Single Image Reflection Removal Based on Deep Residual Learning

Zhixin Xu, Xiaobao Guo, and Guangming Lu(✉)

Harbin Institute of Technology, Shenzhen, China
{xuzhixin,guoxiaobao}@stu.hit.edu.cn,
luguangm@hit.edu.cn

Abstract. We present a deep residual learning approach to address the single image reflection removal problem. Specifically, residual learning exploits the mapping between the observed image and its comparatively simple reflection information, which is then removed to obtain a clear background. Different from other methods that roughly eliminating the reflections and producing the images with remanent sticking, a novel generative adversarial framework is proposed, where the generator is embedded with the deep residual learning, significantly boosting the performance without impairing the intactness of the background by adversarial training. Moreover, a multi-part balanced loss is introduced with comprehensive consideration on the measure of feature similarity as well as the discriminating ability of GAN. It produces the result of high quality by learning the reflection and the background feature simultaneously. Experiments show that the proposed method achieves a state-of-the-art performance.

Keywords: Reflection removal · Residual learning · GAN

1 Introduction and Related Work

Along with the rapid advancement and application of the digital camera, photographing is becoming increasingly prevailing and indispensable in our daily life. However, the image would be corrupted by some additional undesirable parts when is taken in the reflective surroundings such as glass, water, and windows, which exerts considerable degradation on the visual perceptual quality. As shown in Fig. 1, the reflection in the images tends to distract our eyes from the scenes behind the glass. Furthermore, reflection, as a typical kind of noise, would impede both human and computer vision programs from better understanding the scene. Therefore, single image reflection removal is the active and essential research in computer vision community.

Student First Author

The work is supported by the NSFC fund (61332011), Shenzhen Fundamental Research fund (JCYJ20170811155442454, JCYJ20180306172023949), China Postdoctoral Science Foundation (2019TQ0316), and Medical Biometrics Perception and Analysis Engineering Laboratory, Shenzhen, China.

Z. Lin et al. (Eds.): PRCV 2019, LNCS 11858, pp. 253–266, 2019.
https://doi.org/10.1007/978-3-030-31723-2_22

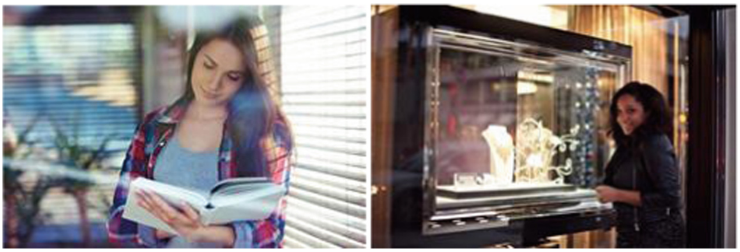


Fig. 1. Captured image samples with reflection in the real world.

Years of studies on solving the reflection removal problem have made some achievements especially on eliminating reflection from non-metallic surfaces by adjusting the assembled component in the camera, e.g. rotating a polarized lens [3,4], moving focus [5], and using a flash [6]. Yet, hardware-based method is usually constrained by the lack of adaptability and flexibility when dealing with various background scenes and tricky reflection sources. In comparison, algorithms that tackle the reflection removal issue are more practical. Conventionally, prior knowledge and hand-crafted features [1,2,7] are utilised to learn the representation or mapping between the input image and the clear background scene. Li *et al.* [7] introduce gradient histogram for the image to construct long tail distribution, assuming that the relative smoothness for reflection and background are different, by which to determine and separate reflection. However, we can not tell the reflection is always smooth in the real case. Arvanitopoulos *et al.* [8] try to solve the problem by suppressing the reflection instead of getting rid of it by manually adjusting the thresholds of Laplacian data fidelity term and an l_0 gradient sparsity term on the output, which is a trade-off between suppressing reflection artifacts and image details. Therefore, it probably causes a degradation on important image details. Although traditional algorithms could basically solve the problem, they still suffer great limitations on coping with challenging situations. Deep learning based methods have greatly mitigated the illogicality between the robustness and adaptiveness compared with traditional methods. CEILNet [9] is the first to address the reflection removal problem with a deep convolutional network, where two cascaded sub-networks are combined, one for edge prediction and the other for image reconstruction. However, features are too many in input images and two-cascaded structure is complicated.

Under the same the precondition that other methods use [7,9], our method is also conducted on the basis of the assumption that the observed image is a compound of a reflection image and a clear background image. This assumption is rational and valid in practice since the reflective source or objects are usually observed and known when people take pictures of a scene. However, they usually focus on reflection removal by suppressing reflection artifacts to restore the background scene only, the learning and separation of the reflection are ignored, which causes a degradation on the result. It is substantial to learn both reflection information and the background scene jointly as they are intertwined in a single observed image.

To alleviate the aforementioned problems, a deep residual learning based single image reflection removal method is proposed. Instead of predicting the background image layer from the observed image directly, the proposed method learns a mapping between the observed image and its reflection layer since the background image is usually intractable for a generator [10] to learn while the reflection is relatively consistent in terms of luminance and color. Intuitively, residual learning is embedded into the proposed generative adversarial framework which provides an effective approach that decouples the reflection information from the entangled observed image [17,18]. Simultaneously, the discriminator is trained to encourage the generated background image, the observed one without residual reflection information, to be more similar to the real background image. Unlike other deep reflection removal network, each part in our proposed method aims at a different target that jointly contributes to solving the reflection removal problem. To make it feasible in practice, the excogitation of loss function is supposed to be balanced, taking both feature similarity and discriminatory ability into consideration. Therefore, the proposed multi-part balanced loss comprises the content loss which measures the similarity between the learned residual information and the real reflection information, a perceptual loss that encourage the feature of the decoupled image to be more consistent with the real background image, and an adversarial loss for improving the discriminating ability of the discriminator. Through experiments, the multi-part balanced loss is proved to be beneficial to eliminating the gradient vanishing and exploding during training while the residual learning with the generative adversarial network is of great ability for both modeling the reflection information and the background image feature.

Our main contributions are as follows:

- A deep residual learning approach is proposed to solve the single image reflection removal problem. It exploits the mapping between the observed image and its reflection information, repressing the intractable issue of learning the complicated feature of the background image directly. Residual learning provides an effective method for deep reflection removal framework, yielding a faster convergence for training the model as well as an enhancement on its performance.
- The design of a novel generative adversarial network is utilized in this task, where the generator is embedded with the residual learning strategy and the discriminator is assigned to distinguish the generated background image from the real background image. The proposed generative adversarial framework for single image reflection removal demonstrates an exceeding performance by keeping the intactness of the background image through adversarial training.
- A multi-part balanced loss as the objective function is proposed, which is composed of the content loss, the perceptual loss, and the adversarial loss. It is of comprehensive consideration on both the measure of feature similarity and the description of the discriminating ability, which entails the network

to produce the background image of high quality by learning the reflection information and the background feature simultaneously.

The experiment results show the effectiveness of the proposed method for single image reflection removal task.

2 Deep Residual Learning Network with GAN

2.1 The Generative Adversarial Framework

The proposed generative adversarial framework is shown in Fig. 2, which can be rendered into two parts, one is a generator embedded with deep residual learning while the other is a discriminator to distinguish the generated background image and its corresponding real one. The captured image is first input into the generator, then the generator is trained to produce its reflection image through residual learning. The difference between the generated reflection image and the real reflection (ground-truth) will be formulated as the content loss which is a part of the objective function for the network. The discriminator is then trained to discriminate the generated background image and the real background (ground-truth), which in turn acts on the generator to produce a reflection image layer that is largely identical to the reality. Hence, residual learning and adversarial training are implemented in the generative adversarial framework. For the target of acquiring a clear background image without any reflection remanent, the features of both generated background image and real background image are extracted from several layers of a pretrained VGG-19 network [12], being reinforced to be similar by introducing a perceptual loss. By learning the mapping of the input image and its reflection information, together with modeling the distribution of the background image, the proposed method enjoys a high efficiency and quality for single image reflection removal. The implementations and details of the proposed method can be referred in the following parts in this paper.

2.2 Residual Learning

Reflection in an observed image usually presents a high intensity of light, emerging in part or full area in the image, pretty confusing when entangled with the background objects. Therefore, it imposes restrictions upon the network to learn the subtle features of the background image, which is normally the reason why extracting the background image directly would not be a favorable solution. Residual learning aims at learning the reflection information from the input image. It is a roundabout method that enables more effective learning to separate the reflection layer from the entangled image without impairing the background scene.

Let G be the trained network for residual learning. I represents the input image while I_b and I_r represent the real background image and the real reflection

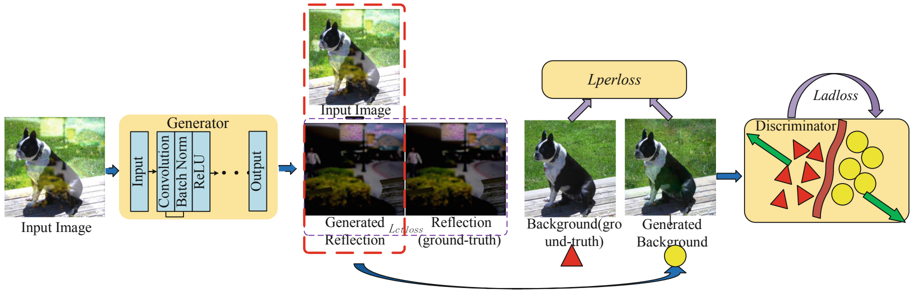


Fig. 2. Overview of the proposed generative adversarial framework. Generator is embedded with deep residual learning to produce reflection image (left). The content loss is conducted between two images in grey dashed box. The generated background image (yellow point) is produced (red dashed box) to be discriminated with the real background image (red triangle) by Discriminator (right). The perceptual loss is also acted as shown. (Color figure online)

image. Let O_b and O_r represent the generated background image and generated reflection image, respectively.

The assumption mentioned above can be interpreted as:

$$I = I_b + I_r \tag{1}$$

And residual learning can be formulated as:

$$O_r = G(I, \theta) \tag{2}$$

where θ is the parameter in the network G.

Combine (1) and (2), we have:

$$O_r = G(I_b, I_r, \theta) \tag{3}$$

Under the assumption mentioned above, the generated background image is formulated by following equation:

$$O_b = I - G(I_b, I_r, \theta) \tag{4}$$

As shown in Fig. 3, a captured image is sent to the network, which learns a mapping between it and its reflection information. The output is a generated reflection image, which is further used to get the generated background image. Experiments reveal that residual learning effectively improves the quality of reflection removal. More details can be found in Sect. 3.

2.3 Design of GAN for Single Image Reflection Removal

The network design of generator of our proposed GAN is shown in Fig. 3. Firstly, a captured image is directly passed into the network with padding in each layer to keep the scale invariant. In the first block, a 9 × 9 convolutional kernel is adopted to enlarge receptive field so as to acquire as much as valuable information. The following two convolutional layers utilize kernels of size 3 × 3. Then its output passes through a series of residual blocks [11], as shown in Fig. 3, which enable the network being extended or squeezed without gradient vanishing and to learn more powerful representation by stacking a different number of them. In the implementation of our case, 36 residual blocks are used in the proposed GAN architecture. Symmetrically, the output is then passed to two convolutional layers with kernel size 3×3 and one with kernel size 9×9. Finally, the output is a residual image captures the reflection information of the input image.

Discriminator plays an important role in the proposed method since the performance, to a large extent, depends on the gradient from it. However, the network of discriminator does not require much complex design but simply a stack of convolutional network layers. Specifically, we adopt eight convolutional layers with a kernel of size 3×3 and an additional convolutional layer with a kernel of size 1×1 to replace fully connection layer. For each interval of the convolutional layer, a batch normalization layer [15] is inserted to normalize the network. Adversarial training is conducive to model the distribution of the generated background image and the real background image when the generator is embedded with residual learning strategy. It is conceivable that the generated background image stems from the captured image without reflection information, which shares a great similarity to the real background image. Therefore, for the latter stages of discriminator training, the challenging cases will dramatically enhance the discriminating ability of discriminator, further providing effective gradient information to instruct generator and the entire model to learn better feature and generate more vivid details coherently.

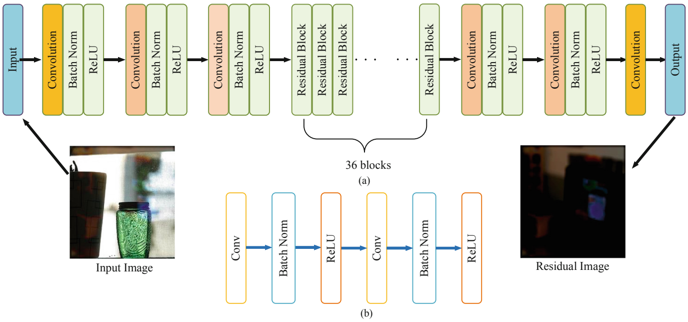


Fig. 3. Network design of generator in proposed GAN.

2.4 Multi-part Balanced Loss

Based on deep residual learning and designed network for single image reflection removal task, an objective function called multi-part balanced loss is proposed. To ensure the gradient information exist both in training the generator and the discriminator, the objective function should balance the network for each part to achieve a different purpose as contributing to the general goal. A content loss measures the similarity between the learned residual information and real reflection information. To encourage the feature of the generated background image to be more consistent with the real background image, a perceptual loss is exploited. Features for perceptual loss is sampled from several layers from a deep pretrained convolutional network, VGG-19 [12]. Along with an adversarial loss which keeps the detail information in the images by improving the discriminating ability of the discriminator, the proposed multi-part balanced loss is composed of the three parts above, instructing the network to produce an image of high quality. Specifically, let G be the generator in our proposed GAN and D be the discriminator. The rest representations are the same as those in Sect. 2.2. For each sample in one mini-batch, the content loss between I_r and O_r can be formulated as:

$$L_{content}(I_r, O_r) = Avg \sum_{i=1}^{w} \sum_{j=1}^{h} MSE(t_{i,j}, o_{i,j}), \tag{5}$$

where w, h represents the weight and height for each sample, $t_{i,j}$ and $o_{i,j}$ represent the target and the predicted value for each pixel, respectively.
Perceptual loss is calculated between the features of I_b and O_b, we indicate perceptual loss by $L_{perceptual}$:

$$L_{perceptual}(I_b, O_b) = \sum_{i=k}^{n} (I_{b,k} - O_{b,k})^2, \tag{6}$$

where k indicate each convolutional layer from the total selected n layers while $I_{b,k}$ and $O_{b,k}$ represent the features for I_b and O_b, respectively.

To implement adversarial training, we adopt binary cross entropy loss for the discriminator as the objective function:

$$L_{adversarial}(I_b, O_b) = log[D(I_b)] + log[D(1 - O_b)], \tag{7}$$

For each time updating discriminator in a mini-batch, the loss is the average of the summation of (6) and (7). The loss for generator is the average of (5) and the second term in (7).
The objective function in the proposed method can be interpreted as:

$$L_{multi-part\ balanced}(G, D) = \min_G \max_D (L_{content} + L_{perceptual} + L_{adversarial}) \tag{8}$$

3 Experiment

In this section, several experiments are conducted on both synthetic and real-world datasets to demonstrate the superiority of the proposed method. On the one hand, we set experiments to show the effectiveness of each component in the proposed method and further analyse the model in terms of quality and quantitative results. On the other, a comparison between the proposed method and state-of-the-art methods is conducted on both synthetic and real-world images. SSIM [14] and PSNR [13] are exploited as metrics to evaluate the generated images.

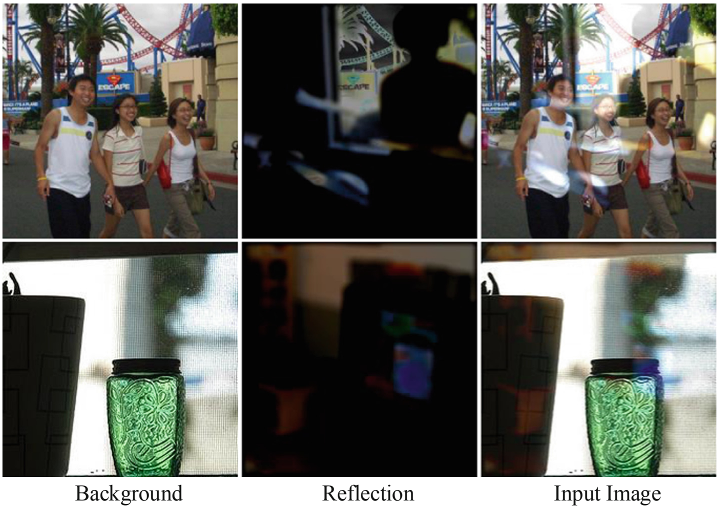


Fig. 4. Synthetic dataset samples. The background scene, the reflection image layer, and the synthetic image (from left to right).

3.1 Dataset

On account of the difficulty of labelling the real-world images, we handcraft a synthetic dataset using the images from PASCAL VOC2012 [16], which is widely adopted by current methods. Based on the valid assumption that the observed image is a compound of the reflection image layer and the clear background image layer, we synthesis one image by adding a reflection layer to a background image. In practice, to make the reflection image more consistent with the true case, a random Gaussian blur is first imposed on it. Subsequently, a clear background image and a randomly selected reflection image are added as an input image. As shown in Fig. 4, the clear background images show more details compared with the reflection images that transmit the moderate light. The synthetic images demonstrate a striking similarity to the real world cases. The synthetic dataset contains 11453 images in total and 500 from which is assigned as a test dataset.

In addition, we test our model on both the synthetic images and real-world images from CEIL dataset [9]. The CEIL real-world image test dataset contains 45 images in total with different degrees of reflection on various background scenes.

3.2 Implementation and Analysis

We implement training the model by deep learning platform PyTorch. It runs on the GPU of NVIDIA Tesla M40 of memory size 24 GB with CUDA version 8.0.44 and CUDNN version 5. The model was trained using Adam optimizer with each mini-batch of 10 and a learning rate of 0.0001. The maximum epoches was set to be 50. The model was tested on 500 pictures from the synthetic dataset.

Table 1. Ablation experiment results on the synthetic test dataset.

Model	SSIM	PSNR
RL	0.8564	22.71
GAN+P	0.8545	23.17
RL+P	0.8841	23.95
RL+GAN+P	**0.8924**	**24.84**

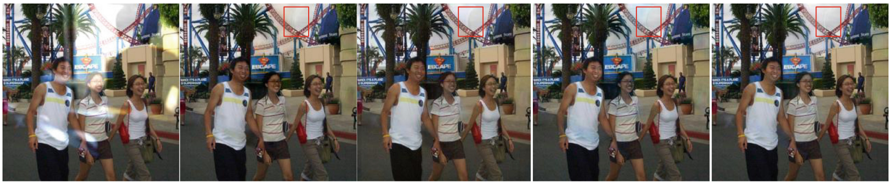


Fig. 5. Qualitative results of ablation experiment with different settings.

To verify the effectiveness of each component in the proposed method. We conduct several ablation experiments in terms of SSIM and PSNR metrics. As shown in Table 1, RL indicates that the model is implemented by residual learning while GAN means the model is trained by adversarial learning in the proposed GAN framework. Letter P represents that the features of predicted and real background images are measured by the perceptual loss function. Simply applying residual learning gains 0.8564 in SSIM and 22.17 in PSNR on average respectively. Comparing the second and the last experiment settings, it shows that deep residual learning contributes to the final results by 0.0379 in SSIM and 1.67 in PSNR, leading a prominent enhancement. In addition, it can be noticed that the proposed GAN is conducive to the performance by comparing the third and the last settings. Perceptual loss evaluates the features of the generated background images and the real background images, which directly benefits the network to produce images of high quality. Therefore, from the results of the first and the third settings, it improves the performance by 0.0277 in SSIM and 1.24 in PSNR averagely.

The qualitative results are shown in Fig. 5, where the visual perception of the proposed method transcends all other settings, which further proves the significance of the deep residual learning and the proposed method.

3.3 Comparison to State-of-the-art Methods

Several experiments are conducted to show both qualitative and quantitative superiorities compared with state-of-the-art methods. Figure 7 compares the visual results on our synthetic datasets, where the proposed method turns out to tackle different complex scenes. As shown in Fig. 8, the proposed method also shows better performance on the images provided by CEILNet [9], where more

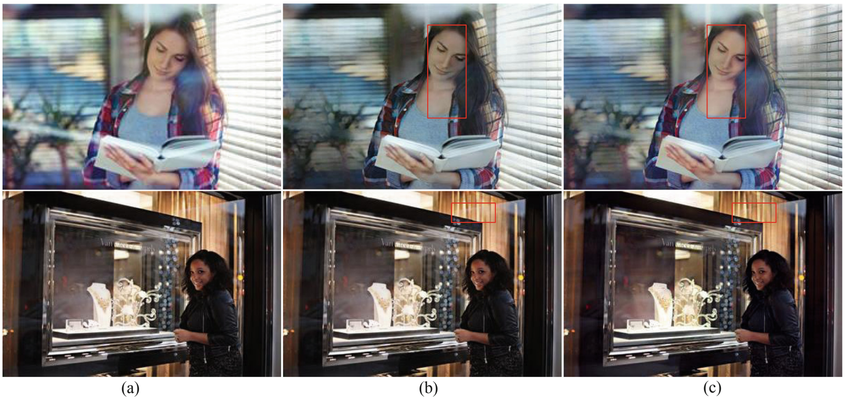


Fig. 6. Comparison of the real-world image with CEILNet.

image details are preserved after reflection removal while other methods leave some remanet artifacts. From top to down, images (a) to (g) show remarkable differences in the marked boxes. The results of the proposed method precede the rest images with clearer background and least reflection remaining.

Samples of real-world image dataset in Fig. 6 further illustrate the exceeding performance especially on fidelity and less distortion of background scene. As we can see, the tone of the objects on the image suffers a degeneration after the process of CEILNet [9], while the result of the proposed method is capable of preserving better color authenticity. From box of the second image, it can be found that the proposed method could deal with the reflection even if it is shown on the confusing background scene. Please zoom in for better view.

Table 2. Quantitative comparison to state-of-the-art methods on images (a) to (g) corresponding to each column in Figs. 7 and 8.

	SSIM				PSNR			
Image	Li [7]	Nikolaos [8]	CEILNet [9]	Ours	Li [7]	Nikolaos [8]	CEILNet [9]	Ours
(a)	0.7379	0.9008	0.9499	**0.9702**	19.7754	23.2607	24.9851	**28.5102**
(b)	0.7497	0.8579	0.8912	**0.9142**	21.4570	20.9500	21.4921	**22.6718**
(c)	0.6434	0.8987	0.9458	**0.9602**	18.0398	21.2330	24.0125	**24.3933**
(d)	0.8359	0.9305	0.9098	**0.9344**	21.6833	22.4701	26.0067	**27.6002**
(e)	0.7261	0.9303	0.9028	**0.9311**	20.0862	23.4430	24.2102	**26.2652**
(f)	0.5784	0.8672	0.9088	**0.9169**	13.2471	22.2741	23.9800	**24.7217**
(g)	0.5522	0.9231	0.8312	**0.9429**	12.8058	25.2152	23.8229	**26.0204**

Table 2 compares the SSIM and PSNR of the images (a) to (g) corresponding to each column in Figs. 7 and 8. Out of fairness, we set the hyperparameters for Nikolaos method [8] to its best adaptiveness for each picture. It is obvious that

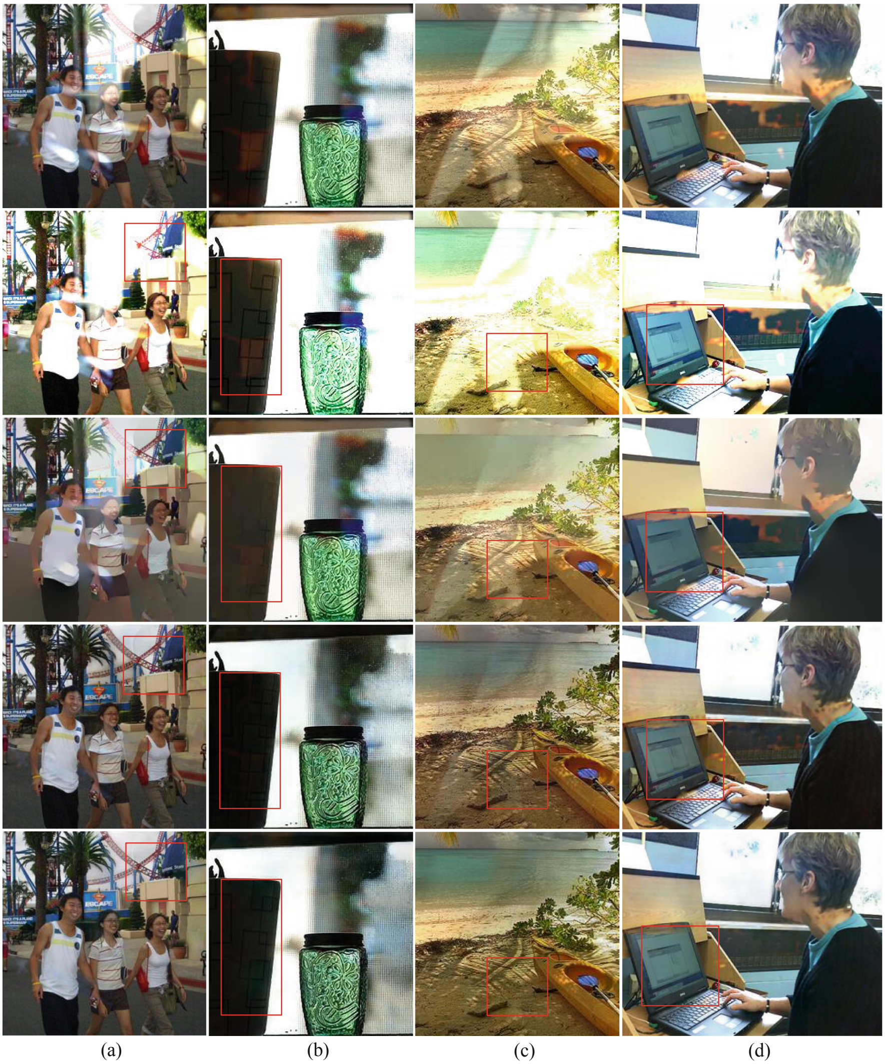
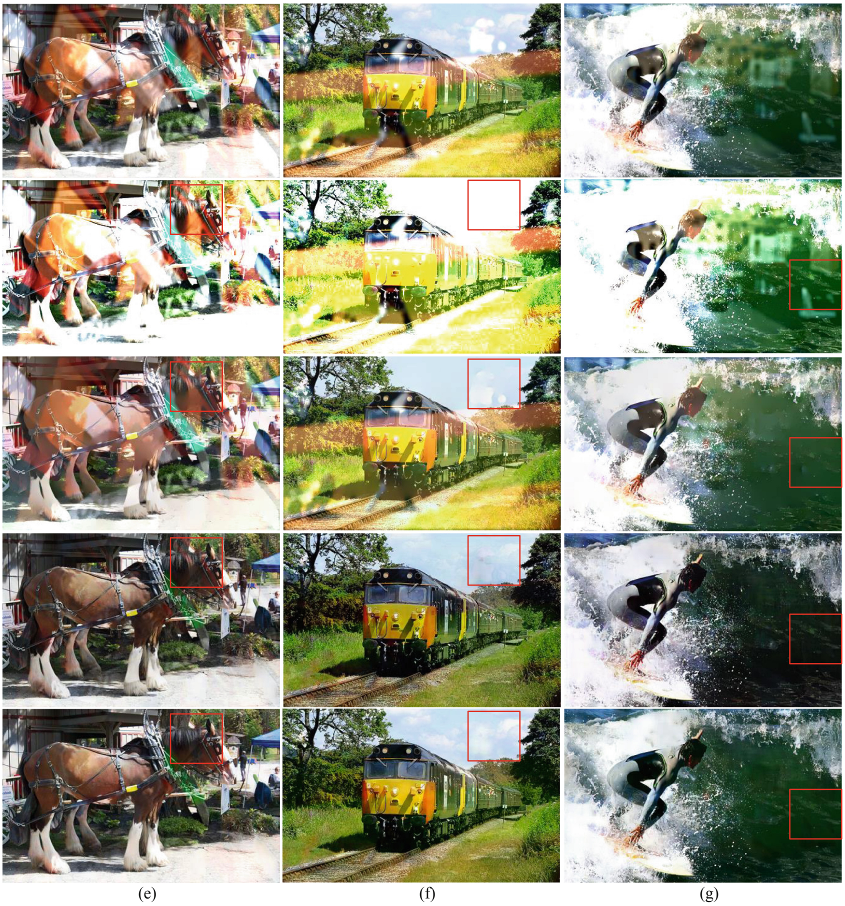


Fig. 7. Reflection removal comparison of the synthetic dataset in this paper with Li, Nikolaos, CEILNet and the proposed method from top to bottom (original images are on the first line).

the evaluation figures of the proposed method outstrip the rest methods, yielding 0.9385 and 25.7404 on average in terms of SSIM and PSNR respectively, which further demonstrates the excellent performance of our proposed method.



(e) (f) (g)

Fig. 8. Comparison of the synthetic images in CEILNet with Li, Nikolaos, CEILNet and the proposed method from top to bottom (original images are on the first line).

4 Conclusion

In this paper, a deep residual learning approach is proposed to address the single image reflection removal problem. Instead of eliminating or suppressing the reflection information from the observed images directly, the proposed method first exploits the mapping between the observed image and its reflection information which is comparatively simple, and then removed it to obtain a clear background image. A novel generative adversarial framework is also proposed that dramatically improves the performance through deep residual learning and adversarial training. Furthermore, a multi-part balanced loss is introduced by considering both the reflection information learning and the background feature similarity measurement simultaneously. The proposed method entails a great performance by keep the intactness of the background scene. Experiments compared with several state-of-the-art methods reveal a significant meaning of deep residual learning and effectiveness of the proposed method on the single image reflection removal task.

References

1. Levin, A., Weiss, Y.: User assisted separation of reflections from a single image using a sparsity prior. IEEE Trans. Pattern Anal. Mach. Intell. **29**(9), 1647–1654 (2007)
2. Han, B.J., Sim, J.Y.: Reflection removal using low-rank matrix completion. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5438–5446 (2017)
3. Kong, N., Tai, Y.-W., Shin, S.Y.: A physically-based approach to reflection separation. In: 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 9–16. IEEE (2012)
4. Schechner, Y.Y., Shamir, J., Kiryati, N.: Polarization-based decorrelation of transparent layers: the inclination angle of an invisible surface. In: Proceedings of the Seventh IEEE International Conference on Computer Vision (ICCV), vol. 2, pp. 814–819. IEEE (1999)
5. Schechner, Y.Y., Kiryati, N., Basri, R.: Separation of transparent layers using focus. Int. J. Comput. Vis. **39**(1), 25–39 (2000)
6. Agrawal, A., Raskar, R., Nayar, S.K., Li, Y.: Removing photography artifacts using gradient projection and flash-exposure sampling. ACM Trans. Graph. (TOG) **24**(3), 828–835 (2005)
7. Li, Y., Brown., M.S.: Single image layer separation using relative smoothness. In: 2014 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2752–2759. IEEE Computer Society (2014)
8. Arvanitopoulos, N., Achanta, R., Susstrunk, S.: Single image reflection suppression. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4498–4506 (2017)
9. Fan, Q., Yang, J., Hua, G., Chen, B., Wipf, D.: A generic deep architecture for single image reflection removal and image smoothing. In: 2017 IEEE International Conference on Computer Vision (ICCV), pp. 3238–3247 (2017)
10. Goodfellow, I., et al.: Generative adversarial nets. In: Advances in Neural Information Processing Systems, pp. 2672–2680 (2014)
11. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)
12. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)
13. Huynh-Thu, Q., Ghanbari, M.: Scope of validity of PSNR in image/video quality assessment. Electron. Lett. **44**(13), 800–801 (2008)
14. Wang, Z., Bovik, A.C., Sheikh, H.R., Simoncelli, E.P.: Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)
15. Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167 (2015)

16. Everingham, M., Eslami, S.A., Van Gool, L., Williams, C.K., Winn, J., Zisserman, A.: The pascal visual object classes challenge: a retrospective. Int. J. Comput. Vis. **111**(1), 98–136 (2015)
17. Li, J., et al.: A probabilistic hierarchical model for multi-view and multi-feature classification. In: Thirty-Second AAAI Conference on Artificial Intelligence (2018)
18. Li, J., et al.: Generative multi-view and multi-feature learning for classification. Inf. Fusion **45**, 215–226 (2019)