DOI: 10.1111/cgf.14944

Pacific Graphics 2023
R. Chaine, Z. Deng, and M. H. Kim
(Guest Editors)

COMPUTER GRAPHICS *forum*
Volume 42 (2023), Number 7

Facial Image Shadow Removal via Graph-based Feature Fusion

Ling Zhang[1], Ben Chen[1], Zheng Liu[2], Chunxia Xiao[3]†

[1]Hubei Key Laboratory of Intelligent Information Processing and Realtime Industrial System,
School of Computer Science and Technology, Wuhan University of Science and Technology, Wuhan, China
[2]School of Computer Science, China University of Geosciences (Wuhan), Wuhan, China
[3]School of Computer Science, Wuhan University, Wuhan, China
{zhling}@wust.edu.cn, {wust_chenben}@163.com, {liu.zheng.jojo}@gmail.com, cxxiao@whu.edu.cn

Abstract

Despite natural image shadow removal methods have made significant progress, they often perform poorly for facial image due to the unique features of the face. Moreover, most learning-based methods are designed based on pixel-level strategies, ignoring the global contextual relationship in the image. In this paper, we propose a graph-based feature fusion network (GraphFFNet) for facial image shadow removal. We apply a graph-based convolution encoder (GCEncoder) to extract global contextual relationships between regions in the coarse shadow-less image produced by an image flipper. Then, we introduce a feature modulation module to fuse the global topological relation onto the image features, enhancing the feature representation of the network. Finally, the fusion decoder integrates all the effective features to reconstruct the image features, producing a satisfactory shadow-removal result. Experimental results demonstrate the superiority of the proposed GraphFFNet over the state-of-the-art and validate the effectiveness of facial image shadow removal.

CCS Concepts

• ***Computing methodologies*** → *Shadow removal; Facial image; Feature fusion;*

1. Introduction

Facial information is one of the many attributes of personal information. When the light source is blocked by objects or other parts of the face, shadows may appear on the face. The low brightness in shadow regions may decrease the quality of the image, reducing the accuracy and effectiveness of some computer vision tasks, such as face biometric identification [ABBR20, XZXH21], facial image editing [DJBY20, JP19], face detection and recognition [GLN*21, AEHM19, WY22], and face modeling [HZLH17, ZLL*20, WLW*20]. Additionally, low-quality images disrupt the aesthetics of the image and do not satisfy the need for visual appreciation. Therefore, it is necessary to recover illumination in the shadow region for the facial image, improving the visibility of the image and enhancing the performance of the image processing tasks.

Facial image shadow removal is a difficult and challenging task. On the one hand, uneven light intensity and different directions of lighting can lead to inconsistent illumination in the shadow regions. On the other hand, the face region and background region in the image may be in different environments, making the two regions have different lighting conditions. Therefore, the proposed method should have a good perception and understanding of the face region and background region. Furthermore, since the human face has rich structural information, including some unique structures such as eyes, nose, mouth and eyebrows, the accuracy of these structures is very important. Thus, the proposed method needs to consider the structure and features of the face, maintaining the realism of the face.

Despite natural image shadow removal has made good progress [WLY18, LYW*21, FZG*21, ZGZ22, WYW*22], these methods generally perform poorly on facial images. The main reason for this situation is that natural images and face images have different properties. Natural shadow image is usually considered as a linear combination of shadow layers and shadow-free image [WYW*22], without fixed structural features. Conversely, face images have unique facial structures, such as eyes, nose, mouth, cheeks, etc. Besides, the effect of subsurface scattering of the face also needs to be considered in the shadow removal process. Without considering the particular properties of the facial image, methods to natural image often bring in color distortion or shadow remnant when they are applied to facial images, as shown in Figure 1(c).

Recently, several facial image shadow removal methods have been proposed [ZBT*20, LHH*22, HXZC21]. Traditional methods often use heuristic algorithms such as light compensation [ZZMC18, HLL*18, DH19] to deal with shadows in the face. Although

† Corresponding to Chunxia Xiao, Email: cxxiao@whu.edu.cn.

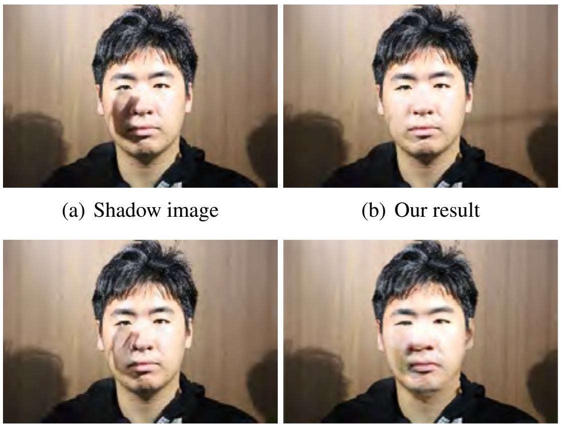

(a) Shadow image
(b) Our result
(c) Result of [WYW*22]
(d) Result of [HXZC21]

Figure 1: *Facial image shadow removal. By fully considering the features of the face image, our method can produce more desirable result.*

these methods can remove shadows in the image, they lack generalization ability and often cause brightness change in non-shadowed regions, which should not be changed for shadow removal task. Learning-based methods are often designed based on pixel-level processing strategies [HXZC21]. Although they can recover illumination in shadow regions, they ignore contextual information and the relationship between different regions, resulting in unstable shadow removal results under different lighting conditions, as shown in Figure 1(d).

To address the above issues, we propose a graph-based feature fusion network (GraphFFNet) for facial image shadow removal. Figure 2 presents the framework of the proposed GraphFFNet. We first employ an image flipper to compute a coarse shadow-less image. Then, we introduce a graph-based convolution encoder (GCEncoder) to extract global contextual relationships unconstrained by spatial position between regions in the coarse shadow-less image. Next, we introduce a feature modulation module to fuse the global topological relation onto the image features extracted by MEncoder features, enhancing the feature representation of the network. By integrating all the effective features, our fusion decoder can reconstruct the image features and produce a satisfactory shadow-removal result, as shown in Figure 2.

In summary, the main contributions of our method are as follows:

- We introduce a new graph-based feature fusion network (GraphFFNet) for facial image shadow removal, which fuses the global and local features from the image to reconstruct the shadow-removal image.
- We apply a GCEncoder to extract graph-level features unconstrained by spatial position and obtain contextual relationships between regions in the image. The coarse shadow-less image computed by the image flipper enables GCEncoder to get more useful illumination features.
- The designed feature modulation module fuses the global topological capability onto the image features, enhancing the feature representation of the network and contributing to high-quality shadow removal result.

2. Related Work

2.1. Image Shadow Removal

Image shadow removal methods can be mainly categorized into two classes: traditional methods [OL09, GPP06, MXZP12, JHK19, BDS*17, YY00, BT06] that rely on prior knowledge and learning-based methods [LCC20, CLZX21, CPS20, WLY18, FZG*21, ZLZX20] that learn the mapping relationship between shadow and shadow-free images in a training dataset.

Early traditional methods addressed this problem by utilizing the underlying physical factors of shadow formation [FHD02,LG08,SL08,VHS17,WTBS07]. Finlayson et al. [FHD02,FDL09] proposed a series of shadow removal methods based on gradient consistency. These methods perform well for simple shadow removal in the domain, but may not yield satisfactory results for darker hard shadows or complex shadows with boundaries, such as tree shadows or inconsistent soft shadows.Another strategy for traditional methods is information transfer, which transfers illumination from non-shadow regions to shadow regions [WHCO08, ZZML13, SL08, XXZC13, XSXM13, ZZX15, ZZX15]. It has found extensive applications in image processing tasks. Wen et al. Although these methods can achieve satisfactory shadow-removal results, their effectiveness depends on the accuracy of texture matching.

Recently, large-scale datasets [QTH*17, WLY18] have been released, enabling the training of deep neural networks for shadow removal [CLZX21, CPS20, WLY18, FZG*21, HJFH19, WTBS07, DLZX19, LYW*21, QTH*17, Sid19, LS19, LYW*21, JST21, WLY18]. Qu et al. [QTH*17] proposed an end-to-end neural network model for automatic shadow removal. The network first utilizes a global network to extract rough global shadow information and then extends two parallel sub-networks from this global network. One sub-network is used to extract color features and other information from the input image, while the other sub-network extracts semantic information. By combining these feature information, the network learns to generate the final shadow-free image. Wang et al. [WLY18] analyzed the relationship between shadow detection and removal and proposed ST-CGAN, a stacked conditional generative adversarial network framework for joint shadow detection and removal. Hu et al. [HFZ*19] addressed the problem of color inconsistency between shadow and non-shadow regions by training on shadow images and their corresponding ground truth shadow-free images and constructing a color compensation mechanism to achieve shadow removal. Hieu et al. [LS19] treated shadow images as a combination of shadow-free images, shadow parameters, and shadow masks, and used a neural network to predict and remove shadows. Hu et al. [HJFH19], to overcome the requirement for paired datasets, employed the concept of CycleGAN and trained their model on unpaired data. By using shadow masks to guide shadow image generation, they solved the problem of multiple shadow images corresponding to a single shadow-free image in CycleGAN

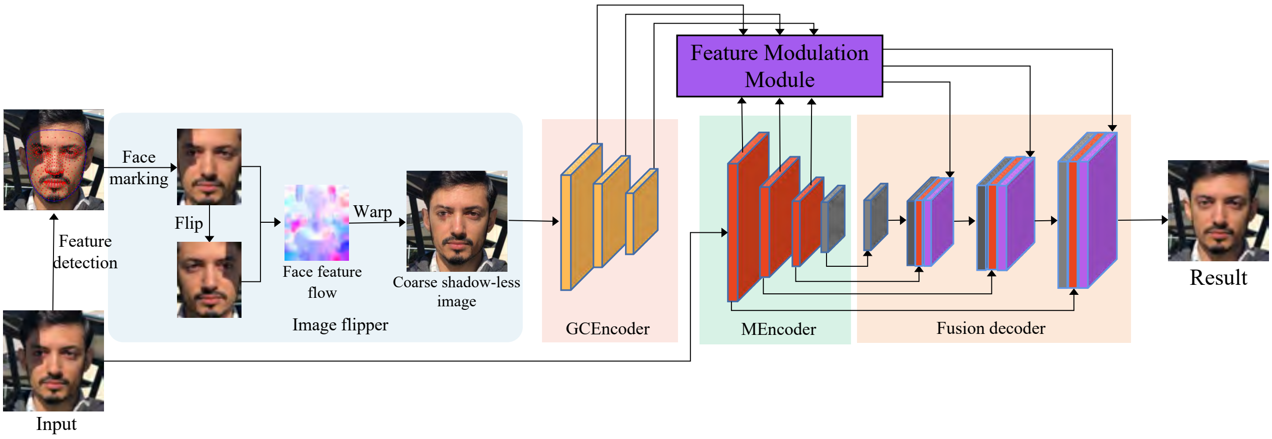


Figure 2: *The framework of the proposed GraphFFNet. We first use an image flipper to compute a coarse shadow-less image, which is fed into the GCEncoder to obtain the global topological relationship between regions in the image. Then, the feature modulation module integrates the global features into the local features from the MEncoder. Finally, the fusion decoder fuses and reconstructs the features, producing a high-quality shadow-removal result for the facial image.*

2.2. Face Relighting

Face relighting mainly focuses on two aspects. One is traditional image processing techniques, such as histogram equalization and color balancing, which adjust the brightness, contrast and other parameters of the image to change the facial illumination effect [CEW06, LYD10]. For example, Faraji et al. [FQ16] combined adaptive homomorphic filtering to generate illumination-invariant features for images with different lighting levels. Zhang et al. [ZTF*09] proposed a method to extract poorly illuminated region features based on the latent structure of facial images. These methods have limited generalizability for color images and do not perform well in complex lighting environments for face relighting.

Second, deep learning techniques are employed to achieve more refined face relighting by learning from a large amount of facial image data [SKCJ18]. For example, Xu et al. [XSHR18] proposed a neural network to edit the illumination of images and reproduce complex lighting effects, but it required capturing five images as input under predefined directional lighting. Calian et al. [CLG*18] employed an inverse rendering approach to decompose outdoor images and relight the facial images, resulting in lighting inconsistencies. To address these issues, we utilize the well-known illumination information on the facial surface in the image to edit the illumination of other facial regions, thereby better preserving the original good lighting information on the face.

3. Methodology

We propose a graph-based feature fusion network (GraphFFNet) for facial image shadow removal. The overall architecture of GraphFFNet is shown in Figure 2. We first use a multi-scale encoder (MEncoder) to extract features from the shadow image. Meanwhile, we apply an image flipper to warp the shadow image based on facial symmetry and obtain a coarse shadow-less image. Then, by using the coarse shadow-less image as input, we employ a graph-based convolution encoder (GCEncoder) to extract the global relationship between regions in the image. Next, we introduce a feature modulation module to fuse the global features from GraphFFNet and the local features from MEncoder. Finally, by integrating the modulated features and the local features, we employ a fusion decoder (FDecoder) to reconstruct the features and produce a high-quality shadow-removal result for the shadow image.

3.1. Encoder Structures

Convolution operations in the network focus mainly on the local features in images, resulting in a limited ability to model global information. For the facial shadow removal task, the convolution operations may lead to insufficient sensitivity to the overall illumination in the face. Motivated by the ability of feature aggregation and transformation for graph convolutional network (GCN), besides a multi-scale encoder (MEncoder) to extract the local features from the image, we also introduce a graph-based convolution encoder (GCEncoder) to extract global features, which can model the spatial relationships between different regions.

We know that the human face has symmetry. But shadows bring in different illumination appearance on the left and right sides of the face. Non-shadow regions have better illumination appearance than in shadow regions. To obtain better facial illumination features, we employ an image flipper to compute a coarse shadow-less image for the shadow image, utilizing the symmetry of the face to transfer the illumination information from the non-shadow regions to the shadow regions.

Image flipper. Image flipper aims to get a coarse shadow-less image. Based on the symmetry of the face, we introduce a face feature flow to measure the spatial variation of illumination between the left and right sides of the face. Then, we use the face feature flow to guide the illumination transfer on the face. Specifically, we first employ the face tagging system constructed by Kartynnik et al. [KAGG19] to detect and calibrate faces, generating a facial geometric model with 468 2D vertices, as shown in Figure 3(b, c). The

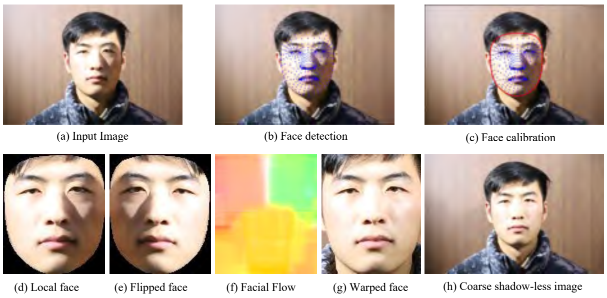


Figure 3: *Image flipper. We first detect and calibrate the face regions (b, c) for the image. Then, we flip the local face (d) to get a flipped face (e). (f) is the optical flow field between (d) and (e). (g) is the illumination transfer result using (f). (h) is the obtained coarse shadow-less image.*

468 feature points are projected onto the facial geometric model, obtaining a local face image F_1, as shown in Figure 3(d). Then, we swap the left and right sides of the face in F_1 utilizing the calibrated feature points and face symmetry, getting a flipped image F_2, as shown in Figure 3(e). Next, we apply Farneback algorithm [Far02] to calculate the face feature flow, which describes the displacement vector of the pixel point moving between F_1 and F_2, as shown in Figure 3(f).

With the computed face feature flow, we can transfer the illumination from the high region in F_2 to the low region in F_1. Concretely, for the pixel point (x,y) in F_1, assuming that its corresponding optical flow vector is (u,v), we copy the illumination of point $(x+u,y+v)$ to point (x,y) to complete the illumination transfer. Thus, we can get a warped face image F_3, as shown in Figure 3(g). To ensure that the illumination is transferred from the high region to the low region, we set the optical flow vector from high illumination to low illumination to $(0,0)$ and only perform the transfer operation on the low illumination region. Finally, F_3 is filled into the original image to obtain the coarse shadow-less image I_{coarse}, as shown in Figure 3(h). The coarse image produced by image flipper is an illumination correction image with fewer shadows (see Figure 4(b)), providing more useful illumination information for the network and contributing to the shadow removal task (see Figure 4(c)).

Multi-scale encoder. Multi-scale encoder (MEncoder) is used to obtain local features from the shadow image. We apply three multi-scale convolution blocks to implement our MEncoder. As shown in Figure 5, each convolution block has two convolution layers followed by a ReLU activation function. After the final convolution layer, a max pooling layer is used to reduce the size of the image, ensuring the size of the CNN features. Then we use a multi-scale pooling block to enlarge the receptive field. The multi-scale pooling block first uses four pooling windows to capture feature information at four scales. With a convolution followed by a ReLU activation function and upsampling operation at each scale, we can resize the extracted features, which are concatenated with the original feature map along the channel dimension. The concatenated features contain more contextual information.

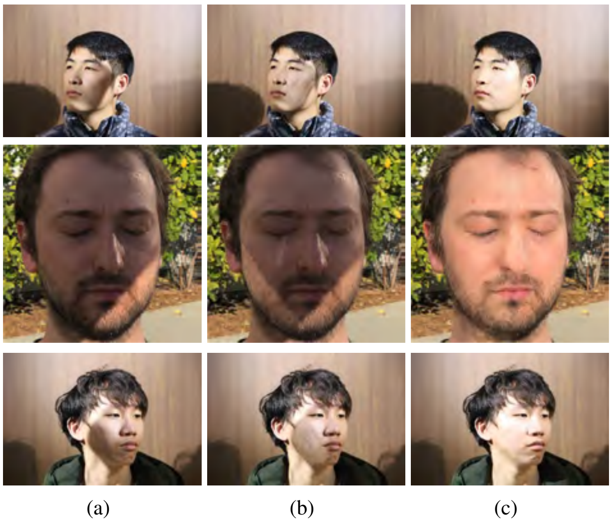


Figure 4: *Coarse shadow-less images and the corresponding shadow-removal images. (a) is the input images. (b) is the coarse shadow-less images produced by image flipper. (c) is the final shadow-removal results.*

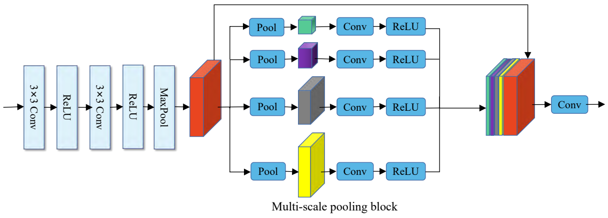


Figure 5: *The structure of multi-scale convolution block in MEncoder.*

Graph-based convolution encoder. The face image is not a regular shaped image, which is generally divided into face region and background region. Usually, the face and the background are in different spatial environments, resulting in inconsistent lighting conditions in the two regions. Moreover, the color and structure of the background region are complex and variable in different images, that are very different from the structure of the human face. Direct feature extraction from the image may introduce too much background information into the model, reducing the feature representation of the face for the network.

Inspired by the global perception capability of graph convolution, we introduce a graph-based convolution encoder (GCEncoder) to extract the features of the image and learn the potential relationships based on the related contents in the image. The image can form a graph structure through content and semantic representations, contributing to a better understanding of face information. Our GCEncoder transforms image into graph structure and uses the node in the graph as a unit for feature extraction and processing. Such treatment can effectively transfer and aggregate feature information from different regions and fuse global contextual information, helping to understand and model face images more com-

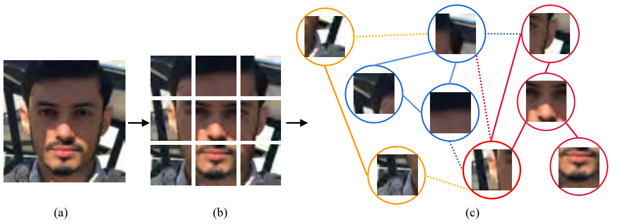


Figure 6: *Graph for the image. We first divide the image (a) into several patches with size of* 4×4 *(b). For each patch, we use KNN algorithm to find the adjacency patches. With the adjacency relationships, we construct a graph for the image (c).*

prehensively. Our GCEncoder uses the coarse shadow-less image I_{coarse} as input to obtain better illumination expression.

Specifically, we first divide I_{coarse} into N patches, as shown in Figure 6(b). Each patch is transformed into a 2-dimensional feature vector. The feature vector can act as a node, and the input image can be represented as a composition of N unordered nodes, that is $V = \{v_1, v_2, \cdots, v_N\}$. For each node v_i, we find K neighbor nodes $N(v_i)$ and add a edge between v_i and v_j for each $v_j \in N(v_i)$. Then, we obtain a graph $G = (V, E)$ for the image, as shown in Figure 6(c), where E denotes the edges between two nodes and each edge has a weight. Our GCEncoder treats the image using graph convolution for the graph G. We perform weighted average for the feature vectors of each node and its neighboring nodes, resulting in a new feature vector for that node. Thus, our GCEncoder can use the global relationship and structure information among nodes to extract feature information from the image.

GCEncoder is a graph convolution-based image pyramid structure. We utilize graph convolution to obtain contextual information at different scales, enabling a better understanding of the semantics and context of the image. Specifically, we first employ K-nearest neighbors (KNN) algorithm [CH67] to compute the adjacency matrix for each node, modeling the feature maps as graph features. For each scale, we introduce a graph-level processing block, as shown in Figure 7, which consists of graph processing module and feed-forward module. The features of neighboring nodes are aggregated and updated to generate a new feature representation for the node, capturing the graph structure and contextual information more effectively. Feed-forward module is a simple multi-layer perception used to enhance feature transformation and alleviate the over-smoothing of graph features.

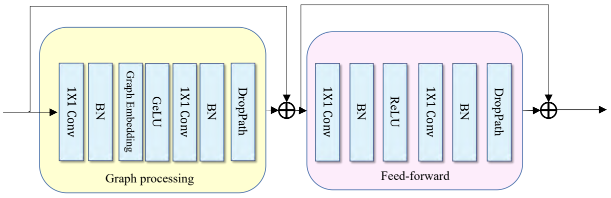


Figure 7: *The structure of graph-level processing block.*

Graph processing module uses the built graph as input. It first applies Conv-BN for each node in the graph, which aims to reduce the channel number of the input features to lower the computation. Then, we apply KNN algorithm to perform graph embedding, which finds K nearest neighbor nodes to update our graph based on feature similarity. Next, we apply GeLU-Conv-BN-DropPath to process graph features and feed them into Feed-forward, improving the performance and generalization of the model. Feed-forward module applies Conv-BN-ReLU-Conv-BN to encode the features. DropPath is used to prevent overfitting. The encoded features are combined with features from the graph processing module, yielding the output of the graph-level processing block.

3.2. Decoder Structures

We have already extracted the features of the image using the encoders. However, due to the inherent localization and equivariance of convolution, MEncoder lacks the ability of contextual understanding, focusing on the local features of the image and ignoring the facial position information. On the contrary, our GCEncoder can get global contextual information by using the graph structure. To better use the extracted features, we introduce a feature modulation module to fuse the global topological features from GCEncoder into the local information from MEncoder, which can obtain a better representation of the features. Finally, we apply a fusion decoder to integrate the different features, reconstructing the features and producing a high-quality shadow-removal result.

Modulation.pdf

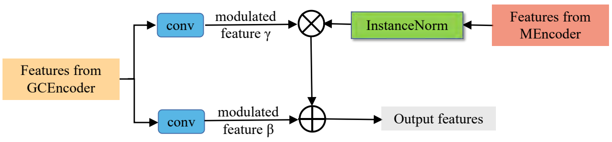


Figure 8: *The structure of feature modulation module.*

Feature modulation module. Inspired by style transfer and conditional image enhancement [JZH*20, WYDL18], we use global features from GCEncoder as conditional information to predict modulation matrices, which are used to modulate local features from MEncoder. With this treatment, we can transfer the global topological capability to local features without compromising the local features.

Figure 8 illustrates the architecture of the feature modulation module, which utilizes features from GCEncoder as conditional features to modulate features from MEncoder at the three scales. We first use two separate convolutions to extract modulation estimation features β and γ. To improve the efficiency of feature modulation, we apply instance normalization to the features from MEncoder, which makes the feature distribution closer to a standard normal distribution, reducing feature bias and variation. This helps enhance the stability and reliability of the features. To extract more discriminative features and filter out noise or redundant information, we employ element-wise multiplication to fuse modulation estimation features γ with the instantiated features, highlighting important features and improving feature discriminability. Subsequently, we combine the fused features with modulation estimation feature β through element-wise addition to obtain the modulation

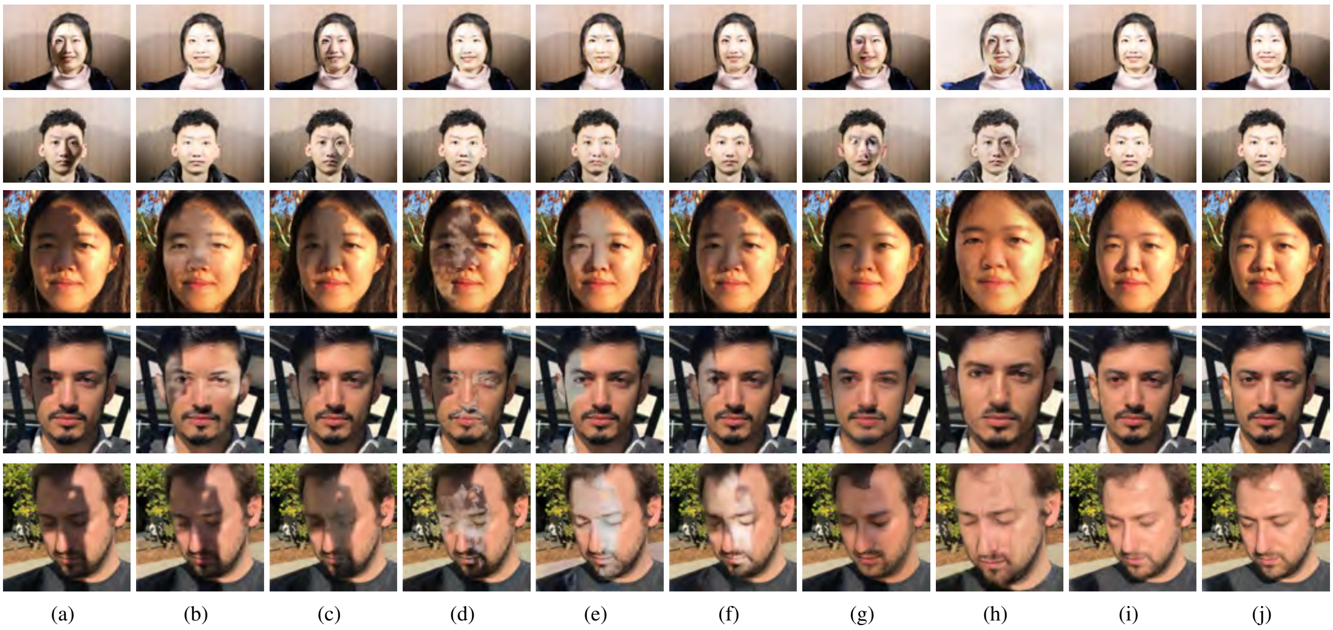


Figure 9: *Visual comparison among state-of-the-art shadow removal results: (a) input images,(b) ST-CGAN [WLY18], (c) G2R-ShadowNet [LYW*21], (d) Auto-Exposure Fusion [FZG*21], (e) SpA-Former [ZGZ22], (f) Style-Guided [WYW*22], (g) He et al. [HXZC21], (h) Blind Removal [LHH*22], (i) our FFShadowNet, and (j) ground-truth images.*

features. Such feature modulation can capture contextual semantic information and enrich the representation of features.

Fusion decoder. To take full advantage of the features, we replace the common decoder with a fusion decoder to reconstruct shadow-free image. We first perform upsampling on the features from MEncoder. Then, we concatenate the upsampled features with modulation features and features from MEncoder along the channel dimension at the corresponding scale. The upsampling and concatenation processes are repeated three times. Then, we use a convolutional layer to reconstruct a high-quality shadow-free image.

3.3. Loss Function

To get a robust parametric mode, we use a loss function to optimize the proposed GraphFFNet for facial image shadow removal. Our loss function L contains two components: visual loss L_{visual} and perceptual loss $L_{percept}$, that is,

$$L = L_{visual} + \alpha L_{percept}, \tag{1}$$

where α is weight parameter. In our experiments, we set $\alpha = 0.5$.

Visual loss is used to evaluate the appearance consistency loss between the predicted shadow-removal result I_{result} and the ground-truth image I_{gt}, which is calculated in the L_1 distance:

$$L_{Visual} = ||I_{result} - I_{gt}||_1. \tag{2}$$

Perceptual loss evaluates the image structure loss between I_{result} and I_{gt}, which is computed using the multi-level features of the VGG19 network,

$$L_{per} = \sum_i ||VGG_i(I) - VGG_i(I_{gt})||_2^2, \tag{3}$$

where $VGG_i()$ represents the i-th layer features of the VGG19 model, and $i \in \{2, 7, 12, 21, 30\}$ [GEB15].

4. Experiments

4.1. Experimental Settings

Implementation Details. Our method is implemented using PyTorch framework. We utilize an NVIDIA GeForce RTX3090 to train our GraphFFNet for 200 epochs. The Adam optimizer with default parameters is used to optimize our model. We set the initial learning rate to 0.0001 and employ the cosine annealing strategy to adjust the learning rate until convergence.

Dataset. We use two datasets as our training dataset to get a better training data. One is FSD dataset constructed by Chunxia Xiao's laboratory, which consists of 2800 pairs of facial shadow and shadow-free images. The other is a new dataset constructed by ourselves, which contains 1612 pairs of facial shadow and shadow-free images. We denote the two datasets as FSD+. We use two test datasets to evaluate our GraphFFNet. One is a new facial shadow test dataset (FSTD) collected by ourselves, which contains 964 facial image pairs. Another test dataset is the dataset proposed by Zhang [ZBT*20] containing 100 image pairs.

Metrics. We evaluate the performance of our GraphFFNet for facial shadow removal using the root mean square error (RMSE) between the shadow-removal result and the corresponding ground

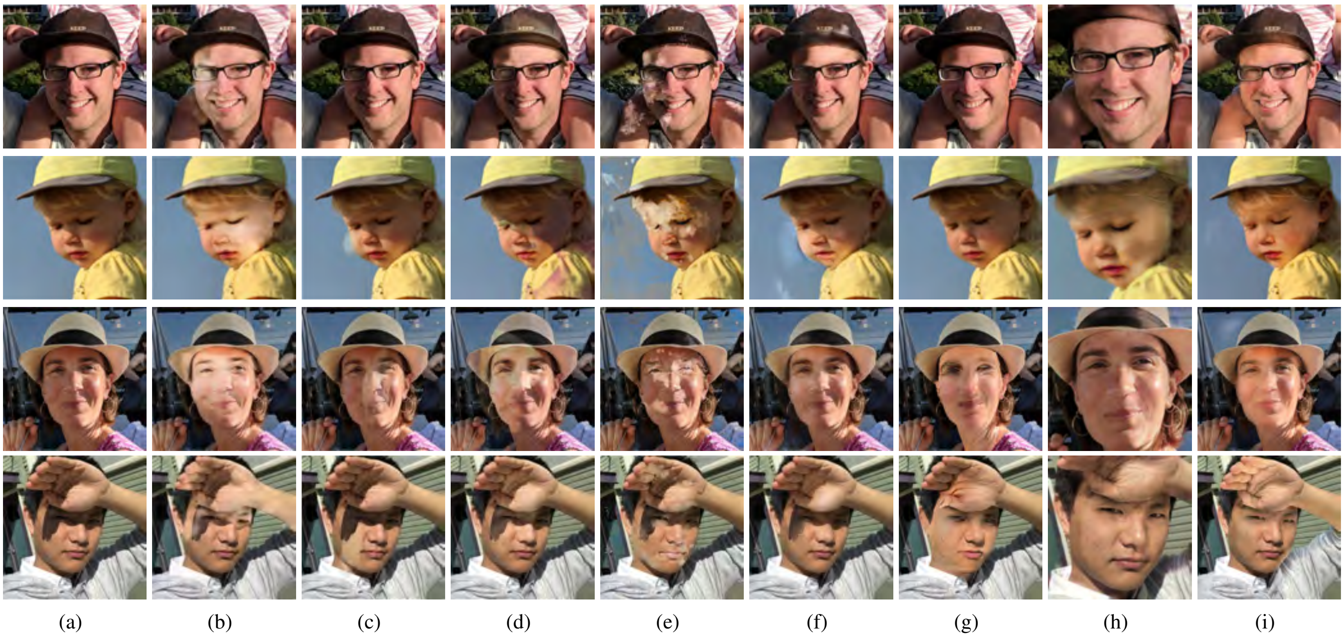


Figure 10: *Visual comparison among state-of-the-art shadow removal results: (a) input images,(b) ST-CGAN [WLY18], (c) G2R-ShadowNet [LYW*21], (d) Auto-Exposure Fusion [FZG*21], (e) SpA-Former [ZGZ22], (f) Style-Guided [WYW*22], (g) He et al. [HXZC21], (h) Blind Removal [LHH*22], and (i) our FFShadowNet.*

truth shadow-free image in the LAB color space. Additionally, we report the peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) in the RGB color space to further assess the performance of the proposed method.

4.2. Comparison with State-of-the-arts

To validate the effectiveness of the proposed GraphFFNet, we compare our results with various state-of-the-art shadow removal methods, including five natural image shadow removal methods [WLY18, LYW*21, FZG*21, ZGZ22, WYW*22] and two facial shadow removal methods [HXZC21, LHH*22]. To make a fair comparison, we train all the learning-based methods on FSD+ dataset using the same hardware. Table 1 summarizes the comparison results using three metrics. From the table, we can observe that, our method achieves the best values in all the metrics among all the comparing methods, indicating the effectiveness of our GraphFFNet.

To further demonstrate the superiority of our method, we provide some visual shadow-removal results for facial images in Figure 9. It can be seen, St-CGAN [WLY18] can remove shadows from images but may cause detail blurring exhibits in the face, as shown in Figure 9(b). SG-ShadowNet [LYW*21] has a weak ability to remove dark shadows, as shown in Figure 9(c). Fu et al. [FZG*21] suffer from desaturation issues during the restoration of skin tones, as shown in Figure 9(d). SpA-Former [ZGZ22] shows insensitivity to shadow regions and fails to remove complex facial shadows effectively, as shown in Figure 9(e). SG-ShadowNet [WYW*22] can remove shadows in the face but introduce artifacts along the shadow boundaries, as shown in Figure 9(f). He et al. [HXZC21] rely on the facial feature prior for shadow removal, resulting in suboptimal performance due to insensitivity to environmental illumination, as shown in Figure 9(g). Liu et al. [LHH*22] directly decompose the shadow removal task into grayscale image shadow removal and image coloring, leading to unstable performance for complex shadows in the face, as shown in Figure 9(h). Comparatively, our GraphFFNet effectively removes shadows in the facial image without fewer artifacts, as shown in Figure 9(i), which is similar to the ground-truth image.

To further validate the robustness and generalization capability of our GraphFFNet, Figure 10 provides some shadow removal results for facial images collected in real-world life, including challenging cases such as heavy shadows, inconsistent illumination and self-shadows on faces. From the results, we find that results produced by our method look more natural with little artifacts in the face.

User study. We conduct a user study to evaluate the visual performance of the proposed GraphFFNet in comparison to several advanced shadow removal methods. We prepare a set of 100 shadow images. Each set contains shadow removal results produced by ST-CGAN [WLY18], G2R-ShadowNet [LYW*21], Fu et al. [FZG*21], SpA-Former [ZGZ22], SG-ShadowNet [WYW*22], He et al. [HXZC21], and Liu et al. [LHH*22], respectively. We randomly select 206 volunteers and provided them with 20 sets of images, asking them to choose the best shadow removal image for each set. From the statistical analysis, we found that 18.93% of the shadow-free images generated by our GraphFFNet are selected as the best. The shadow removal results of ST-CGAN [WLY18], G2R-ShadowNet [WYW*22], Fu et al. [FZG*21], SpA-Former [ZGZ22], SG-ShadowNet [WYW*22],

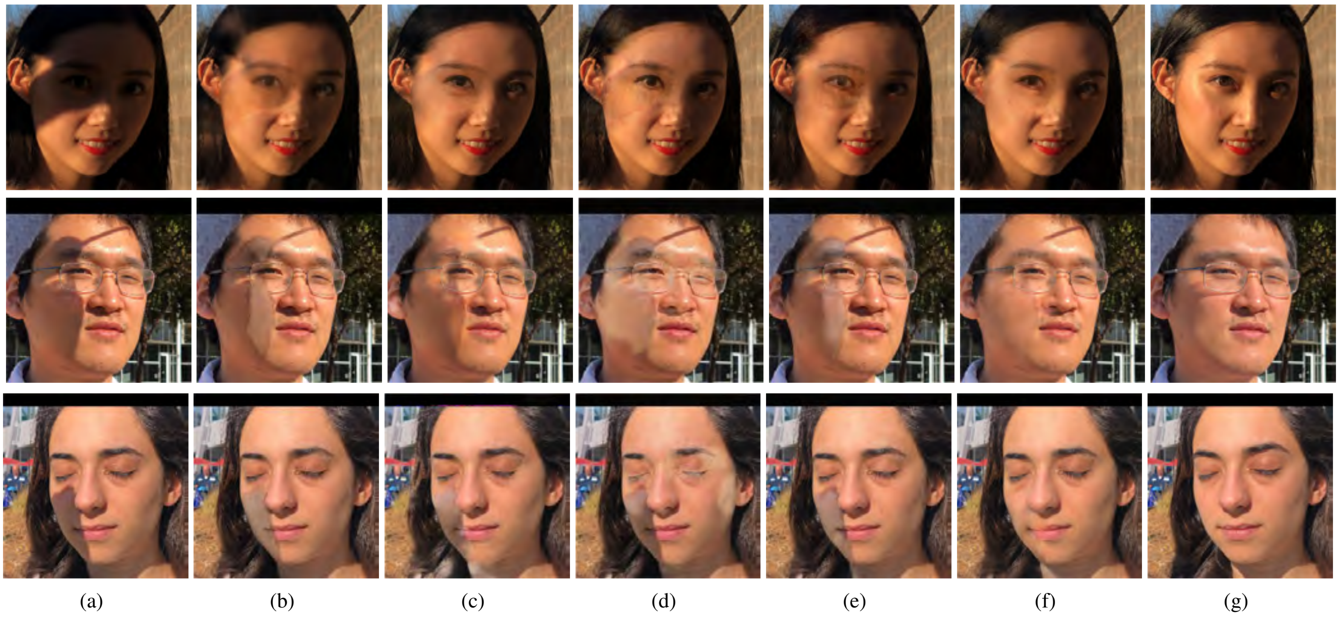


Figure 11: *Visual comparison for ablation study: (a) input images, (b) Base, (c) GraphFFNet$_1$, (d) GraphFFNet$_2$, (e) GraphFFNet$_3$, (f) our GraphFFNet, and (g) ground-truth images.*

Table 1: *Quantitative comparisons of shadow removal on FSTD and Zhang [ZBT*20] datasets in terms of RMSE, PSNR, and SSIM. All the learning-based methods are trained on FSD+ dataset. ↑ means the larger the better while ↓ means the smaller the better.*

Methods	Venue/Year	FSTD			Zhang		
		PSNRąü	SSIMąü	RMSEąý	PSNRąü	SSIMąü	RMSEąý
ST-CGAN [WLY18]	CVPR/2018	30.275	0.954	9.525	19.514	0.825	31.650
G2R-ShadowNet [LYW*21]	CVPR/2021	28.666	0.954	12.401	20.379	0.846	29.817
Fu et al. [FZG*21]	CVPR/2021	31.308	0.963	8.521	19.187	0.798	32.359
SpA-Former [ZGZ22]	CVPR/2022	29.420	0.956	11.009	26.505	0.897	13.710
SG-ShadowNet [WYW*22]	ECCV/2022	32.704	0.975	7.765	23.162	0.863	20.947
He et al. [HXZC21]	CVPR/2021	23.992	0.926	18.903	20.480	0.808	28.059
Liu et al. [LHH*22]	BMVC/2023	21.287	0.838	26.351	19.317	0.725	29.726
Our GraphFFNet	PG/2023	**36.423**	**0.982**	**5.389**	**29.775**	**0.931**	**9.901**

He et al. [HXZC21], and Liu et al. [LHH*22] accounted for 10.68%, 10.19%, 11.17%, 11.65%, 12.62%, 10.68% and 14.08% respectively.

4.3. Ablation Study

To evaluate the performance of different components employed in our GraphFFNet, we conduct ablation experiments by disabling or modifying one specific component. We design four variants:

(1) Base: only MEncoder and a common decoder;

(2) GraphFFNet$_1$: GraphFFNet without image flipper, and GCEncoder uses the shadow image as input;

(3) GraphFFNet$_2$: replace GCEncoder with common encoder without graph-based convolution;

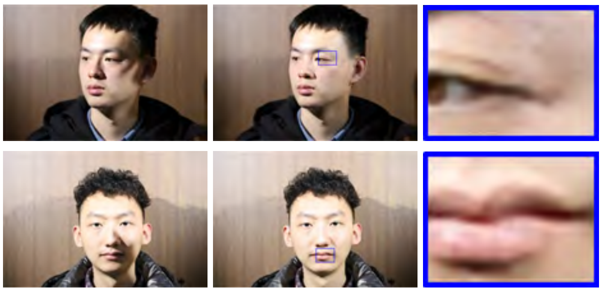
(4) GraphFFNet$_3$: GraphFFNet without feature modulation module, features from GCEncoder are directly connected to the fusion decoder.

We train the four variants on FSD+, and evaluate the results on two test datasets. Table 2 summarizes the quantitative results. From the table, we can observe that all components designed in our method can improve the performance of our method for facial shadow removal. From the table, we can observe that the proposed image filipper, GCEncoder and feature modulation module can help improve the performance of the shadow removal task. The combination leads to the best performance, demonstrating the effectiveness of our GraphFFNet. We also provide some visual results in Figure 11, from which we can ?nd that our GraphFFNet with all the components recovers the best illumination of the shadow-removal regions and looks more realistic.

Ablation of graph convolution. Graph convolution can obtain global contextual information through the topology of the graph, contributing to a better understanding of face information. However, self-attention mechanism also can acquire context-based information. We also conduct an ablation study to examine the effectiveness of graph convolution. We use three different self-attention

Table 2: *Quantitative results of ablation study on FSTD and Zhang datasets using RMSE, PSNR, and SSIM.*

Methods	FSTD			Zhang		
	PSNRąü	SSIMąü	RMSEąý	PSNRąü	SSIMąü	RMSEąý
Base	34.404	0.976	6.564	28.671	0.922	11.059
$GraphFFNet_1$	36.077	0.980	5.532	**29.892**	0.928	9.901
$GraphFFNet_2$	35.148	0.978	6.112	29.441	0.922	10.162
$GraphFFNet_3$	34.359	0.974	6.599	29.064	0.918	10.634
GraphFFNet	**36.423**	**0.982**	**5.389**	29.775	**0.931**	**9.901**



(a) (b) (c)

Figure 12: *Shadow removal results for high resolution images with not less than 3072×2048. (a) is the high resolution shadow images. (b) is our shadow-removal results, and (c) is close-ups for the blue boxes in (b).*

modules (ViT [DBK*20], Non-Local Neural Networks [WGGH18] and DAN [FLT*19]) to replace our GCEncoder in our framework respectively. Table 3 summarizes the numerical results. As can be seen, GraphFFNet using GCEncoder corresponds to better performance, verifying its effectiveness.

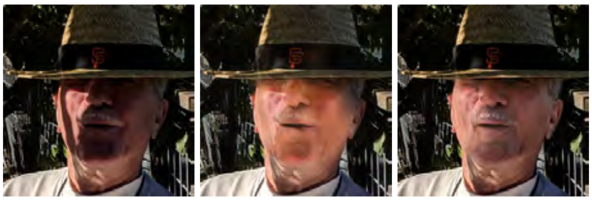
Table 3: *Quantitative results of ablation study on graph convolution.*

Methods	FSD+			Zhang's Datas		
	PSNRąü	SSIMąü	RMSEąý	PSNRąü	SSIMąü	RMSEąý
Using [DBK*20]	35.645	0.979	5.833	28.659	0.922	11.135
Using [WGGH18]	34.810	0.978	6.267	28.832	0.922	10.876
Using [FLT*19]	35.373	0.978	6.112	28.700	0.918	10.933
Our GraphFFNet	**36.423**	**0.982**	**5.389**	**29.775**	**0.931**	**9.901**

4.4. Discussion

Our GraphFFNet also can deal with high resolution image. Figure 12 presents some shadow-removal results. From the results, we observe that our method also works well for high resolution images.

Limitation. Our GraphFFNet can effectively remove shadows in the face images. However, when the shadows are very dark, some high-frequency information on the face like beards, hair and wrinkles may be lost, resulting in blurred details, as shown in Figure 13.



(a) (b) (c)

Figure 13: *Limitation. (a) is the input image. (b) is shadow-removal result produced by [LYW*21], and (c) is our result.*

5. Conclusion

In this paper, we propose a GraphFFNet for facial image shadow removal, which fuses the global topological features into the image features. With the help of the image flipper, our GCEncoder can extract more efficient global contextual features. The feature modulation module fuse global features and local features from MEncoder, enhancing the feature representation capability. Then, we use the fusion decoder to reconstruct features and produce a high-quality shadow removal result for the facial image. Extensive comparisons demonstrate the superiority of our GraphFFNet for facial image shadow removal.

The proposed GraphFFNet is still an image processing method. In the future, we would like to improve it to video-level tasks and apply graph convolution to solve other vision-related problems.

6. Acknowledgments

This work is partially supported by NSFC (No.61902286, No.61972298) and CAAI-Huawei MindSpore Open Fund. It is also supported by Nature Science Foundation of Hubei Province (ZRMS2023000289).

References

[ABBR20] Ammour B., Boubchir L., Bouden T., Ramdani M.: Face–iris multimodal biometric identification system. *Electronics 9*, 1 (2020), 85. 1

[AEHM19] Ali A. A., El-Hafeez T. A., Mohany Y. K.: An accurate system for face detection and recognition. *Journal of Advances in Mathematics and Computer Science 33*, 3 (2019), 1–19. 1

[BDS*17] Bako S., Darabi S., Shechtman E., Wang J., Sunkavalli K., Sen P.: Removing shadows from images of documents. In *Computer Vision–ACCV 2016: 13th Asian Conference on Computer Vision, Taipei, Taiwan, November 20-24, 2016, Revised Selected Papers, Part III 13* (2017), Springer, pp. 173–183. 2

[BT06] Brown M. S., Tsoi Y.-C.: Geometric and shading correction for images of printed materials using boundary. *IEEE Transactions on Image Processing 15*, 6 (2006), 1544–1554. 2

[CEW06] Chen W., Er M. J., Wu S.: Illumination compensation and normalization for robust face recognition using discrete cosine transform in logarithm domain. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 36*, 2 (2006), 458–466. 3

[CH67] Cover T., Hart P.: Nearest neighbor pattern classification. *IEEE transactions on information theory 13*, 1 (1967), 21–27. 5

[CLG*18] Calian D. A., Lalonde J.-F., Gotardo P., Simon T., Matthews I., Mitchell K.: From faces to outdoor light probes. In *Computer Graphics Forum* (2018), vol. 37, Wiley Online Library, pp. 51–61. 3

[CLZX21] Chen Z., Long C., Zhang L., Xiao C.: Canet: A context-aware network for shadow removal. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (2021), pp. 4743–4752. 2

[CPS20] Cun X., Pun C.-M., Shi C.: Towards ghost-free shadow removal via dual hierarchical aggregation network and shadow matting gan. In *Proceedings of the AAAI Conference on Artificial Intelligence* (2020), vol. 34, pp. 10680–10687. 2

[DBK*20] Dosovitskiy A., Beyer L., Kolesnikov A., Weissenborn D., Zhai X., Unterthiner T., Dehghani M., Minderer M., Heigold G., Gelly S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929* (2020). 9

[DH19] Du L., Hu H.: Nuclear norm based adapted occlusion dictionary learning for face recognition with occlusion and illumination changes. *Neurocomputing 340* (2019), 133–144. 1

[DJBY20] Din N. U., Javed K., Bae S., Yi J.: A novel gan-based network for unmasking of masked face. *IEEE Access 8* (2020), 44276–44287. 1

[DLZX19] Ding B., Long C., Zhang L., Xiao C.: Argan: Attentive recurrent generative adversarial network for shadow detection and removal. In *Proceedings of the IEEE/CVF international conference on computer vision* (2019), pp. 10213–10222. 2

[Far02] Farnebäck G.: *Polynomial expansion for orientation and motion estimation*. PhD thesis, Linköping University Electronic Press, 2002. 4

[FDL09] Finlayson G. D., Drew M. S., Lu C.: Entropy minimization for shadow removal. *International Journal of Computer Vision 85*, 1 (2009), 35–57. 2

[FHD02] Finlayson G. D., Hordley S. D., Drew M. S.: Removing shadows from images. In *Computer VisionâĂŤECCV 2002: 7th European Conference on Computer Vision Copenhagen, Denmark, May 28–31, 2002 Proceedings, Part IV 7* (2002), Springer, pp. 823–836. 2

[FLT*19] Fu J., Liu J., Tian H., Li Y., Bao Y., Fang Z., Lu H.: Dual attention network for scene segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2019), pp. 3146–3154. 9

[FQ16] Faraji M. R., Qi X.: Face recognition under varying illuminations using logarithmic fractal dimension-based complete eight local directional patterns. *Neurocomputing 199* (2016), 16–30. 3

[FZG*21] Fu L., Zhou C., Guo Q., Juefei-Xu F., Yu H., Feng W., Liu Y., Wang S.: Auto-exposure fusion for single-image shadow removal. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2021), pp. 10571–10580. 1, 2, 6, 7, 8

[GEB15] Gatys L. A., Ecker A. S., Bethge M.: A neural algorithm of artistic style. *arXiv preprint arXiv:1508.06576* (2015). 6

[GLN*21] Geetha M., Latha R., Nivetha S., Hariprasath S., Gowtham S., Deepak C.: Design of face detection and recognition system to monitor students during online examinations using machine learning algorithms. In *2021 international conference on computer communication and informatics (ICCCI)* (2021), IEEE, pp. 1–4. 1

[GPP06] Gatos B., Pratikakis I., Perantonis S. J.: Adaptive degraded document image binarization. *Pattern recognition 39*, 3 (2006), 317–327. 2

[HFZ*19] Hu X., Fu C.-W., Zhu L., Qin J., Heng P.-A.: Direction-aware spatial context features for shadow detection and removal. *IEEE transactions on pattern analysis and machine intelligence 42*, 11 (2019), 2795–2808. 2

[HJFH19] Hu X., Jiang Y., Fu C.-W., Heng P.-A.: Mask-shadowgan: Learning to remove shadows from unpaired data. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (2019), pp. 2472–2481. 2

[HLL*18] Hu C.-H., Lu X.-B., Liu P., Jing X.-Y., Yue D.: Single sample face recognition under varying illumination via qrcp decomposition. *IEEE Transactions on Image Processing 28*, 5 (2018), 2624–2638. 1

[HXZC21] He Y., Xing Y., Zhang T., Chen Q.: Unsupervised portrait shadow removal via generative priors. In *Proceedings of the 29th ACM International Conference on Multimedia* (2021), pp. 236–244. 1, 2, 6, 7, 8

[HZLH17] Huang R., Zhang S., Li T., He R.: Beyond face rotation: Global and local perception gan for photorealistic and identity preserving frontal view synthesis. In *Proceedings of the IEEE international conference on computer vision* (2017), pp. 2439–2448. 1

[JHK19] Jung S., Hasan M. A., Kim C.: Water-filling: An efficient algorithm for digitized document shadow removal. In *Computer Vision–ACCV 2018: 14th Asian Conference on Computer Vision, Perth, Australia, December 2–6, 2018, Revised Selected Papers, Part I 14* (2019), Springer, pp. 398–414. 2

[JP19] Jo Y., Park J.: Sc-fegan: Face editing generative adversarial network with user's sketch and color. In *Proceedings of the IEEE/CVF international conference on computer vision* (2019), pp. 1745–1753. 1

[JST21] Jin Y., Sharma A., Tan R. T.: Dc-shadownet: Single-image hard and soft shadow removal using unsupervised domain-classifier guided network. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (2021), pp. 5027–5036. 2

[JZH*20] Jiang L., Zhang C., Huang M., Liu C., Shi J., Loy C. C.: Tsit: A simple and versatile framework for image-to-image translation. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part III 16* (2020), Springer, pp. 206–222. 5

[KAGG19] Kartynnik Y., Ablavatski A., Grishchenko I., Grundmann M.: Real-time facial surface geometry from monocular video on mobile gpus. *arXiv preprint arXiv:1907.06724* (2019). 3

[LCC20] Lin Y.-H., Chen W.-C., Chuang Y.-Y.: Bedsr-net: A deep shadow removal network from a single document image. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2020), pp. 12905–12914. 2

[LG08] Liu F., Gleicher M.: Texture-consistent shadow removal. In *Computer Vision–ECCV 2008: 10th European Conference on Computer Vision, Marseille, France, October 12-18, 2008, Proceedings, Part IV 10* (2008), Springer Berlin Heidelberg, pp. 437–450. 2

[LHH*22] Liu Y., Hou A., Huang X., Ren L., Liu X.: Blind removal of facial foreign shadows. 1, 6, 7, 8

[LS19] Le H., Samaras D.: Shadow removal via shadow image decomposition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (2019), pp. 8578–8587. 2

[LYD10] Li Q., Yin W., Deng Z.: Image-based face illumination transferring using logarithmic total variation models. *The visual computer 26* (2010), 41–49. 3

[LYW*21] Liu Z., Yin H., Wu X., Wu Z., Mi Y., Wang S.: From shadow generation to shadow removal. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2021), pp. 4927–4936. 1, 2, 6, 7, 8, 9

[MXZP12] Meng G., Xiang S., Zheng N., Pan C.: Nonparametric illumination correction for scanned document images via convex hulls. *IEEE transactions on pattern analysis and machine intelligence 35*, 7 (2012), 1730–1743. 2

[OL09] Oliveira D. M., Lins R. D.: A new method for shading removal and binarization of documents acquired with portable digital cameras. In *Proceedings of Third International Workshop on Camera-Based Document Analysis and Recognition, Barcelona, Spain* (2009), pp. 3–10. 2

[QTH*17] Qu L., Tian J., He S., Tang Y., Lau R. W.: Deshadownet: A multi-context embedding deep network for shadow removal. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2017), pp. 4067–4075. 2

[Sid19] Sidorov O.: Conditional gans for multi-illuminant color constancy: Revolution or yet another approach? In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops* (2019), pp. 0–0. 2

[SKCJ18] Sengupta S., Kanazawa A., Castillo C. D., Jacobs D. W.: Sfsnet: Learning shape, reflectance and illuminance of facesin the wild'. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2018), pp. 6296–6305. 3

[SL08] Shor Y., Lischinski D.: The shadow meets the mask: Pyramid-based shadow removal. In *Computer Graphics Forum* (2008), vol. 27, Wiley Online Library, pp. 577–586. 2

[VHS17] Vicente T. F. Y., Hoai M., Samaras D.: Leave-one-out kernel optimization for shadow detection and removal. *IEEE Transactions on Pattern Analysis and Machine Intelligence 40*, 3 (2017), 682–695. 2

[WGGH18] Wang X., Girshick R., Gupta A., He K.: Non-local neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2018), pp. 7794–7803. 9

[WHCO08] Wen C.-L., Hsieh C.-H., Chen B.-Y., Ouhyoung M.: Example-based multiple local color transfer by strokes. In *Computer Graphics Forum* (2008), vol. 27, Wiley Online Library, pp. 1765–1772. 2

[WLW*20] Wei Y., Liu M., Wang H., Zhu R., Hu G., Zuo W.: Learning flow-based feature warping for face frontalization with illumination inconsistent supervision. In *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XII 16* (2020), Springer, pp. 558–574. 1

[WLY18] Wang J., Li X., Yang J.: 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* (2018), pp. 1788–1797. 1, 2, 6, 7, 8

[WTBS07] Wu T.-P., Tang C.-K., Brown M. S., Shum H.-Y.: Natural shadow matting. *ACM Transactions on Graphics (TOG) 26*, 2 (2007), 8–es. 2

[WY22] Wang H., Yan W. Q.: Face detection and recognition from distance based on deep learning. In *Aiding Forensic Investigation Through Deep Learning and Machine Learning Frameworks*. IGI Global, 2022, pp. 144–160. 1

[WYDL18] Wang X., Yu K., Dong C., Loy C. C.: Recovering realistic texture in image super-resolution by deep spatial feature transform. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (2018), pp. 606–615. 5

[WYW*22] Wan J., Yin H., Wu Z., Wu X., Liu Y., Wang S.: Style-guided shadow removal. In *Computer Vision–ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23–27, 2022, Proceedings, Part XIX* (2022), Springer, pp. 361–378. 1, 2, 6, 7, 8

[XSHR18] Xu Z., Sunkavalli K., Hadap S., Ramamoorthi R.: Deep image-based relighting from optimal sparse samples. *ACM Transactions on Graphics (ToG) 37*, 4 (2018), 1–13. 3

[XSXM13] Xiao C., She R., Xiao D., Ma K.-L.: Fast shadow removal using adaptive multi-scale illumination transfer. In *Computer Graphics Forum* (2013), vol. 32, Wiley Online Library, pp. 207–218. 2

[XXZC13] Xiao C., Xiao D., Zhang L., Chen L.: Efficient shadow removal using subregion matching illumination transfer. In *Computer Graphics Forum* (2013), vol. 32, Wiley Online Library, pp. 421–430. 2

[XZXH21] Xiong Q., Zhang X., Xu X., He S.: A modified chaotic binary particle swarm optimization scheme and its application in face-iris multimodal biometric identification. *Electronics 10*, 2 (2021), 217. 1

[YY00] Yang Y., Yan H.: An adaptive logical method for binarization of degraded document images. *Pattern recognition 33*, 5 (2000), 787–807. 2

[ZBT*20] Zhang X., Barron J. T., Tsai Y.-T., Pandey R., Zhang X., Ng R., Jacobs D. E.: Portrait shadow manipulation. *ACM Transactions on Graphics (TOG) 39*, 4 (2020), 78–1. 1, 6, 8

[ZGZ22] Zhang X. F., Gu C. C., Zhu S. Y.: Spa-former: Transformer image shadow detection and removal via spatial attention. *arXiv preprint arXiv:2206.10910* (2022). 1, 6, 7, 8

[ZLL*20] Zhou H., Liu J., Liu Z., Liu Y., Wang X.: Rotate-and-render: Unsupervised photorealistic face rotation from single-view images. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (2020), pp. 5911–5920. 1

[ZLZX20] Zhang L., Long C., Zhang X., Xiao C.: Ris-gan: Explore residual and illumination with generative adversarial networks for shadow removal. In *Proceedings of the AAAI Conference on Artificial Intelligence* (2020), vol. 34, pp. 12829–12836. 2

[ZTF*09] Zhang T., Tang Y. Y., Fang B., Shang Z., Liu X.: Face recognition under varying illumination using gradientfaces. *IEEE Transactions on image processing 18*, 11 (2009), 2599–2606. 3

[ZZMC18] Zhang W., Zhao X., Morvan J.-M., Chen L.: Improving shadow suppression for illumination robust face recognition. *IEEE transactions on pattern analysis and machine intelligence 41*, 3 (2018), 611–624. 1

[ZZML13] Zhang Y., Zhao T., Mo Z., Li W.: A method of illumination effect transfer between images using color transfer and gradient fusion. In *2013 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference* (2013), IEEE, pp. 1–6. 2

[ZZX15] Zhang L., Zhang Q., Xiao C.: Shadow remover: Image shadow removal based on illumination recovering optimization. *IEEE Transactions on Image Processing 24*, 11 (2015), 4623–4636. 2