© JAN 2026 | IRE Journals | Volume 9 Issue 7 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I7-1713768

Modelling An Dense Network Model for Moderate Facial Alignment Prediction by Feature Representation

DR. S. BAGHYASHREE[1], K. GAYATHRI[2]
[1, 2]*Information Technology, Anna University College of Engineering and Technology Madurai, India*

Abstract- ***Deep learning approaches are extremely productive and accurate for predicting moderate face alignment and facial landmarks. The model learns sequential analysis during the training process to reduce discrimination between the ground truth value and the shape of the face based on feature representation. While testing, it employs feature representation to identify the shape factors iteratively. Also, when the facial directions and expressions change, the existing learning model cannot acquire superior performance based on the enormous variations among the target and the initial shape. This work proposes a novel multi-stage gradient descent with the ResNet-50 model to preserve higher prediction accuracy on training samples and enhance the testing data accuracy. One sample is provided during training, and multiple samples are provided with changing expressions. During testing, the distance among the face alignment landmarks is evaluated with the optimal selection of ligaments. The simulation is done in MATLAB 2020a environment. The outcomes show that the anticipated model can enhance the conventional approaches' performance and show a better trade-off than other approaches.***

Keywords- ***Face Alignment, Deep Learning, Dense Network Model, Prediction, Accuracy.***

I. INTRODUCTION

Due to the numerous applications, including video surveillance, entertainment, and human-computer interaction, facial landmark localization and tracking in an unrestricted environment have recently attracted much interest [1]. Deep Convolutional Neural Networks (DCNNs), claimed to have Hourglass architecture and resolution-maintained structure, are the focus of the most recent benchmarks for 2D face alignment. The human face's 3D structure is made up by rarely preserved in 2D facial landmark annotations [2]. Despite the almost saturated publically accessible benchmarks, the performance of 2D faces alignment they are not necessarily semantically coherent [3]. It is especially obvious when there are notable position changes and face landmarks.

In contrast, 3D annotations retain a correspondence between postures. In this research, the 3D face landmarks' 2D projections are called "3D annotations." 3D facial landmark projections in [4] are upgraded by incorporating a regression network to determine the depth to complete 3D face landmarks. Due to the variability of facial features, 3D face alignment is highly challenging in uncontrolled situations, significantly due to the camera's occlusion, defocus extreme position, and inadequate resolution [5]. Aside from that, the boundary features could be more trustworthy due to self-occlusion from the significant position variations. Only following the human face's 3D anatomy needs to be preserved contextual information can be used to forecast the occluded landmarks [6].

For 3D face alignment, this research suggests the Cascade Multi-view Hourglass Model (CMHM). Supervision signals from 2D and 3D facial landmarks cascade two Hourglass models. We align this issue properly to increase accuracy when poses change significantly [7]. We have provided three contributions: 1) To increase capacity while maintaining the Hourglass model's computational complexity, instead of using the final bottleneck block, we add a parallel, multi-scale inception resnet block [8]. 2) We create a novel multi-view hourglass model (MHM) that simultaneously estimates semifrontal and profile 2D facial landmarks by using correspondences between frontal and profile face characteristics [9]. 3) Using a cascade technique, we first locate the 2D face landmarks using the Hourglass model with multiple views [10]. After removing the similarity transformation, the 3D facial landmarks are estimated using another Hourglass model. The suggested method provides better results by adopting ResNet50 for facial alignment prediction using the modern deep feature

representation. The research aims to offer superior performance compared to various prevailing approaches and establish the finest trade-off among those approaches.

The work is structured as section 2 provides a wider analysis of diverse approaches. In section 3, the research methodology is elaborated. The experimental outcomes are provided in section 4, with a work summary in section 5.

II. RELATED WORKS

Deep learning refers to artificial intelligence (AI) methods based on neural networks modeled after brain components. (DL). It is a phrase that describes methods for automatically identifying the underlying and basic relationships in a graphical data model. Instead of relying on generating finger features, which may be sporadic and difficult, deep learning techniques require significantly less operator instruction and instead learn appropriate feature representations [11]. Furthermore, DL techniques scale far better than traditional Machine learning when the amount of data increases. An outline of a few important DL ideas is given.

The simplest multilayer perceptron is an artificial neural network (ANN), which has three layers of neurotransmitters: an input node, a dense node, and an output surface. These networks can be categorized as Superficial (Feed-Forward) Artificial Neural Systems because they only receive one hidden unit. A Deep (Feed-Forward) Naive Bayes (DNN), on the other hand, has much more than just a hidden layer [12]. Every hidden neuron in the network is connected to every input, two hidden clusters across networks in a city, concealing each and producing representations of numerous biological neurons. These circuits cannot be directly applied in neuroimaging research because they only accept a single matrix as input.

CNNs, or Convolutional Neural Networks, based on a fundamental mathematical operation known as "transform," were inspired by biological perception and use. Instead of using external neural networks, use 2D arrays as input [13]. The key difference between a DNN and a CNN is that each hidden neuron's output is determined by the input of all synapses in a single layer in a DNN. When compared, this is different in the first case.

Instead, utilizing filters or loudspeakers, a CNN travels across a section of the top image to compute the convolution layers, creating a depth map [14]. The value of each component of the preceding layer of the following element is therefore computed using only a frame of x^2 pixels if the tube's size is x. It immediately impacts the visual field, which may be considered the area in the input vector that impacts a particular CNN property.

The network's multilayer portion is "the extracting features component," while the remaining is "the classifying section." The first learn photographic characteristics to classify the incoming image using the constructed attributes, translated from one array given into the second.

Deep learning-based face alignment techniques have become more popular in recent years. These techniques often consider face alignment a regression issue and discover landmarks employing various deep models in a coarse-tofine progression. Author et al. [15], who also developed the DCNN cascaded CNN (with three levels), employed CNN to identify important points on faces for the first time.

A coarse-to-fine self-encoder network was developed, which depicts the intricate nonlinear mapping between facial appearance and face form. Self-encoding multiple stack networks in a nonlinear cascade allows for the nonlinear description. The model is one possibility in 2016 that can execute face alignment tasks more precisely and roughly forecast the locations of faces and landmarks thanks to its deep convolutional network's three-level cascade topology [15]. However, all these methods fail to provide better prediction outcomes. Thus, the research gap is intended to be filled by the anticipated model.

III. METHODOLOGY

The two main components of this study are gathering data to forecast facial alignment using the ResNet50 model. Measurements, including prediction accuracy, precision, recall, and F1score, evaluate how well the

© JAN 2026 | IRE Journals | Volume 9 Issue 7 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I7-1713768

predicted model performs. Fig 1 depicts the flow of the anticipated model.

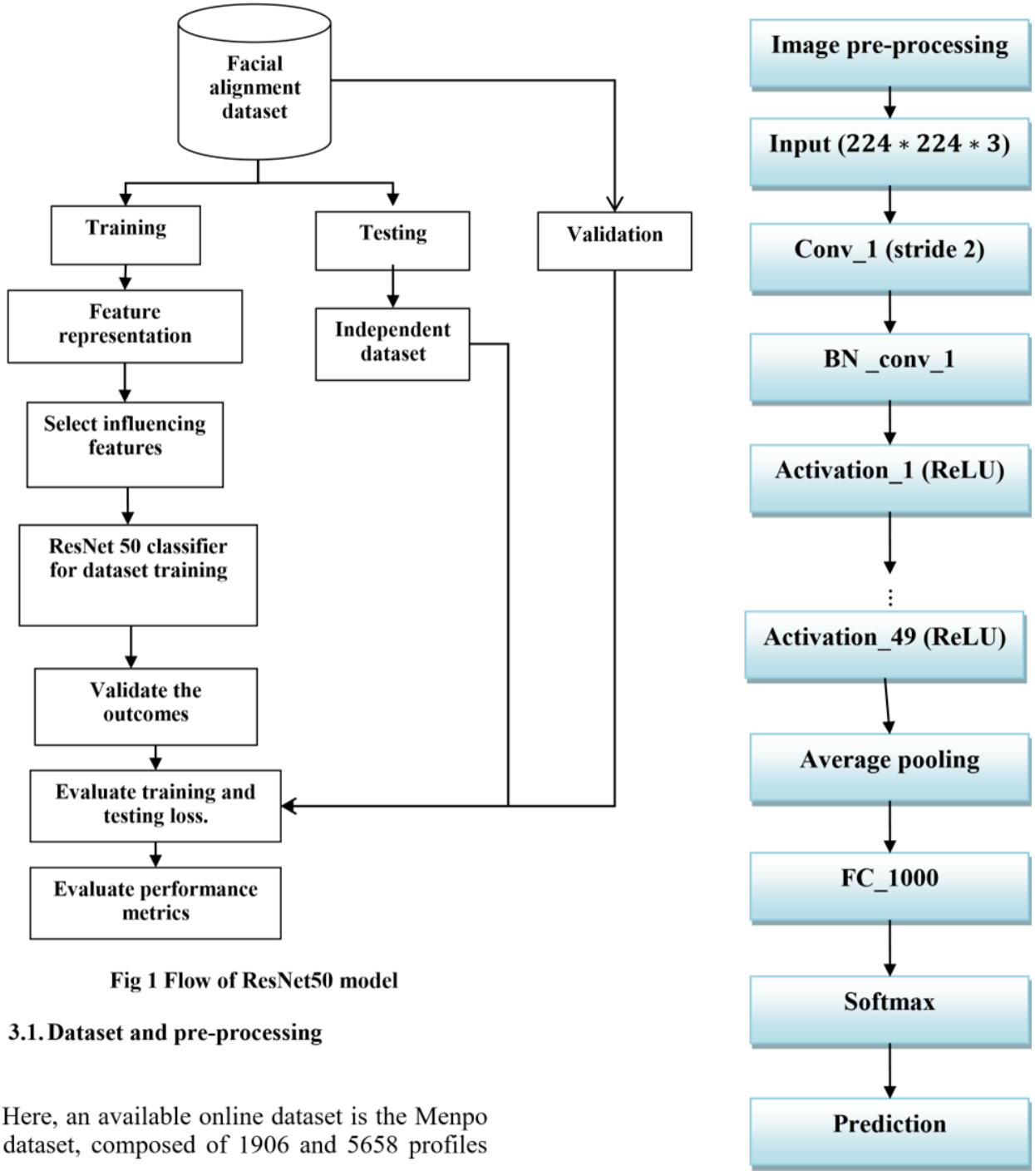

Fig 1 Flow of ResNet50 model

3.1. Dataset and pre-processing

Here, an available online dataset is the Menpo dataset, composed of 1906 and 5658 profiles and semi-frontal face images chosen from the existing ALFW and FDDB datasets. The annotated face images are gathered from unconstraint conditions, showing huge variations in illumination, expression, pose, etc. The profile and semi-frontal faces are annotated here with 39 and 68 2D landmarks. Here, 2D annotation schemes are used for training and testing. Generally, pre-processing is essential for enhancing image quality and performance. Here, augmentation is employed where the image orientations are changed to balance the over-fitting or

© JAN 2026 | IRE Journals | Volume 9 Issue 7 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I7-1713768

under-fitting issues, and the dimensionality is changed based on the requirements.

Fig 2 ResNet50

3.2 ResNet-50

A new CNN model is presented in the framework for effective facial alignment prediction using the supplied input images. Deep features are extracted using one of the deep pre-trained CNN models that have been upgraded and transfer learned. Using the pre-trained ResNet50 model, we developed a brand-new ResNet50-modified CNN architecture, as shown in Fig []. Adding extra layers at the end improves the facial alignment dataset using the ResNet50 model architecture. Low-resolution input photos can its height-to-width proportion varies. Thus, for aIV. comparable action sequence in the described model architecture, the training and testing dataset images are increased to $224 * 224 * 3$. Because of how easy it is to improve ResNet and how much more accurate it might be, It is the best deep-learning architecture. In addition, vanishing gradient problems are recurring problems in which the network's skip connections are fixed. The deep network architecture's temporal complexity increases exponentially as the number of layers rises. A bottleneck design can be used to lessen this complexity. For the development of our framework, we discarded other pre-trained networks with additional layers and went with the ResNet50 pre-trained model. The architecture is explored in more detail below. The ResNet50 architecture has been modified to achieve effective performance forecasting facial alignment. The final three layers of the pretrained model have first changed ResNet50 architecture (completely connected, softmax, and classification layers) to make them more suitable for our classification assignment. A second completely connected layer replaces a layer of the initial pre-trained networks fully connected, whose output size, in our case, symbolizes the two classes, Non-Covid and Covid. The ResNet50 architecture has three more layers: Conv, Batch Normalization, and Activation Relu, which automatically extract robust features from the input images. It is shown in Fig 2. The convolution layer comes first among these layers. The batch normalization layer and activation layer comes next. The 3 layers are added using the procedures below.

1. The newly added Conv layer is connected to the activation 49 relu layers separated from the avg pool layer.
2. The average pool layer is connected to the newly added activation relu layer.
3. The most recent three layers come after the average pool layer and are entirely connected softmax and classification layer. The ResNet50 design is shown in Fig 2 before and after adding new layers. Using this redesigned network, the input images are now processed to gather each image in the collection has features. The network classifier then categorizes the input images. The suggested model was developed to forecast face alignment where the model is optimized with gradient descent.

Numerical results

We have constructed the suggested system for diagnosing face alignment using the Matlab R2020a programming language, an Intel Core i5powered Windows 10 computer with 6 GB of RAM. In addition to the $1e^{-4}$ learning rate, five epochs, and the Adam optimizer, weight updates are performed.

Measures for precision, sensitivity, accuracy, and specificity were used to gauge the productivity and utility of the categorization model. Calculating precision involves comparing the model's accurate predictions to all other projections. The following equations state that to calculate the assessment criteria and four key outcomes— it is necessary to take into account true positives (TP), false positives (FP), true negatives (TN), and false negatives (FN).

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \quad (1)$$

$$Sensitivity = \frac{TP}{TP + FN} \quad (2)$$

$$Specificity = \frac{TN}{TN + FP} \quad (3)$$

Table 1 Comparison of various prevailing approaches

Approaches	Accuracy	Sensitivity	Specificity	Precision
Proposed ResNet-50	97%	98%	95%	94%

© JAN 2026 | IRE Journals | Volume 9 Issue 7 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV9I7-1713768

Google Net	96%	96.5%	96.5%	95%
AlexNet	94%	92%	96%	94%
DenseN et	96.2%	98%	94%	93%
VGG-16	91%	89%	93%	90%
VGG-19	88%	92%	86%	83%
Incepti on v3	96%	97%	95%	94%

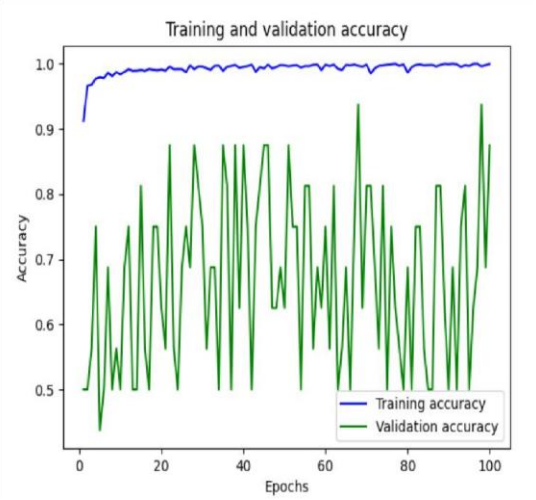

Fig 3 Training and validation accuracy

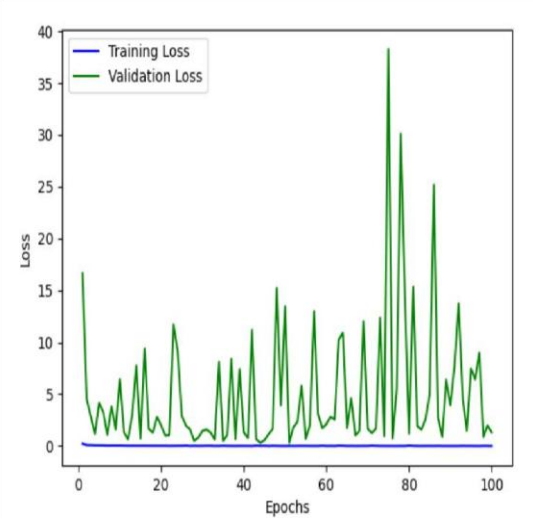

Fig 4 Training and validation loss

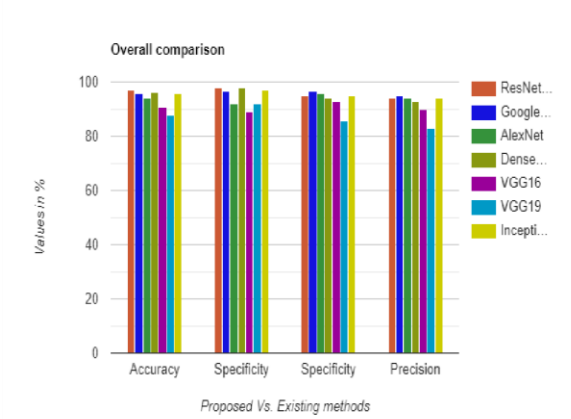

Fig 5 Overall performance comparison Accuracy, a well-known evaluation metric, is used to assess the effectiveness of the suggested solution. When applied to the input dataset, Table 1 demonstrates our suggested model's sensitivity, specificity, and precision scores, which are 97.1%, 98.9%, 95.7%, and 94.5%. According to Table 2, the suggested model gets values of 97.7%, 98.7%, 95.6%, and 97.9%, respectively. Precision, accuracy, sensitivity, and focus were assessed. However, ResNet50 attained 96.8% and 96.8% accuracy levels for the provided input datasets, respectively. The predicted landmarks are shown in Fig 6.

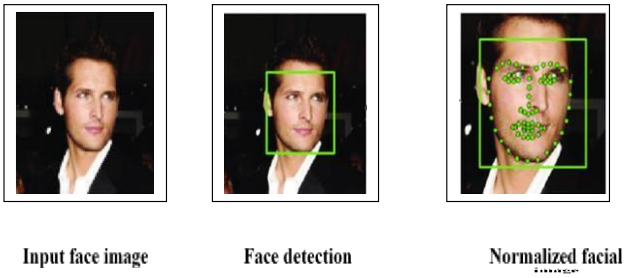

Fig 6 Facial landmarks prediction

V. CONCLUSION

To develop a unique deep transfer learning model for face alignment detection, this paper draws on the pre-trained ResNet50 model and convolutional neural network. The ResNet50 may extract more robust features by adding additional three layers. The three proposed layers have improved the accuracy of the ResNet50 model. Experimental findings demonstrated

the effectiveness of our model as a tool for identifying face alignments.

Studies comparing the two methodologies show that the proposed methodology outperforms the well-known models GoogleNet, AlexNet, DenseNet201, VGG16, VGG19 and InceptionV3 of deep transfer learning. Additionally, the precision of the suggested method outperforms the VGG19 model with the provided input datasets by 10%, respectively. The proposed methodology can improve facial alignment testing because of most medical facilities. The suggested model can therefore serve as an alternative to different facial alignment testing tools.

REFERENCES

[1] Adrian Bulat and Georgios Tzimiropoulos. Two-stage convolutional part heatmap regression for the 1st 3d face alignment in the wild (3dfaw) challenge. In ECCV, pages 616– 624. Springer, 2016.

[2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In CVPR, pages 770–778, 2016

[3] Qingshan Liu, Jiankang Deng, Jing Yang, Guangcan Liu, and Dacheng Tao. Adaptive cascade regression model for robust face alignment. TIP, 26(2):797–807, 2017.

[4] Qingshan Liu, Jing Yang, Jiankang Deng, and Kaihua Zhang. Robust facial landmark tracking via cascade regression. PR, 66:53–62, 2017

[5] Shengtao Xiao, Jiashi Feng, Junliang Xing, Hanjiang Lai, Shuicheng Yan, and Ashraf Kassim. Robust facial landmark detection via recurrent attentive-refinement networks. In ECCV, pages 57–72. Springer, 2016.

[6] Xiangyu Zhu, Zhen Lei, Xiaoming Liu, Hailin Shi, and Stan Z Li. Face alignment across large poses: A 3d solution. In CVPR, pages 146–155, 2016

[7] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Yu Qiao. Joint face detection and alignment using multitask cascaded convolutional networks. SPL, 23(10):1499– 1503, 2016.

[8] Pengfei Xiong, G Li, and Y Sun. 3d face tracking via a two-stage hierarchically attentive shape regression network. In ICCV Workshop, 2017

[9] Yu Liu, Duc Minh Nguyen, Nikos Deligiannis, Wenrui Ding, and Adrian Munteanu. Hourglass-shape network-based semantic segmentation for high-resolution aerial imagery. Remote Sensing, 9(6):522, 2017

[10] Qingshan Liu, Jiankang Deng, and Dacheng Tao. Dual sparse constrained cascade regression for robust face alignment. TIP, 25(2):700– 712, 2016

[11] Hanjiang Lai, Shengtao Xiao, Yan Pan, Zhen Cui, Jiashi Feng, Chunyan Xu, Jian Yin, and Shuicheng Yan. Deep recurrent regression for facial landmark detection. CSVT, 2016 [12] Lee, D., Park, H., & Chang, D. Y. (2015). Face alignment using cascade Gaussian process regression trees (pp. 4204–4212).

[13] Burgos-Artizzu, X.P., Perona, P., Dollár, P.: Robust face landmark estimation under occlusion. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1513–1520 (2013)

[14] Cao, X., Chen, Z., Chen, A., Chen, X., Li, S., Yu, J.: Sparse photometric 3D face reconstruction guided by morphable models. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4635–4644 (2018)

[15] Cao, C., Weng, Y., Zhou, S., Tong, Y., Zhou, K.: Facewarehouse: a 3d facial expression database for visual computing. IEEE Trans. Vis. Comput. Graph. 20(3), 413–425 (2013)