



Highlight Removal from A Single Grayscale Image Using Attentive GAN

Haitao Xu, Qiang Li & Jing Chen

To cite this article: Haitao Xu, Qiang Li & Jing Chen (2022) Highlight Removal from A Single Grayscale Image Using Attentive GAN, Applied Artificial Intelligence, 36:1, 1988441, DOI: [10.1080/08839514.2021.1988441](https://doi.org/10.1080/08839514.2021.1988441)

To link to this article: <https://doi.org/10.1080/08839514.2021.1988441>



© 2022 The Author(s). Published with license by Taylor & Francis Group, LLC.



Published online: 12 Mar 2022.



Submit your article to this journal [↗](#)



Article views: 1251




View related articles [↗](#)



View Crossmark data [↗](#)

Highlight Removal from A Single Grayscale Image Using Attentive GAN

Haitao Xu, Qiang Li, and Jing Chen 

School of Computer Science, Hangzhou Dianzi University, Hangzhou, China

ABSTRACT

The existence of specular highlights hinders high-level computer algorithms. In this paper, we propose a novel approach to remove specular highlights from a single grayscale image by regarding the problem as an image-to-image translation task between the highlight domain and the diffuse domain. We solve this problem by using the generative adversarial network framework, where a generator removes highlights and discriminator judges whether outputs of the generator are clear and highlight-free. Specular highlight removal is intractable as we should remove specular highlights while keeping as many details as possible. Considering the similarity between the highlight image and diffuse image, we adopt an attention-based submodule that generates a mask image, which we call the highlight intensity mask, to locate pixels that contain specular highlights and help the skip-connected autoencoder to remove highlights. A pixel discriminator and Structural Similarity loss are utilized to ensure that more details can be retained in the output images. For training and testing models, we build a grayscale highlight images dataset. It consists of more than a thousand sets of grayscale highlight images with ground truth. Finally, quantitative and qualitative evaluations demonstrate the effectiveness of our method than other contrast generative adversarial network methods.

ARTICLE HISTORY

Received 31 October 2020
Accepted 29 September 2021

Introduction

Most current computer vision algorithms assume the surface of an object is fully diffuse. But the reality is that specular highlights are widely present in real-world objects and specular highlights create obstacles to high-level algorithms. Industrial metal parts are more likely to show specular highlights due to their materials. Therefore, the research on the grayscale image highlight removal of metal parts is of great significance.

According to the dichromatic reflection model (Shafer 1985), the observable intensity $I(p)$ of any pixel p is formed by the linear superposition of the specular reflection component $I_s(p)$ and the diffuse reflection component $I_d(p)$ as:

CONTACT Jing Chen  jingchen.hdu@gmail.com  School of Computer Science, Hangzhou Dianzi University, Hangzhou, Zhejiang, China

© 2022 The Author(s). Published with license by Taylor & Francis Group, LLC.
This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

$$I(p) = I_s(p) + I_d(p) \quad (1)$$

The purpose of specular highlight removal is to separate two intrinsic images from the highlight image, a diffuse image, and a specular image. According to the number of input images, the existing methods can be divided into two categories: single-image-based methods and multiple-image-based methods, but they both have certain limitations (Artusi, Banterle, and Chetverikov 2011). The generative adversarial network (GAN)(Goodfellow et al. 2014) has achieved great success in the field of image synthesis (Huang, Yu, and Wang 2018). Although it is difficult to obtain satisfactory results by applying those GAN models directly to specular highlight removal, it provides a new approach to this problem. That is, we can treat specular highlight removal as an image-to-image translation between the highlight domain and diffuse domain, and then employ conditional GAN methods to solve the problem. Inspired by this, we proposed a novel approach using the generative adversarial network, in which the generator produces the image without specular highlights, and the discriminator determines whether the image contains specular highlights.

We noticed that there are distinctive features within the specular highlight removal. Compared to other image translation tasks, the highlight image and diffuse image show a high degree of similarity as shown in Figure 1. The vast majority of the pixels in them are identical, differing only in a few highlighted areas. The similarity leads to poor results in adversarial training. To address the problem caused by this similarity, we propose to employ an attention module in front of the autoencoder. The attention module is used to predict the specular reflection intensity of

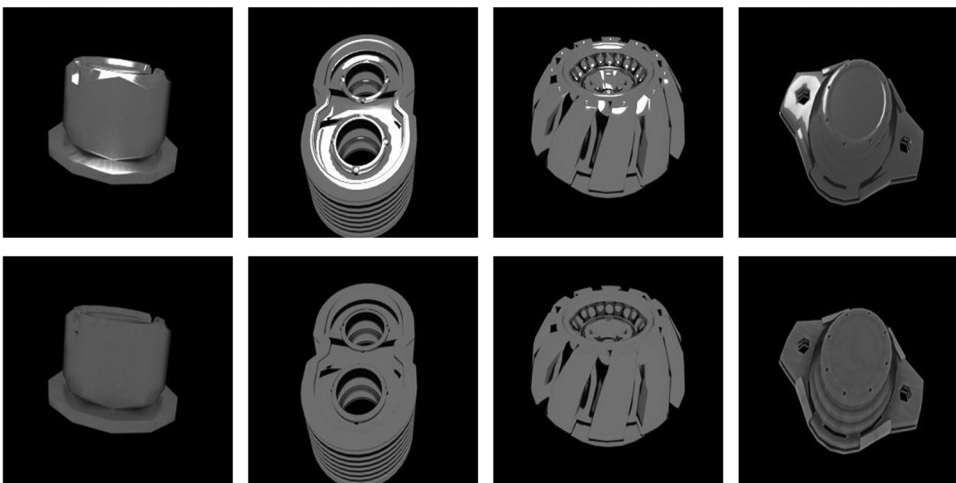


Figure 1. Demonstration of images in specular highlight removal. Top: highlight images, bottom: diffuse images. We need to restore the diffuse image from the highlight image.

each pixel in the highlight image, which is called the highlight intensity mask. It serves as effective auxiliary information to help to remove specular highlights. Skip-connections are exploited in the autoencoder so that some of the information in the highlight image can be conducted directly to the symmetric convolutional layer. For the discriminative network, we apply a pixel discriminator to prevent the accuracy of being affected by excessive receptive fields. The similarity property also causes the MSE between images to always be quite close to zero, reducing the discrimination of MSE as a measure of image similarity, so we adopt Structural Similarity (SSIM) (Z. Wang et al. 2004) as an additional content loss. These targeted designs allow our proposed method to greatly improve the retention of details in the single grayscale image highlight removal problem.

In summary, with the framework of GAN, we propose a novel end-to-end network for the single grayscale image highlight removal. The model consists of a generator with skip-connections and a pixel discriminator. Considering the similarity between the highlight image and the diffuse image, we propose to locate highlight by a highlight intensity mask image, which is generated by an attention module, and use SSIM loss and MSE loss as a combined content loss to train the network. With those features, specular highlights are removed effectively from grayscale images, and details are kept in outputs.

The rest of this paper is organized as follows: Section 2 discusses the related work in the fields of specular highlight removal, and the fields of image applications based on generative adversarial networks. Section 3 describes our proposed model in detail, including both the network structure and the objective function. Section 4 shows our synthetic dataset and training details. Section 5 discusses the results of training and experiments. Finally, Section 6 concludes our paper.

Related Work

Our work involves two topics: specular highlight removal and generative adversarial networks which are briefly discussed in this section.

Specular Highlight Removal

According to the number of input images, specular highlight removal methods can be divided into two main categories: single-image-based method and multiple-image-based method.

Single-image-based Method

The specular highlight removal method based on a single image aims to remove the highlights by only one input image. Klinker et al. analyzed the distribution of pixels and found that diffuse reflection and specular highlights presented a T-shaped distribution in the RGB color space, and achieve the removal of specular highlights (Klinker, Shafer, and Kanade 1988). Later, other scholars extended the color space analysis to UV-space (Schlüns and Teschner 1995) and S-space (Bajcsy, Lee, and Leonardis 1996) based on Klinker's idea. Koirala et al. exploited principal component analysis and histogram equalization method to remove specular highlights quickly (Koirala, Hauta-Kasari, and Parkkinen 2009). Shen et al. discovered the intensity ratio between the maximum values and range values (maximum minus minimums) is independent of surface geometry and applied this rule to achieve real-time highlight removal (Shen and Zheng 2013). Guo et al. proposed a sparse and low-rank reflection model for specular highlight detection and removal with a single input image (Guo, Zhou, and Wang 2018). As this problem is inherently ill-posed, prior knowledge or assumptions on the characteristics of natural images should be exploited to make the problem tractable. Although methods based on prior knowledge have achieved good results, such methods do not always achieve satisfactory solutions in an unconstrained environment.

Multiple-image-based Method

The diffuse reflection does not shift with the viewing angle and the relative position of the light source. According to this characteristic, it is natural to use image sequences from different points of view or multiple light positions to restore the diffuse reflection (Feris et al. 2004; Li and Ma 2006). Woff et al. removed specular highlights by multiple images based on polarization (Wolff and Boulton 1993). Sato et al. made additional use of time-dimensional information to separate the specular highlight component (Sato and Ikeuchi 1994). Xu et al. and Wang et al. utilized light field cameras to assist in removing specular highlights, as they provided depth information (H. Wang et al. 2016; Xu et al. 2015). Wei et al. estimated the angle of the incident light and diffuse reflection component when the geometry of the object is known, and then removed the specular highlights (Wei et al. 2018). These representative methods can achieve some good performance in the removal of specular highlights, but they are dependent on the external environment to control the angle and quantity of light source or need special equipment, such as a light field camera. This requirement greatly limits the usage scenarios of multiple-image-based methods.

Generative Adversarial Network

Since the generative adversarial network (Goodfellow et al. 2014) was proposed in 2014, it has achieved great success in the fields of deep learning (Gui et al. 2020). There are many applications of GAN in computer vision, such as image inpainting (Yeh et al. 2016; Yu et al. 2018), super-resolution (Ding et al. 2019; Xintao Wang et al. 2018b), and image translation (Isola et al. 2017; Zhu et al. 2017). Inspired by the success of GAN in image translation, we utilize it to implement the specular highlight removal. Some scholars have made related attempts before. John Lin et al. used a multi-class discriminator to train the generator to remove the specular highlights from color images (Lin et al. 2019). Funke et al. employed GAN in the endoscope highlights removal (Funke et al. 2018). In general, the research on the combination of GAN and specular highlight removal is still at the initial stage. As discussed in the introduction, this research topic is of great research value, so we propose a GAN-based method to remove specular highlights from a single grayscale image in this paper.

Proposed Method

Overview

In this paper, we define specular highlight removal as an image-to-image translation between the highlight domain and the diffuse domain as shown in Figure 2. The highlight domain is a collection of images with specular highlights and the diffuse domain is composed of diffuse images. With this definition, an input image can be regarded as a random sample from the highlight domain, and a corresponding diffuse image can be found in the diffuse domain. We propose to adopt a generative adversarial network to solve this image translation problem.

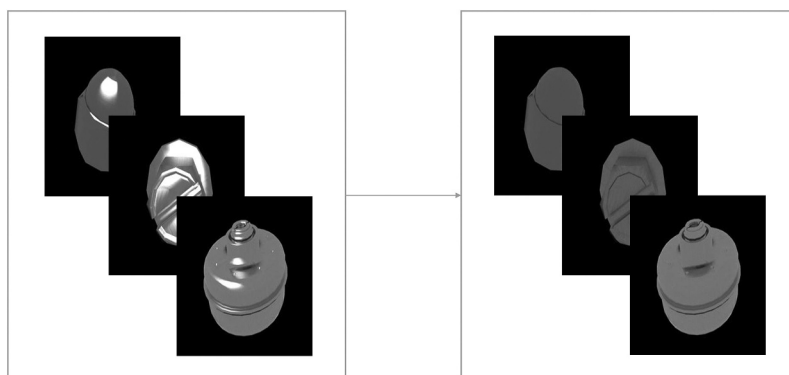


Figure 2. The image-to-image translation in specular highlight removal. The left is the highlight domain, and the right is the diffuse domain.

The structure of our proposed network is shown in [Figure 3](#). It follows the framework of GAN, the network is divided into two sub-networks, a generator, and a discriminator. The generator learns the mapping from the highlight domain to the diffuse domain from the data to remove specular highlights. A discriminator is introduced to determine the way to distinguish the output of the generator from the ground truth (GT). The generator is trained alternately with the discriminator to reach a Nash equilibrium. The process can be expressed as a min-max optimization problem:

$$\min_G \max_D \mathbb{E}_{DI \sim P_{DD}} [\log(D(DI))] + \mathbb{E}_{II \sim P_{HD}} [\log(1 - D(G(II)))] \quad (2)$$

where G stands for the generator and D stands for the discriminator. Input image II and diffuse image DI denote randomly samples from the highlight domain HD and the diffuse domain DD . The structure generator and discriminator will be described in the next part of this section.

The Generator

As shown in [Figure 3](#), the generator is divided into two submodules, one is the attention module and the other is the skip-connected autoencoder.

Attention Module

The attention is widely used in computer vision to locate regions of interest for better feature extraction (Gregor et al. 2015; Zhao et al. 2017). In the same way, our attention module is used to predict the location and intensity of pixels containing specular highlights in the input image and produce the highlight intensity mask as shown in [Figure 4](#). The highlight intensity mask is a single-channel image of the same size as the input, and each pixel in the mask has

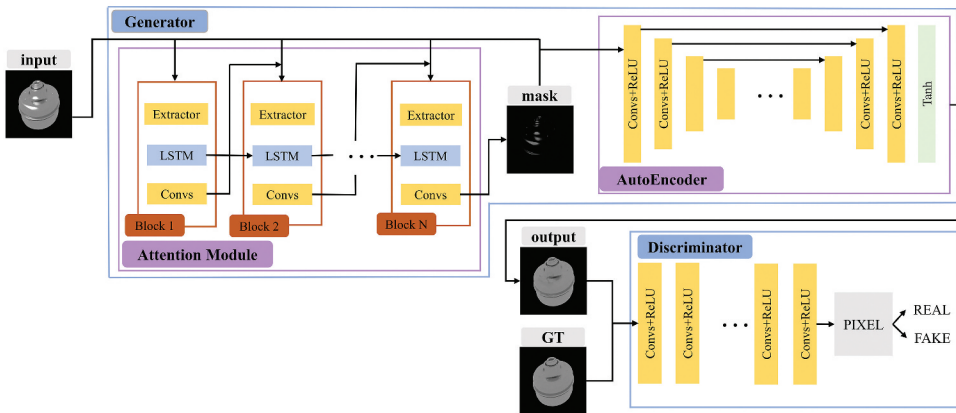


Figure 3. The overall structure of our network. The generator consists of an attention module and autoencoder with skip connections. The discriminator is formed by a series of convolution layers.

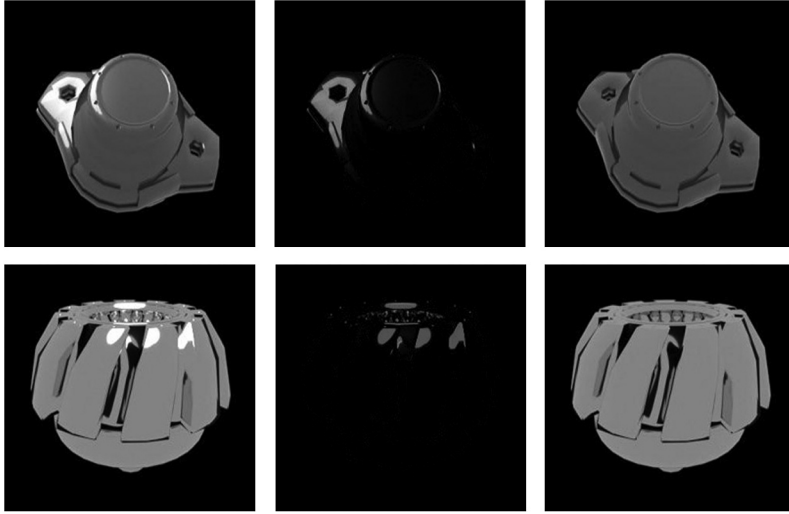


Figure 4. Sample of highlight intensity mask. From left to right are highlight image, highlight intensity mask, and diffuse image which is ground truth.

a value in the range of 0 to 1, which signifies the reflection intensity at the corresponding location of the input. A higher value means a stronger reflection. We consider the mask can help the autoencoder to remove highlights as it suggests the pixels that the autoencoder should focus on by different attention values.

The attention module consists of N recursive blocks as shown in [Figure 3](#). [Figure 5](#) shows the internal structure of each recursive block. There are four residual layers (He et al. 2016) to extract features first in each block. Those short-cut connections in the residual layers ensure that the semantic information is preserved. Next, is the convolutional LSTM unit (Qian et al. 2018), and the last convolution layer is used to adjust the channel to produce the highlight intensity mask.

The convolutional LSTM unit contains an input gate i_t , an output gate o_t , a forget gate f_t , and a cell state C_t , where the t represents time. The relationship between them and the input can be defined as follows:

$$\begin{aligned}
 i_t &= \sigma(W_{xi} \circ X_t + W_{hi} \circ H_{t-1} + W_{ci} * C_{t-1} + b_i) \\
 f_t &= \sigma(W_{xf} \circ X_t + W_{hf} \circ H_{t-1} + W_{cf} * C_{t-1} + b_f) \\
 C_t &= f_t * C_{t-1} + i_t * \tanh(W_{xc} \circ X_t + W_{hc} \circ H_{t-1} + b_c) \\
 o_t &= \sigma(W_{xo} \circ X_t + W_{ho} \circ H_{t-1} + W_{co} * C_{t-1} + b_o) \\
 H_t &= o_t * \tanh(C_t)
 \end{aligned} \tag{3}$$

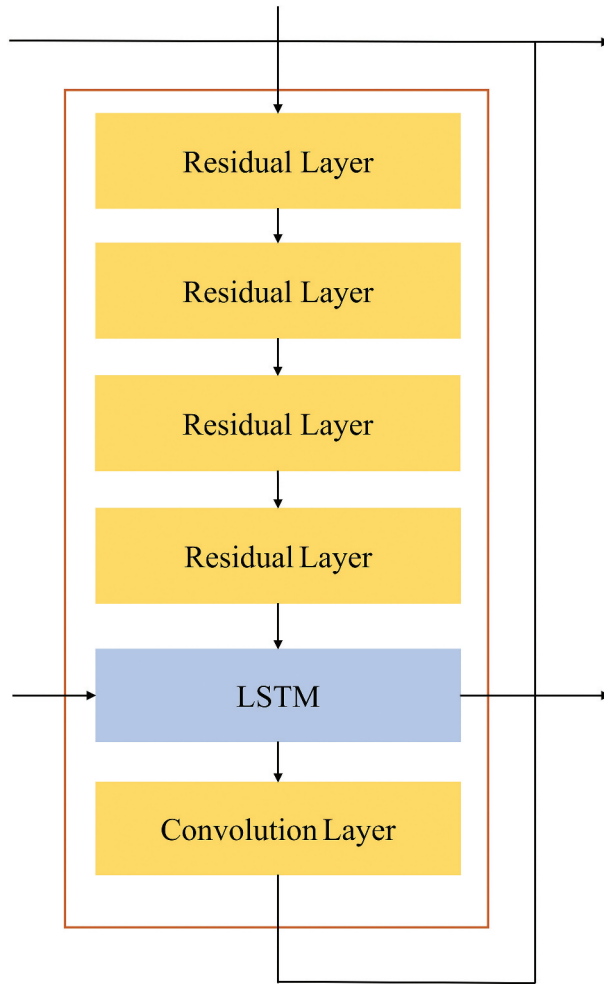


Figure 5. The internal structure of a recursive block in the attention module.

where \circ represents the convolution operation. W and b represent the weights and biases, respectively. X_t represents the features extracted by the residual layer, and H_t represents the output from the previous LSTM unit. The features are fed into the convolutional layers to generate the highlight intensity mask. Each recursive block accepts the input image and the tensor from the previous recursive block as input and outputs a new highlight intensity mask. We initialize the highlight intensity mask to the full 0.5 as the input of the first block and select the output of the last block as the final highlight intensity mask.

Autoencoder

The purpose of our autoencoder is to generate images that are free from specular highlights. The input of the autoencoder is the concatenation of the input highlight image and the mask generated by the last recursive block. We

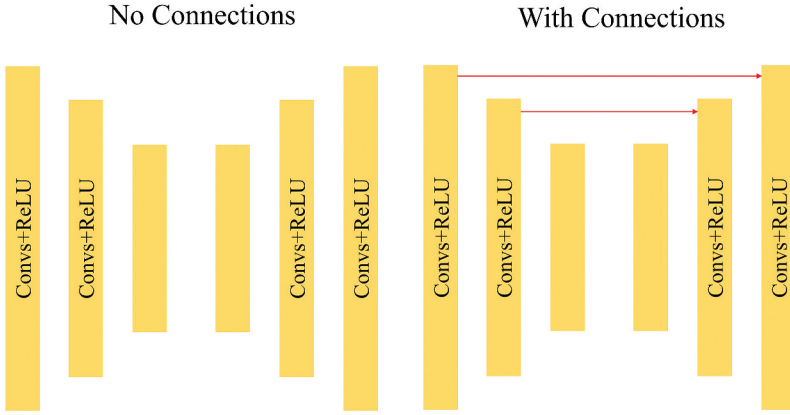


Figure 6. Two choices for the architecture of the autoencoder.

conceive that the diffuse image should restore details from the input image as many as possible. That means information of intensity and structure in the input image needs to be preserved to the output image. For this reason, we adopt a network similar to the U-net (Long, Shelhamer, and Darrell 2015) whose most notable characteristic is the skip connections as shown in Figure 6.

The input image is first passed through a series of down-sampling convolution layers until a bottleneck layer. Then it returns to the original size and channels by up-sampling deconvolution layers. Low-level information can be shared by those skip connections which have been proved to have excellent results in image generation applications (Guan et al. 2020; Chengjia Wang et al. 2018a).

Objective Function

The objective function of the generator consists of four parts: The Mean Square Error (MSE) loss \mathcal{L}_{MSE} , the Structural Similarity (SSIM) loss \mathcal{L}_{SSIM} , the attention module loss \mathcal{L}_{ATT} , and the adversarial loss \mathcal{L}_{GAN} . The λ_* represents the scale factor of different parts. We define the loss function as:

$$\begin{aligned} \mathcal{L}_G = & \lambda_{MSE} \cdot \mathcal{L}_{MSE}(O, T) + \lambda_{SSIM} \cdot \mathcal{L}_{SSIM}(O, T) + \lambda_{ATT} \cdot \mathcal{L}_{ATT}(\{M\}_1^N, MT) \\ & + \mathcal{L}_{GAN}(O) \end{aligned} \quad (4)$$

where O stands for the diffuse image generated by the generator. T indicates the ground truth of the diffuse image. $\{M\}_1^N$ represents the highlight intensity mask generated by 1 to N recursive blocks. MT represents the ground truth of the highlight intensity mask, which is obtained by subtracting the specular highlight image from the corresponding diffuse image and normalized.

The MSE loss \mathcal{L}_{MSE} is widely used to measure the similarity between generated image and ground truth. MSE loss ensures that images generated by the generator are close enough to the ground truth. \mathcal{L}_{MSE} is denoted as:

$$\mathcal{L}_{MSE}(O, T) = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [O(i, j) - T(i, j)]^2 \quad (5)$$

where m and n represent the width and height of the input image, respectively.

The MSE loss converges to zero so quickly that the term no longer contributes to backpropagation. To solve this problem, we introduce the SSIM as the second content loss, which is defined as Equation 6. The inclusion of SSIM loss allows the autoencoder to take not only the loss of the average pixel values, but also information such as contrast, brightness, and structure into account. Experiments demonstrate that we obtain clearer results after using SSIM as a loss term.

$$\mathcal{L}_{SSIM}(O, T) = 1 - \frac{(2\mu_O\mu_T + C_1)(2\sigma_{OT} + C_2)}{(\mu_O^2 + \mu_T^2 + C_1)(\sigma_O^2 + \sigma_T^2 + C_2)} \quad (6)$$

where μ_O and μ_T indicates the mean value of all pixels in the selected window, σ_O and σ_T indicates the variance of the pixel values, σ_{OT} represents covariance between O and T , and C_1 and C_2 are hyper-parameters.

The loss of the attention module \mathcal{L}_{ATT} is defined as Equation 7. Actually, \mathcal{L}_{ATT} calculates the MSE between the highlight intensity mask of each recursive block and the ground truth, where θ denotes the weight.

$$\mathcal{L}_{ATT}(\{M\}_1^N, MT) = \sum_{i=1}^N \theta^{N-i} \mathcal{L}_{MSE}(M_i, MT) \quad (7)$$

Adversarial loss \mathcal{L}_{GAN} is determined by the result of the discriminator and can be written as:

$$\mathcal{L}_{GAN}(O) = \log(1 - D(O)) \quad (8)$$

The Discriminator

The discriminator is employed to distinguish the diffuse image generated by the generator from the ground truth. Discriminators in some GANs compress the features to a number or an N -dimensional matrix (N is much smaller than the original size of the image). Those discriminators include a relatively large receptive field, which will make discriminators capture the overall information while ignoring the highlighted areas that



Figure 7. The structure of the pixel discriminator. Conv2D (a, b) means the input channel of the convolution layer is a, and the output channel is b.

don't account for the majority. Such a discriminator is not suitable for specular highlight removal with high image similarity. So, we adopt a pixel discriminator. Figure 7 shows the specific structure of the pixel discriminator we propose.

The pixel discriminator is composed of convolutional layers with a step size of 1 and a padding of 0. The size of the convolution kernel is (1, 1), and there is no pooling layer in the network. This design ensures that the receptive field is equal to 1, that is, each point in the output can correspond to a certain point in the input. The loss function of the discriminator \mathcal{L}_D is defined as:

$$\mathcal{L}_D = \log(D(T)) + \log(1 - D(O)) \quad (9)$$

where \mathcal{L}_D indicates a criterion that measures the Binary Cross Entropy between the ground truth and the output of the generator.

Training Setting

Highlight Dataset

Similar to the current deep learning methods, our method requires a comparatively large amount of data with ground truth for training. However, there is no such appropriate dataset for specular highlight removal from a grayscale image. Therefore, we construct a synthetic dataset using 3D modeling software. There are 1062 sets of images in this dataset. Each image set contains a specular highlight image and a diffuse image. The dataset contains 85 industrial parts of different sizes and shapes, most of them are assigned different metallic materials, others are assigned composite materials to enhance the richness of the materials. For the rendering, two lighting scenes are constructed, a basic scene with only basic light and a high-lighting

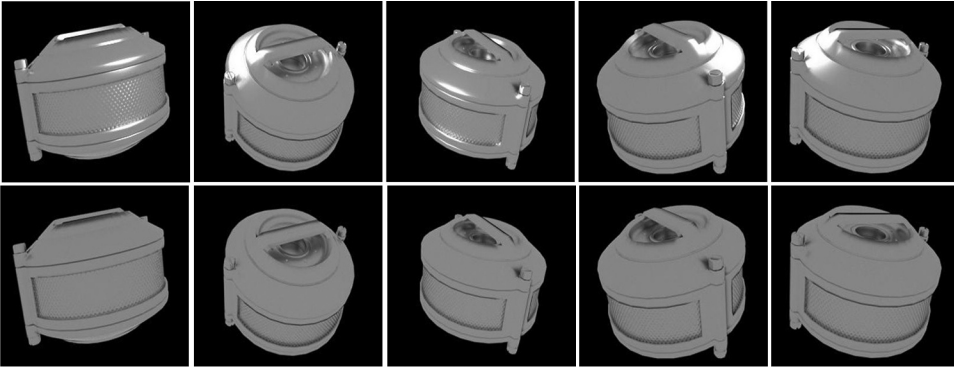


Figure 8. Samples from highlight dataset with 5 different view angles. The top row is highlight images, and the bottom row is diffuse images.

scene with the addition of a globe or flat light. In both scenes, we render the aligned images from 10 different angles as shown in [Figure 8](#). Finally, the rendered images were converted into grayscale for training and testing the models.

Training Details

We implemented our model and comparative models with the PyTorch framework and trained them on an NVIDIA 2080Ti GPU. In training, we adopt Adam optimizer with a batch size of 1 and set a learning rate as 0.0002, the exponential decay rate as $(\beta_1, \beta_2) = (0.6, 0.9)$. We train the generator and discriminator for 200 epochs. The weights of the objective function of the generator are set to $\lambda_{ATT} = 100.0$, $\lambda_{MSE} = 100.0$ and $\lambda_{SSIM} = 10.0$. The hyper-parameters in the SSIM loss C_1, C_2 is set to 0.0001 and 0.0009, and the size of the window is set to 5. The number of recursive blocks N is 4 and the weight θ is 0.5. Theoretically, the larger N makes the better training effect of the attention module, and it also requires more memory.

Experiment Results

MSE Loss and SSIM Loss

Many GANs choose MSE or L1 as the content loss to train neural networks. However, it is not a very wise decision for specular highlight removal, because of the existence of similarity between highlight image and diffuse reflection image, the network can get a good result even if does nothing. In practice, the MSE loss decays very quickly, approaching zero in the training, as shown in [Figure 9](#). As can be seen, both MSE and SSIM show a decreasing trend, indicating that the generator does produce results closer to the ground truth.

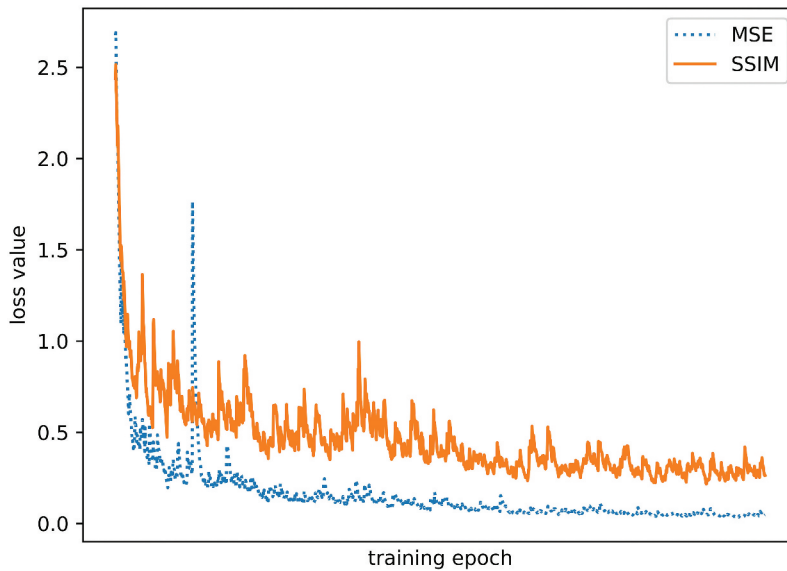


Figure 9. The learning curve of SSIM loss and MSE loss.

But SSIM is still able to maintain a level of 0.4 to 0.5 in the latter part of the training, while MSE is quite close to 0 in the middle of the training, making it unable to contribute to backpropagation. With the addition of SSIM as a loss, the result of the generated images is improved both in quantitative and qualitative evaluation.

Highlight Intensity Mask

Figure 10 shows the highlight intensity mask generated by the attention module at different training steps. It is visualized by the heat map. As the training steps increasing, the highlight intensity mask focuses more and more

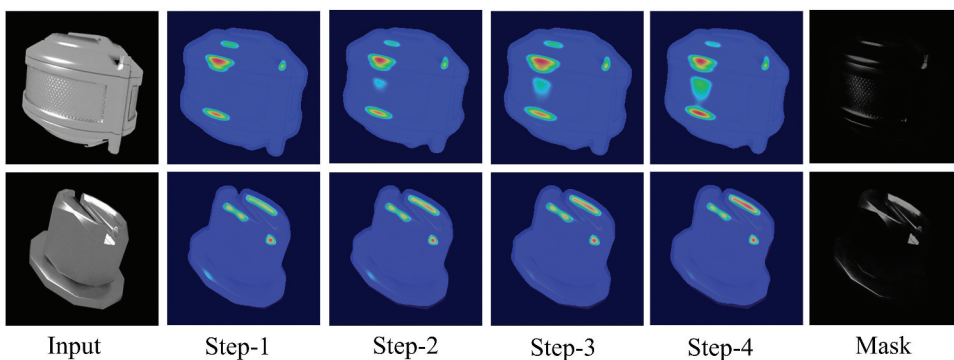


Figure 10. Visualization of the highlight intensity mask generated by our attention module at different training steps.

on the regions that contain specular highlights. It is clear that the mask is an excellent indication of the distribution and intensity of highlight pixels in the highlight image, so it can be used as auxiliary information to help the auto-encoder locate and remove highlights.

Highlight Removal Result

We trained our method and other GAN models on the dataset. ResNet-GAN is the baseline that uses ResNet (He et al. 2016) as the generator. UNet-GAN is similar to ResNet-GAN, but its generator is UNet (Long, Shelhamer, and Darrell 2015). Pix2Pix (Isola et al. 2017) that is a landmark conditional GAN in image translation. It consists of a skip-connected autoencoder and patch discriminator. We also did ablation studies by building networks with parts of features. In Ours-1, we don't exploit the pixel discriminator, but the same patch discriminator as in Pix2Pix. Ours-2 applies MSE loss alone without SSIM loss, and Ours-3 is a model without attention module. Their other parts are the same as the full method.

General Results

To measure the performance, we experiment on the test set with SSIM, Peak signal-to-noise ratio (PSNR), and MSE. They are both important indicators to measure image similarity. The PSNR (in dB) is defined as:

$$PSNR(O, T) = 10\log_{10}\left(\frac{MAX_T^2}{MSE}\right) \quad (10)$$

Where MAX_T is the max value of ground truth, and MSE is the same with \mathcal{L}_{MSE} in Equation 5.

The results of all test images are shown in Table 1, where the MSE is multiplied by 1000 for better comparison. In all metrics, our method gets the best results, which fully demonstrates that our method produces images closer to the ground truth than other methods.

Table 1. The quantitative evaluation result of all test images.

| Method | Metrics | | |
|------------|--------------|---------------|--------------|
| | SSIM | PSNR | MSE |
| Input | 0.972 | 25.288 | 4.522 |
| Resnet-GAN | 0.954 | 27.138 | 2.901 |
| UNet-GAN | 0.982 | 34.905 | 0.557 |
| Pix2Pix | 0.986 | 37.153 | 0.247 |
| Ours-1 | 0.984 | 37.001 | 0.271 |
| Ours-2 | 0.990 | 38.863 | 0.180 |
| Ours-3 | 0.991 | 39.046 | 0.174 |
| Ours | 0.991 | 39.508 | 0.165 |

We notice that the result of UNet-GAN is better than Resnet-GAN, while the only difference between them is that UNet-GAN uses the skip-connected autoencoder. It proves that it is appropriate to employ the skip-connected autoencoder as the generator to deal with the highlight removal problem. With content loss and patch discriminator, Pix2Pix gets an improved grade. It further proves that compared with vanilla GAN, the conditional GAN for image-to-image translation is suitable for highlight removal. Compare to other configurations of our method, Ours-1 (no pixel discriminator), Ours-2 (no SSIM loss), and Ours-3 (no attention), the full model gets the best outcomes, indicating the effectiveness of the combination of those features we proposed.

Cases Study

In this section, we picked five test cases from test images for qualitative and quantitative evaluation. As can be seen in [Figure 11](#), our method is considerably more effective in specular highlight removal compared to

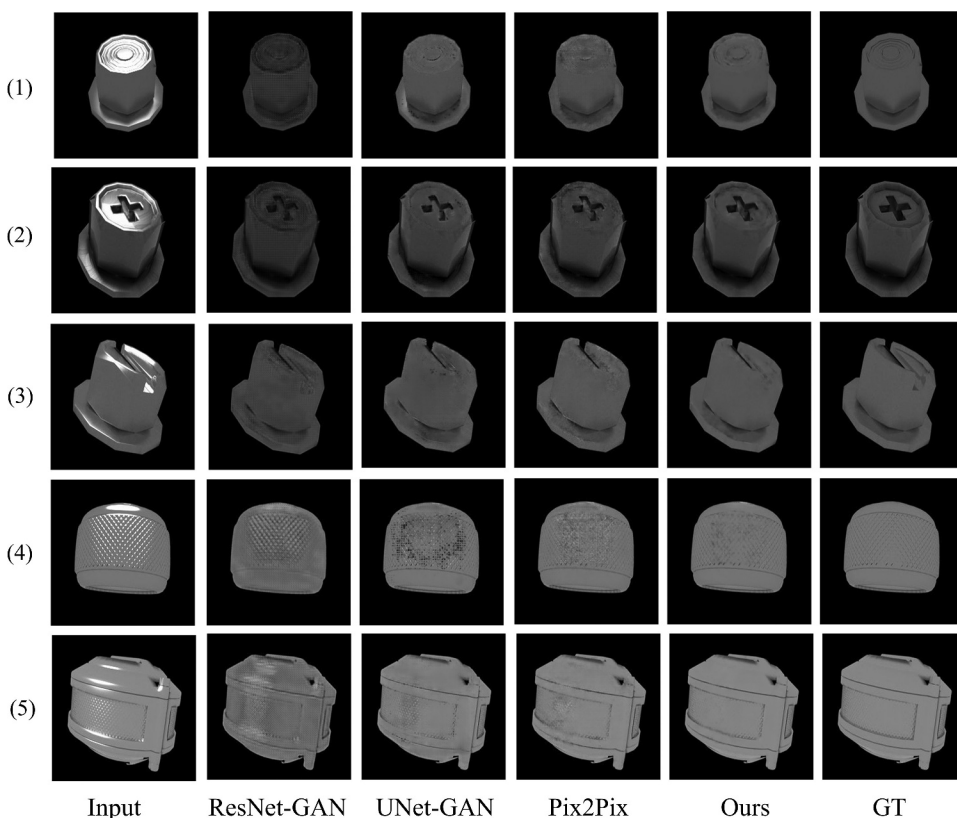


Figure 11. Comparison of 5 test cases of different GAN models in specular highlight removal.

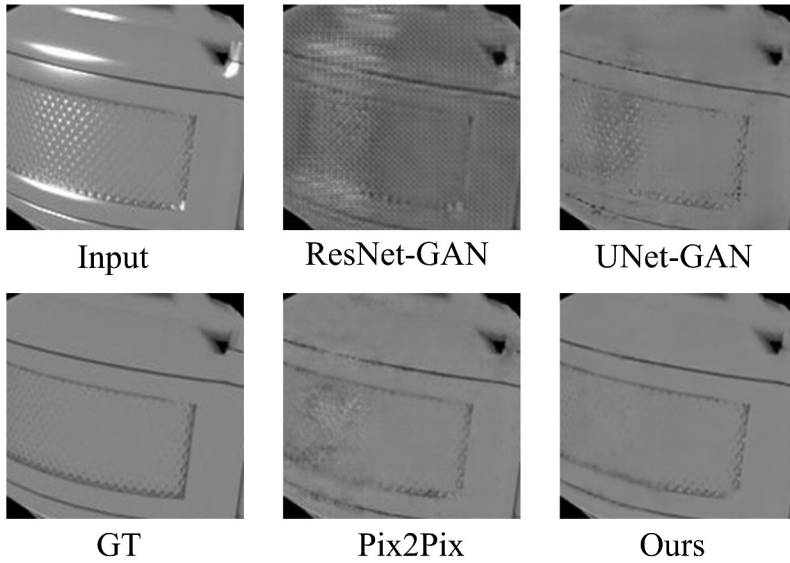


Figure 12. A close look at the comparison between outputs in case (5).

other methods. Our method produces the clearest diffuse images. Especially in the region with texture, the images generated by our method can keep more details, while the images generated by other methods are blurred as shown in [Figure 12](#).

We also made a quantitative evaluation of the five test cases, and the results are shown in [Table 2](#). The quantitative and qualitative evaluations of five test cases demonstrate the performance of our method.

Table 2. The quantitative evaluation result of cases study.

| Test Case | | ResNet-GAN | UNet-GAN | Pix2Pix | Ours |
|-----------|------|------------|----------|---------|---------------|
| (1) | SSIM | 0.940 | 0.981 | 0.982 | 0.989 |
| | PSNR | 22.437 | 34.148 | 36.591 | 37.497 |
| | MSE | 5.70 | 0.384 | 0.219 | 0.177 |
| (2) | SSIM | 0.947 | 0.972 | 0.971 | 0.982 |
| | PSNR | 26.792 | 30.511 | 31.359 | 33.000 |
| | MSE | 2.09 | 0.888 | 0.731 | 0.501 |
| (3) | SSIM | 0.964 | 0.980 | 0.980 | 0.989 |
| | PSNR | 25.958 | 31.516 | 34.307 | 36.420 |
| | MSE | 2.53 | 0.705 | 0.370 | 0.228 |
| (4) | SSIM | 0.922 | 0.937 | 0.963 | 0.983 |
| | PSNR | 27.522 | 28.575 | 35.021 | 38.103 |
| | MSE | 1.76 | 1.38 | 0.314 | 0.154 |
| (5) | SSIM | 0.908 | 0.974 | 0.979 | 0.989 |
| | PSNR | 23.557 | 31.425 | 36.120 | 39.726 |
| | MSE | 4.40 | 0.720 | 0.244 | 0.106 |

Conclusion

In this paper, we regard specular highlight removal as an image-to-image translation between the highlight domain and the diffuse domain. We propose a single-image-based highlight removal method that focuses on removing highlights from a grayscale image. This method utilizes the generative adversarial network, where the generator produces images that is free of specular highlights, while the discriminator is responsible for determining whether images are clear and highlight-free. We take the similarity of the images in specular highlight removal into account, and elaborately design all parts of the network. In generative networks, the attention module is employed for generating the highlight intensity mask to locate highlight pixels. Then highlights are removed by the skip-connected autoencoder which ensures low-level information can be conducted directly. Pixel discriminator and SSIM loss help to train the generator to get results that keep more details. Finally, the effectiveness of the method is proved by quantitative and qualitative evaluation.

Disclosure Statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by Zhejiang Key R & D Plan Project [2021C01114] and the National Science Foundation of China [61703127].

ORCID

Jing Chen  <http://orcid.org/0000-0003-3127-8462>

References

- Artusi, A., F. Banterle, and D. Chetverikov. 2011. A survey of specular removal methods paper presented at computer graphics forum.
- Bajcsy, R., S. W. Lee, and A. Leonardis. 1996. Detection of diffuse and specular interface reflections and inter-reflections by color image segmentation. *International Journal of Computer Vision* 17 (3):241–72. doi:10.1007/BF00128233.
- Ding, Z., X.-Y. Liu, M. Yin, and L. Kong. 2019. Tgan: Deep tensor generative adversarial nets for large image generation. arXiv Preprint arXiv:1901.09953
- Feris, R., R. Raskar, K.-H. Tan, and M. Turk. 2004. Specular reflection reduction with multi-flash imaging paper presented at proceedings. 17th Brazilian Symposium on Computer Graphics and Image Processing
- Funke, I., S. Bodenstedt, C. Riediger, J. Weitz, and S. Speidel. 2018. Generative adversarial networks for specular highlight removal in endoscopic images Paper presented at Medical Imaging 2018: Image-Guided Procedures, Robotic Interventions, and Modeling

- Goodfellow, I., J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. 2014. Generative adversarial nets Paper presented at Advances in neural information processing systems
- Gregor, K., I. Danihelka, A. Graves, D. J. Rezende, and D. Wierstra. 2015. Draw: A recurrent neural network for image generation. arXiv Preprint arXiv:1502.04623
- Guan, S., A. A. Khan, S. Sikdar, and P. V. Chitnis. 2020. Fully dense unet for 2-d sparse photoacoustic tomography artifact removal. *IEEE Journal of Biomedical and Health Informatics* 24 (2):568–76. doi:10.1109/JBHI.2019.2912935.
- Gui, J., Z. Sun, Y. Wen, D. Tao, and J. Ye. 2020. A review on generative adversarial networks: Algorithms, theory, and applications. arXiv Preprint arXiv:2001.06937
- Guo, J., Z. Zhou, and L. Wang. 2018. Single image highlight removal with a sparse and low-rank reflection model Paper presented at Proceedings of the European Conference on Computer Vision (ECCV)
- He, K., X. Zhang, S. Ren, and J. Sun. 2016. Deep residual learning for image recognition Paper presented at Proceedings of the IEEE conference on computer vision and pattern recognition
- Huang, H., P. S. Yu, and C. Wang. 2018. An introduction to image synthesis with generative adversarial nets. arXiv Preprint arXiv:1803.04469.
- Isola, P., J.-Y. Zhu, T. Zhou, and A. A. Efros. 2017. Image-to-image translation with conditional adversarial networks Paper presented at Proceedings of the IEEE conference on computer vision and pattern recognition
- Klinker, G. J., S. A. Shafer, and T. Kanade. 1988. The measurement of highlights in color images. *International Journal of Computer Vision* 2 (1):7–32. doi:10.1007/BF00836279.
- Koirala, P., M. Hauta-Kasari, and J. Parkkinen. 2009. Highlight removal from single image Paper presented at International Conference on Advanced Concepts for Intelligent Vision Systems
- Li, Y., and L. Ma. 2006. Metal highlight spots removal based on multi-light-sources and total variation inpainting Paper presented at Proceedings of the 2006 ACM international conference on Virtual reality continuum and its applications
- Lin, J., M. E. A. Seddik, M. Tamaazousti, Y. Tamaazousti, and A. Bartoli. 2019. Deep multi-class adversarial specular removal Paper presented at Scandinavian Conference on Image Analysis
- Long, J., E. Shelhamer, and T. Darrell. 2015. Fully convolutional networks for semantic segmentation Paper presented at Proceedings of the IEEE conference on computer vision and pattern recognition
- Qian, R., R. T. Tan, W. Yang, J. Su, and J. Liu. 2018. Attentive generative adversarial network for raindrop removal from a single image Paper presented at Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition
- Sato, Y., and K. Ikeuchi. 1994. Temporal-color space analysis of reflection. *JOSA A* 11 (11):2990–3002. doi:10.1364/JOSAA.11.002990.
- Schlüns, K., and M. Teschner. 1995. Fast separation of reflection components and its application in 3d shape recovery Paper presented at Color and Imaging Conference
- Shafer, S. A. 1985. Using color to separate reflection components. *Color Research and Application* 10 (4):210–18. doi:10.1002/col.5080100409.
- Shen, H.-L., and Z.-H. Zheng. 2013. Real-time highlight removal using intensity ratio. *Applied Optics* 52 (19):4483–93. doi:10.1364/AO.52.004483.
- Wang, C., T. Macgillivray, G. Macnaught, G. Yang, and D. Newby. 2018a. A two-stage 3d unet framework for multi-class segmentation on full resolution image. arXiv Preprint arXiv:1804.04341.

- Wang, H., C. Xu, X. Wang, Y. Zhang, B. Peng, and Y. Yang. 2016. Light field imaging based accurate image specular highlight removal. *Plos One* 11 (6):e0156173. doi:[10.1371/journal.pone.0156173](https://doi.org/10.1371/journal.pone.0156173).
- Wang, X., K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. Change Loy. 2018b. Esrgan: Enhanced super-resolution generative adversarial networks Paper presented at Proceedings of the European Conference on Computer Vision (ECCV)
- Wang, Z., A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli. 2004. Image quality assessment: From error visibility to structural similarity. *IEEE Transactions on Image Processing* 13 (4):600–12. doi:[10.1109/tip.2003.819861](https://doi.org/10.1109/tip.2003.819861).
- Wei, X., X. B. Xu, J. W. Zhang, and Y. H. Gong. 2018. Specular highlight reduction with known surface geometry. *Computer Vision and Image Understanding* 168:132–44. doi:[10.1016/j.cviu.2017.10.010](https://doi.org/10.1016/j.cviu.2017.10.010).
- Wolff, L. B., and T. E. Boulton. 1993. Constraining object features using a polarization reflectance model. *Physics-Based Vision: Principles and Practice: Radiometry* 1:167.
- Xu, C., X. Wang, H. Wang, and Y. Zhang. 2015. Accurate image specular highlight removal based on light field imaging Paper presented at 2015 Visual Communications and Image Processing (VCIP)
- Yeh, R., C. Chen, T. Y. Lim, M. Hasegawa-Johnson, and M. N. Do. 2016. Semantic image inpainting with perceptual and contextual losses. arXiv Preprint arXiv:[1607.07539](https://arxiv.org/abs/1607.07539) 2 (3).
- Yu, J., Z. Lin, J. Yang, X. Shen, X. Lu, and T. S. Huang. 2018. Generative image inpainting with contextual attention Paper presented at Proceedings of the IEEE conference on computer vision and pattern recognition
- Zhao, B., X. Wu, J. S. Feng, Q. Peng, and S. C. Yan. 2017. Diversified visual attention networks for fine-grained object classification. *IEEE Transactions on Multimedia* 19 (6):1245–56. doi:[10.1109/Tmm.2017.2648498](https://doi.org/10.1109/Tmm.2017.2648498).