Interim Report of Final Year Project

HKU-Face: A Large Scale Dataset for Deep Face Recognition

Haicheng Wang
3035140108
Haoyu Li
3035141841

COMP4801 Final Year Project
Project Code: 17007
Abstract

Current development of face recognition usually encounters problems with its training dataset because of the small size and human labelling errors. This project proposes a general dataset construction and filtering process in order to tackle the problem efficiently. Several models in the literature are utilized but for the new purpose to filter the original dataset collected from the website. Current results show the impressive effectiveness of automatic filtering and purity enhancement after filtering. Subsequent research and experiment are needed for the further improvement of filtering process with lower false negative rate. After the completion of the project, facial dataset constructions are expected to accelerate with less human effort. Further studies based on it is expected to contribute more to the unsupervised learning in the general object recognition.
Acknowledgments

We want to thank our advisor, Prof. Yu for his supervision and help in providing computational resources. Thanks to Weifeng Ge for his instructive guide through the project. We would also want to thanks all classmates that help us in any perspective of our project.
## Contents

- **Abstract** ii
- **Acknowledgements** iii
- **Contents** iii
- **List of Figures** v
- **List of Tables** vi

### 1 Introduction
- 1.1 Background ................................................. 1
- 1.2 Problem Definition ........................................ 1
- 1.3 Objectives .................................................. 2
- 1.4 Outline of the Report .................................. 3

### 2 Literature review
- 2.1 Theoretical Background ................................ 4
- 2.2 Dataset Construction .................................... 5
- 2.3 Face Recognition Models ................................. 5
- 2.4 Benchmark for Face Recognition ..................... 5

### 3 Methodology and Approach
- 3.1 Original Dataset Construction ...................... 6
- 3.2 Dataset Filtering ......................................... 7
- 3.3 Models in Automatic Filtering ....................... 8
- 3.4 Evaluation of Face Recognition Performance ...... 10

### 4 Current Progress
- 4.1 Data Collection and Preprocessing .................. 11
- 4.2 Dataset filtering .......................................... 12
- 4.3 Major Problem ............................................ 14
- 4.4 Other Potential Difficulties .......................... 15
5 Future Works

5.1 Mitigation Strategy .......................................................... 16
5.2 Schedule ...................................................................... 16

6 Summary ........................................................................ 18

Bibliography ....................................................................... 19
List of Figures

1.1 Illustration of dataset without filtering (in one class). .......................... 2

3.1 Illustration of occupational distribution of celebrity in MS-Celeb-1M[2]. ........ 7
3.2 Initial state and the learning process driven by triplet loss. ....................... 8
3.3 The left column shows 2D and 3D cases for angular softmax loss margin, where the orange color represents the class 1 and green color represents class 2, which are not only classified but also separated by a margin. The corresponding parameters are discussed in SphereFace paragraph. The right-upper graph illustrates the A-softmax margin effect when selecting $m = 4$ after projecting features into 3. The right-lower picture shows the distribution of angles under this setting .................. 9

4.1 Illustration of clustering problem .......................................................... 13
4.2 Purity distribution of benchmark set using newly generated features ............ 14
List of Tables

1.1 Face recognition datasets .................................................... 2

4.1 Statistics of benchmark for auto filtering .......................... 12
4.2 Purity level of benchmark dataset ......................................... 14

5.1 Schedule for subsequent work ............................................... 17
Chapter 1

Introduction

1.1 Background

Face recognition requires computers to recognize people with their identities through the images of faces. It is one of the most representative tasks of Artificial Intelligence and has been receiving great academic and industrial attention for several years. Human beings have already been beaten in some academic benchmarks since the emergence of state-of-the-art models like Gaussian-Face and FaceNet. Face recognition has been applied in various areas like identity authentication in smart home, electronic devices unlocking and procedure recording, benefiting the industry world with its increasing accuracy and low cost.

1.2 Problem Definition

In spite of the outstanding achievements of Face Recognition so far, several limitations continue to exist. Firstly, current academic models are quite sensitive to variations like illumination, different perspective of faces, and potential obfuscation thus they cannot perform well in front of a large number of test cases. Secondly, the industrial companies have to solve the issue of adversarial samples which can be easily designed to crack the face authentication system. As a research project, we focus on the potential mitigation strategies on the first one.

The issue of sensitivity results from the nature of the current models. Almost all of the models are trained by supervised learning which uses dataset annotated by human beings. Two following drawbacks contribute to the problem of sensitivity: (1) limited size of dataset due to limited human effort; (2) accuracy problem resulted from human perceptual bias.

Parkhi, Vedaldi, Zisserman, et al. discuss the limitation on the small training dataset, showing that giant companies hold private face database with larger size of data, while other research institution could only get access to public but much smaller database for training such as CASIA-WebFace, which acts like a barricade to even higher performance (See the comparison in Table 1.1). However, the largest private database is still small compared
CHAPTER 1. INTRODUCTION

2

Figure 1.1: Illustration of dataset without filtering (in one class).

with the size of images on the Internet because it can be inferred impossible to annotate face database in trillion size.

<table>
<thead>
<tr>
<th>Dataset</th>
<th>Availability</th>
<th>identities</th>
<th>images</th>
</tr>
</thead>
<tbody>
<tr>
<td>LFW [6, 7]</td>
<td>public</td>
<td>5K</td>
<td>13K</td>
</tr>
<tr>
<td>YFD [25]</td>
<td>public</td>
<td>1595</td>
<td>3425 videos</td>
</tr>
<tr>
<td>CelebFaces [21]</td>
<td>public</td>
<td>10K</td>
<td>202K</td>
</tr>
<tr>
<td>CASIA-WebFace [28]</td>
<td>public</td>
<td>10K</td>
<td>500K</td>
</tr>
<tr>
<td>MS-Celeb-1M [2]</td>
<td>public</td>
<td>100K</td>
<td>about 10M</td>
</tr>
<tr>
<td>Facebook</td>
<td>private</td>
<td>4K</td>
<td>4400K</td>
</tr>
<tr>
<td>Google</td>
<td>private</td>
<td>8M</td>
<td>100-200M</td>
</tr>
</tbody>
</table>

Table 1.1: Face recognition datasets

In Microsoft paper[2], the second problem has been addressed in an indirect manner by fault-tolerance classifier models. Therefore, training dataset in MS-Celeb-1M[2] is only obtained from popular search engine without any filtering or labeling checking. As Figure 1 shows, these images should belong to one identity but obviously, there are men’s and women’s faces and even a non-face image. The performance of the classifier is highly likely to be better if the dataset were purified.

1.3 Objectives

This project is intended to tackle the issue of sensitivity from the perspective of dataset construction. A general face dataset development procedure will be proposed as well as a
dataset and trained models for the evaluation. The procedure should be able to construct
large-scaled face datasets efficiently with controllable noise. Retrained or fine-tuned on these
datasets, previous models are more likely to reach better performance. What’s more, similar
procedures can be generalized into other image dataset construction processes, which may
yield benefit to other object recognition tasks.

1.4 Outline of the Report

In the rest of this interim report, several key points are discussed. In Chapter 2, Literature
review introduces the theoretical background with essential prerequisite of our project.
Methodology part in Chapter 3 describes how the inspiration from previous work transforms
to our key approach and their implementation details. Current process is shown in the Chap-
ter 4 with empirical evaluations. Difficulties and intended curies are presented as guidelines
for subsequent work, followed by a summary section with future directions.
Chapter 2

Literature review

In this chapter, several key components related to this project are explored. Theoretical background presents the basic task definitions of Face Recognition and Metric Learning. Then methods related to dataset construction, face recognition models and facial benchmarks are elaborated and several existing works are reviewed in details.

2.1 Theoretical Background

Face Recognition  In essence, face recognition is a specific verification task which evaluates the similarity between the query set and the gallery set. Consider gallery set containing training images, \( G = \{g_0, g_1, \cdots, g_n\} \), and denote identity set \( \hat{I} = \{I_1, I_2, \cdots, I_m\} \) where all elements form a partition of \( G \). Closed set face Recognition is to find model \( f(x) \) mapping from testing set \( T \) to identity set \( \hat{I} \), namely to find the identity of every testing image. More generally, an open-set face recognition task removes the constraint that query set must only contain images in training identity. That requires the model to be capable of telling whether it has ”seen” the identity before or not.

Metric learning  Open set problem is more challenging due to its nature which is demanding of general discriminative features. Metric learning shows its superior in this problem. Instead of directly mapping from image instances to identities, metric learning model firstly maps the input to a general feature \( h \) which theoretically lies in a hyperspace manifold\(^\text{[13]}\). Then training a discriminator \( D \) to accomplish the task

\[
D(h_i, h_j) = \begin{cases} 
0, & \text{if } h_i, h_j \text{ do not belong to the same identity} \\
1, & \text{if } h_i, h_j \text{ belong to the same identity}
\end{cases}
\]

The verification task involving a latent space empirically yields better performance as Parkhi, Vedaldi, Zisserman, et al.\(^\text{[18]}\) shows, especially for deep metric learning which harnesses deep neural network to handle the latent feature extraction phase.
2.2 Dataset Construction

Our initial plan was inspired by [18], in which a general procedure of dataset construction is proposed. It provides a well-designed framework to build and evaluate a facial dataset as well as meaningful heuristics for image purification. The data construction part of our project borrows the framework but we try to utilize more advanced feature extraction models and explore filtering automatically. Also we will build new benchmark specifically for the evaluation of the automatic filtering performance which receive less consideration in the literature.

As Lapuschkin et al. [11] shows, the fisher vector used in [18] generally tends to rely more on contextual features rather than overall structural features which have been explored better using deep learning model. This observation shows the feasibility of replacing fisher vectors by deep neural networks.

The loop of automatic filtering and human purifying process indicates the potential potency of a iterative process of handling filtering process. Apart from the iterative collaboration of computer and human effort, purification by iterative automatic filtering and model fine-tuning could be possibly useful.

2.3 Face Recognition Models

This project mainly focus on models related to deep metric learning. Facenet [20] directly maps image instances into Euclidean space which provides convenience for applying direct clustering on the features of faces. Sphereface [14] interprets the embedding extracted in the view of angular distance with effective margins. Furthermore, the general margins in embedding space are elaborated and analyzed by [26]. More details are elaborated in section 3.3.

2.4 Benchmark for Face Recognition

In the literature, LFW [6, 7] is a classical dataset for evaluation of face recognition performance, and Mega-face [9] is a generally new and more challenging benchmark because of more variations and distractors. Recently MS-Celeb-1M [2] has been proposed with more authenticated testing set assisted with knowledge-based annotation.
Chapter 3

Methodology and Approach

Three phases are included in the project. The first phase is a pipeline of the construction of the original dataset. The second one is the key filtering process which involves automatic methods and strategies to remove the noise in the original dataset. With the filtered dataset, the third phase includes model training and testing as the performance evaluation for the dataset. Three phases are elaborated below.

3.1 Original Dataset Construction

Data Crawling

Original image data should be collected from Google image according to a newly designed celebrity list. The celebrity list is expected to contain about 50K identities with more variance. For example, multiple occupations of celebrities are key metrics of identity variation. The MS-Celeb-1M (Figure 3.1) highlights its superiority in occupation variation to [18] whose celebrity list only comes from movies actors and actresses. Therefore, the celebrity list of our dataset should have sufficient variation in occupations. Moreover, we will explore to extend the diversity of nationality of the celebrities if time and resource allow.

Dataset Preprocessing

After getting the raw data from Google image search, one step face preprocessing should be done in order to fit into the domain of face recognition. There are many existing models like ACW[27] and MTCNN[29]. ACW is chosen for its wide application as well as its comparative stability.

According to the celebrity list, top 100 images per identity in Google Image are scratched. Giving credence to search engine ranking is a general technique across research and industrial institutions[7, 6, 2]. During the process of image scratching, the issue of anti-robot program and potential network limitation should be taken into consideration.
CHAPTER 3. METHODOLOGY AND APPROACH

3.2 Dataset Filtering

Model Selection  Based on current trends on face recognition described in the last chapter, it is empirically reasonable to choose FaceNet\(^ {20}\) and SphereFace\(^ {14}\) for the model selection. Both of them are trained models which bring convenience for model evaluation and direct application. At the same time, more models such the ones driven by center loss\(^ {23}\) will be explored for more variations of filters. Using ensemble learning which is basically a fusion of different models, it is more likely to achieve better filtering performance with more distinctive model filters. There are two possible strategies for model fusions. Firstly, features extracted from different models can be concatenated and then classified by a new model. Secondly, a voting system could be constructed and remove the particular item when the majority of models prefer to remove it.

Evaluation of filtering performance  Before conducting the real filtering of the whole dataset, a small but useful benchmark dataset will be built to evaluate the filtering performance of selected models. Concretely, 100 identities are selected randomly from the whole dataset. Then, all images in each identity folder will be labeled into either positive or negative group based on the human judgment on whether they should be filtered out or not. The labeling process should receive enough iteration of examination and cross-checking such that inter-class variations are reserved and ambiguities are reduced to lowest.

Apart from the intention of evaluating existing models filtering performance, another target can be achieved, which is to find appropriate clustering parameters for the best filtering potency of the model. This parameter tuning process can be easily automated by the grid search algorithm\(^ {12}\).

Filtering process  The real filtering process contains two parts: (1) Feature (or embedding) extraction from the original dataset using models selected from last step and (2) Clustering image instances according to the corresponding algorithm with carefully chosen
parameters. Feature extractions can be done by the built-in function in Caffe\[8\] to extract data from any specific layer in the model. The clustering process adopts the method proposed in \[19\], which is a modified version from K-means\[17\], by tuning two key metric \textit{delta} and \textit{density}\[19\].

**Quality control of filtered dataset** The evaluation of filtering is the quality control for the iterative filtering process. After one iteration of filtering, a set of randomly selected data from the output is evaluated according to its remaining noise proportion. A reasonable threshold should be determined for the final pass of the filtering process with empirically rigorous exploration.

### 3.3 Models in Automatic Filtering

**FaceNet**

FaceNet utilizes an innovative loss function called "triplet loss", which calculates the difference using three samples: $Q^a_i, Q^p_i$ and $Q^n_i$. $Q^a_i$ is the query image, which is also called anchor, from a random person. $Q^p_i$ and $Q^n_i$ are the images from the same person (positive) and another person (negative) respectively. \[20\] tries to enforce a margin between positive classes and negative classes by minimizing the loss function. As visualized by figure 3.2, if the metric distance between anchor point and positive point is smaller than that between anchor point and negative point, we are fine with it and set loss to be 0. On the other hand, if the relationship is reversed, we need to add a non-zero loss so that the model can learn to enlarge distance between different classes and lessen distance in same class respectively.

![Figure 3.2: Initial state and the learning process driven by triplet loss.](image)

A more precise mathematical expression for triplet loss is given below. Aside from what we have mentioned above, Triplet Loss tried to solve open-set problem by adding a parameter $\alpha$ to its loss function, which not only classifies images but also enforces a margin between different classes.

$$
\text{Triplet loss} = \sum_i^n \left[ ||f(Q^a_i) - f(Q^p_i)||_2^2 - ||f(Q^a_i) - f(Q^n_i)||_2^2 + \alpha \right] +
$$
Figure 3.3: The left column shows 2D and 3D cases for angular softmax loss margin, where the orange color represents the class 1 and green color represents class 2, which are not only classified but also separated by a margin. The corresponding parameters are discussed in SphereFace paragraph. The right-upper graph illustrates the A-softmax margin effect when selecting $m = 4$ after projecting features into 3. The right-lower picture shows the distribution of angles under this setting.

**SphereFace**

SphereFace\[14\] derives an angular expression from traditional softmax loss function and leads to more discriminative features. Softmax loss is one kind of decision criteria that gives posterior probabilities for each sample.

$$P_i = \frac{\exp(W_i^T x + b_i)}{\sum_j \exp(W_j^T x + b_j)}$$

$$L = \frac{1}{N} \sum_i -\log\left(\frac{\exp(W_i^T x + b_i)}{\sum_j \exp(W_j^T x + b_j)}\right)$$

A general softmax loss posterior probability formula and softmax loss function are given in formula above. It means that the probability of the input data $x$ belonging to class type $i$ equals $P_i$. $W_i^T$ is the i-th column for the weight matrix $W$ in softmax layer and $b_i$ is the
i-th entry for the bias vector. For instance, in binary case, the probability for input to be class 1 is

\[
P_1 = \frac{\exp(W_1^T x + b_1)}{\exp(W_1^T x + b_1) + \exp(W_2^T x + b_2)}
\]

SphereFace further utilizes the property of angular expansion by decomposing \(W_1^T x + b_1\) term into \(|W_1| \cdot |x| \cdot \cos(\theta_i) + b_i\). To simplify calculation, it normalizes weights (\(|W_i^T| = 1\)) and initializes bias to zero (\(b_i = 0\)). Using the same margin setting technique, SphereFace adds another pre-defined parameter \(m\) to its degree term \(\theta_i\). \(\theta_i\) controls the size of angular margin and the author derives a function \(\phi(\theta_{y,i}) = (-1)^k \cos(m \theta_{y,i}) - 2k\) to make it optimizable. \(\theta_{y,i}\) is the angle between \(W_{y_i}\) and \(x_i\) with \(K \in [0, m - 1]\). Omitting some details here, we obtain the final expression of angular softmax loss (A-Softmax) as shown below. The 2-Dimension and 3-Dimension visualizations of SphereFace are shown in Figure 3.3.

\[
L_{ang} = \frac{1}{N} \sum_i \left[-\log\left(\frac{\exp(|x_i| \phi(\theta_{y,i}, i))}{\exp(|x_i| \phi(\theta_{y,i}, i)) + \sum_{j \neq y_i} \exp(|x_i| \cos(\theta_j, i))}\right)\right]
\]

### 3.4 Evaluation of Face Recognition Performance

**Training existing models on the constructed dataset**  FaceNet\cite{20} is a standard model to be retrained or fine-tuned on the newly constructed and filtered dataset. Re-training step basically helps evaluate the quality of the dataset and the improvement of the model performance will act as the evidence of the new datasets effectiveness. To make the evidence more well-rounded and convincing, tests should be conducted on more models including but not limited to contrastive loss\cite{3}, triplet loss\cite{22} and angular distance\cite{14}.

**Designing and training new models**  New models would be explored and proposed based on deeper understanding of current metric learning models, which vary on network architectures and loss designs. Incorporating new feature extraction architectures such as ResNet\cite{4} or DenseNet\cite{5} could bring features with even higher quality in terms of more crucial abstractions and higher sensitiveness. Trying innovative loss design would lead to better target function and appropriate updating directions.

**Testing with existing benchmarks**  After training with filtered data, models should be tested and evaluated by existing face recognition benchmarks. LFW\cite{6, 7} provides a direct reference to most of the models in the literature. MegaFace\cite{9} will evaluate model performances in a challenging environment. MS-Celeb-1M\cite{2} helps to test the models potency on face data in the accurate real world.
Chapter 4

Current Progress

Currently, the project is in the second phase as the schedule specifies so the results section only covers all the output info until the second phase. Difficulties and risks are also discussed followed by the experiments output.

4.1 Data Collection and Preprocessing

It took about 2 weeks to finish all data collection tasks including celebrity selection and image scratching. Then the face data preprocessing, which is face alignment and detection, was completed by four-week automated processing driven by a Bash script.

Python scripts were harnessed to get around 40K identities from the Wiki celebrity page [24] for its advantage in web programming. The detailed celebrity list and image metadata will be available after the publication of the pipeline. Then the image crawling followed by the proposed method in Chapter 3 was conducted subsequently by the joint work of 4 local computers in the instruction lab 311 in Haking Wong Building, HKU. The parallel work reduced the time cost of this task by around 75%.

Particularly, the crawler program solved the issue of anti-robot counterpart and network limitation by employing strategic slowing down. By sleeping around 0.1 second during iterative request sending driven by the browser simulator, we successfully get around the network control in the CS Department and Google reCAPTCHA (an anti-robot system). Detailed code are available on Github[1]

The parallel method was also used in the face alignment and detection process but it only involved in one server computer with multiprocessing. Because the detection software required a special environmental setting, local computers in the lab mentioned above is costly to perform this task. The preprocessing process was found to contribute the filtering process because for some extreme cases like no face or too many faces in one image could be dropped by the failure of single face alignment and detection. Around 30% of images are filtered out in the detection and alignment process.

[1] https://github.com/whcacademy/imageDownloader
4.2 Dataset filtering

As the key of this project, each step of dataset filtering should receive great attention. The project has been in this phase for about 4 months (from Oct. 2017 to Jan. 2018) and several achievements made are described in the following.

Benchmark Construction

The small benchmark, which is the filtering performance indicator, was built by joint work of team members. 100 identities with about 6,000 images (40% of images were filtered out by preprocessing step) were randomly selected, labeled and cross-checked before being used by the model evaluation later. After manual labeling, around 80 of 100 identities are effective and valid. The rest of the identities do not have enough images or only contain too much noise. To guarantee an efficient filtering accuracy measurement process, we deleted these invalid identities. Because of its human-manageable size, labeling process was not affiliated by labeling application and directly scripted by hand, small content maintaining scripts are used to help to simplify the label file so team members only needed to fill one type of tag and leave another type empty to be auto-filled. The detailed statistics of the benchmark is listed in the following table.

<table>
<thead>
<tr>
<th>Image size</th>
<th>Identity size</th>
<th>Average Image per Identity</th>
<th>initial purity level</th>
</tr>
</thead>
<tbody>
<tr>
<td>5200</td>
<td>82</td>
<td>63.4</td>
<td>58%</td>
</tr>
</tbody>
</table>

Table 4.1: Statistics of benchmark for auto filtering

Model Evaluation

Currently, three rounds of filtering process have been conducted using FaceNet pre-trained model. The purity scores have been evaluated on our small benchmark. FaceNet model achieved 99.63% accuracy on LFW dataset, which was previously state-of-the-art score. The model was used in extracting 128-dimensional features from images in benchmark dataset. Following that, an unsupervised clustering algorithm was adopted in these features in order to filter out noisy images.

1. First Round: Cluster-based filtering

In our initial trials of experiments, the filtering strategy was to remove irrelevant clusters generated by density peak clustering algorithm\[19\]. A major cluster was selected to be reserved and images in other clusters were discarded. An example of the clustering output is shown in Figure 4.1. However, the density of red points, which should
be preserved by the filtering process range from low to high values. It reveals that current choice of filtering rule does not satisfy our purification requirement.

2. Second Round: Density-based filtering

With the experience learned from our first filtering experiment, we realized the problem in our method. To filter cluster out is not the correct way for filtering noisy images. Some correct images with high density values were also discarded as they did not belong to the major cluster. More serious problem is that many noisy images with low density blend into the major cluster, which did harm to the purity of dataset. To solve these problems, a density-based filtering method was adopted in our second round of filtering. The features put into clustering algorithm were identical to the first round. After our experiments, we chose the density threshold to be 8 and images with density smaller than this value were removed from dataset. This method improves the mean purity of benchmark set from 58% to 63%. Since the target purity of our dataset is at least above 90%, new filtering approach needs to be proposed.

3. Third Round: Uniformly sized images

To further increase the purity value of our dataset, we put attention to the feature extraction procedures. After face alignment, facial parts were detected and extracted from images in raw dataset. These parts were utilized as refined images and put into FaceNet model for feature extraction. However, these images do not have uniform sizes when feeding into neural network. If the image size was smaller than the input requirement, i.e. $224 \times 224$, then it would be resized to that scale. When image size is greater than $224 \times 224$, the methodology FaceNet used was to conduct random cropping to images. We realized that the quality of generated features was effected since facial part of images may be removed after random cropping. The solution we adopted to this problem was to manually unify the image size. By preserving the central
224×224 square of aligned images, the facial part is reserved. Quality of features should be better without the negative influence by random cropping. In our experiments, using the same clustering algorithm with newly generated features, the purity value on benchmark set can achieve around 90%, which is an evident improvement. The distribution of filtering outcome is visualized by boxplot in Figure 4.2. We can see that more than three quantile of identities have above 90% purity value. In fact, most of these identities have completely purified dataset. Some data points in the graph have low values because their raw datasets are originally jumbled with few correct images. Table 4.2 shows the summary of three purity after filtering. Next step of this project is to experiment on new feature extraction methods such as SphereFace[14]. If several models can extract discriminative features, ensemble methods will be tried for further boosting the result.

<table>
<thead>
<tr>
<th>Original Purity</th>
<th>After first round</th>
<th>After second round</th>
<th>After third round</th>
</tr>
</thead>
<tbody>
<tr>
<td>58%</td>
<td>58%</td>
<td>63%</td>
<td>90%</td>
</tr>
</tbody>
</table>

Table 4.2: Purity level of benchmark dataset

4.3 Major Problem

Although the purity enhancement is quite salient after three round of filtering, there are several possible improvement of the automatic filtering process. Firstly, the purity enhancement is build on the sacrifice of removing a number of positive items which should not be filtered out. Then, the purity after filtering is still improvable and metric learning is quite sensitive to any small noise in the training set. Therefore, the filtering task need more effective rounds of cleaning.

There are several possible causes for the problem.
CHAPTER 4. CURRENT PROGRESS

1. **Face recognition model with less variation** The first round of filtering only incorporated single face recognition model which is the FaceNet\[20\]. A reasonable conjecture is that the model works well on particular set of faces rather than the larger and more varied dataset in our project so that it could falsely remove more positive instances.

2. **Face recognition model with domain adaption problem** The dataset filtering process is similar to another task in the literature called domain adaption\[1\]. This problem in transfer learning usually requires further adaptive training with a new dataset. However, in this project the existing model which training on the source domain is directly utilized without adaption.

4.4 Other Potential Difficulties

Several other potential risks should be concerned and mitigated with proper strategies.

Firstly, the celebrity list is designed by auto crawling from the Wikipedia celebrity list\[24\] and its quality could be questioned. However, since the general pipeline of dataset construction is independent from the celebrity source, the project could be adapted to any other celebrity source to construct dataset with high-quality source identity list like the MS-Celeb-1M\[2\].

Secondly, the development process now could be even advanced over the schedule but the less commitment of both two team members and sometime server issues may slow down the development and experiment progress. But the remaining task is always in a controllable size so it is sure that the project will achieve satisfactory result according to the schedule.

In sum, current and foreseeable difficulties are not serious and corresponding mitigation strategy is well-prepared for each risk of the project.
Chapter 5

Future Works

As mentioned in the last chapter, the main difficulty of the current project is the improvable filtering performance compared to the target. This section includes specific strategies that can be applied in the following months. A detailed schedule is presented at the end of this chapter.

5.1 Mitigation Strategy

The following strategies are specifically designed for the potential causes in the Chapter 4 section 4.3.

1. **More filter models and model fusion** The mechanism of involving more face recognition models is similar to trying several clustering algorithm in the previous point. The variance comes from different model’s nature as well as different training dataset. What’s more, the model fusion

2. **Fine-tuning models with domain adaption techniques** Specific components is going to be added into the current face recognition models to conduct the domain adaption training. For example, the Deep Adaption Networks\(^{15}\) can be appended as the last part of the FaceNet\(^{20}\) to tackle the potential different of source dataset(original dataset the model trained on) and target dataset(our new dataset).

5.2 Schedule

The following is the proposed schedule from Jan 2018 to the end of this academic year. The major research attention will be put in the several iterations of image filtering.
<table>
<thead>
<tr>
<th>Month</th>
<th>Tasks</th>
</tr>
</thead>
</table>
| January, 2018 | (1) Wrapping up and submission for the interim report.  
|             | (2) Adding SphereFace and Center loss model for automatic filtering  
|             | (3) Develop scheme for model fusion and test it in the benchmark dataset |
| February   | (1) Model training and testing, research and development on the new model.  
|            | (2) Paper draft.                                                      |
| March      | (1) Refining of the model designing.                                   
|            | (2) Paper submission.                                                  |
| April      | (1) Wrapping up the whole project.                                    
|            | (2) Submission of the final report and final presentation.             |

Table 5.1: Schedule for subsequent work
Chapter 6

Summary

In this interim report, the detailed information about the project as well as its current process are presented.

Current research trends in the face recognition show the importance of the dataset in terms of its size and variation, which acts as the main motivation of the project. Literature review chapter which covers related works gives hints on the general procedure of constructing a dataset with filtering opportunity. By a careful composition with developed models, the filtering process has shown its first purity enhancement on the benchmark dataset and it is likely to achieve higher filtering effect with further investigation and experiment. The final testing phase after the dataset filtering will be conducted and evaluated in the second semester and the possibility of achieving better recognition performance is expected to be high.

This project mainly focuses on the general dataset building procedure and it is inspirational to find that current models can be utilized as the filters to purify the original dataset with noise. Further study on the accuracy improvement of the filter can be significant for the development of face recognition, especially for the unsupervised learning with higher data requirement.
Bibliography


