iJRASET
International Journal For Research in
Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH
IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 Issue: V Month of publication: May 2024

DOI: https://doi.org/10.22214/ijraset.2024.62662

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

Age and Gender Detection Using Deep Learning

Abhinav Singh[1], Avanish Kumar[2], Anshika Gupta[3], Abhishek Kumar Tiwari[4], Deepak Vishwakarma[5]
[1, 2, 3, 4]Student of Information Technology KIET Group of Institutions, Delhi-NCR
[5]Asst. Prof. Information Technology Department KIET Group of Institutions, Delhi-NCR

Abstract: Facial attributes are crucial in numerous applications such as access control and video surveillance, where demographic data like age and gender can be inferred from facial images. Automatic estimation of age and gender enables tailored content de- livery and personalized services. However, ex- tracting effective features from facial images poses a significant challenge. This paper pro- poses employing Convolutional Neural Net- works (CNNs) for automatic age and gender prediction. CNNs have demonstrated ground-breaking success in face recognition and im- age classification tasks. Leveraging pre-trained deep CNNs, this research aims to estimate age and gender accurately from facial images. The methodology involves utilizing convolution layers to produce a robust and compact output, enhancing the efficiency of age and gender detection systems.

I. INTRODUCTION

Age and gender prediction have be- come one of the more recognized fields in deep learning because of the rise in picture uploads on the internet in to- day's data-driven environment. Although humans are naturally skilled at identifying one another, figuring out gender, and assessing ethnicity, age assessment is nevertheless a challenging task. To underscore the complexity of the issue, consider this: The most used statistic for assessing an individual's age prediction is mean absolute error (MAE). According to research, depending on the database settings, people can estimate the age of an individual over 15 with an MAE 7.2–7.4. This indicates that human forecasts are often wrong by 7.2–7.4 years. The question is, can we do better?[1] Can we automate this problem in a bid to reduce human dependency and simultaneously obtain better results? For these reasons, persons of comparable ages might appear substantially different from one another. Because of this, estimating age is fundamentally a difficult undertaking. This issue is further exacerbated by the non-linear relationship between age and gender and face appearance, as well as the extreme lack of big, balanced datasets with accurate labelling.[13] There are very few such datasets available; the ma- jority are severely skewed, with a large proportion of participants in the 20–75 age range, or they are gender-biased.[14] It is not advisable to use such biased datasets since testing on real-time pic- tures will result in a distribution mis- match and subpar performance. There is an enormous amount of untapped poten- tial in this field of research. The enormous potential that autonomous age and gender prediction offers in a variety of computer science domains, including HCI

Fig. 1. Face Dataset

(Human Computer Interaction), has led to an ever-growing interest in this area. Law enforcement, security management, and forensics are a few possible uses. Using these models with IoT is another useful use.[1] A restaurant may decide to alter its theme, for instance, by calculating the average age or gender of patrons who have come in thus far.

II. PROCEDURE

A. Deep Learning

An artificial intelligence (AI) method called deep learning aims to mimic the human brain by learning from experience. These representations are learned through a technique called training. We must first train the program with a huge number of object photos that we classify into various groups to teach it how to recognize objects. Deep learning-based algorithms take longer to train than conventional machine learning techniques and need a lot more training data. It takes a lot of ef- fort and complexity to identify distinctive features while attempting to identify any item or letter on a picture. Deep learning techniques, which automatically extract significant characteristics from data, can be used to solve issues, in contrast to classical machine learning, where features are collected manually.

Deep learning is the use of numerous hidden layers in a neural network. Once a picture has been taught throughout the network, they can proceed to construct more complex ideas from simpler ones. An image may be taught in the network to understand objects like characters, faces, and so forth by incorporating basic features like form, edges, and corners. Each layer receives a basic attribute as the picture moves across them, progressing to the next one. The network may potentially learn more complicated characteristics as the layers get bigger and combine them to identify the image. Deep learning has found many applications in the field of computer vision. The most significant computer vision applications were found in the fields dealing with face data.

B. Convolutional Neural Networks

Convolutional neural networks (CNNs) are a popular type of machine learning algorithm used for image processing and recognition. They excel at categorizing images by taking them as input and processing them with a given dataset. Comprised of fully connected layers responsible for classifying images after extracting features. CNN utilizes a blend of both supervised and unsupervised learning methods, specifically through a multilayer feed-forward architecture. These distinctive stages consist of numerous layers, each with their own designated functions and objectives. The Convolution layer is a crucial element of the CNN algorithm, responsible for most of its computations. It takes in key components such as the input image, filter, and feature map. For example, when given a human face image as input, the Convolution layer processes its 3D matrix of RGB pixels, defining the image's length, width, and height. Within each layer, we establish filter and stride matrices to aid in the crucial process of feature extraction.

The filter serves as a feature detector, taking on a 2-dimensional form and capable of a variety of sizes, such as a 3 by 3 matrix. This filter acts as a compact representation of numbers, which we use with the input data to pinpoint specific features. The stride dictates how far the filter will move in each direction. The stride value is crucial in determining the size of the output image matrix. A larger stride, such as 4 or 5, will result in a smaller output matrix and potential loss of information and vice versa.

After passing the input image through the convolutional layer, ReLU activation function is applied elementwise to the feature maps produced by the convolutional layer in CNN. Applying ReLU introduces non-linearity to the network and allows it to learn complex patterns and representations in the data. It helps in capturing and amplifying important features while suppressing irrelevant or negative values. Then it is passed through the pooling layer where down sampling of the feature maps happens for faster computation.

The concept of Transfer Learning is application of the knowledge learned by one model to be applied to another. It is used when there is a dearth of suitable learning data. Deep neural network can be trained with previously saved model weights using transfer learning on large datasets. Therefore, a pre-trained model can be improved using large-scale deep transfer learning and limited data.

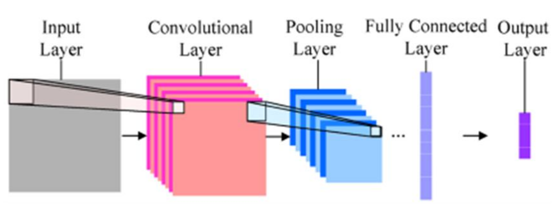


Fig. 2. Layers of Convolutional Neural Network

The UTK Face dataset is a very small dataset to capture the complexity involved in age and gender estimation, so we fo- cused our attention further on leverag- ing transfer learning.[13] Therefore, we are using convolutional blocks of VGG16 pretrained on VGG Face and ResNet50 pre-trained on VGG Face2, as feature extractors.[14] These models are origi- nally proposed for facial recognition, thus can be used for higher level of feature extraction.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue V May 2024- Available at www.ijraset.com

VGG Face is composed of two blocks, each containing layers for batch normalization, spatial dropout with a probability of 0.5, separable convo- lutions layers with 312 filters of size 3x3, maintaining the same padding, and max pooling with kernel size 2x2. The ResNet50 gender consists of only the fully connected system with batch norm, dropout with probability of 0.5, and 128 units with exponential linear unit (ELU) activation. The fully connected system was composed of batch normalization layers, alpha dropout, and 128 neurons with ReLU activation.

The two convolution blocks that make up the VGG face for age estimation are separated by a separable convolution layer that has 312 filters of size 3x3, padding the same so that the dimension remains constant with the ReLU activation function, a batch norm layer, and spatial dropout with keep probabilities of 0.8 and 0.6, respectively. After every con- volution block, a 2x2 kernel max pooling was performed. The three layers of the fully linked system included 648, 312, and 128 neurons each, and their corre- sponding dropout keep probabilities were 0.2, 0.2, and 1.

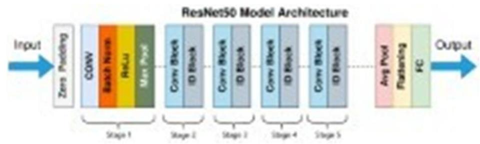


Fig. 3. ResNet50 Model Architecture

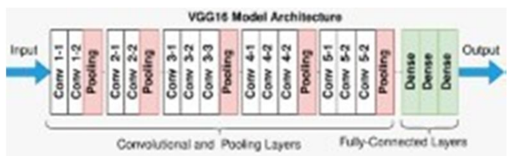


Fig. 4. VGG16 Model Architecture

TABLE I
Network Architecture for Age Estimation

Layer	Filters	Output Size	Kernel Size	Activation
Image	-	180 x 180 x 3	-	-
Separable Conv1	64	180 x 180 x 64	3 x 3	ReLU
Max Pooling	-	90 x 90 x 64	2 x 2	-
Separable Conv2	128	90 x 90 x 128	3 x 3	ReLU
Max Pooling	-	45 x 45 x 128	2 x 2	-
Separable Conv3	128	45 x 45 x 128	3 x 3	ReLU
Max Pooling	-	22 x 22 x 128	2 x 2	-
Separable Conv4	256	22 x 22 x 128	3 x 3	ReLU
Max Pooling	-	11 x 11 x 256	2 x 2	-
Separable Conv5	256	11 x 11 x 256	3 x 3	ReLU
Max Pooling	-	5 x 5 x 256	2 x 2	-
FC1	-	128	-	ReLU
FC2	-	64	-	ReLU
FC3	-	32	-	ReLU
Output	-	1	-	ReLU

III. RESULT

We conducted experiments on Age and gender Recognition, utilizing a Convo- lutional Neural Network (CNN) as our primary algorithm. The CNN underwent training and testing using a dataset com- posed of Human faces images with age and gender mentioned. The following are the pivotal elements of our experiments and the resulting outcomes:

1) *Database:* Our dataset is rich and diverse, encompassing a wide range of faces. It includes multiple types of face, varying with age and gen- der, and a range of perspectives. To prepare and assess our CNN model, the dataset was split into separate training and testing sets.

2) *Model Architecture:* Our team developed a cutting-edge CNN architecture specifically tailored for the purpose of recognizing age and gender. This intricate design comprises several essential components, including convolutional layers, pooling layers, and fully connected layers. To perfect our model, we meticulously fine-tuned important hyperparameters such as the number of layers, filter sizes, and the size of the fully connected layers. This ensured optimal performance and accuracy for our revolutionary Age and Gender detection system.
3) *Training:* The training process was a vital component, as it entailed inputting the training dataset into the CNN model. Through this, the model was able to acquire the ability to accurately distinguish and categorize age and gender based on the visual data provided. Our approach involved carefully selecting suitable loss functions and optimization algorithms, fine-tuning the model's inner workings (weight and biases) to effectively reduce any classification errors.
4) Evaluation: Once the training was complete, we proceeded to assess the model's performance by testing it with a dataset that it had not previously encountered. This step was essential in determining the model's ability to handle unfamiliar data. The evaluation process involved using various metrics such as accuracy, precision, recall, and F1-score, providing a comprehensive measure of the model's performance.

A confusion matrix provides insight into the model's performance, errors, and weaknesses. It breaks down the number of correct and incorrect predictions by each class, and can be used to calculate metrics such as:

- Accuracy: The proportion of predictions that the model classified correctly
- Precision: The proportion of relevant instances among the retrieved instances
- Recall: The proportion of the total amount of relevant instances that were retrieved

Fig. 5. Output Result 1

Fig. 6. Output Result 2

$$\text{Accuracy} = \frac{tp + tn}{tp + fp + fn + tn}$$

$$\text{Sensitivity} = \frac{tp}{tp + fn}$$

$$\text{Specificity} = \frac{tn}{fp + tn}$$

$$\text{Precision} = \frac{tp}{tp + fp}$$

$$\text{F1 - score} = \frac{2.tp}{2.tp + fp + fn}$$

The symbols fp, fn, tp, and tn refer to abbreviations of false positive, false negative, true positive and true negative respectively.

Predicted Values \ Actual Values	Positive (1)	Negative (0)
Positive (1)	TP	FP
Negative (0)	FN	TN

Fig. 7. Confusion Matrix

IV. CONCLUSIONS

Training Accuracy: The CNN demonstrated exceptional learning ability by achieving a training accuracy of 95.75%, showcasing its prowess on the provided dataset.

Test Accuracy: The test accuracy, a crucial measure of how well the model can perform on unfamiliar data, boasted an impressive score of 88.60%. Such a high level of accuracy showcases the model’s exceptional capability in identifying age and gender in real-life situations. Moving forward, our goal is to enhance the system’s capabilities by expanding the range of classes for age and gender and improving the quality of the images. As is typical in machine learning studies, im- proving model quality is a critical and time-consuming process which is achieved by training the deep learning model with huge number of human face images.

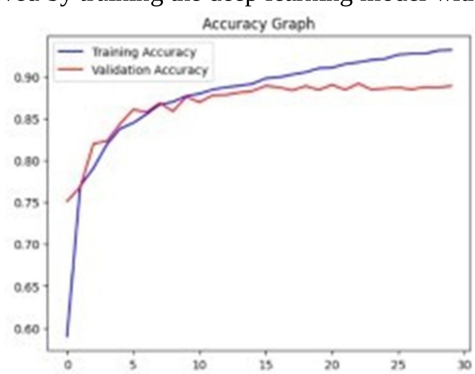


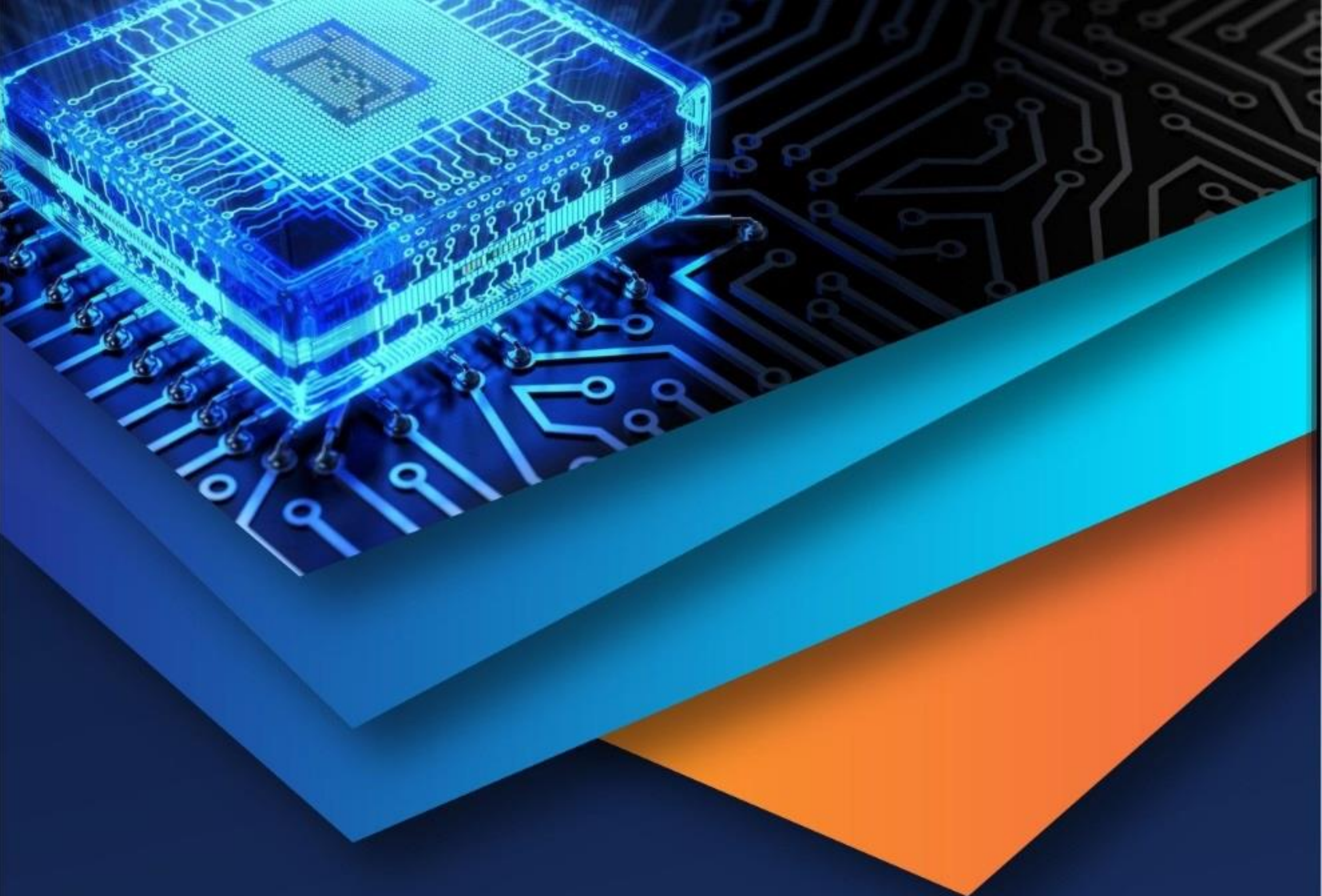
Fig. 8. Accuracy Graph

TABLE II
MINIMUM LOSS VALUE

Age Classification		Gender Classification	
Train	Validation	Train	Validation
0.165	0.374	0.194	0.238

REFERENCES

[1] Amit Dohme, Ranjit Kumar, and Vijay Bhan, “Gender Recognition Through Face Using Deep Learning”, International Conference on Compu- tational Intelligence and Data Science (ICCIDS 2018).
[2] Akash. B. N, Akshay. K Kulkarni, Deek- shith.A and Gowtham Gowda4, “Age and Gender Recognition using Convolution Neural Network”, IJESC, ISSN 2321 3361 Volume 10 Issue No.6.
[3] Anto A Micheal and R Shankar, “Automatic Age and Gender Estimation using Deep Learning and Extreme Learning Machine”, Turkish Journal of Computer and Mathematics Education Vol.12 No.14 (2021), 63- 73.
[4] Shubham Patil, Bhagyashree Patil and Ganesh Tartare, “Gender Recognition and Age Approx- imation using Deep Learning Techniques”, In- ternational Journal of Engineering Research & Technology (IJERT), Vol. 9 Issue 04, April-2020.
[5] SHUBHAM KUMAR TIWARI (1613112045), “AGE AND GENDER DETECTION”, GALGO-TIAS UNIVERSITY, Project Report of Capstone Project- 2.
[6] Sasikumar Gurumurthy, C. Ammu and B. Sreedevi, “Age Estimation and Gender Classifi- cation Based on Face Detection and Feature Ex- traction”, International Journal of Management & Information Technology, ISSN 2278-5612 Vol.4, No.1.
[7] Mahija Kante, Dr. Esther Sunandha Bandaru, Gadilid Manasa, Meghana Emandi and Varanasi Leela Lavanya, “Age and Gender Detection using OpenCV”, INTERNATIONAL JOURNAL OF ADVANCE RESEARCH, IDEAS, AND INNO- VATIONS IN TECHNOLOGY, (Volume 7, Issue 3 - V7I3-2163).
[8] X. Wang, R. Guo, and C. Kambhamettu, “Deeply- learned feature for age estimation,” in Proc. IEEE Winter Conf. Appl. Comput. Vision, 2015, pp. 534–541.
[9] Rothe, Rasmus, Radu Timofte, and Luc Van Gool. "Dex: Deep expectation of apparent age from a single image." Proceedings of the IEEE Inter- national Conference on Computer Vision Work- shops. 2015.
[10] Krizhevsky, Alex, Ilya Sutskever, and Geoffrey E. Hinton. "Imagenet classification with deep con- volutional neural networks." Advances in neural information processing systems. 2012.
[11] G. Guo, Y. Fu, T. S. Huang, and C. R. Dyer, "Locally Adjusted Robust Regression for Human Age Estimation," 2008 IEEE Work- shop on Applications of Computer Vision, Copper Mountain, CO, 2008, pp. 1-6, doi: 10.1109/WACV.2008.4544009.
[12] Angulu, R., Tapamo, J. R., & Adewumi, A. O. (2018). Age estimation via face images: A survey. EURASIP Journal on Image and Video Process- ing, 2018(1). doi:10.1186/s13640-018-0278-6
[13] Akhand, M. A., Sayim, M. I., Roy, S., & Sid- dique, N. (2020). Human Age Prediction from Facial Image Using Transfer Learning in Deep Convolutional Neural Networks. Proceedings of International Joint Conference on Computational Intelligence Algorithms for Intelligent Sys-tems, 217-229. doi:10.1007/978-981-15-3607-6_17
[14] Cao, Q., Shen, L., Xie, W., Parkhi, O. M., & Zisserman, A. (2018). VGGFace2: A Dataset for Recognising Faces across Pose and Age. 2018 13th IEEE International Conference on Auto- matic Face & GestureRecognition (FG 2018). doi:10.1109/fg.2018.00020
[15] H. Han, C. Otto, and A. K. Jain, “Age estima- tion from face images: Human vs. machine per- formance,” in Proc. Int. Conf. BTAS, Jun. 2013,
[16] pp. 1–8.
[17] S. Tamura, H. Kawai, and H. Mitsumoto, “Male/female identification from 8 × 6 very low resolution face images by neural network,” Pattern Recognition, vol. 29, no. 2, pp. 331–335, 1996.
[18] Y. Fu, G. Guo, and T. S. Huang, “Age synthesis and estimation via faces: a survey,” IEEE Transac- tions on Pattern Analysis & Machine Intelligence, vol. 32, pp. 1955–1976, 2010.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY