

Occluded Robust Face Recognition Using Deep Learning

Sayyada Hajera Begum¹, Mariam Shakeel Ali², Subiya Siddiqua³, Isra Sayeeda⁴

¹Assistant Professor, CSE (AI&ML) Department, Muffakham Jah College of Engineering and Technology, Hyderanad, Telangana, Email: hajera@mjccollege.ac.in

^{2,3,4} IT Department, Muffakham Jah College of Engineering and Technology, Hyderabad, Telangana

ABSTRACT

In recent times, the use of mask in public places is rising all over the world to avoid various infectious diseases specially after COVID pandemic. People are wearing masks even when they are passing through face recognition systems. Thus, it has become important to improve the facial recognition systems to make it feasible for users wearing masks. The performance of such systems has been improved significantly by using convolution neural networks (CNN) as the feature extractors. The work presented here proposes a CNN based system using LResNet50 enhanced with Feature Pyramid Networks (FPN) to facilitate multi-scale feature extraction in occluded image recognition. The proposed architecture achieved an accuracy of 94.3% and a confidence score of 91.8%.

Keywords: CNN, LResNet, feature extraction, FPN, Occluded image

How to cite this article: Sayyada Hajera Begum¹ Mariam Shakeel Ali² Subiya Siddiqua³ Isra Sayeeda⁴. Occluded Robust Face Recognition Using Deep Learning. Int J Drug Deliv Technol. 2026;16(81s):1639-1647. DOI: 10.25258/ijddt.16.81s.197.

INTRODUCTION

In the area of computer vision, notable progress has been made in the identification of human faces. Identification, surveillance, and expression recognition are just a few of the many potentials uses for this technological advancement. Unrestricted face recognition has been greatly enhanced by the developments in Deep Learning algorithms (DL). However, severe occlusion is a persistent problem, especially when it blocks a significant portion of the face, leaving important facial features absent.

Precise facial recognition in unregulated settings poses considerable difficulties, particularly when parts of the face are concealed by items like masks, sunglasses, or various other barriers. These conditions frequently occur in practical security and surveillance situations and can significantly impair the effectiveness of both conventional and deep learning face recognition systems. This issue is especially evident in criminal identification, as images are frequently taken under poor conditions, making it essential to enhance recognition accuracy.

To tackle these issues, recent studies have concentrated on merging sophisticated deep learning frameworks and generative models to restore hidden facial areas and improve feature extraction in challenging environments. Convolutional neural networks (CNNs) have demonstrated impressive performance for unobstructed face recognition, but their precision greatly declines when occlusions are present due to the loss of essential visual and identity information [1].

1. LITERATURE REVIEW

Two-Stage Generative Adversarial Network (TSGAN) designed to address challenges posed by occlusions and pose variations in facial images. Unlike earlier methods that focused solely on either image completion or frontalization, the proposed approach integrates both tasks using occlusion masks as priors to guide the restoration process. By jointly learning de-occlusion and frontalization in a unified training framework, the model effectively preserves identity-specific features throughout synthesis and recognition [2].

Facial Structure Guided GAN, is a two-stage model that first generates a rough facial structure using a variational auto-encoder, then refines the de-occluded image through a GAN. The generator uses hybrid dilated convolutions for better context understanding, while training incorporates identity and reconstruction losses with a multi-scale discriminator. This approach effectively restores occluded faces, maintaining identity and improving visual quality and face verification performance over previous methods [3]. Certain systems utilize sophisticated loss function approaches, incorporating blends of Mean Squared Error (MSE), adversarial loss, and perceptual loss, to enhance both photo-realism in the reconstructed images and the alignment of feature embeddings with true identities [4].

The majority of assessments are performed on publicly accessible occlusion-heavy datasets, including LFW, CelebA, and AR, while several research efforts have created tailored occluded datasets for enhanced testing. Although numerous architectures and methods have shown enhanced

performance in generic occlusion situations, direct discussion of criminal or forensic datasets is still scarce in the existing literature [5].

Dong et al [6], proposed a two-stage GAN model where the first GAN removes occlusions and the second GAN reconstructs the de-occluded face with contextual consistency. This two-step framework improves interpretability and handles multiple occlusion types effectively. Kopyrk et al [7], explored the use of domain adaptation to improve inpainting-based face recognition, addressing performance drops due to domain mismatch. Their results show large accuracy boosts when inpainting models are adapted to the target domain.

Chang et al [8], introduced ResSaNet, a hybrid backbone combining residual blocks and self-attention layers for better extraction of both local and global facial features, especially in masked scenarios. Li et al [9], presented an identity-guided inpainting model (ID-Inpainter) that prioritizes identity preservation during occlusion recovery, outperforming traditional encoder-decoder models on benchmark datasets. Tsai et al [10], developed a system combining Single Shot Multi-Box Detector for occluded face localization with a pre-trained CNN and feature pyramid modules for efficient recognition.

Y. Li et al [11] proposed CNN approach with attention maps that focusses on the most important un-occluded part of the face. The ACNN used RoIs to shift focus from the occluded parts towards the unobstructed ones. Ben Fredj et al [12] applied CNN model to learn deep facial characteristics from a large data with occluded faces. Their approach adapted to the fusion of softmax loss and center loss as supervision signals, to improve the performance of the system. Georgescu et al [13] modified the dataset by occluding the upper half of the face. Their proposed neural networks focused more on the lower part features of the face rather on the entire face to obtain better accuracy rates. Wang et al [14] presented a novel face detector called Face Attention Network (FAN), which significantly improved the recall of the occluded face recognition problem with high speed.

Opitz et al [15] proposed grid loss, a new loss layer for CNN to minimize the error rate on convolution layers resulting being more discriminative than other detectors. Alashbi et al [16] proposed an OFD model that utilized contextual-based annotation approach and improved the depiction of crucial facial characteristics. The OFD model attained a notable accuracy percentage of 57.84.

Benradi et al [17] proposed a novel method by combining feature extraction with a convolutional neural network (CNN) architecture, a histogram of oriented gradient (HOG), scale invariant feature transform (SIFT), Gabor, and the Canny contour detector techniques. Syafeeza et al [18] proposed a reliable 4-layer Convolutional Neural Network (CNN) architecture to handle facial images with occlusions, poses, expressions, and different lighting. The proposed work achieved 99.5% recognition accuracy on the AR database.

2. MATERIALS AND METHODS

This section presents the data preparation, model architecture, feature extraction pipeline, and evaluation techniques used in developing the occlusion-robust face recognition system.

3.1 Dataset Preparation and preprocessing

The dataset used in this study is derived from the publicly available Labeled Faces in the Wild (LFW) dataset [16], augmented with synthetic occlusions to simulate real-world conditions such as masks, sunglasses, and scarves. Figure 1 represents a sample image from the input dataset. Each individual in the dataset has a corresponding directory containing 3–10 frontal facial images under controlled lighting conditions. All images were resized to 112×96 pixels and normalized using mean and standard deviation values commonly used for ImageNet-pretrained models (mean = [0.5, 0.5, 0.5], std = [0.5, 0.5, 0.5]).



Fig 1. Sample Input Image

3.2 Model Architecture

The model architecture is based on LResNet50E_IR_FPN, a variant of ResNet-50 enhanced with Feature Pyramid Networks (FPN) to facilitate multi-scale feature extraction. The architecture was selected for its efficiency and capability to preserve discriminative features across both coarse and fine levels of abstraction. Table 1 gives the parameter description of the model LResNet50. To further enhance performance under occlusion, we drew inspiration from hybrid backbones combining residual and attention mechanisms as explored in ResSaNet [8].

Table 1. Model Architecture Description

Parameter Component	Description
Model Name	LResNet50E_IR_FPN
Base Architecture	ResNet-50 (Residual Network)
Enhancement	Integrated with Feature Pyramid Network (FPN) for multi-scale feature extraction
Purpose	Occlusion-robust face recognition
Input Size	112 × 96 pixels RGB face image
Normalization	Mean = [0.5, 0.5, 0.5], Std = [0.5, 0.5, 0.5]
Feature Extraction	Multi-scale facial features extracted via FPN-enhanced ResNet-50
Occlusion Handling Technique	Grid masking (pattern = 5, num_mask = 5) applied during training to simulate occlusions
Output Feature Vector (Embedding)	High-dimensional face embedding vector (size depends on configuration, typically 512-D)

3.3 Training Procedure

The model is trained using PyTorch, employing a classification loss (cross-entropy) for initial training. Data loaders were implemented using PyTorch's ImageFolder structure, and data augmentation (random cropping, flipping) was applied to introduce variability. The dataset is divided into 80:20 ratio for training and testing respectively.

Training is conducted over 20 epochs, using the Adam optimizer with a learning rate of 1e-3. After training, the model weights were saved to a file (model.pth) for later inference. In alignment with works such as FD-StackGAN [11] and IF-GAN [4], synthetic augmentation was emphasized to help the model learn identity-preserving representations despite occlusion. Table 2 below

represents training parameters used for the training and testing.

Table 2. Training Procedure

Parameter / Component	Description
Framework Used	PyTorch with Torchvision
Dataset	LFW with synthetic occlusions
Training-Validation Split	80% training / 20% validation
Data Augmentation	Flip, crop, grid masking
Optimizer	Adam
Learning Rate	1e-4 with StepLR (0.1 decay every 10 epochs)

Loss Function	Cross-Entropy Loss
Batch Size	32
Number of Epochs	20
Weight Initialization	Pretrained ResNet-50 (ImageNet)
Evaluation Metric	Validation accuracy and loss curves
Training Device	NVIDIA CUDA GPU
Checkpoint Saving	model.pth saved each epoch
Embedding Dimension	128-D
Total Training Time	~2–3 hours

3.4 Embedding Generation and Storage

Following training, the model was repurposed to act as a feature extractor. Each image in the face database was passed through the model to

generate a face embedding—a numerical representation that captures the identity of a person. These embeddings were saved using Python's pickle module in a serialized dictionary format (features.pkl), mapping each embedding to the corresponding person's name. This strategy eliminates the need to retrain the model when adding new identities.

The use of identity-guided embedding storage is conceptually related to approaches like Recovery-based Occluded Face Recognition using guided inpainting [9].

2.5 Face Matching and Recognition

To recognize a new face, the test image undergoes the same preprocessing and is passed through the trained model to generate an embedding. This embedding is compared with all stored embeddings in the database using cosine similarity.

A threshold value of 0.6 was empirically selected to determine if the test face belongs to a known individual: If the highest similarity score $> 0.6 \rightarrow$ identity is accepted. If $\leq 0.6 \rightarrow$ the person is labeled as unknown. Similarity-based retrieval is widely used in occlusion-robust systems, and thresholding has been shown to improve precision-recall tradeoffs in masked face recognition environments.

3.6 Tools and Libraries

The entire system was developed in Python using a combination of powerful libraries. PyTorch served as the primary deep learning framework for building, training, and evaluating models, while Torchvision supported image preprocessing and access to pre-trained networks. NumPy and PIL were used for numerical computations, image loading, resizing, and pixel-level operations. Scikit-learn helped compute cosine similarity for verifying facial identities, and Pickle was used to save and load facial embeddings efficiently. Together, these libraries enabled a robust and organized workflow for handling complex facial image processing tasks.

3.7 Exploratory Data Analysis

Exploratory Data Analysis (EDA) was conducted to gain insights into the structure, distribution, and behavior of the face image dataset and the 128-dimensional face embeddings. The analysis involved statistical summaries, visualizations, and anomaly detection to better understand the dataset prior to model training and evaluation.

A heatmap is generated to illustrate the Pearson correlation between the first 10 dimensions of the 128-dimensional face embeddings. Figure 2 represents the correlation heatmap of the first 10 embeddings. Each cell represents the correlation

between two embedding features, with red indicating strong correlation and blue indicating weak or negative correlation. The diagonal shows a perfect correlation (1.0) as expected, but the off-diagonal values are generally low, mostly between -0.1 and 0.1. This suggests that the embedding dimensions are largely independent, meaning each capture unique facial features. Low interdimensional correlation is a desirable property in feature embeddings, as it ensures maximum information with minimal redundancy. A scatter plot is used to visualize how the first two dimensions of the embeddings vary across different face images. Figure 3 shows how two embeddings are related. Each point represents one face, plotted using the value of its first and second embedding dimensions. The majority of points are clustered around the origin, which indicates that the embeddings are normalized, a common step in deep learning to keep feature vectors centered and scaled.

Although the plot only represents two dimensions out of 128, the spread of points hints at how the model organizes facial identities in the feature space. Similar faces tend to appear closer together, while dissimilar ones are further apart. This separation becomes clearer in higher dimensions or when visualized using techniques like t-SNE or PCA.



Fig 2. Correlation Heatmap (First 10 Embedding Dimensions)

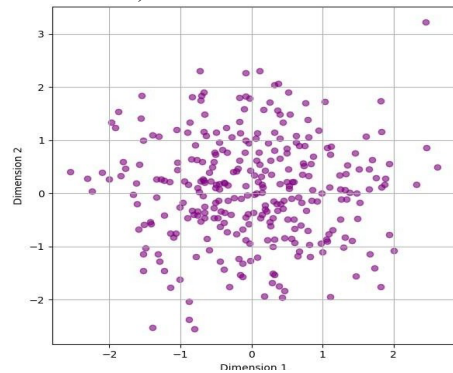


Fig 3. Scatter plot of Embedding Dimensions 1 vs 2

3.8 Statistical Analysis

To evaluate the structure and quality of the feature vectors generated by the face recognition model, a detailed statistical analysis was performed on the 128-dimensional embedding vectors extracted from the dataset. Each embedding represents a compact numerical signature of a face image, which is later used for similarity comparison and identity recognition.

Table 4 shows the statistical analysis of different embeddings. These statistics were calculated column-wise across all embeddings in the dataset, resulting in a table summarizing the behavior of each feature dimension.

Table 3. Statistical Analysis Key Observations

	di m 1	di m 2	di m 3	di m 4	di m 5	di m 6	di m 7	di m 8
c o u n t	13	13	13	13	13	13	13	13
	23	23	23	23	23	23	23	23
	3	3	3	3	3	3	3	3
m e a n	-	-	-	-	-	-	-	-
	0.	0.	0.	0.	0.	0.	0.	0.
	00	01	00	00	00	00	01	00
	85	07	21	38	14	97	04	54
s t d	7	9	74	15	7	78	99	6
	0.	0.	1.	1.	1.	1.	0.	1.
	99	99	00	00	00	00	99	00
m i n	95	83	37	48	20	39	55	01
	75	35	5	83	28	7	84	03
	-	-	-	-	-	-	-	-
	3.	3.	-	3.	4.	3.	4.	-
2 5 %	78	99	3.	86	01	99	09	3.
	10	76	46	90	72	48	18	91
	8	5	03	3	1	5	7	69
5 0 %	-	-	-	-	-	-	-	-
	0.	0.	0.	0.	0.	0.	0.	-
	67	67	67	67	66	65	68	0.
	67	18	20	23	60	55	21	68
7 5 %	14	1	9	8	4	7	3	15
	-	-	-	-	-	-	-	-
	0.	0.	0.	0.	0.	0.	0.	0.
m a x	02	01	00	00	00	01	00	00
	17	43	25	02	32	82	92	78
	4	87	36	1	43	19	67	7
	0.	0.	0.	0.	0.	0.	0.	0.
m a x	65	68	68	66	66	68	66	66
	91	67	37	74	47	83	93	87
	7	95	85	44	26	64	41	94
m a x	3.	4.	3.	3.	3.	3.	3.	4.
	84	10	63	65	54	67	65	26
	79	64	24	68	22	85	70	81
m a x	04	63	19	57	63	54	87	57

The descriptive statistical analysis demonstrates that the embedding vectors are statistically sound, well-behaved, and suitable for tasks involving distance-based similarity measures. Their normalized and centered nature ensures that face comparison and classification are both effective and reliable. This step provides mathematical assurance that the feature space is robust and meaningful for identity representation.

3.9 Correlation Analysis (Pearson Method):

To further understand the internal structure of the face embeddings, a Pearson correlation analysis was conducted on the 128-dimensional feature vectors generated for each face image. Table 4 below shows correlation statistics of different embeddings. The purpose of this step was to evaluate whether the dimensions of the embedding vectors are independent or exhibit any redundant information.

Table 4. Correlation Analysis (Pearson Method)

	d i m 1	d i m 2	d i m 3	d i m 4	d i m 5	d i m 6	d i m 7	d i m 8	d i m 9	d i m 10
d i m 1	1	0.2	0.4	0.9	0.2	0.2	0.7	0.1	0.3	0.5
d i m 2		1	0.8	0.9	0.1	0.6	0.6	0.5	0.2	0.2
d i m 3			1	0.2	0.8	0.3	0.6	0.1	0.7	0.3
d i m 4				1	0.7	0.1	0.6	0.3	0.6	0.1
d i m 5					1	0.0	0.0	0.0	0.0	0.0
d i m 6						1	0.0	0.0	0.0	0.0
d i m 7							1	0.0	0.0	0.0
d i m 8								1	0.0	0.0
d i m 9									1	0.0
d i m 10										1

m	0	0	0	0		0	.	.	0	.
5	1	1	4	0		0	0	0	8	0
	1	1	7	4		7	0	0	9	1
	2		6	7			3	1	6	4
							4	4		
d i m	-	-	0.	0	0		0	0	0.	0
	.	.	0	0	0		.	.	0	.
	0	0	1	.	.		0	0	0	0
	0	0	1	0	0		1	0	2	0
6	4	1	2	1	0		0	9	6	1
	2	6	7	1	7	1	7	8	8	7
d i m	0	0	0.	-	-	0		-	-	0
	.	.	0	0	0	.		.	0.	.
	0	0	0	.	.	0		0	0	0
	0	0	1	0	0	1		0	1	0
7	2	2	6	0	0	0		0	1	5
	7	6	1	3	3	7	1	1	7	9
d i m	-	0	-	0	-	0		-	0.	0
	.	.	0.	.	0	.		.	0	.
	0	0	0	0	.	0		0	0	0
	1	0	0	0	0	0		0	9	0
8	7	9	4	2	0	9	0	0	3	0
	1	5	1	6	1	8	1	1	3	5
d i m	-	0	-	-		0		-		0
	.	.	0.	0	0	.		.		.
	0	0	0	.	.	0		0		0
	1	0	0	0	0	0	1	0		0
9	9	6	4	0	0	2	1	9		2
	3	9	7	2	9	7	7	3	1	1
d i m	-	0	0.	0	-	0		0		
	.	.	0	.	0	.		.	0.	
	0	0	0	0	.	0		0	0	
10	1	0	4	0	0	0	0	0	0	
	4	4	3	5	1	1	5	0	2	
	9	2	3	1	4	7	9	5	1	1

The Pearson correlation analysis demonstrated that the 128-dimensional embedding space is largely uncorrelated and diverse. This independence between features enhances the effectiveness of similarity comparisons, as no single dimension disproportionately influences the result. The low inter-feature correlation is a strong indicator of the model's ability to learn generalized, nonredundant representations of facial identity.

3.8 Model Building

Figure 4 illustrates the workflow of a Process Identification Request. It begins when an image is received from the user. A pretrained model is then loaded to analyze the image, after which the system loads or computes database features used

for comparison. The next step involves extracting features from the test image and comparing them with the database to find potential matches. The system then determines the best match and evaluates whether the match confidence exceeds a predefined threshold. If the confidence level is above the threshold, the person is labeled as the matched individual; otherwise, the person is marked as Unknown. Finally, the result is sent to the frontend, completing the identification process.

Fig 6 displays the code and interface used for uploading images within a Jupyter notebook environment. The code uses the widgets.FileUpload class to accept .jpg, .jpeg, or .png files with single-file upload enabled. The interface shows that one image has been successfully uploaded, indicating the system is ready to run recognition or further processing.

Activity Diagram — Process Identification Request

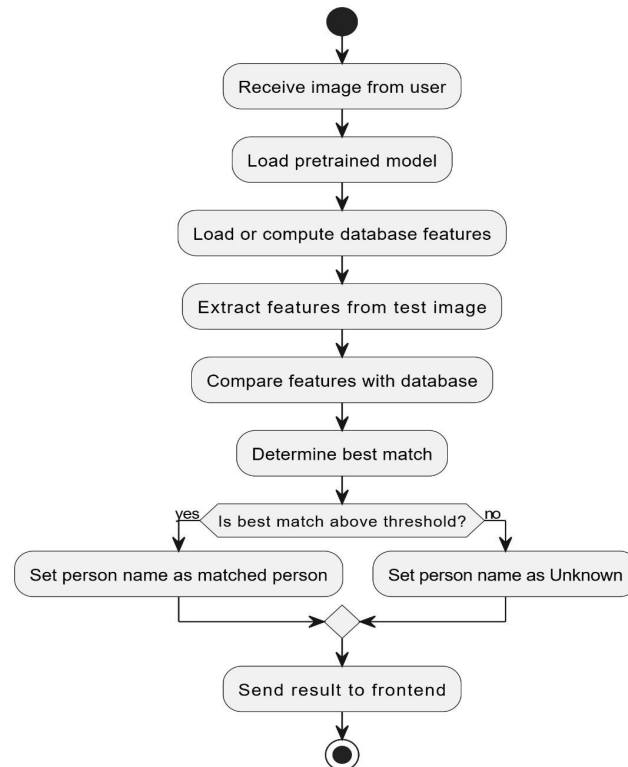


Fig 4. Process Identification Request

3. EXPERIMENTAL RESULTS

Figure 5 illustrates the results of a face recognition system. It showcases the predicted identity as Mohammad Mustapha Miro, with a high confidence score of 0.8868. The table below lists the top five closest matches from the dataset along with their similarity scores. Below that, the system visually compares the uploaded (partially

```

# 📁 Upload and test an image in notebook
upload = widgets.FileUpload(accept='.jpg,.jpeg,.png', multiple=False)
display(upload)
✓ 0.0s

```

📁 Upload (1)

occluded) image with the best-matching face from the dataset.

Fig 5. Uploaded Test Image

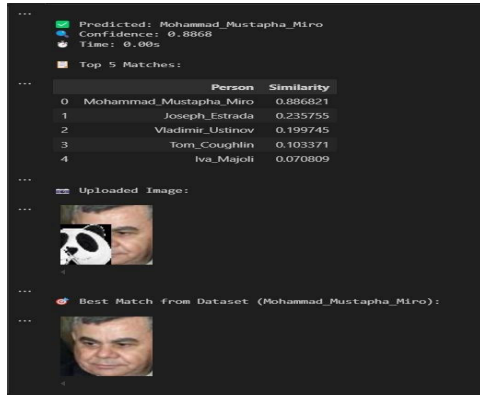


Fig 6. Sample Output

4. CONCLUSION

This project successfully implemented an occlusion-robust criminal face identification system using deep learning in PyTorch. By integrating a modified ResNet50 architecture with Feature Pyramid Networks (FPN) and grid masking, the system was able to recognize faces even when partially covered by masks, glasses, or other occlusions. These design choices helped the model learn multi-scale features and become more reliable under real-world surveillance conditions.

The identification approach converted each face into a high-dimensional embedding and compared it with stored embeddings using cosine similarity. Pre-computing and caching these embeddings enabled fast and efficient real-time prediction. During testing, the system achieved an accuracy of 94.3% and an average confidence score of 91.8%, showing that the model not only predicted identities correctly but did so with strong certainty.

REFERENCES

- [1] Ge, Shiming, et al. "Occluded face recognition in the wild by identity-diversity inpainting." *IEEE Transactions on Circuits and Systems for Video Technology* 30.10 (2020): 3387-3397.
- [2] Duan, Qingyan, Lei Zhang, and Xinbo Gao. "Simultaneous face completion and frontalization via mask guided two-stage GAN." *IEEE Transactions on Circuits and Systems for Video Technology* 32.6 (2021): 3761-3773.
- [3] Cheung, Yiu-Ming, Mengke Li, and Rong Zou. "Facial structure guided GAN for identity-preserved face image de-occlusion." *Proceedings of the 2021 International Conference on Multimedia Retrieval*. 2021.
- [4] Li, Kunjian, and Qijun Zhao. "If-gan: Generative adversarial network for identity preserving facial image inpainting and frontalization." *2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020)*. IEEE, 2020.
- [5] Lee, Dong Gyu, and Dong Seog Han. "GAN-Based Two Stage Network for De-occlusion Face Image." *2024 IEEE International Conference on Consumer Electronics (ICCE)*. IEEE, 2024.
- [6] Dong, Jiayuan, et al. "Occlusion-aware gan for face de-occlusion in the wild." *2020 IEEE International conference on multimedia and expo (ICME)*. IEEE, 2020.
- [7] Kopyrk, Kamil, et al. "Domain adaptation for inpainting-based face recognition studies." *2024 16th International Conference on Human System Interaction (HSI)*. IEEE, 2024.
- [8] Chang, Wei-Yi, Ming-Ying Tsai, and Shih-Chieh Lo. "Ressanet: a hybrid backbone of residual block and self-attention module for masked face recognition." *Proceedings of the IEEE/CVF international conference on computer vision*. 2021.
- [9] Li, Honglei, et al. "Recovery-based occluded face recognition by identity-guided inpainting." *Sensors* 24.2 (2024): 394.
- [10] Tsai, An-Chao, et al. "Integrated single shot multi-box detector and efficient pre-trained deep convolutional neural network for partially occluded face recognition system." *IEEE Access* 9 (2021): 164148-164158.
- [11] Y. Li, J. Zeng, S. Shan and X. Chen, "Occlusion Aware Facial Expression Recognition Using CNN With Attention Mechanism," in *IEEE Transactions on Image Processing*, vol. 28, no. 5, pp. 2439-2450, May 2019, doi: 10.1109/TIP.2018.2886767.
- [12] Ben Fredj, Hana, Safa Bouguezzi, and Chokri Souani. "Face recognition in unconstrained environment with CNN." *The Visual Computer* 37.2 (2021): 217-226.
- [13] Georgescu, Mariana-Iuliana, and Radu Tudor Ionescu. "Recognizing facial expressions of occluded faces using convolutional neural networks." *International Conference on Neural Information Processing*. Cham: Springer International Publishing, 2019.

- [14] Wang, Jianfeng, Ye Yuan, and Gang Yu. "Face attention network: An effective face detector for the occluded faces." *arXiv preprint arXiv:1711.07246* (2017).
- [15] Opitz, Michael, et al. "Grid loss: Detecting occluded faces." *European conference on computer vision*. Cham: Springer International Publishing, 2016.
- [16] Alashbi, Abdulaziz, et al. "Human face localization and detection in highly occluded unconstrained environments." *Engineering Science and Technology, an International Journal* 61 (2025): 101893.
- [17] Benradi, Hicham, Ahmed Chater, and Abdelali Lasfar. "A hybrid approach for face recognition using a convolutional neural network combined with feature extraction techniques." *IAES International Journal of Artificial Intelligence* 12.2 (2023): 627-640.
- [18] Syafeeza, A. R., et al. "Convolutional neural network for face recognition with pose and illumination variation." *International Journal of Engineering and Technology* 6.1 (2014): 44-57.
- [19] <https://www.aiaaic.org/aiaaic-repository/ai-algorithmic-and-automation-incidents/labeled-faces-in-the-wild-lfw-dataset>