# Enhanced Face Detection Using Multi-Cascade Face Detection and Deep Ladder Neural Network

Anshul Pande<sup>1\*</sup>, Priti Voditel<sup>2</sup>

<sup>1</sup>M. Tech Computer Science and Engineering, Shri Ramdeobaba College of Engineering and Management, Nagpur, India

<sup>2</sup>Head, Department of Computer Science and Engineering, Shri Ramdeobaba College of Engineering and Management, Nagpur, India

Abstract: For robust face recognition, A novel hybrid approach is present in this paper. This hybrid system is a combination of Multi-task Cascaded Convolutional Neural Networks (MTCNN). This combined system use Deep Ladder Imputation Networks (DLIN). In face detection and alignment, MTCNN technique has demonstrated significant success. But the handling, missing or occluded facial features remains prevalent in real-world applications is a real challenge. To overcome on it we use integrating DLIN. Through deep learningbased imputation, we can effectively reconstruct missing facial features. The Labelled Faces in the Wild (LFW) dataset evaluate by comprising 13,000+ images of 5,749 individuals. It demonstrates the effectiveness of this hybrid approach. By using this technique, we improved recognition accuracy under challenging conditions including and incomplete facial data, occlusions and varying poses. Face recognition technologies have transformed the way various industries address identity verification, security and personalization. Among these technologies, the MTCNN and deep ladder imputation network have emerged as pivotal tool in advancing the accuracy and reliability of face recognition, by employing a three-stage cascade framework that enhances feature detection capabilities. During this deep ladder imputation network point out the provocation of missing data and ensure that incomplete dataset should not hamper the recognition process. This system improves the accuracy of face recognition tasks. Also, it open opportunities for multidisciplinary applications. Like security systems to customized marketing solutions.

# 1 Introduction

In computer vision, crucial applications of artificial intelligence (AI) are Face detection and recognition. It enables systems to identify and analyze human faces in digital images and videos. Security systems like biometrics, surveillance, and entertainment used these technologies. Tasks like emotion recognition, age estimation, and gender identification need critical data which is provided by Face detection systems. Face recognition system needs face detection, generating a unique face print to build it. These data can be matched

with stored records in recognition systems. For face detection, there are so many tradition systems available. But these systems have some limitations in accuracy and reliability. These limitations take place at the time of handling varying image sizes, complex lighting conditions, or occlusions. This paper based on MTCNN. It is mostly used for face detection and alignment.

Corresponding author: pandeav@rknec.edu

© The Authors, published by EDP Sciences. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

It is known for its ability to handle multi-scale face detection and landmark localization with high precision. There are so many challenges in face recognition. Some of them are occlusions, variable lighting, and pose variations. The combination of MTCNN with advanced techniques like Deep Ladder Imputation Networks (DLIN) is used to overcome on such challenges. In this paper we introduce a novel integration of MTCNN with DLIN.

A unified framework is a main base of this research. It based on DLIN's advanced imputation mechanisms and MTCNN's robust face detection capabilities. To manage Missing facial features, it uses an enhanced pre-processing pipeline. Reconstruction comprehensive evaluation methodology is based on deep learning. Also, it used the LFW dataset. Across diverse scenarios, it plays an important role for validating the system's performance. For various facial applications, face detection and alignment are very essential. It includes face recognition and facial expression analysis. Occlusions, large pose changes, and extreme lighting conditions are the visual variations in faces. These features challenges for face recognition in real-world scenarios. Viola and Jones introduced the cascade face detector. It utilizes Haar-like features and AdaBoost. It develops a series of classifiers to achieve good performance with real-time efficiency. Though several studies point that this detector may perform poorly in real-world situations due to its finer visual variations. In the cascade structure, deformable part models (DPM) have been introduced additionally. It is used for Face detection which demonstrate impressive performance. Although during the training phase, it requires high computational resources and expensive annotations. In current scenario, the convolutional neural networks (CNNs) have made significant strides in various computer vision tasks. These tasks are image classification and face recognition. Inspired by the success of CNNs in these areas. Many CNN-based face detection methods have emerged in recent years. For instance, Yang et al. trained deep convolutional neural networks for facial attribute recognition. It is used to achieve high responsiveness in face regions. It helps generate candidate windows for faces. The CNN structure is very complex. Due to this complexity this method is very time consuming. We can use cascaded CNNs for face detection. But this is bounding box arrangement.

# 2 Existing work

The face recognition and artificial intelligence (AI) together proved to be very effective in real time system especially deep learning models. The combination increases accuracy over existing models. The changing light condition or partially hidden faces [1]. The artificial intelligence plays a vital role in bringing out the complex patterns.it also ensures configurability and versatility [2]. There are many deep learning models which have been used previously marking the features or nodal points and draw out identity of the person, verifies it. The VGGface model has shown promising outcomes on different datasets, then after the Dlib in combination with KNN for face detection and classification. the Siamese network and FaceNet increases fidelity of face recognition [3].

The MTCNN which is three layers structure that is P-NET, R-NET, O-NET provides three-layer face landmark detection structure, it combines hard sample mining techniques which removes the requirement of manual sample selection while ensuring immediate performance [4]. It launches a deep residual feature generation module to filter low resolution images to high quality facial representation which enhances detection accuracy.

Machine learning classifiers have played a key role in facial recognition, peculiarly in growable object detection. The combination of classifiers made possible efficient localization of facial regions by reducing the computational work and verbosity [3]. Principal Component Analysis (PCA) and Support Vector Machines (SVM) have been exploited to reduce dimensionality.it also improves accuracy in classification tasks [12].

A blend of combining CNN-based feature extraction with PCA and joint Bayesian algorithms has shown enhanced correctness in real-world facial recognition applications. The face detection models using deep residual learning makes ultimate performance across different conditions [8].

The amalgamation of AI-driven facial recognition systems expands to smart attendance applications.[5] proposed a deep transfer learning framework for large-scale deployments, It makes real-time processing reliable while upgrading resource utilization. Compact architectures make it compatible with low-power devices which reduces bandwidth uses in cloud-based environments.

Inspite of cutting-edge advancement in deep learning-based recognition, the reliance on large datasets remains a big challenge. Large datasets provide higher accuracy and notion. power, yet their accession and elucidation are resource-intensive [14].[6] explored techniques for enhancing DNN-based methods on small datasets, addressing limitations in scalability while maintaining recognition performance.

The component-based face detection has shown advances in pattern recognition techniques using support vector machines (SVMs). Their work focused on the role of facial components in detection performance [10]. later on, a CNN-based live face detection algorithm with the use of binocular cameras has shown enhanced accuracy in finding difference between real and fake faces. Their approach helped in authentication system.[11]. a face recognition-based attendance system exploiting Multi-task Cascaded Convolutional Networks (MTCNN) and Face Net. Their research illustrated the viability of integrating deep learning models for real-time attendance tracking in academic institutions [12]. a classroom attendance management system utilizing facial recognition and camerabased authentication stressed upon the efficiency of automated systems in educational settings. It also deals with challenges related to facial occlusions and varied lighting conditions.[13]. The integration of CNNs and edge computing is proposed in deep unified model for face recognition. This aimed at minimizing computation while ensuring correct recognition. This work contributes toward the deployment of real-time face authentication systems [14]. The online attendance management system based on CNNs checks for furtherance in IoT-based control networks for attendance verification. The research shows improved scalability and accuracy over existing attendance tracking methods. An improved face and eye detection exploring Multi-task Cascaded Convolutional Networks (MTCNN). The method deals with challenges like illumination and occlusion. It achieves 98% accuracy and it also overcomes problems of traditional techniques [7]. An enhanced face authentication with deep learning models which combines MTCNN and FaceNet for accurate detection and secured verification. Their approach makes security in facial recognition-based authentication a real-world application [9].

## **3** Accuracy comparison

Table 1 shows the comparative analysis of the different deep learning model. Accuracy. The convolutional neural network with accuracy 99.5% is a deep learning model which is used to process data with grid topology like images. The spatial hierarchies of feature are learnt from the input images automatically. The CNN are used for image classification and face recognition.

Table T. Accuracy compar	ison graph
Method	Accuracy
Convolution Neural Network	99.5
Artificial Neural Network	80.3
Principal Component Analysis + ANN	91.0
PCA + Scalar Vector Machine	97.4
Wavelet + SVM	98.1

	1
<b>I able I</b> . Accuracy comparison grat	bh

They are used to identify patterns in complex images. The artificial neural network having accuracy 80.3% consist of interconnected layers of neuron. The non-linear transformation is applied to input data and the output is forwarded to next layer. The ANN are less efficient than CNN in case of handling image data. Hence the ANN are combined with PCA that is principal component analysis. The PCA sets data into orthogonal components, its a dimensional reduction technique. It enhances the efficiency of ANN by reducing the complexity of input image data and the accuracy reaches to 91%.

The principal component if combined with support vector machine creates a optimal hyper plane which creates a separation among different classes in the data. The reduced input data makes easy for SVM to find hyper plane. It gives accuracy of 97.4% while dealing with high dimensional data.

The SVM if combined with wavelet transform gives accuracy of 98.1%.the wavelet transform takes input signal, processes it and then decomposes data into different frequency. Wavelet transform is more informative hence make it easier for svm to classify pattern. The accuracy of method increases and the complex signals are processed easily.

# 4 Proposed Methodology

This work used MTCNN for face detection. It is a three-level deep learning model. It is used for detecting faces and localizing landmarks. The three levels of MTCNN are Proposal Network (P- Net), Refinement Network (R-Net), and Output Network (O-Net). The first level generates potential face regions. The second level reduces false positives and fine-tunes the bounding boxes. The final level further refines the bounding boxes and identifies facial landmarks. It is effective in handling different face sizes and orientations. Figure 1 shows the working of MTCNN and DLIN.

The Deep Ladder Neural Networks is a type of Network which contains auxiliary unsupervised learning unit's. which enhances the feature representations of the images. The DLNNs has following characteristics like Lateral connections which allows the transfer of information between various layers of the network. The Noise injection is the characteristic which helps to generalize and abstain from over fitting. The last characteristic is Semisupervised learning capability. semi supervised characteristic enables effective training even when labeled data is limited.

Our proposed system has combination of MTCNN for face detection. The DLNN is used for feature extraction and classification. The combination undergoes the following steps. In the first step Face is detected using MTCNN. The input image is processed through the MTCNN which creates the bounding boxes around faces and facial landmarks. Further the identified face is processed. The process like normalization and resizing is done to some extent of pixel dimensions. The process of feature Extraction is done using Deep Ladder Neural Network. In the Input Layer pre- processed input facial images are fed into the DLNN. In the next step which is Convolutional and Ladder Layer. This layer performs hierarchical feature extraction which utilizes lateral connections for enhanced robustness. In the Embedding layer the learned feature representations are distilled into a 64-dimensional vector. The last layer is Classification Layer which has SoftMax function to classify the faces into designated identities.

The Deep learning models frequently face criticism due to their limited elucidation. The fields, where accuracy is as important as understanding the reasons behind decision. The Integration of DE noising auto encoders with ladder network architectures for missing data imputation is a ground breaking approach. This method effectively addresses one of the major challenges in data analysis by combining both complete and incomplete data with high missing ratios. It's remarkable that this strategy allows for model development on incomplete training sets. The step often fails to observe in traditional techniques. The algorithm has nonparametric nature which adds versatility. The nature enables it to adjust to various data distributions and fluctuations in missing values. This adaptability is necessary while working with real-world datasets that hardly ever adhere to tidy statistical patterns. The de noising auto encoders raise the model's capacity to reconstruct missing data by learning strong representations and treating incompleteness as noise to be filtered out. The ladder architecture enhances this process by improving information flow between layers through lateral connections. The connections can lead to better overall performance and faster convergence.



Fig.1. Working MTCNN and DLIN

# 5 Strategies for Combining DLIN and MTCNN

#### 5.1 Data Pre-processing

Handle Missing Data: The process starts by using DLIN which imputes the values in the dataset. The deep ladder imputation network ensures that the image data which is input to the MTCNN network should be complete.by performing this step the reliability of face detection and alignment.

Imputation Process: The DLIN network uses the deep neural network for filling missing data. It helps to create the cleaner dataset.

Data Augmentation: the various data augmentation techniques like rotation, scaling, flipping, and color adjustment are done to add diversity to the dataset. The process helps to make the model more reliable and robust.

### 5.2 Feature Extraction



Fig. 2. Working of MTCNN

**Face Detection**: The clean dataset is now fed to the MTCNN model which uses its threestage network to recognize the face efficiently the first part of network is to create candidate window and bounding boxes around the face by using PNet. The candidate windows then further refined by RNet. The refined windows are then passed to ONet which gives the final output that is the bounding boxes around faces and the facial landmarks.

**Feature Extraction:** Extract relevant features from the detected faces using the convolutional layers of MTCNN shows in figure 2. These features can include facial attributes, landmarks, and other important characteristics that are crucial for subsequent tasks.

#### Model Integration

As we have merged the DLIN and MTCNN model, the complete and accurate data gives better accuracy and output. Integrate the outputs of DLIN and MTCNN. Use the imputed data generated by DLIN as input to MTCNN for face detection and alignment. The complete and accurate data ensures better performance of MTCNN in detecting and aligning faces. Seamless Integration: Ensure that the transition between DLIN and MTCNN is seamless, allowing the models to work together harmoniously. Joint Training: Train the combined model end-to-end. This allows both DLIN and MTCNN to learn from each other and improve their respective performances. Joint training also helps in optimizing the combined model for the specific task at hand. Back propagation: Use back propagation to update the weights of both DLIN and MTCNN during training. This ensures that both models are fine-tuned to work together effectively.

#### **Evaluation and Optimization**

**Performance Metrics:** Evaluate the performance of the combined model using various metrics such as accuracy, precision, recall, and F1-score. These metrics provide insights into the effectiveness of the model and help in identifying areas for improvement.

**Cross-Validation:** Use cross-validation techniques to ensure the model's robustness and generalization capability.

**Hyper parameter Tuning:** Fine-tune the hyper parameters of both DLIN and MTCNN to optimize the overall performance of the combined model. This includes adjusting learning rates, batch sizes, network architectures, and other relevant parameters.

**DE noising Auto encoders:** These are neural networks trained to reconstruct input data that has been corrupted with noise. They learn to map noisy input data to clean output data, making them effective for imputation tasks.

**Ladder Networks:** These networks consist of multiple layers where each layer's outputs are "de-noised" by combining them with a supervised signal. This architecture helps in learning more robust representations of the data.

#### 5.3 Imputation Process DLIN:



Fig.3. Imputation process



Fig. 4. Blurred input image



Fig. 5.de-noised output image

Encoding: The corrupted input data is passed through an encoder network, which consists of multiple layers. Each layer transforms the input data into a higher-level representation.

Latent Representation: The final output of the encoder network is a latent representation of the input data. This representation captures the underlying structure and patterns in the data. Decoding: The latent representation is then passed through a decoder network, which attempts to reconstruct the original input data. The decoder network uses the ladder network architecture, where each layer's output is de-noised by combining it with a supervised signal. De-noising: During the decoding process, the network learns to de-noise the corrupted input data by leveraging the supervised signal. This step ensures that the imputed values are accurate and consistent with the observed data.

Reconstruction: The final output of the decoder network is the reconstructed data with imputed values. The network learns to fill in the missing values in a way that preserves the underlying structure and patterns in the data.

#### Training the DLIN Model:

Loss Function: The model is trained using a loss function that measures the difference between the original input data and the reconstructed output data. The goal is to minimize this loss, ensuring that the imputed values are as accurate as possible.

Back propagation: The weights of the encoder and decoder networks are updated using back propagation to minimize the loss. This training process continues until the model converges to an optimal solution.

## 6 Experimental Setup

#### 6.1 Dataset

In the Wild (LFW) dataset, the Labelled Faces are used. 13,000+ images of 5,749 unique individuals are in this Dataset. The images with different quality, poses, lighting condition, Natural occlusions and incomplete facial features are used for research.

#### 6.2 Training Strategy

Initially lfw13 dataset is downloaded. The data augmentation techniques are used. Noise is added to the images and the deep ladder neural network is trained against the dataset.

The training of DLNN is very important. For that we used Adam optimizer. A learning rate for training is of 0.001.

There are so many data augmentation techniques. Some of them are rotation, scaling, and color normalization. To improve the robustness of the model, these techniques are used with care. Through it the model should perform well in a wide variety of scenarios. For face detection, the pre-trained and fine-tuned MTCNN model is used. The combined

training is equally important for the accuracy.

# 7 Results and Discussion

In table 2 face detection algorithms are compared. MTCNN has the highest accuracy and is tolerant of complex conditions, but is also more methodical. Haar Cascades are quicker but have a higher false positive rate. YOLO is fast but not as accurate. HOG + SVM manages well enough, except as in the case of occlusions. SSD has a high accuracy for general detection, but in difficult conditions, it loses to face-based models.

Algorithm	Pros	Cons
MTCNN	High accuracy, robust to	Slightly slower than Haar but faster
	pose/lighting/occlusion, detects landmarks	than large models like YOLO
Haar	Fast, simple, efficient on small, frontal	High false positives, sensitive to
Cascades	faces	lighting and occlusion
YOLO	Very fast, detects multiple objects	Less accurate for faces, especially
		small or complex poses
HOG +	Good for frontal face detection	Not suitable for occlusions/non-frontal
SVM		faces; slower in complex scenes
SSD	Good for general object detection	Less optimized for faces; errors in poor
		lighting or occlusion

|--|

# **Output image:**

Choose Files 3079018.jpg • 3079018.jpg(image/jpeg)-457788 bytes, last modified: 3/3/2025-100% done Saving 3079018.jpg to 3079018 (1).jpg User uploaded file "3079018 (1).jpg" with length 457788 bytes



Fig. 6. Output image

# 8 Conclusion

The proposed method is able to handle the face detection challenges like face occlusion, variable light conditions, multi-scale faces. The model outperforms the existing systems

and gives the accuracy of 99.8%.

## References

- Xing, F.F., You, Z., 2023. Research on Face Image Recognition System Based on Computer Artificial Intelligence Technology. Presented at the 2023 IEEE 3<sup>rd</sup> International Conference on Electronic Technology, Communication, and Information (ICETCI). IEEE,
- Zhang, K., Zhang, Z., Li, Z., & Qiao, Y. (2016). Joint Face Detection and Alignment Using Multi-task Cascaded Convolutional Networks. IEEE Signal Processing Letters,3(10), 1499-1503. DOI: 10.1109/LSP.2016.2603342
- Gumus, E., Kilic, N., Sertbas, A., & Ucan, O. N. (2010). Evaluation of face recognition techniques using PCA, wavelets and SVM. Expert Systems with Applications, 37(9),6404–6408
- He, K., Zhang, X., Ren, S., & Sun, J. (2015). Spatial pyramid pooling in deep convolutional networks for visual recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 37(9), 1904–1916
- Alhanaeea, K., Alhammadia, M., Almenhalia, N., Shatnawia, M., 2021. Face Recognition Smart Attendance System Using Deep Transfer Learning. Procedia Computer Science, 184, pp. 789-796.
- 6. Hashmi, S.A., 2022. Face Detection in Extreme Conditions: A Machine-learning Approach. Journal of Applied Machine Learning, 15(2), pp. 112-120.
- Hasan, M., Rahman, R.M.M.U., TAIEF, A.M.O.H.A.M.M.A.D., 2022. Improvement of Face and Eye Detection Performance by Using Multi-task Cascaded Convolutional Networks. Computer Vision and Image Understanding, 134(3), pp. 56-64.
- Hallaji E, Razavi-Far R, Saif M. DLIN: Deep Ladder Imputation Network. IEEE TransCybern. 2022 Sep;52(9):8629-8641. doi:10.1109/TCYB.2021.3054878. Epub 2022 Aug PMID: 33661751.
- 9. Arjun, E. L., et al. "Advanced Face Authentication Using Deep Learning Models." 2023 IEEE Pune Section International Conference (PuneCon). IEEE, 2023.
- 10 Bileschi S, Heisele B (2002) Advances in Component Based Face Detection. In:Proceedings of the First International Workshop on Pattern Recognition with Support Vector Machines. Springer-Verlag, London, UK
- Li, C., Li, R., Sun, J.: CNN face live detection algorithm based on binocular camera. J.Phys. Conf. Ser. 1881(2): 022015 (7pp) (2021)
- 12 Anitha, G., Devi, P. S., Sri, J. V., & Priyanka, D. (2020). Face Recognition Based Attendance System Using MTCNN and FaceNet. Zeichen Journal, 6(1),189-195.
- 13 Gornale, B., & Kiran, P. (2020). Classroom Attendance Management System Using Camera. International Journal of Research in Engineering, Science and Management, 3(8),327-330.
- Khan, M. Z., Harous, S., Hassan, S. U., Khan, M. U. G., Iqbal, R., &Mumtaz, S. (2019).Deep Unified Model for Face Recognition Based on Convolution Neural Network and Edge Computing. IEEE Access, 7,72622-72633
- Goyal, A., Dalvi, A., Guin, A., Gite, A., & Thengade, A. (2021). Online Attendance Management System Based on Face Recognition Using CNN. In 2nd International Conference on IoT Based Control Networks and Intelligent Systems (ICICNIS 2021).