

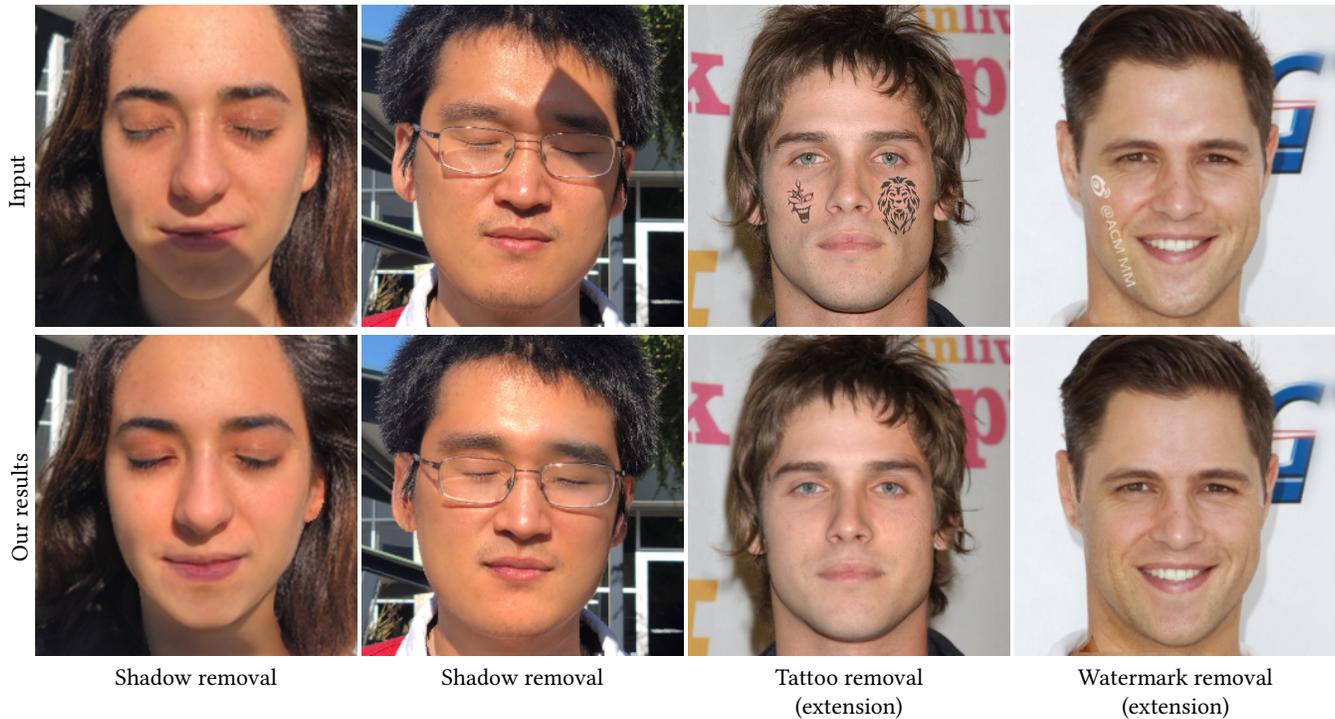
# Unsupervised Portrait Shadow Removal via Generative Priors

Yingqing He\*  
HKUST

Yazhou Xing\*  
HKUST

Tianjia Zhang  
HKUST

Qifeng Chen  
HKUST



**Figure 1: Our portrait shadow removal, tattoo removal, and watermark removal results. We propose the first unsupervised portrait shadow removal method without any training data. Our method can recover high-quality shadow-free portrait images from real-world portrait shadow images. Our proposed method can also be extended to facial tattoo removal and watermark removal as a general framework with only little modification.**

## ABSTRACT

Portrait images often suffer from undesirable shadows cast by casual objects or even the face itself. While existing methods for portrait shadow removal require training on a large-scale synthetic dataset, we propose the first unsupervised method for portrait shadow removal without any training data. Our key idea is to leverage the generative facial priors embedded in the off-the-shelf pretrained StyleGAN2. To achieve this, we formulate the shadow removal task as a layer decomposition problem: a shadowed portrait image is constructed by the blending of a shadow image and a shadow-free

image. We propose an effective progressive optimization algorithm to learn the decomposition process. Our approach can also be extended to portrait tattoo removal and watermark removal. Qualitative and quantitative experiments on a real-world portrait shadow dataset demonstrate that our approach achieves comparable performance with supervised shadow removal methods. Our source code is available at [this repository](#).

## CCS CONCEPTS

• **Computing methodologies** → **Computational photography**: **Image processing**.

## KEYWORDS

Portrait Shadow Removal, Unsupervised Learning, Generative Priors, Image Decomposition

## ACM Reference Format:

Yingqing He, Yazhou Xing, Tianjia Zhang, and Qifeng Chen. 2021. Unsupervised Portrait Shadow Removal via Generative Priors. In *Proceedings of the 29th ACM Int'l Conference on Multimedia (MM '21)*, Oct. 20–24, 2021, Virtual Event, China. ACM, New York, NY, USA, 9 pages. <https://doi.org/10.1145/3474085.3475663>

\*Both authors contributed equally.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

MM '21, October 20–24, 2021, Virtual Event, China.

© 2021 Association for Computing Machinery.

ACM ISBN 978-1-4503-8651-7/21/10...\$15.00

<https://doi.org/10.1145/3474085.3475663>

## 1 INTRODUCTION

Portrait photography is the art of capturing the inherent character of a person with a face in a photograph. Due to the rapid growth of digital cameras and smartphone cameras, portrait photography has also become popular among amateur photographers. However, portrait photographs taken in the wild often suffer from undesirable shadows cast by casual objects or even the face itself due to the lack of professional lighting control or the unpleasing environmental illumination conditions. Although photo editing software such as Adobe Photoshop provides a series of image adjustment operations for post-processing, portrait shadow removal as an effective, high-quality, and automatic application is highly desirable. Therefore, we are interested in designing an automatic and effective algorithm to remove portrait shadows without any user input.

Due to the importance of lighting manipulation to photography, there have been many approaches to remove unpleasing shadows in photographs [1, 5, 7–9, 11, 13, 15, 20, 25, 27, 30, 32, 32, 33]. Recent state-of-the-art shadow removal and portrait shadow removal methods are based on deep learning techniques and trained on large-scale image pairs in a supervised manner [5, 15, 20, 25, 30]. Although it has achieved better performance than traditional methods, those approaches are limited by the nature of their training data which may fail on complex and various real-world images. Moreover, preparing training data is a tedious and challenging task, since the diversity of training pairs for varied background environments, shadows, and subjects are very hard to cover. Meanwhile, different tasks need different kinds of training data also causes inconvenience on both model training and practical usage.

In this paper, we propose the first unsupervised portrait shadow removal method without any paired training data. Our method takes only one single shadowed portrait image as input, then recovers the shadow-free portrait image without any user intervention. The key insight of our method is that off-the-shelf pretrained face generators such as StyleGAN2 [18], own abundant priors on high-quality face appearances and geometries. The prior work [24] also exploits generative priors for versatile image restoration and manipulation tasks including colorization and inpainting. However, their method assumes that the degradation transformation is deterministic (e.g., converting generated RGB image to grayscale for image colorization and a known binary mask for image inpainting). This assumption cannot hold for restoration cases where the image degradation is complex and unknown. In our portrait shadow removal task, the shadow degradation is spatially varying and the shadowed region can be any kinds of shape. Thus, unlike their method, we make the first attempt to exploiting generative priors towards unknown degradation process in order to solve the portrait shadow removal task in an unsupervised manner.

To effectively leverage generative priors with unknown shadowed degradation, we formulate the portrait shadow removal task as an image decomposition problem [36]: a shadowed image is a composite of a full-shadow image and a shadow-free image. Since learning the image layer decomposition from one single image is an ill-posed problem, we observed that exploiting deep generative facial priors can help to reduce the ambiguity of the layer decomposition. For the shadow degradation learning, we exploit a neural network which takes a random noise as input to estimate the

shadow mask. Meanwhile, we optimize a color matrix to obtain a full shadow image from the generated shadow-free image. We note that directly optimizing our framework leads to poor performance due to the layer ambiguity nature of this problem. Thus, we propose a progressive optimization strategy and design effective regularization to clear the ambiguity step-by-step. Extensive experiments show that our unsupervised method can achieve comparable performance with state-of-the-art supervised methods. Moreover, we also extend our method to tattoo removal and watermark removal for portrait images to demonstrate the general ability of our method.

Our contributions can be summarized as follows:

- We present the first unsupervised portrait shadow removal method leveraging the abundant deep generative priors embedded in the pretrained GAN model. Our model can handle unknown degradation process in shadowed images, which cannot be achieved by a set of existing GAN-inversion-based image restoration methods.
- We propose an effective progressive optimization strategy to eliminate the ambiguous nature of the shadow degradation learning from a single input image.
- We demonstrate that our unsupervised method can achieve comparable performance with existing supervised methods. In addition, our method can also serve as a general framework to deal with multiple tasks (i.e., watermark removal and tattoo removal), which cannot be achieved via existing supervised methods.

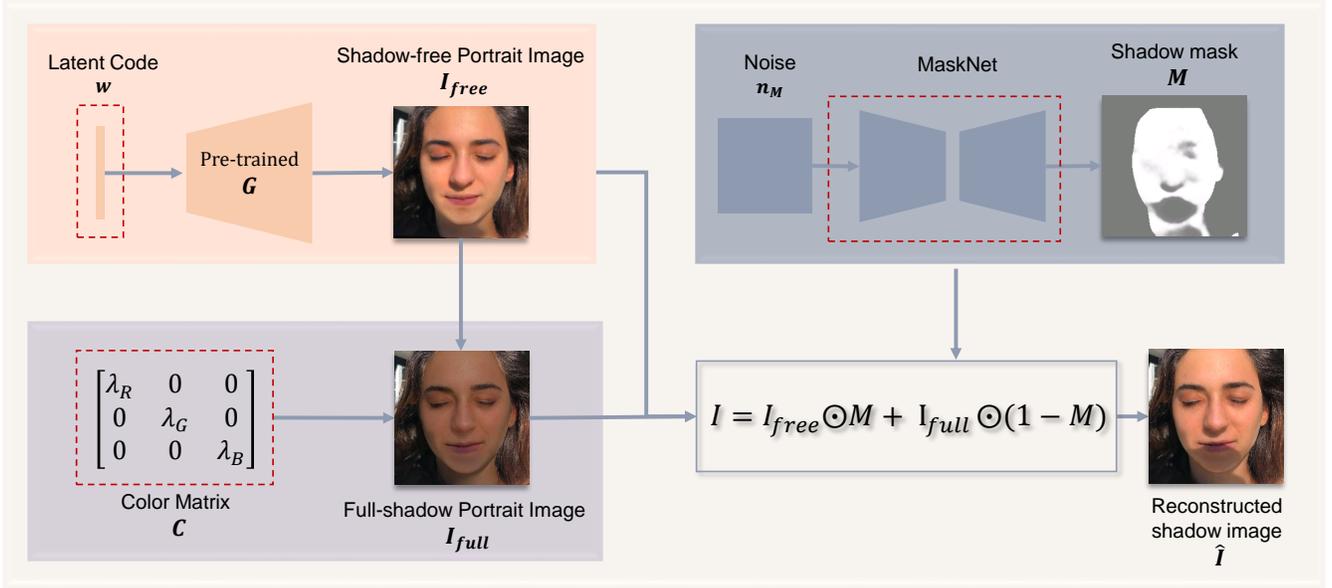
## 2 RELATED WORK

### 2.1 Shadow removal

Traditionally, graphics-based shadow removal approaches identify shadows by manually label shadow regions [1, 11, 27, 32], exploit shadow priors such as illumination discontinuity across shadow edges [2, 26], or find the relationship between shadow and non-shadow regions [13]. Then, shadow removal can be performed by histogram manipulation [19], color transfer [27, 29, 31], illumination modelling and relighting [7–9, 13, 32, 33].

With the development of deep learning in recent years, many works have been proposed to train deep neural networks on large-scale datasets for shadow removal [5, 15, 20, 25, 30]. Deshadow-Net [25] removes shadows in an end-to-end manner to predict a shadow matte. ST-CGAN [30] exploits conditional GAN [16] to generate shadow-free images and shadow masks in a unified framework. Mask-ShadowGAN [14] and ARGAN [6] also exploit generative models [4] to perform shadow removal on unpaired training data or in a semi-supervised manner. Recently [20] formulate the shadow removal task as an image layer decomposition problem similar to ours.

However, all of the previous deep learning approaches need a large amount of paired images (e.g., shadow and shadow-free image pairs) to train their networks which is time-consuming. Besides, collecting training data pairs is tedious and collected training pairs may have inconsistent colors, luminosity, and positions [14]. Most importantly, their method has dubious generalization ability due to the intrinsic limitation caused by their training data which may not be generalized well on portrait images.



**Figure 2: Framework Overview.** We design a multi-path network to effectively decompose a shadowed portrait image into three components: a shadow-free image  $I_{free}$ , a full-shadow image  $I_{full}$ , and a shadow mask  $M$ . We utilize generative priors from the pretrained StyleGAN to generate  $I_{free}$  and optimize a color matrix  $C$  to generate  $I_{full}$ . We propose a MaskNet to learn a shadow mask from sampled noise. Composing  $I_{full}$  with  $I_{free}$  with  $M$  can reconstruct the input shadowed image  $\hat{I}$  (see Eq. 1). Blocks in red dash lines are trainable. Please refer to Section 3 for our detailed design and optimization strategy.

For portrait shadow removal, [36] divide facial shadows into two types: foreign shadows which are cast by foreign objects, and facial shadows which are caused by the face itself. They designed two separate models for each kind of shadow, and two training sets are needed for these two tasks. While our unsupervised portrait shadow removal method can handle the two types of shadow in one single model and only use one single input shadowed portrait image.

## 2.2 Deep generative priors

Since performing shadow and face decomposition in a single-image manner is an ill-posed problem, we exploit the idea of GAN inversion [12, 23, 24, 37] so that we can use a pretrained state-of-the-art GAN model to provide high-quality generative facial priors to guide our clean portrait image reconstruction process.

GAN inversion aims at finding a corresponding latent vector to reconstruct the desired image using a pretrained GAN generator. Thanks to the development of state-of-the-art GAN models, many works proposed to exploit the generative priors to facilitate a set of image restoration and processing tasks. Sachit et al. [23] proposed PULSE which aimed to find a corresponding latent code in the latent space of generative models to generate high-resolution images from a single low-resolution image. Gu et al. [12] increased the amount of latent code and expand the application of GAN priors to multiple image processing and manipulation tasks. Pan et al. [24] proposed DGP to exploit the deep generative priors from the pretrained BigGAN to facilitate a set of image restoration tasks.

However, previous works exploiting GAN priors always assume the image degradation process is known, while our method can learn the spatial-varying and various shapes of shadow degradation from the input single shadowed image, which is a more challenging setting than previous methods.

## 3 METHOD

### 3.1 Overview

We formulate the portrait shadow removal task as an image decomposition problem [36]. To be specific, given an input shadowed portrait image  $I$ , we decompose it into  $I_{free}$ ,  $I_{full}$ , and  $M$ :

$$I = I_{free} \odot M + I_{full} \odot (1 - M), \quad (1)$$

where  $I_{free}$  and  $I_{full}$  denote shadow-free portrait and full-shadow portrait respectively.  $M$  is a shadow mask which denotes the shadow region and intensity.  $\odot$  is the Hadamard product. Decomposing a single shadowed portrait into  $I_{free}$ ,  $I_{full}$  and  $M$  is a highly ill-posed problem. To achieve this, we leverage the generative priors to eliminate the layer ambiguity of the decomposition. To further constrain the decomposition process, we regularize the full shadow image  $I_{full}$  using a color matrix  $C \in [0, 1]^{3 \times 3}$ ,

$$C = \begin{bmatrix} \lambda_R & 0 & 0 \\ 0 & \lambda_G & 0 \\ 0 & 0 & \lambda_B \end{bmatrix}, \quad (2)$$

$$I_{full}(x, y) = C \times I_{free}(x, y), \quad (3)$$

where  $x, y$  denote pixel positions and  $\lambda$  denotes learnable shadow parameters for R, G and B channels. We will firstly introduce our



**Figure 3: Illustration of our method. We formulate the portrait shadow removal as an image decomposition problem. Our unsupervised method takes a single shadow portrait as input and can decompose it into a shadow-free portrait image, a full-shadow portrait image, and a shadow mask.**

network architecture at Section 3.2, and then describe our proposed alternative optimization strategy at Section 3.3. Finally, we introduce how to extend our framework to portrait tattoo removal and watermark removal tasks at Section 3.4.

### 3.2 Network architecture

Following the image decomposition process defined in equation (1), we design customized modules to learn  $I_{free}$ ,  $I_{full}$  and  $M$  respectively, as illustrated in Figure 2. Specifically, we use the pretrained StyleGAN2, denoted as  $G$ , as the branch for recovering underlying shadow-free portrait  $I_{free}$ , using optimization-based GAN inversion method [18]. To reconstruct the full-shadow image  $I_{full}$ , we directly optimize a  $3 \times 3$  color matrix through backpropagation. Note that although the same strategy can be applied to optimize our shadow mask  $M$ , we find it is prone to local minima during the mask  $M$  optimization process. Thus, we adopt another small network *MaskNet*  $f_M$  and take a random initialized noise map  $n_M$  as the input to the network to reconstruct the shadow mask  $M$ . We design the *MaskNet* as an encoder-decoder network with skip connections, similar to Double-DIP [10]. We apply a sigmoid function to *MaskNet* output to further regularize the value of learned mask to be in  $(0, 1)$ . In summary, the StyleGAN2 latent code  $w$ , color matrix  $C$  and the network parameters of *MaskNet*  $f_M$  need to be optimized. We design an effective progressive optimization strategy for this problem, as shown in Sec. 3.3.

### 3.3 Progressive optimization

To produce high-quality shadow removal results, we need to carefully design the optimization process to utilize the GAN priors effectively. Joint optimization produces unpleasing artifacts, as shown in Figure 5. We instead adopt a progressive optimization strategy to guide the image recovering process step by step. We divide the optimization process into three stages, as explained in Algorithm 1.

*Stage 1. Initial face optimization.* We firstly use GAN inversion to project the shadowed image into the StyleGAN2 latent space for  $K$  steps to obtain an initial shadow-free face. To avoid the disturbance of portrait background for GAN inversion quality, we only aim at recovering high-quality human faces through segmenting face parts with a pretrained face parsing model [34]. We use LPIPS loss [35]

as the optimization goal for GAN inversion.

$$\mathcal{L}_{LPIPS} = \|\Phi(I_{free}^{init} \odot S) - \Phi(I \odot S)\|_2, \quad (4)$$

where  $S$  is the segmentation mask for face obtained by [34], and  $\Phi$  is the pretrained VGG-19 network [3, 22, 28, 35]. We also add a regularization term to the StyleGAN2 noise map  $n^G$  such that the image will not be projected into the noise latent space, as shown in [18].

$$\mathcal{L}_{reg} = \sum_{i,j} \mathcal{L}_{reg}^{i,j}. \quad (5)$$

where  $\mathcal{L}_{reg}^{i,j}$  is defined by

$$\mathcal{L}_{reg}^{i,j} = \left( \frac{1}{r_{i,j}^2} \cdot \sum_{x,y} n_{i,j}^G(x,y) \cdot n_{i,j}^G(x-1,y) \right)^2 + \left( \frac{1}{r_{i,j}^2} \cdot \sum_{x,y} n_{i,j}^G(x,y) \cdot n_{i,j}^G(x,y-1) \right)^2, \quad (6)$$

where  $n_{i,j}^G$  is the  $2^j$  times downsampled map of  $i$ -th noise map,  $r_{i,j}$  denotes the resolution of  $n_{i,j}^G$ , and  $x, y$  are pixel positions.

In total, our loss function for *Stage 1* can be formulated as

$$\mathcal{L}_{S_1} = \mathcal{L}_{LPIPS} + \alpha \mathcal{L}_{reg} \quad (7)$$

Following [18], we set  $\alpha$  to  $10^5$  in our experiments.

*Latent Initialization:* In order to ease the optimization process of this stage, we find it helpful to start with a good latent vector  $w$  which approximates the face image  $I \odot S$  well. Thus, instead of randomly initializing a latent vector from the prior distribution, we randomly sample 500 latent vectors fed into the pretrained StyleGAN2 to generate 500 images. Then we use  $\mathcal{L}_{LPIPS}$  to select the best initial value for  $w$ . We empirically set  $K$  to 300 to allow the network to produce the best shadow-free face approximation and to prevent fitting on portrait shadows.

*Stage 2. Color matrix and shadow mask optimization.* After obtaining an approximate shadow-free face in *Stage 1*, we aim at recovering a full-shadow portrait  $I_{full}$  with the color matrix  $C$ , and estimating the shadow mask  $M$  for image blending, following the definition in Equation (1). In this stage, we fix the latent space of the StyleGAN2, and jointly optimize color matrix  $C$  and parameters  $\theta$  of *MaskNet*  $f_M$  to minimize the reconstruction loss, which is defined

---

**Algorithm 1** Progressive optimization

---

**Input:** Shadowed portrait  $I$ , face parsing map  $S$ **Output:** Shadow-free portrait  $I_{free}$ , full-shadow portrait  $I_{full}$ , blending mask  $M$ 

---

*Stage 1 – Initial face optimization.*

---

- 1: Sample 500  $\{z_i\}_{i=1}^{500}$  from Gaussian distribution;
  - 2: Infer  $w$  space latents  $\{w_i\}_{i=1}^{500}$  using  $\{z_i\}_{i=1}^{500}$ ;
  - 3: Select  $w_b$  which minimizes  $\mathcal{L}_{LPIPS}(I, G(w_i))$ ;
  - 4:  $w_b^0 = w_b$ ;
  - 5: **for**  $k = 1$  to  $K$  **do**
  - 6:    $I_{free}^{init} = G(w_b^{k-1})$ ;
  - 7:   Loss =  $\mathcal{L}_{LPIPS}(I_{free}^{init}, I)$ ;
  - 8:   Update  $w_b^{k-1}$  using ADAM algorithm;
  - 9: **end for**
  - 10: **return**  $I_{free}^{init} = G(w_b^K)$ .
- 

*Stage 2 – Color matrix and shadow mask optimization.*

---

- 10: Randomly initialize the MaskNet  $f_M^0$  and the noise map  $n_M$ ;
  - 11:  $M^0 = f_M^0(n_M)$ ;
  - 12: Initialize diagonal element of color matrix to 0.5 to obtain  $C^0$ ;
  - 13: **for**  $p = 1$  to  $P$  **do**
  - 14:    $M^p = f_M^p(n_M)$ ;
  - 15:    $\hat{I} = I_{free}^{init} \odot M^{p-1} + (C^{p-1} I_{free}^{init}) \odot (1 - M^{p-1})$ ;
  - 16:   Loss =  $MSE(\hat{I}, I)$ ;
  - 17:   Update  $C^{p-1}$  and  $f_M^p$  using ADAM algorithm;
  - 18: **end for**
  - 19: **return**  $M^P$  and  $C^P$ .
- 

*Stage 3 – Facial features refinement.*

---

- 19: Initialization:  $C^0 = C^P$ ,  $w^0 = w_b^K$ ,  $I_{free}^0 = I_{free}^{init}$ ;
  - 20: **for**  $q = 1$  to  $Q$  **do**
  - 21:    $I_{free}^{q-1} = G(w^{q-1})$ ;
  - 22:    $\hat{I} = I_{free}^{q-1} \odot M^P + (C^{q-1} I_{free}^{q-1}) \odot (1 - M^P)$ ;
  - 23:   Loss =  $\mathcal{L}_{feat}(\hat{I}, I) + \mathcal{L}_{LPIPS}(\hat{I}, I)$ ;
  - 24:   Update  $C^{q-1}$  and  $w^{q-1}$  using ADAM algorithm;
  - 25: **end for**
  - 26:  $I_{free} = G(w^Q)$ ,  $C = C^Q$ ,  $M = M^P$ ,  $I_{full} = C \times I_{free}$ ;
  - 27: **return**  $I_{free}$ ,  $I_{full}$  and  $M$ .
- 

by the  $\mathcal{L}_2$  distance between reconstructed shadow portrait  $\hat{I}$  and input shadowed portrait  $I$  which served as ground truth:

$$\min_{C, \theta} \|\hat{I}_{C, \theta} - I\|_2^2. \quad (8)$$

In practice, we optimize objective function (8) for only 50 steps to guide the mask  $M$  to learn the blending relationship instead of compensating for the face details. Please note that although there exists some face detail mismatch after the optimization of *Stage 1*, the lighting appearance of shadow-free face  $I_{free}$  is adequate to learn high-quality shadow mask  $M$ , as shown in Figure 3.

*Stage 3. Facial features refinement.* After the previous two optimization stages, we can now coarsely decompose  $I$  into  $I_{free}$  and

$I_{full}$ . However, the face reconstruction results of the first stage may miss perceptually important face details, such as eyeball colors and nose size, since the optimization process is relatively short. Therefore, in this stage, we further improve the projection quality to the StyleGAN2 latent space to refine face details. Besides global perceptual loss applied on the whole face, we also propose facial feature loss to precisely optimize important face components:

$$F = \{\text{nose, eyebrow, eyeball, mouth, glasses}\}.$$

We use the same parse map extracted in *Stage 1* to identify important facial features. The LPIPS losses for face detail refinement are defined as

$$\mathcal{L}_{feat} = \sum_{f \in F} \lambda_f \Phi(f, \hat{f}), \quad (9)$$

where  $f$  is the element defined in  $F$ .

We also optimize color matrix  $C$  in this stage since the change of face detail may influence the estimation of full shadow images. We optimize this stage for 450 iterations, and the total optimization objective is

$$\min_{C, w} \mathcal{L}_{feat} + \mathcal{L}_{LPIPS}. \quad (10)$$

After optimization process, the generator of StyleGAN2 outputs a shadow-free portrait image  $I_{free}$  which is served as final output result.

### 3.4 Extensions

Our framework can also be extended to face tattoo removal and face watermark removal with only minimal modification. Note that this characteristic indicates better versatility of our framework than any existing supervised shadow manipulation methods. To demonstrate this potential, we synthesize faces with tattoos and watermarks respectively and adopt a similar optimization strategy for this layer decomposition problem. In these two tasks, we modify the problem formulation of equation (1) into:

$$I = I_{clean} \odot M + I_{pure} \odot (1 - M), \quad (11)$$

where  $I_{clean}$  and  $I_{pure}$  denote the tattoo-free or watermark-free face and the pure-color face. Here we restrict the mask to be binary for better performance.

$$\mathcal{L}_{binary} = \min(|M - 0|, |M - 1|). \quad (12)$$

## 4 EXPERIMENTS

### 4.1 Experimental setup

*Datasets.* We evaluate our method on a real-world portrait shadow removal dataset which was proposed by [36]. The portrait shadow removal dataset contains 9 subjects and 100 shadowed portrait images in varied poses, shadow shapes, illumination conditions and shadow types which provides a challenging setting for portrait shadow removal task. For tattoo and watermark removal, we synthesize our own data based on CelebA-HQ dataset [21].

*Implementation details.* We implement our method with PyTorch and conduct experiments on the NVIDIA RTX 2080Ti GPU. We use the StyleGAN2 model which is pretrained on FFHQ [17] high-quality face images with resolution  $256 \times 256$ . We set the learning

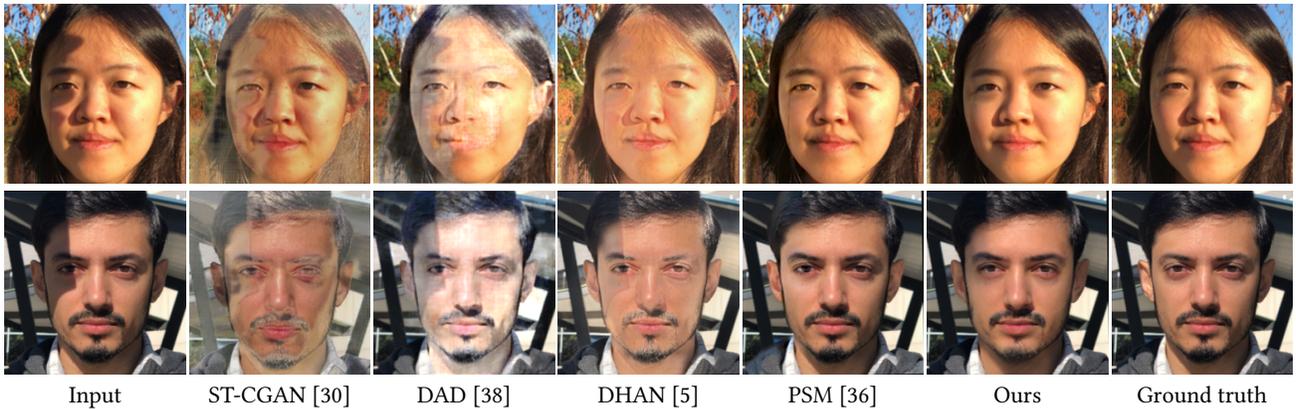


Figure 4: Comparison with baselines. Supervised shadow removal methods ST-CGAN, DAD and DHAN fail at shadow detection or shadow pixel relit, or both. Our unsupervised method can achieve comparable visual performance with state-of-the-art supervised portrait shadow removal method PSM [36].

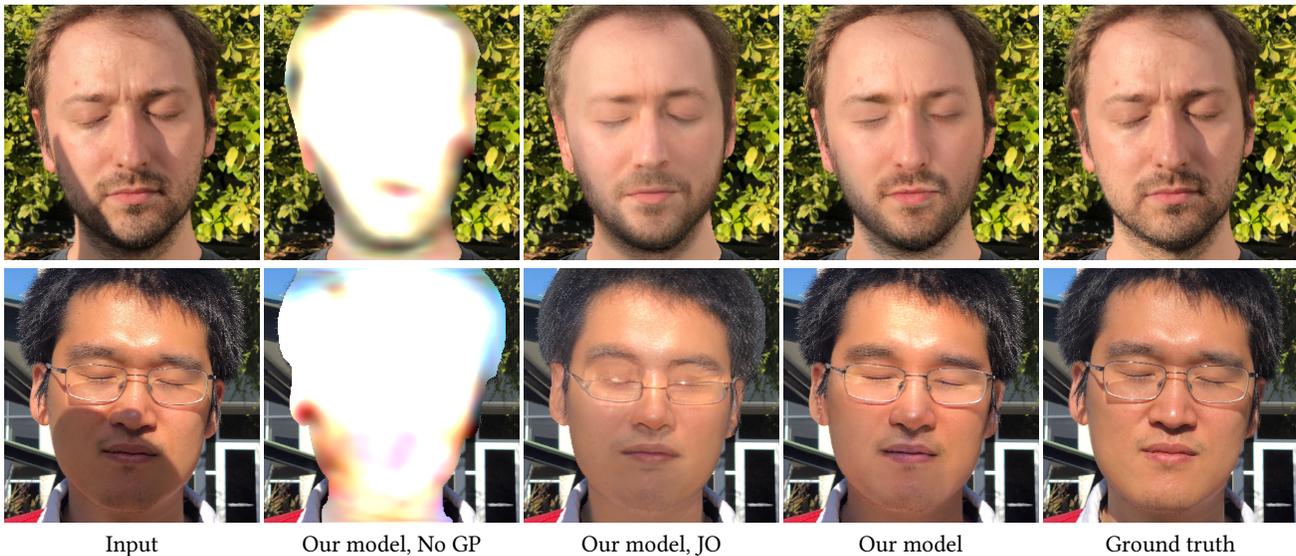


Figure 5: Controlled Experiments Results. Optimizing our framework without pretrained generative priors (No GP), fails at producing meaningful results. Jointly optimizing (JO) our framework will lose important perceptual facial details. Our progressive optimization strategy can effectively leverage the pretrained generative priors thus produce high-quality shadow-free portrait image.

rate to 0.01, 0.001 and 0.01 to project shadow images into StyleGAN2 latent space, optimize MaskNet, and optimize color matrix respectively.

## 4.2 Baselines and controlled experiments

**Baselines:** Since our method is the first unsupervised portrait shadow removal framework, we can only compare our method with a set of state-of-the-art supervised shadow removal methods ST-CGAN [30], DAD [38], and DHAN [5] and portrait shadow removal methods PSM [36] to prove our effectiveness. Our quantitative and qualitative results show that our method achieves superior performance than a set of general shadow removal methods and

Table 1: Quantitative evaluation of our model and baselines on portrait shadow removal.

	Methods	SSIM $\uparrow$	LPIPS $\downarrow$
Supervised methods	ST-CGAN [30]	0.512	0.3031
	DAD [38]	0.603	0.3225
	DHAN [5]	0.629	0.1607
	PSM [36]	<b>0.859</b>	<b>0.0874</b>
Unsupervised methods	DGP [24]	N/A	N/A
	Ours (No GP)	0.707	0.3270
	Ours (JO)	0.811	0.1288
	Ours	<b>0.820</b>	<b>0.1162</b>

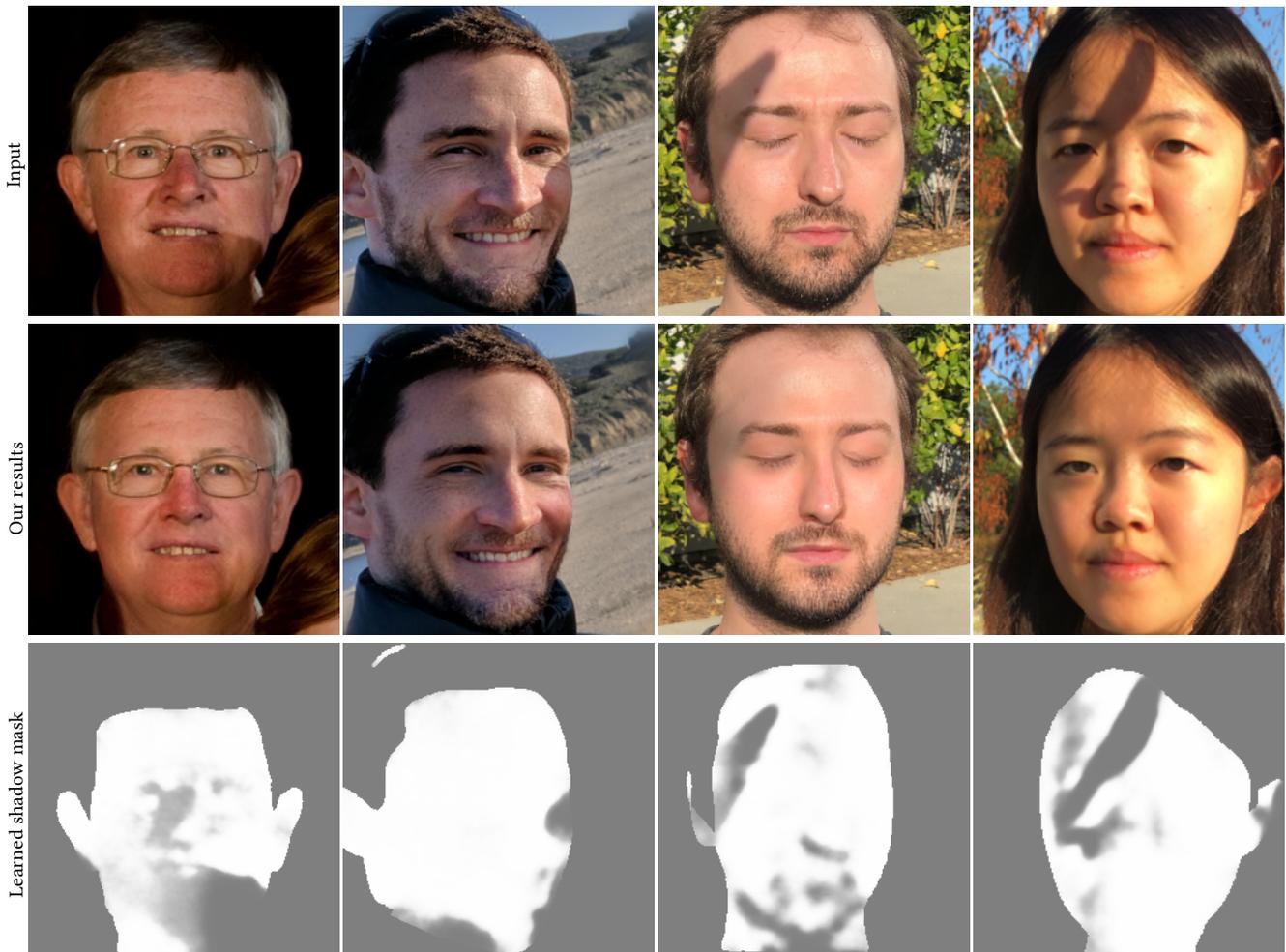


Figure 6: More Results. The bottom row shows the learned shadow mask.

comparable results with the state-of-the-art portrait shadow removal approach.

**Controlled experiments:** We also conduct several controlled experiments to evaluate the effectiveness of our proposed optimization strategies in Figure 5.

(1) **Joint optimization (JO):** Instead of optimizing different terms alternatively, we jointly optimize the shadow-free image, color matrix and shadow mask.

(2) **No generative priors (NO GP):** We conduct another experiment without pretrained StyleGAN2 weights. We keep the same network architecture but randomly initialize the StyleGAN2 weights.

### 4.3 Quantitative results

We use SSIM and LPIPS [35] to quantitatively evaluate our method, since these two metrics mostly reflect perceptual qualities. The results are shown in Table 1. Compared with a series of supervised learning-based method [5, 30, 38], our unsupervised method can achieve superior results in terms of all these evaluation metrics. This is because our method leverages rich generative facial priors which

serve as important guidance to recover the underlying shadow-free portraits. Moreover, the results reflect that the supervised baseline methods own poor generalization ability to other domains such as portraits. Our method, however, is free of generalization issues thanks to its unsupervised nature. Compared with state-of-the-art supervised portrait shadow removal method [36], our method can achieve comparable performance. Moreover, as shown in Section 3.4 and Figure 7, our framework can also be used as other portrait images decomposition tasks such as tattoo removal and watermark removal, which cannot be achieved by [36].

### 4.4 Qualitative results

We also conduct qualitative comparisons between our method and baselines, as illustrated in Figure 4. ST-CGAN [30] and DAD [38] both produce severe artifacts on real-world shadow portraits. Although the shadowed regions can be lightened up, they suffer from unnatural color distortion problems. Besides, the original well-lit face regions also contain distorted color patches. The results indicate that [30, 38] can neither accurately detect shadow regions nor relit shadowed pixels. The recovered shadow-free portrait of

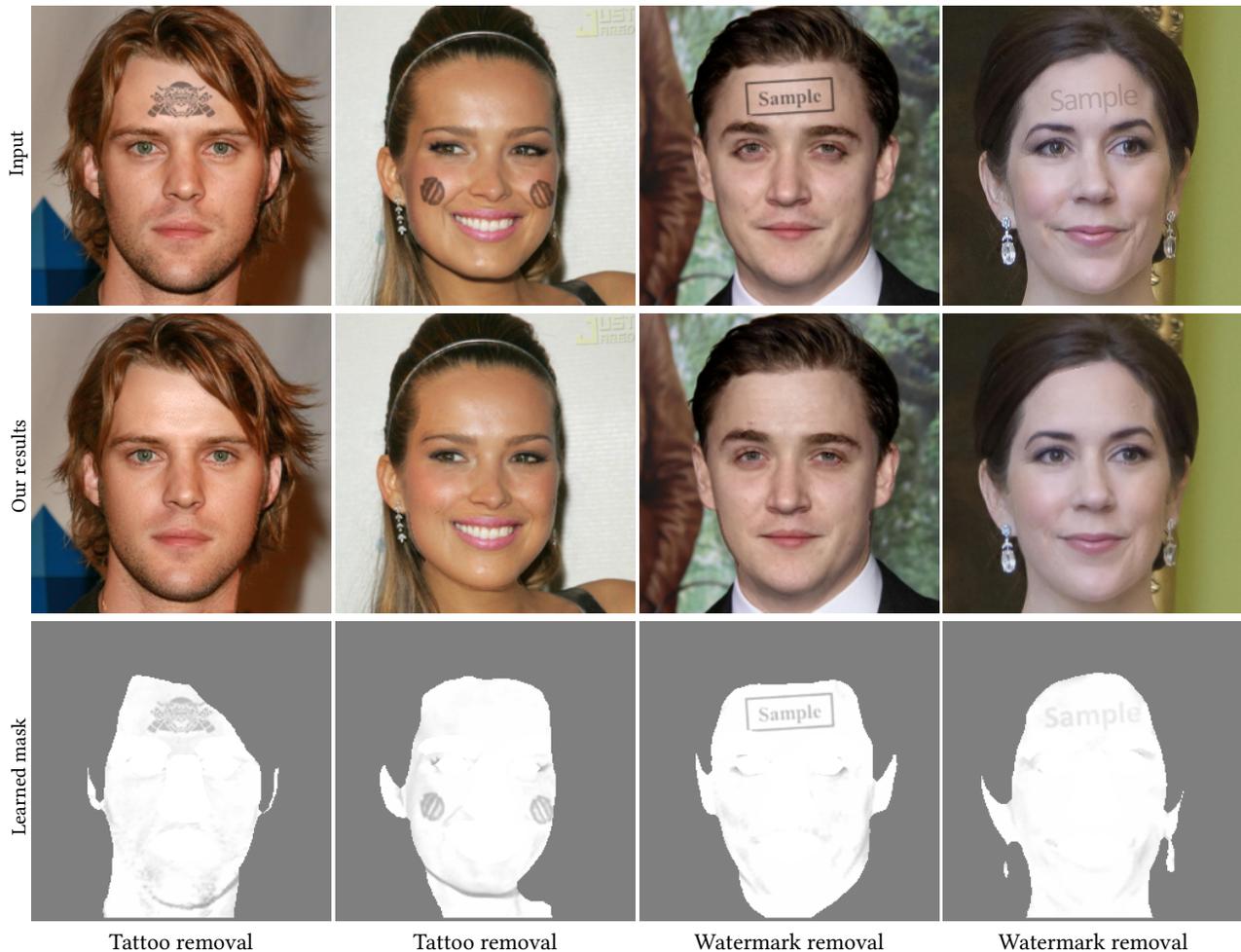


Figure 7: Our facial tattoo and watermark removal results. Our method can recover high-quality clean portrait without any training data. The bottom row shows our learned blending masks.

DHAN [5] differs a lot from ground-truth portraits, especially at no-shadow regions. The results indicate that their method performs well at shadow detection but poor at shadow removal. PSM [36] is a state-of-the-art supervised portrait shadow removal method that is trained on the same dataset as ours. Our method is free of any external training data with the aid of generative priors. Our method can achieve comparable visual results with PSM [36] but own greater application scenarios (e.g., tattoo or watermark removal) than any existing supervised shadow removal methods.

#### 4.5 Limitations

While our method can achieve comparable performance with state-of-the-art supervised methods, we do observe some unpleasant artifacts. For example, when the portrait image contains fine-grained details such as wrinkles and bushy beards (see the first case in Fig. 7), these details may be smoothed in the output image. This is due to the imperfect reconstruction results of GAN inversion, which is also an active research area. Besides, our method may not work well when the portrait accessories or clothing are not

in the training set of StyleGAN2. Thus, more powerful expression ability and GAN inversion techniques of StyleGAN2 can be further explored to improve our face restoration quality.

## 5 CONCLUSION

We proposed the first unsupervised method for portrait shadow removal which needs only one input shadow portrait image. We have shown that the generative priors can be used in this unsupervised layer decomposition setting to handle unknown degradation processes which cannot be accomplished by existing GAN-inversion methods. Meanwhile, we designed progressive optimization techniques to guide the image decomposition and reconstruction process. Then, we achieved comparable performance with existing state-of-the-art supervised-based shadow removal methods, demonstrating the effectiveness of our unsupervised method. Finally, we have shown two extension applications (e.g., portrait tattoo removal and watermark removal) of our method to demonstrate that our method can serve as a unified framework for portrait image decomposition tasks.

## ACKNOWLEDGMENTS

We thank Xuaner Zhang, Jiapeng Zhu and anonymous reviewers for helpful discussions on the paper.

## REFERENCES

- [1] Eli Arbel and Hagit Hel-Or. 2010. Shadow removal using intensity surfaces and texture anchor points. *IEEE transactions on pattern analysis and machine intelligence* 33, 6 (2010), 1202–1216.
- [2] Masashi Baba and Naoki Asada. 2003. Shadow removal from a real picture. In *ACM SIGGRAPH 2003 Sketches & Applications*. 1–1.
- [3] Qifeng Chen and Vladlen Koltun. 2017. Photographic Image Synthesis with Cascaded Refinement Networks. In *IEEE International Conference on Computer Vision*.
- [4] Xi Chen, Yan Duan, Rein Houthoofd, John Schulman, Ilya Sutskever, and Pieter Abbeel. 2016. Infogan: Interpretable representation learning by information maximizing generative adversarial nets. *arXiv preprint arXiv:1606.03657* (2016).
- [5] Xiaodong Cun, Chi-Man Pun, and Cheng Shi. 2020. Towards ghost-free shadow removal via dual hierarchical aggregation network and shadow matting GAN. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 10680–10687.
- [6] Bin Ding, Chengjiang Long, Ling Zhang, and Chunxia Xiao. 2019. Argan: Attentive recurrent generative adversarial network for shadow detection and removal. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*.
- [7] Mark S Drew, Graham D Finlayson, and Steven D Hordley. 2003. Recovery of chromaticity image free from shadows via illumination invariance. In *IEEE Workshop on Color and Photometric Methods in Computer Vision, ICCV'03*. Citeseer, 32–39.
- [8] Graham D Finlayson and Mark S Drew. 2001. 4-sensor camera calibration for image representation invariant to shading, shadows, lighting, and specularities. In *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, Vol. 2. IEEE, 473–480.
- [9] Graham D Finlayson, Steven D Hordley, Cheng Lu, and Mark S Drew. 2005. On the removal of shadows from images. *IEEE transactions on pattern analysis and machine intelligence* 28, 1 (2005), 59–68.
- [10] Yosef Gandelsman, Assaf Shocher, and Michal Irani. 2019. "Double-DIP": Unsupervised Image Decomposition via Coupled Deep-Image-Priors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
- [11] Maciej Gryka, Michael Terry, and Gabriel J Brostow. 2015. Learning to remove soft shadows. *ACM Transactions on Graphics (TOG)* 34, 5 (2015), 1–15.
- [12] Jinjin Gu, Yujun Shen, and Bolei Zhou. 2020. Image processing using multi-code gan prior. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
- [13] Ruiqi Guo, Qieyun Dai, and Derek Hoiem. 2012. Paired regions for shadow detection and removal. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35, 12 (2012), 2956–2967.
- [14] Xiaowei Hu, Yitong Jiang, Chi-Wing Fu, and Pheng-Ann Heng. 2019. Mask-ShadowGAN: Learning to remove shadows from unpaired data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*. 2472–2481.
- [15] Xiaowei Hu, Lei Zhu, Chi-Wing Fu, Jing Qin, and Pheng-Ann Heng. 2018. Direction-aware spatial context features for shadow detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
- [16] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. 2017. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*.
- [17] Tero Karras, Samuli Laine, and Timo Aila. 2019. A style-based generator architecture for generative adversarial networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
- [18] Tero Karras, Samuli Laine, Miika Aittala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. 2020. Analyzing and improving the image quality of stylegan. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
- [19] Liad Kaufman, Dani Lischinski, and Michael Werman. 2012. Content-Aware Automatic Photo Enhancement. In *Computer Graphics Forum*, Vol. 31. Wiley Online Library, 2528–2540.
- [20] Hieu Le and Dimitris Samaras. 2019. Shadow removal via shadow image decomposition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*.
- [21] Cheng-Han Lee, Ziwei Liu, Lingyun Wu, and Ping Luo. 2020. MaskGAN: Towards Diverse and Interactive Facial Image Manipulation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.
- [22] Chenyang Lei, Yazhou Xing, and Qifeng Chen. 2020. Blind video temporal consistency via deep video prior. *Advances in Neural Information Processing Systems* 33 (2020).
- [23] Sachit Menon, Alexandru Damian, Shijia Hu, Nikhil Ravi, and Cynthia Rudin. 2020. PULSE: Self-supervised photo upsampling via latent space exploration of generative models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.
- [24] Xingang Pan, Xiaohang Zhan, Bo Dai, Dahua Lin, Chen Change Loy, and Ping Luo. 2020. Exploiting deep generative prior for versatile image restoration and manipulation. In *European Conference on Computer Vision*.
- [25] Liangqiong Qu, Jiandong Tian, Shengfeng He, Yandong Tang, and Rynson WH Lau. 2017. Deshadownet: A multi-context embedding deep network for shadow removal. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
- [26] Imari Sato, Yoichi Sato, and Katsushi Ikeuchi. 2003. Illumination from shadows. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 25, 3 (2003), 290–300.
- [27] Yael Shor and Dani Lischinski. 2008. The shadow meets the mask: Pyramid-based shadow removal. In *Computer Graphics Forum*, Vol. 27. Wiley Online Library, 577–586.
- [28] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).
- [29] Tomas F Yago Vicente, Minh Hoai, and Dimitris Samaras. 2017. Leave-one-out kernel optimization for shadow detection and removal. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 40, 3 (2017), 682–695.
- [30] Jifeng Wang, Xiang Li, and Jian Yang. 2018. Stacked conditional generative adversarial networks for jointly learning shadow detection and shadow removal. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
- [31] Tai-Pang Wu and Chi-Keung Tang. 2005. A bayesian approach for shadow extraction from a single image. In *IEEE International Conference on Computer Vision*.
- [32] Tai-Pang Wu, Chi-Keung Tang, Michael S Brown, and Heung-Yeung Shum. 2007. Natural shadow matting. *ACM Transactions on Graphics (TOG)* 26, 2 (2007), 8–es.
- [33] Chunxia Xiao, Ruiyun She, Donglin Xiao, and Kwan-Liu Ma. 2013. Fast shadow removal using adaptive multi-scale illumination transfer. In *Computer Graphics Forum*, Vol. 32. Wiley Online Library, 207–218.
- [34] Changqian Yu, Jingbo Wang, Chao Peng, Changxin Gao, Gang Yu, and Nong Sang. 2018. Bisenet: Bilateral segmentation network for real-time semantic segmentation. In *Proceedings of the European conference on computer vision*.
- [35] Richard Zhang, Phillip Isola, Alexei A. Efros, Eli Shechtman, and Oliver Wang. 2018. The Unreasonable Effectiveness of Deep Features as a Perceptual Metric. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*.
- [36] Xuaner Zhang, Jonathan T Barron, Yun-Ta Tsai, Rohit Pandey, Xiuming Zhang, Ren Ng, and David E Jacobs. 2020. Portrait shadow manipulation. *ACM Transactions on Graphics (TOG)* 39, 4 (2020), 78–1.
- [37] Jiapeng Zhu, Yujun Shen, Deli Zhao, and Bolei Zhou. 2020. In-domain gan inversion for real image editing. In *European Conference on Computer Vision*.
- [38] Zhengxia Zou, Sen Lei, Tianyang Shi, Zhenwei Shi, and Jieping Ye. 2020. Deep Adversarial Decomposition: A Unified Framework for Separating Superimposed Images. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 12806–12816.