Single Image Reflection Removal with Absorption Effect

Qian Zheng[1]* Boxin Shi[2,3,4]* Jinnan Chen[1] Xudong Jiang[1] Ling-Yu Duan[2,4] Alex C. Kot[1]

[1]School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore
[2]NELVT, Department of Computer Science and Technology, Peking University, Beijing, China
[3]Institute for Artificial Intelligence, Peking University, Beijing, China
[4]Peng Cheng Laboratory, Shenzhen, China
{zhengqian, exdjiang, eackot}@ntu.edu.sg, {shiboxin, lingyu}@pku.edu.cn, jinnan001@e.ntu.edu.sg

Abstract

In this paper, we consider the absorption effect for the problem of single image reflection removal. We show that the absorption effect can be numerically approximated by the average of refractive amplitude coefficient map. We then reformulate the image formation model and propose a two-step solution that explicitly takes the absorption effect into account. The first step estimates the absorption effect from a reflection-contaminated image, while the second step recovers the transmission image by taking a reflection-contaminated image and the estimated absorption effect as the input. Experimental results on four public datasets show that our two-step solution not only successfully removes reflection artifact, but also faithfully restores the intensity distortion caused by the absorption effect. Our ablation studies further demonstrate that our method achieves superior performance on the recovery of overall intensity and has good model generalization capacity. The code is available at https://github.com/q-zh/absorption.

1. Introduction

When light interacts with a plate glass surface, it can be partly absorbed, reflected, and transmitted. A widely used formation model of a reflection-contaminated image $\mathbf{I}$ is formulated as[1]

$$\mathbf{I} = \mathbf{\Omega}\mathbf{T} + \mathbf{\Phi}\mathbf{R}, \tag{1}$$

where $\mathbf{\Omega}$ and $\mathbf{\Phi}$ represent refractive and reflective amplitude coefficient maps, $\mathbf{T}$ and $\mathbf{R}$ represent transmission and reflection images. Recovering transmission image $\mathbf{T}$ from a single reflection-contaminated image $\mathbf{I}$ is challenging due to its ill-posedness [1]. As can be observed from Equation (1), such ill-posedness is not only caused by unknown content of $\mathbf{T}$ and $\mathbf{R}$, but also by unknown content-free variables of $\mathbf{\Omega}$ and $\mathbf{\Phi}$. Most existing methods rely on image content for single image reflection removal, *i.e.*, explicit priors from image gradients (*e.g.*, [2]) and dictionaries (*e.g.*, [3]), or implicit priors from training data (*e.g.*, [4]). The nature of content-free variables is widely studied to solve the problem of multi-images reflection removal (*e.g.*, [5, 6]) while they are seldom considered in the context of single image reflection removal. Besides, most existing solutions (*e.g.*, single image [7], multi-images [8]) are based on an ideal image formation model that does not take the *absorption effect* (defined in Section 3) into account, *i.e.*, assuming the glass to be thin enough. A recent work solves for $\mathbf{\Omega}\mathbf{T}$ instead of $\mathbf{T}$ as the absorption effect can significantly darken the transmission [9].[2] Since the absorption effect is independent from image content while varying with different colors, thicknesses, or orientations of glass in the real-world, considering the absorption effect can help mitigate the ill-posedness of single image reflection removal.

In this paper, we revisit the formation model of the reflection-contaminated image by taking the absorption effect into account (Section 3). According to the results of Monte Carlo simulation [10], we observe that the absorption effect can be numerically represented by the average of refractive amplitude coefficient map, defined as avg($\mathbf{\Omega}$) (Section 5.1). As the content-free variable avg($\mathbf{\Omega}$) fluctuates in the simulation, we argue that an accurate estimation of avg($\mathbf{\Omega}$) can benefit in solving the problem of single image reflection removal. To this end, we propose a two-step solution to first estimate avg($\mathbf{\Omega}$) and then recover the transmission image $\mathbf{T}$ through two neural networks. To obtain an accurate estimation of avg($\mathbf{\Omega}$) in the first step, we adopt a two-branch training strategy by taking $\mathbf{I}$ and $\mathbf{T}$ as inputs. The core idea is to reduce the influence from image content of transmission and reflection while propagating discriminative features of content-free variable avg($\mathbf{\Omega}$) across the layers of our neural network (Section 4.1). We also constrain the second step with the Lipschitz condition [11] to increase the generalization capacity regarding diverse avg($\mathbf{\Omega}$) (Section 4.2). Our method achieves a superior performance advantage on public datasets. In summary, our contributions are as follows,

- We propose the first formulation to consider the absorption effect in the context of reflection removal. We further show that the absorption effect can be numerically

*Corresponding authors.

[1]As there is no matrix multiplication in this paper, we redefine the matrix multiplication as the element-wise multiplication for simplicity.

[2]Please find our experiments of fitting the absorption effect for real data in the supplementary material.

approximated by the average of refractive amplitude coefficient map.

- We propose a two-step solution, with a two-branch training strategy and the constraint of Lipschitz condition, to solve the problem of single image reflection removal with the consideration of absorption effect. We further analyze how the proposed method facilitates estimating the absorption effect and recovering the transmission image from a single reflection-contaminated image.
- We show by experiments that our method not only successfully removes reflection artifact, but also faithfully restores the intensity distortion caused by the absorption effect. We further demonstrate that our method achieves a superior performance advantage on the recovery of overall intensity and has good model generalization capacity.

2. Related Work

Multi-images reflection removal methods commonly leverage constraints from content-free variables to alleviate the ill-posedness of the solution, *e.g*., polarization angles [8, 5, 12], or reflection disparity [6]. However, priors from content-free variables are seldom considered in the context of a single image. Most single image reflection removal methods impose priors from image content and they can be roughly divided into optimization-based and deep learning-based methods.

Optimization-based methods. Assumptions have been made in the literature to make the problem of single image reflection removal tractable due to its massive ill-posedness. Optimization-based methods exploit the statistics of natural images to make explicit assumptions. For example, Levin and Weiss [13] separate transmission and reflection images based on an image gradient sparsity prior with manual annotations. Li and Brown [1] utilize a smooth image gradient prior since reflection images are likely to be out-of-focus and blurry. Shih *et al*. [14] remove reflection effect based on the observation of ghosting cues. Wan *et al*. [15] perform edge classification for transmission and reflection images through multi-scale depth of fields. Arvanitopoulos *et al*. [2] suppress the reflection by an ℓ_0 penalty on the gradient of recovered transmission images. Wan *et al*. [16] employ both content and image gradient priors to jointly restore missing content and recover transmission images. Yang *et al*. [17] suppress the reflection by solving a partial differential equation. Optimization-based methods generally produce over-smooth results and fail to generalize to various types of reflection in real-world when their assumptions violate.

Deep learning-based methods. Inspired by the unprecedented success achieved by deep convolutional neural networks in versatile low-level vision problems [18], researchers propose several practical data-driven methods to produce robust predictions for transmission images against various types of reflection. Different from optimization-based methods that explicitly transfer the prior knowledge to the exactly formulated constraints, data-driven methods attempt to learn such knowledge from data and are expected to generalize to various types of reflection included in training data. A recent study has shown that directly training an image processing neural network between inputs and outputs can overfit the regression model [19] due to its ill-posedness. Thus, lots of efforts are made to narrow the solution space and reduce the ill-posedness during the optimization of neural networks. Fan *et al*. [20] estimate edge information to guide the recovery of transmission images to better preserve details. Zhang *et al*. [21] further explicitly utilize perceptual information during training to improve the realism of predictions. Wan *et al*. [22] jointly optimize an image gradient estimation network and an image inference network for the transmission layer and capture real reflection images to simulate reflection effects for their training data, and they extend this work by integrating image context information and considering the image gradient level statistics during training [23]. Yang *et al*. [24] use a cascade network structure to jointly recover transmission images and reflection images. Wen *et al*. [7] additionally use a reconstruction error of the reflection-contaminated image to further constrain the estimations. Wei *et al*. [4] restore missing content caused by strong reflection through context encoding modules and use an alignment-invariant loss to address the misalignment in real-world images. Kim *et al*. [25] adopts physically-based rendered images for training. Li *et al*. [26] proposes a cascaded network to iteratively refines the estimation of transmission and reflection images.

Reflection-contaminated image formation models. According to the ways of obtaining reflective amplitude coefficient maps, existing image formation models can be categorized into empirical model [20, 21, 7, 26], spatially-uniform model [23], and data-driven model [7]. Unfortunately, all of them assume an ideal image formation model and do not take the absorption effect into account. Kim *et al*. [25] adopts a physically-based rendering method to render the training data. However, they do not simulate the variation of absorption effect (*i.e*., they only consider the attenuation effect for $\mathbf{R}$ instead of $\mathbf{T}$). In this paper, we revisit the formation model of the contaminated-image formation model by taking the absorption effect into account and develop a single image reflection removal solution based on it.

3. Modeling Absorption Effect

When taking the absorption effect into account, the refractive and reflective amplitude coefficient maps for a plate double-surface glass can be formulated as [27, 28]

$$\mathbf{\Omega} = \frac{(1-\mathbf{P})^2(1-\mathbf{A})}{1-\mathbf{P}^2(1-\mathbf{A})^2},\ \mathbf{\Phi} = \mathbf{P} + \frac{\mathbf{P}(1-\mathbf{P})^2(1-\mathbf{A})^2}{1-\mathbf{P}^2(1-\mathbf{A})^2}. \quad (2)$$

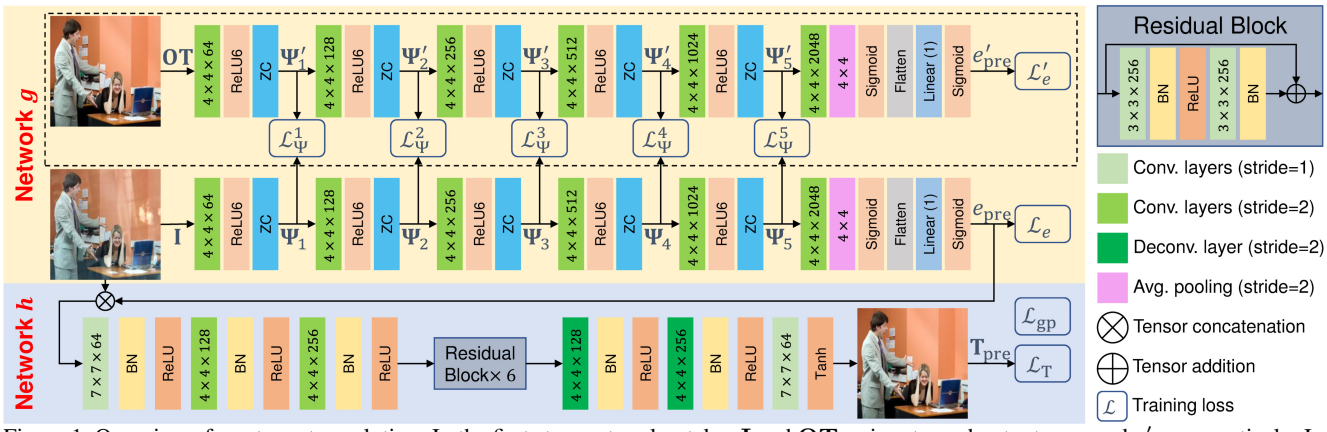


Figure 1. Overview of our two-step solution. In the first step, network g takes $\mathbf{I}$ and $\mathbf{OT}$ as inputs, and outputs e_{pre} and e'_{pre}, respectively. In the second step, network h takes the input which is the concatenation of $\mathbf{I}$ and spatially-replicated e_{pre}, and outputs $\mathbf{T}_{\text{pre}}$. $\mathbf{\Psi}_i$ and $\mathbf{\Psi}'_i$ are outputs of different hidden layers of network g. Note that two branches of network g share the same weights during training, and the one with dotted box is not activated in testing. 'ZC' and 'BN' represent operations of zero-center and batch normalization, respectively.

$\mathbf{A}$ is defined as the *absorption effect* that describes the attenuation of light when travelling through the glass. $\mathbf{P}$ is the reflectivity at the air/glass interface. $\mathbf{A}$ and $\mathbf{P}$ can be formulated according to Beer–Lambert law and Fresnel's equations, respectively

$$\begin{aligned}\mathbf{A} &= 1 - \exp(\frac{-kL}{\cos\mathbf{\Theta}_t}),\\ \mathbf{P} &= \frac{1}{2}((\frac{\cos\mathbf{\Theta} - \kappa\cos\mathbf{\Theta}_t}{\cos\mathbf{\Theta} + \kappa\cos\mathbf{\Theta}_t})^2 + (\frac{\cos\mathbf{\Theta}_t - \kappa\cos\mathbf{\Theta}}{\cos\mathbf{\Theta}_t + \kappa\cos\mathbf{\Theta}})^2),\end{aligned} \tag{3}$$

where k is the attenuation coefficient that represents the color of glass, L is the distance that light travels through the glass which is determined by the thickness of glass, κ is the refractive index, $\mathbf{\Theta}$ is the map of incidence angle regarding the glass, and $\mathbf{\Theta}_t = \arcsin(\frac{1}{\kappa}\sin\mathbf{\Theta})$ according to Snell's law. For a common window glass, k ranges from 4 m^{-1} to 32 m^{-1} [28] and $\kappa = 1.474$ [8]. As introduced in Section 2, existing image formation models assume $\mathbf{A} = 0$, resulting in $\mathbf{\Omega} + \mathbf{\Phi} = 1$. This assumption *ignores* the absorption effect. For a reflection removal method, the unreliable assumption of $\mathbf{\Omega}$ and $\mathbf{\Phi}$ can degrade its performance, especially for the accuracy of overall intensity similarity of $\mathbf{T}$ (defined in Equation (17)). In contrast, we consider a more general formation model with $\mathbf{A} \neq 0$.

Directly solving for $\mathbf{A}$ from $\mathbf{I}$ is a non-trivial task because of the matrix form of $\mathbf{A}$ and the real-world relationship between $\mathbf{A}$ and $\mathbf{I}$ (described by Equation (1)). In this paper, we assume that reflection occurs over a piece of plate glass [29], which is homogeneous, isotropic [30], and fills the whole field of view of the camera (FoV). Based on this assumption, we show that a scalar e, which is the average of $\mathbf{\Omega}$, can be used to numerically approximate the absorption effect in Section 5.1. With such an approximation, $\mathbf{\Omega}$ can be factorized into e multiplying a matrix $\mathbf{O}$,

$$\mathbf{\Omega} = e\mathbf{O}, \tag{4}$$

where all elements in $\mathbf{O}$ tend to be one.

4. Proposed Method

According to the analysis in Section 3, we reformulate the image formation in Equation (1) as

$$\mathbf{I} = e\mathbf{OT} + \mathbf{\Phi R}. \tag{5}$$

As the content-free variable e fluctuates in real-world (shown in Figure 3 (f)), we argue that an accurate estimation of e can benefit in solving the problem of single image reflection removal. We then propose our two-step solution

$$\begin{aligned}&g : \mathbf{I} \rightarrow e,\\ &h_1 : (\mathbf{I}, e) \rightarrow \mathbf{OT}, \quad h_2 : \mathbf{OT} \rightarrow \mathbf{T}.\end{aligned} \tag{6}$$

The first step includes a network g which estimates e from $\mathbf{I}$. The second step consists of network h_1 and h_2, where h_1 recovers $\mathbf{OT}$ from $\mathbf{I}$ with the help of e, and h_2 recovers $\mathbf{T}$ from $\mathbf{OT}$. Figure 1 illustrates the overview of the framework. Note that networks h_1 and h_2 are combined as network h.

4.1. Estimating Absorption Effect e

Since CNNs generally extract features based on the image content, using network g to estimate content-free variable e from $\mathbf{I}$ is not a trivial task. To this end, network g is additionally fed by a paired $\mathbf{OT}$ to focus on content-free features during training (dotted box in Figure 1). The idea is that although $\mathbf{OT}$ and $\mathbf{I}$ have similar image content of transmission, they have quite different e, *i.e.*, $g(\mathbf{I})$ should be e while $g(\mathbf{OT})$ should be 1. Hence, features of content-free variable e are expected to be learned through a supervised manner. Although such a simple scheme benefits in isolating image content features of transmission, it can lead network g to focus on the image content of reflection (*i.e.*, $\mathbf{\Phi R}$). Because discriminative features between $\mathbf{I}$ and $\mathbf{OT}$ include both e and $\mathbf{\Phi R}$. In the following, we show how to mitigate the influence from image content of reflection and focus on e via the design of network g and the loss function $\mathcal{L}_\Psi$.

Figure 2 shows two tuples of $(\mathbf{I}, \mathbf{T}, \mathbf{\Phi R})$ of real data. As can be observed: 1) Strong reflection dominates in sparse

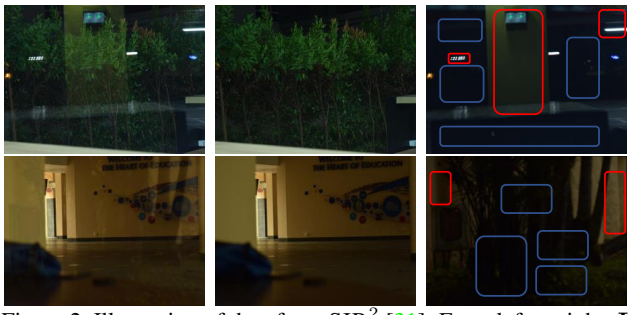


Figure 2. Illustration of data from SIR^2 [31]. From left to right: $\mathbf{I}$, $\mathbf{T}$, and $\mathbf{\Phi R}$. Strong/weak reflection are marked by red/blue boxes.

regions, and the intensity of these regions in $\mathbf{I}$ is much larger than that in $\mathbf{T}$ (red boxes). 2) The remaining weak reflection continuously spreads, and the intensity difference in these regions between $\mathbf{I}$ and $\mathbf{T}$ tends to have close and small values (blue boxes). Based on these observations, the network g is designed as: 1) ReLU6 [32] instead of ReLU [33] is used as the activation function to cut off large values produced by strong reflection. 2) Zero-center (ZC) operation [34] is used to subtract the uniform impact caused by the weak reflection. We also design network g (and h) as a bias-free convolution neural network [35] to better propagate e from the input $\mathbf{I}$.

We then further show how these designs help achieve our objective. Specifically, with the approximation of $\text{ReLU6}(a\mathbf{x}+\mathbf{y}) \approx a\text{ReLU6}(\mathbf{x}) + \text{ReLU6}(\mathbf{y})$, where a is a scalar in the scope of $[0.7, 1]$[3], $\mathbf{x}$ and $\mathbf{y}$ are two tensors, our designs bring the following approximation

$$\mathbf{\Psi}_i \approx e\mathbf{\Psi}'_i + \mathbf{\Delta}_i, \qquad \forall i = 1, 2, 3, 4, 5, \tag{7}$$

where $\mathbf{\Psi}$ is the output of hidden layers as shown in Figure 1, $\mathbf{\Delta}_i$ is defined as

$$\mathbf{\Delta}_i = \begin{cases} \mathbf{\Phi R}, & i = 0 \\ \text{ZC}(\text{ReLU6}(\mathbf{w} * \mathbf{\Delta}_{i-1})), & i = 1, 2, 3, 4, 5, \end{cases} \tag{8}$$

where $\mathbf{w}$ is the learned convolution kernel. ReLU6 [32] cuts off large values (against sparse strong reflection with large intensity values) and ZC [34] subtracts average values (against dense weak reflection with uniform intensity values), thus $\mathbf{\Delta}_i$ tends to be zero with increasing i. As $\mathbf{\Delta}_i$ carries information from $\mathbf{\Phi R}$ as shown in Equation (8), the image content of reflection is mitigated with $\mathbf{\Delta}_i$ approaching zero in deep layers. Besides, Equation (7) suggests that e can be successfully propagated to deep layers.

To better enforce $\mathbf{\Delta}_i$ to be zero during training, we further constrain $\mathbf{\Psi}$ by loss function $\mathcal{L}^i_\Psi$

$$\mathcal{L}_\Psi = \sum_{i=1}^{5} \lambda_i \mathcal{L}^i_\Psi = \sum_{i=1}^{5} \lambda_i \|\text{BCE}(\mathbf{\Psi}_i \oslash \mathbf{\Psi}'_i, e_{\text{gt}})\|, \tag{9}$$

where $\oslash$ is the element-wise division operation, $\text{BCE}(\cdot,\cdot)$ represents the binary cross-entropy loss function [36] applied to each element of matrix $\mathbf{\Psi}_i \oslash \mathbf{\Psi}'_i$ and scalar e_{gt}[4], and λ_i is the weight. We set weight $\{\lambda_i\}$ with increasing numbers $\{0.2, 0.8, 2, 3, 4\}$ as $\mathbf{\Delta}_i$ tends to be zero with increasing i.

In summary, the two-branch training strategy facilitates isolating image content features from transmission. The proposed architecture of network g and the loss function $\mathcal{L}_\Psi$ help mitigate the influence from image content of reflection. These designs help propagate e across the layers of network g and contribute to the accurate estimation of e.

4.2. Recovering Transmission T

Network h_1 is optimized to make equation $h_1(\mathbf{I}, e) = \mathbf{OT}$ hold. Since variable e distributes in a continuous space (*i.e.*, $e \in \mathbb{E}$), we further constrain

$$\forall e \in \mathbb{E}, \quad h_1(\mathbf{I}, e) = \mathbf{OT}, \quad \text{s.t.} \quad \mathbf{I} = f_{\text{I}}(e) = e\mathbf{OT} + \mathbf{\Phi R} \tag{10}$$

As this constraint ensures a range of $e \in \mathbb{E}$ instead of a single value e_0 to satisfy $h_1(\mathbf{I}, e) = \mathbf{OT}$, it is expected to help generalize priors learned from limited training data to real data with diverse absorption effects. In the following, we show that applying the constraint in Equation (10) can be achieved by guaranteeing function '$s(e) = f_{\text{I}}(e), e \in \mathbb{E}$' to be the unique solution of the initial value problem [11]

$$\begin{cases} h_1(s(e), e) = \frac{\text{d}s}{\text{d}e}, \\ \mathbf{I}_0 = s(e_0), \end{cases} \tag{11}$$

where $(\mathbf{I}_0, e_0)$ is from training data.

If network h_1 is trained such that $f_{\text{I}}(e)$ is the unique solution to the initial value problem in Equation (11), we have $\frac{\text{d}s}{\text{d}e} = \mathbf{OT}$. The constraint in Equation (10) is ensured to be held. Otherwise, the derivative of function $s(e)$ is not guaranteed to be unique and $h_1(s(e), e)$ or $h_1(\mathbf{I}, e)$ is not necessarily equal to $\mathbf{OT}$. Thus, the constraint in Equation (10) is not ensured to be satisfied. Obviously, $f_{\text{I}}(e)$ is one of the solutions to the initial value problem in Equation (11) since we train h_1 with data that satisfy $f_{\text{I}}(e)$. Therefore, the key to ensure the constraint in Equation (10) is to ensure the uniqueness of the solution $f_{\text{I}}(e)$.

According to Cauchy-Lipschitz theorem [11] (also called Picard-Lindelöf theorem or Picard existence theorem), only if h_1 satisfies the constraint of Lipschitz condition [37], the uniqueness of solution can be achieved. Suppose $\mathbf{I} \in \mathbb{U}$, the constraint of Lipschitz condition can be expressed as [37]

$$|h_1(\mathbf{I}_1, e) - h_1(\mathbf{I}_2, e)| \leq M|\mathbf{I}_1 - \mathbf{I}_2|, \quad \forall (\mathbf{I}_1, e), (\mathbf{I}_2, e) \in \mathbb{U} \times \mathbb{E}, \tag{12}$$

where M is referred to as a Lipschitz constant. Function h_1 is called as an M-Lipschitz function if it satisfies the constraint in Equation (12). Similar to [38], we set M to 1 in this paper. Recent works realize the constraint of Lipschitz condition for a deep neural network model using gradient penalty [39, 38]. Gulrajani *et al.* [38] prove that a differentiable function is 1-Lipschitz if and only if it has gradients with the norm at most 1 everywhere. They directly constrain the gradient norm of the output of a deep neural network for its input and enforce a soft version of the constraint with a

[3] The scope is determined based on the simulation results in Figure 3 (d).

[4] We have tried to use L1/L2 loss function (*e.g.*, $\|\mathbf{\Psi}_i - e\mathbf{\Psi}'_i\|$), however, the network tends to produce all-zeros feature maps.

penalty on the gradient norm for random samples. Details about the prove and implementation can be found in [38].

We adopt a similar strategy to that in [38] to constrain h_2 as a 1-Lipschitz differentiable function, *i.e.*, a gradient penalty loss function which penalizes the gradient norm to be 1 for random reflection-contaminated image $\hat{\mathbf{I}}$

$$\mathcal{L}_{\text{Con}} = (\|\nabla_{\hat{\mathbf{I}}} h_1(\hat{\mathbf{I}}, \hat{e})\| - 1)^2, \qquad \forall \hat{\mathbf{I}} \in \mathbb{U}, \forall \hat{e} \in \mathbb{E}. \tag{13}$$

$\mathbb{U}$ represents a subspace where reflection-contaminated images distribute. We construct $\hat{\mathbf{I}} = \mathbf{I} + \epsilon_1 \mathbf{OT}$, where $\epsilon_1 \sim \text{U}[-0.1, 0.1]$ and U represents the uniform distribution. That is, we regard $\hat{\mathbf{I}} = (\epsilon_1 + e)\mathbf{OT} + \mathbf{\Phi R}$ as a reflection-contaminated image with absorption effect $\epsilon_1 + e$. Therefore, we have $\hat{\mathbf{I}} \in \mathbb{U}$. We construct $\mathbb{E} = \{\hat{e} | \hat{e} = \epsilon_2 e_{\text{gt}} + (1 - \epsilon_2 e_{\text{pre}})\}$, where $\epsilon_2 \sim \text{U}[0, 1]$. This construction is similar to that in [38], *i.e.*, uniformly sampling along straight lines between data sampled from distributions of ground truth and estimation.

Combining h_1 and h_2. We combine h_1 and h_2 as network h and apply the gradient penalty loss function to h to approximately applying that to h_1. Such an approximation is reasonable because once the gradient of h is penalized, that of h_1 (a component of h) is expected to be penalized.

The gradient penalty loss function guarantees the unique solution to the initial value problem of Equation (11) and ensures the constraint in Equation (10), which helps generalize priors learned from limited training data to real data that are with diverse absorption effects and a variety of scenarios.

4.3. Loss Functions

We use an alternating optimization scheme to train g and h iteratively. We update g once after updating h five times for the better consideration of absorption effect estimation. Loss functions of g and h are as follows

$$\begin{aligned} \mathcal{L}_g &= \mathcal{L}_{\text{T}} + \lambda_e(\mathcal{L}_e + \mathcal{L}'_e) + \sum_{i=1}^{5} \lambda_i \mathcal{L}^i_{\Psi}, \\ \mathcal{L}_h &= \mathcal{L}_{\text{T}} + \lambda_{\text{gp}} \mathcal{L}_{\text{gp}}. \end{aligned} \tag{14}$$

$\mathcal{L}_{\text{T}}$ is the reconstruction loss function and $\mathcal{L}_e$ is the binary cross-entropy loss function [36]

$$\mathcal{L}_{\text{T}} = \mathcal{D}(\mathbf{T}_{\text{pre}}, \mathbf{T}_{\text{gt}}), \quad \mathcal{L}_e = \text{BCE}(e_{\text{pre}}, e_{\text{gt}}), \quad \mathcal{L}'_e = \text{BCE}(e'_{\text{pre}}, e_{\text{gt}}), \tag{15}$$

where $\mathbf{T}_{\text{pre}} = h(\mathbf{I}, e_{\text{pre}}), e_{\text{pre}} = g(\mathbf{I})$, $\mathcal{D}$ is a pre-defined metric that measures the similarity of images $\mathbf{T}_{\text{pre}}$ and $\mathbf{T}_{\text{gt}}$

$$\mathcal{D}(\mathbf{T}_{\text{pre}}, \mathbf{T}_{\text{gt}}) = \ell_{\text{per}} - \lambda_{\text{psnr}} \ell_{\text{PSNR}} - \ell_{\text{SSIM}} - \ell_{\text{SI}}, \tag{16}$$

where λ_{psnr} is set as $\frac{1}{40}$ to balance the values of ℓ_{SSIM} and ℓ_{SI}. Pre-defined metric $\mathcal{D}$ jointly considers commonly adopted metrics in reflection removal. We use a similar implementation of perceptual loss ℓ_{per} as that in [40, 21], which is from the pre-trained VGG-16 [41] models trained on the ImageNet dataset [42]. The Peak Signal-to-Noise Ratio (PSNR) [43] and the Structural SIMilarity (SSIM) [44] are two widely used metrics to measure differences between images. The SSIM is defined with default parameters as those in [44]. The intensity-variant structural SImilarity (SI) [45] focuses only on the structural similarity [31].

Implementation details. Figure 1 shows the network architectures of g and h. We adopt a similar network architecture to that in [46, 47] for h due to its excellent image generation ability. The batch sizes are set to 16. We set $\lambda_e = 0.5, \lambda_{\text{gp}} = 10$ for all experiments. Both of neural networks are trained using Adam solver [48] with $\beta_1 = 0.5$ and $\beta_2 = 0.999$. We set the learning rates for g and h to 0.0001 for the first 100 epochs and decay to 0.00005 for the next 100 epochs. Except where explicitly stated, all our experiments in this paper use the same setup described above.

5. Experiments

Testing data. We use four real datasets for evaluation. As the absorption effect is relevant to factors of glass thickness, orientation, and color (introduced in Section 3), we highlight these factors in the testing datasets[5]. SIR^2 [31] contains 454 testing samples and 120 of them are captured with three different glass thicknesses. We thereby take this subset of SIR^2 [31] as SIR^2-THICK [31] for evaluation. ZC20-ORIEN [50] contains 160 samples captured with five different glass orientations. LY20-DATA [26] contains 220 samples and part of them are captured with two different glass thicknesses and two different glass orientations. ZN18-DATA [21] contains 109 testing samples and part of them are captured with two different glass orientations.

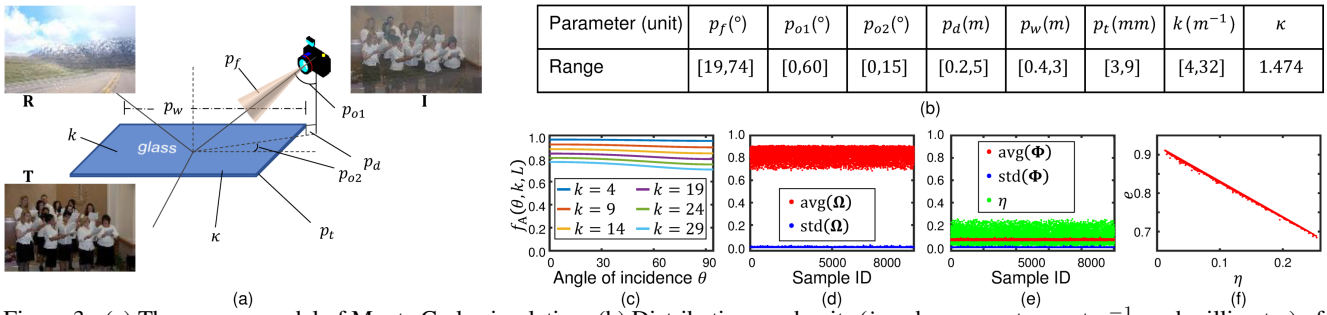
Comparison methods. Seven state-of-the-art single image reflection removal methods, including an optimization-based method, *i.e.*, YM19 [17], and six deep learning-based methods, *i.e.*, ZN18 [21], WS19 [23], WY19 [4], WT19 [7], KH20 [25], LY20 [26], are compared with our method.

Quantitative metrics. We use SSIM [44] and PSNR [43] as error metrics. Besides, we introduce the metric of IS (average of Intensity Similarity index) to evaluate the accuracy of overall intensity similarity. Because one of the key differences between our method and others is the estimation of e, which helps the recovery of $\mathbf{T}$'s overall intensity (according to Equation (5)). The Intensity Similarity index focuses on the intensity similarity between two images $\mathbf{x}$ and $\mathbf{y}$, which is defined as a factor of the SSIM index [44]

$$\text{IS}(\mathbf{x}, \mathbf{y}) = \frac{2\mu_x \mu_y + c}{\mu_x^2 + \mu_y^2 + c}, \tag{17}$$

where μ_x and μ_y are the averages of $\mathbf{x}$ and $\mathbf{y}$, c is a constant which is set as the default value as that in [44].

[5] Since there is no existing dataset with a variation of glass color, we synthesize BLD-COLOR dataset using the physically-based rendering engine Cycles [49]. The proposed method achieves the best performance as compared with other state-of-the-art methods. Details of data synthesizing and results can be found in our supplementary material.



Parameter (unit)	$p_f(°)$	$p_{o1}(°)$	$p_{o2}(°)$	$p_d(m)$	$p_w(m)$	$p_t(mm)$	$k(m^{-1})$	κ
Range	[19,74]	[0,60]	[0,15]	[0.2,5]	[0.4,3]	[3,9]	[4,32]	1.474

Figure 3. (a) The camera model of Monte Carlo simulation. (b) Distributions and units (*i.e.*, degree, meter, meter^{-1}, and millimeter) of input variables for Monte Carlo simulation. (c) Curves of function $a = f_A(\theta, k, L)$, where a is the element of $\mathbf{A}$, θ is the corresponding angle of incidence, and $L = 9\ mm$. (d) and (e) Distributions of 10000 randomly generated η, mean, standard deviations of $\mathbf{\Omega}$ and $\mathbf{\Phi}$. (f) The result of correlation analysis between absorption proxy η and e (or avg($\mathbf{\Omega}$)).

5.1. Synthesizing Training Data

We synthesize our training data based on Equation (1), with 18224 $\mathbf{T}$ from Places365 [51][6], 5552 $\mathbf{R}$ from [22], and 18224 $\{\mathbf{\Omega}, \mathbf{\Phi}\}$ generated according to Monte Carlo simulation [10]. The camera model is displayed in Figure 3 (a). A recent work [50] shows that $\mathbf{\Omega}$ and $\mathbf{\Phi}$ are jointly determined by the horizontal FoV of camera p_f (assuming an input size with ratio $2:3$), refractive index κ, and the orientation of glass (represented by $(\sin p_{o1} \sin p_{o2}, \sin p_{o1} \cos p_{o2}, \cos p_{o1})$). Different from [50] that assumes $\mathbf{A} = 0$, we take the simulation of $\mathbf{A}$ into account. Specifically, we further consider factors of glass width p_w, the distance between camera and glass p_d, glass thickness p_t, and glass attenuation coefficient k. These inputs of the Monte Carlo simulation [10] are generated based on uniform distributions[7]. Figure 3 (b) shows ranges of these variables used in this paper. They are set according to observations in our daily life, *e.g.*, the horizontal FoV is set according to a digital camera (*i.e.*, Canon EOS 5D Mark III), when photographing a glass, the photographer may not stand too close ($< 0.2m$) or too far ($> 5m$).

Numerical approximation. To show that the absorption effect can be numerically represented by the average of $\mathbf{\Omega}$, we additionally generate 10000 $\{\mathbf{\Omega}, \mathbf{\Phi}, \mathbf{A}\}$ for analysis. We observe that all elements of $\mathbf{A}$ tend to be uniform regarding different angles of incidence according to its formulation in Equation (3). This observation can be validated by the curve of $a = f_A(\theta, k, L)$ regarding the variable of θ as shown in Figure 3 (c), where a is an element of $\mathbf{A}$ and θ is the corresponding angle of incidence, k and L are fixed for a given $\mathbf{I}$ (assuming a piece of plate glass [29])[8]. The absorption effect $\mathbf{A}$ hence can be approximately represented by its average η and we define η as the absorption proxy, *i.e.*, $\eta \stackrel{\text{def}}{=} \text{avg}(\mathbf{A})$. We investigate the relationship between η and $\{\mathbf{\Omega}, \mathbf{\Phi}\}$ by plotting the 10000 simulation results of $\{\eta, \text{avg}(\mathbf{\Omega}), \text{std}(\mathbf{\Omega}), \text{avg}(\mathbf{\Phi}), \text{std}(\mathbf{\Phi})\}$[9]. Figure 3 (d) and (e) plot the simulation results. The broad distribution of η indicates the fluctuation of absorption effect in the real-world. In contrast, the narrow distributions of avg($\mathbf{\Phi}$) and std($\mathbf{\Phi}$) indicate the weaker relevance between η and $\mathbf{\Phi}$. All elements of $\mathbf{\Omega}$ tend to be uniform because of the narrow distribution and small values of std($\mathbf{\Omega}$). We thereby define $e \stackrel{\text{def}}{=} \text{avg}(\mathbf{\Omega})$ and focus on the relationship between e and η. Figure 3 (f) displays the result of their correlation analysis. Motivated by the one-to-one mapping relationship of e and η, we use e as a numerical approximation to absorption effect.

Table 1. Average SSIM [44] differences ($\times 10^{-3}$) between re- and pre-trained models. Positive numbers represent improvement is achieved by the re-trained models. '/' represents the pre-trained model of an indicated method is trained using an indicated dataset.

Datasets	ZN18	WS19	WT19	WY19	KH20	LY20
SIR^2-THICK	11.0	5.20	63.0	13.8	9.08	13.8
ZC20-ORIEN	1.97	28.2	24.7	7.20	16.9	7.02
LY20-DATA	6.13	2.77	65.4	58.4	5.60	/
SIR^2	22.9	-12.4	58.7	-6.02	4.26	11.8
ZN18-DATA	/	-13.3	38.7	/	20.0	10.1

5.2. Validation for Image Formation Model

We re-train comparison learning-based methods using the same training data as our method, and compare their performance with their pre-trained models on each testing dataset. As training data used by the re-trained models are generated based on our image formation model, while those used by the pre-trained models are generated by other image formation models (*i.e.*, WT19 [7] adopts the model in

[6]These images are from four scenes, OFFICE, PARKING_GARAGE-INDOOR, RESTAURANT_PATIO and STREET, as glass reflection is more likely to occur in these scenarios. Note that 1776 gray images are excluded.

[7]Normal distributions provide similar observations in following analysis.

[8]As L and k contribute equally to a according to Equation (3), how a varies depending L can be found by how a varies depending k.

[9]'avg' and 'std' output the mean and the standard deviation of a matrix.

Table 2. Comparisons of quantitative results in terms of SSIM, IS, and PSNR on SIR^2-THICK [31], ZC20-ORIEN [50], LY20-DATA [26], SIR^2 [31], and ZN18-DATA [21] for reflection removal. We mark the best and second-best performing methods in red and blue respectively.

Dataset(size)	Metric	Ours	One-branch	*w/o*-Con	ZN18[21]	YM19[17]	WS19[23]	WT19[7]	WY19[4]	KH20[25]	LY20[26]
SIR^2-THICK (120)[31]	SSIM	**0.8965**	0.8877	**0.8940**	0.8494	0.8598	0.8751	0.8687	0.8864	0.8869	0.8641
	IS	**0.9773**	0.9711	**0.9752**	0.9275	0.9520	0.9532	0.9630	0.9646	0.9696	0.9598
	PSNR	**24.05**	22.85	**23.59**	18.91	21.85	20.63	22.03	23.00	23.46	22.02
ZC20-ORIEN (160)[50]	SSIM	**0.8790**	0.8638	0.8663	0.8673	0.8660	0.8244	0.8644	0.8616	**0.8757**	0.8743
	IS	**0.9722**	0.9598	**0.9720**	0.9670	0.9660	0.9142	0.9594	0.9646	0.9712	0.9681
	PSNR	**23.93**	20.42	**23.69**	22.61	23.68	19.26	21.40	23.84	23.48	23.56
LY20-DATA (220)[26]	SSIM	**0.8732**	0.8568	**0.8673**	0.8354	0.8531	0.8420	0.8244	0.8254	0.8480	0.8568
	IS	**0.9552**	0.9428	**0.9503**	0.9410	0.9458	0.9401	0.9368	0.9499	0.9490	0.9414
	PSNR	**23.97**	22.23	**23.72**	23.13	21.93	21.35	20.73	22.41	22.85	23.61
SIR^2 (454)[31]	SSIM	**0.9003**	0.8906	0.8934	0.8703	0.8680	**0.8961**	0.8746	0.8906	0.8916	0.8945
	IS	**0.9756**	0.9688	**0.9733**	0.9267	0.9503	0.9500	0.9594	0.9593	0.9666	0.9589
	PSNR	**24.34**	23.06	**23.90**	19.24	22.20	20.93	22.05	23.35	23.64	22.76
ZN18-DATA (109)[21]	SSIM	**0.7783**	0.7653	0.7669	0.7671	0.7395	0.7663	0.6844	0.7668	0.7507	**0.7691**
	IS	**0.8970**	0.8846	**0.8966**	0.8843	0.8703	0.8956	0.8678	0.8727	0.8808	0.8773
	PSNR	**19.63**	18.32	**19.60**	18.44	18.69	19.04	17.01	19.22	18.84	19.05

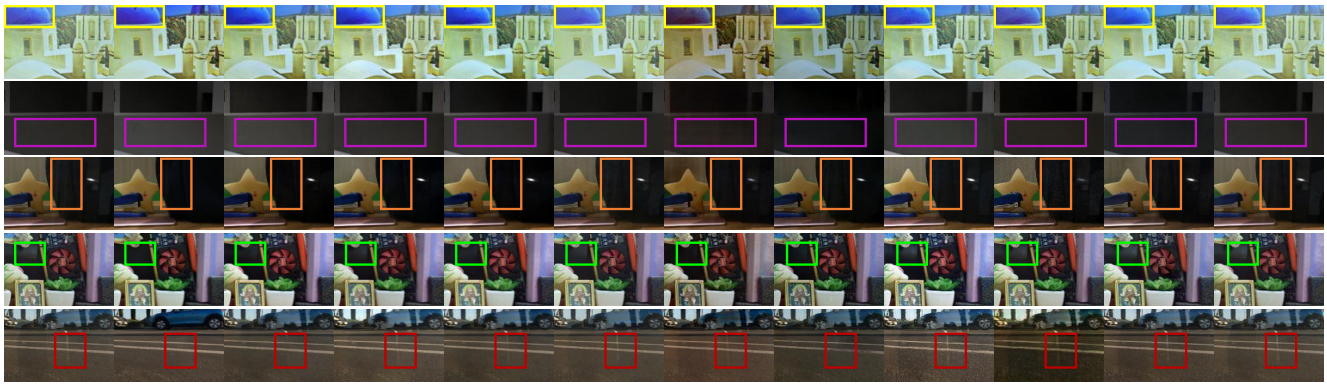


Figure 4. From top to bottom: visual quality comparison of reflection removal for samples from SIR^2-THICK [31], ZC20-ORIEN [50], LY20-DATA [26], SIR^2 [31], and ZN18-DATA [21]. Color boxes highlight noticeable differences. Zoom in for better details.

Equation (1) with $\mathbf{\Omega} + \mathbf{\Phi} = 1$), such performance comparison is to evaluate the effectiveness of our image formation model. Table 1 shows the differences of averaged SSIM [44] between re- and pre-trained models. As can be observed, the re-trained models outperform pre-trained ones with 24 out of 27 cases[10]. Note that the absorption effects of training data in [25] tend to be uniform due to their setting of a fixed glass thickness and color. The performance advantage of the re-trained models demonstrates the effectiveness of our image formation model for single image reflection removal.

5.3. Overall Performance

For a fair comparison with state-of-the-art learning-based methods, we report the better numbers from their re- and pre-trained models for each testing dataset in Table 2. As can be observed, our method achieves the best performance regarding all metrics for all testing datasets. Figure 4 shows the visual quality comparison.[11] As can be observed, YM19 [17] produces over-smooth results, ZN18 [21] recovers $\mathbf{T}$ with color distortion, WS19 [23] predicts dark results, WT19 [7], WY19 [4], KH20 [25], and LY20 [26] fail to remove reflection (third to fifth rows) and produce inaccurate overall intensity similarity (first and second rows). In contrast, our method not only successfully removes reflection artifacts, but also faithfully restores the intensity distortion caused by the absorption effect. The state-of-the-art performance of our method demonstrates the effectiveness of our solution that explicitly considers the absorption effect.

5.4. Ablation Studies

We investigate the effectiveness of each part of our solution[12]. Specifically, we develop two methods which are simplified versions of the proposed method: 1) 'One-branch' method, network g only takes $\mathbf{I}$ as the input and is opti-

[10] As the pre-trained models of ZN18 [21], WY19 [7] use 100 samples from ZN18-DATA [21] for training, the pre-trained model of LY20 [26] uses 200 samples from LY20-DATA [26] for training, we do not compare their results and report performance from re-trained models for these cases.

[11] More examples of visual quality comparison can be found in our supplementary material.

[12] The ablation study of a single step network, *i.e.*, directly regressing $\mathbf{I}$ to $\mathbf{T}$ using network h, can be found in our supplementary material.

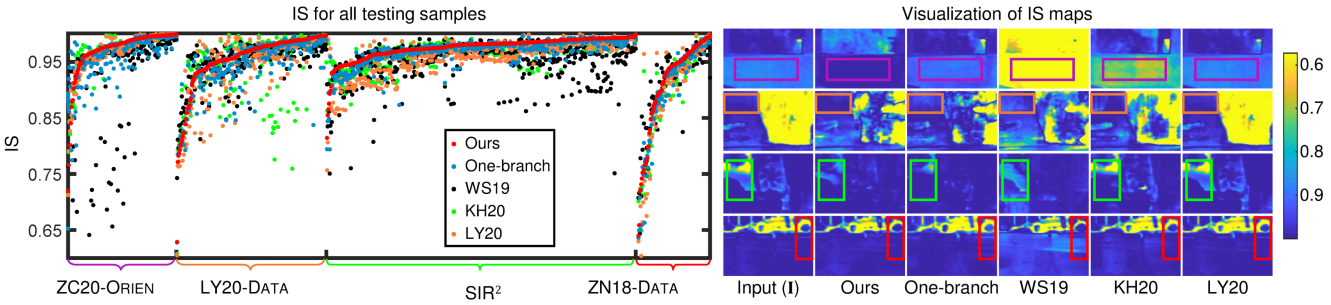


Figure 5. Left: IS results of all testing samples in datasets of ZC20-ORIEN [50], LY20-DATA [26], SIR^2 [31], and ZN18-DATA [21]. Samples of each dataset are ordered with the increasing IS obtained by our method. Right: the visualization of IS index maps between the predicted $\mathbf{T}_{\text{pre}}$ (or the input $\mathbf{I}$) and the ground truth of $\mathbf{T}_{\text{gt}}$. The corresponding images of these IS index maps can be found in Figure 4. Color boxes highlight noticeable differences. Zoom in for better details.

mized without loss function $\mathcal{L}_{\Psi}$ during training. 2) '*w/o*-Con' method, network h is optimized without the constraint in Equation (10) or loss function $\mathcal{L}_{\text{gp}}$. Remaining parts and the training setups are kept unchanged as our method.

Two-branch training to propagate accurate e. The performance comparison between the proposed method and the one-branch method can help validate the effectiveness of two-branch training. As shown in Table 2 and Figure 4, the proposed method outperforms the one-branch method for all testing datasets. Such performance advantage, especially for the metric IS, indicates that two-branch training helps estimate more accurate e so that the overall intensity of $\mathbf{T}$ can be more accurately recovered. We further illustrate the IS distribution for each testing sample in Figure 5 (left). As can be observed, our method has a superior performance advantage for all testing datasets over the one-branch method and other state-of-the-art methods. The visual quality comparison in Figure 5 (right) intuitively shows that the recovered $\mathbf{T}$ from our method is more accurate regarding the overall intensity similarity. Our method contributes to single image reflection removal through the accurate recovery of overall intensity.

Satisfying constraint in Equation (10) to facilitate model generalization capacity. According to the analysis in Section 4.2, satisfying Lipschitz condition of network h helps constrain Equation (10) and facilitates model generalization capacity. Therefore, we evaluate by comparing the performance of the proposed method and *w/o*-Con method. As shown in Table 2 and Figure 4, the proposed method achieves slightly better results as compared with *w/o*-Con method for all testing datasets. To better validate the advantage of our method, we train the proposed method and *w/o*-Con method using a limited size of training data and compare their performance. Because a limited size of training data helps highlight the generalization capacity of a data-driven method. We train these two methods with $\frac{1}{10}$ and $\frac{1}{20}$ of original training data, represented as 'Medium' and 'Small', respectively. We also compare with results based on the original training data (represented as 'Large'). As this constraint is applied based on e, we use IS as the metric. Figure 5 illustrates the performance changing with different training data sizes. As can be observed, the performance advantage of our method is more significant with smaller training data sizes across all testing datasets. This observation verifies that our method is good at learning knowledge from a limited size of training data, which indicates its better generalization capacity compared with *w/o*-Con method.

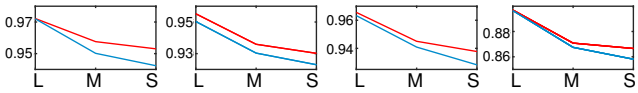


Figure 6. The IS performance comparison of our method (red curves) and *w/o*-Con methods (blue curves) using different sizes of training data (Large, Medium, Small). From left to right: results on testing datasets of ZC20-ORIEN [50], LY20-DATA [26], SIR^2 [31], and ZN18-DATA [21].

6. Conclusion

This paper revisits the formation model of a reflection-contaminated image by taking the absorption effect into account and proposes a two-step solution for single image reflection removal. The state-of-the-art performance achieved by our method verifies that accurately estimated absorption effect benefits the accurate recovery of transmission images.

Limitations. The estimated absorption effect e may be inaccurate as estimating e from $\mathbf{I}$ is an ill-posed problem. This estimation suffers from the ambiguity introduced by unknown scenes (*e.g.*, environment lighting of the scene) and the camera's image signal processor (ISP). This paper holds the assumption that the camera ISP is fixed for all scenes, which may be a strong requirement in real-world. Another limitation is our simplification of the absorption effect, which makes it difficult to directly verify the model by real data.

Acknowledgment

This research was carried out at the Rapid-Rich Object Search (ROSE) Lab, Nanyang Technological University, Singapore, and supported by National Natural Science Foundation of China under Grant No. 61872012, 62088102, and Beijing Academy of Artificial Intelligence (BAAI).

References

[1] Y. Li and M. S. Brown, "Exploiting reflection change for automatic reflection removal," in *Proc. of Internatoinal Conference on Computer Vision*, 2013. 1, 2

[2] N. Arvanitopoulos, R. Achanta, and S. Süsstrunk, "Single image reflection suppression," in *Proc. of Computer Vision and Pattern Recognition*, 2017. 1, 2

[3] R. Wan, B. Shi, A.-H. Tan, and A. C. Kot, "Sparsity based reflection removal using external patch search," in *International Conference on Multimedia and Expo (ICME)*, 2017. 1

[4] K. Wei, J. Yang, Y. Fu, D. Wipf, and H. Huang, "Single image reflection removal exploiting misaligned training data and network enhancements," in *Proc. of Computer Vision and Pattern Recognition*, 2019. 1, 2, 5, 7

[5] P. Wieschollek, O. Gallo, J. Gu, and J. Kautz, "Separating reflection and transmission images in the wild," in *Proc. of European Conference on Computer Vision*, 2018. 1, 2

[6] A. Punnappurath and M. S. Brown, "Reflection removal using a dual-pixel sensor," in *Proc. of Computer Vision and Pattern Recognition*, 2019. 1, 2

[7] Q. Wen, Y. Tan, J. Qin, W. Liu, G. Han, and S. He, "Single image reflection removal beyond linearity," in *Proc. of Computer Vision and Pattern Recognition*, 2019. 1, 2, 5, 6, 7

[8] N. Kong, Y.-W. Tai, and J. S. Shin, "A physically-based approach to reflection separation: from physical modeling to constrained optimization," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 2, pp. 209–221, 2013. 1, 2, 3

[9] C. Lei, X. Huang, M. Zhang, Q. Yan, W. Sun, and Q. Chen, "Polarized reflection removal with perfect alignment in the wild," in *Proc. of Computer Vision and Pattern Recognition*, 2020. 1

[10] C. Z. Mooney, *Monte carlo simulation*. Sage publications, 1997, vol. 116. 1, 6

[11] E. A. Coddington and N. Levinson, Theory of ordinary differential equations. *Tata McGraw-Hill Education*, 1955. 1, 4

[12] Y. Lyu, Z. Cui, S. Li, M. Pollefeys, and B. Shi, "Reflection separation using a pair of unpolarized and polarized images," in *Advances in Neural Information Processing Systems*, 2019. 2

[13] 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. 2

[14] Y. Shih, D. Krishnan, F. Durand, and W. T. Freeman, "Reflection removal using ghosting cues," in *Proc. of Computer Vision and Pattern Recognition*, 2015. 2

[15] R. Wan, B. Shi, T. A. Hwee, and A. C. Kot, "Depth of field guided reflection removal," in *Proc. of Internatoinal Conference on Image Processing*, 2016. 2

[16] R. Wan, B. Shi, L.-Y. Duan, A.-H. Tan, W. Gao, and A. C. Kot, "Region-aware reflection removal with unified content and gradient priors," *IEEE Transactions on Image Processing*, vol. 27, no. 6, pp. 2927–2941, 2018. 2

[17] Y. Yang, W. Ma, Y. Zheng, J.-F. Cai, and W. Xu, "Fast single image reflection suppression via convex optimization," in *Proc. of Computer Vision and Pattern Recognition*, 2019. 2, 5, 7

[18] J. Pan, S. Liu, D. Sun, J. Zhang, Y. Liu, J. Ren, Z. Li, J. Tang, H. Lu, Y.-W. Tai *et al.*, "Learning dual convolutional neural networks for low-level vision," in *Proc. of Computer Vision and Pattern Recognition*, 2018. 2

[19] X. Fu, J. Huang, D. Zeng, Y. Huang, X. Ding, and J. Paisley, "Removing rain from single images via a deep detail network," in *Proc. of Computer Vision and Pattern Recognition*, 2017. 2

[20] Q. Fan, J. Yang, G. Hua, B. Chen, and D. P. Wipf, "A generic deep architecture for single image reflection removal and image smoothing." in *Proc. of Internatoinal Conference on Computer Vision*, 2017. 2

[21] X. Zhang, R. Ng, and Q. Chen, "Single image reflection separation with perceptual losses," in *Proc. of Computer Vision and Pattern Recognition*, 2018. 2, 5, 7, 8

[22] R. Wan, B. Shi, L.-Y. Duan, A.-H. Tan, and A. C. Kot, "CRRN: Multi-scale guided concurrent reflection removal network," in *Proc. of Computer Vision and Pattern Recognition*, 2018. 2, 6

[23] R. Wan, B. Shi, H. Li, L.-Y. Duan, A.-H. Tan, and A. K. Chichung, "CoRRN: Cooperative reflection removal network," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019. 2, 5, 7

[24] J. Yang, D. Gong, L. Liu, and Q. Shi, "Seeing deeply and bidirectionally: A deep learning approach for single image reflection removal," in *Proc. of European Conference on Computer Vision*, 2018. 2

[25] S. Kim, Y. Huo, and S.-E. Yoon, "Single image reflection removal with physically-based training images," in *Proc. of Computer Vision and Pattern Recognition*, June 2020. 2, 5, 7

[26] C. Li, Y. Yang, K. He, S. Lin, and J. E. Hopcroft, "Single image reflection removal through cascaded refinement," in *Proc. of Computer Vision and Pattern Recognition*, June 2020. 2, 5, 7, 8

[27] G. H. Sigel Jr, "Optical absorption of glasses," in *Treatise on Materials Science & Technology*. Elsevier, 1977, vol. 12, pp. 5–89. 2

[28] R. E. Parkin, "A note on the extinction coefficient and absorptivity of glass," *Solar Energy*, vol. 114, pp. 196–197, 2015. 2, 3

[29] W. P. Spence and E. Kultermann, Construction materials, methods and techniques. *Cengage Learning*, 2016. 3, 6

[30] M. Born and E. Wolf, Principles of optics: electromagnetic theory of propagation, interference and diffraction of light. *Elsevier*, 2013. 3

[31] R. Wan, B. Shi, L.-Y. Duan, A.-H. Tan, and A. C. Kot, "Benchmarking single-image reflection removal algorithms," in *Proc. of Internatoinal Conference on Computer Vision*, 2017. 4, 5, 7, 8

[32] A. Krizhevsky and G. Hinton, "Convolutional deep belief networks on cifar-10," *Unpublished manuscript*, vol. 40, no. 7, pp. 1–9, 2010. 4

[33] V. Nair and G. E. Hinton, "Rectified linear units improve restricted boltzmann machines," in *Proc. of International Conference on Machine Learning*. 4

[34] P. T. Komiske, E. M. Metodiev, and M. D. Schwartz, "Deep learning in color: towards automated quark/gluon jet discrimination," *Journal of High Energy Physics*, vol. 2017, no. 1, p. 110, 2017. 4

[35] S. Mohan, Z. Kadkhodaie, E. P. Simoncelli, and C. Fernandez-Granda, "Robust and interpretable blind image denoising via bias-free convolutional neural networks," *International Conference on Learning Representations*, 2020. 4

[36] P.-T. De Boer, D. P. Kroese, S. Mannor, and R. Y. Rubinstein, "A tutorial on the cross-entropy method," *Annals of operations research*, vol. 134, no. 1, pp. 19–67, 2005. 4, 5

[37] V. I. Arnol'd, Ordinary Differential Equations. *Springer*, 1992. 4

[38] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of wasserstein GANs," in *Advances in Neural Information Processing Systems*, 2017. 4, 5

[39] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein GAN," *stat*, vol. 1050, p. 9, 2017. 4

[40] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. of European Conference on Computer Vision*, 2016. 5

[41] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014. 5

[42] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein *et al.*, "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision*, vol. 115, no. 3, pp. 211–252, 2015. 5

[43] Q. Huynh-Thu and M. Ghanbari, "Scope of validity of psnr in image/video quality assessment," *Electronics letters*, vol. 44, no. 13, pp. 800–801, 2008. 5

[44] Z. Wang, E. P. Simoncelli, and A. C. Bovik, "Multiscale structural similarity for image quality assessment," in *The Thrity-Seventh Asilomar Conference on Signals, Systems & Computers, 2003*, vol. 2, 2003, pp. 1398–1402. 5, 6, 7

[45] S.-H. Sun, S.-P. Fan, and Y.-C. F. Wang, "Exploiting image structural similarity for single image rain removal," in *Proc. of Internatoinal Conference on Image Processing*, 2014. 5

[46] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proc. of Internatoinal Conference on Computer Vision*, 2017. 5

[47] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. Choo, "StarGAN: Unified generative adversarial networks for multi-domain image-to-image translation," in *Proc. of Computer Vision and Pattern Recognition*, 2018. 5

[48] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014. 5

[49] Cycles, https://www.cycles-renderer.org/. 5

[50] Q. Zheng, J. Chen, Z. Lu, B. Shi, X. Jiang, K.-H. Yap, L.-Y. Duan, and A. C. Kot, "What does plate glass reveal about camera calibration," in *Proc. of Computer Vision and Pattern Recognition*, 2020. 5, 6, 7, 8

[51] B. Zhou, A. Lapedriza, A. Khosla, A. Oliva, and A. Torralba, "Places: A 10 million image database for scene recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 6, pp. 1452–1464, 2018. 6