FaceX: A 3D Data-driven Morphable Face Synthesis Engine for Facial Expression Analysis

Xiang Li

A thesis submitted for the degree of
Master of Machine Learning and Computer Vision
The Australian National University

November 2020

Except where otherwise indicated, this thesis is my own original work.

Xiang Li
20 November 2020

to my Parents,Shuanghong Li,Wenzi Liu, and my wife, Xiaorong Ma.

Acknowledgments

First of all, I want to thank to my supervisor Tom. I was not sure at the beginning what kind of project I wanted to complete, and in what form my project was presented. Tom gave me many project options. This allowed me to choose this project when I had no idea. In the process of project development, because I am not very familiar with the field, especially unity. Tom gave me a lot of knowledge about software and project progress ideas. At the same time, I was given the channels and methods for obtaining relevant data sets. Evaluate the progress of my project through weekly meetings. At the same time, I answered the problems in my project. Secondly, I want to thank the other students in the laboratory. Actually, the progress of my project is not very smooth. I took a lot of detours and didn't find the right direction until the last few weeks. In this process, the other seniors in the laboratory also gave me great help. Xiaoxiao will explain the content of my project so that I can go deeper into this field faster. At the same time, she answered the questions I asked patiently, even though some of them seemed very silly. I am very grateful to Xiaoxiao for helping me with this project and correcting me in time when I make detours. Other students in the laboratory also helped me a lot in this process. I am very grateful to them. Especially because of the epidemic, I can only communicate online and remotely. They always help me when I need them. Finally, I want to thank my parents and my fiancee. It was they who gave me a lot of spiritual help, and helped me ease my mood when I was depressed and stressed. At the same time, they take great care of me in life so that I can devote myself to learning.

Abstract

Put your abstract here.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Thesis Statement

FaceX is a lightweight, flexible and scalable engine that can be used to generate a variety of high-quality images with diversity of facial expressions and shapes.

1.2 Introduction

In the last few decades, facial expression analysis (FEA) is a challenging task of computer vision and has attracted the interest of more and more researchers. Facial expression has been proven to play important role in understanding human emotion [Mehrabian and Russell, 1974]. Because facial expression is the response of a person's mental state to external stimuli [Cabanac, 2002]. And Ekman classifies human emotions into seven basic categories: Happy, Sad, Surprise, Anger, Disgust, Fear and Neutral based on cross-culture study. In recent years, Compound Emtions(CEs) [Du et al., 2014] and Micro Expressions(MEs) [Ekman, 2006] have also been explored, but achievements are not rich.

With the development of artificial intelligence, especially machine learning (ML), people benefit from artificial intelligence agents adjusting their response according to their emotional state [Adolphs and Andler, 2018]. In this regard, there are braod facial expression applications in different domains like Human-Centred Computing(Hcc) [Cowie et al., Jan./2001], augmented reality (AR) [Chen et al., 2015], virtual reality (VR) [Bekele et al., 2013], automatic driving [Jabon et al., 2011], and gaming [Lankes et al., 2008].

Various types of data can feed the FEA systems. In computer vision, facial images are the mainstream input data type. In addition, electromyography (EMG), electrocardiographic(ECG) and other related physical or chemical signal can be used as input data or auxiliary data as well [Jerritta et al., 2011]. This thesis focus on using facial images taken by sensor to detect expression and feeling of people. Because facial images contain sufficient non-verbal information for FEA [Huang et al., 2019].

We briefly review the development of FEA in computer vision, most of FEA techniques can be defined as either traditional methods or learning-based methods

[Huang et al., 2019].

Traditional methods involves various hand-craft features. Those approaches need to design appropriate feature extraction and choose different classifier for different dataset. Lots of conventional FEA technicals can be summarized into three main steps: preprocessing, feature extraction, facial expression classification. Preprocessing aims to reduce irrelevant redundant information and improve the ability to recognize related information. Image noise reduction, face detection and alignment and image enhancement are popular preprocessing methods. As for feature extraction, Local Binary Pattern(LBP) [Ahonen et al., 2004], Optical Flow [Horn and Schunck, 1981], Gabor feature [Lyons et al., 1998a],and etc. are commonly used in FEA. Another important aspect of traditional FEA methods is selecting classifier to predict facial expression. Support Vector Machine (SVM), k-Nearest Neighbours (kNN), Adaptive Boosting (Adaboost) and etc. classical ML classifier are widely deployed in FEA system.

On the other hand,**Learning-based methods** especially for Deep-Learning (DL) have gradually become mainstream due to the great-leap-forward development of computing resources and datasets. Although the DL method outperforms and is more robust than traditional methods because it reduces the dependence on image preprocessing and hand-craft features, DL requires a huge volume of high-quality data for training compared to traditional methods. According to Roh et al., appropriate high-quality representative training data has become the bottleneck of DL. Regardless of the model, computing resources are no longer the main limitation of the end performance. Data augmentaion can alleviate the problem of insufficient data in some extent, but it is limited to the dataset itself [Shorten and Khoshgoftaar, 2019]. Therefore, we hope to solve this problem with pure synthetic methods. Despite the pure synthetic method greatly increases the diversity of the dataset, how to shrink the domain gap between the synthetic data set and the real data distribution is still the biggest challenge.

In this work, in order to tackle the problem of synthetic facial expression data we need to evaluates some publicly available datasets for FEA. [Khan et al., 2020] summarize that the facial expression dataset can be divided into video dataset and image dataset. Most public datasets only contain less than 10k images like DISFA [Mavadati et al., 2013a] , and some dataset less 200+ images like JAFFE [Lyons et al., 2020]. Only a few data sets contain more than 1 million images like EmotioNet [Benitez-Quiroz et al., 2016a]. However, Some datasets capture a large number of images from the Internet, others are poor resolution, or the distribution of expressions is uneven. Therefore, We propose a facial expression data synthesis engine based on 3D morphable face model. This engine can synthesize a large number of labeled data with different shape, various expressions, and flexible poses. Figure 1.1 show an overview of our engine.

In summary, the contributions of this thesis are as following:

- Briefly summarize the development of FEA recently, and try to use synthetic data to solve the DL data bottleneck. (Section 1.2, Chapter 2)

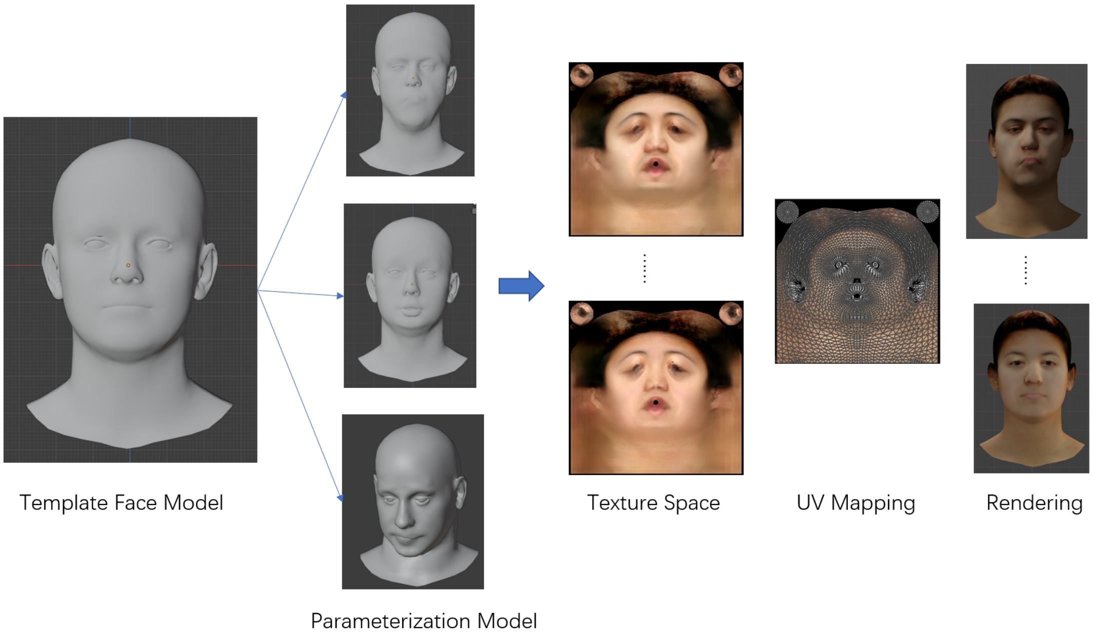


Figure 1.1: Our proposed engine. We are able to generate large-scale facial expression data sets for training deep neural networks. We set the initial face model with different shape, expression and pose parameters, and sample different materials from the texture space to map to the model. Finally render under different lighting parameters.

- Build a complete synthetic data engine, including various face shape, expression, pose and texture. (Chapter 3)
- Measure the performance of DL method on synthetic dataset. (Chapter 4)
- Discusse the potential problems of the synthetic data, and explain possible improvement. (Chapter 6)

1.3 Thesis Outline

The rest of thesis is organized from motivation to building of synthetic facial expression images.

Chapter 2 introduces some details of the traditional FEA method and Deep Learning-based method, and reviews the existing FE datasets, as well as 3D Morphable Face Models (3DMM) that can be used for data synthesis.

Chapter 3 describes the pipeline of the data synthesis engine.

Chapter 4 is the detial of experimental design and synthetic dataset.

Chapter 5 and Chapter 6 are the results of experiments and related analytic demonstration.

Chapter 2

Background and Related Work

This chapter gives a brief overview of Facial Expression Recognition3D morphable face models and facial expression dataset. We will limit our background knowledge on discrete basic categorical model and some static based facial expression recognition techniques we used.

We widely investigated Facial Expression Recognition in Section 2.1 at first. Then, the public available facial expression datasets are described in Section 2.2. Finally, we provide a review of building and applying 3D morphable face models 2.3.

2.1 Facial Expression Recognition

Facial expressions are the most direct and natural carrier of human emotions state[Darwin and Ekman, 2009]. Building an automatic facial expression analysis system has become an urgent need for artificial intelligence. In the field of computer vision and machine learning since the early 20th century, traditional facial expression recognition mainly benefited from the development of handcraft features and the application of classification algorithms; after 2013, due to the large number of facial expression datasets available, facial expression recognition transfers to deep learning gradually.

2.1.1 Traditional methods

Handcraft feature engineering and classification paly vital role in traditional facial expression recognition methods. In this section we will introduce several commonly used features and classifiers. The pipeline shown in Figure 2.1.

2.1.1.1 Feature Extraction

The purpose of feature extraction is to extract the non-pixel information expression in pixels so that the classifier can classify these features. In FER, the major used features are local binary patterns (LBP) [Ahonen et al., 2004],Gabor feature [Lyons et al., 1998a], and Haar-like feature [Viola and Jones, 2001].

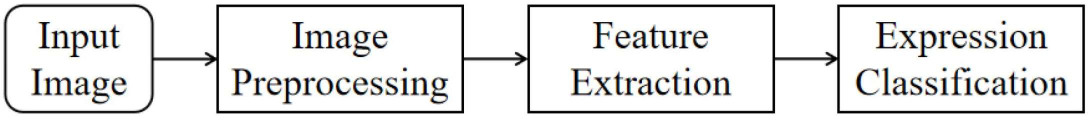


Figure 2.1: **Image Processing** usually aims to reduce redundancy information of related task. **Feature Extraction** involves how to design suitable features for FEA task. **Expression Classification** is how to category features.

Local binary patterns (LBP): is a simple but effective texture operator. It encodes the relationship between each pixel and nearby pixels to binary value. The most important attribute of LBP is its robustness to grayscale changes such as illumination changes. Another important feature is its computational simplicity, which enables real-time analysis of images. A simple way to encode facial expression is shown in Figure 2.2.

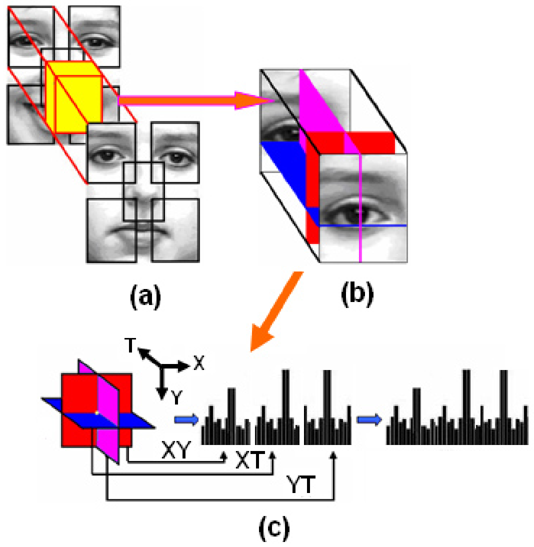


Figure 2.2: Using LBP to encode facial expression.Divide input image into several cells, the local binary features of different cells are calculated respectively. Next, calculate the histogram of each cells. Finally, concatenate the normalized histograms to form the feature vector of input.

In the applications of FER, there are also several promoted algorithms based on LBP. For example, Complete Local Binary Pattern (CLBP)[Zhenhua Guo et al., 2010] achieves better performance than the original LBP. LBP-based Local Directional Pattern (LDP) [Jabid et al., 2010] shows robust to illumination changes. The advantage of LBP methods is low computational complexity and small memory usage demanded. However, noise sensitivity and only focusing on local information are the disadvantages of this method.

Gabor Feature: is another kind of texture feature. In FER, we benefit from the multi-resolution and multi-orientation property of Gabor feature to encode facial expression image [Lyons et al., 1998a]. Mattela and Gupta proposed Gabor-mean-DWT to tackle the dimensional disaster of original Gabor feature. Gabor feature's merits is that it is not sensitive to illumination and direction, and can apply multi-scales. However, the calculation is time-consuming and requires a lot of memory.

Haar-like feature: is designed to be used for target recognition tasks. It combines the lines, borders and other features of the picture. It is also the first instant face detection operation. Compared to most other features, Haar-like feature known for calculation speed. Due to the use of integral images, any size can be calculated in constant time. However, it is sensitive to illumination. If the global region illumination diverse, the Harr-like feature may be hard to describe the local grayscale variation.

2.1.1.2 Classification

After we extract the features, another significant characteristic of the traditional method is the application of classifiers. In machine learning, the most widely used classifiers include: Support Vector Machine (SVM), k-Nearest Neighbours(kNN), Adaptive Boosting (Adaboost) and etc.

Support Vector Machine (SVM): aims to solve the data classification problem in the field of pattern recognition, which is a kind of supervised learning algorithm. The core of SVM is to map linearly inseparable data to a high-dimensional space through the kernel function to achieve linearly separable. Due to the characteristics of the kernel function, we can eliminate the need for complex calculations in high-dimensional space and solve the dimension disaster to a certain extent. In FER, SVM is widely used after different features representation, and has achieved good performance [Michel and El Kaliouby, 2003; Tsai and Chang, 2018; Hsieh, Hsih, Jiang, Cheng, and Liang, 2016; Saeed, Baber, Bakhtyar, Ullah, Sheikh, Dad, and Ali, 2018].

k-Nearest Neighbours(kNN): is the simplest supervised classification algorithm. Typically, kNN is a representative of lazy learning due to no need for training phase. Every new data must be compared with each training data. Unfortunately, kNN has not ability to capture global structure of data. Hence, kNN might involve in local optimal solution or unstable clarification [Dino and Abdulrazzaq, 2019; Wang, Liu, and Zhang, 2015].

Adaptive Boosting (Adaboost)'s: key thought is that classifier will use the sample from last misclassified classifier to train a classifier. AdaBoost aims to find the best training features that are useful for its weak classifiers. After each feature selection, weights will be re-adjusted by the local classification error. It reduces overfitting to a certain extent [Liew and Yairi, 2015; Krishna Gudipati, Ray Barman, Gaffoor, Harshagandha, and Abuzneid, 2016].

There are many other handcraft features and classifer that we have not mentioned in this section, but we can see that the major handcraft features used for facial expression recognition are geometric and texture features. This also shows that if we want to generate facial expressions data, geometry and material will be two important components.

2.1.2 Deep Learning-based Method

Compared with traditional methods, deep learning methods also have three similar steps. But it greatly reduces the model's dependence on image preprocessing and handcraft feature engineering. It also improves the robustness to the environment. In the section, we will outline the existing technologies involved in deep learning in facial expression recognition. Figure 2.3 shows the pipeline of deep learning facial expression recognition system .

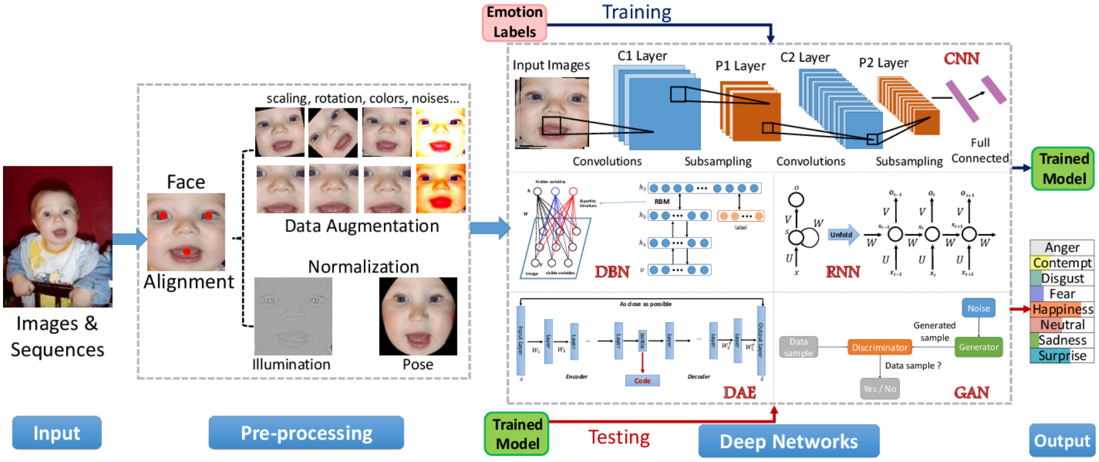


Figure 2.3: A typical architecture for deep learning facial expression recognition [Li and Deng, 2020].

2.1.2.1 Image Preprocessing

Illumination, environment and head pose changes will limit the performance of the training model. Hence, before feeding our data into deep neural network, we should normalize those visual facial semantic information as first.

Face alignment: is widely used in many computer vision tasks of faces. For a given face datasets, the first step is often to detect and crop the face to eliminate irrelevant background and information. ViolaJones(V&J) [Viola and Jones, 2001] is one of widely used frontal face detection, which had built in most popular computer vision library like opencv. After face detection, in order to capture the geometric features, Mollahosseini et al. showed that with landmarks detection feeding into network, the FER performance imporve significantly. That's because it highly reduce the effect of scale and rotation of face image. Of course, many deep neural network has been employed in face detection and landmark detection. Cascaded CNN [Sun et al., 2013] is the most popular technicals due to its inference speed and accuracy. Although Mutlti-task CNN [Zhang et al., 2016] and other multi-task network imporve the performance, we still make compromises on speed and accuracy. In our method, we use Cascaded CNN in dlib to detection face as well.

Face normalization: amis to reduce the effect of diversity of illumination and head pose. *Illumination normalization*: Many deep learning FER system preprocess image by histogram equalization to increase the contrast [Yu and Zhang, 2015; Ebrahimi Kahou, Michalski, Konda, Memisevic, and Pal, 2015; Pitaloka, Wulandari, Basaruddin, and Liliana, 2017; Bargal, Barsoum, Ferrer, and Zhang, 2016]. It shows outstanding performance when the illumination normalization when the background and front face area are under near illumination condition. However, to address the overemphasizing local contrast of using histogram equalization, Kuo et al. proposed a robust histogram equalization with linear combination. And Pitaloka et al. measured the performance of different methods, it concluded that global contrast normalization (GCN) and histogram equalization have achieved the best performance in FER recognition. *Pose normalization*: Variant head pose and occlusion significantly affect the performance of facial expression recognition system. Hassner et al. proposed using a 3D texture reference model to estimate pose, and then project back. Over the same period Sagonas et al. taken advantage of statistical model to localize landmark and estimate pose. In addition, Genrative Adversarial Network (GAN) were employed in front view synthetic and achieve better performance like DR-GAN [Tran et al., 2017], TP-GAN [Yin et al., 2017].

2.1.2.2 Convolutional Neural Network (CNN)

In computer vision and machine learning, DL has shown his capabilities of extracting low or high-level abstract features, which is more effective than handcraft features, and significantly improved performance in many fields, such as objecte detection, face recognition for identity verification and etc.. As for facial expression, Convolutional Neural Network (CNN) are widely used. Fasel and Fasel found that CNN has the ability to tackle pose and scala variation. It also outperformance the traditional multilayer perceptron (MLP).

AlexNet [Krizhevsky et al., 2017],VGG [Simonyan and Zisserman, 2014],GoogleNet [Szegedy et al., 2015], Resnet [He et al., 2016] are the most famous and popular CNN model, which has been explored in facial expression recognition. Moreover, some model proposed to address object detection task also adapt into facial expression recognition, such as Region-based CNN (R-CNN)[Girshick et al., 2013] and Faster R-CNN [Li et al., 2017a]. In order to handle spatio-temporal information, 3D CNN [Ji, Xu, Yang, and Yu, 2012; Tran, Bourdev, Fergus, Torresani, and Paluri, 2014] has been used to capture spatial representation of expression. However, in this thesis, we only focus on static image which doesnot involve spatio-temporal feature.

2.1.2.3 Deep Learning Classification

After learning features from model, the last step of deep learning methods also need to classify these features representation. However, unlike traditional methods, feature extraction and classification are independent of each other. In facial expression recognition, deep learning method can be an end-to-end approach, or just use CNN as a feature extractor and then combined with another classifier. For the first way, in CNN, the most common is to use loss function to minimize the gap between the predicted distribution and the true distribution. We can also use linear SVM [Tang, 2013] or other differentiable classifiers to build end-to-end facial recognition system. In addition to the end-to-end way, independent classifiers such as random forest, adaboost, and etc. can be applied to classify the extracted features from deep learning model [Donahue et al., 2013].

2.2 Facial Expression Dataset

For the design of the a deep facial expression recognition system, it is crucial to have the labeled training data. The training data are sufficient to have variations of environment as well as identify. In this section, we will introduce some public available dataset and basic information of these dataset. We will focus on their environment setting, size, drawbacks and etc..

EmotionNet:[Benitez-Quiroz et al., 2016b] The dataset was released in 2017, with a total of 950,000+ images, including basic expressions, compound expressions, and face action units. With large amount of the image, this dataset cover almost all possible facial expression. All the imges were anontated by the automatic AUs detection tools. However, due to all the images are downloaded from the websites, there will not be of fixed resolution.

AffectNet:[Mollahosseini et al., 2017] The data set was released in 2017. The dataset was collected using 1250 keywords in 6 different languages for retrieval in search engines, and get more than 42000 images. All images are labeled manually. The label type includes basic expression and realted amplitude. The expression type includes 8 basic expressions such as neutral expression, happy, sad, surprised, afraid, disgusted, angry, and contemptuous, as well as expressionless, uncertain, and unmanned. This dataset has provide annotation of searched key for all images.

The extended CohnKanade:[Lucey et al., 2010] This dataset was released in 2010, this database is an extension of the Cohn-Kanade Dataset, which contains 137 video frames of different facial expressions of people. CK+ contains 593 video sequence from 123 objective. The sequences has difference frames from 10 to 60. For sequences, 327 sequences from 118 subject are label with seven different facial expressions. CK+ is all in the lab. This dataset includes the frame information of the video and also gives out the peak of an expression. However, this dataset is different from the wild. When we want to use it in reality, we need to convert it.

Aff-Wild:[Zafeiriou et al., 2017] This dataset contains 248 videos with the length of 30 hours. All the videos in the dataset are recorded arbitrarily. It contains 130 males objects and 70 female objects. The videos are annotated by valence and arousal.Valence and arousal shows the degree of a facial expression, like degree of negative or positive. This dataset is the largest video dataset for facial expression. However due to faces are extracted from the video, numbers of subject are not high.

Oulu-CASIA:[Zhao et al., 2011] This dataset contains images of 80 subjects and six basic face expression, including happiness,surprise, fear, anger, sadness and disgust. National Laboratory of Pattern Recognition Beijing provides 30 Chinese sujects. University of Oulu provides 50 subjects. When they gain the dataset, they used two image light and visible light. So this dataset is popular for images which are captured in different illumination condition. However, this dataset only contains the frontal pose of subjects.

Denver Intensity of Spontaneous Facial Action(DISFA):[Mavadati et al., 2013b] This data set contains videos of 27 subjects. Half are males and half are females. This dataset uses 12 action unit to encode the expression and six expression are identified including surprise, sad, smile, neutral, disgust and neural. The video in this dataset contains 4845 number of frames and also are taken based on their stimulus to an emotive video. This dataset focus on FACS standard to generate video sequence. However, this dataset does not suit for the testing staga.

Japanese Female Facial Expression(JAFFE) Database:[Lyons et al., 1998b] The dataset was released in 1998. The database is a facial expression image captured by a camera by 10 Japanese women who made various expressions according to instructions in an experimental environment. There are a total of 213 images in the entire database, 10 people, all women, and each person makes 7 expressions. These 7 expressions are sad, happy, angry, disgusted, surprised, fearful, and neutral. There are about 20 samples in each group. Figure. Because it captured in the lab, facial expression is very clear. But the background for the image is noisy.

MMI Facial Expression Dataset: [Pantic et al., 2005] This dataset contains over 2900 videos and high-resolution still images of 75 subjects. AUs are completely annotated. This dataset is laboratory-controlled. The 213 sequence are labeled with six basic expression. The sequence in this dataset are onset-apex-offset labeled.

Binghamton University 3D Facial Expression(BU-3DFE): [Yin et al., 2006] This dataset is designed for research on 3D facial expression. It contains 56 females and 44 males subject and 100 in total. Those subjects are from various racial ancestries. Also they have wide age range from 18 to 70. This dataset contains six basic facial expressions. Also, 3D facial models for each sbuject are introduced in the dataset. Also 83 manually annotated facial landmark connected with each model are contained. This dataset is used for multiview 3D facial expression analysis.

FER2013 Face Dataset : [Carrier et al., 2013] This dataset was released in 2013. This dataset contains 35887 face images and includes 28709 training sets, and 3589 verification sets and 589 test sets. All the images are gray scale with 48 pixels plus 48 pixels. This dataset contains seven basic face expressions incluidng fear, anger, sad, surprise, happy, disgust and neutral. Each sample in the dataset has a wide range of age, direction. So it is closed to the real world.

Real-word Affective Database(RAF-DB) : [Li and Deng, 2018] This dataset contains 29672 diverse facial images. All the images are downloade from the website. This dataset contains six basic and eleven compound emotion. The 15339 images from the basic emotion set were separated into two group, including training sample and test sample.

Among these dataset, we can summarize that images from those larger size facial expression datasets are variant resolution, even most of data are low resolution, and inconsistency among different image quality. On the other hand, images from those larger size facial expression datasets are usually generated in the laboratory condition. Therefore, those dataset lack the diversity. In this condition, we aim to solve the quailty of the facial expression image and the diversity of the data images.

2.3 3D Morphable Face Models

One of the most typical ways to synthesize data is to use game engines, such as Unity, Unreal, or modeling tools like Blender, C3D to synthesize data in the virtual physical world.

When we use the 3D engine to synthetic data, we need to find out or construct suitable morphable facial model. Thus, in this section we will investigate 3D Morphable Face Models (3DMM).

A 3D morphable face model is a genertive model for face shape and appearance. 3D Morphable Face Models were first introduced in 1999[Blanz and Vetter, 1999]. And these models were used as a general principled approach to analyse images. There are different ways to compute a 3DMM by modeling. In this part, I will introduce these three different ways.

2.3.1 Shape model

The Shape model is classical modeling approach that uses 3D data. A shape space is traditionally defined as the set of all configuration of n vertices in 3D space with fixed connectivity[Dryden and Mardia, 2016]. Commonly used model are two models. One is the global model that represents variation of the entire face surface. The other is the local model that varation of facial parts.

2.3.2 Expression model

This kind of model captures variation of both identity and expression. Unlike simple linear models which learn through a dataset that has different identity and expression, this model focus on explicitly decouple the influence of identity and expression[Booth et al., 2017]. This is achieve by modeling in separate coefficients.

There are three different kinds of methods. The first one is additive model. This model gives two shape of the subject,including expression and neutral shape. It tranferred expression between subject by adding the offset of the expression[Blanz and Vetter, 1999]. The second is multiplicative model. A common multiplicative model is the concept of the multilinear model, which extends the idea that PCA performs singular value decomposition. The decomposition of the 3D face data into a stack of training data (HOSVD) by performing higher order tensor data to tensor data[Vlasic et al., 2006]. The third is nonlinear model. There also some methods to model facial variation with nonlinear transformation, such as FLAME, an articulated expressive head model that gives nonlinear controlYu et al. [2017].

2.3.3 Appearance model

The appearance model is to capture variation in appearance and illumination. The most common way to build it is to provide statistics on the appearance of the training shape, where the appearance information is usually expressed as a value per vertex or as a texture in the uv space[Booth et al., 2017]. There are two models. One is the linear per-vetex model which is low-dimension texture, the other is linear texture space model which require compatible resolution.

2.3.4 Public Available 3D Morphable Facial Model

Base Face Model(BFM) 2009 : [Paysan et al., 2009] This model is a kind of the shape model. Pascal Paysan used a laser scanner to accurately collect 200 individual data in 2009 to obtain the Basel Face Model dataset. The entire data set contains 200 three-dimensional faces. Among them, 100 were males and 100 were females, and most of them were Caucasian. The age distribution of these data is 8 to 62 years old. Everyone was collected 3 neutral expressions and selected the most natural one.

FaceWarehouse : [Cao et al., 2013] This model is kind of shape and expression model. Cao used Kinect's RGBD camera to capture 150 individuals from 7-80 years old from different ethnic backgrounds. For each person, it collected RGBD data of her different expressions, including neutral expressions and 19 other expressions. FaceWareHouse is widely used in visualization calculations, especially the bilinear face model has excellent performance in estimating face identity and expression in pictures and videos.

Surrey Face Model: [Huber et al., 2016] This model is kind of shape abd expression model. This model is a multi-resolution 3D deformed face model provided by the University of Surrey in the UK.The model contains different grid resolution levels and landmark point annotations, as well as metadata for texture remapping.

Face Learned with an Articulated Model and expression(FLAME) : [Yu et al., 2017] This is kind of shape, expression and head pose model. It contains

3800 individualis for shape, 800 for head pose and 21000 frames for expression. It considers different genders and also a full head model without hair.

Base Face Model(BFM) 2017: [Gerig et al., 2018] This is kind of shape and expression model. It has 200 individuals for shape and appearance and a total of 160 expression scans. Also this is an extension of the BFM 2009. Compare with the BFM 2009, BFM 2019 has multiresolution and was with full head.

Morphable Face Albedo Model:[Smith et al., 2020] This model is an extension of BFM 2017. It contains 73 individuals.The model captured data to provide ground truth for an albedo estimation benchmark using the same fitting pipeline. This model reduces the error in the estimated albedo by nearly 70 percent compared to using the existing base face model

2.4 Summary

In this chapter, we first introduce the traditional methods for FEA. We found that the design features of traditional methods are mainly focused on the geometric features and texture features of the face. This inspired us that for synthesize facial expression data, we should not only pay attention to the geometry information, such as the outline of the face. The material also contains the underlying information of the facial expression.

Moreover, we compared deep learning with traditional methods and summarized the application of deep learning in facial expression recognition.

Finally we summarized the public facial expression dataset and 3d Morphable face models, and explained their potential problems and advantages.

Chapter 3

Synthetic Data Generation

In this chapter, we demonstrate the detial of proposed facial expression image synthetic engine, FaceX. Our engine consist of three components: Face model, expression model and data generation engine. A overview of our engine is shown in Figure 1.1.

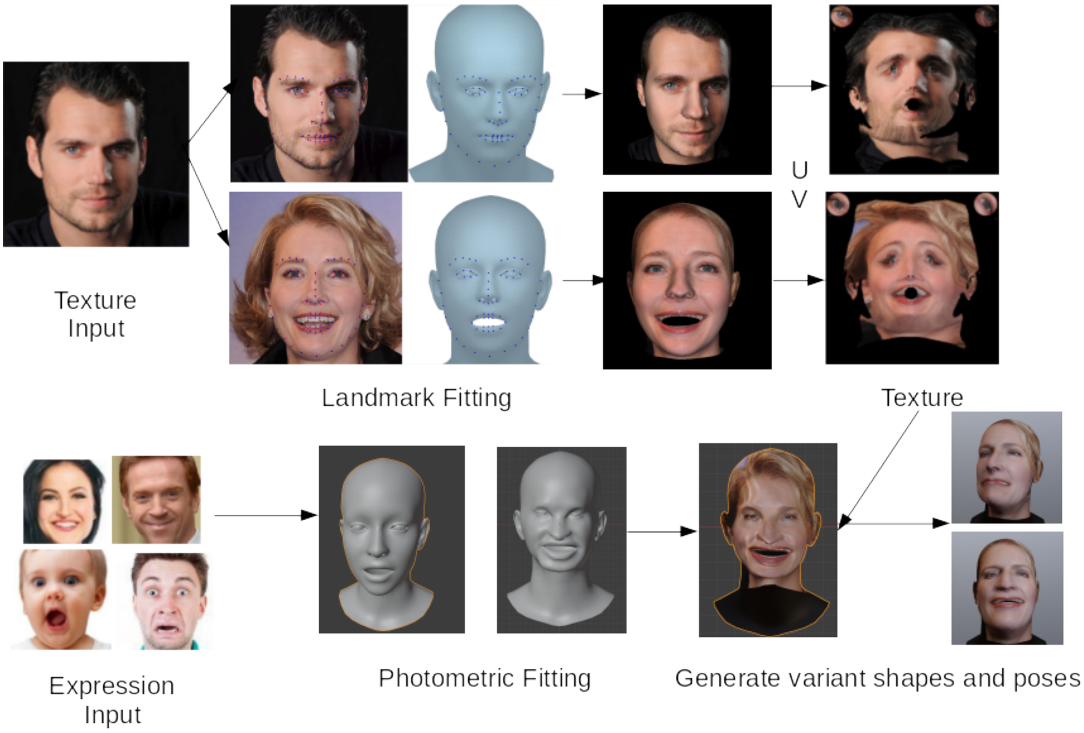


Figure 3.1: Overview of our engine.On the top, we minimize landmark distance to extract the texture from the high resolution image. On the bottom, we extract the face models corresponding to different expressions by minimizing the photometric error, and then put the extracted texture onto the face models. Finally modify the weight of the blendshape to achieve different shape and pose with given facial expression.

We first briefly introduce the setting of 3D morphable face model in section 3.1. Then, we explain how to use 3D facial model to generate faces, texture, and expression in section 3.2 and 3.3. Finally, we build a tools based on Blender [Community, 2018] to generate synthetic data in section 3.4.

3.1 3D Morphable Face Model Setting

According to our introduction of relevant background knowledge in Chapter 2, we tried different public available 3D morphable face model.

At first, we experimented the possibility of utilization the 3D Basel Face Model (BFM) [bfm, 2009]. Although BFM is the most widely used in academic research, we found that if we want to apply the model to facial expression generation, we need plenty of 3d scans mesh to generate blendshape for facial expressions. That means if we want to obtain a accurate expression blendshape, we need a lot of training data. And the model has 53490 vertices, it will take to much time to conduct an experiment on a general PC.

Then, we measured the possibility of Facewarehouse [Cao, Weng, Zhou, Tong, and Zhou, 2014]. As shown in Figure 3.2, Facewarehouse has included 46 facial expression blendshapes, which will save a lot of time for modeling. However, we found that the blendshapes in the dataset was provided separately, and the order of each vertex was also inconsistent, so a lot of model rigging work demanded.

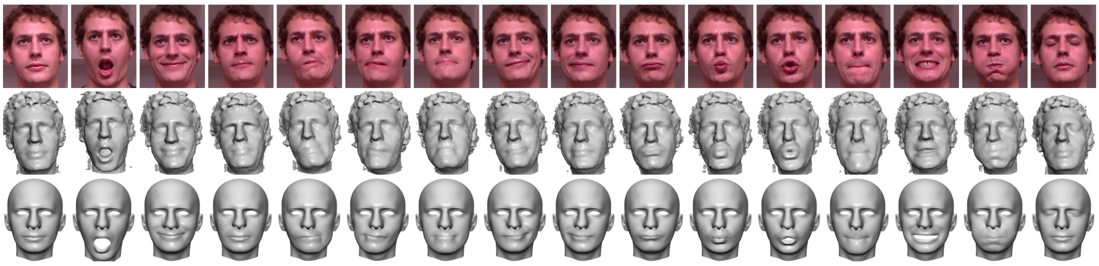


Figure 3.2: Examples from Facewarehouse dataset

Finally, after a few weeks of experimentation, we chose to use Faces Learned with an Articulated Model and Expressions (FLAME) [Li et al., 2017b]. Because this is a lightweight (only 5023 vertices), riging (vertics order fixed) face model. The author also provides 3D mesh static and dynamic 3d landmark. This means that we only need a projection matrix to generate the 2d landmarks of the 3d model.

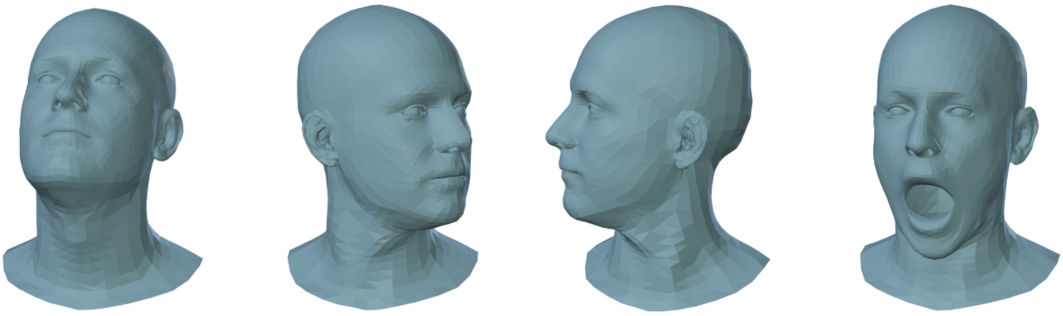


Figure 3.3: Different pose examples from FLAME.

3.2 Face Model

This section will introduce some technical details. A face model usually consists of two components: **shape** and **texture**. With development of 3D modeling technology, we could generate of different identities and expressions by adjusting the parameters of shape and texture.

Shape Model: We use linear combination to describe the shape parameters. FLAME could be formulate a function 3.1:

$$M(\vec{s}, \vec{p}, \vec{e}) : \mathbb{R}^{|\vec{s}|\times|\vec{p}|\times|\vec{e}|} \rightarrow \mathbb{R}^{3N} \tag{3.1}$$

where $N = 5023$ vertices, shape coefficients: $\vec{s} \in \mathbb{R}^{|\vec{s}|}$,pose coefficients: $\vec{p} \in \mathbb{R}^{|\vec{p}|}$,expression coefficientss: $\vec{e} \in \mathbb{R}^{|\vec{e}|}$. Because, we have fixed the order of vertices, so we could perform Principal Component Analysis (PCA). The shape could be consist of the mean shape vector $\bar{M}$ and the linear combination of

3.3 Expression Model

3.4 Data Generation

Chapter 4

Experiment and Result

4.1 Dataset

4.1.1 Dataset for training expression parameters

4.1.2 Pure Synthetic data

12000 images,7 different expression, 2000 subject.

4.2 Evaluation

4.2.1 Face landmarks + HOG with SVM

4.2.2 Different CNN scheme

- Baseline simple CNN
- ResNet FER

4.3 Example of synthetic facial image

4.4 Results

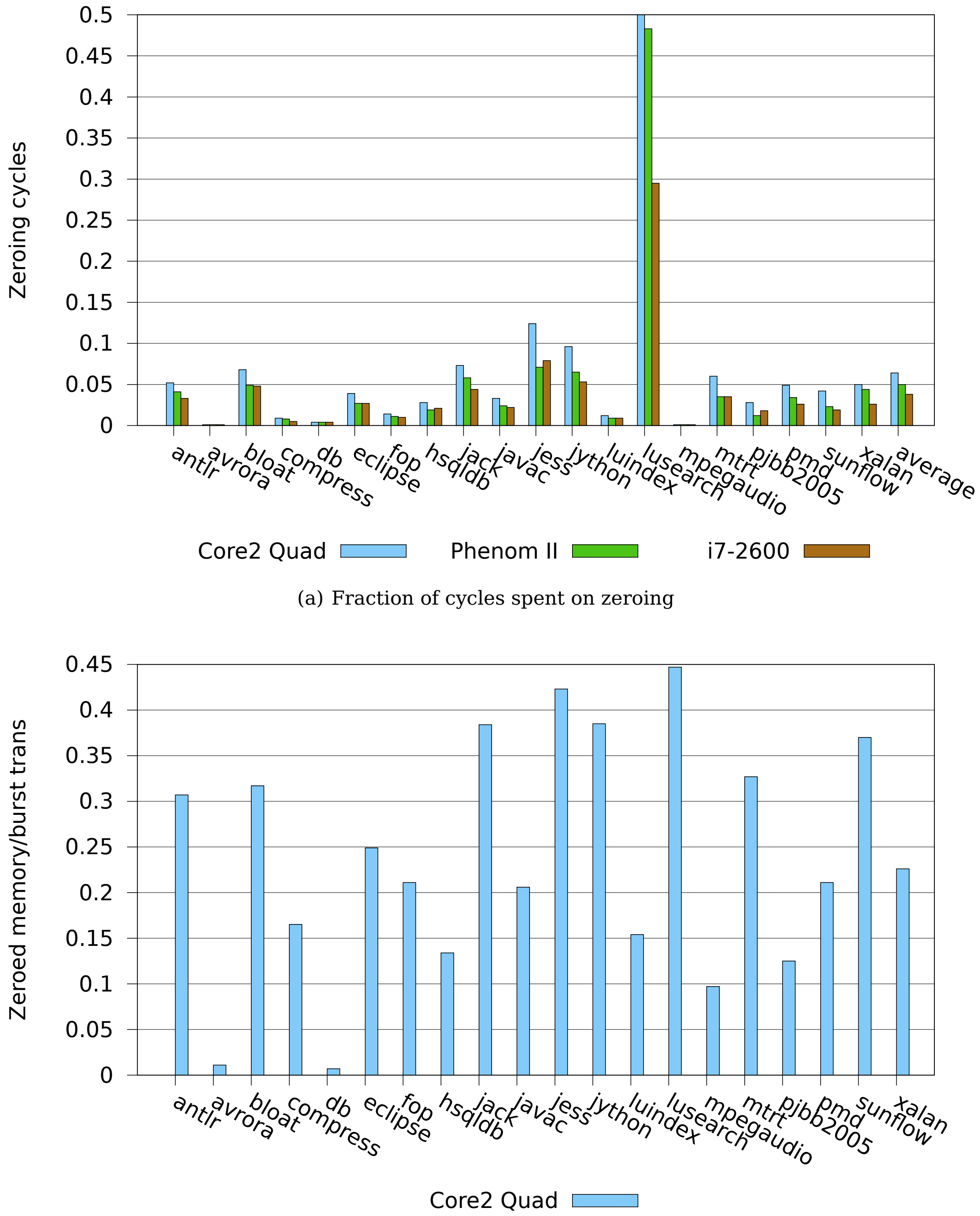
Chapter 5

Results

5.1 Direct Cost

Here is the example to show how to include a figure. Figure 5.1 includes two subfigures (Figure 5.1(a), and Figure 5.1(b));

5.2 Summary



(a) Fraction of cycles spent on zeroing

(b) BytesZeroed / BytesBurstTransactionsTransferred

Figure 5.1: The cost of zero initialization

Chapter 6

Conclusion

Summary your thesis and discuss what you are going to do in the future in Section 6.1.

6.1 Future Work

According to the experiment and related work, our proposed engine and experiment could be extended and imporved in several aspects.

Bibliography

2009. *A 3D Face Model for Pose and Illumination Invariant Face Recognition*. IEEE, Genova, Italy.

Adolphs, R. and Andler, D., 2018. Investigating Emotions as Functional States Distinct From Feelings. (2018), 191–201. doi:10.1177/1754073918765662.

Ahonen, T.; Hadid, A.; and Pietikäinen, M., 2004. Face Recognition with Local Binary Patterns. In *Computer Vision - ECCV 2004* (Eds. T. Kanade; J. Kittler; J. M. Kleinberg; F. Mattern; J. C. Mitchell; O. Nierstrasz; C. Pandu Rangan; B. Steffen; M. Sudan; D. Terzopoulos; D. Tygar; M. Y. Vardi; G. Weikum; T. Pajdla; and J. Matas), 469–481. Springer Berlin Heidelberg, Berlin, Heidelberg. ISBN 978-3-540-21984-2 978-3-540-24670-1. doi:10.1007/978-3-540-24670-1_36.

Bargal, S. A.; Barsoum, E.; Ferrer, C. C.; and Zhang, C., 2016. Emotion recognition in the wild from videos using images. In *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, ICMI ’16, 433–436. Association for Computing Machinery, New York, NY, USA. doi:10.1145/2993148.2997627.

Bekele, E.; Zheng, Z.; Swanson, A.; Crittendon, J.; Warren, Z.; and Sarkar, N., 2013. Understanding How Adolescents with Autism Respond to Facial Expressions in Virtual Reality Environments. *IEEE Transactions on Visualization and Computer Graphics*, (Apr. 2013), 711–720. doi:10.1109/TVCG.2013.42.

Benitez-Quiroz, C. F.; Srinivasan, R.; and Martinez, A. M., 2016a. EmotioNet: An Accurate, Real-Time Algorithm for the Automatic Annotation of a Million Facial Expressions in the Wild. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 5562–5570. IEEE, Las Vegas, NV, USA. doi:10.1109/CVPR.2016.600.

Benitez-Quiroz, C. F.; Srinivasan, R.; and Martinez, A. M., 2016b. Emotionet: An accurate, real-time algorithm for the automatic annotation of a million facial expressions in the wild. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 5562–5570. doi:10.1109/CVPR.2016.600.

Blanz, V. and Vetter, T., 1999. A morphable model for the synthesis of 3d faces. In *Proceedings of the 26th annual conference on Computer graphics and interactive techniques*, 187–194.

Booth, J.; Antonakos, E.; Ploumpis, S.; Trigeorgis, G.; Panagakis, Y.; and Zafeiriou, S., 2017. 3d face morphable models" in-the-wild". In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 5464–5473. IEEE.

Cabanac, M., 2002. What is emotion? *Behavioural Processes*, 60, 2 (Nov 2002), 6983. doi:10.1016/S0376-6357(02)00078-5.

Cao, C.; Weng, Y.; Zhou, S.; Tong, Y.; and Zhou, K., 2013. Facewarehouse: A 3d facial expression database for visual computing. *IEEE Transactions on Visualization and Computer Graphics*, 20, 3 (2013), 413–425.

Cao, C.; Weng, Y.; Zhou, S.; Tong, Y.; and Zhou, K., 2014. FaceWarehouse: A 3D facial expression database for visual computing. *IEEE Transactions on Visualization and Computer Graphics*, (Mar. 2014), 413–425. doi:10.1109/TVCG.2013.249.

Carrier, P.-L.; Courville, A.; Goodfellow, I. J.; Mirza, M.; and Bengio, Y., 2013. Fer-2013 face database. *Universit de Montral*, (2013).

Chen, C.-H.; Lee, I.-J.; and Lin, L.-Y., 2015. Augmented reality-based self-facial modeling to promote the emotional expression and social skills of adolescents with autism spectrum disorders. *Research in Developmental Disabilities*, (Jan. 2015), 396–403. doi:10.1016/j.ridd.2014.10.015.

Community, B. O., 2018. *Blender - a 3D modelling and rendering package*. Blender Foundation, Stichting Blender Foundation, Amsterdam. http://www.blender.org.

Cowie, R.; Douglas-Cowie, E.; Tsapatsoulis, N.; Votsis, G.; Kollias, S.; Fellenz, W.; and Taylor, J., Jan./2001. Emotion recognition in human-computer interaction. (Jan./2001), 32–80. doi:10.1109/79.911197.

Darwin, C. and Ekman, P., 2009. *The Expression of the Emotions in Man and Animals*. Oxford University Press, Oxford ; New York, 4th ed., 200th anniversary ed edn. ISBN 978-0-19-539228-9.

Dino, H. I. and Abdulrazzaq, M. B., 2019. Facial Expression Classification Based on SVM, KNN and MLP Classifiers. In *2019 International Conference on Advanced Science and Engineering (ICOASE)*, 70–75. IEEE, Zakho - Duhok, Iraq. doi:10.1109/ICOASE.2019.8723728.

Donahue, J.; Jia, Y.; Vinyals, O.; Hoffman, J.; Zhang, N.; Tzeng, E.; and Darrell, T., 2013. Decaf: A deep convolutional activation feature for generic visual recognition.

Dryden, I. L. and Mardia, K. V., 2016. *Statistical shape analysis: with applications in R*, vol. 995. John Wiley & Sons.

Du, S.; Tao, Y.; and Martinez, A. M., 2014. Compound facial expressions of emotion. *Proceedings of the National Academy of Sciences*, (Apr. 2014), E1454–E1462. doi:10.1073/pnas.1322355111.

Ebrahimi Kahou, S.; Michalski, V.; Konda, K.; Memisevic, R.; and Pal, C., 2015. Recurrent neural networks for emotion recognition in video. In *Proceedings of the 2015 ACM on International Conference on Multimodal Interaction*, 467–474.

Ekman, P., 1992. An argument for basic emotions. (1992), 169–200. doi:10.1080/02699939208411068.

Ekman, P., 2006. Darwin, Deception, and Facial Expression. *Annals of the New York Academy of Sciences*, (Jan. 2006), 205–221. doi:10.1196/annals.1280.010.

Fasel, B., 2002. Head-pose invariant facial expression recognition using convolutional neural networks. In *Proceedings. Fourth IEEE International Conference on Multimodal Interfaces*, 529–534. doi:10.1109/ICMI.2002.1167051.

Fasel, B., 2002. Robust face analysis using convolutional neural networks. In *Object recognition supported by user interaction for service robots*, vol. 2, 40–43. IEEE.

Gerig, T.; Morel-Forster, A.; Blumer, C.; Egger, B.; Luthi, M.; Schönborn, S.; and Vetter, T., 2018. Morphable face models-an open framework. In *2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018)*, 75–82. IEEE.

Girshick, R.; Donahue, J.; Darrell, T.; and Malik, J., 2013. Rich feature hierarchies for accurate object detection and semantic segmentation.

Hassner, T.; Harel, S.; Paz, E.; and Enbar, R., 2015. Effective face frontalization in unconstrained images. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, (Jun 2015). doi:10.1109/cvpr.2015.7299058. http://dx.doi.org/10.1109/CVPR.2015.7299058.

He, K.; Zhang, X.; Ren, S.; and Sun, J., 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 770–778.

Horn, B. K. and Schunck, B. G., 1981. Determining optical flow. *Artificial Intelligence*, (Aug. 1981), 185–203. doi:10.1016/0004-3702(81)90024-2.

Hsieh, C.-C.; Hsih, M.-H.; Jiang, M.-K.; Cheng, Y.-M.; and Liang, E.-H., 2016. Effective semantic features for facial expressions recognition using SVM. *Multimedia Tools and Applications*, (Jun. 2016), 6663–6682. doi:10.1007/s11042-015-2598-1.

Huang, Y.; Chen, F.; Lv, S.; and Wang, X., 2019. Facial expression recognition: A survey. *Symmetry*, (Sep. 2019), 1189. doi:10.3390/sym11101189.

Huber, P.; Hu, G.; Tena, R.; Mortazavian, P.; Koppen, P.; Christmas, W. J.; Ratsch, M.; and Kittler, J., 2016. A multiresolution 3d morphable face model and fitting framework. In *Proceedings of the 11th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*.

Jabid, T.; Kabir, M. H.; and Chae, O., 2010. Facial expression recognition using Local Directional Pattern (LDP). In *2010 IEEE International Conference on Image Processing*, 1605–1608. IEEE, Hong Kong, Hong Kong. doi:10.1109/ICIP.2010.5652374.

Jabon, M. E.; Bailenson, J. N.; Pontikakis, E.; Takayama, L.; and Nass, C., 2011. Facial expression analysis for predicting unsafe driving behavior. *IEEE Pervasive Computing*, (Apr. 2011), 84–95. doi:10.1109/MPRV.2010.46.

Jerritta, S.; Murugappan, M.; Nagarajan, R.; and Wan, K., 2011. Physiological signals based human emotion Recognition: A review. In *2011 IEEE 7th International Colloquium on Signal Processing and Its Applications*, 410–415. IEEE, Penang, Malaysia. doi:10.1109/CSPA.2011.5759912.

Ji, S.; Xu, W.; Yang, M.; and Yu, K., 2012. 3d convolutional neural networks for human action recognition. *IEEE transactions on pattern analysis and machine intelligence*, 35, 1 (2012), 221–231.

Khan, G.; Samyan, S.; Khan, M. U. G.; Shahid, M.; and Wahla, S. Q., 2020. A survey on analysis of human faces and facial expressions datasets. *International Journal of Machine Learning and Cybernetics*, (Mar. 2020), 553–571. doi:10.1007/s13042-019-00995-6.

Krishna Gudipati, V.; Ray Barman, O.; Gaffoor, M.; Harshagandha; and Abuzneid, A., 2016. Efficient facial expression recognition using adaboost and haar cascade classifiers. In *2016 Annual Connecticut Conference on Industrial Electronics, Technology & Automation (CT-IETA)*, 1–4. IEEE, Bridgeport, CT, USA. doi:10.1109/CT-IETA.2016.7868250.

Krizhevsky, A.; Sutskever, I.; and Hinton, G. E., 2017. Imagenet classification with deep convolutional neural networks. *Communications of the ACM*, 60, 6 (2017), 84–90.

Kuo, C.-M.; Lai, S.-H.; and Sarkis, M., 2018. A compact deep learning model for robust facial expression recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2121–2129.

Lankes, M.; Riegler, S.; Weiss, A.; Mirlacher, T.; Pirker, M.; and Tscheligi, M., 2008. Facial expressions as game input with different emotional feedback conditions. In *Proceedings of the 2008 International Conference in Advances on Computer Entertainment Technology - ACE '08*, 253. ACM Press, Yokohama, Japan. doi:10.1145/1501750.1501809.

Li, J.; Zhang, D.; Zhang, J.; Zhang, J.; Li, T.; Xia, Y.; Yan, Q.; and Xun, L., 2017a. Facial expression recognition with faster r-cnn. *Procedia Computer Science*, 107 (2017), 135–140.

Li, S. and Deng, W., 2018. Reliable crowdsourcing and deep locality-preserving learning for unconstrained facial expression recognition. *IEEE Transactions on Image Processing*, 28, 1 (2018), 356–370.

Li, S. and Deng, W., 2020. Deep Facial Expression Recognition: A Survey. *IEEE Transactions on Affective Computing*, (2020), 1–1. doi:10.1109/TAFFC.2020.2981446.

Li, T.; Bolkart, T.; Black, M. J.; Li, H.; and Romero, J., 2017b. Learning a model of facial shape and expression from 4D scans. *ACM Trans. Graph.*, (Nov. 2017). doi:10.1145/3130800.3130813.

Liew, C. F. and Yairi, T., 2015. Facial Expression Recognition and Analysis: A Comparison Study of Feature Descriptors. *IPSJ Transactions on Computer Vision and Applications*, (2015), 104–120. doi:10.2197/ipsjtcva.7.104.

Lucey, P.; Cohn, J. F.; Kanade, T.; Saragih, J.; Ambadar, Z.; and Matthews, I., 2010. A complete dataset for action unit and emotion-specified expression. In *Computer Vision and Pattern Recognition Workshops (CVPRW), 2010 IEEE Computer Society Conference on*, 94–101.

Lyons, M.; Akamatsu, S.; Kamachi, M.; and Gyoba, J., 1998a. Coding facial expressions with Gabor wavelets. In *Proceedings Third IEEE International Conference on Automatic Face and Gesture Recognition*, 200–205. IEEE Comput. Soc, Nara, Japan. doi:10.1109/AFGR.1998.670949.

Lyons, M.; Akamatsu, S.; Kamachi, M.; and Gyoba, J., 1998b. Coding facial expressions with gabor wavelets. In *Proceedings Third IEEE international conference on automatic face and gesture recognition*, 200–205. IEEE.

Lyons, M. J.; Kamachi, M.; and Gyoba, J., 2020. Coding Facial Expressions with Gabor Wavelets (IVC Special Issue). *arXiv:2009.05938 [cs]*, (Sep. 2020). doi:10.5281/zenodo.4029679.

Mattela, G. and Gupta, S. K., 2018. Facial Expression Recognition Using Gabor-Mean-DWT Feature Extraction Technique. In *2018 5th International Conference on Signal Processing and Integrated Networks (SPIN)*, 575–580. IEEE, Noida. doi:10.1109/SPIN.2018.8474206.

Mavadati, S. M.; Mahoor, M. H.; Bartlett, K.; Trinh, P.; and Cohn, J. F., 2013a. DISFA: A Spontaneous Facial Action Intensity Database. *IEEE Transactions on Affective Computing*, (Apr. 2013), 151–160. doi:10.1109/T-AFFC.2013.4.

Mavadati, S. M.; Mahoor, M. H.; Bartlett, K.; Trinh, P.; and Cohn, J. F., 2013b. Disfa: A spontaneous facial action intensity database. *IEEE Transactions on Affective Computing*, 4, 2 (2013), 151–160.

Mehrabian, A. and Russell, J. A., 1974. *An approach to environmental psychology*. M.I.T. Press. ISBN 9780262130905.

Michel, P. and El Kaliouby, R., 2003. Real time facial expression recognition in video using support vector machines. In *Proceedings of the 5th International Conference on Multimodal Interfaces - ICMI '03*, 258. ACM Press, Vancouver, British Columbia, Canada. doi:10.1145/958432.958479.

Mollahosseini, A.; Chan, D.; and Mahoor, M. H., 2016. Going deeper in facial expression recognition using deep neural networks. In *2016 IEEE Winter Conference on Applications of Computer Vision (WACV)*, 1–10. IEEE, Lake Placid, NY, USA. doi:10.1109/WACV.2016.7477450.

Mollahosseini, A.; Hasani, B.; and Mahoor, M. H., 2017. Affectnet: A database for facial expression, valence, and arousal computing in the wild. *IEEE Transactions on Affective Computing*, 10, 1 (2017), 18–31.

Pantic, M.; Valstar, M.; Rademaker, R.; and Maat, L., 2005. Web-based database for facial expression analysis. In *2005 IEEE international conference on multimedia and Expo*, 5–pp. IEEE.

Paysan, P.; Knothe, R.; Amberg, B.; Romdhani, S.; and Vetter, T., 2009. A 3d face model for pose and illumination invariant face recognition. In *2009 Sixth IEEE International Conference on Advanced Video and Signal Based Surveillance*, 296–301. Ieee.

Pitaloka, D. A.; Wulandari, A.; Basaruddin, T.; and Liliana, D. Y., 2017. Enhancing cnn with preprocessing stage in automatic emotion recognition. *Procedia computer science*, 116 (2017), 523–529.

Roh, Y.; Heo, G.; and Whang, S. E., 2019. A Survey on Data Collection for Machine Learning: A Big Data – AI Integration Perspective. *arXiv:1811.03402 [cs, stat]*, (Aug. 2019).

Saeed, S.; Baber, J.; Bakhtyar, M.; Ullah, I.; Sheikh, N.; Dad, I.; and Ali, A., 2018. Empirical Evaluation of SVM for Facial Expression Recognition. *International Journal of Advanced Computer Science and Applications*, (2018). doi:10.14569/IJACSA.2018.091195.

Sagonas, C.; Panagakis, Y.; Zafeiriou, S.; and Pantic, M., 2015. Robust statistical face frontalization. In *2015 IEEE International Conference on Computer Vision (ICCV)*, 3871–3879. doi:10.1109/ICCV.2015.441.

Shorten, C. and Khoshgoftaar, T. M., 2019. A survey on Image Data Augmentation for Deep Learning. *Journal of Big Data*, (Dec. 2019), 60. doi:10.1186/s40537-019-0197-0.

Simonyan, K. and Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, (2014).

Smith, W. A.; Seck, A.; Dee, H.; Tiddeman, B.; Tenenbaum, J. B.; and Egger, B., 2020. A morphable face albedo model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 5011–5020.

Sun, Y.; Wang, X.; and Tang, X., 2013. Deep Convolutional Network Cascade for Facial Point Detection. In *2013 IEEE Conference on Computer Vision and Pattern Recognition*, 3476–3483. IEEE, Portland, OR, USA. doi:10.1109/CVPR.2013.446.

Szegedy, C.; Liu, W.; Jia, Y.; Sermanet, P.; Reed, S.; Anguelov, D.; Erhan, D.; Vanhoucke, V.; and Rabinovich, A., 2015. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1–9.

Tang, Y., 2013. Deep learning using linear support vector machines. *arXiv preprint arXiv:1306.0239*, (2013).

Tran, D.; Bourdev, L.; Fergus, R.; Torresani, L.; and Paluri, M., 2014. Learning spatiotemporal features with 3d convolutional networks.

Tran, L.; Yin, X.; and Liu, X., 2017. Disentangled representation learning gan for pose-invariant face recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, 1415–1424.

Tsai, H.-H. and Chang, Y.-C., 2018. Facial expression recognition using a combination of multiple facial features and support vector machine. *Soft Computing*, (Jul. 2018), 4389–4405. doi:10.1007/s00500-017-2634-3.

Viola, P. and Jones, M., 2001. 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*, I–511–I–518. IEEE Comput. Soc, Kauai, HI, USA. doi:10.1109/CVPR.2001.990517.

Vlasic, D.; Brand, M.; Pfister, H.; and Popovic, J., 2006. Face transfer with multilinear models. In *ACM SIGGRAPH 2006 Courses*, 24–es.

Wang, X.-H.; Liu, A.; and Zhang, S.-Q., 2015. New facial expression recognition based on FSVM and KNN. *Optik*, (Nov. 2015), 3132–3134. doi:10.1016/j.ijleo.2015.07.073.

Yin, L.; Wei, X.; Sun, Y.; Wang, J.; and Rosato, M. J., 2006. A 3d facial expression database for facial behavior research. In *7th international conference on automatic face and gesture recognition (FGR06)*, 211–216. IEEE.

Yin, X.; Yu, X.; Sohn, K.; Liu, X.; and Chandraker, M., 2017. Towards large-pose face frontalization in the wild. In *Proceedings of the IEEE international conference on computer vision*, 3990–3999.

Yu, R.; Saito, S.; Li, H.; Ceylan, D.; and Li, H., 2017. Learning dense facial correspondences in unconstrained images. In *Proceedings of the IEEE International Conference on Computer Vision*, 4723–4732.

Yu, Z. and Zhang, C., 2015. Image based static facial expression recognition with multiple deep network learning. In *Proceedings of the 2015 ACM on international conference on multimodal interaction*, 435–442.

Zafeiriou, S.; Kollias, D.; Nicolaou, M. A.; Papaioannou, A.; Zhao, G.; and Kotsia, I., 2017. Aff-wild: Valence and arousal'in-the-wild'challenge. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 34–41.

Zhang, K.; Zhang, Z.; Li, Z.; and Qiao, Y., 2016. Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks. *IEEE Signal Processing Letters*, (Oct. 2016), 1499–1503. doi:10.1109/LSP.2016.2603342.

Zhao, G.; Huang, X.; Taini, M.; Li, S. Z.; and PietikäInen, M., 2011. Facial expression recognition from near-infrared videos. *Image and Vision Computing*, 29, 9 (2011), 607–619.

Zhenhua Guo; Lei Zhang; and Zhang, D., 2010. A Completed Modeling of Local Binary Pattern Operator for Texture Classification. *IEEE Transactions on Image Processing*, (Jun. 2010), 1657–1663. doi:10.1109/TIP.2010.2044957.