



Evaluating the Effectiveness of Convolutional Neural Networks (CNN) in Face Recognition Systems

Prepared by

Mohameed mushahhin alatawi

Contents

Introduction	2
Paper 1: Face Mask Recognition System Using CNN Model (Kaur et al., 2022)	2
Paper 2: Human Face Recognition Using CNN and Augmented Dataset (Lu et al., 2021).....	4
Paper 3: Performance Evaluation of CNNs and VGG in Real-Time Face Recognition (Dar & Palanivel, 2021).....	6
Paper 4: Real-Time Face Recognition with CNN (Pranav & Manikandan, 2020)	8
Paper 5: Pooling Layer Benchmarking in CNNs (Zafar et al., 2024)	9
Paper 6: CNN Review for Face Recognition (Hassan & Abdulazeez, 2021).....	11
Comparative Summary Table	13
Findings Synthesis	14
Conclusion.....	14
References.....	15

Introduction

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.

Paper 1: Face Mask Recognition System Using CNN Model (Kaur et al., 2022)

Kaur et al. (2022) made available a CNN-based model that aims at determining whether a person is wearing a face mask in an image or video feed, conceived during the COVID-19 pandemic, wherein the decision to force people to wear face masks became a matter of public health concern. This model combines a number of the machine learning tools and libraries, i.e., TensorFlow, Keras, OpenCV, and Scikit-Learn, to come out with a real-time detection system whereby the system is optimized in terms of dealing with a high level of accuracy as far as lighting and environmental conditions are concerned. There is a major data set that was unfortunately employed, which has been facial images labeled as either no mask or mask and contrast data augmentation techniques, including certain flipping and rotation, were applied to improve the generalization and model performance on previously unseen data. The CNN architecture to be proposed has several convolution and pooling layers that extract features and fully connected layers to do final classification. The model uses dropout layers to prevent overfitting, and the hyperparameter adjustment was used in order to optimize learning rates and the batch sizes used, so that the training process was sound. To evaluate, the main parameters of accuracy, precision, recall and F1-score were used, as well as confusion matrices to see a detailed analysis of model predictions. They registered an impressive performance of the system, which was over 95 percent accurate and used this to prove that it had strong detection capabilities in real time,

even in environments that had varying crowding and lighting levels. Such findings validate the effectiveness of CNNs in restricted binary recognition problems under the bigger umbrella of face recognition problems.

Critically speaking, the Strengths of the paper are evident both in its application and practical implementations. The real-time functionality of the model is particularly interesting considering the possibility of deployment in such a public surveillance setting as hospitals and transport terminals. Further, the data augmentation with the dropout regularization is the sign of knowing the major traps of deep learning, namely overfitting on small or biased data. Strong performance measures also enhance the paper, providing the transparency and the numerical standards that would be useful to other researchers who might also be inclined to take up such an exercise. The scalability of the model also is another important advantage; the integration of the model with the well-known image processing libraries guarantees its use with various hardware and software environments.

But, the research does not lack limitations. Of greatest significance, its classification system is tightly constrained to the two-way mask vs. no-mask label, whereas despite the benefits of using it in the context of surveillance during a pandemic, it has little to no applicability to the cases of general face recognition, e.g., identity checking or demographic assignment. Moreover, the results of the model depend much on the quality of clear images inserted into it. Most of them could fail accuracy due to weak resolutions, occlusions (sunglasses, hands), and head angle extremities, which are recognized in the discussion section but not thoroughly experimentally evaluated or addressed in the implementation. Although the authors propose the future research in which the enhanced models utilize the combination of CNNs and alternative methods of artificial intelligence to reduce the mentioned shortcomings, the current work is limited in its potential.

Finally, Kaur et al. (2022) present a purposively and effectively implemented application of CNNs in the subdomain of face mask detection suggesting the flexibility and robustness of the CNNs, as well as the application of the technology, in the context of real-time surveillance. Despite the lack of contextual relevance and the positioning of the work itself as the less-than-holistic solution to

face recognition, the work helps gain the insights into ways of optimizing CNNs to perform binary tasks related to the detection of facial attributes. The publication also provides the valuable roadmap of CNN implementation at the top of broadly supported machine learning libraries and thus can be easily implemented and extended in further studies. The fact that the authors reveal the limitations of the system and give their recommendations on how to improve the system provides the study with additional academic credibility. Even in the confined focus of the studies, contributions are robust evidence to the power of CNNs under real world computer vision tasks.

Paper 2: Human Face Recognition Using CNN and Augmented Dataset (Lu et al., 2021)

Lu, Song, and Xu (2021) deal with one of the biggest issues when it comes to using CNNs in face recognition: insufficient and lack of diversity in the training datasets. As the researchers are aware, numerous CNN models require large and highly varied datasets to learn and generalize efficiently, so they offer a workable alternative, in which a CNN-based face recognition model is paired with a rigorous data augmentation approach. The idea behind them is to augment a fairly small dataset 400 grayscale images of human faces representing 40 people and form a much larger variety of training data by using augmentations including horizontal flip, pixel shifts, scaling (90-% to 110-percent), and rotation (20-degrees to negative 20-degrees). Since the model architecture is based on the LeNet-5 architecture, it has 2 convolutional layers (input: C1, C2), two layers of mean-pooling C (input: S1, S2), and a fully connected layer that generates 40 identity classes. This convolutional-pooling alternating structure guarantees the extraction and compression of high-level features and the simple activation function and backpropagation facilitate to optimize the network to correctly classify.

Systematic image transformation led to augmented data, that, in turn, provided the exponential growth in the training data samples, which was reported to improve the CNN generalization and minimize overfitting, which is also one of the most frequent drawbacks of working with small datasets in deep learning. The researchers confirmed the working of the model at diverse epoch/dataset, showing a steady rise in recognition accuracy and a simultaneous drop in mean squared error (MSE) accomplished during the training iterations. In particular, the model received

a recognition accuracy of up to 99.5% (Figure 3), which is much higher than that of ordinary methods such as Artificial Neural Networks (ANN), Principal Component Analysis with ANN (PCA+ANN), the conventional support vectors as PCA+SVM and Wavelet+SVM. The findings demonstrate the promise of properly designed data augmentation to increase the performance of CNNs in limited data settings.

When analyzing the research critically, the major strength of the research is in the convincing argument that data augmentation has the capacity to circumvent datasets constraints encountered during the use of CNN. This approach also provides a repeatable system to boost model performance without any new raw labeled data, in practice where gathering big labeled collections of faces can be constrained either by privacy issues or logistics. The systematic appraisal to sample size and counting periods of the study gives valuable information in relation to model development and optimization. What is more, the authors compare the suggested model of CNN to other conventional methods, thus putting it aptly into the context of the superiority of CNNs in facial feature extraction and recognition.

However, there are also limitations that are associated with the study. This study is too narrow in that it uses the ORL dataset and thus the results cannot be generalized much. The fact that the ethnicity of the people included in the dataset lacks diversity, as does the pose testing variation, the lighting conditions, and even the amount of noise on recordings, also means that it was not applicable to real-life situations where such variables are not regulated. In addition, the implementation is performed in a fully offline nature and the study is not tested in real-time environments which do not answer the question on how the model performs on dynamic input conditions or under the extreme conditions presented by edge devices. Also, the CNN model no matter how well it performed in the confines of the experiment failed to discuss the computing budget of training on the augmented dataset, which is likely to have a heavy frontier in practice.

In conclusion, the paper by Lu et al. (2021) provides a rather interesting solution to a practical issue with face recognition based on CNN: the lack of data. Using dataset augmentation they show that the robustness and accuracy of the model can be improved dramatically even when a small base dataset is provided. Although the limited experimental parameters and small dataset

that the study follows limit its effectiveness, it can still be recommended as a useful point of reference in regard to researchers who have to handle limited data. The evidence demonstrating the fact that CNNs, behind the narrative of the work, embraces such crucial concepts as the involvement of the death cap in the emergence of the timeless, the sensation of nothingness (nothingness), the sense of loss. even in their relatively simple architectures, outperform traditional face recognition techniques when appropriately trained and tuned. As such, this paper contributes meaningfully to the evolving body of literature on enhancing CNN performance in low-data environments.

Paper 3: Performance Evaluation of CNNs and VGG in Real-Time Face Recognition (Dar & Palanivel, 2021)

The article of Dar and Palanivel (2021) studies how deep learning methods could be applied to real-time face recognition systems to analyze the results of a Model developed based on CNN and the approach of transfer learning with pre-training based on the VGG-16 model. The dataset applied in this paper consists of 1056 real-time captured images of 24 persons as well as the standard Olivetti face dataset. The framework of the evaluation of the performance included three steps: construction of an image database, training the model with CNN and VGG structure, and a complete check of the performance based on macro-averaged precision, recall, F1-score, and k-fold cross-validation. An Increase of its accuracy to 99.37% using attention modules and softmax classifier enhanced VGG-16 model provided a better output compared to the basic CNN architecture, which performed with an accuracy of 69.09%. Medium architectures (VGG3, VGG7) came at 75.71 percent and 96.53 percent respectively as architectures demonstrated a steady rise in performance levels as complexity and depth of training increased.

The main contribution of the study is in its in real-time use and verified performance setting up a working benchmark to be used with face recognition systems deployed in the real world. The authors address the shortcomings of small datasets by the use of transfer learning and exploit the credibility of generalization of the pre-trained models, which were trained on a massive facial image repository. More importantly, this research highlights the advantage of having feature extraction systems which are automated within CNN-based applications and that where pre-

existing feature engineering systems are used, several pipelines will be necessary, complicating calculations and introducing more possible errors. Multiple evaluation metrics also increase the strength of the results denoted in the paper since the efficiency of the model will not be judged based on accuracy rather on the overall quality of the classification.

However, despite these strengths there is also the shortcoming of the study. One of its important constraints is failure to solve the illumination problem, which is a notorious issue in face recognition activities. They can severely distort the features of the face, and this is why though the VGG-16 model demonstrated good results with ideal lightning, they did not analyze how it would behave in more diverse environmental settings. Also although VGG-16 was able to attain a near human rate, its complexity makes it inefficient to implement in real-time especially in mobile or embedded platforms where system resources are limited. These concerns are identified in the study but the authors do not suggest and/or evaluate the use of solutions, e.g. lightweight CNN architectures (e.g. MobileNet), or hardware-accelerated approaches. The second gap is the lack of interpretability; the authors do not explain what particular characteristics or layers in VGG-16 were the most influential to the model which is a missed chance at steering iteration of the architecture.

To sum up, the article by Dar and Palanivel (2021) thoroughly proved the effectiveness of the models based on the transfer learning, in terms of their recognizing face photos in real time and providing high levels of accuracy. The findings support the practical application of deep CNN architecture but also create awareness that there should be a balance between the performance and efficiency. Although the paper does not discuss the solutions to deployment on limited devices or under different lighting conditions, it creates a good background that can be further extended in this area. Its empirical evidence and methods can be used by researchers and developers in order to achieve better and more effective face recognition systems that can be deployed using scalable and real-time frameworks.

Paper 4: Real-Time Face Recognition with CNN (Pranav & Manikandan, 2020)

Pranav and Manikandan (2020) contain a detailed and realistic experiment on the design and benchmark of an actual-time face recognition system with the Convolutional Neural Networks (CNNs). This is driven by the requirement of swift, reliable faces recognition systems that can be robust in real-time, particularly in surveillance, access and interactive systems. The scientists created a model based on CNN, and it was first trained on the AT&T (ORL) face dataset before being tested in real-time with the help of live camera input. They constitute feature extraction convolutional layers, dimensionality reduction layers (max pooling) and finally a classification layer (layers made of fully connected nodes). Notably, the authors have included parameter tuning practices and the strategies of data augmentation to maximize the capacity of the model to learn and generalize different data inputs. They obtained outstanding recognition rates, namely 98.75 percent on the standard data set and 98.00 percent in real-time mode, which proves the efficiency and flexibility of their suggested framework.

The fact that the paper empirically shows the strength and resilience of CNNs even in the environments not intended during the development of the models is one of the major strengths the paper has. Contrary to most laboratory-based studies, this experiment extends the gap between a theory design of a model and its implementation. The fact that the camera input into the testing process is real time lends more reality and relevance to the research that also goes a long way in boosting the credibility of the study. The system performance was very stable as it shifted gears into a dynamic environment which was not controlled. Further, the authors stress the necessity of hyperparameter tuning practices, which was a vital key to achieving the highest degree of accuracy and stability of the model by modifying hyperparameter values such as the learning rate, filters size, a number of layers, and the drop-out rates values. Artificial growing methods such as image rotation and image scaling were applied to extend the size of the dataset and minimize the effect of overfitting the model as well as enhance its capability to generalize using different facial inputs.

Although the study has good strength, there are limitations, which should be mentioned. First, the dataset that has been used, the AT&T dataset, is composed of gray scale images of only 40 people in rather scarce lighting conditions, facial expression, and pose. Consequently, it does not exactly illustrate what one can expect in reality where facial images are highly non-homogeneous in terms of resolution, amount of background clutter, ethnicity, age, and accessories (e.g., glasses or hats). The authors recognize this shortcoming although they fail to provide an empirical experiment with the more challenging datasets like LFW or CASIA-WebFace. Moreover, the article does not clearly mention effects of higher environmental complex elements, e.g. shadows, occlusion, background distractions etc. on the model performance. They are some of the important variables in surveillance and access control cases where the real-world input is not always ideal.

Moreover, even though the CNN model is efficient and optimized to be used in the research, the paper does not investigate the implications of increased recognition accuracy and computational cost-effectiveness. Such a lack is of particular issue when looking at deployment to edge environments, e.g. embedded systems or on mobile systems, where the hardware resources available are constrained. There is no detail given about inference time, memory or model size; inference time is often critical in production, model size and memory will all determine the viability of deep learning models in real time applications. The authors fail to compare their solution with more recent lightweight CNN designs or other methods of similar or improved level of accuracy with less computational load, such as transfer learning.

Paper 5: Pooling Layer Benchmarking in CNNs (Zafar et al., 2024)

Zafar et al. (2024) explore one of the most important, but least heard, details of CNN architecture, pooling layers. Although CNNs are renowned to represent features at different levels in picture data, their performance can depend greatly on architectural design, in particular, pooling forms utilized to downsample spatial size and preserve needed features. It is a systematic assessment of two popular pooling techniques, max pooling and average pooling in varying datasets and CNNs. The three common models used by the authors are LeNet, AlexNet, and ResNet, where the authors test their models on the common benchmark datasets, such as MNIST, CIFAR-10, and

CIFAR-100. They evaluate the effectiveness of pooling strategies with a broad range of experimentations in different settings of the batch size, learning rate and dropout option. This indicates that max pooling is especially useful in structured data where there is high contrast such as that in the case of CIFAR-10 whereas average pooling is more effective in low contrast settings, which are noisy e.g. MNIST.

One of the great and important contributions of this study is that it is methodologically rigorous. The experimental design provided by the study is highly effective as it assesses pooling approaches within different variables and models. It is especially important to the researchers and practitioners interested in empirical findings that help them make architectural choices in CNN design. The obtained results are not only statistically significant (with p-values support the observed trends), but also have practical insight: e.g., the conclusion about max pooling being most preferable in operations demanding acute preservation of features makes sense given its typical application as high-resolution object detection. On the same note, the effectiveness of average pooling in denoising is also confirmed, bringing some valuable advice to the use of applications involving noisy data. The visualizations and performance measures that can be used to compare the results to the accuracy in various pooling conditions can also be provided which only can support the validity of the conclusions made in the analysis.

Nevertheless, in spite of the experimental richness of the research, it contains some serious limitations especially related to the fact that it is relevant to the field of face recognition. Although the provided datasets, namely MNIST, CIFAR-10 and CIFAR-100, are common in computer vision benchmark evaluations, they lack an image of human faces. In this respect, the non-specificity of the results in relation to face recognition applications is not really ascertained. The type of objects and their distribution in pictures is rather specific and might be more complicated compared to simpler, center-centered objects the use of which might be found in these data. In addition pooling methods may not react equally to two aspects which are subtle differences in facial shapes, texture and expression which may be more involved in facial recognition tasks.

The other disadvantage is that there is lack of details on the computational cost or efficiency trade-off between the various types of pooling, which is sensitive in a resource-constrained environment. As an example, the better max pooling might retain, useful feature-wise, the size of activation mapping might be bigger and memory consumption can be higher too. These are not examined in the study, even though they play an essential role in real-time or embedded applications. More advanced variants of pooling e.g. global pooling, stochastic pooling and adaptive pooling, some of which already achieved success in the literature due to their increased flexibility and results were not discussed in the paper either.

Paper 6: CNN Review for Face Recognition (Hassan & Abdulazeez, 2021)

The article by Hassan and Abdulazeez (2021) provides an in-depth literature review of using deep learning in face recognition, more precisely Convolutional Neural Networks (CNNs). In their paper, they review the development, structure, approaches and the behavior of different CNN-based architectures namely VGG16, AlexNet, and MobileNet on popular datasets of faces like LFW, CASIA-WebFace, and Mega-face. The authors review the implementing architecture of CNNs in detail, covering convolutional layers to extricate features, pooling layers to reduce dimensionality to the space, and fully connected layers to generate classification, and point out the collaboration of the component to train to learn high-dimensional patterns of faces. Hybrid techniques (like CNNs with other types of machine learning techniques like PCA, SVM, LBP) are also evaluated in the review and it is said that despite the fact hybrids used to yield better performance in the past, nowadays they are beaten by the end-to-end CNNs since the former can learn features hierarchically.

It is also one of the main strong points that this paper is broad in gathering various works into an organized story. The survey will reach back to the establishment of basic CNN structures and the current development of transfer learning and lightweight networks, providing a broad overview of the community. It talks through an example, such as how models such as VGG16 are able to achieve high accuracy of recognitions through their biggest contributions of having deep layers and small filters in use whereas MobileNet sacrifices some recognitions in order to increase speed and efficiencies, thus usable on mobile and embedded devices. Also, the paper highlights critical

problems of face recognition, which include lighting, occlusion, and pose changes, and discusses how different CNNs tried to combat them. The consideration of the future study directions such as the promise of 3D face recognition and implementation of CNNs with hardware optimization to real-time use cases are strategic features of the review.

However, the paper also has certain limitations especially because it is a secondary study. As opposed to empirical studies allowing obtaining quantitative confirmation, this review does not introduce novel empirical data and standards. Its results are a secondary source and though detailed are not compared in a systematized fashion within a framework of equal evaluation. There is therefore, a trade-off between the usefulness of the insights, which are interpretive rather than conclusive in nature. Moreover, the review at certain points mentions more dated hybrid models such as PCA+SVM, when compared to newer inventions such as attention mechanisms, residual networks and vision transformers which have not been discussed thoroughly.

Hassan and Abdulazeez (2021) present an informative review in an organized way that will help the new researchers entering the field of CNN-based face recognition. The paper has a robust theoretical background, places the existing different CNN architectures in context, and solves the real-life aspects of the processing speed and environmental variability. Though it does not offer any innovative experimentation or reflect on the latest changes pertaining much to the topic, it is still a useful source of information in terms of how the CNNs have developed thus far, and how they are expected to change the field of the face recognition technology.

Comparative Summary Table

Paper	Year	Authors	Methodology	Key Findings	Weaknesses
1. Face Mask CNN	2022	Kaur et al.	CNN with Keras, TensorFlow	95%+ real-time accuracy; strong binary classification	Sensitive to occlusion; limited to masks
2. Augmented CNN	2021	Lu et al.	CNN + data augmentation	99.5% accuracy on ORL dataset	Lacks real-time and multi-dataset validation
3. CNN vs. VGG	2021	Dar & Palanivel	VGG-16 + Transfer Learning	VGG-16: 99.37% accuracy in real-time	High resource consumption; no illumination handling
4. Real-Time CNN	2020	Pranav & Manikandan	CNN with live inputs + tuning	98.00% accuracy in real-time	Limited diversity in dataset
5. Pooling Layers	2024	Zafar et al.	Pooling variant benchmarking	Max pooling better in contrast; average in noise	Indirect to face recognition
6. CNN Review	2021	Hassan & Abdulazeez	Comparative literature review	CNNs > hybrids for occlusion, accuracy	No empirical study; lacks modern methods

Findings Synthesis

The effectiveness of CNNs to recognize faces is proved by all the six studies with averaged accuracy within the systems more than 95%. Nevertheless, the performance is to a large extent influenced by methodological diversity (transfer learning, data augmentation, pooling optimization). Although there was research that showed better outcomes using transfer learning (Dar & Palanivel, 2021) and aggressive augmentation (Lu et al., 2021), there were other papers based on the problem of binary classification in terms of pandemic use-cases (Kaur et al., 2022).

There was contrast in real-time performance. Such as, simple CNN based networks had a hard time again VGG-16, which demonstrates the significance of depth of design and pre-training. In addition, Hassan & Abdulazeez (2021) highlighted the resistance to lighting and occlusion, referring to such future directions as the lightweight CNNs and 3D recognition.

The optimisation of pooling strategy as displayed in Zafar et al. (2024) introduces an architectural tuning layer that can be underestimated yet effective in a limited setting.

Conclusion

CNNs have transformed the face recognition technology in terms of providing us higher accuracy and real-time as well as flexibility. This review of the literature shows that despite their consistent relative performance, architectures of CNN such as VGG-16 can be improved dramatically through augmentation and transfer learning as well as fine tuning of architecture.

Nevertheless, there are still constraints. Real time application accuracy is yet not free of environmental factors like lighting and occlusion. Future research should focus on more robust models, 3D recognition integration, and lightweight architectures for mobile deployment.

The reviewed papers offer comprehensive insight into both foundational principles and innovative strategies in CNN-based face recognition, establishing a strong theoretical and practical foundation for future exploration in the field.

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>