A Multi-Task Network for Joint Specular Highlight Detection and Removal

Gang Fu¹, Qing Zhang², Lei Zhu³, Ping Li⁴, and Chunxia Xiao¹,*
¹School of Computer Science, Wuhan University, China
²School of Computer Science and Engineering, Sun Yat-sen University, China
³Department of Applied Mathematics and Theoretical Physics, University of Cambridge, UK
⁴Department of Computing, The Hong Kong Polytechnic University, China

Abstract

Specular highlight detection and removal are fundamental and challenging tasks. Although recent methods have achieved promising results on the two tasks by training on synthetic training data in a supervised manner, they are typically solely designed for highlight detection or removal, and their performance usually deteriorates significantly on real-world images. In this paper, we present a novel network that aims to detect and remove highlights from natural images. To remove the domain gap between synthetic training samples and real test images, and support the investigation of learning-based approaches, we first introduce a dataset with about 16K real images, each of which has the corresponding ground truths of highlight detection and removal. Using the presented dataset, we develop a multi-task network for joint highlight detection and removal, based on a new specular highlight image formation model. Experiments on the benchmark datasets and our new dataset show that our approach clearly outperforms state-of-the-art methods for both highlight detection and removal.

1. Introduction

Specular highlight, as a common physical phenomenon in the real world, often presents as bright spot on shiny material surfaces when illuminated. Highlight detection and removal have long been fundamental problems in computer vision. The reason is twofold. First, detecting where the highlight is allows us to infer the light direction, scene geometry [19] and camera location. Second, removing the effect from highlight can help improve the performance of many vision tasks, such as object detection [14], intrinsic image decomposition [2], and tracking [9]. Note that, for simplicity, we refer highlight to specular highlight in this paper, except stated definitely.

Early works detect highlight by treating the brightest pixels in an image as highlights [41, 18], which has low accuracy since it may mistake pixels with high intensities as highlight. As for highlight removal, traditional methods are often based on optimization [13, 20, 7], clustering [29] and filtering [37, 38] etc, and may fail to handle large-scale highlight removal due to lack of image semantics. More recent methods on highlight detection and removal are mostly deep learning-based. Although learning-based highlight detection and removal methods have achieved remarkable progress [28, 39, 6], they basically have two limitations. First, these methods are typically trained on synthetic data or a very small number of real data, so that may...
not work well for real images due to the domain gap between training and test images. Second, they are either designed for highlight detection or removal, and are not able to be employed for joint highlight detection and removal.

To enable effective training and comprehensive evaluation for highlight detection and removal, we in this paper introduce a large-scale dataset for highlight detection and removal. It contains about 16K real images that cover a wide range of scenes, subjects, and lighting conditions. Each image in the dataset has the corresponding highlight detection, removal, and highlight intensity images. Based on the presented dataset and a new region-aware highlight image formation model, we develop a deep learning framework for joint highlight detection and removal. Particularly, a multi-task network with multiple Dilated Spatial Contextual Feature Aggregation (DSCFA) modules is designed to harvest contextual features of different scales, for accurately detecting and removing highlights of varying sizes.

To sum up, our contributions are as follows:

- We present the first large-scale highlight detection and removal dataset with about 16K real images.
- We develop a multi-task network for jointly detecting and removing highlights from natural images.
- Experimental results on benchmark datasets and our new dataset show that our network performs favorably against the previous methods for both highlight detection and removal.

2. Related work

Highlight detection. Most of existing detection methods [18, 22] are typically based on different forms of a thresholding scheme. Based on an assumption that only a small portion of a scene contains highlights, Zhang et al. [41] formulated highlight detection as Non-negative Matrix Factorization (NMF) [41] problem. Although these methods are efficient and easy to implement, they often incorrectly detect pixels with white appearance or high intensities as highlights. To overcome this issue, Fu et al. [6] recently proposed a deep learning-based network leveraging context contrast feature for detection.

Single-image highlight removal. Tan et al. [30] proposed a method based on chromaticity analysis without requiring any geometrical information. Yang et al. [37] utilized low-pass filter to propagate information from the diffuse pixels to the specular pixels. The method proposed by Kim et al. [13] is based on an observation that the dark channel usually provides an approximate highlight-free image. Liu et al. [20] presented a two-step saturation-preserving method, where it first produces an over-saturation highlight-free image, and then corrects the saturation. Akashi et al. [1] formulated the highlight removal problem as a sparse non-negative matrix factorization model. Guo et al. [10] presented a sparse and low-rank reflection model for highlight removal. Shen et al. [27] used the idea of intensity ratio to estimate the specular fractions of the image. Li et al. [16, 17] presented a method for removing specular highlight in facial images. Recently, Shi et al. [28] proposed a deep learning-based method to handle the non-Lambertian object-level intrinsic decomposition problem. Yi et al. [39] designed a unified framework for joint intrinsic image decomposition and highlight separation. However, they fail to handle images with complex illuminations and textures. In comparison, we present a unified framework for jointly learning highlight detection and removal, which can effectively remove highlights while preserving saturation.

Multi-image highlight removal. Some methods are proposed to remove highlight from multiple images. By assuming that surface geometry is known, Wei et al. [35] leveraged the principal component analysis to separate highlights and estimate the position of light source. Feris et al. [4] removed highlight by solving a Poisson equation in the gradient domain. Guo et al. [10] proposed RPCA, which focuses on removing specular reflection from superimposed multiple images. Despite methods in this category produce promising highlight removal results, the requirement of multiple images limits their applicability.

3. Dataset Preparation

3.1. Background

Existing works [37] often use the cross-polarization technique to capture images without highlights in a rigorous laboratory environment. Table 1 reports the number of existing datasets made by this technique. These image pairs are very few and are object-level without background for quality assurance. In fact, it is difficult to control the cross-polarization process in general settings to produce high-quality highlight-free images for everyday objects. Also, it is very difficult and impractical (taking a lot of manpower) to precisely annotate many highlight regions often with dotted and very thin shapes in a real image. In contrast, we propose a semi-automatic method to construct a large-scale real dataset with about 16K Specular Highlight Image Quadruples (SHIQ), by leveraging multi-illumination sequences to produce ground truths.

3.2. Building Dataset

We first discuss how we obtain an initial set of multi-illumination sequences, then describe the pipeline of the task performed on these images and chosen high-quality region pairs. Figure 2 shows the overall workflow of our data creation. In the following, we will describe the steps in detail.

Stage 1: Collecting multi-illumination images. We col-
lect multi-illumination image sequences from the MIW dataset [23], which consists of 1016 scenes, each photographed under 25 predetermined lighting directions for a total of 25,000 high-resolution images. The MIW dataset contains many everyday shiny materials on which characteristically appear highlights.

Stage 2: Obtaining highlight-free images. We choose the state-of-the-art RPCA method proposed by Guo et al. [11] to generate highlight-free images in the MIW dataset.

Stage 3: Screening high-quality regions. As RPCA may fail to produce satisfactory highlight removal results for sequences with complex illumination variations, we thus screen high-quality input/output image pairs for the highlight regions rather than the entire image. To this end, we first randomly cropped each image pair into overlapping image patches of size $k \times k$ with a step size of $l$, where $k$ and $l$ are set as 200 and 50, respectively. In total, we produced 3,197,250 image pairs. Next, each subject was shown with 63K image pairs (i.e., input and its corresponding highlight-free, highlight detection task. Here, elements in the masks are binary values, where “1” indicates highlight regions and “0” indicates non-highlight regions. Finally, (d), (h), (f) and (i) in Figure 2 (i.e., input and its corresponding highlight-free, highlight as well as highlight mask images) are an example quadruples in our dataset. Our dataset can be used simultaneously for highlight detection and removal tasks.

4. Highlight Image Formation

The dichromatic reflection model [26] has been commonly used in the highlight removal field. This model formulates a color image (denoted as $I$) as a linear combination of diffuse component (i.e., highlight-free layer, denoted as $D$) and specular component (i.e., highlight layer, denoted as $S$):

$$I = D + S.$$  \hspace{1cm} (1)

Based on this model, highlight removal could be considered as a two-signal separation problem. This means that given the observation $I$, how to separate the highlight-free layer $D$ from the highlight layer $S$ by relying on their unique characteristics. Existing highlight removal methods based on Equation (1) have two issues. First, the existing methods based on Equation (1) without distinguishing the highlight and non-highlight regions would suffer from the saturation ambiguity problem [13]. Namely, the less saturated surface colors are incorrectly treated as highlights to be removed, leading to color distortion in the non-highlight regions, for example, a white material surface (see Figure 1). Second, highlights in real-world scenes usually have a wide range of intensity values, and have different spatial distributions. So, nearly all traditional optimization-based methods [13, 20] leverage a smoothness prior and do not effectively model $S$ to produce satisfactory results. In essence, the main reason for the above issues lies in the inherent ambiguity of modeling $D$ and $S$.

To address the above issues, we present a generalized
highlight image formation model, expressed as

\[ I = D + M \otimes S, \quad (2) \]

where \( \otimes \) denotes the element-wise multiplication, and \( M \) denotes the highlight mask to indicate the locations of individually visible highlights. The above highlight image formation model has two desirable advantages for learning-based highlight removal methods: (1) it provides additional position information for the network to learn about highlight regions; (2) it allows a new highlight removal framework to first detect highlights, and subsequently to operate differently on the highlight and non-highlight regions, benefiting for producing saturation-preserving results with natural-looking appearances.

5. Joint Highlight Detection and Removal

We design a multi-task network for Joint Specular Highlight Detection and Removal (JSHDR), based on the inverse problem in Equation (2). To accurately detect and remove highlights of varying sizes, we further propose a Dilated Spatial Contextual Feature Aggregation (DSCFA) module.

5.1. Multi-Task Network for Joint Highlight Detection and Removal

According to Equation (2), \( D, S, \) and \( M \) are inherently correlated, and thus computing \( S \) and \( M \) benefits for the estimation of \( D \). Motivated by this observation, we develop a multi-task convolutional neural network with DSCFA modules to jointly predicting \( D, S, \) and \( M \) in an end-to-end manner. Figure 5 shows the schematic illustration of the developed network. As an encoder-decoder framework, our network first passes an input image into a series of DSCFA modules (see Section 5.2) of an encoder and decoder framework to extract highlight features \( F \). Then, \( M, S \) and \( D \) are predicted in a sequential order from \( F \):

1. \( M \) is estimated by using a convolutional block with “\( \text{Conv}(3 \times 3) \rightarrow \text{Conv}(3 \times 3) \rightarrow \text{Conv}(3 \times 3) \)” on \( F \);
2. \( S \) is estimated by applying another convolutional block with three \( 3 \times 3 \) convolutions on the concatenation of \( [F, M] \);
3. \( D \) is estimated by feeding the concatenation of \( [F, M, S, I-MS] \) into a convolutional block consisting of three \( 3 \times 3 \) convolutions.

5.2. Dilated Spatial Contextual Feature Aggregation Module

Figure 6 shows the schematic illustration of the proposed DSCFA module, which extracts and aggregates multi-scale dilated spatial contextual information simultaneously for detecting and removing highlights at varied region sizes by developing a series of DSCFA blocks.

DSCFA block. Contextual information has been demonstrated to be useful for highlight detection [6] and low-level image processing [3]. Our DSCFA block learns dilated spatial features (DSF) to extract and aggregate dilated contextual features from four directions. To achieve this, we first replace the common convolution of the spatial CNN module [24] with dilated convolutions to enlarge the receptive field for more contextual information, and then obtain four features (i.e., DSCNN_L, DSCNN_R, DSCNN_D, DSCNN_U) along with four directions. Figure 7 shows a dilated spatial module, which learns a DSCNN_D, DSCNN_U, DSCNN_L, and DSCNN_R from input features. Spatial aggregation from four directions adopt slice-by-slice convolutions within feature maps from downward, upward, rightward, and leftward directions, thus enabling rich message passing between pixels across rows and columns in a layer. Then, given an input feature map, we first apply a \( 3 \times 3 \) convolution and a ReLU Layer to produce a new feature map \( H \). Based on \( H \), we apply two branches to learn contextual features. The first branch is the whole procedure of Figure 7 while the second branch is to replace the order of feature learning at four directions, where DSCNN_L, DSCNN_R,
Figure 5. The pipeline of our joint highlight detection and removal network. Our network applies an encoder-decoder structure with DSCFA modules to extract highlight features $F$ from the input highlight image $I$. Based on $F$, $M$, $S$ and $D$ are then subsequently predicted to perform the joint highlight detection, estimation and removal.

Figure 6. The schematic illustration of the proposed DSCFA module. The input features are pass through four parallel DSCFA blocks, and the output features of four DSCFA blocks are performed weighted fusion to produce multi-scale dilated spatial contextual features. In each DSCFA block (dark red dashed box), input features are fed to two parallel dilated spatial convolutions with opposite convolution orders to obtain abundant contextual infromation with different highlight characteristics.

Figure 7. Illustrations of a DSCFA block. The DSCNN module with suffix ‘D’, ‘U’, ‘R’, and ‘L’ indicates DSCNN with downward, upward, rightward, and leftward directions respectively.

DSCNN_D and DSCNN_U are obtained from $H$. After that, we concatenate features from two branches and apply a $3 \times 3$ convolution and a ReLU layer to obtain the output features of our DSCFA block; as shown in Figure 6.

**DSCFA module.** Highlight regions in an image often have a wide range of region sizes. Figure 4 shows an example, where highlights of the input image cover from a dotted region to a very long thin shape across a whole object. Note that the DSCFA block with a given dilation rate $n$ tends to extract contextual information from a receptive field of a fixed size. Hence, it suffers from two issues. First, the receptive fields are maybe larger to detect small highlight regions, thereby incurring a false positive result due to much incurred noise. Second, the receptive fields are sometimes small for inferring a large highlight region. As a result, the target highlight regions are partially detected and cannot be removed completely from the input highlight image due to insufficient contextual information to eliminate it. To overcome this issue, we devise a multi-scale contextual module, namely DSCFA module, to harvest contextual information from multiple receptive fields at varied scales. In detail, as shown in Figure 6, we feed the input features into four parallel DSCFA blocks to obtain four DS features and then use a convolutional block with “Conv(3×3) → ReLU → Conv(3×3) → ReLU → Conv(1×1)” to learn an attention map with four channels to weight features from four...
DSF estimation, and highlight removal. The definition of highlight detection, highlight intensity prediction losses of our DSCFA module:

\[
\mathcal{L} = \lambda_s \Phi_{L2}(\mathbf{S}, \mathbf{S}) + \lambda_d \Phi_{L2}(\mathbf{D}, \mathbf{D}) + \lambda_m \Phi_{BCE}(\mathbf{M}, \mathbf{M}),
\]

where \(\lambda_s\), \(\lambda_d\), and \(\lambda_m\) are the weighting parameters, we empirically set them as \(\lambda_s = \lambda_d = \lambda_m = 1\) for all experiments. \(\mathbf{M}\) and \(\mathbf{M}\) are the predicted highlight detection map and the associate ground truth. \(\mathbf{S}\) and \(\mathbf{S}\) are the predicted highlight intensity map and the associate ground truth. \(\mathbf{D}\) and \(\mathbf{D}\) are the predicted highlight removal result and the associate ground truth.

5.3. Network Training

The total training loss \(\mathcal{L}\) of our network consists of three prediction losses of highlight detection, highlight intensity estimation, and highlight removal. The definition of \(\mathcal{L}\) is given by:

\[
\mathcal{L} = \lambda_s \Phi_{L2}(\mathbf{S}, \mathbf{S}) + \lambda_d \Phi_{L2}(\mathbf{D}, \mathbf{D}) + \lambda_m \Phi_{BCE}(\mathbf{M}, \mathbf{M}),
\]

where \(\lambda_s\), \(\lambda_d\) and \(\lambda_m\) are the weighting parameters, we empirically set them as \(\lambda_s = \lambda_d = \lambda_m = 1\) for all experiments. \(\mathbf{M}\) and \(\mathbf{M}\) are the predicted highlight detection map and the associate ground truth. \(\mathbf{S}\) and \(\mathbf{S}\) are the predicted highlight intensity map and the associate ground truth. \(\mathbf{D}\) and \(\mathbf{D}\) are the predicted highlight removal result and the associate ground truth.

5.4. Implementation details

We randomly split our dataset into 12K quadruples for training and 4K quadruples testing. We have implemented JHSDR in PyTorch on a PC equipped with NVIDIA GeForce GTX 2080Ti. The Adam optimizer [15] is used to train our network. We use 100 epochs to train our network with batch size of 8, and the whole training process requires about 3 days. The initial learning rate is \(2 \times 10^{-5}\) and we then divide the learning rate by 10 after 20 epochs. We also use Huawei MindSpore platform to partly verify our JHSDR. What’s more, we adopt the highlight attenuation and boosting editing [25] as a data augmentation to produce more training images with diverse highlights, enabling our network to better address weak and strong highlights.

6. Experiments

6.1. Datasets

Since our dataset SHIQ has the associate ground truths of the highlight detection and highlight removal, we use SHIQ to evaluate highlight detection and removal, respectively. Apart from our SHIQ, we also include a common SRW [6] to evaluate different highlight detection methods (available at https://github.com/fu123456/SHDNet). Regarding highlight removal, we first create a dataset (namely CLH) by collecting testing images from existing highlight removal works [30, 27, 37, 39], and introduce an existing synthesized LIME dataset [21]. Overall, we use SRW and our SHIQ as two datasets for highlight detection evaluation, while CLH, LIME, and our SHIQ are considered as three datasets for highlight removal evaluation.

6.2. Compared methods and Metrics

Highlight detection. We compare our method with the state-of-the-art highlight detection methods including two traditional methods of NMF [41] and ATA [18], and a deep learning-based method of SHDN [6]. To evaluate the highlight detection performance quantitatively, we follow two commonly used metric terms [12] including the accuracy and the balance error rate (BER). Higher value of accuracy and lower value of BER indicate better detection results.

Highlight removal. We compare our method against traditional methods [27, 37, 30, 1, 28, 10, 36] and recent CNN-based methods [28, 39]. PSNR and SSIM [34] are adopted to quantitatively compare different highlight removal methods. In general, larger PSNR and SSIM scores indicate better removal results.
Figure 10. Visual comparison of our method against state-of-the-art highlight removal methods on real-world images from the Internet.

(a) Input (b) Shen [27] (c) Yang [37] (d) Akashi [1] (e) Yama. [36] (f) Tan [30] (g) Guo [10] (h) Shi [28] (i) Yi [39] (j) Ours

Figure 11. Rating percentage distribution in the user study.

6.3. Comparison with SOTA Highlight Detectors

Quantitative comparison. Table 2 reports the accuracy and BER scores of different highlight detectors on SRW [6] and our collected dataset SHIQ. Apparently, our method achieves the best quantitative results on the accuracy and BER metrics, demonstrating that our network can more accurately detect highlight regions.

Visual comparison. Figure 9 visually compares highlight detection maps produced by our network and state-of-the-art methods. Specifically, the traditional methods [18, 41] often wrongly detect white text texture as highlight, since they fail to semantically distinguish highlight regions from white material surfaces. In addition, the method proposed in [6] fails to locate weak highlight regions. By contrast, our method can more accurately locate spatially-varying highlight regions, and our results are more consistent with the ground truths.

6.4. Comparison with SOTA Highlight Removal Methods

Quantitative comparison. Table 2 reports PSNR and SSIM values of different methods on three datasets (i.e., our SHIQ, CLH, and LIME). It shows that our network has larger PSNR and SSIM scores than all the compared methods.

Visual comparison. In Figure 8, we show the highlight removal results of different methods on our dataset. From the results, we can see that traditional methods based on optimizations and color analysis either fail to effectively remove highlight, or produce color/shading distortion. Regarding two CNN-based methods, Shi et al. [28] suffers from the over-despecular problem and thus is not able to preserve shading information, thereby resulting in unnatural-looking results. Yi et al. [39] tends to leave parts of highlight in the highlight removal results. Fortunately, our method can produce high-quality results, which are considerably similar to the ground truths.

User study. To further evaluate the performance of our network on real-world highlight images, we collect 200 images and conduct a user study to evaluate the results of different methods. Specifically, we download 200 test images from Pinterest by searching with keywords like jade, sculpture and mask. Then, we test all methods on 200 collected images to produce their results of estimating underlying highlight-free counterparts, and recruit 20 students to rate different results, which are listed for rating in a random
Ours has a superior PSNR and SSIM performance over \( M_1, M_2, \) and \( M_3 \), showing that progressively adding the number of DSCFA blocks improves the highlight removal performance, and our network with four DSCFA blocks has the best performance. Figure 12 shows visual results of our network, \( M_4, M_5, \) and \( M_6 \), showing that our network has the best performance of highlight removal. (3) Our network has a better PSNR and SSIM performance over \( M_7 \), showing that the effectiveness of DSCFA over a basic convolution. In the nutshell, it indicates that with the help of the designed multi-task scheme and the DSCFA module, our method can effectively remove highlight while preserving shading/saturation very well, thereby leading to natural-looking results.

**Limitations.** Our method has two limitations. First, our method is less able to effectively remove highlights in colored lighting. Figure 13 presents two examples where our method is less able to effectively remove highlights in colored lighting. (b) and (c) are two composited color versions of (a).

Table 4 lists the PSNR and SSIM results of our network and seven baselines. From the results, we have the following observations: (1) Our network has a superior PSNR and SSIM performance over \( M_1, M_2, \) and \( M_3 \), showing that additional highlight detection and estimation in the multi-task learning of our network help our network to better recover the highlight-free results. (2) The higher PSNR and SSIM scores of our method than \( M_4 \) and \( M_5 \), and \( M_6 \) demonstrate that progressively adding the number of DSCFA blocks improves the highlight removal performance, and our network with four DSCFA blocks has the best performance. Figure 12 shows visual results of our network, \( M_4, M_5, \) and \( M_6 \), showing that our network has the best performance of highlight removal.

2. Our method has two limitations. First, our method is less able to effectively remove highlights in colored lighting. Second, our method and even state-of-the-art inpainting methods are not able to recover text textures since there are no meaningful and reliable contextual cues to help restore them.

6.5. Discussions

**Ablation study.** To validate the effectiveness of the major components of our network, we accordingly modify our network to construct seven baselines:

- \( M_1 \): JSHDR w/o highlight detection.
- \( M_2 \): JSHDR w/o highlight detection and estimation.
- \( M_3 \): JSHDR with only highlight estimation.
- \( M_4 \): use only one DSCFA block in the DSCFA module of our JSHDR.
- \( M_5 \): use two DSCFA blocks in the DSCFA module of our JSHDR.
- \( M_6 \): use three DSCFA blocks in the DSCFA module of our JSHDR.
- \( M_7 \): replace DSCFA modules of our JSHDR with simple convolution operations.

Table 4 lists the PSNR and SSIM results of our network and seven baselines. From the results, we have the following observations: (1) Our network has a superior PSNR and SSIM performance over \( M_1, M_2, \) and \( M_3 \), showing that additional highlight detection and estimation in the multi-task learning of our network help our network to better recover the highlight-free results. (2) The higher PSNR and SSIM scores of our method than \( M_4 \) and \( M_5 \), and \( M_6 \) demonstrate that progressively adding the number of DSCFA blocks improves the highlight removal performance, and our network with four DSCFA blocks has the best performance. Figure 12 shows visual results of our network, \( M_4, M_5, \) and \( M_6 \), showing that our network has the best performance of highlight removal. (3) Our network has a better PSNR and SSIM performance over \( M_7 \), showing that the effectiveness of DSCFA over a basic convolution. In the nutshell, it indicates that with the help of the designed multi-task scheme and the DSCFA module, our method can effectively remove highlight while preserving shading/saturation very well, thereby leading to natural-looking results.

7. Conclusion

In this paper, we have presented a large-scale real dataset with about 16K image quadruples that covers a diversity of real-world highlight scenes. Besides, based on the classic dichromatic reflection model, a new region-dependent highlight image formation model is proposed for highlight detection, which provides useful information for highlight removal. Based on this model, we proposed a multi-task convolution network for joint highlight detection and removal. Extensive experiments illustrate that in comparison to previous methods, our method can effectively handle spatially-varying highlights, while preserving shading well.

In the future, we will incorporate an illumination color estimation module into our network and extend our method to handling colored-illumination scenes. Moreover, we will take our method as a pre-processing step for intrinsic image decomposition [5, 8] and recoloring [40], and take our DSCFA as a common feature extraction module for deraining [32, 33] and dehazing [31].

Acknowledgments

We would like to thank Wenxi Gan and Siyuan Yang for their hard work in building the dataset. This work was partly supported by the National Key Research and Development Program of China (2017YFB10002600), the Key Technological Innovation Projects of Hubei Province (2018AAA062), NSFC (NO. 61972298), and CAAI-Huawei MindSpore Open Fund.
References

[18] Ranyang Li, Junjun Pan, Yaqing Si, Bin Yan, Yong Hu, and Hong Qin. Specular reflections removal for endoscopic image sequences with adaptive-rpca decomposition. IEEE Transactions on Medical Imaging, 39(2):328–340, 2019. 1, 2, 6, 7


