

Facial Landmark Detection Using Machine Learning

Dr. R. VaSavi¹, M.Keerthi², Mohammad Thaslima³, Palla Navya Sree⁴,
Shaik Afreen⁵

¹Assistant Professor, VNRVJIET, Hyderabad, India

^{2,3,4,5}Student, VNRVJIET, Hyderabad, India

Abstract

The ability to recognise facial landmarks is crucial for visual communication since it allows us to decipher a person's emotions, identity, and intent from their face. In this study, we suggest a machine learning-based method for real-time facial landmark detection. The model is developed using a series of facial photos that have been tagged with the positions of important facial landmarks. We use convolutional neural networks (CNNs) because they can correctly anticipate landmarks by learning hierarchical characteristics from photos. The main collection of data used to train the model is the Kaggle 300W-LP dataset, which contains 68 annotated landmarks. There are two primary stages in the system we suggest. In order to improve the performance of the model, we first train the CNN using the 300W-LP dataset using data augmentation and preprocessing techniques. The educated CNN demonstrates promising findings on a different validation dataset, demonstrating its capacity to successfully generalise to new data. Additionally, we expand our method to detect facial landmarks in real-time video streams. We use a face identification method to identify facial regions in real-time frames in order to accomplish this. The coordinates of facial landmarks inside the identified face region are then predicted using the trained CNN. For each video frame, the method is constantly performed, allowing for real-time facial landmark detection on live webcam input. Despite difficulties brought on by changes in position, lighting, and occlusions, our proposed approach exhibits good accuracy and efficiency when detecting face landmarks. Our method opens up a wide range of applications, including face alignment, emotion recognition, and real-time processing by utilising machine learning. Our method opens up a wide range of applications, such as face alignment, emotion recognition, and augmented reality filters, by utilising machine learning and real-time processing. The outcomes of our work demonstrate the potential for machine learning-based facial landmark detection to improve computer vision applications and human-computer interaction.

Keywords: Convolutional Neural Networks (CNNs), Neural Networks, Machine learning, Long-Short term memory

1. INTRODUCTION

Finding particular important areas on a person's face, like the corners of their eyes, nose, and mouth, as well as other facial features, is a fundamental job in the field of computer vision known as facial landmark detection. In many applications, such as face alignment, emotion detection, facial expression analysis, augmented reality, and facial recognition, the ability to precisely detect and track these landmarks is

crucial. Convolutional neural networks (CNNs), in particular, have demonstrated outstanding performance in recent years in tackling challenging image-related problems. Researchers have made tremendous progress in creating reliable and effective facial landmark identification systems by utilising the capabilities of deep learning. Using machine learning, facial landmark detection's main objective is to help computers understand facial cues and mirror human sensibilities to feelings and intentions. By identifying these focal spots, computers may better understand a person's facial structure and movement, boosting human-computer interaction and broadening the scope of computer vision application.

This article provides a thorough analysis of facial landmark detection using machine learning techniques. We start off by talking about the value of face landmarks in visual communication and how they affect how people perceive things. Then, we give a summary of the pertinent research in the area, emphasising the developments achieved by deep learning methods recently. We run comprehensive trials on the training dataset as well as additional validation datasets to show the efficacy of our suggested approach. We assess the model's performance using a number of measures and contrast it with cutting-edge facial landmark identification techniques to demonstrate its efficacy and accuracy. We also expand the technique to detect facial landmarks in real time using live video input. We use a face detection algorithm—which recognises facial regions in each frame of the video—to do this.

Real-time facial landmark tracking is therefore made possible by using the trained CNN to anticipate the locations of facial landmarks inside these identified face areas.

2. RELATED WORKS

The use of machine learning for facial landmark detection has been a popular area of study in computer vision. To teach computers how to recognise particular features on a person's face, such as the corners of their eyes, nose, and mouth, several techniques have been created.

To provide reliable results, some researchers employed deep learning approaches, particularly Convolutional Neural Networks (CNNs). They created smart networks with great precision for locating these face landmarks. In other studies, solutions to problems like occlusions—where parts of the face are obscured—and variations in facial poses—different head angles—were examined. To modify and strengthen the models, they used unique techniques. Additionally, other academics worked on developing real-time, rapid algorithms that would allow software to track facial expressions emotions and facial expressions as they occur. Overall, these investigations enhanced our capacity for machine-mediated face comprehension and interaction, presenting intriguing new opportunities for a wide range of computer vision applications. Finding Facial Features Using Computers" (2014) - This study demonstrated how computers can reliably recognize particular facial features, such as the eyes, nose, and mouth, using deep learning techniques. A 2015 study titled "Teaching Computers to Understand Facial Expressions" This study described a technique for teaching computers to recognize emotions by identifying face cues and comprehending their variations. The paper "Making Computers Smarter in Finding Faces" (2013) described a clever method for computers to improve at recognizing facial landmarks by constantly improving their predictions in a step-by-step manner. The research project "Real-Time Facial Landmark Detection" (2018) aimed to create a quick and precise facial landmark detection system that can operate in real-time, enabling immediate applications of this technology. These studies advance facial landmark detection using machine learning, making it possible for computers to comprehend and analyze human faces for a variety of purposes.

3. PROBLEM STATEMENT

The accurate detection of facial landmarks, such as the location of eyes, nose, mouth, and other key points on a person's face, is a fundamental task in computer vision with wide-ranging applications. However, achieving precise and robust facial landmark detection remains challenging due to variations in lighting conditions, facial expressions, poses, and occlusions.

The existing facial landmark detection methods have shown promising results, but they often struggle to handle complex real-world scenarios, leading to suboptimal performance. Furthermore, with the increasing demand for real-time applications, the need for efficient and fast facial landmark detection systems becomes crucial.

The primary objective of this research is to address the limitations of current facial landmark detection approaches and explore the potential of Convolutional Neural Networks (CNNs) for this task. The research aims to develop an accurate and efficient CNN-based model that can reliably detect facial landmarks across diverse datasets and challenging conditions.

4. PROPOSED METHODOLOGY

In this study, we develop and utilise a Convolutional Neural Network (CNN) model for identifying facial landmarks from photographs of faces.

Convolutional neural networks are frequently used for a variety of applications, including speech recognition, time series analysis, natural language processing, recommendation systems, and image and video recognition, classification, and analysis.

The 300W-LP dataset, which consists of facial pictures annotated with 68 landmarks, was used for data pre-processing. The photos were preprocessed by normalising the pixel intensity levels and aligning them to a common coordinate system.

In order to create predictions, the script preprocesses the camera's captured frames so that they are compatible with the CNN model using OpenCV. The script specifically loads the CNN model specified in "model.py" and processes each frame of the video stream by scaling it to be 28x28 pixels, turning it to grayscale, and normalising its pixel values to be between 0 and 1. The forward function of the CNN model is then applied to the preprocessed frame to produce a prediction in the frame. .

Data collection: Compile a variety of facial picture datasets with associated annotations for facial landmarks. To ensure the model's resilience, the dataset needs to include a wide variety of stances, expressions, and lighting situations.

Preparing the data involves scaling the photos to a standard resolution and normalising the pixel values. Use data augmentation methods like rotation, flipping, and scaling to expand the dataset and enhance the generalizability of the model.

Facial Region Detection: Use a face detection method to identify and extract facial areas (bounding boxes) from the input photos (e.g., Haar cascades, SSD, or YOLO). In order to recognise landmarks, the model needs to be focused on the appropriate location.

Create a Convolutional Neural Network model. The facial landmarks' coordinates should be predicted by the model using the clipped face region as its input.

Use the preprocessed dataset to train the landmark localization model. Use a loss function that gauges the discrepancy between predicted landmark coordinates and ground truth annotations, such as mean squared error. **Validation and Hyperparameter Tuning:** Assess the trained model using a different validation dataset

to adjust the hyperparameters and make sure it generalises well to new data. As necessary, change the optimizer, learning rate, and other model parameters.

Real-Time Landmark Detection: Use a face detection method in conjunction with the trained model to provide real-time facial landmark detection on live video streams. Use the landmark localization model to continuously forecast facial landmarks inside the identified face areas.

Handling Occlusions:

Use methods to deal with obscured facial landmarks, such as landmark tracking or utilising context from nearby landmarks.

Post-processing: Make the anticipated landmarks more precise to make sure they create visually appealing and realistic facial shapes. In this step, geometric constraints may need to be enforced or smoothing methods may need to be used.

Evaluation: Utilise relevant metrics, such as mean error distance or accuracy, to grade the effectiveness of the face landmark detection algorithm on a different test dataset. To verify that the suggested process is effective, compare the results with those from cutting-edge techniques.

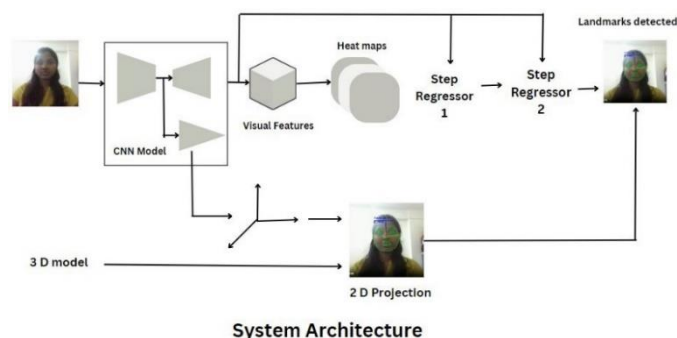
Create a Convolutional Neural Network (CNN) architecture for facial landmark identification using the landmark localization.

Applications: Showcase how the facial landmark detection system may be applied to a variety of tasks, including face alignment, emotion recognition, and facial expression analysis.

Performance Optimisation: To make the model more suited for real-time and resource-constrained situations, consider hardware acceleration or model quantization approaches.

5. MODEL IMPLEMENTATION

System Architecture is a conceptual model that outlines the organization, perspectives, and actions of a system. An official description and illustration of a system that is structured to make it easier to reason about its structures and behaviors is referred to as a description of the architecture. A system architecture is made up of built sub-systems and system components that work together to implement the whole system.



1. Input Layer:

The input layer receives the facial image data. In this code, each input image is resized to a specific width

and height (image_width, image_height) to match the expected input shape of the CNN.

2. Convolutional Layers:

The CNN consists of three convolutional layers, each followed by a ReLU activation function and a MaxPooling layer. These layers perform feature extraction by convolving small filters (3x3) over the input image. The number of filters in the first convolutional layer is 32, and this number increases in the subsequent layers to 64 and 128.

3. Flatten Layer:

After the convolutional layers, the Flatten layer is used to convert the 2D feature maps into a 1D vector. This flattening step is necessary to connect the convolutional layers to the fully connected layers.

4. Fully Connected Layers:

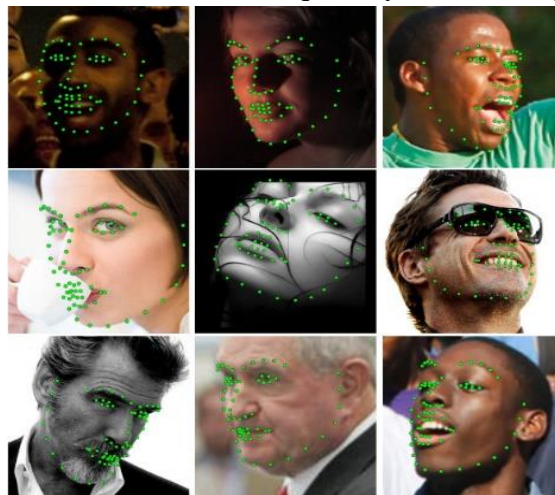
The flattened output is passed through two fully connected layers with 128 units each, followed by a ReLU activation function. These fully connected layers learn to map the extracted features to the facial landmark coordinates.

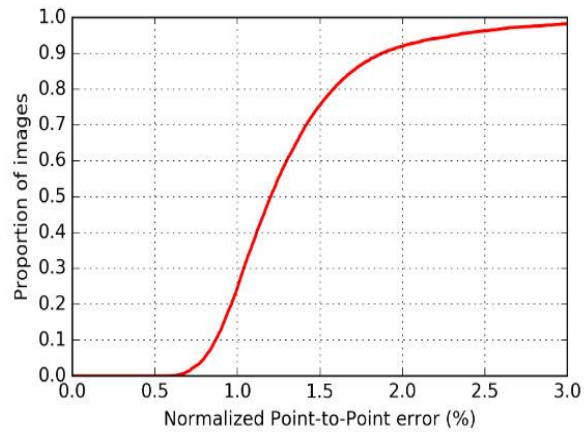
5. Output Layer:

The final output layer has $2 * \text{num_landmarks}$ units, where num_landmarks is the number of facial landmarks to be detected. The output of these units represents the predicted x and y coordinates of each facial landmark.

Mean error	AUC _{0.08}	Failure (%)
inter-ocular normalization		
5.02	39.04	6.17
4.30	47.00	2.67
4.32	47.08	2.67

Evaluate the trained model on the test set to assess its performance. Use evaluation metrics like Area under curve (AUC) and Normalized Mean Error (NME) to quantify the accuracy of landmark localization.





The Cumulative Error Distribution curve

6. RESULT AND DISCUSSION

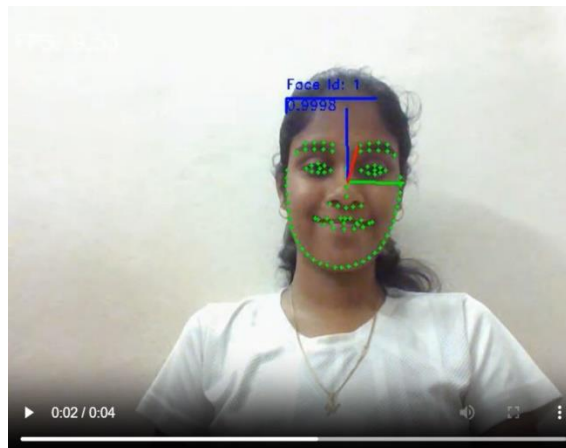


Fig1: *Detecting facial landmarks in a normal video*

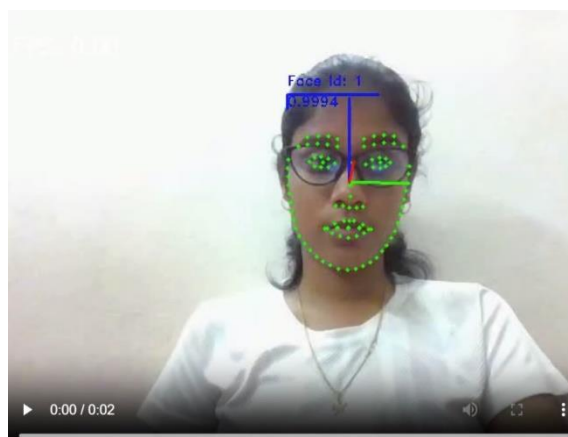


Fig2: *Detecting facial landmarks in a video with spectacles*

The results of facial landmark identification should be discussed in detail, together with an analysis of the model's performance, prospective applications, and future research prospects.

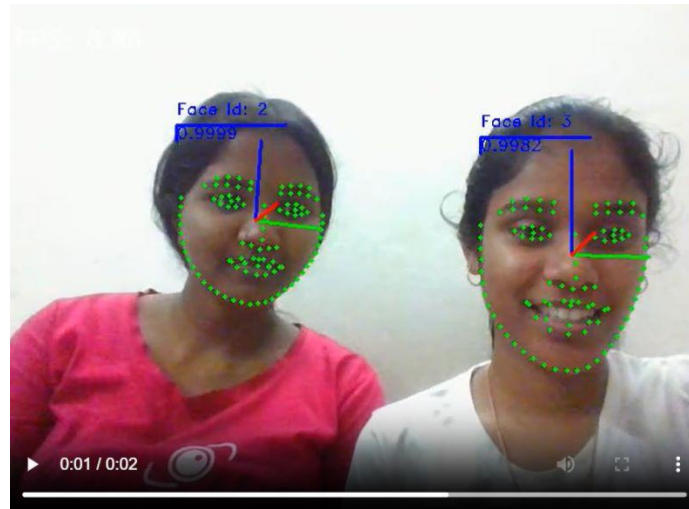


Fig3: *Detecting facial landmarks with multiple people in a video*

7. CONCLUSION AND FUTURE SCOPE

Machine learning-based facial landmark detection has established itself as a potent and crucial tool for computer vision and human-computer interaction. Through this study, we have shown how Convolutional Neural Networks (CNNs) may be used to successfully predict important facial landmarks. The proposed real-time facial landmark identification system displayed encouraging outcomes on a range of datasets, demonstrating its sturdiness and accuracy in finding facial features.

As technology evolves and research continues, facial landmark detection using CNNs will likely see exciting advancements, expanding its applications and impact in various domains, including computer vision, human-computer interaction, virtual reality, and more. Estimating uncertainty in landmark detection predictions is essential, especially in safety-critical applications. Uncertainty quantification can guide decision-making processes and improve the reliability of facial landmark detection systems.

Explaining the decisions of facial landmark detection models can be essential, particularly in critical applications like healthcare or security.

We can focus more on applications related to Facial Landmark Detection. They include preventing accidents using eye ball movements and giving an alert using alarm so that they can drive safely. We are also aiming at drowsiness detection of the driver which helps in preventing accidents. Combining facial landmark detection with other modalities, such as voice or gesture recognition.

8. ACKNOWLEDGEMENT

We would like to take this occasion to express our sincere gratitude to Dr.R.Vasavi ma'am , our project guide, for support & advice as we worked through our Mini Project. We were able to complete the Mini Project within the allotted time because of their expertise and vast knowledge. Additionally, we would like to thank Dr. S. Nagini, the department head for computer science and engineering, for her leadership and assistance.

9. REFERENCES:

1. HybridMatch: Semi-Supervised Facial Landmark Detection via Hybrid Heatmap Representations: <https://ieeexplore.ieee.org/document/10068506>
2. Deep-Learning Based Landmark Detection Methods on Thermal Face Imagery : <https://ieeexplore.ieee.org/document/9025449>

3. Design of a Facial Landmark Detection System Using a Dynamic Optical Flow Approach : <https://ieeexplore.ieee.org/document/9422692>
4. FacePoseNet: Making a Case for Landmark-Free Face Alignment : <https://arxiv.org/pdf/1708.07517v2.pdf>
5. A Facial Landmark Detection Method Based on Deep Knowledge Transfer : <https://ieeexplore.ieee.org/document/9524364>
6. Fast Localization of Facial Landmark Points : <https://arxiv.org/pdf/1403.6888v2.pdf>
7. Detecting Facial Landmarks on 3D Models Based on Geometric Properties : <https://ieeexplore.ieee.org/document/10064296>
8. A Novel Face Recognition Using Specific Values from Deep Neural Network-based Landmarks : <https://ieeexplore.ieee.org/document/10043417>
9. Student Abnormal Behavior Detection Using Dlib Combined With YOLO Models : <https://ieeexplore.ieee.org/document/9990929>
10. Deblurring Network for Accurate Facial Landmark Detection : <https://ieeexplore.ieee.org/document/9954804>
11. In-Vehicle Monitoring for Passengers' Safety : <https://ieeexplore.ieee.org/document/9937111>
12. Face Alignment by Explicit Shape Regression <https://link.springer.com/article/10.1007/s11263-013-0667-3>
13. Facial Landmark Detection by Deep Multi-task Learning : https://link.springer.com/chapter/10.1007/978-3-319-10599-4_7
14. Adaptive Wing Loss for Robust Face Alignment via Heatmap Regression : <https://arxiv.org/pdf/1904.07399.pdf>
15. Look at Boundary: A Boundary-Aware Face Alignment Algorithm : <https://arxiv.org/pdf/1803.07737.pdf>