

A Review on Age and Gender Recognition using various datasets and deep learning models

Premy P Jacob
M.Tech, CSE Dept.
Mangalam College of Engineering
Ettumanoor, Kottayam, India

Dr. K. John Peter
Prof., CSE Dept.
Mangalam College of Engineering
Ettumanoor, Kottayam, India

Abstract—Age and gender, two facial attributes play a major role in the society. An automatic age and gender recognition have a vast number of real-world applications includes a customer service, the priority voting system, medical diagnosis, the human computer interaction. Deep learning techniques are commonly used in most researches and achieved to improve the performance too. Implementing different deep learning models and evaluating the improvement in accuracy leads to further researches. The main purpose of paper is, to conduct a detailed examination of age and gender recognition through various datasets and deep learning models. In this paper, explain the progress made by, highlighting the contributions addressed, the models and dataset used and evaluating the suggested approach with the results obtained.

Keywords- Age and gender classification, Deep learning, Deep neural network, Datasets.

I. INTRODUCTION

Age and gender classification is the task of identifying a person's age and gender from an image or video. Age and gender play a major role in identifying a person. With the advent of social media, there is growing interest in automatic classification of age and gender by facial images. Age and gender are the most basic facial qualities in social interaction. Human face contains characteristics that determine identity, age, gender, emotions, ethnicity. Therefore, the process of age and gender assessment is an important step for many applications. Some real-world applications are visual surveillance, electronic customer, crowd behavior analysis, online advertisement, item recommendation, law enforcement, prevent juveniles from purchasing banned drugs from shops, prevent children's from browsing harmful websites, forensic, anti-aging treatment, beauty products production, movie role casting etc. With the help of human eye, it is difficult to estimate age because from the middle age to old age, the facial features become key time-varying due to skin transformation. In adolescent, due to growth. Age and gender identification becomes a common open challenge for researchers because of some common problems.

So, computer vision steps forward to solve all of these problems. Computer vision is an artificial intelligence (AI) that enables computers and systems to derive meaningful information from digital images, videos, and other visual inputs, and to take action and make recommendations based on that information. So, need of an efficient model for age and gender estimation tasks is very necessary.

In most of the past researches, individually designed features with the models. Machine learning models are popularly used. Machine learning is the study of computer algorithms that can be improved automatically through the use of experience and data. These machine learning models i.e., individually designed features behave deficient on some datasets, and it fails to produce expected results. In some datasets it produces better results too. Due to inadequacy on some dataset, researches to make progress into convolutional neural network based deep learning models. In deep learning train model with large amounts of data and complex algorithms, and automatically extract features from facial images. Some neural network based deep learning models are: VGG16, VGG19, Xception, ResNet50V2, ResNet101V2, InceptionV3, Inception, ResNetV2, MobileNet, MobileNetV2, DenseNet121, DenseNet169, DenseNet201, NASNetMobile, EfficientNetB0, EfficientNetB1, EfficientNetB2, EfficientNetB3, EfficientNetB4 etc.

Deep learning simply a subset of machine learning. While comparing with simple machine learning concepts deep learning works like to imitate how humans think and learn. Deep learning makes the processes easier and faster. Facial age and gender recognition is one of the greatest applications of deep learning. CNN (Convolutional neural network) simply has one or more convolutional layers and has some specific functions too. It some great functions that it automatically detects important features, pre-processing is less dependent, need of human effort is less without any lose on quality, and it also reduces the number of parameters.

Some of the common drawbacks of age and gender recognition are:(a) Dataset problem that affect the performance, machine learning depends on quality of data. Mislabelled data excessive noise can cause the models to start learning wrong things.(b) Traditional Classification algorithms can't learn the complicated nonlinear relationship in image data.(c) Deep Neural Network extract features in images is not much efficient and accurate i.e. particular model extract only one type of feature.(d) Minor change in alignment which affects the performance.(e) Problem of misclassification(f) Problem of occlusion, pose, illumination, resolution, facial expression etc. in images.

The rest of this study is organized as follows. Section II literature survey on Age and gender recognition approaches. Section III provides details of the previous research models performances and its comparison. The conclusions are described in Section IV.

II. LITERATURE SURVEY

In 2017, [2] k. Zang, et al. proposed Residual network of Residual network (ROR) for automatic prediction of age and gender from face images of unconstrained condition. The architecture is for high resolution facial images age and gender classification. Two mechanisms such as pretrained by gender and training with weighted loss layer, and used to improve the performance of age estimation. In order to future improve the performance and alleviate overfitting problem, ROR is pretrained on ImageNet, then it is fine-tuned on IMBD-WIKI 101dataset for further learning the features of face images and finally tuned on Adience dataset. High accuracy is achieved for gender classification task works well for high resolution facial images. Age estimation accuracy of 67.37% and gender estimation accuracy of 93.27%. Lower age estimation accuracy and ROR model is slower than other models make this challenging. Accuracy of age estimation sometimes affect due to minor change in alignment. Issue of dataset i.e., Adience. Works well on some specific features only.

In 2018,[3] Philip smith, et al., transfer learning is employed to tackle the issue of recognizing a person’s age and gender from an image using deep CNNs. Transfer learning to use VGG19 and VGGFace pretrained models are used to increase the efficiency. Training techniques such as input standardization, data augmentation, label distribution age encoding is compared. Dataset used is MORPIL.VGGFace produce better result than VGG19. VGGFace takes far fewer epochs to fit the training data VGGFace gender prediction accuracy of 98.68% of 4.1 years. Age recognition produces MAE of 4.1 years due to female characteristics. Gender prediction is largely based on the absence and presence of long hair, tilt of head. Larger and more dataset is necessary. Mislabelled data and excessive noise can cause model to start learning wrong things. The major problem caused by this research is due to dataset i.e., MORP II. In this dataset few noisy variations: heads are tilted in different direction. Most images of age ranges from 16 to 77. Most of them are male images and black colored.

In 2019, [4] Ningning Yu, et al. proposed an ensemble learning used for facial age estimation within non-ideal facial imagery in Fig[1] . The method consists of mainly image preprocessing, feature extraction, and age predication. Separately, the input face image is preprocessed in RGB Stream, Luminance Modified Stream, and YIQ Stream. Three different pretrained DCNNs equipped with softmax are wont to implement feature extraction and age estimation because the weak classifiers. Finally, the ensemble learning module fuses the three weak classifiers to get a more accurate estimation. Dataset used is IMBD-WIKI from Wikipedia. Three stream method improves the performance. To generate more accurate classification, ensemble learning to fuse the weak classifiers. To evaluate the performance of classification, some evaluation indexes such as AEM (Exact match) and AEO (An error one age category). With ensemble method, AEM of 45.57% and AEO of 88.20%. Need an efficient global search method is necessary. Require an algorithm based on dataset which cover more data. DCNN requires large time for feature extraction and training. The performance is affected by issues of dataset where IMBD-WIKI, is affected by ambient illumination and complex background situation, which increases difficulty of estimation.

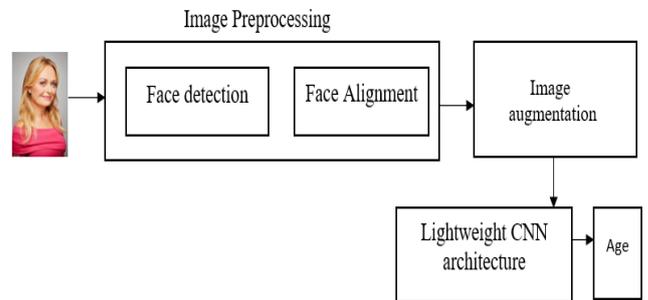


Fig.1. Block diagram for Lightweight CNN Age estimation, Olatunbosun, et al. [5].

In 2020, [5] Olatunbosun, et al. proposed a Lightweight Convolutional neural network for real and apparent age estimation of human faces in Fig [1]. Real and apparent age estimation has numerous real-world applications such as medical diagnosis, forensic, facial beauty product production. CNN model is larger, more complex, too large network parameters and layers, training time is long, huge training dataset which increases computation cost and storage overhead. So, proposed to design a lightweight CNN layer of fewer layers to estimate real and apparent age. Input is real-world face image. First step is, image preprocessing i.e., face detection and alignment. Then, followed by image augmentation where random scaling, random horizontal flipping, color channel shifting, standard color jittering, random rotation and also generate an alter copies of every training image. Next step, is estimate real and apparent age using light CNN model. FG-NET, MORP-II, APPA-REAL are the datasets used. FGNET produces MAE of 3.05, MORP II produce MAE of 2.31, APPA-REAL produce MAE of 4.94. Model with MORP II dataset improves the performance. Light CNN has low training time. Need for more robust and quality image processing algorithm because it detects the unfiltered image faster. Lighter CNN with few parameters is necessary. Non-frontal face image reduces the problem of age classification of unfiltered face. Issue of dataset is very challenging, and noisy variation of these datasets affect the performance.

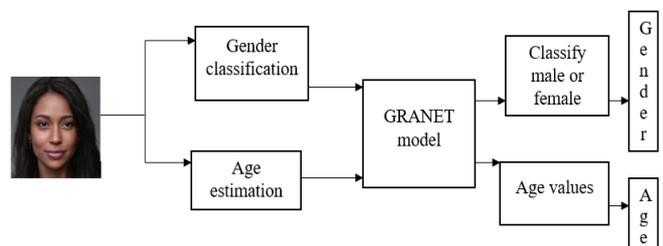


Fig.2. Block diagram of GRANET model, Avishek Garain, et al. [1]

In 2021 [1] Avishek Garain, et al. proposed a model GRANET (Gated Residual Attention Network) for classification of age and gender from facial images in Fig [2]. Some common shortcomings of previous researches are: higher MAE, lower age estimation accuracy, Gender classification accuracy is not at all high, Performance is impacted by minor change in alignment, some model works well on high resolution images but others not. So suggested multiple attention blocks are gated together to form the model. FG-NET, AFAD, Wikipedia, UTKFace, AdienceDB are the dataset used for this model. Among other datasets UTKFace performs better and produce gender recognition accuracy of 99.2% and age recognition accuracy of 93.7%. AFAD produce age estimation result of 3.10 MAE.FGNET produce 3.23 MAE.Wikipedia Age of 5.45 MAE, UTKFace of 1.07 MAE, AdienceDB of 10.57 MAE. Adience produce age estimation result of 65.1% and gender classification of 81.4%. There exist some drawbacks too for this research: Images of kids is very difficult to classify. Problem of misclassification, Occurrence of wrong prediction, Need more intelligence when images are obstructive partially viewed. Issue of dataset is the main concern which affect the performance. UTKFace dataset better because it overcomes the variability in illumination, pose, occlusion, resolution.

A. AGE AND GENDER RECOGNITION TECHNOLOGY

Deep Neural Network and Transfer learning: Neural network structures like human brain, and it contains nodes. It consists of input layer, hidden layer and output. Images as input and multiple hidden layers the node multiply the inputs with random weights, calculate them and pass to output layer. Some major algorithms used in deep learning are: Convolutional Neural Networks (CNNs), Long Short-Term Memory Networks (LSTMs), Recurrent Neural Networks (RNNs). Convolutional Neural Networks (CNNs) have multiple layers that process and extract features from data. CNN has a convolution layer that has several filters to perform the convolution operation. Next, it has a ReLU layer to perform operations on elements. Then, it has pooled layer where pooling is down-sampling operation that reduces the dimensions of the feature map. Last, fully connected layer forms when the flattened matrix from the pooling layer is fed as an input, which classifies and identifies the images. Long Short-Term Memory Networks (LSTMs) are a type of Recurrent Neural Network (RNN) that can learn and memorize long-term dependent features. Recurrent Neural Networks (RNNs) which allow the outputs from the LSTM to be fed as inputs to the current phase. Transfer Learning is machine learning method where the major task is reusing a pre-trained model as the starting point for a model on a new task. Some high accuracy transfer learning models are VGG, Inception, Xception, ResNet.

B. DATASETS USED IN AGE AND GENDER RECOGNITION

The training testing of model is very important factor in deep learning. Datasets play a major role in training and testing. The availability of different age and gender recognition datasets are great support for researchers. Each of the datasets has different features too.

IMBD-WIKI DATASET: The largest face dataset with gender, a name and age information. It contains 500 thousand images of faces. In total 460,723 face images from 20,284 celebrities from IMBD and 62,328 from Wikipedia, thus total of 523,051. Some problems of the dataset are: all images are different size, some ages are invalid, there are more male faces than female faces.

MORP II DATASET: A facial age estimation dataset which contains 55,134 facial images of 13,617 subjects ranging from 16 to 77 years old. Images of 84.6% are males and 77.2% are of black.

IMAGENET: An image database organized according to WorldNet hierarchy. ImageNet contains more than 20,000 categories such as a strawberry, a balloon and several other objects.

FG-NET AGING DATASET: The dataset consists of 1002 images of 82 different subjects with their ages varying from a baby to 69 years old. Images from the photographs of the personal collections. Some challenges are: the quality of images, the quality of photographic paper, affect some Variability in the quality, an illumination, the resolution, the viewpoint, an expression, presence of an occlusion in the form of facial hairs, spectacles, hats.

AFAD DATASET: The Asian Face Age Dataset for evaluating the performance of various age estimation. It contains more than 160K facial images along with their corresponding age labels. The dataset has been designed for an age estimation on Asian faces. There are labeled prints in the AFAD dataset with the ages varying from 15 to 40. The AFAD dataset was constructed by collecting shots of users from a particular social network.

WIKIPEDIA AGE DATASET: The publicly available dataset contains facial images. Images of various celebrities are available in this dataset. The images which were the date when the photo was taken was removed as they will not have any age information in them. In total, 62,328 face images from 20,284 celebrities were obtained from Wikipedia.

UTKFace DATASET: The large-scale dataset which contains face images with a very long span range from 0 to 116. It consists of more than the 20k face images. It contains labels of a face age, gender and ethnicity. The images cover huge variation in illumination, pose, occlusion, facial expression, resolution etc. The dataset contributes advantage over face detection, age estimation to age progression or regression and landmark localization.



Fig.3.Sample images of UTKFace dataset.

ADIENCE DATASET: The Adience dataset consists of pics taken through a digital digicam from a phone or tablets. The images of the dataset seize severe variations, together with a severe blur (low-resolution), occlusions, out-of-plane, pose variations, expressions.

BEFA: Bias Estimation in face analytics (BEFA) contains 13431 test images where illustrates age (baby, child, teenagers, young, adults, middle age and senior), gender (male and female) and ethnicity (white, black, Asian, Indian).

III. COMPARISON AND DISCUSSION OF PREVIOUS RESEARCHES

Table 1: Comparative study of age and gender recognition based on previous researches.

Model	Datasets	Advantages	Disadvantages
CNN [13]	Adience dataset	High accuracy in gender estimation task i.e,86.8%. Reducing the chances of overfitting because reduced number of parameters.	Lower age recognition accuracy i.e,50.7% due to its simple design.
ROR [12]	IMBD-WIKI dataset, ImageNet dataset.	High accuracy achievement in gender classification i.e,93.24%. Works well on high resolution images	Lower age detection accuracy i.e,67.34%.
CNN [11]	WEAFD dataset	Gender classification accuracy is quite good i.e,88%. Labelled face images for classification.	Age detection accuracy is very low i.e,38%.
MTCNN [7]	UTKFace dataset, BEFA dataset	Produces high gender classification accuracy i.e., for UTKFace dataset is 98.23%, and for BEFA dataset it is 93.72%.	Due to limited amount of facial attributes. Lower accuracy of the age classification task for UTKFace dataset is 70.1% and for the BEFA dataset is 71.83%.
VGG19, VGGFace [3]	MORP-II dataset	VGGFace is better. High gender classification accuracy i.e,98.7%.	Age estimation produces higher MAE (error) of 4.1 years. Trivial change affects the prediction task.

CNN [10]	Adience dataset	Produces better gender classification accuracy i.e,88%. The model focusses on useful and essential features.	Lower age detection accuracy i.e,61%.
LMTCNN [9]	Adience dataset	Gender classification accuracy is good i.e,85%.	Lower age detection accuracy i.e,44%. Larger size does not work well with unconstrained faces.
CNN [8]	UTKFace dataset	High accuracy for gender classification i.e,94.1%.	Age estimation produces higher MAE (error) of 5.44 years.
GRANET [1]	FG-NET dataset, AFAD dataset, Wikipedia dataset, UTKFace dataset, AdienceDB dataset.	UTKFace performs better and produce gender recognition accuracy of 99.2%. Produce less age estimation MAE of 1.07.	Produces age recognition accuracy of 93.7%. Problem of misclassification.

Currently there are many existing research models for recognizing the age and the gender. While comparing with some early researches performance of gender recognition is quite good, but the age estimation is not well. While comparing some researches model using UTKFace Dataset, Facenet model produce gender recognition accuracy of 91.2% and age estimation accuracy of 56.9%. Next Finetuned Facenet (FFNet) produce gender recognition accuracy of 96.1% and age estimation accuracy of 64%. Then [7] Multitask cascaded convolutional neural networks (MTCNN) produces accuracy rates on gender and age is 98.23% and 70.1%. Then [6] Residual attention network (RAN) model produce accuracy on gender is 97.5% and age estimation accuracy is 85.4%. RAN model on AFAD dataset age estimation MAE of 3.42 and FG-NET dataset age estimation MAE of 4.05. In the latest research, using [1] GRANET (Gated Residual Attention Network) model on five publicly available datasets. MAE over FG-NET, AFAD, Wikipedia, UTKFace, Adience are 3.23,3.10,5.45,1.07,10.57 respectively. UTKFace produces better gender recognition accuracy of 99.2% and age estimation accuracy of 93.7%. Table 1 illustrates the comparative study of previous researches.

From the above researches proves that deep learning models are better and produces effective performances. Finally, approaches an assumption from among all the datasets that UTKFace is better one. UTKFace dataset covers large illumination, pose, occlusion, facial expression, resolution etc.

IV. CONCLUSION

The facial age and gender recognition is a complex task. But it was very important to the society where, it has many real-world applications. Age and gender recognition with individually designed features and the machine learning fail. In the traditional classification technique cannot learn non-linear relationship in images. So, the advent of deep learning models is very significant. Deep learning models performed better from previous researches and also improves the performance. In gender classification tasks most of the researches performs better and achieves good results. But an improvement is very necessary in the case of an age estimation. The single model extracts only one type of features, and it strongly affects the performance. So, need to introduce an ensembling technique using various deep learning models to improve the performance on age and gender recognition in future.

REFERENCES

- [1] A Garain, B Ray, "GRANet: A Deep Learning Model for Classification of Age and Gender From Facial Images" *IEEE Access*, vol.9, pp.85672-85689, 2021.
- [2] K. Zhang, C. Gao, L. Guo, M. Sun, X. Yuan, T. X. Han, Z. Zhao, and B. Li, "Age group and gender estimation in the wild with deep RoR architecture," *IEEE Access*, vol. 5, pp.22492-22503, 2017.
- [3] P. Smith and C. Chen, "Transfer learning with deep CNNs for gender recognition and age estimation," in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Dec. 2018, pp. 2564-2571.
- [4] Ningning yu, l. qian, y. huang, and yuan wu, "Ensemble Learning for Facial Age Estimation Within Non-Ideal Imagery" *Facial, IEEE Access*, vo.7, pp.97938-97948, 2019.
- [5] Olatunbosun agbo-ajala and Serestina viriri "A Lightweight Convolutional Neural Network for Real and Apparent Age Estimation in Unconstrained Face Images", *IEEE Access*, vol.8, pp.162800-162808, 2020.
- [6] F. Wang, M. Jiang, C. Qian, S. Yang, C. Li, H. Zhang, X. Wang, and X. Tang "Residual attention network for image classification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 3156-3164.
- [7] A. Das, A. Dantcheva, and F. Bremond, "Mitigating bias in gender, age, and ethnicity classification: A multi-task convolution neural network approach," in *Proc. Eur. Conf. Comput. Vis. (ECCV) Workshops*, 2018, pp. 1-13.
- [8] A.V. Savchenko, "Efficient facial representations for age, gender and identity recognition in organizing photo albums using multi-output ConvNet," *PeerJ Comput. Sci.*, vol. 5, p. e197, Jun. 2019, doi: [10.7717/peerj-cs.197](https://doi.org/10.7717/peerj-cs.197).
- [9] J. H. Lee, Y. M. Chan, T. Y. Chen, and C. S. Chen, "Joint estimation of age and gender from unconstrained face images using lightweight multitask CNN for mobile applications," in *Proc. IEEE Conf. Multimedia Inf. Process. Retr. (MIPR)*, Apr. 2018, Art. no. 17877533.
- [10] S. Hosseini, S. H. Lee, H. J. Kwon, H. I. Koo, and N. I. Cho, "Age and gender classification using wide convolutional neural network and Gabor filter," in *Proc. Int. Workshop Adv. Image Technol.*, 2018, pp. 1-3, doi:[10.1109/IWAIT.2018.8369721](https://doi.org/10.1109/IWAIT.2018.8369721).
- [11] N. Srinivas, H. Atwal, D. C. Rose, G. Mahalingam, K. Ricanek, and D. S. Bolme, "Age, gender, and _ne-grained ethnicity prediction using convolutional neural networks for the East Asian face dataset," in *Proc. 12th IEEE Int. Conf. Autom. Face Gesture Recognit. (FG)*, May 2017, pp. 953-960, doi: [10.1109/FG.2017.118](https://doi.org/10.1109/FG.2017.118).
- [12] K. Zhang, C. Gao, L. Guo, M. Sun, X. Yuan, T. X. Han, Z. Zhao, and B. Li, "Age group and gender estimation in the wild with deep RoR architecture," *IEEE Access*, vol. 5, pp.22492-22503, 2017, doi:[10.1109/ACCESS.2017.2761849](https://doi.org/10.1109/ACCESS.2017.2761849).
- [13] G. Levi and T. Hassner, "Age and gender classification using convolutional neural networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2015, pp. 34-42, doi: [10.1109/CVPRW.2015.7301352](https://doi.org/10.1109/CVPRW.2015.7301352).