UNIVERSITY OF TARTU
Institute of Computer Science
Software Engineering Curriculum

Abdulateef Olamide Alli

Extremely low quality image face recognition

Master's Thesis(30ECTS):

Supervisor

Prof. Gholamreza Anbarjafari

Tartu, 2019

Abstract

Extremely low quality image face recognition

Image processing and analysis has evolved over the years into providing practical solutions to everyday challenges. The birth of new solutions and proposals also create new challenges usually surrounding the new innovations.

Existing face recognition algorithms have performed well and they have been deployed into solutions such as social media image tagging, mobile phone facial bio-metric authentication, immigration border control face matching among other solutions. The existing algorithms have been able to perform well in these scenarios because the quality of image from these use cases are usually of high quality with high resolution (HR) [1]. In other possible application of face recognition such as city camera surveillance, airport security surveillance and other related scenarios where image stream quality cannot be directly controlled or manipulated, it becomes imperative to seek a more robust solution that can deal with face recognition regardless of the frame size, lighting condition, race, age, pose and other varying factor that can significantly change the way the images are perceived by existing algorithms.

The goal of this thesis is to identify and test alternative methods of performing face recognition task in extremely low quality images.

Keywords: Face Recognition, Super Resolution, Extremely Low Quality Image

CERCS: T111

Näo tuvastamine eriti madala kvaliteediga pildil

Aastate jooksul on piltide töötlemine ja analüüs arenenud pakkudes nüüd igapäevastele väljakutsetele praktilisi lahendusi. Uute lahenduste ja ettepanekute sünd toob kaasa ka uusi väljakutseid, mis on paratamatult seotud innovaatiliste uuendustega. Olemasolevad näotuvastuse algoritmid on hästi toiminud ja neid on muu hulgas rakendatud sellistes lahendustes nagu sotsiaalmeedia kujutise märgistamine, mobiiltelefoni näo biomeetriline autentimine ja sisserände piirikontrolli näotuvastus. Põhjus miks need algoritmid on suutnud eelnimetatud stsenaariumides hästi toimida tuleneb sellest, et kasutuskõlblike kujutiste kvaliteet on tavaliselt kõrge eraldusvõimega [1].

Teistes näidetes kus näotuvastus vajalikuks osutub nagu linna turvakaamerad, lennujaama kaamerad ja muud situatsioonid kus kujutise salvestuskvaliteeti ei saa kontrollida või manipuleerida, muutub jõulisema lahenduse leidmine pea kohustuslikuks, et oleks võimalik nägu tuvastada sõltumata kaadri suurusest, valgusoludest, rassist, vanusest, kehaasendist või muudest varieeruvatest faktoritest, mis võivad oluliselt muuta algoritmide võimet kujutistest aru saada.

Käesoleva töö eesmärk on tuvastada ja testida alternatiivseid meetodeid näotuvastusülesannete täitmiseks äärmiselt madala kvaliteediga piltides.

Märksõnad: Näotuvastus, superresolutsioon, äärmiselt madal kvaliteet.

CERCS: T111

Contents

List of Tables

List of Figures

Chapter 1

Introduction

1.1 General Overview

Image processing and analysis has evolved over the years into providing practical solutions to everyday challenges [2]. The birth of new solutions and proposals also create new challenges usually surrounding the new innovations.

Face recognition as the name implies seeks to solve the problem of identifying human face(s) from an image, stream of images or a video input. In order to achieve face recognition, a series of step needs to have been carried out. Some of the steps involved in face recognition includes Image understanding/Feature extraction, face(s) detection, face comparison with subject(s) of interest [3, 4, 5, 6, 7, 8, 9, 10, 11].

Existing face recognition algorithms have performed well and they have been deployed into solutions such as social media image tagging, mobile phone facial bio-metric authentication, immigration border control face matching among other solutions. The existing algorithms have been able to perform well in these scenarios because the quality of image from these use cases are usually of high resolution (HR) and they are mostly frontal based images with little or no variations. In other possible application of face recognition such as city camera surveillance, airport security surveillance and other related scenarios where image stream quality cannot be directly controlled or manipulated, it becomes imperative to seek a more robust solution that

can deal with face recognition regardless of the frame size, lighting condition, race, age, pose and other varying factor that can significantly change the way the images are perceived by existing algorithms [12, 13, 14, 15, 16].

Face recognition in regular sized images seems to be entirely different from Face recognition in Extremely low quality images because the target often is not only to get features of faces or comparing existing features with queried features from low resolution (LR) images. In extremely low quality image face recognition, the consideration includes improving the quality of the queried image to an appreciable extent such that it is possible to get discriminating features to compare against.

1.2 Problem Description

The problem with face recognition in extremely low resolution image is described.

Images I are often retrieved from captures taken at a distance, this distance often mean the face of interest will only be a fraction $1/f$ of the actual image captured. These images usually have some noise N in them and the face-image(s) i of interest are often distributed across the actual-image I in smaller resolutions coupled with the noise. The small scale of the face-image i also means that some vital information will be missing from the actual face of interest extracted from the whole image I. An equation that mimics the face image i can be stated as follows:

$$i = \frac{1}{f}(I) + N \tag{1.1}$$

In order to perform face recognition for face image i, we need a predictive model M that is able to discriminate between faces regardless of the quality in image size and noise N introduced. If the face recognition task using our model M is $f(M)$, the result of prediction we seek should be equal when the image is of high resolution $_{(hr)}$ and low resolution $_{(lr)}$.

$$f(M_{lr}) === f(M_{hr}) \tag{1.2}$$

There have been some research works that propose solutions to a related problem, but they often fail to produce best result with extremely low quality images even after they introduce several new steps and become resource intensive (time and computation wise).

1.3 Objectives

The main objective of this research as described through out this master's thesis is to identify and test alternative methods of performing face recognition task when the input image has extremely low quality in its presentation. Other objectives include:

- Research and document the evolution of face recognition. The goal is to understand the challenges that were surmounted and review the improvements made through the evolution.
- Research and Review state of the art in face recognition for extremely low quality images.
- Understand how low quality images are used and the techniques for improving them during image processing tasks. The goal is to identify if this techniques are at all helpful in creating better results with face recognition.
- Identify relevant data-source for face recognition tasks in order to compare results.

1.4 Scope

Face recognition is a broad topic in the field of computer vision as it entails a number of independent researchable areas. A wide variety of publications are available on the traditional application and improvements of face recognition methods. However, most of these research works are focused on high resolution or regular resolution images.
The research in this thesis is limited to reviewing and researching alternative methods engaged when performing face recognition in Extremely low quality images using the popularly available dataset as indicated in later chapters.

Chapter 2

Background Theory

2.1 Evolution of Face Recognition

In order to perform face recognition task, a system has to first complete the task of face detection which entails understanding the image and identification of face or faces, this identification should ideally also include a knowledge about certain attributes of each face(s) reported before it becomes possible to make prediction on the subject or subjects in the image provided.

Early years of research into face-detection saw the creation of a frontal face detection system in real-time video stream. In [17], P. Viola and M. Jones proposed a detection system that does not work directly with image intensity. An integral image representation was devised in order to compute features rapidly at larger scale. Since the Harr-like features are usually too large, a method for constructing a classifier was created by selecting fewer features in an iterative process that is regarded as a feature selection process, after which they apply an incrementally complex classifier in a cascade structure to increase the speed of detection by increasingly focusing attention on more promising regions of the image.

Face recognition should be possible even when frontal face images are not available as input, moving beyond the success of fast frontal face detection, According to [18] and [19], the authors reported the effectiveness of tree-structured model in capturing global elastic deformation as they are also easy to optimize. In [19], Ramanan et al.(2012) were able to achieve a good result

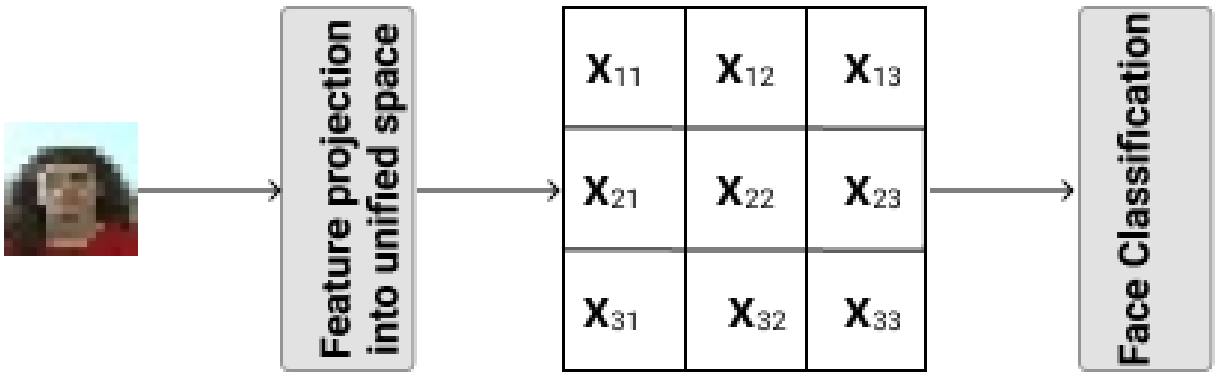

Figure 2.1: A generic workflow for face recognition without super resolution of the input LR image.

in face detection, pose estimation and landmark localization by engaging a fully supervised training for the model where positive images with landmark labels and negative images without faces were provided as training data.

2.2 Face Recognition on Low Quality Images

To solve the problem of low resolution (LR) images in practical face recognition task, these poor quality images need to be understood or at least show a true representation of the original image we intend to query for. Often times, these LR images are preprocessed using Super Resolution (SR) to enhance the images into a better quality before running recognition algorithms on them.

Traditional two-step methods perform super-resolution(SR). Sophisticated SR algorithms are time-consuming and therefore reported not suitable for real-time applications [12]. Subsequent sub-chapters discusses the evolution by categorizing the techniques used in processing low resolution images.

2.2.1 LR-Image face recognition without super resolution

Some research into LR face recognition did not involve Super-resolving LR images. Instead of performing Super-resolution other methods attempted includes image synthesis and learned

Figure 2.2: A two step workflow for face recognition which involves super resolution of the input LR image.

mapping. In [20], low resolution input images were projected onto a decision boundary of support vector data description (SVDD), after which a synthesis of facial images which were obtained from the projection that was done and then face recognition is performed. Figure 2.1 displays a generic workflow that provides an insight into the expectation for image projection into a unified plane.

Coupled Locality Preserving Mapping (CLPM)

A LR face recognition method that is based on coupled mappings (CM) was proposed in [12], the approach taken was to project the data-points from high resolution (HR) and LR feature spaces into a unified feature space by coupled mappings. As depicted in Algorithm 1, this method involves 2 phases, the first phase is referred to as the offline phase, this includes learning the coupled mappings and transformation on HR images and a second phase which is online and consists of querying LR image transformations and feature matching.
According to [12] the problem statement was defined as finding the distance measure between low resolution face image and a high resolution counterpart.

$$d_{ij} = D(I_i, h_j) \tag{2.1}$$

In equation 2.1, d_{ij} represents the distance function for the LR images and $D(I_i, h_j)$ represents the distance function for the HR images.

Algorithm 1 CLPM Algorithm

procedure CLPM - *Offline Phase*

trainingInput ← (LR_{image} + LR_{label} and corresponding HR_{image} + HR_{label})

learnCoupledMappingMatrices(trainingInput)

transformImageToUnifiedFeatureSpace($HR_{enrolledimage}$)

Output: Mapping of LR - HR

procedure CLPM - *online Phase*

queryInput ← ($\text{LR}_{queryimage}$)

transformImageToUnifiedFeatureSpace(queryInput)

inferClassLabelUsingNearestNeighborClassifier()

Output: Class label for query image

Deep coupled ResNet (DCR)

In [21], deep coupled ResNet (DCR) model was used for face recognition. The model consists of a trunk network and two branch networks as shown in Algorithm 2. The trunk network is trained with three significantly different resolutions of face images which are used for the extraction of discriminative features that are robust to the resolution change. The two branch networks work as resolution-specific coupled mappings (CM) used to transform HR and corresponding LR features into a space where difference is minimized. They are trained by HR images and images of the targeted LR.

Algorithm 2 DCR Algorithm

procedure DCR - *Trunk network*
trainingInput ← (Images with varying resolution)
extractDiscriminativeFeatures(trainingInput)

Output: discriminativeFeatures

procedure DCR - *Branch network*
learnCoupledMapping()
projectMappingIntoUnifiedSubspace()

Output: DCR Model

2.2.2 LR-Image face recognition with super resolution

Super Resolution is broadly a set of image processing techniques employed to improve the quality of images by regenerating lost details and adding a clearer perspective to poor quality images.

To deal with Extremely low quality image face recognition, some algorithms enhance the image using super resolution techniques before attempting to predict on the actual image, below are some of the techniques used in super resolving images. In Figure 2.2, we depict a workflow for a traditional two step method involved in face recognition.

Relationship Learning based Super Resolution (RLSR)

Some algorithms that are developed in more recent times are variants of SR.Relationship learning based SR (RLSR) was proposed in [14] to solve the problem of very low resolution images in face recognition. Since visual image quality is important for human-based recognition, a new data constraint was proposed and a new RLSR algorithm was developed for good visual quality image reconstruction. On the other hand, machine-based face recognition will require the extraction of discriminative features, a discriminative constraint was designed and a discriminative SR (DSR) algorithm was proposed. This approach uses linearity clustering to ensure training image pairs have relationships that are linear and in turn reduces the complexity of learning in RLSR and DSR. Algorithm 3 describes a high level algorithm used for RLSR implementation.

Algorithm 3 RLSR Algorithm

procedure RLSR

trainingSet ← imagePairs(LR, HR)

clusteredSet ← *linearityClustering(trainingSet)*

model ← RLSRLearning(*clusteredSet*)

Output: RSLR Model

Ultra Resolution-Discriminative Generative Network (UR-DGN)

Another approach proposes a method to "ultra-resolve" very low resolution images directly[22] regardless of the unexpected variations that might be present in the input images. They introduced a pixel-wise regularization term into a generative model and used the feedback from a discriminative network to make an up-sampled face similar to a real one.

According to [22], the work was inspired by the generative adversarial network (GAN) which consists of two topologies "A generative network G that is designed to learn the distribution of the training data samples and generate a new sample similar to the training data, and a discriminative network D that estimates the probability that a sample comes from the training dataset rather than G. Algorithm 4 shows how the UR-DGN is trained.

Algorithm 4 UR-DGN Algorithm

1: **procedure** UR-DGN
2: *trainingSet* ← imagePairs(LR, HR)
3: $N \leftarrow$ minibatch
4: $K \leftarrow$ iterations
5: **while iter $<$ K do**
6: *Choose a minibatch from image pairs (l, h), i = 1,.... N*
7: *Generate minibatch of HR from l*I
8: *Update parameters of discriminative network D*
9: *Update parameters of generative network G*
10: **end while**

Output: UR-DGN model

Super-Resolution Convolutional Neural Network (SRCNN)

SRCNN was proposed by [23] as a method that learns an end to end mapping between the different resolution of images. SRCNN was reported to achieve fast speed even on CPU. Algorithm 5 depicts steps to perform super resolution with SRCNN, a low resolution image will be upscaled to a target size using bicubic interpolation.

To perform super resolution with SRCNN, a low resolution image will be upscaled to a target size using bicubic interpolation. After upscaling, a relationship mapping is learned and this consists of 3 operations which are Patch extraction and representation, Non-linear mapping then Reconstruction.

This method was compared by [23] to sparse-coding-based method without the non-linear mapping.

Algorithm 5 SRCNN Algorithm

1: **procedure** SRCNN
2: *trainingInput* ← bicubicUpscaled(LR)
3: *patchExtraction(traininitInput)*
4: *nonLinearMapping(trainingInput, HR)*
5: *imageReconstruction*

Output: SRCNN model

Fast Super-Resolution Convolutional Neural Networks (FSRCNN)

A method for "Accelerating the Super-Resolution Convolutional Neural Network" was proposed by [24] to improve the performance of SRCNN with respect to high computational cost and it was subsequently named FSRCNN. Although the FSRCNN was inspired by the success of SRCNN, FSRCNN as shown in Algorithm 6 has a redesigned architecture that learns its mapping directly from a low-resolution image without the need for interpolation(as it exists in SRCNN). The changes FSRCNN introduced includes the following:

- Learns mapping directly from LR Image
- changed mapping layer to shrink input feature dimension before mapping and expand afterwards.
- Using smaller filter sizes but more mapping layers
- Deconvolution layer at the end of the network.

Algorithm 6 FSRCNN Algorithm

1: **procedure** FSRCNN
2: *trainingInput* ← LR-Image
3: *featureExtraction()*
4: *Shrinking()*
5: *Mapping()*
6: *Expanding()*
7: *Deconvolution for Upsampling*

Output: FSRCNN model

Efficient Sub-Pixel Convolutional Neural Network (ESPCN)

This method was proposed by [25], it is a neural network architecture where "feature maps are extracted in LR space".

According to [25], "it introduces an efficient sub-pixel convolution layer which learns an array of upscaling filters to upscale the final LR feature maps into HR output", this is depicted in Algorithm 7.

In ESPCN, LR images are upscaled in the last stage thereby reducing the computation requirements because the network uses small sized feature maps.

Algorithm 7 ESPCN Algorithm

1: **procedure** ESPCN
2: *trainingInput* ← LR-Image
3: *featureExtractionLayers()*
4: *subPixelConvolutionLayer()*

Output: ESPCN model

Deep Laplacian Pyramid Super-Resolution Network (LapSRN)

The LapSRN proposed in [26] takes an LR Image as its input and it performs a progressive prediction of its residual images. Algorithm 8 attempts to state the flow of LapSRN. This

architecture consists of two branches, a feature extraction branch and an image reconstruction branch. [26] describes that its feature extraction branch has convolutional layers using one transposed convolutional layer to upsample the extracted features, with its output connected to two layers. One of these layers is used for reconstructing the residual image and the other is used for extracting features at a finer level. This method also aims to reduce the size of the feature maps by extracting features directly from the low-resolution space just like [24] and [25].

Algorithm 8 LapSRN Algorithm

procedure LapSRN

trainingInput ← LR-Image

loop:

for *progressivePredictionOfResidual* **do**

featureExtractionLayer()

imageReconstructionLayer()

goto *loop*

Output: LapSRN model

Chapter 3

Methodology

One of the objectives in this thesis is to identify a great method or combination of methods that will prove capable of delivering a higher accuracy in predicting the faces available in an input image which has a low resolution. This chapter outlines the series of experiments and operations carried out to simulate low resolution images and how results were derived for comparison.
In a bid to accomplish the objectives set forth by this thesis, the following activities were carried out and summarized in Figure 3.1.

- Source & analyze data set.
- Construct model to test & record baseline results.
- Re-sample images in scales (Up-sample and Down-sample).
- Retrain and predict with different scales using constructed CNN model.
- Compare results with different scales & baseline.

3.1 Sourcing data

Data was sourced by reviewing articles online and visiting sites managed by Computer Vision research groups and organizations. The details about the origin of sourced data which was used

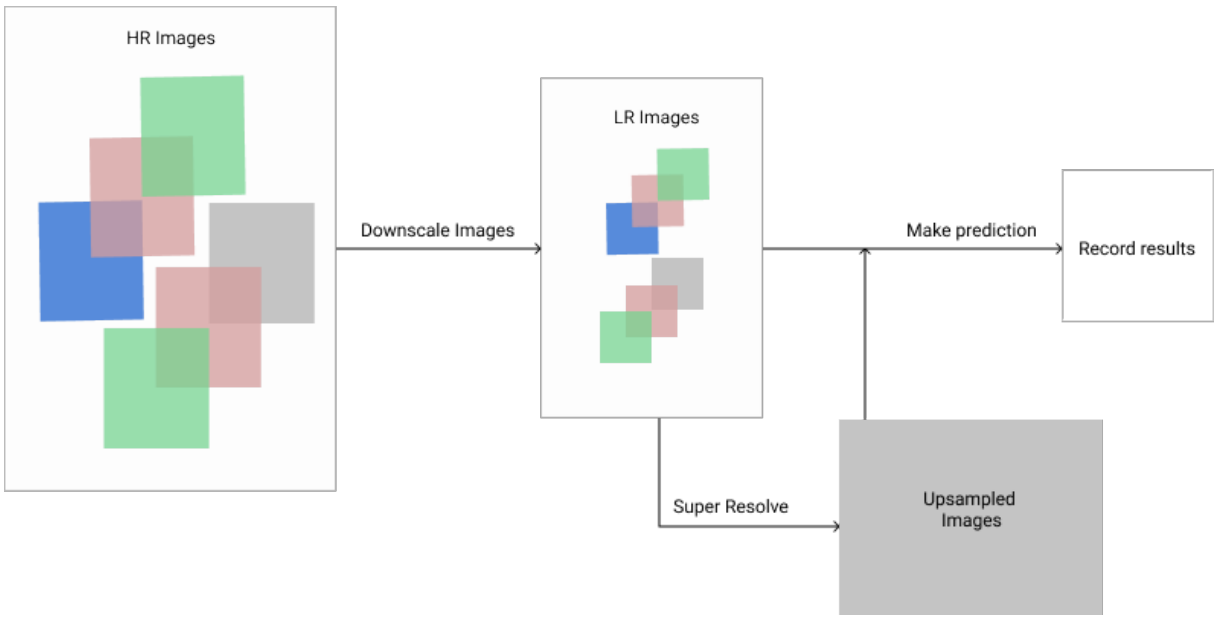

Figure 3.1: Methodology - Original HR images are downsampled and used to train and test a prediction model. The same images are then Super resolved in an attempt to get back the original quality image before next round of training.

in the experiments reported in this thesis has been outlined in a later chapter. However, it is important to mention that some level of pre-processing and data augmentation was carried out to create realistic results and a robust model while training the convolutional neural networks. Multiple databases were engaged in every experiment carried out during this thesis and the results have been presented in a later chapter.

3.2 Face Recognition

Some inspirations drawn from the success achieved in [27] led to the adoption of vgg-face pre-trained weights for preliminary experiments.

In order to use the vgg model, the full VGG16 layered architecture was reconstructed, the pretrained weights were loaded into the model and we performed face recognition by retrieving the features from the model using 2 face images and comparing the Euclidean distance between them. The Euclidean distance measure is shown in equation 3.1.

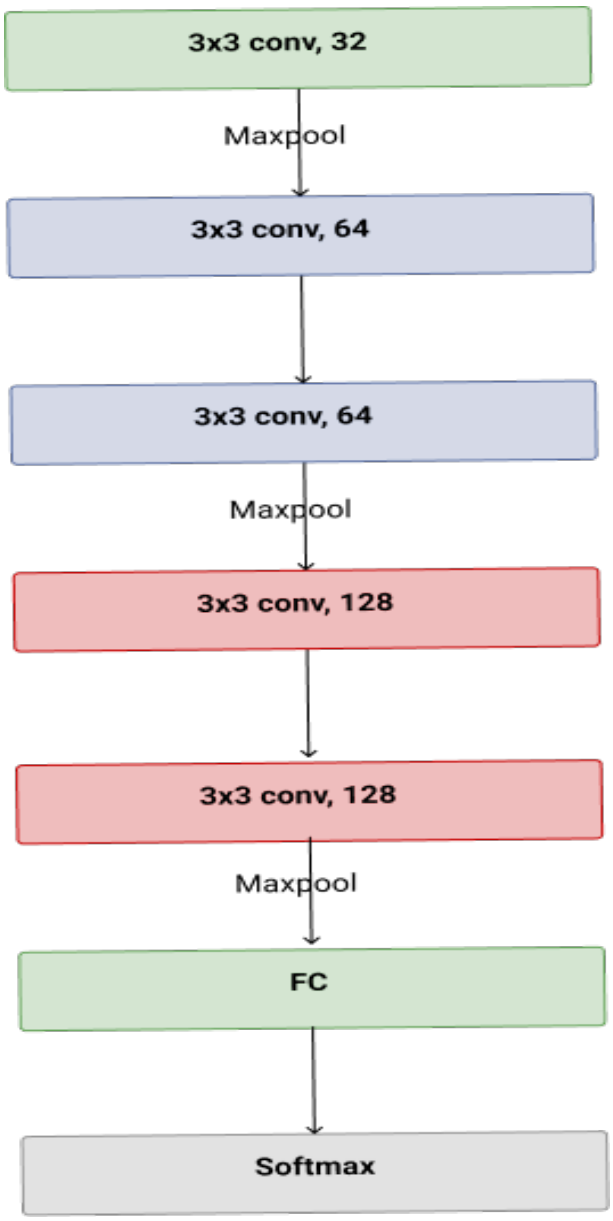

Figure 3.2: CNN Architecture used for face recognition.

$$d(known, query) = \sqrt{\sum_{i=1}^{n}(known_i - query_i)^2} \tag{3.1}$$

After performing this experiment and getting excellent results in accuracy, it became obvious that more tests were required to validate the squeaky clean results because the initial accuracy beats the expectation. It turns out that the initial experiment was performing so good because there was not much comparison with samples in other class. Subsequent tests carried out involved comparing faces of the same sample and randomly comparing the same known face

with faces that belong to other classes.

In other face recognition experiments reported, a simple custom convolutional neural network was built and trained from scratch using a keras library in python. The custom network is made up of 5 convolutional layers with varying sizes of filters and kernel. The architecture of the neural network is shown in Figure 3.2. This network was designed to look like the VGG but with fewer layers.

Algorithm 9 Face Recognition Process

```
1: procedure FACE RECOGNITION PRE-PROCESSING AND LEARNING
2:     if source == Video then
3:         frames ← cutVideo(Vid)
4:     faces ← faceDetection(frames)
5:     augFaces ← augmentByScale(faces)
6:     faces ← faces + alignFaces(augFaces)
7:     faces ← faces + superResolve(faces)
8:     xData ← []
9:     yLabel ← []
10:    i ← 0
11:    loop:
12:    if numpyArray(faces[i].data) then
13:        xData[i] ← faces[i].data
14:        xLabel[i] ← faces[i].label
15:        i ← i + 1.
16:        goto loop
17:        close;
18:    cnn:
19:    features ← featureExtraction(xData, yLabel)
20:    learningNetwork(features)
```

Algorithm 9 states the steps engaged in performing the face recognition task. The significant

addition to the process is related to augmenting the image frames and also randomly aligning faces in the pre-processing stage while training our small convolutional neural network.
We also came across an awesome face recognition implementation in python by [28], this implementation is based on the dlib's state-of-the-heart face recognition algorithms [29] which was claimed to achieve an accuracy as high as 99.38% on the Labelled faces in the wild (LFW) image dataset. This implementation was also used to gather early results and to have a basic idea of how well existing implementations will behave when faced with our problem statement.

3.3 Re-sampling Images (Down-sampling and Up-sampling with DCSCN)

The input full sized images were downsized to provide adequate training data for the purpose of this thesis. The images in our dataset came from different sources, therefore they have different original dimensions. However, four dimensions were used for experiments carried out in this report. The dimensions used are 128 x 128, 64 x 64, 32 x 32 and 16 x 16. Creating multiple scales makes it easy to compare and contrast results for adequate analysis. These dimensions were chosen because they qualify as a range from low to extremely low resolution dimensions when images are concerned.

Algorithm 10 DCSCN Algorithm

1: **procedure** DCSCN
2: *trainingInput* ← LR-Image
3: *preprocessingLayers()*
4: *featureExtraction()*
5: *upscalingNeuralNetwork()*

Output: DCSCN model

In order to simulate a situation where a low quality image was processed to retrieve a higher quality version, Deep CNN with Skip Connection and Network in Network (DCSCN) [30]

was adopted. DCSCN has fewer neural network layers, it is made up of feature extraction and reconstruction network, it combines Deep CNNs with Skip connection layers for image reconstruction. Algorithm 10 outlines a simplified process structure of the model as stated in [30].

3.4 Evaluation

After recording the results on accuracy from prediction using the down-sampled images(128px, 64px, 32px and 16px) as training and test input data, an idea was conceived to use the images with specific dimensions and mix them all up during training before predictions were evaluated. The purpose of mixing up these dimensions was to make sure the trained networks are able to identify the differences in faces even if they have been partially restored without a neural network based super resolution. The results from this experiment was quite promising as revealed in a later chapter of this report.

It seemed important to consider other methods of evaluation besides only accuracy after the first few experiments were carried out, so we also included reports about the confusion matrix for experiments carried out in order to allow us calculate other important metrics in addition to accuracy such as True positive rate, True negative rate and precision

Chapter 4

Data

The datasets used in this thesis are all open source datasets available publicly online. These datasets have been engaged in several computer vision experiments which makes them perfect for this use case. The ICV NAO dataset consists of both Video and Still Image data.

Table 4 presents a detailed information about the various datasets used or considered for the experiments. The table shows the number of samples available in each dataset and the count of unique faces(classes) available in them.

It was interesting to see that some the images available were mistaken because of the background noise, example of such is shown in Figure 4.1.

Figures 4.2, 4.3 and 4.4 show sample images from geotech, icv and feret databases respectively. This images show the visual difference from left with scale (16 X 16) to right (scale 128 X 128).

Dataset	**Description**	**Sample size**	**Classes**
ICV NAO Face Image	Data gathered by ICV research group	310	31
ICV NAO Face Video	Data gathered by ICV research group	17 video files	17
FERET	Facial Recognition Technology program	500	50
GEOTECH	Georgia Tech Face Database	750	50
LFW	Labeled Faces in the Wild	13000	1680

Table 4.1: List of Databases considered.

Figure 4.1: Sample images from the GEOTECH database shows two different people but the face recognition model thinks they are the same because of the noise introduced by the background shelf

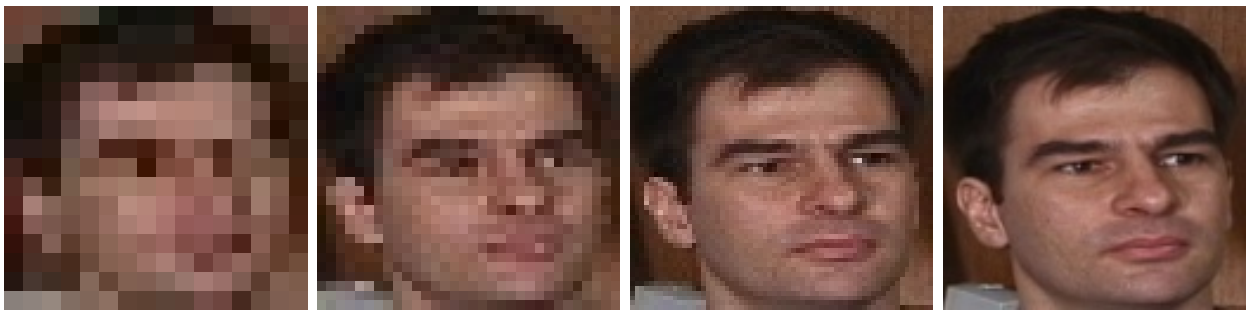

Figure 4.2: Sample face image from GEOTECH database with varying quality (16x16, 32x32, 64x64, 128x128 respectively)

Figure 4.5 shows four samples from a single sample class, these images shows how much visual difference one person can have based on makeup or hair style, this is how robust the LFW database is with its samples. Figure 4.6 displays two frames cut out from video streams recorded in the icv database, faces in such samples are clearly a tiny fraction of the whole image (1.125% of total image area). Figure 4.7 shows the result after face alignment from sample image in figure 4.6.

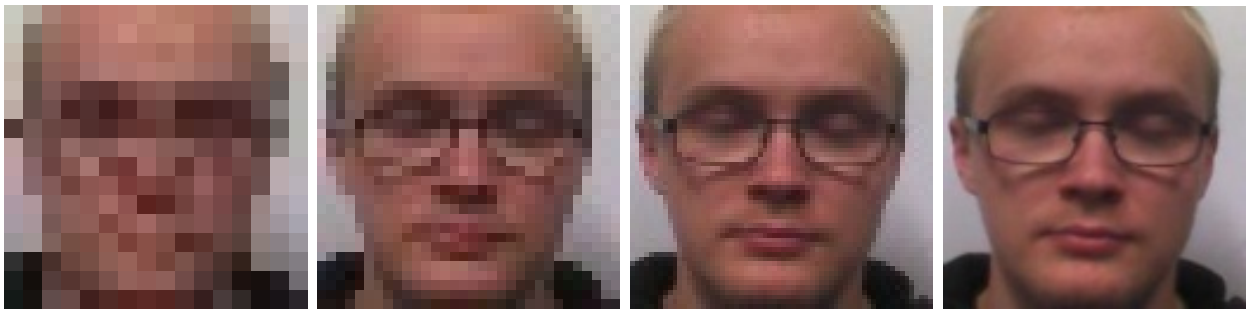

Figure 4.3: Sample face image from ICV image database with varying quality (16x16, 32x32, 64x64, 128x128 respectively)

Figure 4.4: Sample face image from FERET image database with varying quality (16x16, 32x32, 64x64, 128x128 respectively)

Figure 4.5: Sample face image from LFW image database showing the significant variation possible from just one sample class

Figure 4.6: Sample frames extracted from two videos recorded in the ICV database

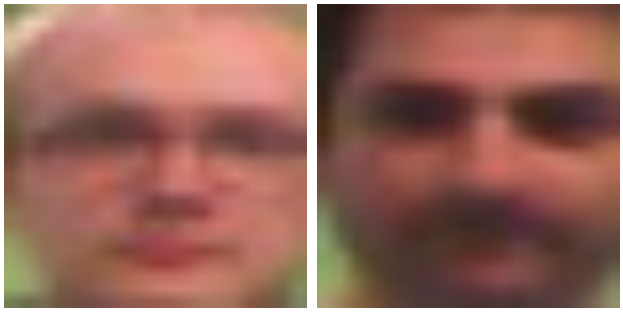

Figure 4.7: Aligned and cropped version of sample frames extracted from two videos recorded in the ICV database

4.1 Data Manipulation

Although most of the data available in our databases were taken in controlled situation, it was still necessary to perform some level of manipulation on the images before extracting features and before converting them into representing arrays. Manipulations to the images include resizing to achieve test in different scales, face alignment in images that have background noise and some data cleaning.

Resizing

This thesis is concerned with face recognition and its sensitivity to varying scales especially very low resolution face images, it was therefore natural that the images used in experiments had to be scaled to fit the interest of our experiments. Also considering that our dataset have different origins and thus different dimensions from their origin, our starting point for all experiment was 128px X 128px, which explains another reason why we needed to resize images to fit.

Face alignment and image cropping

Face alignment involves face detection and localization. Figure 4.6 shows two samples from frames extracted by cutting a video stream collected in the ICV video dataset. Images like these two are true representation of what is possible in a surveillance scenario, in order to perform face recognition on scenes like this, we need to perform face alignment, to answer two questions (a. Is there a face in this frame? and b. Where is the face located?). Once we know the location of the face(s) in this image, we then need to crop the faces ahead of subsequent operations to be carried on the image.

Data Cleaning

The images available in the LFW dataset automatically introduced bias for some of the sample class because there were some classes with as little as 1 image in its sample collection. To deal

with this, we scanned through the dataset and resulted in keeping only images with 3 or more images in its sample class.

The iCV video dataset also introduced its own challenges. This challenge was based on the fact that the image frames were retrieved by cutting video sequences into image frames, which meant a lot of the images might have no face in them at all, we did not see the need to train our model with such, since the task at hand is not based on face detection.

Chapter 5

Results and Analysis

This chapter explains the results achieved from all experiments carried out using same scale input/training data and mixed input/training data. All images used in these experiments have original dimensions bigger than 128 x 128, therefore the images were downscaled to the desired dimensions for the test.

Tables 5.1, 5.2 reports the preliminary results recorded from running the prediction model on our dataset. These results show a failure in terms of accuracy of prediction, with the accuracy getting worse as the resolution decreases. However, the VGG-face algorithm shows a more promising result as shown in Table 5.2.

Table 5.1: Results presented here are from experiments carried out by making predictions using a python code which was implemented around the dlib face recognition algorithm.

	Geo Tech	**Feret**	**ICV**	**LFW**
Original x 3	0.736	0.602	0.812	0.
128 x 128 x 3	0.532	0.398	0.813	0.
64 x 64 x 3	0.506	0.297	0.709	0.
32 x 32 x 3	0.333	0.178	0.585	0.
16 x 16 x 3	0.292	0	0.177	0.

Table 5.2: Results presented here are from experiments carried out by making predictions using VGG-face pre-trained weights and euclidean distance calculation on output layer results.

	Geo Tech	**Feret**	**ICV**	**LFW**
Original x 3	0.92	0.80	0.85	0.82
128 x 128 x 3	0.91	0.73	0.90	0.77
64 x 64 x 3	0.89	0.74	0.84	0.77
32 x 32 x 3	0.87	0.66	0.78	0.61
16 x 16 x 3	0.51	0.50	0.51	0.54

Table 5.3 shows the computation of True positive rates, True negative rates and Precision using equations 5.1, 5.2 and 5.3 with values from confusion matrix computed in Table 5.4.

$$\text{True Positive Rate} = \textit{Recall} = \frac{TruePositive}{TruePositive + FalseNegative} \tag{5.1}$$

$$\text{True Negative Rate} = \frac{TrueNegative}{TrueNegative + FalsePositive} \tag{5.2}$$

$$Precision = \frac{TruePositive}{TruePositive + FalsePositive} \tag{5.3}$$

Table 5.3: Metrics computed from the confusion matrix after making predictions on images with original dimension.

	Geo Tech	**Feret**	**ICV**	**LFW**
TPR	1	0.97	1	0.93
TNR	0.83	0.62	0.71	0.69
Precision	0.86	0.72	0.77	0.77

Table 5.4: Confusion matrix for results on prediction using the vgg-face library with image database using original dimension without image pre processing or scaling.

	GEOTECH		FERET		ICV		LFW	
Predicted/Actual	**+ve**	**-ve**	**+ve**	**-ve**	**+ve**	**-ve**	**+ve**	**-ve**
Pred. Positive	750	124	486	191	310	91	7099	2133
Pred. Negative	0	626	14	309	0	219	507	4646

After recording the results shown in Table 5.1, 5.4 and 5.2, it was time to run different batches of experiments which are documented in this chapter.

5.1 Same scale results

This experiment involves initial down scaling of images and then super resolving same images to an appropriate scale needed for the test, the images were also aligned to contain only face images before the training and evaluation process. Tables in this section are results on accuracy for prediction evaluation carried out on the FERET, GeoTech and iCV databases, the model was trained and tested in batches using the same scale per batch.

The iCV NAO database is a real representation of the problem we seek to solve with low resolution image face recognition. The images generated have an original dimension of 200 x 200 pixels, but the face within the image is at its best 65 x 65 pixels in dimension. In the case of iCV NAO database, it was important to align the faces within the image to reduce the background environment noise while training and predicting.

Tables 5.5, 5.6, 5.7 and 5.8 shows reports gathered on experiments which were carried out using the iCV NAO Face database, iCV Video database, FERET's database and Geo Tech database respectively. These test was carried out using four different scales of data as presented in the table, and the result reveals a much better result that our baseline on these dataset. However, the results here can not be considered as perfect because the accuracy still deteriorates as the image resolution quality reduces.

Table 5.5: Prediction performance on ICV NAO Face Dataset. The results presented here are from experiments carried out using still images captured from a humanoid robot. The images were aligned ahead of the experiments.

Dataset	**Loss**	**Accuracy**	**Validation loss**	**Validation Acc**
128 x 128 x 3	0.0491	0.9731	0.0151	1.0000
64 x 64 x 3	0.1100	0.9596	0.0034	1.0000
32 x 32 x 3	0.1736	0.9378	2.6352	0.5000
16 x 16 x 3	0.6324	0.7818	0.8794	0.7400

Table 5.6: Prediction performance on iCV NAO Video Dataset. The results presented here are from experiments carried out using video frames captured from a humanoid robot. In total, there are 17 videos. The video frames were extracted and after face alignment, an average of 160 images per class were obtained for the experiment.

Dataset	**Loss**	**Accuracy**	**Validation loss**	**Validation Acc**
128 x 128 x 3	0.1262	0.9571	0.1112	0.9858
64 x 64 x 3	0.1383	0.9534	0.3254	0.9148
32 x 32 x 3	0.2137	0.9287	1.3340	0.7074
16 x 16 x 3	0.3787	0.8696	0.3075	0.9062

Table 5.7: Prediction performance on FERET Dataset. The FERET Database has cropped face images of different facial angle, the results presented here were achieved by down scaling the images which were originally 250x250.

Dataset	**Loss**	**Accuracy**	**Validation loss**	**Validation Acc**
128 x 128 x 3	0.1881	0.9246	0.1425	0.9400
64 x 64 x 3	0.2159	0.9242	0.6466	0.8500
32 x 32 x 3	0.4612	0.8310	0.3418	0.8400
16 x 16 x 3	1.1343	0.6196	1.4477	0.5800

Table 5.8: Prediction performance on GEOTECH Dataset. The GeoTech Database also has cropped face images, the images were down-scaled to desired (128, 96, 32 and 16) dimensions ahead of training and evaluation.

Dataset	**Loss**	**Accuracy**	**Validation loss**	**Validation Acc**
128 x 128 x 3	0.0346	0.9896	0.0273	0.9933
64 x 64 x 3	0.0967	0.9589	0.7789	0.7467
32 x 32 x 3	0.2848	0.9011	0.0416	0.9933
16 x 16 x 3	0.6324	0.7818	0.8794	0.7400

5.2 Cross Evaluation results

Table 5.9: Cross evaluation result on a model trained with GEOTECH image database.

GEOTECH Dataset Results			
GEOTECH - 128 x 128			
Dimension	64 x 64	32 x 32	16 x 16
Accuracy	0.967	0.883	0.687
GEOTECH - 64 x 64			
Dimension	128 x 128	32 x 32	16 x 16
Accuracy	0.857	0.924	0.723
GEOTECH - 32 x 32			
Dimension	128 x 128	64 x 64	16 x 16
Accuracy			
GEOTECH - 16 x 16			
Dimension	128 x 128	64 x 64	32 x 32
Accuracy	0.548	0.664	0.781

Accuracy results from tests carried out on the FERET, GeoTech and iCV databases are reported in a tabular format. The model was trained using specific scales but cross evaluated in batches

using different scales of image per batch. Table 5.9 presents the results achieved on Geotech's face database using dimensions of 128 x 128, 64 x 64, 32 x 32 and 16 x 16. Accuracy shows that the closer the dimension of the test set to its training set the better the accuracy. Similarly, Table 5.2 also shows the results achieved carrying out same experiments on FERET's database.

FERET Dataset Results			
FERET - 128 x 128			
Dimension	64 x 64	32 x 32	16 x 16
Accuracy	0.906	0.738	0.389
FERET - 64 x 64			
Dimension	128 x 128	32 x 32	16 x 16
Accuracy	0.950	0.638	0.292
FERET - 32 x 32			
Dimension	128 x 128	64 x 64	16 x 16
Accuracy	0.858	0.916	0.598
FERET - 16 x 16			
Dimension	128 x 128	64 x 64	32 x 32
Accuracy	0.406	0.546	0.842

Table 5.10: Cross evaluation result on a model trained with FERET image database.

Chapter 6

Conclusion

Based on the experiments carried out so far, it became obvious that the quality of prediction through accuracy degraded in most cases as the quality of the image provided reduced.
However, in a situation whereby there was cross examination of training and test scales, the results of prediction for test scales closer to the training scale seemed to have better results than others.
This led to a conclusion that synthesizing image scales and super-resolving while mixing up the image quality in terms of scale and face alignment during training significantly improves the accuracy of our prediction.
By default, existing models can do a very good job at keeping up the True positive rates of face recognition, but the use case of any concrete implementation that adopts these models will determine if it is a good fit or not. In an airport or city surveillance system where law enforcement agencies are looking out for offenders, it might be permissible to only focus on the True Positive Rate and allowing some tolerance for False Positives.

6.1 Summary of Thesis Achievements

This thesis report serves as a clear documentation of the image processing techniques used in enhancing the quality of images ahead of performing a facial recognition task. We also

documented the consideration for improvements and evolution in the space. Ultimately, this report includes a proposal towards enhancing predictions on extremely low resolution images, by tweaking the training process with specific addition of randomly scale-synthesized and face aligned images.

6.2 Applications

Commercial Off-the-shelf(COTS) face recognition software are already deployed in a lot of systems. The Application of Extremely low quality image face recognition can not be downplayed especially when it relates to security surveillance. The use of security cameras that have implemented the COTS can only get better if the software is able to perform its face recognition task regardless of the quality of the frame provided through the video stream.
Another interesting application will be in the development of smart glasses for the visually impaired. These glasses are expected to provide as much information as is needed by the user and therefore the software implementation will need to be robust enough in identifying objects regardless of the variables that could affect image quality.

6.3 Future Work

As impressive as it is to identify faces after they have been super resolved, certain challenges are still imminent with face recognition tasks. As reported in [31], face recognition research has not taken into consideration the degradation of these models when the time difference between the training/face enrollment and face query becomes more by the day. It might be beneficial if there are some networks which are able to learn the possible facial variations with time elapsed and create a synthesized image for training based on these expectations.

Bibliography

[1] C. Ding and D. Tao, "Trunk-branch ensemble convolutional neural networks for video-based face recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 4, pp. 1002–1014, 2018.

[2] E. N. Malamas, E. G. Petrakis, M. Zervakis, L. Petit, and J.-D. Legat, "A survey on industrial vision systems, applications and tools," *Image and vision computing*, vol. 21, no. 2, pp. 171–188, 2003.

[3] M. A. Turk and A. P. Pentland, "Face recognition using eigenfaces," in *Proceedings. 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 586–591, IEEE, 1991.

[4] W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM computing surveys (CSUR)*, vol. 35, no. 4, pp. 399–458, 2003.

[5] X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang, "Face recognition using laplacianfaces," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, no. 3, pp. 328–340, 2005.

[6] H. Demirel and G. Anbarjafari, "Pose invariant face recognition using probability distribution functions in different color channels," *IEEE Signal Processing Letters*, vol. 15, pp. 537–540, 2008.

[7] R. Jafri and H. R. Arabnia, "A survey of face recognition techniques.," *Jips*, vol. 5, no. 2, pp. 41–68, 2009.

[8] G. Anbarjafari, "Face recognition using color local binary pattern from mutually independent color channels," *EURASIP Journal on Image and Video Processing*, vol. 2013, no. 1, p. 6, 2013.

[9] T. Uiboupin, P. Rasti, G. Anbarjafari, and H. Demirel, "Facial image super resolution using sparse representation for improving face recognition in surveillance monitoring," in *2016 24th Signal Processing and Communication Application Conference (SIU)*, pp. 437–440, IEEE, 2016.

[10] M. Wang and W. Deng, "Deep face recognition: A survey," *arXiv preprint arXiv:1804.06655*, 2018.

[11] A. Bolotnikova, H. Demirel, and G. Anbarjafari, "Real-time ensemble based face recognition system for nao humanoids using local binary pattern," *Analog Integrated Circuits and Signal Processing*, vol. 92, no. 3, pp. 467–475, 2017.

[12] B. Li, H. Chang, S. Shan, and X. Chen, "Low-resolution face recognition via coupled locality preserving mappings," *IEEE Signal processing letters*, vol. 17, no. 1, pp. 20–23, 2010.

[13] A. Sharma and D. W. Jacobs, "Bypassing synthesis: Pls for face recognition with pose, low-resolution and sketch," in *CVPR 2011*, pp. 593–600, IEEE, 2011.

[14] W. W. Zou and P. C. Yuen, "Very low resolution face recognition problem," *IEEE Transactions on image processing*, vol. 21, no. 1, pp. 327–340, 2012.

[15] M. Bellantonio, M. A. Haque, P. Rodriguez, K. Nasrollahi, T. Telve, S. Escalera, J. Gonzalez, T. B. Moeslund, P. Rasti, and G. Anbarjafari, "Spatio-temporal pain recognition in cnn-based super-resolved facial images," in *Video Analytics. Face and Facial Expression Recognition and Audience Measurement*, pp. 151–162, Springer, 2016.

[16] A. Clapés, O. Bilici, D. Temirova, E. Avots, G. Anbarjafari, and S. Escalera, "From apparent to real age: gender, age, ethnic, makeup, and expression bias analysis in real age estimation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 2373–2382, 2018.

[17] P. Viola and M. Jones, "Robust real-time face detection," in *null*, p. 747, IEEE, 2001.

[18] G.-S. Hsu, K.-H. Chang, and S.-C. Huang, "Regressive tree structured model for facial landmark localization," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 3855–3861, 2015.

[19] X. Zhu and D. Ramanan, "Face detection, pose estimation, and landmark localization in the wild," in *Computer Vision and Pattern Recognition (CVPR), 2012 IEEE Conference on*, pp. 2879–2886, IEEE, 2012.

[20] S.-W. Lee, J. Park, and S.-W. Lee, "Low resolution face recognition based on support vector data description," *Pattern Recognition*, vol. 39, no. 9, pp. 1809–1812, 2006.

[21] Z. Lu, X. Jiang, and A. C. Kot, "Deep coupled resnet for low-resolution face recognition," *IEEE Signal Processing Letters*, 2018.

[22] X. Yu and F. Porikli, "Ultra-resolving face images by discriminative generative networks," in *European Conference on Computer Vision*, pp. 318–333, Springer, 2016.

[23] C. Dong, C. C. Loy, K. He, and X. Tang, "Image super-resolution using deep convolutional networks," *IEEE transactions on pattern analysis and machine intelligence*, vol. 38, no. 2, pp. 295–307, 2016.

[24] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *European Conference on Computer Vision*, pp. 391–407, Springer, 2016.

[25] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1874–1883, 2016.

[26] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Fast and accurate image super-resolution with deep laplacian pyramid networks," *IEEE transactions on pattern analysis and machine intelligence*, 2018.

[27] O. M. Parkhi, A. Vedaldi, A. Zisserman, *et al.*, "Deep face recognition.," in *bmvc*, vol. 1, p. 6, 2015.

[28] A. Geitgey, "Face recognition implementation." `https://github.com/ageitgey/face_recognition`, 2018. [Online; accessed 19-January-2019].

[29] B. Amos, B. Ludwiczuk, M. Satyanarayanan, *et al.*, "Openface: A general-purpose face recognition library with mobile applications," *CMU School of Computer Science*, vol. 6, 2016.

[30] J. Yamanaka, S. Kuwashima, and T. Kurita, "Fast and accurate image super resolution by deep cnn with skip connection and network in network," in *Neural Information Processing*, pp. 217–225, Springer, 2017.

[31] L. Best-Rowden and A. K. Jain, "Longitudinal study of automatic face recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 40, no. 1, pp. 148–162, 2018.