See Through the Windshield from Surveillance Camera

Daiqian Ma[1,2], Yan Bai[2], Renjie Wan[3], Ce Wang[2], Boxin Shi[2,4], Ling-Yu Duan[1,2,4*]
The SECE of Shenzhen Graduate School, Peking University, Shenzhen, China[1]
The National Engineering Lab for Video Technology, Peking University, Beijing, China[2]
School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore[3]
The Peng Cheng Laboratory, Shenzhen, China[4]
{madaiqian,yanbai,wce,shiboxin,lingyu}@pku.edu.cn,rjwan@ntu.edu.sg

ABSTRACT

This paper attempts to address the challenging task of seeing through the windshield images captured by surveillance cameras in the wild. Such images usually have very low visibility due to heterogeneous degradations caused by blur, haze, reflection, noise etc., which makes existing image enhancing methods inapplicable. We propose a windshield image restoration generative adversarial network (WIRE-GAN) to restore and enhance the visibility of windshield images. We adopt the weakly supervised framework based on the generative model, which has effectively released the request of paired training data for a specific type of degradation. To generate more semantically consistent results even in extreme lighting conditions, we introduce a novel content-preserving strategy into the proposed weakly-supervised framework. To make the image restoration more reliable, the WIRE-GAN network constructs a sort of content-aware embedding space and enforces the constraint of the restored windshield images being closer to the original input in the embedding space. Moreover, we collect a large-scale windshield image dataset (WIRE dataset) to validate the advantage of our method in improving the image quality, and further evaluate the impact of windshield restoration on the vehicle ReID performance.

KEYWORDS

Windshield restoration; heterogeneous degradations; generative adversarial network

ACM Reference Format:

Daiqian Ma, Yan Bai, Renjie Wan, Ce Wang, Boxin Shi, Ling-Yu Duan. 2019. See Through the Windshield from Surveillance Camera. In *Proceedings of the 27th ACM International Conference on Multimedia (MM '19), October 21-25, 2019, Nice, France.* ACM, New York, NY, USA, 9 pages.
https://doi.org/10.1145/3343031.3351077

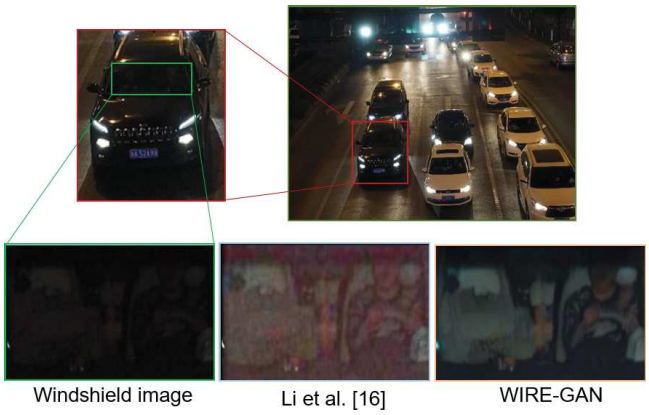

Figure 1: Surveillance cameras capture vehicles in various unconstrained condition (top right), which results in windshield images with low visibility and high noise (top left). Our method (bottom right) is able to see through the low-visibility windshield images with superior performance over existing methods (bottom middle).

1 INTRODUCTION

Vehicle is an important target for surveillance cameras, and the scene behind the windshield is of great interests to surveillance video analysis for the benefit of public security control and criminal investigation. Windshield images are usually captured by diverse types of surveillance cameras under various unconstrained scenarios. The uncontrollable imaging conditions bring numerous sources of distortions to images, such as high noises from the surveillance cameras under low light conditions, strong reflections from the glass, motion blur caused by fast-moving vehicles, and so on, which make seeing through the windshield challenging for either human observers or computer vision systems (Figure 1 upper row).

Windshield images from surveillance cameras suffer from heterogeneous distortions due to different types of degradation (*e.g.*, blurring effects, low-visibility in bad weather, high noise in low light condition, *etc.*), which makes its enhancement and restoration unique and challenging. Given the origin of distortion, enhancing and restoring a low-quality image is usually efficient when applying a well-designed specific solution, such as deblurring [14], dehazing [29] or denoising [4], namely, each type of method is designed with handcrafted prior according to the particular degradation model [27, 32], or with implicit distribution prior learned from a large amount of data [13, 30]. Figure 1 gives a typical example of restoring windshield image of low lighting condition. As shown in the lower row in Figure 1, the state-of-the-art method [16] fails to provide a clear enough view through the windshield, in which the restored image is with blurry and noisy effects, while

*Ling-Yu Duan is the corresponding author.

MM '19, October 21–25, 2019, Nice, France

ACM ISBN 978-1-4503-6889-6/19/10…$15.00
https://doi.org/10.1145/3343031.3351077

Low-quality windshields High-quality windshields

Figure 2: Sample images from our WIRE image dataset. The left part is the low-quality windshields with various types of degradations. The right part is the high-quality windshields with visually clean appearances.

our method (WIRE-GAN) performs better. Although deep learning based methods have shown promising results in a variety of challenging scenarios, the majority of those methods target a single type of degradation origin, which brings about the heavy reliance of generating supervision data complying to a specific type of problem.

Clearly, generating synthetic paired training data to simulate surveillance windshield images in a supervised manner is infeasible due to the heterogeneous degradation conditions. The acquisition of large amount of real data is also too costly to be implemented. To alleviate this problem, an option is to apply existing weakly supervised GAN-based methods [6, 34] directly. However, owing to their unconstrained generative training manner, the GANs incline to generate images resembling the set of target domain on the whole, but may incur serious deflection for an individual subject. This is inappropriate for our task, since surveillance application requires consistency between the restored subject and the original one.

In this paper, we propose a unified weakly supervised framework for **W**ind-shield **I**mage **R**estoration and **E**nhancement, named "**WIRE**-GAN". Our method has successfully released the requirement from image restoration or enhancement tasks sticking to one type of degradation and supervision by using a weakly-supervised generative adversarial mechanism, thereby handling the diversity of degradation phenomena simultaneously. Specially, to generate more semantically consistent restoration results, we introduce a content-aware discriminator, which attempts to preserve the content of input windshield by aggregating the windshield images with the same vehicle ID but different appearances in a content-aware embedding space, and further enforce the restored windshield images to be located closer to the original input in such embedding space. Our major contributions are summarized as follows:

- To the best of our knowledge, we are the first to tackle this challenging but valuable problem of seeing through windshield images from real-world surveillance cameras with heterogeneous image degradations.
- We propose to reconsider the problem of image restoration from the perspective of content preservation. By incorporating a content-aware discriminator into a unified domain translation framework, we are able to maintain better consistency between the restored and the original content. Moreover, the capability of reliable visual feature preserving further contributes to the windshield ReID task.
- We collect the first large-scale windshield image dataset with 176,425 unpaired image samples, which is expected to facilitate the research of windshield image restoration and a variety of real-world image restoration and enhancement tasks in a weakly supervised manner.

2 RELATED WORK

Various types of image restoration solution have been designed according to the degradation type of original images, such as super-resolution [15], deblurring [14], dehazing [29], denoising [4] and so on. Generally, each type of problem comes up with a specialized model and solution. Existing works can be categorized into non-learning based methods and learning-based methods.

Non-learning based Methods. Most of the conventional approaches are non-learning optimization-based and various hand-designed priors are applied to exploit the properties of the degradation factors (*e.g.*, noise, blur, reflection, *etc.*). For example, Nikolaos *et al.* [1] proposed an approach to suppress reflections based on a Laplacian data fidelity term and an l_0 gradient sparsity term. Yair *et al.* [27] presented a multi-scale weighted nuclear norm minimization method for image restoration and Zhang *et al.* [32] employed low-rank tensor factor analysis for tensors generated by grouped image patches to restore the degraded images. However, non-learning based methods usually have limited generalization ability as the hand-designed priors may not apply in real-world cases.

Learning based Methods. Recent success has been achieved by deep learning approaches. Kligvasser *et al.* [13] proposed an xUnit structure to learn a spatial activation function for efficient image restoration and Yu *et al.* [30] devised a joint learning scheme to craft a tool chain for image restoration by deep reinforcement learning. However, a common property of these works is that most of them target at restoring a single type of degradation. It is worthy to mention that our goal is not limited to any specific degradation origin, and we aim to restore the real-world low-quality windshield

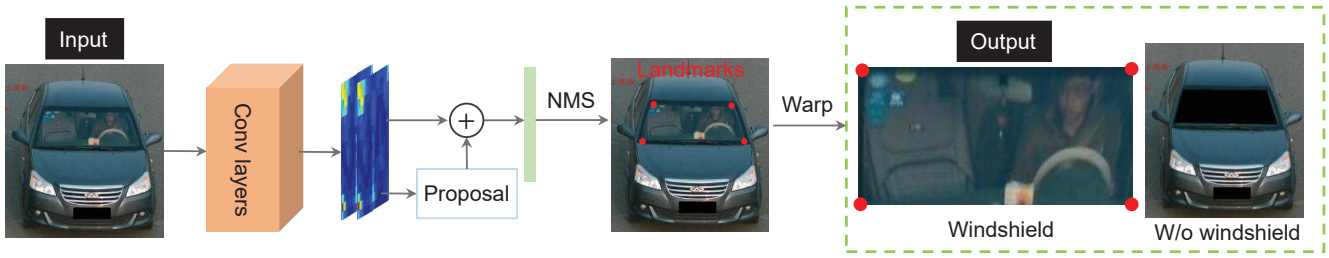

Figure 3: The dataset pre-processing pipeline for cropping the windshields, which applies a landmark regression structure with regional proposal network. NMS means non maximum suppression and warp represents the warp transformation.

images with unknown mixture of various types of degradation to high-quality windshield images.

In addition, the proposed new problem of windshield restoration is also related to image-to-image translation networks using generative models [8–10, 18, 34]. Isola *et al.* [10] proposed the first unified framework for image-to-image translation based on conditional GANs, which has been extended to generate high-resolution images by Wang *et al.* [24]. Recent studies have also attempted to learn image translations without supervision. Note that this type of problem is inherently ill-posed and requires additional constraints. Some works have ever enforced the translation to preserve certain properties of the source domain data, such as pixel values [20], pixel gradients [3], semantic features [22], class labels [3], or pairwise sample distances [2]. Another typical constraint for enforcing the translation is the cycle consistency mechanism [12, 34], that is when we translate an image to the target domain and do the inverse translation process from the target to the source domain, we should obtain exactly the same image as the original input. These works show remarkable performance in general image translation tasks but cannot be directly applied to the proposed windshield restoration task, because they can not consistently keep the semantically meaningful content during the converting procedure. Thus, we propose to incorporate a sort of content preserving mechanism to fulfill the request of content-aware image translation in our task.

3 WIRE DATASET

We collect a new dataset for training models and evaluating the performance of Windshield Image Restoration and Enhancement task, named "WIRE" dataset. The WIRE consists of two parts: low-quality windshield images taken by cameras in toll-gates and high-quality windshield images taken with cameras of higher resolution and lower noise. We extract windshield landmarks from the surveillance camera images using a regional proposal network, and then rectify the windshield regions to a rectangular coordinate, as shown in Figure 3. Sample images are given in Figure 2.

Low-quality windshield images (see the example images on the left part of Figure 2) are obtained from existing vehicle dataset [17], which is captured from multiple real-world surveillance cameras, and a dominant portion of this dataset is front-view vehicle images. We select all the front-view images with a pretrained classifier and then locate and crop out the regions of windshields with the pipeline mentioned above. The total number of low-quality images is 166,290, including 12,305 training IDs and 11,890 testing IDs.

Low-quality windshield images contain various levels and types of image degradation. From Figure 2, the first row of the low-quality windshields mainly shows the condition of extreme low lighting, which makes the surveillance scene content almost invisible from the dark environment at night; the second row illustrates how reflections of the glass contaminate the interior content of vehicles under surveillance cameras; and the third row shows a set of highly blurred pictures, due to the combined effect of the out-of-focus blur from surveillance cameras, motion blur from fast-moving vehicles, and in some cases bad atmosphere conditions like haze. The left two samples in the last row reveal the existence of high specular reflection, also there are some generally clear windshield images as shown in the right half of this row. It is also worth noting that although the exhibited samples are contaminated by one dominant type of degradation phenomenon, most of the real-world surveillance images actually show various phenomena when observing closely, which poses a unique challenge to our task.

High-quality windshield images are captured by setting up the camera to face the road at the overpass. The camera is Canon EOS-1D X Mark II, the lens used is EF70-200mm F2.8 L IS II USM, and we take the auto-focusing strategy. We locate and crop the windshields from these images and manually check the quality of each windshield. The number of high-quality windshield images is 10,135, including 1722 different vehicle IDs.

4 PROPOSED METHOD

WIRE-GAN aims at learning the content-aware translation from low-quality windshield images to high-quality windshield images using unpaired supervision. Given an ID-annotated dataset X from source domain and an ID-annotated dataset Y from target domain, our goal is to train a reliable restoration model that translates from source domain to target domain but keeps the content details in the original images. As shown in Figure 4, we propose a unified content-aware weakly supervised framework, it consists of two generators G and F, two adversarial discriminators D_H and D_L, and one content-aware discriminator D_C. In this section, we first present the domain translation framework and then introduce the details of the content-aware discriminator.

4.1 Domain Translation

To bridge the source domain (low-quality windshield images) and target domain (high-quality windshield images), we use two generator-discriminator pairs: $\{G, D_H\}$ and $\{F, D_L\}$, which aims at producing samples indistinguishable from those in the target (source) domain, respectively. G generates samples from low-quality domain to high-quality, and F inversely. The whole training is guided by four losses ($\mathcal{L}_{Hadv}$, $\mathcal{L}_{Ladv}$, $\mathcal{L}_{cyc}$ and $\mathcal{L}_{ide}$). For generator G and its associated discriminator D_H, the adversarial loss is as follows:

$$\begin{aligned}\mathcal{L}_{Hadv}(G, D_H, X, Y) = &\mathbb{E}_{y\sim p_y}[\log D_H(y)] \\ &+ \mathbb{E}_{x\sim p_x}[\log(1 - D_H(G(x)))],\end{aligned} \tag{1}$$

where x is the sample from the source domain, y is the sample from the target domain, and p_x and p_y denote the sample distributions in the source and target domain, respectively. The adversarial loss $\mathcal{L}_{Ladv}$ for generator F and its associated discriminator D_L can be similarly defined as follows:

$$\begin{aligned}\mathcal{L}_{Ladv}(F, D_L, X, Y) = &\mathbb{E}_{x\sim p_x}[\log D_L(x)]. \\ &+ \mathbb{E}_{y\sim p_y}[\log(1 - D_L(F(y)))].\end{aligned} \tag{2}$$

To guarantee that the learned function can map an individual input x_i to the desired output y_i and further reduce the space of possible mapping functions, we embed the cycle-consistency [12, 34] into the mapping process as follows:

$$\begin{aligned}\mathcal{L}_{cyc}(G, F) = &\mathbb{E}_{x\sim p_x}||x - F(G(x))||_1 \\ &+ \mathbb{E}_{y\sim p_y}||y - G(F(y))||_1.\end{aligned} \tag{3}$$

Since the generator G and F have a high degree of freedom to change the tint of input images, we employ the target domain identity constraint to restrict such freedom as an auxiliary term for image-to-image translation. Target domain identity constraint is used to regularize the generator to be the identity matrix on the samples from target domain, which is formulated as:

$$\begin{aligned}\mathcal{L}_{ide}(G, F) = &\mathbb{E}_{x\sim p_x}||x - F(x)||_1 \\ &+\mathbb{E}_{y\sim p_y}||y - G(y)||_1.\end{aligned} \tag{4}$$

The identity constraint helps better preserve the color consistency between the input and output.

4.2 Content-Preserving Strategy

The proposed domain translation framework is to initially recover the background scenes behind the windshield. However, due to the lack of specific constraint on the content, it may generate results inclined to the average distribution in the target domain, which are unsuitable for surveillance purposes and high-level recognition problem because of the loss of content details. To make the domain translation more reliable, we propose a content-aware discriminator to enforce additional constraints in the feature embedding space.

The content-aware discriminator is committed to constructing a content related embedding feature space to aggregate the images with the same content (i.e. the vehicle ID) under different degraded conditions. Thus, the constraints in the embedding space will help the generator in learning to keep the contents while improving the visual quality. The training details on the content-preserving strategy are discussed below.

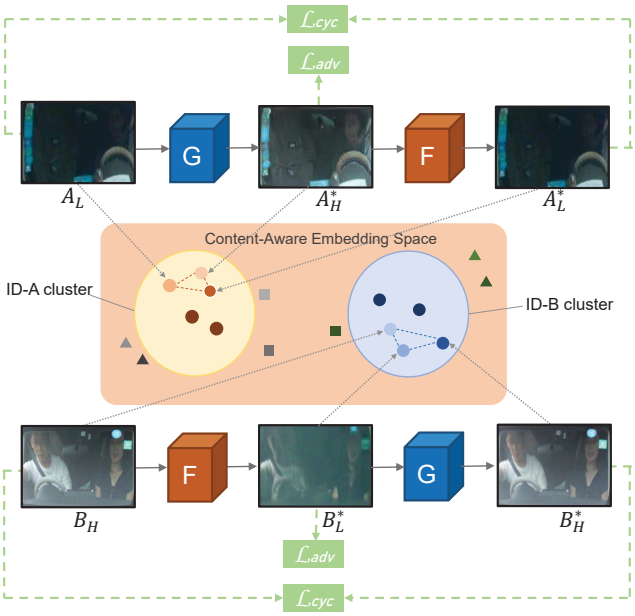

Figure 4: Illustration of the proposed WIRE-GAN. There are two generators G and F to generate target domain images. Here A and B represent different IDs, and A_H, A_L denote the high-quality and low- quality windshield images with ID-A (similar for ID-B), respectively. * represents the generated images in the training procedure. Two cycle consistent losses are formulated as $F(G(A_L)) \approx A_L$ and $G(F(B_H) \approx B_H$.

Content-aware embedding space. First, we introduce how the content-aware discriminator learns a better content-aware embedding space with a triplet mechanism. The idea is to pull windshields with same ID closer and push windshields with different IDs farther apart. The specific loss for the discriminator D_C is formulated as:

$$\begin{aligned}\mathcal{L}_C^D = &\mathbb{E}_{x\sim p_x}\max\{||f(x^a) - f(x)|| + \alpha_1 - ||f(x^a) - f(x^n)||, 0\} \\ &+\mathbb{E}_{y\sim p_y}\max\{||f(y^a) - f(y)|| + \alpha_1 - ||f(y^a) - f(y^n)||, 0\},\end{aligned} \tag{5}$$

where x^a and x^n represent the anchor sample (sample with the same ID) and the negative sample (sample with different ID) from source domain X, f denotes VGGM network, y^a and y^n are the anchor sample and the negative sample from target domain Y, α_1 denotes the threshold for the distance metric. Moreover, we adopt a multi-loss training strategy containing softmax loss, which has been demonstrated to be effective in vehicle ReID. tasks [25, 33]. As illustrated in Figure 5, the windshield images with different degradations but the same ID are pulled closer while the windshields with different IDs are pushed away in such a embedding space.

Content consistent loss. Then we explain the constraints in the content-aware embedding space to help the generator preserve the contents. Concretely, we constrain the generated $G(x)$ located in the neighboring margin specified via distance threshold α_2 as:

$$||x - G(x)||_2 \leq \alpha_2, ||y - F(y)||_2 \leq \alpha_2, \tag{6}$$

The parameter α_2 is the maximum distance from x, which aims to constrain $G(x)$ and $F(y)$ to locate within a certain distance from x. Without this constraint, $G(x)$ and $F(y)$ would tend to generate a content-irrelevant windshield image. The embedding loss for $G(x)$ and $F(y)$ can be formulated as:

$$\begin{aligned}\mathcal{L}_C^G =& \mathbb{E}_{x\sim p_x} \max\{||f(x) - f(G(x))|| - \alpha_2, 0\} \\ &+ \mathbb{E}_{y\sim p_y} \max\{||f(y) - f(F(y))|| - \alpha_2, 0\}.\end{aligned} \tag{7}$$

4.3 Overall Objective Function.

The overall objective function for WIRE-GAN is formulated as

$$\begin{aligned}\mathcal{L}_{total} = \mathcal{L}_{Hadv} + \mathcal{L}_{Ladv} + \lambda_1\mathcal{L}_{cyc} + \lambda_2\mathcal{L}_{ide} \\ + \lambda_3\mathcal{L}_C^G + \mathcal{L}_C^D,\end{aligned} \tag{8}$$

where $\lambda_t, t \in \{1, 2, 3\}$ controls the relative importance of these losses. The first four losses belong to the domain translation framework introduced in Section 4.1, and the last two losses are from the content-aware discriminator introduced in Section 4.2.

Details of network architecture. For the content-aware discriminator, we use VGGM [21] as the base network. The input for loss function layer is the last fully-connected layer *fc7* with the dimension of 1024. The content-aware discriminator serves as a feature extractor, and the output of L_2 normalization layer is treated as the feature representation for content consistency. For the D_L and D_H discriminator, we use 70×70 PatchGANs [10], which aim to classify whether 70×70 overlapping image patches are real or fake. As for the generators, we adopt the architecture from Johnson *et al.* [11]. This network contains two stride-2 convolutions, six residual blocks [7], and two fractionally strides convolutions with stride $\frac{1}{2}$. The input image size is 128×256 and instance normalization [23] is utilized for better training.

5 EXPERIMENTS

We conduct both qualitative and quantitative experiments on the proposed WIRE dataset. We first introduce the implementation details, then the qualitative analyses are performed, including the visualization of these methods and a user study. Finally, we analyze the performance gain to the high-level task windshield ReID.

5.1 Implementation Details

Training strategy. The model is implemented with PyTorch[1]. We feed the image with size 256×128 into the network and we adopt the alternative-training strategy. For each stage, we first train the generator once, and then train the D_L, D_H and D_C discriminators each for once separately. The learning rate is set to 2×10^{-4} for the first 100 epochs and we linearly decay the learning rate to 0 over the next 100 epochs.

Parameters setting. Regarding the weights of loss functions, we set $\lambda_1 = 10$ and $\lambda_2 = 0.5$ as in [34]. The λ_3 is set as 2 to control the content-consistency during the training procedure. Regarding the parameters for distance constraint, we set $\alpha_1 = 0.6$ and $\alpha_2 = 0.5$. These three parameters aim to learn a better embedding space to constraint the content information.

[1]http://pytorch.org

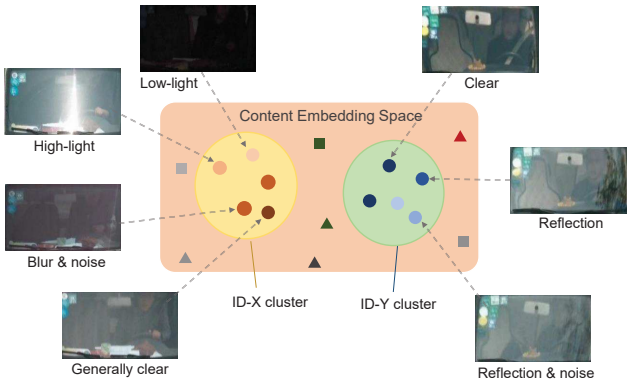

Figure 5: Illustration of the embedding space with content-aware discriminator. The low-light windshields, high-light windshields, blur windshields and clear windshields of the same ID are pulled closer while the windshields with different IDs are pushed away.

Figure 6: Three ReID experiments (w/o windshield ReID, windshield ReID, vehicle ReID) are used to verify the importance of the windshield for recognizing the identity.

Experiment setting. As it is infeasible to obtain high quality ground truth of windshield images from real-world surveillance cameras, the numerical evaluations based on the classical error metrics (*e.g.*, PSNR or SSIM) do not apply. Thus, we propose to evaluate the performance in the following two ways:

(1) **Windshield ReID evaluation.** We use the restored windshields as query images in the windshield ReID task to investigate whether the restored windshields can be used as more reliable input to benefit the high-level task. In this part, we adopt the mean Average Precision (mAP) and the top-K match rate as the error metrics. The definition of AP is:

$$AP = \frac{\sum_{k=1}^{n} P(k) \times gt(k)}{N_{gt}}, \tag{9}$$

where k is the rank in the sequence of retrieved windshields, n is the number of retrieved windshields, and N_{gt} is the number of relevant windshields. $P(k)$ is the precision at cut-off k in the recall list and $gt(k)$ indicates whether the k-th recall image is correct or not. The mAP is defined as:

$$mAP = \frac{\sum_{q=1}^{Q} AP(q)}{Q}, \tag{10}$$

where Q is the number of total query images.

Figure 7: Examples of windshield restoration results compared with various non-learning image restoration methods (the first and second row are processed with deblurring method [26] and denoising refinement [5]; the third and fourth row are processed with reflection removal methods [1] and denoising refinement [5]; the bottom two rows are restored through low lighting enhancement method [16] followed by BM3D denoising [5]), CycleGAN[34], SPGAN [6], in which the red and blue boxes indicate regions with notable differences.

(2) **Qualitative evaluation.** We study the visual quality of the restored windshields by comparing our method against baselines and then we conduct a perceptual study to evaluate the restoration quality from three learning based methods.

5.2 Windshield ReID Evaluation

The Importance of the Windshield for Recognizing Identity. Since a complete vehicle image can be disentangled as a windshield image and a w/o windshield image, we conduct three kinds of vehicle ReID experiments with different query inputs including the complete vehicle image, the windshield image, and the w/o windshield image, to investigate the richness of discriminative visual information for recognition of the three kinds of input images. As shown in Figure 6, the mAP results of vehicle ReID, windshield ReID, and w/o windshield ReID are 0.781, 0.723 and 0.568, respectively. It shows that the windshield plays a more important role than the vehicle body for recognizing the identity. The windshield contains rich identity-representative features which contribute to the moderate performance drop in vehicle ReID compared to the complete vehicle image query, thus it is meaningful to preserve such information during the mapping from low-quality windshield to high-quality windshields. Therefore, we propose a content-preserving strategy and in the following experiments, we use the performance of restored windshield image in the windshield ReID task as an evaluation metric. The visualization of these

Table 1: Performance comparisons on WIRE dataset. QE represents query expansion. R=k means the top-k accuracy..

Settings	Small-scale data			Large-scale data		
Methods	mAP	R=1	R=5	mAP	R=1	R=5
GoogLeNet[28]	-	49.88	67.18	-	43.40	63.86
HDC+Contrastive[31]	71.5	70.12	79.05	64.5	60.48	67.25
Baseline	72.3	69.78	79.66	66.8	62.97	68.18
CycleGAN[34]	65.5	58.02	72.65	62.8	56.36	68.84
SPGAN[6]	70.3	67.03	76.25	65.4	60.47	70.98
Ours	73.1	71.35	81.35	67.6	63.82	72.22
CycleGAN (QE)	70.2	62.52	78.43	65.0	59.05	72.79
SPGAN (QE)	72.9	70.14	79.91	67.4	62.06	73.35
Ours (QE)	**74.8**	**73.72**	**81.57**	**68.3**	**67.31**	**73.81**

identity-representative regions is analyzed in Section 5.4.

Windshield ReID Evaluation. The results in the visual quality evaluation and user study evaluation have shown the advantages of our proposed method for human perception. In order to fully investigate wheth-er our method can benefit the high-level task, we conduct a Windshield ReID experiment. For the test dataset, we use two test subsets of different sizes from the proposed WIRE dataset, *i.e.*, 6493 images with 729 IDs in small size, and 19777 images with 2178 IDs in large size. We first train a baseline model [17] with softmax loss and triplet loss, then we put the images restored by CycleGAN [34], SPGAN [6] and WIRE-GAN as query images into the windshield retrieval evaluation task, and report the average performance on the test sets. Moreover, we apply query expansion (QE) to further investigate whether the restored image is reliable enough to improve the retrieval task. The QE strategy is similar to [19], we use the features extracted from the restored image to get another rank with Euclidean distance, then we do a rank fusion with these two ranks and calculate the performance of final rank.

The performance comparison on WIRE dataset is listed in the Table 1. From the comparison among non-QE methods, it is obvious that our method achieves the best performance. The performances of CycleGAN [34] and SPGAN [6] are even lower than the original windshield images, it shows that the examples restored by these two methods are not reliable enough to improve the retrieval results. For the comparison among QE methods, the performance gain of our method is better than other two methods, it shows that the restored image by our method is not only reliable, but also more discriminative to facilitate the retrieval task.

5.3 Qualitative Evaluation

Visual Quality Comparison. As shown in Figure 7, our method largely enhances the visibility of the scene behind the windshield, with clearer facial outlines of passengers (see the red regions in Figure 7) and also keeps the consistency of detailed visual information such as those small objects. (see the blue regions in Figure 7) . The three non-learning based methods [5, 16, 26] cannot effectively enhance the visibility of the background scenes (*e.g.*, the results

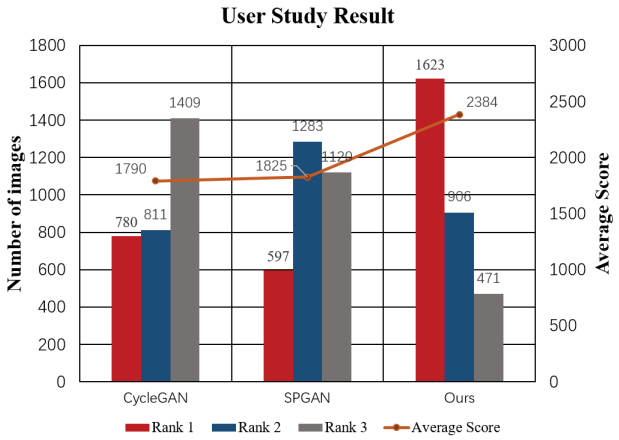

Figure 8: User study results on the WIRE dataset with three learning-based methods (CycleGAN [34], SPGAN [6] and WIRE-GAN (ours)). The statistics are obtained by collecting the ranking results from 30 participants and 100 images.

obtained by the non-learning based method in the third row of Figure 7). In some cases, they even introduce new artifacts to the input images (*e.g.*, the color shift in the sixth row of Figure 7). The weakly supervised methods (CycleGAN [34] and SPGAN [6]) can generate much better results than the non-learning based methods, however, when compared with our method, they still introduce new artifacts into the final estimated results (see the content distortions in the third and fourth row of Figure 7).

User Study Evaluation. As there's no well-established error metric specifically developed for the windshield restoration task, we conduct a restoration quality user study and invite 30 participants to evaluate the quality of 100 images randomly selected from the windshield dataset. Due to the obviously degenerated results from non-learning methods, in this user study we only focus on learning based approaches (CycleGAN [34], SPGAN [6], and WIRE-GAN). The user study is conducted with the following procedures:

(1) The participants are first trained with the common windshield images to gain a general sense on this task.
(2) Each participant is asked to view four images at a time, with the leftmost image showing the input degraded windshield image followed by three restored images by different methods, which are displayed in a random order. They are asked to rank the restoration quality without any time constraint. This test is performed on 100 groups of such quadruples.
(3) The average score ϕ for a certain method is calculated from the ranking as $\phi_k = \frac{1}{N}\sum_i\sum_j(N - rank_{i,j,k} + 1)$, where N is the total number of evaluated methods and i, j, k indicate the i-th participant, j-th group of images and k-th method, respectively.

The results in Figure 8 show that the rank-1 value of our method is even higher than the sum of the rest two methods and the rank-3 value of our method is obviously smaller, which demonstrates the

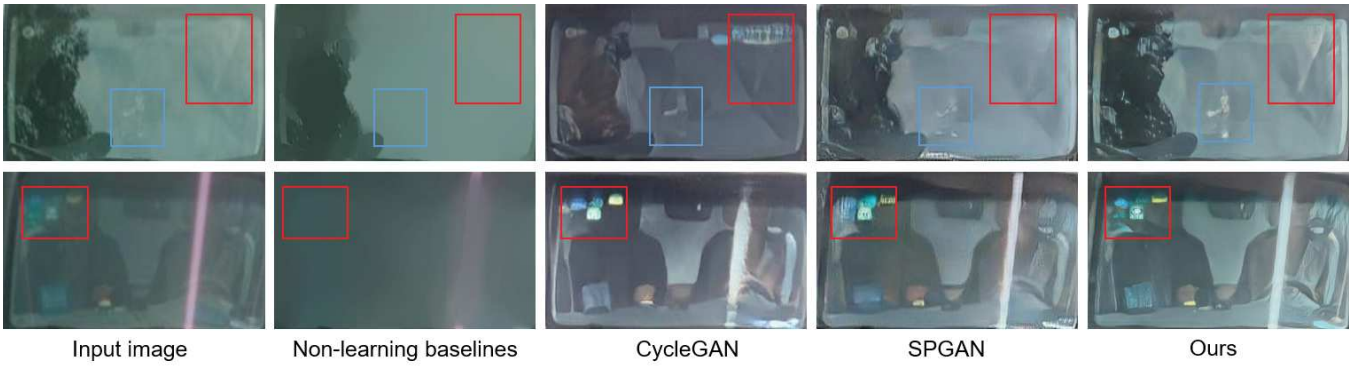

Figure 9: Failure cases with strong reflections and saturation caused by specular highlights, compared with non-learning specialized methods [1], CycleGAN [34] and SPGAN [6], in which the red and blue boxes indicate notable differences.

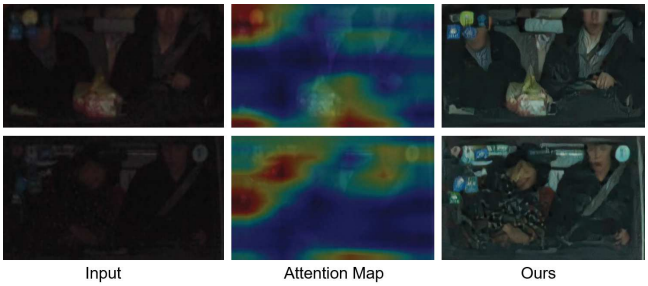

Figure 10: Visualization of the attention map from the content-aware discriminator. Here the annual inspection tags (top left corner) and the face of driver are the most representative regions and WIRE-GAN is able to maintain these ID-discriminative details and improve the visual quality.

superior perceptual quality of our method.

5.4 Visualization Analysis

As introduced above, we propose a content-aware discriminator to constrain the visual features of the restored windshield image consistent with the original input at the instance level. The attention map of such a discriminator is also of interest to investigate, because we can find which regions are more crucial in the retrieval of windshield image to distinguish the ID information.

As shown in Figure 10, the attention map is extracted from the last convolutional layers (pool5) which undergoes the embedding layer to generate final feature representation. The region with the most attention on the windshield is the top left corner where the annual inspection tags are pasted. Besides, it can be observed that the regions of the drivers' face and the accessories put behind the windshield are also highly informative for the discriminator. The preservation of the discriminative visual features during restoration explains why the restored results of our proposed method lead to retrieval performance improvement. This attention distribution of our discriminator is close to that of the human observers when identifying a vehicle merely with the windshield region to some extent. When human observers were asked to tell the differences between two windshield images, they would also probably first identify obvious different regions like annual inspection tags or drivers' faces. Since our content-aware discriminator has similar attentive regions as human, it may also benefit the human observers to more easily acquire the subtle details in windshield for human-in-the-loop analysis.

6 CONCLUSION AND DISCUSSION

We propose a new problem of windshield image restoration. By leveraging the weakly supervised learning framework based on the generative model, we propose a content-aware discriminator to learn the embedding space, which can effectively keep the content consistency of the restored windshield images. Over the newly collected large-scale windshield image dataset, our method has reported encouraging performance in terms of the perceptual quality and the accuracy of windshield ReID.

Limitations & future works. Although our method is supposed to cover a broad range of image degradation phenomena, the performance may drop when it comes to the specular highlight caused by glass reflections as shown in Figure 9. The limited enhancement in those difficult cases can be mainly attributed to the fact that some essential parts of the windshields are concealed by the reflection light, thus all the methods cannot perform inpainting in those regions properly. However, even in such extreme conditions, our method still recovers more information than others. The generalization of our method is yet to be improved when handing more diverse and challenging scenes in the future.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China under Grant 61661146005, Grant U1611461, and 61872012, in part by the Shenzhen Municipal Science and Technology Program under Grant JCYJ20170818141146428, and in part by the National Research Foundation, Prime Minister's Office, Singapore, through the NRF-NSFC Grant, under Grant NRF2016NRF-NSFC001-098.

REFERENCES

[1] Nikolaos Arvanitopoulos, Radhakrishna Achanta, and Sabine Süsstrunk. 2017. Single Image Reflection Suppression. In *Proc. of Computer Vision and Pattern Recognition (CVPR).*

[2] Sagie Benaim and Lior Wolf. 2017. One-sided unsupervised domain mapping. In *Advances in Neural Information Processing Systems (NIPS).*

[3] Konstantinos Bousmalis, Nathan Silberman, David Dohan, Dumitru Erhan, and Dilip Krishnan. 2017. Unsupervised Pixel-Level Domain Adaptation With Generative Adversarial Networks. In *Proc. of Computer Vision and Pattern Recognition (CVPR).*

[4] Jingwen Chen, Jiawei Chen, Hongyang Chao, and Ming Yang. 2018. Image Blind Denoising With Generative Adversarial Network Based Noise Modeling. In *Proc. of Computer Vision and Pattern Recognition (CVPR).*

[5] Kostadin Dabov, Alessandro Foi, Vladimir Katkovnik, and Karen Egiazarian. 2007. Image denoising by sparse 3-D transform-domain collaborative filtering. *IEEE Transactions on Image Processing* 16, 8 (2007).

[6] Weijian Deng, Liang Zheng, Qixiang Ye, Guoliang Kang, Yi Yang, and Jianbin Jiao. 2018. Image-Image Domain Adaptation With Preserved Self-Similarity and Domain-Dissimilarity for Person Re-Identification. In *Proc. of Computer Vision and Pattern Recognition (CVPR).*

[7] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proc. of Computer Vision and Pattern Recognition (CVPR).*

[8] Xun Huang, Ming-Yu Liu, Serge Belongie, and Jan Kautz. 2018. Multimodal Unsupervised Image-to-Image Translation. In *Proc. of European Conference on Computer Vision (ECCV).*

[9] Andrey Ignatov, Nikolay Kobyshev, Radu Timofte, Kenneth Vanhoey, and Luc Van Gool. 2017. WESPE: weakly supervised photo enhancer for digital cameras. *arXiv preprint arXiv:1709.01118* (2017).

[10] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. 2017. Image-to-image translation with conditional adversarial networks. *arXiv preprint* (2017).

[11] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. 2016. Perceptual losses for real-time style transfer and super-resolution. In *Proc. of European Conference on Computer Vision (ECCV).*

[12] Taeksoo Kim, Moonsu Cha, Hyunsoo Kim, Jung Kwon Lee, and Jiwon Kim. 2017. Learning to discover cross-domain relations with generative adversarial networks. In *Proceedings of the 34th International Conference on Machine Learning-Volume 70*. JMLR. org, 1857–1865.

[13] Idan Kligvasser, Tamar Rott Shaham, and Tomer Michaeli. 2018. xUnit: Learning a Spatial Activation Function for Efficient Image Restoration. In *Proc. of Computer Vision and Pattern Recognition (CVPR).*

[14] Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, and Jiri Matas. 2017. DeblurGAN: Blind Motion Deblurring Using Conditional Adversarial Networks. *arXiv preprint arXiv:1711.07064* (2017).

[15] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew P Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. 2017. Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network.. In *Proc. of Computer Vision and Pattern Recognition (CVPR)*, Vol. 2.

[16] Mading Li, Jiaying Liu, Wenhan Yang, Xiaoyan Sun, and Zongming Guo. 2018. Structure-Revealing Low-Light Image Enhancement Via Robust Retinex Model. *IEEE Transactions on Image Processing* 27, 6 (2018).

[17] Hongye Liu, Yonghong Tian, Yaowei Wang, Lu Pang, and Tiejun Huang. 2016. Deep relative distance learning: Tell the difference between similar vehicles. In *Proc. of Computer Vision and Pattern Recognition (CVPR).*

[18] Ming-Yu Liu, Thomas Breuel, and Jan Kautz. 2017. Unsupervised image-to-image translation networks. In *Advances in Neural Information Processing Systems (NIPS).*

[19] M. Saquib Sarfraz, Arne Schumann, Andreas Eberle, and Rainer Stiefelhagen. 2018. A Pose-Sensitive Embedding for Person Re-Identification With Expanded Cross Neighborhood Re-Ranking. In *Proc. of Computer Vision and Pattern Recognition (CVPR).*

[20] Ashish Shrivastava, Tomas Pfister, Oncel Tuzel, Joshua Susskind, Wenda Wang, and Russell Webb. 2017. Learning From Simulated and Unsupervised Images Through Adversarial Training. In *Proc. of Computer Vision and Pattern Recognition (CVPR).*

[21] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).

[22] Yaniv Taigman, Adam Polyak, and Lior Wolf. 2017. Unsupervised cross-domain image generation. *International Conference on Learning Representations (ICLR)* (2017).

[23] Dmitry Ulyanov, Andrea Vedaldi, and Victor S. Lempitsky. 2016. Instance Normalization: The Missing Ingredient for Fast Stylization. *CoRR* abs/1607.08022 (2016).

[24] Ting-Chun Wang, Ming-Yu Liu, Jun-Yan Zhu, Andrew Tao, Jan Kautz, and Bryan Catanzaro. 2018. High-Resolution Image Synthesis and Semantic Manipulation With Conditional GANs. In *Proc. of Computer Vision and Pattern Recognition (CVPR).*

[25] Yandong Wen, Kaipeng Zhang, Zhifeng Li, and Yu Qiao. 2016. A discriminative feature learning approach for deep face recognition. In *Proc. of European Conference on Computer Vision (ECCV).*

[26] Li Xu, Shicheng Zheng, and Jiaya Jia. 2013. Unnatural l0 sparse representation for natural image deblurring. In *Proc. of Computer Vision and Pattern Recognition (CVPR).*

[27] Noam Yair and Tomer Michaeli. 2018. Multi-Scale Weighted Nuclear Norm Image Restoration. In *Proc. of Computer Vision and Pattern Recognition (CVPR).*

[28] Linjie Yang, Ping Luo, Chen Change Loy, and Xiaoou Tang. 2015. A large-scale car dataset for fine-grained categorization and verification. In *Proc. of International Conference on Computer Vision (ICCV).*

[29] Xitong Yang, Zheng Xu, and Jiebo Luo. 2018. Towards perceptual image dehazing by physics-based disentanglement and adversarial training. In *The Thirty-Second AAAI Conference on Artificial Intelligence (AAAI).*

[30] Ke Yu, Chao Dong, Liang Lin, and Chen Change Loy. 2018. Crafting a Toolchain for Image Restoration by Deep Reinforcement Learning. In *Proc. of Computer Vision and Pattern Recognition (CVPR).*

[31] Yuhui Yuan, Kuiyuan Yang, and Chao Zhang. 2017. Hard-aware deeply cascaded embedding. In *Proc. of International Conference on Computer Vision (ICCV).*

[32] Xinyuan Zhang, Xin Yuan, and Lawrence Carin. 2018. Nonlocal Low-Rank Tensor Factor Analysis for Image Restoration. In *Proc. of Computer Vision and Pattern Recognition (CVPR).*

[33] Xiaofan Zhang, Feng Zhou, Yuanqing Lin, and Shaoting Zhang. 2016. Embedding label structures for fine-grained feature representation. In *Proc. of Computer Vision and Pattern Recognition (CVPR).*

[34] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. 2017. Unpaired image-to-image translation using cycle-consistent adversarial networks. *Proc. of International Conference on Computer Vision (ICCV)* (2017).