Single Image Reflection Removal via Deep Feature Contrast

Lumin Liu,
School of Information Engineering, Xinyang Agriculture and Forestry University
No.1 North Circular Road, Pingqiao District, Xinyang, 464000.
China

Received: July 6, 2021. Revised: December 25, 2021. Accepted: January 10, 2022. Published: January 12, 2022.

Abstract- Removing undesired reflection from a single image is in demand for computational photography. Reflection removal methods are gradually effective because of the fast development of deep neural networks. However, current results of reflection removal methods usually leave salient reflection residues due to the challenge of recognizing diverse reflection patterns. In this paper, we present a one-stage reflection removal framework with an end-to-end manner that considers both low-level information correlation and efficient feature separation. Our approach employs the criss-cross attention mechanism to extract low-level features and to efficiently enhance contextual correlation. To thoroughly remove reflection residues in the background image, we punish the similar texture feature by contrasting the parallel feature separation networks, and thus unrelated textures in the background image could be progressively separated during model training. Experiments on both real-world and synthetic datasets manifest our approach can reach the state-of-the-art effect quantitatively and qualitatively.

Keywords- Reflection removal, Computational photography, Deep learning, Image restoration.

I. Introduction

Reflection is a common phenomenon in daily photography, especially when capturing images through transparent mediums like glass. The background object may be shielded, blurred or overexposed by undesired reflection. Methods for reflection removal can be applied in many fields, including intelligent vision systems, traffic recording cameras, and so on. As the popularity of vision devices in daily life, the degraded image might make most vision systems inoperable. Hence, practical reflection removal approaches are extremely in demand.

Main challenge in reflection removal stems from accurately recognizing reflection constituents of the degraded image. Traditionally, reflection-contaminated image **I** is mixed with linearly weighted background **B** and reflection image **R**. Hereby the reflection removal task could be transferred into an image decomposition task. Generally, the definition of reflection degradation is as follows[1-3]:

$$\mathbf{I} = \alpha \cdot \mathbf{B} + \beta \cdot (\mathbf{K} \otimes \mathbf{R}) + \mathrm{n}, \tag{1}$$

where **K** means the Gaussian blur kernel, $\otimes$ represents the convolution operation, α and β denote the proportion of background and reflection contributes, and n is the noise item.

Traditional solutions for reflection removal are mostly based on computational priors. One usual method is to predict the background edge firstly[6-8], and then contrasts the predicted background edge and contaminated-image to obtain reflection-free background through iterations. However, the limitation is obvious that it's challenging to predict accurate background edge without enough prior knowledge, which might cause many background texture details lost. Even if some recent user-assisted methods restrict this case[6, 9], the process of reflection removal becomes less practical in most scenarios as well.

Recently, the reflection removal task is in focus again thanks to the fast development of deep learning. Generative adversarial network[10], variation auto-encoder[11], and other models have already made headway so that the visual quality of image synthesis has been significantly elevated. In the beginning, plain convolution neural networks(CNNs) are used to obtain effective features[12] or predict background edge[13]. However, these multi-stage methods still follow traditional steps to remove reflection, so there are many distinct reflection residues can be seen from the restoration results. Another line of reflection removal is often in an end-to-end manner, which mainly includes two types. One is based on coarse-to-fine framework to first generate coarse reflection and background image to facilitate the predicted background[3, 14]. However, it often lets CNN models automatically learn potential contrast cues of background and reflection images without explicit instruction, as well as a huge scale of computation requires more inference time. The other attempts to generate background image directly by asking for extra misaligned images[5] or designing extra models to simulate the formation process of reflection[15] for training more degenerated images. As illustrated in Figure 1, we set an example to display restoration results of various types of state-of-the-art reflection removal methods on the real-world contaminated-image, which reveals present methods' weakness of background texture preservation and reflection patterns recognizing.

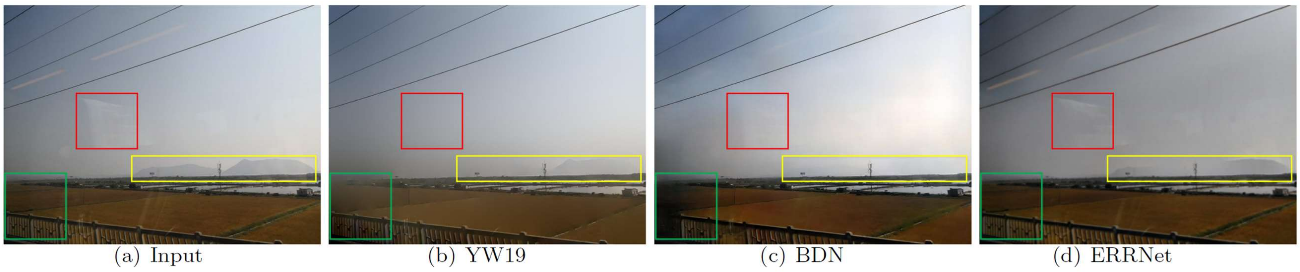


Fig. 1: An intuitive comparison of restoration results from various methods on a real-world reflection-contaminated photograph. Specifically, Computational priors based YW19[4] loses more vein details like the area of grass in the photograph. Without direct instruction from reflection image, multi-stage method BDN[3] causes more undesired artifacts. Even though ERRNet[5] has good preservation of background information, obvious reflection residues are left.

In this paper, we propose an end-to-end reflection removal framework, which sufficiently integrates multi-level contextual features and employs predicted reflection image as crucial contrast cues. We respectively propose Contextual Supplement Network(CSN), Restoration Backbone Network (RBN), and Feature Separation Network (FSN) as main components. The CSN strengthens feature pixel correlations from object edges and supplements more low-level cues. The RBN is designed to integrate multi-level contextual features for the final separation process. Afterward, parallel FSNs utilize separated features and feature contrast cues to measure global texture similarity between predicted background and reflection layer. Thus, the FSNs can effectively guide the output background image for removing various reflection patterns. Overall contribution of our research summarizes as below:

- We propose an end-to-end framework for single image reflection removal. The framework includes three efficient functional components and effective multi-level feature fusion for improving the performance of reflection removal.
- We especially reinforce feature extraction and separation process to sufficiently take advantage of multi-level semantic information's correlation and difference as critical cues to remove reflection.
- We carefully design texture discrepancy loss to more thoroughly exclude undesired constituents in the background or reflection layer by punishing similar texture of global features.
- Extensive experiments demonstrate the superiority of our framework. Both quantitative and qualitative results in real-world and synthetic datasets show our method achieves state-of-the-art performance.

II. Related Work

Multiple-view methods. Reflection removal is an extremely ill-posed problem, some methods ask for a sequence of images in various viewpoints to restrict solution domain. To eliminate reflection in a sequence of frames, multiview approaches usually regard camera motion [16-21] as crucial cues by assuming discrepant motion of the background layer and reflection layer. Methods[2, 22] estimate optical-flow to align objects in different frames, and thus the model can better separate reflection and background images.

Non-learning approaches. Practically, most real scenarios require removing reflection from a single image. Conventional solutions to restore images with reflection usually leverages handcrafted priors. Yan et al.[23] leverage gradient sparsity prior to predict the background gradient map, and then employ the gradient map again to facilitate the reconstruction of the background image. Considering the prior of Depth-of-Field(DoF), Wan et al.[8] first predict the confidence map of DoF to generate the gradient map of background image. To more accurately predict the background image, Levin et al.[6] require for user-interaction for obtaining reflection regions. Then, Fan et al.[9] improve the performance of user-interaction based method by respectively preserving vein and structure. Inspired by Laplacian fidelity term[25], Arvanitopoulos et al.[24] remove relfection by penalizing the weak edges. Soon, Yang et al.[4] improve the target function of reflection modeling and employ convex optimization to efficiently remove reflection constituents.

Deep learning approaches. With the fast development of deep learning[47, 48], the performance of image reflection removal is significantly improved. Fan et al.[13] firstly introduced a two-stage deep learning architecture that primarily predicts objective gradient map from the input image. Next, they utilized both origin input and predicted gradient map to recover sharp background layer. Besides, it also proposed an effective method to synthesize contaminated image from large-scale image dataset. Thereafter, another two-stage methods is proposed by Yang et al.[3], which adopted step-wise prediction for reflection and background image as well. Li et al.[26] employed the gradient confidence map to facilitate reflection removal and background image synthesis. Then, Wan et al.[27] proposed a novel architecture with two cooperative sub-networks, among which multi-level information could aggregate interactively aggregated. Considering global various features, Zhang et al.[28] introduced dilated convolution and perceptual loss to improve the quality of restored background images. Later, Wei et al.[5] proposed a novel approach to employ mis-aligned image pairs for model training. Recently, a

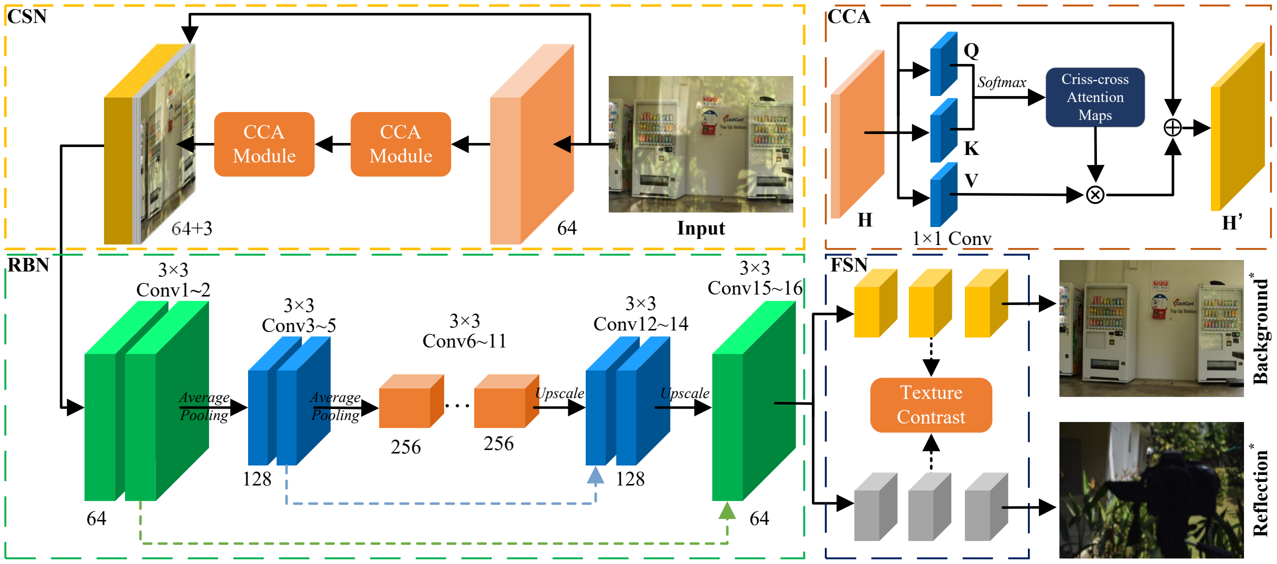


Fig. 2: Overall framework of our proposed reflection removal method. Main designed components consist of Contextual Supplement Network(CSN), reflection Restoration Backbone Network(RBN) and Feature Separation Network(FSN). In criss-cross attention(CCA) module, **H** and **H'** are input and output feature maps, as well as **Q**, **K** and **V** are extracted feature maps from 1×1 convolution in CCA module.

novel type of methods pays attention to modeling reflection degradation[15, 29]. Thus, it can generate training images as close as the real-world images with reflection, and improves the model robustness and effectiveness.

III. Method

In this section, we would like to introduce the main framework of our reflection removal method. At first, we elucidate how our functional components work in the overall proposed architecture. Then, we enumerate the loss functions during the training phase. Lastly, we supplement the specific implementatal details for model training.

A. *Network architecture*

As illustrated in Figure 2, our proposed reflection removal architecture consists of three main components, Contextual Supplement Network(CSN), Restoration Backbone Network(RBN), and Feature Separation Network(FSN).

Contextual Supplement Network. Some researches[8, 13, 14, 27] supplement low-level information to augment dereflection effect through additional subnetwork or multi-stage architecture, mainly because high frequency proportion in low-level feature is a reliable prior for reflection removal. Global pixel-wise correlation facilitates to restore high quality background images[30], and thus we employ the criss-cross attention in CSN to cement global low-level information and pixel-wise correlation. As an efficient evolution of non-local neural network[31], we employ Criss-Cross Attention (CCA) module, and its effectiveness has been proved for semantic segmentation[32]. As illustrated in Figure 2, the input feature H is mapped into **Q**, **K** and **V** by 1×1 convolutions. The CCA module includes two steps, the affinity and the aggregation. The affinity operation calculates pixel-wise correlation of all channels between feature maps **Q** and **K**:

$$d_{i,u} = \mathbf{Q}_u \cdot \mathbf{\Omega}_{i,u}, \tag{2}$$

where $\mathbf{Q}_u$ denotes a vector of pixels in all channels at position u in spatial dimension of feature maps **Q**, $\mathbf{\Omega}_{i,u}$ is the i-th element in the vector of position u's row or column in spatial feature maps **K**, and $d_{i,u}$ is the calculated correlation degree. Then attention map **A** is generated from $d_{i,u}$ via softmax function. Residual feature aggregation operation defines as below:

$$H'_u = \sum_{i\in|\Phi|} \mathbf{A}_{i,u}\mathbf{\Phi}_{i,u} + H_u, \tag{3}$$

where H_u and H'_u note input and output feature map at position u, $\mathbf{\Phi}_{i,u}$ is feature map vectors in **V**, and $\mathbf{A}_{i,u}$ is scalar value of attention map obtained from softmax function after $d_{i,u}$. According to CCA module's peculiarity, long-term contextual information can be well enriched. However, to gain sufficiently global correlation information, at least two recurrently connected CCA modules are necessary. Therefore, our CSN component can aggregate global corresponding between low-level contextual features. Notably, CCA module is computationally cheap, so it consumes little resource to deal with larger feature maps, as well as effectively aggregating low-level information according to pixel-wise correlation.

Restoration Backbone Network. To obtain reflection features from input reflection-contaminated image, we employed a common U-Net[33] as RBN's architecture shown in Figure 2. This subnetwork combines H' and I as input to obtain deep separable features $F_c = f(I, H')$. In main RBN, altogether 16 layers include 3×3 convolution filters, average pooling, leakyReLU[34] and symmetric skip-connections. Moreover, we achieve upscaling by bilinear interpolation with

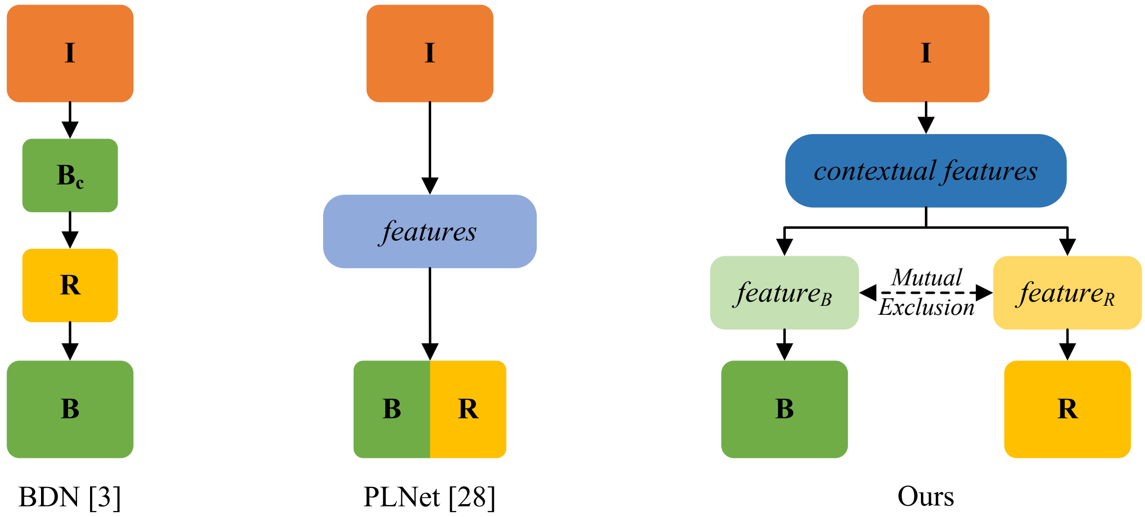


Fig. 3: Main strategies of different reflection separation based reflection removal methods. BDN[3] step-wise generates reflection and background images from coarse to fine. PLNet[28] simultaneously obtains reflection and background images. Our method utilizes the mutex relationship of texture between the reflection and background branches.

3×3 convolution and leakyReLU operations. The batch-normalization layer ignores absolute difference on image contrast and generates more undesired artifacts[35], so we forbid all the batch-norm layers to better suit our task.

Feature Separation Network. Main strategies of image decomposing based reflection removal methods are displayed in Figure 3. The third is the main strategy of our concurrent Feature Separation Network(FSN). Motivated by the low efficiency of most methods that use CNNs to implicitly learn reflection and background feature. We design the concurrent FSNs which are explicitly supervised to punish similar texture by feature contrast cues. Standard low-level feature is extracted from the shallow layers of pretrained VGG-19[36] to determine the target boundary. Therefore, we can finally obtain cleaner low-level features to respectively predict B^* and R^* through parallel FSNs $\mathcal{H}_B$ and $\mathcal{H}_R$:

$$\begin{aligned} B^* &= \mathcal{H}_B(f(I, H')), \\ R^* &= \mathcal{H}_R(f(I, H')). \end{aligned} \tag{4}$$

Additionally, we adopt residual learning $(I - B)$[37] so that our network can reduce the amplitude of variation to avoid artifacts while testing.

B. *Loss function*

Content loss. Only employing $\mathcal{L}_1$ distance to measure absolute error between groundtruth B_{GT} and predicted background image B^* might cause undesired artifacts at times. Hence, we introduce perceptual loss[38] that focuses on differences of multi-level semantic features between predicted background layer and groundtruth to enhance image details. Multi-level features are extracted from five layers of pretrained VGG-19[36] $\mathcal{F}_l$. The weights of various feature sizes are normalized by coefficients α_l. Combined with pixel-wise loss, more high frequency details can be preserved. Content loss is defined as follows:

$$\mathcal{L}_{content} = \mathcal{L}_{pixel} + \mathcal{L}_{feat}, \tag{5}$$

$$\mathcal{L}_{pixel} = ||B^* - B_{GT}||_1, \tag{6}$$

$$\mathcal{L}_{feat} = \sum_l \alpha_l ||\mathcal{F}_l(B^*) - \mathcal{F}_l(B_{GT})||_1. \tag{7}$$

Gradient loss. Precise gradient is critical prior for image restoration[13, 23, 27, 28, 39]. Our gradient loss is inspired from traditional prior of blurry image edges, which mainly includes two parts, total variation loss and exclusion loss. We introduce total variation loss[40] as regularization item to enhance spatial smoothness of output background and to restrain high-frequency artifacts. Gradient loss is defined as follows:

$$\mathcal{L}_{grad} = \mathcal{L}_{tv} + \mathcal{L}_{excl}, \tag{8}$$

$$\mathcal{L}_{tv} = ||\nabla_x B^*||_2^2 + ||\nabla_y B^*||_2^2, \tag{9}$$

where ∇_x and ∇_y are image gradients in x and y directions. Besides, gradient exclusion loss was proved useful for overlap image decomposition tasks, which can more thoroughly separate overlapped areas based on edge blurry prior.

$$\mathcal{L}_{excl} = \sum_l \sum_{n=1}^{N} ||\Psi(g_B^{\downarrow n}(I, \theta), g_R^{\downarrow n}(I, \theta)||_F, \tag{10}$$

$$\Psi(B, R) = tanh(\lambda_B |\nabla B|) \odot tanh(\lambda_R |\nabla R|), \tag{11}$$

where λ_T and λ_R are the normalization factors, g and n are downsampling operation parameters, and $\odot$ notes element-wise multiplication. More details about parameters definition are expatiated in original introduction[28].

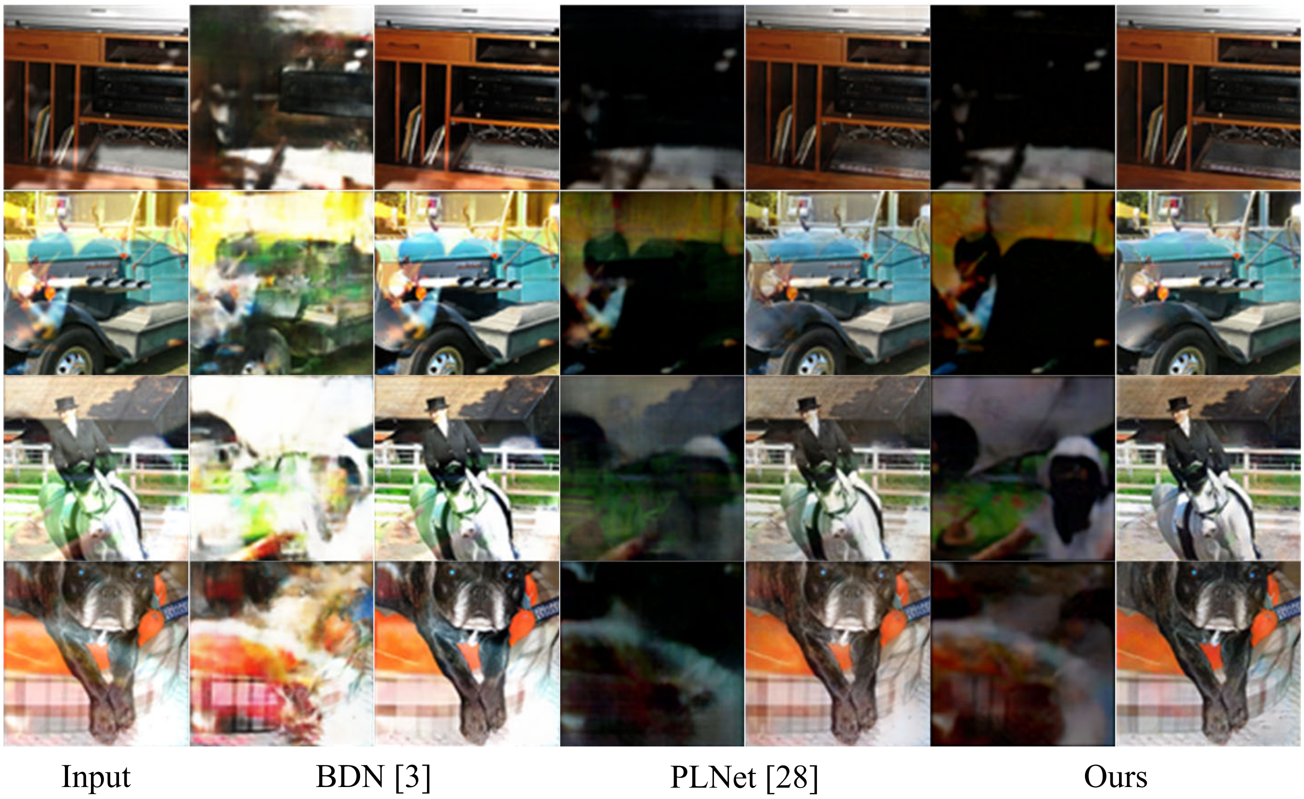


Fig. 4: Visual comparison of the separated background layer and reflection layer with BDN[3] and PLNet[28] on validation dataset of synthetic images[13]. Our method can restore cleaner and higher quality results from the input degraded image.

Adversarial loss. In order to preserve more vein and texture detail, additional supervisor helps much to improve the visual effect[41]. This can also relieve color degradation case and make the restoration results more like the realistic reflection-free images. Definition of adversarial loss is defined as follows:

$$\mathcal{L}_{adv} = -\mathbb{E}_{B^* \backsim \mathbb{P}_{\mathrm{G}}}[D(B^*, I)], \tag{12}$$

$$\mathcal{L}_D = -\mathbb{E}_{B_{\mathrm{GT}} \backsim \mathbb{P}_{\mathrm{data}}}[D(B_{\mathrm{GT}}, I)] + \mathbb{E}_{B^* \backsim \mathbb{P}_{\mathrm{G}}}[D(B^*, I)]. \tag{13}$$

We employ a similar discriminator proposed by the *pix2pix*[41].

Texture contrast loss. To effectively utilize mutually-exclusive features in parallel FSNs, we contrast the texture similarity of global features between background and reflection branches. Following the global texture extraction in image synthesis task[46], our designed texture contrast loss function penalizes those global texture features that exist in both branches at the same time by performing supervision. We infer locally extracted features belong to only one layer and output of deep separable features from RBN are ready to conclude final predictions. And we employ texture contrast loss to instruct the final FSNs to obtain cleaner predictions rather than simply receive B and R from the last convolution layer's six channels as PLNet[28].

Firstly, we protrude global texture feature through the Gram matrix, and then measure intersection proportion in a contrast manner by cosine similarity. In detail, we put the groundtruth of B and R into a pretrained VGG-19 and then take out 64 channels features in 'layer1_2' to gain metric boundary d_1. Analogously, we calculate the intersection proportion d_2 of parallel FSNs feature. Finally, we minimize the distance of d_1 and d_2 to restrain similar texture features. Texture contrast loss formulates as follows:

$$\mathcal{L}_{tc} = \frac{1}{c}\sum_{i=1}^{N} ||\mathcal{G}_i(B_{GT}, R_{GT}) - \mathcal{G}_i(B^*, R^*)||_1, \tag{14}$$

$$\mathcal{G}_i(B, R) = cos(\phi(F_{B,i}), \phi(F_{R,i})), \tag{15}$$

where ϕ denotes Gram matrix, cos is cosine similarity, C is feature channels and additionally $F_i = F_i(f(I, H))$ notes the i-th channel of RBN's output feature maps.

To sum up, our final loss function is formulated as:

$$\mathcal{L} = \lambda_1\mathcal{L}_{content} + \lambda_2\mathcal{L}_{grad} + \lambda_3\mathcal{L}_{adv} + \lambda_4\mathcal{L}_{tc}, \tag{16}$$

where the coefficients of each loss are empirically set as λ_1=1, λ_2=λ_3=0.01, λ_4=0.1.

IV. Experiments

This section shows our experiment results and both quantitative and qualitative comparisons with recent advanced reflection removal methods including BDN[3], PLNet[28], ERRNet[5], and YW19[4]. Among them, most methods are learning-based and one is the state-of-the-art non-learning based. Methods are tested on both

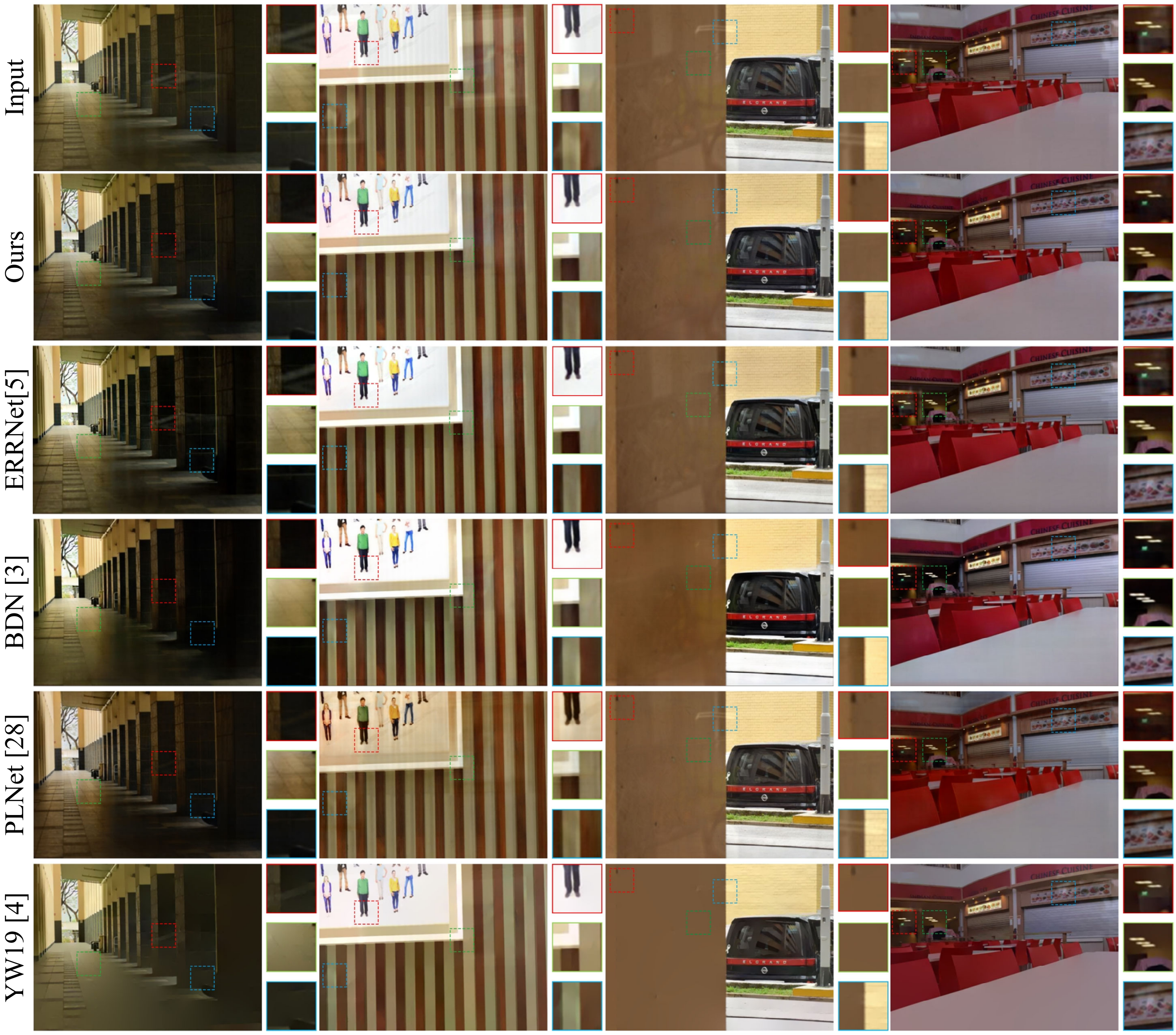


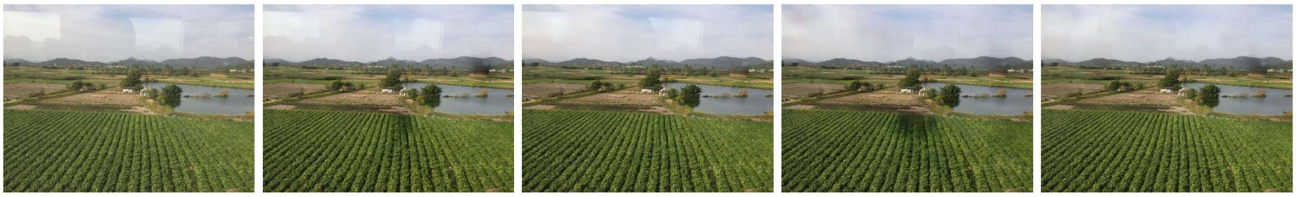
Fig. 5: Visualization of background restoration results from five reflection removal methods on four randomly selected test images in SIR^2 Wild dataset[1]. Our method keeps good fidelity of background information and removes reflection more thoroughly. Besides, we magnify the areas with more reflections with bounding boxes.

synthetic and real-world datasets to reveal our method's superiority. As for synthetic images, we qualitatively compare the results of image decomposition based methods, through which both reflection and background layers are predicted. Next, we compare our method on four real-world test datasets with various type methods to display our method practicality in real-life scenes. Additionally, we adopted PSNR, SSIM, LMSE[42], and NCC[2] metrics to more thoroughly evaluate different methods' results. Among them, higher values for PSNR, SSIM and NCC and lower LMSE mean higher quality of predicted background image.

A. Results comparison

Here we would like to show our results and compare them with state-of-the-art methods on both visual quality and quantitative metrics. For synthetic datasets, Figure 4 qualitatively shows four randomly selected results of 100 standard synthetic test images[13]. Compared with recent image decomposition based methods[3, 28], our model decomposes reflection-contaminated input into higher quality of prediction of reflection and background layers. Obviously, our method leaves fewer reflection residues on the predicted background layer, as well as capturing a more precise reflection layer.

To deeply explain our model's reflection-free effect, we employ four real-world datasets for evaluation. Specific visual quality comparison with recent reflection removal methods is shown in Figure 5 and Figure 7. In detail, our method is not merely better for detail fidelity of the background, but also diminishes reflection residues. For instance, our method exhibits prominent background preservation, and creates fewer artifacts on SIR^2 Wild dataset[1] in the second row of Figure 5. Both the pillar and floor details are preserved better than other state-



(a) Input (b) RBN-only (c) w/o CSN (d) w/o FSN (e) Complete

Fig. 6: Visualization of restoration results from four variants of our method in ablation experiments on a randomly selected real-world test image.

Table 1: Performance of different methods for image reflection removal on four real-world test datasets in terms of *PSNR*, *SSIM*, *LMSE* and *NCC*. Note that best results are red boldly and the second best are blue boldly.

Dataset	Metrics	YW19[4]	BDN[3]	PLNet[28]	ERRNet[5]	Ours
Real20	PSNR	17.99	18.97	21.91	**22.68**	**22.84**
	SSIM	0.642	0.740	**0.787**	**0.798**	**0.798**
	NCC	0.758	0.794	**0.896**	**0.877**	0.867
	LMSE	0.027	0.032	**0.021**	**0.022**	**0.022**
Solid	PSNR	20.39	22.73	22.63	**24.69**	**24.92**
	SSIM	0.817	0.853	0.874	**0.891**	**0.886**
	NCC	0.953	0.978	0.963	**0.982**	**0.979**
	LMSE	0.007	**0.005**	0.006	**0.005**	**0.004**
Postcard	PSNR	20.04	20.71	16.82	**21.85**	**21.60**
	SSIM	0.791	0.856	0.797	**0.878**	**0.876**
	NCC	0.846	**0.913**	0.884	0.880	**0.917**
	LMSE	0.008	0.007	0.007	**0.005**	**0.006**
Wild	PSNR	21.39	22.34	21.50	**24.45**	**24.80**
	SSIM	0.822	0.821	0.829	**0.859**	**0.879**
	NCC	0.798	0.794	0.896	**0.904**	**0.927**
	LMSE	0.009	**0.007**	0.008	**0.006**	**0.006**

of-the-art methods of the first column. In fact, the decomposition based method PLNet[28] is outstanding for reflection removal. PLNet[28] transforms picture tone leading many image color saturation changes. Moreover, YW19[4] also appears the problem of global pixel transformation of the input image. Compared with learning based methods, some high-frequency details of background might be lost or blurred. Next, to show our model's generalization performance, Figure 7 displays visual results of other existing real-world images with reflection[4, 13, 28]. Our method can restore interfered images to high-quality background images. Interestingly, our model can recognize the glass edge in the third row of Figure 7 and preserve it, rather than remove it like ERRNet[5]. It also implies our method can more accurately capture reflection regions and preserve background details.

Table 1 shows a quantitative comparison between state-of-the-art methods on the existing real-world test datasets in terms of four evaluation metrics. The quantitative results prove our method can be comparable with the most state-of-the-art results for image reflection removal. In terms of PSNR, our method achieves 0.35 dB higher than competing methods on Wild dataset. Table 2 also displays that our method is of great efficiency while inferring. However, we are willing to acknowledge that relatively uncommon scenes of SIR^2 postcard dataset are still challenging and have more progress space to enhance performance.

Table 2: Comparison of model efficiency of different methods from the number of trainable parameters and inference time. Note that all input images are tested with the resolution of 256×256 on a same GPU.

Method	Trainable params	Inference time
ERRNet[5]	45.4M	0.074s
BDN[3]	75.2M	0.066s
Ours	**5.06M**	**0.012s**

B. *Training dataset*

Following the synthesis method proposed by PLNet[28], we generated 7000 synthetic triplets $\{I_t, B_t, R_t\}_t^N$ by synthesizing images from the PASCAL VOC dataset[43] for model training. In addition, we supplemented real-world training images proposed by PLNet[28], and cropped the real-world images into 3000 training pairs. Finally, we obtained 10000 image pairs as training dataset.

To evaluate different methods effectively, we utilized four real-world test datasets. The SIR^2[1] is a collection of images for reflection removal, which can be divided into three datasets: (1) 200 images captured in indoor scenes with solid objects. (2) 199 challenging triplets that are captured from overlapped postcards. (3) 55 wild scenes in daily life reflection condition. For comparison, three datasets are respectively named as Solid, Postcard, and Wild. Besides, we name the 20 real-world test images in PLNet[28] as Real20.

C. *Implemental details*

We implement our experiments through PyTorch. The image scale is randomly cropped into 256×256. Besides, training images employ flipping and random resizing for augmentation. Afterward, synthetic and real-world datasets are loaded with adjustable proportion to avoid over-fitting. Our adjustable learning rate changes from 2e-4 to 1e-5 along with training epochs. We adopt Adam[44] optimizer, batchsize 4, and altogether 120 epochs were trained from scratch using weight initializing method[45]. All experiments are implemented with 4 Titan RTX GPUs on distributed mode.

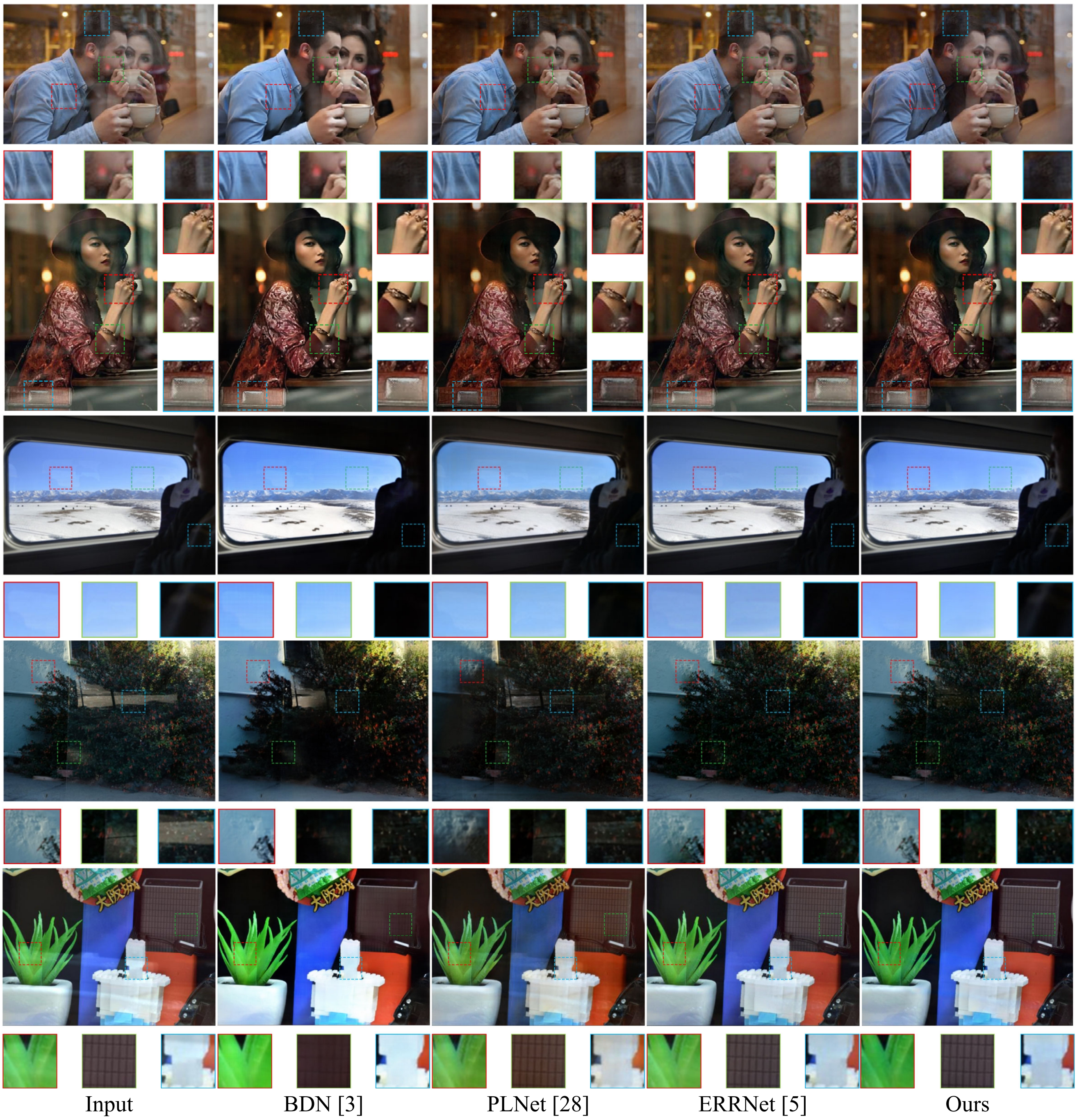


Fig. 7: Visualization of restoration results on five randomly selected images from real-world test datasets. We compare our method with state-of-the-art results of decomposition based methods BDN[3] and PLNet[28], as well as non-decomposition based method ERRNet[5]. Our method obtains higher quality background images.

D. Ablation study

To further explore the actual effect of the loss function and network component, we perform a series of ablation study experiments. Compared components are three subnets of CSN, RBN, and FSN. As quantitative results displayed in Table 3, it's obvious to conclude that the complete model performs best.

Qualitative evaluations of restoration results from our model's variants in Figure 6 demonstrate specific reflection removal effects by a random selected real-world image. The initial parameters of each experiment are the

Table 3: Quantitative performance of four ablation variants of our method on both synthetic dataset and real-world Wild dataset in terms of *PSNR* and *SSIM*. Note that RBN-only and w/o FSN forbid $\mathcal{L}_{tc}$ while training.

Dataset	Metrics	RBN-only	w/o FSN	w/o CSN	Complete
Real20[28]	PSNR	19.97	21.09	22.06	**22.84**
	SSIM	0.750	0.769	0.787	**0.798**
Synthesis[13]	PSNR	22.73	23.22	24.17	**24.80**
	SSIM	0.843	0.860	0.866	**0.879**

same. The increasingly better visual and quantitative results in ablation study can demonstrate that the proposed framework and the loss function are of crucial effectiveness to deal with reflection removal task.

V. Conclusion

In this paper, we present an end-to-end framework to restore reflection-contaminated images. Inspired by conventional solutions, we design three functional components of CSN, RBN and FSN respectively to efficiently utilize multi-level information. To thoroughly exclude reflection residues, we propose texture contrast loss to measure texture intersection proportion between branches. Extensive experiments demonstrate that our proposed framework can effectively work on both synthetic and real-world datasets.

References

[1] Wan R, Shi B, Duan L Y, et al. Benchmarking single-image reflection removal algorithms[C]//Proceedings of the IEEE International Conference on Computer Vision. 2017: 3922-3930.

[2] Xue T, Rubinstein M, Liu C, et al. A computational approach for obstruction-free photography[J]. ACM Transactions on Graphics (TOG), 2015, 34(4): 1-11.

[3] Yang J, Gong D, Liu L, et al. Seeing deeply and bidirectionally: A deep learning approach for single image reflection removal[C]//Proceedings of the european conference on computer vision (ECCV). 2018: 654-669.

[4] Yang Y, Ma W, Zheng Y, et al. Fast single image reflection suppression via convex optimization[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 8141-8149.

[5] Wei K, Yang J, Fu Y, et al. Single image reflection removal exploiting misaligned training data and network enhancements[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 8178-8187.

[6] Levin A, Weiss Y. User assisted separation of reflections from a single image using a sparsity prior[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2007, 29(9): 1647-1654.

[7] Levin A, Zomet A, Weiss Y. Separating reflections from a single image using local features[C]//Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004. IEEE, 2004, 1: I-I.

[8] Wan R, Shi B, Hwee T A, et al. Depth of field guided reflection removal[C]//2016 IEEE International Conference on Image Processing (ICIP). IEEE, 2016: 21-25.

[9] Heydecker D, Maierhofer G, Aviles-Rivero A I, et al. Mirror, mirror, on the wall, who's got the clearest image of them all?—A tailored approach to single image reflection removal[J]. IEEE Transactions on Image Processing, 2019, 28(12): 6185-6197.

[10] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[J]. Advances in neural information processing systems, 2014, 27.

[11] Kingma D P, Welling M. Auto-encoding variational bayes[J]. arXiv preprint arXiv:1312.6114, 2013.

[12] Chandramouli P, Noroozi M, Favaro P. Convnet-based depth estimation, reflection separation and deblurring of plenoptic images[C]//Asian Conference on Computer Vision. Springer, Cham, 2016: 129-144.

[13] Fan Q, Yang J, Hua G, et al. A generic deep architecture for single image reflection removal and image smoothing[C]//Proceedings of the IEEE International Conference on Computer Vision. 2017: 3238-3247.

[14] Li C, Yang Y, He K, et al. Single image reflection removal through cascaded refinement[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2020: 3565-3574.

[15] Wen Q, Tan Y, Qin J, et al. Single image reflection removal beyond linearity[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 3771-3779.

[16] Farid H, Adelson E H. Separating reflections and lighting using independent components analysis[C]//Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149). IEEE, 1999, 1: 262-267.

[17] Gai K, Shi Z, Zhang C. Blind separation of superimposed moving images using image statistics[J]. IEEE transactions on pattern analysis and machine intelligence, 2011, 34(1): 19-32.

[18] Guo X, Cao X, Ma Y. Robust separation of reflection from multiple images[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2014: 2187-2194.

[19] Li Y, Brown M S. Exploiting reflection change for automatic reflection removal[C]//Proceedings of the IEEE international conference on computer vision. 2013: 2432-2439.

[20] Nandoriya A, Elgharib M, Kim C, et al. Video reflection removal through spatio-temporal optimization[C]//Proceedings of the IEEE International Conference on Computer Vision. 2017: 2411-2419.

[21] Sarel B, Irani M. Separating transparent layers through layer information exchange[C]//European Conference on Computer Vision. Springer, Berlin, Heidelberg, 2004: 328-341.

[22] Yang J, Li H, Dai Y, et al. Robust optical flow estimation of double-layer images under transparency or reflection[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016: 1410-1419.

[23] Yan Q, Xu Y, Yang X, et al. Separation of weak reflection from a single superimposed image[J]. IEEE Signal Processing Letters, 2014, 21(10): 1173-1176.

[24] Arvanitopoulos N, Achanta R, Susstrunk S. Single image reflection suppression[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 4498-4506.

[25] Xu L, Lu C, Xu Y, et al. Image smoothing via L 0 gradient minimization[C]//Proceedings of the 2011 SIGGRAPH Asia conference. 2011: 1-12.

[26] Li T, Lun D P K. Single-image reflection removal via a two-stage background recovery process[J]. IEEE Signal Processing Letters, 2019, 26(8): 1237-1241.

[27] Wan R, Shi B, Duan L Y, et al. Crrn: Multi-scale guided concurrent reflection removal network[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2018: 4777-4785.

[28] Zhang X, Ng R, Chen Q. Single image reflection separation with perceptual losses[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 4786-4794.

[29] Ma D, Wan R, Shi B, et al. Learning to jointly generate and separate reflections[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 2444-2452.

[30] Buades A, Coll B, Morel J M. A non-local algorithm for image denoising[C]//2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). IEEE, 2005, 2: 60-65.

[31] Wang X, Girshick R, Gupta A, et al. Non-local neural networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2018: 7794-7803.

[32] Huang Z, Wang X, Huang L, et al. Cc-net: Criss-cross attention for semantic segmentation[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2019: 603-612.

[33] Ronneberger O, Fischer P, Brox T. U-net: Convolutional networks for biomedical image segmentation[C]//International Conference on Medical image computing and computer-assisted intervention. Springer, Cham, 2015: 234-241.

[34] Maas A L, Hannun A Y, Ng A Y. Rectifier nonlinearities improve neural network acoustic models[C]//Proc. icml. 2013, 30(1): 3.

[35] Wang X, Yu K, Wu S, et al. Esrgan: Enhanced super-resolution generative adversarial networks[C]//Proceedings of the European conference on computer vision (ECCV) workshops. 2018: 0-0.

[36] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition[J]. arXiv preprint arXiv:1409.1556, 2014.

[37] Fu X, Huang J, Zeng D, et al. Removing rain from single images via a deep detail network[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2017: 3855-3863.

[38] Johnson J, Alahi A, Fei-Fei L. Perceptual losses for real-time style transfer and super-resolution[C]//European conference on computer vision. Springer, Cham, 2016: 694-711.

[39] Punnappurath A, Brown M S. Reflection removal using a dual-pixel sensor[C]//Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. 2019: 1556-1565.

[40] Rudin L I, Osher S, Fatemi E. Nonlinear total variation based noise removal algorithms[J]. Physica D: nonlinear phenomena, 1992, 60(1-4): 259-268.

[41] Isola P, Zhu J Y, Zhou T, et al. Image-to-image translation with conditional adversarial networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2017: 1125-1134.

[42] Grosse R, Johnson M K, Adelson E H, et al. Ground truth dataset and baseline evaluations for intrinsic image algorithms[C]//2009 IEEE 12th International Conference on Computer Vision. IEEE, 2009: 2335-2342.

[43] Everingham M, Van Gool L, Williams C K I, et al. The pascal visual object classes (voc) challenge[J]. International journal of computer vision, 2010, 88(2): 303-338.

[44] Kingma D P, Ba J. Adam: A method for stochastic optimization[J]. arXiv preprint arXiv:1412.6980, 2014.

[45] He K, Zhang X, Ren S, et al. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1026-1034.

[46] Gatys L A, Ecker A S, Bethge M. Image style transfer using convolutional neural networks[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 2414-2423.

[47] Luqman Hakim, Muhammad Ihsan Zul, Implementation of Discrete Wavelet Transform on Movement Images and Recognition by Artificial Neural Network Algorithm, WSEAS Transactions on Signal Processing, ISSN / E-ISSN: 1790-5052 / 2224-3488, Volume 15, 2019, Art. 18, pp. 149-154

[48] Jose Augusto Cadena Moreano, Nora Bertha La Serna Palomino, Efficient Technique for Facial Image Recognition with Support Vector Machines in 2D Images with Cross-Validation in Matlab, WSEAS Transactions on Systems and Control, ISSN / E-ISSN: 1991-8763 / 2224-2856, Volume 15, 2020, Art. 18, pp. 175-183

Creative Commons Attribution License 4.0 (Attribution 4.0 International , CC BY 4.0)
This article is published under the terms of the Creative Commons Attribution License 4.0

https://creativecommons.org/licenses/by/4.0/deed.en_US