Joint Demosaicing and Super-Resolution (JDSR): Network Design and Perceptual Optimization

Xuan Xu, Yanfang Ye, and Xin Li, *Fellow, IEEE*

***Abstract*—Image demosaicing and super-resolution are two important tasks in color imaging pipeline. So far they have been mostly independently studied in the open literature of deep learning; little is known about the potential benefit of formulating a joint demosaicing and super-resolution (JDSR) problem. In this paper, we propose an end-to-end optimization solution to the JDSR problem and demonstrate its practical significance in computational imaging. Our technical contributions are mainly two-fold. On network design, we have developed a Residual-Dense Squeeze-and-Excitation Networks (RDSEN) supported by a pre-demosaicing network (PDNet) as the pre-processing step. We address the issue of *spatio-spectral attention* for color-filter-array (CFA) data and discuss how to achieve better information flow by concatenating Residue-Dense Squeeze-and-Excitation Blocks (RDSEBs) for JDSR. Experimental results have shown that significant PSNR/SSIM gain can be achieved by RDSEN over previous network architectures including state-of-the-art RCAN. On perceptual optimization, we propose to leverage the latest ideas including *relativistic discriminator* and pre-excitation perceptual loss function to further improve the visual quality of textured regions in reconstructed images. Our extensive experiment results have shown that Texture-enhanced Relativistic average Generative Adversarial Network (TRaGAN) can produce both subjectively more pleasant images and objectively lower perceptual distortion scores than standard GAN for JDSR. Finally, we have verified the benefit of JDSR to high-quality image reconstruction from real-world Bayer pattern data collected by NASA Mars Curiosity.**

***Index Terms*—Color imaging, Joint image demosaicing and super-resolution (JDSR), residual-dense squeeze-and-excitation network (RDSEN), perceptual optimization.**

I. Introduction

IMAGE demosaicing and single image super-resolution (SISR) are two important image processing tasks to the pipeline of color imaging. Demosaicing is a necessary step to reconstruct full-resolution color images from so-called Color filter Array (CFA) such as Bayer pattern. SISR is a cost-effective alternative to more expensive hardware-based solution (i.e., optical zoom). Both problems have been extensively yet separately studied in the literature - from model-based methods [1], [2], [3], [4], [5], [6], [7], [8], [9] to learning-based approaches [10], [11], [12], [13], [14], [15], [16], [17], [18]. Treating demosaicing and SISR as two independent problems may generate undesirable edge blurring as shown in Fig. 1. Moreover, the processes of demosaicing and SISR can be integrated and optimized together from a practical application point of view (e.g., digital zoom for smartphone cameras such as iPhone 11 pro max, Google Pixel 4 and Huawei P30).

X. Xu and X. Li are with the Lane Department of Computer Science and Electrical Engineering, West Virginia University, Morgantown, WV, 26505 USA. Yanfang Ye is with Department of Computer and Data Sciences, Case Western Reserve University, Cleveland, OH 44106. E-mail: xuxu@mix.wvu.edu, xin.li@mail.wvu.edu, yanfang.ye@case.edu

HR DemoNet+RCAN (4x) RDSEN (4x)

Fig. 1. Comparison of JDSR output to separately demosaic-super-resovle output. Left to right: a) HR image (ground-truth); b) 4× upscaling output by concatenating state-of-art demosaicing method DemoNet [19] with SISR method RCAN [18] (separated approach); c) 4× upscaling output of our proposed RDSEN networks (joint approach).

Inspired by the success of joint demosaicing and denoising [19], we propose to study the problem of joint image demosaicing and super-resolution (JDSR) in this paper and develop a principled solution leveraging latest advances in deep learning to computational imaging. We argue that the newly formulated JDSR problem has high practical impact (e.g., to support the mission of NASA Mars Curiosity and smartphone applications). The problem of JDSR is intellectually appealing but has been under-researched so far. The only existing work we can find in the open literature is a recently published paper [20] which contained a straightforward application of ResNet [21] and considered the scaling ratio of two only. As demonstrated in Fig. 1, our optimized solution to JDSR can achieve significantly better visual quality than the brute-force approach.

The motivation behind our approach is mainly two-fold. On one hand, rapid advances in deep residual learning have offered a rich set of tools for image demosaicing and SISR. For example, DenseNet [22] has been adapted to fully exploit hierarchical features for the problem of SR in SRDenseNet [23] and residual dense network (RDN) [17]; residual channel attention network (RCAN) [18] allows us to develop much deeper networks (over 400 layers) with squeeze-and-excitation (SE) blocks [24] than previous works (e.g., [14], [25]). Inspired by RDN and RCAN, our previous [26] presented a spatial color attention mechanism (SCAN) to further improve the SISR performance on real world SR dataset. However, to the best of our knowledge, the issue of *spatio-spectral attention* mechanism has not been explicitly addressed for raw CFA data

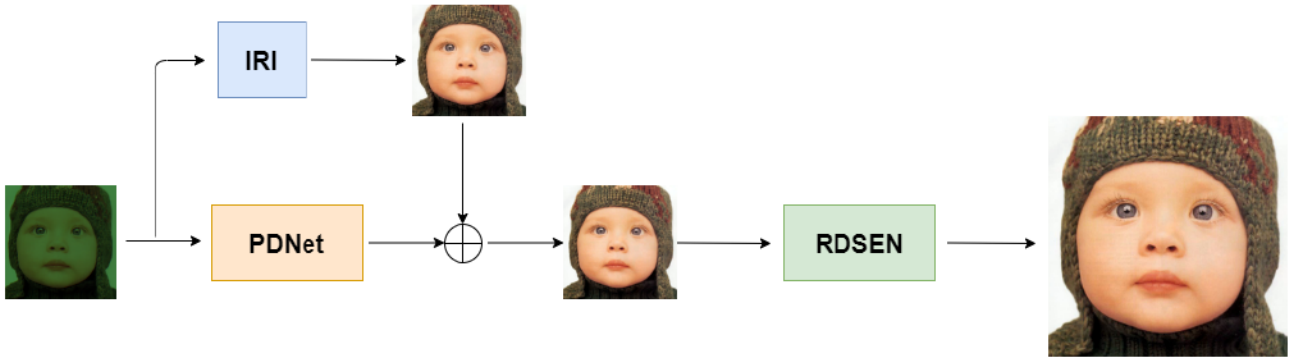

Fig. 2. Overview of proposed RDSEN with PDNet network architecture, $\oplus$ means element-wise sum.

in the open literature. How to design a network architecture for *jointly* exploiting spatial and spectral dependency in Bayer patterns deserves a systematic study.

On the other hand, we propose to optimize the *perceptual quality* for JDSR because that is what really matters in various real-world applications (e.g., to support the mission of NASA to Mars). Generative adversarial network (GAN) [27] is arguably the most popular approach toward perceptual optimization and has demonstrated convincing improvement for SISR in SRGAN [15]. It has also been widely observed that the training of GAN suffers from stability issue which could have catastrophic impact on reconstructed images. There has been a flurry of latest works (e.g., Relativistic average GAN (RaGAN) [28], enhanced SRGAN (ESRGAN) [16] and perception-enhances SR (PESR) [29]) showing the potential of relativistic discriminator in stabilizing GAN and improving visual quality of SISR images. However, the issue of perceptual optimization has not been addressed in previous works on joint demosaicing-and-denoising (JDD) at all [19], [30], [31]. For the first time, we aim at studying the potential of GAN-based models on perceptual optimization for the JDSR problem, which has practical significance when no ground-truth (reference image) is available.

Overall, our contributions are summarized as follows:

- Network design: we propose a concatenation of pre-demosaicing network (PDNet) and Residual-Dense Squeeze-and-Excitation Networks (RDSEN) for JDSR. The former takes a model-based demosaicing result via iterative-residual interpolation (IRI) [32] as the surrogate target to facilitate deep residue learning for pre-demosaicing. Then a novel concatenation of Residual-Dense Squeeze-and-Excitation Block (RDSEB) modules is designed to facilitate information flow between the intermediate demosaicing result and the final reconstruction. Through the combination of long and short skip connections, we manage to train RDSEN more efficiently than existing RCAN while still achieving better performance.
- Perceptual optimization: we have leveraged the latest advance RaGAN [28] from SISR to JDSR and studied the choices of perceptual loss function for JDSR. In addition to improved stability, we have found that Texture-enhanced RaGAN (TRaGAN) with a before-activation perceptual loss function can produce visually more pleasant results. We argue that the issue of perceptual optimization is particularly important for JDSR because it has been largely overlooked in the existing literature of JDD.
- Simulation study and real-world application: We have conducted extensive stimulation study to demonstrate the superiority of our network to other competing approaches. When compared against the current state-of-the-art RCAN [18], our RDSEN has achieved significant improvement on both objective (up to 1.2dB in terms of PSNR on McM dataset) and subjective qualities. We have also applied the proposed RDSEN+TRaGAN solution to raw Bayer pattern data collected by the Mast Camera (Mastcam) of NASA Mars Curiosity Rover. Our experimental results have shown visually superior high-resolution image reconstruction can be achieved at the scaling ratio as large as 4.

II. Related Works

Both image demosaicing and super-resolution have been studied in decades in the open literature. In this section, we review image demosaicing and image super-resolution approaches separately and focus on deep learning based methods.

A. Image Demosaicing

Existing approaches toward image demosaicing can be classified into two categories: model-based methods [1], [2], [3], [4] and learning-based methods [10], [11], [13]. Model-based approaches rely on hand-crafted parametric models which often suffer from lacking of the generalization capability to handle varying characteristics in color images (i.e., the potential model-data mismatch). Recently, deep learning methods show the advantages in image demosaicing field. Inspired by single image super-resolution model SRCNN [33], DMCNN [34] utilized super-resolution based CNN model and ResNet [21] to investigate image demosacing problem. CDM-CNN [35] introduced to apply residual learning [21] with a two-phase network architecture which firstly recovers green channel as guidance prior and then uses this guidance prior to reconstruct the RGB channels. Besides to explore image demosacing methods only, there are several works studying joint image demosaicing and denoising (JDD) problem. Dong *et.al.* [36] developed a deep neural network with generative adversarial networks (GAN) [27] and perceptual loss functions to solve JDD problems. Inspired by classical image regularization and majorization-minimization optimization, Kokkinos

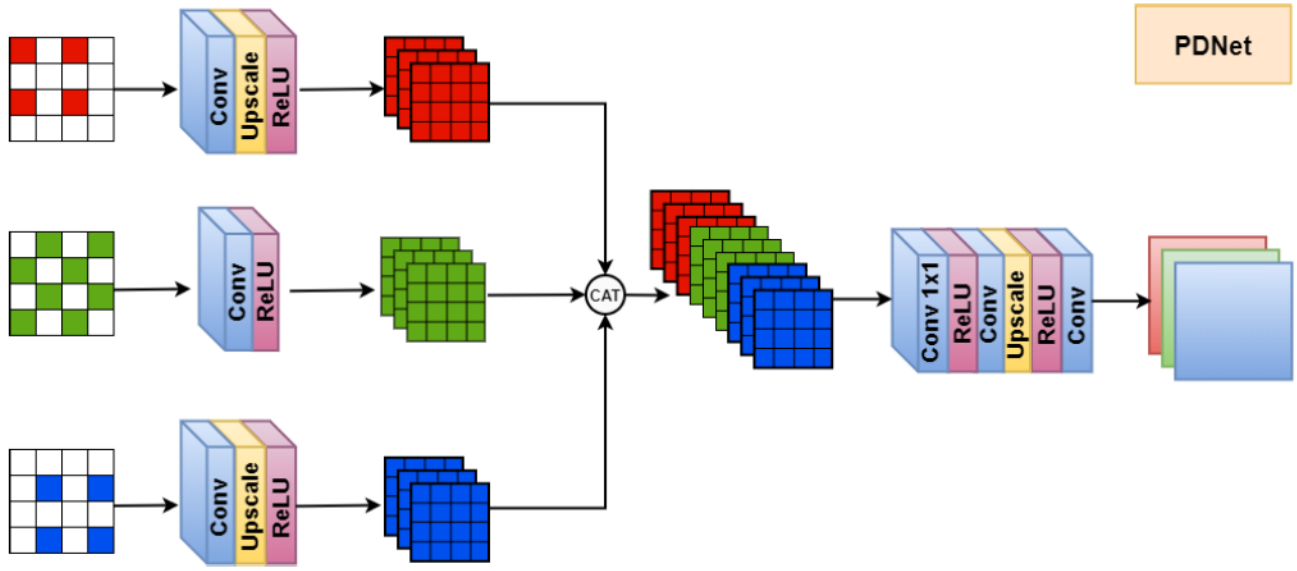

Fig. 3. Structure of PDNet, 'CAT' is feature concatenation.

and Lefkimmiatis [12] proposed a deep neural network to solve JDD problem. Deep learning based image demosaicing techniques have shown convincingly improved performance over model-based ones on several widely-used benchmark dataset (e.g., Kodak and McMaster [37]). However, the issue of suppressing spatio-spectral aliasing has not been addressed in the open literature as far as we know.

B. Image Super-resolution

Model-based approaches towards SISR [5], [6], [7], [8], [9] suffer from notorious aliasing artifacts and edge blurring. Recently, deep learning-based approaches have advanced rapidly. SRCNN [33] first introduced deep learning based method to solve single image super-solution task with three convolutional layers and achieved much better performance than model based methods. Benefit by concept of ResNet [21], VDSR [14] firstly trained 20 layers deep networks with long residual connection which can only learn more high-frequency information and increase the convergence speed. EDSR [25] proposed to integrate several resblocks and remove batch-normalization layer, which can save GPU memory, stack more layers and make networks wider [38], to further improve SISR performance. LapSRN [39] proposed to super-resolve LR image several times to save GPU memory and achieve better performance.

Most recent advances include SRDenseNet [23] which applied denseNet [22] to solve SISR task, RDN [17] which utilized ResNet and DenseNet to create residual dense block (RDB). Through local feature fusion, the proposed RDB can allow larger growth rate to boost the performance. RCAN [18] first introduced attention mechanism inspired by SENet [24] to calibrate feature maps and proposed residual in residual structure to achieve a very deep convolutional networks which achieved new state-of-art performance for SISR task. Besides objective measures such as PSNR/SSIM [40], SRGAN [15] introduced a novel generative adversarial networks (GAN) [27] based architecture to optimize the perceptual quality of SR images, benefit by GAN, SRGAN can reconstruct more textures from low-res images. An enhanced version of SRGAN named ESRGAN [16] using relativistic average GAN (RaGAN) was developed in [28] as well as [29] which can recover more realistic super-resolved image compared with SRGAN. By contrast, the problem of JDSR has been under-researched so far with the only exceptions of [41], [20], and [42].

III. Network Design: Residual-Dense Attention

The hierarchy of our network design goes like: Overview of proposed network (Fig. 2) → PDNet subnetwork (Fig. 3) → RDSEN (Fig. 4) → RDSEB with channel attention (Fig. 5).

A. Pre-demosaicing Network

One challenging issue in JDSR is that not only high-frequency components but also two-third of color pixels are missing. This issue can lead undesirable distortion or artifact in the reconstructed full-resolution color image. Inspired by recent work CBDNet [43], we have designed a pre-demosaicing network (PDNet) for initially demosaicing the Bayer pattern as a pre-processing step to reduce the gap between LR CFA data and HR color image. As shown in Fig. 2 before the RDSEN module, we have adopted a model-based demosaicing method called iterative-residual interpolation (IRI) [32] to generate an intermediate demosaicing result, which will be used as the input to the refinement module. This intermediate demosaicing results will be refined by PDNet as shown in Fig. 3 (conceptually similar to ResNet [21]). In the PDNet, we opt to separately process Red, Green, and Blue channels. For Red and Blue channel, we use a convolution layer with stride of 4 to shrink the corresponding Bayer pattern and then upscale them with a factor of 2; for Green channel we shrink it by a convolution layer with stride of 2. This is because the Red and Blue channels each contains one-fourth information and G

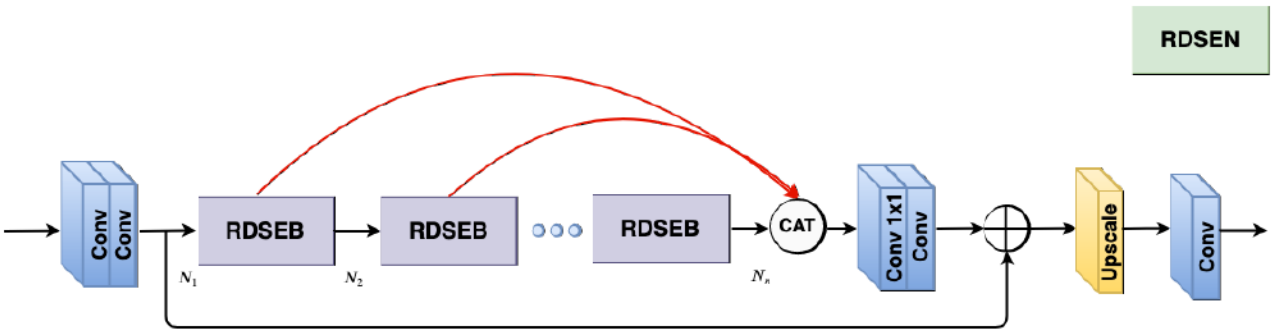

Fig. 4. Structure of RDSEN, 'CAT' is feature concatenation and $\oplus$ denotes element-wise sum respectively.

channel contains one-half information. Then we concatenate RGB feature maps and fused them with a 1×1 kernel of convolution layer. Finally we upscale the fused feature maps back to the same size as input CFA data.

B. Residual-Dense Squeeze-and-Excitation Network

Channel attention mechanism has been successfully applied to both high-level (e.g., SENet [24] and LS-CNN [44]) and low-level (e.g., RCAN [18]) vision tasks. A channel attention module first squeezes the input feature map and then activates one-time reduction-and-expansion to excite the squeezed feature map. Such strategy is not optimal for recovering missing high-frequency information in SISR when the network is very deep; meanwhile, JDSR problem requires simultaneous recovery of incomplete color information across Red, Green, Blue channels, which requires extra attention toward the dependency in the spectral domain. How to generalize the channel attention mechanism from spatial-only to joint *spatio-spectral* has remained one of open problems in the design of attention-based networks.

As discussed in [18], high-frequency components often correspond to regions in an image such as textures, edges, corners and so on. Conventional layers have limited capability of exploiting contextual information outside the *local* receptive field especially due to the missing data in Bayer patterns. To overcome this difficulty, we propose to design a new Residual-Dense Squeeze-and-Excitation Network (RDSEN) as shown in Fig. 4 and Fig. 5. The proposed RDSEN is designed to implement a *deeper* and *wider* spatio-spectral channel attention mechanism for the purpose of more effectively suppressing spatio-spectral aliasing in LR Bayer patterns.

Unlike SENet [24] and RCAN [18] (using residual block to stack with channel attention module), RDSEN based on multiple RDSEB blocks attemps to fuse both *local and global* residual-dense attention information to assure more faithful information recovery when the network gets deeper and wider. As shown in Figs. 4 and 5, we have kept both long skip and short skip connections like RCAN in order to make the overall training stable and facilitate the information flow both inside and outside the RDSEN module. Although similar idea of local feature fusion existed in residual dense block of RDN [17], our *hybrid* design - i.e., the RDSEB block combining the ideas from RDN and RCAN - is novel because it represents an alternative approach to strike an improved tradeoff between cost (in terms of network parameters) and performance (in terms of visual quality).

Our design of concatenating RDSEB modules also has its merit from the perspective of exploiting joint spatio-spectral attention for JDSR. Spatio-spectral channel attention mechanism in the proposed RDSEB module can help to recalibrate input features via channel statistics [24] across different spectral bands. In SISR, residual-in-residual attention or dense connection operation might be sufficient for capturing channel-wise dependencies for LR color images; however our JDSR task aims at recovering two-third of missing data in spectral bands in addition to the missing high-frequency information. We have experimentally verified that such design of deeper and wider networks [38] based on concatenation of multiple RDSEB modules indeed helps the boosting of our JDSR performance.

C. Residual-Dense Squeeze-and-Excitation Block

The key to deeper and wider networks lies in the design of RDSEB module - i.e., how to use short skip connection and multiple concatenations after channel attention mechanism to assure faithful information recovery both inside and outside RDSEB modules? As shown in Fig. 5, we propose a Residual-Dense Squeeze-and-Excitation Block (RDSEB) in which the channel size can be expanded *step by step* (see Fig. 5). The key advantages of this newly designed RDSEB include: 1) the reduced channel descriptor can be smoothly activated *multiple times* and therefore more faithful information across spatio-spectral domain is accumulated; 2) dense-connection can increase the network *depth and width* without running into the notorious vanishing-gradient problem [45]; 3) both information flow and network stability, which are important to a principled solution to JDSR, can be jointly improved by introducing dense connections to SE residual blocks (so we can train even deeper than RCAN [18]).

More specifically, to implement the CA block, we first apply global average pooling to *squeeze* input feature maps. Let us denote the input feature maps by $\mathbf{U} = [u_1, u_2, ..., u_C]$, which contains C feature maps with the dimension of $H \times W$. Then the global average pooling output $z \in \mathbb{R}^C$ can be calculated

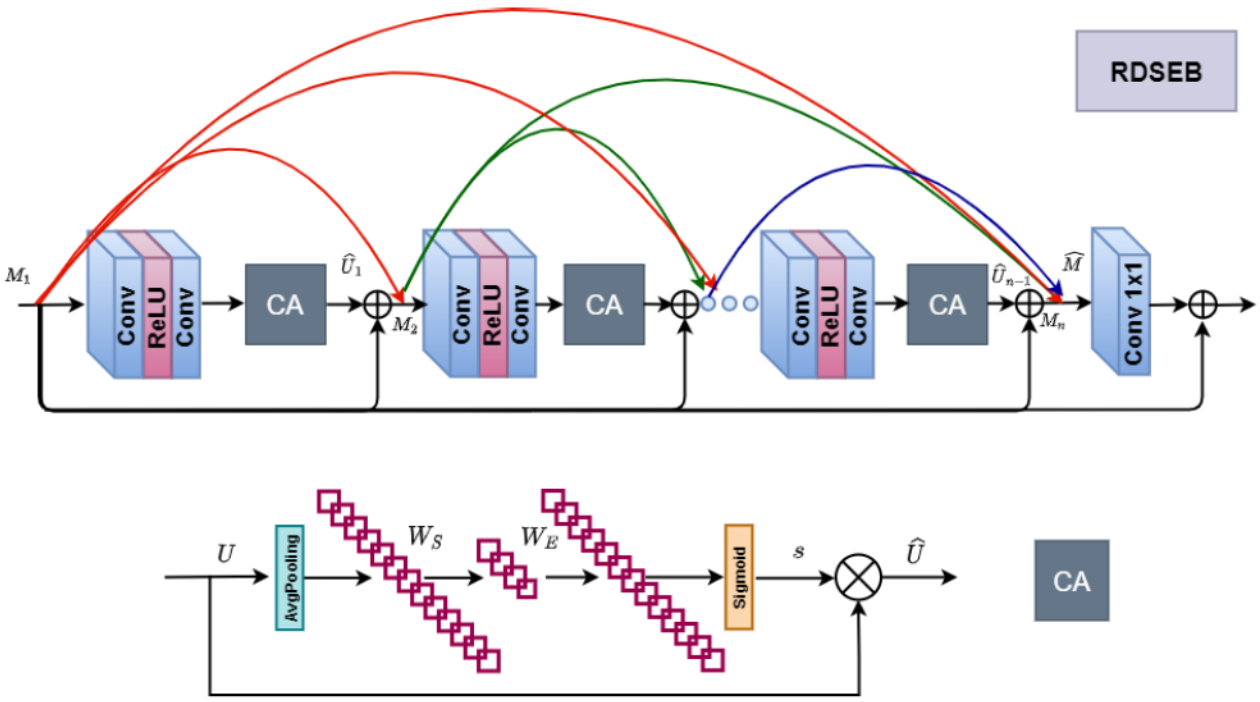

Fig. 5. Flowchart of Residual-Dense Squeeze-and-Excitation Block (RDSEB) and Channel Attention (CA) module ($\otimes$ denotes element-wise product).

by:

$$z_C = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} \mathbf{u}_C(i,j) \tag{1}$$

where z_C is the c-th element of z, $u_C(i,j)$ is the pixel value of the c-th feature at position (i,j) from input feature maps. Then we propose to implement a simple gating mechanism as adopted by previous works including SENet [24] and RCAN [18]:

$$s = \sigma(\mathbf{W}_E(\delta(\mathbf{W}_S(z)))) \tag{2}$$

where σ refers to a sigmoid function, δ denotes the ReLU function, both $\mathbf{W}_S$ and $\mathbf{W}_E$ are Conv layers with weights $\mathbf{W}_S \in \mathbb{R}^{1\times1\times\frac{C}{r}}$ and $\mathbf{W}_E \in \mathbb{R}^{1\times1\times C}$, r is the scaling ratio to reduce the dimension of z (details about this hyperparameter controlling the tradeoff between the capacity and the complexity can be found in SENet [24]).

In order to achieve deeper and wider channel attention, we propose a novel strategy of connecting each output of channel attention block not only with short skip connection (residual) but also dense-connection as shown in Fig. 5. To formalize this problem, define U as the input feature map to CA module, the rescaled input feature map $\hat{\mathbf{U}}$ can be expressed as:

$$\hat{\mathbf{U}} = s \cdot \mathrm{U} \tag{3}$$

where '$\cdot$' stands for element-wise product.

Finally, to implement dense-connection, we define M_1 as the input feature map of RDSEB block. Then the output feature map $\hat{M}$ can be written as the following equations:

$$M_i = \hat{\mathbf{U}}_{i-1} + M_1, \text{ where } i \in [2,n] \tag{4}$$

$$\hat{M} = [M_1, M_2, ..., M_n] \tag{5}$$

where $[M_1, M_2 ... M_n]$ refers to the concatenation of feature maps, $\hat{\mathbf{U}}_{i-1}$ is the corresponding output of CA module at the $i-1$-th stage as shown in Fig. 5. With the new RDSEB block, we can train a deeper and wider network thanks to the improved information flow.

IV. Perceptual Optimization: Relativistic Discriminator and Loss Function

A. Texture-enhanced Relativistic average GAN (TRaGAN)

The discriminator D in standard GAN [27] only estimates the probabilities of real/fake images, and the interaction between generator and discriminator is interpreted as a two-player minimax game. It can be expressed as $D(x) = \sigma(C(x))$, where σ is sigmoid function, $C(x)$ is non-transformed layer, x is the input image. Such idea has been successfully applied to the problem of SISR such as SRGAN [15] in which the super-resolved image (fake version) is compared against the ground-truth (real version). In other words, discriminator D serves as a judge for perceptual optimization of generator.

Unlike standard GAN, relativistic average GAN (RaGAN) [28] can make the discriminator D to estimate the probability based on both real and fake images, making a real image more realistic than a fake one (on the average). According to [28], RaGAN can not only generate more realistic images but also stabilize the training progress. Recently, the benefit of RaGAN over conventional GAN has been demonstrated for SISR in

Method	Scale	Set5	Set14	B100	Manga109	McM	PhotoCD
		PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
FlexIPS[46]+RCAN[18]	x2	35.18/0.9387	31.24/0.8776	31.00/0.8647	30.32/0.9199	34.80/0.9301	43.02/0.9610
DemoNet[19]+RCAN[18]	x2	35.92/0.9458	32.27/0.8971	31.38/0.8823	35.50/0.9590	35.34/0.9362	43.53/0.9642
RDSR[20]	x2	36.29/0.9485	32.56/0.9008	31.56/0.8850	36.14/0.9625	35.90/0.9423	43.74/0.9655
RCAN [18]	x2	36.54/0.9499	32.74/0.9032	31.68/0.8878	36.65/0.9643	36.18/0.9445	43.91/0.9661
RDSEN (ours)	x2	**37.40/0.9575**	**32.91/0.9128**	**32.00/0.8972**	**36.86/0.9716**	**37.38/0.9565**	**44.70/0.9716**
FlexISP+RCAN	x3	31.21/0.8731	28.55/0.7884	27.31/0.7310	27.58/0.8647	31.25/0.8661	40.32/0.9402
DemoNet+RCAN	x3	32.16/0.9030	29.24/0.8137	28.42/0.7801	30.75/0.9112	31.65/0.8739	40.74/0.9445
RDSR	x3	33.05/0.9103	29.54/0.8211	28.61/0.7859	31.69/0.9225	32.21/0.8842	40.90/0.9458
RCAN	x3	33.24/0.9125	29.67/0.8241	28.69/0.7882	32.06/0.9267	32.42/0.8874	41.11/0.9469
RDSEN (ours)	x3	**33.75/0.9218**	**29.91/0.8337**	**28.84/0.7993**	**32.14/0.9330**	**33.21/0.9032**	**41.60/0.9521**
FlexISP+RCAN	x4	29.57/0.8376	26.94/0.7177	26.68/0.6896	26.69/0.8427	27.78/0.7651	38.28/0.9201
DemoNet+RCAN	x4	30.33/0.8596	27.58/0.7488	26.94/0.7081	27.81/0.8590	29.49/0.8187	38.67/0.9243
RDSR	x4	30.87/0.8712	27.91/0.7589	27.16/0.7151	28.86/0.8800	30.10/0.8328	38.81/0.9258
RCAN	x4	31.04/0.8746	27.98/0.7613	27.20/0.7175	29.12/0.8856	30.24/0.8367	39.01/0.9271
RDSEN (ours)	x4	**31.63/0.8863**	**28.26/0.7725**	**27.39/0.7284**	**29.28/0.8903**	**30.74/0.8523**	**39.41/0.9317**

TABLE I
PSNR/SSIM COMPARISON AMONG DIFFERENT COMPETING METHODS. **BOLD** FONT INDICATES THE BEST RESULT AND UNDERLINE THE SECOND BEST.

[16] and [29]. Here we propose to leverage the idea of RaGAN to JDSR and demonstrate how relativistic discriminator can work with the proposed RDSEN (generator) for the purpose of perceptual optimization when no ground-truth (reference image) is available. Note that this issue has been overlooked in the literature of not only SISR (e.g, RDN [17], RCAN [18]) but also JDD (e.g., [19], [30], [31]).

To implement RaGAN, we represent the real and fake images by x_r and x_f respectively; then we can formulate the output of a modified discriminator $\hat{D}$ for RaGAN by:

$$\hat{D}(x_r) = \sigma(C(x_r) - \mathbb{E}_{x_f}[C(x_f)]) \tag{6}$$

$$\hat{D}(x_f) = \sigma(C(x_f) - \mathbb{E}_{x_r}[C(x_r)]) \tag{7}$$

where $\mathbb{E}_{x_f}$ and $\mathbb{E}_{x_r}$ are the expectation functions. It follows that the discriminator loss function L_D^{RaGAN} and adversarial loss function L_G^{RaGAN} can be written as:

$$L_D^{RaGAN} = -\mathbb{E}_{x_r}[\log(\hat{D}(x_r)] - \mathbb{E}_{x_f}[\log(1 - \hat{D}(x_f))] \tag{8}$$

$$L_G^{RaGAN} = -\mathbb{E}_{x_r}[\log(1 - \hat{D}(x_r))] - \mathbb{E}_{x_f}[\log(\hat{D}(x_f)] \tag{9}$$

It has been observed that the class of texture images is often more difficult for SISR due to spatial aliasing [29]. One way of achieving better texture reconstruction is through attention mechanism at the image level - i.e., to emphasize (i.e., increase the weight) difficult samples and overlook (i.e., down-weighting) easy ones. Such idea of weighting can be conveniently incorporated into the RaGAN package because the PyTorch implementation allows an optional weight input. More specifically, we propose to consider the following weighted function with a new hyperparameter γ tailored for *T*exture enhancement:

$$\begin{aligned} L_G^{TRaGAN} = &- \sum_i (\hat{D}(x_r))^\gamma \log(1 - \hat{D}(x_r)) \\ &- \sum_i (1 - \hat{D}(x_f))^\gamma \log(\hat{D}(x_f)) \end{aligned} \tag{10}$$

B. Perceptual Loss Function

We have implemented the following perceptual loss function based on [47], [15], [16], [29]. With a pre-trained VGG19 model [48], we can extract high-level perceptual features of both high-resolution (HR) and SR images from the 4-th convolutional layer of VGG19 before the activation function is applied. Inspired by [16], we propose to extract high-level features before the activation function layer because it can further improve the performance. Let's define perceptual loss as L_{vgg} and L_1-norm distance as L_1. Then the total loss for our generator L_G can be formulated as follows:

$$L_G = L_{vgg} + \lambda_1 L_G^{TRaGAN} + \lambda_2 L_1 \tag{11}$$

where coefficients λ_1 and λ_2 are used to balance different loss terms. $L_{vgg} = \Phi(f(SR), f(HR))$. Φ denotes the mean-squared error function (MSE), $f(SR)$ and $f(HR)$ are the high-level features extracted from the output of the 4^{th} convolution layer of VGGNet before the pooling. Note that although similar loss functions were considered in previous studies including [16] and [29], their experiments include synthetic low-resolution images only. In this paper, we will demonstrate the effectiveness of the proposed perceptual optimization for JDSR on both synthetic and real-world data next.

V. EXPERIMENTAL RESULTS

A. Implementation details

In our proposed RDSEN networks, we set the number of RDSEB blocks as 16; and each block includes 6 residual-dense SE modules. Most kernel size of Conv layers is 3×3 with 64 filters ($C = 64$) except those described in particular: the Conv layers in CA modules and Conv layers marked as '1×1' with a 1×1 kernel size. The reduction ratio is $r = 16$. The upscale module we have used is the same as [49]. The last layer filter is set to 3 in order to output super-resolved color images. For the discriminator setting, we have implemented the same discriminator network structure as SRGAN [15]. All kernel size of Conv layers is 3×3.

In our PyTorch implementation of RDSEN, we first randomly crop the Bayer patterns as small patches with the size of 48×48, and crop the corresponding HR color images, with a batch size of 16; then we augment the training set by standard

Methods	Scale	Set5	Set14	B100	Manga109	McM	PhotoCD
FlexISP[46]+RCAN[18]	x2	4.16	4.14	3.34	4.97	3.51	5.42
DemoNet[19]+RCAN[18]	x2	4.13	3.76	3.31	3.99	3.48	5.59
RDSEN (ours)	x2	4.17	3.81	3.28	4.07	3.27	5.65
RDSEN_GAN (ours)	x2	3.41	2.95	**2.34**	3.53	2.59	4.85
RDSEN_TRaGAN (ours)	x2	**3.06**	**2.90**	2.35	**3.45**	**2.52**	**4.72**
FlexISP[46]+RCAN[18]	x3	6.98	5.70	6.18	5.43	5.14	6.42
DemoNet+RCAN	x3	6.31	5.18	4.97	4.63	5.19	6.61
RDSEN (ours)	x3	5.71	4.74	4.48	4.53	4.57	6.52
RDSEN_GAN (ours)	x3	3.78	2.94	2.39	3.44	2.60	4.96
RDSEN_TRaGAN (ours)	x3	**3.58**	**2.81**	**2.36**	**3.37**	**2.44**	**4.78**
FlexISP[46]+RCAN[18]	x4	7.42	6.63	6.30	5.28	7.15	6.88
DemoNet+RCAN	x4	7.21	6.23	6.28	5.43	6.22	7.04
RDSEN (ours)	x4	6.18	5.94	5.92	5.00	5.68	6.87
RDSEN_GAN (ours)	x4	4.50	3.31	2.84	3.65	2.84	5.01
RDSEN_TRaGAN (ours)	x4	**4.24**	**3.11**	**2.55**	**3.45**	**2.72**	**4.44**

TABLE II
OBJECTIVE PERFORMANCE COMPARISON AMONG DIFFERENT METHODS IN TERMS OF PERCEPTUAL INDEX (THE LOWER THE BETTER). **BOLD** INDICATES THE BEST RESULT AND UNDERLINE THE SECOND BEST.

geometric transformations (flipping and rotation). Our model is trained and optimized by ADAM [50] with $\beta_1 = 10^{-8}$, $\beta_2 = 0.999$, and $\epsilon = 10^{-8}$. The initial learning rate is set to 1×10^{-4}, the decay factor is set to 5, which decreases the learning rate by half after [$80k$, $120k$, $150k$, $180k$] steps; the L_1 loss function is applied to minimize the error between HR and SR images. To train GAN-based networks, we have used the trained RDSEN to initialize the generator of GAN to get a better initial SR image for discriminator. The same learning rate and decay strategies are adopted here. λ_1 and λ_2 in Eq. (11) are set to 5×10^{-3} and 1×10^{-2} respectively as [16].

Because the codes of RDSR [20] are not publicly available, we have tried our best to reproduce RDSR using PyTorch while keeping the batch size (16), patch size (64×64) and the number of residual blocks (24) exactly the same as used by the original work [20]. The learning rate and decay steps in RDSR implementation are the same as those in our RDSEN. This way, we have striven to make the experimental comparison against RDSR [20] as fair as possible.

B. Training Dataset

In our experiment, we have used DIV2K dataset [51] as the training set, which includes 800 images (2K resolution). For testing, we have evaluated both popular image super-resolution benchmark datasets including Set5 [5], Set14 [52], B100 [53], and Manga109 [54], and popular image demosaicing datasets such as McMaster [37] and Kodak PhotoCD. To pre-process training and testing data, we downsample original high-resolution images by a factor of 2×, 3×, 4× using Bicubic interpolation then generate the 'RGGB' Bayer pattern. Based on previous work [34] and our own study (refer to next paragraph), supplying three-channels separately as the input (instead of the mosaicked single-channel composition) works better for the proposed network architecture. All experiments are implemented using PyTorch framework [55] and trained on NVIDIA Titan Xp GPUs. As an indicator of the overall computational complexity, the training time of our RDSEN

Method	RDN	RCAN	RDSEN
Time (s)	128	160	130

TABLE III
TRAINING TIME COMPASSION OF RCAN, RDN AND PROPOSED RDSEN, PER EPOCH

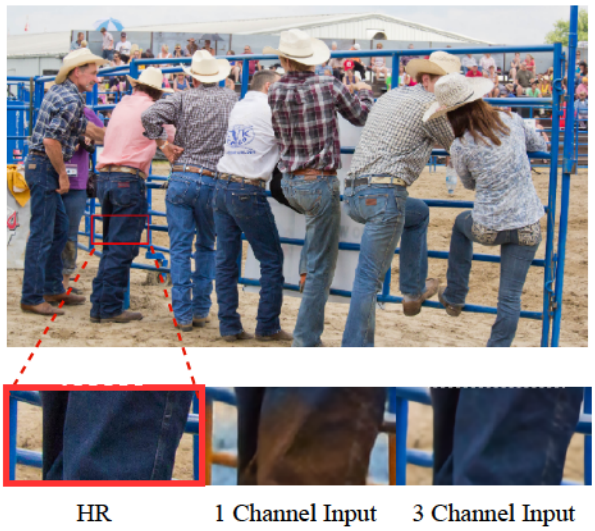

Fig. 6. Visual comparison of training data effect, the bottom images, from left to right, are HR image, SR image generated by one-channel feature map (raw Bayer-pattern), SR image generated by three-channel feature map (R,G,B with zero padding for the missing pixels).

lies somewhere between that of RDN [17] and RCAN [18] as shown in Table. III. We have verified for all competing networks, it takes around 1000 epochs to reach the convergence.

Note that we have to be careful about four different spatial arrangements of Bayer patterns [56]) in our definition of feature maps. One can either treat the Bayer pattern like a gray-scale image (one-channel setting) which ignores the important spatial arrangement of R/G/B; or take spatial arrangement as a priori knowledge and pad missing values across R,G,B bands by zeroes (three-channel setting). As shown in Fig. 6, the former has the tendency of producing

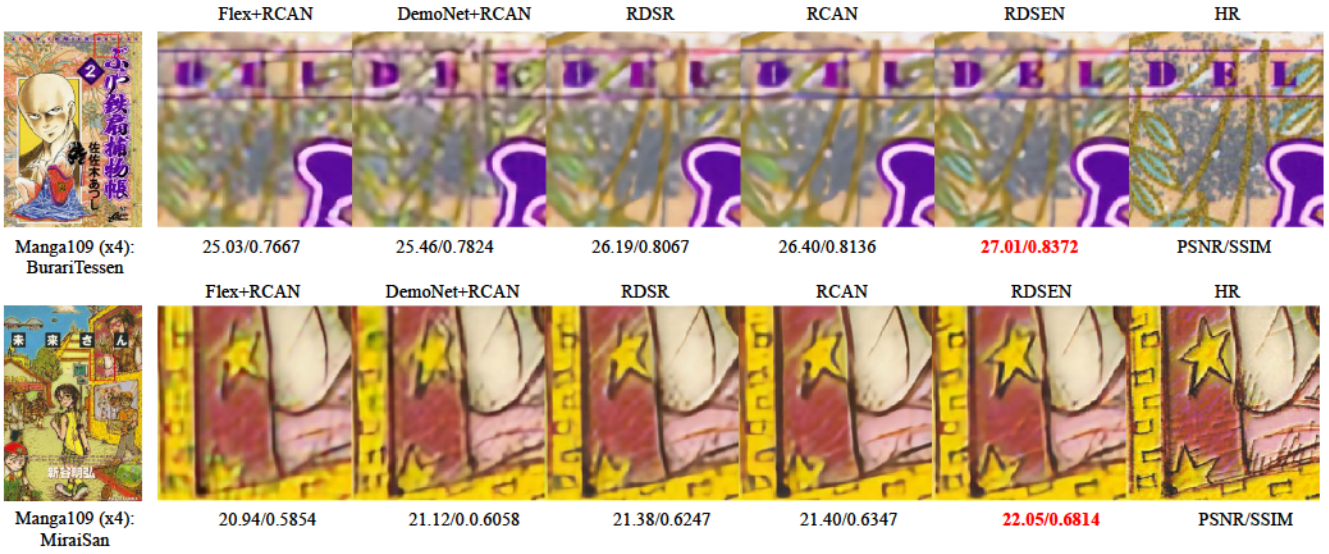

Fig. 7. Visual results among competing approaches for Manga109 dataset at a scaling factor of 4.

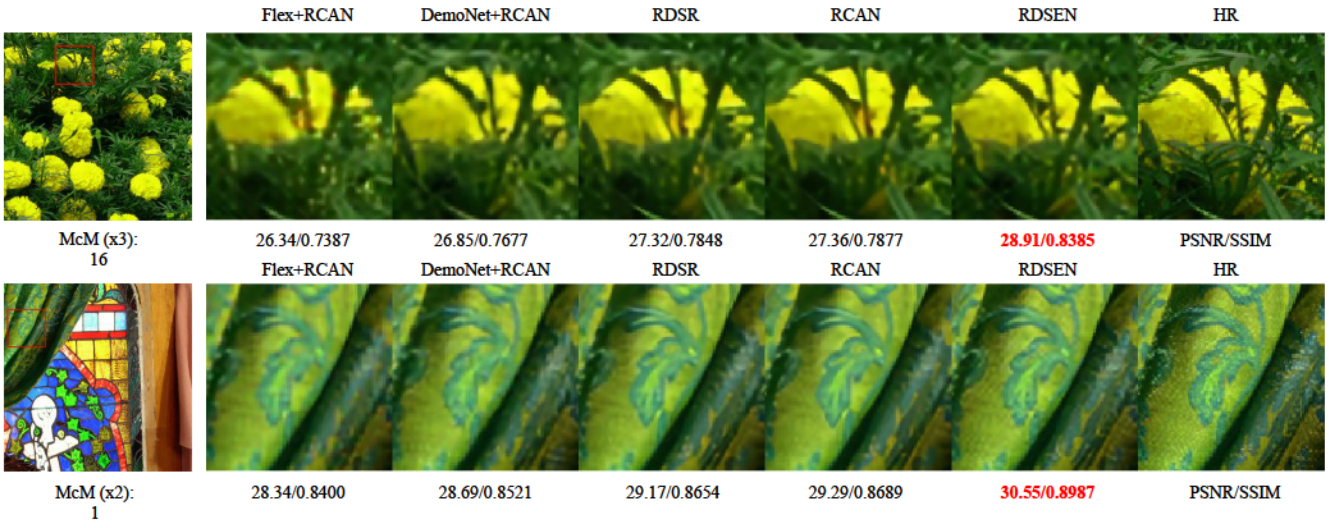

Fig. 8. Visual results among competing approaches for McM atasets at a scaling factor of 2 and 3.

Fig. 9. Visual results among competing approaches for Set14 and B100 datasets at a scaling factor of 3 and 2.

color misregistration artifacts, which suggests the latter works better. Our experimental result has confirmed a similar finding previously reported in [34].

C. PSNR/SSIM Comparisons

We have compared our methods against four benchmark methods: separated (brute-force) approaches Flex [46] + RCAN [18] and DemoNet [19] + RCAN [18], recently published literature RDSR [20], and state-of-the-art SR approach RCAN [18]. To evaluate the results of DemoNet [19] + RCAN [18] approach, we first demosaiced the LR mosaiced images by using a pre-trained demosaicing network DemoNet to get LR color images, then super-resolved them by applying a pre-trained RCAN model. Note that we have used the pre-trained DemoNet and RCAN weights provided by the authors on GitHub.

Table I shows PSNR/SSIM comparison results for scaling factors of $2\times$, $3\times$ and $4\times$. It can be seen that our RDSEN method perform the best for all datasets and scale factors. We observe that significant PSNR/SSIM gains (up to $1.2dB$) over previous state-of-the-art. Since PSNR/SSIM metrics do not always faithfully reflect the visual quality of images, we have also included the subjective quality comparison results for images "BurariTessen" and "MiraiSan" in Fig. 7. For the first row of Fig. 7, it can be readily observed that for the top of the letters, only our RDSEN can faithfully recover text details; brute-force approaches (Flex+RCAN and DemoNet+RCAN), RDSR and RCAN have produced severe blurring artifacts; for the second row, only our method can reconstruct the yellow stars faithfully. Taking another example, Fig. 8 shows the comparison at two other scaling factors ($3\times$ and $2\times$). For "McM(x3)_16", we observe that all approaches contain color artifacts between the flower and grass, but our RDSEN method can recover more realistic details than other competing approaches; for "McM(x2)_1", pattern recovered by RDSEN appears to have the highest quality and most detailed textures. For more visual comparison, see Fig. 9 which shows more convincing visual comparison among various competing approaches (please zoom in for detailed evaluation).

D. Perceptual Index (PI) Comparisons

Most recently, a new objective metric called Perceptual Index (PI) [57] has been developed for perceptual SISR (e.g., the 2018 PIRM Challenge [58]). The PI score is defined by

$$\mathrm{PI} = \frac{1}{2}((10 - \mathrm{MA}) + \mathrm{NIQE}) \tag{12}$$

where MA denotes a no-reference quality metric [59] and NIQE referred to Natural Image Quality Evaluator [60]. Note that the lower PI score, the better perceptual quality (i.e., contrary to SSIM metric [40]). Objective comparison of competing JDSR methods in terms of PI is shown in Table II. We have observed that GAN-based methods produce the lowest PI scores for all datasets and scaling factors. Fig. 10 provides the visual comparison with image "IMG0019" ($4\times$). It can be observed that GAN-based methods can recover sharper edges and overcome the issue of over-smoothed regions. Additionally, TRaGAN is capable of achieving even lower PI scores than the standard GAN. Fig. 11 shows another two results to demonstrate the advanced ability to recover texture details of GAN based methods, especially of TRaGAN.

E. Ablation Studies

To demonstrate the effect of proposed RDSEB module, we study the networks: 1) only based on ResNet; 2) ResNet with channel attention module (RCAN); 3) ResNet with proposed Dense connected Squeeze-and-Excitation modules (RDSEN). All three networks are trained under same setting for fair comparison. The general SR benchmark datasets are used, scale factor is 2. From Table. IV, we have found that ResNet has similar performance to more advanced RCAN. But when compared with our proposed RDSEN, the PSNR/SSIM performance of RCAN and ResNet are much lower than RDSEN; the proposed RDSEN has the best performance on all benchmark datasets.

F. Performance on the Real-world Data

Finally, we have tested our proposed JDSR technique on some real-world data collected by the Mastcam of NASA Mars Curiosity. The raw data are 'RGGB' bayer pattern sized by 1600×1200. Due to hardware constraints, the left camera and the right camera of Mastcam have different focal lengths (the left is about 3 times weaker than the right). To compensate such a "lazy-eye" effect on raw Bayer patterns, it is desirable to develop a joint demosaicking and SR technique with at least a scaling factor of 3 (in order to support high-level standard stereo-based vision tasks such as 3D reconstruction and object recognition). Our proposed JDSR algorithm is a perfect fit for this task, which shows the great potential of computer vision and deep learning in deep space exploration.

The visual comparison results are shown in Fig. 12 for a scaling factor of 4. It can be seen that brute-force approach (Flex+RCAN) suffers from undesired artifacts especially around the edge of rocks. Our proposed RDSEN method can overcome this difficulty but the result appears over-smoothed. RDSEN_GAN improves the visual quality to some degree - e.g., more fine details are present and sharper edges can be observed. Replacing GAN by TRaGAN can further improve the visual quality not only around the textured regions (e.g., roads and rocks) but also in the background (e.g., terrain appears visually clearer and sharper). Fig. 13 shows the visual comparison among Flex+RCAN, RDSEN, RDSEN_GAN and RDSEN_TRaGAN approaches. The raw image is captured by the right eye of NASA Mast Camera. The scale factor is 4 (please zoom in to get a better view).

VI. Conclusion

In this paper, we proposed to study the problem of joint demosaicing and super-resolution (JDSR) - a topic has been underexplored in the literature of deep learning. Our solution consists of a new residual-dense squeeze-and-excitation network for image reconstruction and an improved GAN with relativistic discriminator and new loss functions for

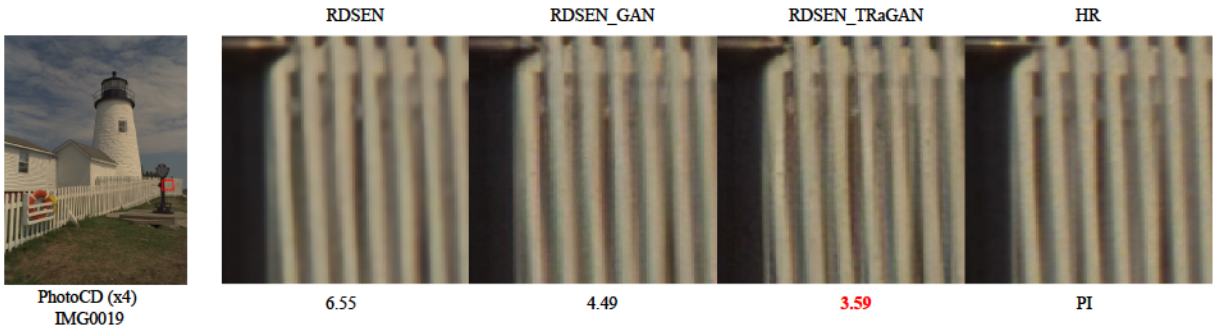

Fig. 10. Visual comparison results among competing approaches for PhotoCD dataset at a scaling factor of 4.

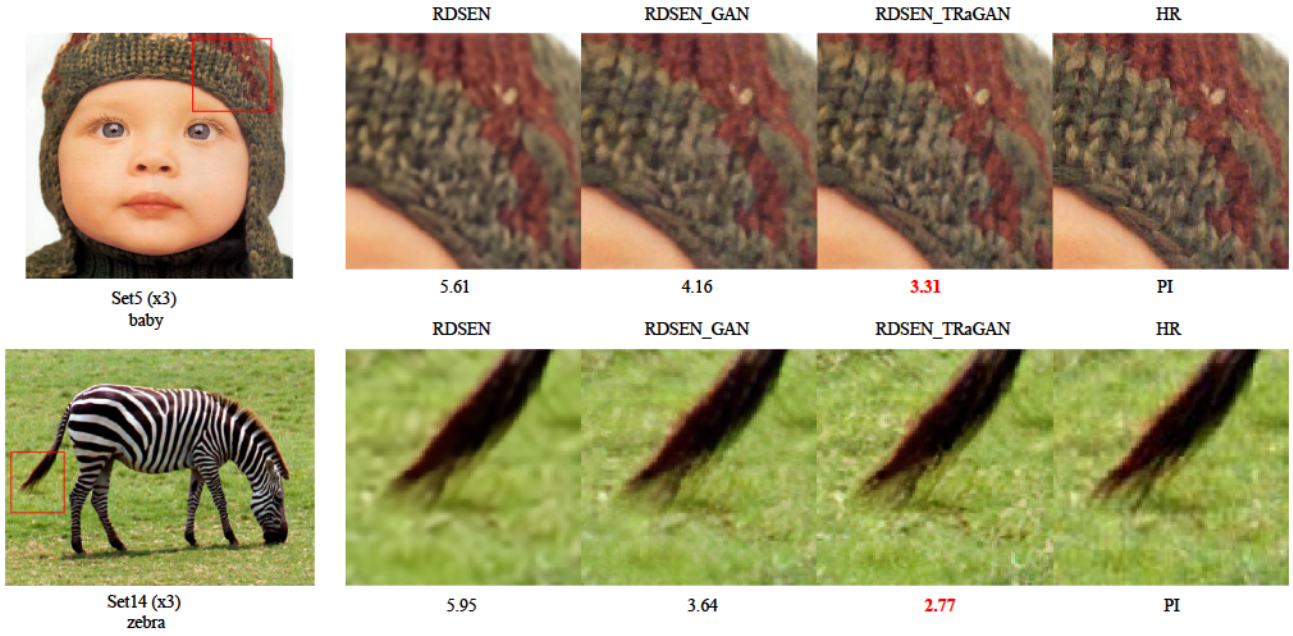

Fig. 11. Visual comparison results among competing approaches for Set5 and Set14 datasets at a scaling factor of 3.

Method	Scale	Set5	Set14	B100	Manga109	McM
		PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
ResNet	x2	36.48/0.9498	32.71/0.9030	31.67/0.8876	36.48/0.9642	36.11/0.9443
RCAN	x2	36.54/0.9499	32.74/0.9032	31.68/0.8878	36.65/0.9643	36.18/0.9445
RDSEN (ours)	x2	**37.40/0.9575**	**32.91/0.9128**	**32.00/0.8972**	**36.86/0.9716**	**37.38/0.9565**

TABLE IV
ABLATION STUDY FOR RESNET, RESNET WITH CA (RCAN) AND RESNET WITH PROPOSED RDSEN. **BOLD** FONT INDICATES THE BEST RESULT.

texture enhancement. Compared with naive network designs, our proposed network can stack more layers and be trained deeper and wider by newly designed RDSEB block. This is because RDSEB makes multiple residual-dense connection on channel descriptor to allow more faithful information flow. Additionally, we have studied the problem of perceptual optimization for JDSR. Our experimental results have verified that TRaGAN can generate more realistically-looking images (especially around textured regions) and achieve lower PI scores than standard GAN. Finally, we have evaluated our proposed method (RDSEN_TRaGAN) on real-world Bayer patterns collected by the Mastcam of NASA Mars Curiosity Rover, which supports its superiority to naive network design (e.g., Flex+RCAN) and the effectiveness of perceptual optimization. Another potential application of JDSR in practice is the digital zoom feature in smartphone cameras. Our solution to JDSR offers a cost-effective alternative to the existing hardware-based solutions (e.g. periscope camera design to achieve optical zoom).

ACKNOWLEDGMENT

The authors would like to thank Dr. Chiman Kwan for supplying real-world Bayer pattern collected by NASA Mars Curiosity. This work is partially supported by the NSF under grants IIS-2027127, IIS-1951504, CNS-1940859, CNS-1946327, CNS-1814825 and OAC-1940855, the DoJ/NIJ un-

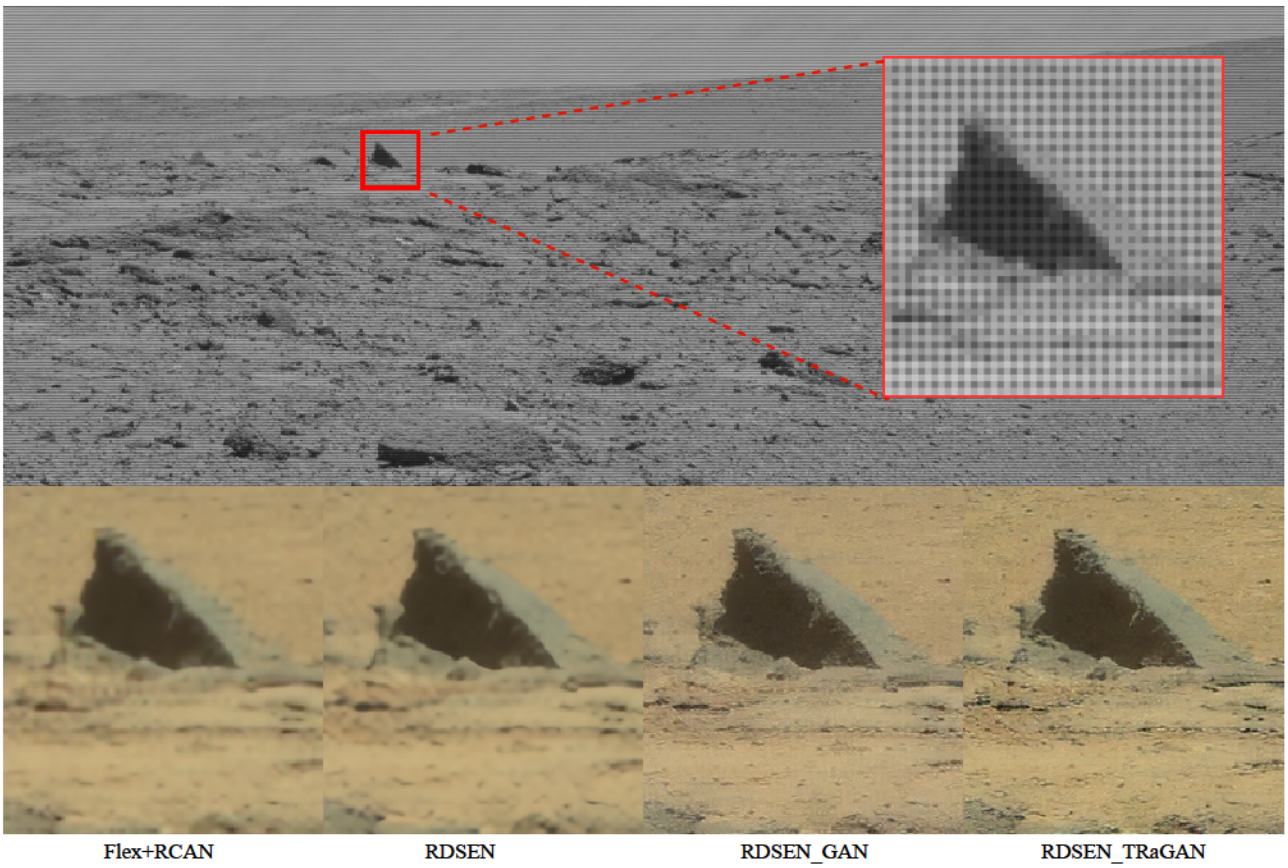

Fig. 12. Visual quality comparison of JDSR results on real-world Bayer pattern collected by NASA Mars Curiosity (4×).

der grant NIJ 2018-75-CX-0032, and the WV Higher Education Policy Commission Grant (HEPC.dsr.18.5).

REFERENCES

[1] L. Zhang and X. Wu, "Color demosaicking via directional linear minimum mean square-error estimation," *IEEE Transactions on Image Processing*, vol. 14, no. 12, pp. 2167–2178, 2005.

[2] X. Li, B. Gunturk, and L. Zhang, "Image demosaicing: A systematic survey," in *Visual Communications and Image Processing 2008*, vol. 6822. International Society for Optics and Photonics, 2008, p. 68221J.

[3] X. Li, "Demosaicing by successive approximation," *IEEE Transactions on Image Processing*, vol. 14, no. 3, pp. 370–379, 2005.

[4] W. Ye and K.-K. Ma, "Color image demosaicing using iterative residual interpolation," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5879–5891, 2015.

[5] M. Bevilacqua, A. Roumy, C. Guillemot, and M. L. Alberi-Morel, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," 2012.

[6] H. Chang, D.-Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in *Computer Vision and Pattern Recognition, 2004. CVPR 2004. Proceedings of the 2004 IEEE Computer Society Conference on*, vol. 1. IEEE, 2004, pp. I–I.

[7] R. Timofte, V. De Smet, and L. Van Gool, "Anchored neighborhood regression for fast example-based super-resolution," in *Proceedings of the IEEE international conference on computer vision*, 2013, pp. 1920–1927.

[8] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE transactions on image processing*, vol. 19, no. 11, pp. 2861–2873, 2010.

[9] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *International conference on curves and surfaces*. Springer, 2010, pp. 711–730.

[10] F.-L. He, Y.-C. F. Wang, and K.-L. Hua, "Self-learning approach to color demosaicking via support vector regression," in *Image Processing (ICIP), 2012 19th IEEE International Conference on*. IEEE, 2012, pp. 2765–2768.

[11] O. Kapah and H. Z. Hel-Or, "Demosaicking using artificial neural networks," in *Applications of Artificial Neural Networks in Image Processing V*, vol. 3962. International Society for Optics and Photonics, 2000, pp. 112–121.

[12] F. Kokkinos and S. Lefkimmiatis, "Deep image demosaicking using a cascade of convolutional residual denoising networks," in *The European Conference on Computer Vision (ECCV)*, September 2018.

[13] J. Sun and M. F. Tappen, "Separable markov random field model and its applications in low level vision," *IEEE transactions on image processing*, vol. 22, no. 1, pp. 402–407, 2013.

[14] J. Kim, J. Kwon Lee, and K. Mu Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1646–1654.

[15] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. P. Aitken, A. Tejani, J. Totz, Z. Wang *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network." in *CVPR*, vol. 2, no. 3, 2017, p. 4.

[16] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. C. Loy, "Esrgan: Enhanced super-resolution generative adversarial networks," in *The European Conference on Computer Vision Workshops (ECCVW)*, September 2018.

[17] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[18] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *The European Conference on Computer Vision (ECCV)*, September 2018.

[19] M. Gharbi, G. Chaurasia, S. Paris, and F. Durand, "Deep joint demosaicking and denoising," *ACM Transactions on Graphics (TOG)*, vol. 35, no. 6, p. 191, 2016.

[20] R. Zhou, R. Achanta, and S. Süsstrunk, "Deep residual network for joint

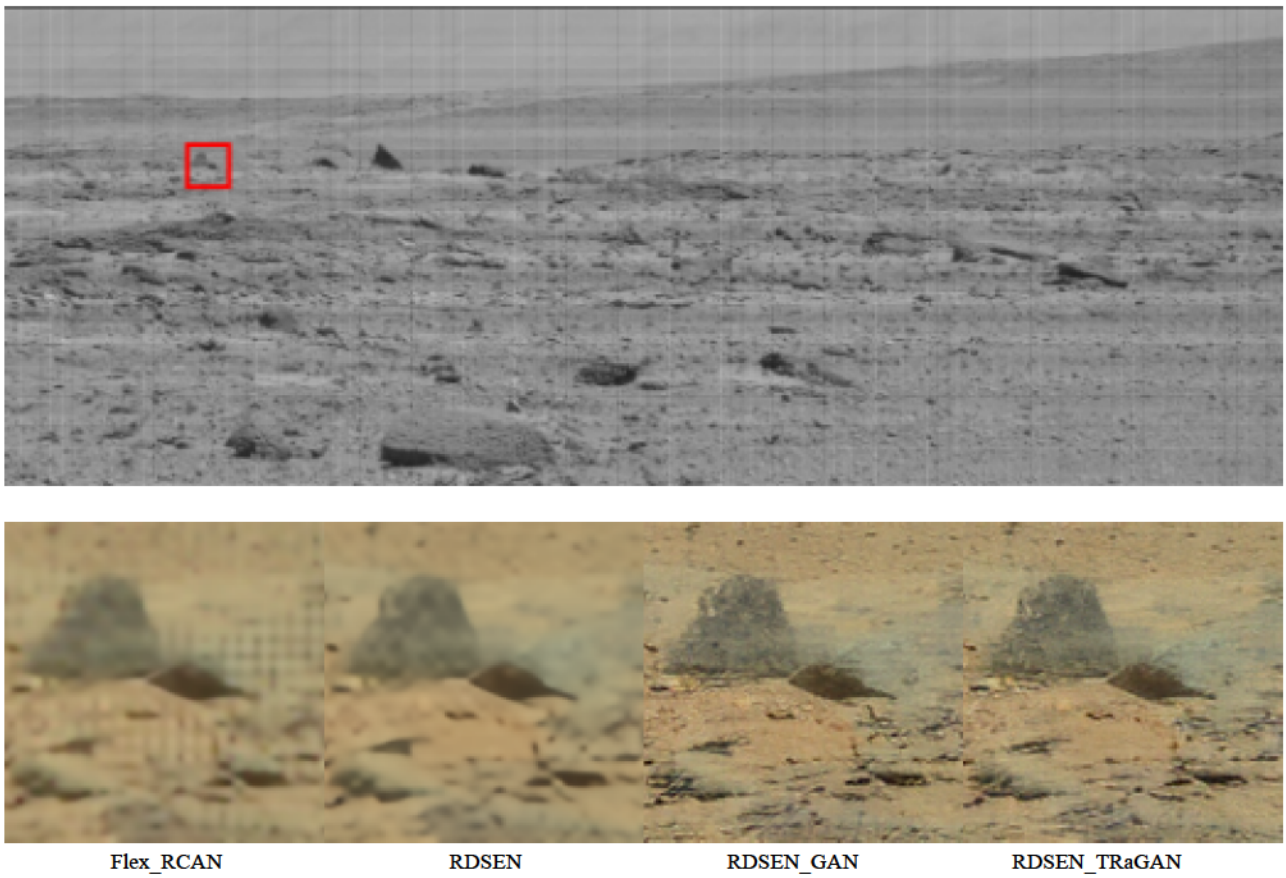

Fig. 13. More visual quality comparison of JDSR results on real-world Bayer pattern collected by NASA Mars Curiosity.

demosaicing and super-resolution," in *Color and Imaging Conference*, vol. 2018, no. 1. Society for Imaging Science and Technology, 2018, pp. 75–80.

[21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[22] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, July 2017.

[23] 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. 4799–4807.

[24] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

[25] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, "Enhanced deep residual networks for single image super-resolution," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, July 2017.

[26] X. Xu and X. Li, "Scan: Spatial color attention networks for real single image super-resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019, pp. 0–0.

[27] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets," in *Advances in neural information processing systems*, 2014, pp. 2672–2680.

[28] A. Jolicoeur-Martineau, "The relativistic discriminator: a key element missing from standard GAN," *arXiv preprint arXiv:1807.00734*, 2018.

[29] T. Vu, T. M. Luu, and C. D. Yoo, "Perception-enhanced image super-resolution via relativistic generative adversarial networks," in *The European Conference on Computer Vision (ECCV) Workshops*, September 2018.

[30] F. Kokkinos and S. Lefkimmiatis, "Deep image demosaicking using a cascade of convolutional residual denoising networks," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 303–319.

[31] ——, "Iterative joint image demosaicking and denoising using a residual denoising network," *IEEE Transactions on Image Processing*, vol. 28, no. 8, pp. 4177–4188, 2019.

[32] W. Ye and K.-K. Ma, "Color image demosaicing using iterative residual interpolation," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5879–5891, 2015.

[33] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *European conference on computer vision*. Springer, 2014, pp. 184–199.

[34] N.-S. Syu, Y.-S. Chen, and Y.-Y. Chuang, "Learning deep convolutional networks for demosaicing," *arXiv preprint arXiv:1802.03769*, 2018.

[35] R. Tan, K. Zhang, W. Zuo, and L. Zhang, "Color image demosaicking via deep residual learning," in *2017 IEEE International Conference on Multimedia and Expo (ICME)*. IEEE, 2017, pp. 793–798.

[36] W. Dong, M. Yuan, X. Li, and G. Shi, "Joint demosaicing and denoising with perceptual optimization on a generative adversarial network," *arXiv preprint arXiv:1802.04723*, 2018.

[37] L. Zhang, X. Wu, A. Buades, and X. Li, "Color demosaicking by local directional interpolation and nonlocal adaptive thresholding," *Journal of Electronic imaging*, vol. 20, no. 2, p. 023016, 2011.

[38] V. Singh, K. Ramnath, S. Arunachalam, and A. Mittal, "Going much wider with deep networks for image super-resolution," in *The IEEE Winter Conference on Applications of Computer Vision*, 2020, pp. 2343–2354.

[39] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Deep laplacian pyramid networks for fast and accurate superresolution," in *IEEE Conference on Computer Vision and Pattern Recognition*, vol. 2, no. 3, 2017, p. 5.

[40] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[41] P. Vandewalle, K. Krichane, D. Alleysson, and S. Süsstrunk, "Joint demosaicing and super-resolution imaging from a set of unregistered

aliased images," in *Digital Photography III*, vol. 6502. International Society for Optics and Photonics, 2007, p. 65020A.
[42] G. Qian, J. Gu, J. S. Ren, C. Dong, F. Zhao, and J. Lin, "Trinity of pixel enhancement: a joint solution for demosaicking, denoising and super-resolution," *arXiv preprint arXiv:1905.02538*, 2019.
[43] S. Guo, Z. Yan, K. Zhang, W. Zuo, and L. Zhang, "Toward convolutional blind denoising of real photographs," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 1712–1722.
[44] Q. Wang and G. Guo, "Ls-cnn: Characterizing local patches at multiple scales for face recognition," *IEEE Transactions on Information Forensics and Security*, 2019.
[45] Y. Bengio, P. Simard, and P. Frasconi, "Learning long-term dependencies with gradient descent is difficult," *IEEE transactions on neural networks*, vol. 5, no. 2, pp. 157–166, 1994.
[46] F. Heide, M. Steinberger, Y.-T. Tsai, M. Rouf, D. Pajak, D. Reddy, O. Gallo, J. Liu, W. Heidrich, K. Egiazarian *et al.*, "FlexISP: A flexible camera image processing framework," *ACM Transactions on Graphics (TOG)*, vol. 33, no. 6, p. 231, 2014.
[47] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *European Conference on Computer Vision*. Springer, 2016, pp. 694–711.
[48] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, 2014.
[49] 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 Conference on Computer Vision and Pattern Recognition*, 2016, pp. 1874–1883.
[50] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.
[51] E. Agustsson and R. Timofte, "NTIRE 2017 Challenge on Single Image Super-resolution: Dataset and Study," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, July 2017.
[52] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *International conference on curves and surfaces*. Springer, 2010, pp. 711–730.
[53] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Computer Vision, 2001. ICCV 2001. Proceedings. Eighth IEEE International Conference on*, vol. 2. IEEE, 2001, pp. 416–423.
[54] Y. Matsui, K. Ito, Y. Aramaki, A. Fujimoto, T. Ogawa, T. Yamasaki, and K. Aizawa, "Sketch-based manga retrieval using manga109 dataset," *Multimedia Tools and Applications*, vol. 76, no. 20, pp. 21 811–21 838, 2017.
[55] A. Paszke, S. Gross, S. Chintala, G. Chanan, E. Yang, Z. DeVito, Z. Lin, A. Desmaison, L. Antiga, and A. Lerer, "Automatic differentiation in pytorch," in *NIPS-W*, 2017.
[56] B. K. Gunturk, J. Glotzbach, Y. Altunbasak, R. W. Schafer, and R. M. Mersereau, "Demosaicking: color filter array interpolation," *IEEE Signal processing magazine*, vol. 22, no. 1, pp. 44–54, 2005.
[57] Y. Blau and T. Michaeli, "The perception-distortion tradeoff," in *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.
[58] Y. Blau, R. Mechrez, R. Timofte, T. Michaeli, and L. Zelnik-Manor, "2018 PIRM Challenge on Perceptual Image Super-resolution," *arXiv preprint arXiv:1809.07517*, 2018.
[59] C. Ma, C.-Y. Yang, X. Yang, and M.-H. Yang, "Learning a no-reference quality metric for single-image super-resolution," *Computer Vision and Image Understanding*, vol. 158, pp. 1–16, 2017.
[60] 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, 2013.