Single-stage Face Detection under Extremely Low-light Conditions

Jun Yu
University of Science and Technology of China
Hefei, China
harryjun@ustc.edu.cn

Xinlong Hao
University of Science and Technology of China
Hefei, China
haoxl@mail.ustc.edu.cn

Peng He*
University of Science and Technology of China
Hefei, China
hp0618@mail.ustc.edu.cn

Abstract

Face detection has been well studied for many years. One remaining challenge is to detect faces from low-light images. The brightness of the image captured under extremely low-light conditions could be very low and the contrast will be severely reduced. It is easy to cause confusion during feature extraction and affects the performance of face detection. In this paper, we propose a single-stage low-light face detection method. First, we design an improved MSRCR method to increase the image quality under the condition of ensuring that the colors of the image are not distorted. It shows better enhancement effect than other methods in the DARK FACE dataset, especially the low-resolution face details are well preserved. There are a number of small, blurred and partially occluded faces. To address this, the Pyramidbox algorithm is a very effective face detection algorithm. Moreover, we conduct multi-scale tests to further develop the performance of the model and integrated the results through Soft-NMS method to obtain final results. Integrating these techniques, this paper has achieved high accuracy and obtained excellent results in the face detection task of the DARK FACE dataset.

1. Introduction

Face detection is a research hotspot in the field of computer vision and has received more and more attention in recent years [13, 16, 35]. The related research has been maturely applied in security, finance, transportation and other industries and fields [5, 43]. As the development of technology, given an ideal image and without the influence of complex illumination, low resolution and occlusion, the face detection algorithm has been able to guarantee high accuracy and stability.

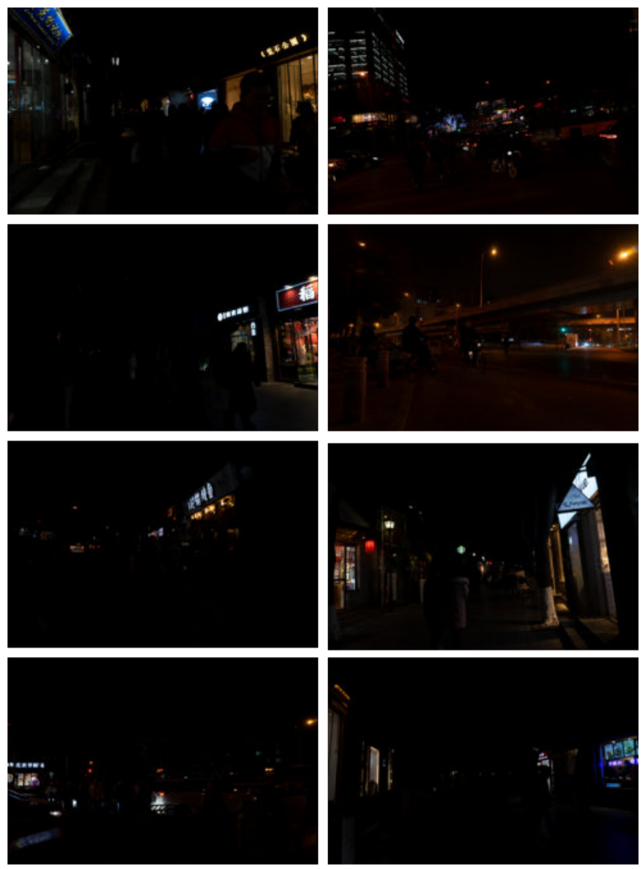


Figure 1. The images provided in DARK FACE dataset. They are captured during the nighttime, at teaching buildings, streets, bridges, overpasses, parks etc., all labeled with bounding boxes of human face.

However, the continuous expansion of application scenarios also puts forward higher requirements on the face detection technology. One of the challenging problems is

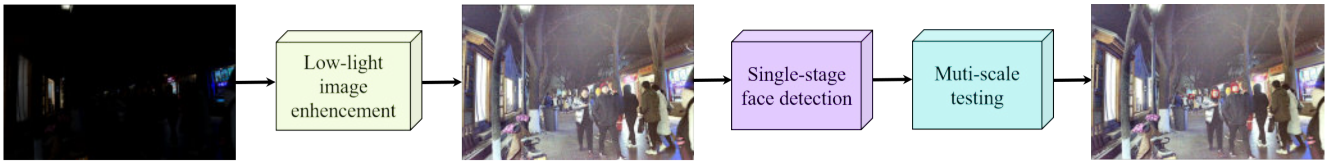


Figure 2. The flow diagram of low-light face detection method under extremely low-light conditions proposed in this paper. It includes three parts: low-light image enhancement, single-stage face detection and multi-scale testing.

the face detection in extremely low-light scenes. Images captured under extremely low-light conditions belong to the common type of low-quality images. The brightness of the image is very low and will reduce the contrast, so it is easy to cause confusion during feature extraction and hurt the performance of face detection. In addition, when processing low-light images obtained in real scenes, serious noise pollution and uneven illumination must be considered. The former may destroy the basic structure of the face in the image, while the latter may interfere with the effectiveness of spatial feature extraction and bring unavoidable troubles.

In order to systematically study the performance of face detection algorithms under extremely low light conditions, a challenging benchmark named DARK FACE [41, 36] was recently constructed. As shown in Figure 1, DARK FACE dataset provides 6,000 real-world low light images captured during the nighttime, at teaching buildings, streets, bridges, overpasses, parks etc. They are all labeled with bounding boxes of human face, as the training and validation sets. The current advanced face detection algorithm has also been tested on DARK FACE dataset. For example, DSFD [18] produces an mAP of 15.3%, in a sharp contrast to above 90% on the hard subset of the popular WIDER FACE [39] benchmark, which shows that it remains extremely challenging to detect faces under low-light conditions.

Faced with the drawbacks of low-light images, image enhancement technology is an effective processing method. At present, retinex theory is one of the most popular methods in the field of low-light image enhancement [17]. There are many algorithms based on retinex theory and the representative method is multi-scale retinex with color restoration (MSRCR) [15]. The MSRCR method proposes to estimate the illumination of the input image using Gaussian surround filtering of different scales and conduct enhancement by applying color restoration followed by linear stretching to the logarithm of reflectance [24]. This method has shown great ability in providing dynamic range compression, color restoration and preserving most of the details. Based on the MSRCR method, this article will seek the best image enhancement method to improve the problem of face detection algorithm failure caused by low-light.

The Pyramidbox algorithm is a very effective face detection algorithm [33]. Compared with previous methods, it proposes an anchor-based context assisted method to introduce supervised information on learning contextual features for small, blurred and partially occluded faces. The Low-level Feature Pyramid Networks (LFPN) and Context-sensitive Prediction Module are designed in order to merge contextual features and facial features better and handle faces with different scales well in a single shot. The algorithm has strong adaptability in images with different characteristics, maintaining high accuracy and stability. We choose the Pyramidbox algorithm as the core idea of the low-light face detection algorithm, and adjust the loss function according to the characteristics of the data set to obtain the detection network that best matches the image.

Driven by the above motivations, we propose a single-stage face detection method under extreme low-light conditions. It contains three modules, as shown in Figure 2. First, based on the MSRCR method, the image enhancement module is designed to complete the preprocessing of the low-light images in the dataset. Then the Pyramidbox algorithm is used as the core module of face detection and the classification loss function and regression loss function are modified for image features. Moreover, multi-scale testing is used to improve the overall model detection performance as there are a large number of small targets in images. Experiment results show that the design of these three modules has achieved a great improvement in the face detection results under extremely low-light conditions.

To summarize, we make the following contributions:

1) An effective method to solve the problem of low-light face detection: single-stage low-light face detection method based on the improved MSRCR method,

2) The Pyramidbox algorithm and multi-scale testing scheme are proved to be effective for face detection under extremely low-light conditions,

3) The advanced performance of the model is verified on the DARK FACE dataset and has obtained leading experimental results.

2. Related Work

The focus of this paper is to propose a face detection method for low-light images and it is closely related to two aspects: low-light image enhancement and face detection.

Low-light image enhancement. Low-light image enhancement has always been an important means to improve image perception quality [20]. The common enhancement methods can be divided into two categories [37]: 1) image restoration based on physical models; 2) image enhancement based on image processing techniques. For the first category, by establishing and inverting the image degradation process to obtain the best estimate of the clear image [12, 31], and the second category directly improves contrast and highlights details by global or local pixel processing, regardless of the cause of color cast and image degradation [15]. Many effcient solutions are frequently designed based on the retinex theory [17]. It assumes an image as a combination of a reflectance map that reflects the physical characteristic of scene objects and a spatially smooth illumination map. Based on this theory, algorithms were designed to focus on resolving the ambiguity between illumination and reflectance by imposing certain priors on a variational model based on empirical observations [7, 8, 11, 19]. There are many algorithms based on retinex theory such as single-scale retinex (SSR), multi-scale retinex (MSR) and multi-scale retinex with color restoration (MSRCR) [15].

In this paper, we constantly adjust and improve the MSRCR algorithm to find a low-light image enhancement method that best matches the DARK FACE dataset, which not only improves the perception quality of the original image, but also ensures that the subsequent face detection algorithm can exert more stable performances.

Face detection. Face detection is the most fundamental and essential task in face-related applications. The most representative and outstanding work in the early is utilizing AdaBoost algorithm with Haar-Like features to train a cascade of face and non-face classifiers [35]. Then a large number of subsequent works were proposed for the improvement of cascade detectors [44, 28, 3, 38]. With the rapid development of convolutional neural networks(CNN), a lot of progress for face detection has been made in recent years. The object detection algorithm designed based on CNN has achieved great success in this field, including R-CNN [9, 10], SSD [22], YOLO [29] and their extensions. In addition, for situations that detect faces in uncontrolled environment, the anchor-based detection framework can play a greater advantage, for example: WIDER FACE [1], SSH [26] and S^{3}FD [42] develop scale-invariant networks to detect faces with different scales.

Refer to previous design ideas for detectors, the PyramidBox algorithm was proposed to handle the hard face detection problem [33]. Its advantage is reflected in improving the utilization of contextual information. By designing the PyramidAnchor and Low-level Feature Pyramid Network to combine adequate high-level context semantic feature and low-level facial feature together, which allows the PyramidBox algorithm to predict faces of all scales in a single shot. In this paper, we use the PyramidBox algorithm as the core module of face detection. After image enhancement, the sample characteristics of images such as noise, image resolution, and contrast have changed a lot. We adjusted the PyramidBox Loss Layer with reference to facal loss [21] and balanced L1 loss [27], and finally got the face detector we need.

3. Proposed Method

The method we propose is divided into three modules: the MSRCR low-light image enhancement module, the PyramidBox face detection module and the multi-scale test module. They are combined together to form the low-light single-stage face detection model.

3.1. MSRCR low-light image enhancement module

In most cases, retinex processing on images will cause serious saturation reduction problems, resulting in color distortion [15]. The advantage of the MSRCR method is reflected in the restoration of colors and maintaining the color consistency before and after image enhancement. The single-scale retinex is given by:

$$R_i(x,y) = \log I_i(x,y) - \log[F(x,y) * I_i(x,y)] \qquad (1)$$

where $R_i(x,y)$ is the retinex output, $I_i(x,y)$ is the image distribution in the ith spectral band and $F(x,y)$ is the surround function.

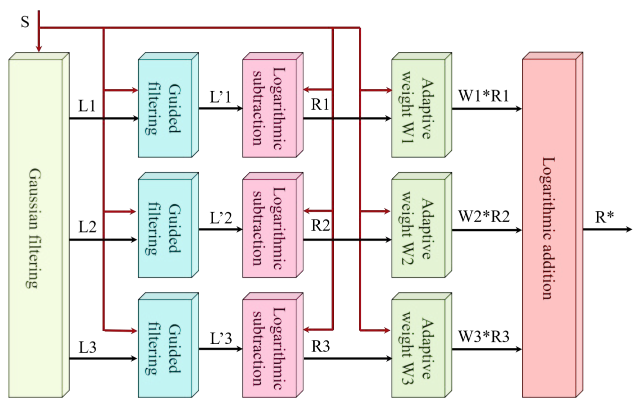


Figure 3. The pipeline of the improved MSRCR algorithm.

For the MSRCR method, we consider a simple colorimetric transform. It is a method to create a relative color space, and in so doing becomes less dependent than raw spectrophotometry on illuminant spectral distributions:

$$I'_i(x,y) = \frac{I_i(x,y)}{\Sigma_{i=1}^{S} I_i(x,y)} \qquad (2)$$

for the ith color band, and S is the number of spectral channels ($S = 3$ for the RGB color space).

Then we can get the output of MSRCR as:

$$R_{MSRCR_i}(x,y) = C_i(x,y)\Sigma_{n=1}^{N}\omega_n R_{n_i} \qquad (3)$$

where

$$C_i(x,y) = f[I'_i(x,y)] \qquad (4)$$

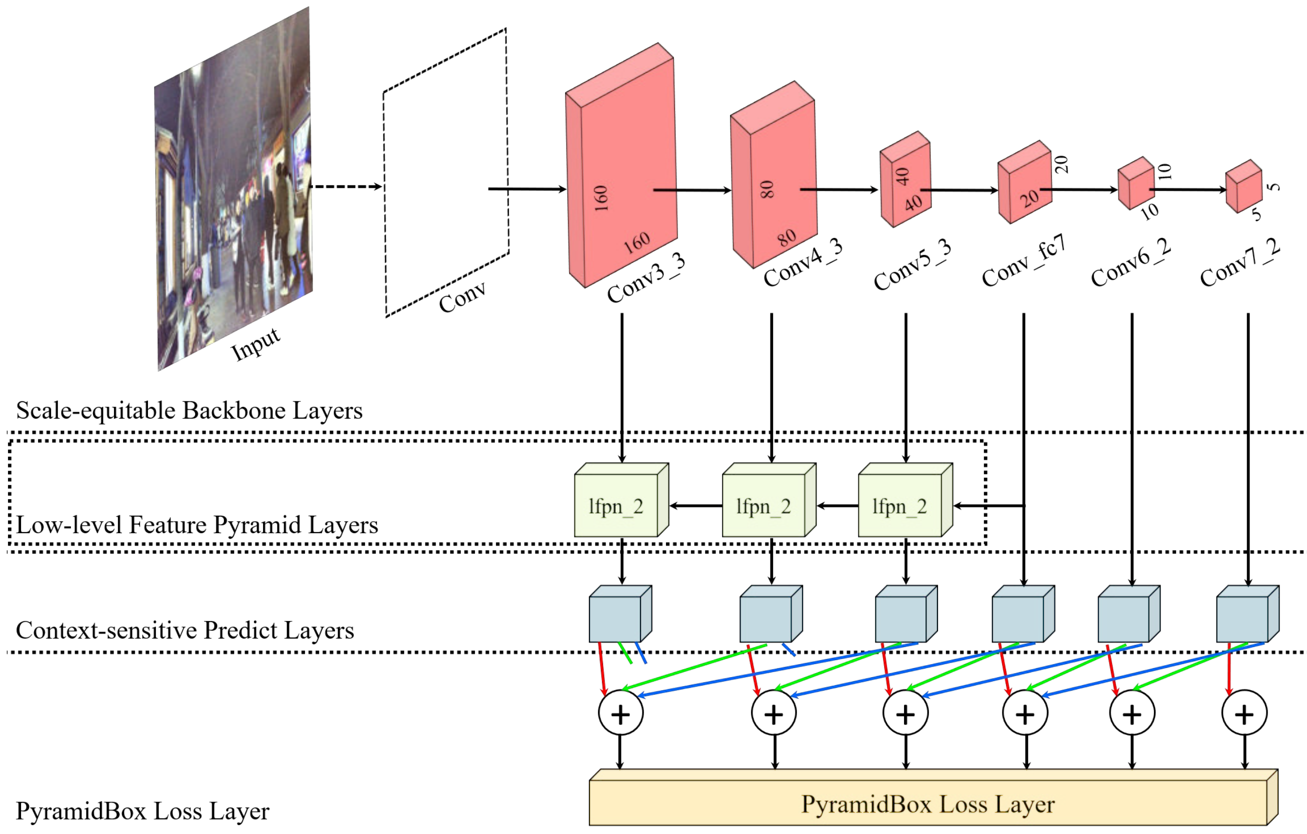


Figure 4. The architecture of the PyramidBox algorithm. It consists of Scale-equitable Backbone Layers, Low-level Feature Pyramid Layers, Context-sensitive Predict Layers and PyramidBox Loss Layer.

is the ith band of the color restoration function (CRF) in the chromaticity space, N is the number of scales, ω_n is the weight associated with the nth scale, R_{n_i} is the ith component of the nth scale and R_{MSRCR_i} is the ith spectral band of the multiscale retinex with color restoration.

The advantage of the MSRCR algorithm lies in the global enhancement of the image but there are obvious drawbacks in the application process [24], for example: the weight of each scale are fixed values without rich gradient information; the traditional retinex algorithm takes it for granted that the space illumination changes slowly, but the brightness of image always mutates in the actual scene; the Gaussian filtering does not have good performance in edge preservation. It is necessary to improve the MSRCR to adapt to low-light images collected under natural conditions. The improved MSRCR algorithm pipeline is shown in Figure 3.

We use the Gaussian filtering to obtain the rough illumination components of the original image and accurate illumination components are acquired by guided filtering. With improved Sobel edge detector, we optimize weight selection for different scales of reflectance estimation.

The rough illumination components $L_{n,i}(n \in [1, N], i \in \{R, G, B\}$ representing the three color channels) and the original image have the relationship as:

$$L_{n,i}(x, y) = S_i(x, y) * M_n(x, y) \tag{5}$$

where $M_n(x, y)$ is the surround function, which is related to the scales of Gaussian filtering.

The accurate illumination components $L'_{n,i}$, the original image S_i and the rough illumination components $L_{n,i}$ have the relationship with guided filtering as:

$$L'_{n,i}(x, y) = F_{guided}(S_i(x, y), L_{n,i}(x, y), r_n, \epsilon) \tag{6}$$

where F_{guided} is the guided filtering function, S_i is the guidance image, $L_{n,i}$ is the filtering input image, r_n is the scale and ϵ is a regularization parameter.

Then the logarithmic reflectance image of can be rewritten as:

$$R_i^*(x, y) = \Sigma_{n=1}^{N} W_{n,i} \log[S_i(x, y)] - \log[L'_{n,i}(x, y)] \tag{7}$$

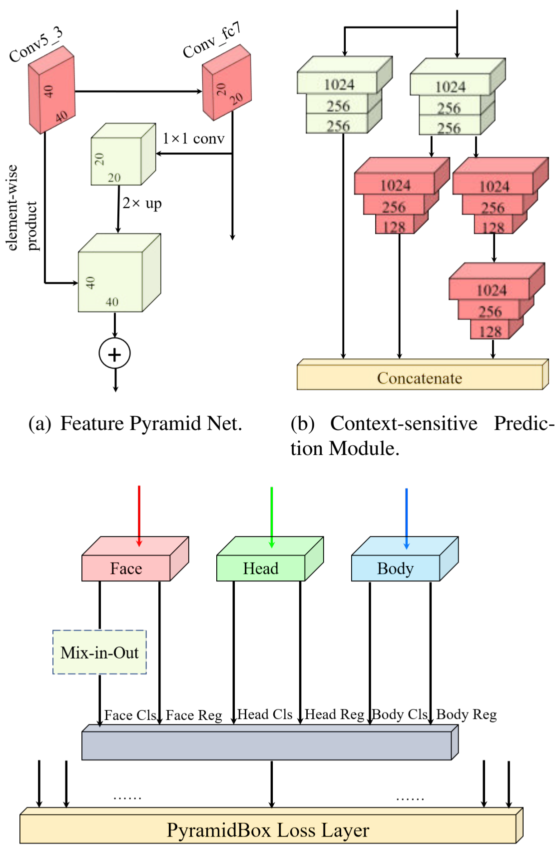
where $W_{n,i}$ is the parameter related to the Sobel edge detector. The improved MSRCR method makes up for the inherent shortcomings of the MSRCR and exerts an excellent level in the image enhancement of the DARK FACE dataset. Especially in the preservation of face details, it has advantages that other methods cannot match.

3.2. PyramidBox face detection module

Anchor-based object detection frameworks with sophisticated design of anchors have been proved effective to handle faces of variable scales [30, 22, 26, 42]. The architecture of PyramidBox, as shown in Figure 4, uses the same extended VGG16 backbone and anchor scale design as S^3FD [42], which can generate feature maps at different levels and anchors with equal-proportion interval. Low-level FPN is added on this backbone and the Context-sensitive Predict Module is used as a branch network from each pyramid detection layer to get the final output [33]. The details of each component in the architecture are as follows.

1) Scale-equitable Backbone Layers.

Use the base convolution layers and extra convolutional layers in S^3FD [42] as backbone layers, which keep layers of VGG16 from conv1_1 to pool5, then convert fc6 and fc7 of VGG16 to conv_fc layers. Then add more convolutional layers to make it deeper and the feature pyramid of the original image con be extracted in this layers.



(a) Feature Pyramid Net.

(b) Context-sensitive Prediction Module.

(c) PyramidBox Loss Layer.

Figure 5. Details in the architecture of PyramidBox algorithm.

2) Low-level Feature Pyramid Layers.

The low-level feature with highresolution plays a key role to improve the performance of face detector to handle faces of different scales. Many state-of-the-art works construct different structures in the same framework to detect faces with variant size, where the high-level features are designed to detect large faces while lowlevel features for small faces [26, 42, 40]. Faces that are small, blurred and occluded have different texture feature from the large, clear and complete ones, so high-level features play a very limited role in the small faces detection. At the same time, the noise impact of high-level features cannot be ignored, as they are usually extracted from regions with little face texture.

In the PyramidBox algorithm, by constructing the Low-level Feature Pyramid Network (LFPN) and the top-down structure from a middle layer, receptive field would be close to the half of the input size, instead of the top layer. The structure details of each LFPN block are shown in Figure 5(a).

3) Context-sensitive Predict Layers.

In order to jointly enjoy the gain of wider and deeper network [32], Context-sensitive Predict Module is the most essential part in the PyramidBox algorithm. SSH [26] increases the receptive field by placing a wider convolutional prediction module on top of layers with different strides and DSSD [6] adds residual blocks for each prediction module, as shown in Figure 5(b). In the Context-sensitive Predict Module, the PyramidBox algorithm replaced the convolution layers of context module in SSH by the residual-free prediction module of DSSD to reap all the benefits of the DSSD module approach while remaining rich contextual information from SSH context module. At the same time, it is novel to apply the max-out background label on both positive and negative samples to reduce the false positive (FP) rate of small negatives and improve the problem of small, blurred, or occluded faces that are prone to appear in images in low-light environments. The design details are shown in Figure 5(c).

4) PyramidBox Loss Layers.

There are a series of pyramid anchors to supervise the task of classification and simultaneously for each target face. In the PyramidBox Loss Layers, We determined in the experiment to use focal loss for classification and balanced L1 loss for regression. Subsequent experimental data proves that this configuration is reasonable and shows the best performance on the DARK FACE dataset.

3.3. Multi-scale test module

The size of the image will greatly affect the performance of the model. For low-light face data collected under extremely low-light conditions, neither the size of the target face nor the quality of the image can be well guaranteed. Therefore, the input of multi-scale images can very cleverly average the differences and output multiple results of models can be obtained by adjusting the input size of the image.

Soft-NMS [2] is an algorithm which decays the detection scores of all other objects as a continuous function of their overlap. For the multiple output of models, the basic idea of the Soft-NMS is used for further processing to get the final prediction result.

In the object detection task, if a bounding box has a very high overlap with detection box M, it should be assigned a very low score, while if it has a low overlap, it can maintain its original detection score. Out of this consideration, Soft-NMS propose a modification to the traditional greedy NMS algorithm [4, 34, 2]. It decrease the detection scores as an increasing function of overlap instead of setting the score to zero as in NMS. Soft-NMS can better process the overlapping detection areas, especially the detection results of multiple models in the same area, and improve the accuracy of the final result.The following pseudocode Algorithm 1 shows the procedures of the Soft-NMS.

Algorithm 1: The pipline of the Soft-NMS algorithm[2].

```
Input: B = {b_1, .., b_N}, S = {s_1, .., s_N};
B is the list of initial detection boxes;
S contains corresponding detection scores.
begin
    D ← {} ;
    while B ≠ empty do
        m ← argmax S ;
        M ← b_m;
        D ← D ∪ M;
        B ← B − M;
        for b_i in B do
            s_i ← s_i f(iou(M, b_i));
        end
    end
    return D, S;
end
```

4. Experiments

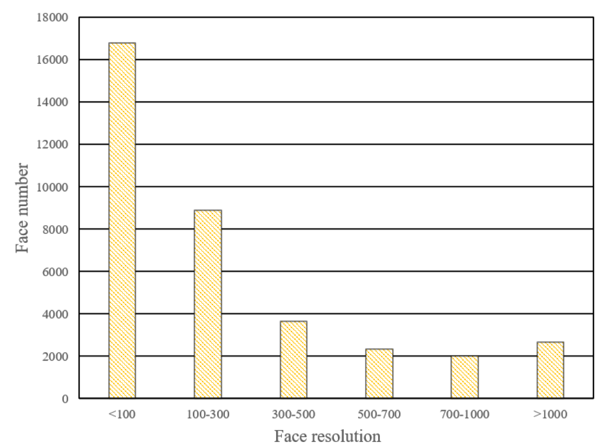
4.1. Setup

1) Dataset and metric

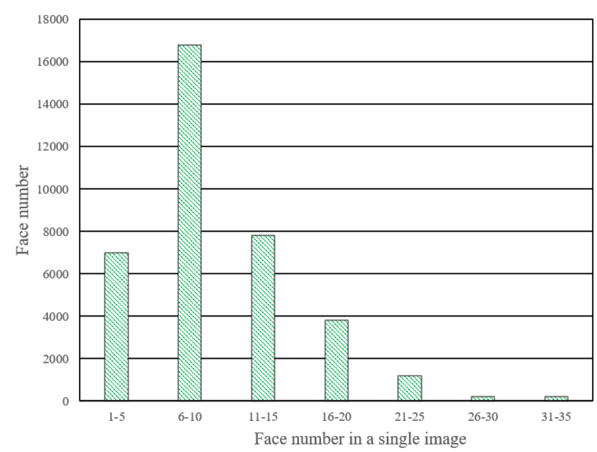
The DARK FACE dataset is the most commonly used low-light face detection dataset and we use it as our testbed. It is composed of 6,000 images taken in under-exposure condition where human faces are annotated by human with bounding boxes for training and validation, and 9,000 images taken with the same equipment in the similar environment without human annotations. It also provide a unique set of 789 paired lowlight/normal-light images captured in controllable real lighting conditions which can be used as parts of the training data. The training and evaluation set includes 43,849 annotated faces [41, 36]. Table 1 presents a summary of a few common low-light face detection image datasets.

Table 1. Comparison of low-light face detection image datasets.

Dataset	Training		Testing	
	Image	Face	Image	Face
ExDark[23]	400	-	209	-
UFDD[25]	-	-	612	-
DARK FACE	**6,000**	**4,3849**	**4,000**	**37,711**



(a) Face resolution distribution.



(b) Face number distribution.

Figure 6. Face resolution and face number distribution in training set.

For the images in the dataset, the resolution is 1,080 $\times$ 720. Figure 6 shows the distribution of the number of faces and resolution in the 6,000 images of the training and 4,000 for testing (since the original test set is withheld, we use the publicly available 100 test images as our test set). There are usually 1 to 20 annotated faces in a single image and the annotated faces have large scale variance, ranging from 1×2

to 335×296, but majority are concentrated in the small area. From the statistics, we know that the final performance of the algorithm depends largely on the detection results of small resolution face in the DARK FACE dataset.

Face detection performance is usually measured by mean Average Precision (mAP), and it is also the most important basis for us to evaluate algorithms.

2) Optimization

As for the parameter initialization, the size of the input image is 480×480, batch size is set to 8 and our PyramidBox algorithm uses the pre-trained model on WIDER FACE [39] to initialize the weights. The initial learning rate lr is set to 10^{-4} and it is reduced by 10 times every 30 epochs, with a total of 100 epochs are trained.

4.2. The effect of low-light image enhancement

We test the improved MSRCR method in the DARK FACE dataset. At the same time, the RetinexNet [36] and EnlightenGAN [14] method are used to compare their performance on mAP(0.5) indicator. The image processed by the three methods is shown in Figure 7. Through the comparison, it is easy to see that the improved MSRCR method greatly improves the clarity of the image while ensuring that the image color is not distorted. By observing the human face area, the improved MSRCR method can still exert a powerful improvement effect on faces with small resolution, which is a great help for subsequent face detection tasks under extreme conditions.

Table 2 shows the face detection mAP(0.5) obtained by using different enhancement method. The data in the table shows that the improved MSRCR method can play the best enhancement effect on the DARK FACE dataset.

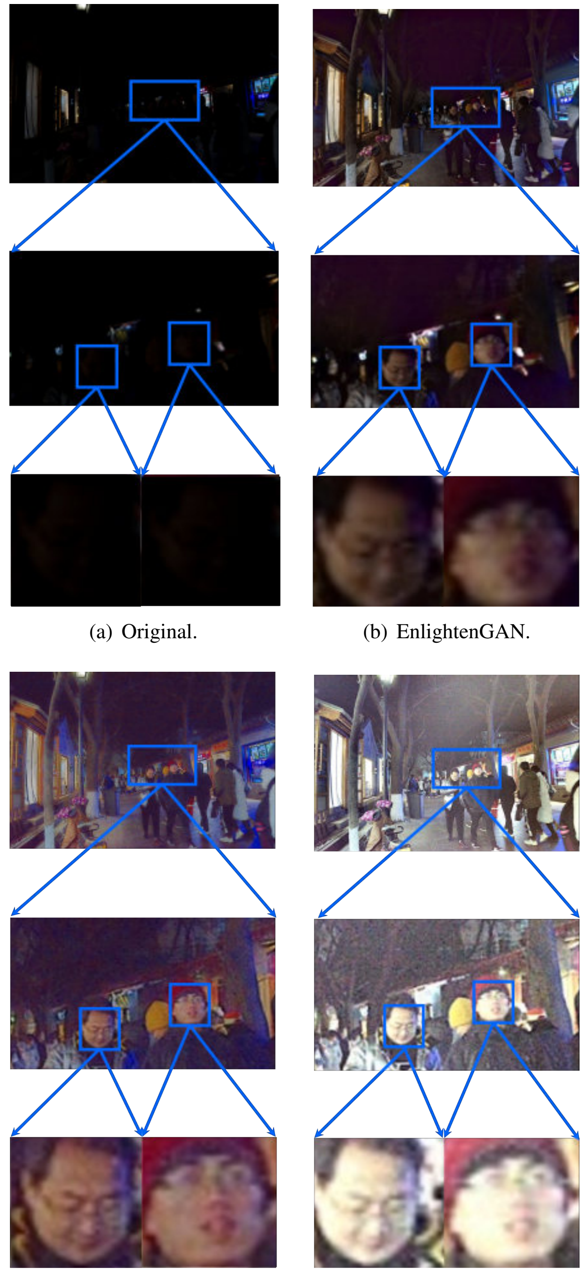
Table 2. Comparison of mAP(0.5) obtained by different low-light image enhancement methods.

Method	RetinexNet	EnlightenGAN	MSRCR	**improved MSRCR**
mAP(0.5)	69.8	72.7	75.6	**76.2**

4.3. The choice of loss functions

Different loss functions will have a great impact on the training process and results of the model. When detecting faces with very low resolution, the original loss function of the PyramidBox method can no longer guarantee the best performance of the model. For different tasks, it is necessary to test and find the better loss functions for the face detection under extremely low-light conditions.

In the original PyramidBox algorithm, cross entropy loss and smooth L1 loss are respectively the classification and regression loss functions. On the basis of preprocessing the image, we choose cross entropy loss and focal loss [21] as the loss function of the classification task, smooth L1 loss



(a) Original. (b) EnlightenGAN. (c) RetinexNet. (d) improved MSRCR.

Figure 7. Three different low-light image enhancement methods. Enlarge the concentrated area of the face to compare the effects of different methods.

and balanced L1 loss [27] as the loss function of the regression task, cross-validate and obtain the optimal model from it. Table 3 shows the test results of the four groups of experiments.

According to the data in the table, we finally choose focal loss as the loss function of the classification task and balanced L1 loss as the loss function of the regression task to ensure that the model can perform best in low-resolution

Figure 8. The final detect results. Enlarging the dense areas of faces, it can be seen that low-resolution faces can also be well detected.

Table 3. Compare the mAP(0.5) obtained by four different loss function selection methods.

Classification loss function	cross entropy loss	cross entropy loss	focal loss	**focal loss**
Regression loss function	smooth L1 loss	balanced L1 loss	smooth L1 loss	**balanced L1 loss**
mAP(0.5)	76.2	77.1	77.9	**79.2**

face detection tasks.

4.4. Multi-scale testing and Soft-NMS

There are various face sizes in the dataset. The multi-scale testing method can use images of different scales to make full use of training samples and equalize the face detection results of different scales as much as possible, so as to ensure that faces with small resolutions are used to further improve the performance of the model. The Soft-NMS method [2] is used to integrate the test results of different scales in the most reasonable way, and finally get the output results we need.

Table 4. Comparison of mAP(0.5) obtained by different scale selection.

Scale selection	0.5-1-2	1-1.5-2	**2-2.5-3**	2-2.3-2.5 -2.7-3	2-2.5 -2.75-3
mAP(0.5)	78.0	77.9	**82.3**	81.8	82.0

It can be seen in Table 4 that if different scales are selected for testing, the results of mAP(0.5) are also different. In experiments, we determined the optimal scale choice and obtained the final detection results on the DARK FACE dataset. Part of the result is shown in Figure 8.

The leaderboard is illustrated in Table 5, where we show top 3 contestants in the UG2+ 2021 Track1.2 - Face Detection in the Low-Light Condition. The result proves the effectiveness of our method again.

Table 5. Leaderboard of UG2+ 2021 Track1.2 - Face Detection in the Low-Light Condition in the test stage.

Team	TAL-ai	DeepBlueAI	New Horizons	**Ours**
mAP(0.5)	74.89	71.65	69.71	**82.3**

5. Conclusion

This paper proposes an advanced face detection method under extremely low-light conditions. The highlights are mainly reflected in: the use of improved MSRCR method to enhance low-light face images, use the PyramidBox method for face detection with the innovative loss functions, multi-scale testing and use the Soft-NMS method to complete the results integration. Integrating these techniques, this paper achieved high accuracy and obtained excellent results in the face detection task of the DARK FACE dataset.

6. Acknowledgement

This work was supported by the Anhui Province Key Research and Development Program (202104a05020007) and USTC Research Funds of the Double First-Class Initiative (YD2350002001).

References

[1] Adrian Barbu, Nathan Lay, and Gary Gramajo. Chapter 6 - face detection with a 3d model. In *Academic Press Library in Signal Processing, Volume 6*, pages 237–259. 2018. 3

[2] Navaneeth Bodla, Bharat Singh, Rama Chellappa, and Larry S. Davis. Soft-nms — improving object detection with one line of code. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 5562–5570, 2017. 6, 8

[3] S. Charles Brubaker, Jianxin Wu, Jie Sun, Matthew D. Mullin, and James M. Rehg. On the design of cascades of boosted ensembles for face detection. 77(1–3):65–86, 2008. 3

[4] Navneet Dalal and Bill Triggs. Histograms of oriented gradients for human detection. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1, pages 886–893 vol. 1, 2005. 6

[5] Guodong Ding, Shanshan Zhang, Salman Khan, Zhenmin Tang, Jian Zhang, and Fatih Porikli. Feature affinity-based pseudo labeling for semi-supervised person re-identification. *IEEE Transactions on Multimedia*, 21(11):2891–2902, 2019. 1

[6] Chengyang Fu, Wei Liu, Ananth Ranga, Ambrish Tyagi, and Alexander C. Berg. Dssd: Deconvolutional single shot detector. *CoRR*, abs/1701.06659, 2017. 5

[7] Xueyang Fu, Delu Zeng, Yue Huang, Yinghao Liao, Xinghao Ding, and John Paisley. A fusion-based enhancing method for weakly illuminated images. *Signal Processing*, 129:82–96, 2016. 3

[8] Xueyang Fu, Delu Zeng, Yue Huang, X.-P Zhang, and Xinghao Ding. A weighted variational model for simultaneous reflectance and illumination estimation. pages 2782–2790, 06 2016. 3

[9] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 580–587, 2014. 3

[10] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Region-based convolutional networks for accurate object detection and segmentation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(1):142–158, 2016. 3

[11] Xiaojie Guo, Yu Li, and Haibin Ling. Lime: Low-light image enhancement via illumination map estimation. *IEEE Transactions on Image Processing*, 26(2):982–993, 2017. 3

[12] Kaiming He, Jian Sun, and Xiaoou Tang. Single image haze removal using dark channel prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(12):2341–2353, 2011. 3

[13] Vidit Jain and E. Learned-Miller. Fddb: A benchmark for face detection in unconstrained settings. 2010. 1

[14] Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, and Zhangyang Wang. Enlightengan: Deep light enhancement without paired supervision. *IEEE Transactions on Image Processing*, 30:2340–2349, 2021. 7

[15] Daniel J. Jobson, Zia ur Rahman, and Glenn A. Woodell. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE Transactions on Image Processing*, 6(7):965–976, 1997. 2, 3

[16] Brendan F. Klare, Ben Klein, Emma Taborsky, Austin Blanton, Jordan Cheney, Kristen Allen, Patrick Grother, Alan Mah, Mark Burge, and Anil K. Jain. Pushing the frontiers of unconstrained face detection and recognition: Iarpa janus benchmark a. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1931–1939, 2015. 1

[17] Edwin H. Land. An alternative technique for the computation of the designator in the retinex theory of color vision. *Proceedings of the National Academy of Sciences of the United States of America*, 83(10):3078–3080, 1986. 2, 3

[18] Jian Li, Yabiao Wang, Changan Wang, Ying Tai, Jianjun Qian, Jian Yang, Chengjie Wang, Jilin Li, and Feiyue Huang. Dsfd: Dual shot face detector. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5055–5064, 2019. 2

[19] Mading Li, Jiaying Liu, Wenhan Yang, Xiaoyan Sun, and Zongming Guo. Structure-revealing low-light image enhancement via robust retinex model. *IEEE Transactions on Image Processing*, 27(6):2828–2841, 2018. 3

[20] Jinxiu Liang, Jingwen Wang, Yuhui Quan, Tianyi Chen, Jiaying Liu, Haibin Ling, and Yong Xu. Recurrent exposure generation for low-light face detection, 07 2020. 2

[21] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollar. Focal loss for dense object detection. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 2999–3007, 2017. 3, 7

[22] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C. Berg. Ssd: Single shot multibox detector. In *Computer Vision – ECCV 2016*, pages 21–37. Springer International Publishing, 2016. 3, 5

[23] Yuen Peng Loh and Chee Seng Chan. Getting to know low-light images with the exclusively dark dataset. *Computer Vision and Image Understanding*, 178:30–42, 2019. 6

[24] Jinxiang Ma, Xin Fan, Jianjun Ni, Xifang Zhu, and Chao Xiong. Multi-scale retinex with color restoration image enhancement based on gaussian filtering and guided filtering. *International Journal of Modern Physics B*, 31:1744077, 05 2017. 2, 4

[25] Hajime Nada, Vishwanath Sindagi, He Zhang, and Vishal M. Patel. Pushing the limits of unconstrained face detection: a challenge dataset and baseline results. *arXiv preprint arXiv:1804.10275*, 2018. 6

[26] Mahyar Najibi, Pouya Samangouei, Rama Chellappa, and Larry S. Davis. Ssh: Single stage headless face detector. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 4885–4894, 2017. 3, 5

[27] Jiangmiao Pang, Kai Chen, Jianping Shi, Huajun Feng, Wanli Ouyang, and Dahua Lin. Libra r-cnn: Towards balanced learning for object detection, 04 2019. 3, 7

[28] Minh-Tri Pham and Tat-Jen Cham. Fast training and selection of haar features using statistics in boosting-based face detection. In *2007 IEEE 11th International Conference on Computer Vision*, pages 1–7, 2007. 3

[29] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 779–788, 2016. 3

[30] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 39(6):1137–1149, 2017. 5

[31] Sebastian Salazar-Colores, Juan-Manuel Ramos-Arreguin, Jesus-Carlos Pedraza-Ortega, and J Rodriguez-Resendiz. Efficient single image dehazing by modifying the dark channel prior. *J Image Video Proc*, 66, 2019. 3

[32] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander A. Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. In *AAAI*, 2017. 5

[33] Xu Tang, Daniel K. Du, Zeqiang He, and Jingtuo Liu. Pyramidbox: A context-assisted single shot face detector. In *Computer Vision – ECCV 2018*, pages 812–828, Cham, 2018. Springer International Publishing. 2, 3, 5

[34] Paul Viola and Michael Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, volume 1, pages I–I, 2001. 6

[35] Paul Viola and Michael Jones. Robust real-time face detection. In *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, volume 2, pages 747–747, 2001. 1, 3

[36] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. Deep retinex decomposition for low-light enhancement. In *British Machine Vision Conference*, 2018. 2, 6, 7

[37] Bin Xie, Fan Guo, and Zixing Cai. Improved single image dehazing using dark channel prior and multi-scale retinex. In *2010 International Conference on Intelligent System Design and Engineering Application*, volume 1, pages 848–851, 2010. 2

[38] Bin Yang, Junjie Yan, Zhen Lei, and Stan Z. Li. Aggregate channel features for multi-view face detection. In *IEEE International Joint Conference on Biometrics*, pages 1–8, 2014. 3

[39] Shuo Yang, Ping Luo, Chen Change Loy, and Xiaoou Tang. Wider face: A face detection benchmark. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 5525–5533, 2016. 2, 7

[40] Shuo Yang, Yuanjun Xiong, Chen Change Loy, and Xiaoou Tang. Face detection through scale-friendly deep convolutional networks. 06 2017. 5

[41] Wenhan Yang, Ye Yuan, Wenqi Ren, Jiaying Liu, Walter J. Scheirer, and Zhangyang Wang. Advancing image understanding in poor visibility environments: A collective benchmark study. *IEEE Transactions on Image Processing*, 29:5737–5752, 2020. 2, 6

[42] Shifeng Zhang, Xiangyu Zhu, Zhen Lei, Hailin Shi, Xiaobo Wang, and Stan Z. Li. S3fd: Single shot scale-invariant face detector. In *2017 IEEE International Conference on Computer Vision (ICCV)*, pages 192–201, 2017. 3, 5

[43] Jian Zhao, Lin Xiong, Yu Cheng, Yi Cheng, Jianshu Li, Li Zhou, Yan Xu, Jayashree Karlekar, Sugiri Pranata, Shengmei Shen, Junliang Xing, Shuicheng Yan, and Jiashi Feng. 3d-aided deep pose-invariant face recognition. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence*, pages 1184–1190. International Joint Conferences on Artificial Intelligence Organization, 7 2018. 1

[44] Qiang Zhu, Mei-Chen Yeh, Kwang-Ting Cheng, and S. Avidan. Fast human detection using a cascade of histograms of oriented gradients. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 2, pages 1491–1498, 2006. 3