electronics

Article

DeFFace: Deep Face Recognition Unlocked by Illumination Attributes

Xiangling Zhou, Zhongmin Gao, Huanji Gong and Shenglin Li *

School of Software Engineering, Chongqing University of Posts and Telecommunications,
Chongqing 400065, China; s221201040@stu.cqupt.edu.cn (X.Z.); s221201008@stu.cqupt.edu.cn (Z.G.);
s231231017@stu.cqupt.edu.cn (H.G.)
* Correspondence: lisl@cqupt.edu.cn

Abstract: General face recognition is currently one of the key technologies in the field of computer vision, and it has achieved tremendous success with the support of deep-learning technology. General face recognition models currently exhibit extremely high accuracy on some high-quality face datasets. However, their performance decreases in challenging environments, such as low-light scenes. To enhance the performance of face recognition models in low-light scenarios, we propose a face recognition approach based on feature decoupling and fusion (DeFFace). Our main idea is to extract facial-related features from images that are not influenced by illumination. First, we introduce a feature decoupling network (D-Net) to decouple the image into facial-related features and illumination-related features. By incorporating the illumination triplet loss optimized with unpaired identity IDs, we regulate illumination-related features to minimize the impact of lighting conditions on the face recognition system. However, the decoupled features are relatively coarse. Therefore, we introduce a feature fusion network (F-Net) to further extract the residual facial-related features from the illumination-related features and fuse them with the initial facial-related features. Finally, we introduce a lighting-facial correlation loss to reduce the correlation between the two decoupled features in the specific space. We demonstrate the effectiveness of our method on four real-world low-light datasets and three simulated low-light datasets. We retrain multiple general face recognition methods using our proposed low-light training sets to further validate the advanced performance of our method. Compared to general face recognition methods, our approach achieves an average improvement of more than 2.11 percentage points on low-light face datasets. In comparison with image enhancement-based solutions, our method shows an average improvement of around 16 percentage points on low-light datasets, and it also delivers an average improvement of approximately 5.67 percentage points when compared to illumination normalization-based methods.

Keywords: general face recognition; low-light face recognition; feature decoupling; feature fusion

check for updates

Citation: Zhou, X.; Gao, Z.; Gong, H.; Li, S. DeFFace: Deep Face Recognition Unlocked by Illumination Attributes. *Electronics* **2024**, *13*, 4566. https://doi.org/10.3390/electronics13224566

Academic Editor: Manohar Das

Received: 15 September 2024
Revised: 29 October 2024
Accepted: 14 November 2024
Published: 20 November 2024

1. Introduction

With the rise of deep-learning technology, deep learning has been widely applied in the field of face recognition, leading to significant breakthroughs in face recognition technology. The accuracy of existing general face recognition models [1–5] on high-quality face datasets (e.g., LFW [6], CFP-FP [7], AgeDB-30 [8], CALFW [9], and CPLFW [10]) is mostly above 95%, and even as high as 99%, indicating that face recognition technology has become quite mature under ideal conditions. These models are designed to extract robust facial features from constrained facial images, thereby enhancing the performance of face recognition models. However, they have not fully addressed the challenges in some constrained scenarios. Their performance in these constrained scenarios (e.g., varying lighting, complex backgrounds, and dynamic environments), particularly in low-light conditions, is still less than ideal. The general classic face recognition method, such as ArcFace [5] fails to achieve correct matching in low-light facial images, leading to a significant decrease in performance.

Therefore, general face recognition models exhibit poor feature extraction capabilities in constrained environments.

Currently, many efforts [11–16] are focused on developing methods to address face recognition in low-light scenarios, continuously improving the performance of face recognition systems in such conditions. However, many of these methods lack generalizability and are particularly dependent on the training dataset. Based on the above, we aim to develop a face recognition method that extracts robust facial features in low-light environments, thereby improving the feature extraction performance of face recognition systems under low-light conditions. Existing solutions to improve the accuracy of low-light facial image recognition mainly include three approaches: (1) Methods based on image enhancement techniques [11,14,17,18]. Image enhancement techniques (e.g., histogram equalization, adaptive histogram equalization, gamma correction, or deep-learning-based methods) are first used to enhance low-light facial images, and this process is considered a data preprocessing step. The enhanced images are then transmitted to the face recognition system for subsequent recognition, as shown in Figure 1a. However, these methods may not only result in the over-brightening of certain regions of the image and amplification of noise but also rely on multiple stages, making the process cumbersome, and the enhanced image may even reduce the recognition performance of the face recognition system. (2) Methods based on illumination normalization [12,16,19,20], such as using the Retinex theory [21] to normalize illumination variations by decomposing facial images into illumination and reflectance components, which helps maintain the visual quality of the enhanced image under different lighting conditions, as shown in Figure 1b. The Retinex method primarily targets these two components for processing. However, during the identification of these components, there is a risk of losing facial texture. (3) Methods based on NIR devices [22–25]. By capturing facial images using near-infrared (NIR) spectra, effective imaging can be achieved under low-light conditions, as shown in Figure 1c. However, NIR is limited by skin reflectance, has limited anti-interference capabilities, and requires additional equipment costs.

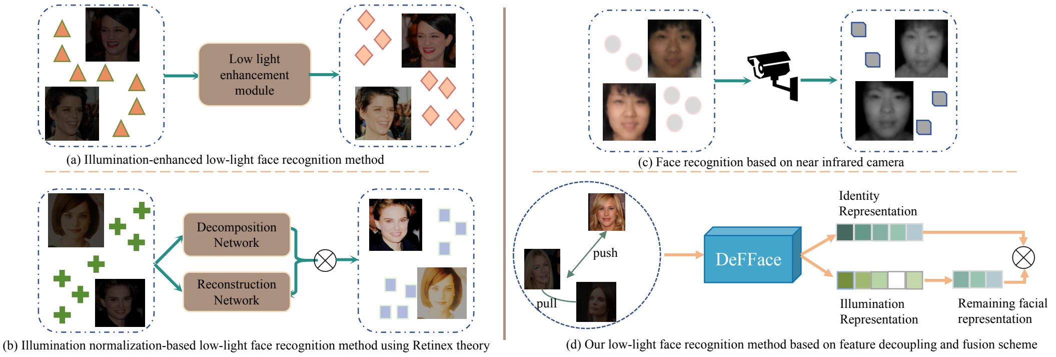
Low light enhancement module
(a) Illumination-enhanced low-light face recognition method
(c) Face recognition based on near infrared camera
Decomposition Network
Reconstruction Network
(b) Illumination normalization-based low-light face recognition method using Retinex theory
push
pull
DeFFace
Identity Representation
Illumination Representation
Remaining facial representation
(d) Our low-light face recognition method based on feature decoupling and fusion scheme

Figure 1. Diagrams of different approaches for low-light face image recognition. (**a**) Illumination-enhanced low-light face recognition method. (**b**) Illumination normalization-based low-light face recognition method using Retinex theory. (**c**) Face recognition based on near-infrared camera. (**d**) Our Low-light face recognition method based on feature decoupling and fusion scheme (DeFFace).

To address these issues, we aim to identify illumination features that are independent of facial features. In Retinex theory [21], an image is decomposed into reflectance and illumination components, allowing the human eye to maintain stable color perception under varying lighting conditions. Retinex enhances images by estimating and removing the illumination component while preserving the reflectance. Inspired by this theory, we propose DeFFace, a low-light face recognition method based on feature decoupling and fusion, as shown in Figure 1d. DeFFace significantly improves face recognition accuracy in low-light environments by decoupling facial-related features from illumination-related features. This technology is highly applicable in areas such as security surveillance and identity verification, providing accurate and reliable recognition even in challenging low-light conditions. The widespread adoption of the DeFFace model will enhance public safety and privacy protection, delivering superior service experiences in intelligent security systems and digital identity management. Specifically, we decompose facial image features into two components: facial-related features and illumination-related features. By applying constrained supervision to the illumination-related features, we effectively mitigate the impact of lighting variations. Simply applying enhancement techniques to low-light facial images can lead to issues such as overexposure; therefore, we address lighting effects from the perspective of feature enhancement.

Our main idea is to extract facial-related features from low-light images that are unaffected by lighting conditions. To achieve this, we introduce a novel feature decoupling module (D-Net) that operates at the feature level. D-Net decouples the image into facial-related features and illumination-related features, addressing the poor performance of face recognition systems in low-light scenarios from a novel perspective. This approach eliminates the need for additional preprocessing steps and the associated risks. In D-Net, we also introduce an illumination triplet loss based on the traditional triplet loss. The illumination triplet loss controls the decoupled illumination-related features, eliminating the influence of lighting factors in the image. Additionally, to further obtain purer facial-related features, we introduce a feature fusion module (F-Net). F-Net extracts the remaining facial-related features from the illumination-related features and fuses them with the facial-related features from D-Net. Finally, we introduce a lighting-facial correlation loss, reducing their correlation in the feature space.

In summary, the contributions of this paper can be summarized as follows:

- We propose DeFFace to mitigate the effects of illumination factors on face recognition systems through feature enhancement to enhance the recognition performance of face recognition systems in low-light environments.
- We introduce a D-Net that decouples the image into facial-related features and illumination-related features. Additionally, we introduce a new illumination triplet loss to control the illumination-related features in the decoupling module using unpaired identity ID training. We also introduce an F-Net to further obtain pure facial-related features and introduce a lighting–facial correlation loss to reduce the correlation between the two features.
- We validate our method against existing low-light face recognition methods and several general face recognition methods on four real-world low-light datasets and three simulated low-light datasets. The experimental results indicate that our method achieves compelling results, demonstrating its effectiveness and superiority.

2. Related Work

In the field of face recognition research, privacy and security issues are increasingly gaining attention. The collection and use of facial data inherently raise concerns about personal privacy, especially when such data can be easily exploited if proper security measures are not in place. Moreover, adversarial attacks on face recognition systems have become a growing concern, where attackers can modify or forge images to manipulate the system into returning incorrect recognition results, thereby threatening both data privacy and system security [26]. Recent studies, such as Baia et al. [27], have shown how emotion adversarial attacks (EAAs) can be used to shield privacy by intentionally obstructing emotion recognition systems. This demonstrates the importance of integrating adversarial defense mechanisms to enhance privacy protections in facial recognition systems. Therefore, when designing face recognition models, it is crucial to consider not only their performance under varying lighting conditions but also their robustness against adversarial attacks and their potential impact on privacy [28]. Although this study focuses on enhancing recognition performance in low-light environments through feature decoupling and fusion techniques, we recommend that future research incorporates adversarial defense strategies to improve model robustness and protect user privacy while ensuring system security.

In this section, we revisit prior work on general face recognition methods and low-light face recognition techniques.

2.1. General Face Recognition

Face recognition, as a crucial research area in computer vision, has witnessed significant advancements. Presently, some deep-learning-based face recognition methods exhibit outstanding performance on various general high-quality face datasets (e.g., LFw [6], CFP-FP [7], and CPLFW [10]) by employing innovative loss functions [2–5,29–31]. However, their performance tends to degrade on low-quality face datasets (e.g., IJB-S [32] and Tiny-Face [33]). Recent research [1,2] has optimized face recognition performance by focusing on image quality, emphasizing the relationship between feature norm and image quality, and making significant progress in enhancing the accuracy of low-quality face images. Although these deep-learning face models have improved face recognition accuracy to some extent in constrained conditions, facial matching under varying lighting conditions still poses challenges.

2.2. Low-Light Face Recognition

Some existing face recognition models exhibit poor generalization ability in low-light scenarios. Many studies aim to improve the accuracy of face recognition under low-light conditions. The main solutions are as follows.

Low-light face recognition based on VIS images can be roughly divided into the following two main categories: (1) Low-light face recognition methods based on image enhancement techniques. Lu and Gan [18] applied Gaussian filtering to handle pixel noise at the bottom to mitigate low-light effects and introduced histogram enhancement (HE) for low-light image enhancement. However, traditional enhancement methods suffer from issues such as over-enhancement and noise amplification, which can even damage face recognition systems. Bendjillali et al. [17] improved the contrast-limited adaptive histogram equalization (CLAHE) algorithm for face image enhancement, which can handle local contrast more delicately compared to traditional HE methods. Fan et al. [11] created three low-light face image datasets and designed an FC task-driven low-light enhancement network (Low-FaceNet) to enhance low-light images before recognition. Le and Kakadiaris [14] proposed the SeLENet model, emphasizing that existing image enhancement methods have not been optimized for face images. The authors designed a decomposition and reconstruction network based on the Retinex theory and used a semi-supervised learning method with spherical harmonic coefficients to enhance and reconstruct lighting conditions. Hu et al. [13] also proposed a low-light face recognition technique called FIN-GAN based on the Retinex theory and GAN. However, the authors introduced too many loss functions during implementation, which increased computational costs and optimization difficulty. Among these enhancement methods, most aim to produce an enhanced version of the low-light image, essentially a normal brightness image. However, determining whether an image has normal or low brightness is highly subjective and difficult to quantify with data. (2) Methods based on feature enhancement. Fang and Li [34] demonstrated the effectiveness of this approach under illumination variations by using a sparse representation classification method and Gabor wavelet transform, adaptively selecting the optimal atoms for each feature to be effectively represented during classification. Huang and Chen [12,16] obtained feature images between the original and enhanced images through a designed feature restoration module, exploring the generation of robust feature images to enhance the performance of deep face recognition models in low-light environments. Wang et al. [20] employed Generative Adversarial Networks (GANs) to address the image pairing problem, enhancing the quality of low-light face images through adversarial training. Miao and Wang [19] proposed a knowledge-guided representation decoupling network and illumination offset loss for low-light face recognition, where the loss helped maintain the consistency of decoupled illumination-related features. (3) Based on NIR-VIS images. Infrared image capture is one of the technical solutions for low-light face recognition, as shown in Figure 1c. There are also many scholars who have conducted research on NIR-based solutions. Lezama et al. [22] proposed a face recognition method based on near-infrared and visible light (NIR-VIS) technology, which processes input data through cross-spectral hallucination and low-rank embedding, facilitating the analysis of features from different spectral images. This approach offers a new solution for face recognition in low-light scenarios. Miyamoto et al. [35] proposed a joint feature distribution alignment learning (JFDAL) method for NIR-VIS and VIS-VIS face recognition. This approach addresses domain discrepancies between NIR and VIS images while maintaining the discriminative power of VIS features. He et al. [25] proposed an adversarial cross-spectral face completion method for NIR-VIS face recognition, using GANs with texture inpainting and pose correction to generate high-resolution VIS images from NIR inputs, improving synthesis accuracy and quality. However, NIR face recognition requires additional hardware (NIR cameras), making it potentially more efficient and cost-effective to work within the VIS image domain. Table 1 summarizes the main methods for low-light face recognition introduced in this section, including general FR approaches, NIR-VIS FR approaches, and VIS image-based approaches. It presents the advantages and limitations of these techniques in low-light face recognition scenarios.

Table 1. Comparison of different low-light face recognition methods.

Option	Author	Method	Database	Limitation	Advantage
General FR	Schroff et al. [36]	FaceNet	LFW/Youtube Faces DB	Low recognition rate	Processing time
	Wang et al. [29]	CosFace	LFW/AgeDB	Illumination	Measure of angle
	Deng et al. [5]	ArcFace	LFW/IJB-B	Illumination	Excellent class spacing
	Huang et al. [3]	CurricularFace	LFW/CFP-FP	Complexities of CNN	Large-scale dataset
	Kim et al. [1]	AdaFace	IJB-S/TinyFace	Computationally complex	Robustness
NIR-VIS FR	Lezama et al. [22]	–	CASIA NIR-VIS 2.0	Quality of generated VIS images	No need for retraining
	He et al. [25]	–	Oulu-CASIA	Reliance on paired data for training	Multi-scale discriminator
	Miyamoto et al. [35]	SFDAL	Oulu-CASIA	Performance trade-off between NIR and VIS	Preserved VIS domain performance
VIS FR	Bendjillali et al. [17]	–	CMU-PIE	Effectiveness dependent on enhancement techniques	Deep-learning architectures
	Fan et al. [11]	Low-Face Net	LFW*	Too many pre-trained models	Improved visual quality
	Huang and Chen [12]	FRNetFuse	SoF	Low recognition rate	Robustness
	Miao and Wang [19]	–	CASIA NIR-VIS 2.0	Robustness	Illumination offset loss
	Hu et al. [13]	FIN-GAN	ExtendedYaleB	Performance on in-the-wild data	Reduced computational complexity

Overall, among the three main approaches discussed above, our method offers several advantages. Compared to the NIR-based approach, we do not require additional near-infrared devices, which reduces cost. Compared to image enhancement techniques, we do not need to generate intermediate enhanced images, avoiding issues like overexposure and noise that may arise from enhanced images, and allowing us to focus directly on improving the performance of low-light face recognition systems. Finally, compared to general face recognition methods, our approach significantly enhances the robustness of generic face recognition models in low-light scenarios. Additionally, existing image enhancement techniques are mostly designed for processing scene-type images, and their performance when handling face-type images has been less than ideal. Chen et al. [37] pointed out that existing enhancement techniques might have resulted in “pseudo artifacts” when enhancing faces. This not only affected visual effects but also interfered with face recognition systems. Additionally, some solutions [11,38–43] that used image enhancement before face recognition suffered from the problem of excessive enhancement, which could have adversely affected the performance of face recognition systems. In response, this paper processes low-light facial images in the feature space, decoupling the images into illumination-related features and facial-related features. By processing the decoupled features, it avoids the issues caused by enhancing images in the pixel domain. Additionally, an end-to-end approach is employed to simplify the training steps and reduce computational resources.

3. Methodology

In this section, we provide a detailed description of the construction of our network, as schematically illustrated in Figure 2, and its details. Sections 3.1 and 3.2 elaborate on the main content of our decoupling and fusion modules. Section 3.3 introduces the algorithmic process of mining low-light image triplets and Section 3.4 covers the loss terms.

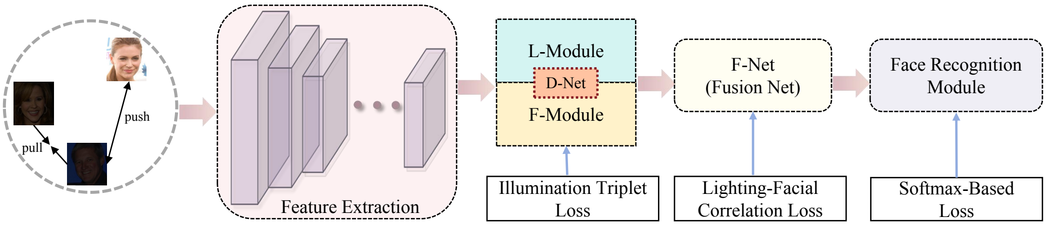

Figure 2. The DeFFace architecture diagram primarily consists of four parts: backbone, D-Net, F-Net, and face recognition. The D-Net is constrained by the illumination triplet loss, the F-Net is constrained by the lighting–facial correlation loss, and the face recognition module is constrained by the Softmax-based loss.

3.1. Feature Decoupling Module

Image features contain rich information, and in low-light facial images, these features include both facial and illumination characteristics. In Section 2.2, we have reviewed the relevant work on feature processing. Fan et al. [11] simply computed the entire image features, crudely comparing the features of low-light images with those of normal-light images. They treated the features of normal and low-light images as a whole for learning, neglecting the complexity of image features and focusing solely on learning the overall features. Miao and Wang [19] used two feature encoders to extract facial and illumination features from low-light images and then used a generator to produce images that transition from one lighting condition to another. The authors considered different categories of features within the images. However, the model relies on the effectiveness of the generator, and if the generated images are not realistic enough, it may actually harm the performance of the face recognition system. In response, this paper introduces a feature decoupling module (D-Net) to effectively learn facial features and separate them from illumination-related features, thereby eliminating the influence of illumination on facial features. Related work [44–48] has shown that feature decomposition is effective. D-Net avoids the adverse effects brought by image enhancement techniques while also eliminating the influence of lighting factors. This approach enhances the accuracy of low-light face recognition from a novel perspective, rather than relying on traditional simple enhancement methods. The following is the definition and description of D-Net.

D-Net. The core objective of the D-Net model is to decouple the input face image features into illumination-related features and facial-related features without relying on illumination labels, ensuring that these two types of features can be independently extracted and learned. Specifically, the input dataset, data, includes face images captured under low-light and normal-light conditions, with each image labeled with its identity ID. The role of the decoupling module in D-Net is to separate the illumination information from the identity features through a specific network structure, enabling the model to maintain a high recognition rate despite significant lighting variations while also accurately modeling illumination characteristics, as shown in Figure 3a. During the training phase, the decoupling module learns from these identity-labeled images. D-Net consists of two parallel sub-modules: the L-module (lighting feature decoupling module) and the F-module (facial feature decoupling module). These two sub-modules focus on extracting different features—while the L-module is responsible for capturing illumination-related information, the F-module focuses on extracting features related to face identity. When a face image passes through the backbone network, these features are simultaneously fed into both the L-module and the F-module. In this process, the 512-dimensional input features are passed through the consecutive convolutional layers of both sub-modules, gradually extracting more abstract feature information. The L-module primarily learns illumination features, such as changes in brightness and lighting direction, while the F-module focuses on capturing face identity details, like facial contours and textures. To enhance the discriminative ability of the features, each sub-module's feature extraction layers apply the ReLU activation function. This nonlinear activation function introduces nonlinearity into the model, enhancing its ability to learn complex features. Particularly in the context of decoupling illumination and facial features, ReLU increases the network's sensitivity to subtle differences. By the end of this process, D-Net successfully separates the illumination information and the identity information from the input image, forming two independent feature representations. Through this dual-module decoupling structure, D-Net not only effectively handles variations in lighting conditions but also ensures that facial features remain consistent across different illumination environments. This feature decoupling enables D-Net to achieve efficient and accurate face recognition under varying lighting conditions. This process is represented by Equation (1).

$$\forall I, F_i = D\text{-}Net(I), \tag{1}$$

where I is the input facial image, and F is the decoupled 512-dimensional feature. When $i = 1$, F is the decoupled 512-dimensional facial-related feature, and when $i = 2$, F is the decoupled illumination-related feature. D-Net represents the decoupling module.

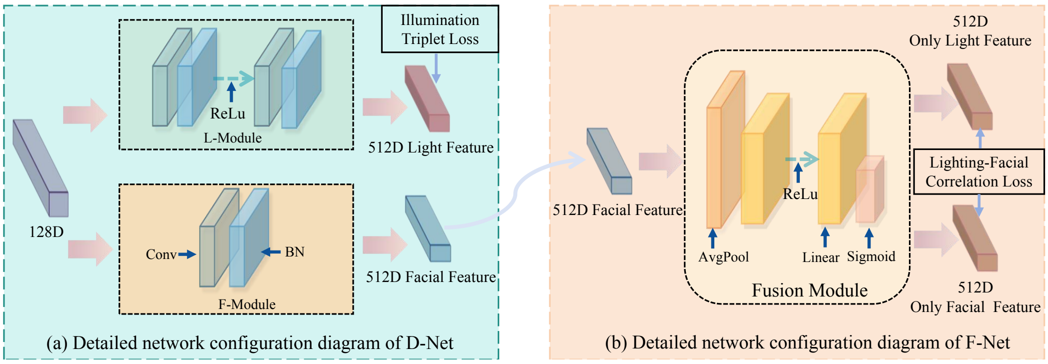

Figure 3. Diagram showing the detailed network configuration of the D-Net and F-Net. Subfigure (**a**) presents the detailed configuration of D-Net, and subfigure (**b**) presents the detailed configuration of the F-Net.

3.2. Feature Fusion Module

In Section 3.1, we focus on the decoupling of two essential features: the illuminance-related feature and the facial-related feature. It is important to note that the illuminance-related feature often retains some residual facial characteristics, which can lead to a loss of important identity information during processing. To mitigate this issue and preserve the integrity of facial features, we have developed a novel feature fusion module, referred to as F-Net, as illustrated in Figure 3b. The primary objective of F-Net is to effectively integrate facial features with illuminance features by applying channel-level weighting to the input illumination feature map. This process begins with a subnetwork that consists of global average pooling followed by fully connected hierarchical sequences. Through this architecture, F-Net learns the global statistical information of each channel within the input feature map. As a result, it generates a specific weight for each channel, which signifies the relevance of that channel to the extraction of facial features. Once the channel weights are determined, they are applied to the input illumination feature map. This multiplication process segregates the features that are associated with the face, emphasizing the channels that contribute most significantly to facial recognition. The learned channel weights enable the model to prioritize features that are crucial for accurately identifying faces while diminishing the influence of less relevant features. The next step involves recalibrating both the facial features and the illumination features by performing element-wise multiplication with the original feature map. This recalibration process results in two distinct sets of feature representations—one for facial features and one for illumination features—both of which possess enhanced anti-jamming capabilities. This means that even in challenging lighting conditions, the model can maintain its ability to recognize faces effectively, as depicted in Equation (2).

$$Y = X \otimes \sigma(W_2 \ \mathrm{ReLu}(W_1 \ \mathrm{AvgPool}(X))), \tag{2}$$

where X denotes the input illumination feature map, and Y denotes the residual facial features in the illumination feature map, while W_1 and W_2 represent two linear transformations. W_1 is employed to decrease the number of channels, whereas W_2 is used to recuperate the number of channels.

3.3. Mining of Low-Illumination Facial Triplet Images

To enable the model to better learn illumination characteristics, this paper proposes an illumination face image triplet based on illumination characteristics, using an unpaired identity ID training method inspired by the traditional triplet construction concept.

3.3.1. Synthesize Data

In Section 3.1, we outline a given input face image set, data = {low, normal}, in which every image possesses an identity ID within the two segments of the set. For the normal subset, acquiring the images inside is relatively straightforward; however, for the low subset, obtaining facial images under differing lighting conditions with the same ID presents a significant challenge. Hence, this paper adheres to the low-light image simulation technique introduced in [49], utilizing a combination of gamma and linear transformations to mimic low-light images. The detailed expression of the equation is provided in Equation (3).

$$I_m^n = \beta * (\alpha * I_{m'}^n)^\gamma, n \in \{R, G, B\}, \tag{3}$$

where α and β denote linear transformations. $f(x)^\gamma$ represents a gamma transformation. The range for the three parameters is as follows: $\alpha \subset [0.9, 1]$, $\beta \subset [0.5, 1]$, and $\gamma \subset [1.5, 3.5]$. Hence, for each identity ID, we possess images under normal illumination as well as facial images under non-paired random illumination variations.

3.3.2. Mining of Triplet Facial Images Under Low Illumination

We aim to bring same-class low-light face samples closer together in the feature space while increasing the distance between different-class low-light face samples and normal-light face samples. To achieve this, we reorganize the initial training data by dividing them into two pools: a normal-light face image pool and a low-light face image pool. We define the following: for any low-light sample (anchor), we randomly select a different low-light sample from the low-light image pool as the positive sample (positive), and randomly select a sample from the normal-light pool as the negative sample (negative). The triplet can be expressed by Equation (4).

$$T_i = \{Anchors_m, Positives_n, Negative_k\}, \tag{4}$$

where i represents the ith triplet, m and n denote any images in the low-light pool, and $m \neq n$. k represents any image in the normal-light pool. Refer to Algorithm 1.

Algorithm 1 Random triplet image generation process

```
Input: low_image_pool, normal_image_pool, input_size(height, width)
Output: Random triple $T_i$
# Create two pools of images: one in low-light and one in normal light
used_as_anchor = set(), used_as_negative = set()
available_anchors = low_image_pool() # Filter available anchor
if not available_anchors then
    used_as_anchor.clear()
    available_anchors = low_image_pool - used_as_anchor
end if
available_positives = low_image_pool() # Filter available positives
if not available_positives then
    used_as_anchor.clear()
    available_positives = (low_image_pool - used_as_positive) - {anchor_path}
    used_as_positive) - {anchor_path}
end if
negative_path = random.choice(list(normal_image_pool)) # Filter available negatives
return {triple table}
```

3.4. Loss Function

To effectively reduce the impact of illumination in facial recognition systems and improve the model's recognition performance in low-light conditions, we design a loss mechanism to control the decoupled illumination-related features and facial-related features.

3.4.1. Illumination Triplet Loss

We aim to implement a loss function that controls lighting features in the decoupling module. As mentioned before, we use the samples in the low image pool as the anchor and positive, and those in the normal image pool as the negative, to create a lighting triplet. After going through the decoupling module, the lighting triplet will yield lighting-related and facial-related features, where the decoupled lighting feature is supervised by the illumination triplet loss. We calculate the Euclidean distance between the anchor and the positive and negative samples and set a boundary η value of 0.2 to ensure there is a sufficient gap between the samples. Our goal is to make the decoupled anchor's lighting feature more similar to the lighting feature of the low-light positive sample, and further from the lighting feature of the opposite normal-light negative sample. This way, we can eliminate the influence of lighting. The formula is expressed by Equation (5).

$$\mathcal{L}_{light} = max(0, \parallel A_m - P_k \parallel^2 - \parallel A_m - N_n \parallel + \eta^2), \tag{5}$$

where A represents the anchor, and A_m denotes any low-illumination facial sample in the low-light pool. P is the positive sample, and P_k denotes any low-illumination facial sample in the low-light pool, with m not equal to k. N is the negative sample, and N_n denotes any normal brightness sample in the normal-light pool.

3.4.2. Lighting–Facial Correlation Loss

For the decoupled illumination features and facial features, we consider that there might be some linkage in the feature space. Thus, we introduce a new loss function, lighting–facial correlation loss, to reduce the correlation between these two features in the feature space. The formulation of the introduced loss function is shown in Equation (6).

$$\mathcal{L}_{lf} = \sum_i (e^{cos_i^2} - 1), \tag{6}$$

where cos_i represents the cosine similarity between the lighting feature and facial feature of the ith sample.

3.4.3. Softmax-Based Loss

In the field of face recognition, most loss functions are designed based on the Softmax function. Since the features learned from the original Softmax loss might lack the ability to distinguish some challenging face samples, we adopt Kim et al. [1]'s proposed AdaLoss, which is represented as shown in Equation (7). AdaLoss is an adaptive quality margin loss function that adaptively adjusts the loss margin based on the quality of sample features during training, thus improving the model's ability to distinguish classified samples. AdaLoss incorporates an adaptive angular margin and an adaptive additive term, efficiently adjusting the loss margin for each sample. This allows the model to better focus on the differences between high-quality and low-quality samples during training.

$$\mathcal{L}_{ada} = -\log \frac{exp(f(\theta_{y_i}, m))}{exp(f(\theta_{y_i,m})) + \sum_{j \neq y_i}^{n} exp(scos\theta_j)}, \tag{7}$$

where $f(x)$ is the boundary function regarding the boundary m, and θ_j is the angle between the feature vector and jth classifier weight vector. The adaptive boundary function for inter-class and intra-class adjustment proposed by Kim et al. [1] can handle some low-quality samples in the data. If the image quality is high, hard samples are emphasized during training; if the image quality is low, hard samples are not emphasized. We utilize the adaptive boundary function proposed by AdaFace for sample prediction and classification. The boundary function f is formulated by Equation (8).

$$\begin{aligned} f(\theta_j, m) &= scos(\theta_j + G_{ag}) - G_{ad}, \quad j = y_i, \\ f(\theta_j, m) &= scos(\theta_j), \quad j \neq y_i, \end{aligned} \tag{8}$$

where G_{ag} represents the angle function and G_{ad} represents the additional margin function. It can be expressed by Equation (9).

$$G_{ag} = -m. \widehat{\| \mathcal{Z}_i \|}, G_{ad} = m. \| \mathcal{Z}_i \| + m, \tag{9}$$

where $\| \hat{\mathcal{Z}}_i \|$ is the feature vector. When $\| \hat{\mathcal{Z}}_i \|$ = -1, the G_{ag} function turns into ArcFace. When $\| \hat{\mathcal{Z}}_i \|$ = 0, it turns into CosFace. When $\| \hat{\mathcal{Z}}_i \|$ = 1, it turns into a shifted negative angular margin.

3.4.4. Total Loss

The overall loss function formula is shown in Equation (10):

$$\mathcal{L}_{total} = \alpha * \mathcal{L}_{light} + \beta * \mathcal{L}_{lf} + \gamma * \mathcal{L}_{Ada}, \tag{10}$$

where α represents the weight of illumination triplet loss, β is the weight of lighting–facial correlation loss, and γ is the weight of AdaLoss.

4. Experiments

In this section, we introduce a comparison of our model's recognition performance on low-light face datasets with low-light face recognition methods and some general face recognition methods. We visualize the recognition results of our method and compare them with general face recognition methods. Ablation studies are presented to verify the effectiveness of each module in our model.

4.1. Implementation Details

4.1.1. Training Set

We follow the low-light simulation synthesis scheme proposed by [49] to synthesize our training set. We select 2500 identities from the CASIA-WebFace [50] dataset for training, with each ID having eight normal/low-light images. The synthesized training set is referred to as LowCASIA-Train, and some of the data for low-light face synthesis are shown in Figure 4. The first and second rows are the low-light face image set and the normal illumination face image set, respectively. The third row comprises the triplets generated from the two sets: the anchor is a low-light face sample, the positive is a low-light face sample of a different person, and the negative is a normal illumination face sample. Our method is based on training with unpaired data, meaning we do not require the face identity IDs in each triplet to be the same. This is a random outcome.

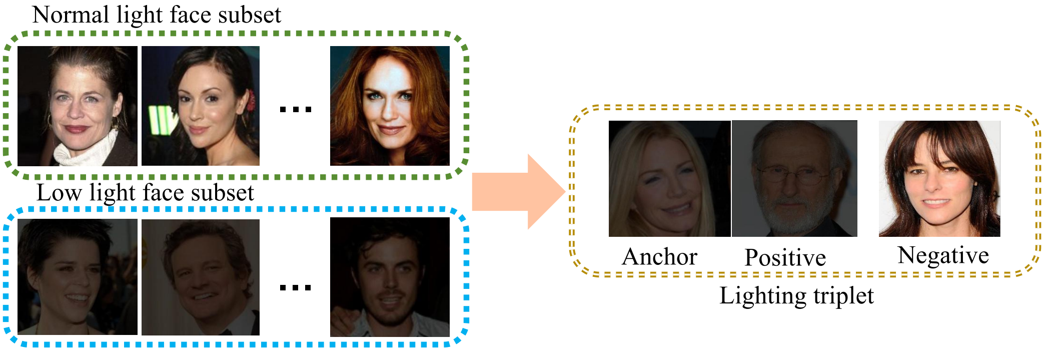

Figure 4. Examples from the LowCASIA-Train, where green boxes indicate well-illuminated facial areas, blue boxes denote low-light facial areas, and yellow boxes represent randomly selected lighting triples.

4.1.2. Validation Set

We evaluate the performance of our proposed model on several publicly available low-light and varying illumination face datasets, including four real low-light face datasets: Extended-YaleB [51], CASPEAL [52], CMU-PIE [53], and SoF (Specs on Faces) [54], and three simulated low-light datasets: LFW* [6], LaPa-Test [55], and CASIA-Test [56]. We briefly introduce the details of the dataset in Table 2 and present some examples from the dataset in Figure 5.

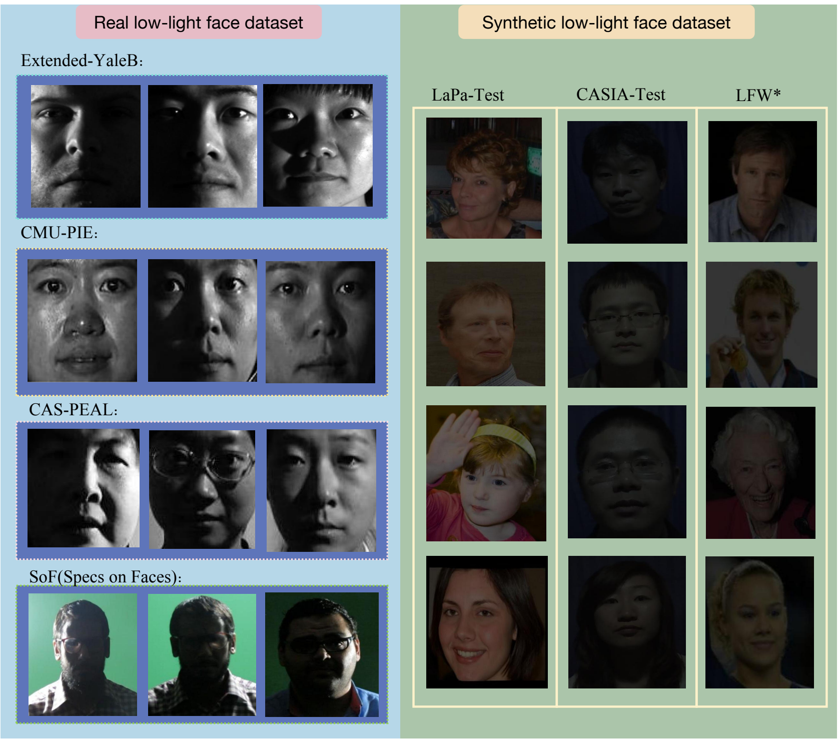

Figure 5. Examples from the validation set. LFW* indicates the Low-light version of LFW.

Table 2. Description of validation set.

Real/Synthetic	Name	Total	Resolution	RGB/Gray
Real	Extended-YaleB	1426	128 × 128	Gray
Real	CASPEAL	2242	128 × 128	Gray
Real	CMU-PIE	1632	128 × 128	Gray
Real	SoF	2662	128 × 128	RGB
Synthetic	LFW*	13,233	128 × 128	RGB
Synthetic	CASIA-Test	500	128 × 128	RGB
Synthetic	LaPa-Test	1789	128 × 128	RGB

4.1.3. Training Details

We trained the model on PyTorch 1.10.0 using an NVIDIA RTX 2080 Ti GPU (12 vCPU Intel(R) Xeon(R) Platinum 8255C CPU @ 2.50 GHz), and we used Mobilefacenets as the backbone for feature extraction. The batch size was set to 64, and the model was trained using the SGD algorithm with a momentum of 0.9 and a weight decay of 1×10^{-4}. The input face image resolution to the network was 128 × 128. The training was set to 80 epochs, with an initial learning rate of 1×10^{-4}, which decayed by a factor of 10 at the 20th and 40th epochs. For the values of α, β, and γ in the total loss, we set $\alpha = 0.1$, $\beta = 0.1$, and $\gamma = 1$.

4.2. Comparisons with State-of-the-Art Methods

Based on the review in Section 2.2, we will compare our method with several existing low-light face recognition approaches. Specifically, we will compare it with illumination normalization methods, image enhancement methods, and several general face recognition methods.

To validate the effectiveness of our proposed method and ensure a fair comparison with existing illumination normalization methods, we test the face recognition accuracy of our model on three real-world low-light facial datasets. As shown in Table 3, our method outperforms these illumination normalization methods, achieving the best performance across all three datasets. Our method outperforms the second-best approach by 2.8 percentage points on the YaleB dataset and by 1.2 percentage points on the CAS-PEAL dataset, while achieving the best performance on the CMU-PIE dataset.

Table 3. Comparisons of the face recognition rate (%) with illumination normalization schemes.

Methods	Extended-YaleB	CAS-PEAL	CMU-PIE
Original Image	79.1	83.8	99.7
SQI [57]	81.2	84.3	98.2
Retinex-Net [41]	85.2	85.4	98.2
Pixpix-GAN [58]	51.3	70.2	92.4
Cycle-GAN [59]	78.9	82.8	98.0
Supervised CycleGAN [60]	80.4	84.3	98.5
AJGAN [15]	86.8	87.4	99.9
FIN-GAN [13]	93.4	92.4	99.9
Ours	96.2	93.2	100

Our method does not rely on image enhancement techniques. To investigate why our method still outperforms image enhancement-based methods in low-light scenarios. we conduct a quantitative comparison between our method and these approaches on both a real low-light dataset and a synthetic low-exposure dataset. As shown in Table 4, these low-light face recognition methods based on image enhancement techniques exhibit varying degrees of performance degradation. RetinexNet-enhanced images achieve only 34.29% recognition performance on LFW*, while the best enhancement method, Low-Face Net, achieves only a 75.63% recognition rate on LFW* after enhancement. Our method outperforms the best Low-Face Net by 9.92%. On the real low-light face dataset SoF, we also outperform the best method by 14.27%, indicating that enhancing low-light images as a preprocessing step for face recognition systems is not very feasible.

To further illustrate the effectiveness of our method, we compare our approach with some general face recognition methods. For a fair comparison, the general face recognition methods were trained using the same training data and hyperparameters as our approach. The backbone used was Mobilefacenets, and the training set was LowCASIA-Train. According to the results shown in Table 5, on the LFW* dataset, our performance is 85.55%, which is 1.47% higher than the second-best. On the CASIA-Test dataset, our performance is 91.18%, which is 2.8% higher than the second-best. On the LaPa-Test dataset, our best performance is 92.38%, which is very close to the second-best at 92.36%. Our method performs excellently across all three datasets, with the most significant improvement on the CASIA-Test dataset. This demonstrates the effectiveness of our method on low-light face datasets. Additionally, to further illustrate the poor performance of general face recognition methods in low-light scenarios and the superiority of our method in such conditions, we used the publicly available ArcFace model (ArcFace+) for testing on these three low-light datasets. As shown in the first row of Table 5, ArcFace+ performed 4.39% worse on LFW*, 12.45% worse on CASIA-Test, and 1% worse on LaPa-Test compared to our method. This also indicates that general face recognition models perform poorly in low-light scenarios, falling short compared to specialized low-light face recognition algorithms.

Table 4. Comparisons of the face recognition rate (%) with image enhancement method.

Methods	LFW*	SoF
LIME [42]	70.33	72.95
RetinexNet [41]	34.29	64.04
SSIENet [43]	70.46	78.56
RUAS [40]	66.71	39.16
Zero-DCE [38]	63.98	77.05
Zero-DCE++ [39]	69.11	49.20
Low-Face Net [11]	75.63	83.99
Ours	85.55	98.26

Table 5. Comparisons of the face recognition rate (%) with general face recognition methods.

Method	LFW*	CASIA-Test	LaPa-Test
ArcFace^{+} **[5]**	81.16	78.73	79.87
CenterLoss [30]	83.63	88.38	91.47
CosFace [29]	81.89	85.60	91.24
ArcFace [5]	83.29	87.98	90.67
CurricularFace [3]	84.05	86.78	92.06
AdaFace [1]	83.95	85.95	92.36
ElasticFace [4]	84.08	86.80	91.46
Ours	85.55	91.18	92.38

Our method demonstrates significant advantages in low-light face recognition, excelling not only in experimental results but also in methodological innovation. First, the proposed feature decoupling and fusion network effectively captures facial features under low-light conditions, enhancing the model's adaptability to varying illumination levels through its optimized architecture. This structured approach to feature extraction allows us to bypass the need for image enhancement preprocessing, avoiding the distortions and performance degradation often seen in traditional methods. Additionally, the consistent performance of our method across various low-light datasets highlights its broad adaptability to different lighting conditions, further enhancing the model's generalizability. These innovations ensure that our method not only surpasses state-of-the-art techniques in accuracy but also offers greater stability and robustness when dealing with complex low-light environments.

4.3. Exploratory Experiments

Exploration of different low-light training set versions. To explore the impact of training set configurations on our method, we created three types of synthetic low-light datasets for training. We also considered variations in image resolution and the number of IDs to provide a more comprehensive comparative analysis. Although we propose three types of training sets with different low-light levels, in Table 5, we use CASIA-TrainA as the unified training set for quantitative analysis with general face recognition models. Our method still achieves the best performance, indicating that our approach does not rely on a specific training set. We adjust the LowCASIA-Train proposed in Section 4.1.1, resulting in three versions: LowCASIA-TrainA, LowCASIA-TrainB, and LowCASIA-TrainC. A represents the current version we are using, B indicates fewer low-exposure images than A, $\gamma \subset [2.0, 4.0]$, and C denotes more low-exposure images than A, $\gamma \subset [1.0, 3.0]$. We conduct investigations into the effects of these different dataset versions under consistent model structures and other hyperparameter settings. The experiment results are shown in Table 6.

Table 6. The impact of different versions of low-light training sets on the face recognition rate (%) of DeFFace.

TrainData	Resolution	IDs	Total	LFW*	LaPa-Test	CASIA-Test
LowCASIA-TrainA	128 × 128	8	20 k	85.55	92.38	91.18
	250 × 250	8	20 k	80.23	88.25	87.89
	128 × 128	16	40 k	83.07	91.80	89.97
	250 × 250	16	40 k	81.58	89.72	88.43
LowCASIA-TrainB	128 × 128	8	20 k	83.72	92.15	87.55
	250 × 250	8	20 k	79.11	90.01	90.71
	128 × 128	16	40 k	81.79	88.51	89.92
	250 × 250	16	40 k	79.08	88.79	88.63
LowCASIA-TrainC	128 × 128	8	20 k	84.63	92.26	89.57
	250 × 250	8	20 k	83.44	91.29	90.32
	128 × 128	16	40 k	83.92	90.16	89.71
	250 × 250	16	40 k	82.51	89.23	90.28

According to Table 6, the variations in brightness levels, resolutions, and the number of IDs in the three different versions of the training sets have different impacts on the experimental results. The LowCASIA-TrainA version, configured with a resolution of 128 × 128 and 8 IDs, demonstrated the best performance. In this study, we use LowCASIA-TrainA as the training set for all experiments, with a total of 20,000 samples.

Exploration of D-Net setting. The D-Net plays a crucial role in our approach. In this section, we conduct an experiment to investigate the impact of the number of convolutional layers in the decoupling module on the model's performance. First, by setting the number of conv layers in the two sub-modules of the decoupling module from 1 to 6 (if greater than 6, the model will not converge), we explore the model's face recognition rate on the LFW* and CASIA-Test datasets. As shown in Figure 6 left, we can see that the performance is best when the number of convolutional layers in the decoupling module is two, and the performance decreases as the number of layers increases. This also demonstrates the significant impact of configuring different convolutional layer counts in the decoupled module in our approach. Since the decoupling module consists of two parallel sub-modules, we also explore the face recognition rate of different conv layer combinations in different sub-modules on the LFW* and CASIA-Test datasets. Here, we select parallel modules with three layers or fewer for combination. According to Figure 6 right, in these six combinations of different layers of L-modules and F-modules, the combination of a one-layer L-module and a two-layer F-module performs the best, which is also the best D-Net configuration in our proposed approach.

Exploration of D-Net output dimension. To explore the influence of the final output dimension of the last layer of D-Net in our approach, we compare the impact of different final layer output dimensions of our approach in D-Net on the face recognition rates of the

LFW*, LaPa-Test, and CAS-PEAL datasets, with the decoupling module scheme fixed as a two-layer conv F-module and a one-layer conv L-module as previously described. As shown in Table 7, the model performed best when the output dimension was 512, and performance did not continue to improve with increasing dimensions across the three datasets. Therefore, we chose 512-D as our optimal configuration.

Exploration of backbone setting. To study the impact of different feature extractors on our model's performance, we conduct multiple experiments using several common convolutional neural network architectures as the model's backbone. These architectures included ResNet [61], VGG [62], iResNet [63], and Mobilefacenets [64]. We trained each different backbone network using the same data and parameters and tested them on LFW*, CASIA-Test, and CAS-PEAL to ensure the comparability of the results. As shown in Table 8, among all the feature extraction networks, Mobilefacenets performed the best in our approach. Based on the experimental results, we chose Mobilefacenets as the final selection.

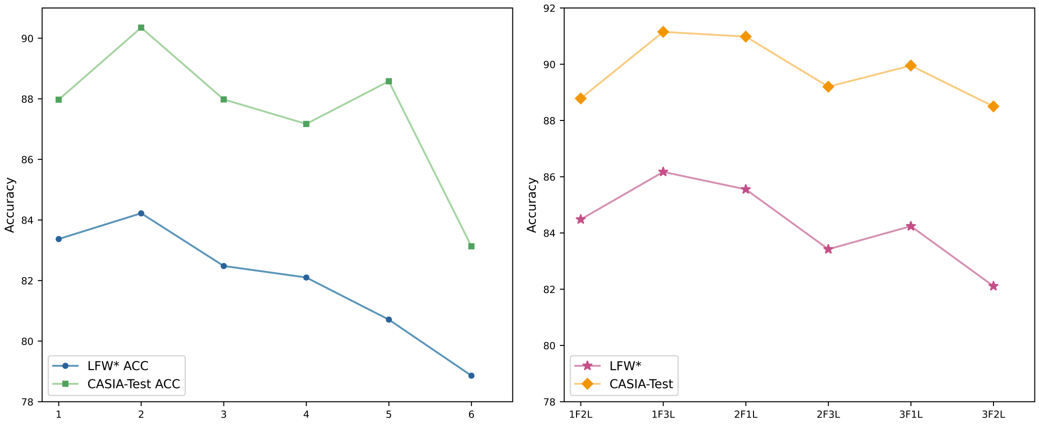

Figure 6. Left depicts the performance of decoupling sub-modules with identical layer configurations, while right illustrates the performance of decoupling sub-modules with varying numbers of layers.

Table 7. The impact of the D-Net output dimension on the face recognition rate (%) of DeFFace.

Output Dimension	LFW*	LaPa-Test	CAS-PEAL
256	80.83	92.26	90.99
512	85.55	92.38	93.17
1024	83.28	92.83	91.35
2046	83.22	79.13	92.44

Table 8. The impact of the backbone setting on the face recognition rate (%) of DeFFace.

Backbone	LFW*	CASIA-Test	CAS-PEAL
ResNet50	64.45	86.78	90.99
ResNet100	63.58	84.78	91.08
iResNet100	56.09	81.13	91.26
VGG16	56.14	80.33	92.24
Mobilefacenets	85.55	91.18	93.17

Exploratory classification loss. We conduct a detailed exploration of the effectiveness of our chosen classification loss function, AdaLoss, by comparing its performance against various commonly used face recognition loss functions within our framework. Classification loss plays a crucial role in guiding the optimization of neural networks and directly influences the performance of face recognition models. Therefore, we evaluate several widely used loss functions, including ArcFace loss, CurricularFace loss, and CosFace loss, to benchmark the performance of AdaLoss in our proposed low-light face recognition model. AdaLoss introduces an adaptive weighting mechanism that dynamically adjusts the contribution of harder samples during training, effectively addressing the imbalance between easy and hard samples in face recognition tasks. This adaptability is especially important in low-light conditions, where the complexity of facial features can vary significantly. By focusing more on challenging samples, AdaLoss enhances the model's ability to learn discriminative features in difficult lighting scenarios. Through a series of experiments, we compare the performance of AdaLoss against other loss functions across several datasets, including LFW* and LaPa-Test. As shown in Table 9, AdaLoss consistently outperforms other loss functions. Specifically, when compared to ArcFace loss, AdaLoss improves the recognition rate by 1.92% on LFW* (from 83.63% to 85.55%) and 0.26% on LaPa-Test (from 92.12% to 92.38%). Similarly, when benchmarked against CosFace loss, AdaLoss demonstrates improvements of 2.83% on LFW* (from 82.72% to 85.55%) and 0.60% on LaPa-Test (from 91.78% to 92.38%).

These experiments confirm that AdaLoss provides significant advantages over other classification loss functions, particularly in scenarios where lighting conditions are sub-optimal. Its ability to dynamically balance sample difficulty contributes to the overall improvement of face recognition accuracy, making it an essential component of our proposed model.

Table 9. Comparison of different loss functions on LFW and LaPa-Test.

Loss	LFW*	LaPa-Test
ArcFace loss [5]	83.63	92.12
CosFace loss [29]	82.72	91.78
CurricularFace loss [3]	83.26	91.29
AdaFace loss [1]	85.55	92.38

Exploration of complexity. In the context of face recognition systems, especially under challenging conditions such as low-light environments, the complexity of the proposed model plays a critical role in its practical applicability. The complexity of the model is measured by factors such as computational cost, model size, inference time, and the number of parameters. The purpose of this section is to explore the key factors contributing to the complexity of the DeFFace model and their impact on real-world applications. While DeFFace provides significant performance improvements in low-light face recognition, its increased computational cost, model size, and inference time must be carefully evaluated to ensure practical usability. By investigating optimization techniques, future research can address these complexities and ensure that the DeFFace model remains effective and scalable across various environments. As shown in Table 10, the comparison reveals that our model achieves an effective balance between complexity and efficiency. With 1.43 M parameters and 310.85 M FLOPs, it is significantly more lightweight than FRNetFuse [12] and CurricularFace [3], which have much larger model sizes and higher computational costs. Despite its smaller size, our model maintains competitive performance, with a reasonable inference time of 0.006 s and 176.16 FPS, making it suitable for real-time low-light face recognition tasks.

Table 10. Comparison of parameters, FLOPs, FPS, and inference time for different models.

Method	Number of Parameters	FLOPs	FPS	Inference Time (s)
Low-faceNet [11]	5.18 M	300.99 M	295.98	0.013
Zero-DCE [38]	4.80 M	1297.61 M	770.55	0.004
ArcFace [5]	1.20 M	2275.74 M	182.23	0.011
FRNetFuse [12]	32.67 M	29,964.08 M	/	0.015
CurricularFace [3]	65.18 M	12,117.20 M	76.91	0.013
Ours	1.43 M	310.85 M	176.16	0.006

In this section of exploratory experiments, we conduct a detailed investigation into the low-light training set versions, the D-Net configuration, the D-Net output dimension, and the feature extraction backbone network. The goal is to determine the optimal hyperparameter settings for each component to ensure the best possible performance of the

DeFFace model in low-light face recognition tasks. We explore three different versions of the low-light training set: LowCASIA-TrainA, LowCASIA-TrainB, and LowCASIA-TrainC. Based on the experimental results, LowCASIA-TrainA demonstrates the best performance across all test sets, so we select it as the final training set version for all experiments, as shown in Figure 4. Next, by adjusting the number of convolutional layers in the decoupling module, we determine that the optimal configuration consists of a combination of one low-light decoupling CNN layer and two face decoupling CNN layers, as shown in Figure 6. This becomes the final D-Net setting in our method. In terms of the D-Net output dimension, the experiments reveal that an output dimension of 512 yields the best performance across various datasets, as shown in Table 7. Thus, 512-D is chosen as the final output dimension of the D-Net. Additionally, we compare the performance of ResNet50, ResNet100, iResNet100, VGG16, and Mobilefacenets as feature extractors. Among these, Mobilefacenets achieves the best results on all test sets, as shown in Table 8. As a result, we select Mobilefacenets as the final feature extraction backbone network.

Through these exploratory experiments, we successfully identify the optimal hyperparameter settings for the DeFFace model, as shown in Table 11, ensuring its best performance in low-light face recognition tasks.

Table 11. Optimal parameter configuration.

TrainData	D-Net Setting	D-Net Output Dimension	Backbone
LowCASIA-TrainA	1-layer L-module 2-layer F-module	512-D	Mobilefacenets

4.4. Qualitative Visualization

To qualitatively evaluate the effectiveness of our proposed method for low-light face recognition, we conducted a visualization comparison against two state-of-the-art methods on the LFW* dataset. Given a low-light image, we obtained the rank-10 results in the dataset through similarity scores. Figure 7 presents a comparative visualization of the rank-10 results for a sample low-light image using our method and two general face recognition models. The results are arranged in descending order of similarity scores.

DeFFace. Figure 7a illustrates the results obtained by our proposed DeFFace method. Clearly, our method can retrieve the most relevant matches. In rank-10 matching, the first four retrievals are correct, and the confidence scores are higher than those of the other two methods. The overall performance is noticeably superior to other methods, showing outstanding performance in low-light environments.

ArcFace*. The second row, shown in Figure 7b, presents the results of ArcFace*. ArcFace* was trained using our training dataset LowCASIA-Train. Five out of the top ten matching results are correct matches. However, the top-ranked match is not correct. Additionally, our method demonstrates significantly higher confidence levels compared to ArcFace*. This indicates that although ArcFace* shows improvements in visual quality and matching success rate, its performance under extreme low-light conditions remains suboptimal.

ArcFace. The third row, shown in Figure 7c, displays the top ten matching results retrieved by the ArcFace model. Here, ArcFace is tested directly using the open-source ArcFace-trained model. The ArcFace model failed to correctly match the relevant faces under low-light conditions, with all top ten results being incorrect matches, highlighting the model's limitations in handling extreme low-light conditions.

The qualitative comparison clearly shows that our method outperforms existing face recognition methods in low-light face recognition. The matches retrieved by our method exhibit higher relevance, indicating that our approach effectively addresses the challenges posed by low-light conditions. This is further supported by the quantitative results presented in Section 4.2, where our method significantly outperforms the other scheme in terms of recognition accuracy.

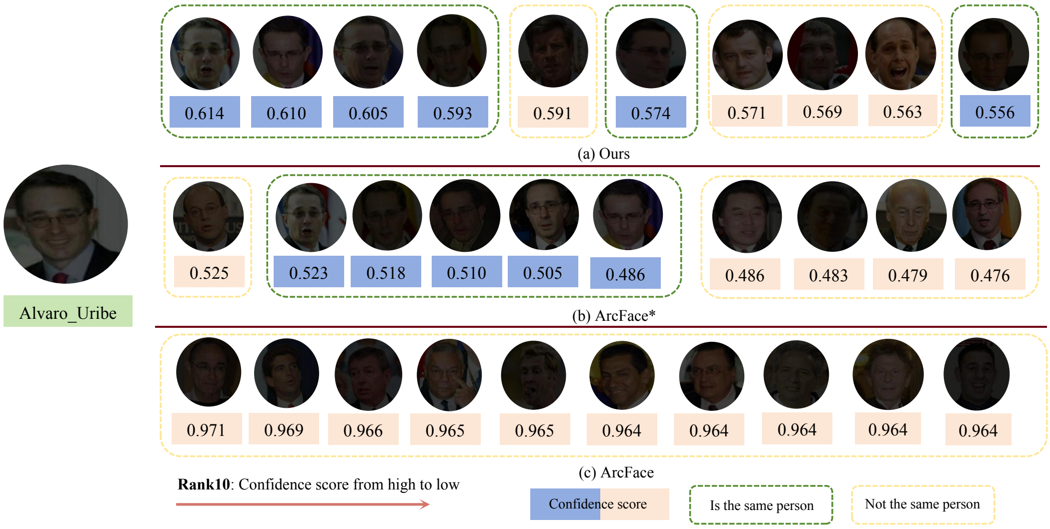

Figure 7. The visualized results of rank-10 retrieval on the low-light face dataset LFW* using our method and the ArcFace method are presented. Using the person on the far left as an example, the green dashed box indicates a match as the same person, the yellow dashed box indicates a different person, and the blue and orange text boxes represent confidence scores. We display the top rank-10 visualization results from high to low.

4.5. Ablation Study

In this section, we further analyze the reasons behind the superior performance of DeFFace in low-light face recognition. We conduct ablation experiments to verify the effectiveness of our proposed D-Net, F-Net, illumination triplet loss, and lighting–facial correlation loss, and their impact on model performance. Here, we present the face recognition rate on LFW* and CASIA-Test. To verify the impact of the decoupling module, we use Mobilefacenets [64] and AdaLoss [1] as the baseline, retraining with cross-entropy loss while discarding the relevant component modules and losses in our approach. Then, we add the decoupling module and illumination triplet loss to the baseline. As shown in Table 12, after incorporating D-Net and the lighting-triple loss, our approach achieves a 0.72% improvement on LFW* and a 1.22% improvement on CASIA-Test. Next, we add the F-Net to the combination of the baseline, D-Net, and the lighting-triple loss, achieving an improvement of 0.22% on LFW* and 2.83% on CASIA-Test. This indicates that the F-Net can effectively extract residual facial-related features from coarse illumination-related features and fuse them with the facial-related features in the D-Net. Finally, adding the lighting–facial correlation loss resulted in an improvement of 0.66% on LFW* and 1.18% on CASIA-Test. The lighting–facial correlation loss can effectively reduce the correlation between disentangled facial-related features and illumination-related features in the feature space. Our optimal configuration is above the baseline, adding D-Net and illumination triplet loss ($\mathcal{L}_{light}$), F-Net, and lighting–facial correlation loss ($\mathcal{L}_{lf}$) to the baseline.

To verify the advantages of unpaired training datasets, we do not rely on paired data for training; we conduct additional experiments using paired data training under the optimal configuration. In the final row of the table, we present the results of these experiments. We replace the standard training data with paired data, where each pair consists of one low-light and one normal-light image of the same subject. Contrary to our expectations, the performance of paired data training is not as effective as unpaired data training. Specifically, the performance on LFW* showed a slight decrease, and a similar trend is observed on the CASIA-Test dataset. This suggests that while paired data training can provide some benefits in terms of consistency between illumination conditions, it may also introduce limitations due to the restricted variability in training data. The outstanding performance is also due to our proposed D-Net and multiple loss constraints. Therefore, our unpaired data training strategy remains the optimal approach for enhancing low-light face recognition performance.

Table 12. Ablation study of the DeFFace on the LFW* and CASIA-Test dataset.

Baseline	D-Net + $\mathcal{L}_{\text{light}}$	F-Net	$\mathcal{L}_{lf}$	Paired Data	LFW*	CASIA-Test
✓					83.95	85.95
✓	✓				84.67	87.17
✓	✓	✓			84.89	90.00
✓	✓	✓	✓		85.55	91.18
✓	✓	✓	✓	✓	85.44	87.60

5. Conclusions and Future Work

In this paper, we propose DeFFace, an innovative unpaired end-to-end solution for low-light face recognition. Leveraging the decomposition concept from the Retinex theory, DeFFace decouples facial-related and illumination-related features through a feature decoupling module (D-Net) and refines the facial features with a feature fusion module (F-Net). To further reduce the correlation between these features, we introduce a lighting-facial correlation loss. Our method supports accurate ID matching using a single low-light image and enables large-scale recognition tasks. Experimentally, DeFFace outperformed general face recognition methods by approximately 2.11%, image enhancement-based solutions by around 16%, and illumination normalization methods by 5.67%. These results demonstrate the effectiveness and robustness of DeFFace, even in the absence of paired normal-light and low-light images.

While the focus of this paper is on enhancing low-light face recognition through feature decoupling and fusion, it is crucial to consider the robustness of the underlying model, as this directly affects the system's resilience to adversarial attacks. Although adversarial robustness is not the central topic of this work, we recognize that face recognition systems are susceptible to such manipulations, which could pose significant threats to data privacy and security. Future research could explore incorporating defense mechanisms against adversarial attacks to strengthen the model's robustness, as securing facial recognition systems against these threats is increasingly important in this field.

Moreover, while we have demonstrated the effectiveness of our method under low-light conditions, our evaluation of the model's performance in scenarios involving partial occlusion and image blurriness is limited. Expanding the test datasets to include such complex, real-world conditions would be a valuable direction for future research, ensuring the model's reliability in a wider range of challenging environments.

Author Contributions: Conceptualization, X.Z., Z.G. and H.G.; methodology—origin drawing, X.Z. and Z.G.; software, X.Z. and S.L.; validation, X.Z. and H.G.; formal analysis, S.L. and Z.G.; investigation, X.Z.; data curation, X.Z.; writing—original draft preparation, X.Z.; writing—review and editing, X.Z., Z.G. and S.L.; visualization, X.Z.; supervision, H.G. All authors have read and agreed to the published version of this manuscript.

Funding: This research received no external funding.

Data Availability Statement: The datasets used in this study are either publicly available or can be accessed upon request: LFW*, LaPa-Test, and CASIA-Test datasets are publicly available at https://github.com/fanyihua0309/Low-FaceNet (accessed on 29 October 2024). SoF dataset is publicly accessible at https://sites.google.com/view/sof-dataset (accessed on 29 October 2024). CAS-PEAL, CMU-PIE, and Extended-YaleB datasets can be obtained from their respective official websites. The LowCASIA-TrainA dataset used for training is available upon request from the corresponding author.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Kim, M.; Jain, A.K.; Liu, X. Adaface: Quality adaptive margin for face recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 18750–18759.
2. Meng, Q.; Zhao, S.; Huang, Z.; Zhou, F. Magface: A universal representation for face recognition and quality assessment. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 14225–14234.
3. Huang, Y.; Wang, Y.; Tai, Y.; Liu, X.; Shen, P.; Li, S.; Li, J.; Huang, F. Curricularface: Adaptive curriculum learning loss for deep face recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 5901–5910.
4. Boutros, F.; Damer, N.; Kirchbuchner, F.; Kuijper, A. Elasticface: Elastic margin loss for deep face recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 1578–1587.
5. Deng, J.; Guo, J.; Xue, N.; Zafeiriou, S. Arcface: Additive angular margin loss for deep face recognition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–17 June 2019; pp. 4690–4699.
6. Huang, G.B.; Mattar, M.; Berg, T.; Learned-Miller, E. Labeled faces in the wild: A database forstudying face recognition in unconstrained environments. In Proceedings of the Workshop on Faces in 'Real-Life' Images: Detection, Alignment, and Recognition, Marseille, France, 17 October 2008.
7. Sengupta, S.; Chen, J.C.; Castillo, C.; Patel, V.M.; Chellappa, R.; Jacobs, D.W. Frontal to profile face verification in the wild. In Proceedings of the 2016 IEEE Winter Conference on Applications of Computer Vision (WACV), Lake Placid, NY, USA, 7–10 March 2016; pp. 1–9.
8. Moschoglou, S.; Papaioannou, A.; Sagonas, C.; Deng, J.; Kotsia, I.; Zafeiriou, S. Agedb: The first manually collected, in-the-wild age database. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 51–59.
9. Zheng, T.; Deng, W.; Hu, J. Cross-age lfw: A database for studying cross-age face recognition in unconstrained environments. *arXiv* **2017**, arXiv:1708.08197.
10. Zheng, T.; Deng, W. Cross-pose lfw: A database for studying cross-pose face recognition in unconstrained environments. *Beijing Univ. Posts Telecommun. Tech. Rep.* **2018**, *5*, 5.
11. Fan, Y.; Wang, Y.; Liang, D.; Chen, Y.; Xie, H.; Wang, F.L.; Li, J.; Wei, M. Low-FaceNet: Face Recognition-driven Low-light Image Enhancement. *IEEE Trans. Instrum. Meas.* **2024**, *73*, 5019413. [CrossRef]
12. Huang, Y.H.; Chen, H.H. Deep face recognition for dim images. *Pattern Recognit.* **2022**, *126*, 108580. [CrossRef]
13. Hu, Y.; Lu, M.; Xie, C.; Lu, X. FIN-GAN: Face illumination normalization via retinex-based self-supervised learning and conditional generative adversarial network. *Neurocomputing* **2021**, *456*, 109–125. [CrossRef]
14. Le, H.A.; Kakadiaris, I.A. SeLENet: A semi-supervised low light face enhancement method for mobile face unlock. In Proceedings of the 2019 International Conference on Biometrics (ICB), Crete, Greece, 4–7 June 2019; pp. 1–8.
15. Han, X.; Yang, H.; Xing, G.; Liu, Y. Asymmetric joint GANs for normalizing face illumination from a single image. *IEEE Trans. Multimed.* **2019**, *22*, 1619–1633. [CrossRef]
16. Huang, Y.H.; Chen, H.H. Face recognition under low illumination via deep feature reconstruction network. In Proceedings of the 2020 IEEE International Conference on Image Processing (ICIP), Abu Dhabi, United Arab Emirates, 25–28 October 2020; pp. 2161–2165.
17. Bendjillali, R.I.; Beladgham, M.; Merit, K.; Taleb-Ahmed, A. Illumination-robust face recognition based on deep convolutional neural networks architectures. *Indones. J. Electr. Eng. Comput. Sci.* **2020**, *18*, 1015–1027. [CrossRef]
18. Lu, Q.; Gan, P. Low-Light Face Recognition and Identity Verification Based on Image Enhancement. *Trait. Du Signal* **2022**, *39*, 513. [CrossRef]
19. Miao, X.; Wang, S. Knowledge Guided Representation Disentanglement for Face Recognition from Low Illumination Images. In Proceedings of the 30th ACM International Conference on Multimedia, Lisbon, Portugal, 10–14 October 2022; pp. 6655–6663.
20. Wang, Z.; Deng, W.; Ge, J. FIE-GAN: Illumination Enhancement Network for Face Recognition. In Proceedings of the Pattern Recognition and Computer Vision: 4th Chinese Conference, PRCV 2021, Beijing, China, 29 October–1 November 2021; Proceedings, Part III 4; Springer: Berlin/Heidelberg, Germany, 2021; pp. 214–225.
21. Brainard, D.H.; Wandell, B.A. Analysis of the retinex theory of color vision. *JOSA A* **1986**, *3*, 1651–1661. [CrossRef]
22. Lezama, J.; Qiu, Q.; Sapiro, G. Not afraid of the dark: Nir-vis face recognition via cross-spectral hallucination and low-rank embedding. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 6628–6637.
23. Li, S.Z.; Chu, R.; Liao, S.; Zhang, L. Illumination invariant face recognition using near-infrared images. *IEEE Trans. Pattern Anal. Mach. Intell.* **2007**, *29*, 627–639. [CrossRef]
24. Yi, D.; Liu, R.; Chu, R.; Lei, Z.; Li, S.Z. Face matching between near infrared and visible light images. In Proceedings of the Advances in Biometrics: International Conference, ICB 2007, Seoul, Republic of Korea, 27–29 August 2007; Springer: Berlin/Heidelberg, Germany, 2007; pp. 523–530.
25. He, R.; Cao, J.; Song, L.; Sun, Z.; Tan, T. Adversarial cross-spectral face completion for NIR-VIS face recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *42*, 1025–1037. [CrossRef] [PubMed]

26. Vakhshiteh, F.; Nickabadi, A.; Ramachandra, R. Adversarial attacks against face recognition: A comprehensive study. *IEEE Access* **2021**, *9*, 92735–92756. [CrossRef]
27. Baia, A.E.; Biondi, G.; Franzoni, V.; Milani, A.; Poggioni, V. Lie to me: Shield your emotions from prying software. *Sensors* **2022**, *22*, 967. [CrossRef] [PubMed]
28. Massoli, F.V.; Carrara, F.; Amato, G.; Falchi, F. Detection of face recognition adversarial attacks. *Comput. Vis. Image Underst.* **2021**, *202*, 103103. [CrossRef]
29. Wang, H.; Wang, Y.; Zhou, Z.; Ji, X.; Gong, D.; Zhou, J.; Li, Z.; Liu, W. Cosface: Large margin cosine loss for deep face recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 5265–5274.
30. Wen, Y.; Zhang, K.; Li, Z.; Qiao, Y. A discriminative feature learning approach for deep face recognition. In Proceedings of the Computer vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; Springer: Berlin/Heidelberg, Germany, 2016; pp. 499–515.
31. Liu, W.; Wen, Y.; Yu, Z.; Li, M.; Raj, B.; Song, L. Sphereface: Deep hypersphere embedding for face recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 212–220.
32. Kalka, N.D.; Maze, B.; Duncan, J.A.; O'Connor, K.; Elliott, S.; Hebert, K.; Bryan, J.; Jain, A.K. Ijb–s: Iarpa janus surveillance video benchmark. In Proceedings of the 2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS), Redondo Beach, CA, USA, 22–25 October 2018; pp. 1–9.
33. Cheng, Z.; Zhu, X.; Gong, S. Low-resolution face recognition. In Proceedings of the Computer Vision—ACCV 2018: 14th Asian Conference on Computer Vision, Perth, Australia, 2–6 December 2018; Revised Selected Papers, Part III 14; Springer: Berlin/Heidelberg, Germany, 2019; pp. 605–621.
34. Fang, L.; Li, S. Face recognition by exploiting local Gabor features with multitask adaptive sparse representation. *IEEE Trans. Instrum. Meas.* **2015**, *64*, 2605–2615. [CrossRef]
35. Miyamoto, T.; Hashimoto, H.; Hayasaka, A.; Ebihara, A.F.; Imaoka, H. Joint feature distribution alignment learning for NIR-VIS and VIS-VIS face recognition. In Proceedings of the 2021 IEEE International Joint Conference on Biometrics (IJCB), Shenzhen, China, 4–7 August 2021; pp. 1–8.
36. Schroff, F.; Kalenichenko, D.; Philbin, J. Facenet: A unified embedding for face recognition and clustering. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 815–823.
37. Chen, C.; Chen, Q.; Xu, J.; Koltun, V. Learning to see in the dark. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 3291–3300.
38. Guo, C.; Li, C.; Guo, J.; Loy, C.C.; Hou, J.; Kwong, S.; Cong, R. Zero-reference deep curve estimation for low-light image enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 1780–1789.
39. Li, C.; Guo, C.; Loy, C.C. Learning to enhance low-light image via zero-reference deep curve estimation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 4225–4238. [CrossRef]
40. Liu, R.; Ma, L.; Zhang, J.; Fan, X.; Luo, Z. Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 10561–10570.
41. Wei, C.; Wang, W.; Yang, W.; Liu, J. Deep retinex decomposition for low-light enhancement. *arXiv* **2018**, arXiv:1808.04560.
42. Guo, X.; Li, Y.; Ling, H. LIME: Low-light image enhancement via illumination map estimation. *IEEE Trans. Image Process.* **2016**, *26*, 982–993. [CrossRef]
43. Zhang, Y.; Di, X.; Zhang, B.; Wang, C. Self-supervised image enhancement network: Training with low light images only. *arXiv* **2020**, arXiv:2002.11300.
44. Tran, L.; Yin, X.; Liu, X. Disentangled representation learning gan for pose-invariant face recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1415–1424.
45. Wu, X.; Huang, H.; Patel, V.M.; He, R.; Sun, Z. Disentangled variational representation for heterogeneous face recognition. In Proceedings of the AAAI Conference on Artificial Intelligence, Honolulu, HI, USA, 27 January–1 February 2019; Volume 33, pp. 9005–9012.
46. Liu, Y.; Wei, F.; Shao, J.; Sheng, L.; Yan, J.; Wang, X. Exploring disentangled feature representation beyond face identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 2080–2089.
47. Yang, Z.; Liang, J.; Fu, C.; Luo, M.; Zhang, X.Y. Heterogeneous face recognition via face synthesis with identity-attribute disentanglement. *IEEE Trans. Inf. Forensics Secur.* **2022**, *17*, 1344–1358. [CrossRef]
48. Peng, X.; Yu, X.; Sohn, K.; Metaxas, D.N.; Chandraker, M. Reconstruction-based disentanglement for pose-invariant face recognition. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 1623–1632.
49. Lv, F.; Li, Y.; Lu, F. Attention guided low-light image enhancement with a large scale low-light simulation dataset. *Int. J. Comput. Vis.* **2021**, *129*, 2175–2193. [CrossRef]
50. Yi, D.; Lei, Z.; Liao, S.; Li, S.Z. Learning face representation from scratch. *arXiv* **2014**, arXiv:1411.7923.

51. Georghiades, A.S.; Belhumeur, P.N.; Kriegman, D.J. From few to many: Illumination cone models for face recognition under variable lighting and pose. *IEEE Trans. Pattern Anal. Mach. Intell.* **2001**, *23*, 643–660. [CrossRef]
52. Gao, W.; Cao, B.; Shan, S.; Chen, X.; Zhou, D.; Zhang, X.; Zhao, D. The CAS-PEAL large-scale Chinese face database and baseline evaluations. *IEEE Trans. Syst. Man Cybern. Part A Syst. Hum.* **2007**, *38*, 149–161.
53. Sim, T.; Baker, S.; Bsat, M. The CMU Pose, Illumination and Expression database of human faces. In *Technical Report CMU-RI-TR-OI-02*; Carnegie Mellon University: Pittsburgh, PA, USA, 2001.
54. Afifi, M.; Abdelhamed, A. Afif4: Deep gender classification based on adaboost-based fusion of isolated facial features and foggy faces. *J. Vis. Commun. Image Represent.* **2019**, *62*, 77–86. [CrossRef]
55. Liu, Y.; Shi, H.; Shen, H.; Si, Y.; Wang, X.; Mei, T. A new dataset and boundary-attention semantic segmentation for face parsing. In Proceedings of the AAAI Conference on Artificial Intelligence, New York, NY, USA, 7–12 February 2020; Volume 34, pp. 11637–11644.
56. Dong, J.; Wang, W.; Tan, T. Casia image tampering detection evaluation database. In Proceedings of the 2013 IEEE China Summit and International Conference on Signal and Information Processing, Beijing, China, 6–10 July 2013; pp. 422–426.
57. Wang, H.; Li, S.Z.; Wang, Y. Generalized quotient image. In Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, Washington, DC, USA, 27 June–2 July 2004; CVPR 2004; Volume 2, p. II.
58. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-to-image translation with conditional adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1125–1134.
59. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–27 October 2017; pp. 2223–2232.
60. Chen, T.; Zhai, X.; Ritter, M.; Lucic, M.; Houlsby, N. Self-supervised gans via auxiliary rotation loss. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 12154–12163.
61. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
62. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
63. He, T.; Zhang, Z.; Zhang, H.; Zhang, Z.; Xie, J.; Li, M. Bag of tricks for image classification with convolutional neural networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 558–567.
64. Chen, S.; Liu, Y.; Gao, X.; Han, Z. Mobilefacenets: Efficient cnns for accurate real-time face verification on mobile devices. In Proceedings of the Biometric Recognition: 13th Chinese Conference, CCBR 2018, Urumqi, China, 11–12 August 2018; Proceedings 13; Springer: Berlin/Heidelberg, Germany, 2018; pp. 428–438.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.