



Evaluating CNN Architectures for Real-Time Face Recognition: A Focus on Accuracy and Practical Deployment

Prepared by

Mohammed mushshin alatawi

supervisor

Abstract

The goal of this study is to explore the use of Convolutional Neural Networks (CNNs) in face recognition systems geared towards real-time applications, with the aim to understand the trade-offs among the accuracy, robustness, and computational complexity/consumption. Although CNNs do well at learning hierarchical representations in a facial image, it is also not easy to deploy them in dynamic real world scenarios like light, pose, and occlusion variability as well as device limitations. By comparing two CNNs, namely VGG-16, MobileNet and LeNet this experiment is aimed at analyzing the performance of the models considering some of the primary parameters, namely recognition accuracy, inference time, and resource utilization. Along with that, strategies to improve the generalization and adaptability, e.g. data augmentation, transfer learning, and pooling layer optimization are studied. The work also presents a literature review-based criticism of previous works in an effort to establish gaps in comparative benchmarking, real time testing, and in scaling models of edge devices. Simulated experiments in close real time conditions are performed to determine the model reliability against changes of input. The findings demonstrate the fundamental space between complexity and reality of implementation. Finally, the paper presents some strategic points on the choice and setup of CNN models optimal in terms of real-time face recognition where embedded technology or mobile computing is involved. These results are meant to guide the developers and researchers in a direction of designing efficient and scalable face recognition systems with real-life applications.

Keywords:

Convolutional Neural Networks, Real-Time Face Recognition, Model Optimization

Contents

Chapter 1	3
1.1 Introduction.....	3
1.2 Background.....	4
1.3 Research Problem Statement	7
1.4 Research Objectives	8
1.5 Research Scope.....	8
Chapter 2	10
2.1 Literature review	10
2.2 Comments of Literature Review and research gap	17
Chapter 3	20
3.1 Methodology	20
3.2 Expected Outcomes.....	21
3.3 Gant chart.....	22
References	23

Chapter 1

1.1 Introduction

In the current security, authentication and surveillance world face recognition technology has emerged as an inseparable part of face recognition systems. Out of a variety of ways that were adopted to facilitate an automated face recognition, CNNs have proven to be the most effective in realising this due to their capability to learn hierarchical and discriminative features directly through raw image data despite their learning time taking a longer space. Nevertheless, practical implementation of CNN-based face recognition system in real-time environment continues to be limited by a number of technical issues, including environment variability, resource constraints and cross validation in a variety of data and user conditions. Such obstacles become very important especially in real world application where speed and efficiency is desirable e.g. security systems, mobile applications, and surveillance technologies.

Several studies have also identified weaknesses and advantages of CNNs in the context where they are being used to recognize faces. To illustrate, the research conducted by Kaur et al. (2022) applied a CNN based model to identify face mask during COVID-19 pandemic with more than 95 percent accuracy. Although the system delivered great performance with binary classification tasks on different environmental features, it has a narrow focus as it only dealt with mask detection and was sensitive to occlusions. Conversely, Lu et al. (2021) revealed the effect that data augmentation would have when used on small datasets with 99.5% accuracy in the ORL dataset. Nonetheless, it should be noted that their model was not validated in real-time or during varied conditions and thus, its usefulness was insufficient.

Dar and Palanivel (2021) also considered more sophisticated CNNs, namely, VGG-16, and used transfer learning in real-time faces recognition. They developed an averagely complex VGG-16 model with high accuracy of 99.37%, surviving the simple CNNs. However, the fact that the computational cost was high and it was not robust in different lighting conditions meant that it was questionable as to whether it could run on resource-constrained devices. Pranav and

Manikandan (2020) used a more lightweight strategy by training a simple CNN model using real-time camera feed as inputs.

Their model managed to achieve 98.00 and demonstrated the ability to perform under live environments, yet, it was quite weak in terms of its overall application because of the covered limited dataset diversity.

In that regard, the present experiment addresses the idea of comparing the quality of both deeper (such as VGG-16) as well as lighter (such as MobileNet, LeNet) CNNs regarding real-time accuracy, resilience, and computational performance. It is a necessary comparison to know about trade-offs between model complexity and practically deploy the model on edge devices, or on mobile platforms.

On the whole, these works cumulatively emphasise the success of CNNs in face recognition yet also expose crucial disconnects between accuracy reported in lab settings and real-world applications. The target of this research is to assess the efficiency of the different CNN architectures in real-time based face recognition systems, not necessarily the accuracy but also the robustness and efficiency, the ability to adapt. Based on comparing the performance of different models including VGG-16, LeNet and MobileNet, and assessing their compatibility using a mobile or embedded platform, the study aims at proposing efficient approaches of developing scalable and reliable face recognition technologies.

1.2 Background

Introduction to Face Recognition and Deep Learning

Face recognition is an extension of computer vision and biometric authentication concerned with identifying or authenticating a person on the basis of facial features. It has found a wide application in the real world amalgamating into situations of surveillance, security systems, mobile authentication, law enforcement and even on providing individualised user experience on digital devices. The use of traditional face recognition methods was on the hands crafted feature and shallow learning models, this was very ineffective in situations where lighting, facial expression, occlusion changed.

The performance of face recognition systems has increased lately with the concept of deep learning, especially Convolutional Neural Networks (CNNs). The CNNs is a special form of artificial NN that is implemented to handle grid-structured data (images). They are pre-trained to learn hierarchical feature representations straightforwardly on raw image pixels whereby models can identify edges, textures, and complex facial patterns across several levels (Hassan & Abdulazeez, 2021).

The CNN in Face Recognition

The modern face recognition systems have grown to be based on CNNs because of their being able to generalize well off large datasets. A common CNN framework consists of convolutional layers to extract features, pooling layers to decrease dimensions and fully connected layers to classify. The model allows examining only relevant facial features and reduces noise and irrelevant data with this architecture (Zafar et al., 2024).

But on the other hand, even though CNNs work well in controlled settings, they cannot be used in a real-time system easily. Among them are the requirements to make inferences rapidly, to be resistant to environmental noise, and to be hardware-efficient in terms of communication speeds, memory consumption and so on, particularly in the mobile and embedded applications. There are also other limitations related to the performance of any CNN-based system, which imply the quality and variety of the training data, the depth of the architecture, and adoption of optimization methodologies like data augmentation or transfer learning.

Challenges in Real-Time Face Recognition

Although with large accuracy in the lab, a lot of models of CNN fail in practice. Lighting changes, pose, occlusion (e.g. sunglasses, masks) and low-resolution inputs (e.g. phone cameras) can drastically reduce performance. Additionally, bigger and more complicated architectures like VGG-16, which are more powerful, may have the need of substantial computational resources, thus not being as functional to real-time or on-device implementation (Dar & Palanivel, 2021).

In order to overcome those issues, scientists have studied lightweight frameworks such as MobileNet and LeNet, which provide good balance between accuracy and efficiency. Additionally,

transfer learning (the use of pre-trained weights of large datasets) and data augmentation (application of artificially larger sizes of datasets of transformed images) have proven very promising in constructing better model generalization (Lu et al., 2021).

Optimization with CNN based Face recognition

Research by pioneers aims at discovering ways in which CNN can be harnessed to enhance face recognition systems in various ways:

- **Data Augmentation:** Adding Data either by rotating, flipping or scaling images to the training data can give the model an ability to generalize. Lu et al. (2021) also showed that this technique with a small-sized data set leads to a better recognition outcome of 99.5% accuracy, which outlines the potential of such types of technique in low-resource environments.
- **Transfer Learning:** Pre-trained models such as VGG-16 work well in Face Recognition one can fine-tune them by using these smaller and task-specific data sets. Dar and Palanivel (2021) demonstrated that this direction enhanced the degree of recognition in real-time considerably but, on the downside, its computational needs rose.
- **Optimization on Pooling Layer:** The structure of pooling layers-serving to minimize space- is also essential. According to Zafar et al. (2024), max pooling should be used with high-contrast images and average pooling should be used with noisy inputs. This means that even minute architectural choices can affect model robustness.
- **Hyperparameter Tuning:** At training time, it is important to tune hyperparameters including learning rates, dropout rates, and the size of the filters, in the case of real-time, they will have a substantial influence on both training stability and model accuracy (Pranav & Manikandan, 2020).

Practical Applications and Limitations

In practice, CNN-based face recognition is applicable in such real-world applications as health monitoring, smart surveillance, or mobile authentication. To give one of the examples, CNNs were implemented in face mask recognition systems during the COVID-19 pandemic, and achieved

more than 95 percent classification accuracy in real-time scenarios (Kaur et al., 2022). Nevertheless, even though this indicates practical value in CNNs, these models tend to be task-specific and lack generalizability to the areas related to face recognition, e.g., identity verification or demographic analysis.

In addition, CNNs can be slow on occluded and poor quality images particularly within unconstrained scenes. The research by Hassan and Abdulazeez (2021) highlights the necessity of models that resist pose variation, lighting differences, and partial visibility of the face, which can define the possible way of future development of 3D face recognition and hardware-aware CNN design.

1.3 Research Problem Statement

Face recognition has become one of the most important technologies in the current security, surveillance and user authentication systems. Contrary to this, Convolutional Neural Networks (CNNs) have outperformed other methods because of the auto extraction and learning of hierarchical features of facial images (Hassan & Abdulazeez, 2021). Notwithstanding the high accuracy of face recognition systems under the controlled settings of a laboratory, according to Dar and Palanivel (2021), face recognition systems based on CNN frequently become less accurate when evaluated under the real-world, real-time conditions involving variable lighting conditions, occlusions, changed in position, and complex backgrounds.

Although models consisting of VGG-16 and lightweight architectures with data augmentation have an accuracy level of above 95% on the benchmark datasets (Lu et al., 2021; Kaur et al., 2022), they perform poorly when the deployment environment is dynamic and uncontrolled (Pranav & Manikandan, 2020). This weakness has mainly been pointed out to be a result of poor generalization, biases in data sets, and overfitting because of lack of training diversity. In addition, the upper complexity of these models poses another severe obstacle to implementing them using mobile, embedded, or edge devices with limited resource availability (Dar & Palanivel, 2021).

Despite recent advances on transfer learning, data augmentation, and optimized pooling layers (CNN robustness), more detailed and comparative study of CNN architectures in the study of

practical real-time restrictions is necessary. There is a definite research gap when it comes to determining which CNN models best strike the clever equilibrium of accuracy, efficiency, and deployment possibility given different environmental and hardware environments.

Therefore, this study aims to analyze, compare, and optimize a set of CNN architectures for real-time face recognition, focusing on key performance factors and offering practical recommendations for scalable and efficient implementation in real-world applications.

1.4 Research Objectives

1. Review and compare existing CNN architectures (e.g., VGG-16, LeNet, MobileNet) commonly used in face recognition, analyzing their structural characteristics and performance.
2. Identify key challenges that affect CNN performance in real-time face recognition scenarios, including occlusion, lighting variation, limited training data, and hardware constraints.
3. Evaluate the effectiveness of enhancement techniques such as data augmentation and transfer learning in improving CNN generalization and recognition accuracy.
4. Measure and compare computational efficiency, including inference speed, memory consumption, and real-time applicability across different platforms (e.g., embedded systems, mobile devices).
5. Provide practical recommendations for selecting or designing CNN architectures optimized for real-time face recognition tasks in real-world conditions.

1.5 Research Scope

Despite recent advances on transfer learning, data augmentation, and optimized pooling layers (CNN robustness), more detailed and comparative study of CNN architectures in the study of practical real-time restrictions is necessary. There is a definite research gap when it comes to determining which CNN models best strike the clever equilibrium of accuracy, efficiency, and deployment possibility given different environmental and hardware environments.

The given research is aimed at assessing the level of improvement in the varying architectures of the Convolutional Neural Network (CNN) applied to the real-time face recognition systems and contrasting them. The paper evaluates both heavyweight models (e.g. VGG-16) and lightweight models (e.g. MobileNet, LeNet), to evaluate their utility in the real world where a diligent balance between accuracy, invariance, and computational-efficiency is required.

It covers:

- Architectural assessment of a few CNN models that have been involved in the face recognition activities.
- Accuracy, inference speed, robustness to occlusions/lighting changes, and hardware adaptation requirements as the performance metrics on to which it will be compared.
- Use of techniques to enhance models like data augmentation, transfer learning and pooling techniques to enhance model generalization.
- Testing and testing of models in simulated real-time conditions and on different computing systems (ex., edge devices, mobile systems).
- Comparative analysis in order to give insight on the trade-offs in the model complexity and the potential practical deployment.

Nevertheless, the research does not attempt to introduce a new CNN architecture itself, as well as it contains no hardware-based implementation and optimization (e.g. FPGA design, GPU design). Furthermore, the use of non-deep learning methods of face recognition (PCA, LBP or hybrid SVM) falls outside the realm of interest of this study and is only mentioned within the context of literature.

Focusing on the practical and scalable CNN models that can be applied to real-time face recognition adds some degree of practicality to the given topic of study as the compiled research aims to provide beneficial knowledge to the researcher and developers in applying the deep learning systems in real-world applications like face recognition in terms of surveillance, mobile, embedded systems.

Chapter 2

2.1 Literature review

The face recognition has been developed Long ago as a speculative notion to the reality of a powerful and fundamental aspect of security and surveillance systems today. The strongest innovations have revolved around the Convolutional Neural Networks (CNNs), the type of deep learning model that is known to perform well when it comes to recognizing images. These can learn non-linear representations of faces due to their hierarchical abilities to extract features, something that standard machine learning models can not do so well in. The following paper is a literature review of the six chosen studies based on the file you uploaded, it evaluates the articles and determines their usefulness, usability and findings concerning the task of face recognition using CNN-based systems.

Systematic Trends in CNN-based Face Recognition

In a more complicated approach, Barcic, Grd, and Tomicic (2023) carried out a systematic literature review of 150 scholarly articles on topics related to CNN-based face recognition systems. They did this to reveal the existing tendencies, evaluation indicators, and research issues related to convolutional neural networks in the field. The review suggests that VGG, ResNet, and MobileNet are those most common implementations of CNN architectures and also denotes that there is an increasing trend in using transfer learning, multimodal fusion, and attention mechanisms. Besides, the most frequently applied datasets, such as LFW, CASIA-WebFace, and MegaFace, are discussed and partially related to the aspects of model generalization, robustness, and scalability. The paper highlights emerging issues in the field, especially in the areas of data bias, fairness, privacy, and having adaptive models that can consider the variability of the actual real world environments. Among the main contributions of the review, the fact that it records the steps of transitioning methods based on traditional face recognition methods to the methods that are based on deep learning and the usage of CNNs to extract features more efficiently and reliably can be mentioned. It also suggests promising areas in future research, specifically lightweight model deployment and better generalization with regards to various datasets. On the whole, the

review is an essential starting point to analyze the evolution of CNNs in the context of the face recognition setting and to indicate the tension between accuracy and real-time response and ethical issues.

CNN Emotion Recognition in real-time

Hakim et al. (2024) proposed a real-time facial emotion recognition model that supports situations of extracting emotional expressions identified in live videos, or live streaming, which include happiness, sadness, and surprise. They trained their model on the FER2013 dataset and got accuracy of about 85%. Image preprocessing methodologies of normalization and resizing were used in the work as well as the Open CV library which is used to manage real-time video and processing of frames. Having established that the primal goal of the research is emotion recognition and not identity recognition, it is important to note that the CNN-based research is very applicable in dynamic visual environment analysis.

Using real-life issues like background clutter, variable lighting and occlusion, the model shows the possibility to implement CNNs on life situations. The authors developed a rather light CNN model that can be used in real time and balances between the accuracy and the number of computations and thus is a good choice in mobile and embedded systems. The article is valuable to the general body of research on face recognition that shows the effectiveness of CNN in uncontrollable and real-time contexts. Its results can be used to validate the idea that CNN-based systems under the appropriate optimization may potentially be employed even in scenarios that deals with more than a one-dimensional picture or even in the world as in the case of mental-health, customer service encounters, and adaptive user interfaces.

CNNs to face liveness detection

Koshy and Mahmood (2019) aimed at improving CNN in face liveness detection which is a much needed feature in avoiding spoofing attacks in biometric authentication systems. They study the possibility to optimize CNNs to differentiate between real and fake face data, by utilizing the NUAA liveness dataset to verify the findings. The authors enhanced feature representation and fine-tuned a large architecture using deep CNN by incorporating texture-based nonlinear diffusion preprocessing and obtained a perfect classification accuracy (100%) on the Inception v4

architecture training. Various architectures have different outcomes, such as CNN-5, ResNet50 among others are compared and it is found that Inception v4 with diffusion preprocessing is the most accurate and resistant. This shows that the CNNs can be applicable beyond face recognition domain into security-sensitive tasks where it is imperative to be able to distinguish between genuine inputs. The model has been tested across the typical performance measures, including the accuracy and false acceptance rates, which also proves its reliability. The research is of great value within the field of anti-spoofing technologies because it demonstrates that preprocessing procedures can be utilized synergistically along with deep learning models to improve the detection performance. Regardless of the extreme focus that the paper places on identity recognition, the pertinent architecture innerings, and preprocessing processes are fluid to distinctively analogous issues of CNN-based face recognition applications, especially where real-time, high-security user verification is involved.

Comparative Analysis of CNN Architectures and Databases

Nemavhola, Chibaya, and Viriri (2025) give an extensive analysis of CNN architectures and how they perform on various representative face recognition datasets that have gained popularity like ORL, LFW, CASIA-WebFace, MegaFace. They have contrasted contemporary architectures such as VGG16, FaceNet and MTCNN and they noted that multitask learning and ensemble models performance was advantageous. According to the results, it can be concluded that VGG16 and FaceNet score over 98 per cent accuracy on large-scale data sets compared to traditional machine learning models used and the older CNN architectural designs. One of the main contributions of the study is the elaboration on the characteristics of datasets used in the study, i.e. their resolutions, diversity, and labeling quality and their implications to model training and generalization. The authors indicate that old data, as commonly as applied, restrict the generalizability of CNNs in practical environments. Also, they stress the importance of transfer learning and data augmentation in enhancing the performance, particularly when the computational resources or data set is scarce. With these novel ideas and possibilities, the paper concludes by providing future research recommendations, which included combining CNNs with 3D facial modeling, hybrid ensemble frameworks, and considerations of ethical concerns,

including bias and privacy. This paper offers a superior guide in the choice and performance alignment of CNN architectures depending on data set and setting.

Face Liveness Detection: Performance and Generalization

Khairnar et al. (2025) have carried out an analytic discussion comparing eight readily available CNN models to solve face liveness detection, noting accuracy and operational efficiency as the key factors. The model, viz., DenseNet201, MobileNetV2, ResNet50, and VGG19 was evaluated on three benchmark datasets, viz., NUAA, Replay-Attack, and SiW-MV2. DenseNet201 produced the most accurate result of 98.5 percent on the NUAA dataset, and at the same time, MobileNetV2 showed the most efficient execution in terms of the highest speed (only 15 ms), and minimum memory consumption (45 MB). The cross-dataset validation is a significant strength of this study since it ensures that, indeed, such models as MobileNetV2 and DenseNet201 can be a better generalization and could be a good choice when they are ready to be used in real-time or mobile applications. The study points to the trade off between complexity and feasibility of deployment of a model, which argue strongly in favor of light weight architectures within embedded systems. Moreover, the paper discusses the issue of spoofing attacks, namely, the necessity of the strong CNN solutions which should be able to work in the limited hardware conditions applying the high accuracy. The authors are able to contribute to the building of practical useful face authentication systems by finding a compromise between the quality of detection and the resources that systems require. The results can be greatly utilized in areas that need to have safe on device facial comparison like in the banking, mobile identification, smart surveillance.

Face Mask Recognition Program With the CNN Model

Although not related to a face mask, the system created by Kaur et al. (2022) based on CNN represents a very useful study about CNN efficiency in a real-time, binary classification system against the background of the COVID-19 pandemic. They trained their system on datasets of facial images with tags indicating mask or no-mask so that data augmentation methods, rotation, and flipping confused the system. CNN structure used had layers of convolutional and pooling which were trained to extract features and then had layers of full connection to ensemble feature

extraction followed by fully connected layers to classify followed by dropout to avoid overfitting. Using machine learning libraries like TensorFlow, Keras, OpenCV and Scikit-learn, the implementation could correctly identify or assess they were executed above 95 percent of the time within real-time settings with various lighting conditions and crowds. The metrics used to demonstrate performance of the models were precision, recall and F1-score, as well as confusion matrices to help demonstrate accuracy of models. Although the system was designed to specifically detect the mask, its simplicity of use in real-time and reliability of performance under visual noise underscores the flexibility of CNNs in such areas as surveillance and health monitoring. Yet, the restricted locality of the classification will not allow it to be applicable in wider face recognition applications such as identity verification. Nevertheless, this study showcases how CNNs can be effectively trained and optimized using limited data for specific facial attributes, providing a useful blueprint for deploying CNNs in focused, real-time face analysis systems.

Human Face Recognition Using CNN and Augmented Dataset

Lu, Song, and Xu (2021) addressed the insufficient training data sets issue in face recognition by employing data augmentation models to improve generalization performance of a CNN-based model. The authors used a small corpus of 400 grayscale image faces that were samples of 40 individuals in the ORL database to execute some of these augmentation methods like horizontal flip, shifting of pixels, rotation, and scaling. They used CNN architecture of LeNet-5 and had two convolution layers, two mean-pooling layer, and completely connected classification layer. The augmentation actually spurred them into a bigger dataset, leading to better generalization and lower overfitting. Recognition of the model reached its highest accuracy rate of 99.5 percent beating the conventional methods such as PCA+ANN and PCA+SVM. What they found is that over training epochs the recognition accuracy steadily increased and the mean squared error steadily decreased. It was a conclusion of the authors that, even with rather elementary architecture, CNNs are capable of achieving results better than those provided by classical methods given suitable additions to training. Though the model itself was not tested in real-time or on larger and more various datasets, this study makes a strong argument in support of data augmentation as

an inexpensive, high-reward approach to enhancing CNN performance in limited environments, academic or otherwise resource-poor.

Performance Evaluation of CNNs and VGG in Real-Time Face Recognition

Dar and Palanivel (2021) examined how transfer learning and CNNs that were trained in real-time face recognition systems could apply and noted that VGG-16 and its modified versions could be used. Their experimental context relied on the usage of a dataset consisting of 1 056 real-time and captured images as well as the Olivetti face dataset. The different metrics of performance evaluation were applied, including accuracy, precision, and recall rate, F1-score, and k-fold cross-validation. The extreme VGG-16 model also established an impressive accuracy of 99.37%, which was far much better compared to the simpler CNN architectures whose accuracy was between 69 and 96%. The study also showed that transferring the use of pre-trained weights of large datasets to improve the accuracy of smaller dataset-specific tasks. Nevertheless, its great size and complexity led to concerns that the model could not be deployed on resource-poor systems like mobile or embedded systems as suggested by the high accuracy. How the study would do under variable environmental conditions i.e. poor lighting or occlusion was not explicitly discussed. In any case, its results reinforce the advantages of transfer learning in enhancing the performance of face recognition and introduce an operational benchmark in the usage of future systems in real time. It is believed that the research indicates an accuracy trade-off between deep models and deployment efficiency and invites further research on exploring lightweight alternatives that can be used in real-time settings.

CNN Face Recognition on Real-Time

The purpose of the empirical research conducted by Pranav and Manikandan (2020) was to formulate a model of a real-time face recognition system obtained through a CNN architecture and to test it. First, their model was trained by using the AT&T (ORL) dataset, and then they tested their model using live camera inputs to determine the success of the model in the real world. The CNN model consisted of conventional elements: convolutional layers combined features, the process of max pooling reduced dimensions, and fully connected layers were made with classification. 81 The model obtained 98.75 percent of accuracy on benchmark data and 98.00

percent in live test which presented the ease at which the system responded under real time conditions. The methods of hyperparameter tuning and data augmentation were also included by the authors to increase the effective learning and generalization. The study also had limitations irrespective of its high performance and all these limitations included the homogeneity of the ORL dataset that is devoid of lighting, posing, and ethnic diversities. Additionally, it did not report any numerical measure that was associated to inference time or memory usage, which is rather challenging to analyze the scalability of the model to edge devices. Nevertheless, the experiment realistically closes the gap between theory and practice on the design and training of CNNs and the practical application of even simplest one comprising only a few convolution layers to deliver a high-accuracy level of real-time face recognition by providing a supporting training and tuning strategy.

Pooling Layer Benchmarking in CNNs

A comprehensive benchmarking analysis was carried out by Zafar et al. (2024) and the article was published in 2024 in an effort to provide a clear picture on how variants of pooling layers namely max pooling and average pooling can affect the performance of CNN in different datasets and architectures. Testing these pooling methods with popular CNN models like LeNet, AlexNet, and ResNet with typical classifications like MNIST, CIFAR-10, and CIFAR-100 upon them, the paper has achieved positive results. The experiments demonstrated that max pooling is more suited to operate in high-contrast labeled datasets such as CIFAR-10 whereas average pooling is more suited to operate in noisy environments such as MNIST. To obtain complete results, batch sizes, learning rates, and drop out schemes were changed by the researchers. Although the paper is intended to study some other aspects not specifically related to the face recognition, its findings can be widely used to optimize the CNN architecture in these contexts. The work highlights that aggregation of the decisions may have a considerable influence on the accuracy and robustness of the models relative to the type of input data. Nonetheless, the paper does not comment on the facial image considerations particular to related problems such as occlusion, slight variations of features, and real-time requirements. Moreover, there were no investigations of trade-offs in each pooling type, related to, for example, the memory usage or latency of inference. The results

still provide useful recommendations in the context of CNN design that it might be wise to use suitable pooling strategies depending on the targeted use and the particularity of data.

Face recognition CNN Review

Hassan and Abdulazeez (2021) conducted a thorough literature review that observed the development, architecture, and performance of the CNN architectures in use in face recognition systems. The authors compared different types of architectures like VGG16 and AlexNet and MobileNet networks across the popular face datasets, LFW, CASIA-WebFace, and Megaface. Another major conclusion of the review (which stems, in particular, not to mention the merits of CNN) is that end-to-end CNNs are more effective than hybrid two-component models featuring the integration of CNNs with classical algorithms such as PCA, SVM or LBP.

The authors pointed out that deeper and more complex networks like VGG16 have the benefit of accordance since they are expected to render high levels of accuracy through their ability to extract more complex features whereas the lightweight models like MobileNet are more convenient in real-time applications and embedded environments as they are less resource-intensive. Another issue in face recognition that is covered in the paper are critical challenges such as changes in lighting, pose, and occlusion and how they are handled by the various CNNs. In addition, the authors present areas of future research including the combination of 3D face recognition and hardware-aware CNN optimization and on-device processing. Even though the review does not present any new empirical data, it provides systematic overviews of the available literature and presents strategic pointers in the choice of CNN architectures in accordance with the requirements of the application. That is why it could be used as an effective source of information by researchers and developers in the area of face recognition.

2.2 Comments of Literature Review and research gap

The research reviewed literature gives an in-depth and multi-faceted picture on the usage of Convolutional Neural Networks (CNNs) in face recognition systems. It encompasses a large body of content on architecture design, the diversity of the datasets, training approaches, real-time application, and some others like face liveness detection and emotion recognition. Articles like

those by Barcic et al. (2023) and Hassan & Abdulazeez (2021) provide systematized summaries of existing CNN models, datasets, and evaluations, which creates a sound theoretical basis of understanding the trends of CNN development.

Both deep and lightweight CNN models have been proved to be viable in face recognition both in deployed and constrained environments (e.g., Dar & Palanivel, 2021; Pranav & Manikandan, 2020; Lu et al., 2021). Moreover, studies specializing in optimization methods, including data augmentation, transfer learning, and pooling strategies (Zafar et al., 2024; Lu et al., 2021), could serve as a good source of information about the aspects of improving their generalization and efficiency.

A notable strength of the literature is the diversity of models evaluated, ranging from traditional CNNs to pre-trained architectures like VGG16, Inception, and MobileNetV2. Additionally, several studies addressed challenges such as environmental variability, occlusion, and computational constraints, which are highly relevant to real-world face recognition applications. However, most of the research tends to focus either on ideal datasets or specific, isolated tasks like mask detection or liveness detection, rather than offering a comprehensive comparison of CNN models under real-time, dynamic conditions.

Research Gap

- **Lack of Comparative Evaluation Across CNN Models:**

While many studies assess individual CNN architectures (e.g., VGG16, MobileNet, LeNet), there is a lack of direct, head-to-head comparisons under identical experimental conditions that evaluate accuracy, speed, memory usage, and robustness simultaneously.

- **Insufficient Testing in Real-Time, Unconstrained Environments:**

Most existing research uses controlled or ideal datasets (e.g., ORL, FER2013), which do not reflect the variability in lighting, background, pose, and occlusion found in real-world scenarios.

- **Limited Integration of Optimization Techniques:**

Although data augmentation, transfer learning, and pooling strategies are individually explored, few studies analyze their combined effect on model performance, especially in real-time applications.

- **Neglect of Computational Constraints on Edge Devices:**

Few studies examine inference time, memory consumption, and deployment feasibility of CNN models on mobile or embedded platforms, which is critical for real-time face recognition applications.

- **Overemphasis on Accuracy Without Addressing Efficiency Trade-Offs:**

Several studies report high accuracy but overlook the practical trade-offs between model complexity, latency, and scalability needed for deployment in resource-constrained systems

Chapter 3

3.1 Methodology

Among the notable strong points of the literature is that it consists of evaluations of different models - traditional CNN to pre-trained like VGG16, Inception, and MobileNetV2. Moreover, numerous works were dedicated to such tasks as environmental variability, occlusion, and computation limitation that are the most pertinent to practical face recognition applications. Nevertheless, the larger part of the research is aimed either at optimal datasets or at some special, independent tasks such as mask detection or liveness detection, but not at the general comparison of CNN models in the real-life, dynamic scenario.

In order to meet the research objectives aforementioned, the study shall be associated with a comparative analytical approach that integrates literature-based evaluation and experimental benchmarking of CNN models. The research procedure is organized in the following way:

1. Literature Review

A systematic review of the recent research studies will be used to determine and comprehend popular CNN architectures (e.g., VGG-16, MobileNet, LeNet) utilized in face recognition, their design, and shortcomings and advantages. Such methods as data augmentation, pooling strategies and transfer learning will be investigated as well.

2. Selection The preparation of data set and the selection of model.

A representative collection of CNN models is what will be chosen to compare. The generalization of the model will be increased by preprocessing publicly available image datasets of facial images (e.g. ORL, LFW or a real-time acquired custom dataset) via augmentation.

3. Training and assessment of the model

A consolidated training protocol will be followed in training the selected CNN models to be able to compare them. All the models will be put to the test in simulated real-time settings with different conditions (e.g. lighting, occlusion). The assessment criteria will be:

- Accuracy
- Precision, Recall, F1-Score
- In:ference Time (swiftness)
- Computational usage and memory usage

4. optimization and analysis

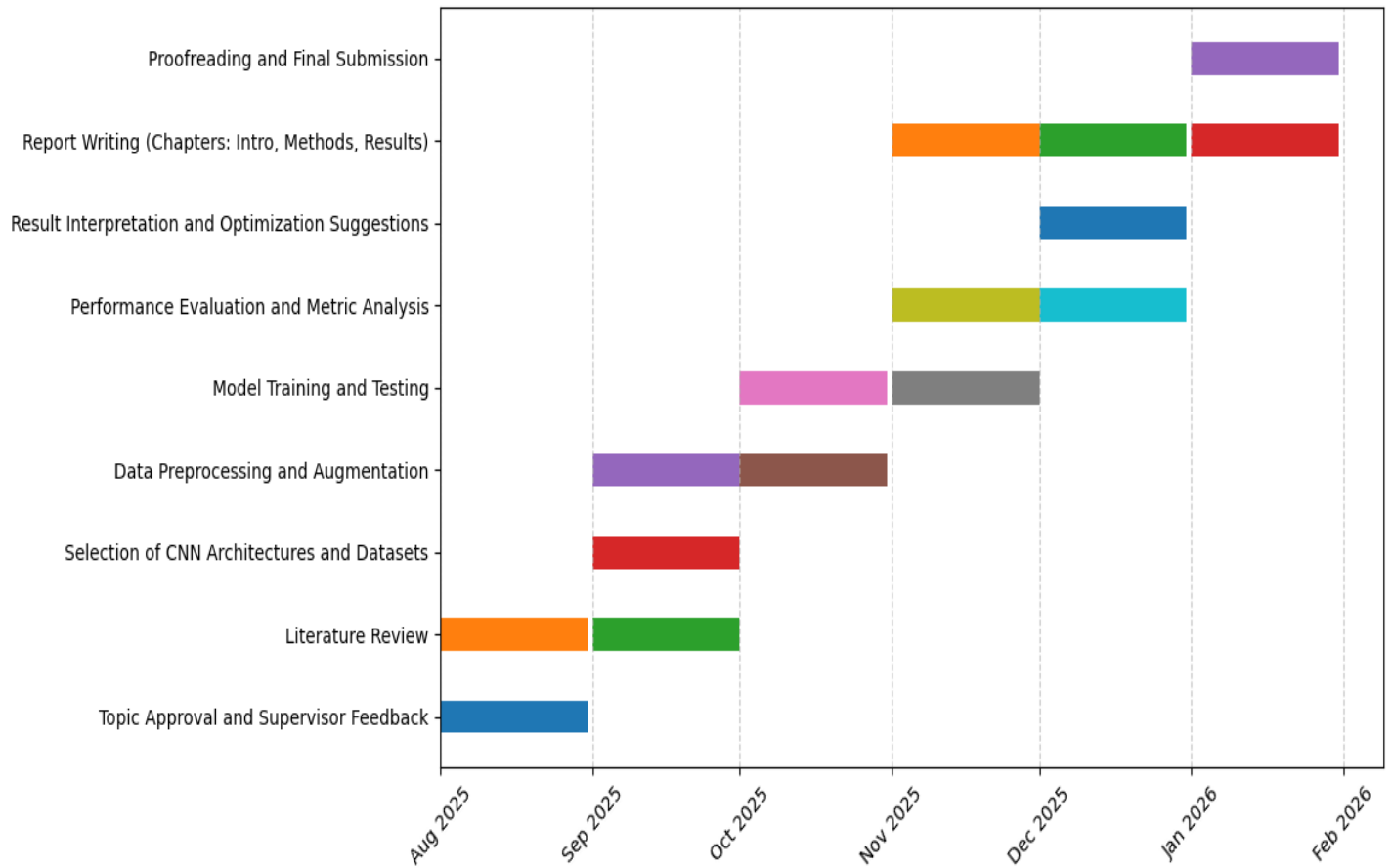
Outputs will be evaluated to determine the best model(s) both in regard to accuracy and the ability to deploy at real-time. It will give recommendations on appropriate model set-ups based on the various real-life conditions.

3.2 Expected Outcomes

The following outcomes are likely to be discovered in this research:

- Comparison of the performance of major CNN architectures applied in face recognition systems, with indication of the accuracy-robustness-computational efficiency tradeoffs.
- Discovery of important architectural and training methods (e.g., transfer learning, data augmentation) to further boost CNNs performance under real-time conditions.
- Useful knowledge on how to decide which models can be deployed on a resource beanie device like a mobile phone or an embedded system.
- A list of guidelines to researchers and developers on implementing and choosing CNN models given the application requirements, such as its implementation into surveillance systems, biometrics authentication, and mobile applications.
- Contribution to the literature through bridging the gap between theoretical model performance and actual, real time performance of face recognition systems based on CNNs.

3.3 Gant chart



References

- Dar, S. A., & Palanivel, S. (2021). Performance evaluation of convolutional neural networks (CNNs) and VGG on real-time face recognition system. *Advances in Science, Technology and Engineering Systems Journal*, 6(2), 956–964. <https://dx.doi.org/10.25046/aj0602109>
- Hassan, R. J., & Abdulazeez, A. M. (2021). Deep learning convolutional neural network for face recognition: A review. *International Journal of Science and Business*, 5(2), 114–127. <https://doi.org/10.5281/zenodo.4471013>
- Kaur, G., Sinha, R., Tiwari, P. K., Yadav, S. K., Pandey, P., Raj, R., Vashisth, A., & Rakhra, M. (2022). Face mask recognition system using CNN model. *Neuroscience Informatics*, 2(3). <https://doi.org/10.1016/j.neuri.2021.100035>
- Lu, P., Song, B., & Xu, L. (2021). Human face recognition based on convolutional neural network and augmented dataset. *Systems Science & Control Engineering*, 9(sup2), 29–37. <https://doi.org/10.1080/21642583.2020.1836526>
- Pranav, K. B., & Manikandan, J. (2020). Design and evaluation of a real-time face recognition system using convolutional neural networks. *Procedia Computer Science*, 171, 1651–1659. <https://doi.org/10.1016/j.procs.2020.04.177>
- Zafar, A., Saba, N., Arshad, A., Alabrah, A., Riaz, S., Suleman, M., Zafar, S., & Nadeem, M. (2024). Convolutional neural networks: A comprehensive evaluation and benchmarking of pooling layer variants. *Symmetry*, 16, 1516. <https://doi.org/10.3390/sym16111516>
- Barcic, E., Grd, P., & Tomicic, I. (2023). Convolutional Neural Networks for Face Recognition: A Systematic Literature Review. <https://doi.org/10.21203/rs.3.rs-3145839/v1>
- Hakim, G., Simangunsong, G., Ningrat, R., Rabika, J., Rusafni, M., Giri, E., & Mindara, G. (2024). Real-Time Facial Emotion Detection Application with Image Processing Based on Convolutional Neural Network (CNN). *International Journal of Electrical Engineering, Mathematics and Computer Science*, 1(4), 27–36. <https://doi.org/10.62951/ijeemcs.v1i4.123>

Koshy, R., & Mahmood, A. (2019). Optimizing Deep CNN Architectures for Face Liveness Detection. *Entropy*, 21(4), 423. <https://doi.org/10.3390/e21040423>

Nemavhola, A., Chibaya, C., & Viriri, S. (2025). A Systematic Review of CNN Architectures, Databases, Performance Metrics, and Applications in Face Recognition. *Information*, 16(2), 107. <https://doi.org/10.3390/info16020107>

Khairnar, S., Gite, S., Pradhan, B., Thepade, S. D., & Alamri, A. (2025). Optimizing CNN Architectures for Face Liveness Detection: Performance, Efficiency, and Generalization across Datasets. *CMES - Computer Modeling in Engineering and Sciences*, 143(3), 3677–3707. <https://doi.org/10.32604/cmes.2025.058855>