Benchmarking Shadow Removal for Facial Landmark Detection

Lan Fu[1], Qing Guo[2†], Felix Juefei-Xu[3], Hongkai Yu[4], Yang Liu[5], Wei Feng[6], Song Wang[1]

[1]University of South Carolina, USA
[2]IHPC and CFAR, Agency for Science, Technology and Research (A*STAR), Singapore
[3]New York University, USA [4]Cleveland State University, USA
[5]Nanyang Technological University, Singapore [6]Tianjin University, China

***Abstract*—Facial landmark detection is a very fundamental task and its accuracy plays a significant role for many downstream face-related vision applications. In practice, the facial landmark detection can be affected by a lot of natural degradations. One of the most common and important degradations is the shadow caused by light source being blocked with an external occluder. While many advanced shadow removal methods have been proposed to restore the image quality in recent years, their effects on facial landmark detection are not well studied. For example, it remains unclear whether the shadow removal could enhance the robustness of facial landmark detection to diverse shadow patterns or not. In this work, for the first time, we construct a novel benchmark (*i.e.*, SHAREL) to link the two independent but relatable tasks (*i.e.*, shadow removal and facial landmark detection). In particular, SHAREL covers diverse face shadows with different intensities, sizes, shapes, and locations. Moreover, to mine hard shadow patterns against facial landmark detection, we propose a novel method (*i.e.*, adversarial shadow attack), which allows us to construct a challenging subset of the benchmark for a comprehensive analysis. With the constructed benchmark, we conduct extensive analysis on three state-of-the-art shadow removal methods and three landmark detectors. We observed a highly positive correlation between shadow removal and facial landmark detection tasks, which probably will provide insight to improve the robustness of the facial landmark detection in the future.**

***Index Terms*—Face shadow, shadow removal, Facial Landmark detection**

I. Introduction

Facial landmark detection (1; 2; 3) is a fundamental step for numerous facial related applications, *e.g.*, face recognition and verification (4; 5), 3D face reconstruction (6), and safety-critical applications, *e.g.*, deepfake detection (7; 8). While recent deep-learning techniques bring us continuously improved landmark-detection performance, most of them are designed to handle images of "clean faces". However, in real-world applications, face images usually contain image degradations, such as noise (9; 10), shadow (11; 12), and haze (13), which degrades the aesthetic quality of images directly and may further affect the performance of landmark detectors.

As a natural phenomenon, shadow is very common on face images – in practice, the light to any face region can be occluded by surrounding objects or by part of itself. This is especially true for portrait images captured in the wild with unconstrained environments. As shown in Fig. 1(a), portrait shadow happens in two kinds of scenes: foreign shadow and facial shadow. Foreign shadow appears when there is an external occluder (*e.g.*, a tree or a hat brim) blocking the light source to reach out to the subjects' face. Foreign shadow presents an arbitrary 2D shape in the natural image, relying on the shape of the occluder and position of the light source. In contrast, facial shadow casting on the face by the face itself due to the facial geometry presents a small space of 2D shapes when natural lighting is not perfectly uniform (17).

The foreign shadow effects on facial landmark tasks are under-explored, although there are works (11; 12) exploring illumination invariance for face recognition which cannot be simply extended to facial landmark detection. Such works are mainly designed for facial shadows caused by the intensity and position of the light source in the indoor scene. Note that foreign shadows are almost always distracting compared to facial shadows. Image intensity edges appear in most of foreign shadow scenes, which are uncorrelated to facial geometry and will obfuscate facial 3D structure. By contrast, the intensity edges introduced by facial shadows are more likely to be helpful for inferring the shape of face. Therefore, we aim to remove the foreign shadow entirely. In this work, we set the research scope to tackle foreign shadows. As a result of shadow cast, spatial-variant illumination and color distortion in the shadow region (18) degrade the image quality and undermine the image features significantly. As shown in Fig. 1(b), shadowed faces hurt the image quality with large root mean square errors (RMSEs), and present unreasonable and much deteriorated landmark locations at the eyebrows (See Case1) and mouth (See Case2), as measured by much degraded NME scores.

An intuition way to alleviate the performance loss caused by shadow is to restore the underlying shadow-free image utilizing current state-of-the-art (SOTA) shadow removal methods. However, there are two challenges posing to such a solution: ❶ The interplay between light, occluder, and the subject directly affects the shadow appearance. As a result, in the

This research is supported by the National Research Foundation, Singapore, and DSO National Laboratories under the AI Singapore Programme (AISG Award No: AISG2-GC-2023-008). †Corresponding author: Qing Guo (tsingqguo@ieee.org).

979-8-3503-5409-6/24/$31.00 ©2024 IEEE
DOI 10.1109/CAI59869.2024.00059

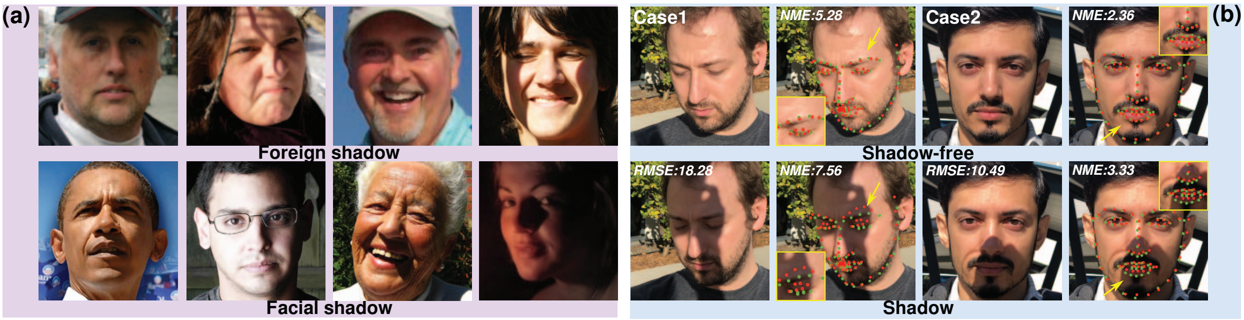

Figure 1: Illustrations of (a) various shadow scenes on facial landmark detection benchmarks (14; 15) and (b) effects of foreign shadow on image quality and facial landmark detection (16). Red: prediction. Green: ground truth. RMSE measures the image degradation caused by shadow, and NME evaluates the detection error.

real world, shadow patterns are significantly diverse, which increases the difficulty of shadow removal algorithms. ❷ Even though shadow removal methods could obtain high visual-quality images with lower RMSE, the landmark detection performance may even get worse compared to that of shadow images due to the potential domain shift between landmark detection and image quality enhancement. Existing works (haze (19) and rain (20) removal) demonstrate that visual quality improvement benefits little or even hurts the high-level perception task performance. All above facts motivate us to answer two basic questions: how shadow affects the landmark detection, and whether shadow removal can benefit the robustness of landmark detectors.

To this end, for the first time, we propose to link the two seemingly independent but intrinsically related tasks, *i.e.*, shadow removal and facial landmark detection, by constructing a totally novel dataset and benchmark. There are few studies on this topic in both communities before this work. Note that, constructing such a benchmark is challenging and not trivial since the shadow patterns are not exhaustive, and *existing facial landmark detection benchmarking techniques* (14; 15) *collecting natural images cannot meet the requirements*: ❶ There are about less than 2% of data in each benchmark (14; 15) presenting foreign shadow scenes, which is not enough for a comprehensive study on the robustness evaluation. ❷ For those foreign shadow samples, as shown in Fig. 1 (a), though with increasing shadow intensity, they primarily exhibit less abrupt edges and their shadow patterns are limited.

To alleviate these challenges, we propose novel solutions to ensure the comprehensiveness: ❶ We employ the physical model of shadow and synthesize facial shadow images by considering four common factors (*i.e.*, intensity, size, shape, and location) with three severities, ❷ We investigate the shadow from the perspective of adversarial attack and propose a totally new attack (*i.e.*, adversarial shadow attack) to identify shadow patterns that are more challenging to landmark detection. ❸ We introduce a real-world shadow face dataset for verifying the generalization ability of facial landmark detectors. With these elaborated designs, we can quantitatively and systematically study the effect of shadows on the facial landmark detection.

Overall, we summarize our contributions as follows:

- We construct a synthetic shadow-face dataset by comprehensively considering shadow intensity, size, shape, and location to analyze the effects of shadow on image visual quality and facial landmark detection.
- We proposed a novel adversarial shadow-face dataset to cover more challenging shadow patterns for facial landmark detectors. We also explore a real shadow dataset to verify and improve the facial landmark detectors' robustness.
- With the three subsets, togethter as SHAREL, we comprehensively and quantitatively study the effects of shadow and shadow removal to image visual quality and the performance of facial landmark detection.

II. Datasets Construction

A. Overview

Natural shadow presents diverse shadow patterns in the wild due to the influences of occluders and light sources. For example, different light occluders can lead to diverse shadow appearances with different sizes and shapes. In addition, the illumination level, material of occluders and object surface where shadow casts determine the reflection and scattering of the light, which may affect the intensity at the shadow region. Nevertheless, enumeration of all permutations formulating patterns is not practical due to dynamic and complex scenes. To alleviate this issue and analyze the effects of shadow and shadow removal on facial landmark detection extensively, we propose three dataset construction strategies: ❶ We follow the well-known physical shadow model to synthesize shadowed faces on the clean facial landmark detection dataset (*i.e.*, 300W (14)) and consider four factors (*i.e.*, intensity, size, shape, and location) with three severities (See Sec. II-B). ❷ To mine hard shadow images that affect landmark detection easily, we think this problem from the perspective of adversarial attack and propose a novel synthesis method (*i.e.*, *adversarial shadow attack*) in Sec. II-C. ❸ To address the potential shifting problem between synthesized shadow faces and the real ones, we introduce 100 real shadow face images as a subset of the whole dataset (See Sec. II-D).

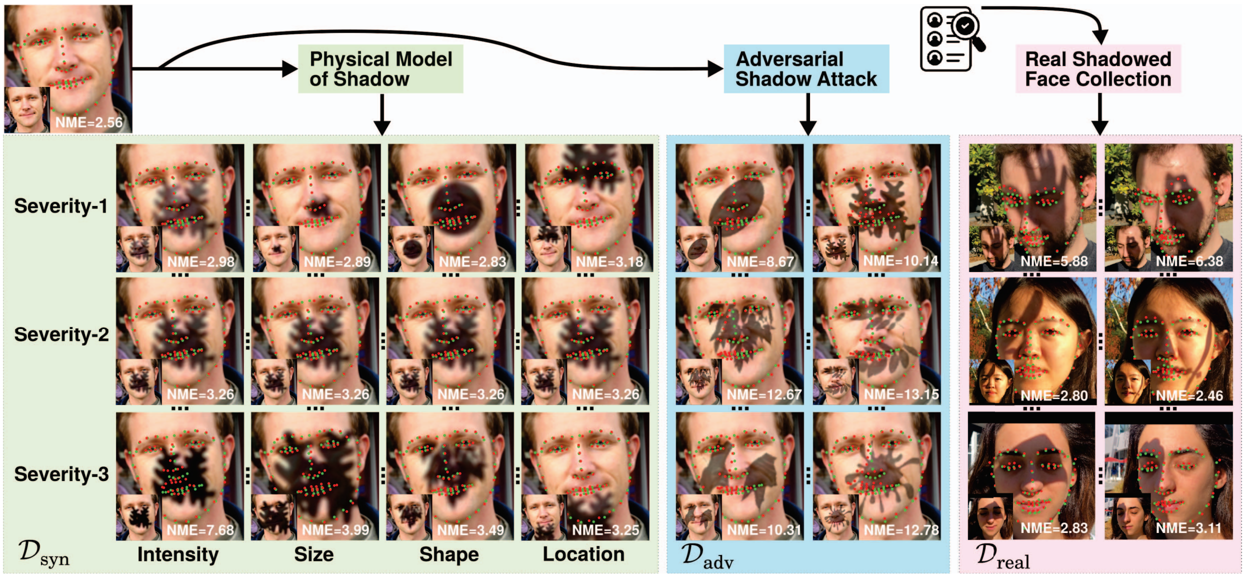

Figure 2: Three dataset construction strategies including physical model-based synthesis (Sec. II-B), adversarial shadow attack (Sec. II-C), and real shadowed face collection (Sec. II-D). Green: ground truth. Red: prediction. NME measures the landmark detection performance. The lower, the better.

B. Synthetic Shadowed Faces

Physical model of shadow. We adopt the well-known and widely used physical model of shadow in (21). Specifically, following the illumination and reflectance formulation of an image (21), we can represent a clean (*i.e.*, shadow-free) image captured under a single primary light source as

$$\mathbf{I}_p^{\text{cln}} = \mathbf{L}_p \mathbf{R}_p = (\mathbf{L}_p^{\text{d}} + \mathbf{L}_p^{\text{a}})\mathbf{R}_p, \tag{1}$$

where $\mathbf{I}_p^{\text{cln}}$, $\mathbf{L}_p$, and $\mathbf{R}_p$ are pixel intensity, illumination, and reflectance at the p-th pixel, respectively. The illumination stems from two sources, *i.e.*, the direct illumination $\mathbf{L}^{\text{d}}$ and the ambient illumination $\mathbf{L}^{\text{a}}$. When an occluder appears in front of the light source, the direct illumination disappears while the ambient illumination is also affected. We can represent the p-th shadowed pixel as

$$\mathbf{I}_p^{\text{shd}} = \alpha \mathbf{L}^{\text{a}} \mathbf{R}_p = \alpha(\mathbf{I}_p^{\text{cln}} - \mathbf{L}_p^{\text{d}} \mathbf{R}_p), \tag{2}$$

where α is a scalar and determines the attenuation of the ambient illumination, which is caused by the occluder. With a clean image $\mathbf{I}^{\text{cln}}$ and a dark image $\mathbf{I}^{\text{shd}}$, we can represent an image $\mathbf{I}$ containing a shadow region, following (22; 23), as

$$\mathbf{I} = \mathbf{I}^{\text{shd}} \odot \mathbf{M} + \mathbf{I}^{\text{cln}} \odot (1 - \mathbf{M}), \tag{3}$$

where $\mathbf{M}$ is a binary map that defines the shadow region and is determined by the occluder. To generate more realistic shadow, we reformulate Eq. (3) to

$$\mathbf{I} = \mathbf{I}^{\text{shd}} \odot \rho(\mathbf{D} \odot \mathbf{M}) + \mathbf{I}^{\text{cln}} \odot (1 - \rho(\mathbf{D} \odot \mathbf{M})), \tag{4}$$

where ρ models light scattering and spatial variation and $\mathbf{D}$ is a face depth map. Note that, the images of living faces have face-like depth information, which are critical for anti-spoofing application. $\mathbf{D}$ can make generated shadow more realistic, which is not considered in previous shadow models (22; 23). Moreover, to generate realistic shadow pattern, we borrow the implementation in (22; 24) and use the function ρ to render the depth-aware mask (*i.e.* $\mathbf{D} \odot \mathbf{M}$) to become a shadow matte image by modeling the light scattering beneath human skin and modeling the spatial variation of the shadow via a spatially-varying blur. Please find more details in (22).

Then, we can substitute Eq. (2) into Eq. (4) and get

$$\begin{aligned} \mathbf{I} = &\ \text{Shadow}(\mathbf{I}^{\text{cln}}, \mathbf{M}, \alpha) \\ = &\ (1 - (1 - \alpha)\rho(\mathbf{D} \odot \mathbf{M}))\mathbf{I}^{\text{cln}} + \alpha\beta\rho(\mathbf{D} \odot \mathbf{M}), \end{aligned} \tag{5}$$

where $\beta = -\mathbf{L}^{\text{d}}\mathbf{R}$ representing the response of the camera to the reflected direct illumination and the ambient attenuation α does not depend on the light source (*e.g.*, wavelength) (21). Moreover, as demonstrated in (23), β is a three-channel vector and can be estimated from the α via a linear transformation.

Overall, *given a clean face image* $\mathbf{I}^{\text{cln}}$*, a shadow map* $\mathbf{M}$*, a depth map* $\mathbf{D}$*, and the* α*, we can synthesize a shadowed face* $\mathbf{I}$. In practice, we use the 3DDFA-V2 (25) to predict the depth map from the clean image.

Synthesized shadows with different factors and severities. To cover extensive shadow patterns in the real world, we generate shadowed faces for a clean face image from four factors: intensity, size, shape, and location.

i. *Intensity.* The illumination level and material of object surfaces determine the reflection and scattering of light, resulting in shadow with diverse intensities. We model the shadow intensity via the parameter α in Eq. (5) since it directly models the relationship between shadowed pixels and illuminated pixels. α is about in range [0.0, 1.0) for realistic shadow scene (23). We uniformly sample α from ranges $[0.8, 1.0)$, $[0.4, 0.6)$, $[0.0, 0.2)$, for light, medium and heavy shadows. The lower α, the heavier the shadow. For different shadow intensity level design, we want to quantify how much texture and content degradation shadow brings, and how that affects visual quality and landmark detection. We present three kinds of intensities for the same face in Fig. 2.

ii. *Size.* The size of an occluder blocking the light and position of the light source directly affect the area of the shadow (*i.e.*, shadow size). We model shadow size via the number of non-zero pixels in $\mathbf{M}$ in Eq. (5) and consider three different severities, *i.e.*, small, medium, and large shadow regions. Intuitively, large-size shadow will degrade image quality more than small-size shadow because face-related information (*e.g.*, structure) becomes less. Given a specified shadow shape, we can set the shadow areas (*i.e.*, number of non-zero pixels in $\mathbf{M}$) to take up $10\% \sim 20\%$, $45\% \sim 55\%$, and $80\% \sim 90\%$ areas of the face images by rescaling the shadow region in $\mathbf{M}$, which corresponds to three severities, *i.e.*, small, medium, and large shadow regions. We show the three different shadow sizes for the same face in Fig. 2.

iii. *Shape.* Occluders with different 3D geometrical shapes and the lights with different positions relative to the same occluder also affect the shadow shapes. We represent the shadow shape via the shadow mask in $\mathbf{M}$ in Eq. (5). To cover diverse shadow shapes, we collect a silhouette dataset containing 132 shapes of natural objects, and classify them into three levels by a shape complexity metric defined in (26), which is denoted as E. The shape complexity metric considers two aspects during measurement, *i.e.*, the distance distribution of the contour points of a shape to its centroid and the smoothness of the contour. Intuitively, if the complexity of a shape is low, the shape may tend to be a circle or has smooth contour. We present three shapes for the same face in the Fig. 2, their complexity values are 0.04, 0.10, and 0.15 from severity 1 to 3. With the collected silhouette dataset, we first calculate the shape complexity for each collected shape. Then, we sort all shapes according to the complexity and evenly divide them into three severities, *i.e.*, low, medium, and high complexities.

iv. *Location.* We further consider the shadow position in the face image due to the facial geometry. For example, facial landmarks include clues of eyebrows, eyes, nose, jaw, and mouth. Shadow degradation to different parts of the facial structure will help quantitatively recognize the importance of each structural information to landmark detection. We shift the centroid point of the shadow mask in $\mathbf{M}$ to the center of the three regions.

Synthetic shadowed face subset $\mathscr{D}_{\text{syn}}$. With the above synthesis strategies, given a clean face image, we can generate three shadowed faces for each factor, which corresponds to three severities. We have $3^4 = 81$ shadowed faces across all factors and severities for each clean image. Then, based on the facial landmark dataset 300W (14) that contains 689 clean face images for testing landmark detectors, we can generate a larger dataset with $81 \times 689 = 55,809$ shadowed images. We present some examples in Fig. 2. Although the constructed dataset covers diverse shadow patterns, it cannot represent all possible situations, in particular, the hard cases that SOTA landmark detectors cannot address. To alleviate this issue, we further propose a novel adversarial attack in Sec. II-C to mine the hard shadow patterns.

C. Adversarially Shadowed Faces

Given an image, adversarial attack is to calculate an imperceptible noise-like perturbation under the guidance of a targeted deep model, and then add it to the image. As a result, the corrupted image can mislead the targeted model easily. We can regard the adversarial attack as a way to mine hard noise patterns that cannot be addressed by the targeted deep model. Here, we propose a novel attack method, *i.e.*, *adversarial shadow attack*, and further extend it to generate hard shadow patterns that are able to fool the landmark detectors. Therefore, we can evaluate the shadow robustness.

Intuitively, we can tune the physical parameters of shadow model, *i.e.*, $\text{Shadow}(\mathbf{I}^{\text{cln}}, \mathbf{M}, \alpha)$ in Eq. (5), like the α and $\mathbf{M}$ under the supervision of landmark detectors to cover different shadow patterns with different intensities, sizes, shapes, and locations. Specifically, given a clean face image $\mathbf{I}^{\text{cln}}$ and a pre-trained landmark detector $\varphi(\cdot)$ we want to evaluate, we can: 1) First use Eq. (5) to synthesize the shadowed image, and feed it to $\varphi(\cdot)$. 2) We get the detection results and calculate the loss according to the ground truth (*i.e.*, $\mathbf{y}$). 3) We tune the physical variables $\mathbf{M}$ and α iteratively to maximize the landmark detection loss. As a result, the synthesized face can fool the detection easily while maintaining the physical properties of the shadow. We can formulate the above process by

$$\arg\max_{\mathbf{M},\alpha,\vartheta} \mathscr{J}(\varphi(\text{Shadow}(\mathbf{I}^{\text{cln}}, \text{Aff}_{\vartheta}(\mathbf{M}), \alpha)), \mathbf{y}), \tag{6}$$

subject to $\|\mathbf{M} - \mathbf{M}_0\|_{\text{p}} < \varepsilon_{\text{M}}, \|\alpha - \alpha_0\| < \varepsilon_{\alpha}, \|\vartheta - \vartheta_0\|_{\text{p}} < \varepsilon_{\vartheta}$,

where $\mathscr{J}(\cdot)$ is the loss function of landmark detection.

Different from the raw synthesis function in Eq. (5), we conduct the affine transformation (*i.e.*, $\text{Aff}_{\vartheta}(\cdot)$) on $\mathbf{M}$ before feeding it for synthesis, which allows us to mine more shadow shapes with a given shadow mask. The ϑ contains six affine parameters. Like general adversarial attack methods, we set the L_{p} norm to $\mathbf{M}$, α, and ϑ to force the optimization space within a ball of ε_{M}, ε_{α}, and ε_{ϑ}, around their initialization (*i.e.*, $\mathbf{M}_0$, α_0, and ϑ_0), respectively.

To solve the Eq. (6), we follow the general adversarial attack methods: ❶ We set $\mathbf{M}_0$, α_0, and ϑ_0, and get the initial synthesized image. ❷ We feed the generated image to the landmark detector $\varphi(\cdot)$ and calculate the loss. ❸ We conduct back-propagation and get the gradients of $\mathbf{M}$, α, and ϑ w.r.t. the loss function. ❹ We calculate the sign of the gradients and use them to update the three variables by multiplying the gradients with three step sizes. ❺ We generate a new synthesized image and loop step-2 to step-4 for a number of iterations. In terms of the initialization, we select $\mathbf{M}_0$ from the collected 132 silhouette images and set α_0 to be 0.8. Then, we initialize ϑ_0 as $\begin{bmatrix} 1.0 & 0.0 & 0.0 \\ 0.0 & 1.0 & 0.0 \end{bmatrix}$, $\text{Aff}_{\vartheta}(\mathbf{M}) = \mathbf{M}$ during initialization. We set the step size of α, ϑ, and $\mathbf{M}$ as 0.01, 0.02, and 0.0012, respectively. The number of iterations is set to be 40. we use ∞ norm for L_{p}, and set ε_{α}, ε_{ϑ}, and ε_{M} as 0.4, 0.8, and 0.048, respectively. As a result, adversarial shadow images can present more hard shadow patterns against landmark detectors, as shown in Fig. 2, the NMEs in $\mathscr{D}_{\text{adv}}$ could be over 10 compared to around 3 in $\mathscr{D}_{\text{syn}}$.

Adversarially shadowed face subset $\mathscr{D}_{\text{adv}}$. With the above method, given a landmark detector and the 300W dataset, we first conduct attack for each image, and then evaluate the detector on the adversarially shadowed faces. Thus, for each detector, we have an exclusive new version of 689 adversarially shadowed face images to evaluate their robustness.

D. Real Shadowed Faces

Real shadowed face subset $\mathscr{D}_{\text{real}}$. To verify the shadow effect on visual quality and landmark detection in the real-world scenario, we introduce a real-world shadow portrait dataset (22). However, this dataset lacks facial landmark annotations for landmark detection evaluation. We first obtain pseudo ground truth by a SOTA pre-trained HRNet (27), and then refine it manually as the final landmark ground truth. Finally, we have 9 subjects and 100 pairs of shadowed and shadow-free portrait images captured in the outdoor scenes with varied face poses, shadow shapes, and illumination conditions. Figure 2 presents some examples.

III. Shadow Removal & Landmark Detection Benchmark (SHAREL)

A. Setups

Datasets. As introduced in Sec. II, our main data is constructed based on the landmark detection benchmark 300W (14). 300W contains $3,148$ face images for training and 689 images for testing, where almost all images in 300W are clean (shadow-free). Each image is labeled with 68 landmarks. We construct SHAREL based on the testing dataset of 300W. We add shadow patterns to the 300W and get $\mathscr{D}_{\text{syn}}$; we propose adversarial shadow attack and obtain $\mathscr{D}_{\text{adv}}$ for each landmark detector; we collect real shadowed faces (*i.e.*, $\mathscr{D}_{\text{real}}$) to further enrich our dataset. Finally, our dataset $\{\mathscr{D}_{\text{syn}};\mathscr{D}_{\text{adv}};\mathscr{D}_{\text{real}}\}$ has $\{55,809;689;100\}$ shadowed and shadow-free image pairs (total $56,598$ pairs) that are labeled with 68 landmarks.

We additionally construct $\{\mathscr{D}^{\text{t}}_{\text{syn}},\mathscr{D}^{\text{t}}_{\text{adv}}\}$ from a randomly selected subset (1,500 clean images) of 300W training set for training shadow removal models. Each of $\{\mathscr{D}^{\text{t}}_{\text{syn}},\mathscr{D}^{\text{t}}_{\text{adv}}\}$ contains 1,500 shadow-free and shadowed image pairs. For $\mathscr{D}^{\text{t}}_{\text{syn}}$, each clean image uniformly selects a severity for each factor to generate the shadow image. $\mathscr{D}^{\text{t}}_{\text{adv}}$ follows the same shadow generation way of $\mathscr{D}_{\text{adv}}$.

Metrics. To clarify the shadow and deshadow effect on image quality, we adopt the Root Mean Square Error (RMSE) metric in LAB color space for evaluation, similar to (18; 28; 29). For facial landmark detection evaluation, we adopt Normalized Mean Error (NME) metric with inter-ocular distance as normalization strategy following (16; 27; 30). Both the lower, the better.

Evaluated methods. With our SHAREL, we can evaluate the quality restoration capability of the shadow removal methods and the detection accuracy of facial landmark detectors on different shadow or deshadowed patterns. We first analyze three SOTA facial landmark detectors, *i.e.*, SAN (16), HRNet (27), and LUVLi (30), under different shadow patterns. All landmark detectors are pre-trained on clean face images. Further, we utilize three SOTA deep shadow removal methods, *i.e.*, MaskShadow-GAN (29), SP+M-Net (28), and AEFNet (18), to handle the shadowed faces in SHAREL and discuss whether and how these methods can help improve landmark detection performance. All shadow removal algorithms are trained on dataset $\mathscr{D}^{\text{t}}_{\text{syn}}$ and $\mathscr{D}^{\text{t}}_{\text{adv}}$ separately for fair comparison, and shadow removal models trained on $\mathscr{D}^{\text{t}}_{\text{syn}}$ are also utilized to test on real data.

B. Evaluation Results and Discussion

How does shadow affect visual quality and facial landmark detection? In Fig. 3(a-c), we report the RMSEs of shadow images and landmark detection results with NMEs in $\{\mathscr{D}_{\text{syn}},\mathscr{D}_{\text{adv}},\mathscr{D}_{\text{real}}\}$ to identify the shadow degradation on image quality and detection performance. Fig. 3(d-g) report the shadow pattern analysis on $\mathscr{D}_{\text{syn}}$ with four factors. The detector adopted in (d-g) is SAN (16). The results show that: ❶ Compared with shadow-free images, shadow images have high RMSEs since the shadow harms the image quality significantly. More intense the shadow degradation, worse the visual quality. For example, the RMSE of shadow and shadow-free images of large-size with 15.52 is higher than that of small-size with 2.74 in $\mathscr{D}_{\text{syn}}$ (Fig. 3(e)). Intensity, size, and location, instead of shape, are dominant factors affecting the shadow degradation. ❷ According to the NME results, we observe that: the performance of all landmark detectors drops when shadow appears in images and hard shadow pattern, *i.e.*, higher-severity shadow and adversarial shadow, hurts the detection task most. Specifically, the landmark detector SAN (16) achieves 4.05 NME on clean images of $\mathscr{D}_{\text{adv}}$, while the NME of shadow images increases by 152.3% to 10.22 (Fig. 3(b)). In $\mathscr{D}_{\text{syn}}$, heavy-intensity shadow achieves 6.26 NME with 54.7% performance drop compared to NME of clean images, while the performance loss caused by light-intensity shadow is 10.2% by SAN (16) (Fig. 3(d)).

In summary, shadow hurts the image quality and landmark detection significantly. Higher-severity presents high degradation capacity, that is, two tasks suffer from larger performance loss with increasing RMSEs and NMEs.

How does shadow removal affect visual quality and facial landmark detection? We perform shadow removal on shadow images, and present RMSEs and NMEs of shadow-removed images in $\{\mathscr{D}_{\text{syn}},\mathscr{D}_{\text{adv}},\mathscr{D}_{\text{real}}\}$ to evaluate the effectiveness of shadow removal methods. The results are shown in the Fig. 3. We can observe that: ❶ Shadow removal methods present different capabilities on the image quality enhancement (Fig. 3(a-c)). To be specific, SP+M-Net (28) and AEFNet (18) can enhance the image quality significantly in all subsets. MaskShadow-GAN (29) further hurts the quality in the subsets $\{\mathscr{D}_{\text{syn}},\mathscr{D}_{\text{real}}\}$ while achieving counterpart result in $\mathscr{D}_{\text{adv}}$. The

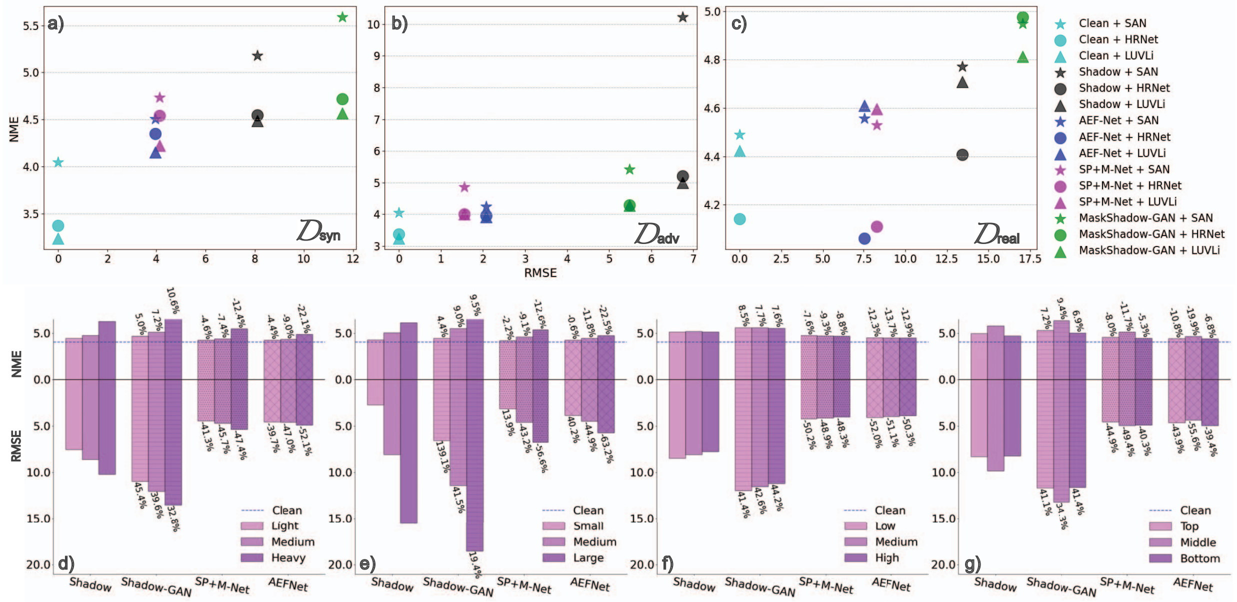

Figure 3: Shadow removal and landmark detection performance on SHAREL. (a-c): shadow removal (RMSE) and landmark detection (NME) results of $\{\mathscr{D}_{\text{syn}}, \mathscr{D}_{\text{adv}}, \mathscr{D}_{\text{real}}\}$ subsets, respectively. Each color represents results on shadow-free images (*e.g.*, Clean+*), shadow images (*i.e.*, Shadow+*), and shadow-removed images with three shadow removal methods (*e.g.*, AEFNet/SP+M-Net/MaskShadow-GAN(Shadow-GAN)). Different icon shapes represent different landmark detectors. (d-g): shadow pattern analysis of landmark detection (NME) and shadow removal (RMSE) results of $\mathscr{D}_{\text{syn}}$ for **intensity** (d), **size** (e), **shape** (f), and **location** (g). Blue dash line represents the result on clean images by the pre-trained landmark detector SAN (16). Each group represents results on shadow images (*i.e.*, Shadow), and shadow-removed images with three shadow removal methods (*e.g.*, Shadow-GAN/SP+M-Net/AEFNet). Each color represents a severity type. Relative performance gain, *i.e.*, the percent of NME/RMSE drops, after shadow removal compared to shadow images is listed.

former mainly stems from that MaskShadow-GAN, *i.e.*, a GAN-based image translation method, introduces artifacts during training. The reason why MaskShadow-GAN performs better on $\mathscr{D}_{\text{adv}}$ may be that shadow pattern generated by MaskShadow-GAN overlaps with that of $\mathscr{D}_{\text{adv}}$ since both of them are generated in an adversarial training way. ❷ Higher-severity shadow pattern achieves much larger relative gain for image quality enhancement. For example, large-size shadow-removed images via AEFNet acquire 63.2% visual quality improvement compared to 40.2% quality degradation of small-size shadow in $\mathscr{D}_{\text{syn}}$ (Fig. 3(e)). The latter further quality degradation stems from the over smoothing of current shadow removal methods. In addition, $\mathscr{D}_{\text{adv}}$ also achieves much larger gain compared to $\mathscr{D}_{\text{syn}}$ by SAN after shadow removal via SP+M-Net and AEFNet (Fig. 3(a-b)). ❸ The same performance gain trend, of shadow removal methods and higher-severity shadow pattern, presents in the landmark detection evaluation. In Fig. 3(e), after shadow removal via AEFNet, the large-size shadow pattern obtains the highest 22.5% NME decreasing compared to 0.6% of small-size shadow. $\mathscr{D}_{\text{adv}}$ achieves 58.5% detection improvement compared to 13.0% of $\mathscr{D}_{\text{syn}}$ by SAN (Fig. 3(a-b)).

In summary: ❶ *Current SOTA shadow removal methods can effectively improve the image quality and landmark detection simultaneously.* ❷ *Higher-severity achieves much larger performance gain after shadow removal for image quality and landmark detection.* ❸ *There is a positive correlation between shadow removal and landmark detection tasks. However, such positive correlation does not always exist in computer vision tasks,* e.g.*, between deraining and object detection (20), and between haze removal and classification (19).*

IV. Conclusion

We have proposed a shadow-removal benchmark dataset (*i.e.*, SHAREL) to explore the mutual influence of shadow removal and facial landmark detection tasks. We first proposed three strategies to construct the benchmark. Based on physical shadow model, we synthesize the shadowed faces considering four factors (*i.e.*, intensity, size, shape, and location) with three severities to cover diverse shadow patterns. We also proposed an adversarial shadow attack as hard shadow patterns to make the landmark detection fail easily. Real shadowed face dataset for landmark detection is to reduce the distribution shift with synthetic data. With SHAREL, we explored the shadow and shadow-removal effects to visual quality and landmark detection comprehensively. We observed that there is a highly positive correlation between shadow removal and the facial landmark detection task, especially, when degradation level is higher. We believe the proposed benchmark dataset and the positive correlation between shadow removal and facial landmark detection will provide insight to boost the robustness of facial landmark detection.

References

[1] Y. Wu and Q. Ji, "Facial landmark detection: A literature survey," *International Journal of Computer Vision (IJCV)*,

vol. 127, no. 2, pp. 115–142, 2019.

[2] Z. Zhang, P. Luo, C. C. Loy, and X. Tang, "Facial landmark detection by deep multi-task learning," in *ECCV*. Springer, 2014, pp. 94–108.

[3] F. Juefei-Xu and M. Savvides, "An image statistics approach towards efficient and robust refinement for landmarks on facial boundary," in *International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, 2013, pp. 1–8.

[4] X. Zhu, Z. Lei, J. Yan, D. Yi, and S. Z. Li, "High-fidelity pose and expression normalization for face recognition in the wild," in *CVPR*, 2015, pp. 787–796.

[5] Y. Liu, F. Wei, J. Shao, L. Sheng, J. Yan, and X. Wang, "Exploring disentangled feature representation beyond face identification," in *CVPR*, 2018, pp. 2080–2089.

[6] F. Liu, D. Zeng, Q. Zhao, and X. Liu, "Joint face alignment and 3d face reconstruction," in *ECCV*, 2016, pp. 545–560.

[7] T. Zhao, X. Xu, M. Xu, H. Ding, Y. Xiong, and W. Xia, "Learning self-consistency for deepfake detection," in *ICCV*, 2021, pp. 15 023–15 033.

[8] Y. Li, X. Yang, P. Sun, H. Qi, and S. Lyu, "Celeb-df: A large-scale challenging dataset for deepfake forensics," in *CVPR*, 2020, pp. 3207–3216.

[9] L. Chen, J. Pan, J. Jiang, J. Zhang, Z. Han, and L. Bao, "Multi-stage degradation homogenization for super-resolution of face images with extreme degradations," *IEEE Transactions on Image Processing (TIP)*, vol. 30, pp. 5600–5612, 2021.

[10] J. Jiang, J. Ma, C. Chen, X. Jiang, and Z. Wang, "Noise robust face image super-resolution through smooth sparse representation," *IEEE Transactions on Cybernetics*, vol. 47, no. 11, pp. 3991–4002, 2016.

[11] W. Zhang, X. Zhao, J.-M. Morvan, and L. Chen, "Improving shadow suppression for illumination robust face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 41, no. 3, pp. 611–624, 2018.

[12] C.-H. Hu, J. Yu, F. Wu, Y. Zhang, X.-Y. Jing, X.-B. Lu, and P. Liu, "Face illumination recovery for the deep learning feature under severe illumination variations," *Pattern Recognition (PR)*, vol. 111, p. 107724, 2021.

[13] H. Nada, V. A. Sindagi, H. Zhang, and V. M. Patel, "Pushing the limits of unconstrained face detection: a challenge dataset and baseline results," in *International Conference on Biometrics Theory, Applications and Systems (BTAS)*, 2018, pp. 1–10.

[14] C. Sagonas, G. Tzimiropoulos, S. Zafeiriou, and M. Pantic, "300 faces in-the-wild challenge: The first facial landmark localization challenge," in *ICCVW*, 2013, pp. 397–403.

[15] W. Wu, C. Qian, S. Yang, Q. Wang, Y. Cai, and Q. Zhou, "Look at boundary: A boundary-aware face alignment algorithm," in *CVPR*, 2018, pp. 2129–2138.

[16] X. Dong, Y. Yan, W. Ouyang, and Y. Yang, "Style aggregated network for facial landmark detection," in *CVPR*, 2018, pp. 379–388.

[17] J. Zhang, X. Zeng, M. Wang, Y. Pan, L. Liu, Y. Liu, Y. Ding, and C. Fan, "Freenet: Multi-identity face reenactment," in *CVPR*, 2020, pp. 5326–5335.

[18] L. Fu, C. Zhou, Q. Guo, F. Juefei-Xu, H. Yu, W. Feng, Y. Liu, and S. Wang, "Auto-exposure fusion for single-image shadow removal," in *CVPR*, 2021, pp. 10 571–10 580.

[19] Y. Pei, Y. Huang, Q. Zou, X. Zhang, and S. Wang, "Effects of image degradation and degradation removal to cnn-based image classification," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 43, no. 4, pp. 1239–1253, 2019.

[20] M. Hnewa and H. Radha, "Object detection under rainy conditions for autonomous vehicles: A review of state-of-the-art and emerging techniques," *IEEE Signal Processing Magazine (SPM)*, vol. 38, no. 1, pp. 53–67, 2020.

[21] Y. Shor and D. Lischinski, "The shadow meets the mask: Pyramid-based shadow removal," *Computer Graphics Forum (CGF)*, vol. 27, no. 2, pp. 577–586, 2008.

[22] X. Zhang, J. T. Barron, Y.-T. Tsai, R. Pandey, X. Zhang, R. Ng, and D. E. Jacobs, "Portrait shadow manipulation," *ACM Transactions on Graphics (TOG)*, vol. 39, no. 4, pp. 78–1, 2020.

[23] N. Inoue and T. Yamasaki, "Learning from synthetic shadows for shadow detection and removal," *IEEE Transactions on Circuits and Systems for Video Technology (TCSVT)*, vol. 31, no. 11, pp. 4187–4197, 2020.

[24] P. Hanrahan and W. Krueger, "Reflection from layered surfaces due to subsurface scattering," in *Conference on Computer Graphics and Interactive Techniques (CGIT)*, 1993, pp. 165–174.

[25] J. Guo, X. Zhu, Y. Yang, F. Yang, Z. Lei, and S. Z. Li, "Towards fast, accurate and stable 3d dense face alignment," in *ECCV*, 2020, pp. 152–168.

[26] Y. Chen and H. Sundaram, "Estimating complexity of 2d shapes," in *Workshop on Multimedia Signal Processing (MSPW)*, 2005, pp. 1–4.

[27] J. Wang, K. Sun, T. Cheng, B. Jiang, C. Deng, Y. Zhao, D. Liu, Y. Mu, M. Tan, X. Wang *et al.*, "Deep high-resolution representation learning for visual recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)*, vol. 43, no. 10, pp. 3349–3364, 2020.

[28] H. Le and D. Samaras, "Shadow removal via shadow image decomposition," in *ICCV*, 2019, pp. 8578–8587.

[29] X. Hu, Y. Jiang, C.-W. Fu, and P.-A. Heng, "Mask-shadowgan: Learning to remove shadows from unpaired data," in *ICCV*, 2019, pp. 2472–2481.

[30] A. Kumar, T. K. Marks, W. Mou, Y. Wang, M. Jones, A. Cherian, T. Koike-Akino, X. Liu, and C. Feng, "Luvli face alignment: Estimating landmarks' location, uncertainty, and visibility likelihood," in *CVPR*, 2020, pp. 8236–8246.