FACIAL LANDMARKS DETECTION AND EXPRESSION RECOGNITION IN THE DARK

A Thesis

by

QIYU WANG

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

MASTER OF SCIENCE

Chair of Committee,	Anxiao Jiang
Committee Members,	Jiang Hu
	Tim Davis
Head of Department,	Miroslav Begovic

December 2020

Major Subject: Computer Engineering

ABSTRACT

Facial landmark detection has been widely adopted for body language analysis and facial identification task. A variety of facial landmark detectors have been proposed in different approaches, such as AAM, AdaBoost, LBF and DPM. However, most detectors were trained and tested on high resolution images with controlled environments. Recent study has focused on robust landmark detectors and obtained increasing excellent performance under different poses and light conditions. However, it remains an open question about implementing facial landmark detection in extremely dark images. Our implementation is to build an application for facial expression analysis in extremely dark environments by landmarks. To address this problem, we explored different dark image enhancement methods to facilitate landmark detection. And we designed landmark correctness methods to evaluate landmarks' localization. This step guarantees the accuracy of expression recognition. Then, we analyzed the feature extraction methods, such as HOG, polar coordinate and landmarks' distance, and normalization methods for facial expression recognition. Compared with the existing facial expression recognition system, our system is more robust in the dark environment, and performs very well in detecting happy and surprising.

ACKNOWLEDGMENTS

I would like to express my gratitude to my research advisor Professor Anxiao Jiang for his guidance and support throughout the course of this research.

I would also like to appreciate Professor Jiang Hu and Professor Tim Davis for their help and being my committee members.

Thanks also to all my colleagues and the faculty and staff of the Department of ECEN, the Department of CSCE and Texas A&M University for their help throughout my Master of Science program.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supported by a thesis committee consisting of Professor Anxiao Jiang of the Department of ECEN & CSCE, Professor Hu Jiang of the Department of ECEN & CSCE and Professor Tim Davis of the Department of CSCE.

The Karolinska Directed Emotional Faces (KDEF) dataset was provided by Daniel Lundqvist and his colleagues. The Japanese Female Facial Expression (JAFFE) Dataset was offered by Michael Lyons, Miyuki Kamachi, and Jiro Gyoba. The CK/CK+ dataset was provided by Jeffrey Cohn and Tian Y.

OpenCV landmark detection API was provided by Laksono Kurnianggoro.

Funding Sources

Graduate study was supported by the fellowship from Texas A&M University.

NOMENCLATURE

KDEF	The Karolinska Directed Emotional Faces
NMF	Non-negative Matrix Factorization
HDR	High dynamic range
LIME	Low-Light Image Enhancement
GAN	Generative Adversarial Network
WGAN	Wasserstein GAN
HOG	Histogram Oriented Gradients
AAM	Active Appearance Model
LBF	Local Binary Feature
SVM	Support Vector Machine
RCPR	Robust Cascaded Pose Regression
GSF^2EFR	Geometric Shaped Facial Feature Extraction for Face Recognition
SIFT	Scale Invariant Feature Transform
MLP	Multilayer Perceptron

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

1. INTRODUCTION

It has been suggested that body language constitutes more than 60 percent of what people communicate, learning to understand body languages has great potentials in many areas such as intelligent human-computer interaction, emotion recognition and behavior prediction. With the achievements in object detection and tracking, human body language analysis has attracted a lot of attentions, especially facial expression analysis. Facial expressions convey the emotional state and physical sensations of an individual. There have been many applications, such as patient monitoring system and computer entertainments, which analyzed human facial expressions. Due to the complexity of facial expressions, especially those inappreciable micro-expressions, researchers designed different patterns to represent human facial features. While facial expression recognition has been implemented in a multitude of approaches, such as LBF(local binary pattern)[10], non-negative matrix factorization(NMF)[53] and sparse learning[54], we noticed the recent achievements in facial landmark detection could provide an alternative method for facial expression analysis. Since most of existing facial landmark detectors are based on 2D image input, they suffer from variant environments, especially illumination. This problem is inevitable in practical application environments.

Our study is to build an application for facial expression recognition in extremely dark environment. This implementation can be categorized into four parts: dark image processing, landmark detection, feature extraction and facial expression detection. On each step, we utilized cutting-edge techniques to boost the performance. The biggest barrier for our study is the limitation of dataset. There is no existing dataset with extremely dark images with annotated facial landmarks or expressions. This is because it's difficult for the professional annotators to annotate on extremely dark image. Due to the limitation of dataset, we didn't design a new facial landmark detection model, but utilized dark image enhancement techniques. We compared different landmark detectors' performance in the processed images and added correctness methods to promote accuracy. We leveraged different patterns for facial landmarks, and trained a machine learning model to

predict facial expressions.

2. RELATED WORK

2.1 Dark Image Enhancement

Commonly, images captured in dark environments challenge computer vision techniques due to low contrast and high ISO noise. To address this problem, many dark image enhancement approaches have been proposed. Contrast stretching and histogram equalization [3] are two basic enhancement techniques to dark images. These two methods enlarge the contrast by redistributing intensities in a larger histogram range. [39] proposes adaptive histogram equalization to overcome the challenge of unbalanced enhancement in different areas. Several histograms corresponding to different sections of the image are computed and used to redistribute the illumination.

Retinex[29][35][38] is the theory of human color vision proposed by Edwin Land to account for color sensations. The dark images are decomposed to reflection map and illumination map. Reflection component preserves the texture and illumination map preserves the lightness information. [16] imposes regulation terms to Retinex to stabilize the output illumination. Low-light image enhancement(LIME)[20] further proposes a structure prior to the illumination map as the reference for enhancement.

Recently, deep learning has been widely used in computer vision. It has been proved successful in many areas, such as super-resolution[14], denoising[28][45] and deblurring[2]. Also, many learning-based approaches were proposed for dark image enhancement. [44] utilizes convolutional neural network[27] to estimate the transmission rate in different areas of the image. LL-Net[32] designed a deep auto encoder to learn joint denoising and enhancement on the patchs. Retinex-Net[47] coordinates the Retinex theory[29] and deep neural networks. It uses the neural network for reflection/illumination decomposition.

More recently, GANs[18] have accomplished great achievements in image synthesis[21] and segmentation[52]. Retinex-GAN[42] decomposed paired dark-bright images into reflection map and illumination map by a neural network, and GAN was responsible to generate a new illumi-

nation map and construct a new image. EnlightenGAN[22] achieves impressive visual results with unpaired training images. EnlightenGAN extracts content features by VGG, and adds self-attention to generator to constrict texture features. EnlightenGAN randomly selected 5 patches and implemented sub-discriminators to handle variant enhancement rates in different areas. [11] introduced Wasserstein GAN(WGAN)[4] and individual batch normalization to CycleGAN for image enhancement.

2.2 Facial Landmark Detection

The goal of facial landmark detection is to detect key-points in human faces. The first step of facial landmark detection is to localize face regions in the image. Then, predict facial landmarks in the face region. There are three major approaches for facial landmark detection: HOG [24], ASM[30] and LBF[10]. HOG divides image into numbers of blocks. Gradients are calculated within an image per block. Constituted a pixel map by the magnitude and direction of change in the pixels within the block. After extracting features by HOG, trains SVM model to predict the location of landmarks. ASM has a template for facial landmarks. First, suggest a tentative shape by adjusting the shape points' location. Second, conform the tentative shape to a global shape model. Repeat these two steps. LBF learns a feature mapping function to generate local binary features. Given the features and target shapes, update the mapping function by learning regression.

Although the lab-controlled landmark detection systems have achieved high accuracy, the real-world appliance faces many challenges. [7] and [50] improve landmark localization under occlusion by multi-pose regression. [36] efficiently constructs strong representations to disentangle highly nonlinear relationships between images and shapes. Style-Aggregated Network [15] was proposed to generate robust landmarks in variant image styles. [49] utilizes boundary lines as the geometric structure of a human face to help facial landmark localization. FAN[6] explores facial landmark detection to 3D.

2.3 Facial Feature Extraction

Facial feature extraction is the process of extracting face component features, such as eyes, nose, lips, etc, from face image. Facial feature extraction plays an important role in face recognition, face tracking and expression analysis. [48] extracts the biometric features of the face and the K-mean method is used to cluster the face features. Scale Invariant Feature Transform (SIFT)[19][1] has sparingly been used in 3D face recognition. Recently, [5] focuses on eyes detection and designs Geometric Shaped Facial Feature Extraction for Face Recognition (GSF^2EFR) to identify the person by finding the center and corners of the eye using eye detection and eye localization modules.

2.4 Facial Expression Recognition

Facial expression recognition systems can be widely applied to various research areas. [25] proposed a simple and effective CNN to extract facial expression features. [37] proposed learning-based approach to build the saliency map for each emotion respectively. [41] designed a frame pattern for facial landmarks and used it to extract facial features and predict facial expressions. In order to reduce errors, [41] used SVD to extract principle components for each feature respectively and designed a pattern to eliminate noise. Khan[26] proposed a framework to detect facial expressions based on landmarks, but didn't investigate in feature normalization methods.

3. METHODOLOGY

Our objective is to build an application for facial expression recognition in the dark. As far as we know, there is no existing facial landmark detection model performing well in very dark environments. SAN[15] proposes a robust landmark detection model in multi-style images, but it can't handle extremely dark images well. Our first approach was to retrain a landmark detector to overcome the challenge of lack of illumination. However, we didn't find existing dataset with people in very dark image with both annotated landmarks and expression. SAN utilized PhotoShop to convert annotated face images into a variety of styles and created a paired dataset, but we tended to use natural dark images, since landmark detection model might learn to remove artificial features, instead of handing variant dark environments. Due to lack of dataset that fulfills our requirements, we decided not to retrain a landmark detection model by artificial en-darking images, but further looked into how factors affects facial landmark detection and introduce image enhancement technique to facilitate landmark detection.

3.1 Dark Image Process

3.1.1 Landmark Detectors in Dark

Due to lack of illumination, most pixels of a dark image are in a narrow low-intensity range and landmark detection models can't perform very well. Before we process the input dark image, we tested facial landmark detectors in the dark.

There is no existing dataset with face in the dark environments. Therefore, we recorded three video clips. In each video clip, a person presents different expressions with random occlusion in the dark environment. We set a basic illumination levels for each clip individually. During each video clip, the lightness is adjusted subtly. We extracted one frame every 5 seconds.

Table 3.1 shows the affects of illumination to three basic and popular landmark detectors, openpose(CNN based)[8] landmark detector, dlib landmark detector(HOG based)[51] and OpenCV LBF landmark detector(LBF based)[10]. Extract 90 frames from dark videos with very low light,

78 with medium low light and 85 from lightly low light. While the lightness decreases, the rate of frames that can be detected keeps decreasing.

Lightness Level	Image	openpose Landmark Detector	dlib Landmark Detector	OpenCV LBF Landmark Detector
1		37/90	49/90	18/90
2		67/78	67/78	48/90
3		81/85	85/85	78/90

Table 3.1: Detecting Rate in 3 Levels

3.1.2 Contrast Stretching and Histogram Equalization

An intuitive method to help landmark detection is to stretch the contrast and enlarge intensities difference. Contrast Stretching simply 'stretches' the range of intensity to a larger range. The average value of R,G,B channels is regarded as the intensity.

$$I_{out} = (I_{in} - I_{lowest})\frac{255}{I_{highest} - I_{lowest}} + I_{lowest}$$

I_{out} and I_{in} represent the output intensity and input intensity. I_{lowest} and $I_{highest}$ are the minimum intensity and the maximum intensity of all pixels in the input image.

In dark images, pixels' intensities intend to cluster in a small range, histogram equalization stretches the distribution of intensities globally. Adaptive histogram equalization considers the

variants in different parts of the image. By dividing the image into some small blocks, Adaptive histogram equalization stretches the intensities in every block instead of the entire image.

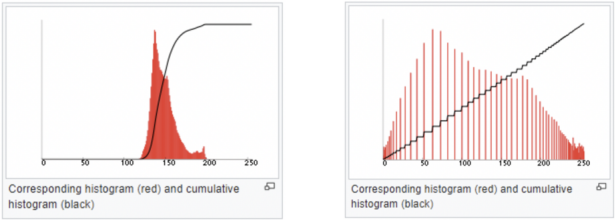


Figure 3.1: Histogram Equalization[43]

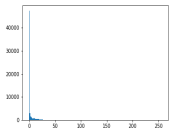
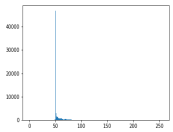
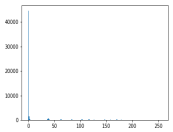
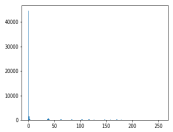
	Dark Image	Contrast Stretching	Histogram Equalization	Adaptive Histogram Equalization
Image				
Histogram				
dlib's Face Landmarks	49/90	56/90	72/90	70/90

Table 3.2: Landmark Detection Comparison.

From our observation, histogram equalization and adaptive histogram equalization are helpful to landmark detection, but their improvements are limited.

According to Retinex theory, a given image can be decomposed into two different maps: the reflection map R and the illumination map L. The reflection map R preserves the texture of the reflective object and the illumination map L determines the brightness. We selected the easiest

way for decomposition. Since the frequency of L is much higher than the frequency of R, we used a low-pass filter to remove L. Then, applied a new illumination map with higher brightness to the reflected map R and constructed the new image. Furthermore, implementing different scales of Guassian filters could maintain high image fidelity and compress the dynamic range of the image.

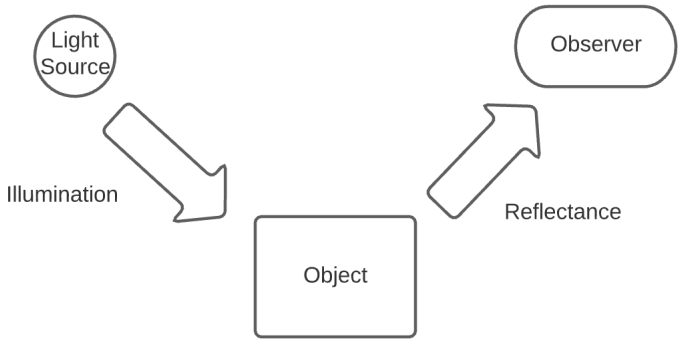


Figure 3.2: Retinex Theory

3.1.3 Learning-based Dark Image Enhancement

Because of lack of paired database, we designed a CycleGAN-based dark enhancement model. Besides enlighteness, this model should be able to reduce noise spots on the output image and sharpen the edges. Inspired by histogram equalization, we added edge histogram loss to remove haze and blur on the output image. To preserve the textures, we leveraged VGG to extract texture features of the input and output image and restrict texture loss.

We assembled an unpaired dataset with 1000 low light images and 1000 high light images. Those images were random selected from Exdark [31] and LOL[46] datasets.

We used U-Net for generator and PatchGAN for discriminator. Our U-Net generator is implemented with 10 convolutional blocks. At the downsampling stage, each block consists of two 4×4 convolutional layers, followed by LeakyReLu and a batch normalization layer. At the upsampling

stage, we replaced the standard deconvolutional layer by one upsampling layer plus one convolutional layer to reduce the checkerboard artifacts. VGG16 was used to extract texture features, and wasserstein distance was used for edge histogram loss.

Our model is trained for 50 epochs. We use the Adam optimizer with the learning rate of 1e-4 and the batch size is set to be 32.

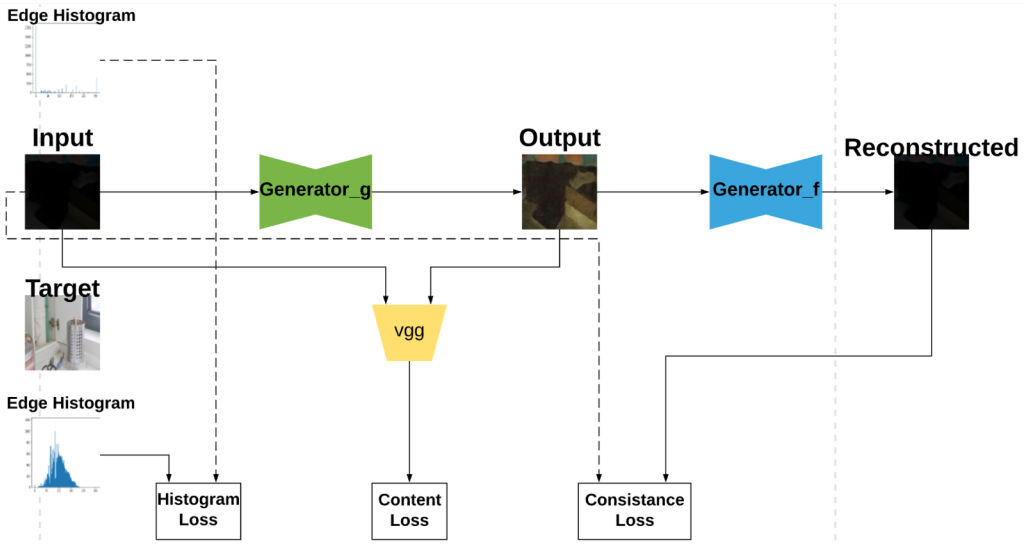


Figure 3.3: Architecture of my CycleGAN-based model

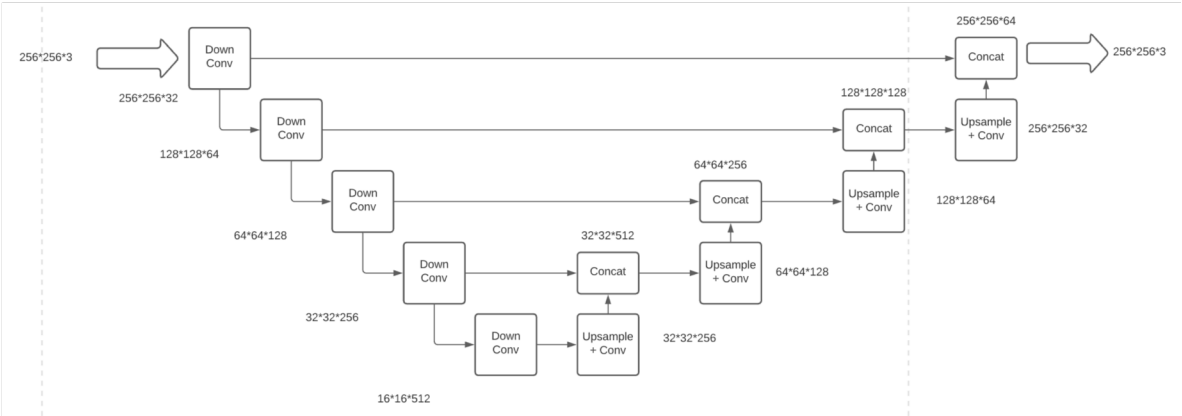


Figure 3.4: Architecture of my CycleGAN Generator

	Input	Retinex	LIME	RetinexNet	EnlightenGAN	My Model
Image						
dlib	49/90	81/90	65/90	78/90	79/90	73/90

Table 3.3: Comparison of Different Enhancement Approaches.

According to the results above, Retinex, RetinexNet and EnlightenGAN have the best results for landmark detection. LIME tends to bring disorder on the image, RetinexNet blurs the image, EnlightenGAN has the best visual effects. My model tends the over-enlighten the image, but eliminate noise spots than others. This is not as helpful to landmark detection as expected, but will help landmark correction.

3.2 Landmark Detection

After dark image enhancement, we need to process the enhanced image to facilitate landmark detection.

- Convert RGB image to gray scale.
- Implement fast NL-Means method to remove noise spots.
- Sharpen the edges.

From our observation, color provides no useful information for landmark detection. The only information landmark detectors needed is the intensity and gradient. So gray-scale image is easier for landmark detector to process. NL-Means method can remove reduce ISO noise effectively. Sharper edges can facilitate landmark detection a lot. The first step of facial landmark detection is face localization, then predict the locations of landmarks in the facial area.

Due to the noise brought from darkness, the face detection is quite unstable. A common problem is it probably locates the face in the incorrect region. Therefore, I added the skin detector

(a) Before preprocess (b) After Preprocess

Figure 3.5: Image Processing

to avoid this kind of error. After the face detector finds the face area on the image, the skin detector will discriminate all pixels from this area using skin detector. If more than half pixels are discriminated as skin, we discriminate this area as face.

Below is the conditional statements to discriminate skin and not-skin:[9]

$$R > 80 \quad \text{and} \quad G > 40 \quad \text{and} \quad B > 20$$

$$maxR, G, B - minR, G, B > 15$$

$$|R - G| > 15$$

$$R > G \quad \text{and} \quad R > B$$

And from our test, these statements can detect the main skin colors in Figure 3.5. We observed there is still a problem about how to detect dark skin in very dark environment, but our detector has covered a large range of skin colors.

Skin detector is helpful to avoid incorrect face detection failure cases in Table 3.5: incorrect location in the dark environment, mask or painting.

From our observation, the locations of landmarks are not convincing in some cases. To further stabilize the location of landmarks, we applied landmark correctness, mainly for eyebrows and lips.

Color	HEX	RGB
	#c58c85	rgb(197, 140, 133)
	#ecbcb4	rgb(236, 188, 180)
	#d1a3a4	rgb(209, 163, 164)
	#a1665e	rgb(161, 102, 94)
	#503335	rgb(80, 51, 53)
	#592f2a	rgb(89, 47, 42)

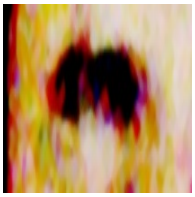
Figure 3.6: Main Skin Colors[12]

(a) Location Error (b) Color Error (c) Stretch Error

Figure 3.7: Face Detection Failure Cases

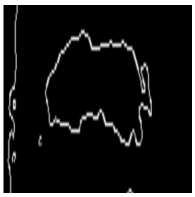
The facial landmark correction has four steps:

- Extract facial feature regions by landmarks.
- Smooth Edges by Gaussion filter.
- Utilized k-means to obtain sharp edges.
- Calculate the confidence of each landmark by the distance from the landmark to the edges.



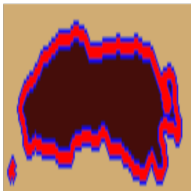
(a) Eyebrow

(b) K-means Classification

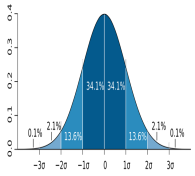


(c) Edge Detection

Figure 3.8: Edge Detecting



(a) Eyebrow Region



(b) Confidence Distribution[13]

Figure 3.9: Confidence Distribution.

For each landmark, use the 1/10 of the distance from the landmark to its adjacent landmark as σ. X coordinate is the distance from the landmark to the edge, Y is the confidence of the landmark. If the distance between one landmark and the edge exceeds 2σ, we discriminate it as failure. If the average distance for all lanmarks to the edges exceeds σ, the detection of this facial image is failure. In Table 3.7 (a), the red region is where the landmarks locate at [-σ, σ], the blue region is where the landmarks locate at [-2σ, -σ] and [σ, 2σ].

3.3 Feature Extraction

It's unnecessary to consider all 68 landmarks, for example, the leftmost landmark and the right landmark seems no movement in all emotions. The Facial Action Coding System(FACS)[40] was introduced by Carl-Herman Hjortsjö in 1970 to represent the relation between facial muscle movements and emotions. Then, it was subsequently developed further by Paul Ekman, and Wallace Friesen. Referring to FACS, I selected left eye, left eyebrow, right eye, right eyebrow, nose, lip and jaw are generally to describe a human face expression. Figure 3.7 shows the selected points. There are different feature extraction methods for facial landmarks, we examined HOG feature

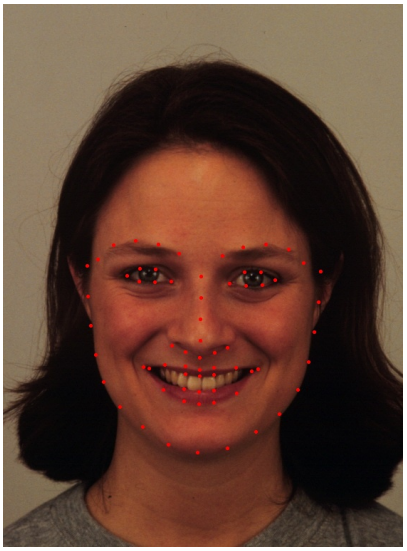


Figure 3.10: Detected Landmarks

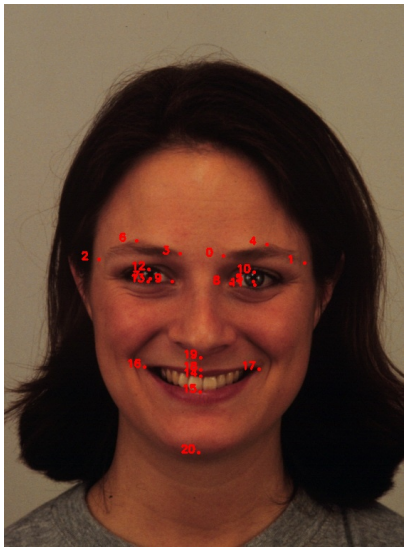


Figure 3.11: Selected Landmarks

descriptor, polar coordinate and points distance.

3.3.1 HOG Feature Descriptor

HOG splits the image into many small blocks, and uses the histogram of gradient of each block as the feature of that block respectively. we used the HOG the blocks where the selected landmarks are located at as the features. But later, HOG features were found too subtle to reflect useful information for facial expressions.

3.3.2 Polar Coordinate

Use the vector starting at the nose tip and ending at the eyebrow center as the polar axis. Construct 20 feature vectors starting from the nose tip and ends at other landmarks. Each feature vector is represented by two parameters, magnitude and angle.

3.3.3 Points Distances

We consider the distances between all distinct pairs of landmark points. Some pairs, like the inner corner points of two eyes, contains no useful information. But some pairs, like the point on the left corner of lips and the right corner of lips, are very helpful for expression recognition. We

Figure 3.12: Polar Coordinate.

selected 14 feature distances in total.

3.3.4 Normalization

Due to the variant of face shapes and distance from camera to face, normalization is necessary. We used the distance from the nose lip to the eyebrows' center as the normalization distance. All distances are normalized by this distance.

Before we trained machine learning model. All features must be normalized to [0, 1]. Suppose the feature vector $f = (f_1, f_2, f_3, f_4 ... f_n)$, normalize it to $\widetilde{f} = (\widetilde{f_1}, \widetilde{f_2}, \widetilde{f_3}, \widetilde{f_4}, ... \widetilde{f_n})$ by:

$\widetilde{f_i} = \frac{((f_i - \mu_i)/2\sigma_i)+1}{2}$, i = 1, 2, 3, ..., n,

Name	Distance
Left Eyebrow Length	Dis(0, 1)
Right Eyebrow Length	Dis(2, 3)
Left Outer Eyebrow Height	Dis(4, 10)
Right Outer Eyebrow Height	Dis(6, 12)
Left Inner Eyebrow Height	Dis(0, 8)
Right Inner Eyebrow Height	Dis(3, 9)
Left Eye Height	Dis(10, 11)
Right Eye Height	Dis(12, 13)
Mouth Height	Dis(14, 15)
Mouth Width	Dis(16, 17)
Left Lip Height	Dis(5, 17)
Right Lip Height	Dis(7, 16)
Upper Lip Height	Dis(18, 19)
Jaw Height	Dis(20, 19)

Table 3.4: Selected Features.

Name	Distance
Eyes Center	mid(5, 7)
Norm Distance	Dis(19, mid(5, 7))

Table 3.5: Selected distance for normalization.

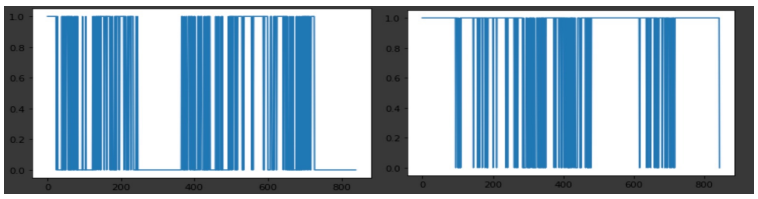
where μ_i and σ_i are mean and standard deviation of the ith feature across the training data. This normalization methods can guarantee 98% of features are located in [0,1], and we crop the features that exceed this range to 0 or 1.

3.4 Facial Expressions Recognition

Since the selected feature is simple, we trained poly SVM, Random Forest Tree and MLP model for expression recognition.

4.1 Image Processing

To analyze the performance of image processing, we took a dark video record in 60 seconds with 827 frames. Table 4.1 compares landmark detection before and after processing. The rate of detection has been promoted significantly.



(a) 322/827 (b) 724/827

Figure 4.1: Facial Landmark Detection on Video Sample

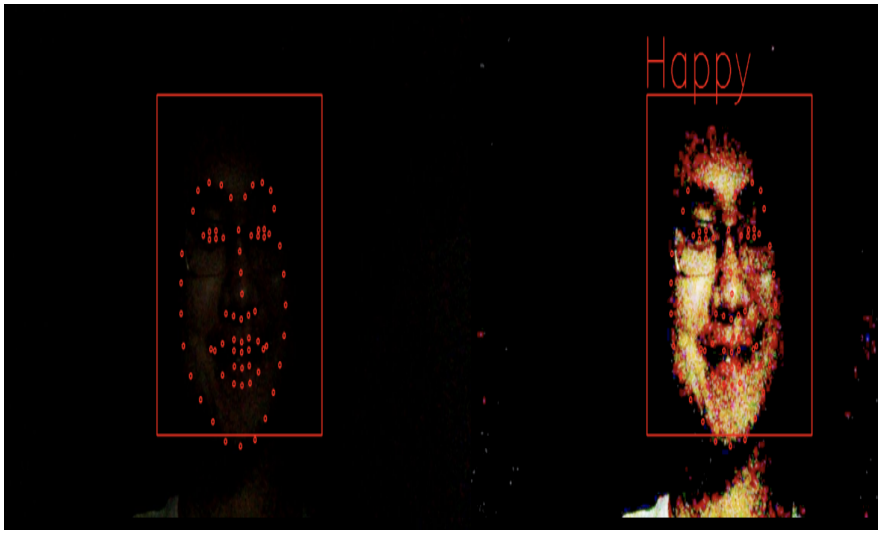
4.2 Expression Recognition

4.2.1 Dataset

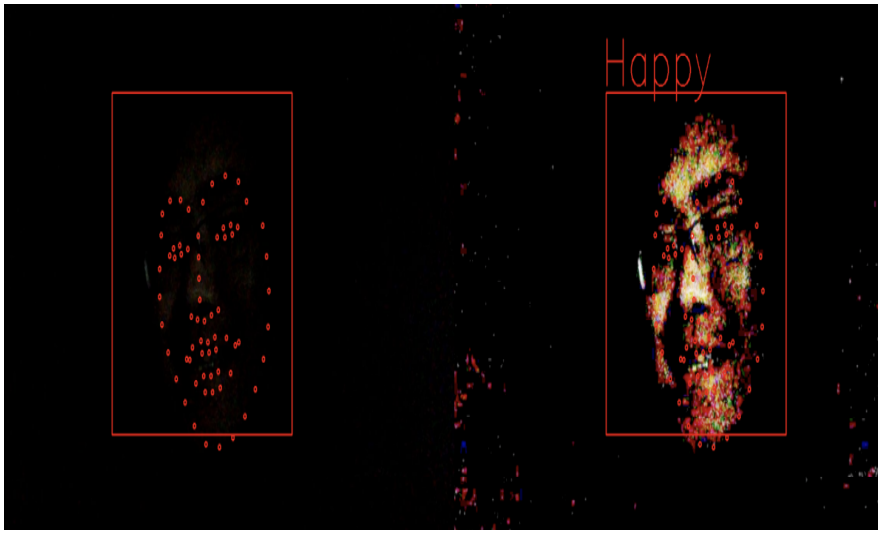
KDEF[17] dataset has 4900 images from 70 actress. Every participate expresses 7 emotions: neutral, happy, angry, afraid, disgusted, sad and surprised. Every expression is taken from 5 angles: full left, half left, straight, half right, full right. In our implementation, we focused on frontal face.

JAFFE[34] dataset has 213 images from 10 Japanese actress. Every participate expresses 7 emotions: neutral, happy, angry, afraid, disgusted, sad and surprised. Each image has averaged semantic ratings on 6 emotion adjectives by 60 Japanese viewers. All images are taken from the front.

CK/CK+[23][33] recorded the facial behavior of 210 adults. Participants were instructed by an experimenter to perform a series of 23 facial displays.



(a) Happy with Frontal Face



(b) Happy with half left profile

Figure 4.2: Expression Recognition Sample

We trained the model by KDEF dataset. The dataset is split into two parts; 90% for training and 10% for testing. JAFFE and CK/CK+ dataset are used for cross-dataset testing.

4.2.2 Results

4.2.3 Discussion

From our test, points distance is a better method than polar coordinate for facial expression recognition. And if two methods are used together, the accuracy will be boosted. SVM performs

Expression	Polar Coordinate			Points Distance			Polar Coordinate + Points Distance		
	Poly SVM	Random Forest	MLP	Poly SVM	Random Forest	MLP	Poly SVM	Random Forest	MLP
afraid	0.72	0.86	0.50	0.70	0.62	0.74	0.67	0.62	0.62
angry	0.70	0.67	0.56	0.65	0.65	0.62	0.71	0.77	0.89
disgusted	0.68	0.61	0.65	0.77	0.65	0.75	1.00	0.80	1.00
happy	0.92	0.92	0.91	1.00	0.92	1.00	1.00	1.00	1.00
neutral	0.65	0.61	0.56	0.78	0.71	0.70	0.85	0.75	0.64
sad	0.67	0.68	0.65	0.67	0.55	0.70	0.89	0.67	0.70
surprised	0.87	0.81	0.67	0.91	0.85	0.82	0.82	0.82	0.82
average	0.74	0.74	0.64	0.78	0.71	0.76	0.85	0.78	0.83

Table 4.1: Accuracy for Polar Coordinate and Points Distance Features.

Truth/Prediction	afraid	angry	disgusted	happy	neutral	sad	surprised
afraid	13	1	2	1	4	1	4
angry	3	16	4	0	2	1	0
disgusted	1	3	19	1	0	1	0
happy	0	0	1	24	0	1	0
neutral	0	1	0	0	20	4	0
sad	1	2	2	0	5	16	0
surprised	0	0	0	0	0	0	26

Table 4.2: Polar Coordinate SVM Confusion Matrix.

Truth/Prediction	afraid	angry	disgusted	happy	neutral	sad	surprised
afraid	13	0	2	1	4	1	5
angry	0	17	6	0	2	1	0
disgusted	0	4	17	1	0	3	0
happy	1	0	1	22	1	1	0
neutral	0	2	0	0	22	1	0
sad	1	3	3	0	6	13	0
surprised	0	0	0	0	4	1	21

Table 4.3: Polar Coordinate Random Forest Confusion Matrix.

Truth/Prediction	afraid	angry	disgusted	happy	neutral	sad	surprised
afraid	8	1	2	1	3	0	11
angry	1	19	2	0	3	1	0
disgusted	3	8	11	1	0	1	1
happy	1	0	1	21	1	2	0
neutral	1	2	0	0	19	2	1
sad	2	4	1	0	8	11	0
surprised	0	0	0	0	0	0	24

Table 4.4: Polar Coordinate MLP Confusion Matrix.

Truth/Prediction	afraid	angry	disgusted	happy	neutral	sad	surprised
afraid	19	2	0	0	1	2	2
angry	0	20	2	0	0	4	0
disgusted	1	4	20	0	0	0	0
happy	1	0	1	23	0	1	0
neutral	1	2	2	0	18	2	0
sad	0	3	1	0	4	18	0
surprised	5	0	0	0	0	0	21

Table 4.5: Points Distance Poly SVM Confusion Matrix.

Truth/Prediction	afraid	angry	disgusted	happy	neutral	sad	surprised
afraid	16	1	1	2	1	1	4
angry	3	13	6	0	1	3	0
disgusted	1	4	20	0	0	0	0
happy	1	0	1	23	0	1	0
neutral	0	0	0	0	17	8	0
sad	0	2	3	0	5	16	0
surprised	4	0	0	0	0	0	22

Table 4.6: Points Distance Random Forest Confusion Matrix.

Truth/Prediction	afraid	angry	disgusted	happy	neutral	sad	surprised
afraid	17	1	0	0	3	0	5
angry	3	15	4	0	1	3	0
disgusted	0	4	21	0	0	0	0
happy	0	0	1	24	0	1	0
neutral	0	2	0	0	21	2	0
sad	1	2	2	0	5	16	0
surprised	2	0	0	0	0	1	24

Table 4.7: Points Distance MLP Confusion Matrix.

	Model	KDEF	JAFFE	CK/CK+
Polar Coordinate	SVM	0.74	0.63	0.57
	Random Forest	0.74	0.60	0.53
	MLP	0.64	0.58	0.55
Points Distance	SVM	0.78	0.67	0.62
	Random Forest	0.71	0.60	0.55
	MLP	0.76	0.67	0.64
Polar Coordinate + Points Distance	SVM	0.85	0.72	0.64
	Random Forest	0.74	0.60	0.55
	MLP	0.81	0.72	0.68

Table 4.8: Cross-dataset Test.

the best among SVM, Random Forest and MLP. We find happy and surprised are easier to recognize than other expressions. The model tends to confuse sad/neutral. Some philosophy study doesn't count afraid as one of the seven main expressions, since it is more complicated than others. But it is still an important expression in our real life, we still take afraid into consideration in our application. The cross-dataset test shows the accuracy decreases in other dataset. To build a robust facial expression recognition application, a big dataset is necessary. And, the normalization methods may be improved. It can be adjusted according to the face shapes or other geometric features.

To evaluate the features we extract from points distance, we used ANOVA and Random Forest to analyze the association between features and expressions. Higher value means higher importance. Both ANOVA and Random Forest regards mouth as the most important feature for expression recognition. It matches our assumption, since the movement of mouth is obvious than other facial features. Besides mouth, eye and eyebrow are measured as the most important feature. This also matches our life experiences. Eyebrow is a very flexible feature on the face. By observing the related position between eyebrows and eyes, we can recognize human's expression. Eyebrow height is regarded as useless information. We think it is because even though eyebrow is flexible, its weight has no change for different expressions.

Name	ANOVA F-value	Random Forest
Left Eyebrow Length	6.1837	0.0347
Right Eyebrow Length	5.6474	0.0346
Left Outer Eyebrow Height	183.4206	0.0912
Right Outer Eyebrow Height	188.7228	0.0818
Left Inner Eyebrow Height	125.2920	0.0609
Right Inner Eyebrow Height	129.4476	0.0642
Left Eye Height	204.5092	0.0794
Right Eye Height	220.0162	0.0897
Mouth Height	175.4912	0.1294
Mouth Width	229.9945	0.1256
Left Lip Height	125.1850	0.0537
Right Lip Height	133.1062	0.0638
Upper Lip Height	66.1808	0.0539
Jaw Height	41.5754	0.0372

Table 4.9: Points Distance and its ANOVA F-value and Random Forest importance score.

5. SUMMARY

In this study, we explore various techniques for landmark detection and expression recognition in dark environments. We examine different computer vision techniques for dark image enhancement, such as contrast stretching, histogram equalization, Retinex and learning-base approaches to facilitate landmark detection. We designed a CycleGAN-based dark image enhancement model that sharpens edges and reduces ISO noise. Then, we used skin detection and edge detection to improve the accuracy of landmark location. With the detected landmarks, we trained a landmark-based expression recognition model and built a real-time expression recognition application in the dark. This application has the potential in smart home, patient monitoring and surveillance. A limiting factor to this study is there is no existing dataset with annotated expressions and facial landmarks in dark environments. We recorded some dark video clips with variant face angles, pose, backgrounds and brightness, but it couldn't simulate practical environment very well. The facial expression is quite subjective to people. For example, in the similar facial expression, some people look neutral and other people look sad. The low accuracy of neutral and sad approves this idea. The features are extracted by the euclidean distance of landmarks and polar coordinates, to handle different face shapes and face distances, we introduced normalization to promote the accuracy in our test. However, to handle more complicated environment, like face angle, pose and occlusion, new normalization methods need to be proposed.

REFERENCES

[1] Vinay A et al. “Two Novel Detector-Descriptor Based Approaches for Face Recognition Using SIFT and SURF”. In: *Procedia Computer Science* 70 (2015). Proceedings of the 4th International Conference on Eco-friendly Computing and Communication Systems, pp. 185–197. ISSN: 1877-0509. DOI: https://doi.org/10.1016/j.procs.2015.10.070. URL: http://www.sciencedirect.com/science/article/pii/S1877050915032342.

[2] Fatma Albluwi, Vladimir A. Krylov, and Rozenn Dahyot. “Image Deblurring and Super-Resolution Using Deep Convolutional Neural Networks”. In: Sept. 2018, pp. 1–6. DOI: 10.1109/MLSP.2018.8516983.

[3] Shaikh Allayear et al. “Human Face Detection in Excessive Dark Image by Using Contrast Stretching, Histogram Equalization and Adaptive Equalization”. In: 7 (Dec. 2018), pp. 3984–3989. DOI: 10.14419/ijet.v7i4.13713.

[4] Martin Arjovsky, Soumith Chintala, and Léon Bottou. *Wasserstein GAN*. 2017. arXiv: 1701.07875 [stat.ML].

[5] S. R. Benedict and J. S. Kumar. “Geometric shaped facial feature extraction for face recognition”. In: *2016 IEEE International Conference on Advances in Computer Applications (ICACA)*. 2016, pp. 275–278.

[6] Adrian Bulat and Georgios Tzimiropoulos. “How far are we from solving the 2D & 3D Face Alignment problem? (and a dataset of 230,000 3D facial landmarks)”. In: *International Conference on Computer Vision*. 2017.

[7] Xavier Burgos-Artizzu, Pietro Perona, and Piotr Dollár. “Robust Face Landmark Estimation under Occlusion”. In: Dec. 2013, pp. 1513–1520. DOI: 10.1109/ICCV.2013.191.

[8] Z. Cao et al. “OpenPose: Realtime Multi-Person 2D Pose Estimation using Part Affinity Fields”. In: *IEEE Transactions on Pattern Analysis and Machine Intelligence* (2019).

[9] D. N. Chandrappa, M. Ravishankar, and D. R. RameshBabu. “Face detection in color images using skin color model algorithm based on skin color information”. In: *2011 3rd International Conference on Electronics Computer Technology*. Vol. 1. 2011, pp. 254–258.

[10] Tianyuan Chang et al. *Facial Expression Recognition Based on Complexity Perception Classification Algorithm*. 2018. arXiv: 1803.00185 [cs.CV].

[11] Yu-Sheng Chen et al. “Deep Photo Enhancer: Unpaired Learning for Image Enhancement from Photographs with GANs”. In: *Proceedings of IEEE International Conference on Computer Vision and Pattern Recognition (CVPR 2018)*. Salt Lake City, June 2018, pp. 6306–6314.

[12] colorswall. “Colors in palette”. In: 2018. URL: https://colorswall.com/palette/2513/.

[13] National Cancer Institute Dietary Assessment Primer Section Name. National Institutes of Health. “Learn More about Normal Distribution”. In: URL: https://dietassessmentprimer.cancer.gov/learn/distribution.html. (accessed: 15.09.2020).

[14] Chao Dong et al. *Image Super-Resolution Using Deep Convolutional Networks*. 2015. arXiv: 1501.00092 [cs.CV].

[15] Xuanyi Dong et al. *Style Aggregated Network for Facial Landmark Detection*. 2018. arXiv: 1803.04108 [cs.CV].

[16] X. Fu et al. “A Weighted Variational Model for Simultaneous Reflectance and Illumination Estimation”. In: *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Los Alamitos, CA, USA: IEEE Computer Society, June 2016, pp. 2782–2790. DOI: 10.1109/CVPR.2016.304. URL: https://doi.ieeecomputersociety.org/10.1109/CVPR.2016.304.

[17] Ellen Goeleven et al. “The Karolinska Directed Emotional Faces: A validation study”. In: *Cognition and Emotion* 22.6 (2008), pp. 1094–1118. DOI: 10.1080/02699930701626582. eprint: https://doi.org/10.1080/02699930701626582. URL: https://doi.org/10.1080/02699930701626582.

[18] Ian J. Goodfellow et al. *Generative Adversarial Networks*. 2014. arXiv: 1406.2661 [stat.ML].

[19] H. Guo, K. Zhang, and Q. Jia. “2.5D SIFT Descriptor for Facial Feature Extraction”. In: *2010 Sixth International Conference on Intelligent Information Hiding and Multimedia Signal Processing*. 2010, pp. 723–726.

[20] Xiaojie Guo. “LIME: A Method for Low-light IMage Enhancement”. In: *CoRR* abs/1605.05034 (2016). arXiv: 1605.05034. URL: http://arxiv.org/abs/1605.05034.

[21] He Huang, Philip S. Yu, and Changhu Wang. *An Introduction to Image Synthesis with Generative Adversarial Nets*. 2018. arXiv: 1803.04469 [cs.CV].

[22] Yifan Jiang et al. *EnlightenGAN: Deep Light Enhancement without Paired Supervision*. 2019. arXiv: 1906.06972 [cs.CV].

[23] Takeo Kanade, Jeffrey Cohn, and Ying-Li Tian. “Comprehensive Database for Facial Expression Analysis”. In: *Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition (FG’00)*. Mar. 2000, pp. 46–53.

[24] V. Kazemi and J. Sullivan. “One millisecond face alignment with an ensemble of regression trees”. In: *2014 IEEE Conference on Computer Vision and Pattern Recognition*. 2014, pp. 1867–1874.

[25] Fuzail Khan. *Facial Expression Recognition using Facial Landmark Detection and Feature Extraction via Neural Networks*. 2018. arXiv: 1812.04510 [cs.CV].

[26] Fuzail Khan. *Facial Expression Recognition using Facial Landmark Detection and Feature Extraction via Neural Networks*. 2020. arXiv: 1812.04510 [cs.CV].

[27] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. "ImageNet Classification with Deep Convolutional Neural Networks". In: *Advances in Neural Information Processing Systems 25*. Ed. by F. Pereira et al. Curran Associates, Inc., 2012, pp. 1097–1105. URL: http://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks.pdf.

[28] W. Kumwilaisak et al. "Image Denoising With Deep Convolutional Neural and Multi-Directional Long Short-Term Memory Networks Under Poisson Noise Environments". In: *IEEE Access* 8 (2020), pp. 86998–87010.

[29] E. H. Land. "The Retinex Theory of Color Vision SCIENTIFIC AMERICAN". In: 2009.

[30] Thai Hoang Le and Truong Nhat Vo. *Face Alignment Using Active Shape Model And Support Vector Machine*. 2012. arXiv: 1209.6151 [cs.CV].

[31] Yuen Peng Loh and Chee Seng Chan. "Getting to Know Low-light Images with The Exclusively Dark Dataset". In: *Computer Vision and Image Understanding* 178 (2019), pp. 30–42. DOI: https://doi.org/10.1016/j.cviu.2018.10.010.

[32] Kin Gwn Lore, Adedotun Akintayo, and Soumik Sarkar. "LLNet: A Deep Autoencoder Approach to Natural Low-light Image Enhancement". In: *Pattern Recognition* 61 (Nov. 2015). DOI: 10.1016/j.patcog.2016.06.008.

[33] P. Lucey et al. "The Extended Cohn-Kanade Dataset (CK+): A complete dataset for action unit and emotion-specified expression". In: *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Workshops*. 2010, pp. 94–101.

[34] Michael Lyons, Miyuki Kamachi, and Jiro Gyoba. *The Japanese Female Facial Expression (JAFFE) Dataset*. The images are provided at no cost for non- commercial scientific research only. If you agree to the conditions listed below, you may request access to download. Zenodo, Apr. 1998. DOI: 10.5281/zenodo.3451524. URL: https://doi.org/10.5281/zenodo.3451524.

[35] John McCann. “Retinex Theory”. In: *Encyclopedia of Color Science and Technology*. Ed. by Ming Ronnier Luo. New York, NY: Springer New York, 2016, pp. 1118–1125. ISBN: 978-1-4419-8071-7. DOI: 10.1007/978-1-4419-8071-7_260. URL: https://doi.org/10.1007/978-1-4419-8071-7_260.

[36] X. Miao et al. “Direct Shape Regression Networks for End-to-End Face Alignment”. In: *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 2018, pp. 5040–5049.

[37] Shervin Minaee and Amirali Abdolrashidi. *Deep-Emotion: Facial Expression Recognition Using Attentional Convolutional Network*. 2019. arXiv: 1902.01019 [cs.CV].

[38] A. S. Parihar and K. Singh. “A study on Retinex based method for image enhancement”. In: *2018 2nd International Conference on Inventive Systems and Control (ICISC)*. 2018, pp. 619–624.

[39] Stephen M. Pizer et al. “Adaptive Histogram Equalization and its Variations”. In: 39.3 (Sept. 1987), pp. 355–368. DOI: http://doi.acm.org/10.1145/29040.29046.

[40] Emily B. Prince, Katherine B. Martin, and D. Messinger. “Facial Action Coding System”. In: 2015.

[41] Anwar Saeed et al. “Frame-Based Facial Expression Recognition Using Geometrical Features”. In: *Advances in Human-Computer Interaction* 2014 (Apr. 2014), pp. 1–13. DOI: 10.1155/2014/408953.

[42] Yangming Shi, Xiaopo Wu, and Ming Zhu. *Low-light Image Enhancement Algorithm Based on Retinex and Generative Adversarial Network*. June 2019.

[43] Shreenidhi Sudhakar. “Histogram Equalization”. In: 2017. URL: https://towardsdatascience.com/histogram-equalization-5d1013626e64.

[44] Li Tao et al. “Low-light image enhancement using CNN and bright channel prior”. In: Sept. 2017, pp. 3215–3219. DOI: 10.1109/ICIP.2017.8296876.

[45] Chunwei Tian et al. *Deep Learning on Image Denoising: An overview*. 2020. arXiv: 1912.13171 [eess.IV].

[46] Chen Wei et al. *Deep Retinex Decomposition for Low-Light Enhancement*. 2018. arXiv: 1808.04560 [cs.CV].

[47] Chen Wei et al. "Deep Retinex Decomposition for Low-Light Enhancement". In: *CoRR* abs/1808.04560 (2018). arXiv: 1808.04560. URL: http://arxiv.org/abs/1808.04560.

[48] Pengcheng Wei et al. "Research on face feature extraction based on K-mean algorithm". In: *EURASIP Journal on Image and Video Processing* 2018 (Dec. 2018). DOI: 10.1186/s13640-018-0313-7.

[49] Wayne Wu et al. "Look at Boundary: A Boundary-Aware Face Alignment Algorithm". In: *CVPR*. 2018.

[50] Yue Wu and Qiang Ji. "Robust Facial Landmark Detection under Significant Head Poses and Occlusion". In: *CoRR* abs/1709.08127 (2017). arXiv: 1709.08127. URL: http://arxiv.org/abs/1709.08127.

[51] Wenhui Zhang et al. "Gray-Edge-HOG feature based cascaded learning for facial landmark detection". In: *MATEC Web of Conferences* 189 (Jan. 2018), p. 10023. DOI: 10.1051/matecconf/201818910023.

[52] X. Zhang et al. "SegGAN: Semantic Segmentation with Generative Adversarial Network". In: *2018 IEEE Fourth International Conference on Multimedia Big Data (BigMM)*. 2018, pp. 1–5.

[53] R. Zhi et al. "Graph-Preserving Sparse Nonnegative Matrix Factorization With Application to Facial Expression Recognition". In: *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 41.1 (2011), pp. 38–52.

[54] L. Zhong et al. "Learning active facial patches for expression analysis". In: *2012 IEEE Conference on Computer Vision and Pattern Recognition*. 2012, pp. 2562–2569.