

# Perception-Preserving Convolutional Networks for Image Enhancement on Smartphones

Zheng Hui<sup>1</sup>, Xiumei Wang<sup>1</sup>, Lirui Deng<sup>2</sup>, and Xinbo Gao<sup>1</sup>

School of Electronic Engineering, Xidian University, Xi'an, China  
[zheng.hui@aliyun.com](mailto:zheng.hui@aliyun.com), [wangxm@xidian.edu.cn](mailto:wangxm@xidian.edu.cn), [xbgao@mail.xidian.edu.cn](mailto:xbgao@mail.xidian.edu.cn)  
Department of Computer Science and Technology, Tsinghua University  
[Rayal.deng@outlook.com](mailto:Rayal.deng@outlook.com)

**Abstract.** Although the configuration of smartphone cameras is getting better and better, the quality of smartphone photos still cannot match DSLR camera photos due to the limitation of physical space, hardware and cost. In this work, we present a fast and accurate image enhancement approach based on generative adversarial nets, which elevates the quality of photos on smartphones. We propose the lightweight local residual convolutional network to learn the mapping between ordinary photos and DSLR-quality images. To make the generated images look real, we introduce the perception-preserving measurement error, which comprises content, color, and adversarial losses. Especially, the content loss is constituted of contextual and SSIM losses, which maintains the natural internal statistics and the structure of images. In addition, we introduce the knowledge transfer strategy to ensure the high performance of the proposed network. The experiments demonstrate that our proposed method produces better results compared with the state-of-the-art approaches, both qualitatively and quantitatively. The code is available at <https://github.com/Zheng222/PPCN>.

**Keywords:** Image enhancement · Perception-preserving measurement error · Knowledge transfer.

## 1 Introduction

Continuous improvement for the quality of tiny camera sensors and lens makes smartphone photography come into vogue. However, from the viewpoint of aesthetics, photos captured by mobile phones still cannot attain the DSLR-quality because of their compact sensors and lens. Larger sensors are conducive to improving image quality, reducing noise and shooting night scenes. In order to automatically translate the low-quality mobile phone pictures into the high-quality images, Andrey *et al.* [11] propose an end-to-end deep learning approach uses a composite perceptual error function that combines content, color, and texture losses, where the content loss is simply defined as the VGG loss based on the ReLU activation layers of the pre-trained 19-layer VGG network described in [25]. The authors also present a weakly-supervised approach in [12] to overcome the requirement of matched input/target training image pairs. Though the

above methods have achieved remarkable results, they still have the deficiencies to be addressed. One of the limitations of the existing CNN-based methods is that researchers always trying to deepen the generator network to reach better performance, which leads to a substantial computational cost and memory consumption which will further bring increasing power consumption. Therefore, these methods are not conducive to real mobile phone applications. The other cause is the artifacts and amplified noises appeared on the processed images in [11], which affects the user experience.

To tackle these issues, we propose a novel CNN-based image enhancement approach, which introduces the teacher-student information transfer to boost the performance of the compact student network and contextual loss that proposed in [23,22] to preserve the nature of images. Moreover, we combine adversarial (GAN) [9], color, total variation losses to learn photo-realistic image quality. Finally, to guarantee the structural preservation of the enhanced images, we employ the SSIM loss as the constraint term.



**Fig. 1.** DPED image enhanced by our method.

The main contributions of the perception-preserving CNN are summarized as follows:

- We propose a novel compact network for single image enhancement as illustrated in Fig. 3, which adopts 1-D separable kernels and dilated convolutions to expand the network receptive field.
- We exploit knowledge transfer to promote the performance of the student network.
- We employ contextual and SSIM losses to maintain the nature of the image.
- The effective network architecture for single image super-resolution is devised as shown in Fig. 2, which can fast super-resolve the low resolution images.
- Our proposed method achieves superior performance compared with the state-of-the-art methods.

## 2 Related work

The problem of image quality enhancement is part of the image-to-image translation task. In this section, we introduce several related works from the image transformation field.

### 2.1 Image enhancement

We build our solution upon recent advances in image-to-image translation networks. Ignatov et al. [11] propose an end-to-end enhancer achieving photo-realistic results for arbitrary image resolutions by combining content, texture and color losses. However, it still has its disadvantages, such as slower inference speed, results with artifacts (color deviations and too high contrast levels) and noises. The authors also present WESPE [12], a weakly supervised solution for the image quality enhancement problem. This approach is trained to map low-quality photos into the domain of high-quality photos without requiring labeled data, only images from two different domains are needed.

### 2.2 Image super-resolution

Single image super-resolution aims to recover the visually pleasing high-resolution (HR) image from a low resolution (LR) one. Dong et al. [4,5] first exploit a three-layer convolutional neural network, named SRCNN, to approximate the complex nonlinear mapping between the LR image and the HR counterpart. To reduce computational complexity, the authors propose a fast SRCNN (FSRCNN) [6], which adopts the transposed convolution to execute upscaling operation at the output layer. Kim et al. [15] present a very deep super-resolution network (VDSR) with residual architecture to achieve eminent SR performance, which utilizes broader contextual information with a larger model capacity. Lai et al. propose the Laplacian pyramid super-resolution network (LapSRN) [17] to progressively reconstruct the sub-band residuals of high-resolution images. Tai et al. [26] present a deep recursive residual network (DRRN), which employs the parameters sharing strategy. The authors also propose a very deep end-to-end persistent memory network (MemNet) [27] for image restoration task, which tackles the long-term dependency problem in the previous CNN architectures. The aforementioned approaches focus on promoting the objective evaluation index, while Ledig et al. [18] achieve the photo-realistic results on super-resolution task by using a VGG-based loss function [14] and adversarial networks [9].

### 2.3 Image deraining

Rain is a common weather in our life. Since it can affect the line of sight, it is a significant task to remove the rain and recover the background from rain images for post image processing. Recently, several deep learning based deraining methods achieve promising performance. Fu et al. [7,8] first introduce deep learning methods to the deraining problem. Yang et al. [30] design a deep recurrent dilated network to jointly detect and remove rain streaks. Zhang et al. [34] propose a density-aware image deraining method with the multi-stream densely connected

network for jointly rain-density estimation and deraining. Li et al. [19] design a scale-aware multi-stage recurrent network that estimates rain streaks of different sizes and densities individually.

## 2.4 Contextual loss

Mechrez et al. [23,22] design a loss function that can measure the dissimilarity between a generated image  $x$  and a target image  $y$ , represented by feature sets  $X = \{x_i\}$  and  $Y = \{y_i\}$ , respectively. Let  $A_{ij}$  denote the affinity between features  $x_i$  and  $y_j$ . The Contextual loss is defined as:

$$\mathcal{L}_{CX}(x, y) = -\log \left( \frac{1}{M} \sum_j \max_i A_{ij} \right) \quad (1)$$

The affinities  $A_{ij}$  are defined in a way that promotes a single close match of each feature  $y_i$  in  $X$ . To implement this, first the Cosine distances  $d_{ij}$  are computed between all pairs  $x_i, y_j$ . The distances are then normalized:  $\tilde{d}_{ij} = d_{ij} / (\min_k d_{ik} + \epsilon)$  (with  $\epsilon = 1e-5$ ), and finally the pairwise affinities  $A_{ij} \in [0, 1]$  are defined as:

$$A_{ij} = \frac{\exp(1 - \tilde{d}_{ij}/h)}{\sum_l \exp(1 - \tilde{d}_{il}/h)} = \begin{cases} \approx 1 & \text{if } \tilde{d}_{ij} \ll \tilde{d}_{il} \quad \forall l \neq j \\ \approx 0 & \text{otherwise} \end{cases} \quad (2)$$

where  $h > 0$  is a bandwidth parameter.

## 2.5 Knowledge transfer

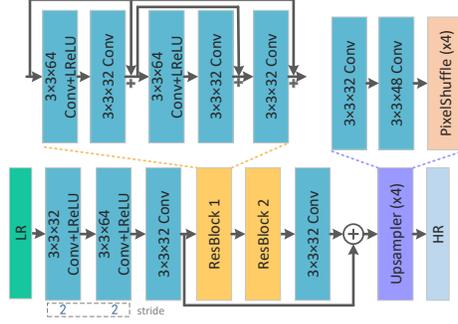
This line of research aims at distilling knowledge from a complicated teacher model into a compact student model without performance drop. Recently, Zagoruyko et al. [32] present several ways of transferring attention from one network to another over several image recognition datasets. Yim et al. [31] propose a novel approach to generate distilled knowledge from the DNN, which determines the distilled knowledge as the flow of the solving procedure calculated with the proposed FSP matrix.

## 3 Proposed method

In this section, we first describe the proposed solution for single image super-resolution (SR) task and then introduce the image quality enhancement on smartphones.

### 3.1 Single image super-resolution

As shown in Fig. 2, the presented SR method first adopts two convolutional layers with stride 2 to reduce the resolutions of feature maps. This way can



**Fig. 2.** The schematics of the proposed network for image super-resolution.

dramatically decrease the computational cost during the testing phase. The following operations are two residual blocks, each of them consists of two residual modules and one transition convolution. Finally, we employ a global residual for fast model optimization and an upsampler that is composed of two convolutions with  $3 \times 3$  kernels and the sub-pixel convolution [24].

When it comes to the loss function, mean absolute error (MAE) and structural similarity index (SSIM) loss are applied to our SR methods. Given a training set  $\{I_{LR}^i, I_{HR}^i\}_{i=1}^N$ , which contains  $N$  LR inputs and their counterparts. The  $L_1$  loss can be formulated as follows:

$$\mathcal{L}_{MAE} = \frac{1}{N} \sum_{i=1}^N \|I_{HR}^i - G(I_{LR}^i)\|_1, \quad (3)$$

where  $G$  denotes the proposed SR network. In addition, SSIM loss is as follows:

$$\mathcal{L}_{SSIM} = \frac{1}{N} \sum_{i=1}^N 1 - SSIM(I_{HR}^i, G(I_{LR}^i)), \quad (4)$$

where,

$$SSIM(x, y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \cdot \frac{2\sigma_{xy} + C_2}{\sigma_x^2 + \sigma_y^2 + C_2}, \quad (5)$$

where  $\mu_x$ ,  $\mu_y$  are the mean,  $\sigma_{xy}$  is the covariance of  $x$  and  $y$  and  $C_1$ ,  $C_2$  are constants. Therefore, the total loss can be expressed as

$$\mathcal{L}_{total} = \mathcal{L}_{MAE} + 25\mathcal{L}_{SSIM} \quad (6)$$

### 3.2 Single image enhancement

For image quality enhancement, we devote to adjusting the contrast, suppressing noises and enhancing the image details. Considering that the time performance

is a vital aspect of image processing on smartphones with limited computational sources, the enhancer must be lightweight and efficient. Moreover, since the resolutions of inputs are arbitrary, the model should be the fully convolutional network. Thus, we prune our generator (student) as much as possible. In Fig. 3, the upper model indicates teacher generator with more convolution filters and the below one denotes student generator that is more compact. This topological structure is conducive to elevate the quantitative and qualitative performances of student generator without increasing parameters and computational cost.

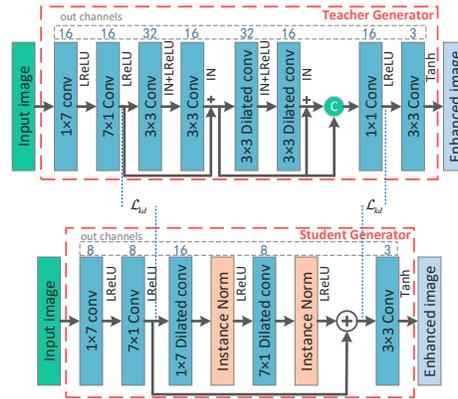


Fig. 3. The structure of the proposed generator for image enhancement.

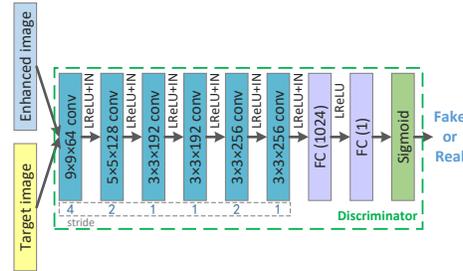


Fig. 4. The structure of the proposed discriminator for image enhancement.

Another core of this method is loss functions. In consideration of making the enhanced picture more photo-realistic, we follow the practice of Ignatov et al. [11], i.e., assume the overall perceptual image quality can be resolved into three portions: i) content quality, ii) texture quality and iii) color quality.

**Content loss** Inspired by [23,22], we choose contextual loss based on layer 'conv4\_2' of the VGG-19 network [25]. In addition, to perverse the structural

information of images, SSIM loss mentioned in Eq. 4 is also utilized. Thus, the content loss can be defined as

$$\mathcal{L}_{content} = \frac{4}{N} \sum_{i=1}^N \mathcal{L}_{CX} (G(I_{input}^i), I_{target}^i) + 25\mathcal{L}_{SSIM}, \quad (7)$$

where  $I_{input}^i$  and  $I_{target}^i$  constitute the training pairs  $\{I_{input}^i, I_{target}^i\}_{i=1}^N$ ,  $G$  represents the generator for image quality enhancement.

**Texture loss** Image texture quality is addressed by an adversarial discriminator as depicted in Fig. 4, which simply consists of 6 convolutional layers with leaky ReLU, 2 fully connected layers, and a sigmoid function. Following the way in [11,12], this discriminator is applied to grayscale images and is trained to identify the authenticity of a given image. The texture loss is defined as:

$$\mathcal{L}_{texture} = - \sum_i \log D(G(I_{input}^i)), \quad (8)$$

where  $D$  is the discriminator as illustrated in Fig. 4.

**Color loss** Image color quality is measured by MSE function that is trained to minimize the difference between the blurred versions of the low-quality input  $I_{input}$  and the high-quality target  $I_{target}$ . The blurred input can be expressed as

$$I_{input.b} = \sum_{k,l} I_{input}(i+k, j+l) \cdot G_{k,l}, \quad (9)$$

where  $G_{k,l} = A \exp\left(-\frac{(k-\mu_x)^2}{2\sigma_x} - \frac{(l-\mu_y)^2}{2\sigma_y}\right)$  indicates Gaussian blur with  $A = 0.053$ ,  $\mu_{x,y} = 0$ , and  $\sigma_{x,y} = 3$  proposed in [11,12]. Therefore, color loss can be written as:

$$\mathcal{L}_{color} = \|I_{input.b} - I_{target.b}\|_2^2. \quad (10)$$

**Tv loss** To suppress noises of the generated images we add a total variation loss [2] defined as follows:

$$\mathcal{L}_{tv} = \frac{1}{CHW} \|\nabla_x G(I_{input}) + \nabla_y G(I_{input})\|, \quad (11)$$

where  $C, H, W$  are the dimensions of the enhanced image  $G(I_{input})$ .

**Kd loss** The knowledge distillation loss is used to boost the performance of student model and is defined as follows:

$$\mathcal{L}_{kd} = \sum_{j \in \mathcal{J}} \left\| \frac{Q_S^j}{\|Q_S^j\|_2} - \frac{Q_T^j}{\|Q_T^j\|_2} \right\|_2, \quad (12)$$

where  $Q_S^j = \text{vec}\left(F\left(A_S^j\right)\right)$  and  $Q_T^j = \text{vec}\left(F\left(A_T^j\right)\right)$  are respectively the  $j$ -th pair of student and teacher mean feature maps in vectorized form, and  $F(A) = \frac{1}{C} \sum_{i=1}^C A_i$ .

**Sum of losses** We formulate the total loss as the weighted sum of aforementioned losses as:

$$\mathcal{L}_{total} = 10\mathcal{L}_{content} + \mathcal{L}_{texture} + \mathcal{L}_{color} + 2 \times 10^3 \mathcal{L}_{tv} + 75\mathcal{L}_{kd}. \quad (13)$$

## 4 Experiments

### 4.1 Datasets

**Image super-resolution task** For the instructions of the Perceptual Image Restoration and Manipulation (PIRM) challenges on Perceptual Enhancement on Smartphones<sup>1</sup> [13], we use the DIV2K dataset [1,28,29], which consists of 1000 high-quality RGB images (800 training images, 100 validation images, and 100 test images) with 2K resolution. HR image patches from HR images with the size of  $384 \times 384$  are randomly sampled for training. An HR image patch and its corresponding LR image patch are treated as a training pair.

For testing, we evaluate the performance of our network on five widely used benchmark datasets: Set5 [3], Set14 [33], BSD100 [20], Urban100 [10], and Manga109 [21].

**Image enhancement task** As for image enhancement task, we use the DPED dataset [11], which contains patches of size  $100 \times 100$  pixels for CNN training (139K, 160K and 162K pairs for BlackBerry, iPhone, and Sony, respectively). In this work, according to the illustration of the challenge, we consider only a sub-task of improving images from a very low-quality iPhone 3GS device. As for testing, we use the 400 patches provided by challenge<sup>2</sup>.

### 4.2 Implementation and Training Details

**Image super-resolution task** We randomly extract 16 LR RGB patches with the size of  $96 \times 96$  and interpolate them bicubically with the upscaling factor of 4. We augment LR patches with a random horizontal flip and 90 degree rotation. Experimentally, we set the initial learning rate to  $5 \times 10^{-4}$  and decreases by the factor 5 for every 1000 epochs ( $5 \times 10^4$  iterations). The Adam optimizer [16] with  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$  is used to train our model.

<sup>1</sup> <http://ai-benchmark.com/challenge.html>

<sup>2</sup> <https://github.com/aiff22/ai-challenge>

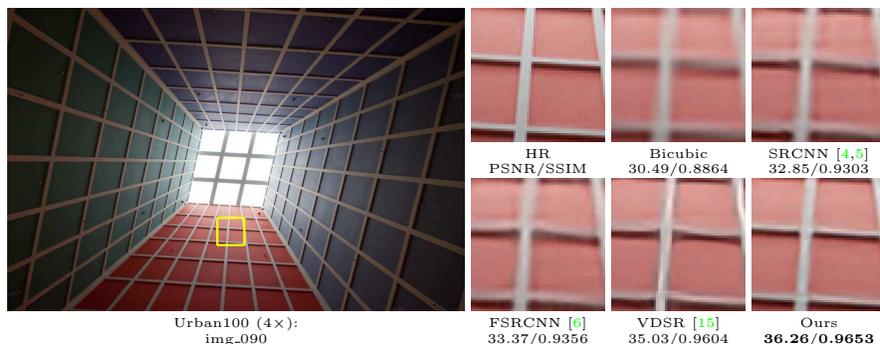
**Image enhancement task** Drawing on the experience of [11], we take 50 image patches with the size of  $100 \times 100$  as inputs. The learning rate is initialized to  $5 \times 10^{-4}$  for all layers and decreases by the factor 10 for every  $10^4$  iterations. We use the Adam optimizer [16] with  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ , and  $\epsilon = 10^{-8}$  for training. To improve the performance of the student, we first train the teacher with the same training hyper-parameters and then use it to guide the training of the student network by using Eq. 12.

All the experiments are implemented in the platform Ubuntu 16.04 operation system, TensorFlow 1.8 development environment, 3.7GHz Intel i7-8700k CPU, 64 GB memory and Nvidia GTX1080Ti GPU.

### 4.3 Comparison with baseline methods

**Table 1.** Quantitative evaluation results in terms of PSNR and SSIM. **Red** and **blue** colors indicates the best and second best methods, respectively.

Method	Scale	Set5		Set14		B100		Urban100		Manga109	
		PSNR	SSIM								
Bicubic	$\times 4$	28.42	0.8104	26.00	0.7027	25.96	0.6675	23.14	0.6577	24.89	0.7866
SRCNN [5]	$\times 4$	30.48	0.8628	27.50	0.7513	26.90	0.7101	24.52	0.7221	27.58	0.8555
FSRCNN [6]	$\times 4$	30.72	0.8660	27.61	0.7550	26.98	0.7150	24.62	0.7280	27.90	0.8610
VDSR [15]	$\times 4$	<b>31.35</b>	<b>0.8838</b>	<b>28.01</b>	<b>0.7674</b>	<b>27.29</b>	<b>0.7251</b>	<b>25.18</b>	<b>0.7524</b>	<b>28.83</b>	<b>0.8870</b>
Ours	$\times 4$	<b>31.37</b>	<b>0.8835</b>	<b>28.11</b>	<b>0.7698</b>	<b>27.24</b>	<b>0.7246</b>	<b>25.18</b>	<b>0.7542</b>	<b>28.93</b>	<b>0.8847</b>



**Fig. 5.** Visual comparison for  $4\times$  SR on Urban100 dataset.

**Image super-resolution task** To evaluate the performance of our proposed SR network, we use two baseline approaches SRCNN [4,5] and VDSR [15]. Tab. 1 shows the average PSNR and SSIM values on five benchmark datasets with the scaling factor of 4. From this table, we can see that the proposed method performs favorably against benchmark results. Tab. 2 indicates our solution better leverages the execution speed and the performance. In Fig. 5, it is obvious that



Fig. 6. Visual comparison for 4× SR on B100 dataset.

Table 2. Performance of our method on track A.

Track A:Image Super-Resolution							
Team	PSNR	MS-SSIM	CPU, ms	GPU, ms	Score A	Score B	Score C
SRCNN-Baseline	27.21	0.9552	3239	205	5.33	7.77	5.93
Rainbow (Ours)	28.13	0.9632	654	56	<b>12.84</b>	<b>14.92</b>	<b>13.91</b>

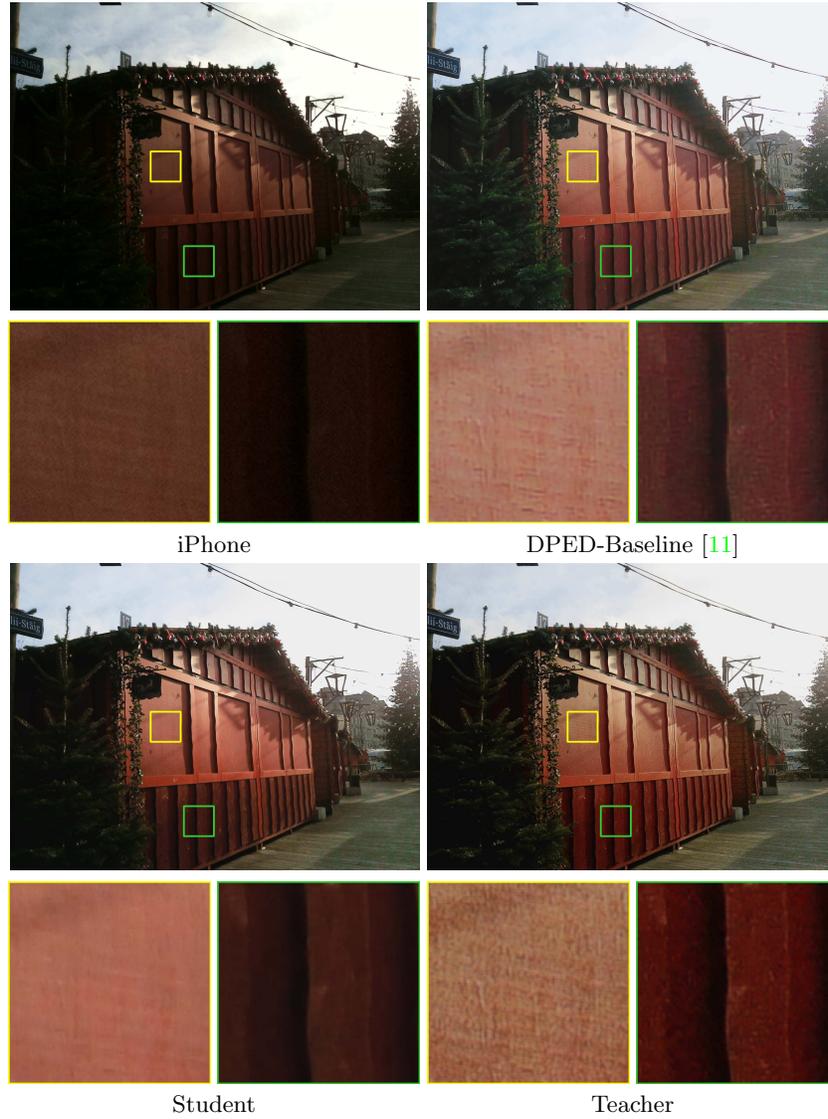
Table 3. The effectiveness of knowledge transfer.

Method		PSNR	MS-SSIM	CPU, ms
Ours	Teacher	<b>22.96</b>	<b>0.9299</b>	1182
	Student	22.54	0.9244	392
	Student w/o knowledge transfer	22.34	0.9229	445
	Student with L1 and VGG losses <sup>3</sup>	22.37	0.9231	478
Others	SRCNN-Baseline	21.32	0.9030	1832
	DPED-Baseline	22.17	0.9204	10701
	ResNet_8.32	22.38	0.9156	3226

the fidelity of geometric structure in our result is superior to the other methods.

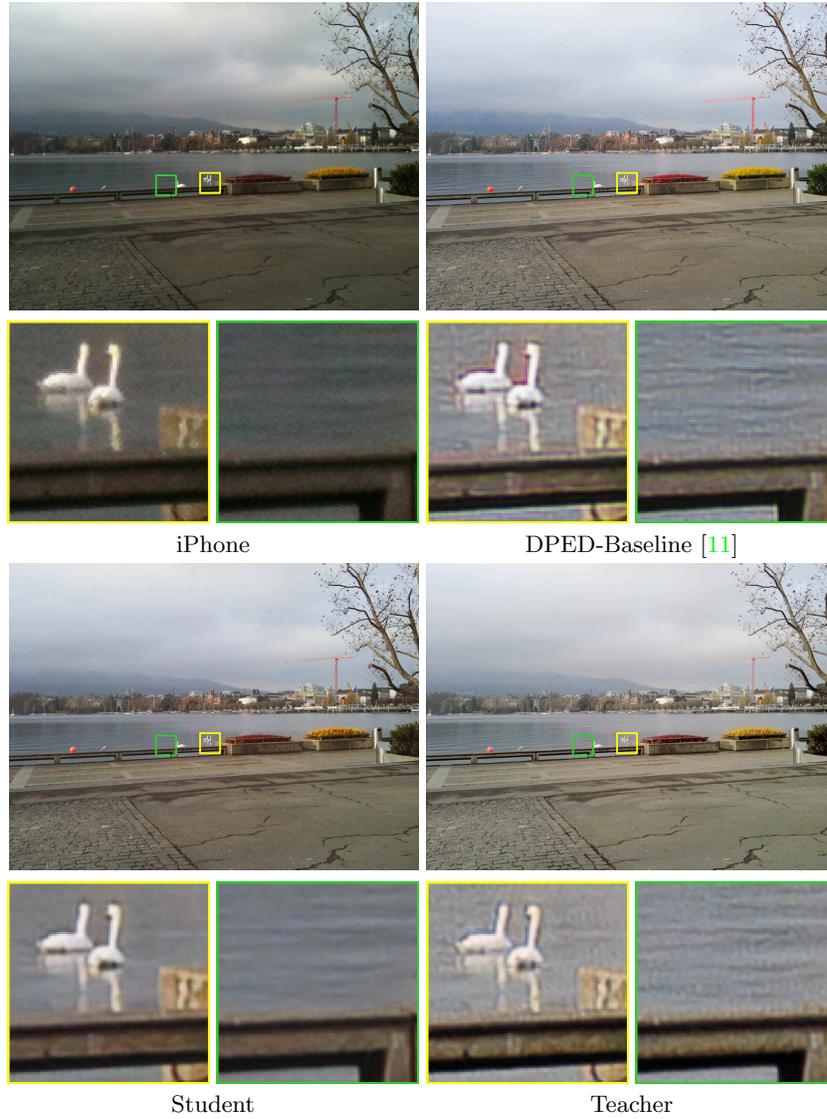
**Image enhancement task** In order to better transfer the model to practical application, we must weigh the performance and the speed of image enhance-

<sup>3</sup> This is the final submitted model in the challenge, where L1 loss can be expressed as  $\mathcal{L}_{l1} = \frac{1}{N} \sum_{i=1}^N \|I_{t\text{arget}}^i - G(I_{input}^i)\|_1$  and VGG loss is formulated as  $\mathcal{L}_{vgg} = \|\phi_l(I_{t\text{arget}}) - \phi_l(G(I_{input}))\|_2^2$ .



**Fig. 7.** Visual qualitative comparison of the “7” image in DPED test images.

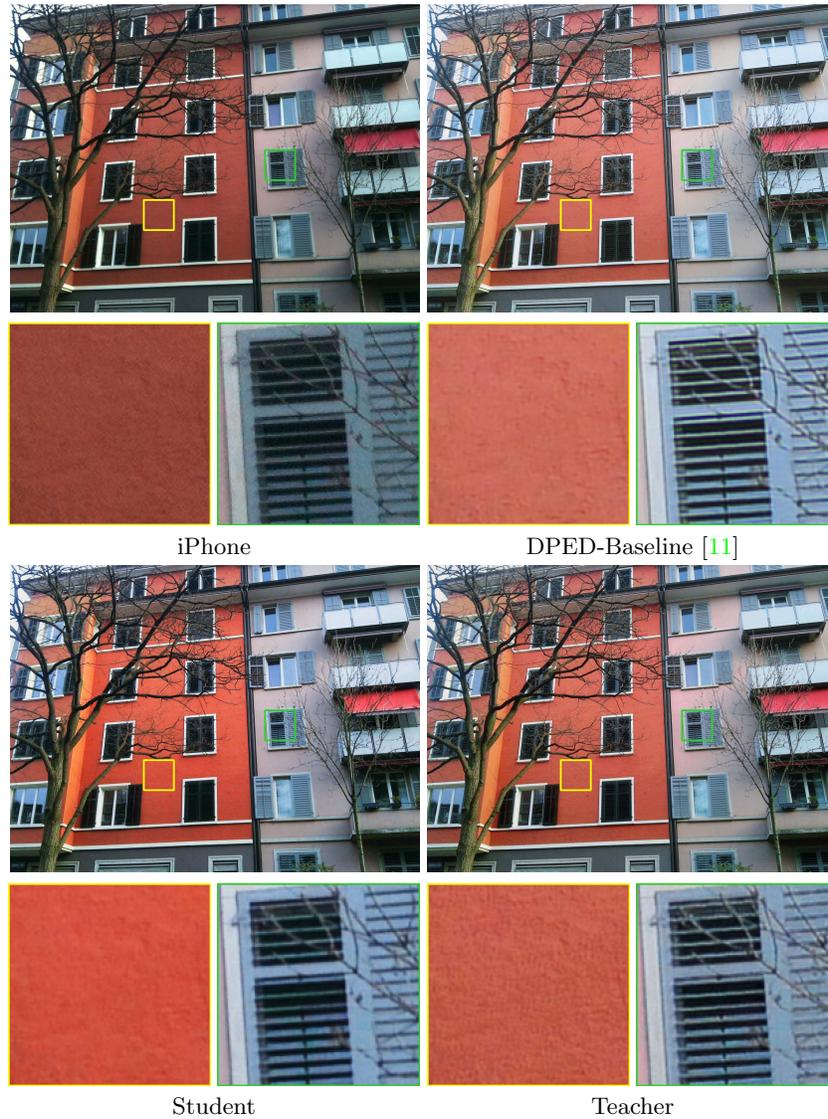
ment. From Tab. 3, the teacher network achieves high performance in terms of PSNR and MS-SSIM, but the execution speed is slightly slow. It is worth noting that the student model with L1 and VGG losses is our submitted version in the challenge. We experimentally find that when removing these two losses, the performance of the proposed student net can be prominently improved as shown in the third row of Tab. 3. Considering time testing, three student models have the same computational complexity, and the differences in Tab. 3 are caused by



**Fig. 8.** Visual qualitative comparison of the “8” image in DPED test images.

test errors. In Fig. 7, the generated result of the teacher model performs more realistic and the wood grain is clearer. But in terms of color saturation, the student network performs better. The DPED [11] produces color deviations in Fig. 8, whereas the student model successfully suppresses this typical artifact.

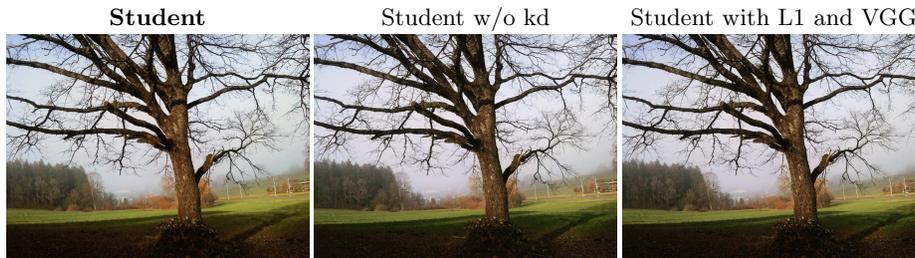
Previous results of our student model with L1 and VGG losses is shown in Tab. 4, which ranks 2nd in the challenge. Trained with losses in Eq. 13, we improve our model as shown in Tab. 3.



**Fig. 9.** Visual qualitative comparison of the "10" image in DPED test images.

#### 4.4 Ablation study

**Effectiveness of knowledge transfer** To demonstrate the effectiveness of the proposed knowledge transfer, we remove Kd loss in the training of student model while the network structure and other losses remain unchanged. Tab. 3 shows the effectiveness of knowledge transfer. From the visual assessment as shown in



**Fig. 10.** Result images for three student models.

**Table 4.** Performance of our method on track B.

Team	Track B:Image Enhancement							
	PSNR	MS-SSIM	MOS	CPU, ms	GPU, ms	Score A	Score B	Score C
SRCNN-Baseline	21.31	0.8929	2.295	3274	204	3.22	2.29	3.49
DPED-Baseline	21.38	0.9034	2.4411	20462	1517	2.89	4.9	3.32
Rainbow (Ours)	21.85	0.9067	2.5583	828	111	<b>13.19</b>	<b>16.31</b>	<b>16.93</b>

Fig. 10, the image generated by the student model is more saturated in color and more expressive.

#### 4.5 Limitations

Although visually realistic, the reconstructed images may contain emphasized high-frequency noise (see generated image of Teacher model in Fig. 8). It’s remarkable that the produced image of Student model successfully suppresses the noises, but the result appears smooth (Fig. 9).

## 5 Conclusions

In this paper, we propose the perception-preserving convolution network (PPCN) to enhance the image quality. Specifically, we devise a novel lightweight architecture that directly maps the low-quality images to the DSLR-quality counterparts to adapt to the environment with limited resource. To attain a more realistic visual effect, we introduce contextual and SSIM losses as the content loss. Furthermore, to improve the ability of the network, we adopt the knowledge transfer strategy, which enables the student model to learn information from the pre-trained teacher network. In addition, we propose a compact network for super-resolution task. Extensive experiments demonstrate the effectiveness of our proposed models.

## Acknowledgment

This work was supported in part by the National Natural Science Foundation of China under Grant 61472304, 61432914 and U1605252, in part by the Fundamental Research Funds for the Central Universities, and in part by the Innovation Fund of Xidian University.

## References

1. Agustsson, E., Timofte, R.: Ntire 2017 challenge on single image super-resolution: Dataset and study. In: CVPRW. pp. 1122–1131 (2017)
2. Aly, H.A., Dubois, E.: Image up-sampling using total-variation regularization with a new observation model. TIP **14**(10), 1647–1659 (2005)
3. Bevilacqua, M., Roumy, A., Guillemot, C., Alberi-Morel, M.L.: Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In: B-MVC (2012)
4. Dong, C., Loy, C.C., He, K., Tang, X.: Learning a deep convolutional network for image super-resolution. In: ECCV. pp. 184–199 (2014)
5. Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. TPAMI **38**(2), 295–307 (2015)
6. Dong, C., Loy, C.C., Tang, x.: Accelerating the super-resolution convolutional neural network. In: ECCV. pp. 391–407 (2016)
7. Fu, X., Huang, J., Ding, X., Liao, Y., Paisley, J.: Clearing the skies: A deep network architecture for single-image rain removal. TIP **26**(6), 2944–2956 (2017)
8. Fu, X., Huang, J., Zeng, D., Huang, Y., Ding, X., Paisley, J.: Removing rain from single images via a deep detail network. In: CVPR. pp. 1715–1723 (2017)
9. Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y.: In: NIPS. pp. 2672–2680 (2014)
10. Huang, J.B., Singh, A., Ahuja, N.: Single image super-resolution from transformed self-exemplars. In: CVPR. pp. 5197–5206 (2015)
11. Ignatov, A., Kobyshev, N., Vanhoey, K., Timofte, R., Gool, L.V.: Dslr-quality photos on mobile devices with deep convolutional networks. In: ICCV. pp. 3277–3285 (2017)
12. Ignatov, A., Kobyshev, N., Vanhoey, K., Timofte, R., Gool, L.V.: Wespe: Weakly supervised photo enhancer for digital cameras. In: CVPRW. pp. 804–813 (2018)
13. Ignatov, A., Timofte, R., et al.: Pirm challenge on perceptual image enhancement on smartphones: Report. In: ECCVW (2018)
14. Justin, J., Alexandre, A., Li, F.F.: Perceptual losses for real-time style transfer and super-resolution. In: ECCV. pp. 694–711 (2016)
15. Kim, J., Lee, J.K., Lee, K.M.: Accurate image super-resolution using very deep convolutional networks. In: CVPR. pp. 1646–1654 (2016)
16. Kingma, D.P., Ba, J.: Adam: A method for stochastic optimization. In: ICLR (2014)
17. Lai, W.S., Huang, J.B., Ahuja, N., Yang, M.H.: Deep laplacian pyramid networks for fast and accurate super-resolution. In: CVPR. pp. 624–632 (2017)
18. Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A., Tejani, A., Totz, J., Wang, Z., Shi, W.: Photo-realistic single image super-resolution using a generative adversarial network. In: CVPR. pp. 4681–4690 (2017)
19. Li, R., Cheong, L.F., Tan, R.T.: Single image deraining using scale-aware multi-stage recurrent network. In: arXiv:1712.06830 (2017)
20. Martin, D., Fowlkes, C., Tal, D., Malik, J.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In: ICCV. pp. 416–423 (2001)
21. Matsui, Y., Ito, K., Aramaki, Y., Fujimoto, A., Ogawa, T., Yamasaki, T., Aizawa, K.: Sketch-based manga retrieval using manga109 dataset. Multimedia Tools and Applications **76**(20), 21811–21838 (2017)

22. Mechrez, R., Talmi, I., Shama, F., Zelnik-Manor, L.: Maintaining natural image statistics with the contextual loss. In: arXiv:1803.04626 (2018)
23. Mechrez, R., Talmi, I., Zelnik-Manor, L.: The contextual loss for image transformation with non-aligned data. In: ECCV (2018)
24. Shi, W., Caballero, J., Huszár, Totz, J., Aitken, A.P., Bishop, R., Rueckert, D., Wang, Z.: Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: CVPR. pp. 1874–1883 (2016)
25. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. In: ICLR (2015)
26. Tai, Y., Yang, J., Liu, X.: Image super-resolution via deep recursive residual network. In: CVPR. pp. 3147–3155 (2017)
27. Tai, Y., Yang, J., Liu, X., Xu, C.: Memnet: A persistent memory network for image restoration. In: ICCV. pp. 3147–3155 (2017)
28. Timofte, R., Agustsson, E., Gool, L.V., Yang, M.H., Zhang, L., et al: Ntire 2017 challenge on single image super-resolution: Methods and results. In: CVPRW. pp. 1110–1121 (2017)
29. Timofte, R., Gu, S., Wu, J., Gool, L.V., Yang, M.H., et al: Ntire 2018 challenge on single image super-resolution: Methods and results. In: CVPRW. pp. 852–863 (2018)
30. Yang, W., Tan, R.T., Feng, J., Liu, J., Guo, Z., Yan, S.: Deep joint rain detection and removal from a single image. In: CVPR. pp. 1357–1366 (2017)
31. Yim, J., Joo, D., Bae, J., Kim, J.: A gift from knowledge distillation: Fast optimization, network minimization and transfer learning. In: CVPR. pp. 4133–4141 (2017)
32. Zagoruyko, S., Komodakis, N.: Paying more attention to attention: improving the performance of convolutional neural networks via attention transfer. In: ICLR (2017)
33. Zeyde, R., Elad, M., Protter, M.: On single image scale-up using sparse-representations. In: Curves and Surfaces. pp. 711–730 (2012)
34. Zhang, H., Patel, V.M.: Density-aware single image de-raining using a multi-stream dense network. In: CVPR. pp. 695–704 (2018)