

# Joint Generative Learning and Super-Resolution For Real-World Camera-Screen Degradation

Guanghao Yin, Shouqian Sun, Chao Li, Xin Min

**Abstract**—In real-world single image super-resolution (SISR) task, the low-resolution image suffers more complicated degradations, not only downsampled by unknown kernels. However, existing SISR methods are generally studied with the synthetic low-resolution generation such as bicubic interpolation (BI), which greatly limits their performance. Recently, some researchers investigate real-world SISR from the perspective of the camera and smartphone. However, except the acquisition equipment, the display device also involves more complicated degradations. In this paper, we focus on the camera-screen degradation and build a real-world dataset (Cam-ScreenSR), where HR images are original ground truths from the previous DIV2K dataset and corresponding LR images are camera-captured versions of HRs displayed on the screen. We conduct extensive experiments to demonstrate that involving more real degradations is positive to improve the generalization of SISR models. Moreover, we propose a joint two-stage model. Firstly, the downsampling degradation GAN(DD-GAN) is trained to model the degradation and produces more various of LR images, which is validated to be efficient for data augmentation. Then the dual residual channel attention network (DuRCAN) learns to recover the SR image. The weighted combination of L1 loss and proposed Laplacian loss are applied to sharpen the high-frequency edges. Extensive experimental results in both typical synthetic and complicated real-world degradations validate the proposed method outperforms than existing SOTA models with less parameters, faster speed and better visual results. Moreover, in real captured photographs, our model also delivers best visual quality with sharper edge, less artifacts, especially appropriate color enhancement, which has not been accomplished by previous methods.

**Index Terms**—Camera-screen degradation, generative Learning, single image super resolution.

## I. INTRODUCTION

THE single super-resolution (SISR) is an elementary low-level vision task, which aims at the reconstruction of the high-resolution (HR) image from its low-resolution (LR) observation [1]. The SISR has high practical values to enhance the quality of image to promote human visual experience, which has been applied in medical imaging [2], satellite image enhancement [3] and facilitating other high-level tasks [4].

The SISR is a seriously ill-posed inverse problem because of ill-conditioned registration, unknown degraded operators and multiple correspondence from a specific LR input to a crop of HR images [5], [6]. Generally, the researches of SISR focus on learning the pixel and texture prior information from the paired HR and LR exemplar images [1], [7]–[9].

Guanghao Yin, Shouqian Sun, Xin Min, Chao Li are with the Key Laboratory of Design Intelligence and Digital Creativity of Zhejiang Province, Zhejiang University, Hangzhou 310027.

E-mail: {ygh\_zju, ssq, superli, minx}@zju.edu.cn

Existing SISR solutions can be divided into three types: interpolation-based methods, reconstruction-based methods and learning-based methods [10]. Early interpolation-based solutions, such as bicubic interpolation [11] and Lanczos resampling [12], have the fast speed but produce yield poor results. Reconstruction-based solutions utilize complicated prior knowledge to restrict the reconstruction [13], [14]. Learning-based solutions utilize machine learning models to mine the relationships from the LR-HR pairs. Since the classical SRCNN [15] has been proposed, deep convolutional neural network (CNN) based SISR methods are continually bringing prosperous improvement in terms of reconstruction accuracy [16]–[22].

However, the SISR research with complicated degradation still lacks effective exploration [10]. Existing deep learning based methods are suffering limitations of generalization and robustness in real-world degradations [23]–[25] because those models are well-designed for synthetic downsampling, such as bicubic interpolation (BI) [26]. For example, it can be seen in Fig. 11 that the state-of-the-art models, ESRGAN [27] and RCAN [21], are sensitive to the high-frequency Moiré pattern, although they have been trained with camera-screen degraded data. The popular SISR datasets with paired high-quality HR and artificial LR images lead to the over-fitting of DNN models on the certain degradation. There are two possible solutions that can be explored: (1) involving more LR images, which are more accordant with the complicated degradations in real-world conditions; (2) improving the representation ability of model to synthetically handle more complicated degradations. The recent trend of collecting real-world images [23]–[25] and generating multiple simulated data [26], [28], [29] is very positive, since it involves more degraded images and makes the resulting trained models perform better on real data.

In this paper, we attempt to explore whether camera-screen degradation could effectively improve the performance and generalization of SISR models. Different from the film days, digital photos are directly shown by the display screen and people would like to use their image acquisition device to record contents on the screen for convenience. In this real-world scene, we found that the camera-screen degraded image was more complicated with noise, blur, corruption and over-exposure under the joint influence of camera and screen, as shown in Fig. 1. The estimation of camera degradation is non-uniform which cannot count on synthetic kernel estimation methods [25], [29]. As the degradation is jointly influenced by camera and screen, the uniform solution becomes much more complicated. It should be characterized by obtaining real LR-HR pairs. Therefore, we establish a dataset named as

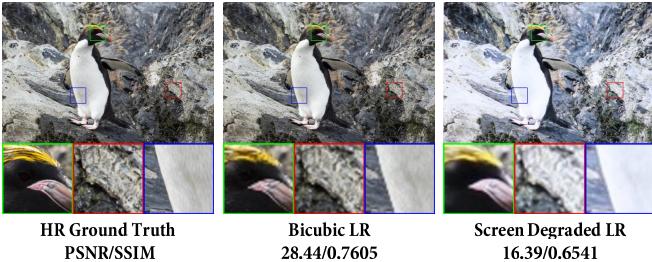


Fig. 1. Visual comparisons between the LR image (X4) with bicubic downsampling and the camera-screen degradation (both are displayed after interpolation). The camera-screen example is much degraded with noise, blur, corruption and overexposure, which is quantitatively verified by PSNR/SSIM.

Cam-ScreenSR, which contained the degradation from both the image acquisition and display device. The HR ground truths of the Cam-ScreenSR are from DIV2K dataset [30]. And the LR images are the corresponding photographs captured from a monitor by the camera. At first, we just established the training/testing sets with the same camera and monitor. However, the same image appears significantly different when displayed and captured using different camera-display combinations [31], as shown in Fig. 3. Therefore, two more testing sets were collected with different equipment to validate that our data acquisition solution was not strict only to the specific screen and specific camera. It should be emphasized that our exploration focuses on the SISR task. Recovering the photographs from the camera-screen degradation is a more sophisticated task than what researchers usually call "pure super resolution". The existing SISR methods only restore an image that has been prefiltered by some kernel and then downsampled. It really limits the applications of SISR. The camera-screen degraded SISR task includes things like denoising, sharpening the edge, fixing color distortion, and so on, which each have a long history of study in image processing. However, if we attempt to improve the practicability of SISR solution in real-world scene, it's inevitable to involve complicated degradations. It should be encouraged, not strictly separating research areas of image restoration.

To handle the camera-screen degradation, we also propose a joint generative learning and super-resolution model, as illustrated in Fig. 5. The proposed model contains two networks: (1) The downsampling degradation GAN (DD-GAN) is used to learn the camera-screen degradation and generate more degraded LR images for data augmentation (the results shown in Fig. 7). The DD-GAN focuses on overcoming the time-consuming and inefficient problem of large scale HR-LR manual acquisition. (2) The dual residual channel attention network (DuRCAN) recovers the mixed real captured and generated LR images. As existing pure SISR model can't handle complicated degradations, we involve the dual residual learning inspired by [32]. The channel attention mechanism is applied to exploit the inter-channel relationship and adaptively reweights channel-wise features. We figure out that the dual residual blocks focus on recovering clearer textures and the channel attention blocks conduct the color calibration. Besides, similar to the Laplace operation commonly used in image processing [33], our solution additionally involves a Laplacian

loss to sharpen the edge and smooth the noise.

Systemic ablation experiments have been conducted in the Cam-ScreenSR. And we compared our joint model with other SISR state-of-the-arts, which were all trained with the Cam-ScreenSR for fair comparisons. The comparisons were also conducted for the pure SISR task. The Cam-ScreenSR-trained models were finetuned with typical BI DIV2K training set and evaluated on popular SISR testing sets (Set5 [34], Set14 [35], BSD100 [36], Urban100 [9]). Moreover, the restoration of real-world photographs proved that our model could appropriately conduct color enhancement because of the camera-screen degraded data and channel attention mechanism, which has not been accomplished by previous SISR tasks. The excellent improvements in those experiments validate the great robustness and generalization of our solution. Compared with existing SOTA models, the proposed method can produce better visual results with less parameters and faster speed. The results also prove that involving more complicated degradation is helpful to boost development for SISR task.

In summary, the contributions of our paper are:

- First involving the camera-screen degradation and proposing a data acquisition strategy to establish the Cam-ScreenSR dataset, which is proved to be helpful for typical and real-world SISR tasks.
- Proposing a downsampling degradation GAN(DD-GAN), which learns the degradation from real-captured data and generates more LR images to replace the time-consuming manual acquisition.
- Proposing the dual residual channel attention network (DuRCAN), which is a controllable model to jointly restore the high-resolution details and enhance the color from the degraded images.
- Adding a Laplacian loss to sharpen the edge and smooth the noise.

## II. RELATED WORK

### A. Deep Learning Based Single Image Super-Resolution.

As a long-standing problem, early solutions for SISR task utilized the prior statistics [37]–[39] or exemplar patches [1], [40]. Due to the superior performance of the pioneer SRCNN model [15], deep learning-based methods have become the hotspots to tackle the ill-posed SISR problem. Then, researchers focused on designing deeper network structure with larger receptive field, such as VDSR [18], DRCN [41]. To utilize hierarchical features from different layers, many recent models also apply residual connections and dense blocks to mine the different frequency information from weight layers, such as SRDenseNet [20], EDSR [19], RCAN [21]. After various novel architectures and training strategies have been proposed, the SISR performance gets continuously improved, such as Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity (SSIM) values.

Some researchers noticed that the PSNR-oriented solutions tended to output over-smoothed results without sufficient details [27], [42]. Therefore, several explorations have been conducted to pursue visually pleasing results. The milestones, such as SRGAN [42] and ESRGAN [27], combined the

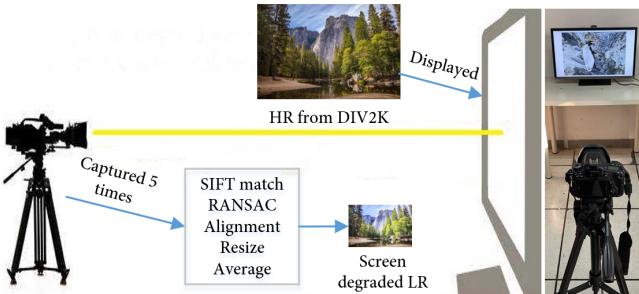


Fig. 2. The calibrated acquisition equipment for the camera-screen degraded LR images. Without changing other conditions, the LR training set and one testing set are collected with Samsung S27R350 + Canon 760D camera. And the other two LR testing sets are collected with Dell IN2020M + iPhone 11, Lenovo X1 laptop + Huawei P30. The latter two testing versions are used to validate the generalization of our model, seen in Section V.

adversarial loss and GAN framework to optimize the model in a feature space instead of pixel space [43]. Hence, those perceptual-driven SISR models can produce more photo-realistic results, which visually contain more irregular noises to rich high frequency details.

However, all those aforementioned works are trained to restore the limited artificial degradation. As mentioned before, the LR image in real-world SISR suffers more complicated degradations and all these existing models trained on synthetic datasets has poor ability to handle them [23]–[25].

### B. Real-world Datasets for Single Image Super-resolution

The synthetic datasets have been widely used for training and evaluating the SISR solutions, including Set5 [34], Set14 [35], BSD100 [36], Urban100 [9] and DIV2K [30]. However, the SISR models trained with simulated data deliver poor results when applied to real LR images [44]. It really limits the practicability of SISR models for real applications. To overcome the limitations of uniform downsampling, some recent works address on capturing paired LR and HR photographs in real-world scenes. To the best of authors' knowledge, only three real-world SISR datasets have been established. Chen *et al.* [23] employed camera/smartphone from the perspective of camera lenses and conducted data rectification to get aligned LR-HR paired CameraSR dataset. Zhang *et al.* [24] addressed on the optical zoom functionality of the camera to establish the SR-Raw dataset. Cai *et al.* [25] established a larger benchmark dataset (RealSR), where the LR-HR pairs on the same scene were captured by adjusting the focal length of a digital camera. Different from existing camera-based strategies, our Cam-ScreenSR is the first attempt from the perspective of the acquisition and display device. We also attempt to prove that involving more complicated camera-screen degradations is indeed valuable for SISR task.

### C. Residual Learning for Single Image Super-resolution

The network depth is of crucial importance to the representation ability [45], [46]. However, the stacked deep network suffers the notorious problem of vanishing/exploding gradients. After He *et al.* [47] creatively proposed the concept of

residual learning, the residual block became widely used in computer vision [48]–[50]. The residual connection provides a shortcut path to transfer the gradient of the error during back-propagation, which can effectively ease the training of deep networks (seen more details in [47]).

In SISR task, various researchers utilized the residual block as the basic unit to construct deep models for easier and more stable training [19]–[21]. Kim *et al.* [41] involved the residual skip-connection from input to the reconstruction layer, which could effectively supervise their recursive SISR model. Tong *et al.* [20] utilized the dense residual connections to reuse features from different layers and channels. Haris *et al.* [51] conducted iterative up-downsampling and used residual connections to project features. Recently, Liu *et al.* [32] proposed the concept of dual residual learning for noise removal, motion blur removal, haze removal, raindrop removal and rain-streak removal, where the dual residual connections provide more path to deliver features between the paired large- and small-size convolution kernels. The different combinations of dual kernels also provide various receptive fields for different resolution. Hence, we refer to the dual residual convolution operation to structure the basic block for camera-screen SISR task.

### D. Attention Mechanism

In human proprioceptive systems, attention generally provides a guidance to focus on the most informative components of an input [52]. In neural networks, attention mechanism is effective to mine the long-range feature correlations in channel- and spatial-wise, which can guide models to reweight features and focus on more useful parts. Recently, the superiority of attention models has been proved in various tasks, ranging from image classification [52]–[54] to language translation [55]. As [21] explains, in SISR task, the channel-wise features from different frequency are more informative for HR reconstruction. Therefore, we only involve the channel attention block to decrease the parameters of model. Moreover, we have explored that the channel attention can provide the ability of color calibration for our SISR solution with camera-screen degradation, seen in Section V.

## III. DATA ACQUISITION STRATEGY

To capture realistic camera-screen degradation, we display the original images of DIV2K dataset on a Samsung S27R350 monitor. The resolution of screen keeps the maximum  $1920 \times 1080$  with the 16:9 aspect ratio. To maintain the picture quality of the original source, the monitor is set to Standard mode (Brightness: 30, Contrast: 75, Sharpness: 64). The HR images are fullscreen displayed with Microsoft photo viewer. For image acquisition device, we utilize a DSLR camera (Canon 760D) to capture the camera-screen degraded LR images. The resolution of Canon 760D is  $6000 \times 4000$  and we capture LR observation at minimum 18mm focal length. Similar to the settings in [25], the camera is set to aperture priority mode and the ISO value is set to the lowest level to alleviate noise. The camera focuses on the center of monitor. The white balance and exposure are set to automatic mode.

TABLE I  
CAMERA AND SCREEN SPECIFICATIONS FOR CAM-SCREENSR DATASET.

Cam-ScreenSR	Camera	Screen	Resolution
Training Set	Canon 760D	Samsung S27R350	$1920 \times 1080$
testing set 1	Canon 760D	Samsung S27R350	$1920 \times 1080$
testing set 2	Huawei P30	Lenovo X1	$3840 \times 2160$
testing set 3	iPhone 11	Dell IN2020M	$1600 \times 900$



Fig. 3. Cam-ScreenSR examples: Our dataset contains one camera-screen degraded training set and three testing sets. Each column corresponds to a different camera-display pair. The same image appears significantly different with different camera-display combinations. (Best viewed as zoomed-in PDF.)

Not only used for training, the Cam-ScreenSR training data also guides our DD-GAN to produce various degraded LRs. However, Camera properties (spectral sensitivity, radiometric function, spatial sensor pattern) and display properties (spatial emitter pattern, spectral emittance function) cause the same image to appear significantly different when displayed and captured using different camera-display hardware [31]. Therefore, to validate that our solution is not strict only to the specific equipment, we added two more testing sets. The added display devices are Dell IN2020M ( $1600 \times 900$ ) and Lenovo X1 laptop ( $3840 \times 2160$ ), which are set to Standard mode similar to the Samsung S27R350 monitor. And we capture two versions of testing sets by the smartphone iPhone 11 and Huawei P30. Referring to [25], the configurations of the smartphone are similar to that for DSLR camera by using the ProCam software. To avoid less-effective repetition, we just present the training data acquisition strategy in the follow.

As shown in Fig. 2, the monitor is put in front of the clear background and the camera is mounted on a stable tripod at a distance of about 1 meter from the screen. To minimize the spatial misalignment and lens distortion, we utilize the mapping equipment to calibrate the camera lens parallel to the monitor and adjust camera to the same height of the screen center. For keeping stabilization, the camera connection software is used to control the shutter remotely. Different from zooming the camera lens in [23]–[25], our image collection strategy maintains the focal lengths which can avoid lens distortions at different focal lengths. Therefore, the fixed focal length and calibrated spatial position make our image pair alignment more easier. Each HR image is five continuous captured. Similar to CameraSR [23], we conduct SIFT key-points match [56] between the original HR images

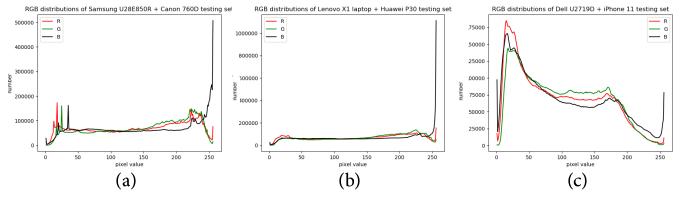


Fig. 4. The distributions of RGB channels of three Cam-ScreenSR testing sets. The figures in (a), (b), (c) separately correspond to the Samsung S27R350 + Canon 760D, Dell IN2020M + iPhone 11 and Lenovo X1 laptop + Huawei P30 testing datasets.

and the five captured LRs. Then, we utilize RANSAC [57] to filter mismatched coordinates and estimate the homography. After getting the alignment parameters, we obtain five aligned images and average them to the one. Finally, the interpolation with scale factor is conduct to produce low resolution images. According to [23], the smoothing effects from the interpolation is not critical for LR images. In this paper, we typically choose the scale factor 4 to avoid the less-effective experiment repetition, similar to previous articles [27], [42].

For data rectification, we conduct cropping, alignment, interpolation, but not luminance adjustment. The intensity variation is an important characteristic of camera-screen degradation. In practice, when facing LED shiner from different monitors, the aperture was auto-enlarged. Then, the captured photographs have different degrees of exposure and color distortion (as shown in Fig. 3). This situation should not be ignored because it's consistent with human pupil dilation. Moreover, the camera-screen degraded data with color distortion plays an important role to the PSNR/SSIM indexs and also guides the model for color correcting, which greatly expands the practicability of our solution (seen in Section. V).

We define the data acquired by Samsung + Canon 760D, Lenovo X1 + Huawei P30 and Dell + iPhone 11 combinations as the testing set 1, 2, 3. The RGB channel distributions of three testing sets are illustrated as Fig. 4. Compared to testing set 3, the distributions of testing set 1 and 2 versions are relatively similar, because those two monitors have higher resolution and better color revivification degree. When LED shinners of the monitor are more luminous than the environment, the aperture of the camera will be auto-enlarged to receive more light. It explains the images of testing set 1 and 2 are lighter than HR images, and the one of testing set 3 is darker, as illustrated in Fig. 3.

## IV. PROPOSED NETWORK

### A. Overview

As described in Section III, we have established the camera-screen degraded super-resolution dataset (Cam-ScreenSR), which consists of paired LR-HR images  $\{Y, X_{LR}\}$ . To avoid the less-effective experiment repetition, our dataset focuses on SISR with scale factor 4. The size of  $N$  HR ground truths ( $Y = \{Y_1, \dots, Y_n\}$ ) is  $h \times w$ , and the paired LR ( $X_{LR} = \{X_{LR1}, \dots, X_{LRn}\}$ ) is  $\frac{h}{4} \times \frac{w}{4}$ . Previous SR formulations only consider the influence of camera [25], [26], [29]. The camera-screen degradation  $D_{SR}(\cdot)$  is a comprehensive function with

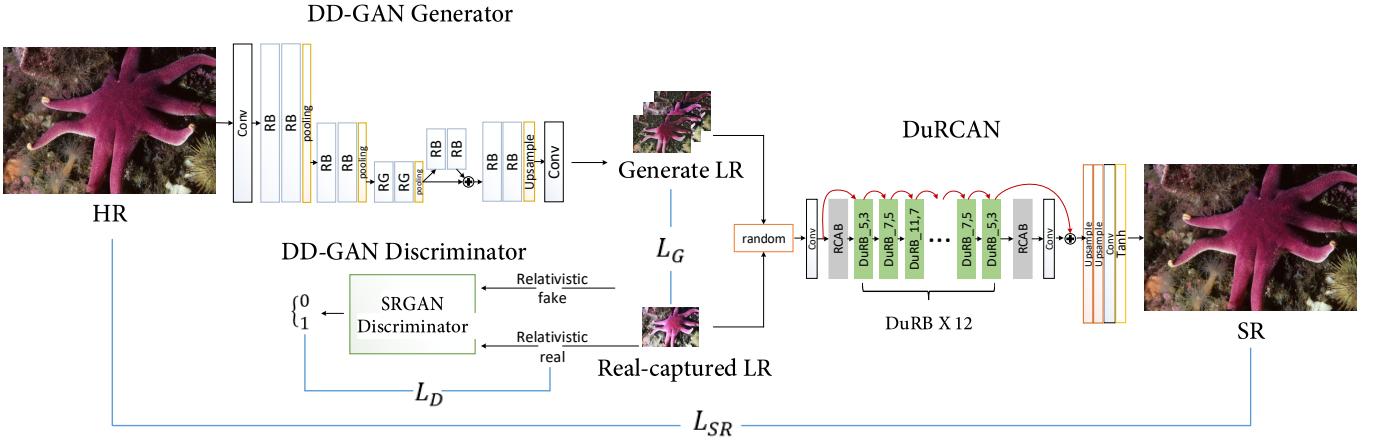


Fig. 5. Overview of the proposed model. The number of upsample blocks can be adjusted for other scale factors. After joint training, the parameters of DD-GAN and DuRCAN can be finetuned for other situations.

noise, blur, luminance corruption and downsampling from acquisition and display device. As expressed in [44], we can define the camera-screen degradation  $D_{SR}(\cdot)$  as:

$$X_{LR} = D_{SR}(Y) \quad (1)$$

$$= ((f(Y * k_1) \downarrow^1 + n_1) * k_2) \downarrow^2 + n_2, \quad (2)$$

where  $f(\cdot)$ ,  $k_1$ ,  $\downarrow^1$ ,  $n_1$  denote the image distortion, degraded kernel, downsampling operator, additive noise arose from the screen respectively and  $k_2$ ,  $\downarrow^2$ ,  $n_2$  are those from the camera. Obviously, simplifying the noise and downsampling components like previous blue-kernel estimated SISR methods [58], [59] will underestimate the camera-screen degradation. Considering it is difficult to derive an numerical solution for Eq. 2, we directly learn the SR restoration with real camera-screen degraded photographs.

The overall architecture of our model is illustrated in Fig. 5. The data collection with various combinations of camera/monitor is a huge workload. Therefore, we utilize the generative learning to simulate camera-screen degradation from limited real-captured data and conduct data augmentation. The HRs ( $Y = \{Y_1, \dots, Y_n\}$ ) from training set are fed into the DD-GAN. The generator  $G_{\Theta_1}$  produces generated LR images ( $X_{GLR}$ ) as:

$$X_{GLR} = G_{\Theta_1}(Y). \quad (3)$$

Referring to the Relativistic GAN [60], the discriminator  $D_{\Theta_2}$  predicts the probability that a real LR image  $x_{LR}$  is relatively more realistic than the average of generated fake images  $x_{GLR}$ , which guides the generator to produce more realistic outputs. Following Goodfellow *et al.* [61], the DD-GAN is optimized to solve the adversarial min-max problem:

$$\min_{\Theta_1} \max_{\Theta_2} \mathbb{E}_{Y \sim p_{train}(Y)} [\log(D_{\Theta_2}(X_{LR}, G_{\Theta_1}(Y)))] + \mathbb{E}_{X_{LR} \sim p_G(Y)} [\log(1 - D_{\Theta_2}(G_{\Theta_1}(Y), X_{LR}))]. \quad (4)$$

Then, the real captured LR  $X_{LR}$  and generated  $X_{GLR}$  are randomly sent to the SR restoration network DuRCAN. We can obtain the SR images  $\hat{Y}$  as:

$$\hat{Y} = S_{\Theta_3}(X), X = \{X_{LR}, \gamma X_{GLR}\}, \quad (5)$$

where  $\gamma$  is the mixing rate.

It should be emphasized the DD-GAN and DuRCAN are jointly trained where the failure outputs from the unbalanced generator can promote the robustness of DuRCAN. After training, the DuRCAN is used to restore LR images.

### B. High-to-Low Downsampling Degradation GAN

The DD-GAN conducts a high-to-low generating, which simulates the process of camera-screen degradation to get more synthetic LR images for data augmentation. Previous work [28] in face super-resolution learns the artificial degradation pattern by concatenating noise vectors with unpaired HR images which easily causes mode collapse. Different from the aforementioned model, our DD-GAN directly utilizes real captured image pairs.

1) *Generator*: The generator relies on a encoder-decoder architecture for downsampling degradation. Given the input HR images  $Y$ , the generator of DD-GAN output the synthetic LR images  $X_{GLR}$  to augment the limited real captured LR images. The downsampling and degradation process is modeled as two-stages: (1) the contracting subnet encodes the features of HR inputs. (2) the expansive subnet decodes the internal features from the contracting subnet to inversely generate camera-screen degraded LR images.

As illustrated in Fig. 5 and Fig. 6 (a), the typical Res-block [47] with MaxPooling operation is the stacked unit. Firstly, the single convolutional layer extracts shallow features from the input HR images. Then, those features are processed by the contracting subnet, which consists of 3 repeated groups. Each residual group contains two Res-blocks followed by a  $2 \times 2$  MaxPooling operation for downsampling. Although maxpooling is not recommended in SR task for reducing image details [41], it's suitable in the reverse high-to-low downampling degradation [28]. Two stacked Res-blocks conduct "bottom" feature extraction. After that, the expansive subnet decodes concatenated features from previous layers and corresponding layers of contracting subnet with the PixelShuffle upsample-blocks [17], the number of which is calculated by  $N - \log_2 S$  where  $N$  is number of contracting

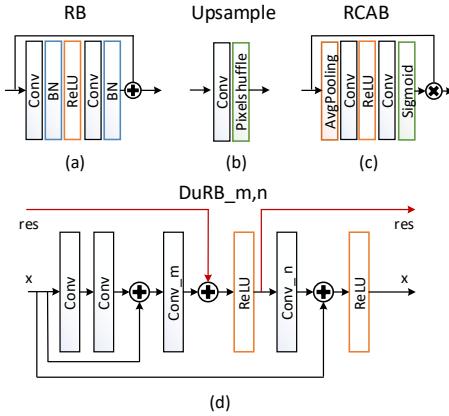


Fig. 6. The unit architectures of (a) the residual block (RB); (b) upsample block; (c) residual channel attention block (RCAB) and (d) dual residual block with the small- and large-kernel sizes  $[T_m, T_n]$  (DuRB<sub>m,n</sub>).

layers and  $S$  is the scale factor. We attempt to output  $\times 4$  LR images. Hence, the contracting subnet consists of one upsample-block and one convolutional layer to get 3-channel output.

2) *Discriminator*: The discriminator of standard GAN estimates the probability that the input image is real, which guides the generator to increase the probability that fake data is as real as ground truth. However, our DD-GAN tries to generate more various degraded images during the generative learning. If we apply standard GAN, the output will fall into the specific degradation pattern similar to the Cam-ScreenSR training data. Therefore, we enhance the discriminator with the relativistic label [60], which can effectively decrease the realistic probability of real data during the training. The average evaluation of relativistic discriminator predicts the probability that a real image is relatively more realistic than all fake data in a batch, which formulated as

$$D_{\Theta_2}(X_{LR_i}, X_{GLR}) \rightarrow 1, \quad (6)$$

$$D_{\Theta_2}(X_{GLR_i}, X_{LR}) \rightarrow 0. \quad (7)$$

Specifically for our task, we follow the architecture of SRGAN discriminator [42] and enhance it with the average evaluation from relativistic discriminator [60]. The paired fake generated data  $X_{GLR}$  and real-captured data  $X_{LR}$  are fed into the SRGAN discriminator  $C(\cdot)$  to predict the probability. Then the output of target type subtracts the average of the opposite type in the mini-batch, followed with a Sigmoid function, which is formulated as:

$$D_{\Theta_2}(X_{LR}, X_{GLR}) = \delta(C(X_{LR}) - \mathbb{E}_{\mathbb{Q}}[C(X_{GLR})]), \quad (8)$$

$$D_{\Theta_2}(X_{GLR}, X_{LR}) = \delta(C(X_{GLR}) - \mathbb{E}_{\mathbb{P}}[C(X_{LR})]), \quad (9)$$

where  $\delta(\cdot)$  is the Sigmoid function and  $C(\cdot)$  is the SRGAN discriminator and  $\mathbb{E}$  represents the average operation of the data in the mini-batch and  $\mathbb{P}, \mathbb{Q}$  respectively denote the distribution of real and fake data. The examples of generated LR images produced by the DD-GAN are shown in Fig. 7.

### C. Low-to-High Dual Residual Channel Attention Network

Previous CNN-based SISR approaches [15], [18], [20], [21], [41], [62] have achieved impressive results on synthetic

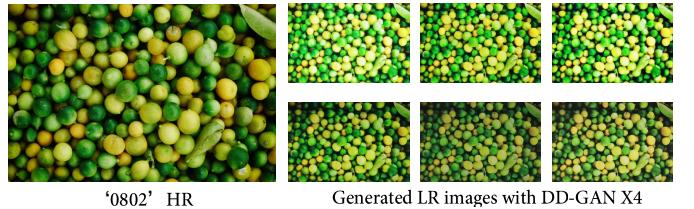


Fig. 7. Examples of LR images produced by our DD-GAN trained with real-captured Cam-ScreenSR training dataset. It can be seen that the DD-GAN tends to generate more varieties of degradation which is benefit for improving the performance of SR models (described in Section V).

datasets. However, those models have the poor ability to handle complicated degradation patterns in real-world (seen in Fig. 8 and Fig. 11), which motivates us to propose a novel low-to-high model with great generalization and robustness.

As shown in Fig. 5, the DuRCAN consists of five components: the shallow feature extraction, the deep feature extraction based on dual residual group (DuRG), two residual channel attention blocks at the beginning (RCAB\_bg) and the end (RCAB\_ed) of DuRG and the final upscale reconstruction to output the SR image. The number of channels is set to 64 for the internal layers.

Firstly, one convolutional layer  $H_{SF}(\cdot)$  extracts shallow feature  $F_{SF}$  from the LR input  $X$  as:

$$F_{SF} = H_{SF}(X). \quad (10)$$

Then, the residual channel attention block (RCAB\_bg) reweights the shallow feature and focuses on more useful parts in channel-wise dimensions. As Fig. 6 (c) shows, the average-pooling operation firstly aggregates the channel information to get the channel average-pooled feature  $F_{avg}^{h/4 \times w/4 \times c}$ . Then, the descriptor is fed into a multi-layer perceptron (MLP) with two convolutional layers and a ReLU function. In order to reduce parameter overhead, the convolutional layer  $W_0$  conducts channel reduction, where the channels of processed features are decreased by the reduction ratio  $\gamma$  to  $F_{avg_0}^{h/4 \times w/4 \times c/\gamma}$ . And the latter layer  $W_1$  recovers the feature shape. The sigmoid activation function  $\delta(\cdot)$  generates the normalized channel attention weight  $W^{h/4 \times w/4 \times c}$  between 0 to 1. After that, the original shallow feature is reweight by the multiplication with the channel attention weight. Overall, the reweighted channel-wise feature  $F_{CA1}$  is computed as:

$$\begin{aligned} F_{CA1} &= \delta(MLP(Pool(F_{SF}))) * F_{SF} \\ &= \delta(W_1(W_0(F_{avg}))) * F_{SF}. \end{aligned} \quad (11)$$

Later, the dual residual group focuses on the SR reconstruction, which consists of 6 dual residual blocks (DuRB). As illustrated in Fig. 6 (d), the DuRB receives features and residual components  $[x_i, Res_i]$  from previous layer and outputs the processed  $[x_{i+1}, Res_{i+1}]$  to the next DuRB. It consists of two parts: (1) The residual unit with two stacked convolutional layers  $C(\cdot)$  processes feature  $x_i$  from the previous  $i$ th DuRB, which is formulated as  $x_{ci} = C(x_i) + x_i$ . (2) Two paired convolutional layers  $[C_m(\cdot), C_n(\cdot)]$  focus on the SR reconstruction with different kernel sizes  $[T_m, T_n]$ . The ReLU operators are followed after two layers. Without

a pyramid structure [63], the alternate kernels provide the different receptive fields to conduct reconstruction. Moreover, the alternate convolutions with large and small kernel also conduct coarse- and fine-grained feature extraction from multi-degraded LR images. During conducting the convolution, the dual residual connections not only involve the residual messages  $Res_i$  from previous DuRB, but also provide paths to deliver features from the first kernel to the latter one and deliver residual components to the next DuRB. It can greatly increases potential interactions between each block. The whole processing can be formulated as:

$$Res_{i+1} = C_m(C(x_i) + x_i) + Res_i, \quad (12)$$

$$x_{i+1} = C_n(C_m(C(x_i) + x_i) + Res_i) + x, \quad (13)$$

where the ReLU function is omitted in the equations. We set the kernel size of 12 DuRBs  $\{[T_1^l, T_1^s], \dots, [T_{12}^l, T_{12}^s]\}$  as  $\{[5,3], [5,3], [7,3], [7,5], [11,5], [11,7], [11,7], [11,5], [7,5], [7,3], [5,3], [5,3]\}$ . As [19], [27] point out, when the distributions of the training and testing sets differ a lot, batch normalization (BN) tends to introduce unpleasant artifacts and limit the generalization ability. Therefore, we don't involve the BN operation for our network.

Before upsampling to the original size of HR image, the reconstructed features are reweighted by another residual channel attention block (RCAB\_ed) and processed by one convolutional layer  $C_{ed}(\cdot)$ . Finally, after adding the 12th residual component  $Res_{12}$  from the last DuRB, the reweighted feature  $F_{CA2}$  are up-sampled using PixelShuffle  $PS(\cdot)$  [17]. And the last convolution layer  $C_{la}(\cdot)$  followed with Tanh function  $Tanh(\cdot)$  outputs the 3-channel SR image. The output SR image is computed as:

$$\hat{Y} = \text{Tanh}(C_{la}(PS(C_{ed}(F_{CA2}) + Res_{12}))). \quad (14)$$

Moreover, in ablation experiments, we find the DuRBs pay attention on restoring the details from complicated degradations. As the depth of DuRB increases, the SR image contains clearer textures and less artifacts. The channel attention blocks (RCAB\_bg and RCAB\_ed) focus more on color calibration. Removing the RCABs has acceptable influence on the SR definition, but greatly limits the ability of color calibration. Hence, for different scenes, the DuRCAN can be finetuned specifically. With the support of sufficient computing resources, the deeper DuRG or other novel model can be involved to improve the SR definition. And the targeted finetuning of the RCAB can control the ability of color enhancement.

#### D. Loss Functions

In this section, we will describe the loss functions in details.

*1) DD-GAN Loss Function:* To avoid overfitting in the specific degradation pattern, we apply the label smoothing [64] for discriminator. The distinguishing labels of real and fake data are not static as 1 and 0, but randomly sampled from the uniform distribution  $U(0, \alpha)$  and  $U(\beta, 1)$ , where  $\alpha$  and  $\beta$  are near 0 and 1. The BCE loss  $L_{BCE}(\cdot)$  is set to evaluate the distance between distinguishing label and predicted probability

from the discriminator. Therefore, the parameters of DD-GAN discriminator is optimized by discriminator loss  $L_D$  as:

$$L_D = \mathbb{E}_{\mathbb{P}}[L_{BCE}(a, D_{\Theta_2}(X_{LR}, X_{GLR}))] + \mathbb{E}_{\mathbb{Q}}[L_{BCE}(b, D_{\Theta_3}(X_{GLR}, X_{LR}))], \quad (15)$$

where  $\mathbb{P}, \mathbb{Q}$  respectively denote the distribution of real and fake data, and  $a, b$  are the random values in the range of  $U(0, \alpha)$  and  $U(\beta, 1)$ .

For generator, the loss function  $L_G$  is the combination of the content loss  $L_{con}$  and adversarial loss  $L_G^a$ . The content loss  $L_{con}$  consists of a perceptual loss  $L_{vgg19\_54}$  and a pixel-wise loss  $L_1(\cdot)$ , which is consistent with the previous ESRGAN [27]. The adversarial loss  $L_G^a$  is a symmetrical form with discriminator loss (Eq. 15) as:

$$L_G^a = \mathbb{E}_{\mathbb{P}}[L_{BCE}(b, D_{\Theta_2}(X_{LR}, X_{GLR}))] + \mathbb{E}_{\mathbb{Q}}[L_{BCE}(a, D_{\Theta_3}(X_{GLR}, X_{LR}))]. \quad (16)$$

Taking the adversarial training, the parameters of DD-GAN generator  $G_{\Theta_2}$  is optimized by generator loss  $L_G$  as:

$$L_G = L_{con} + \lambda L_G^a \quad (17)$$

where  $\lambda$  is the the coefficient to balance two loss terms. It should be noted that the follow-up Laplacian loss is not involved for DD-GAN, because the synthetic noises should be retained to get more degraded LR images.

*2) DuRCAN Restoration Loss Function:* Previous CNN-based SISR models [15], [19], [21] commonly utilize the pixel-oriented loss. Considering non-uniform noises greatly pollute the low-frequency area and edges of Cam-ScreenSR images are degraded, we add the Laplacian loss  $L_{lap}$ , which is inspired by the Laplace operation in image processing [33] to sharpen the high-frequency edge and smooth the noises in low-frequency area. The Laplacian loss is defined on the 2D Laplace operator to minimize the  $L_1$  distance between the filtering images of generated image  $S_{\Theta_3}(X)$  and ground truth  $Y$ . Hence, the restoration loss function  $L_{SR}(\cdot)$  is a weighted combination of  $L_1$  loss and  $L_{lap}$  loss as:

$$L_{SR} = L_1 + \eta L_{lap} \\ = L_1 + \eta \frac{1}{w_l h_l} \sum_{i=1}^{w_l} \sum_{j=1}^{h_l} |\kappa_{lap}(Y)_{i,j} - \kappa_{lap}(S_{\Theta_3}(X))_{i,j}|, \quad (18)$$

where  $L_1$  loss is the main part and  $\eta$  is the coefficient to balance two loss terms.  $\kappa_{lap}(\cdot)$  denotes the filter with second order differential Laplace kernel [33],  $w_l, h_l$  represent the size of filtering image. The effectiveness of Laplacian loss is demonstrated in Section V-G.

## V. EXPERIMENTS

For easier sorting through detailed results, we conduct comparison experiments and ablation analysis. We first introduce the datasets involved for our comparison experiments in Section V-A and our training setup in Section V-B. Then, we train our model and several state-of-the-art methods (SOTA) on the Cam-ScreenSR training set. The cross-camera-screen evaluations are conducted in Section V-C to prove the generalization

TABLE II

AVERAGE PSNR AND SSIM RESULTS ON THREE CAMERA-SCREEN DEGRADED TESTING DATASETS WITH SCALE FACTOR 4. **TEXT** AND **TEXT** INDICATE THE BEST AND THE SECOND BEST PERFORMANCE.

Algorithm	Scale	Parameters	Testing Set 1 Samsung S27R350 + Canon 760D PSNR/SSIM	Testing Set 2 Lenovo X1 + Huawei P30 PSNR/SSIM	Testing Set 3 Dell IN2020M + iPhone 11 PSNR/SSIM
Bicubic	X4	-	16.24 / 0.6401	16.82 / 0.6622	19.31 / 0.6904
SRCNN [15]	X4	15K	18.55 / 0.6048	18.37 / 0.6926	13.43 / 0.5615
VDSR [18]	X4	665K	20.51 / 0.7025	19.76 / 0.6983	13.96 / 0.5842
EDSR [19]	X4	43090K	21.93 / 0.7086	21.32 / 0.7054	14.19 / 0.5933
ESRGAN [27]	X4	17000K	23.68 / 0.7140	23.91 / 0.7122	14.53 / 0.6014
RCAN [21]	X4	15592K	23.87 / 0.7155	23.33 / 0.7116	14.44 / 0.6045
DuRCAN (Ours)	X4	5453K	24.34 / 0.7224	24.07 / 0.7181	14.98 / 0.6178
SRCNN [15] + DD-GAN	X4	5352K	17.83 / 0.6858	18.62 / 0.6922	19.84 / 0.6909
VDSR [18] + DD-GAN	X4	6002K	19.97 / 0.7044	21.82 / 0.7068	20.63 / 0.7030
EDSR [19] + DD-GAN	X4	49077K	21.29 / 0.7051	23.48 / 0.7125	21.13 / 0.7061
ESRGAN [27] + DD-GAN	X4	22987K	23.74 / 0.7120	24.20 / 0.7207	21.50 / 0.7033
RCAN [21] + DD-GAN	X4	21580K	23.81 / 0.7148	24.03 / 0.7179	21.86 / 0.7075
DuRCAN + DD-GAN(Final,Ours)	X4	11440K	<b>24.82 / 0.7271</b>	<b>24.51 / 0.7240</b>	<b>22.19 / 0.7103</b>

and effectiveness of our solution. The comparisons between models with and without DD-GAN have also been conducted. Next, in Section V-D, we attempt to prove that camera-screen degradation can effectively improve the performance for typical SISR task. Hence, we finetune and evaluate those Cam-ScreenSR-trained models in typical bicubic interpolation (BI) datasets. Moreover, we conduct the qualitative evaluations on real-world photographs to compare our model against other SOTAs in Section V-E. And the comparisons of computational cost is presented in Section V-F. Finally, we conduct ablation studies in Section V-B to clearly present the effects of dual residual blocks, residual channel attention blocks and Laplacian loss.

### A. Datasets

We utilize the HR images of DIV2K [30] as the ground truths and collect the corresponding LR images with camera-screen degradation for X4 SISR task. The original DIV2K is divided to the training (ID: 0001-0800) and testing sets (ID: 0801-0900). For training set, the LR image is captured with Samsung S27R350 + Canon 760D. And for more general validation, the testing sets has three versions with different camera-screen combinations, as shown in Table I. To prove complicated camera-screen degradation is efficient for typical BI SISR task, the camera-screen trained models, including our DuRCAN and other SOTAs, are finetuned with the original DIV2K training set [30] and tested on popular BI datasets: Set5 [34], Set14 [35], BSD100 [36], Urban100 [9]. Moreover, the real-captured photographs by Huawei P30 will be used to validate the generalization of our approach in real-world scene. Following previous works [15], [18]-[21], [27], [41], [42], the evaluation metrics in our work are PSNR and SSIM [65] indices. The SISR results are evaluated using the Y channel in the YCbCr space.

### B. Training Setup

Both the camera-screen SISR, typical BI SISR and ablation analysis apply the setups in this section. The LR training

images of Cam-ScreenSR and typical BI datasets are randomly cropped into  $48 \times 48$  with mini-batch size 16. The 800 training images are randomly rotated by  $90^\circ, 180^\circ, 270^\circ$  and horizontally flipped for data augmentation.

To balance the distribution of LR inputs, we set the random rate between real LRs  $X_{LR}$  and generated LRs  $X_{GLR}$  with 4 : 1, where the mixing rate in Eq. 5 is  $\gamma = 0.25$ . The upper and lower limits of discriminator label smoothing in Eq. 15 are set as  $\alpha = 0.2, \beta = 0.8$ . To balance different loss terms, the coefficients of generator and restoration loss functions are set as  $\lambda = 1 \times 10^{-3}$  in Eq. 17 and  $\eta = 6 \times 10^{-3}$  in Eq. 18.

We select several typical state-of-the-art (SOTA) SISR models for comparison, the opensource codes of which have been released, including: SRCNN [15], VDSR [18], EDSR [19], ESRGAN [27] and RCAN [21]. The proposed DD-GAN and DuRCAN are jointly trained and the SISR testing only uses the trained DuRCAN. The learning rate is fixed at  $10^{-4}$  and halved every 50000 iterations. We use Adam [66] ( $\beta_1 = 0.9, \beta_2 = 0.999$ ) to optimize parameters of our network. All the experiments were conducted on NVIDIA Titan Xp GPUs.

### C. SISR Models Trained on Camera-Screen Degradation

As we mentioned before, what researchers usually call "super resolution" task is purely to restore an image that has been prefiltered by some kernel and then downsampled. But our work focuses on SISR with more complicated camera-screen degradation. To the best of authors' knowledge, there is no previous solution in this task to compare. Considering our model still attempt to restore images for higher resolution, we compare the proposed method with typical state-of-the-art (SOTA) SISR models including: SRCNN [15], VDSR [18], EDSR [19], ESRGAN [27] and RCAN [21]. We trained those models on Cam-ScreenSR training dataset and tested them on three testing datasets with different camera-screen combinations to evaluate the robustness of SISR methods. Moreover, in order to validate that the generated LR images from DD-GAN is benefit for avoiding overfitting in specific degradation, we separately train the models with and without



Fig. 8. Visual comparisons of Img-0825 from the testing set 2, where upscaling factor is 4. We present 6 better performed models with DD-GAN data augmentation. The HR and Bicubic baselines are also shown in the first and second columns. Our joint model delivers best visual quality with sharper edge, less artifacts and more appropriate color enhancement.

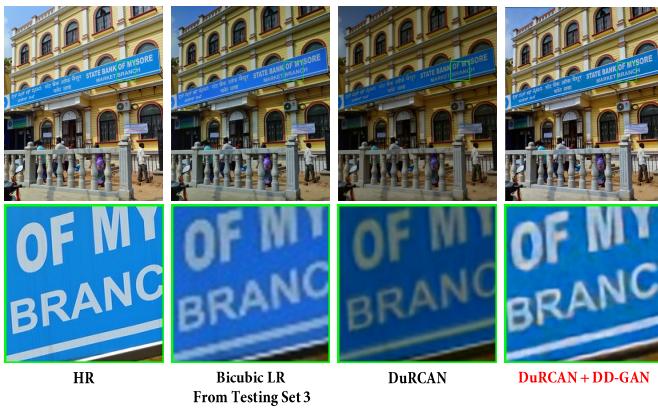


Fig. 9. Visual comparisons (X4) on the testing set 3 with Dell IN2020M and iPhone 11. The first and second rows show the '0891' HR and bicubic baselines. The latter two rows show the results of DuRCAN separately trained with and without DD-GAN.

DD-GAN. The experiments of those existing methods are conducted with the released opensource codes.

1) *Quantitative Analyses*: As the quantitative results shown in Table II, we can make some analyses as follows:

The 1st row in Table II presents the bicubic baselines of three testing sets. The baseline of testing set 1 is lower than testing set 2 because the Lenovo X1 monitor with higher resolution provides clearer LR images. As Fig. 4 shows, the distributions of testing set 1 and 2 are relatively similar, which explains the slight gap between the baselines in those two testing sets. Both the RGB distributions of Fig. 3 and visual examples in Fig. 4 reveal that the images of testing set 3 suffer less overexposure and color distortion. Hence, the third bicubic result is the higher than other two results.

The 2nd to 7th rows in Table II show the results where SISR models are trained without DD-GAN. Our proposed DuRCAN outperforms the SOTA models. It is also clear that the PSNR/SSIM indexs of all the SISR models on testing set 1 has great improvements than bicubic baseline. Because the same camera-screen equipment is used, the training set and testing set 1 have the consistent degradation. However, learning the specific degradation pattern leads to the overfitting of CNN models, which results in the great performance deterioration on other degraded LR images. The bicubic baselines reveal that the LR images of testing set 2 have higher quality than testing set 1. However, the SISR models trained

without DD-GAN bring less improvements on testing set 2. When testing on testing set 3 with much different degradation pattern, all the CNN-based SISR models produce worst results, even lower than the bicubic baseline in the first row. Overfitting on specific degradation greatly deviates the learning representation of DNN models and limits the generalization in real-world applications. When the variety of collected data is limited, exploring the effective data augmentation is necessary.

The 8th to 13th rows in Table II show the results where SISR models are jointly trained with DD-GAN. Because of the big gap between the HRs and real camera-screen degraded images, the proposed DD-GAN can generate more various LRs, as shown in Fig. 7. Except on testing set 3, the performance of most SOTA models + DD-GAN are slightly decreased, which means the generalization of those models is not enough to handle complicated degradation. After the DD-GAN generating more LR images with various degradation, the performance of our DuRCAN on both three testing sets get significant improvement than DD-GAN without DD-GAN and also achieves the best in Table II. Specifically, the ESRGAN is not seriously influenced by data distribution, and utilizes the adversarial learning to enrique texture details by adding high-frequency noises. However, those uncontrollable noises greatly increase the training difficulty and pollute the output image, which leads to the less evaluation indexs compared with our DuRCAN and also limits the robustness in real-world images (seen in Section V-E). The generated data enlarges the variety of image degradation and the dual residual convolution of our DuRCAN has great ability to handle those complicated degradations. Hence, the PSNR/SSIM growth of our model on testing set 3 is remarkable. Moreover, under the premise of better performance, the parameters of our DuRCAN is much less than existing SOTA models, like ESRGAN [27] and RCAN [21], which proves the superiority of our model in real-world degradation.

2) *Qualitative Analyses*: In Fig. 8, we show the bicubic baselines and visual comparisons of better SISR results with SISR models + DD-GAN on testing set 2 (Lenovo X1 + Huawei P30). Specifically, the visual comparisons of DuRCAN with and without DD-GAN on testing set 3 are illustrated in Fig. 9, which validates the effectiveness of our proposed DD-GAN. It can be clearly seen that our model delivers best visual quality with sharper edge, less artifacts, especially appropriate color enhancement.

TABLE III

AVERAGE PSNR AND SSIM RESULTS ON FOUR TYPICAL BICUBIC INTERPOLATION DATASETS WITH SCALE FACTOR 4. THE SYMBOL "+" REPRESENTS USING THE CAMERA-SCREEN PRETRAINED MODEL IN SECTION V-C. TEXT AND TEXT INDICATE THE BEST AND THE SECOND BEST PERFORMANCE.

Model	Scale	Parameters / Size	Set5 [34] PSNR / SSIM	Set14 [35] PSNR / SSIM	BSD100 [36] PSNR / SSIM	Urban100 [9] PSNR / SSIM
Bicubic	X4	-	28.42 / 0.8104	26.00 / 0.7027	25.96 / 0.6675	23.14 / 0.6577
SRCCNN [15]	X4	15K	30.48 / 0.8628	27.50 / 0.7513	26.90 / 0.7101	24.52 / 0.7221
VDSR [18]	X4	665K	31.35 / 0.8830	28.02 / 0.7680	27.29 / 0.7251	25.18 / 0.7524
EDSR [19]	X4	43090K	32.46 / 0.8968	28.80 / 0.7876	27.71 / 0.7420	26.64 / 0.8033
ESRGAN [27]	X4	17000K	32.60 / 0.9002	28.88 / 0.7896	27.76 / 0.7432	26.73 / 0.8072
RCAN [21]	X4	15592K	32.63 / 0.9002	28.87 / 0.7889	27.77 / 0.7436	26.82 / 0.8087
DuRCAN (Ours)	X4	5453K	32.61 / 0.8996	28.85 / 0.7884	27.74 / 0.7429	26.84 / 0.8091
SRCCNN+ [15]	X4	59KB	30.50 / 0.8643	27.59 / 0.7518	26.96 / 0.7151	24.60 / 0.7233
VDSR+ [18]	X4	2.55MB	31.39 / 0.8835	28.10 / 0.7686	27.33 / 0.7261	25.27 / 0.7540
EDSR+ [19]	X4	164MB	32.58 / 0.8984	28.84 / 0.7882	27.79 / 0.7431	26.78 / 0.8073
ESRGAN+ [27]	X4	63.8MB	<u>32.63</u> / 0.9005	28.89 / 0.7894	27.80 / 0.7433	26.84 / 0.8081
RCAN+ [21]	X4	59.7MB	<b>32.70</b> / <b>0.9007</b>	28.90 / 0.7896	<b>27.81</b> / <b>0.7439</b>	26.87 / 0.8099
DuRCAN+ (Final,Ours)	X4	20.8MB	32.60 / 0.8982	<b>28.93</b> / <b>0.7900</b>	27.64 / 0.7415	<b>26.92</b> / <b>0.8116</b>

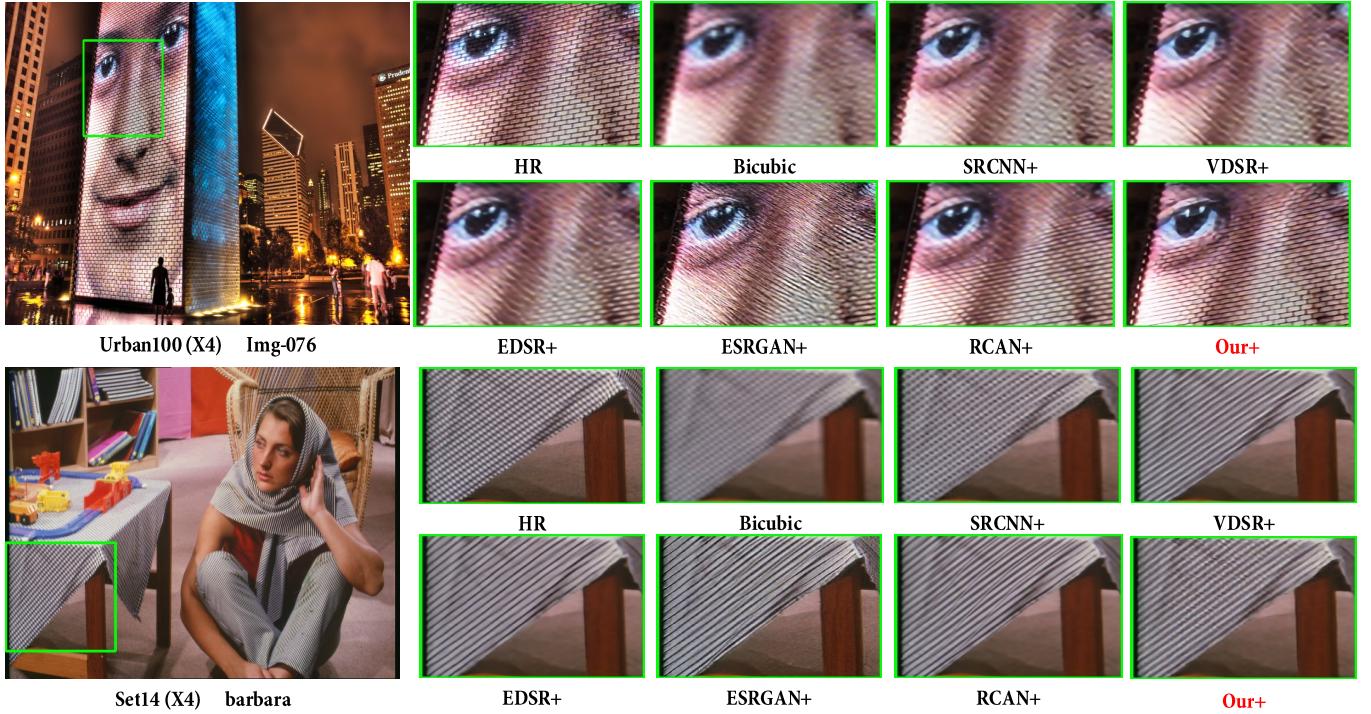


Fig. 10. Examples of visual comparison (X4) on BI degradation. We present the results from finetuned models. Our DuRCAN restores more details, such as the tablecloth in Set14 "barbara" and the crosshatched pattern in Urban100 Img-076.

In this subsection, both the quantitative results and qualitative visual comparisons validate that: (1) the proposed DuRCAN has superiority to handle complicated real-world degradation; (2) the generative learning of proposed DD-GAN effectively enlarges the variety of degradations from limited real-captured data and the generated LR images from DD-GAN greatly enhance the robustness of SISR models.

#### D. SOTA Comparisons on Typical Bicubic Datasets

In Section V-C, we have proved that our proposed model has superiority over existing SISR models on real camera-screen degradation. It should be noticed that all those SOTAs were originally well-designed for typical "super resolution" task. Hence, we train all the methods on the original DIV2K

dataset and evaluate them on four typical bicubic interpolation datasets to present a systemic comparison and verify the generalization of our joint model, including Set5 [34], Set14 [35], BSD100 [36] and Urban100 [9]. Besides directly citing the SISR results from the original papers of SRCCNN [15], VDSR [18], EDSR [19], ESRGAN [27] and RCAN [21], we utilize the pretrained camera-screen models in Section V-C to initialize the weights of parameters and finetune them on the typical BI DIV2K dataset.

1) *Quantitative Analyses:* As the quantitative results shown in Table III, we can see that after using the pretrained weights from more complicated camera-screen degradation, the performance of all the models get improved (seen in the 8th to 13th rows of Table III). This provides a new attempt for



Fig. 11. Examples of visual comparison (X4) on raw camera-screen degradation. The real-world images outside our dataset are not pre-processed. The RCAN and ESRGAN trained with Cam-ScreenSR are sensitive to the high-frequency Moiré pattern and can easily get corrupt by real-world disturbance. And our DuRCAN is more stable to produce the best results with less artifacts, sharper edges and better color enhancement.

the improvement on SISR tasks that the SISR model can be appropriately pretrained with complicated degraded images and then be finetuned in specific scene. Moreover, training with (DuRCAN+) and without (DuRCAN) pretrained initialization, our models both get competitive results among those well-designed SOTA methods for BI degradation. Although achieving a little bit lower performance on Set5 and BSD100 datasets, the DuRCAN outperforms on Set14 and Urban100 datasets. It is worth noting that the parameter quantities of the state-of-the-art perceptual-driven model ESRGAN and pixel-oriented model RCAN are 17000K and 15592K in x4 scale, while the parameter quantities of DuRCAN are only 5453K in x4 scale, which is about one-third (1/3) of those of ESRGAN and RCAN. With fewer parameters, our DuRCAN achieves the competitive and even better performance than those well-designed BI SISR models, which validates the effectiveness of our method.

2) *Qualitative Analyses*: In Fig. 10, we present the bicubic baselines and 6 better performed models with pretrained weights on two examples, including the Set14 "barbara" and Urban100 Img-076. Benefiting from the network structure and added Laplacian loss, many of the regular patterns that are undersampled and incorrectly reconstructed in the other methods are dealt with very well by our DuRCAN, such as the tablecloth in Set14 barbara and the crosshatched pattern in Urban100 Img-076. The Set14 and Urban100 datasets contain more regular graphics. The better quantitative and qualitative results on those two datasets validate that our method has superiority to enrique more details for regular graphics with sharper edges and less artifacts.

In this subsection, the experimental results validate that: (1) not only for complicated camera-screen degradation, the DuRCAN is also competitive with less parameters for typical bicubic interpolation SISR task; (2) for better performance, the SISR model can be appropriately pretrained with complicated

degraded images and then be finetuned in specific scene.

#### E. Qualitative Evaluations on Real-world Photographs

To further validate the generalization capability, we compare our model against SOTA models on more general real-world scenes. Since there are no ground-truth for the real-captured image, we conduct the perceptual judgement. Two scenes, including raw camera-screen degradation and landscape photographs captured by smartphone, are presented as follows. All the photographs are the original versions without data rectification.

1) *Raw Camera-Screen Degradation*: To further validate the generalization capability of our Cam-ScreenSR dataset and proposed joint model, we should compare our model with other models on raw camera-screen degraded images outside our dataset. We randomly selected high-quality images from the Google search and displayed them on the Lenovo X1 laptop. The degraded images are captured by an iPhone 11 smartphone and are directly fed into the trained models in Section V-C without data rectification. We selected three better performed methods trained on Cam-ScreenSR dataset for visual comparison, including ESRGAN [27], RCAN [21] and our proposed DuRCAN.

The visual examples of three models and bicubic baselines are presented in Fig. 11. As the visual results show, the models trained with our Cam-ScreenSR have advantages to handle the noises, color distortion and blurs influenced by the screen and camera. Without the image rectification in Section III, including the alignment, interpolation and average with continuous shoots, the raw images are more degraded. The third row in Fig. 11 reveals the RCAN and ESRGAN trained with Cam-ScreenSR are sensitive to the high-frequency Moiré pattern and can easily get corrupt by real-world disturbance. Compared to previous well-designed SOTAs for BI



Fig. 12. Examples of visual comparison for x4 SR images on landscape photographs captured by an iPhone 11 smartphone. We frozed the DuRBs and slightly finetuned the RCABs of Cam-ScreenSR trained models with BI images. Our DuRCAN delivers more comfortable visual results, especially with the excellent color enhancement.

TABLE IV

COMPARISONS ON PSNR/SSIM VALUES, MODEL PARAMETERS (PYTORCH-VERSION) AND AVERAGE INTERFACE TIME. THE TESTING SET 1 (SAMSUNG S27R350 + CANON 760D) WITH SCALE FACTOR  $\times 4$  IS USED FOR MEASUREMENT. ALL THE RUNNING TIME IS CALCULATED BY A NVIDIA TITAN XP GPU.

	SRCNN [15]	VDSR [18]	EDSR [19]	ESRGAN [27]	RCAN [21]	DuRCAN
Para.	<b>15K</b>	665K	43090K	17000K	15592K	5453K
Sec.	4.073	<b>0.721</b>	2.163	2.575	1.659	1.071
PSNR	17.83	19.97	21.29	23.74	23.81	<b>24.82</b>
SSIM	0.6858	0.7044	0.7051	0.7148		<b>0.7271</b>

degradation, our DuRCAN has great robustness to handle more complicated situations in real-world degradation.

2) *Landscape Photographs Captured by Smartphone*: The super-resolution is an useful application for mobile phones to provide more comfortable visual experience for customers. To estimate the reliability and practicability of our method for real-captured images, we also evaluate our model on landscape photographs captured by Smartphone. In this scene, we captured real-world landscape photographs with an iPhone 11. We also compare the proposed DuRCAN with two better performed SOTA methods, including ESRGAN [27] and RCAN [21]. To control the color enhancement appropriately, the Cam-ScreenSR trained DuRCAN was finetuned with BI degraded images slightly. Specifically, we froze the parameters of dual residual block and finetuned two residual channel attention blocks. The network training was early stopped after 100 iterations.

The visual examples of three models and bicubic baselines are presented in Fig. 12. As the visual results show, not only recovering more details, our finetuned model also enriches the image color appropriately. After finetuning the targeted residual channel attention blocks, our DuRCAN can conduct color enhancement to produce bluer sea and greener grass for example, while keeping the restoration ability. With the premise of recovering sufficient details in photographs, the

appropriate color enhancement can provide more comfortable visual experience in real-world photography.

#### F. Comparisons on Computational Cost

For fair comparison, we use the 6 Cam-ScreenSR trained models in Section V-C, including SRCNN [15], VDSR [18], EDSR [19], ESRGAN [27], RCAN [21] and proposed DuRCAN, to evaluate the runtime on the computer with 2.2 GHz Intel i7 CPU and 1 NVIDIA Titan Xp GPU. The PSNR/SSIM values, model parameters and average interface time on testing set 1 (Samsung S27R350 + Canon 760D) are listed in Table IV. It's clear that SRCNN has fewest parameters but achieve worst reconstruction performance with much slower running speed compared with other methods. Although VDSR recovers SR images with the fastest speed, this method still produces worse SR results than complicated models with more parameters. The proposed DuRCAN can achieve superior PSNR/SSIM values with faster reconstruction speed than ESRGAN and RCAN.

#### G. Ablation Study

As discussed in Section IV, our joint solution contains four main components, including residual channel attention blocks (RCAB), dual residual blocks (DuRB), the added Laplacian

TABLE V

SISR RESULTS OF MODELS WITH (DURCAN) AND WITHOUT (*Base*) RESIDUAL CHANNEL ATTENTION BLOCKS ON THREE CAMERA-SCREEN DEGRADED TESTING DATASETS. THE DD-GAN DATA AUGMENTATION IS ALSO JOINTLY APPLIED.

Cam-ScreenSR Testing Set	Scale	<i>Base</i> PSNR / SSIM	<i>Base</i> + RCABs (DuRCAN) PSNR / SSIM
1	X4	24.78 / 0.7265	<b>24.82 / 0.7271</b>
2	X4	23.82 / 0.7146	<b>24.51 / 0.7240</b>
3	X4	21.47 / 0.7013	<b>22.19 / 0.7103</b>

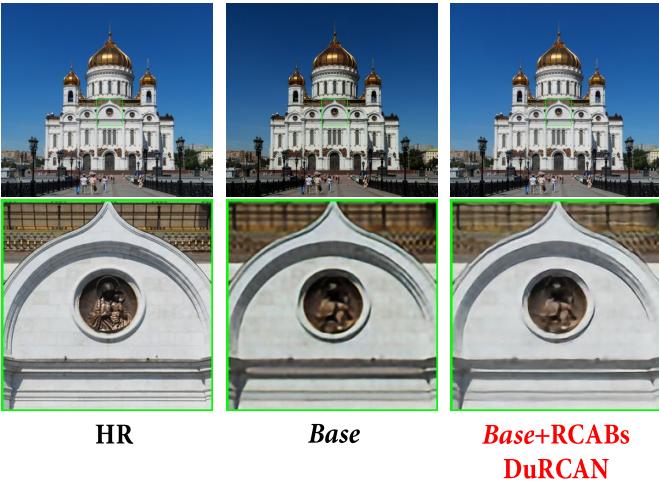


Fig. 13. After removing RCAB, the SISR result of DIV2K Img-0879 produced by *Base* model has acceptable noises and artifacts, but it has poor color distortion, which means RCABs focus more on the color calibration.

loss and the data augmentation with downsampling degradation GAN (DD-GAN). We have validated the effectiveness of DD-GAN in Section V-C. Therefore, we will conduct ablation experiments of the rest three components as follows.

1) *Residual Channel Attention Blocks (RCAB)*: To analyse the effects of residual channel attention mechanism in our DuRCAN, firstly, we compare the quantitative performance between the DuRCANs with and without RCABs. The RCABs removed DuRCAN (*Base*) and intact DuRCAN were trained on camera-screen degraded training set and evaluated on three testing sets with DD-GAN.

The quantitative results are listed in Table V. It can be seen that the quantitative results of *Base* model and DuRCAN are similar in testing set 1. In testing set 2 and 3, the intact DuRCAN outperforms *Base* model, which verifies the effectiveness of RCABs. As we have described in Section III, the LR images of training set and testing set 1 are degraded with same equipment. But the degradation patterns of testing set 2 and 3 are different from training set. It means that after involving the residual channel attention mechanism, the generalization of our model with different camera-screen combinations is greatly enhanced.

Moreover, the visual evaluation can clearly present the effects of RCABs. Specifically, we remove the RCABs from the intact trained DuRCAN and list the visual results in Fig. V. The RCAB-removed result has acceptable noises and artifacts,

TABLE VI

SISR RESULTS OF DURCAN WITH DIFFERENT DURB CONFIGURATIONS ON THREE CAM-SCREENSR TESTING DATASETS. THE DD-GAN DATA AUGMENTATION IS ALSO JOINTLY APPLIED.

Model	Para.	Sec.	Dual kernel size
DuRCAN-6_s	1978K	0.4756	[3, 3], [5, 3], [7, 5], [7, 5], [7, 3], [5, 3]
DuRCAN-6	3518K	0.6996	[5, 3], [7, 5], [11, 7], [11, 7], [11, 5], [7, 5]
DuRCAN-12	5453K	1.071	[5, 3], [5, 3], [7, 3], [7, 5], [11, 5], [11, 7], [11, 7], [11, 5], [7, 5], [7, 3], [5, 3]
DuRCAN-18	9878K	1.529	[5, 3], [5, 3], [5, 3], [7, 5], [7, 5], [7, 5], [11, 7], [11, 7], [11, 7], [11, 7], [11, 7], [11, 7], [11, 5], [11, 5], [11, 5], [7, 5], [7, 5], [7, 5]
Model	Scale	Camera-Screen Testing Set	PSNR / SSIM
DuRCAN-6_s	X4	1 2 3	23.89 / 0.7129 23.95 / 0.7163 20.89 / 0.6822
DuRCAN-6	X4	1 2 3	24.21 / 0.7205 24.26 / 0.7207 21.45 / 0.6984
DuRCAN-12	X4	1 2 3	<b>24.82 / 0.7271</b> <b>24.51 / 0.7240</b> 22.19 / 0.7103
DuRCAN-18	X4	1 2 3	<b>24.84 / 0.7275</b> <b>24.56 / 0.7242</b> <b>22.21 / 0.7104</b>

but it has poor color distortion compared with HR image and SR result of DuRCAN, which reveals the color adjustment ability of *Base* model is greatly weakened. Therefore, the residual channel attention blocks of our model have the limited influence on the SR definition, but focus more on the color calibration. It also provides us a way to specifically finetune the RCABs to control the color enhancement ability of our model.

2) *Dual Residual Blocks (DuRB)*: For SISR task, receptive field determines whether the ability of model is good enough to explore the relationships of neighbor pixels and recover the missing contextual information [41]. As the pooling operation will discard the image details, existing SISR models focus on increasing the network depth and enlarging the convolutional kernel size [15], [18]–[21], [27], [41], [42]. The configuration of dual residual blocks (DuRBs) determines the receptive field of our DuRCAN. Therefore, we conduct the ablation comparisons between different depth and kernel size settings. We changed the depth of DuRBs  $d = 6, 12, 18$ , named as DuRCAN-6, DuRCAN-12 (our proposed model) and DuRCAN-18 to evaluate their performance respectively. And we also structured a DuRCAN-6 with smaller kernel sizes, named as DuRCAN-6\_s. The model configurations of dual kernel size are listed in Table VI. All those four models were evaluated in three camera-screen degraded testing sets.

The quantitative results are shown in Table VI. We can see that keeping the same depth of DuRBs, the DuRCAN-6

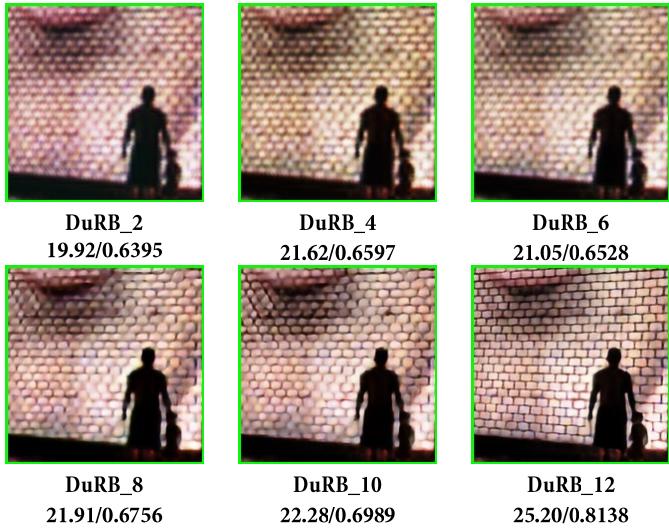


Fig. 14. As the depth of DuRB increases, the X4 SR results of Urban100 Img-076 broadly become clearer and contain more textures, which reveals the dual residual operations pay attention on restoring the details.

TABLE VII  
SISR RESULTS OF DURCAN WITH DIFFERENT LOSS FUNCTIONS ON CAM-SCREENSR TESTING SETS. THE DD-GAN DATA AUGMENTATION IS ALSO JOINTLY APPLIED.

Cam-ScreenSR Testing Set	Scale	DuRCAN + $L_1$ PSNR / SSIM	DuRCAN + $L_{SR}$ PSNR / SSIM
1	X4	24.03 / 0.7191	<b>24.82 / 0.7271</b>
2	X4	23.84 / 0.7123	<b>24.51 / 0.7240</b>
3	X4	20.95 / 0.7028	<b>22.19 / 0.7103</b>

performs better than DuRCAN-6\_s in all three testing sets, which reveals the larger kernel size is effective for our DuRCAN structure. With different depth, the deeper DuRCAN-18 performed better than the shallow DuRCAN-6 and DuRCAN-12. It should be noticed that when increasing the depth of DuRBs, the performance of DuRCAN-18 gets marginal improvement than DuRCAN-12, but the model parameter and execution time are greatly increased. Considering that the DuRCAN are jointly trained with generative learning network DD-GAN, we choose the configuration of DuRCAN-12 as our proposed method to balance the computing cost and SISR performance.

Moreover, we visualize the recovering features of the DuRB to verify the effects of stacked DuRBs. In order not to involve extreme color distortion, the BI trained DuRCAN in Section V-D is applied. We recover the output features from every two layers of DuRBs to get the SR results. As Fig. 14 shows, the deeper DuRB produces the better SR result. Combining the previous ablation experiment of residual channel attention mechanism, it can be seen that the dual residual operations pay more attention on restoring the details from complicated degradations. Without the limitations of computing resources and execution time, the deeper DuRG can be involved to improve the SR definition.

3) *Laplacian Loss*: It has been widely acknowledged that an image consists of the high-frequency and low-frequency

TABLE VIII  
SISR RESULTS OF DURCAN WITH DIFFERENT LOSS FUNCTIONS ON TYPICAL BI TESTING SETS. THE DD-GAN DATA AUGMENTATION IS ALSO APPLIED.

Typical BI Testing Set	Scale	DuRCAN + $L_1$ PSNR / SSIM	DuRCAN + $L_{SR}$ PSNR / SSIM
Set5 [34]	X4	32.57 / 0.8973	<b>32.60 / 0.8982</b>
Set14 [35]	X4	28.85 / 0.7890	<b>28.93 / 0.7900</b>
BSD100 [36]	X4	27.53 / 0.7369	<b>27.64 / 0.7415</b>
Urban100 [9]	X4	26.82 / 0.8077	<b>26.92 / 0.8116</b>

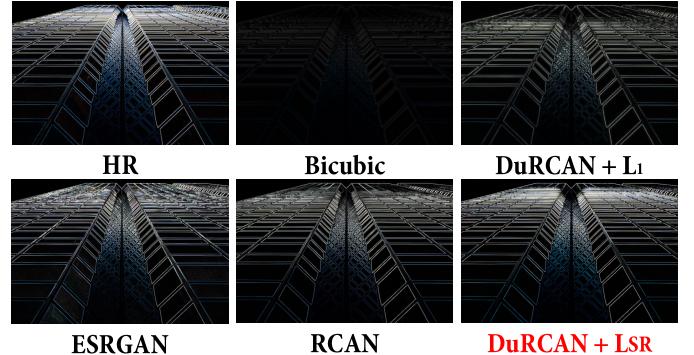


Fig. 15. Visual comparisons of the edges processed by second order differential Laplacian operation. The DuRCAN trained with combined  $L_{SR}$  loss can reconstruct sharper edges with more texture details and less noises.

messages. The low-frequency messages deliver the basic gray level of the image. And the high-frequency messages deliver the sharp changes of the pixel intensity, which reveal more details of the image [67]. Therefore, it's important to supervise our model to restore the high-frequency edges. Inspired by the second order differential Laplace operation commonly used in image processing [33], we involve the high-frequency supervised Laplacian loss function to minimize the  $L_1$  distance between the HR and SR image processed by Laplacian operation. As Eq 18 presents, the Laplacian Loss  $L_{lap}$  and  $L_1$  loss are weighted combined as  $L_{SR}$ . To demonstrate the effectiveness of Laplacian loss, we compared the performance of the DuRCAN with  $L_1$  and  $L_{SR}$  on Cam-ScreenSR dataset respectively. The DD-GAN data augmentation was applied to correspond with the training method in Sec V-C. We also fine-tuned the models on typical BI DIV2K dataset and evaluated their performance on Set5 [34], Set14 [35], BSD100 [36], Urban100 [9] datasets.

The quantitative results on camera-screen and typical BI degraded datasets are respectively shown in Table VII and Table VIII. After adding the Laplacian loss, the performance of DuRCAN get improved on both Cam-ScreenSR and typical BI datasets. As visual examples shown in Fig 1, the camera-screen degradation contains less high-frequency information than bicubic downsampling, because of much more blurs, noises and color distortion. Therefore, the quantitative improvements on Cam-ScreenSR dataset are greater than those on BI datasets after involving the high-frequency supervised loss  $L_{lap}$ .

Moreover, we visualize the edges of different methods processed by second order differential Laplacian operation in Fig. 15, including the bicubic baseline, ESRGAN [27],

RCAN [21] and our DuRCAN trained with  $L_1$  loss and  $L_{SR}$  loss. It can be seen that the bicubic SR image contains few high-frequency edges. The RCAN and DuRCAN trained with  $L_1$  loss models outperform the bicubic baseline. Although ESRGAN enriches texture details by adding high-frequency noises, those uncontrollable noises severely limit the robustness in real-world images (seen in Fig. 11 and Fig. 12). The DuRCAN trained with combined  $L_{SR}$  loss can reconstruct sharper edges with more texture details. All the qualitative and quantitative results prove that the added Laplacian loss can efficiently supervise our proposed joint model to smooth the noises and sharpen the edges of SISR results.

## VI. DISCUSSION AND FUTURE WORK

Restoring the low-resolution images from the camera-screen degradation is a more sophisticated task. It's different from the pure "super-resolution", where the images are prefiltered by downsampling kernels. The camera-screen SISR task has to jointly solve problems such as denoising, sharpening the edge, fixing color distortion, etc.. Although those degradations have a long research history in image processing, they are inevitable for real-world SISR application. We have proved that involving them is benefit for the performance of SISR models. As our model focuses on SISR tasks, it's normal to choose previous SISR models for comparisons and we also conduct fair experiments on the same datasets to validate the effectiveness of our model. We believe that the joint solution for camera-screen SISR scene is valuable and should be encouraged, because the user can directly get the high-resolution output without multi-step processing.

In different real-world environments, there exists more extreme camera-screen degradation. In order not to decrease the stability of the proposed joint model by more extreme degradation, our data acquisition strategy partly simplifies the degradation: the repeated photo capturing partly weakens the stroboscopic effect; the image rectification and multiple image averaging partly weaken the noises, such as Moiré patterns; the monitors are set to the Standard mode to not cause extreme color distortion, and etc.. In the future work, more camera-screen combinations, more extreme degradation can be involve to expand our dataset and explore the novel model for better generalization. Moreover, although our Cam-ScreenSR dataset and joint model focus on the SISR task, they can be used in other image restoration tasks, such as image denoising, deblurring and color correction. We also leave it as our future work.

## VII. CONCLUSION

There exists a long standing problem that the SISR model trained with synthetic degradation has poor generalization on real-world image. In this paper, we made an first attempt to involve the degradation of camera-screen device for SISR task and proposed a data acquisition strategy to establish a baseline, Cam-ScreenSR dataset. Although the camera-screen degradation is complicated, our proposed joint model has the great ability to handle those degradations, which produces better visual results than previous SOTA models with sharper

edge, less artifacts and appropriate color enhancement. We believe that our decent SISR solution will provide clearer visual experience for general users when they use their mobile phones to record contents on screens for convenience, simplicity and efficiency. Meanwihle, the appropriate color enhancement can also be accomplished, which is more easily perceived by human eyes.

## ACKNOWLEDGMENT

The authors would like to thank...

## REFERENCES

- [1] D. Glasner, S. Bagon, and M. Irani, "Super-resolution from a single image," in *2009 IEEE 12th international conference on computer vision*. IEEE, pp. 349–356.
- [2] J. Yang, Z. Wang, Z. Lin, S. Cohen, and T. S. Huang, "Coupled dictionary training for image super-resolution," *IEEE Transactions on Image Processing*, vol. 21, no. 8, pp. 3467–3478, 2012.
- [3] H. Demirel and G. Anbarjafari, "Discrete wavelet transform-based satellite image resolution enhancement," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 49, no. 6, pp. 1997–2004, 2011.
- [4] B. K. Gunturk, A. U. Batur, Y. Altunbasak, M. H. Hayes, and R. M. Mersereau, "Eigenface-domain super-resolution for face recognition," *IEEE Transactions on Image Processing*, vol. 12, no. 5, pp. 597–606, 2003.
- [5] 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.
- [6] M. Bertero and P. Boccacci, *Introduction to inverse problems in imaging*. CRC press, 1998.
- [7] K. I. Kim and Y. Kwon, "Single-image super-resolution using sparse regression and natural image prior," *IEEE transactions on pattern analysis and machine intelligence*, vol. 32, no. 6, pp. 1127–1133, 2010.
- [8] G. Freedman and R. Fattal, "Image and video upscaling from local self-examples," *ACM Transactions on Graphics (TOG)*, vol. 30, no. 2, p. 12, 2011.
- [9] J.-B. Huang, A. Singh, and N. Ahuja, "Single image super-resolution from transformed self-exemplars," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 5197–5206.
- [10] W. Yang, X. Zhang, Y. Tian, W. Wang, J. Xue, and Q. Liao, "Deep learning for single image super-resolution: A brief review," *IEEE Transactions on Multimedia*, vol. 21, no. 12, pp. 3106–3121, 2019.
- [11] R. Keys, "Cubic convolution interpolation for digital image processing," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 29, no. 6, pp. 1153–1160, 1981.
- [12] C. E. Duchon, "Lanczos filtering in one and two dimensions," *J. Appl. Meteorol.*, vol. 18, no. 8, pp. 1016–1022, 1979.
- [13] S. Dai, M. Han, W. Xu, Y. Wu, Y. Gong, and A. K. Katsaggelos, "Soft-cuts: A soft edge smoothness prior for color image super-resolution," *IEEE Transactions on Image Processing*, vol. 18, no. 5, pp. 969–981, 2009.
- [14] X. a. Q.Yan, Y.Xu, "Single image superresolution based on gradient profile sharpness," *IEEE Transactions on Image Processing*, vol. 24, no. 10, pp. 3187–3202, 2015.
- [15] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 2, pp. 295–307, 2015.
- [16] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *European conference on computer vision*. Springer, 2016, pp. 391–407.
- [17] 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.
- [18] 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.
- [19] B. Lim, S. Son, H. Kim, S. Nah, and K. Mu Lee, "Enhanced deep residual networks for single image super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2017, pp. 136–144.

- [20] 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.
- [21] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 286–301.
- [22] T. Dai, J. Cai, Y. Zhang, S.-T. Xia, and L. Zhang, "Second-order attention network for single image super-resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 11065–11074.
- [23] C. Chen, Z. Xiong, X. Tian, Z.-J. Zha, and F. Wu, "Camera lens super-resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 1652–1660.
- [24] X. Zhang, Q. Chen, R. Ng, and V. Koltun, "Zoom to learn, learn to zoom," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 3762–3770.
- [25] 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," *arXiv preprint arXiv:1904.00523*, 2019.
- [26] K. Zhang, W. Zuo, and L. Zhang, "Learning a single convolutional super-resolution network for multiple degradations," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 3262–3271.
- [27] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. Change Loy, "EsrGAN: Enhanced super-resolution generative adversarial networks," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 0–0.
- [28] A. Bulat, J. Yang, and G. Tzimiropoulos, "To learn image super-resolution, use a gan to learn how to do image degradation first," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 185–200.
- [29] R. Zhou and S. Susstrunk, "Kernel modeling super-resolution on real low-resolution images," in *Proceedings of the IEEE International Conference on Computer Vision*, 2019, pp. 2433–2443.
- [30] 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 *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2017, pp. 114–125.
- [31] E. Wengrowski and K. J. Dana, "Light field messaging with deep photographic steganography," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2019, pp. 1515–1524.
- [32] X. Liu, M. Suganuma, Z. Sun, and T. Okatanai, "Dual residual networks leveraging the potential of paired operations for image restoration," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 7007–7016.
- [33] A. Rosenfeld, *Digital picture processing*. Academic press, 1976.
- [34] M. Bevilacqua, A. Roumy, C. Guillemot, and M. L. Alberi-Morel, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," 2012.
- [35] 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.
- [36] P. Arbelaez, M. Maire, C. Fowlkes, and J. Malik, "Contour detection and hierarchical image segmentation," *IEEE transactions on pattern analysis and machine intelligence*, vol. 33, no. 5, pp. 898–916, 2010.
- [37] H. A. Aly and E. Dubois, "Image up-sampling using total-variation regularization with a new observation model," *IEEE Transactions on Image Processing*, vol. 14, no. 10, pp. 1647–1659, 2005.
- [38] Z. Xiong, X. Sun, and F. Wu, "Robust web image/video super-resolution," *IEEE transactions on image processing*, vol. 19, no. 8, pp. 2017–2028, 2010.
- [39] L. He, H. Qi, and R. Zaretzki, "Beta process joint dictionary learning for coupled feature spaces with application to single image super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2013, pp. 345–352.
- [40] W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example-based super-resolution," *IEEE Computer graphics and Applications*, no. 2, pp. 56–65, 2002.
- [41] J. Kim, J. Kwon Lee, and K. Mu Lee, "Deeply-recursive convolutional network for image super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1637–1645.
- [42] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang *et al.*, "Photo-realistic single image super-resolution using a generative adversarial network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4681–4690.
- [43] J. Johnson, A. Alahi, and L. Feifei, "Perceptual losses for real-time style transfer and super-resolution," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2016, pp. 694–711.
- [44] N. Efrat, D. Glasner, A. Apartsin, B. Nadler, and A. Levin, "Accurate blur models vs. image priors in single image super-resolution," in *Proceedings of the IEEE International Conference on Computer Vision*, 2013, pp. 2832–2839.
- [45] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *ICLR*, 2015.
- [46] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.
- [47] 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.
- [48] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," *Communications of The ACM*, vol. 60, no. 6, pp. 84–90, 2017.
- [49] G. Huang, Z. Liu, L. V. Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2261–2269.
- [50] K. He, G. Gkioxari, P. Dollar, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 2980–2988.
- [51] M. Haris, G. Shakhnarovich, and N. Ukita, "Deep back-projection networks for super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 1664–1673.
- [52] J. Hu, L. Shen, S. Albanie, G. Sun, and E. Wu, "Squeeze-and-excitation networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2019.
- [53] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 6450–6458.
- [54] S. Woo, J. Park, J. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," pp. 3–19, 2018.
- [55] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," pp. 5998–6008, 2017.
- [56] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.
- [57] M. A. Fischler and R. C. Bolles, "Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography," *Communications of the ACM*, vol. 24, no. 6, pp. 381–395, 1981.
- [58] K. Zhang, W. Zuo, and L. Zhang, "Deep plug-and-play super-resolution for arbitrary blur kernels," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2019, pp. 1671–1681.
- [59] J. Gu, H. Lu, W. Zuo, and C. Dong, "Blind super-resolution with iterative kernel correction," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2019, pp. 1604–1613.
- [60] A. Jolicoeur-Martineau, "The relativistic discriminator: a key element missing from standard gan," in *ICLR 2019 : 7th International Conference on Learning Representations*, 2019.
- [61] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," pp. 2672–2680, 2014.
- [62] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 2472–2481.
- [63] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Deep laplacian pyramid networks for fast and accurate super-resolution," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017, pp. 5835–5843.
- [64] R. Müller, S. Kornblith, and G. E. Hinton, "When does label smoothing help?" in *Advances in Neural Information Processing Systems*, 2019, pp. 4696–4705.
- [65] Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli *et al.*, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [66] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *ICLR*, 2014.
- [67] R. C. Gonzalez, R. E. Woods, and S. L. Eddins, *Digital image processing using MATLAB*. Pearson Education India, 2004.