DOI: 10.21917/ijivp.2022.0387

SINGLE IMAGE REFLECTION REMOVAL WITH SEGMENTATION

Rashmi Chaurasiya and Dinesh Ganotra
Department of Physics, Indira Gandhi Delhi Technical University for Women, India

Abstract

Removal of reflection is of high importance to reclaim the original background image. Several attempts have been made to separate reflection from background. A number of approaches are based on assuming certain conditions about the reflective material (glass) and type of reflection. Humans can separate familiar objects easily due to the understanding of the objects in scene, same analogy is applied here. In this paper, additional information of segmentation map is utilized rather than using a single reflection image as input. Estimated segmentation map corresponds to the composite image. Our aim is to investigate the efficacy of segmentation map in reflection removal approaches. Proposed method performs adequately on real-world images and suppresses the reflection components in background effectively.

Keywords:

Reflection Removal, Deep Learning, Semantic Guidance

1. INTRODUCTION

. Reflection removal is not a new problem statement. Several attempts have been made to solve it. This arises when a glass pane comes in between the camera and the desired scene. Additional glass pane results in the superposition of the resultant scene with the desired background. It is an ill-posed problem where each reflection image is corrupted with any random real scene. This scene is distributed haphazardly in the background scene that makes some of the regions in the background brighter and some darker. Sometimes it completely obscures the background that even a human cannot distinguish between the two layers of transmission (T) and reflection (R).

Reflection removal approaches can be broadly classified into two types. There are approaches relying on handcrafted priors or optimization-based approaches and the other is learning type approaches based on deep convolutional neural networks. The first type of method often requires user assistance or some prior assumption.

Mathematically this problem can be viewed as a layer separation problem, where composite image (I) is the linear combination of background layer/transmission layer and reflection layer.

$$I = B + R \tag{1}$$

In the above equation, two variables (R and B) have to be estimated with a single variable I. There exist infinite solutions for this problem. That makes reflection removal an ill-posed problem.

The idea behind the proposed work is to provide an understanding of the objects in the scene so they can be differentiated based on which layer they belong to. This same system is utilized by humans as well to separate objects that is in case of a human face - eyes, nose and ears belong to the face so anything other than that on that location does not belong to the same layer, that has to be the additional noise (reflection), hence the two layers reflection and transmission can be separated that way.

Fig.1. Resultant images of RAGNet [2] and ours for the same input

Usage of semantic information (segmentation map) [1] in reflection removal is not new, however it is not one of the popular approaches. This paper is dedicated to evaluate the robustness of semantic guidance in reflection removal, where the performance is compared on the existing datasets. To fulfill this task, a region aware guidance [2] is utilized, it is a very recent method in reflection removal that shows outstanding results and surpasses all the existing performances. Here in this work RAGNet's [2] network structure is utilized as our base model. With same training conditions performance evaluation can be performed easily, Fig.1 compares the result of RAGNet [2] and ours on real images.

Current work aims to validate whether the performance of RAGNet [2] can be improved with the usage of semantic information.

The rest of the paper is organized as follows, section 2 catalogues related methods categorized as prior and optimization-based methods and learning methods. Section 3 covers the proposed methodology that includes dataset creation, network structure and loss function. Section 4 shows the result of proposed work on various synthetic as well as on real images. Section 5, 6 explain the limitation and future work and conclusion section.

Contribution of this paper are as follows:

- Proposed work enquires the efficacy of segmentation maps in reflection removal. Here an additional input of segmentation map is provided to Li et al. [2]'s network with all the other parameters at same values.
- In this work segmented images are used as an additional input that works as a significant feature. They assist the method in the separation of targeted objects in the scene from its noisy background.
- Existing works target on predicting the noise i.e., reflection in the image first then estimate the background in the second

step. This work focuses on locating the object of interest in the image first.

- Performance of ours and Li et al. [2] is analyzed qualitatively and quantitatively to examine the proposed claim.

In this work it is shown that additional input does not degrades the performance of the base network's [2] performance. Compared performance shows similar SSIM values for both of the methods. For images having objects recognized by segmentation model, our work rather improves the performance of Li et al. [2].

The rest of the paper is organized as follows section two list down the existing methods and categorized them as prior based and learning based methods, section three explains the proposed methodology that includes the subsections of dataset creation, network parameters and loss function, section four compares the result of the proposed work with the existing state of the art. In the end section five and six concludes this work with limitation and future work and conclusion.

2. RELATED WORK

There exist several reflection removal methods, these methods can be classified based on the number of input images fed or on the nature of the method. In terms of the number of inputs, there are single image-based methods and multiple image-based methods. The other way to organize these methods is based on their methodology, for instance, if they are solved with optimization-based method or learning based method as stated in equation (1) reflection removal from a single image is a highly ill-posed problem. It requires human intervention and several priors to solve the problem. Levin and Weiss's [3] method allows the user to assist the solution by manually marking the reflection area in the image. Multiple image methods require more than one image in input, these images can be images in pairs with flash and no flash [4], images captured at different viewpoints [5]-[7], images through a polarizer with different orientation [8]-[10] and a video sequence [11]. Reflection removal methods can be classified in two main categories, single image-based method and multiple image-based method. Here methods are organized as traditional optimization-based methods and learning based methods.

2.1 PRIOR AND OPTIMIZATION BASED METHODS

These techniques utilize assumptions, additional priors etc. The most common assumption [12] is to consider the reflection layer being blurrier than the transmission layer. This is based on the fact that reflected objects are at different distances from the camera that lead to different blur levels of R and T. So, the two layers can be separated on the basis of their gradient distribution. Then a probabilistic model to regularize the gradients of the two layers is proposed. In [3] user annotations are used to distinguish the two layers; user assistance is later combined with gradient sparsity prior by Levin et al. [13]. Wan et al. [14] used a multi-scale depth of field-guided map to classify the edges belonging to reflection and transmission. Arvanitopoulos et al. [15] proposed an optimization function, where the Laplacian data fidelity term is introduced to secure the fine details of the transmission image. Shih et al. [16] studied ghosting cues, which is a categorical phenomenon when the glass has a certain thickness, and employed a patch-predicated GMM prior to model the natural image for reflection removal.

2.2 LEARNING BASED

There is an emerging trend of applying deep convolutional neural networks for reflection removal, Fan et al. [17] were the first in this edition. In their [17] model they estimated the edge map of background which later followed by the reconstruction of background in the color domain. [18] followed the same footprints and generated a co-operative sub-network rather than a two-stage approach. Yang et al. [19] presented Bidirectional Network (BDN) to predict background and reflection layer sequentially, they used B to estimate R and used R to estimate B. Recent methods [20] [21] adopted a combination of loss functions rather than using a single mean squared loss (MSE), a combination of perceptual [22], adversarial [23], MSE is used. Perceptual loss provides a comparison of multi-stage features of a deep neural network pre-trained on ImageNet [25]. Apart from these, existing methods also experimented with various network architectures as convolutional neural network [26], encoder-decoder [21] and LSTM neural networks [27]. Lack of the real-world aligned training data is also a big concern in this field, to solve this problem Wei et al. [28] proposed a method to utilize un-aligned data. Due to the practicability in real life, majority of the learning-based methods are single image-based method.

3. PROPOSED METHODOLOGY

3.1 DATASET CREATION

3.1.1 Training Data:

Dataset utilized in this paper is created synthetically, for which COCO dataset [29] is used. Half set is assigned as reflection layer and other half is assigned as background. After some preprocessing like center crop, resizing and discarding faulty images, reflection is added using ERRNet [28]. Then segmentation maps of these created reflection images are generated using Deep Lab V3+ [30] pre-trained model. This model has total 21 classes of objects out of which 20 classes are for objects and one for background.

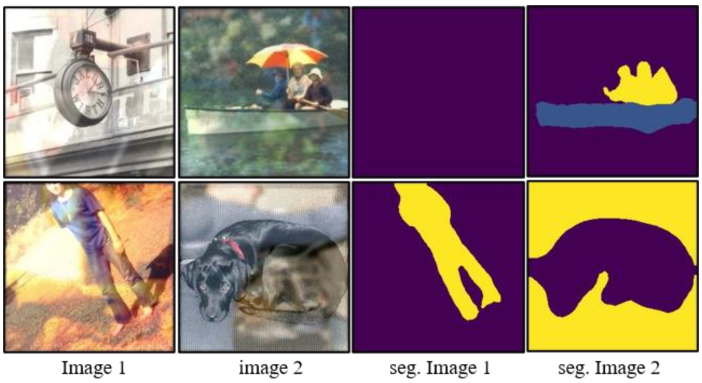
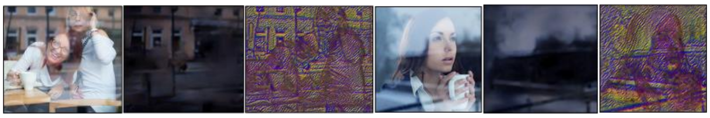


Fig.2. Created Set

This pre-trained model produces most of the images as blank images as most of the objects do not belong to listed classes in pre-trained model. Since segmentation maps work as an important feature in this work, segmentation need to be done precisely. To address this problem another training set is created using PASCAL VOC2012 [31] dataset. This dataset contains images that belong to the object classes in DeepLabV3 [30] model, precise segmentation can be ensured this way. Sample from both of these training set is shown in Fig.2 where first row shows the segmented (referred as seg. Image) images for COCO [29] dataset and second row show images on VOC2012 [31] dataset. It is visible from the images that segmentation for COCO [30] dataset is not accurate. This defeats the purpose of this work, which is to determine the efficacy of segmentation in reflection removal.

3.1.2 Evaluation Data:

Three subsets from SIR2 [32] dataset for testing, these datasets include Wild, postcard and solid image sets with images 55, 199 and 200 respectively. In all three subsets of SIR2, three image are available observed image, reflection and GT background Apart from that real45 [17] image set is used for real images. This dataset does not have their corresponding GT images so performance can be compared only through visual mode.



Input G_R output $G_O$output Input $G_R$output $G_O$output

Fig.3. Output comparison of G_R and G_O module.

3.2 NETWORK PARAMETERS

3.2.1 Structure:

Two images as inputs are provided in the network structure (Fig.4). One is synthesized reflection image and second is its corresponding segmented image. Images are fed to the network as a six-channel input, first three channels correspond to RGB image and next three channels are for segmentation image. Network structure used here is adapted from a recent work [2], which is a two-stage network. It's GR module estimates reflection layer and GT module predicts background image with the estimated reflection layer. In this work GR module is replaced with GO where GO directs the network about the targeted object. Li et al. [2]'s GR module predicts reflection layer and ours GO module predicts the location of targeted object in the image. Output of both GR and GO is shown in Fig.3, it is clearly visible that GR highlights the reflection and GO highlights the targeted object.

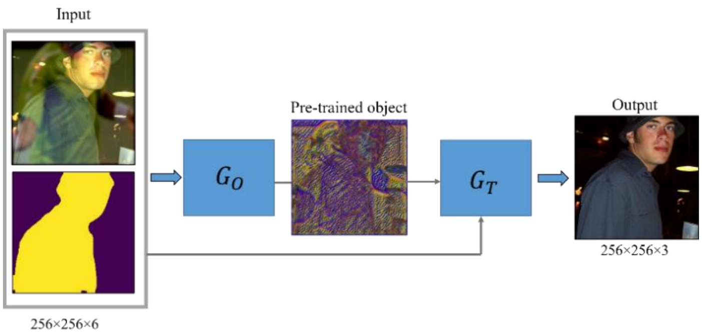


Fig.4. Network Structure

In Fig.4, GO is a U-Net [33] model and GT is a combination of two encoders and one decoder. Working of GT is same as [2] where it produces background using two parameters observed image and estimated object layer. Learning rate set at 0.001 for total 500 epochs.

3.2.2 Loss Function:

Originally Li et al. used five losses reconstruction, perceptual [22], adversarial [23], exclusion [20] and mask loss, here only first three losses are used. Reconstruction loss calculates pixel loss between the predicted background and ground truth (GT) background. Perceptual loss uses VGG19 [34] network to compare features $\varphi(T')$ and $\varphi(T)$ in the selected five feature layers. $\varphi(T')$ and $\varphi(T)$ corresponds to predicted and GT background.

Adversarial loss is the generator discriminator loss, where whole network is considered as generator and an additional four-layer network acts as discriminator.

Loss function used here is given by eq. (2)

$$Loss\ Function = Reconstruction_{loss} + 0.2*F_{loss} + 0.01*A_{loss} + E_{loss} \quad (2)$$

where F_{loss} refers feature loss, A_{loss} is adversarial loss and E_{loss} is exclusion loss. Here exclusion loss suppresses reflection components from the output.

4. RESULTS

4.1 QUANTITATIVE MEASUREMENT

Proposed method is compared against the state-of-the-art methods including Zhang et al. [20], ERRNet [28], IBCLN [27] and RAGNet [2]. Results are compared on three sub-datasets from SIR2 [32], number of images in these test sets are given in the first column of Table.1 along with their names. PSNR (in dB) and SSIM are used for quantitative measurement (shown in Eq.(2) and Eq.(3)), resultant values are shown in the format of average PSNR/average SSIM. Although segmentation map for COCO [29] dataset is not accurate for most of the images, its performance is closer to [2] in accordance with the SSIM values. Results for this work is shown in sixth column as "ours 1".

For VOC2012 [31] dataset (shown in seventh column as "ours 2") proposed method achieved similar SSIM (bolded) values with RAGNet [2] and performed better than our previous studied case "ours 1". It also surpasses the other state-of-the-art methods in SSIM and PSNR listed in the Table.1. However, it could not outperform the PSNR values of RAGNet [2].

$$PSNR = 20\log_{10}\frac{(255)^2}{MSE} \quad (2)$$

$$SSIM = \frac{(2\mu_s\mu_{s'} + C_1)(2\sigma_{ss'} + C_2)}{(\mu_s^2 + \mu_{s'}^2 + C_1)(\sigma_s^2 + \sigma_{s'}^2 + C_2)} \quad (3)$$

4.2 QUALITATIVE MEASUREMENT

Fig.5 represents result on four images from real45 [17] real world dataset compared visually. Where and input, output of RAGNet [2] and our dataset are shown, ground truth is not available for this dataset. All the four resultant images points to the conclusion that the main object is more highlighted in our images than RAGNet [2]' resultant images. That is, reflected scene in background is more suppressed in our images. This is owing to the fact that our focus is on locating the object of interest in the image rather than locating the reflection area.

5. LIMITATION AND FUTURE WORK

The approach formulated here is specifically designed to provide an understanding of the scene in reflection removal. Images segmented

Table.1. Average PSNR and SSIM values of our results.

	Zhang [20]	**ERRNet [28]**	**IBCLN [27]**	**RAGNet [2]**	**Ours1**	**Ours2**
Solid200 (199)	23.37/ 0.875	24.85/ 0.894	24.88/ 0.893	26.16/ 0.90	25.35/ 0.90	25.20/ 0.90
Postcard (199)	21.07/ 0.80	24.16/ 0.847	24.71/ 0.886	25.52/ 0.86	23.79/ 0.86	24.22/ 0.86
Wild55 (55)	16.81/ 0.797	21.99/ 0.874	23.39/ 0.875	23.67/ 0.87	21.78/ 0.86	21.64/ 0.87

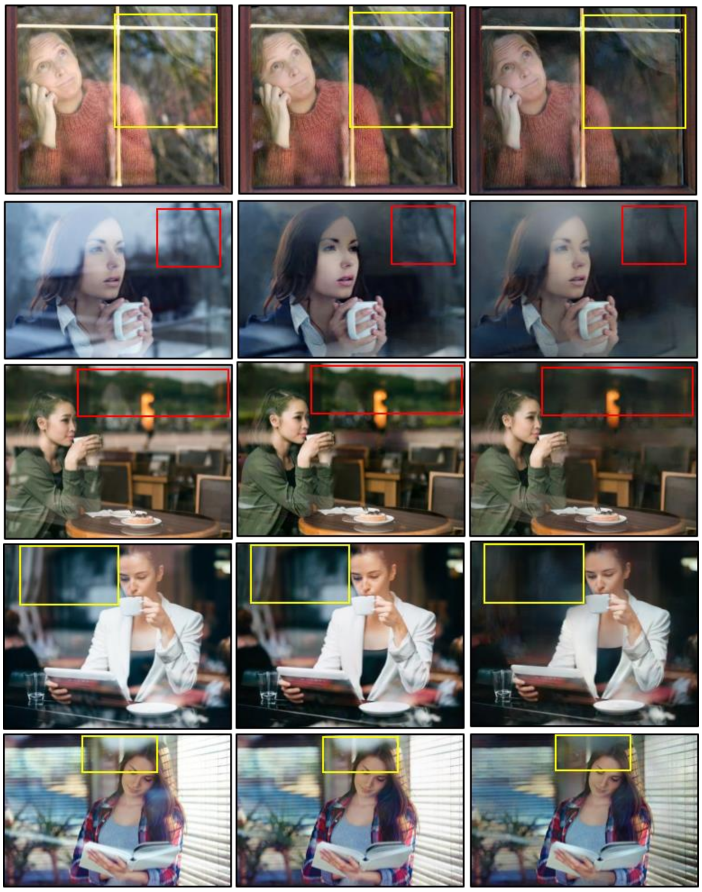


Fig.5. Real test set results

accurately leads to the improved performance than [2] on real45 test set. For images having the object other than the specified objects in Deep Lab V3+ [30] model, proposed work performs weakly than RAGNet (as shown in Table.1). However, result on real images proves that using semantic information in reflection removal maintains similar performance as original base model [2]. The future scope includes additional refinement in segmentation model and more efforts in this direction to validate the proposed aim.

6. CONCLUSION

Proposed work is primarily focused on two aims, first to improve the recent published [2] method's performance and second is to determine the efficacy of segmentation in reflection removal. Outcomes of this paper are compared with all the existing recent works. Performance evaluation suggests that this work has relatively similar performance to the base model (RAGNet) with additional benefit on real test set images. However, this additional advantage provided here is limited to the classes of segmentation map. That is, when an object in the image is one of objects that Deep Lab V3+ [28] recognizes only then it can provide correct segmentation map.

REFERENCES

[1] Y. Liu, Y. Li and F. Lu, "Semantic Guided Single Image Reflection Removal", *Proceedings of International Conference on Recent Trends in Computer Science*, pp. 1-7, 2019.

[2] Li Yu, "Two-Stage Single Image Reflection Removal with Reflection-Aware Guidance", *Proceedings of International Conference on Image Processing*, pp. 1-9, 2021.

[3] A. Levin and Y. Weiss, "User Assisted Separation of Reflections from a Single Image using a Sparsity Prior", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 29, No. 9, pp. 1647-1654, 2007.

[4] A. Agrawal, R. Raskar and Y. Li, "Removing Photography Artifacts using Gradient Projection and Flash Exposure Sampling", *ACM Transactions on Graphics*, Vol. 24, No. 3, pp. 828-835, 2005.

[5] K. Gai, Z. Shi and C. Zhang, "Blind Separation of Superimposed Moving Images using Image Statistics", *Proceedings of IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 19-32, 2012.

[6] Y. Li and S.M. Brown, "Exploiting Reflection Change for Automatic Reflection Removal", *Proceedings of IEEE International Conference on Computer Vision*, pp. 2432-2439, 2013.

[7] X. Guo, X. Cao and Y. Ma, "Robust Separation of Reflection from Multiple Images", *Proceedings of International Conference on Computer Vision and Pattern Recognition*, pp. 2187-2194, 2014.

[8] N. Kong, Y. Tai and J.S. Shin, "A Physically-Based Approach to Reflection Separation: from Physical Modeling to Constrained Optimization", *Proceedings of IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 36, No. 2, pp. 209-221, 2014.

[9] Y.Y. Schechner, J. Shamir and N. Kiryati, "Polarization and Statistical Analysis of Scenes Containing a Semireflector", *Journal of the Optical Society of America*, Vol. 17, No. 2, pp. 276-284, 2000.

[10] B. Sarel and M. Irani, "Separating Transparent Layers through Layer Information Exchange", *Proceedings of International Conference on Electronics and Computer Vision*, pp. 328-341, 2004.

[11] T. Xue, M. Rubinstein and W.T. Freeman, "A Computational Approach for Obstruction-Free Photography", *Proceedings of ACM Transactions on Graphics*, Vol. 34, No. 4, pp. 1-11, 2015.
[12] Y. Li and M.S. Brown, "Single Image Layer Separation using Relative Smoothness", *Proceedings of International Conference on Electronics and Computer Vision*, pp. 2752-2759, 2014.
[13] A. Levin, A. Zomet and Y. Weiss, "Learning to Perceive Transparency from the Statistics of Natural Scenes", *Proceedings of International Conference on Advances in Neural Information Processing Systems*, pp. 1271-1278, 2003.
[14] R. Wan, B. Shi, A.H. Tan and A.C. Kot, "Depth of Field Guided Reflection Removal", *Proceedings of International Conference on Electronics and Computer Vision*, pp. 21-25, 2016.
[15] N. Arvanitopoulos, R. Achanta and S. Susstrunk, "Single Image Reflection Suppression", *Proceedings of International Conference on Computer Vision and Pattern Recognition*, pp. 1752-1760, 2017.
[16] Y. Shih, D. Krishnan and W.T. Freeman, "Reflection Removal using Ghosting Cues", *Proceedings of International Conference on Computer Vision and Pattern Recognition*, pp.3193-3201, 2015.
[17] Q. Fan, J. Yang and D. Wipf, "A Generic Deep Architecture for Single Image Reflection Removal and Image Smoothing", *Proceedings of IEEE International Conference on Computer Vision*, pp. 1-13, 2017.
[18] R. Wan, B. Shi and A.C. Kot, "CRRN: Multi-Scale Guided Concurrent Reflection Removal Network", *Proceedings of International Conference on Computer Vision and Pattern Recognition*, pp. 4777-4785, 2018.
[19] J. Yang, D. Gong, L. Liu and Q. Shi, "Seeing Deeply and Bidirectionally: A Deep Learning Approach for Single Image Reflection Removal", *Proceedings of International Conference on Electronics and Computer Vision*, pp. 654-669, 2018.
[20] X. Zhang, R. Ng and Q. Chen, "Single Image Reflection Separation with Perceptual Losses", *Proceedings of International Conference on Electronics and Computer Vision and Pattern Recognition*, pp. 4786-4794, 2018.
[21] Z. Chi, X. Wu and J. Gu, "Single Image Reflection Removal using Deep Encoder-Decoder Network", *Proceedings of International Conference on Electronics and Computer Vision*, pp. 1-13, 2018
[22] J. Johnson, A. Alahi and L. Fei-Fei, "Perceptual Losses for Real-Time Style Transfer and Super-Resolution", *Proceedings of European Conference on Computer Vision*, pp 694-711, 2016.
[23] D. Lee, M.H. Yang and S. Oh, "Generative Single Image Reflection Separation", *Proceedings of International Conference on Electronics and Computer Vision*, pp. 78-88, 2018
[24] H. Zhang, K. Dana, J. Shi and A. Agrawal, "Context Encoding for Semantic Segmentation", *Proceedings of International Conference on Computer Vision and Pattern Recognition*, pp. 7151-7160, 2018.
[25] O. Russakovsky, J. Deng and M. Bernstein, "ImageNet Large Scale Visual Recognition Challenge", *International Journal of Computer Vision*, Vol. 115, No. 3, pp. 211-252, 2015.
[26] Y. Chang and C. Jung, "Single Image Reflection Removal Using Convolutional Neural Networks", *IEEE Transactions on Image Processing*, Vol. 28, No. 4, pp. 1954-1966, 2019.
[27] C. Li, Y. Yang and J. Hopcroft, "Single Image Reflection Removal through Cascaded Refinement", *Proceedings of International Conference on Computer Vision and Pattern Recognition*, pp. 3562-3571, 2019.
[28] K. Wei and H. Huang, "Single Image Reflection Removal Exploiting Misaligned Training Data and Network Enhancements", *Proceedings of International Conference on Computer Vision and Pattern Recognition*, pp. 8170-8179, 2019.
[29] D. Fleet, D. "Microsoft COCO: Common Objects in Context", *Proceedings of International Conference on Conference on Computer Vision*, pp. 90-98, 2014.
[30] L.C. Chen, Y. Zhu and H. Adam, "Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation", *Proceedings of International Conference on Conference on Computer Vision*, pp. 1-13, 2018.
[31] M. Everingham and A. Zisserman, "The PASCAL Visual Object Classes Challenge", Available at http://www.pascal-network.org/challenges/VOC/voc2012/workshop/index.html, Accessed at 2012.
[32] R. Wan, B. Shi and C.A. Kot, "Benchmarking Single-Image Reflection Removal Algorithms", *Proceedings of International Conference on Conference on Computer Vision*, pp. 441-447, 2017.
[33] O. Ronneberger, P. Fischer and T. Brox, "Unet: Convolutional Networks for Biomedical Image Segmentation", *Proceedings of International Conference on Conference on Medical Image Computing and Computer-assisted Intervention*, pp. 234-241, 2015.
[34] S. Liu and W. Deng, "Very Deep Convolutional Neural Network based Image Classification using Small Training Sample Size", *Proceedings of International Conference on Conference on Pattern Recognition*, pp. 730-734, 2015.