# Gender Recognition And Age Estimation Using Deep Neural Network

By

**Roshni Tasnim (C151222)**

**Afsana Akther Chowdhury(C151226)**

**A thesis report submitted in partial fulfillment of the requirement for the degree of Bachelor of Science in Computer Science and Engineering.**


**Department of Computer Science and Engineering**

**International Islamic University Chittagong**

# Gender Recognition And Age Estimation Using Deep Neural Network

By

**Roshni Tasnim (C151222)**

**Afsana Akther Chowdhury(C151226)**

**A thesis report submitted in partial fulfillment of the requirement for the degree of Bachelor of Science in Computer Science and Engineering.**

**Approved by:**

| | |
|---|---|
| **Supervisor**<br>**Mr. Md. Khaliluzzaman**<br>Assistant Professor,<br>Dept. of CSE, IIUC<br>January, 2020 | **Tanveer Ahsan**<br>Associate Professor and Chairman<br>Dept. of CSE, IIUC<br>January, 2020 |

# Declaration

We hereby declare that except where specific reference is made to the work of others, the contents of this dissertation are original and have not been submitted in consideration for any other degree or qualification in this or any other university. This dissertation is our own work and contains nothing which is the outcome of work done in collaboration with others, except as specified in the text and acknowledgements.

_____

**Roshni Tasnim (C151222)**
**January, 2020**

_____

**Afsana Akther Chowdhury (C151226)**
**January, 2020**

# Acknowledgement

First and foremost, we would like to express our heartiest gratitude to the Almighty Allah (SWT).

It has been a great opportunity for us to be students of the Department of Computer Science and Engineering which is one of the prestigious departments of the International Islamic University Chittagong. We would like to express our sincere gratitude to our department, professors and staffs as well.

We would like to express our deep and sincere gratitude to our thesis supervisor, Mr. Md. Khaliluzzaman, Assistant Professor of Department of Computer Science and Engineering, International Islamic University Chittagong (IIUC) for providing us with the valuable guidance, inspiration, direction, encouragement throughout the research work, which would not be successful without his supervision.

We are thankful to all of our esteemed professors and course teachers for their scholarly teaching, guidance, support, and inspiration.

Last but not least, we are also extremely grateful to our parents who have not lonely given us the support and encouragement to study but also provided the inspiration to do everything possible. Without their immense support, this research work would not be successful.

# Abstract

Our research estimates the facial attributes age and gender from images which are taken in challenging and wild conditions. This problem is concerned less attention in the field of face recognition. Because of the growth of online social networking websites and social media, extracted information from images with a level of accuracy is quite useful in person re-identification. We propose a convolutional neural network based model which predicts age and gender of multiple attributes. The model is evaluated on the Adience benchmark. Our main contribution is using a unique dataset where age and gender labeled images acquired by smart phones and other devices and uploaded without filtering in the image repositories and correcting the whole dataset on the basis of age and gender attributes. These images are more challenging than other face photo benchmark. A CNN model for far view image is used as a based line method in order to compare with ours. Experiments show that our model outperforms some of the popular methods.

# Table of Contents

# Figure Appendix

# Table Appendix

# Chapter 1

# Introduction

## 1.1 Introduction:

Gender recognition and age estimation plays an important role in computer vision. There are many significant applications of computer vision like human-computer interaction, visual surveillance and security control. Automatic gender recognition has increased to an extension in various software and hardware because of the growth of online and social networking websites and social media. How about the performance of already exist system with face pictures are not excellent in comparison with the result of face recognition. With the advancement of machine learning, learning and classification is more easier nowadays. Deep learning and convolutional neural network become an important tool in computer vision application like gender recognition and age estimation image resolution is very low, so when amount of data is limited to performing any experiments related to computer vision in that case it is good to take an advantage from an already train deep convolutional neural network.

The identification of gender and age can be used for a significant number of application like:

1) Robotic applications that require information about gender and age
2) Human computer interaction surveillance systems
3) automated analysis of gender behavior and biometrics
4) control of access in important areas

## 1.2 Challenges for gender recognition and age estimation:

1) Appearance of hurdle objects in the range of camera.
2) Back views based gender prediction is also challenging.
3) Account of disparity and the variety of pedestrian clothing the intraclass differences occur among different images for the same attribute.
4) Gender and age identification is multi-label classification preferably a multiclass and a single class classification problem. The multi-class classification algorithm is not relevant

because the attributes are not completely simultaneous. So multi-label class classification is suitable because it has already owned method for these challenges.

Deep learning and convolutional neural networks have shown unbelievable performance in the past few years the handcrafted features methods extract features automatically from images. It has done incredible performance when it came to computer vision such as image classification and recognition. From the first layer to the last layer they can learn features to high level abstract features because it together several layers.

### 1.3 Contributions of this research work:

In our research work, we proposed a convolutional network for gender and age from close view images from adience dataset. The main contributions and goals of our research work are:

- Proposed a model which is based on deep learning to classify the age and gender from images as a multiclass classification task on the basis of the relationship between the classes.
- Compare the performance using adience dataset with Mini VGGnet for age and gender detection.
- By changing parameters, we create 10 model to accurate our model by comparing performance.

# Chapter 2

# Theoretical Background

## 2.1 Deep Learning:

Deep learning is a part of machine learning which concerned with algorithms inspired by the structure and function of the brain. Where there is some lacking of machine learning deep learning shines there. Deep learning refers deep artificial neural networks and somewhat less frequently to deep reinforcement learning. Deep learning's capability is differ from some aspects from traditional shallow machine learning. Deep learning is capable of using raw data and can automatically learn the features required to perform the specific identification task. This learning ability is based on stacking several non-linear models as a stack of multiple layers that convert the raw input data into a higher label more abstract representation. Each successive layer uses the output from the previous layer as input for the next layer.

Deep learning can take messy and broadly unlabeled data and make useful predictions. But its drawback is it requires a big amount of data to train.

## 2.2 Regular Neural Network:

Regular neural networks are modeled like human brain which are designed to recognize patterns. They can take decisions like human brain understand the real-world things.

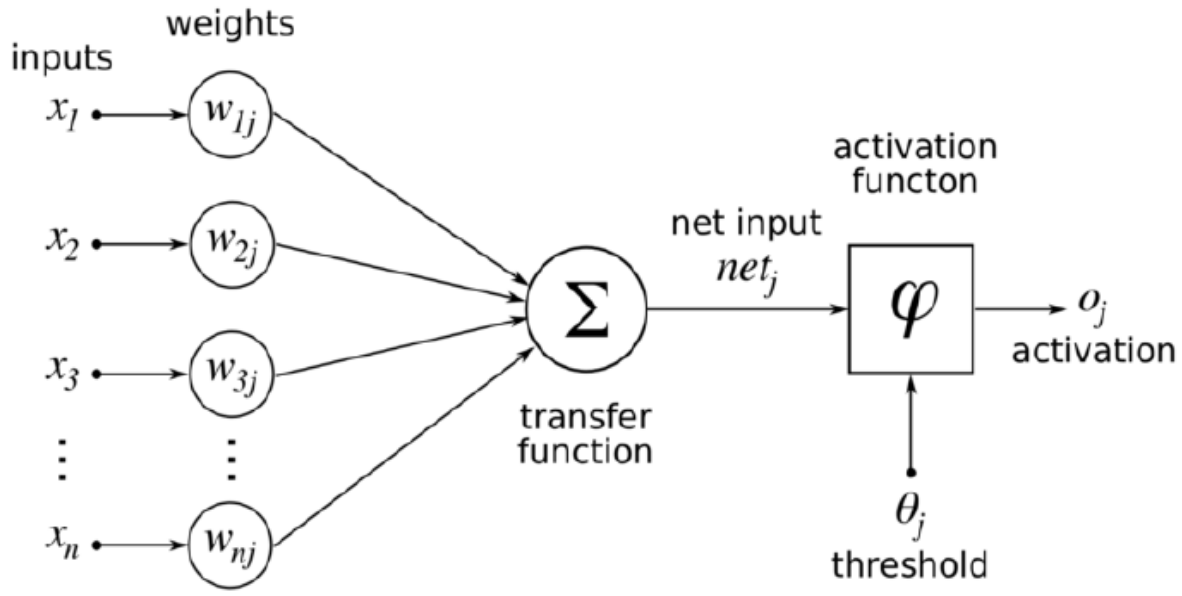



Figure 1: Structure of Regular Network

The figure shows the structure of a perceptron which gives binary output based on a threshold by taking a unit of weighted inputs.

A neural network is a set of connected neurons designed in layers:

- **Input layer:** It contains all initial input for further processing.
- **Hidden layer:** This layer is in between input and output layers. It capture all complexity with every layer by discovering relationships between feature in the input.
- **Output layer:** It is the last layer of neurons which produces given output for the program.

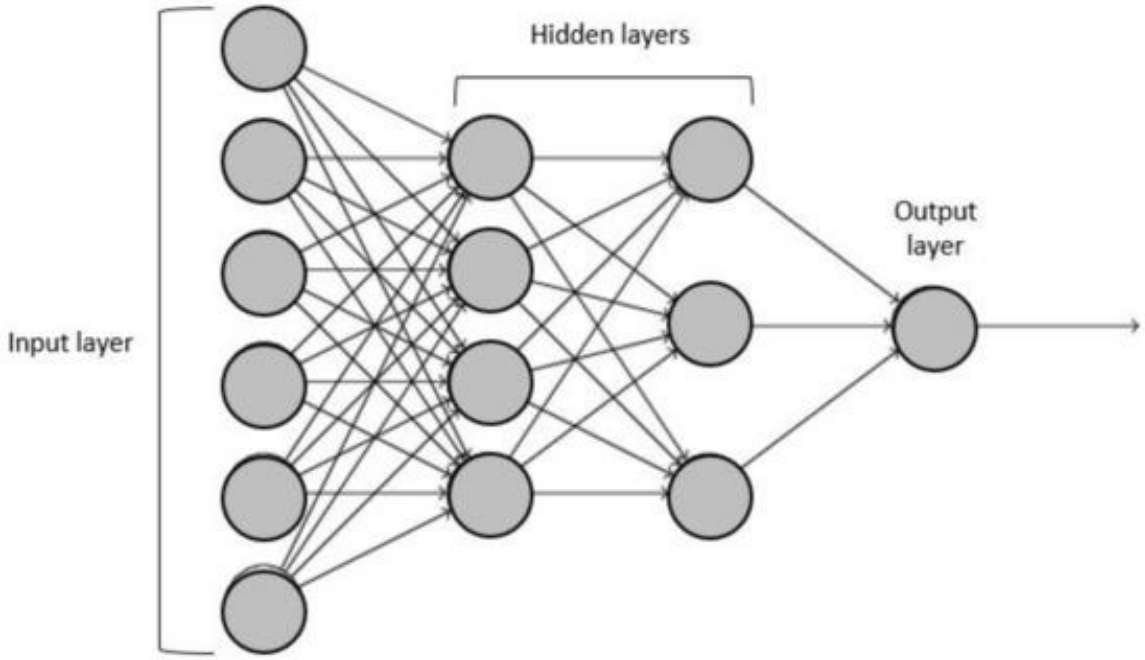



Figure 2 : Structure of layers in Neural Network

**2.3 Convolutional Neural Network (CNN):**

Convolutional neural network is a part of deep neural network. Convolution neural network is a very incredible and effective method in regions such as image identification and classification. They take input which made up with images and by compelling the architecture in a more practical way it gives proper output. It's a regularized versions of multilayer perceptron. Multiplayer perceptron refer to fully connected network.

There are three types of convolution layer:

1) **Convolution**: Convolution is focused to extract features using a small squares filters of input data which is an n-dimensional matrix. It move the whole original image by n pixels and for every position between two matrices. Then compute multiplication and add the all multiplication outputs to get the final output matrix.

   The three-dimension matrix is called a 'filter, 'or 'kernel,' or 'feature detector' in convolutional neural network terminology and moving the filter over the whole image and computing the dot product is called the 'Convolved Feature' or 'Activation Map' or the 'Feature Map' is formed the matrix. Filters play an important role to detect the features from the original input image.

2) **Non-Linearity (ReLU):** ReLU stands for Rectified Linear Unit and is a non-linear operation. It puts zero in the features map by replacing all the negative values. It is an element-wise operation. Introducing the non-linearity in our ConvNet is the main purpose of the ReLu since most of the real-world data we would want our ConvNet to learn would be non-linear.
3) **Pooling**: Pooling layer is used to reduce the size of representation to speed the computation as well as make some of the features it detects a bit more robust.
4) **Fully connected**: Fully connectedness makes the networks prone to overfitting. Data it uses a softmax activation function in the final output layer. The main purpose of the fully connected layer is using these features for classifying the input image into different classes based on the training dataset.

### 2.4 Classification Loss Function:

Loss function measures the performance of a classification model output is a probability value which is in between 0 and 1. It also called cost function. The loss function we are going to use in binary cross entropy loss. It also called sigmoid cross entropy loss. It is a sigmoid activation plus a cross entropy loss. It computed the loss for every CNN output class is not affected by other component values that's why we used it in multi label classification. The insight of an element belonging to a certain class should not influence the decision for another class.

The mathematical formula of binary cross entropy is

$$\mathrm{CE} = -\sum_{i=1}^{C'=2} t_i \log\big(f(s_i)\big) = -t_1 \log\big(f(s_1)\big) - (1 - t_1)\log(1 - f((s_1)) \quad \text{.....(1)}$$

Where,

$C$ = the independent binary classification problems

$s_1$ = the score

$t_1$ = the groundtruth label for class

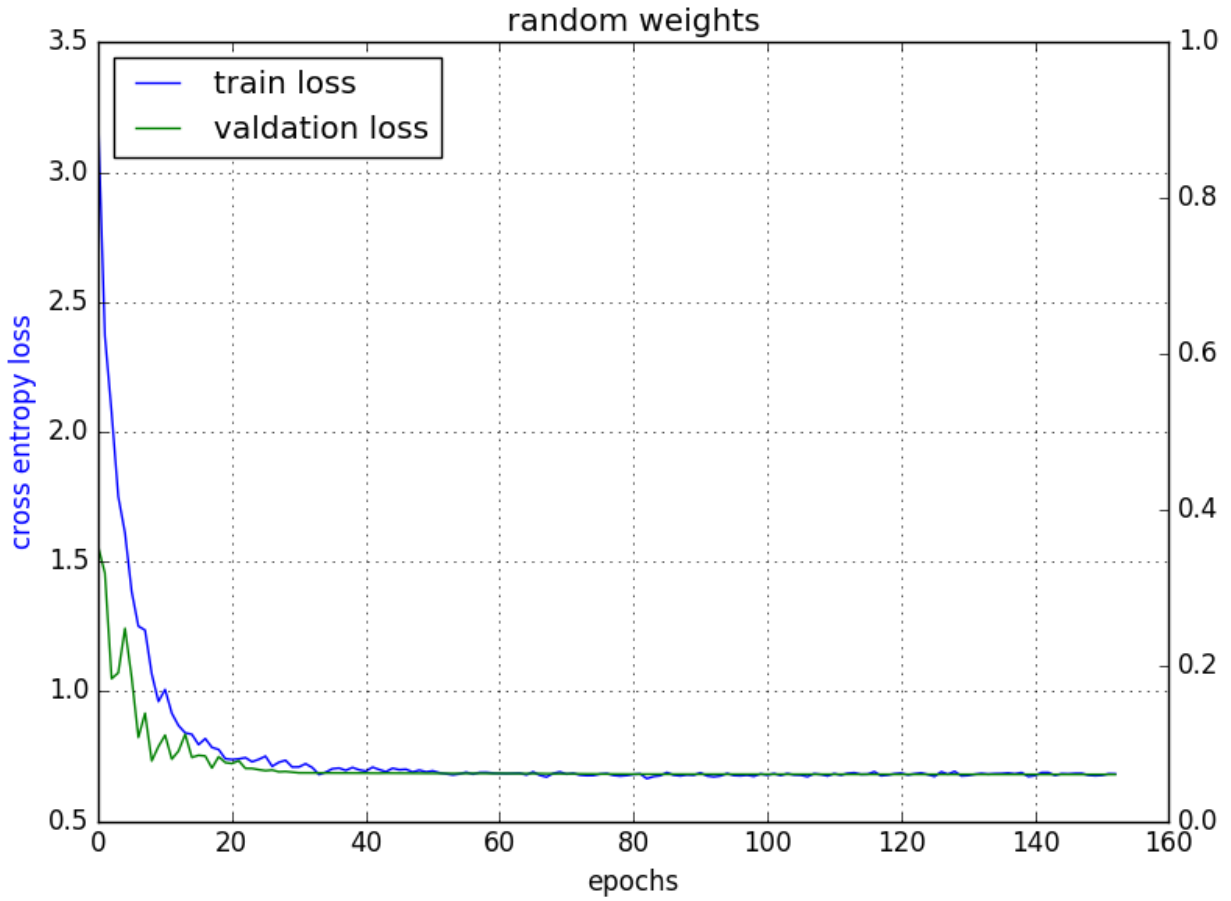



Figure 3 : Binary Cross Entropy Loss Curve

**2.5 Optimization Algorithm:**

For minimizing the loss function and make correct prediction as much as possible we need used optimization Algorithm. Optimization algorithm change and shape the model by updating the parameters and loss function.

**Gradient Descent:** Gradient Descent is a iterative optimization algorithm that used to minimized the function value. It updates the parameters from the negative direction of the gradient of the function to reach the local/global minima. Most of the time this function is loss function. Gradient Descent is most popular optimization algorithm.

A simple gradient descent algorithm is as follows:

1) Obtain a function to minimize F(x).
2) For start the descent from initialize a value of a x.
3) For determine how much step to descend specify a learning rate.
4) Find the derivative of x.
5) Multiply the derivative with learning rate.
6) Update the value of x with new value of x.
7) If stopping condition satisfied the stop otherwise keep repeating from step 4 until satisfied the condition.

The following formula updates the parameters-

$$\theta = \theta - \gamma \nabla F(\theta) \quad \text{.....(2)}$$

Where $\theta$ denotes parameters, $\gamma$ denotes learning rate and $\nabla F(\theta)$ denotes gradient of loss function $F(\theta)$.

We are going to discuss how gradient descent evolved using different strategies and why they were improved.

**a) Standard batch/Vanilla Gradient Descent:** Vanilla gradient descent is a basic method of gradient which is without bells. Here Vanilla means Standard. It computes the gradient of whole dataset only once. It updates weight frequently that's why convergence will get very slow.

$$\nabla f = 1n \sum I \nabla loss\ (xi) \quad \text{.....(3)}$$

b) **Stochastic Gradient Descent:** Stochastic means a process or a system. It is linked with a random probability. It reaches convergence much faster. It finds maximums and minimums by iteration for every single training example. We used stochastic gradient descent for our model. It has much efficiency and ease of implementation.

It has some drawbacks too. It is computationally more expensive because it requires number of hyper parameters like regularization parameter and number of iteration.

c) **Mini-batch gradient descent:** Mini batch has the good parts of batch and stochastic gradient descent. It splits the training dataset into small batchs which used to calculate model error and update model coefficients. In minibatch parameters are updated after computing the gradient of error

# Chapter 3

# Literature Review

### 3.1 Existing works on gender recognition and age estimation:

## Age and Gender Classification using Convolutional Neural Networks [1]

This would be the base paper of our thesis. In this paper they propose a simple convolution net architecture that can be used even when the amount of learning rate is limited. They earn same achievement with a similar network where there is limited availability of accurate age and gender labels in existing face data sets.

**Network Architecture:** Images are rescaled to 256x256 and a crop of 227x227. The three subsequent convolutional layers are described below:

1. 96 filters of 3x7x7 pixels followed with a ReLU, a max pooling of maximum value of 3x3 regions with 2px strides and a local response normalization layer applied in the first convNet.
2. 256 filters of size 96x5x5 pixels is followed by ReLU, a max pooling layer and a local response normalization .
3. The 3rd convNet works on 256x14x14 blob by applying 384 filters of size 256x3x3 pixels followed by ReLU and a max pooling layer.

Fully Connected layers: In the 3 fully connected layer, the first one takes the output of the 3rd convNet which contains 512 neurons, followed by a ReLU and a dropout and the last two also receives 512 dimensional input, followed by a ReLU and a dropout layer.

Finally, the output of the last FC is fed into a soft-max layer which estimates a probability for each class.

**Parameters:**

1. Weights in all layers are initialized with random values using zero mean Gaussian with standard deviation of 0.01.
2. Do not used pre-trained models for initialization of network.

**Hyper Parameters:**

1. 3 convolution layers and 2 fully-connected layers
2. 2 dropout layers with a dropout ration of 0.5.
3. Data augmentation by taking random crop of 227x227 pixels from the 256x256 input images
4. Gradient descent with batch size of fifty images.
5. Initial learning rate e^-3, reduced to e^-4 after 10k iterations.

**Merits:**

1. Improved age and gender classification results can be gained using much smaller size of contemporary unconstrained image sets labeled for age and gender.
2. The simplicity of this model proves that more deep network using more training data will be able to improve result.

**Demerits:**

1. They choose a smaller network to reduce the risk of overfitting which made it shallow compared to other network.

## DeepPose: Human Pose Estimation via Deep Neural Networks[2]

In this paper, a DNN regressors is introduced which gains high precision pose estimations and have done their work on four academic benchmarks of diverse real-world images. They propose a cascade of DNN-based pose predictors allows for increased precision of joint localization based on the full image starting with an initial pose estimation where DNN based regressors refines the joint predictions by using higher resolution sub-images. The network learns filters capture pose properties at coarse scale. Instead of a classification loss, a linear regression is trained on the last network layer to predict a pose vector by minimizing L2 distance between the prediction and the true pose vector. A cascade of pose regressors is trained initially as estimating an initial pose outlined in previous section and each subsequent stage uses the predicted joint locations to focus on the relevant parts of the image, sub-images are cropped around the predicted joint from previous stage and the pose displacement regression for this joint is applied on this sub-image. Simulated predictions are generated in training.

**Model:**

1. Conv layer(55x55x96)-LRN(Local Response Normalization Layer)-P(Pooling Layer)-C(27x27x256)-LRN-P-C(13x13x384)-C(13x13x384)-C(13x13x256)-P-F(Fully Connected Layer)(4096)-F(4096).
2. Filter size: In first two C layer 11x11 and 5x5, in remaining three is 3x3.
3. Input image of 220x220, stride of 4 and total number of parameters of the model is 40M.

**Hyper Parameters:**

1. Hidden layers: 7
2. Mini-batch: 128
3. Learning rate: 0.0005
4. Data augmentation
5. Dropout regularization for Fully Connected layer: 0.6

**Merits:**

1. DNN is capable of capturing the full context of each body joint.
2. This approach is substantially simpler to formulate than methods based on graphical models.
3. Percent of detected joints(PDJ) metric is used instead of Percentage of Correct Parts(PCP), which allows to vary the threshold for distance between prediction and ground truth and acts as a localization precision.
4. A generic convolutional neural network can be applied to the different task of localization.

## **Deep Learning Face Representation from Predicting 10,000 Classes** [3]

In this paper a set of Deep hidden Identity features(DeepID) is learned through Deep learning where these are taken from last hidden layer neuron activations of deep Convolutional networks(ConvNets). Highly compact 160 dimensional DeepID is acquired at the end of the cascade that contain rich identity information and directly predict 10,000 classes of identity. When DeepID learned, as classifiers to recognize classes in the training set and configured to keep reducing the neuron numbers along the feature extraction hierarchy and these Deep

ConvNets gradually form compact identity related features in the top layers with only a small number of hidden neurons. They use method proposed at Sun et al. to detect facial point and trained 60 ConvNet, each of which extracts two 160 dimensional DeepID vectors where features are extracted from 60 face patches. They use the joint Bayesian technique and also train a neural network based on DeepID for face verification.

**Model:**

1. Deep Convolutional Network: ConvNet contains convolutional layers (with max pooling) to extract features hierarchically, followed by the fully connected DeepID layer and the softmax output layer indicating identity classes. They use ReLU non-linearity for hidden neurons where weights are locally shared to learn and weights in 3rd layer are shared in every 2x2 regions while in the 4th layer, they are unshared. The last hidden layer of DeepID is fully connected to both 3rd and 4th layer where multi scale features are used. Stochastic gradient descent is used with gradients calculated by back propagation. The input is 39x31xk for rectangle and 31x31xk for square patches.

**Hyper Parameters:**

1. Classifying all the identities makes full use of neural network to extract features for face recognition where it implicitly adds a strong regularization to ConvNets.
2. Face Verification Network: The neural network contains one input layer taking the DeepID, one locally connected layer, one fully connected layer and a single output neuron indicating face similarities. The 2nd hidden layer is fully connected to the 1st hidden layer where the hidden neurons are ReLU and the output neuron is sigmoid.

**Merits:**

1. When the number of training identities increases, the verification performance steadily gets improved.

2. Faces of the same identity tend to have more commonly activated neurons than those of different identities.

3. The bypass connections between the third convolutional layer and the last hidden layer reduces the possible information loss in the 4th convolutional layer as the 4th layer contains too few neurons.

## How Transferable are CNN-based Features for Age and Gender Classification?[4]

In this paper, they have explored transferability of generic Alex Net-like architecture and domain specific VGG-Face CNN model and are employed and fine-tuned with the Audience dataset prepared for age and gender classification in uncontrolled environments. Task specific GilNet CNN model has been utilized and used as a baseline method in order to compare with transferred models.

1. **Task Specific GilNet CNN model:** GilNet CNN has Convolution layers followed by fully connected layers and each convolution layer is activated by a ReLU and then maxpooled. The output of the first two convolution layers is normalized by Local Response Normalization (LRN) approach. FC6 and Fc7 with an output size of 512 and both vectors are regularized by dropout.

**Input**: 3 channels of 256x256 resolution image, cropping pathes of 227x227 .

**Hyper parameters:**

1. Hidden layers: 3
2. Fully connected layers: 3
3. Dropout for training: 0.5
4. 50 samples of input are fed into the network

**SVM Classification on CNN features:**

CNN features are extracted from FC7 and the dimensional feature vector of 512 is obtained for each training image. All feature vectors are then used for training SVM classifiers in grid-search manner. Both linear and RBF kernels are combined and ending up 14 SVM models for training/testing.

**2. Generic Alex Net-like CNN Model:**

AlexNet-like CNN model has convolutional layers followed by fully connected layers (FC6,FC7 and FC8). Each convolutional layer is activated by ReLU and then maxpooled followed by LRN. FC6 and FC7 has 4096 output. Softmax has a probabilistic output of 1000 for each object category. FC7 of size 4096 of feature vector is used for feature extraction and is fed into SVM classifier.

**Input:**

1. 3 channels, 256x256 resolution image
2. Cropped with patches of 227x227

**Hyper paramets**:

1. Hidden layers: 5
2. Fully Connected layers: 3
3. Batch size: 256
4. Dropout: 0.5
5. Learning rate in the last fully connected layer is set to a higher value (x10) than the learning rates of other layers.

**3. Domain Specific VGG-Face CNN model:** VGG-Face CNN has 16 layers and was trained on 2.6M facial images of 2622 identities. First two blocks, one convolution layer is followed by another one and the output of each layer is activated by a ReLU. At the end of the block . The output is max pooled. The last convolution layer is followed by FC6, FC7 and FC8. FC6 and FC7 have an output of 4096 where FC8 is responsible for classification and has output of 2622.

**Input:**

1. 3 channels of 224x224 resolution image

**Hyper Parameters:**

1. Hidden layers: 5
2. Dropout: 0.5(FC6,FC7)
3. Batch-size: 64

**SVM Classification on CNN feature:**

The output of FC7 layer of 4096 dimensional feature vector is used as the CNN feature and is fed into SVM classifiers with the same grid search manner applied for previous models.

**Framework:** Caffe deep learning framework

**Hyper parameters:**

1. Momentum: 0.9
2. Weight decay: 0.0005
3. Gamma: 0.1
4. Initial learning rate: 0.001

**Training:** VGG-Face Net was fine tuned in 30000 iterations while AlexNet-like and GilNet models are fine tuned/trained in 50000 iterations. There were two models for each pre-trained model with a total of 6 models. For grid-search of the SVM models linear and Radial Basis Function (RBF) kernels were used with varying cost values ranging from 10^-3 to 10^3 with a factor of 10, C= {0.001,0.01,0.1,1,10,100,1000} ending up 14 different SVM classifiers per task for each pre-trained model.

**Merits:**

1. Ft-VGG Face (Fine tuned) model can learn rapidly in just 1000 iterations (~4 epoch) with an accuracy ~80% for age classification and in gender classification this model learns very fast by having accuracies ~95% in just 1000 iterations with RBF kernel of C = 100.

2. Transferring a deep CNN model can have better classification performance than training a task specific model from scratch in case of availability of limited data.

3. Transferring from a closed domain is more useful than transferring from generic.

## A Convolutional Neural Network for Pedestrian Gender Recognition[5]

This paper propose a discriminatively trained convolutional neural network for gender classification of pedestrians that can be relatively straight forward architecture and minima preprocessing of the images which achieved 80.4% accuracy on dataset containing full body images in both front and rear views.

**Network Architecture:**

This proposed architecture is comprise of 7 layers. Images were cropped to 54x108 by removing border pixel before resizing down to 40x80.

1) First layer contains 10 features maps and uses 5x5 filters.

2) After down sampling by 2x2 max pooling feature maps is 18x38 in second layer.

3) Third layer contains 20 feature maps is in 14x34.

4) Fourth layer contains 7x17 feature maps using 2x2 max pooling.
5) All unit of the feature maps are connected to each of the 25 neuron unit in fifth layer.
6) The output layer has two units for binary classification.
Hyperbolic tangent activations in both convolution and hidden layers.

**Parameter:**

1) The total member of free parameters that are learnt by training 64,857.

2) Weights are randomly initialized from a uniform distribution range $[-\sqrt{(\frac{6}{f})}, \sqrt{(\frac{6}{f})}]$, f is the total number of input and output connections

3) During each iteration weights were updated by backpropagation.

4) Each unit of a feature map shares the same set of weights for the filter.

**Hyper parameter:**

1) Images are cropped to 54x108 before resizing down to 40x80.

2) Images are converted to grayscale and in the range [0,1] which used as input image.

**Merits:**

1) Used relatively small number of feature maps, thus require less computational intensity.

2) Weight sharing reduces the number of trainable parameters to achieve better generalization ability.

**Demerits:**

1) Some images are misclassified.

2) Only fully supervised training is used.

## Part-Wise Pedestrian Gender Recognition Via Deep Convolutional Neural Network[6]

This paper proposed a deep convolutional neural network to improve the accuracy of prediction. In this model, first parsed the images with existing deep de-compositional neural network into full body and upper body images. Then both type images goes through a CNN model. This method produces better accuracy of the frontal view abdominal area class which is 83.3%. The training performance is measured in validation error. The minimum VE achieved by this network is 0.078%.

**Network Architecture:**

The CNN is composed several layers. Four convolutional layer's filter blank size is 9x9x1x25, 7x7x25x50, 7x7x50x72, 4x4x75x100 respectively. Two fully connected layer, four pooling layers having stride 2. Two ReLU layers used after second and third pooling layers. One dropout layer is used to reduce overfitting. The last layer act as classifier corresponds to softmax layer to

predict the gender across frontal ,back and mixed views. The training of the network follows gradient descent function.

**Hyper Parameter:**

1. Four convolution layer, two fully connected layer, two ReLU layers.
2. One dropout layer with dropout rate 0.5.
3. The learning rate and momentum is 0.001 and 0.9 respectively.
4. The size of mini-batches is 128.Total epochs is 50.
5. Input image size is 112x92x1,where 1 shows the gray channel of the image.

**Merits:**

1. Performance rate is increased a bit higher than existing methods.
2. The input vectors for training are provided in the form of mini-batches to avoid burden on the system.
3. Dropout layer reduces overfitting.

**Demerits:**

1. The mean accuracy for the frontal perspective of abdominal area prediction is lower than the mean accuracy for the full-body frontal view.

## Gender Recognition from Body[7]

This papers contributions are three-fold. First attempt is to investigate gender recognition form still human body images which are taken from frontal and back views. Second, it relax the fixed view constraint and show the possibility to train a flexible classifier for mixed view case. And last, it shown its robustness to small alignment errors.

**Model:**

To remove affects of visual variations, this paper consider to employ edge maps instead of raw pixels. After image representation two classification model are used for part-based gender recognition which are

1. **Adaboost:**

   It is iteratively constructs optimal weak classifiers from reweighting training samples in each round. output is a weighted additive model combining all these weak classifiers into a strong one.

2. **Random Forest:**

   It try to maximize the independence of features, while keeping the independence classifiers s accurate as possible.

**Part-based Gender Recognition:**

It use grid sampling to partition the images into patches. One patch corresponds to some human body which can provide useful cues for recognizing gender. To combine these cues it adopt the ensemble learning algorithm which constructs a succession of weak classifiers. In each round, algorithm first selects the most discriminative patch position and computes the optimal classifier only using the patch features of the selected position. To build a combined classifier with the weighted voting of the recognition results based on different body parts.

**Gender Recognition from mixed view:**

This paper suggests that the classifiers of both share some similarity. It train PBGR algorithm on the non-fixed view images. It takes images randomly form both frontal and back views. The overall accuracy is $75.0 \pm 2.9\%$ which is very similar to the fixed view accuracy.

**Robustness Analysis:**

It train the model using original data while testing the model with misaligned images. Accuracy remains same when misalignment is 5% of the body width and the accuracy only drops to 69% subject to 15% misalignment.

**Hyper Parameter:**

1. Edge map is obtained by Canny's method.
2. Each body figure is divided into 3x3 blocks. Each block is represented by a HOG feature as a vector of length 8 describing the gradient in 8 orientations.
3. Grid is 6x9 in the experiment.

4. Final dataset is composed of 53% back and 47% frontal view images.

**Merits:**

1. This approach is robust to tolerate small misalignment errors.
2. It can build a more flexible classifier without degrading the accuracy.
3. Random forest are robust to overfitting and are also very fast to train and test.

## Pedestrian Attribute Recognition At Far Distance[8]

This paper contributes in two aspects. First, it release a new pedestrian attribute dataset which is by far the largest and most diverse of its kind. Second it proposed a noble method by exploiting the neighborhood information among image samples.

**Pedestrian Attribute Dataset:**

First, it removes erroneous or duplicated copies from each datasets. Each image is newly labeled with 61 binary and 4 multi-class attributes. This dataset can be used in visual surveillance research on pedestrian tracking, detection re-identification and activity analysis.

**Methods:**

**Baseline 1:**

SVM with intersection kernel reduces the time and space complexities of the traditional linear kernel SVM.

**Baseline 2:**

It exploit the context of neighboring images by Markov Random Field. Each random variable corresponds to an images and the relation between two variables corresponds to the similarity between images. When largest variations are presented an alternative is to employ the random forest which is adopted as unsupervised. The RF can be learned using the pseudo two-class method. One is original unlabeled test samples and another class is created by sampling at random from the univariate distributions of the unlabeled test samples.

**Parameter:**

1. A 2784-dimensional feature vector is obtained for each image.

**Hyper Parameter:**

1. Build a K-NN sparse graph by limiting the number of neighbhors for each node. And it set k=5 in experiment.
2. Horizontally partitioned the images into six strips and then extracted feature channels and each is described by bin-size of 16.

**Merits:**

1. Release a largest and most diverse dataset for PETA dataset.
2. Emphasis to mitigating the visual ambiguity of appearance features.
3. Model is capable of a accurate detecting subtle attributes which may otherwise misdetected from single image.

## An All-In-One Convolutional Neural Network for Face Analysis[9]

This paper propose a novel CNN that simultaneously performs face detection, pose estimation, age estimation, gender recognition, smile detection and face identification and verification. It design a Multi-task learning(MTL) framework for training which regularizes the parameters of the network.

**Method:**

**Multi-task Learning:**

In MTL lower layers of CNN are shared among all tasks and input domain. Lower layer shares learn features common to a general set of face analysis where more specific in upper layers to individual tasks

**Network:**

It consists of seven convolution layers and three fully connected layers. They divide the task into groups where subject-independent tasks (face detection pose Estimation and smile prediction) are done in first, third and fifth convolution layers. It attached two convolution layer and a pooling layer respectively to these layers. A dimensionality reduction layer is followed by a fully connected layer. Gender and age estimation are branched out from the sixth convolution layer after performing the max pooling operation. It keep the seventh convolutional layer unshared to adapt in specifically to the face recognition task. Gender recognition, age estimation and face recognition are trained using separate sub-network and at test time, the sub-networks are fused together into a single all in one CNN.

For training of general recognition MORPH, IMDM+WIKI datasets are used images are aligned using facial key points. To evaluate gender recognition performance on large scale Celeb faces attributes and Chalear learn faces of the world datasets. And to evaluate accuracy on in all the methods for gender accuracy and achieved state-of-the-art performance for both Gender and smile classification.

**Hyper parameters:**

1. A consistent feature map size of 6x6.
2. A dimensionality reduction layer reduce the number of feature maps to 256. It followed by a fully connected layer of dimension 2048.
3. Dimension of each fully connected layers of dimension 512.

**Merits:**

1. This model is able to learn robust features for distinct task.
2. This approach saves both time and memory because its simultaneously solve the tasks and requires the storage of a single CNN model.

## DAGER: Deep Age, Gender And Emotion Recognition Using Convolutional Neural Networks[10]

This paper describes the details of Sighthound's fully automated age, gender and emotion recognition system. It present deep networks which are not only computationally inexpensive but also outperform competitive methods on several benchmarks. It present large datasets.

**System Model:**

First deep model is trained on a large dataset of four million images for the task of face recognition. It's facial attributes recognizers used to fine-tune network for real age estimation, apparent age estimation, gender recognition and emotion recognition. It feeds data in network after pre-processing. Then trained network using four million images of over 40,000 identities.

It compared gender recognition model on the Adience benchmark with other leading methods. The Adience benchmark contains 17,492 faces labeled with their corresponding gender. It compare its method with a couple of commercial APIs and Sightbound gets 91.00% accuracy.

## Identifictaion Of Pedestrian Attributes Using Deep Network[11]

This paper proposed a simple network architecture for attributes identification. Also it proves that tailored network perform better than general purpose network like AlexNet, ResNet, inception network. After experiments the results shows that between 5 types of machine learning methods ,this network works well with SVM classifier compared with ResNet and Inception network.

**Network:**

Proposed network consist of three convolution layers and ten task specific classifiers like gender,top-clothing,boots,hat,backpack,bag,handbag,shoes,upper-body color and lower body color. This classification of attributes follows the description of DukeMTMC – attributes

dataset. This proposed network utilize multi-task learning strategy. The categorical cross-entrophy loss is utilized.

**Hyper parameter**:

1. Takes RGB image with size 128x64 pixel.
2. Two convolution layer use 5x5 filter , ReLU activation function and max-pooling size 2x2.
3. For each classifier, an independent hidden layer with 32 neuron utilized. The amount of neurons at final layer is two for gender, seven for lower-body color.

**Merits**:

1. The system is trained so that the resulting features in flatten layer are distinguishable by the classification layers for all the task.
2. The network utilize smaller amount of parameters, so that it requires smaller computational loads.

**Demerits:**

1. Due to self-occlusion or color uncertainty some attributes cannot be well identified.

### 3.2 Pros and Cons of Convolutional Neural Network:

**Pros:**

1. CNS are capable of learning features on their own.It replaces manual feature engineering and does it automatically which saves time, cost and effort.
2. Parameter sharing is an important feature of CNN.CNNs take advantage of local spatial coherence in the input.
3. The main advantage is accuracy in image recognition problems.
4. CNNs are more efficient in terms of complexity and memory.
5. An already train CNN model can be used by fine-tuning for a specific task and thus saves time and memory.

**Cons:**

1. CNNs has to take a large dataset for training so it's mostly computationally expensive and has overfitting problems.
2. It needs a good GPU to train faster.
3. Hyper-parameter tuning is non-trival.

### 3.3 Existing Methodologies to be compared to Proposed Network:

**Age and Gender Classification using Convolutional Neural Networks [1]:** In this paper they used three convolutional layer and two fully connected layer with small number of neurons. They rescaled images to 256x256 and a crop of 227x227.Their desire is to reduce overfitting with their limited learning rate. They used gradient descent with batch size of fifty images. They choose a smaller network to reduce the risk of overfitting which made it shallow compared to other network. In our work, we used flipping and rotating augmentation. Our batch size is hundred and image size is 75x75 with 3 RGB channel.

# Chapter 4

# Methodology

## 4.1 Multi-label Classification:

In machine learning, multiclass or multinomial classification is the problem of classifying instances into one of three or more label.

Multilabel classification makes the assumption that each sample is assigned to one and only one label.

There are two basic approaches for multiclass classification problems. One involves constructing and combining several binary classifiers and the other involves considering all data in one optimization formulation. The latter approach the formulation to solve multiclass support vector machine problems in one step has a variables proportional to the number of classes. It is computationally more expensive to solve multiclass problem than a binary problem with the same amount of data.

In neural networks, multilabel perceptrons provide a natural extension to the multilabel problem. Instead of just having one neuron in the output layer with binary output, one could have N binary neurons leading to multilabel classification.

**Multilabel Classification Transformation:**

We used transformation to binary techniques. It can be categorized into one versus Rest and One versus One. The techniques developed based on reducing the multilabel problem into multiple binary problems can also be called problem of transformation techniques.

### 4.2 Data Preprocessing

#### 4.2.1 Adience Dataset

The dataset we are going to use in our research work is Adience dataset. It contains 19322 images which resolution are 816 by 816,horizontal and vertical resolution is 96 dpi and 24bit. Images are close-view image. Images are divided into 30 different age class and 3 gender class.

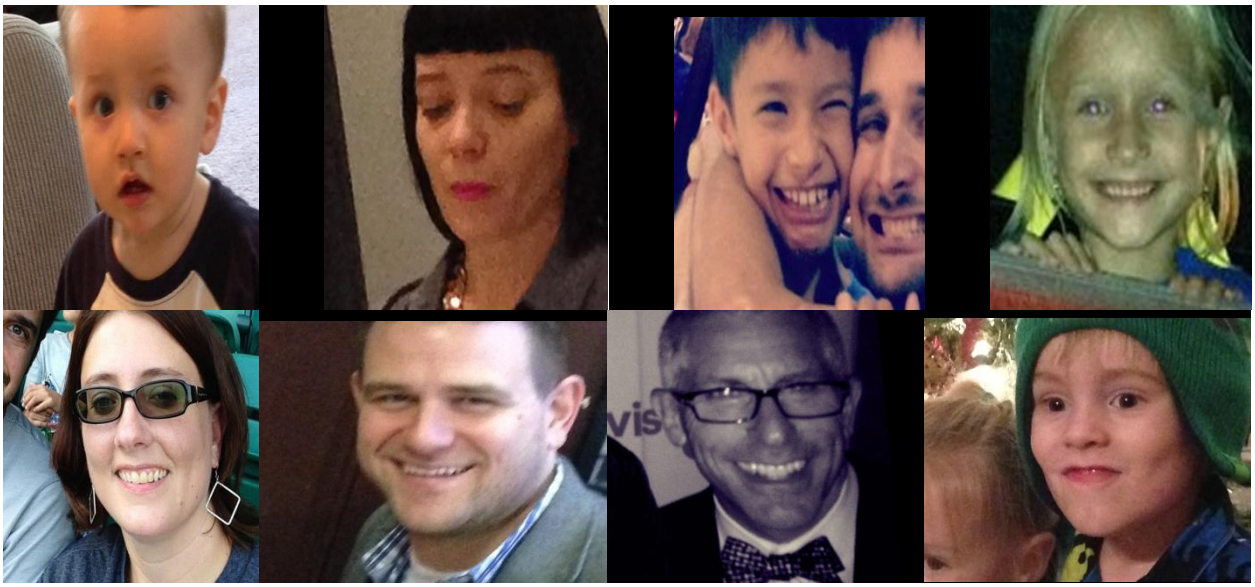


Figure 4 : Some image from different attribute class from Adience Dataset

From those class, we arrange age classes into 8 class and 2 gender class. Our age classes are

- **0-2**
- **4-6**
- **8-12**
- **15-20**
- **25-32**
- **38-43**
- **48-53**
- **60-100**

Our gender classes are:

- **Male**
- **Female**

We first divided images into age classes and then according to the age divided images into two gender class.

**Adience dataset URL:** https://www.kaggle.com/ttungl/adience-benchmark-gender-and-age-classification

The dataset is officially available at the link above for research purpose only.

### 4.2.2 Downloaded Files Information

The Adience dataset was available as a zip file containing all the images of all datasets and label text files that contained the labels for corresponding images. But the images were arranged in 5 different folders for 5 datasets and 5 csv file for the respective images. So, this required a lot of preprocessing steps to fix the labels and image files before using them.

### 4.2.3 Label and Image Files Preprocessing:

**a) Image Rearranging to the class:** Between 30 age class we create 8 age class and from 3 gender class we take 2. We include all the age class in those 8 class.

**b) Image Converting:** First read all the csv file convert the image into channel 1 image.

**c) Image Scaling:** After converting the image read the csv file and divide with 255.0, minus 0.5 and multiply 2.0 for scaling the image.

**d) Image Resizing:** After scaling resize the original image into 75 by 75 image and give the image id from 1 to increase plus 1 for next image.

**e) Create Separate Pickle Files:** After Resizing the image, create 5 pickle file for each of 5 csv file. Then read the pickle files and make a directory.

**f) Create Separate label Files:** By reading the csv file the image divided into different folder which are divided into firstly as age class and then into male and female class.

**g) Strip Newline:** When reading every label line from the final label text file, the newline characters should be stripped from the end of every line to prevent extra labels from showing up.

**h) Image Renaming:** We rename image according to the label of age class and gender and then add a random number and all the image is in jpg format. Image file name is like : **(0-2)_m_503.jpg** which means age is 0 to 2 and it is a male.

**i) Image Augmentation:** We use ImageDataGenerator for image augmentation. Here rotate range is 25, width-shift-range is 0.1, right-shift-range is 0.1, shear-range 0.2, zoom-range is 0.2 and flip the image horizontaly. For augmentation the number of image is increased double.

Figure 5 : Some Preprocessed image from all classes

### 4.5 Proposed CNN Model

We experimented with different CNN model configurations. All of the experimented models contained filters of 3x3 size. We have trained the baseline model with our dataset and named it as NewNet11.The training and testing accuracy for the different configurations are given below.

Table 1 : Test and Training Accuracy for Different Model Configuration Table

| Model | Filter | Dense Layer | Parameters | Train Accuracy(%) | Test Accuracy (%) |
|---|---|---|---|---|---|
| NewNet 1 | 64-128-256 | 1024 - 0.5<br>512 - 0.3 | 76,665,226<br>76,663,306 | 93.46 | 91.24 |
| NewNet 2 | 64-128-256 | 1024 - 0.5<br>512 - 0.4 | 76,665,226<br>76,663,306 | 93.39 | 89.19 |
| NewNet 3 | 32-64-128-256 | 1024 - 0.5<br>512 - 0.3 | 17,700,554<br>17,698,570 | 94.39 | 92.16 |
| NewNet 4 | 64-128-256 | 1024 - 0.5<br>256 - 0.3 | 76,399,242<br>76,397,834 | 93.54 | 91.84 |
| NewNet 5 | 64-128-256 | 1024 - 0.5<br>256 - 0.4 | 76,399,242<br>76,397,834 | 93.28 | 91.57 |
| NewNet 6 | 64-64-128-256 | 1024 - 0.5<br>256 - 0.4 | 17,454,026<br>17,452,490 | 94.50 | 92.15 |
| NewNet 7 | 32-64-128-256 | 1024 - 0.5<br>256 - 0.4 | 17,434,570<br>17,433,098 | 94.55 | 91.90 |
| **NewNet 8** | **32-64-128-256** | **1024 - 0.5<br>256 - 0.3** | **17,434,570<br>17,433,098** | **94.60** | **92.24** |
| NewNet 9 | 32-64-128-256 | 1024 - 0.4<br>512 - 0.3 | 17,700,554<br>17,698,570 | 94.56 | 91.85 |
| NewNet 10 | 32-64-128-256 | 1024 - 0.4<br>256 - 0.3 | 17,434,570<br>17,433,098 | 94.39 | 92.15 |
| NewNet 11 | 64-128-256 | 1024 - 0.4<br>1024 - 0.3 | 77,197,194<br>77,194,250 | 93.17 | 91.78 |
| MiniVGG Net | 32-32-64-64 | 512-0.5 | 10,690,858<br>10,689,450 | 91.01 | 91.11 |

From all these network NewNet3, NewNet8 and NewNet10 performed better. But we select model 8 which is named here as NewNet 8 because of it gives higher training and testing accuracy with small number of parameter where all other takes huge number of parameter and gives low training and testing accuracy. So our proposed model is NewNet 8.

Now we will discuss the proposed convolutional network architecture. It consists of four convolutional layers and two fully connected layers which use the extracted features from the previous convolutional layers and learn from it and thus performs the classification at the output layer.

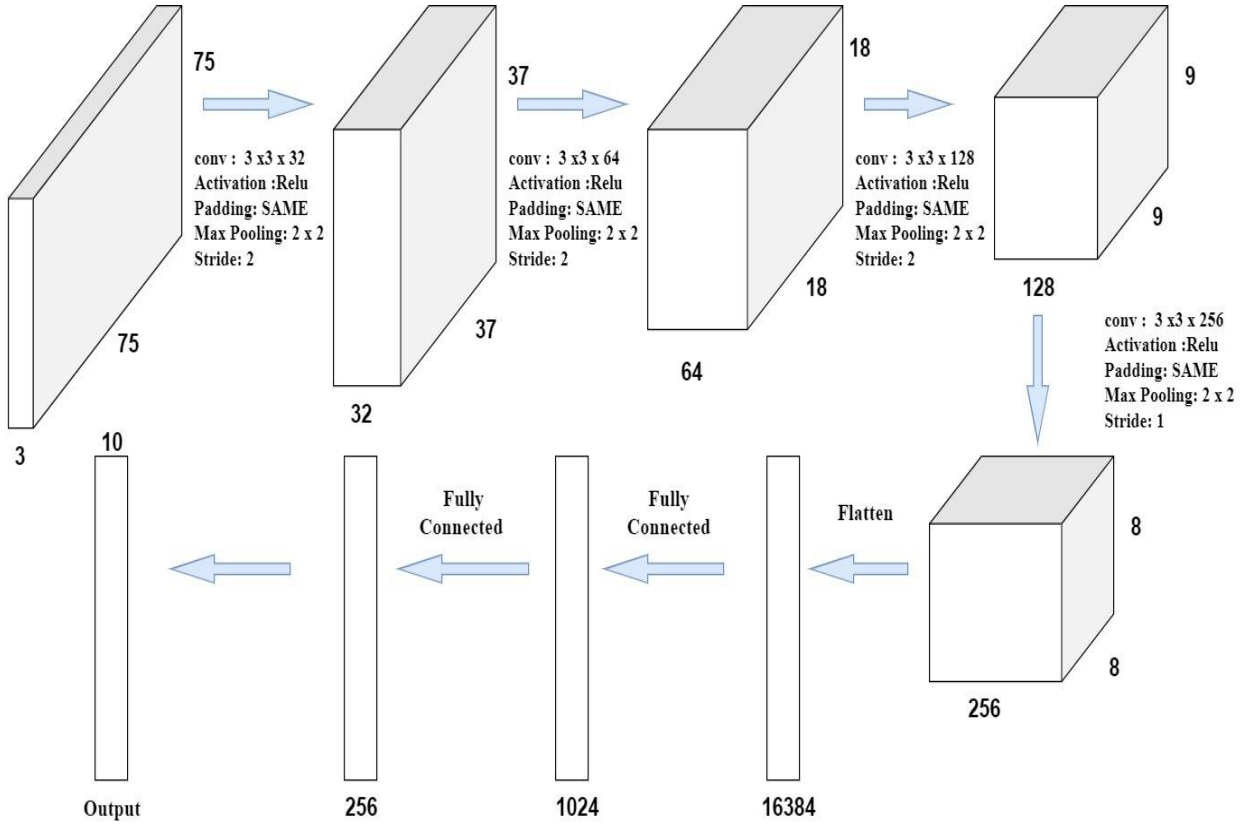



Figure 6 : Proposed Convolutional Neural Network model architecture

close view RGB images of size 75 x 75 with 3 channels will be input to this convolutional neural network model.

**1st Conv Layer:** The first convolutional layer consists of 32 filters of size 3 x 3 and 3 channels each. Here the SAME padding is used i.e. the input images are zero padded in such a way that the filters convolve over every pixel of the input image. For SAME padding the output image after convolution is the same as the input image. The 32 filters convolve the input images of size 75 x 75. Because of the SAME padding, the output feature maps after convolution are of size 75 x 75 and 32 channels.

**ReLu:** The output feature map is run through a ReLU activation function to introduce non-linearity. This activation function turns any negative value into zero. So, it makes every value non-negative.

**1st Max pool:** The max pooling filter is of size 2 x 2 and it is used to reduce the dimensionality of the feature maps. This max pooling filter moves by a stride of 2. And the output image is 37x37x32.

**2nd Conv Layer:** The second convolutional layer consists of 64 filters of size 3 x 3 and 64 channels each. Here the SAME padding is also used. 64 filters convolve the input images of size 37 x 37 and the output feature maps after convolution is of the size 37 x 37 and 64 channels.

**ReLu:** Relu activation function is also applied to the feature maps.

**2nd Max Pool:** Max pooling filter of size 2 x2 which moves by a stride of 2, perform max pooling operation on the feature maps and output after this are images of size 18 x 18 and 64 channels.

**3rd Conv Layer:** The second convolutional layer consists of 128 filters of size 3 x 3 and 128 channels each. Here the SAME padding is also used. 128 filters convolve the input images of size 18 x 18 and the output feature maps after convolution is of the size 18 x 18 and 128 channels.

**ReLu:** Relu activation function is also applied to the feature maps.

**3rd Max Pool:** Max pooling filter of size 2 x2 which moves by a stride of 2, perform max pooling operation on the feature maps and output after this are images of size 9 x 9 and 128 channels.

**4th Conv Layer:** The 3rd convolutional layer is where things work a bit differently. Here the filters are of size 3 x 3. This is used for convolving the features from pedestrian images in a vertical manner as all the images are of vertically shown pedestrians. This non-symmetrical size of the filter helps in finding discriminative features from the pedestrian images to detect the attributes. The SAME padding is also used here. So, after convolution, the output future maps are of the size 9 x 9 and 256 channels.

**Relu:** Relu activation function is also applied to the output feature maps.

**4th Max Pool:** Max pooling filter of size 2 x2 which moves by a stride of 1, perform max pooling operation on the feature maps and output after this are images of size 8 x 8 and 256 channels.

**FC – 1024:** The result of the last pool layer is flattened into a vectorized form containing 16,384 elements (8 * 8 * 256) which are followed by a fully connected layer of 1024 neurons. Every element of the vector is fully connected to the 1024 neurons of this layer. Fully connected layers are used to use the final features extracted from the previous convolution and pooling layers to classify the image into the respective classes in the output layer. These FC layers learn from the features.

**Dropout:** A dropout layer is used for dropping out the neurons for preventing the model from overfitting to the training data. We applied dropout on fully connected layers. There is a total of two dropouts used in this model. The dropouts are

Dropout 1= 0.5

Dropout 2= 0.3

**Batch Normalization:** The batch normalization is used to improve the model performance and for stability. It normalizes the input of each layer. Here, batch normalization is 100.

**FC-256:** Another fully connected layer is used to improve the learning and increase the accuracy of the prediction. This layer consists of 1024 neurons.

**Output Layer:** FC-256 is fully connected to the output layer which consists of 10 nodes representing 10 class of the dataset. The attribute score for each of the attribute is then computed in the 10 nodes respectively.

This is how the proposed CNN model shall work with the far view images we provide to this model.

# Chapter 5

# Experiments and Results

## 5.1 Dataset and experimental settings

As we discussed in previous, the dataset we used to train and test our model is Adience dataset. It contains close-view images. Adience dataset contains 19,322 images labeled with their gender and age. Faces are divided into 5 folds. Ages are divided with different group which is 0-2, 4-6, 8-13, 15-20, 25-32, 38-43, 48-53 & 60+ and gender are divided into Male and Female class. Original dataset has 30 age label but we cut the label 'none' and blank . And include other age label into our 8 age class.It captures variations in pose, light, appearance and more. We use 12,272 of images from dataset for training and 3068 for testing. Images original size is 816 by 816 and we used those image by resizing into 75 by 75. We train our dataset with fitGenerator.

## 5.2 Experimental Tools and Environment

For doing our research work efficiently and correctly from preprocess our dataset to evaluating the performance of our model, we used some packages, tools and development environment.

### 5.2.1 Programming Language

We need python language for converting easily our ideas into code. Python is a programming language which is interpreted, high-level. It has the largest collection of packages for implementing machine learning algorithm. It is the most mature and well-supported among programming language in the area of machine learning. To achieve greater productivity with systematic efforts. For implementing computer vision python allows develops to automate tasks that involve visualization. We used many packages and libraries of python like numpy, scikit, opencv, tensorflow, pytorch, pandas and so on.

### 5.2.2 Integrated Development Environments

**Anaconda:**

Anaconda is a free and open source distribution in python. It is also R distribution. It's main focus to provide everything for data science and machine learning application. It have more than 150 packages included with python core language.it also provides anaconda cloud. It is a environment manager. Conda is the package management system of Anaconda. Anaconda comes with a navigator which install and update package in environment and also search packages in local anaconda and anaconda cloud.

**Spyder IDE:**

Spyder is an open source, cross platform Integrated Development Environment. It is a numberical computing environment credit goes to ipython and python libraries like scipy, numpy, pandas etc. They have convenient data transformation functions that will same time. It provides tools for data inspection. It's highlights syntax, provide python and ipython console and so on. All our data and image preprocessing codes were run in spyder IDE.

**Jupyter:**

The jupyter notebook is an open-source web based interface that allows you to create and share documents that contain live code, equations, visualizations and all. The power of jupyter notebook is gives live code and visualization in one place. It's become more easier to share code to others. Jupyter notebook supports 40 programming languages including python, R. It's specially tailored for working with deep learning libraries like Tensorflow, keras and others.

### 5.2.3 Libraries:

**Keras:**

Keras is a high-level libraray which is an interface that wraps multiple frameworks like Theano , TensorFlow and CNTK. Kears is the fastest track when we need to quickly train and test a model from standard layers. Keras can define a network, compile network, fit network, evaluate network and use it to make prediction. Keras has strong multi-GPU support also provide

distributed training support. Keras provide the facilities to write custom building blocks easily. It minimizes the number of user action and provide feedback upon user error.

**Scikit-learn:**

Scikit-learn is an open-source, commercially usable simple and efficient tool in machine learning. It uses extensively python libraries like matploltib, plotly, numpy, pandas, scipy etc for high performance linear algebra and array operations.

**Numpy:**

Numpy is the library for computing in python. It provides a high-performance multidimensional array object and tools for working with arrays.

**Tensorflow:**

Tensorflow is an open source library for numerical computation and large-scale machine learning. It makes easier to acquire data, training model, solving predictions and refining future. It is created by google. It makes machine learning and deep learning models and algorithm together.

## 5.3 Implementation Details

In our proposed model we used Adience dataset for evaluating our model using keras Python library. It contains all the method for optimize and loss function.

- We used Adam optimizer and binary cross entropy for loss function.
- Same padding used for every convolutional and max pooling layer.
- Batch size is 100 and number of epoch is 50 for every experimental run.
- Dropouts are 0.5 and 0.3 after two fully connected layer.
- Batch normalization are used for every convolutional layer.

We preprocessed the Adience dataset and augmented the dataset. Then we use 12,272 of those image for training and 3068 for testing purpose.

### 5.4 Experimental Results

As we discussed before we experiment 11 model with our dataset and we get different accuracy with different number of parameter. We build 10 model from our own with several number of convolutional layer and parameter. From all those, we select our proposed model as NewNet 8 because of its high training and testing accuracy.

#### 5.4.1 Classification Report Comparison between 11 Model

Table 2: Classification report Comparison

| Network | Class | 0-2 | 4-6 | 8-12 | 15-20 | 25-32 | 38-43 | 48-53 | 60-100 | M | F | Accuracy (%) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| NewNet 1 | **Precision(%)** | 73 | 79 | 74 | 56 | 62 | 57 | 52 | 83 | 92 | 90 | 78 |
| | **Recall(%)** | 91 | 62 | 72 | 67 | 78 | 48 | 18 | 54 | 88 | 93 | |
| | **F1-score(%)** | 81 | 69 | 73 | 61 | 68 | 52 | 26 | 65 | 90 | 91 | |
| NewNet 2 | **Precision(%)** | 77 | 70 | 79 | 88 | 63 | 38 | 32 | 50 | 83 | 92 | 72 |
| | **Recall(%)** | 72 | 69 | 51 | 22 | 61 | 73 | 26 | 61 | 92 | 82 | |
| | **F1-score(%)** | 75 | 69 | 62 | 35 | 62 | 50 | 29 | 55 | 87 | 87 | |
| NewNet 3 | **Precision(%)** | 84 | 86 | 66 | 70 | 64 | 66 | 47 | 81 | 89 | 96 | 80 |
| | **Recall(%)** | 85 | 66 | 82 | 52 | 83 | 52 | 45 | 54 | 96 | 89 | |
| | **F1-score(%)** | 85 | 74 | 73 | 60 | 72 | 58 | 46 | 65 | 92 | 92 | |
| NewNet 4 | **Precision(%)** | 85 | 78 | 72 | 65 | 63 | 62 | 49 | 67 | 89 | 94 | 79 |
| | **Recall(%)** | 81 | 73 | 75 | 57 | 79 | 47 | 44 | 63 | 94 | 89 | |
| | **F1-score(%)** | 83 | 75 | 74 | 61 | 70 | 54 | 46 | 65 | 91 | 91 | |
| NewNet 5 | **Precision(%)** | 78 | 85 | 69 | 51 | 67 | 60 | 58 | 73 | 89 | 93 | 79 |
| | **Recall(%)** | 89 | 62 | 76 | 73 | 71 | 56 | 26 | 49 | 93 | 90 | |
| | **F1-score(%)** | 83 | 71 | 73 | 60 | 69 | 58 | 36 | 59 | 91 | 91 | |
| NewNet 6 | **Precision(%)** | 91 | 79 | 77 | 72 | 62 | 59 | 49 | 64 | 94 | 92 | 80 |
| | **Recall(%)** | 76 | 72 | 69 | 48 | 79 | 62 | 42 | 71 | 91 | 95 | |
| | **F1-score(%)** | 83 | 75 | 73 | 58 | 69 | 60 | 45 | 67 | 92 | 93 | |
| NewNet 7 | **Precision(%)** | 88 | 83 | 71 | 71 | 64 | 68 | 40 | 62 | 87 | 95 | 79 |
| | **Recall(%)** | 84 | 71 | 77 | 50 | 84 | 44 | 38 | 68 | 95 | 87 | |
| | **F1-score(%)** | 86 | 77 | 74 | 59 | 72 | 55 | 39 | 65 | 91 | 91 | |
| NewNet 8 | **Precision(%)** | 88 | 79 | 69 | 61 | 68 | 63 | 62 | 71 | 92 | 93 | 81 |
| | **Recall(%)** | 79 | 75 | 79 | 70 | 75 | 58 | 34 | 54 | 93 | 93 | |
| | **F1-score(%)** | 83 | 77 | 74 | 65 | 71 | 60 | 44 | 61 | 92 | 93 | |
| NewNet 9 | **Precision(%)** | 73 | 81 | 79 | 73 | 66 | 68 | 51 | 51 | 95 | 88 | 80 |
| | **Recall(%)** | 90 | 67 | 70 | 60 | 79 | 53 | 47 | 77 | 86 | 95 | |
| | **F1-score(%)** | 81 | 74 | 74 | 66 | 72 | 60 | 49 | 61 | 90 | 92 | |
| NewNet 10 | **Precision(%)** | 93 | 74 | 82 | 71 | 60 | 65 | 46 | 67 | 90 | 95 | 80 |
| | **Recall(%)** | 70 | 80 | 71 | 49 | 83 | 49 | 49 | 62 | 95 | 90 | |

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | **F1-score(%)** | 80 | 77 | 76 | 58 | 70 | 55 | 47 | 64 | 92 | 92 | |
| NewNet 11 | **Precision(%)** | 74 | 79 | 60 | 59 | 72 | 58 | 64 | 61 | 92 | 92 | 79 |
| | **Recall(%)** | 84 | 65 | 81 | 71 | 68 | 54 | 17 | 70 | 91 | 93 | |
| | **F1-score(%)** | 79 | 71 | 69 | 62 | 70 | 56 | 27 | 65 | 92 | 92 | |
| MiniVGG Net | **Precision(%)** | 82 | 77 | 79 | 71 | 55 | 48 | 41 | 70 | 93 | 89 | 76 |
| | **Recall(%)** | 79 | 69 | 48 | 41 | 78 | 57 | 26 | 59 | 88 | 94 | |
| | **F1-score(%)** | 80 | 73 | 60 | 52 | 64 | 52 | 32 | 64 | 90 | 92 | |
| | **support** | 250 | 396 | 411 | 357 | 824 | 501 | 174 | 155 | 1464 | 1604 | 6136 |

The table shows all the networks precision, recall, F1-score and support in percentage. Support is same for all the network. For NewNet 8 we get precision for age label (0-2),(4-6),(8-12),(15-20),(25-32),(38-43),(48-53),(60-100) is 88%, 79%, 69%, 61%, 68%, 63%, 62%, 71% respectively and for gender label Female and male is 93%, 92%..Recall for age label (0-2),(4-6),(8-12),(15-20),(25-32),(38-43),(48-53),(60-100) is 79%, 75%, 79%, 70%, 75%, 58%, 34% respectively and for gender label Female and male is 93% and 93%. F1-score for age label (0-2), (4-6),(8-12),(15-20),(25-32),(38-43),(48-53),(60-100) is 83%, 77%, 74%, 65%, 71%, 60%, 44%, 61% respectively and for gender label Female and male is 93% and 92%. This values are compare with all other network which is better than other. That's why we choose NewNet 8.

The following graph shows the training and validation accuracy for 11 model and MiniVGG Net.


Figure 7: Training and Validation Accuracy Curve for 11 model and MiniVGG Net

The graph shows that, all the training and validation accuracy for 11 network and miniVGG net. The training accuracy for NewNet1, NewNe2, NewNet3, NewNet4, NewNet5, NewNet6, NewNet7, NewNet8, NewNet9, NewNet10, NewNet11, MiniVGG Net is 93.46%, 93.39%, 94.39%, 93.54%, 93.28%, 94.50%, 94.55%, 94.60%, 94.56%, 94.39%, 93.17%, 91.01% respectively. The testing accuracy for NewNet1, NewNe2, NewNet3, NewNet4, NewNet5, NewNet6, NewNet7, NewNet8, NewNet9, NewNet10, NewNet11, MiniVGG Net is 91.24%, 89.19%, 92.16%, 91.84%, 91.57%, 92.15%, 91.90%, 92.24%, 91.85%, 92.15%,91.78%, 91.11%. From all this we see that accuracy is high for NewNet 8.

The following graph shows the training and validation loss for 11 model and MiniVGG Net.


Figure 8 : Training and Validation Loss Curve for 11 model and for MiniVGG Net

The graph shows that, all the training and validation loss for 11 network and miniVGG net. The training accuracy for NewNet1, NewNe2, NewNet3, NewNet4, NewNet5, NewNet6, NewNet7, NewNet8, NewNet9, NewNet10, NewNet11, MiniVGG Net is 16.23%, 16.49%, 13.94%, 15.98%, 16.45%, 13.68%, 13.46%, 13.41%, 13.68%, 13.91%, 16.60%, 21.16% respectively. The testing accuracy for NewNet1, NewNe2, NewNet3, NewNet4, NewNet5, NewNet6, NewNet7, NewNet8, NewNet9, NewNet10, NewNet11, MiniVGG Net is 23.12%, 29.39%, 21.20%, 20.85%, 21.75%, 21.12%, 22.45%, 20.45%, 22.83%, 21.17%, 22.06%, 21.79%. From all this we see that NewNet 8 gives less loss then other network.

The following curve shows the training and validation loss for NewNet 8 which is our proposed model.


Figure 9 : Training and Validation Loss Curve for proposed model

The following curve shows the training and validation accuracy for NewNet 8 which we choose as our proposed model.


Figure 10 :Training and Validation Accuracy Curve for proposed model

Our model gives 94.60% accuracy for training and 92.24% accuracy for testing with four convolutional layer and two fully connected layer with small amount of parameter which is better then all 11 model.

The following graph shows the confusion matrix of our model.

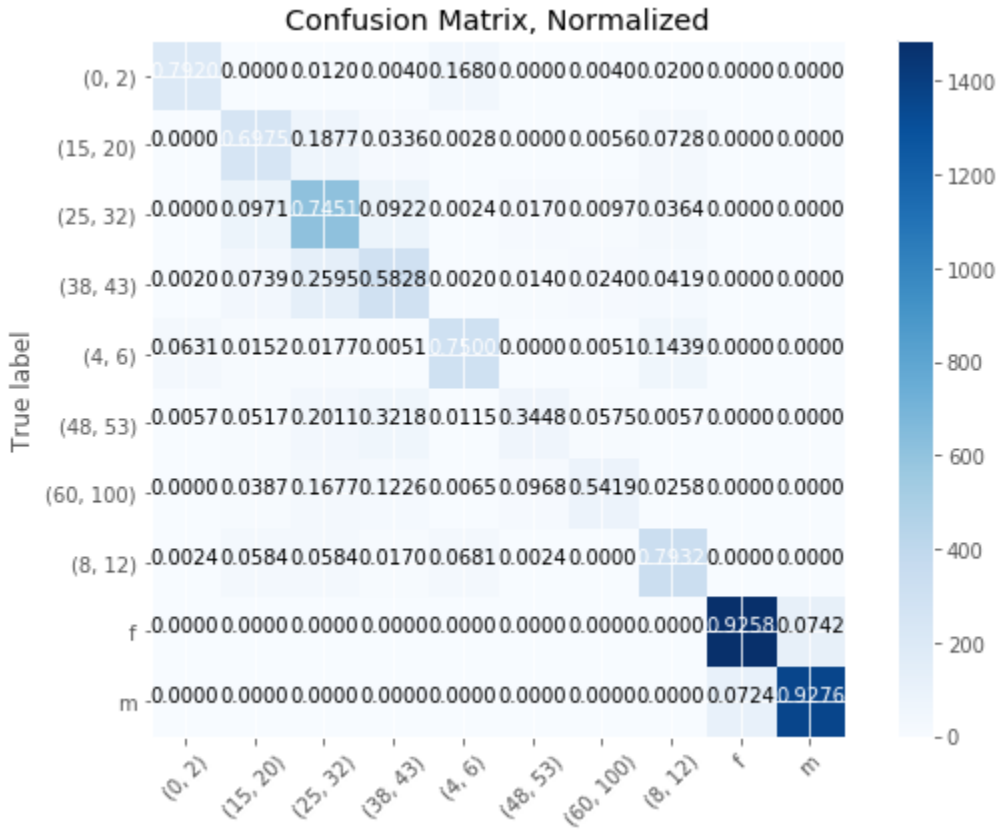


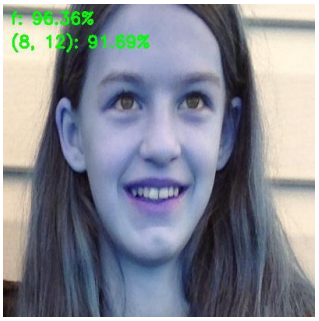

Figure 11 : Confusion Matrix of our proposed model

From Figure, it is noticed that the (8-12) year old age label is estimated with the highest accuracy 79.32%. (48-53) and (60-100) age labels classified with lowest accuracies 34.48% and 54.19% respectively. Label 6 and label 7 is highly misclassified with label 4 and label 3 which are (38-43) age groups. For label 6,misclassify rate is with label 4 is 32.18% and with label 3 is 20.11%. For label 7, misclassify rate with label 4 is 12.26% and with label 3 is 16.77%. These result might be a result of the difference in subject numbers between labels 6 and 7 and the label 3 and label 4. (0-2) age label is estimated with 79.20% and (4-6) age is estimated with 75% accuracy. (15-20) label is misclassified with label 3 of 18.77%. the number of subjects in label 5 is very large compare to other labels. For gender label male is classified with highest accuracy with 92.76% and gender label female is classified with accuracy 92.58%. Male and female gender misclassified with each other 7%.

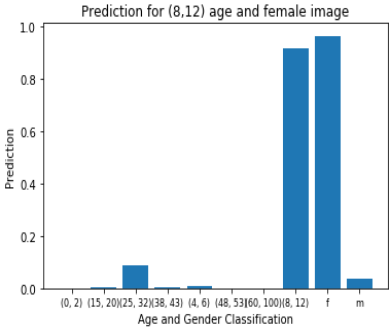
### 5.4.2 Some Processed example with our proposed model

The following sample images is collected from the image of different class which are classified correctly.

(a)

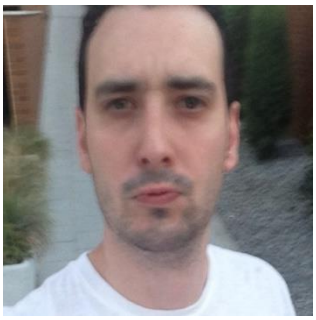



(b)

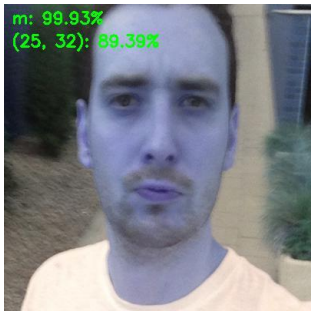



(c)

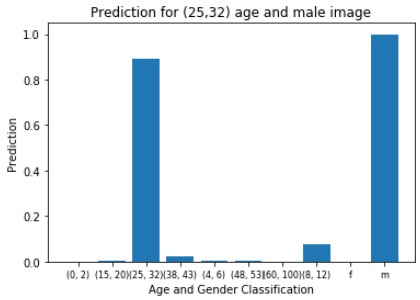
Sample #1



(a)

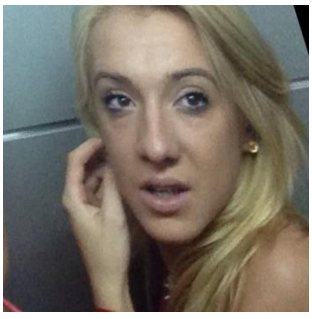



(b)



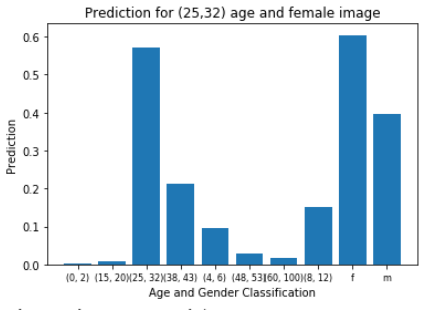

(c)

Sample #2



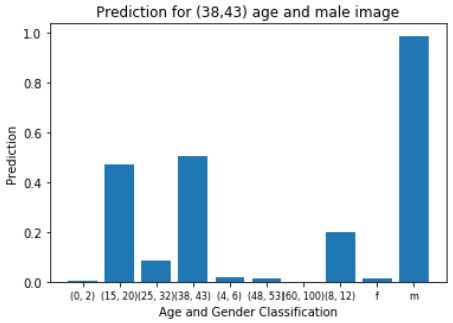
(a)


(b)



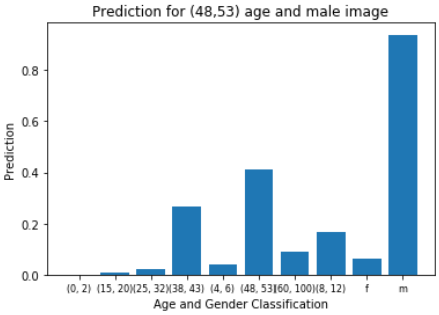

(c)

Sample #3

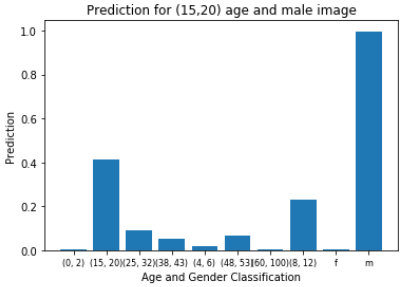

(a) (b) (c)

Sample #4


(a) (b) (c)

Sample #5

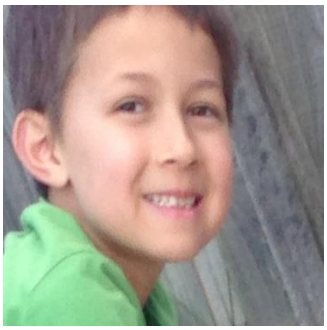

(a) (b) (c)

Sample #6

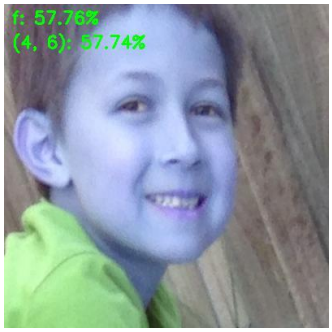
Figure 9 : (a) Original Image, (b) Tested Image, (c) histogram of the image. Some samples of images with correct identification of every attribute with their histogram

Sample 1 is an image where age label (8-12) and gender label Female. Sample 2 is an image where age label (25-32) and gender label Male. Sample 3 is an image where age label (25-32) and gender label Female. Sample 4 is an image where age label (38-43) and gender label Male. Sample 5 is an image where age label (48-53) and gender label Male. Sample 6 is an image where age label (15-20) and gender label Male. All the image predicted correctly for both age and gender attribute.

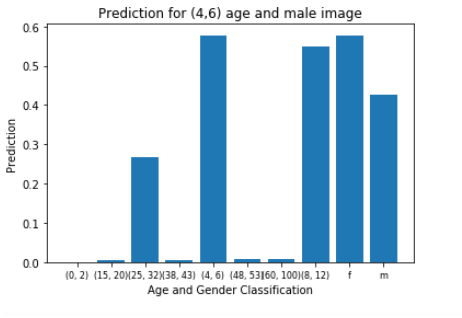
The following shows the image from different class which are misclassified .



(a)




(b)

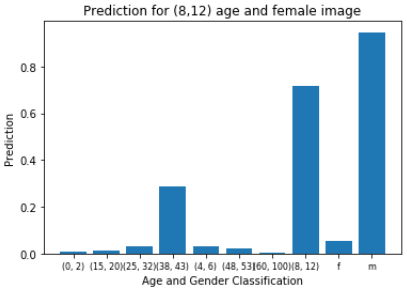



(c)

Sample #7

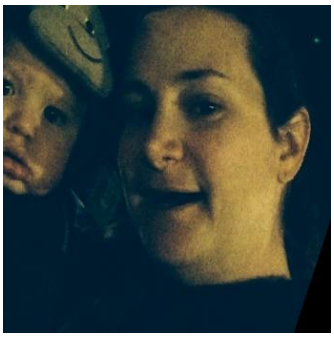
(a)


(b)



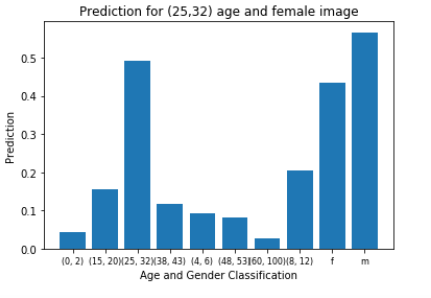

(c)

Sample #8



(a)


(b)

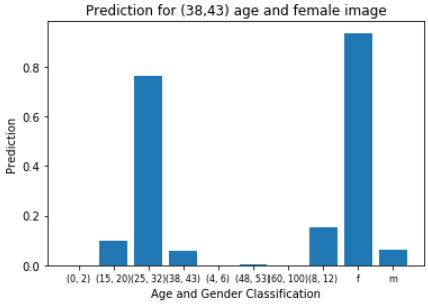



(c)

Sample #9

(b)

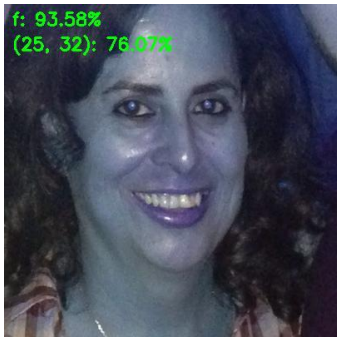

(b)



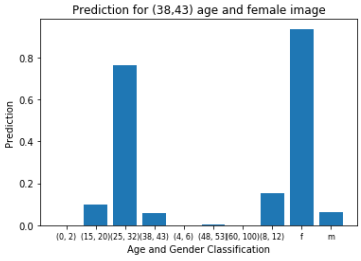

(c)

Sample #10

(a)

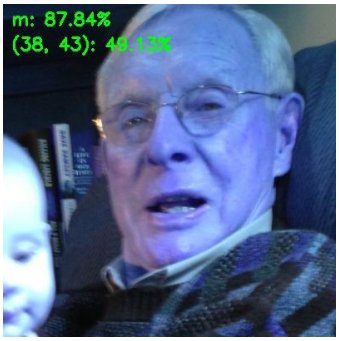


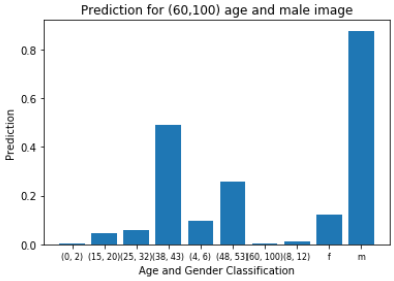

(b)




(c)

Sample #11



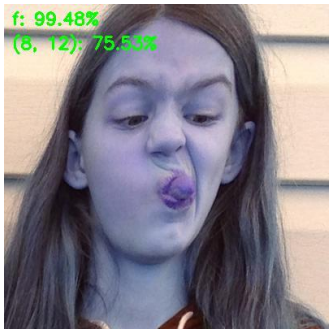

(a) (b) (c)

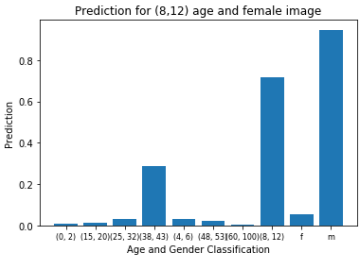
Sample #12

Figure 10 : (a) Original Image, (b) Tested Image, (c) histogram of the image. Some samples with the most incorrect identification of attributes with their histogram

We provide a few examples of both gender and age misclassifications. Sample 7,8,9 are misclassified in geder prediction. Sample 10,11,12 are misclassified in age prediction. From all of these results we see that our model can predict gender label more accurately than the age label. These show that many of the mistakes made by our system are due to extremely challenging viewing conditions of some of the Adience benchmark images. Most notable are mistakes caused by blur or low resolution and occlusions.

The following images are difficult but our model classified its correctly age and gender.




(a) (b) (c)

Sample #11


(a) (b) (c)

Sample #12

Figure 11 : (a) Original Image, (b) Tested Image, (c) histogram of the image. Some difficult samples which are correctly classified with their histogram

Sample 11 is an image of (8-12) Female image. Sample 12 is an image of (25-32) Male image. Though the image are difficult to predict but our model prdict age and gender both correctly.

### 5.4.3 Comparision of our proposed model with State of the Art:

In [4], they've used Adience dataset and their model is generic Alex-Net like architecture and domain specific VGG-face CNN model. In addition task specific GilNet CNN model has also been utilized and used as a baseline method in order to compare with transfer model. In [20], a new CNN based method where leveraging Networks of residential networks. ROR model is pre-trained on image net firstly and then used to fine-tuned on the IMDB-WIKI101 dataset for further learning and finally used to fine tune on Adience dataset. Finally the ROR-152+IMDB-WIKI101 with two mechanisms achieve new state of the art results on Adience benchmark. Age cross validation result is for two mechanisms where exact accuracy 67.34±3.56 and 1- off accuracy 97.51±0.67 and gender cross validation result is for ROR-34+IMDB-WIKI 93.24±1.77.

VGG-16 and MiniVGG Net gives 71.5±3.225 and 91.11 accuracy respectively with Adience Dataset.

Table 3 : Comparison with state of the art results of age and gender classification on Adience

| Method | Exact Accuracy (%) |
|---|---|
| VGG-16[20] | 71.5±3.225 |
| ROR-34+IMDB-WIKI [20] | 79.99 ±2.23 |
| GilNet[1] | 68.75±3.35 |
| Ft- VGG-Face+SVM [4] | 74.95 |
| MiniVGG Net | 91.11 |
| Baseline Network | 91.78 |
| NewNet 8 | 92.24 |

We use MiniVGG Net, the baseline Network on the Adience dataset. The results are shown in the table 3 along with VGG-16[20], ROR-34+IMDB-WIKI [20], GilNet[1], Ft- VGG-Face+SVM [4] state of the art. Our proposed model NewNet8 achieves a competitive accuracy 92.24% by only pretraining on which outperforms other networks and some state of the art.

### 4.6 Discussion

#### 4.6.1 Model Results for 10 attribute

- The proposed model when predicting 10 attribute from Adience dataset it's training accuracy is 94.60% and testing accuracy is 92.24%.
- It gave comparatively less accuracy in predicting age.
- It gave 92% accuracy for male and 93% accuracy for predicting female.
- The proposed model gives accurate prediction mostly for (25-32), (8-12) class for both male and female class and (4-6) for female (15-20),(38-43), (48-53) for male class.

# Chapter-6

# Conclusion and Future Work

## 6.1 Conclusion

In this research work, we proposed a convolutional neural network for classifying gender and age from Adience dataset. We achieved 92.24% accuracy on the dataset which is used for predicting 8 attributes of age and 2 attributes of gender. Except a few attributes our model is affecting age and gender

## 6.2 Future Work

For future work ,We plan to implement improvements to our CNN architecture and explore the use of unsupervised pre-training to improve it's classification performance.

# References

1. Levi, G. and Hassncer, T. (2015). Age and gender classification using convolutional neural networks. 2015 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW).

2. Toshev, A. and Szegedy, C. (2014). DeepPose: Human Pose Estimation via Deep Neural Networks. 2014 IEEE Conference on Computer Vision and Pattern Recognition.

3. Sun, Y., Wang, X. and Tang, X. (2014). Deep Learning Face Representation from Predicting 10,000 Classes. 2014 IEEE Conference on Computer Vision and Pattern Recognition.

4. Ozbulak, G., Aytar, Y. and Ekenel, H. (2016). How Transferable Are CNN-Based Features for Age and Gender Classification?. 2016 International Conference of the Biometrics Special Interest Group (BIOSIG).

5. Ng, C., Tay, Y. and Goi, B. (2013). A Convolutional Neural Network for Pedestrian Gender Recognition. Advances in Neural Networks – ISNN 2013, pp.558-564.

6. Raza, M., Chen Zonghai, Rehman, S., Ge Zhenhua, Wang Jikai and Bao Peng (2017). Part-Wise Pedestrian Gender Recognition Via Deep Convolutional Neural Networks. 2nd IET International Conference on Biomedical Image and Signal Processing (ICBISP 2017).

7. Cao, L., Dikmen, M., Fu, Y. and Huang, T. (2008). Gender recognition from body. Proceeding of the 16th ACM international conference on Multimedia - MM '08.

8. DENG, Y., Luo, P., Loy, C. and Tang, X. (2014). Pedestrian Attribute Recognition At Far Distance. Proceedings of the ACM International Conference on Multimedia - MM '14.

9. Ranjan, R., Sankaranarayanan, S., Castillo, C. and Chellappa, R. (2017). An All-In-One Convolutional Neural Network for Face Analysis. 2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017).

10.. Dehghan, A., Ortiz, E., Shu, G. and Masood, S. (2019). *DAGER: Deep Age, Gender and Emotion Recognition Using Convolutional Neural Network*. [online] arXiv.org. Available at: https://arxiv.org/abs/1702.04280 [Accessed 2 Jun. 2019].

11. Kurnianggoro, L. and Jo, K. (2017). Identification of pedestrian attributes using deep network. *IECON 2017 - 43rd Annual Conference of the IEEE Industrial Electronics Society*.

12. Raza, M., Sharif, M., Yasmin, M., Khan, M., Saba, T. and Fernandes, S. (2018). Appearance based pedestrians' gender recognition by employing stacked auto encoders in deep learning. *Future Generation Computer Systems*, 88, pp.28-39.

13. Antipov, G., Baccouche, M., Berrani, S. and Dugelay, J. (2017). Effective training of convolutional neural networks for face-based gender and age prediction. *Pattern Recognition*, 72, pp.15-26.

14. Xing, J., Li, K., Hu, W., Yuan, C. and Ling, H. (2017). Diagnosing deep learning models for high accuracy age estimation from a single image.

15. Simonyan, K. and Zisserman, A. (2020). *Very Deep Convolutional Networks for Large-Scale Image Recognition*. [online] arXiv.org. Available at: https://arxiv.org/abs/1409.1556 [Accessed 7 Jul. 2019].

16. Zhu, J., Liao, S., Lei, Z. and Li, S. (2017). Multi-label convolutional neural network based pedestrian attribute classification. *Image and Vision Computing*, 58, pp.224-229.

17. Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, 61, pp.85-117.

18. Wolfshaar, J., Karaaba, M. and Wiering, M. (2015). Deep Convolutional Neural Networks and Support Vector Machines for Gender Recognition. *2015 IEEE Symposium Series on Computational Intelligence*.

19. Anon, (2020). [online] Available at: https://www.researchgate.net/publication/338171619_Pedestrian_Age_and_Gender_Identification_from_Far_View_Images_Using_Convolutional_Neural_Network [Accessed 1 Jan. 2020].

20. Zhang, K., Gao, C., Guo, L., Sun, M., Yuan, X., Han, T., Zhao, Z. and Li, B. (2017). Age Group and Gender Estimation in the Wild With Deep RoR Architecture. *IEEE Access*, 5, pp.22492-22503.

21. Qawaqneh, Z., Mallouh, A. and Barkana, B. (2020). *Deep Convolutional Neural Network for Age Estimation based on VGG-Face Model*. [online] arXiv.org. Available at: https://arxiv.org/abs/1709.01664 [Accessed 18 Dec. 2019].