SEMINAR REPORT

ON

# GRA_Net: A Deep Learning Model for Classification of Age and Gender from Facial Images

Submitted by

**VIPINDAS K**

**(KSD18CS089)**

*To the APJ Abdul Kalam Technological University in partial fulfillment of the requirements for the award of the degree of*

**BACHELOR OF TECHNOLOGY**

*In*

**COMPUTER SCIENCE AND ENGINEERING**

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

**LBS COLLEGE OF ENGINEERIING**

**KASARAGOD – 671542, KERALA**

JANUARY 2022

# DECLARATION

*“I hereby declare that this submission is my own work and that, to the best of my knowledge and belief, it contains no material previously published or written by another person nor material which has been accepted for the award of any other degree or diploma of this Institute or other Institute of higher learning, except where due acknowledgement has been made in the text.”*

| | |
|---|---|
| PLACE: KASARAGOD | NAME: VIPINDAS K |
| DATE: 12/01/2022 | REG NO: KSD18CS089 |

# CERTIFICATE

*This is to certify that the seminar report entitled* **— "GRA_Net: A Deep Learning Model for Classification of Age and Gender from Facial Images"** *submitted by* **VIPINDAS K** *to the APJ Abdul Kalam Technological University in partial fulfillment of the requirements for the award of the degree of* **Bachelor of Technology in Computer Science and Engineering** *is a bonafide record of the work done by him under my supervision and guidance*.

| | | |
|---|---|---|
| **Mr. Sarith Divakar M** | **Mr. Sudeesh KP** | **Mrs. Smitha Mol MB** |
| Assistant Professor | Assistant Professor | Department of CSE |
| Department of CSE | Department of CSE | (Head of Department) |
| (Guide) | (Seminar Coordinator) | |

Place: Kasaragod
Date: 12/01/2022

# ACKNOWLEDGEMENT

It is a really a momentous opportunity and privilege to express my deep sense of Gratitude to all those who have us to accomplish this task. I express humble God Almighty for his incessant blessing on us during this seminar.

I have taken efforts in this seminar. However, it would not have been possible without the kind support and help of many individuals and organizations. I would like to extend its sincere to all of them.

I sincerely thank its principal **Dr. MOHAMMED SHEKOOR T** for providing me facilities in order to go ahead with its seminar development. I express its sincere gratitude to head of the department **Dr. SMITHAMOL M B** and I also express heartiest gratitude to seminar coordinator **Mr. SUDEESH KP** in computer science and engineering for their valuable advice and guidance. I would always oblige for the helping hands of all other staff members of the department and all my friends and well-wishers who directly or indirectly contributed to this venture.

Last but not least, I am indebted to God almighty for being the guiding light throughout this work and for helping me to complete the same within the stipulated time.

**VIPINDAS K**

# ABSTRACT

The problem of gender and age identification has been addressed by many researchers, however, the attention given to it compared to the other related problems of face recognition in particular is far less. Any language in the world has a separate set of words and grammatical rules when addressing people of different ages. The decision associated with its usage, relies on its ability to demarcate these individual characteristics like gender and age from the facial appearances at one glance. With the rapid usage of Artificial Intelligence (AI) based systems in different fields, we expect that such decision making capability of these systems match as much as to the human capability.

To this end, in this work, we have designed a deep learning based model, called GRA_Net (Gated Residual Attention Network), for the prediction of age and gender from the facial images. This is a modified and improved version of Residual Attention Network where we have included the concept of Gate in the architecture. Gender identification is a binary classification problem whereas prediction of age is a regression problem. We have decomposed this regression problem into a combination of classification and regression problems for achieving better accuracy.

Experiments have been done on five publicly available standard datasets namely

FG-Net, Wikipedia, AFAD, UTKFAce and AdienceDB. Obtained results have proven its effectiveness for both age and gender classification, thus making it a proper candidate for the same against any other stateof-the-art methods.

# CONTENTS

# List of Figures

## List of Abbreviations

| | |
|---|---|
| CNN | : Convolutional Neural Network |
| GRA_Net | : Gated Residual Attention Network |
| FG-NET | : Face and Gesture Recognition Network |
| LSTM | : Long Short-Term Memory |
| GRU | : Gated Residual Unit |

# 1. INTRODUCTION

Age identification and gender classification play a pivotal role in its social lives. Every language in the world reserves different salutations for men and women, and very often different vocabularies are used when addressing elders compared to young people. These customs are largely dependent on one's ability to estimate these individual traits of a person: age and gender, which are obtained from the facial appearances. A vast number of application developers, especially after the growth in social media and social networks, are indulging themselves with automatic age identification. Age and gender are the most fundamental facial qualities in social interaction. Human's face contains features that determine identity, age, gender, emotions, and the ethnicity of people. Among these features, age and gender identification can be especially helpful in several real-world applications including visual surveillance, medical diagnosis (premature facial aging), human-computer interaction system, access control or soft biometrics, demographic information collection, law enforcement, marketing intelligence, etc. it is necessary to identify the age and the gender from the facial images of the human beings. However, several problems in age and gender identification are still considered as open challenges to the researchers. Despite the progress, the computer vision community keeps making continuous improvement by introducing the new techniques that advance the state-of-the-art, age and gender predictions from the unfiltered real-life facial images are yet to meet the need of the commercial and real-world applications. Thus a robust and accurate method for the age and the gender identification tasks becomes an absolute necessity. In the past few years, in order to intensify the ability to identify these attributes from facial images, many methods have been put forward. Previous researchers approached the task of estimating age or classifying gender from face images using individually designed

feature vectors with statistical models and machine learning based models. However, even though a lot of these models have been designed, the individually designed features behave inadequately on the benchmark datasets having unconstrained images. Hence, in the more recent years, researchers have been started exploring the domain of convolutional neural network (CNN) based deep learning architectures for the task of age and gender prediction, which help in automatic feature extraction from the input images. To improve such feature extraction ability, use of the long short-term memory (LSTM) units inserted between Residual Networks (ResNets) has also been found. In this paper, we have proposed an architecture which harnesses the capabilities of both classification and regression to solve the tasks of age estimation and gender prediction from the facial images. It is having a backbone support of Residual Attention Network modified with the addition of a new parameter called ’Gate’ similar to the concept of gates in Gated Residual Units (GRUs), the only difference being the use of the ’Gate’ on the higher level of components rather than applying it on individual units of the architecture. Impressive results obtained on various standard datasets confirm the credibility and precision of the architecture.

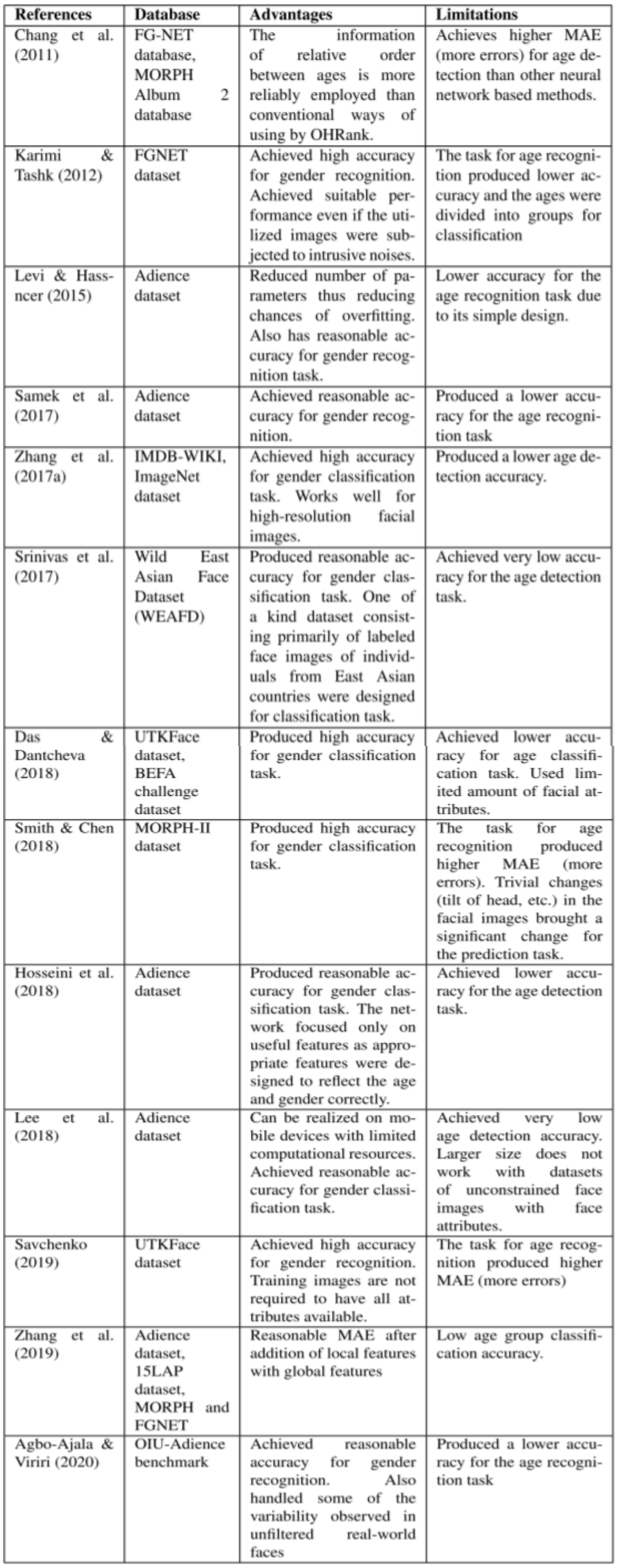
## 2. LITERATURE SURVEY

In the past few years, there have been several attempts to estimate the individual traits: age and gender from the facial images. A variety of features consisting both of facial shape and textures have been used by the researchers for the estimation of age and gender from facial images. Some of those features are: mean pixel values of channels, energy and entropy of filtered images, Histogram of Oriented Gradients (HOG) etc. Simple features such as density of edges obtained from an image using an edge detector have also been applied. For the FGNET dataset, the method achieves an accuracy of 95% for the gender recognition task and an accuracy of 79.31% for class wise prediction of ages i.e., groups of 1 to 12 years, 13 to 40 years and 41 to 80 years. A supervised appearance model (sAM) that improves on active appearance model (AAM) by replacing PCA with partial least-squares regression. The sAM model is used as a feature extractor for age and gender estimation from the facial images. As a viable alternative, thus, researchers have started applying various deep learning models for the said tasks. In the recent years, deep learning based models have shown reassuring performance in the field of age and gender identification, especially on unfiltered face images. Recently, methods such as simple CNN, MobileNet CNN, LSTM of recurrent neural network (RNN) architectures have been used by various researchers to estimate age and gender from facial images. A simple CNN based architecture that could be used for limited amount of learning data. The network consists of only three convolutional layers and two fully-connected layers with a very less number of neurons. The proposed method achieves highest accuracy of 86.8% for gender estimation task and highest accuracy of 50.7% for exact age estimation task on Adience dataset. Results of two CNN architectures are given, showing potential design considerations that promotes region based feature extraction thus optimizing the network. Age and gender accuracies of 38% and 88% respectively are achieved by the method on Wild East Asian Face Dataset (WEAFD). The WEAFD is a new and unique dataset consisting mainly of labeled facial images of individuals from East

Asian countries. A hierarchy of deep CNNs is tested, which classifies subjects on the basis of gender. A gender recognition accuracy of 98.7% and an MAE of 4.1 years are achieved by the proposed method on MORPH-II dataset. For the CNN network, Gabor filter responses are used as inputs. back-propagation for an end-to-end architecture have been used in order to learn the weighting of Gabor-filter responses. The proposed method achieves age and gender estimation accuracies of 61% and 88% respectively. The LMTCNN uses depth-wise separable convolution to reduce the model size and save the inference time. The method achieves age and gender recognition accuracies of 44% and 85% respectively on Adience dataset. Reference [6] has proposed a two-stage approach, where at first, the CNN predicts age and gender and also extracts facial representations suitable for face identification by using a modified MobileNet. At the second stage, the extracted facial representations are grouped using hierarchical agglomerative clustering technique. The proposed method achieves 94.1% gender recognition accuracy, and 5.44 MAE on UTKFace dataset. In the same year,  have proposed a Multi-Task CNN (MTCNN) with joint dynamic loss weight adjustment towards classification of age and gender from the facial images. The mean classification accuracy of the gender classification task for UTKFace dataset is 98.23%, and for BEFA challenge dataset it is 93.72%. The accuracy of the age classification task for UTKFace dataset is 70.1% and for the BEFA challenge dataset is 71.83%.

# 3. DISCUSSION ON THE PAST RESEARCHES

Comparative study of some age and gender prediction methods proposed in last few years is shown in FIGURE 1. In this project tried to find out the strengths and weaknesses of past works done in this domain over the years, and harnessing the newer resitsces and algorithms



| References | Database | Advantages | Limitations |
|---|---|---|---|
| Chang et al. (2011) | FG-NET database, MORPH Album 2 database | The information of relative order between ages is more reliably employed than conventional ways of using by OHRank. | Achieves higher MAE (more errors) for age detection than other neural network based methods. |
| Karimi & Tashk (2012) | FGNET dataset | Achieved high accuracy for gender recognition. Achieved suitable performance even if the utilized images were subjected to intrusive noises. | The task for age recognition produced lower accuracy and the ages were divided into groups for classification |
| Levi & Hassncer (2015) | Adience dataset | Reduced number of parameters thus reducing chances of overfitting. Also has reasonable accuracy for gender recognition task. | Lower accuracy for the age recognition task due to its simple design. |
| Samek et al. (2017) | Adience dataset | Achieved reasonable accuracy for gender recognition. | Produced a lower accuracy for the age recognition task |
| Zhang et al. (2017a) | IMDB-WIKI, ImageNet dataset | Achieved high accuracy for gender classification task. Works well for high-resolution facial images. | Produced a lower age detection accuracy. |
| Srinivas et al. (2017) | Wild East Asian Face Dataset (WEAFD) | Produced reasonable accuracy for gender classification task. One of a kind dataset consisting primarily of labeled face images of individuals from East Asian countries were designed for classification task. | Achieved very low accuracy for the age detection task. |
| Das & Dantcheva (2018) | UTKFace dataset, BEFA challenge dataset | Produced high accuracy for gender classification task. | Achieved lower accuracy for age classification task. Used limited amount of facial attributes. |
| Smith & Chen (2018) | MORPH-II dataset | Produced high accuracy for gender classification task. | The task for age recognition produced higher MAE (more errors). Trivial changes (tilt of head, etc.) in the facial images brought a significant change for the prediction task. |
| Hosseini et al. (2018) | Adience dataset | Produced reasonable accuracy for gender classification task. The network focused only on useful features as appropriate features were designed to reflect the age and gender correctly. | Achieved lower accuracy for the age detection task. |
| Lee et al. (2018) | Adience dataset | Can be realized on mobile devices with limited computational resources. Achieved reasonable accuracy for gender classification task. | Achieved very low age detection accuracy. Larger size does not work with datasets of unconstrained face images with face attributes. |
| Savchenko (2019) | UTKFace dataset | Achieved high accuracy for gender recognition. Training images are not required to have all attributes available. | The task for age recognition produced higher MAE (more errors) |
| Zhang et al. (2019) | Adience dataset, 15LAP dataset, MORPH and FGNET | Reasonable MAE after addition of local features with global features | Low age group classification accuracy. |
| Agbo-Ajala & Viriri (2020) | OIU-Adience benchmark | Achieved reasonable accuracy for gender recognition. Also handled some of the variability observed in unfiltered real-world faces | Produced a lower accuracy for the age recognition task |

**Fig 1**

# 4. MOTIVATION AND CONTRIBUTIONS

From the above discussion and FIGURE 1, it is clear that the previous methods have a common shortcoming of higher MAE and lower accuracy mainly for the task of age estimation. The accuracy of the gender classification of the methods is also not as high as expected to bridge the gap between the human level and the machine level errors. Minor changes in alignment of the face in the images degrade the performance for some of the methods like the work proposed. However, the method worked quite well in classifying the gender from the images. Some methods work well on higher resolution images while some other methods have more reliably used the information of relative order between ages for the purpose of correct age identification. Some of the works are computationally efficient enough to be employed on mobile devices with limited computational resitsces. This makes the method very useful in practical scenarios. However, the most important requirement of such method is the precision level which can match to the human's ability. Keeping in mind the strengths and weaknesses of the previous works, contributions in this work:

1) An architecture harnessing the capabilities of classification and regression for Age identification purposes.

2) Same architecture capable of performing the separate task of gender classification thus ensuring the model's versatility.

3) Introduction of the new concept of Gates for Residual Attention Network used as a backbone of the architecture.

4) Handling the poor performance caused by minor changes in facial orientation by applying attention masks through various channels covering as many combinations as possible.

5) Evaluated on 5 datasets having images of people belonging to different ethnic groups and various background.

6) Achieved lower MAE and higher identification accuracy for age and impressive performance in gender classification. The overall workflow of its architecture is shown in Figure 2.

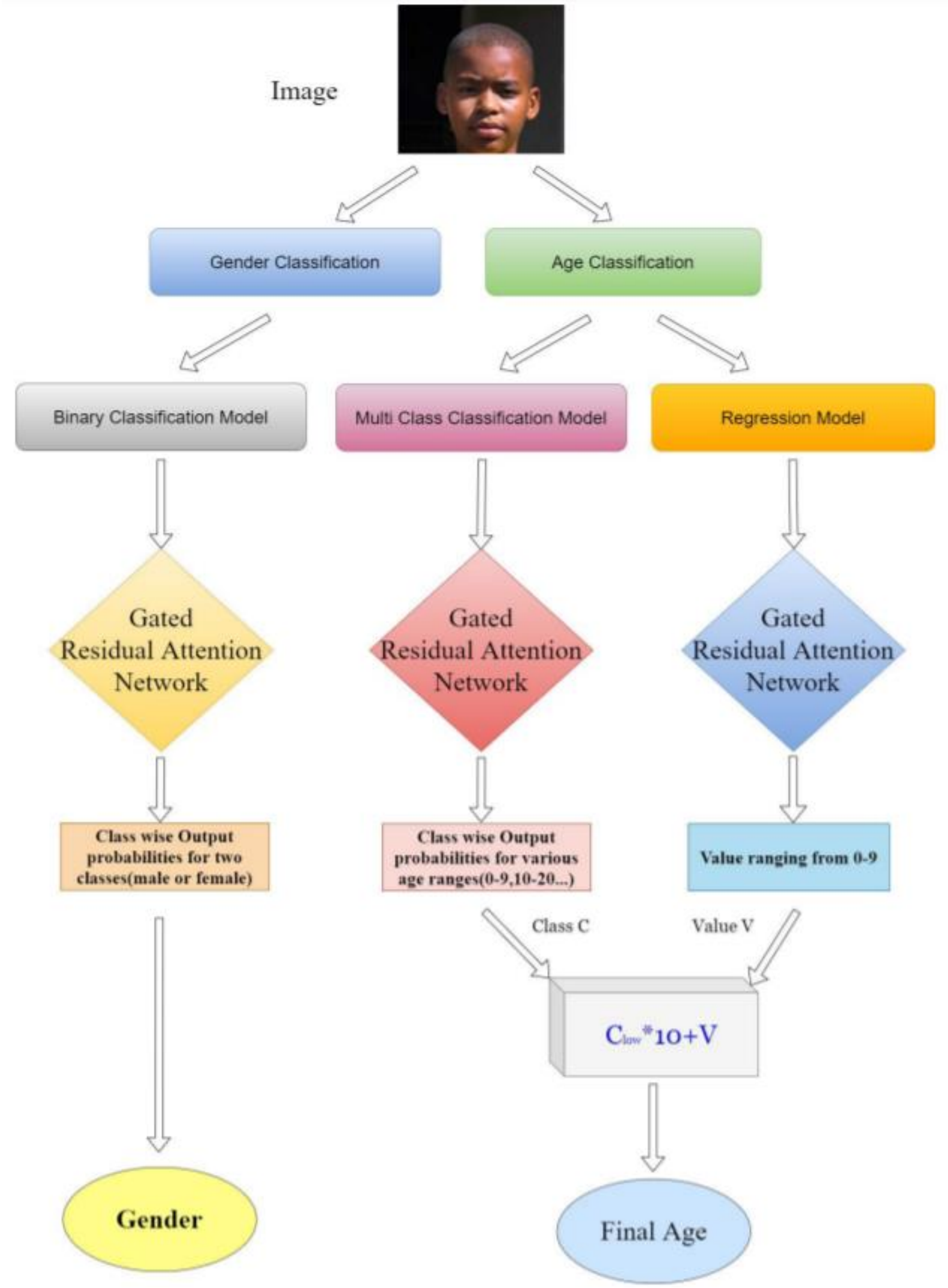


Image
Gender Classification
Age Classification
Binary Classification Model
Multi Class Classification Model
Regression Model
Gated Residual Attention Network
Gated Residual Attention Network
Gated Residual Attention Network
Class wise Output probabilities for two classes(male or female)
Class wise Output probabilities for various age ranges(0-9,10-20...)
Value ranging from 0-9
Class C
Value V
$C_{low}*10+V$
Gender
Final Age

**Fig 2**

# 5. DATASETS USED IN THE PRESENT WORK

## 5.1. FG-NET AGING DATASET

The FG-NET Aging dataset was developed as part of the project FG-NET (Face and Gesture Recognition Network). The dataset consists of 1002 images of 82 different subjects with their ages varying from newborns FIGURE 2. Schematic diagram depicting the flowchart of the proposed model. to 69 years old. However, there is class imbalance resulting from over population of images belonging to ages between zero to 40 years in the database. The images with individuals at their more recent ages were the ones for which digital images were available. In most of the cases, the images were collected by scanning photographs of subjects found in personal collections. Some challenging issues related to this dataset are mainly the quality of images depends on the photographic skills of the photographer and the quality of the imaging equipment that the photographer has used. Also, the quality of photographic paper and printing along with the condition of photographs also have a great impact on the dataset. Thus, the face images in the dataset display considerable variability in quality, illumination, resolution, viewpoint and expression. Another challenge to work upon the dataset is the presence of occlusion in the form of spectacles, facial hair and hats in a number of images. In particular, information about the age, gender, expression, pose, image quality and occlusions like moustaches, beards, hats or spectacles was recorded.

## 5.2. AFAD DATASET

The Asian Face Age Dataset (AFAD) was developed for evaluating the performance of various age estimation models and architectures. It contains more than 160K facial images along with their corresponding age labels. This dataset has been designed for estimation of age on Asian faces, so all the facial images are of Asian people. It is to be noted that the AFAD is the largest dataset for age estimation till date. It is a perfect and well suited benchmark dataset to evaluate how deep learning methods can be adopted for age estimation. There are 164,432 labeled photos in the AFAD dataset with the ages varying from 15 to 40. The AFAD dataset was built by collecting photos of users from a particular social network called RenRen Social Network (RSN). The RSN is a social media platform in China which is widely used by Asian students belonging to all levels of education be it middle school, high school, undergraduate, or graduate students. Even after leaving from school, most of the people still access the platform in order to connect and keep in touch with their old classmates. So, the age of the RSN users belongs to a wide range varying from 15-years to more than 40-years old.

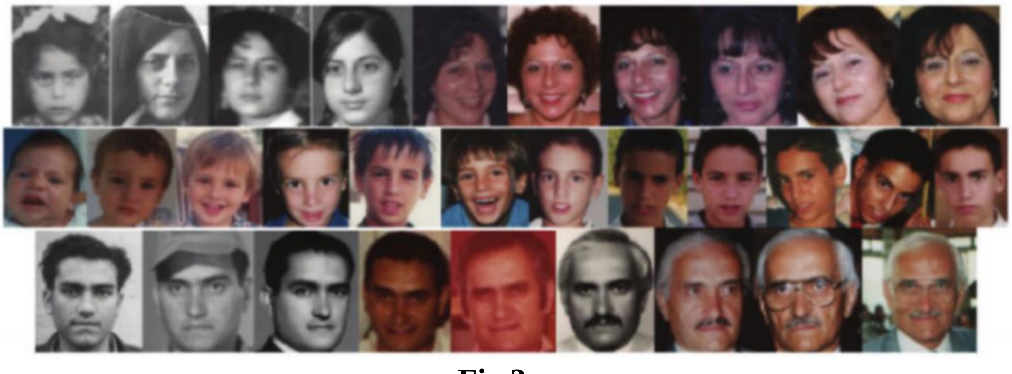


**Fig 3**

## 5.3. WIKIPEDIA AGE DATASET

Public availability of the datasets comprising of face images are a challenging issue. If available, it is often of small to medium size and rarely exceeds tens of thousands of images. Moreover, collection of age information for them is quite a challenging task. So, a large dataset  of faces of various celebrities was collected for this purpose. The most popular 100,000 actors as listed on the IMDb website were taken and various metadata related to that person like date of birth, name and gender were (automatically) crawled from their profiles. This metadata was used to crawl all profile images from the pages of people from Wikipedia. The images which were not timestamped (the date when the photo was taken) were removed as they will not have any age information in them. It was assumed that the images with single faces were likely to show the actors and that the timestamp and date of birth are correct, thus enabling proper assignment of the biological (real) age to each such image. In total, 62,328 face images from 20,284 celebrities were obtained from Wikipedia.

## 5.4. UTKFace DATASET

The UTKFace dataset is a large scale dataset which consists of face images with a very long age span ranging from 0 to 116 years old subjects. The dataset consists of more than 20K face images with metadata annotations of age, gender, and ethnicity. The images cover large variation in illumination, occlusion, pose, facial expression, resolution, etc. The dataset finds its use in a variety of tasks ranging from face detection, age estimation to age progression/regression and landmark localization, etc.

## 5.5. AdienceDB DATASET

The images present in AdienceDB dataset were crawled from Flickr.com albums, obtained by automatic upload from smartphones. After downloading the photos from the Flickr site, they were processed by first running the Viola and Jones face detector on them. Many faces in the albums appeared at different roll angles and to avoid missing such faces, the process of detecting faces was applied to each and every image, rotated 360◦ degrees in steps of 5◦ increments. Finally, the images were manually labeled for gender, age and identity using both the images themselves and any available contextual meta information like image tags and associated text or additional photos in the same album etc.

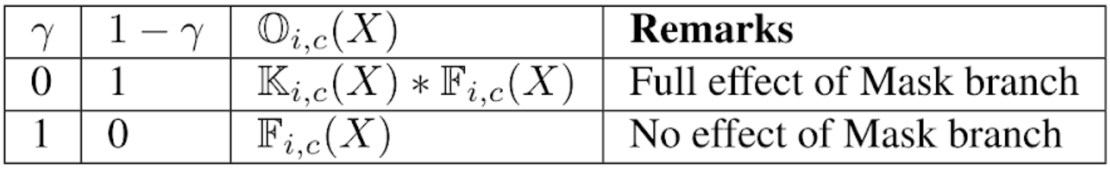
# 6. METHODOLOGY

## 6.1. GATED RESIDUAL ATTENTION NETWORK

The Residual Attention Network ([41]) used here has been constructed by combining and stacking multiple Attention blocks. Every block is further subdivided into two branches namely the Mask branch and the Trunk branch. The function of the Trunk branch is to perform feature processing, and it can easily be incorporated into any state-of-the-art network architectures. In this architecture, we have used pre-activation Residual Unit and ResNeXt with gated activation as its Gated Residual Attention Network's basic unit to construct Attention block. For a given Trunk branch's output P(X) with input X, the Mask branch makes use of a bottom-up top-down approach  for learning the same size mask K(X) that softly adds weights to the output features P(X). The bottom-up top-down approach tries to mimic the fast feed-forward (top-down) and feedback attention (bottom up) mechanisms. The output mask serves as the control gate for neurons of the Trunk branch similar to that of Highway Network . The output of Attention O is given as shown in Eq. 1. Oi,c(X) = Ki,c(X) ∗ Pi,c(X) (1) where, i ranges over all spatial positions and c ∈ {1, . . . ,C} is the index of the channel. The attention mask that is present in the Attention blocks, not only serves in the feature selection procedure during forward inference, but also plays a vital role as a gradient update filter during back propagation. In the soft mask branch, the gradient of mask for the input feature is shown in Eq. 2. ∂K(X, θ) ∂φ P(X, φ) = K(X, θ) ∂P(X, φ) ∂φ (2) where, θ are the mask branch parameters and φ are the trunk branch parameters. This property makes Attention blocks robust to noisy labels. Mask branches have the property to prevent wrong gradients (from noisy labels) and to update Trunk parameters. Instead of stacking Attention blocks to model the architecture, a simpler approach is to make use of a single network branch in order to generate a soft weight mask which is similar to spatial transformer layer. In its model, features from different layers need to be modeled automatically by the different attention masks to capture the most effective features for age and estimation. Use of a single mask branch requires an exponential number of channels in order to capture all combinations of the different factors. A single Attention block only modifies the features once. If such scenarios arise where the modification fails on some parts of the image, the subsequent network modules are not able to get a second chance. The Gated Residual Attention Network alleviates the aforementioned problems. In Attention block, every Trunk branch has its own mask branch with the purpose to learn attention that is specialized for its features. Besides, for complex images, the incremental nature of stacked network configuration can gradually refine attention.

## 6.2. GATED RESIDUAL ATTENTION LEARNING

Just stacking up of Attention blocks in a naive manner which may lead to the performance drop. This mainly occurs due to two major reasons. Firstly, the dot product with mask range in between zero to one repeatedly results in degradation of the value of features in the deeper layers. Secondly, the soft mask can potentially break the good property of Trunk branch like the identical mapping of Residual Unit. We have used gated residual attention learning to ease the aforementioned problems. Similar to ideas in residual learning, if a soft mask unit can be designed so as to serve as identical mapping, the performances may be no worse than its counterpart without attention. Thus, we use a modified version of output O of Attention block as given in Eq. 3. $Oi,c(X) = (\gamma + (1 - \gamma ) * Ki,c(X)) * Fi,c(X)$ (3) where, $K(X)$ is in the range of [0,1], with $K(X)$ approximating 0, $O(X)$ will approximate original features $F(X)$. The values $\gamma$ and $(1 - \gamma )$ signify the Gates of range [0,1] and are trainable parameters, controlling the effects of Mask branch and features generated by Deep Convolutional Networks on output Oi,c(X). We call this method as Gated Residual Attention learning which is different from residual learning and the architecture therefore is termed as Gated Residual Attention Network (GRA_Net). Figure 5 shows the schematic working of a Gated Attention block. In the original ResNet, residual learning is formulated as Oi,c(X) = X +Fi,c(X), where Fi,c(X) approximates the residual function. In its formulation, Fi,c(X) represents the features generated by Deep Convolutional Networks. The mask branches, K(X), play the role of feature selector which primarily aim to keep the good features and suppress any kind of noises from the Trunk features. Additionally, stacking up Attention blocks plays the role of backing up Gated Residual Attention learning by its incremental nature. This learning has the ability to keep good properties of original features, but also provides them the ability to bypass soft mask branch and forward to the top layers in order to weaken the mask branch's feature selection ability. Stacked Attention blocks can lead to gradual refinement of the feature maps. As shown in Figure 7, the features become more and more useful with increasing depths. By using the Gated Residual Attention learning, increasing depth of the network can improve performance consistently.



| $\gamma$ | $1-\gamma$ | $\mathbb{O}_{i,c}(X)$ | **Remarks** |
|---|---|---|---|
| 0 | 1 | $\mathbb{K}_{i,c}(X) * \mathbb{F}_{i,c}(X)$ | Full effect of Mask branch |
| 1 | 0 | $\mathbb{F}_{i,c}(X)$ | No effect of Mask branch |

**Fig 4**

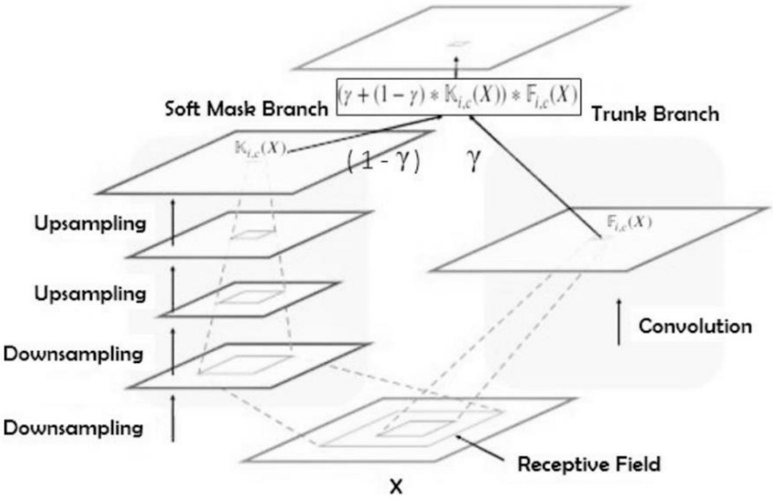


**Fig 5**

## 6.3. SOFT MASK BRANCH

By making use of the previous attention mechanism concept as present in Deep Belief Network (DBN) like Restricted Boltzmann Machines(RBM), the mask branch contains fast feed-forward sweep and top-down feedback steps. The feed-forward operation manages to quickly collect the global information available in the whole image, while the latter one tries to combine the same with the original feature maps. In the CNN, the two steps unfold into bottom-up and top-down fully convolutional architectures. From the input, repetitive applications of max pooling is done in order to increase the receptive field with a rapid rate after a small number of Gated Residual Units. As soon as the lowest resolution is reached, the global information is then expanded by a symmetrical top-down architecture in order to guide the input features in each pixel position. Linear interpolation tries to up sample the output after some Residual Units as seen in Figure 6. The number of bi-linear interpolation applied is the same in number as max pooling to make sure that the output size is the same as the input feature map. Then after two consecutive 1 X1 convolution layers, a sigmoid layer normalizes the output range to [0, 1]. Skip connections have also been added between the top-down and bottom-up components with the motive of capturing information from different scales without any loss of useful information from the knowledge of previous layers. The complete module is illustrated in Figure 6. The main aim of the mask branch is to improve the Trunk branch features in place of solving a complex problem directly.

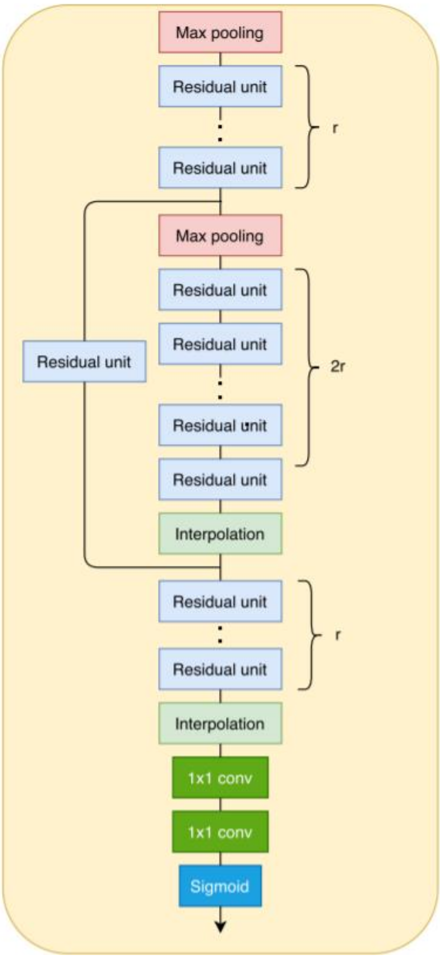


**Fig 6**

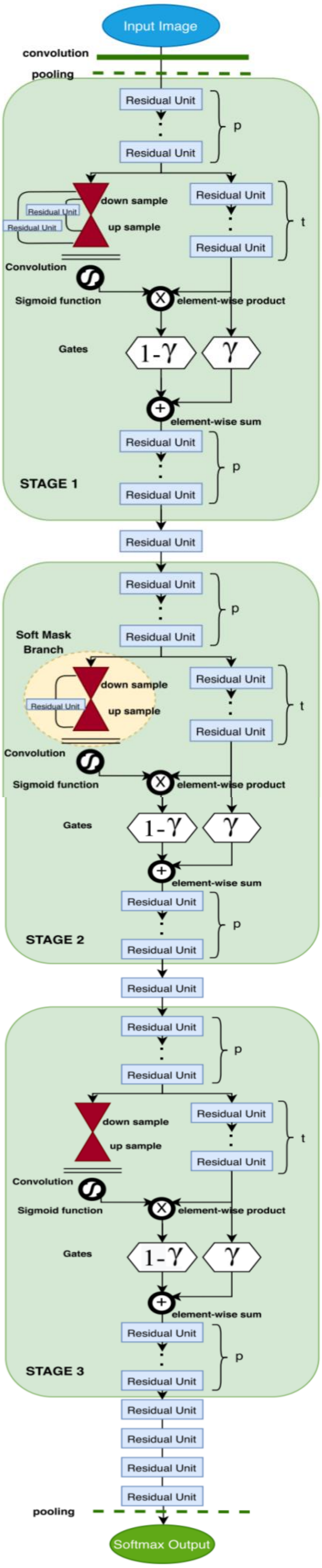

Input Image
convolution
pooling
Residual Unit
p
down sample
up sample
Residual Unit
t
Convolution
Sigmoid function
element-wise product
Gates
1-γ
γ
element-wise sum
STAGE 1
Soft Mask Branch
STAGE 2
STAGE 3
pooling
Softmax Output

**Fig 7**

## 6.4. SPATIAL AND CHANNEL ATTENTION

In this work, attention provided by the mask branch changes continuously by adapting with the features of the Trunk branch. However, attention in the mask branch can be restricted by making changes in the normalization step of the activation function before soft mask output. We have made use of 3 types of activation functions corresponding to Channel attention, Mixed attention and Spatial attention. The Mixed attention f1 (refer Eq. 4) without any additional constraints makes use of a sigmoid function for each channel and each spatial position. Channel attention f2 (refer Eq. 5) applies L2 normalization within all channels for each and every spatial position in order to reduce spatial information. Spatial attention f3 (refer Eq. 6) executes normalization within the feature map from each channel and then sigmoid activation in order to retrieve a soft mask related to spatial information only. f1(Xi,c) = 1 1 + exp(−Xi,c) (4) f2(Xi,c) = Xi,c ||Xi || (5) f3(Xi,c) = 1 1 + exp(−(Xi,c − meanc)/stdc) (6) where, i ranges over all spatial positions and c ranges over all channels. meanc and stdc denote the average and standard deviation of feature map from c th channel respectively. Xi denotes the feature vector at the i th spatial position.

## 6.5. REGRESSION

In this work, attention provided by the mask branch changes continuously by adapting with the features of the Trunk branch. However, attention in the mask branch can be restricted by making changes in the normalization step of the activation function before soft mask output.

## 6.6. LABEL DIVISION

The division of age label into various sub-labels is shown in Figure 8.

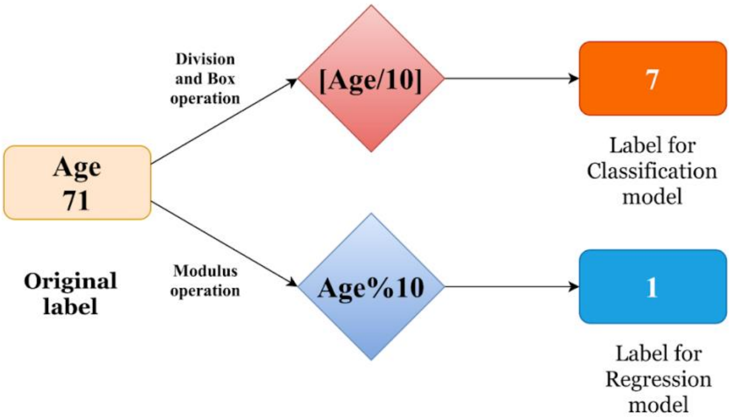

Division and Box operation
[Age/10]
7
Label for Classification model
Age 71
Original label
Modulus operation
Age%10
1
Label for Regression model

**Fig 8**

# 7. ANALYSIS OF RESULTS

## 7.1. PRE-PROCESSING AND PARAMETER SETTINGS

1) GENDER IDENTIFICATION Facial images are first scaled and resized to 32 × 32 keeping the RGB channels intact. These images are then passed dynamically by an ImageDataGenerator object to the model for training purposes. The output layer consists of a softmax layer with two units depicting ‘‘Male’’ and ‘‘Female’’ classes. As this is a binary classification problem, so loss function used here is binary cross entropy. Optimization is done using Nadam (Adam with Nesterov momentum) optimizer with a learning rate of 0.001. The final model consisted of 33 million trainable parameters.
2) AGE IDENTIFICATION Similar to the process of gender identification, the images are first scaled and resized to 32 × 32 keeping the RGB channels intact. Thereafter, these images are passed by an ImageDataGenerator object but to two different models, one for classifying the age category, and another one for classifying the regression value of exact age for the subject. As the classification model is used here a multi-class classification problem, so loss function used is categorical cross entropy and optimizer used is same as that used in gender identification. This model also consists of 33 million trainable parameters

## 7.2EVALUATION METRICS

For age estimation problem, in order to guarantee the accuracy of its algorithm and provide fair comparison with available state-of-the-art models, MAE  is taken as the evaluation metric, which minimizes the error between the estimated age and the ground truth label. MAE(J (X ) is given by Eq. 7. J (X ) = 1 M X M i=1 |Y˜ i − Yi | (7) where, Y˜ i = True Age and Yi = Predicted Age for i th data point. Apart from MAE, the test accuracy for some datasets has also been measured for comparison purposes. For gender classification, measured the test accuracy for the datasets which have labels for the same. As gender classification is basically a binary classification problem, so the evaluation metric ‘‘accuracy’’ is defined as shown in Eq. 8. Accuracy = (tp + tn) (tp + tn + fp + fn) (8) where, tp = True positive; fp = False positive; tn = True negative; fn = False negative.

## 8. CONCLUSION

Identifying the age and gender of the individuals we come across in its daily lives has an important role in its social lives too. For example, languages used to salute for men and women are very often different, and the words used to address the elders and the young are different too. We human beings are able to evaluate the individual’s age and gender just from their facial appearances. Nowadays, many smart applications that include visual surveillance, medical diagnosis, and marketing intelligence, need to evaluate the same for individuals using their facial images. Here lies the importance of a robust and efficient methodology for gender and age estimation through the computing devices. To this end, in this paper, proposed a deep learning based model, named GRA_Net, for the purpose of age (regression problem) and gender (a binary classification problem) prediction from the facial images. It considered the age prediction problem as a combination of classification and regression problems. Its proposed model has been evaluated on five publicly available standard datasets. Obtained results for both the tasks have been able to outperform many state-of-the-art methods. There are still some rooms for improvement of the proposed model which can be done in the near future. For example, we have to do more work while identifying the gender of kids as in the tender age both the male and female individuals have many commonalities in the facial features. Also, its model needs to be more intelligent in estimating the age and gender when images are obstructive, partially viewed, bearing hat/glass/wig, and wearing some unusual make-up etc. Even facial images of different provinces of the world have different characteristics.

# 9. REFERENCES

[1] GRA_Net: A Deep Learning Model for Classification of Age and Gender From Facial Images: https://ieeexplore.ieee.org/document/9446083

[2] Deep Learning: https://searchenterpriseai.techtarget.com/definition/deep-learning-deep-neural-network

[3] CNN: https://towardsdatascience.com/a-comprehensive-guide-to-convolutional-neural-networks-the-eli5-way-3bd2b1164a53

[4] O. Agbo-Ajala and S. Viriri, ‘‘Deeply learned classifiers for age and gender predictions of unfiltered faces,’’ Sci. World J., vol. 2020, pp. 1–12, Apr. 2020, doi: 10.1155/2020/1289408.

[5] V. Badrinarayanan, A. Handa, and R. Cipolla, ‘‘SegNet: A deep convolutional encoder-decoder architecture for robust semantic pixelwise labelling,’’ 2015, arXiv:1505.07293. [Online]. Available: http:// arxiv.org/abs/1505.07293

[6] A. M. Bukar, H. Ugail, and D. Connah, ‘‘Automatic age and gender classification using supervised appearance model,’’ J. Electron. Imag., vol. 25, no. 6, Aug. 2016, Art. no. 061605, doi: 10.1117/1.JEI.25.6. 061605.

[4] K.-Y. Chang, C.-S. Chen, and Y.-P. Hung, ‘‘Ordinal hyperplanes ranker with cost sensitivities for age estimation,’’ in Proc. CVPR, Jun. 2011, pp. 585–592.

[5] K.-Y. Chang, C.-S. Chen, and Y.-P. Hung, ‘‘A ranking approach for human ages estimation based on face images,’’ in Proc. 20th Int. Conf. Pattern Recognit., Aug. 2010, pp. 3396–3399.

[6] 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.