Mirror, Mirror, on the Wall, Who's Got the Clearest Image of Them All?—A Tailored Approach to Single Image Reflection Removal

Daniel Heydecker, Georg Maierhofer, Angelica I. Aviles-Rivero, *Graduate Student Member, IEEE*, Qingnan Fan, Dongdong Chen, Carola-Bibiane Schönlieb, and Sabine Süsstrunk, *Fellow, IEEE*

***Abstract*— Removing reflection artefacts from a single image is a problem of both theoretical and practical interest, which still presents challenges because of the massively ill-posed nature of the problem. In this paper, we propose a technique based on a novel optimization problem. First, we introduce a *simple* user interaction scheme, which helps minimize information loss in the reflection-free regions. Second, we introduce an H^2 fidelity term, which preserves fine detail while enforcing the global color similarity. We show that this combination allows us to mitigate the shortcomings in structure and color preservation, which presents some of the most prominent drawbacks in the existing methods for reflection removal. We demonstrate, through numerical and visual experiments, that our method is able to outperform the state-of-the-art model-based methods and compete with recent deep-learning approaches.**

***Index Terms*— Reflection suppression, image enhancement, optical reflection.**

I. Introduction

THIS paper addresses the problem of single image reflection removal. Reflection artefacts are ubiquitous in many classes of images; in real-world scenes, the conditions are often far from optimal, and photographs have to be taken in which target objects are covered by reflections and artefacts appear in undesired places. This does not only affect amateur photography; such artefacts may also arise in documentation in museums and aquariums, or black-box cameras in cars (see Fig. 1). It is therefore unsurprising that the problem of removing reflection artefacts is of great interest, from both practical and theoretical points of view.

Although it is possible to reduce reflection artefacts by the use of specialized hardware such as polarization filters [1]–[3], this option has several downsides. Firstly, even though the use of hardware can have a significant effect on removing the reflection, it only works when certain capture conditions are fulfilled, such as Brewster's angle [4]. In practice, it is difficult to achieve optimal capture conditions, which results in residual reflections [5], [6]. As a result, post-processing techniques are often needed for further improvement of the image. Moreover, for the purposes of amateur photography, the use of specialized hardware is expensive, and consequently less appealing.

As an alternative to the use of specialized hardware, a body of research has established a variety of computational techniques. These can be divided in those that use *multiple images*, and those that use a *single image*. The former techniques employ images from various view points (e.g. [7]–[10]), with the aim of exploiting temporal information to separate the reflection artefacts from the observed target, while for the latter, carefully selected image priors are used to obtain a good approximation of the target object, for example [11]–[14].

Although the use of multiple images somewhat mitigates the massively ill-posed problem created by the reflection removal formulation, the success of these techniques requires multiple images from several viewpoints and their performance is strongly conditional on the quality of the acquired temporal information. Moreover, in practice, acquisition conditions are non-optimal, which often results in image degradation, causing occlusions and blurring in the images. Therefore, either many images or post-processing are needed, which strongly restricts the applicability and feasibility of these methods to a typical end-user. These constraints make single-image methods a focus of great attention to the scientific community, since it is appropriate for most users, and this is the approach which we will take in this paper.

Mathematically, an image $\mathbf{Y}$ containing reflection artefacts can be represented as a linear superposition [15] as:

$$\mathbf{Y} = \mathbf{T} + \mathbf{R}, \tag{1}$$

where $\mathbf{T}, \mathbf{R}$ are $n \times m$ matrices representing the transmission layer and reflection layer, respectively. Therefore, the goal of

Manuscript received August 4, 2018; revised January 25, 2019 and April 12, 2019; accepted May 30, 2019. Date of publication July 1, 2019; date of current version September 4, 2019. This work was supported by the CMIH University of Cambridge. The work of D. Heydecker and G. Maierhofer was supported by the U.K. Engineering and Physical Sciences Research Council (EPSRC) through the University of Cambridge Centre for Doctoral Training and the Cambridge Centre for Analysis under Grant EP/L016516/1. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Chia-Kai Liang. *(Daniel Heydecker, Georg Maierhofer, and Angelica I. Aviles-Rivero are co-first authors). (Corresponding author: Daniel Heydecker.)*

D. Heydecker, G. Maierhofer, A. I. Aviles-Rivero, and C.-B. Schönlieb are with the DAMTP and DPMMS, University of Cambridge, Cambridge CB3 0WA, U.K. (e-mail: dh489@cam.ac.uk; gam37@cam.ac.uk; ai323@cam.ac.uk; cbs31@cam.ac.uk).

Q. Fan is with the Computer Science and Technology School, Shandong University, Shandong 266237, China (e-mail: fqnchina@gmail.com).

D. Chen is with Microsoft Cloud and AI, Beijing 100080, China (e-mail: cddlyf@gmail.com).

S. Süsstrunk is with the School of Computer and Communication Sciences, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland (e-mail: sabine.susstrunk@epfl.ch).

This article has supplementary downloadable material at http://ieeexplore.ieee.org, provided by the authors.

Digital Object Identifier 10.1109/TIP.2019.2923559

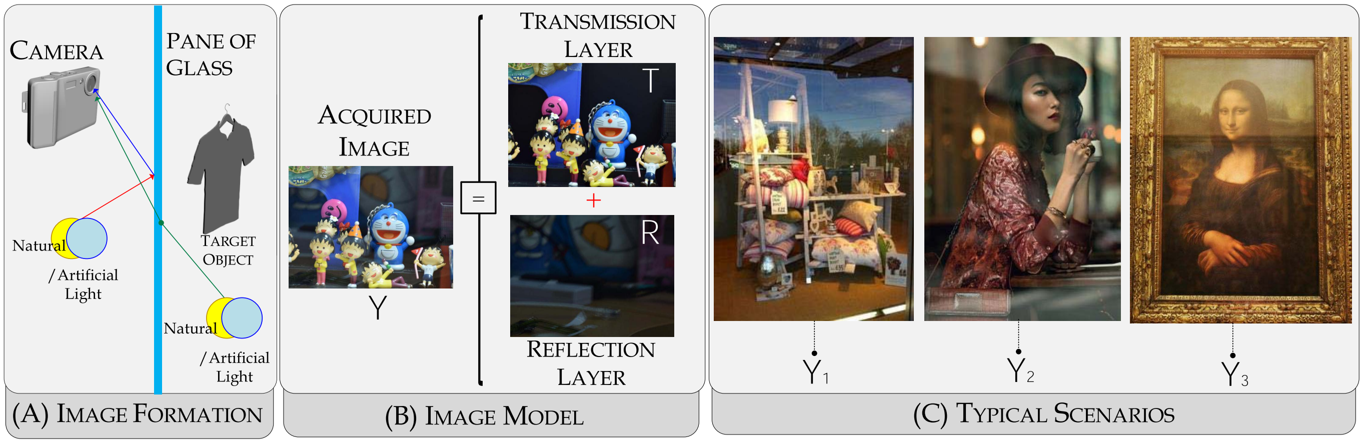


Fig. 1. (A) An illustration of the image formation in which a target object captured through a pane of glass will have reflection artefacts. (B) Based on the image model, an acquired image (**Y**) can be decomposed into two layers: Transmission (**T**) and Reflection (**R**). (C) images ($\mathbf{Y}_{1,2,3}$) show a set of typical situations where there is no option but to take the picture through a pane of glass such as store display or in museums.

a reflection suppression technique is to approximate **T** from the acquired image **Y**.

Although the body of literature for single-image reflection removal has proven promising results, this remains an open problem, and there is still potential for further enhancements. We consider the problem of how to get a better approximation of **T**.

In this work, we propose a new approach, closely related to [14], and inspired by the observation that even *low-level* user input may contain a lot of information. Our technique relies on additional information, which gives the rough location of reflections. In our experiments, this is given by user-input; in principle, this could be done by an algorithmic or machine-learning technique. We recast the reflection removal problem as an optimization problem which is solved iteratively, by breaking it up into two more computationally tractable problems. Compared to existing solutions from the literature, we achieve a better approximation of **T** from a well-chosen optimization problem, which preserves image details and eliminates global color shifts. Our contributions are as follows:

- We propose a computationally tractable mathematical model for single-image reflection removal, in which we highlight:
  - A *simple and tractable* user interaction method to select reflection-heavy regions, which is implemented at the level of the optimization problem as a spatially aware prior term. We show that this improves the retention of detail in reflection-free areas.
  - A combined H^2 *fidelity term*, which combines L^2 and Laplacian terms. We show that this combination yields significant improvements in the quality of the color and structure preservation.

  We establish that the resulting optimization problem can be solved efficiently by half-quadratic splitting.
- We validate the theory with a range of numerical and visual results, in different scenes and under varying capture conditions.
- We demonstrate that the combination of our fidelity term and prior term leads to a better approximation of **T** than state-of-the-art model based techniques, and can compete with the most recent deep-learning (DL) techniques.

II. Related Work

The problem of image reflection removal has been extensively investigated in the computer vision community, in which solutions rely on using multiple images and single image data, alone or in combination with specialized hardware. In this section, we review the existing techniques in turn.

A number of techniques have been developed which use information from multiple images to detect and remove reflections. These include the use of different polarization angles [3], [5], [6], [16], [17], adjustment of focal and flash settings [1], [2], [18], and the uses of relative motion and coherence [7], [8], [19]–[25]. A recent technique [26] seeks to improve on these methods by seeking to match the *transmitted* layer, while other techniques may erroneously match the reflected layer. Each of these techniques requires particular modeling hypotheses to be met, and advantageous capture conditions which may not be feasible in practice.

We now review the related works in single image techniques, as they are most applicable to everyday capture. A commonality of these techniques is the choice of a sparse gradient prior, which imposes a preference for output transmission layers **T** with few strong edges.

A user-intervention method was proposed in [11], which labels gradients as belonging to either transmission or reflection layer. They then propose to solve a constrained optimization problem, with prior distribution given by the superposition of two Laplace distributions. A similar optimization problem is used by [13], which replaces user-intervention labeling by a depth-of-field based inference scheme, while [27] relies on ghosting artefacts.

Our work is most closely related to the optimization-based models and techniques of [12], [14]. The authors of [12] propose a smooth gradient prior on the reflection layer, and a sparse gradient prior on the transmission layer. This approach was adapted by Arvanitopoulos *et al.* in [14], who proposed a Laplacian-based fidelity term with a novel sparse gradient prior. This preserves (Gestalt) continuity of structure, while also reducing loss of high-frequency detail in the transmission layer. The algorithm they propose is both more effective, and more computationally efficient, than the other techniques discussed above.

The application of deep learning to reflection removal was pioneered by Fan *et al.* in [28]. In this work, the authors propose a deep neural network structure, which firstly predicts the edge map and then separates the layers. This technique outperforms the algorithmic approach of [12]. Further work in this direction was made by Zhang *et al.* [29], who use a fully convolutional neural network with three loss terms, which help to ensure preservation of features and pixel-wise separation of the layers. Wan *et al.* [30] seek to use a loss function inspired by human perception to estimate the gradient of the transmission layer, and use this to concurrently estimate the two layers using convolutional neural networks, and Jin *et al.* [31] proposes a convolutional neural network with a resampling strategy, to capture features of global priors, and avoid the ambiguity of the average color. Most recently, Yang *et al.* [32] propose a bidirectional deep learning-scheme based on a cascade neutral network. This method first estimates the background layer **T**, then uses this to estimate the reflected layer **R**. Finally, the estimate on **R** is used to improve the estimate of **T**.

The philosophy of our approach is similar to that of [11]. Motivated by the principle that *humans are good at distinguishing reflections*, both our work and [11] seek to exploit further user input to assist an algorithmic technique. However, we emphasize that we are the first to propose a *simple and tractable* user interaction scheme: in evaluating our user interaction scheme in Section IV/E3, we will see that our user interaction scheme requires very little effort from the user, and that our algorithm performs well with even very crude selection. By contrast, the algorithm of [11] requires much more effort, and a much more detailed input.

III. Proposed Method

This section contains the three key parts of the proposed mathematical model: (i) the combined Laplacian and L^2 fidelity term, (ii) a *spatially aware* prior term, given by user input, and (iii) a computationally tractable solution, using quadratic splitting, to the resulting optimization problem.

Although the model for an image with reflection artefacts described in (1) is widely-used, our solution adopts the observation of [1], [12], [14] that the reflection layer is less in focus and often blurred, which we formalize as follows:

Observation 1: In many cases, the reflected image will be blurred, and out of focus. This may be the case, for instance, if the reflected image is at a different focal distance from the transmitted layer. Moreover, reflections are often less intense than the transmitted layer.

Based on this observation, the image model [1], [12] which we adapt is

$$\mathbf{Y} = w\mathbf{T} + (1-w)(\mathbf{k} \star \mathbf{R}), \tag{2}$$

where $\star$ denotes convolution, w is a weight $w \in [0, 1]$ that controls the relative strength of reflections, and $\mathbf{k}$ is a blurring kernel.

A. Fidelity and Prior Terms

We begin by discussing the prior term. Loss of some detail, in reflection heavy regions, is to be expected, and is a result of the ill-posed nature of reflection suppression. We seek to use low-level user input to reduce the loss of detail *in reflection-free regions*, motivated by the following observation:

Observation 2: In many instances, the reflections are only present in a region of the image, and it is easy for an end user to label these areas. In regions where reflections are not present, all gradients in **Y** *arise from* **T***, and so should not be penalized in a sparsity prior. Moreover, in certain instances, it may be particularly important to preserve fine detail in certain regions.*

For instance, for photographs containing a window, the reflections will only occur in the window, and not elsewhere in the image. To this end, we propose to incorporate a *region selection function* $\boldsymbol{\phi}$, taking values in [0, 1], into a *spatially aware prior:*

$$P(\boldsymbol{\phi}, \mathbf{T}) = \sum_{i,j} \phi_{ij} 1[\nabla_x T_{ij} \neq 0 \text{ or } \nabla_y T_{ij} \neq 0]. \tag{3}$$

Here, 1[..] denotes the indicator function for the set of indexes (i, j) where one of the gradients $\nabla_x \mathbf{T}$, $\nabla_y \mathbf{T}$ is nonzero. We assume that the region selection function $\boldsymbol{\phi}$ is given by the user, along with the input. Although this is philosophically similar to the user intervention method of [11], our approach is drastically less effort-intensive: rather than labeling many edges, it is sufficient to (crudely) indicate which regions contain reflections. The practicalities of our technique will be discussed in Subsection C below. We will show that, by choosing $\phi_{ij} \approx 1$ on reflection-heavy regions and $\phi_{ij} \approx 0$ elsewhere, we can minimize the loss of detail in reflection-free areas. Without this, we would see a 'flattening' effect, where large areas are wrongly given the same value and contrast is decreased, as gradients belonging to the transmitted layer **T** are wrongly suppressed. This removes visual cues, such as minor color variation, which indicate depth, and leads to visually unpleasant and unrealistic-seeming output in which objects appear 'flat'. Examples of this will be highlighted in the experimental results. We also note that a naïve attempt to apply the approach of [14] to a region of the image produces noticeable color shifts at the boundary of the selected region, which our spatially aware prior term avoids.

We now consider the fidelity term, seeking to build on the Laplacian fidelity term proposed by [14]; this choice of fidelity term penalizes over-smoothing, and enforces consistency in fine details. Although this improves on the L^2 fidelity term of Xu *et al.* [33], one can still observe significant 'flattening' effects, as described above. Moreover, we also note that for any constant matrix **C** the Laplacian is invariant under the transformation $\mathbf{T} \mapsto \mathbf{T} + \mathbf{C}$. As a result, the algorithm proposed by [14] risks producing global color shifts; at the level of the optimization problem, this reflects the non-uniqueness of minimizers. To eliminate this possibility, we propose a combined H^2 fidelity term:

$$d_\gamma(\mathbf{T}, \mathbf{Y}) = \|\Delta\mathbf{T} - \Delta\mathbf{Y}\|_2^2 + \gamma\,\|\mathbf{T} - \mathbf{Y}\|_2^2, \tag{4}$$

where $\Delta\mathbf{T}$ is the discrete Laplacian defined as $\Delta\mathbf{T} = \nabla_{xx}\mathbf{T} + \nabla_{yy}\mathbf{T}$, and γ is a positive parameter controlling the relative importance of the two terms. We will see, in numerical experiments, that this leads to results with more natural, saturated colors, and which are consequently more visually pleasing. We remark that other kernel filters are possible which would play the same role of measuring structure, such as the discrete gradient ∇, or more complicated elliptic second-order operators; we use the Laplacian for the following reasons. Firstly, the Laplacian penalizes loss of high-frequency detail more strongly than first order operators such as ∇, as can be seen by moving to Fourier space, and so our choice will preserve high-frequency details well. Secondly, the Laplacian is a simple measure of structure, and which is invariant under the (natural) symmetry of rotation.

Combining the prior and fidelity terms, as defined in (3) and (4), our optimization problem is therefore

$$\mathbf{T}^* = \operatorname{argmin}_{\mathbf{T}} \left\{ \|\Delta\mathbf{T} - \Delta\mathbf{Y}\|_2^2 + \gamma\,\|\mathbf{T} - \mathbf{Y}\|_2^2 + \lambda P(\boldsymbol{\phi}, \mathbf{T}) \right\}. \tag{5}$$

Here, λ is a regularization parameter to be chosen later. The reader is invited to compare this optimization problem to the similar problem of (localized) L^0 image smoothing, but to note the important difference of having a fidelity term including the image Laplacian. In the next section, we will detail how the proposed optimization problem can be solved in a tractable computational manner by using quadratic splitting.

B. Solving the Optimization Problem

We solve the optimization problem introduced in (5) by half-quadratic splitting. We introduce auxiliary variables $\mathbf{D}^x, \mathbf{D}^y$ as proxies for, respectively, $\nabla_x\mathbf{T}$ and $\nabla_y\mathbf{T}$. For ease of notation, we write $\mathbf{D}$ for the pair $[\mathbf{D}^x, \mathbf{D}^y]$, and similarly $\nabla\mathbf{T}$ for the pair $[\nabla_x\mathbf{T}, \nabla_y\mathbf{T}]$. This leads to the auxiliary problem:

$$\begin{aligned}\mathbf{T}^*, \mathbf{D}^* = \operatorname{argmin}_{\mathbf{T},\mathbf{D}} \Big\{ &\|\Delta\mathbf{T} - \Delta\mathbf{Y}\|_2^2 + \gamma\,\|\mathbf{T} - \mathbf{Y}\|_2^2 \\ &+ \lambda P(\boldsymbol{\phi}, \mathbf{D}) + \beta\|\mathbf{D} - \nabla\mathbf{T}\|_2^2 \Big\}\end{aligned} \tag{6}$$

where $\beta \in \mathbb{R}_{>0}$ is a penalty parameter yet to be chosen, and we use the shorthand

$$P(\boldsymbol{\phi}, \mathbf{D}) = \sum_{i,j} \phi_{ij}\, 1[D^x_{ij} \neq 0 \text{ or } D^y_{ij} \neq 0]. \tag{7}$$

Notice that in the limit $\beta \to \infty$ the auxiliary penalty term ensures that we recover the solution to the original optimization problem (5). Hence, we may approximately solve the optimization problem (6) by splitting into two more computational tractable problems. We alternate between optimizing over $\mathbf{T}$ and $\mathbf{D}$, while keeping the other fixed; at the same time, we increment β so that, after a large number of steps, $\mathbf{D}$ is a good approximation of $\nabla\mathbf{T}$. We give details on the solution of each sub-problem below, and the full solution is presented in Algorithm 1.

Algorithm 1 Our Proposed Method

```
Start from $\mathbf{T} \leftarrow \mathbf{Y}$ and $\beta = \beta_{\mathrm{min}}$;
while $\beta \le \beta_{\max}$ do
    Optimise over $\mathbf{D}$, for the current value of $\mathbf{T}$:
        Set $(D^x_{ij}, D^y_{ij}) = \begin{cases} (0,0) \text{ if } |(\nabla_x T_{ij}, \nabla_y T_{ij})|_2^2 \le \frac{\lambda_{ij}}{\beta}; \\ (\nabla_x T_{ij}, \nabla_y T_{ij}) \text{ otherwise;} \end{cases}$
    Using ADAM [34] and (12), find the minimum $\mathbf{T}^\star$ of (8), and replace $\mathbf{T} \leftarrow \mathbf{T}^\star$;
    Increment $\beta \leftarrow \kappa\beta$;
end while
return $\mathbf{T}$.
```

1) Sub-Problem 1 (Optimization Over $\mathbf{T}$*):* For a fixed $\mathbf{D}$, we wish to optimize:

$$\begin{aligned}\mathbf{T}^* = \operatorname{argmin}_T \Big\{ &\|\Delta\mathbf{T} - \Delta\mathbf{Y}\|_2^2 + \gamma\,\|\mathbf{T} - \mathbf{Y}\|_2^2 \\ &+ \beta\,\|\mathbf{D} - \nabla\mathbf{T}\|_2^2 \Big\}.\end{aligned} \tag{8}$$

The objective function is now quadratic in $\mathbf{T}$. We note that the discrete gradient ∇ and the discrete Laplacian Δ are both linear maps which take an $m \times n$ image matrix to an array of size $2 \times m \times n$ and $m \times n$ respectively. We can therefore view these linear maps as tensors, and use index notation to describe their action on an image (T_{ij}) as follows:

$$(\nabla_\mu \mathbf{T})_{ij} = \sum_{k,l} \nabla^\mu_{ijkl} T_{kl};\ 1 \le i \le m,\ 1 \le j \le n,\ \mu \in \{x, y\} \tag{9}$$

and similarly:

$$(\Delta\mathbf{T})_{ij} = \sum_{k,l} \Delta_{ijkl} T_{kl};\quad 1 \le i \le m,\ 1 \le j \le n. \tag{10}$$

With this notation, we can write the objective function as:

$$\begin{aligned} &F_1(\mathbf{T}, \mathbf{D}) \\ &= \beta \sum_{\substack{1 \le i \le m \\ 1 \le j \le n \\ \mu \in \{x,y\}}} \left(D^\mu_{ij} - \sum_{1 \le k \le m, 1 \le l \le n} \nabla^\mu_{ijkl} T_{kl} \right)^2 \\ &\quad + \sum_{i \le m, j \le n} \left(\left(\sum_{k \le m, l \le n} \Delta_{ijkl}(T_{kl} - Y_{kl}) \right)^2 + \gamma\,(T_{ij} - Y_{ij})^2 \right) \end{aligned} \tag{11}$$

We observe that this is quadratic, and in particular smooth, in the components T_{ij}. Using the summation convention, we compute the gradient:

$$\begin{aligned}\frac{\partial}{\partial T_{ij}} F_1(\mathbf{T}, \mathbf{D}) = &\ 2\Delta_{abij}\Delta_{abkl}(T_{kl} - Y_{kl}) + 2\gamma\,(T_{ij} - Y_{ij}) \\ &+ 2\beta\nabla^\mu_{abij}(\nabla^\mu_{abkl} T_{kl} - D^\mu_{ab}).\end{aligned} \tag{12}$$

We use this computation, together with ADAM [34], a first-order gradient descent method in stochastic optimization, to efficiently optimize over $\mathbf{T}$.

2) Sub-Problem 2 (Optimization Over $\mathbf{D}$*):* For a fixed $\mathbf{T}$, the optimization problem in $\mathbf{D}$ is given by

$$\mathbf{D}^* = \operatorname{argmin}_{\mathbf{D}} \left\{ \beta \left\| \mathbf{D} - \nabla \mathbf{T} \right\|_2^2 + \lambda P(\boldsymbol{\phi}, \mathbf{D}) \right\}. \tag{13}$$

Although the objective function, F_2, is neither convex nor smooth, due to the L^0 prior term, we observe that it separates as

$$F_2(\mathbf{T}, \mathbf{D}) = \sum_{i,j} \Big[\beta \left(|D^x_{ij} - \nabla_x T_{ij}|^2 + |D^y_{ij} - \nabla_y T_{ij}|^2 \right) + \lambda \phi_{ij} 1 \left((D^x_{ij}, D^y_{ij}) \neq 0 \right) \Big]. \tag{14}$$

By explicitly solving the separated problems for each pair (D^x_{ij}, D^y_{ij}), it is straightforward to see that *a* solution to (14) is given by

$$(D^x_{ij}, D^y_{ij}) = \begin{cases} (0,0) & \text{if } |(\nabla_x T_{ij}, \nabla_y T_{ij})|_2^2 \leq \dfrac{\lambda \phi_{ij}}{\beta}; \\ (\nabla_x T_{ij}, \nabla_y T_{ij}) & \text{otherwise.} \end{cases} \tag{15}$$

Moreover, this minimizer is unique, provided that none of the edges are in the boundary case $|(\nabla_x T_{ij}, \nabla_y T_{ij})|_2^2 = \frac{\lambda \phi_{ij}}{\beta}$.

Hence, the optimization (13) removes gradients below the *local* threshold $\frac{\lambda \phi_{ij}}{\beta}$. We will show, in numerical experiments, that this has the effect of smoothing *only* the selected regions, while keeping the strong edges which force continuity of structures, as was described in Section II.

The overall procedure of our method, in which previous individual steps are combined to solve the original optimization problem (5), is listed in Algorithm 1.

C. User Interaction Scheme

We describe the user interaction scheme, and how the region selection function ϕ_{ij} may be obtained in practice. We recall that ϕ is responsible for passing information about the location of reflection into the algorithm, and that it takes values in the range $[0, 1]$ with

- ϕ_{ij} close to 1 if a reflection is present at pixel (i, j) and
- ϕ_{ij} close to 0 if no reflection is present at pixel (i, j).

In practice a user, or an arbitrary instance that can recognize rough locations of reflections, is given an image, as in left-side of Fig. 2, and selects the regions in which reflections are present. A possible result can be seen in the middle part of Fig. 2, where the values of ϕ_{ij} are displayed as the grey-values in the image. This selection is then fed into our algorithm together with the input image to produce the reflection removed output as shown at right side of Fig.2.

In the absence of user interaction, we default to $\phi_{ij} \equiv 1$; that is, we assume reflections are present throughout the image.

It is noteworthy that the way this selection is performed is very simple and requires little effort. This makes it suitable for a range of applications, from an amateur human user, to algorithms that can recognize reflections, even in a very crude manner. For our experiments, the selection was performed by creating an overlay image in a raster graphics editor, where white regions are marked with a rough brush on top of reflections. This process can be performed in a matter of seconds for each image. The results can, of course, improve with increasing selection quality, but even a rough selection produces significant improvements over no selection; see Section IV/E3 for experiments and discussion. Examples of region selection in practice are included in Section IV of the supplementary material.

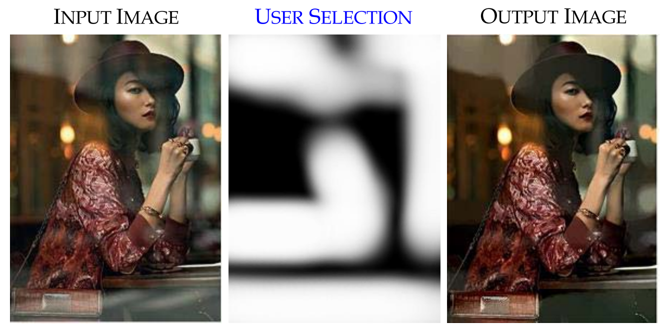


Fig. 2. From left to right. Input image, visualization of the user interaction in practice and output image with our technique.

D. Performance Reasoning of Parameters

Our procedure uses two parameters λ, γ, and an auxiliary parameter β in intermediary optimization steps. We think of β as a coupling parameter, which determines the importance of the texture term in comparison to the coupling to the auxiliary variable $\mathbf{D}$. At later iterations, β is large and the coupling is strong, which justifies the use of $\mathbf{D}$ as a proxy for $\nabla \mathbf{T}$.

The parameter λ determines the relative importance of preserving the structure versus preserving the texture. In terms of the model described above, it controls the importance of the penalty term $P(\boldsymbol{\phi}, \mathbf{T})$ against the Laplacian $\| \Delta \mathbf{T} - \Delta \mathbf{Y} \|_2^2$. In regions where $\lambda \phi_{ij}$ is comparatively large, the sparsity of edges is much more important than the texture. Therefore, any edges which do not enforce structure will be washed out, and the region is smoothed during the optimization over $\mathbf{D}$. On the other hand, in regions where $\lambda \phi_{ij}$ is comparatively small, the texture term dominates, and only very few edges are removed. In terms of the algorithm, this corresponds to controlling the edge threshold $\frac{\lambda \phi_{ij}}{\beta}$. This is illustrated in the supplementary material.

We also give an interpretation of why it is natural to increase β in this way. In the first stages of the iteration, β is very small, and so the threshold keeps only the largest magnitude edges, and sets most edges of reflection-heavy areas to 0. After each iteration, β increases and the threshold $\frac{\lambda_{ij}}{\beta}$ decreases, and so the next iteration will preserve more edges. Hence, in reflection-heavy areas, we include edges in decreasing order of magnitude; this corresponds to looking at strongly-defined structures first, and then considering incrementally weaker structure. This is illustrated in the supplementary material.

We give a theoretical basis for excluding the limiting regimes of either $\gamma \ll 1$ or $\gamma \geq 1$. In the regime where

$\gamma \ll 1$, we may consider a step of the gradient descent to be a step of 'uncorrected' gradient descent, with $\gamma = 0$, followed by a small correction $\gamma(\mathbf{Y} - \mathbf{T})$ to correct color shift. For this reason, if $\gamma \ll 1$ is too small, our algorithm will not adequately correct for color shifts. On the other hand, if $\gamma > 1$, then the L^2 term dominates the Laplacian term, and we expect blurring and loss of texture, as discussed in [14].

IV. Experimental Results

In this section, we describe in detail the range of experiments that we conducted to validate our proposed method.

A. Data Description

We evaluate the theory using the following three datasets. Firstly, we use real-world data from the SIR^2 benchmark dataset [35]. The dataset is composed of 1500 images with size of 400×540, and provides variety in scenes with different degrees of freedom in terms of aperture size and thickness of the glass. These variations allow us to test the respective algorithms in the presence of different effects, such as reflection shift. Moreover, it provides a ground truth that permits for quantitative evaluation. We also use the Berkeley dataset from [29], which contains 110 real image pairs (reflection and transmission layer) whose characteristics can be founds in [29]. Finally, we also use a selection of 'real-world' images from [28], for which ground truths are not available. All measurements and reconstructions were taken from these datasets.

B. Evaluation Methodology

We design a four-part evaluation scheme, where the evaluation protocol for each part is as follows.

(E1) The first part is a visual comparison of our method against AR17 [14]. We remark that in the case $\gamma = 0, \phi \equiv 1$, our method reduces to that of AR17; this comparison therefore shows that the changes made to the objective function fulfill their intended purposes.

(E2) The main part of the evaluation is to compare our solution to the state-of-the-art methods. In (E2a) we compare to state-of-the-art algorithmic techniques LB14 [12], SH15 [27], AR17 [14], using FAN17 [28] as a benchmark. (E2b) is an evaluation against more recent advances in deep-learning FAN17 [28], WAN18 [30], ZHANG18 [29] and YANG18 [32] on both real-world images and the Berkeley dataset. We present both numerical comparisons, averaged over the SIR^2 and Berkely datasets in (E2a, E2b) respectively, and visual comparisons for a range of selected images from all three datasets.

(E3) We evaluate the impact of the user input, and show the results of our method with no region selection, with crude region selection and with more detailed region selection. This will justify our claim that crude region selection is sufficient to minimize loss of detail in reflection-free areas, but offers a substantial qualitative improvement on *no* region selection.

(E4) Finally, we demonstrate that, by comparison to the existing user interaction approach of Levin and Weiss [11], we produce better results whilst requiring less effort from the end-user.

We address our scheme from both qualitative and quantitative points of view. The former is based on a visual inspection of the output $\mathbf{T}$, and the latter on the computation of three metrics: the structural similarity (SSIM) index [36], the Peak Signal-to-Noise Ratio (PSNR) and the inverted Localized Mean Squared Error (sLMSE). Explicit definition of the metrics can be found in Section VI of the Supplemental Material.

C. Parameter Selection

For each of the approaches LB14 [12], SH15 [27] and AR17 [14], we use the available codes from each corresponding author, and set the parameters as described in the corresponding paper. For FAN17 [28], we assumed a given trained network and with parameters set as described in that paper.

For our approach, we set the values of the ADAM method as suggested in [34]. For our technique, we set $\lambda = 2e - 3$, $\beta_{\max} = 1e5$ and $\kappa = 2$ and $\gamma = 0.012$. The choices of $\lambda, \beta_{\max}, \kappa$ follow [14] for analogous parameters, which is consistent with the reasoning in Subsection III-D. γ was chosen based on experimental results for a range of images disjoint from the test dataset, with a range of test values following the discussion in Subsection III-D. The effect of different choices of γ, which validates this choice, is discussed further in Section VII of the Supplementary Material.

D. Results and Discussion

We evaluate our proposed method following the scheme described in Section IV-B.

(E1). We begin by evaluating our method against AR17 [14]. We ran both approaches on the complete solid objects category of the dataset. In Fig. 3, we show four output examples with different settings (Aperture value F=$\{11, 32\}$ and thickness of glass TG=$\{3, 10\}$). Visual assessment agrees with the theory of our approach, in which we highlight the elimination of color shifts and the preservation of the image details. Most notably, we see that our approach enforces global color similarity and avoids blurring effects produced by the outputs of AR17 [14]; see, for example, outputs (A), (C) and (D). The detail in Fig. 3 highlights these effects, in particular in (A) the blur and color loss effects in the *Winnie the Pooh* toy, in (C) the loss of edge details in the shirt collar (left toy) and the neck (white toy), and in (D) a blurring effect in the toy's legs. In the detail of output (B), it can be seen that AR17 [14] fails to preserve the shadows and the color saturation of the floral pattern. This is further reflected in the numerical results, where our method reported higher values for the three evaluation metrics.

Overall, we noticed that often AR17 [14] fails to penalize color shifts, due to the translation invariance of the Laplacian fidelity term. It also tends to produce blurring effects in reflection-free parts of the image, which our approach is able to prevent through our spatially aware technique.

(E2a). We now evaluate our approach against the model-based state-of-the-art methods (LB14 [12], SH15 [27],

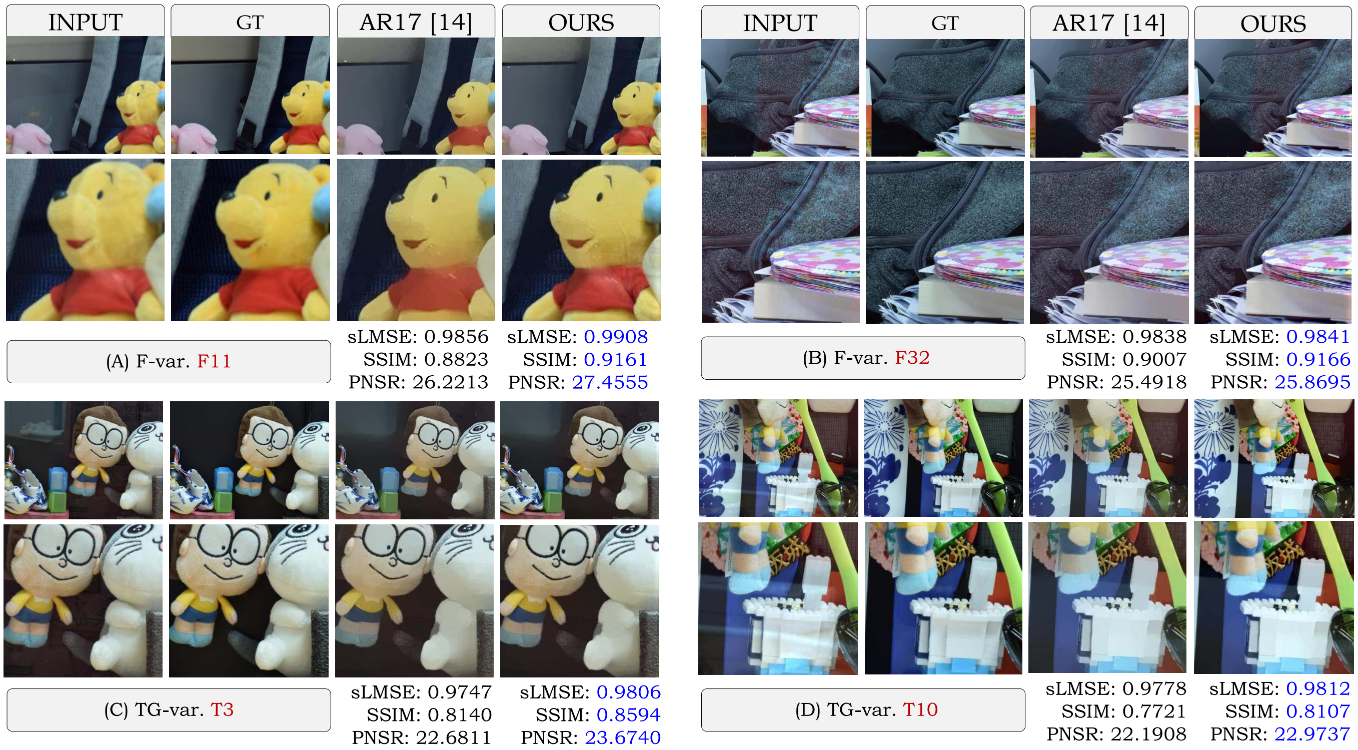


Fig. 3. (E1). Examples of the output, along with ground truth, of our approach compared against AR17 [14]. The examples with varying settings such as the focus in (A) and (B) and the glass thickness in (C) and (D). The three evaluation metrics of the reflection-free image are computed using the ground truth.

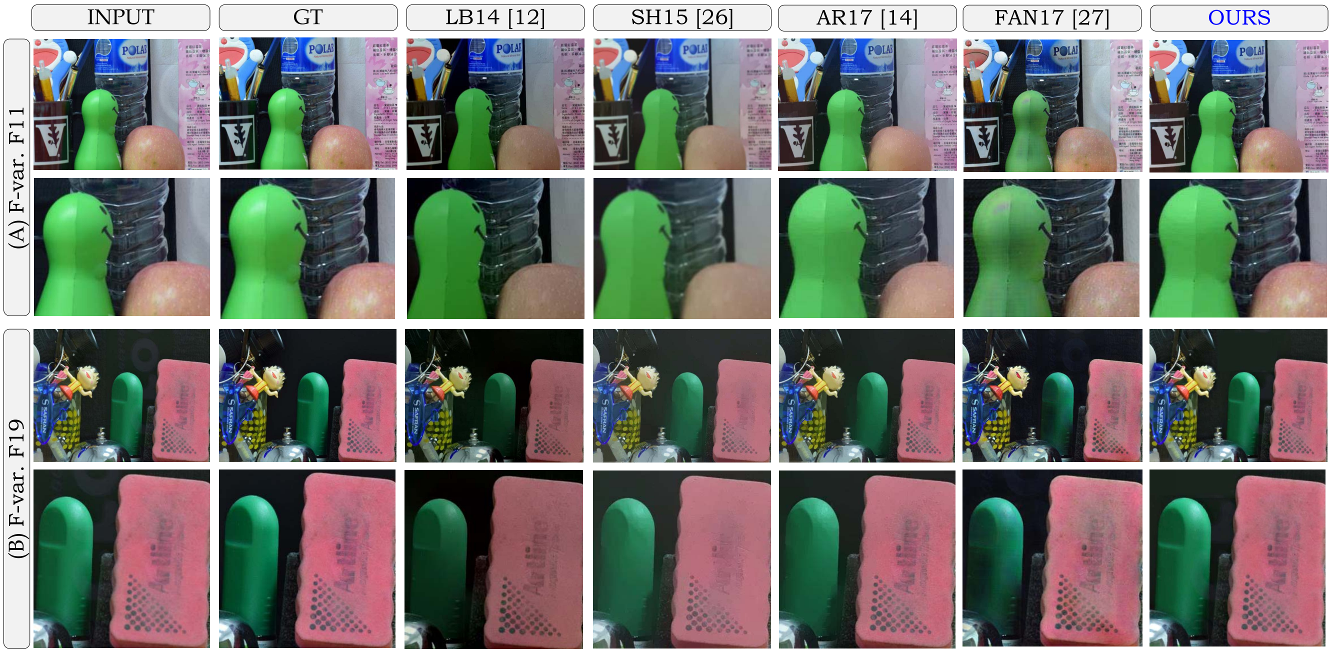


Fig. 4. (E2a). Visual comparison against the state-of-the-art of model-based approaches (including FAN17 [28] as baseline for comparison). The selected frames show variations in shape, color and texture to appreciate the performance of the compared approaches. Overall, our approach gives a better approximation of $\mathbf{T}$ by preserving color and structure quality while keeping fine details. Details are better appreciated on screen.

AR17 [14], and include FAN17 [28] as a baseline of comparison) using the full solid objects category of the SIR^2 dataset. As discussed above, we may view the results of AR17 as those of our algorithm in the special case $\gamma = 0$, and without user interaction ($\phi \equiv 1$) to evaluate the effect of these changes. We emphasize that results for our algorithm were generated with user interaction, *as this is a key part of our technique.*

TABLE I

(E2a). Measures Averaged Over All Images in the Solid-Object Dataset [35]

F-var.	sLMSE			SSIM			PNSR		
	F11	F19	F32	F11	F19	F32	F11	F19	F32
LB14 [12]	0.835	0.832	0.833	0.784	0.804	0.791	21.659	21.869	21.678
SH15 [27]	0.901	0.852	0.874	0.779	0.813	0.765	21.642	22.046	21.620
AR17 [14]	0.983	0.984	**0.984**	0.820	0.825	0.824	22.748	22.705	22.851
FAN17 [28]	0.981	0.982	0.982	**0.854**	0.859	0.851	**23.262**	23.853	23.432
OURS	**0.984**	**0.986**	**0.984**	0.852	**0.866**	**0.854**	23.254	**23.907**	**23.649**

TG-var.	sLMSE			SSIM			PNSR		
	TG3	TG5	TG10	TG3	TG5	TG10	TG3	TG5	TG10
LB14 [12]	0.834	0.833	0.834	0.718	0.811	0.805	21.605	21.981	21.850
SH15 [27]	0.915	0.889	0.917	0.779	0.820	0.765	21.682	22.546	21.620
AR17 [14]	0.983	**0.984**	0.982	0.820	0.825	0.824	22.748	22.705	22.851
FAN17 [28]	0.981	0.981	0.981	**0.850**	**0.852**	0.852	**23.415**	23.403	23.470
OURS	**0.984**	**0.984**	**0.984**	0.846	0.851	**0.861**	23.374	**23.421**	**23.507**

We show the output of the selected methods and our proposed one for four chosen images along with the ground truth in Fig. 4. By visual inspection, we observe that outputs generated with LB14 [12] are darker than the desired output; see, for instance, the detail of (A). Moreover, LB14 fails to preserve texture and global color similarity, as is apparent in (A) on the surface of the apple, and (B) on the pink block. By contrast, our approach was able to keep the details on both cases. Moreover, we observed that both SH15 [27] and AR17 [14] tend to have a noticeable color shift and a significant loss of structure, as is visible on (B) the green pole. In particular, we highlight the green pole in (B), in which only our approach was clearly able to maintain the fine details.

We observe that the deep learning based solution FAN17 [28] shows good edge preservation, but often fails to correctly reproduce color and texture, and produces noticeable artefacts. This will be discussed further in (E2b). Overall, out of the evaluated model-based single-image reflection removal techniques, our approach consistently yields the most visually pleasing results. These observations are confirmed by further examples in Section II of the Supplemental Material.

For a more detailed quantitative analysis, we report the global results in Table I. The displayed numbers are the average of the image metrics across the whole body of ‘solid-object’ files in the dataset, in order to understand the general behavior and performance of the algorithms.

We observe that both AR17 [14] and our approach outperform the remaining algorithms with respect to sLMSE. With respect to SSIM and PNSR, we also achieve significant improvements over most state-of-the-art techniques, most notably over the similar technique AR17 [14]. The only other approach evaluated here which performs similarly well is the deep learning approach FAN17 [28]. As was discussed above, a closer look at single images shows occasional difficulties of this approach, and the more reliable performance of our model-based method.

(E2b). Having extensively compared our new method to model-based approaches in (E2a), we now present a detailed comparison against recent advances in single-image reflection removal based on deep-learning. We compare against FAN17 [28], WAN18 [30], ZHANG18 [29] and YANG18 [32] on both the Berkeley dataset and real-world images.

Having used FAN17 [28] as a benchmark for comparison in (E2a), we first present a further comparison of this method against our technique. Indeed, from Table I, it may appear that FAN17 produces output of a similar quality to our technique. However, we notice that the outputs displayed in Fig. 4 suggest that our method produces *visually nicer* results; to validate this, we present further experiments in Fig. 5. The images displayed are two cases from the SIR^2 dataset, in which we observe difficulties similar to those in Fig. 4. In Fig.s 4A, 5A, FAN17 has wrongly identified a specular reflection in the transmitted layer as belonging to the reflected layer, producing unpleasant artefacts. We also highlight incomplete reflection removal in the examples in Fig. 5, false-color effects in Fig.s 4 and 3B, and unwanted color flattening in Fig. 5A.

Next, in Table II we present the similarity measures which are computed as the average over all images in the Berkeley dataset. With respect to sLMSE, our method outperforms all other techniques, in particular FAN17 [28], WAN18 [30] and YANG18 [32] by a significant margin. With respect to SSIM and PNSR, our method performs similarly well, and places second behind ZHANG18 [29].

To further analyze the performance of the techniques on this dataset, we present a visual comparison of a selection of interesting cases from the Berkeley dataset in Fig. 6, including the values of the similarity metrics to the ground truths. We observe that FAN17 [28] displays poor color retention in the first image and introduces displeasing artefacts in the second, and similarly WAN18 [30] somewhat darkens the colors of the first image, and displays incomplete removal of the reflection in the second. YANG18 [32] induces a significant amount of blurring, which is visible on the roll of tape in the first image, and the door in the third. In the second and third images, ZHANG18 [29] performs very well both visually and numerically, which is consistent with the strong numerical results reported in Table II, but performs very

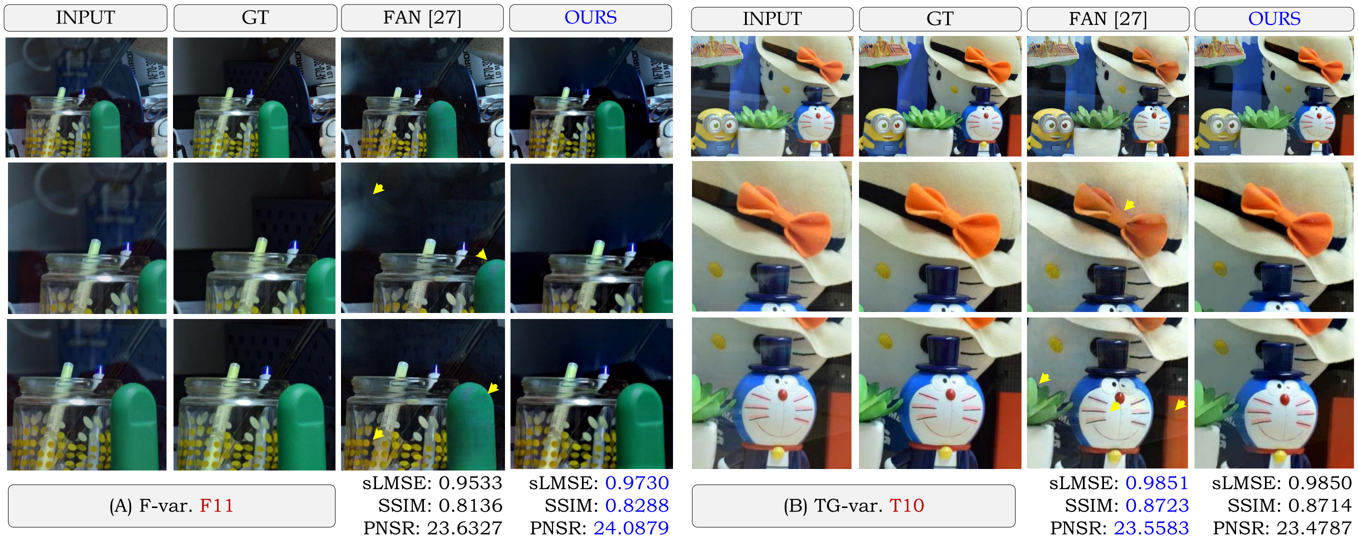


Fig. 5. (E2b). Two interesting cases in which we visually and numerically compare our approach against the work of Fan *et al.* [28]. We emphasize that even in cases when the metrics are higher for FAN17 [28], the output from our algorithm appears visually more appealing and natural. We highlight the false color effects (see bow in (B)), loss of fine details (see green object in (A)) and reflection artefacts (see yellow markers in both) in the output of FAN17. Details are better appreciated on screen.

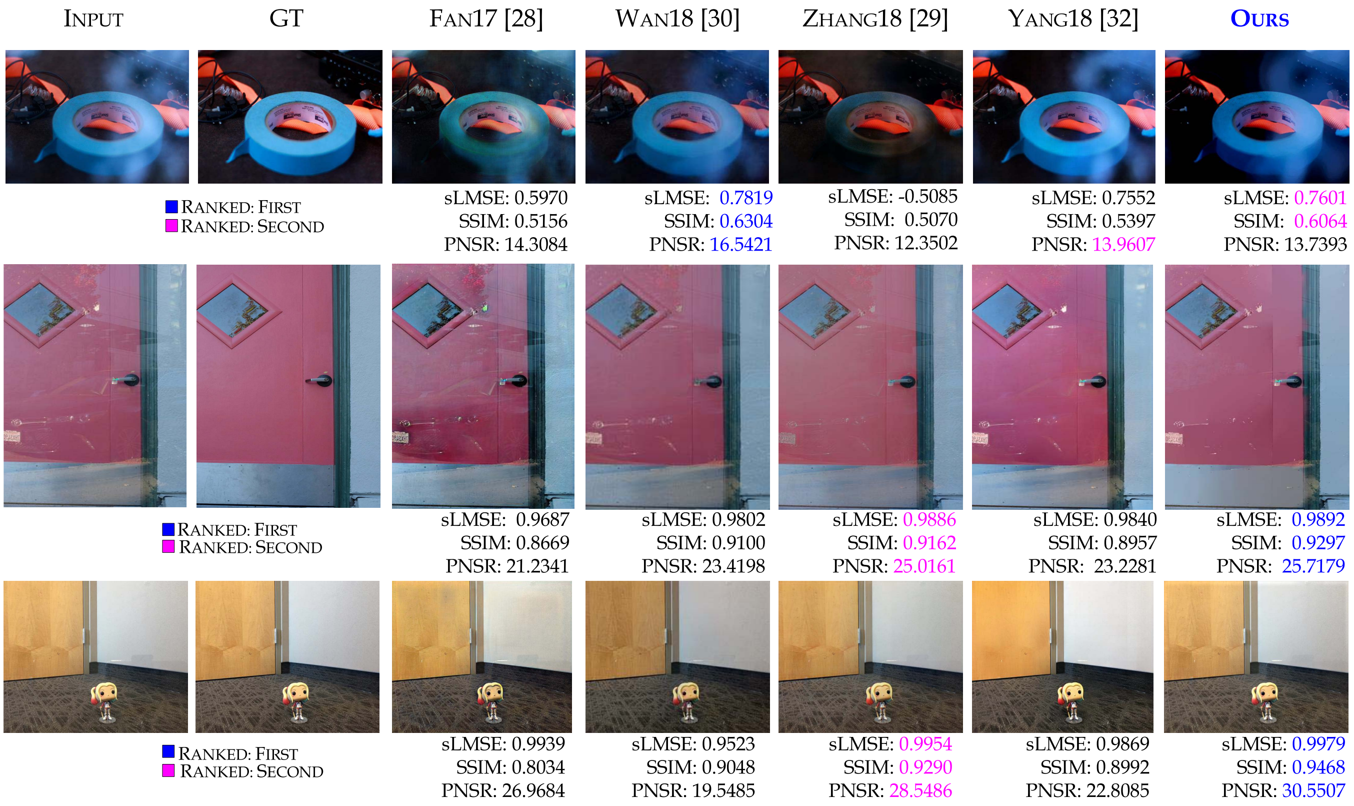


Fig. 6. (E2b). Visual and numerical comparison of our technique vs. Deep-learning techniques on a selection of images from the Berkley dataset. Details are better appreciated on screen.

badly on the first image. Therefore, although ZHANG18 [29] performs very well *on average*, this performance is highly inconsistent. In each image, our method readily competes with the best performing DL technique in terms of similarity metrics, but also is able to preserve structure and color, while still removing a comparable amount of the reflections. We note that, in the second displayed image, our algorithm is unable to completely remove the reflection, but that the resulting suppression is comparable to that of ZHANG18 [29], and better than the competing DL techniques.

TABLE II

(E2b). Numerical Comparison of Our Technique vs. Deep-Learning Techniques for the Entire Berkley Dataset. The Numerical Values Are Computed as the Averages of the Similarity Metrics Over All Images

The Berkley Dataset					
	FAN17 [28]	WAN18 [30]	ZHANG18 [29]	YANG18 [32]	**OURS**
sLMSE	0.8407	0.8090	**0.8638**	0.8398	**0.8647**
SSIM	0.7022	0.6982	**0.7923**	0.6911	**0.7315**
PNSR	18.2989	18.300	**21.6203**	17.8673	**18.7833**

🟦 Ranked First 🟥 Ranked Second

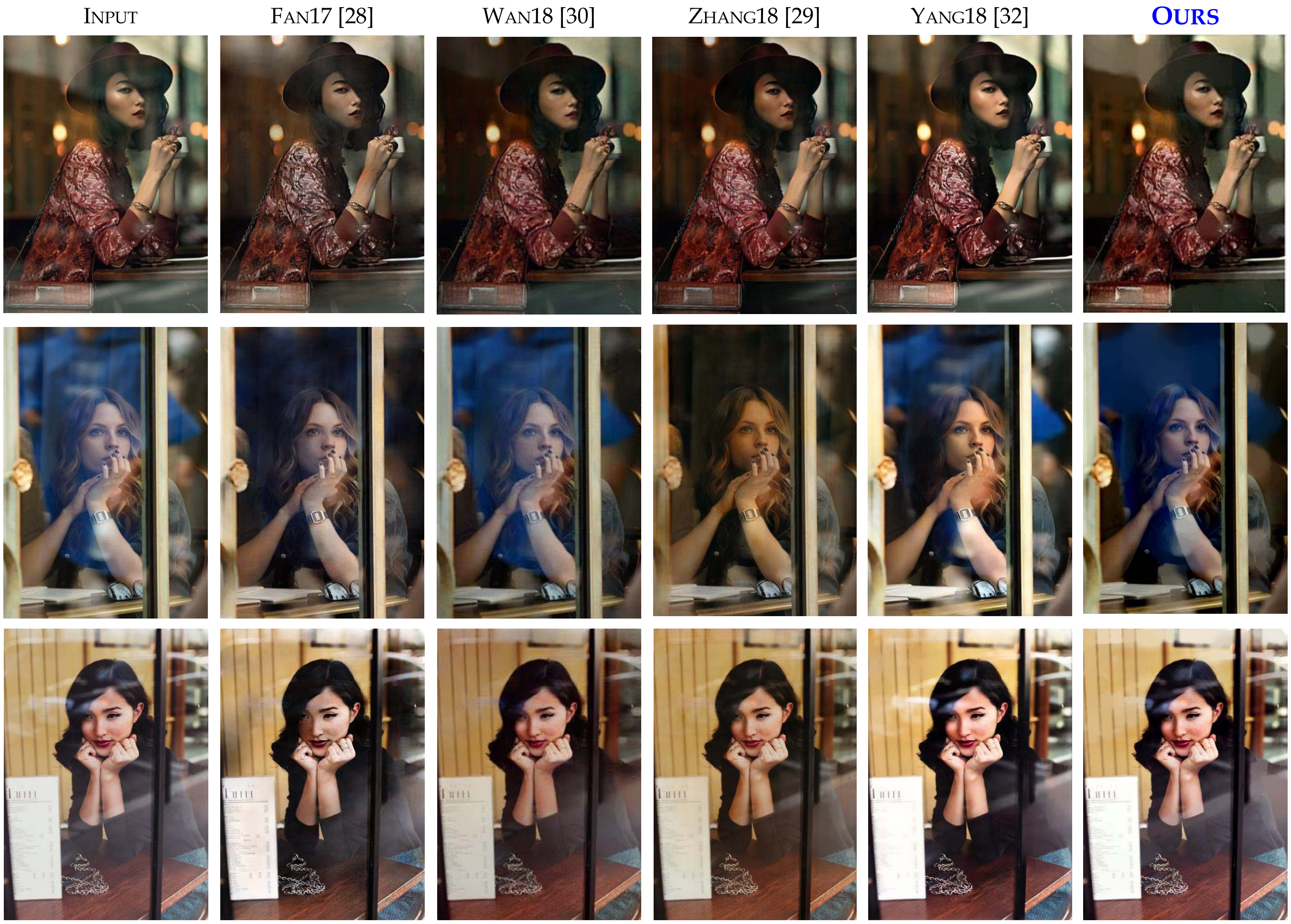


Fig. 7. (E2b). Comparison of our technique vs. Deep-Learning techniques on real-world images. We note that our technique is able to suppress the reflections while avoiding the flattening effect visible in the outputs on FAN17 [28], and avoiding color shifts such as those produced by FAN17 and YANG18, which visibly undersaturate the skin tone in the first image. This is an example of our motivation in Observation 2: color flattening on the skin is much more noticeable than the same effect on the props. Images are from the real-world dataset [28] and no ground truths are available.

Finally, we test all of the DL methods on a selection of real-world images in Fig. 7. These images are from the real-world dataset [28], where no ground truth is available, and so a numerical evaluation is impossible here; however, the results will allow us to evaluate the qualitative performance of our technique against competing techniques for real-world images. We observe that most of the competing methods suffer from poor color preservation, which is especially visible in ZHANG18 [29] with respect to the skin color in middle and upper image, and incomplete removal of the reflections. In FAN17 [28] especially we notice the introduction of artefacts on the arms in the top picture and near the head in the bottom one. Our method, while not completely removing the reflections, still ensures good preservation of color and important structure, and produces outputs of similar visual quality to the competing DL techniques. Additional experiments, which further validate this conclusion, may be found in Sections III, VIII of the Supplementary Material.

The above comparisons demonstrate that *at this point in time, our model-based method readily competes with deep learning in terms of output quality*. The authors note that traditionally, deep learning has achieved ground breaking

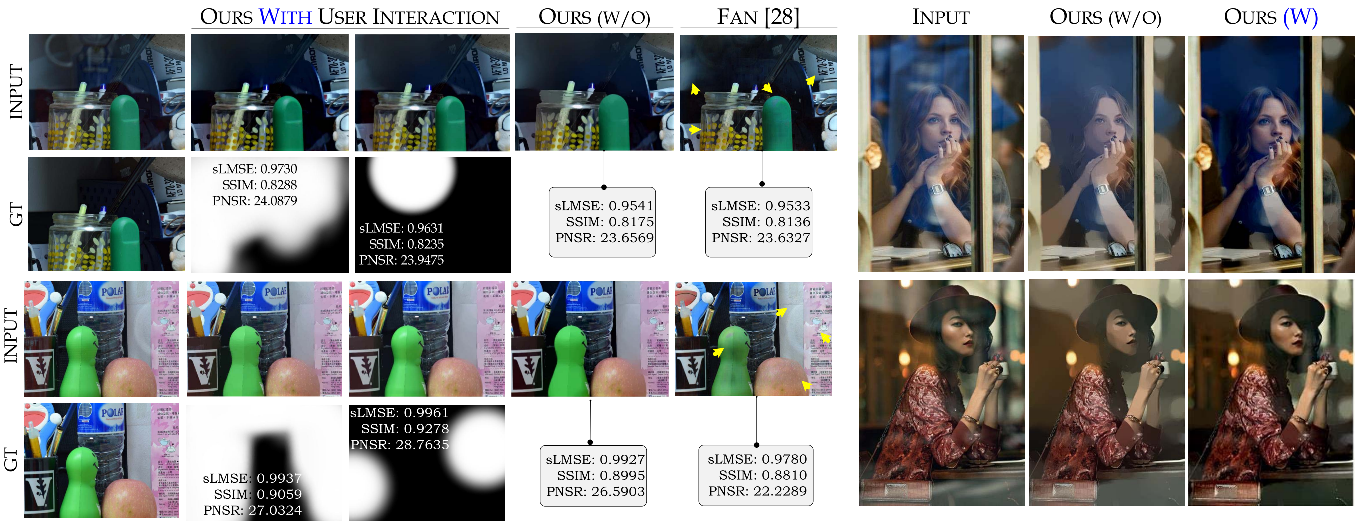


Fig. 8. (E3). From left to right: The impact of the user-interaction on the outputs computed by OUR approach (with and without user interaction), with FAN [28] as a benchmark. Examples of cases where region selection leads to noticeable qualitative improvements in avoiding flattening.

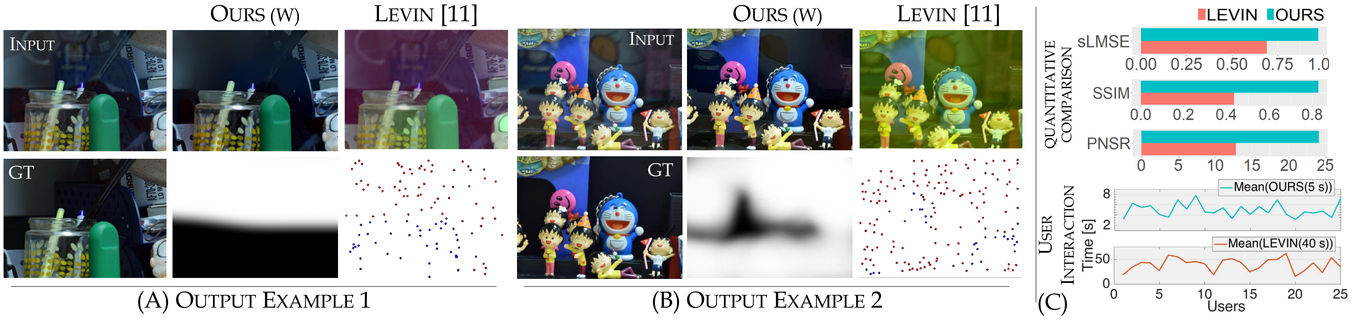


Fig. 9. (E4). (A-B): Visual comparison of the user-interaction schemes in LEVIN [11] and OURS based on a specific example (C): Quantitative comparison of the two schemes on the solid object corpus of the SIR^2 dataset, and user-interaction time based on a selection of images from this dataset.

success in tasks involving labeling or classification [37], [38]. The good visual results generated by deep network usually benefit from the statistical information covered in the large body of training samples. However, a plain fully convolutional neural network does not impose the same kind of rigid and intuitive constraints as model-based approaches; for example, piecewise smoothness is not enforced. Such a limitation in the deep network results in inconsistent reflection removal within a single image, as seen in Fig.s 5, 6, 7. While in this paper the deep-learning based techniques provide an important benchmark, their classification as 'single-image' techniques raises definitional issues that might be interesting for the community to discuss. This discussion can be found in Section V of the Supplemental Material.

(E3). In Fig. 8, we analyze the impact of the user-interaction, again including FAN17 [28] as a baseline for comparison. In the first subfigure, we present the results of our approach without region selection, and with both crude and detailed region selection. Without region selection, there is noticeable blurring and flattening: see, for example, the green object in the first example and the apple in the second. Even with very crude region selection, our technique is able to mitigate these to produce a visually better result which outperforms the result of FAN17 [28]. A more refined region selection, as displayed in the first subcoloumn, leads to an additional small improvement but demonstrates that the quality of our approach is not strongly dependent on a highly detailed region selection. In the second subfigure, we show the result of our technique with and without region selection on two examples from the real-world dataset where region selection makes a substantial visual difference to the output. In both cases, without region selection, the output has a lot of color flattening on the skin of the model, leading to a very unnatural and unrealistic output. We therefore conclude that *even very crude selection of the reflection regions results in good reflection removal, and that crude region selection noticeably improves on no region selection*. This justifies our claim of a providing a simple and effective user-interaction scheme.

(E4). We also compare our method to the existent user-interaction by Levin and Weiss [11]. We demonstrate that in comparison, our method produces qualitatively and quantitatively better results, while requiring significantly less effort from the end-user. This underlines one of the main messages of this paper, that *we provide a simple user-interaction method,* *which gives a significant improvement in the quality of the output*.

In Fig. 9 we compare the amount of user interaction required and the quality of the resulting output for both methods. Firstly, in the bottom half of (A-B), the user-interaction for both methods is shown. For our method, the user is asked to determine the location of reflections in the image by marking the rough location in white; several examples of this user-selection are provided in Section IV of the Supplemental Material. In Levin's approach, the user is asked to select foreground gradients in red and background gradients in blue. We can also see the corresponding output of the algorithm, which can be visually observed to be significantly improved using our method.

In Fig. 9 (C) we compare the specific effort of user-interaction between Levin and Weiss [11] and our proposed method. For this we asked a group of 25 colleagues to perform the user-interaction on both schemes and try to achieve the best quality removal as quickly as possible. We observe that, on average, our approach took our colleagues around 5 seconds per image, while Levin's method required around 40 seconds, an increase of around 700%. The corresponding quantitative results can be seen in the upper half of Fig. 9 (C). The numerical values are the metrics averaged over the entire output from 25 users working on the solid-object dataset. In particular each user was given 6 different settings (3 types of focus and 3 types of thickness) of reflections for each of the 20 images in the dataset, and was then asked to perform the user selection for both methods. We see that the similarity metrics are significantly improved using our new method. This shows that our method *requires significantly less effort from the end user than other existent approaches*, while at the same time significantly improving the quality of reflection removal.

V. Conclusion

This paper addresses the challenging problem of single image reflection removal. We propose a technique in which two novelties are introduced to provide reflection removal of higher quality. The first is an *spatially aware prior term, exploiting low-level user interaction*, which tailors reflection suppression to preserve detail in reflection-free areas. The second is an H^2 *fidelity term*, which combines advantages of both L^2 and Laplacian fidelity terms, and promotes better reconstruction of faithful and natural colors. Together, these result in better preservation of structure, detail and color. We demonstrate the potential of our model through quantitative and qualitative analyses, in which it produces better results than all tested model-based approaches and readily competes with recent deep learning techniques. Future work might include the use of deep learning techniques to automatically select regions, which would avoid the need for user interaction, while preserving many of the advantages of our technique.

References

[1] Y. Y. Schechner, N. Kiryati, and R. Basri, "Separation of transparent layers using focus," *Int. J. Comput. Vis.*, vol. 39, no. 1, pp. 25–39, 2000.

[2] A. Agrawal, R. Raskar, S. K. Nayar, and Y. Li, "Removing photography artifacts using gradient projection and flash-exposure sampling," *ACM Trans. Graph.*, vol. 24, no. 3, pp. 828–835, 2005.

[3] N. Kong, Y.-W. Tai, and J. S. Shin, "A physically-based approach to reflection separation: From physical modeling to constrained optimization," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 2, pp. 209–221, Feb. 2014.

[4] A. Lakhtakia, "General schema for the brewster conditions," *Optik*, vol. 90, no. 4, pp. 184–186, 1992.

[5] H. Farid and E. H. Adelson, "Separating reflections and lighting using independent components analysis," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Jun. 1999, pp. 262–267.

[6] N. Kong, Y. W. Tai, and S. Y. Shin, "High-quality reflection separation using polarized images," *IEEE Trans. Image Process.*, vol. 20, no. 12, pp. 3393–3405, Dec. 2011.

[7] R. Szeliski, S. Avidan, and P. Anandan, "Layer extraction from multiple images containing reflections and transparency," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2000, pp. 246–253.

[8] B. Sarel and M. Irani, "Separating transparent layers through layer information exchange," in *Proc. Eur. Conf. Comput. Visioan (ECCV)*, 2004, pp. 328–341.

[9] S. N. Sinha, J. Kopf, M. Goesele, D. Scharstein, and R. Szeliski, "Image-based rendering for scenes with reflections," *ACM Trans. Graph.*, vol. 31, no. 4, pp. 100–101, 2012.

[10] X. Guo, X. Cao, and Y. Ma, "Robust separation of reflection from multiple images," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 2187–2194.

[11] A. Levin and Y. Weiss, "User assisted separation of reflections from a single image using a sparsity prior," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 29, no. 9, pp. 1647–1654, Sep. 2007.

[12] Y. Li and M. S. Brown, "Single image layer separation using relative smoothness," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2014, pp. 2752–2759.

[13] R. Wan, B. Shi, T. A. Hwee, and A. C. Kot, "Depth of field guided reflection removal," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2016, pp. 21–25.

[14] N. Arvanitopoulos, R. Achanta, and S. Süsstrunk, "Single image reflection suppression," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 1752–1760.

[15] H. Barrow, J. Tenenbaum, A. Hanson, and E. Riseman, "Recovering intrinsic scene characteristics," *Comput. Vis. Syst*, vol. 2, p. 3–26, 1978.

[16] Y. Y. Schechner, J. Shamir, and N. Kiryati, "Polarization and statistical analysis of scenes containing a semireflector," *J. Opt. Soc. Amer. A, Opt. Image Sci.*, vol. 17, no. 2, pp. 276–284, 2000.

[17] N. Kong, Y.-W. Tai, and S. Y. Shin, "A physically-based approach to reflection separation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2012, pp. 9–16.

[18] Y. Y. Schechner, N. Kiryati, and J. Shamir, "Blind recovery of transparent and semireflected scenes," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, Jun. 2000, pp. 38–43.

[19] Y. Li and M. S. Brown, "Exploiting reflection change for automatic reflection removal," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2013, pp. 2432–2439.

[20] C. Sun, S. Liu, T. Yang, B. Zeng, Z. Wang, and G. Liu, "Automatic reflection removal using gradient intensity and motion cues," in *Proc. ACM Multimedia Conf.*, 2016, pp. 466–470.

[21] T. Xue, M. Rubinstein, C. Liu, and W. T. Freeman, "A computational approach for obstruction-free photography," *ACM Trans. Graph.*, vol. 34, no. 4, 2015, Art. no. 79.

[22] A. Nandoriya, M. Elgharib, C. Kim, M. Hefeeda, and W. Matusik, "Video reflection removal through spatio-temporal optimization," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 2430–2438.

[23] S. M. Z. A. Shah, S. Marshall, and P. Murray, "Removal of specular reflections from image sequences using feature correspondences," *Mach. Vis. Appl.*, vol. 28, nos. 3–4, pp. 409–420, 2017.

[24] C. Simon and I. K. Park, "Reflection removal for in-vehicle black box videos," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 4231–4239.

[25] J. Y. Cheong, C. Simon, C.-S. Kim, and I. K. Park, "Reflection removal under fast forward camera motion," *IEEE Trans. Image Process.*, vol. 26, no. 12, pp. 6061–6073, Dec. 2017.

[26] B.-J. Han and J.-Y. Sim, "Glass reflection removal using co-saliency-based image alignment and low-rank matrix completion in gradient domain," *IEEE Trans. Image Process.*, vol. 27, no. 10, pp. 4873–4888, Oct. 2018.

[27] Y. Shih, D. Krishnan, F. Durand, and W. T. Freeman, "Reflection removal using ghosting cues," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3193–3201.

[28] Q. Fan, J. Yang, G. Hua, B. Chen, and D. Wipf, "A generic deep architecture for single image reflection removal and image smoothing," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 3258–3267.

[29] X. Zhang, R. Ng, and Q. Chen, "Single image reflection separation with perceptual losses," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2018, pp. 4786–4794.

[30] R. Wan, B. Shi, L.-Y. Duan, A.-H. Tan, and A. C. Kot, "CRRN: Multi-scale guided concurrent reflection removal network," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2018, pp. 4777–4785.

[31] M. Jin, S. Süsstrunk, and P. Favaro, "Learning to see through reflections," in *Proc. IEEE Int. Conf. Comput. Photography (ICCP)*, May 2018, pp. 1–12.

[32] J. Yang, D. Gong, L. Liu, and Q. Shi, "Seeing deeply and bidirectionally: A deep learning approach for single image reflection removal," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2018, pp. 675–691.

[33] L. Xu, C. Lu, Y. Xu, and J. Jia, "Image smoothing via L0 gradient minimization," *ACM Trans. Graph.*, vol. 30, no. 6, 2011, Art. no. 174.

[34] D. Kingma and J. Ba, "Adam: A method for stochastic optimization," in *Proc. Int. Conf. Learn. Represent. (ICLR)*, 2014, pp. 1–15.

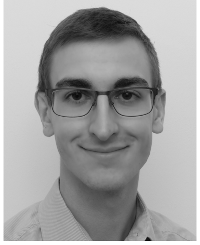
[35] R. Wan, B. Shi, L.-Y. Duan, A.-H. Tan, and A. C. Kot, "Benchmarking single-image reflection removal algorithms," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 3942–3950.

[36] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

[37] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2012, pp. 1097–1105.

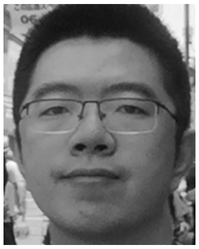
[38] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, p. 436, 2015.

Daniel Heydecker received the B.A. and master's degrees in mathematics from the University of Cambridge in 2017, where he is currently pursuing the Ph.D. degree with the Cambridge Centre for Analysis, Department of Pure Mathematics and Mathematical Statistics (DPMMS), with a focus on the probability of systems with many degrees of freedom.



Georg Maierhofer received the B.A. and master's degrees in mathematics from the University of Cambridge in 2017, where he is currently pursuing the Ph.D. degree with the Cambridge Centre for Analysis, Department of Applied Mathematics and Theoretical Physics (DAMTP). His research interests include approximation theory, numerical integration, hybrid numerical-asymptotic methods, and topics in numerical and applied analysis more broadly.

Angelica I. Aviles-Rivero (GS'18) received the Ph.D. degree from the Polytechnic University of Catalonia, Spain, in 2017. She is currently a Research Associate with the Department of Pure Mathematics and Mathematical Statistics (DPMMS), University of Cambridge, U.K. Her research lies at the intersection of computational mathematics, computer vision, and machine learning for applications to large-scale real-world problems. Her central research is to develop new data-driven algorithmic techniques that allow computers to gain high-level understanding from vast amounts of data, with the aim of aiding the decisions of users from multiple disciplines. This line of research has allowed her to work with a wide range of various data types, including medical imaging, computational photography, computer graphics, and remote sensing, to name a few. She is also an SIAM, SMF, ISMRM, and ACM member.



Qingnan Fan is currently pursuing the Ph.D. degree with Shandong University, China. He is also a Research Intern with the Advanced Innovation Center for Future Visual Entertainment, Beijing Film Academy. His research interests mainly include computational photography, 3D vision, and deep learning.

Dongdong Chen is currently a Researcher with Microsoft Cloud and AI. Before that, he was a joint Ph.D. between his university and Microsoft Asia. His research interests mainly include style transfer, image generation, image restoration, low-level image processing, and object detection.

Carola-Bibiane Schönlieb received the Ph.D. degree from the University of Cambridge in 2009. After one year of postdoctoral activity at the University of Göttingen, Germany, she became a Lecturer at Department of Applied Mathematics and Theoretical Physics (DAMTP) in 2010 and was promoted to Reader in 2015 and Professor in 2018. She is currently a Professor of applied mathematics with the DAMTP, University of Cambridge. She is also the Director of the EPSRC Centre for Mathematical and Statistical Analysis of Multimodal Clinical Imaging, Head of Cambridge Image Analysis Group, and the Director of the Cantab Capital Institute for Mathematics of Information, University of Cambridge. Since 2011, she has been a fellow with the Jesus College, Cambridge. Her current research interests focus on variational methods, partial differential equations and machine learning for image analysis, image processing, and inverse imaging problems.

Sabine Süsstrunk (M'03–SM'09–F'17) leads the Images and Visual Representation Lab (IVRL), EPFL, Switzerland. She has published over 150 scientific papers, of which seven have received the best paper/demos awards. She holds ten patents. Her research areas are in computational photography, color computer vision and color image processing, image quality, and computational aesthetics. She is a fellow of the IS&T. She was a recipient of the IS&T/SPIE 2013 Electronic Imaging Scientist of the Year Award and the IS&T's 2018 Raymond C. Bowman Award.