IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A Comparative Study Of Cnn And Transfer Learning Based Models For Face Recognition

[1]Sumedha Arya, [2]Nirmal Gaud

Abstract:
Facial Recognition (FR) is a technique that helps recognize people's faces and has wide applications. New methodologies based on computer vision, deep learning, and transfer learning have been used to perform face recognition. However, limitations still exist. This study proposes a comprehensive methodology that evaluates multiple Convolutional Neural Network (CNN) and transfer learning-based models, including VGG16, InceptionV3, and CBAM-enhanced CNNs, us- ing a Kaggle celebrity face dataset. The VGG16-based model achieved the highest validation accuracy (88.65%) and lowest loss (0.4267), while a custom CNN (256 units, 128x128) offered a balance of accuracy (85.13%) and compu- tational efficiency. Challenges such as overfitting and computational cost were observed, particularly in regularized models. Future work aims to enhance generalization through data augmentation, hyperparameter tuning, and model ensemble techniques to address scalability and real-world distortions.

Index Terms: Facial Recognition (FR), Deep Learning, 3D Face Modeling, OpenCV, LS-SIFT Descriptor

1 Introduction

Facial Recognition (FR) is a prominent biometric technique. It is applied in security, healthcare, and identity verifica- tion. Recent advances have changed FR from traditional to advanced deep learning and hybrid approaches, addressing challenges such as illumination, pose variations, occlusions, and aging [Gururaj et al., 2024, Abdelbar et al., 2024]. Various surveys emphasize the role of large-scale datasets, generative AI for 3D modeling, and tools like OpenCV and, Python, in FR. However, they also highlight unresolved issues such as cross-sensor compatibility, ethical concerns, and robustness [Gururaj et al., 2024, Abdelbar et al., 2024]. Optimization techniques have gained focus, which reduce com- putational complexity while maintaining high accuracy in resource-constrained environments [Ouloul et al., 2025, Shi et al., 2024]. Novel approaches, such as the augmentation method by [Malakar et al., 2024] and the LS-SIFT descrip- tor, further improve accuracy and robustness in diverse datasets [Malakar et al., 2024, Lin and Otoya, 2024]. Despite progress, challenges such as scalability, real-world distortions, and computational efficiency still persist [Ho et al., 2024, Zhu et al., 2025, Robbins et al., 2024].

This paper describes the various face recognition techniques, with latest algorithms. After that, we have proposed our research methodology on face recognition using different CNN and transfer learning-based models.

2 Literature Review

Facial Recognition (FR) is a crucial technique in biometrics, including other applications such as security, healthcare, and identity verification. The survey conducted by the authors highlights face recognition techniques from conventional to sophisticated hybrid and deep learning approaches [Gururaj et al., 2024]. They categorize FR methods to address challenges such as illumination variations, pose differences, occlusions, and aging. Their review further emphasizes the importance of large-scale datasets and the integration of generative AI for 3D facial modeling. Moreover, they recognize ethical considerations and privacy-preserving mechanisms for better FR systems development [Gururaj et al., 2024].

In similar research, the authors conducted a systematic review of FR techniques, focusing on security, identity ver- ification, and autonomous systems [Abdelbar et al., 2024]. Their work extends the survey by [Gururaj et al., 2024], which classifies FR methods into different advanced techniques, including appearance-based and hybrid models in deep learning and 3D facial recognition [Gururaj et al., 2024, Abdelbar et al., 2024]. Both studies highlighted the prominent role of OpenCV and Python in building FR-based applications. However, despite technological progress, core issues such as cross-sensor compatibility, ethical concerns, and robustness remain unsolved [Gururaj et al., 2024, Abdelbar et al., 2024].

Recent developments in FR systems have increasingly emphasized optimization techniques to increase performance with low-cost embedded architectures, reducing computational complexity while maintaining high accuracy. The authors [Ouloul et al., 2025], proposed an embedded FPGA-based FR system utilizing the centrally overlapped blocks-local binary pattern (COB-LBP) descriptor, achieving high recognition rates and real-time processing [Ouloul et al., 2025]. Their research highlights the trade-offs between hardware limitations and algorithm robustness. The COB-LBP method demonstrates significant effectiveness in resource-constrained environments [Ouloul et al., 2025].

The authors [Shi et al., 2024] introduced a lightweight model, LighterFace, integrated with CSPNet and ShuffleNetv2 with a Global Attention Mechanism (GAMAttention) [Shi et al., 2024]. Their approach results in an 85.4% reduction in computational load and a 66.3% increase in processing speed compared to YOLOv5 while maintaining 90.6% detection accuracy. Designed for edge devices such as the Raspberry Pi, LighterFace showcases its practical viability in community security applications [Shi et al., 2024].

A comprehensive review by the authors discusses how face recognition systems process collected facial images using automated equipment [Li et al., 2016]. They discussed important evaluation criteria and benchmark databases for researchers to assess system performance. The authors identified existing limitations in practical applications and suggested future solutions, including the development of specialized cameras with high image quality and 3D techniques to handle facial rotation and occlusion. Their future work focuses on advancements that could enhance capabilities in image filtering, reconstruction, and denoising [Li et al., 2016].

The authors [Malakar et al., 2024] introduced a novel approach that improves the accuracy of recognition by aug- mentation, occluding the lower portion of the face rather than reconstructing the entire facial structure [Malakar et al., 2024]. Their approach preserves individual identity more effectively than conventional inpainting methods. The pro- posed methodology integrates CNN-based feature matching with SURF-based geometric alignment, producing realistic full-face images from masked inputs and achieving 4-6% higher accuracy than existing techniques. The proposed model reduces computational complexity while maintaining robustness across datasets such as LFW and CASIA-WebFace [Malakar et al., 2024].

The authors [Lin and Otoya, 2024] introduced LS-SIFT, an innovative learned descriptor that enhances traditional SIFT by incorporating landmark-specific nonlinear transformations, thereby increasing the robustness of images [Lin and Otoya, 2024]. Their framework integrates head pose classification for visible landmark selection with a novel Mahalanobis Similarity (MS) as the base learner model. It achieved remarkable accuracy on the CMU-PIE dataset and over 94% Rank-1 accuracy on Multi-PIE datasets. The two-stage methodology of the system outperformed conventional SIFT by 4-8% across multiple benchmarks while preserving robustness [Lin and Otoya, 2024].

The authors [Sohail et al., 2024] proposed a YOLO-V5-based real-time face matching framework that analyzes dif- ferent facial orientations [Sohail et al., 2024]. It utilizes a multi-pose pattern recognition technique where facial features are aligned according to spectral similarity in specific regions. This approach effectively minimizes computational com- plexity while preserving high accuracy. In addition, the system presents an innovative error function that integrates execution time, accuracy loss, and identity loss [Sohail

et al., 2024].

The authors [Ho et al., 2024] performed a systematic review of 70 studies published between 2018 and 2023, analyz- ing Principal Component Analysis (PCA)-based methods in modern contexts [Ho et al., 2024]. This research categorizes Eigenface applications, benchmarking datasets, and outlines implementation pipelines. The technique focuses on exist- ing challenges of sensitivity to lighting, occlusions, and scalability limitations. The study also integrates traditional and contemporary research for variants of the eigenface, surveillance, and emotion detection. For the future, they proposed focusing on dataset-specific performance benchmarking with deep learning architectures [Ho et al., 2024].

The introduction of Ghost modules by the authors addressed the challenges of employing cost-effective linear trans- formations for feature maps [Alansari et al., 2023]. This technique optimizes the computational cost of the transformation process. The authors proposed GhostFaceNets—a series of lightweight models with attention mechanisms to enhance feature representation. When evaluated on benchmark datasets, GhostFaceNets achieved high performance with lower computational requirements compared to conventional CNNs [Alansari et al., 2023].

The authors conducted a survey on FR methods for traditional and deep learning-based algorithms [Wang et al., 2022]. They conducted an analysis to check the performance of these algorithms under various conditions such as illumination, pose, and occlusion. The research highlighted a shift from hand-made features to end-to-end neural networks in FR methods. They also examined the influence of large-scale datasets on model generalization. In addition, the authors explore emerging trends such as self-supervised learning and bias mitigation, along with open challenges in cross- domain robustness and computational efficiency [Wang et al., 2022].

The authors proposed a novel technique to identify the relationship between feature norms and image quality in FR [Gim and Sohn, 2024]. According to them, higher feature norms show more similarity to those learned by deep learning models. They introduced the Noise Direction Regularization (NDR) technique, which comprises noise samples identified by feature norms to enhance FR performance in low-resolution images. However, the approach currently detects only a subset of noise samples, with future work aimed at improving precision [Gim and Sohn, 2024].

The authors proposed DEFOG, a cross-age FR algorithm that enhances deep learning with an attention mechanism and Arcface loss to address challenges for age-related facial changes [Zhu et al., 2025]. By integrating Retinaface for face detection and an improved Resnet-50 model with attention mechanisms, the method extracts robust facial features. It achieved higher accuracy and robustness across diverse age groups. Despite being effective, the large-scale feature extraction network of the model presents challenges for deployment in resource-constrained systems [Zhu et al., 2025]. The authors introduced DaliID, a robust method for FR and person re-identification under real-world image distor- tions [Robbins et al., 2024]. The approach employs a novel distortion-adaptive technique with atmospheric distortion augmentation and an easy-to-hard adaptive weighting schedule. Additionally, a magnitude-weighted feature fusion of parallel distortion-adaptive and clean backbones improves performance across high- and low-quality images. DaliID achieves state-of-the-art results on seven benchmarks, including IJB-S, TinyFace, and MSMT17, and introduces new long-distance (750+ meters) datasets for evaluating realistic distortions. Future work aims to explore self-paced curricu-lum learning and diverse backbone combinations [Robbins et al., 2024].

3 Research Methodology

This research methodology consists of data collection, preprocessing, face detection, model development, training, and evaluation.

3.1 Data Collection and Preprocessing

The dataset is taken from Kaggle, containing images of celebrity faces. A CSV file of metadata with image IDs and corresponding celebrity labels is provided for annotation. Images are loaded and CSV file is read to map image IDs to celebrity labels. Then, images are resized to 224x224 pixels for models requiring this input or 128x128 pixels for the CBAM-enhanced model to standardize input dimensions. Pixel values are normalized to the range [0, 1] by dividing by 255 to improve model convergence.

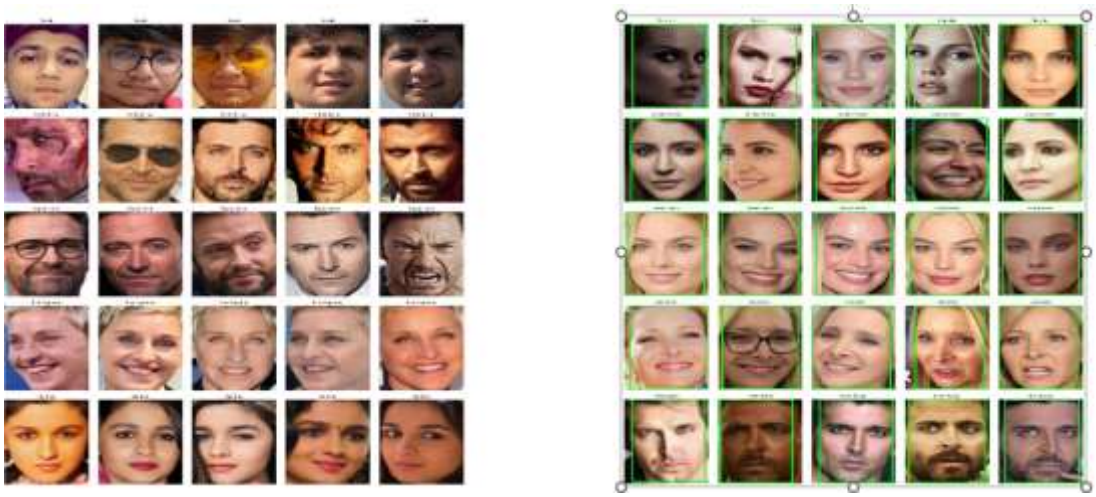

Figure 1: celebrity faces and celebrity faces recognized by mtcnn

3.2 Face Detection

A deep learning-based face detection model, called MTCNN is used, that detects facial bounding boxes with higher accuracy. It ensures precise face extraction for model training.

3.3 Model Development

Distinct models were developed and evaluated to compare their performance in face recognition. They are as follows:

1. **Baseline CNN:** A simple CNN with three convolutional layers (32, 64, 128 filters), each followed by max-pooling. Includes a dense layer with 512 units (ReLU activation), dropout (0.5), and a softmax output layer. Input size: 128 × 128 × 3. No regularization applied.
2. **Custom CNN (256 units, Regularized):** A deeper CNN with four convolutional layers (32, 64, 128, 256 filters), followed by max-pooling. Includes a dense layer with 256 units (ReLU activation) with L1 (0.001) and L2 (0.001) regularization, batch normalization, dropout (0.7), and a softmax output layer. Input size: 128 × 128 × 3.
3. **Custom CNN (128 units, Regularized):** Similar to the 256-unit variant but with a dense layer of 128 units. Includes L1 (0.001) and L2 (0.001) regularization, batch normalization, dropout (0.5), and a softmax output layer. Input size: 128 × 128 × 3.
4. **InceptionV3-Based Model:** A transfer learning model using pre-trained InceptionV3 (with frozen weights), fol- lowed by global average pooling, a dense layer with 512 units (ReLU activation), dropout (0.5), and a softmax output layer. Input size: 224 × 224 × 3. No regularization applied.
5. **CBAM CNN:** A CNN with three convolutional layers (32, 64, 128 filters) enhanced with Convolutional Block Attention Modules (CBAM), followed by max-pooling. Includes a dense layer with 512 units (ReLU activation), dropout (0.5), and a softmax output layer. Input size: 224 × 224 × 3. No regularization applied.
6. **VGG16-Based Model:** A transfer learning model using pre-trained VGG16 (with frozen weights), followed by flattening, a dense layer with 512 units (ReLU activation), batch normalization, dropout (0.5), and a softmax output layer. Input size: 224 × 224 × 3. Uses data augmentation for regularization.
7. **Custom CNN (512 units, 224×224):** A CNN with four convolutional layers (32, 64, 128, 256 filters), followed by max-pooling. Includes a dense layer with 512 units (ReLU activation), batch normalization, dropout (0.7), and a softmax output layer. Input size: 224 × 224 × 3. No regularization applied.
8. **Custom CNN (512 units, BatchNorm):** A CNN with three convolutional layers (32, 64, 128 filters), followed by max-pooling. Includes a dense layer with 512 units (ReLU activation), batch normalization, dropout (0.5), and a softmax output layer. Input size: 128 × 128 × 3. No regularization

applied.

9. **Custom CNN (256 units, 128×128):** A CNN with four convolutional layers (32, 64, 128, 256 filters), followed by max-pooling. Includes a dense layer with 256 units (ReLU activation), batch normalization, dropout (0.5), and a softmax output layer. Input size: 128 × 128 × 3. No regularization applied.

Each model is designed to balance accuracy, computational efficiency, and generalization, with variations in depth, input resolution, and regularization strategies to address overfitting and enhance performance.

3.4 Training and Validation

The face recognition dataset is split into training and validation sets, with images resized to the respective input sizes based on the model requirements. Adam optimizer is used for all models with default learning rate, except for VGG16, which uses learning rate reduction. Batch size is 32 for all models. Epochs varied by models from 20 to 40. Data augmentation is applied to VGG16 to enhance generalization.

4 Results and Discussion

4.1 Validation Accuracy and Loss

The key observations from the Table are:

- The VGG16-Based Model achieves the highest validation accuracy (0.8865) and lowest validation loss (0.4267).
- The Custom CNN (256 units, 128x128) performs strongly with a validation accuracy of 0.8513 and a low valida- tion loss of 0.5633, making it a lightweight but effective model. However, it is showing some overfitting.
- The CBAM CNN yields the lowest validation accuracy (0.6634).
- Regularized Custom CNN models (256 and 128 units) show moderate performance (0.7730 and 0.7495 validation accuracy), but high validation losses (2.2034 and 1.9012) suggest overfitting despite regularization.

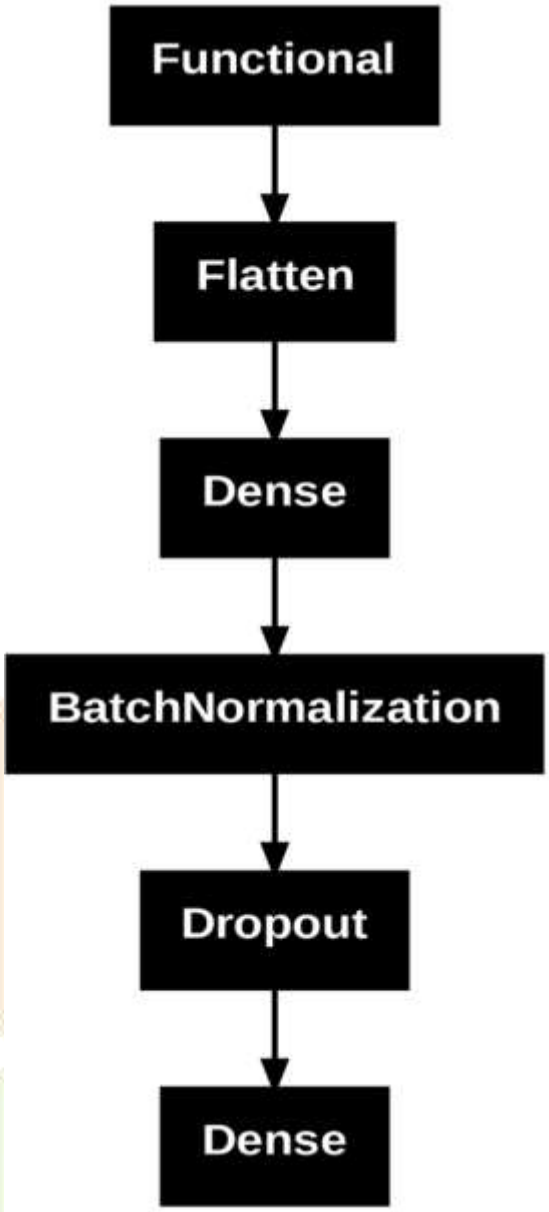

Figure 2: vgg16 transfer learning model architecture

4.2 Training Performance

The Key Observations of the training performance of all the models is as follows:

- The Custom CNN (512 units, BatchNorm), Custom CNN (512 units, 224x224), and Custom CNN (256 units, 128x128) achieve perfect training accuracy (1.0000) with extremely low training losses (0.0042, 0.0079, 0.0121), but their validation accuracies (0.8180–0.8513) indicate overfitting.
- The Baseline CNN shows high training accuracy (0.9645) and low training loss (0.0998) with the fastest training time (~1.21s per epoch), making it computationally efficient.
- The VGG16-Based Model has the longest training time (~24–32s per epoch) due to its deep architecture and data augmentation, but its validation performance justifies the computational cost.
- The InceptionV3-Based Model balances training accuracy (0.9079) and computational efficiency (~5.81s per epoch), but its validation accuracy (0.7319) is lower than VGG16.

The research methodology prioritizes models based on validation accuracy, validation loss, and computational effi- ciency for face recognition selection. VGG16-Based Model is selected as the primary model due to its superior vali- dation accuracy (0.8865) and low validation loss (0.4267). Despite its high computational cost (~24–32s per epoch),

Table 1: validation accuracy and loss for all models

Model	Val. Accuracy	Val. Loss	Epochs	Input Size	Regularization
Baseline CNN	0.6947	1.7243	20	128 × 128 × 3	None
Custom CNN (256 units, Regularized)	0.7730	2.2034	40	128 × 128 × 3	L1=0.001; L2=0.001
Custom CNN (128 units, Regularized)	0.7495	1.9012	40	128 × 128 × 3	L1=0.001; L2=0.001
InceptionV3-Based Model	0.7319	0.9929	20	224 × 224 × 3	None
CBAM CNN	0.6634	1.7996	20	224 × 224 × 3	None
VGG16-Based Model	**0.8865**	**0.4267**	30	224 × 224 × 3	Data Augmentation
Custom CNN (512 units, 224 × 224)	0.8395	0.6096	20	224 × 224 × 3	None
Custom CNN (512 units, BatchNorm)	0.8180	0.6641	20	128 × 128 × 3	None
Custom CNN (256 units, 128 × 128)	0.8513	0.5633	20	128 × 128 × 3	None

Table 2: training performance of all models

Model	Final Training Accuracy	Final Training Loss	Training Time per Epoch (s)
Baseline CNN	0.9645	0.0998	~1.21
Custom CNN (256 units, Regularized)	0.9133	1.6053	~4.60
Custom CNN (128 units, Regularized)	0.9530	1.1221	~4.57
InceptionV3-Based Model	0.9079	0.3426	~5.81
CBAM CNN	0.9440	0.1793	~2.29
VGG16-Based Model	0.8599	0.5344	~24–32
Custom CNN (512 units, 224 × 224)	1.0000	0.0079	~4.63
Custom CNN (512 units, BatchNorm)	1.0000	0.0042	~1.41
Custom CNN (256 units, 128 × 128)	1.0000	0.0121	~1.44

its robustness, enhanced by data augmentation and pre-trained weights, makes it ideal for high-accuracy applications. Secondary Choice could be Custom CNN (256 units, 128x128), offering a validation accuracy of 0.8513, low validation loss (0.5633), and fast training time (~1.44s per epoch).

5 Conclusion and Future Work

In conclusion, we will extend the data augmentation to all models to improve generalization, as demonstrated by VGG16. Also, optimize learning rates, batch sizes, and regularization parameters using grid search for hyper-parameter tuning. We will also refine CBAM CNN by increasing training epochs or simplifying attention modules to better suit the dataset. Top-performing models can also be combined to potentially improve accuracy. This research ensures flexibility for future improvements and adaptability to varying computational constraints.

References

1. Abdelbar et al. Face recognition using python and opencv: Review. *International Journal of Scientific Research Publications*, 14(1):71–71, January 2024. doi: 10.29322/IJSRP.14.01.2024.p14508.
2. M. Alansari et al. Ghostfacenets: Lightweight face recognition model from cheap operations. *IEEE Access*, 11, 2023. doi: 10.1109/ACCESS.2023.3266068.
3. T. Gim and K.-A. Sohn. Regularization using noise samples identified by the feature norm for face recognition. *IEEE Access*, 12, 2024. doi: 10.1109/ACCESS.2024.3453030.
4. H. L. Gururaj et al. A comprehensive review of face recognition techniques, trends, and challenges. *IEEE Access*, 12: 107903–107903, 2024. doi: 10.1109/ACCESS.2024.3424933.
5. H.-T. Ho et al. Face detection using eigenfaces: A comprehensive review. *IEEE Access*, 12, 2024. doi: 10.1109/ ACCESS.2024.3435964.
6. L. Li et al. A review of face recognition technology. *IEEE Access*, 4:1–XX, 2016. doi: 10.1109/ACCESS.2016.2600527.
7. S. D. Lin and P. E. Linares Otoya. Ls-sift: Enhancing the robustness of sift during pose-invariant face recognition by learning facial landmark specific mappings. *IEEE Access*, 12, 2024. doi: 10.1109/ACCESS.2024.3406911.
8. S. Malakar, W. Chiracharit, and K. Chamnongthai. Masked face recognition with generated occluded part using image augmentation and cnn maintaining face identity. *IEEE Access*, 12, 2024. doi: 10.1109/ACCESS.2024.3446652.
9. M. I. Ouloul, Z. Moutakki, A. Amghar, and K. Afdel. Low-cost embedded facial recognition system based on overlapped local binary pattern. *e-Prime-Advances in Electrical Engineering, Electronics and Energy*, 11:100924, 2025. doi: 10.1016/j.prime.2025.100924.
10. W. Robbins, G. Bertocco, and T. E. Boult. Daliid: Distortion-adaptive learned invariance for identification—a robust technique for face recognition and person re-identification. *IEEE Access*, 12, 2024. doi: 10.1109/ACCESS.2024. 3385782.
11. Y. Shi et al. Lighterface model for community face detection and recognition. *Information*, 15(4):215, 2024. doi: 10.3390/info15040215.
12. M. Sohail et al. Deep learning based multi pose human face matching system. *IEEE Access*, 12, 2024. doi: 10.1109/ ACCESS.2024.3366451.
13. X. Wang et al. A survey of face recognition. arXiv preprint arXiv:2212.13038, 2022.
14. Zhu et al. Defog: Deep learning with attention mechanism enabled cross-age face recognition. *Tsinghua Science and Technology*, 30(3):1342–1358, June 2025. doi: 10.26599/TST.2024.9010107.