Name: M.R.T.Herath
Student Reference Number: 10820763

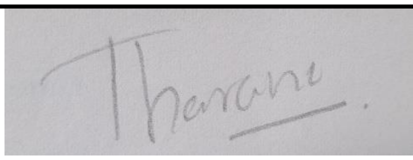
IN PARTNERSHIP WITH PLYMOUTH UNIVERSITY

Module Code: PUSL3123	Module Name: AI And Machine Learning
Coursework Title:	
Deadline Date: 2024/ 01/03	Member of staff responsible for coursework: Dr. Neamah Al Naffakh
Programme: BSc(Hons) Software Engineering	
Please note that University Academic Regulations are available under Rules and Regulations on the University website www.plymouth.ac.uk/studenthandbook.	

Group work: please list all names of all participants formally associated with this work and state whether the work was undertaken alone or as part of a team. Please note you may be required to identify individual responsibility for component parts.

10820763- Manawaduge Rovindu Tharana Herath
10820822 - Biyanwilage Shihan Gaurawa Ferando
10818178 - Warnakulasooriya Fernando

We confirm that we have read and understood the Plymouth University regulations relating to Assessment Offences and that we are aware of the possible penalties for any breach of these regulations. We confirm that this is the independent work of the group.

Signed on behalf of the group: 
Individual assignment: ***I confirm that I have read and understood the Plymouth University regulations relating to Assessment Offences and that I am aware of the possible penalties for any breach of these regulations. I confirm that this is my own independent work.*** Signed :
Use of translation software: failure to declare that translation software or a similar writing aid has been used will be treated as an assessment offence. I *have used/not used translation software. If used, please state name software…………………………………………………………
Overall mark _____% Assessors Initials ______ Date_________

*Please delete as appropriateSci/ps/d:/students/cwkfrontcover/2013/14

Table of Contents

Acknowledgement

Completion of this would not have been possible without the participation and support of many people whose names may not be mentioned. Their contribution is sincerely appreciated and gratefully acknowledged. However. Our team would especially like to express their deep appreciation and indebtedness to the following.

We thank you dear Dr. Neamah Al Naffakh for the module lecture on AI and Machine Learning. And all the other lecturers who have helped us with their endless support, kind, and understanding spirit during our project.

To all friends and others who in one way or another shared their support, either morally, physically,

Thank You very much

Introduction

Pattern recognition is one of the cutting-edge technological developments, revolutionizing the ways in which we authenticate and identify people, especially in the field of biometric face recognition. Pattern recognition is a multidisciplinary field that combines computer science, artificial intelligence, and image processing. Because of its inherent complexity and ability to recognize faces accurately due to its distinctive features along with uniqueness, the human face has emerged as a focal point for biometric identification. Techniques for pattern recognition use complex algorithms to examine the features, structures, and patterns of the face to offer a trustworthy method of authentication. It is essential to understand the development of reliable and effective face recognition systems.

This review covers the fundamental ideas of pattern recognition and provides insight into the various methods such as neural networks, machine learning, and deep learning models along with their utilizations in previous related works that are employed in biometric face recognition and how valid and appropriate those employed data collection feature extraction and system evaluation have done. Understanding these approaches helps us to get better understanding on pattern recognition's overview, performance, issues and potential applications in biometric face recognition systems.

Literature Review

There are countless research papers have issued throughout years related for facial recognition which will be selected and discussed as below.

A deep learning architecture created especially for face recognition tasks is the **VGGFace model**. In contrast to general-purpose models like ResNet, VGGFace is specifically designed to be highly effective at identifying faces in images (Haertle and Millis, 2014). Although, the datasets and source were not clearly mentioned, the model uses data augmentation tactics like images in different angles, different light conditions and variety of backgrounds. Also, utilizes common datasets for face recognition include Labeled Faces in the Wild (LFW) and the VGGFace dataset itself. For feature extraction it uses CNN specially VGGFace capture hierarchical architectures allows to detect facial representations and facial recognition. VGGFace is evaluated by measuring how well it performs on benchmarks for face recognition. The model's accuracy in identifying specific people can be evaluated using metrics like accuracy, precision and recall. Protocols for face verification and identification, like those in the LFW dataset, can be utilized to assess how well the model performs in various contexts. Usually, tables with a summary of the model's accuracy and face discrimination capabilities are used to display the results. Therefore, VGGFace model is capable of efficiently extracts and classifies features for accurate identification, and its performance is evaluated on face recognition datasets.

FaceNet is an earlier approach of face recognition which adopts a face recognition deep learning model that maps face photos into a condensed Euclidean space (Schroff, Kalenichenko and Philbin, 2015). It examines how to train a model using triplet loss and assesses the model using common benchmarks. This model employs two core architectures which is Zeiler & Fergus like network (Zeiler and Fergus, 2014) which was developed aiming enhancement of interpretability of learned features in CNNs. **The Zeiler & Fergus network** has introduced a deconvolutional layer to visualize learned features by mapping them back to the input pixel space. This layer allows researchers to understand the hierarchical representations captured by the network at different layers, revealing specific features or patterns. The network's architecture includes multiple convolutional and pooling layers, but the deconvolutional layer sets it apart for visualization purposes. Therefore, it could pave the way for subsequent research aimed at simplifying the inner workings of deep neural networks and making them more transparent and accessible for analysis which could be expected from FaceNet as is directly based on these architectures.

FaceNet model is based on four training datasets considering exceptions like Labelled Faces in the Wild (LFW) and YouTube Faces, where other evaluation tasks were considered. After that it was used with experimental datasets which includes 100 – 200 million thumbnails of 8 million individuals. Input sizes ranged was from 96x96 pixels to 224x224 pixels. Since it has used YouTube as the data collection source it could have collected variety of faces allowing to represent factors like diversity, variety and draw connections from real-world scenarios. The model has achieved classification accuracy of 98.87%±0.15 with fixed center crop and 99.63%±0.09 with additional face alignment. It has also performed well on YouTube Faces DB, with classification accuracy of 95.12%±0.39 for the first hundred frames per video (Schroff, Kalenichenko and Philbin, 2015). The FaceNet approach has some limitations and potential drawbacks despite the thorough evaluation methodology. Furthermore, concerns regarding the model's robustness and generalization abilities are raised by the lack of thorough explanations of data augmentation techniques and the reasoning behind selecting 100M–200M training face thumbnails. A more thorough analysis that incorporates more performance metrics and sheds light on how particular design decisions affect the model's overall efficacy would enhance the evaluation.

DeepID is another approach of face recognition which uses deep neural networking architecture known as DeepID3. It was based on architectures including VGG net and GoogLeNet. For this model 300,000 training sample were used along 6000 face pairs as validation dataset and the model included with 10 – 15 non-linear feature extraction layers (Sun *et al.*, 2015). Accuracy was presented as 99.52% and the evaluation also explores common false positives and negatives among various DeepID series algorithms, shedding light on the model's performance on specific challenging instances. However, this model is lacking on factors like contrasting on ages in face pairs and labeling. Regardless its short comes, it holds larger accuracy and holds higher diversity as it gains dataset from geographically different locations. Additionally, the trade-off between model complexity and performance improvement needs careful consideration, especially given the model's large-scale training data requirements.

Another research paper discusses on **ResNet** and how to improve deep convolutional neural networks (CNNs) with large residual networks validating with Imagenet which is a large-scale visual recognition database in order to improve its applicability extends to various vision and non-vision problems. Classification dataset includes 1.28 million training dataset and 50,000 validation dataset and tested 0n 100,000 dataset reportedly (He *et al.*, 2016). Similarly, feature extraction happens with layers included within it. Model included with 18-layer and 34-layer residual networks (ResNets). With the hierarchical representation given by ResNet enables it to automatically extract and utilize relevant features for the given classification task. It exhibits a lower training error, suggesting that it can generalize well to validation data. These results highlight residual learning's ability to mitigate optimization issues brought on by deeper network penetration. In order to evaluate the ResNet system, a variety of classification tasks are evaluated, with an emphasis on top-1 and top-5 error rates (He *et al.*, 2016). These metrics assess the model's predictive accuracy in terms of class label accuracy within the top one or top five predictions, respectively. In order to verify that the model can generalize to new data, it is tested on distinct validation and test sets after being trained on the extensive ImageNet dataset.

Another research was targeted on advancing face recognition through the introduction of the **ArcFace** loss function. The 5.8 million faces in the MS1M dataset are used by the authors to test and train their suggested method. Although the size of the dataset is impressive, the paper does not go into great detail about how diverse the data is or if it represents a representative sample of identities, which raises concerns about bias and generalization (Deng *et al.*, 2019). The research presents the ArcFace loss, a feature extraction technique created specially to improve discriminative feature learning. An additive angular margin is incorporated into this loss function, encouraging a greater angular separation between features that belong to distinct identities. Although the paper shows that the suggested loss is useful in increasing feature separability, a deeper examination of the loss's interpretability and sensitivity to hyperparameters would be beneficial. ArcFace's classifier is built using an Angular Margin Loss in conjunction with a fully connected layer. The classifier's decision boundaries are directly impacted by the loss function selection, making it an important decision. Nevertheless, neither the decision-making process's fairness considerations nor any potential biases brought about by this choice are covered in great detail in this paper. The paper evaluates ArcFace's performance on benchmarks like Labeled Faces in the Wild, but suggests a more comprehensive evaluation of its practical utility in real-world scenarios, considering factors like computational efficiency, potential biases, and failure cases.

The goal of the Next Research study known as **DeepFace** was to introduce a deep neural network architecture in order to improve face verification performance. The authors utilize an internal dataset comprising four million facial images, each of which belongs to approximately 4,000 people, for the purpose of data collection (Lu and Tang, 2015). Despite the large size of the dataset, the paper would benefit from a deeper analysis of its diversity, possible biases, and the extent to which it accurately represents a variety of demographic and ethnic groups. Inspired by the classical LeNet-5 model, DeepFace uses a deep convolutional neural network (CNN) architecture for feature extraction. The authors suggest a multitask learning strategy in which the CNN is trained for both face attribute classification and face verification simultaneously. Despite the novelty of this approach, the paper does not provide a comprehensive analysis of the interpretability of the learned features and the potential trade-offs between verification and attribute classification tasks. Additionally, the impact of feature extraction on the model's performance under varying conditions and image quality could be explored more extensively. DeepFace is a verification classifier using a neural network with locally connected and fully connected layers. However, the paper lacks a detailed analysis of interpretability and fairness in the decision-making process. Despite achieving human-level performance on the Labeled Faces in the Wild dataset, a more comprehensive evaluation including real-world scenarios and diverse datasets would enhance the understanding of DeepFace's practical utility. Examining the model's robustness to lighting, pose, and demographic factors and analyzing computational efficiency would provide a more nuanced assessment of its strengths and limitations.

A multi-task learning strategy for face detection and alignment is the main focus of the paper "Joint Face Detection and Alignment using **Multi-task Cascaded Convolutional Networks**," (Zhang *et al.*, 2016). Study will discuss critically at the essential components of feature extraction/selection, classifiers, data collection, and system evaluation in this analysis. To address face detection and alignment simultaneously, the authors use a multi-task cascaded convolutional network

(MTCNN). The reference provided does not specifically mention the dataset that was used for training and assessment. To guarantee the robustness and generalizability of the suggested model, a varied dataset containing labeled facial images would be necessary, though, considering the nature of the tasks (identification and alignment). MTCNN uses cascaded networks, where each network is in charge of a particular function (landmark localization, detection). Hierarchical feature extraction using convolutional layers enables the model to learn discriminative features at various scales. The convolutional nature of the network naturally permits the extraction of pertinent facial features during training, even though the feature selection procedure is not specifically described in the paper. The information supplied does not specifically address the classifier used in MTCNN. For both face detection and landmark alignment tasks, it is reasonable to assume that softmax classifiers or similar activation functions are used at relevant stages of the network, as this approach is based on convolutional neural networks (CNNs) (Zhang *et al.*, 2016). The paper emphasizes system evaluation through experiments on benchmark datasets, utilizing common face detection metrics like precision, recall, and F1 score, and assessing landmark alignment accuracy. The paper should detail specific datasets, considering variations in pose, illumination, and expression.

Face recognition holds several challenges including issues inherits as subject of image procession such as lighting / illumination, image background and poses of objects. Another study showcases on **Face Recognition Algorithms in Unconstrained Environments** which discusses impact of unconstrained environmental factors like illumination or poses addressing common challenges for face recognition models. Data collection of this model involved collecting images of unconstrained environments accounting factors such as illumination, pose, and expression, making face recognition more challenging including labeled Faces in the Wild (LFW), IJB-C, and MegaFace (Yousefi and Kschischang, 2014). The research will discuss the methodologies used for feature extraction and selection, including deep feature learning and classifiers like support vector machines (SVMs), k-nearest neighbors (KNN), and deep neural networks. The main focus will be on evaluating face recognition systems under unconstrained conditions, using metrics like accuracy, precision, recall, and ROC curves. The paper may also explore the impact of illumination changes, occlusions, and pose variations on algorithm performance. The authors may compare their proposed algorithms against existing methods to demonstrate their effectiveness in handling the challenges posed by unconstrained environments. The study examines nanostructures with single electrons, revealing slow steady state buildup and coherent processes. The population of dots is sensitive to energy changes, especially due to interaction-induced renormalization effects. Researchers propose a method to detect these effects using bias dependence. The study uses numerically exact results and Born-Markov theory (Moy, Hope and Savage, 1999) to analyze quantum systems' time evolution over extended time scales, including steady states and decay times of coherent charge oscillations (Yousefi and Kschischang, 2014).

Discussion

Seemingly discussed research papers have adopted several tactics to overcome several practical issues of face recognition model either it has related for datasets or their conditions like occurrences of different environmental factors, quality and other. Following section will discuss essential aspects of different researches and their targeted projects along with their identical differences comparing to others.

The goal of the Zeiler & Fergus network is to comprehend and visualize CNN activations. Though not specifically made for facial recognition, some conclusions can be made. To facilitate feature extraction, the network analyzes features at various layers. Its applicability to features unique to faces, however, may be restricted. The interpretability component of the suggested system has an impact since it offers a visualization tool for comprehending learned features. Because of their task-specific architectures, specialized networks like FaceNet or DeepFace may provide more immediate advantages for face recognition. FaceNet uses deep neural networks to recognize faces. Gathering data is essential, and FaceNet makes use of a sizable dataset that includes a wide variety of face photos. The diversity of the model aids in its generalization under different conditions.

A deep CNN is used for feature extraction, embedding faces into a hyperspace. FaceNet's triplet loss guarantees a margin of separation between dissimilar faces and the embeddings of similar faces (Schroff, Kalenichenko and Philbin, 2015). High accuracy is displayed in the evaluation, proving the triplet loss and deep embeddings' efficacy. FaceNet's success can be attributed to its use of large-scale datasets and its resilience to a variety of conditions representing factors like geographical representation.

Deep convolutional networks are introduced by DeepID for face recognition. A labeled dataset is used for data collection, and hierarchical feature extraction is made possible by deep architectures. In order to improve generalization, DeepID places a strong emphasis on learning identity-related representations (Sun *et al.*, 2015). Through network training, the method implicitly incorporates feature selection. The assessment shows competitive outcomes. The effectiveness of DeepID can be assigned to its unique architecture, which prioritizes identity-related features; however, the paper does not provide a thorough examination of how learned features can be interpreted. DeepFace emphasizes diversity in its training by utilizing a large dataset. Hierarchical features are extracted by the deep architecture, and attribute classification is integrated into multitask learning.

The influence is found in the model's capacity to accommodate changes. The study, however, might offer more details on how feature representations are impacted by multitask learning. The system's effectiveness is demonstrated by human-level performance on the LFW dataset. The robustness of the system is enhanced by multitask learning and dataset diversity.

The degradation issue in deep networks is addressed by ResNet. The system ensures wide coverage by gathering data for ImageNet. By using residual learning for feature extraction, vanishing gradient problems are mitigated. The system is greatly affected, demonstrating how well residual learning handles deep architectures. The assessment shows that performance gets better as depth increases. The secret to ResNet's success is its capacity to train extremely deep networks, which helps it get past optimization obstacles and boosts accuracy.

The VGG network is used by VGGFace to recognize faces. The simplicity of the VGG architecture helps the model. Data is gathered using a labelled face dataset. In deep convolutional layers, feature extraction is based. The effect is on readability and simplicity, but VGGFace may not be as good at capturing finer details as more sophisticated models. The evaluation's findings indicate competitive performance. VGGFace's straightforward architecture and competitive accuracy make it a viable choice for face recognition tasks.

For face recognition, ArcFace adds an angular margin loss. Datasets with labels on faces are used in data collection. Deep CNNs are used in feature extraction, with a particular emphasis on integrating margin-based loss. The effect is to strengthen discriminative features by emphasizing angular margins. The evaluation shows cutting-edge outcomes. ArcFace's contribution is to increase embeddings' discriminative power, which guarantees greater class separability. Its influence is especially noticeable when obtaining outstanding outcomes on benchmark datasets.

Above discussion proves importance of evaluation of deep learning models as they need to address their short comes and resolve them to improve them further.

Conclusion

To sum up, the research on face recognition that has been chosen, such as FaceNet, Zeiler & Fergus Network, DeepID, DeepFace, ResNet, VGGFace, and ArcFace, shows how far the field has come. These methods achieve remarkable accuracy in face recognition tasks by utilizing creative loss functions, complex neural network architectures, and large-scale datasets. FaceNet uses triplet loss to embed faces into a hyperspace and emphasizes diversity in data collection, demonstrating the power of specialized architectures for facial feature extraction. Despite not being specifically made for face recognition, the Zeiler & Fergus Network offers insightful information about how CNN activations are visualized, which improves the interpretability of learned features. Specific features are crucial for precise face recognition, as demonstrated by DeepID's emphasis on identity-related representations and multitask learning; however, more research on feature interpretability may make the system more useful.

The robustness of the suggested system is highlighted by DeepFace's use of a large dataset, diversity considerations, and multitask learning, all of which help to achieve human-level performance. The fact that ResNet was able to solve the degradation issue in deep networks emphasizes how important architecture design is for solving optimization problems. The architecture simplicity of VGGFace, along with its competitive accuracy, emphasizes the trade-off between model complexity and efficacy in face recognition. With the addition of an angular margin loss, ArcFace has made significant progress toward improving embeddings' discriminative power and achieving state-of-the-art performance. Even though these studies contribute significantly, there are some drawbacks. One persistent challenge is making learned features interpretable, particularly in more complex models. Furthermore, responsible deployment requires a thorough grasp of the biases and real-world implications inherent in these models. By addressing these issues, facial recognition technology will develop and improve even more, guaranteeing its moral and practical application in a wide range of settings.

The authors have shown that Multi-Tablated Convolutional Neural Network (MTCNN) can effectively detect faces and align landmarks. The model's effectiveness is attributed to its robust data collection, allowing it to generalize across diverse facial attributes. However, further discussion on feature selection mechanisms and datasets used for evaluation is needed. Despite these limitations, the paper contributes significantly to face analysis by proposing a joint framework that streamlines face detection and landmark alignment. Future work could focus on refining feature extraction processes and providing more transparency about datasets and evaluation metrics. At the end, each research paper has addressed countless practical and logical issues of face recognition models and improved it further to overcome those issues which become reason to achieve flexile facial recognition in several products and implementations available for us.

References

Deng, J. *et al.* (2019) ‘Arcface: Additive angular margin loss for deep face recognition’, in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 4690–4699.

Haertle, R. and Millis, A.J. (2014) ‘Formation of nonequilibrium steady states in interacting double quantum dots: When coherences dominate the charge distribution’, *Physical Review B*, 90(24), p. 245426.

He, K. *et al.* (2016) ‘Deep residual learning for image recognition’, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778.

Lu, C. and Tang, X. (2015) ‘Surpassing human-level face verification performance on LFW with GaussianFace’, in *Proceedings of the AAAI conference on artificial intelligence*.

Moy, G.M., Hope, J.J. and Savage, C.M. (1999) ‘Born and Markov approximations for atom lasers’, *Physical Review A*, 59(1), p. 667.

Schroff, F., Kalenichenko, D. and Philbin, J. (2015) ‘Facenet: A unified embedding for face recognition and clustering’, in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 815–823.

Sun, Y. *et al.* (2015) ‘Deepid3: Face recognition with very deep neural networks’, *arXiv preprint arXiv:1502.00873* [Preprint].

Yousefi, M.I. and Kschischang, F.R. (2014) ‘Information transmission using the nonlinear Fourier transform, Part II: Numerical methods’, *IEEE Transactions on Information Theory*, 60(7), pp. 4329–4345.

Zeiler, M.D. and Fergus, R. (2014) ‘Visualizing and understanding convolutional networks’, in *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part I 13*, pp. 818–833.

Zhang, K. *et al.* (2016) ‘Joint face detection and alignment using multitask cascaded convolutional networks’, *IEEE signal processing letters*, 23(10), pp. 1499–1503.