adv. Neural Inf. Process. Syst.*, 2018, pp. 700–709.

[7] R. Zhang, P. Isola, A. A. Efros, E. Shechtman, and O. Wang, “The unreasonable effectiveness of deep features as a perceptual metric,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 586–595.

[8] K.-Y. Lin and G. Wang, “Hallucination-IPA: No-reference image quality assessment via adversarial learning,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 732–741.

[9] A. Lugmayr et al., “Aim 2019 challenge on real-world image super-resolution: Methods and results,” in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, 2019, pp. 3575–3583.

[10] A. Lugmayr et al., “NTIRE 2020 challenge on real-world image super-resolution: Methods and results,” in *Proc. Adv. Neural Inf. Process. Syst. Workshops*, 2020, pp. 2058–2076.

[11] C. Chen, Z. Xiong, X. Tian, Z. Zha, and F. Wu, “Camera lens super-resolution,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 1652–1660.

[12] J. Cai, H. Zeng, H. Yong, Z. Cao, and L. Zhang, “Toward real-world single image super-resolution: A new benchmark and a new model,” in *Proc. IEEE Int. Conf. Comput. Vis.*, 2019, pp. 3086–3095.

[13] Y. Mei et al., “Image super-resolution using very deep residual channel attention networks,” in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 286–301.

[14] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, “Residual dense network for image restoration,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 43, no. 7, pp. 2480–2495, Jul. 2020.

[15] R. Timofte, E. Agustsson, L. Van Gool, M.-H. Yang, and L. Zhang, “NTIRE 2017 challenge on single image super-resolution: Methods and results,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, 2017, pp. 114–125.

[16] Y. Wei et al., “Unsupervised real-world image super-resolution via domain-distance aware training,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2020, pp. 13385–13394.

[17] C. Dong, C. C. Loy, K. He, and X. Tang, “Learning a deep convolutional network for image super-resolution,” in *Proc. Eur. Conf. Comput. Vis.*, 2014, pp. 184–199.

[18] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2016, pp. 770–778.

[19] Y. Luo et al., “EBSR: Feature enhanced burst super-resolution with deformable alignment,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2021, pp. 471–478.

[20] Y. Zhang et al., “Image super-resolution using very deep residual channel attention networks,” in *Proc. Eur. Conf. Comput. Vis.*, 2018, pp. 286–301.

[21] Z. Luo et al., “ESRGAN: Enhanced super-resolution generative adversarial networks,” in *Proc. Adv. Neural Inf. Process. Syst. Workshops*, 2018, pp. 63–79.

[22] X. Hu et al., “Meta-USR: A unified super-resolution network for multiple degradation parameters,” *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 9, pp. 4151–4165, Sep. 2021.

[23] Y. Mei et al., “Image super-resolution with cross-scale non-local attention and exhaustive self-exemplars mining,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 3118–3126.

[24] K. Jiang, Z. Wang, P. Yi, and J. Jiang, “Hierarchical dense recursive network for image super-resolution,” *Pattern Recognit.*, vol. 107, 2020, Art. no. 107475.

[25] A. Shocher, N. Cohen, and M. Irani, “Zero-shot super-resolution using deep internal learning,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2018, pp. 3118–3126.

[26] S. Bell-Kligler, A. Shocher, and M. Irani, “Blind super-resolution kernel estimation using an internal-GAN,” in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 3086–3095.
[29] M. Fritsche, S. Gu, and R. Timofte, “Frequency separation for real-world super-resolution,” in Proc. IEEE Int. Conf. Comput. Vis. Workshops, 2019, pp. 3599–3608.

[30] Y. Pang et al., “FAN: Frequency aggregation network for real image super-resolution,” in Proc. Eur. Conf. Comput. Vis., 2020, pp. 468–483.

[31] X. Ji et al., “Real-world super-resolution via kernel estimation and noise injection,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops, 2020, pp. 1914–1923.

[32] L. Liu, B. Liu, H. Huang, and A. C. Bovik, “No-reference image quality assessment based on spatial and spectral entropies,” Signal Process.: Image Commun., vol. 29, no. 8, pp. 856–863, 2014.

[33] A. Mittal, R. Soundararajan, and A. C. Bovik, “Making a “completely blind” image quality analyzer,” IEEE Signal Process. Lett., vol. 20, no. 3, pp. 209–212, Mar. 2013.

[34] Y. Blau, R. Mechrez, R. Timofte, T. Michaeli, and L. Zelnik-Manor, “The 2018 PRIM challenge on perceptual image super-resolution,” in Proc. Eur. Conf. Comput. Vis. Workshops, 2018, pp. 334–355.

[35] C. Ma, C.-Y. Yang, X. Yang, and M.-H. Yang, “Learning a no-reference quality metric for single-image super-resolution,” Comput. Vis. Image Understanding, vol. 158, pp. 1–16, 2017.

[36] M. Lebrun, M. Colom, and J. Morel, “Multiscale image blind denoising,” IEEE Trans. Image Process., vol. 24, no. 10, pp. 3149–3161, Oct. 2015.

[37] M. Lebrun, A. Buades, and J. M. Morel, “A nonlocal Bayesian image denoising algorithm,” SIAM J. Imag. Sci., vol. 6, no. 3, pp. 1665–1685, 2013.

[38] K. Zhang, W. Zuo, and L. Zhang, “FFDNet: Toward a fast and flexible solution for CNN-based image denoising,” IEEE Trans. Image Process., vol. 27, no. 9, pp. 4608–4622, Sep. 2018.

[39] S. Guo, Z. Yan, K. Zhang, W. Zuo, and L. Zhang, “Toward convolutional blind denoising of real photographs,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2019, pp. 1712–1722.

[40] J. Chen, J. Chen, H. Chao, and M. Yang, “Image blind denoising with generative adversarial network based noise modeling,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2018, pp. 3155–3164.

[41] S. Anwar and N. Barnes, “Real image denoising with feature attention,” in Proc. IEEE Int. Conf. Comput. Vis., 2019, pp. 3155–3164.

[42] K. Simonyan and A. Zisserman, “Very deep convolutional networks for discriminative localization,” in Proc. Eur. Conf. Comput. Vis., 2016, pp. 2921–2929.

[43] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” in Proc. Int. Conf. Learn. Representations, 2015, arXiv:1412.6980.

[44] Y. Yuan et al., “Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops, 2018, pp. 701–710.

[45] J. Chen, J. Chen, H. Chao, and M. Yang, “Image blind denoising with generative adversarial network based noise modeling,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., 2018, pp. 3155–3164.

[46] S. Anwar and N. Barnes, “Real image denoising with feature attention,” in Proc. IEEE Int. Conf. Comput. Vis., 2019, pp. 3155–3164.

[47] K. Simonyan and A. Zisserman, “Very deep convolutional networks for discriminative localization,” in Proc. Eur. Conf. Comput. Vis., 2016, pp. 2921–2929.

[48] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” in Proc. Int. Conf. Learn. Representations, 2015, arXiv:1412.6980.

[49] Y. Yuan et al., “Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops, 2018, pp. 701–710.

Hao Li received the B.E. degree from the School of Information Technology, Guangdong University of Technology, Guangzhou, China, where he is currently working toward the M.S. degree. His research interests include computer vision and machine learning.

Jinghui Qin received the B.E. and Ph.D. degree from the School of Data and Computer Science, Sun Yat-Sen University, Guangzhou, China. He is currently a Research Scientist with Sun Yat-sen University. His research interests include computer vision and machine learning.

Zhijing Yang received the B.S. and Ph.D. degrees from the Mathematics and Computing Science, Sun Yat-Sen University, Guangzhou China, in 2005 and 2008, respectively. He was a Visiting Research Scholar with the School of Computing, Informatics, and Media, University of Bradford, Bradford, U.K., between July and December 2009, and a Research Fellow with the School of Engineering, University of Lincoln, Lincoln, U.K., between January 2011 and January 2013. He is currently a Professor with the School of Information Engineering, Guangdong University of Technology, Guangzhou, China. He has authored or coauthored more than 60 peer-review journal and conference papers. His research interests include time-frequency analysis, image retrieval, machine learning, and pattern recognition.

Pengxu Wei received the B.S. degree in computer science and technology from the China University of Mining and Technology, Beijing, China, in 2011, and the Ph.D. degree from the School of Electronic, Electrical, and Communication Engineering, University of Chinese Academy of Sciences, Beijing, China, in 2018. She is currently a Research Scientist with Sun Yat-Sen University, Guangzhou, China. Her research interests include computer vision and machine learning.

Jinshan Pan (Member, IEEE) received the Ph.D. degree in computational mathematics from the Dalian University of Technology, Dalian, China, in 2017. He is currently a Professor with the School of Computer Science and Engineering, Nanjing University of Science and Technology, Nanjing, China. His research interests include image deblurring, image/video analysis and enhancement, and related vision problems.

Liang Lin (Senior Member, IEEE) is currently a Full Professor with Sun Yat-sen University, Guangzhou, China. He is the Excellent Young Scientist of the National Natural Science Foundation of China. He has authored or coauthored more than 100 papers in top-tier academic journals and conferences. He has been an Associate Editor for the IEEE TRANSACTION HUMAN-MACHINE SYSTEMS, The Visual Computer, and Neurocomputing. He was the Area/Session Chair of numerous conferences, including CVPR, ACCV, and ICMR. He was the recipient of the Best Paper Runners-Up Award at ACM NPAR 2010, Google Faculty Award in 2012, Best Paper Diamond Award at IEEE ICME 2017, Annual Best Paper Award at Pattern Recognition 2018, and Best Paper Nomination at ICCV 2019. He is a fellow of IET.

Yukai Shi received the Ph.D. degrees from the School of Data and Computer Science, Sun Yat-sen University, Guangzhou, China, in 2019. He is currently a Lecturer with the School of Information Engineering, Guangdong University of Technology, Guangzhou, China. His research interests include computer vision and machine learning.