National College of Ireland

Person Identification Using Landmarks and Deep Learning Techniques

MSc Research Project
Data Analytics

Arundev Vamadevan
Student ID: x22144421

School of Computing
National College of Ireland

Supervisor: Dr. Christian Horn

National College of Ireland

MSc Project Submission Sheet

School of Computing

Student Name:	Arundev Vamadevan
Student ID:	x22144421
Programme:	MSc Data Analytics **Year:** 2022-23
Module:	MSc Research Project
Supervisor:	Dr. Christian Horn
Submission Due Date:	14 December 2023
Project Title:	Person Identification Using Landmarks and Deep Learning Techniques
Word Count:	……**PageCount 19**

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.
ALL internet material must be referenced in the bibliography section. Students are required to use the Referencing Standard specified in the report template. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action.

Signature: ……

Date: ……

PLEASE READ THE FOLLOWING INSTRUCTIONS AND CHECKLIST

Attach a completed copy of this sheet to each project (including multiple copies)	□
Attach a Moodle submission receipt of the online project submission, to each project (including multiple copies).	□
You must ensure that you retain a HARD COPY of the project, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer.	□

Assignments that are submitted to the Programme Coordinator Office must be placed into the assignment box located outside the office.

Office Use Only	
Signature:	
Date:	
Penalty Applied (if applicable):	

Person Identification Using Landmarks and Deep Learning Techniques

Arundev Vamadevan
X22144421

Abstract

This research focuses on person identification through the detection of landmarks from facial images. This study uses VGG-16 model for facial landmark extraction and the Support Vector Machines (SVM) classifier for person identification. The models are trained with dataset VGGFace-2 which contains a large collection of different personalities with large variations in pose, ethnicity, age and illumination. Multi-Task cascaded Convolutional Neural Network (MTCNN) is used for face detection from the images. VGG-16 model is validated with a Root Mean Squared Error (RMSE) of 7.02 for landmark extraction and SVM identified persons from images with an accuracy of 70 %. The five facial key points selected are landmark coordinates of eyes, nose and distance between lip to lip. Histogram of Oriented Gradients (HOG) is used to calculate the feature descriptors which is used to train the SVM model. Images and facial landmarks are transformed using augmentation techniques. A comprehensive and detailed evaluation is conducted and the result shows a better performance in landmark extraction with deep learning models.

Keywords: Landmark detection, VGG-16, SVM, MTCNN, VGGFace-2

1 Introduction

This section contains the project background and motivation towards the selection of this project topic and the research which intended to done. Also, the section explains about the importance of face recognition and person identification using facial landmarks and various deep learning techniques.

1.1 Project Background and Motivation for the Topic

Person identification is gaining importance in various domains day by day as the automation of face detection will help various domains and applications in various ways. Apart from the security applications, face recognition is also used for scientific, marketing and industrial areas (Chacua, 2019). Early techniques used for landmark detection and person identification are only suitable for detection in controlled environments. However, with the advancements in deep learning and machine learning technologies, many advanced techniques like neural networks are available for detection in uncontrolled situation with high accuracy and person identification is becoming a hot topic in today's world (Khabarlak, 2021.). But the dense facial landmark detection in very challenging environments like partial occlusion, pose difference, illumination and backgrounds. This automated facial recognition technique is differing from other traditional biometric systems which has many limitations.

Facial landmarks such as eyes, nose and mouth are the important key feature points on the face which can be used for different tasks such as face recognition, person identification, tracking a person, detection of gaze and emotion detection. To detect a person correctly in various challenging situation, a very fast and accurate landmark system is required. Convolutional Neural Networks (CNN) showing very strong performance on image classification and detection of objects (Kim, 2020). Even though CNN can perform well on extracting facial landmarks from facial area which covers facial contours, eyes, nose, and mouth with 68 facial points, for more accurate results pre-trained CNN models such as Residual Network (ResNet) and VGG-16 which has more layers added

can be used. Even though facial landmark detection has many applications, it still faces many challenges like detecting landmarks under various illumination conditions such as occlusions or shadows, detecting landmarks from faces with different skin tones or facial features. Also, facial landmark detection is very sensitive to posture, movements and facial expressions. Therefore, the algorithm should be accurate robust enough to handle all these situations to accurately extracting the landmarks. Another technique that can be done to improve accuracy is the use of image augmentation, which increases the size of the training data and thereby improving accuracy.

In summary, the advancements in the areas of machine learning and deep learning gave rise to improvements in performance of facial landmark detection and numerous applications using landmark detection for person identification. It also introduced new areas of research in the areas of computer vision and image processing and analysis. Many challenges are there in analysis and detection of landmarks and addressing challenges will lead to new developments in the area of face recognition and person identification. A deep study is required on these challenges with landmark detection and the unexplored techniques. Further research in these areas will definitely beneficial to all the domains that uses facial image recognition and person identification for various applications.

1.2 Project Requirement Specifications

The research question mainly focuses on the feasibility and analysis of various deep learning techniques for the accurate detection of facial landmarks from images and usage of machine learning techniques for person identification tasks. CNN based transfer learning technique VGG-16 is designed and trained with large dataset containing images of very different varieties of people. But the small computational error exists with original landmarks used for training the deep learning model may cause landmarks predicted by CNN based pre-trained model with variance in some pixels which affects the performance of the classifier that used to identify the person. This research also includes a deep learning model to train large dataset with wide varieties of images and landmarks and machine learning classifier to identifying the person from predicted landmarks.

RQ: "*Can deep learning facial landmark extraction models improve or enhance face identification task by machine learning models?"*

Sub-RQ:" *To what extend a face recognition or person identification model (SVM) can be used for successfully perform face identification with landmarks predicted by a deep learning model and directly from images?****"***

The performance of the machine learning classifier may degrade if the original landmarks used for trained them are having any kind of computational error or impurities. If such case exists, how much successfully these models can identify a person if we directly classify them with images instead of landmarks predicted by deep learning models.

The remaining parts of the report is divided and structured as follows: Section 2 contains literature review, Section 3 contains methodology, Section 4 is design, Section 5 is implementation of various techniques. Section 6 is result evaluation and Section 7 is conclusion and discussion.

2 Related Work

In this section, a detailed literature review on the above topic and associated key arears are studied and presented. As the study is on topic Landmark Detection and People Identification using deep learning, the literature review has been done in areas: facial landmark detection, facial landmark detection using deep learning techniques and people identification using machine learning.

2.1 Face Detection Using Landmarks

Detecting face and facial landmarks from an image with high precision is a challenging task, especially when processing images with different poses, illumination and with occlusions. Another challenge that the studies face is detecting landmarks for people with different skin tones and different facial features. Also, the detection of facial landmark is highly sensitive to facial movements and postures.

Face detection is in high demand in today's world since it has a wide range of applications in different areas which are sensitive to attack or where high security is needed. If the face of particular person can be detected from the small input or glimpse of a portion of face, then it will help the system in very effective ways than the traditional methods we are using like bio matric person identification which requires close contact with the system to detect the person and in terms of security it's not robust. (Saleem, 2023) proposed a method which uses facial landmarks extracted and the region of interest is matched with the person. The facial features used are the Euclidean distance between different facial key points like eyes, nose and distance between lip to lip. The method is implemented with 82.4 % accuracy. But the study didn't outline the other methods that can be used for face detection, various metrics that can be used for calculating the error and how much the system effective in uncontrolled situations like illumination, occlusion and head movement. (Wu Y. a.,2019) done a study on various techniques that can be used in such situations and automatically detect these key points. Based on the ways they use shape information, study categories them into three classes such as holistic methods, constrained local model and regression-based methods. The study did a detailed comparison of performance of various techniques in both controlled and wild datasets under different conditions. While most of the algorithms performing well on the controlled situations, study shed light on various algorithms that could be used in-the-wild conditions in real world situations with high accuracy.

(Wu Y. a., 2015) proposed a method which addresses these challenges detecting facial landmarks from images with extreme occlusion and poses. The study proposes a robust cascade regression framework which can predict the location of the occlusion and can estimate the occlusion with the help of a supervised regression method and which updates landmark visibility and improves robustness. The experimental results show the model performed well in various uncontrolled situations especially with severe occlusions and head poses. But the method needs more improvement in performance of conditions like illumination changes and images with low resolution. Another major issue when consider facial analysis is the sequential performance of different tasks such as landmark detection, pose estimation and face deformation review and analysis. (Wu Y. G., 2017) proposes a method which addresses these issues by performing these tasks simultaneously. Study uses learning-based landmark detection and iteratively updates location of facial landmarks, occlusion, pose and deformation until convergence. The joint relationship with various tasks is definitely an advantage and will improve performance. The method has another advantage that it can have integration with 3D annotations. But there could be some limitations when processing images with non-rigid facial expressions and poses.

2.2 People Identification Using Landmarks with Deep Learning Techniques

Recent improvements in deep learning and computer vision techniques also influenced the performance of facial landmark detection. Convolutional neural networks are an example as it can identify and detect the complex patterns in facial key points and the variations in them. The use image augmentation also improved the performance and efficiency as it increases the training dataset size and thereby help accurate detection of landmarks.

(Chacua, 2019) proposed a people identification method which can perform on both controlled and uncontrolled situations from extracting facial landmarks using CNN and classification of person using SVM. The real time testing with 128 samples gave an accuracy of 96% in controlled situations and 71.43% in uncontrolled situations. The approach was effective for classifying known and unknown persons. But the limitation of the model is the difficulty in locating the areas of face with occlusion or illumination. (Hannane, 2020) proposed a method for landmark detection using a binary-hierarchical and cascaded regression approach. This model divides the image into multiple patches which is non-overlapping. Then the regressor will detect and refine the landmark within each

patch. After testing with various datasets, this method showed better accuracy and efficiency for images taken various conditions. The model is efficient and robust but it shows some limitations in identifying landmarks from images with heavy occlusion. (Kim, 2020) proposed another method using EMTCNN which can extract 68 facial landmark points real time as well. The model uses an increased filter size by augmenting with dilated convolution and CoordConv techniques. The study also did a comparative analysis between four different models MTCNN, EMTCNN, Augmented MTCNN and Dlib, out of four augmented EMTCNN performed better with higher accuracy.

(Hsu, 2020) proposed a hybrid loss function, and a discrimination network to improve the communication between landmarks in the pixel wise classification model. The model is tested with six facial landmark datasets and pixel-wise classification model with hybrid loss function performed better than other approaches. Another problem faces by face recognition is Single Sample Per Person. In cosem cases the sampkle size will be one per person and in t that cases, the efficiency of the detection system decreases because of the very limited sample and resource and also the difference in illumination, pose and resolution of operational domain and enrolment domain. (Abdelmaksoud, 2020) proposed a solution for this problem. 3D reconstruction of face in different poses and situation which will overcome the problem with SSPP. Also face illumination transfer techniques are used to tackle with illumination problem and Super Resolution GAN is used to overcome the low-resolution problem. The validation and final testing results confirmed that the proposed method performed well than the traditional deep learning methods in terms of accuracy and robustness. Large intrinsic variance in images is another problem faced by facial landmark detection. (Dong, 2018) proposed a style-aggregated method which deals with the problem of intrinsic variance in images. The original images are transformed into different style aggregated images by a generative adversarial module which are very robust to environment changes. Then both the original and style aggregated images are used to train the landmark detector. After the validation of methods, this approach enhanced the performance than other traditional deep learning algorithms.

One of trend which can be seen in person identification using deep learning techniques is using only one feature like face. But both face and hand can be used for identifying a person using landmark extracted. (Kabisha, 2022) proposed a method which uses a VGG-16 model for face recognition and a CNN model for hand gesture recognition. The models are experimented with two customized dataset and gained an accuracy of 98% for both face and hand recognition of a person. (Pan, 2018) proposed a ResNet based model for person reidentification. The approach succeeded in reducing computational complexity by using a genetic algorithm. First, a set of related images are selected based of landmarks and then these images are promoted to reach to initial retrieval result. The model was very successful against background variation, illumination and pose. One of the problem faces in person reidentification using landmarks is the large disparity happens at pre-training tasks, image classification and person identification which limits the CNN performance for reidentification of person. (Matsukawa, 2016) proposed a method to fine tuning the CNN features by pedestrian attribute dataset. Also, they proposed new labels for additional classification loss function. After experimenting the model with four, person re-identification datasets, the model exhibit significant performance improvement. (Teoh, 2021) proposed a CNN based face recognition and identification system which uses CNN for training and feature extraction. The model is trained with large datasets and tested with real time video using computer vision techniques. But the model struggled with images taken in low intensity light and performed well in high intensity light. Classifier performed with an accuracy of 91.7 % in recognizing from images and 86.7 % for recognizing from video.

2.3 Literature Review on People Identification Using Machine Learning Techniques

Face recognition task can also do with machine learning algorithms as they can do pattern recognition. Once we have the landmark features extracted, they can be used to train the machine learning algorithms and use them to identify a particular person. The challenges when doing facial analysis in real world situations are partial occlusion, pose and posture variety, illumination etc. deep learning techniques such as CNN, transfer learning models like ResNet and VGG-16 can perform well on these situations to an extend but if we are using machine learning algorithms for the same task, how well they can perform is a question mark. (Sharma, 2020) did a study on the same problem.

They experimented different machine learning algorithms like Linear Discriminant Analysis, Multilayer Perceptron, Naïve Bayes and Support Vector Machine along with Principal Component Analysis for face recognition task. The study achieved an accuracy 97% for PCA and 100% for LDA. The problem with machine learning techniques facing while doing face recognition in large datasets are the speed. To get more enhanced performance, we need to train with more samples and there is a chance that computational time it takes will increase. (Chen, 2017) proposed a study on facial recognition with machine learning algorithms that has high accuracy in recognition and high computational speed. The model was built by combining different algorithms like PCA and SVM which gave over 95 % accuracy and PCA with KNN achieved a best-balanced result between computational speed and accuracy.

(Filali, 2018) proposed another research on comparison boosting algorithms and Gabor filter for the task of facial detection. They used Haar-AdaBoost, LBP-AdaBoost boosting algorithms and GF-SVM, GF-NN Gabor filter techniques. The techniques were compared and the result showed that both in terms of detection time and detection rate, boosting algorithms were performed well. EG Amaro et. al. proposed a method which can detect faces from video frames and generate a face dataset. These face images are filtered and pre-processed and a collection of machine learning techniques are applied by using these face images as input to detect faces and person. This approach is suitable for analyzing large volume of data on which face labels are not available. (Pandey, 2019) proposed a study with a modified CNN architecture by adding two normalization operations to the layers. One provides acceleration of the network. This model can extract facial features and classify with Softmax classifier. This approach enhanced the performance of recognition and classification.

2.4 Conclusion

After carefully reviewing the literatures above, a brief idea of the research happened in face analysis using landmarks, person identification using landmarks and various deep learning techniques and person identification using machine learning techniques etc. are given. Also from the review, we can see that not enough research has been done in the area of facial recognition with landmark using various transfer learning techniques with images having very challenging situations like pose, occlusion, gender, ethnicity, profession and age. A novel approach using deep learning techniques like ResNet and VGG-16 to extract landmarks in these various challenging situations and using these extracted landmarks to identifying a person is not been explored much. The research will focus on identifying a person with landmarks extracted from the combined and effective use of deep learning and machine learning techniques.

3 Research Methodology

The motivation of the study is to perform facial landmark extraction and person identification in images with different personalities using deep learning techniques. After successful extraction of facial features from the images, which is used for the identification of the personality. But the study showed person identification using the predicted features are not performed well due to the systematic error with dataset and mostly with some pixels difference which causes the model to identify the person wrongly. And later the person identification is done by using feature descriptors computed by Histogram of Gradients on the images and these feature descriptors are given as input to the SVM classifier for person identification task and the identification was performed well.

3.1 Dataset Description

This research uses the dataset, VGG FACE-2 – A large dataset of facial images which can be used for identifying faces and person across various ages, poses, profession and ethnicity. The dataset was introduced by (Cao, 2018). The dataset consists of around 3.31 million of images of 9131 persons with an approximate average of 362 samples for each person. The dataset is in compressed format and is about 40 GB in size. It can be downloaded from academictorrents.com. This dataset is very

commonly used for face related task as it has a wide variety of samples with different poses, ethnicity, illumination, profession and age. The dataset also comes with an annotation file which consists of facial landmarks for each sample (Ground truth landmarks). Landmarks consists of ten coordinates which represents five facial key points such as two eyes, nose and two for lips respectively. For the research purpose, the study used 6402 samples of 20 personalities which is used for training both deep learning and machine learning approaches. In the study, it was required to check each image to verify the precision of landmark which is available with the dataset and the study doesn't have the capacity to do the process in the total data.

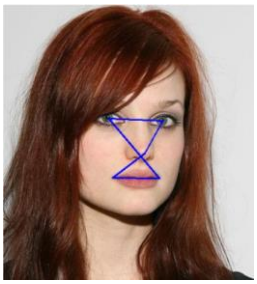

Fig 1: Original Image with Landmarks

Fig 1 depicts a sample of originl image with landmarks displayed on it by drawing a polygon. One sample image is loaded from the corresponding folder and respective landmarks are fetched from annotation file. Then a polygon is drawn by connecting the landmarks using computer vision techniques.

3.2 Process Flow Diagram

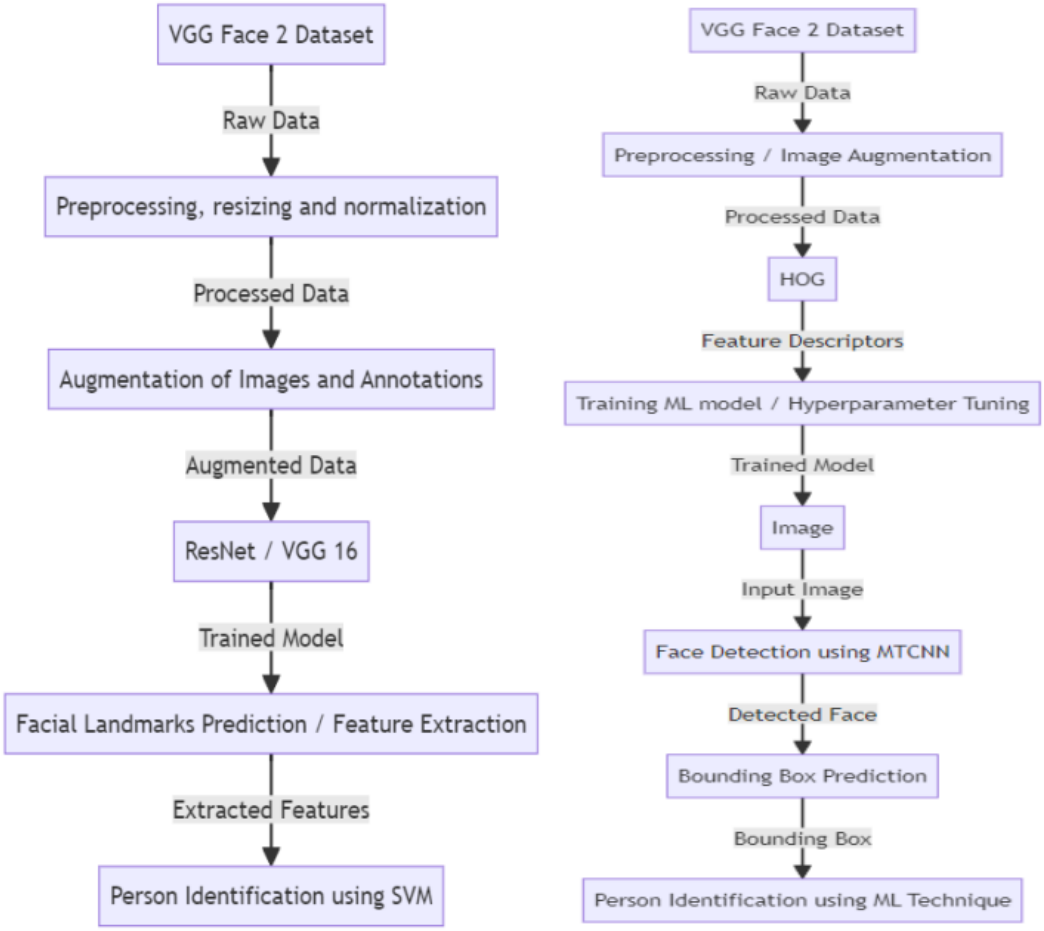

Fig 2: a) Person Identification Using Landmarks b) Person Identification using ML Technique

Fig 2 explains the process flow of proposed study. Fig 2 a) It starts from collecting dataset VGG Face-2 from the respective source, then preprocessing the images according to the requirement of the deep learning model build for facial landmark prediction. Then augmentation applied on both the pre-processed images and the landmarks coordinates in the csv file. Then the combined data is trained with a deep learning model which is best suited for the study to predict the facial landmarks. The study tried two models, ResNet and VGG-16 and finally selected VGG-16 after trying various experiments and it built a more stable and consistent model. After extracting the facial features, it used for identifying the person using SVM classifier

Fig 2 b) explains the process flow of person identification using the classifier model and is built with ML techniques. The images are pre-processed and Histogram of Gradient (HOG) is used as feature descriptor and the computed features and Region of Interest (ROI) are passed to SVM for training. Once training is done, an unknown person's image is given, the models and identifies the person as known or unknown and face detection is done with MTCNN and bounding box of ROI is predicted. Finally, the image with bounding box around face and identified name is displayed.

3.3 Image Augmentation

3.3.1 Person Identification using landmarks

Image augmentation is a technique which can be used in image related deep learning tasks. The technique will increase the size of the dataset by applying some transformations on the original images. These slight modifications will bring diversity in training data and helps to reduce overfitting while training and make the model more robust to variations in the input data. Also, the augmentation will help the model to learn to identify landmarks under various poses and facial expressions and various lighting conditions. Here one simple image augmentation techniques, Image Flipping is used to make the samples selected more diverse in nature.

For flipping an image with facial landmarks, we need to flip both image and landmarks. To flip the image with landmarks, first we need to calculate the centre of face or midpoint of the landmarks. Then we need to calculate the mirror point for each landmark, by subtracting the landmark's x-coordinate from the mid point's x-coordinate and keep the y-coordinate unchanged. The new landmark coordinated will help to keep the landmarks in correct position after flipping the image horizontally.

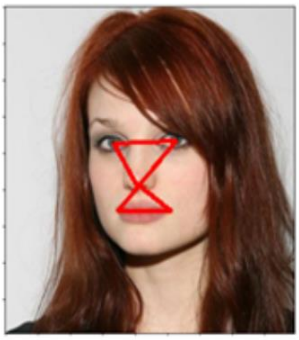

Fig 3 a) Flipped Image

Fig 3 shows the flipped version of origaninal image.

The landmark cordinates for the flipped images are also transformed using the same logic and write back to the csv file, so that these coordinates will be intact position after the image transformation.

3.3.2 Person identification using ML Technique

The same preprocessing and augmentation steps are done for the above approach are reated for for the ML model as well.

3.4 Data Preprocessing

3.4.1 Person identification using landmarks

VGG Face-2 is the dataset which is used for training, testing and validation of the VGG-16 deep learning model which is under CNN architecture. Several experiments have done for finding the best model for landmarks extraction and prediction. Different samples from the dataset are used in various stages based on the requirements of the dataset and algorithms used. According to the requirement of the model used for landmark prediction, the input layer shape is adjusted to 128 x 128 pixels and to get the sample in same shape and size, cropping and resizing is done using computer vision techniques. To get all pixels in a range of 0 and 1, pixel normalization also done by dividing with 255.

3.4.1 Person identification using ML Technique

VGG Face-2 is the dataset which used for training, testing and validating the ML model for face detection and person identification. Different samples of dataset are used for cross validation and hyper parameter tuning during training. The samples are converted to grey scale equivalent as it will be well suited for computer vision tasks. Then the images are resized to 128 x 128 pixels as it compatible with the size of trained images in VGG-16 model. Then the histogram of gradients is computed for each image to identify the face within a window or region of interest (ROI). The image which is used for person identification also needs to do all these preprocessing steps and then MTCNN is used for the face identification tasks.

3.5 Evaluation Metrics

3.5.1 Person identification using landmarks

To find out the accuracy of the predicted landmarks, calculate the distance between predicted landmarks and ground truth landmarks. It is a comparison of vectors of real numbers and one of the ways to determine accuracy is to calculate the relative error by comparing the difference between corresponding elements to the magnitude of those elements. Then the aggregate error for all images in the testing dataset can be calculated by calculating error for all landmarks in each image and aggregating it using any of the various available statistical measures like mean error, median error etc. To quantify the average squared difference between corresponding elements, we can use measures like Mean Squared Error (MSE) or Root Mean Squared Error (RMSE). MSE penalizes the larger error more and results in a larger impact in the overall MSE value. RMSE uses the same unit as the data.

$$MSE = \Sigma (y_i - p_i)^2/n$$

$$RMSE = sqrt [\Sigma (y_i - p_i)^2/n]$$

3.5.2 Person identification using ML Technique

To evaluate the performance of the classifier, the study used accuracy metric.

Accuracy = (TP + TN) / (TP + FP + TN + FN)

Accuracy can be defined as the ratio between the true predictions with total predictions. Where, TP is True Positive, TN is True Negative, FP is False Positive and FN is False Negative.

4 Design Specification

4.1 Face Detection

For detecting the region of interest (ROI) and face, two approaches named HOG (Albiol, 2008) and MTCNN (Xiang, 2017) were used.

4.1.1 Histogram Of Gradients (HOG)

HOG is a feature descriptor which commonly used in computer vision tasks for object detection. HOG features calculated for the image is given as input for the SVM model to recognize the person. The process starts with computing the gradient of image intensity using some filters in both horizontal and vertical direction. The image is divided into a number of cells and a histogram of gradient direction is calculated for each cell. Then the neighbouring cells are grouped together into blocks which accounts for changes in lighting and contrast. The blocks are normalized and merged together to get the final result. Thus, the required pattern of the face is obtained and this pattern is used to train the classifier, SVM. During training, SVM will be able to identify images with faces and images without faces, which is grouped into positive samples and negative samples. A window size of (64, 128), cell size of (8,8), block size of (16,16), block stride of (8,8) and bin size of 9 is used in implementation and the window moves through the entire image using the sliding window approach.

4.1.2 Multi-Task CNN

MTCNN is a deep learning algorithm which can efficiently identify faces and facial key points in an image. The algorithm first predicts a bounding box and a face or non-face probabilities using a convolutional network (P-Net), which scans the entire image in sliding window approach to identify the face and refines the bounding box coordinates accordingly. MTCNN can further refines the identified face region and localize facial landmarks in addition to face detection and bounding box regression. MTCNN is used here because it is robust to pose and illumination variations and it can perform all tasks simultaneously.

4.2 Model Building

4.2.1 Person identification using landmarks

Facial landmark detection is a challenging task and various deep learning techniques such as ResNet (Mandal, 2021) and VGG 16 (Aung, 2021) have achieved remarkable results in various object detection and classification tasks. Also, for extracting various facial features in challenging situations like pose, illumination and occlusion, we need a powerful CNN based model. That is the reason that the study used two models to experiment and select one stable and robust model from ResNet and VGG-16.

ResNet-50 Model

Residual Network (ResNet) is a kind of CNN introduced by (He, 2016) which can be used for powerful image processing and computer vision applications. ResNet-50 is 50 layers deep and trained on more than million images from the database ImageNet. It has 48 convolutional layers, one maxpool layer and an average pool layer.

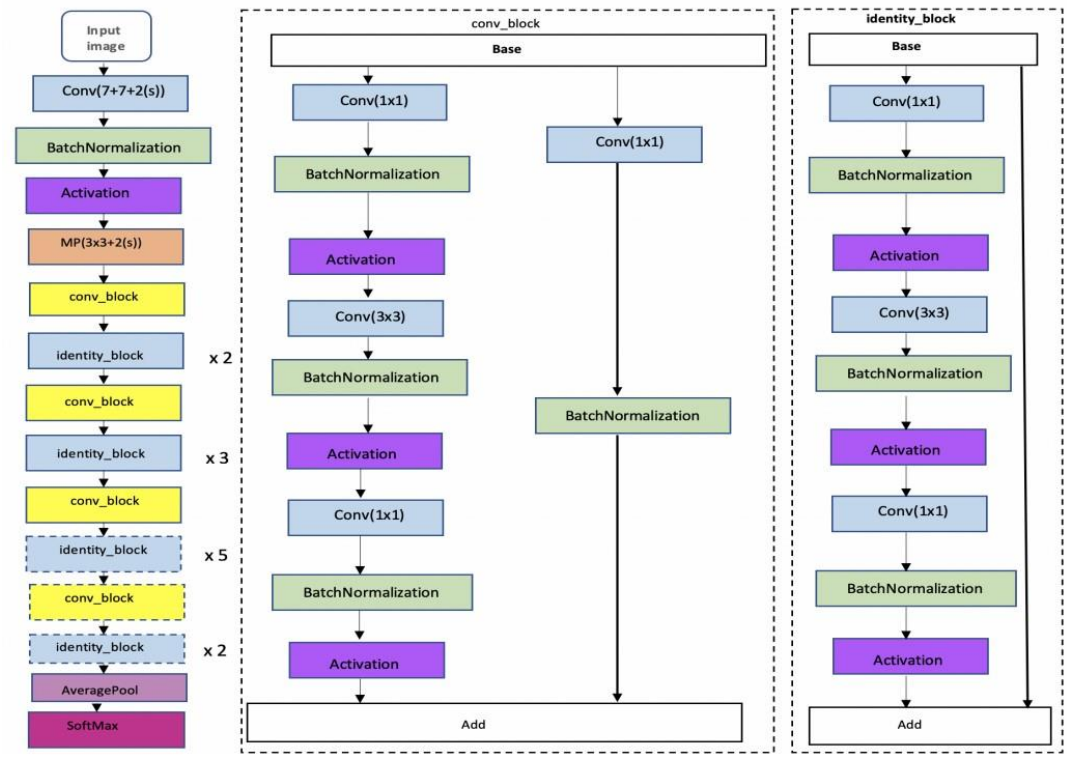

Fig 4: ResNet – 50 architecture (Wu Y. a.,2019)

Fig 4 explains the general scheme of ResNet – 50 architecture. It uses a design called bottleneck, the residual block which uses 1 x 1 convolutions. This design is helpful to reduce the matrix multiplications and parameters. This feature provides ResNet-50 a faster training of each layer. Transfer learning with a pretrained machine learning model means reusing it for another similar tasks for faster training and development and it helps to achieve higher performance even with smaller dataset

VGG-16 Model

VGG-16 is a pretrained CNN with 16 layers deep and trained on ImageNet dataset which is having more than a million images. This model was proposed by (Simonyan, 2014). VGG model can achieve a veery high test accuracy since it was trained on ImageNet with more than 14 million images and over 1000 classes of objects. VGG uses multiple smaller layers instead of a large single layer which improves the decision functions and helps the neural network to converge quickly. Also, VGG uses smaller convolutional filter which has the capability to reduce the overfitting during training. Since VGG 3 x 3 filter, it is optimal size and can capture image's spatial features.

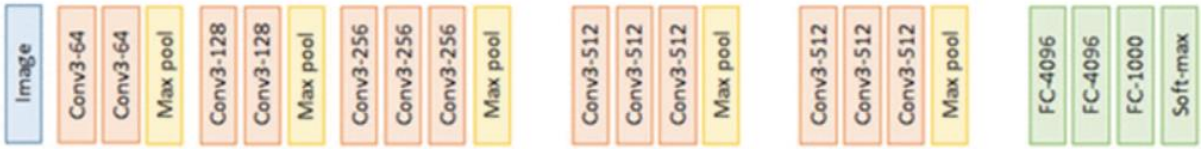

Fig 5: VGG-16 architecture (Wu Y. 2015)

4.2.2 Person Identification using SVM

Support Vector Machines is a simple machine learning algorithm which produces significant performance with less computation resources. SVM classifies the datapoints by separating them with

hyperplane in N-dimensional space with maximum margin. This maximum margin can ensure that the future data points that needs to be classified can be done with more confidence and also if we made any error in the location of boundary, then the chances for misclassification will be less. Also, SVM can avoid local minima and can provide better classification.

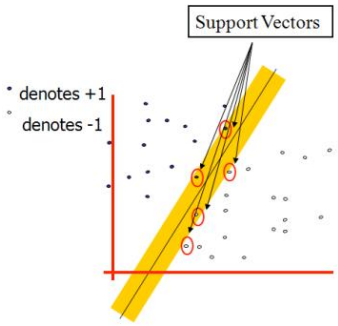

Fig 6: Representation of Support Vectors (Jakkula, 2006)

5 Implementation

5.1 Implementation of person identification using landmarks

5.1.1 Implementation of ResNet.

A ResNet model is created with 6402 images and 10 % of augmented images. The top layers are freezed and two dense layers are added with 512 neurons and relu activation function under the supervision of mean squared error metric. Weights are assigned from ImageNet dataset and the model is trained for 20 epochs and batch size of 35 with Adam optimizer with a learning rate of 0.001. After the validation, the performance of the model was measured with an RMSE of 16.33, but need to be optimized for better performance. For improving the performance of the model, data set need to be increased for the training process, so that the model can generalize the landmarks better. So, the model is trained with 6402 original images and the augmented images of the whole original images. All the layers and parameters kept same. After the training and validation of the model, its performance is measured with an RMSE of 7.60. Then the model is experimented with an epoch of 50 to see the difference in performance. But after the training of the model and validation, the performance measured with RMSE was 7.54. A small improvement could see but the model performance was consistent. Also, model was experimented with increasing classes in ResNet architecture, but the result in performance was same. RMSE value of some of the test images are calculated and plotted in a histogram to check the difference in values and to get an idea of how wrong the data or predictions are.

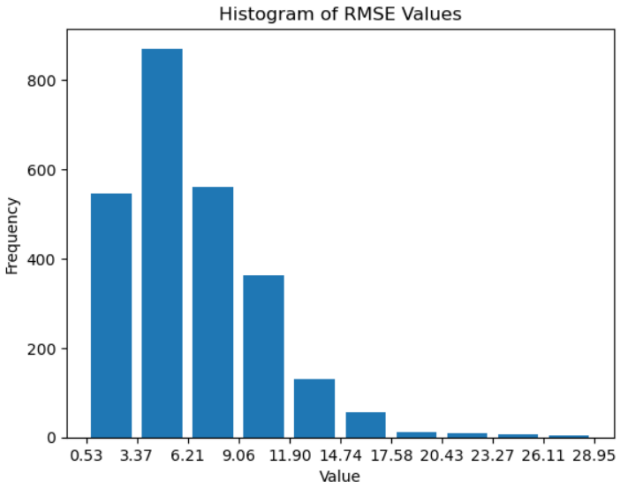

Fig 7 : Histogram of RMSE values

Fig 7 depicts the Histogram of RMSE values of some of the testing images and the analysis shows the presence of systematic error which could be caused from data or model. To identify the same, the model is experimented with various samples of data including only original images without using augmented, only augmented images without using original, only flipped images, only translated images but all of the experiments showed almost similar results. The input size of images is fed to ResNet are also changed to 224 x 224 and cross validation of dataset samples are done. But the performance was varied only in negligible values.

5.1.2 Implementation of VGG-16

A VGG-16 model is created with weights imagenet and the top layers are freezed. Two dense layers are added with 512 and 10 neurons respectively. The model is trained with 6402 original images of 20 personalities and its augmented images under the supervision of mean squared error loss function and Adam optimizer with a learning rate of 0.001. The training is done with 40 epochs and a batch size of 64. After the model tested with testing dataset, it measured with an RMSE of 7.02 and result was stale for various samples of dataset.

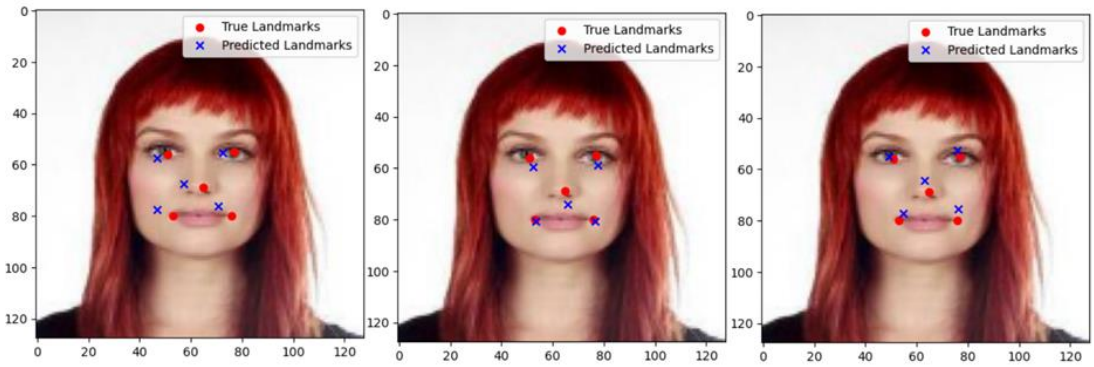

Fig 8: Landmarks prediction by a) CNN model, b) ResNet model, c) VGG-16 model

Fig 8 depicts the visual inspection of facial landmarks predicted by CNN, ResNet and VGG-16 models. Both models performed well on training and testing data. But when we consider the best model with both performance and consistency, VGG-16 is selected as the model for facial landmarks prediction for person identification.

5.1.4 Person identification using landmarks

The features extracted from VGG-16 model is used for the person identification. To predict the landmarks, an SVM model is built by training with facial landmarks downloaded from VGG Face-2 dataset. Model's accuracy was 24 % after enhancing with grid search cross validation and hyper parameter tuning. The SVM model is then used for identifying a person using the predicted landmarks from VGG-16 model and the actual landmarks available in csv file. But result was as shown in fig 9, SVM model wrongly identified the person with predicted landmarks and correctly classified with the original landmarks.

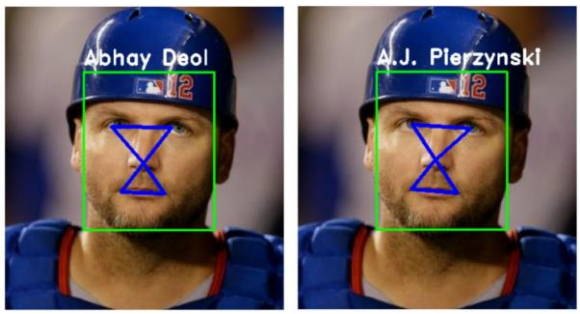

Fig 9: a) Wrongly classified by SVM from predicted landmarks b) Correctly classified by SVM with ground truth landmarks

Then the identification is done with the ground truth landmarks of all the 6402 images, and after prediction, the model identified the different persons with an accuracy of 60.90 %. But the performance with predicted landmarks is very less.

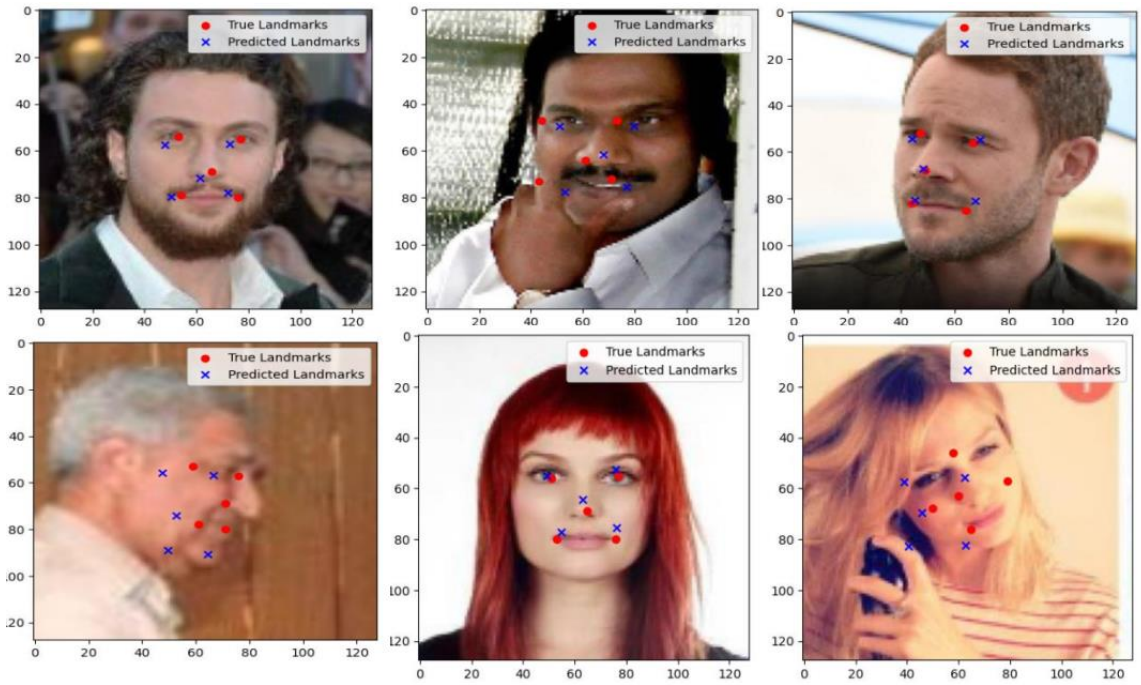

Fig 10: Samples shows actual and predicted landmarks marked on facial key points

Fig 10 shows some of the sample images from test dataset which is marked with actual landmarks and predicted landmarks. From the images, it is evident that a number of images are wrongly labelled and landmarks not accurate. This might be responsible for the problem with VGG-16 model to learn the landmark positions correctly and poor performance in the landmark prediction, and which result in less performance and wrong person identifications by SVM model. The systematic error with landmarks might be reason for the under performance of the VGG-16 model and the study didn't have the capacity to check the data in total. In the literature also there is no hint that there might be a systematic error in labelling the images.

Therefore, the study decided to try the research with another approach, person identification using Machine Learning (ML) Technique. Instead of landmarks, feature descriptors calculated from the images can be used for training the ML model. Feature descriptors of the images can be calculated using technique.

5.2 Implementation of Person Identification using ML Technique

Person identification is the process when the model is presented with an image of an unknown person, the model responds with its best possible estimate of the particular person's identity from the collection of known persons. The algorithm will predict the most similar individual from the known persons database. To get a benchmark of comparison and select the best classifier, the study did a comparison of various algorithms Logistic Regression, Decision Tree, Random Forest, SVC, Naïve Bayes, XG Boost, K Nearest Neighbor.

The dataset used was the same used for training deep learning models for landmark prediction. Each image is labelled with corresponding personality manually and the images are undergone preprocessing and augmentation steps. Then the images and labels are separated into predictor and target variables. Feature descriptors are calculated for each image and these calculated feature descriptors are given as the input to the ML models. Then the data is split into training and testing dataset in 80:20 proportion. Then a 3-fold cross validation is applied on the dataset to get the cross-validation score for each model. The performance of each model is shown in Table1,

Table 1
Comparison of various ML models

Sl No	Models	Accuracy (%)
1	Logistic Regression	62.53
2	Decision Tree	17.70
3	Random Forest	45.57
4	SVM	66.21
5	Naïve Bayes	50
6	XG Booster	58.71
7	KNN	48.86

Based on nature and complexity of the dataset, suitability for multiclass classification task and the performance on the above test, SVM is selected for building the final model of person identification classifier.

To enhance the performance of the SVM model, hyper parameter tuning is applied with Grid Search Cross validation technique with different values of regularization, gamma and kernel. After finding the best parameters for SVM using hyper parameter tuning, model is trained with best estimators and the accuracy was enhanced to 70.60 %.

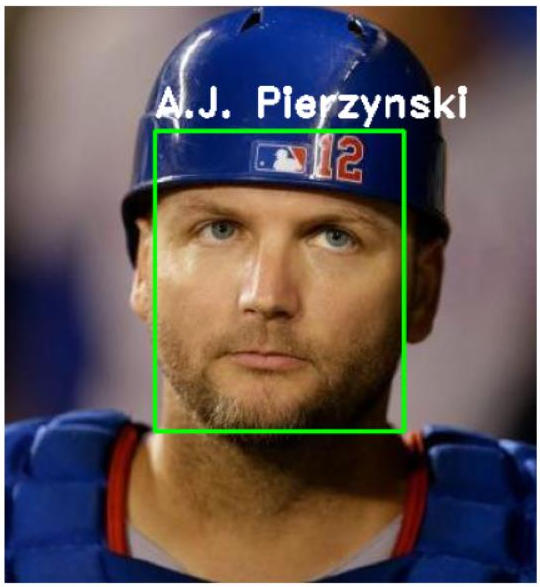

Fig 11: Person identified by the SVM model

Fig 11 shows the person identified correctly by the SVM model

6 Evaluation

6.1 Person Identification Using Landmarks

VGG-16 model evaluation was performed using VGGFace-2 dataset for the facial landmark extraction with an RMSE value of 7.02. The model was trained for 40 epochs and the graph of Mean Squared Error for training and testing phase is depicted in Fig 12 below,

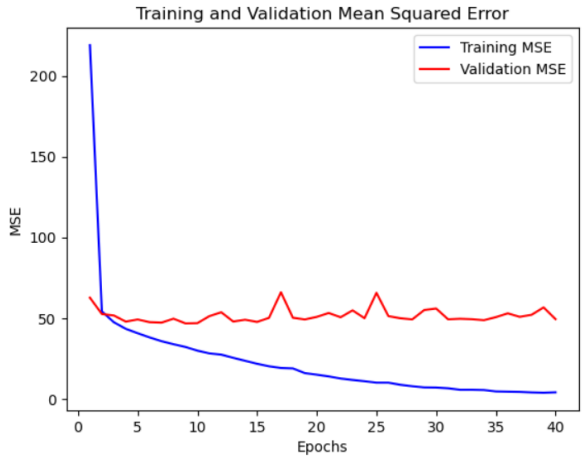

Fig 12: Graph of Training and Validation MSE of VGG-16 feature extraction model

From the graph, we can observe that after the second epoch, the MSE dropped from over 250 to 50 and then gradually decreased per epoch. But the validation MSE was overall consistent throughout the epochs and shows the consistent performance of the model.

The SVM model for person identification trained with ground truth landmarks are having an accuracy of 24 % due to the systematic error in landmarks. But the prediction of persons with ground truth landmarks shows an accuracy of 60 %.

6.2 Person Identification Using SVM

SVM model evaluation is performed on VGGFace-2 dataset for person identification. The model was evaluated with an accuracy of 70% as shown in Table 2. The confusion matrix for the model evaluation is also shown in below Fig 13,

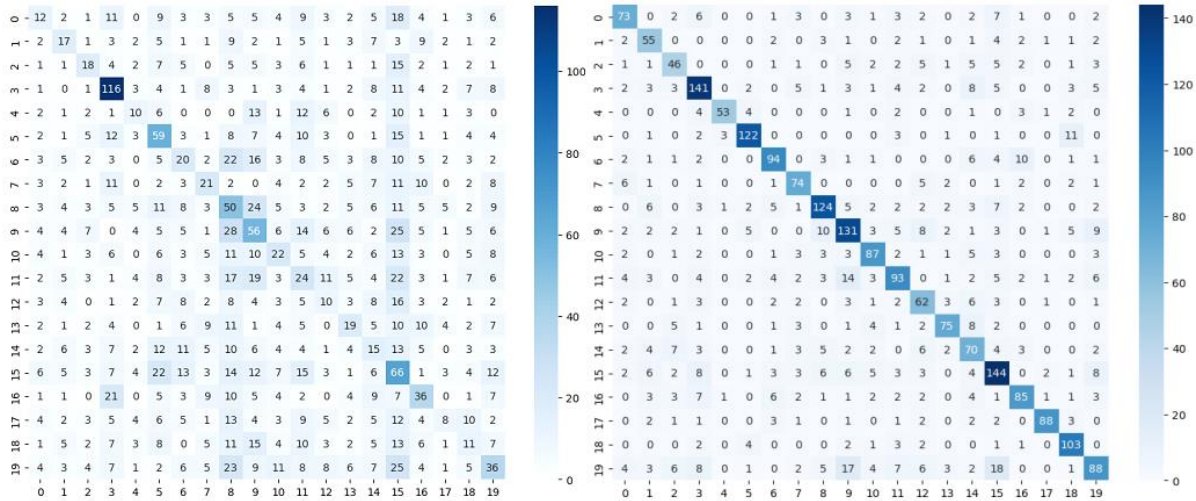

Fig 13 a) Confusion matrix of SVM person identification using landmarks b) Confusion matrix of person identification using SVM with images

6.3 Discussion

Detection of landmarks comes with many complex challenges like pose, illumination and occlusion, and dealing them properly is important to get maximum performance from the model for accurate landmark detection and prediction. The landmarks of face images provided with the datasets are displayed with many anomalies. Around 40% of the test images displayed with marked ground truth landmarks were differed by many pixels. So many mislabeled images were there and that might be

responsible for the problems in learning the landmark positions by VGG-16 model and the underperformance of SVM model for person identification, trained with landmarks predicted by this model. But with original landmarks some of them identified correct persons.

7 Conclusion and Future Work

Various models are implemented and validated successfully and the results obtained are enabled us identify and answer the research question. Different models implemented are shown reasonably good results. More fine tuning with the models and the dataset with proper landmarks and images without anomalies will definitely improve the performance and accuracy of both landmark detection and person identification. The pre-trained models used in research are performed well especially VGG-16. MTCNN is used for face detection and Histogram of Gradients is used for calculating feature descriptors for SVM. There has been lot of researches going on in this area, as referred in literature review. Many advanced models are available which performs well on large datasets of images with challenges, but implementation of those models is currently out of scope for this research. The future work can be researching on these advanced and sophisticated models with complex datasets and fine tune them accordingly to get best performance with facial image datasets.

8 Acknowledgement

I would like to thank my supervisor Dr. Christian Horn for the guidance, support, timely reviews and motivation throughout the semester to complete the research successfully.

References

Khabarlak, K. and Koriashkina, L., 2021. Fast facial landmark detection and applications: A survey. *arXiv preprint arXiv:2101.10808*. Journal of Computer Science & Technology, vol. 22, no. 1, pp. 12–41

Chacua, B., Garcia, I., Rosero, P., Suárez, L., Ramirez, I., Simbaña, Z. and Pusda, M., 2019, November. People identification through facial recognition using deep learning. In *2019 IEEE Latin American Conference on Computational Intelligence (LA-CCI)* (pp. 1-6). IEEE.

Kim, H.W., Kim, H.J., Rho, S. and Hwang, E., 2020. Augmented EMTCNN: A fast and accurate facial landmark detection network. *Applied Sciences*, *10*(7), p.2253

.

Saleem, S., Shiney, J., Shan, B.P. and Mishra, V.K., 2023. Face recognition using facial features. *Materials Today: Proceedings*, *80*, pp.3857-3862.

Wu, Y. and Ji, Q., 2019. Facial landmark detection: A literature survey. *International Journal of Computer Vision*, *127*, pp.115-142.

Wu, Y. and Ji, Q., 2015. Robust facial landmark detection under significant head poses and occlusion. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 3658-3666).

Wu, Y., Gou, C. and Ji, Q., 2017. Simultaneous facial landmark detection, pose and deformation estimation under facial occlusion. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 3471-3480).

Hannane, R., Elboushaki, A. and Afdel, K., 2020. A divide-and-conquer strategy for facial landmark detection using dual-task CNN architecture. *Pattern Recognition*, *107*, p.107504.

Hsu, C.F., Lin, C.C., Hung, T.Y., Lei, C.L. and Chen, K.T., 2020. A detailed look at cnn-based approaches in facial landmark detection. *arXiv preprint arXiv:2005.*08649. Published in arXiv.org 8 May 2020

Abdelmaksoud, M., Nabil, E., Farag, I. and Hameed, H.A., 2020. A novel neural network method for face recognition with a single sample per person. *IEEE Access*, *8*, pp.102212-102221.

Dong, X., Yan, Y., Ouyang, W. and Yang, Y., 2018. Style aggregated network for facial landmark detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 379-388).

Kabisha, M.S., Rahim, K.A., Khaliluzzaman, M. and Khan, S.I., 2022. Face and Hand Gesture Recognition Based Person Identification System using Convolutional Neural Network.

Pan, N., 2018, October. Better Person Re-identification Using ResNet Model and Re-ranking Strategy. In *IOP Conference Series: Materials Science and Engineering* (Vol. 435, No. 1, p. 012002). IOP Publishing.

Matsukawa, T. and Suzuki, E., 2016, December. Person re-identification using CNN features learned from combination of attributes. In *2016 23rd international conference on pattern recognition (ICPR)* (pp. 2428-2433). IEEE.

Teoh, K.H., Ismail, R.C., Naziri, S.Z.M., Hussin, R., Isa, M.N.M. and Basir, M.S.S.M., 2021, February. Face recognition and identification using deep learning approach. In *Journal of Physics: Conference Series* (Vol. 1755, No. 1, p. 012006). IOP Publishing.

Sharma, S., Bhatt, M. and Sharma, P., 2020, June. Face recognition system using machine learning algorithm. In *2020 5th International Conference on Communication and Electronics Systems (ICCES)* (pp. 1162-1168). IEEE.

Chen, J. and Jenkins, W.K., 2017, August. Facial recognition with PCA and machine learning methods. In *2017 IEEE 60th international Midwest symposium on circuits and systems (MWSCAS)* (pp. 973-976). IEEE.

Filali, H., Riffi, J., Mahraz, A.M. and Tairi, H., 2018, April. Multiple face detection based on machine learning. In *2018 International Conference on Intelligent Systems and Computer Vision (ISCV)* (pp. 1-8). IEEE.

Amaro, E.G., Nuño-Maganda, M.A. and Morales-Sandoval, M., 2012, February. Evaluation of machine learning techniques for face detection and recognition. In *CONIELECOMP 2012, 22nd International Conference on Electrical Communications and Computers* (pp. 213-218). IEEE.

Pandey, I.R., Raj, M., Sah, K.K., Mathew, T. and Padmini, M.S., 2019. Face Recognition Using Machine Learning. *Int. Res. J. Eng. Technol*, *6*, pp.3772-3776.

Cao, Q., Shen, L., Xie, W., Parkhi, O.M. and Zisserman, A., 2018, May. Vggface2: A dataset for recognising faces across pose and age. In *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)* (pp. 67-74). IEEE.

Albiol, A., Monzo, D., Martin, A., Sastre, J. and Albiol, A., 2008. Face recognition using HOG–EBGM. *Pattern Recognition Letters*, *29*(10), pp.1537-1543.

Xiang, J. and Zhu, G., 2017, July. Joint face detection and facial expression recognition with MTCNN. In *2017 4th international conference on information science and control engineering (ICISCE)* (pp. 424-427). IEEE.

Mandal, B., Okeukwu, A. and Theis, Y., 2021. Masked face recognition using resnet-50. *arXiv preprint arXiv:2104.08997*. Published in arXiv.org 19 April 2021

Aung, H., Bobkov, A.V. and Tun, N.L., 2021, May. Face detection in real time live video using yolo algorithm based on Vgg16 convolutional neural network. In *2021 International conference on industrial engineering, applications and manufacturing (ICIEAM)* (pp. 697-702). IEEE.

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* (pp. 770-778).

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

Jakkula, V., 2006. Tutorial on support vector machine (svm). *School of EECS, Washington State University*, *37*(2.5), p.3.