International Journal of Basic and Applied Sciences, 14 (3) (2025) 311-316

International Journal of Basic and Applied Sciences

Website: www.sciencepubco.com/index.php/IJBAS
https://doi.org/10.14419/0jt7m549
Research paper

POSE-Inclusive Face Recognition: Addressing The Influence of Face Angle in Person Identification

C. J. Harshitha *, R. K. Bharathi, Rakshitha P.

Dept. of Computer Applications JSS Science and Technology University Mysuru, India
**Corresponding author E-mail: harshi.cj@jssstuniv.in*

Received: June 18, 2025, Accepted: July 17, 2025, Published: July 27, 2025

Abstract

Face recognition remains a critical task in computer vision, especially in applications involving authentication, surveillance, and access control. However, real-world challenges like pose variations and occlusions continue to hinder recognition performance. This paper presents a pose-inclusive face recognition system that integrates Multi-task Cascaded Convolutional Networks (MTCNN) for face detection and alignment with ResNet-50 for feature extraction and classification. The Labeled Faces in the Wild (LFW) dataset is used, and images undergo data augmentation and normalization to simulate pose diversity. Experimental evaluation demonstrates that ResNet-50 significantly outperforms traditional CNN models, achieving an accuracy of 99.60%. The proposed approach ensures robust and scalable performance in uncontrolled environments.

Keywords: *Occluded Face Recognition; MTCNN; CNN.*

1. Introduction

Face recognition is a critical task in computer vision for wide-ranging applications of biometrics like authentication systems, surveillance systems, identity verification, human-computer interaction, and many more. However, one of the main challenges in deploying face recognition systems in real-world scenarios is the variability in face orientation, also known as pose variation. Non-frontal faces result in partial occlusion of key facial features such as eyes, nose, or jawline, significantly degrading the performance of traditional recognition systems that are often optimized for frontal face images.

To address this, we propose a deep learning-based face recognition pipeline that explicitly treats pose variation as a form of occlusion. Our system integrates Multi-task Cascaded Convolutional Networks (MTCNN) for robust face detection and alignment, followed by ResNet50, a 50-layer deep residual network, for feature extraction and classification. MTCNN leverages three cascaded stages—P-Net, R-Net, and O-Net—to localize faces and key landmarks (eyes, nose, mouth), enabling consistent face alignment before recognition. The aligned faces are then processed by ResNet50, which uses residual learning to overcome vanishing gradients and extract discriminative embeddings from high-dimensional input images.

We use the Labelled Faces in the Wild (LFW) dataset, known for its unconstrained face images, to train and evaluate our model. To simulate real-world conditions and improve generalization, we apply data augmentation by rotating faces within a ±45° range. Each face is normalized using a pixel intensity rescaling method by ensuring uniform intensity distribution and stability during training. Through this design, the proposed system aims to deliver consistent face recognition performance despite the challenges posed by pose variation and partial occlusion.

2. Literature review

The core challenge of face recognition lies in accurately identifying or verifying individuals based on facial features. In earlier approaches, face recognition systems primarily relied on frontal face images, where the subject was positioned directly in front of the camera. However, with the advancement of deep learning and convolutional neural networks (CNNs), recognition systems have become more accurate and robust, allowing for higher performance under real-world conditions.

In particular, the development of embeddings such as FaceNet has revolutionized face recognition by creating a unified embedding space, where face identities are encoded as vectors, and distances between these vectors correlate with identity [4]. Despite these advances, face recognition systems still struggle with challenges such as pose variations and occlusions. These problems significantly affect the recognition process in uncontrolled environments, where faces may appear at different angles or be partially obscured by masks or other objects [1].

One of the most challenging issues in face recognition is dealing with pose variations. Faces captured from different angles (yaw, pitch, and roll) exhibit distortions in facial features, making it difficult for traditional models to match them to known identities. Side views of faces present significant challenges because key features such as the eyes, nose, and mouth are often not visible or are heavily occluded,

which reduces recognition accuracy [1]. Studies like those of Cao et al. [13] and Naser et al. [3] emphasize the importance of addressing yaw poses, as they cause significant distortions in the face's appearance.

One potential solution to this problem is to use deep learning techniques to learn pose-invariant representations of faces. MTCNN (Multi-task Cascaded Convolutional Networks) can be used for face detection and alignment, particularly in situations where faces are captured in non-frontal views. MTCNN is effective in detecting faces in various poses and angles by localizing the key facial landmarks, such as the eyes, nose, and mouth, in an efficient manner [23]. Once these landmarks are identified, the face can be aligned to a canonical pose, allowing for more accurate recognition, even when pose variations are present.

Occlusions, such as the presence of glasses, masks, or other objects covering parts of the face, pose another significant challenge for face recognition. Masks that obscure the nose and mouth, with two critical features for identification, can drastically reduce the performance of face cognition models [3]. Moreover, the combination of occlusions and pose variations further complicates the problem. For instance, wearing glasses and a mask simultaneously can obscure both the eyes and the mouth, which are essential for distinguishing individuals in a face recognition system [17]. Researchers have proposed various strategies to mitigate the impact of occlusions, such as identity-diversity inpainting [14], which reconstructs the occluded parts of the face, and multi-task contrastive learning methods like those used in FocusFace [16], which are designed to handle masked faces.

Recent research suggests that pose variation can be considered a form of occlusion. This perspective stems from the fact that, like occlusions, pose variations can obscure key facial features. By treating pose variation as an occlusion problem, researchers can apply similar strategies for handling occlusions to address pose-related challenges in face recognition [12]. For example, the use of pose-adaptive loss functions in models like PoseFace [11] has proven effective in improving recognition accuracy by focusing on the most critical features of the face, even when some are occluded by pose distortions. These pose-adaptive loss functions are computationally light yet provide good robustness to extreme and moderate pose variations and are thus appropriate for real-time applications. They may, however, continue to perform poorly when too many facial features are absent.

Another promising approach is the integration of 3D face models into recognition systems. By using 3D facial data, models can generate synthetic frontal views from side or angled profiles, effectively compensating for the loss of information caused by pose variations or occlusions [12]. Although 3D models are very robust to extreme poses, they are computationally expensive and demand high-quality 3D data and are thus less suitable for use on resource-limited devices.

Additionally, attention mechanisms have shown success in enhancing face recognition systems under pose and occlusion challenges. Tsai and Yeh [10] introduced the Pose Attention Module (PAM), which uses attention mechanisms to focus on the most discriminative parts of the face. By attending to regions of the face that are less affected by pose and occlusion, PAM helps improve the robustness of the recognition system, making it less sensitive to variations in pose. In comparison to 3D reconstruction, attention mechanisms such as PAM are computationally efficient and flexible but can be heavily dependent on observable facial features, thereby restricting their performance in the event of extreme occlusion.

Incorporating depth information is another advanced technique for handling pose variations and occlusions. Depth-based approaches have been shown to enhance face recognition by providing supplementary data about the geometry of the face. For example, Hu [6] combines fine-level facial depth generation with RGB-D complementary feature learning to improve recognition accuracy, especially when faces are captured from varying angles or are partially occluded by masks or glasses. However, the success of depth-based approaches is subject to the presence of depth sensors, whose integration might be impractical under many real-world scenarios.

Furthermore, ResNet-based architectures, such as those used in FaceNet, have proven to be effective in addressing the challenge of pose variation. ResNet's deep residual learning allows for the efficient learning of complex features, even when the input image contains distortions due to pose variations. By learning residual mappings, ResNet can capture subtle differences in facial features, which helps mitigate the impact of pose and occlusion [4]. Using ResNet in conjunction with MTCNN for face detection and alignment can significantly improve performance in real-world settings where both pose variations and occlusions are common.

Kim et al. [7] also propose a KeyPoint Relative Position Encoding method, which encodes the relative positions of key facial features. This technique helps the model focus on
the spatial relationships between facial points, which remain relatively constant even under pose variations or occlusions. By incorporating this method into a deep learning framework, face recognition systems can become more resilient to pose and occlusion challenges.

Face recognition systems continue to evolve, with significant advancements made in handling pose variations and occlusions. By adopting novel techniques such as pose-adaptive loss functions, depth-based reconstruction, and attention mechanisms, researchers have improved the robustness of face recognition models. Furthermore, integrating MTCNN for face alignment and ResNet for feature learning is effective in mitigating the challenges posed by pose and occlusion. As the field progresses, these methods will continue to play a crucial role in enhancing the accuracy and reliability of face recognition systems, particularly in uncontrolled environments [2].

3. Methodology

The proposed face recognition system is designed to address the challenges posed by pose variations in unconstrained environments. As in Fig.1, the methodology involves several key stages: dataset preparation and preprocessing,
face detection and alignment, feature extraction, and classification. Together, these components form an end-to-end pipeline capable of learning pose-invariant facial representations and accurately identifying individuals.

To begin with, the system utilizes the Labeled Faces in the Wild (LFW) dataset, which contains over 13,000 face images of 5,749 individuals captured in real-world settings. Due to computing and hardware constraints, the research used a subset of 100 classes from the original LFW database. The subjects were chosen to maintain a balance between performance measurement and resource limitations. Efforts were made to include people with different ethnicities, genders, and combinations of facial poses to reduce possible biases and enhance generalizability. This was to ensure that the chosen subset still had some diverse characteristics of the larger dataset. The images are stored in a structured directory format, with each subfolder containing multiple samples of the same person. Preprocessing these images is critical for improving the consistency and quality of training data. Initially, all faces are detected and cropped using Multi-task Cascaded Convolutional Networks (MTCNN), which isolates facial regions while eliminating background noise. After detection, pixel-wise normalization is performed using min-max scaling to ensure uniform pixel intensities. The normalization is expressed as in eqation-1.

$$I_{norm} = \frac{I - I_{min}}{I_{max} - I_{min}} \quad (1)$$

Where I represents the input pixel intensity, and $I_{max} = 255$ and $I_{min} = 0$ represents the maximum and minimum pixel values, respectively. To further enhance generalization, data augmentation is applied by rotating each cropped image randomly within a range of 45° to +45°, generating five variations per image. This helps simulate pose differences commonly encountered in practical scenarios.

For face detection and alignment, MTCNN is employed due to its high precision in identifying facial regions and key landmarks such as the eyes, nose, and mouth corners. It operates in three sequential stages: the Proposal Network (P-Net) generates initial face candidate regions, the Refinement Network (R-Net) filters out false positives and improves bounding box precision, and the Output Network (O-Net) refines these results and localizes facial landmarks for alignment.

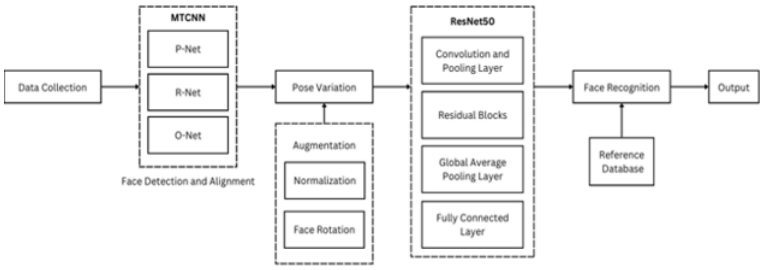


Fig. 1: Pipeline of the Proposed Work.

This can be seen in Fig.2. The detection process is governed by a multi-task loss function in equation 2

$$L_{total} = L_{cls} + \lambda_1 L_{bbox} + \lambda_2 L_{landmark} \tag{2}$$

Where L_{cls} represents the classification loss, L_{bbox} the bounding box regression loss, and $L_{landmark}$ the facial landmark localization loss. The hyperparameters λ_1 and λ_2 control the contribution of each term. The detected and aligned faces are then passed to the feature extraction stage.

The ResNet50 architecture (Fig.3) includes an initial 7×7 convolution layer, followed by multiple bottleneck residual blocks, and ends with a global average pooling layer that reduces the spatial dimensions of the feature maps. The resulting feature vector is passed through a fully connected layer, and the final classification is performed using the Softmax function in equation 3

$$P(y_i) = \frac{e^{z_i}}{\sum_j e^{z_j}} \tag{3}$$

Where z_i is the output logit for class i, and the denominator normalizes the logits into a probability distribution over all classes. The images in the dataset were resized to 224×224 and passed to a ResNet-50 model.

The training was done using the Adam optimizer (learning rate: 1×10^{-4}, batch size: 32) with cross-entropy loss for 25 epochs.

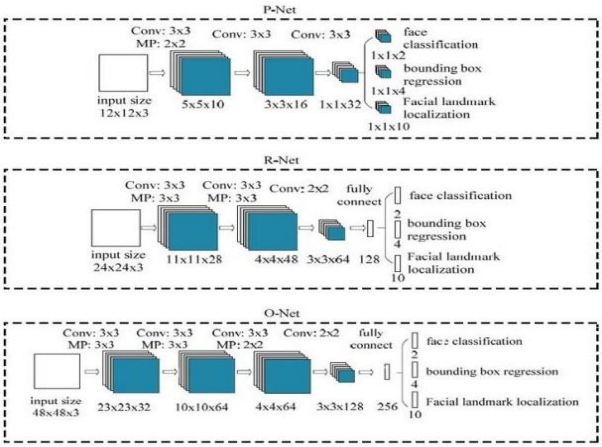


Fig. 2: Architecture of MTCNN.

This of this training are shown in Table 1.

Fig.4 shows the final recognition pipeline. It begins with input images from the LFW dataset, which are pre-processed through detection, alignment, normalization, and augmentation. The aligned and enhanced images are fed into ResNet50, which extracts high-level discriminative features. These features are then classified into one of the 100 identities using a Softmax classifier. The entire system is evaluated using standard performance metrics such as accuracy, precision, recall, and F1-score to assess its robustness and effectiveness under pose-induced occlusions.

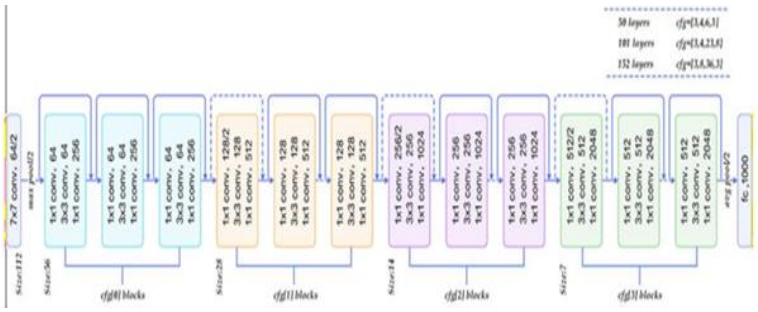


Fig. 3: Architecture of ResNet50.

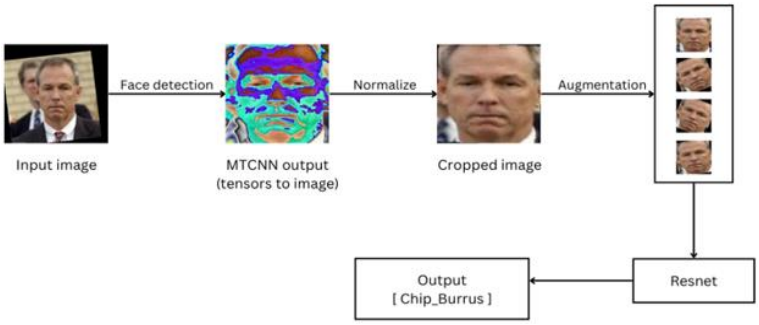


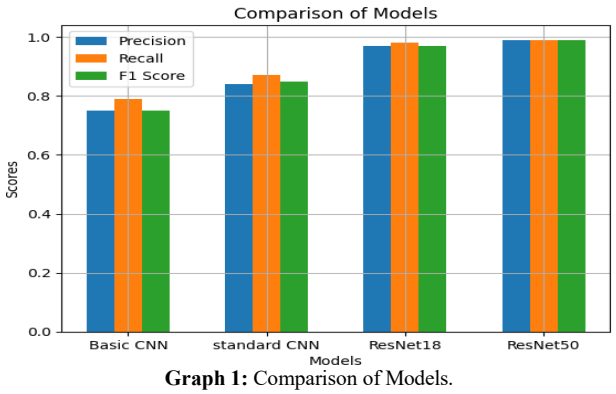
Fig. 4: Visualization of the Proposed Work.

4. Results and discussion

The proposed face recognition model was evaluated using a subset of the Labeled Faces in the Wild (LFW) dataset, comprising 100 individuals. After preprocessing, which included face detection using MTCNN, normalization, and data augmentation via random rotations between 45° to +45°, the dataset was split into training and testing sets with an 80:20 ratio. The model was trained using the ResNet50 architecture for feature extraction and classification. The evaluation focused on analyzing the performance of the model under varying pose conditions and its ability to maintain high recognition accuracy despite angular variations introduced during augmentation. To measure the system's effectiveness, standard evaluation metrics including accuracy, precision, recall, and F1-score were employed. Accuracy was computed as the ratio of correctly predicted face identities to the total number of predictions. Precision and recall provided insight into the correctness and completeness of the classifier, while the F1-score offered a balanced measure of both. This can be visualized in Table 1.

Table 1: Comparison of Model Scores

Model	Precision	Recall	F1 Score
Basic CNN	0.75	0.79	0.75
Standardized CNN	0.84	0.87	0.85
Resnet18	0.9742	0.9801	0.9755
Resnet50	0.9930	0.9960	0.9943



Graph 1: Comparison of Models.

The model achieved a training accuracy of 100% and a test accuracy of 99%, indicating strong generalization capability despite the pose variations present in the test data. This high accuracy confirms that the combination of MTCNN and ResNet50 is effective in learning pose-invariant representations. In addition to quantitative results, visualizations were used to analyze feature separability and model interpretability. A confusion matrix (Fig.5) was plotted to illustrate the distribution of true versus predicted labels. Most diagonal elements showed strong confidence in correct predictions, while the few off-diagonal elements indicated confusion between similar-looking faces. This visualization highlighted specific identities that require more discriminative learning, potentially addressable through fine-tuning or incorporating additional contextual cues.

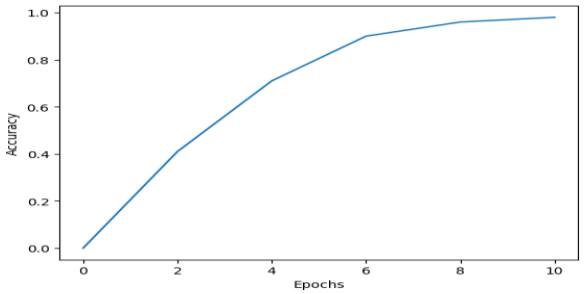


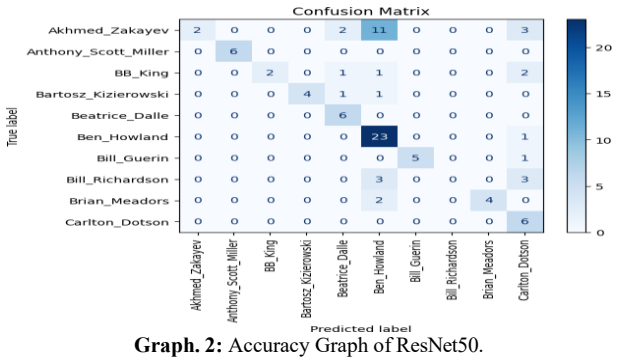
Fig. 5: The Confusion Matrix Illustrates the Performance of the Trained Face Recognition Model Across 10 Classes. Darker Shades Indicate a Higher Number of Correct Classifications. Misclassifications are Visible, especially between "Akhmed_Zakayev" And "Ben_Howland," This Suggests That There Is a Possibility of Similarity in Facial Features or pose/Lighting Distortions.

The performance was also benchmarked against a baseline CNN model without residual connections. While the baseline model showed satisfactory training performance, it suffered from overfitting and delivered a lower test accuracy of approximately 86%. This comparison demonstrated the superiority of deep residual networks in learning generalizable face embedding's, particularly in the presence of pose distortions.

Table 2: Performance Comparison of Model Architectures

Model	Accuracy
Basic CNN	79.28%
Standardized CNN	87.65%
Resnet18	98.01%
Resnet50	99.60%

In this study, we evaluated multiple deep learning architectures for face recognition, including Basic CNN, Standardized CNN, ResNet18, and ResNet50. The experimental results demonstrated that deeper networks with residual connections significantly outperform traditional CNN-based models. While the Basic CNN achieved a moderate accuracy of 79.28%, the Standardized CNN improved to 87.65%. However, ResNet architectures, particularly ResNet50, achieved the highest accuracy of 99.60%, showcasing the effectiveness of deeper models in extracting complex facial features. Furthermore, metrics such as precision, recall, and F1 score reinforced the superiority of ResNet- based models, proving their robustness in minimizing false positives and false negatives.



Graph. 2: Accuracy Graph of ResNet50.

5. Conclusion

The findings of this study highlight the advantages of using deep residual networks in face recognition applications. The ability of ResNet models to learn deeper feature representations without suffering from vanishing gradients makes them highly suitable for real-world implementations in security, authentication, and surveillance systems. Although traditional CNNs provide a reasonable performance baseline, their limitations become evident when compared to more advanced architectures. Future work could explore further optimizations, such as transfer learning with large-scale datasets, model pruning for efficiency, and integration with attention mechanisms to enhance recognition performance even further.

Fig. 6: Shows the Qualitative Analysis that was Conducted by Displaying Correctly and Incorrectly Classified Images Alongside their Ground Truth Labels. This Helped Understand Model Behavior in Edge Cases.

The findings of this study highlight the advantages of using deep residual networks in face recognition applications. The ability of ResNet models to learn deeper feature representations without suffering from vanishing gradients makes them highly suitable for real-world implementations in security, authentication, and surveillance systems. Although traditional CNNs provide a reasonable performance baseline, their limitations become evident when compared to more advanced architectures. Future work could explore further optimizations, such as transfer learning with large-scale datasets, model pruning for efficiency, and integration with attention mechanisms to enhance recognition performance even further.

Additionally, handling extreme pose variations (e.g., yaw angles > 45°) remains a critical challenge, since important facial components are lost or severely distorted. Future directions may extend to the use of 3D face models or pose-adaptive embeddings to effectively address such scenarios. Another compelling direction is recognition under combined occlusions—e.g., masks and glasses—that cover most discriminative areas. Methods such as partial feature learning, identity-aware inpainting, or multi-modal inputs could be helpful. New research, such as Naser et al. [3], also points towards the necessity for standardized benchmarks and datasets tailored to test face recognition in these real-world, more challenging scenarios, providing key guidance for future experimentation.

References

[1] Santemiz, P., Spreeuwers, L. J., & Veldhuis, R. N. J. (2024). A survey on automatic face recognition using side-view face images. IET Biometrics, 2024(1), Article 7886911. https://doi.org/10.1049/2024/7886911.

[2] Rusia, M. K., Singh, D. K., & Ansari, M. A. (2024). A novel deep transfer learning-based approach for face pose estimation. Cybernetics and Information Technologies, 24(2), 105–119. https://doi.org/10.2478/cait-2024-0018.

[3] Naser, O. A., Syed Ahmad, S. M., Samsudin, K., & Hanafi, M. (2024). Enhancing 2D face recognition systems: Addressing yaw poses and occlusions with masks, glasses, and both. Advances in Artificial Intelligence and Machine Learning, 4(3), 2545–2574. https://doi.org/10.54364/AAIML.2024.43149.

[4] Schroff, F., Kalenichenko, D., & Philbin, J. (2015). FaceNet: A unified embedding for face recognition and clustering. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 815–823. https://doi.org/10.1109/CVPR.2015.7298682.

[5] Na, I. S., Tran, C., Nguyen, D., & Dinh, S. (2020). Facial UV map completion for pose-invariant face recognition: A novel adversarial approach based on coupled attention residual UNets. Journal of Ambient Intelligence and Humanized Computing, 11(6), 2461–2472. https://doi.org/10.1186/s13673-020-00250-w.

[6] Hu, W. (2023). Improving 2D face recognition via fine-level facial depth generation and RGB-D complementary feature learning. arXiv.2305.04426arXiv.

[7] Kim, M., Su, Y., Liu, F., Jain, A., & Liu, X. (2024). KeyPoint relative position encoding for face recognition. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 244–255. https://doi.org/10.1109/CVPR52733.2024.00031.

[8] Ruan, S., Tang, C., Zhou, X., Jin, Z., Chen, S., Wen, H., Liu, H., & Tang, D. (2020). Multi-pose face recognition based on deep learning in unconstrained scenes. Applied Sciences, 10(13), 4669. https://doi.org/10.3390/app10134669.

[9] Essel, J. K., Mensah, J. A., Ocran, E., & Asiedu, L. (2024). On the search for efficient face recognition algorithm subject to multiple environmental constraints. Heliyon, 10(7), Article e28568. https://doi.org/10.1016/j.heliyon.2024.e28568.

[10] Tsai, E.-J., & Yeh, W.-C. (2021). PAM: Pose Attention Module for pose-invariantfacerecognition.arXiv.

[11] Meng, Q., Xu, X., Wang, X., Qian, Y., Qin, Y., Wang, Z., Zhao, C., Zhou, F., & Lei, Z. (2021). PoseFace: Pose-invariant features and pose-adaptive loss for face recognition. arXiv.

[12] Pitteri, G., Munaro, M., & Menegatti, E. (2017). Depth-based frontal view generation for pose invariant face recognition with consumer RGB-D sensors. In Intelligent Autonomous Systems 14 (pp. 925–937). Springer. https://doi.org/10.1007/978-3-319-48036-7_67.

[13] Cao, K., Rong, Y., Li, C., Tang, X., & Loy, C. C. (2018). Pose-robust face recognition via deep residual equivariant mapping. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018, 1085–1094.

[14] Ge, S., Li, C., Zhao, S., & Zeng, D. (2020). Occluded face recognition in the wild by identity-diversity inpainting. IEEE Transactions on Circuits and Systems for Video Technology, 30(10), 3387–3397. https://doi.org/10.1109/TCSVT.2020.2967754.

[15] Cen, F., & Qi, G.-J. (2018). Dictionary representation of deep features for occlusion-robust face recognition. IEEE Access, 6, 73925–73934.

[16] Neto, P. C., Boutros, F., Pinto, J. R., Damer, N., Sequeira, A. F., & Cardoso, J. S. (2021). Focus Face: Multi-task contrastive learning for masked face recognition. Proceedings of the 16th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2021), 1–8. https://doi.org/10.1109/FG52635.2021.9666792.

[17] Noyes, E., Davis, J. P., Petrov, N., Gray, K. L. H., & Ritchie, K. (2021). The effect of face masks and sunglasses on identity and expression recognition with super-recognizers and typical observers. Royal Society Open Science, 8(3), Article 201169. https://doi.org/10.1098/rsos.201169.

[18] Naser, O. A., Mumtazah, S., Hanafi, M., & Samsudin, K. (2024). Enhancing 2D face recognition systems: Addressing yaw poses and occlusions with masks, glasses, and both. Multimedia Tools and Applications, 83(5), 1–26. https://doi.org/10.54364/AAIML.2024.43149.

[19] Deng, J., Guo, J., Xue, N., & Zafeiriou, S. (2019). ArcFace: Additive angular margin loss for deep face recognition. Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 4690–4699. https://doi.org/10.1109/CVPR.2019.00482.

[20] Cament, L. A., Galdames, F. J., Bowyer, K. W., & Pérez, C. A. (2015). Face recognition under pose variation with local Gabor features enhanced by Active Shape and Statistical Models. Pattern Recognition, 48(11), 3371–3384. https://doi.org/10.1016/j.patcog.2015.05.017.

[21] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 770–778. https://doi.org/10.1109/CVPR.2016.90.

[22] Huang, G. B., Jain, V., & Learned-Miller, E. (2007). Labeled faces in the wild: A database for studying face recognition in unconstrained environments. Proceedings of the IEEE International Conference on Computer Vision (ICCV), 1–8.

[23] Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. (2016). Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Processing Letters, 23(10), 1499–150. https://doi.org/10.1109/LSP.2016.2603342.