Journal Pre-proof

Unsupervised learning non-uniform face enhancement under physics-guided model of illumination decoupling

Xin Ding, Jing Fang, Zhongyuan Wang, Qiong Liu, You Yang, Zhenyu Shu

PII: S0031-3203(25)00014-7
DOI: https://doi.org/10.1016/j.patcog.2025.111354
Reference: PR 111354

To appear in: *Pattern Recognition*

Received date : 28 December 2023
Revised date : 13 October 2024
Accepted date : 7 January 2025

Please cite this article as: X. Ding, J. Fang, Z. Wang et al., Unsupervised learning non-uniform face enhancement under physics-guided model of illumination decoupling, *Pattern Recognition* (2025), doi: https://doi.org/10.1016/j.patcog.2025.111354.

Unsupervised Learning Non-uniform Face Enhancement Under Physics-guided Model of Illumination Decoupling

Xin Ding[a], Jing Fang[b], Zhongyuan Wang[c], Qiong Liu[a,*], You Yang[a] and Zhenyu Shu[d]

[a]School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan, 430074, China
[b]School of Computer Science, Northwestern Polytechnical University, Xian, 710127, China
[c]School of Computer Science, Wuhan University, Wuhan, 430072, China
[d]NingboTech University, Ningbo, 315100, China

ARTICLE INFO

Keywords:
Deep Learning
Non-uniform Low-light Images
Face Enhancement
Unsupervised Learning

ABSTRACT

Dynamic lighting conditions pose challenges for capturing well-illuminated facial images, resulting in difficulties for effective recognition. Previous enhancement methods have relied on paired facial datasets with varying brightness levels to improve images quality. However, acquiring high-quality pairs of facial images under different lighting conditions is quite challenging, and these approaches often struggle to perform effectively in scenarios with highly complex lighting. In response to these limitations, this paper introduces a NUL-Face dataset, which consists of non-uniformly illuminated face images without matching pairs. And we propose an unsupervised non-uniform illumination face enhancement algorithm based on a physics-guided Model. Firstly, we perform the model analysis and establish a two-phase iterative enhancement process, incorporating a consistency constraint on the results. The framework guides low-quality images toward approximating a target high-quality uniform brightness output through training. It leverages illumination analysis to iteratively update the images and enhance facial details. Furthermore, by incorporating constraint terms, our model effectively decouples the influence of intricate illumination conditions on the images, facilitating their reconstruction across diverse states. Finally, we employ a discriminative network to ensure the naturalness of the enhanced results. Experimental results demonstrate the superiority of our proposed method over existing alternatives, underscoring its effectiveness in tackling complex lighting conditions and enhancing non-uniformly illuminated face images.

1. Introduction

Due to complex lighting factors, capturing facial images at night or in backlit scenes is typically challenged by low light. Poorly exposed images can also result in low quality and unclear facial details. To improve the image quality [35], various enhancement methods have been proposed to address this problem.

In some cases, images tend to be non-uniform in brightness. In practice, controlling illumination is often unfeasible. Without compensating for this uneven lighting, achieving satisfactory recognition results becomes impossible [14]. As shown in Fig 1(a), a part of the face is affected by the shadow due to the particular lighting condition of the shooting, resulting in poor subjective recognition, which in some cases may affect subsequent processing. Existing

Xin Ding, Liu Qiong and You Yang are with the the School of Electronics Information and Communications, Huazhong University of Science and Technology. Jing Fang is with the School of Computer, Northwestern Polytechnical University. Zhongyuan Wang is with the School of Computer, Wuhan University. Zhenyu Shu is with the School of Computer, and Data Engineering, NingboTech University. This work was supported in part by the National Natural Science Foundation of China under Grants 61872321, 62172356.

*Corresponding author
xding07@163.com (X. Ding); jingfang@whu.edu.cn (J. Fang); wzy_hope@163.com (Z. Wang); q.liu@hust.edu.cn (Q. Liu); yangyou@hust.edu.cn (Y. Yang); shuzhenyu@nit.zju.edu.cn (Z. Shu)
ORCID(s): 0000-0002-9541-3021 (X. Ding)

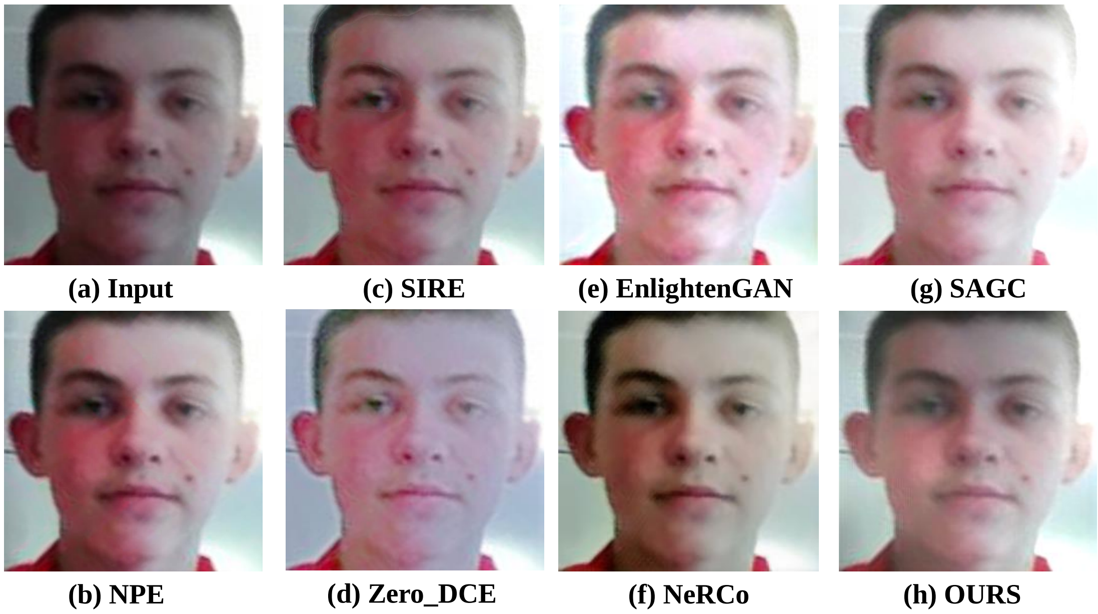


Figure 1: The input face exhibits non-uniform brightness due to backlighting. (a) The non-uniform input face, (b) the result generated by Zero-DCE [9], (c) the result generated by EnlightenGAN [10], (d) the result generated by SAGC [3], (e) the result generated by NPE [30], (f) the result generated by SIRE [6], (g) the result generated by NeRCo [34], (h) our result.

uniform brightness enhancement algorithms use equalization criteria to enhance different regions of the image, and may therefore produce under-exposed or over-exposed regions [31].

To enhance the non-uniform image, Wang *et al.* [30] propose the model Naturalness Preserved Enhancement (NPE) method based on brightness analysis. In Fig. 1(b), the enhanced result improves the subjective performance, but still suffers from the low light. Fu *et al.* [6] introduce Simultaneous Illumination and Reflectance Estimation (SIRE) for non-uniform image enhancement, which is also based on image brightness estimation. His enhanced image in Fig. 1(c) is still not good enough. When traditional methods are used for this problem, the reconstruction performance is still not good enough. Then, many deep learning methods [20, 29] are proposed for the non-uniform low-light enhancement. These methods require many input/output image pairs to train the networks. The low-quality faces such as Fig 1(a) are in the special brightness state which is difficult to obtain for the training pairs. Then, some methods of face relighting are proposed for data simulation. Nestmeyer *et al.* [21] explore an end-to-end deep learning architectures that both delight and displease an image of a human face based on the BRDF model [13].

Instead of directly learning an image-to-image mapping as in previous methods, these solutions introduce unsupervised networks and illumination analysis (e.g. Zero-Reference Deep Curve Estimation, Zero-DCE [9]). Jiang *et al.* [10] propose an unsupervised generative adversarial network (EnlightenGAN) that can be trained without low/normal brightness image pairs and performs well on various real-world test images. However, these methods, which use high-low adaptive enhancement or multi-exposure fusion enhancement to achieve non-uniform low-light image enhancement, can result in poor foreground enhancement or background overexposure. As demonstrated in Fig. 1(d) and (e), the image cannot be effectively reconstructed due to unnatural detail reconstruction and the improved faces exhibit unnatural colors. The recent proposed unsupervised methods Neural Representation for Cooperative Low-light Image Enhancement (NeRCo) [34] and semantics-aware low-light enhancement model (SAGC) [3] also enhance the non-uniform low-light images. Fig. 1(f) shows a face affected by under-exposure, while Fig. 1(g) depicts a face suffering from overexposure. Interestingly, the input face in Fig. 1(a) retains good details in well-lit regions, whereas all the mentioned methods alter these intensities, affecting the representation.

Furthermore, to better reconstruct face images under different lighting conditions, two types of methods are proposed, including portrait relighting [38] and shadow removal [4]. The former relies on supervised training datasets to implement the enhancement of face images with different lighting conditions. The latter removes shadows by detecting and analyzing them, thereby realizing the enhancement of face images. However, due to the peculiarities of the backlit illumination state, it is difficult for these methods to reconstruct the ideal enhancement results when the foreground image is in an extremely low-light state. The performance of existing work shows that image enhancement is poor without prior knowledge. However, their results are also not good enough.

In this paper, we introduce an unsupervised augmentation method for non-uniform low-light face images without corresponding high-quality training pairs. Our approach begins with data analysis and the proposal of a physics-guided model for image decoupling. Through this model, we aim to achieve an optimal representation for the subsequent stages and establish a two-phase iterative enhancement process that incorporates consistency constraints on the results. Following this, we present a face illumination decoupling network framework that guides low-quality inputs toward approximating a high-quality uniform brightness image through training, designed to iteratively update the face reconstruction while focusing on enhancing details. Finally, we employ a discriminative network to constrain the naturalness of the enhanced results.

The main **contributions** of this work are summarized as follows:

1) We built a dataset **NUL-Face** [1] by collecting the non-uniform illumination face images from the website. The dataset contains non-uniform illumination images without training pairs. Our unsupervised learning model for face enhancement is trained on the dataset.

2) We propose the physics-guided model to enhance the non-uniform illumination face images without training pairs. Our network is based on model analysis and illumination decoupling to update the face images.

[1] https://github.com/jeanfang/NUL-Face.git.

(a) Faces (b) Simulated Faces (c) Shadings (d) Albedos (d) Residual Terms

Figure 2: The high-quality face images are decomposed into albedos, shadings and the corresponding residuals. (a) The faces. (b) The simulated non-uniform faces. (c) The shadings separated from the faces. (d) The albedos of the faces, (e) The residual terms.

3) Extensive experiments demonstrate the effectiveness of our method compared to the existing methods. For unsupervised training on the randomly selected non-uniform illumination face images, our model performs well on the non-uniform illumination face images in reality.

The remainder of the paper is organized as follows. In Section 2, we introduce some related works. Our method is proposed in Section 3. The experiments on synthetics and real-world low-quality images are conducted in Section 4. In Section 5, we conclude our work.

2. Related Work

2.1. Portrait Relighting

To solve the backlit, many works [21] have been proposed to improve or generate the exposure state of the image. In Portrait Shadow Manipulation (PSM), Zhang *et al.* [38] present a computational approach that gives casual photographers some of this control, thereby poorly-lit portraits to be relit post-capture in a realistic and easily controllable way. In [28], Wang *et al.* formulate the single image relighting task and propose a novel Deep Relighting Network (DeepRelight) with scene reconversion, shadow prior estimation, and re-renderer to form the required estimation under the target light source. He *et al.* [32] present a hybrid parametric neural relighting (PN-Relighting) framework for single portrait relighting.

However, the experimental results of the existing work show that it is difficult to reconstruct ideal enhancement results by relighting algorithms due to the special nature of the backlit illumination state, or when the foreground image is in the extremely low lighting conditions, where the image enhancement is poor in the absence of prior knowledge.

2.2. Non-uniform Low-light Enhancement

Low-illumination image enhancement approaches [9] amplify illumination and improve the visibility of dark images. They are classified mainly into two categories: Retinex decomposition-based [12], and deep learning-based [36].

The above illumination enhancement methods can not be applied to the non-uniform illumination scene directly, and then some algorithms [29] are proposed to address the problem. Instead of directly learning an image-to-image mapping as in previous work, some methods introduce the illumination analysis (Zero-Reference Deep Curve Estimation (Zero-DCE) [9]) or multiple exposure illumination [18] for non-uniform illumination images enhancement adaptively. In [33], a non-local similarity decomposition model based on the Retinex theory is proposed to obtain high-quality enhanced results for non-uniform low-light images. To address the high complexity, Zhou *et al.* [39] design a surround function, Adaptive Surround Function, to estimate the illumination map, which has the conventional surround function form and can be trained end to end for low-light enhancement.

Wang *et al.* [29] introduce intermediate illumination in our network to associate the input with the expected enhancement result (DeepUPE), which augments the network's capability to learn complex photographic adjustment from expert-retouched input/output image pairs. In [26], focusing on the fact that the intensity histogram of a backlit image shows a characteristic bimodal distribution, Ueda *et al.* propose an image enhancement method for single backlit images using histogram specification. Liu *et al.* [17] propose a Locally-Adaptive Embedding Network, to realize high-quality low-light image enhancement with locally-adaptive kernel selection and feature adaptation for multi-distribution issues. Yang *et al.* [34] introduce the semantic-orientated supervision method NeRCo with priors for non-uniform low-light images from the pre-trained vision-language model, which encouraged results to meet subjective expectations, finding more visual-friendly solutions.

When using high and low brightness adaptive enhancement or multiple exposure fusion enhancement algorithms for enhancement, the experimental results show that the adaptive optimization is often insufficient, resulting in poor foreground enhancement or background over-exposure. And because of the insufficient reconstruction of details, image reconstruction cannot be achieved effectively.

2.3. Unsupervised Image Enhancement

Benefiting from the development of unsupervised learning methods in image processing [16], many unsupervised methods have been proposed to address the task of low-light enhancement. Fu *et al.* [7] designed an illumination-aware attention module based on Generative Adversarial Network (GAN) [8] that enhanced the feature extraction of the network to address the problems of image noise and color bias, as well as improve the visual quality for the unpaired low/normal-light images. Ye *et al.* [11] introduced an unsupervised method that integrates a layer decomposition network and a light-effects suppression network for low-light enhancement (LDLES-Net). Chen *et al.* [3] proposed a semantics-aware yet unsupervised low-light enhancement model (SAGC) that utilized gamma correction and was guided by high-quality reference images and the inherent semantic information for image enhancement. Sun *et al.* [25] proposed an unsupervised Multi-Branch with High-Frequency Enhancement Network which contain an multi-Branch network and a high-frequency components enhancement module for image enhancement.

Although these methods solve the problem of unpaired data, the quality of the enhanced images is limited, e.g., these methods enhance the holistic brightness of the input but cause over-exposure in the regions with relatively high brightness. Besides, these methods cannot deal with the noise well simultaneously, which has a significant impact on the visual quality of the enhanced images.

3. Method

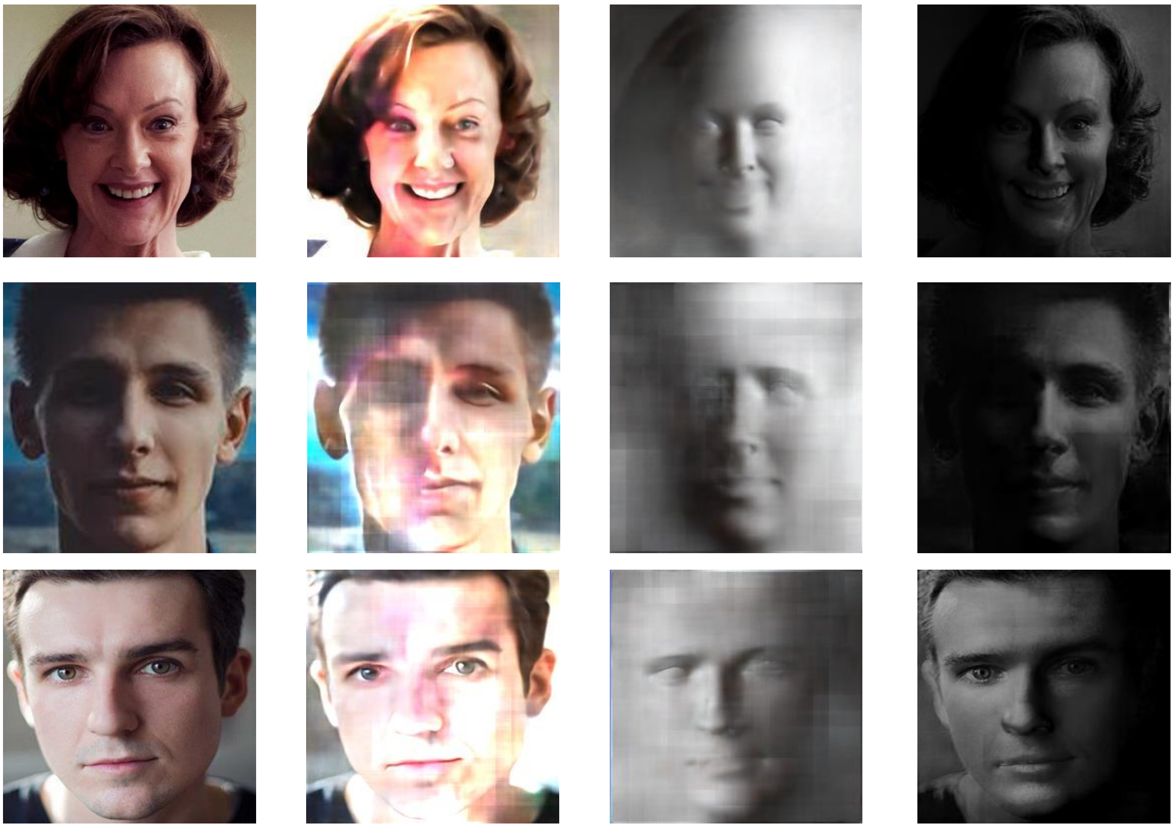
3.1. Motivation

As mentioned above, it is difficult to obtain aligned low-light and target face pairs for complex, non-uniformly illuminated face images. The few known corresponding face datasets also fail to achieve full illumination coverage. Therefore, the network models trained on such datasets are difficult to apply to different real-world scenarios. Due to the lack of brightness analysis of complex low-light faces, known low-light enhancement algorithms face challenges in achieving both detail enhancement and uniform illumination representation for non-uniform low-light faces.

To address these issues, we attempt to achieve non-uniform illumination face enhancement without high-quality and low-quality pairs and realize image enhancement with illumination decoupling based on the 3D Morphable Model (3DMM) [1]. Inspired by [21], our solution introduces an optimal representation model for face illumination

$$L = a \cdot s. \tag{1}$$

This simplified model can be seen as a diffuse decomposition into two face features albedo a and shading s as in Fig. 3, and L represents the target high-quality face image. In this model, the albedo and shading can be considered separately as the original face and the light feedback.



(a) Faces (b) Albedos (c) Shadings (d) Residual Terms

Figure 3: The non-uniform low-light face images are decomposed into the albedos, shadings, and the corresponding residual terms. (a) The low-quality faces, (b) the abledos of these faces, (c) the divided shadings, and (d) the residual terms.

Of course, such a simple model implements feature decomposition in an idealized state with residual values $res = L - a \cdot s$. However, under uniform illumination, the residual values for most areas of the face are extremely small in Fig. As the illumination changes and the face is under non-uniform illumination, it is difficult for the model to decompose the original face, leading to increasingly large residual values.

In our scenario, the initial input faces are non-uniform and low-light. Given a low-light face image L_0 with unknown illumination, we assume it to be a high-quality face to apply the above model and decompose it into two sub-distributions a_0 and s_0. However, since the initial face is non-uniform, we introduce a residual term res_0 in the decomposed model.

$$L_0 = (a_0 + r) \cdot s_0 = a_0 \cdot s_0 + res_0, \tag{2}$$

where r means the light-varying residual, and the residual term $res_0 = r * s_0$ is determined by the varying illumination and can be nearly seen as the difference between the low-light and target natural images. In Fig. 3, res_0 also reflects the illumination distribution.

The distributions of three non-uniform images are shown in Fig. 3. The image in the first row is from the test set of [40] generated by 3DMM, where the face images are in a non-uniform illumination state. Due to the correlation of the data, the trained model reconstructs albedo and shading accurately. However, the reconstruction of the decoupled information is not accurate enough due to the non-uniform illumination. We define the difference res_0 as the residual term. By analyzing the information of res_0, we find that it preserves more high-frequency information about the face, while decoupling the illumination state of the original image more clearly. The images in the second and third rows are random faces independent of the dataset, and under the same training model, the albedo and shadow can be decoupled, but their image information and sharpness are less accurate. However, the residual term retains the same representation of the illumination state.

Our solution to improve the non-uniform face images is based on information updating and illumination decoupling. First, we update the input L_0 to the image L_1 with the constraint of the illumination distribution $||res_0 + r_1 \cdot s_0||_2^2$ as:

$$J(L_1, r_1) = \arg min\{||L_1 - (L_0 + r_1 \cdot s_0)||_2^2 + ||res_0 + r_1 \cdot s_0||_2^2\}, \tag{3}$$

where r_1 is the compensation illumination to optimize the face image $L_1 = L_{ini} + r_1 \cdot s_0$. It is reconstructed under the constraint of the illumination decoupling term res_0. After obtaining the image L_1, we replace L_{ini} with the new image in the generations a_1 and s_1.

$$L_1 = a_1 \cdot s_1 + res_1. \tag{4}$$

Then we update the input image L_1 to the image L_2 with the residual term:

$$J(L_2, r_2) = \arg min\{||L_2 - (L_1 + r_2 \cdot s_1)||_2^2 + ||res_1 + r_2 \cdot s_1||_2^2\}, \tag{5}$$

where r_2 is also the compensation lighting to optimize the face image $L_2 = L_1 + r_2 \cdot s_1$.

As the illumination parameters r_1 and r_2 are updated iteratively, the information in the face image becomes more and more similar to the target image with uniform illumination, and the generations of different iterations become closer to each other, as $L_1 \approx L_2$. Then the generation tends towards the target.

3.2. Networks

To realize the reconstruction for the final result, we design an iterative CNN network in Fig. 4. The generation network follows the analysis of the face enhancement with the illumination distribution and updates the face result without the training pairs. We enhance the low-light input L_0, and generate the result L_2 with the intermediate result L_1.

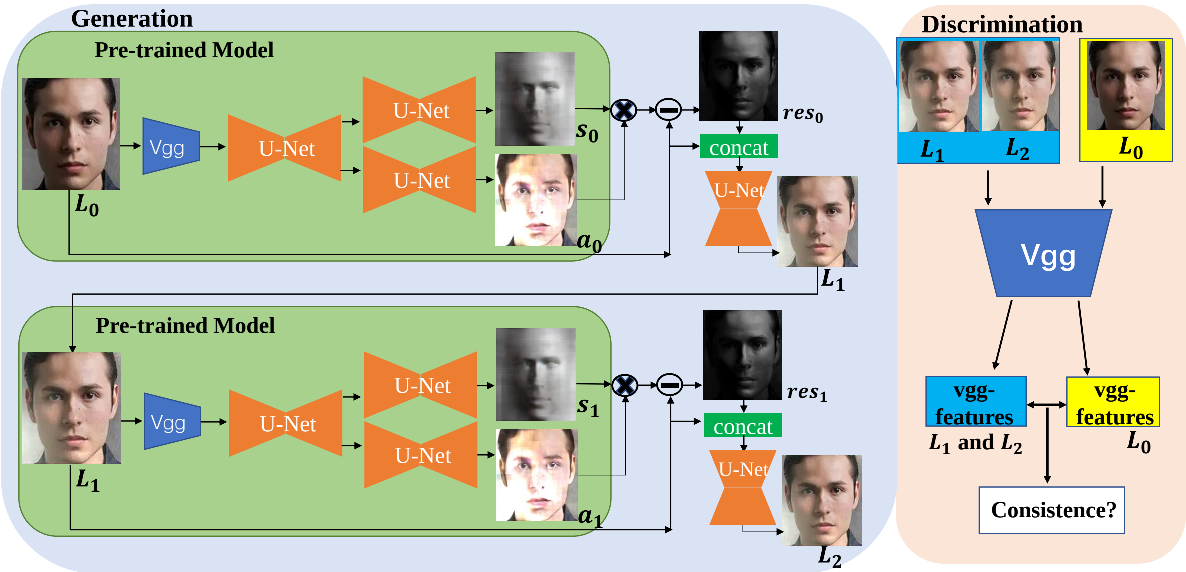


Figure 4: The framework of our proposed non-uniform low-light face enhancement algorithm. The generation involves two steps to update the results. The network in each step has the same network structure, which decomposes the input image L into two sub-distributions: albedo a and shading s. It then generates the residual term res for illumination decoupling. Finally, the residual term res is concatenated with the input image L to feed the enhanced face image network. The VGG discriminative network transfers the generations and the input image to the features and maintains their feature consistency to constrain the image quality of the generations.

As in Eq. (1), the input L_0 is decomposed into features with network modules. The network modules are included by VGG [23] and U-Net [22]. The input is fed into VGG and U-Net for facial features. Then these features are split into two components a_0 and s_0 by two U-Net modules. For better generations, this part of the network is pre-trained with the datasets from [40]. Then we generate the results res_0 according to Eq. (2). The generation of res_0 can be used to estimate the varying illumination s_1. So we feed in L_0 and res_0 for the intermediate result L_1 according to Eq. (3).

As shown, the estimation for the generation is based on the iterative update. After obtaining the result L_1, we feed it into the network in the second stage for the final result L_2 with the residual result res_1, which is generated from the sub-distributions a_1 and s_1. The iterative training from the initial image L_0 to the intermediate result L_1 and then on to the generation L_2 is extended. If the output L_2 and the intermediate result L_1 are close enough, it means that the network has been sufficiently trained and the input image has been accurately updated.

In order to achieve training for reinforcement without training pairs, the proposed network must be constrained from three perspectives.

The constraint of the training goal. In this work we build an iterative framework with two generations L_1 and L_2. As an argument of motivation, the result should be shared with little varying light, which makes the generation L_1 tend to L_2 during the training.

The constraint of the convergence direction. The goal of the network is to realize $L_1 \approx L_2$, but we should also control the convergence of the improvement. Especially the input L_0 is in the complex ill-posed brightness, then we introduce the analysis of the illumination for the convergence direction.

The image quality constraint. The generations of the network may contain the artifacts during training. Then we constrain the improved quality with the image characteristics. Then the discriminative network in Fig. 4 is proposed to constrain the consistency between the generations and the natural high quality images for better image quality.

During training, the above constraints can be set as different loss functions for the well enhancement.

3.3. Loss Functions

The loss functions for the training goal. To obtain better face images, we set different loss functions for the unsupervised training. The first loss function is proposed to limit the sufficient optimization, making the generation L_2 approach to the intermediate results L_1,

$$loss_0 = ||L_2 - L_1||_2^2. \tag{6}$$

In the unpaired setting, the second loss function follows the self-feature preserving loss to constrain the VGG [23] feature distance between the low-light input and its enhanced normal light output images by the VGG model Φ.

$$loss_1 = ||\Phi(L_0) - \Phi(L_1)||_2^2 + ||\Phi(L_0) - \Phi(L_2)||_2^2. \tag{7}$$

The third loss function makes the generations contain the facial features as the input. We utilize the model to generate the parsing features P as [24] from the generations L_1 and L_2, and constrain the distance between them and the input low-light image.

$$loss_2 = ||P(L_0) - P(L_1)||_2^2 + ||P(L_0) - P(L_2)||_2^2. \tag{8}$$

The loss functions for the convergence direction. The fourth loss function is to enhance the low-light region with the brightness analysis. According to [10], we adopt image intensity maps, which decompose each low-light face into two parts of low-light M_l (Fig. 5(d)) and high-light M_h (Fig. 5(c)), as binary attention maps. The loss function is set as follows:

$$loss_{3a} = ||L_1 \cdot M_l/sum(M_l) - L_0 \cdot M_h/sum(M_h)||_2^2, \tag{9}$$

$$loss_{3b} = ||L_2 \cdot M_l/sum(M_l) - L_0 \cdot M_h/sum(M_h)||_2^2. \tag{10}$$

Our goal is to make the intensities in the low-light part approach the intensities in the high-light part. Therefore, the fifth loss function is proposed to constrain the generations L_1 and L_2, which maintain the intensities in the high-illumination region of the input L_0, as

$$loss_4 = ||L_1 \cdot M_h - L_0 \cdot M_h||_2^2 + ||L_2 \cdot M_h - L_0 \cdot M_h||_2^2. \tag{11}$$

The loss functions for the image quality. To preserve the monotonicity relationships between neighboring pixels for image smoothness, we introduce the sixth loss function to the generations with the **tv** loss function $loss_{tv}$ in [9]:

$$loss_5 = loss_{tv}(res_0) + loss_{tv}(res_1) + loss_{tv}(L_1) + loss_{tv}(L_2). \tag{12}$$

The last loss function acts as the discriminator of GAN. We take the model of the discriminator D from the generations L_1 and L_2 and constrain the distance between them and the input low-light image.

$$loss_6 = ||D(L_0) - D(L_1)||_2^2 + ||D(L_0) - D(L_2)||_2^2. \tag{13}$$

The above loss functions all contribute to the final results, and the total loss function is expressed as follows:

$$loss_t = \sum_{i=0}^{6} \alpha_i \cdot loss_i, \tag{14}$$

where α_i is the balance parameter determined by experimental analysis.

3.4. Implementation Details

Training Setting: the architecture of the module follows Fig. 4. The model is trained using the ADAM optimizer [15]. To maintain the stability of the global loss and to balance these loss values, we set the balance parameter α_3 to 10, α_6 to 0.1 and other parameters to 1 after experimental verification. We change these parameters in the ablation study experiments. The learning rate is set to $5e-5$. The batch size is set to 1, and the number of training epochs is set to 50 during training. In every 10 epochs, we validate the performance of the enhancement and choose the best one as the test model. Our experiments are implemented on Pytorch using NVIDIA RTX 2080Ti GPUs.

4. Experiments

4.1. Databases

We have experimentally verified the effectiveness of our algorithm on the self-built non-uniform low-light face image dataset NUL-Face, using a reverse image search engine, as shown in Fig. 5. The dataset comprises 2000 face

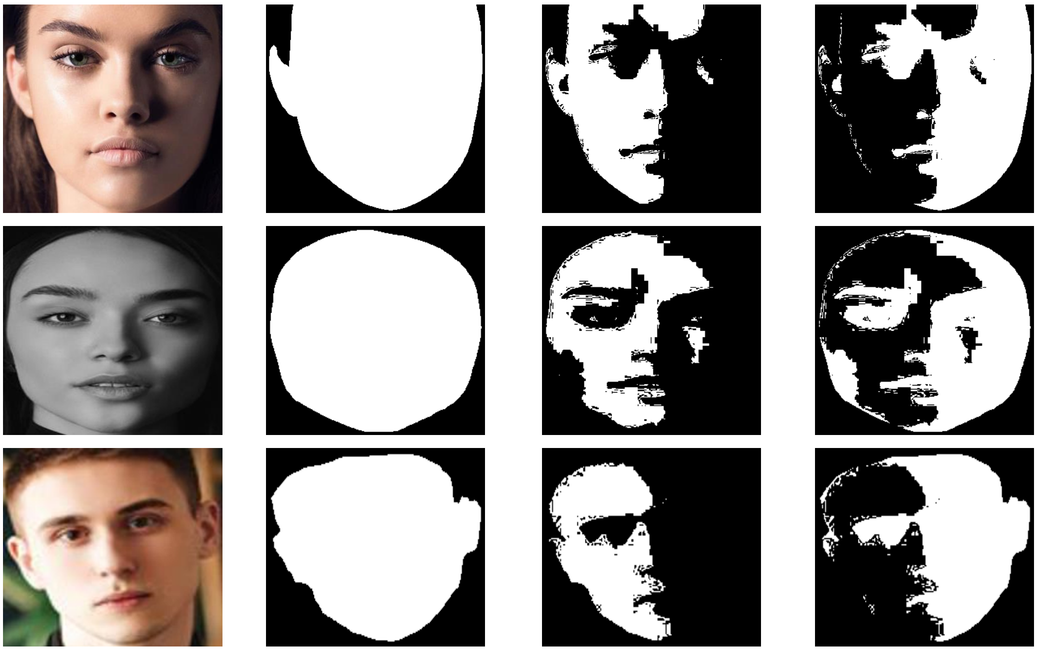


Figure 5: The images from our face database NUL-Face, including (a)face images, (b)face masks, (c)high-illumination, and (d)low-illumination binary maps.

images for training and 50 for testing, all of which are downloaded from the website. Notably, neither the training nor testing images have corresponding high-quality counterparts. Furthermore, we evaluated the performance on the 27 faces cropped from the MIT-Adobe dataset [2]. The images all are sized as 224×224.

To analyze the illumination states of the low-quality images as Fig. 5(a), we estimate the high-illumination binary maps M_h (Fig. 5(c)) and low-illumination binary maps M_l (Fig. 5(d)) based on the face mask. The face mask as Fig. 5(b) is generated by the parsing map learning method [37].

4.2. Ablation Study

Effect of loss functions: in this part, we provide an ablation study on the effect of the different loss functions, presented in Sect. 3.3. During the experiment, we set each of the different balance parameters α_i as zero, and remove its constraint on the training network. And then we test their effects. When we set the balance parameter $\alpha_0 = 0$, it means that we get rid of the constraint $loss_0$, which is represented as without α_0. Others are similar. As shown in Fig. 6, the different loss functions have their effects on the generations.

The input face in Fig. 6(a) suffers from non-uniform illumination, and part of the face has low brightness. The result without α_0 in Fig. 6(b) is slightly over-exposed with artifacts and streaks. This is because the generation without

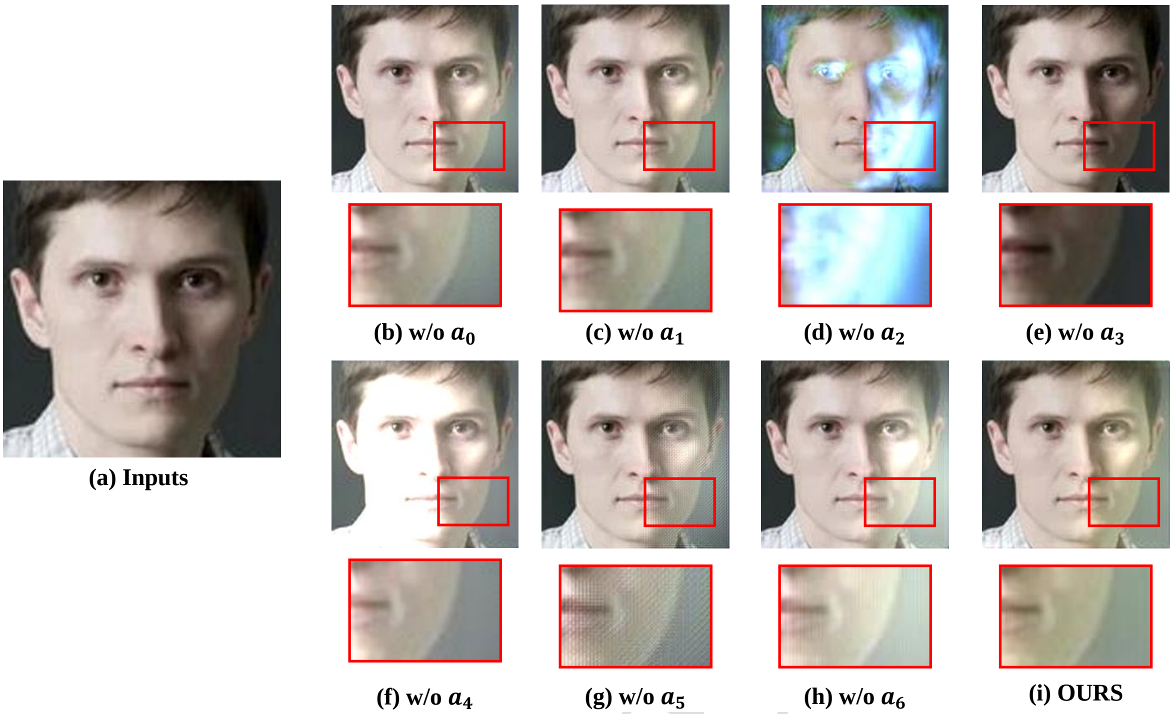


Figure 6: The results of the ablation study. The unsupervised network is constrained by different loss functions, then we set the corresponding balance parameter to zero to test its impact. (a) The input image, (b) the result without α_0, (c) the result without α_1, (d) the result without α_2, (e) the result without α_3, (f) the result without α_4, (g) the result without α_5, (h) the result without α_6, (i) our result. The boxes in red are zoomed in for more details.

$loss_0$ cannot promote the good brightness state. The result without α_4 in Fig. 6(f) is more over-exposed, especially in the high-illumination region of the input. It shows that the loss function $loss_4$ really preserves the original high-illumination region of the face. The results without α_1 (Fig. 6(c)), α_5 (Fig. 6(g)) and α_6 (Fig. 6(h)) suffer from the artifacts and streaks, representing that these loss functions provide the constraint of the image details and quality. The result without α_2 in Fig. 6(d) is in a bad situation, meaning that the VGG feature distance is important during the enhancement. And the result without α_3 is similar to the input face, showing that the illumination analysis provides the convergence direction for the training. Without $loss_3$, the network cannot even enhance the input face. The above analysis demonstrates the validity of our different loss functions.

Furthermore, we also test the objective metrics of the generations in the ablation study. For our testing images that have no corresponding reference targets, we utilize the non-reference metrics Perception-based Image Quality Evaluator (PIQE) [27] and Contrast-Distorted Images Quality (CEIQ) [5]. We also introduce the metric LOE [30] to show the uniform performance on low-light facial images.

The values of LOE correspond to uniform performance. Our work focuses on restoring faces with consistent lighting, where LOE is crucial among the validation metrics. In Tab. 1, the restoration without α_6 demonstrates

Table 1
The average non-reference metrics **PIQE** [27], **CEIQ** [5] and **LOE** [30] of the ablation study on the dataset NUL-face.

Algorithms	↓ PIQE	↑ CEIQ	↑ LOE
inputs	42.30	3.35	/
w/o α_0	20.95	**3.41**	834.94
w/o α_1	17.93	3.40	900.24
w/o α_2	27.56	3.39	305.23
w/o α_3	28.07	3.35	116.53
w/o α_4	17.48	3.08	888.34
w/o α_5	**8.89**	3.39	949.03
w/o α_6	24.29	3.37	**1451.46**
Ours	21.49	3.38	1356.44

the best performance, while the result incorporating all loss functions ranks second. This indicates that the GAN's discriminator, which distinguishes between generated images and the input L_0, can undermine uniform performance. However, from Fig. 6(h) it can be seen that the result without α_6 is overexposed. Consequently, the GAN's discriminator as the constraint ensures a balance between restoring low-light images and preventing overexposure in the final output. In summary, our proposed algorithm achieves the best overall results when all loss functions are included.

The non-reference metric PIQE is used to validate the image quality of these generations. The lower the value of PIQE, the better the result. From Tab. 1, the metric without α_5 has the best performance. PIQE, as the metric, usually validate to the whole image, which may ignore the details. From Fig. 6(g), the result without α_5 from the heavy streaks in the red box. Though its metric PIQE is better, it does not mean that the constraint of $loss_6$ make no sense.

The non-reference metric CEIQ also reflects the quality of these generated images. A higher CEIQ value indicates better results. CEIQ focuses on the contrast of the image, based on the premise that a high-contrast image is often more similar to its contrast-enhanced counterpart. The result obtained without α_0 is the best, possibly because the non-uniformity improves CEIQ. The values obtained through different loss functions are very close, and our method ranks among the top results.

In terms of these metrics, our result is not the best. However, this may be due to the instability of the non-reference. And the subjective results show our superiority and our comprehensive performance of the metrics is optimal.

4.3. Comparison with methods

The related non-uniform low-light enhancement algorithms include the traditional algorithms NPE [30], SRIE [6], and the framework of deep learning algorithms, Zero-DCE [9], EnlightenGAN [10], SAGC [3], LDLES-Net [11] and NeRCo [34]. All the deep learning algorithms are retrained and fine-tuned on the NUL-Face and MIT-Adobe datasets.

The proposed method has been evaluated through objective and subjective experiments. In Fig. 7, visual comparison results of NPE, SRIE, Zero-DCE, EnlightenGAN, SAGC, LDLES-Net, NeRCo, and our proposed method are

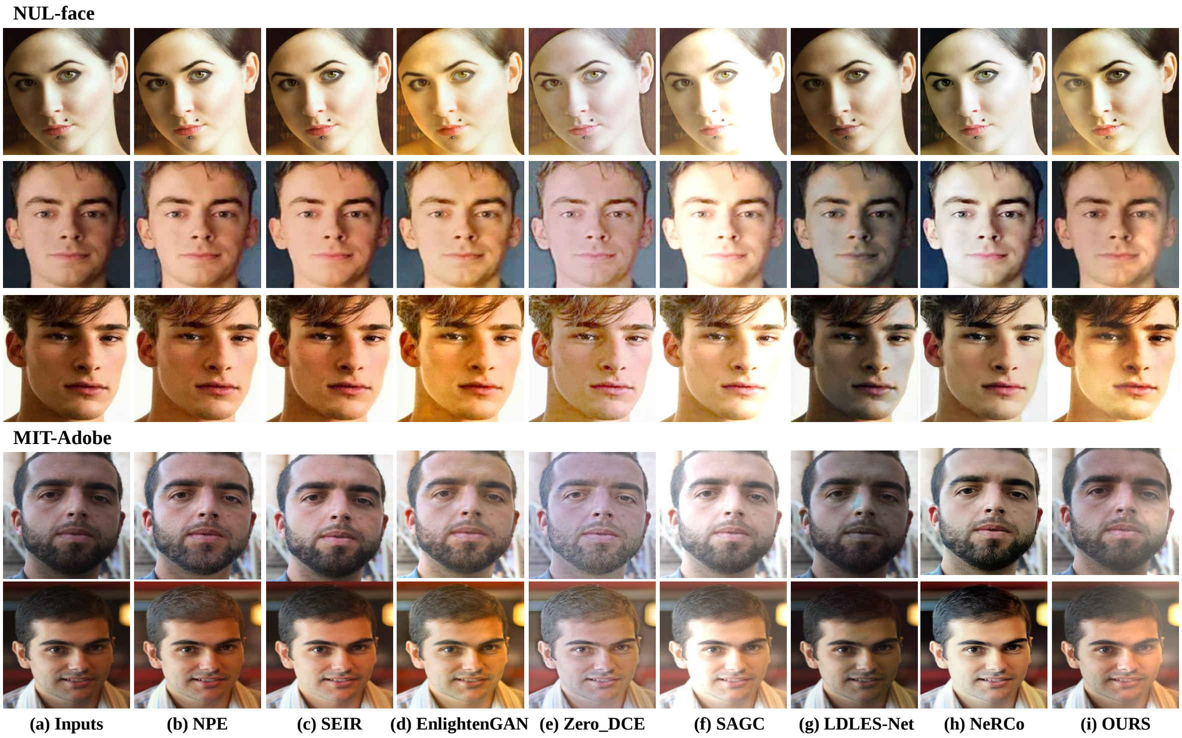


Figure 7: The results of real non-uniform low-light face images from the NUL-Face and MIT-Adobe datasets generated by the comparison methods NPE [30], SRIE [6], Zero-DCE [9], EnlightenGAN [10], SAGC [3], LDLES-Net [11], NeRCo [34] and ours.

showcased. It can be observed that NPE and SRIE, the traditional algorithms, exhibit subpar performance. Although facial images are recovered, the details in dark regions still suffer from low light. LDLES-Net generates results that also suffer from low illumination, even worse than the input images. Unsupervised deep learning methods, Zero-DCE and EnlightenGAN, produce better reconstruction results but have limitations. Both Zero-DCE and EnlightenGAN outputs suffer from incorrect colors. SAGC can enhance low-light images, but the results are overexposed and affect the details. NeRCo shows much better performance with good facial details. However, there is an obvious color difference between the generated result and the input image, with overexposure or color distortion of the originally well-lit areas. In contrast, our method surpasses these approaches in terms of details and illumination. The colors in our reconstructed images closely match the intensity of the high-illumination region. By incorporating brightness analysis, our model can be effectively applied to most real non-illuminated face images.

The quantitative results are shown in Tab. 2, where we have used non-reference metrics such as BRISQUE [19], PIQE, CEIQ, and LOE [30] to evaluate the image quality of our proposed method and the baselines. The best results are highlighted in bold. LOE scores are used to assess the uniformity of image illumination, and is crucial among the validation metrics. As this metric requires a comparison to the original inputs, it cannot be computed for the input

Table 2
The average non-reference metrics **BRISQUE** [19], **PIQE**, **CEIQ** and **LOE** of our results and the comparison methods NPE [30], SRIE [6], EnlightenGAN [10], Zero-DCE [9] on the real test faces of the datasets NUL-face and MIT-Adobe.

Datasets	NUL-face				MIT-Adobe			
Metrics / Algorithms	↓ BRISQUE	↓ PIQE	↑ CEIQ	↑ LOE	↓ BRISQUE	↓ PIQE	↑ CEIQ	↑ LOE
Inputs	190.47	42.30	3.35	/	172.88	14.40	3.12	/
NPE [30]	178.31	43.49	3.38	457.42	166.98	8.26	3.26	730.85
SIRE [6]	37.09	50.66	3.40	222.15	151.40	14.08	3.32	204.07
EnlightenGAN [10]	36.24	44.12	3.33	955.90	152.25	21.27	3.34	999.75
Zero-DCE [9]	33.30	38.59	3.16	455.41	151.40	**7.58**	3.16	597.34
SAGC [3]	136.67	40.63	3.05	559.02	177.73	9.93	3.13	303.18
LDLES-Net [11]	146.19	40.30	3.17	614.91	150.24	16.78	3.09	388.35
NeRCo [34]	146.78	36.80	**3.44**	**4399.11**	179.23	26.58	**3.53**	3501.86
Ours	**28.26**	**21.49**	3.38	1356.44	**149.72**	12.18	3.33	**3694.41**

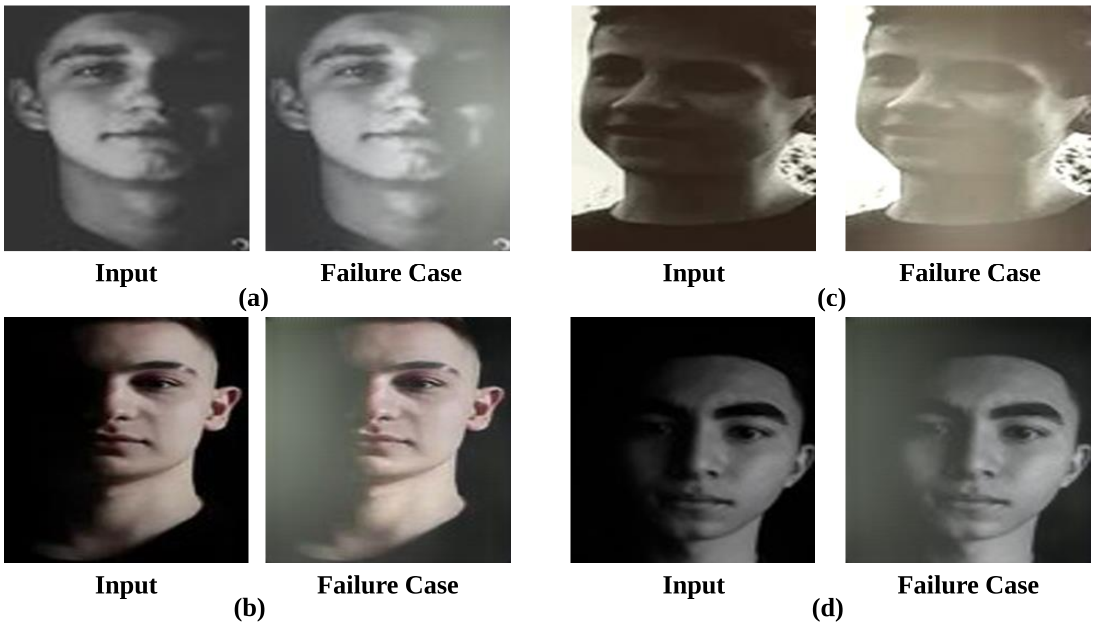


Figure 8: Our failure cases of the low-quality images.

images. The LOE scores indicate that our results have better performance in terms of uniform image illumination. Although NeRCo outperforms our method in the NUL-face dataset, this is primarily due to the overexposure in their images, which benefits their LOE scores. In contrast, our approach strikes a balance between effectively restoring low-light images and preventing overexposure in the final output.

BRISQUE and PIQE metrics are utilized to evaluate image quality, and our proposed method outperforms existing approaches, as evidenced by the higher scores we achieved. This demonstrates the effectiveness of our solution. This highlights the effectiveness of our solution. CEIQ measures enhancement performance based on natural scene statistics. In this context, the results from EnlightenGAN and Zero-DCE, which are generated using deep learning techniques, yield low CEIQ values. NeRCo achieves the best CEIQ performance across both datasets and also scores highest on the

LOE metric for the NUL-face dataset; however, its performance in other metrics is inferior to ours. While CEIQ tends to favor images with high contrast, NeRCo's tendency for overexposure may contribute positively to its CEIQ scores. The values obtained through different loss functions are quite similar, and our method consistently ranks among the top results. Overall, our approach achieves high scores across nearly all metrics, establishing it as the leading performer on these two datasets.

4.4. Details and Failure Cases

During training, there are some details of the images that need to be taken into account. Since our network is unsupervised, the convergence of the training process is very sensitive. Then it has high demands on the learning rate. When the learning rate is set as $1e-4$ (in many deep learning methods, this learning rate tends to good performance), the training of the network is not good with, and the generations are extremely poor.

Furthermore, there are some failure cases in our work as shown in Fig. 8. The low-quality faces suffer from the extremely poor situations. In Fig. 8 (a) and (b), the low exposure leads to a complete missing of details in some regions, making it impossible to improve the quality regardless of any adjustments made to the facial state. In Fig. 8(c), the uneven facial features and the high contrast between the foreground and background complicate effective enhancement. In Fig. 8(d), the complex lighting results in varying degrees of under-exposure; while some areas, though dark, can still convey information through enhancement, other regions are so dark that recovery is impossible. This ultimately hinders the full enhancement of the face. In such complex situations, our unsupervised network is unable to sufficiently improve these images.

5. Conclusion

Our work focuses on overcoming the poor quality of faces in non-uniform, low-light face images. Complex illumination makes it difficult to identify image details, and existing methods face challenges in effectively enhancing face images. We propose an unsupervised non-uniform low-light face enhancement algorithm that trains the network with low-quality face images without corresponding high-quality images, and then propose an unsupervised non-uniform low-light face enhancement algorithm to achieve iterative updating and detail enhancement of face information based on brightness analysis. We introduce illumination decoupling to the network to constrain the direction of network convergence. The experiments on the real low-quality face images show that our approach outperforms the state-of-the-art methods. Of course, our work may encounter failed enhancement results in some extreme scenarios, such as extreme shadows leading to the missing of facial information or complex lighting causing enhancement to fail. In the future, we will introduce generative models or more sophisticated enhancement models to address these failure scenarios.

References

[1] Blanz, V., Vetter, T., 1999. A morphable model for the synthesis of 3d faces, in: Processing on the International Conference on Computer Graphics and Interactive Techniques, pp. 187–194.

[2] Bychkovsky, V., Paris, S., Chan, E., Durand, F., 2011. Learning photographic global tonal adjustment with a database of input / output image pairs, in: The Twenty-Fourth IEEE Conference on Computer Vision and Pattern Recognition.

[3] Chen, Y.H., Pan, F.C., Liao, Y.C., Kao, J.H., Wang, Y.C.F., 2023. Semantics-aware gamma correction for unsupervised low-light image enhancement, in: the IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 1–5.

[4] Cun, X., Pun, C.M., Shi, C., 2020. Towards ghost-free shadow removal via dual hierarchical aggregation network and shadow matting gan, in: Processing of the AAAI Conference on Artificial Intelligence, IEEE. pp. 10680–10687.

[5] Fang, Y., Ma, K., Wang, Z., Lin, W., Fang, Z., Zhai, G., 2014. No-reference quality assessment of contrast-distorted images based on natural scene statistics. the IEEE Signal Processing Letters 22, 838–842.

[6] Fu, X., Liao, Y., Zeng, D., Huang, Y., Zhang, X.P., Ding, X., 2015. A probabilistic method for image enhancement with simultaneous illumination and reflectance estimation. the IEEE Transactions on Image Processing 24, 4965–4977.

[7] Fu, Y., Hong, Y., Chen, L., You, S., 2022. Le-gan: Unsupervised low-light image enhancement network using attention module and identity invariant loss. Knowledge-Based Systems 240, 108010.

[8] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., Bengio, Y., 2014. Generative adversarial nets. Processing of the Advances in Neural Information Processing Systems 27.

[9] Guo, C., Li, C., Guo, J., Loy, C.C., Hou, J., Kwong, S., Cong, R., 2020. Zero-reference deep curve estimation for low-light image enhancement, in: Processing of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 1780–1789.

[10] Jiang, Y., Gong, X., Liu, D., Cheng, Y., Fang, C., Shen, X., Yang, J., Zhou, P., Wang, Z., 2021. Enlightengan: Deep light enhancement without paired supervision. the IEEE Transactions on Image Processing 30, 2340–2349.

[11] Jin, Y., Yang, W., Tan, R.T., 2022. Unsupervised night image enhancement: When layer decomposition meets light-effects suppression, in: European Conference on Computer Vision, Springer. pp. 404–421.

[12] Jobson, D.J., Rahman, Z., Woodell, G.A., 1997. Properties and performance of a center/surround retinex. the IEEE Transactions on Image Processing 6, 451–462.

[13] Kajiya, J.T., 1986. The rendering equation, in: Proceedings of the Annual Conference on Computer Graphics and Interactive Techniques, pp. 143–150.

[14] Kao, W.C., Hsu, M.C., Yang, Y.Y., 2010. Local contrast enhancement and adaptive feature extraction for illumination-invariant face recognition. Pattern Recognition 43, 1736–1747.

[15] Kingma, D.P., Ba, J., 2014. Adam: A method for stochastic optimization. arXiv: Learning .

[16] Lin, Y.B., Tseng, H.Y., Lee, H.Y., Lin, Y.Y., Yang, M.H., 2023. Unsupervised sound localization via iterative contrastive learning. Computer Vision and Image Understanding 227, 103602.

[17] Liu, X., Ma, W., Ma, X., Wang, J., 2023. Lae-net: A locally-adaptive embedding network for low-light image enhancement. Pattern Recognition 133, 109039.

[18] Lv, F., Liu, B., Lu, F., 2020. Fast enhancement for non-uniform illumination images using light-weight cnns, in: Proceedings of the 28th ACM International Conference on Multimedia (ACM MM), pp. 1450–1458.

[19] Mittal, A., Moorthy, A.K., Bovik, A.C., 2012. No-reference image quality assessment in the spatial domain. the IEEE Transactions on Image Processing 21, 4695–4708.

[20] Mustapha, A., Oulefki, A., Bengherabi, M., Boutellaa, E., Algaet, M.A., 2017. Towards nonuniform illumination face enhancement via adaptive contrast stretching. Multimedia Tools and Applications 76, 21961–21999.

[21] Nestmeyer, T., Lalonde, J.F., Matthews, I., Lehrmann, A., 2020. Learning physics-guided face relighting under directional light, in: Processing of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5124–5133.

[22] Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation, in: Processing of the International Conference on Medical Image Computing and Computer-Assisted Intervention.

[23] Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 .

[24] Smith, B.M., Zhang, L., Brandt, J., Lin, Z., Yang, J., 2013. Exemplar-based face parsing, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

[25] Sun, H., Luo, Z., Ren, D., Du, B., Chang, L., Wan, J., 2024. Unsupervised multi-branch network with high-frequency enhancement for image dehazing. Pattern Recognition 156.

[26] Ueda, Y., Moriyama, D., Koga, T., Suetake, N., 2020. Histogram specification-based image enhancement for backlit image, in: Processing of the IEEE International Conference on Image Processing, pp. 958–962.

[27] Venkatanath, N., Praneeth, D., Bh, M.C., Channappayya, S.S., Medasani, S.S., 2015. Blind image quality evaluation using perception based features, in: the National Conference on Communications, IEEE. pp. 1–6.

[28] Wang, L.W., Siu, W.C., Liu, Z.S., Li, C.T., Lun, D.P., 2020. Deep relighting networks for image light source manipulation, in: Proceedings of the European Conference on Computer Vision, Springer. pp. 550–567.

[29] Wang, R., Zhang, Q., Fu, C.W., Shen, X., Zheng, W.S., Jia, J., 2019a. Underexposed photo enhancement using deep illumination estimation, in: Processing of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 6849–6857.

[30] Wang, S., Zheng, J., Hu, H.M., Li, B., 2013. Naturalness preserved enhancement algorithm for non-uniform illumination images. IEEE Transactions on Image Processing 22, 3538–3548.

[31] Wang, Y., Cao, Y., Zha, Z.J., Zhang, J., Xiong, Z., Zhang, W., Wu, F., 2019b. Progressive retinex: Mutually reinforced illumination-noise perception network for low-light image enhancement, in: Processing of the ACM Multimedia, pp. 2015–2023.

[32] Wang, Y., He, K., Zhou, T., Yao, K., Li, N., Xu, L., Yu, J., 2023. Free-view face relighting using a hybrid parametric neural model on a small-olat dataset. International Journal of Computer Vision , 1–20.

[33] Wu, Y., Song, W., Zheng, J., Liu, F., 2021. Non-uniform low-light image enhancement via non-local similarity decomposition model. Signal Processing: Image Communication 93, 116141.

[34] Yang, S., Ding, M., Wu, Y., Li, Z., Zhang, J., 2023a. Implicit neural representation for cooperative low-light image enhancement, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 12918–12927.

[35] Yang, S., Zhou, D., Cao, J., Guo, Y., 2023b. Lightingnet: An integrated learning method for low-light image enhancement. the IEEE Transactions on Computational Imaging 9, 29–42.

[36] Yang, W., Wang, S., Fang, Y., Wang, Y., Liu, J., 2020. From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 3060–3069.

[37] Yu, C., Wang, J., Peng, C., Gao, C., Yu, G., Sang, N., 2018. Bisenet: Bilateral segmentation network for real-time semantic segmentation, in: Proceedings of the European conference on computer vision, pp. 325–341.

[38] Zhang, X., Barron, J.T., Tsai, Y.T., Pandey, R., Zhang, X., Ng, R., Jacobs, D.E., 2020. Portrait shadow manipulation. ACM Transactions on Graphics 39, 78–1.

[39] Zhou, F., Sun, X., Dong, J., Zhu, X.X., 2023. Surroundnet: Towards effective low-light image enhancement. Pattern Recognition 141, 109602.

[40] Zhou, H., Hadap, S., Sunkavalli, K., Jacobs, D.W., 2019. Deep single-image portrait relighting, in: Processing of the IEEE International Conference on Computer Vision, pp. 7194–7202.

Title Pages:

Unsupervised Learning Non-uniform Face Enhancement Under Physics-guided Model of Illumination Decoupling

Authors:
Xin Ding[a], Jing Fang[b], Zhongyuan Wang[c], Qiong Liu[a*], You Yang[a] and Zhenyu Shu[d]

Affiliations
a School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan, 430074, China
b School of Computer Science, Northwestern Polytechnical University, Xian, 710127, China
c School of Computer Science, Wuhan University, Wuhan, 430072, China
d NingboTech University, Ningbo, 315100, China

Corresponding author: Qiong Liu
Emails: q.liu@hust.edu.cn

Highlights:

Real-world non-uniform low-light face enhancement

Physics-guided Model based on the brightness state of low-quality images for illumination decoupling

The unsupervised network following the theoretical analysis of image decoupling

The self-built dataset of the real-world non-uniform low-light faces

The excellent performance of our method on the real-world images compared with the recent methods

Author Lists:

Xin Ding received the B.E. degree from Electronic Information School of Huazhong University of science and technology in 2012, and M.S., and Ph.D. degrees from the school of computer science, Wuhan University respectively in 2016 and in 2022. He is now working as postdoc. at Huazhong University of Science and Technology. His research interests focus on computer vision and Image Processing.

Fang Jing received her B.S degree in communication engineering from Jianghan University and M.S degree in electronic information engineering from Wuhan University. She received Ph.D. degree in National Engineering Research Center for Multimedia Software, School of Computer, Wuhan University, Wuhan, China. He is now working as postdoc. at Northwestern Polytechnical University. Her research interests include image super resolution and image reconstruction.

Zhongyuan Wang received the Ph.D. degree in communication and information system from Wuhan University, Wuhan, China, in 2008. Dr. Wang is now an associate professor with School of Computer, Wuhan University, Wuhan, China. He is currently directing three projects funded by the National Natural Science Foundation Program of China. His research interests include video compression, image processing, and multimedia communications, etc.

Qiong Liu received the Ph.D. degree in computer science from the School of Computer Science, Wuhan University, Wuhan, China, in 2008. She has worked as a Post-Doctoral Fellow with the Department of Automation, Tsinghua University, from 2010 to 2012. Since 2009, she has been with the School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan, and the Division of Intelligent Media and Fiber Communications, Wuhan National Laboratory for Opto-Electronics, Wuhan, where she is currently a Full Professor with the Department of Information Engineering. She has authored or coauthored more than 50 technical articles and authorized more than 30 patents. Her research interests include cross-discipline researches between artificial intelligence and 3D computer vision, including 3D visual communication, human–robot interaction, human intention, and interactive visual applications.

You Yang received the Ph.D. degree in computer science from the Institute of Computing Technology, Chinese Academy of Sciences, Beijing, China, in 2009. From 2009 to 2011, he was a Post-Doctoral Fellow with the Department of Automation, Tsinghua University. From 2011 to 2013, he was the Chief Scientist with Sumavision Research. Since 2013, he has been with the School of Electronic Information and Communications, Huazhong University of Science and Technology, Wuhan, China, and the Division of Intelligent Media and Fiber Communications, Wuhan National Laboratory for Optoelectronics, Wuhan, where he is currently a Full Professor and the Vice Dean. He has authored or coauthored over 110 peer-reviewed articles and authorized 41 domestic and U.S. patents. His research interests include 3D visual communications and its applications, including multiview imaging systems, 3D/VR/AR content processing and visual communications, human–machine interaction techniques, and interactive visual applications.

Zhenyu Shu got his Ph.D. degree in 2010 at the Zhejiang University, China. He is now working as a full professor at NingboTech University. His research interests include computer graphics, digital geometry processing and machine learning. He has published over 30 papers in international conferences or journals.

Declaration of Interest Statement:

We declare that we have no financial and personal relationships with other people or organizations that can inappropriately influence our work, there is no professional or other personal interest of any nature or kind in any product, service and/or company that could be construed as influencing the position presented in, or the review of the manuscript.