(Established under Galgotias University Uttar Pradesh Act No. 14 of 2011)

AGE AND GENDER DETECTION

Project Report Of Capstone Project- 2

Submitted by

SHUBHAM KUMAR
TIWARI(1613112045)

in partial fulfilment for the award of the degree
of

BACHELOR OF TECHNOLOGY
IN

COMPUTER SCIENCE AND ENGINEERING WITH SPECIALIZATION OF

DATA ANALYTICS

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING

Under the Supervision
of

Dr. SANJAY KUMAR, M.Tech., Ph.D., Professor

APRIL / MAY-2020

SCHOOL OF COMPUTING AND SCIENCE AND ENGINEERING

BONAFIDE CERTIFICATE

Certified that this project report "AGE AND GENDER DETECTION" is the bonifide work of "SHUBHAM KUMAR TIWARI(1613112045)" who carried out the project work under my supervision.

SIGNATURE OF HEAD	SIGNATURE OF SUPERVISOR
Dr. MUNISH SHABARWAL,	Dr SANJAY KUMAR
PhD (Management),	M tech(CS), PhD(CS)
PhD (CS) Professor & Dean,	Professor
School of Computing Science & Engineering	School of Computer Science & Engineering

TABLE OF CONTENTS

Objective

To build a gender and age detector that can approximately guess the gender and age of the person (face) in a picture using **Deep Learning** on the Audience dataset.

Abstract

Nowadays research has explored to extracting auxiliary information from various biometric techniques such as fingerprints, face, iris, palm, voice etc. This information contains some features like gender, age, beard, moustache, scars, height, hair, skin color, glasses, weight, facial marks, tattoos etc. All this information contributes more and more during identification. The major changes that come across face recognition is to find age & gender of the person. This paper contributes a significant survey of various face recognition techniques for finding the age and gender. The existing techniques are discussed based on their performances. This paper also provides future directions for further research. Age and gender, two of the key facial attributes, play a very foundational role in social interactions, making age and gender estimation from a single face image an important task in intelligent applications, such as access control, human-computer interaction, law enforcement, marketing intelligence and visual surveillance.

This project is based upon computer vision the various terminologies used to process image and detect age and gender of the person from the image.

A Convulational Neural work is a deep neural network (DNN) widely used for the purposes of image recognition and processing and NLP. Also known as a ConvNet, a CNN has input and output layers, and multiple hidden layers, many of which are convolutional. In a way, CNNs are regularized multilayer perceptron.

Computer Vision is the field of study that enables computers to see and identify digital images and videos as a human would. The challenges it faces largely follow from the limited understanding biological vision. A fast and efficient gender and age estimation system based on facial images is developed. There are many methods have

been proposed in the literature for the age estimation and gender classification. However, all of them have still disadvantage such as not complete reflection about face structure, face texture. This technique applies to both face alignment and recognition and significantly improves three aspects Within a given database, all weight vectors of the persons within the same age group are averaged together. A range of an age estimation result is 15 to 70 years old, and divided into 13 classes with 5 years old range. Experimental results show that better gender classification and age estimation. age and gender classification has become applicable to an extending measure of applications, particularly resulting to the ascent of social platforms and social media. Regardless, execution of existing strategies on real-world images is still fundamentallymissing, especially when considered the immense bounced in execution starting late reported for the related task of face acknowledgment. In this paper we exhibit that by learning representations through the use of significantConvolutional Neural Systems (CNN), a huge augmentation in execution can be acquired on these errands. To this end, we propose a direct Convolutional Neural System engineering can be used despite when the measure of learning data is limited. We survey our procedure on the recent Adience benchmark for age and gender estimation and demonstrate it to radically outflank current state-of-the-art methods.

Introduction

Age and gender, two of the key facial attributes, play a very foundational role in social interactions, making age and gender estimation from a single face image an important task in intelligent applications, such as access control, human-computer interaction, law enforcement, marketing intelligence and visual surveillance, etc. The enhancing of raw images that are received from the camera sources, from satellites, aircrafts and the pictures captured in day-to-day lives is called image processing. The images have been processed through many different techniques and calculations have been made on the basis and analysis of the studies. There is a need of analyzing and studying the digitally formed images. There are two main and very common steps followed for image processing. The improvement of an image such that the resulted image is of greater quality and can be used by other programs, is called image enhancement . The other technique is the most sought after technique used for extraction of information from an image. There is a division of the image into certain number of parts or objects so that the problem is solved. This process is called segmentation. A neural network consists of many simple and similar compressing elements. It is a system with inputs and outputs. There are a number of internal parameters called weights. An artificial neural network is made of set of processing elements which are also known as neurons or nodes. These nodes are interconnected. Training in ANN is done through the track of the examples. There are various such methods that fail to produce appropriate results. For each class, an essential rule called the characteristic rule is generated. This set of rules is also called as differentiating rules. A systematic method which is used to train multilayer artificial neural networks is known as back propagation. It is also considered as a gradient method where the gradient of the error is evaluated by considering the weights of the given inputs. The detection of the data available in the images is very important. The data that the image contains is to be changed and modified for the detection purposes. There are various types of techniques involved for detection as well as the removal of the problem. In a Facial detection technique: The expressions that the faces contain hold a lot of information. Whenever a person interacts with the other person, there is an involvement of a lot of expressions . The changing of expressions helps in calculating certain parameters. Age estimation is a multi-class problem in which the years are classified into classes. People with different ages have different facials, so it is difficult to gather the images. Various age detection methods are used. The preprocessing is applied to the image. Features are the extracted from the neural network through the convolution network. Based on the trained models the image is then classified to one of the age classes. Features are extracted from the images for further processing. The features are processed further and sent to the training systems. The databases provide a study to the features and help in completing the face detection for proving the age detection of the person in the image. Age and gender assume essential parts in social between activities. Dialects hold distinctive greetings and grammar rules for men or women, and frequently diverse vocabularies are utilized while tending to senior citizens compared to youngsters. In spite of the essential parts these characteristics play in our everyday lives, the capacity to consequently assess them precisely and dependably from face image is still a long way from addressing the requirements of business applications. This is especially puzzling while considering late claims to super-human capacities in the related errand of face recognition. .Past ways to deal with assessing or ordering these properties from face images have depended on contrasts in facial feature dimensions or "customized" face descriptors. Most have utilized characterization plans composed especially for age or gender orientation estimation undertakings, including and others. Few of these past strategies wereintended to handle the numerous difficulties of unconstrained imaging conditions . In addition, the machine learning strategies utilized by these frameworks did not completely abuse the huge quantities of image cases and information accessible through the Internet keeping in mind the end goal to enhance characterization capacities.In this paper we endeavour to close the gap between automatic face recognition abilities and those of age and gender classification techniques. To this end, we take after the fruitful sample set around late face recognition frameworks: Face recognition systems portrayed in the most recent couple of years have demonstrated that gigantic advancement can be made by the utilization of profound convolutional neural networks (CNN) . We show comparative additions with basic system engineering, composed by considering the somewhat constrained accessibility of precise age and gender classification names inexisting face information sets.

Technologies Used in the Estimation Process And Related Work

Image Proccessing

Vision processing incorporates human perception and intelligence which makes the field most interesting to the research community as it can mimic human behaviour in the computer system by means of video surveillance system, integrating more intelligence to machines such as robots, as well as in ecology, biometrics and medical applications. Interestingly, recent NASA's mission "Curiosity" on Mars, sending valuable images and information of Mars environment in a secure communication channel, transmitted images also need to processed exhaustively to find out any vital information about Mars. Hardware designs for image and video processing is used for faster performance rather than software, to meet the requirements of the end users, keeping its market relevancy and at the same timesecurity is another concern, so the necessity to communicate these media data securely among multiple platforms after processing to enhance human perception and satisfaction in which our focus lies. The basic 4 steps in image processing domain are pre-processing, segmentation, feature extraction and recognition and those has been keeping their strong importance in research mostly in the case of software implementation and very few implemented on hardware. Initial pre-processing step is carried out to enhance the quality of the original image by removing noise, unbalanced brightness etc as common interfering elements followed by segmentation where images are separated from the background into various elements with properties. Next in the feature extraction stage, extraction is performed on every detected object to reduce its information to a list of parameters storing in memory. Finally in the recognition stage a set of signals are generated using this list which constitute the upper level of processing assigning a specific meaning to every detected object. In this paper we focused on image thresholding which is mainly used in the pre-processing and segmentation stages respectively, where our implementation is performing well enough in comparison to existing work (compared below), followed by secured transmission of the image data between multiple FPGA platforms and to the best of our knowledge this design belongs to a class of advanced implementation.

CNN For age and gender estimation

Gathering a substantial, marked image preparing set for age and gender estimation from social network image archives requires either access to individual data on the subjects showing up in the images, which is regularly private, or is tediousto physically name . Information sets for age and gender estimation from true social network images are in this way moderately constrained in size and in a matter of seconds no match in size with the much larger image arrangement information sets (e.g. the Image net dataset). Over fitting is normal issue, when machine learning construct strategies are utilized as a part of image accumulations. This issue is exacerbated while considering profound convolutional neural network systems. because of their enormous quantities of model parameters. Care should in this way be taken with a specific end goal to stay away from over fitting under such circumstances. A.Network ArchitectureOur proposed system design is utilized all through our tests for both age and gender classification order. It is delineated that The system contains just three convolutional layers and two completely associated layers with little number of neurons. This, by correlation with the much bigger models connected, for instance. Our decision of a system outline is spurred both from our longing to lessen the danger of over fitting and in addition the way of the issues we are endeavoring to unravel: age grouping on the Adience set requires recognizing eight classes; gender classification needs just two classes. This contrasted with, e.g., the ten thousand personality classes used to prepare the system utilized for face acknowledgment as a part Each of the three shading channels is handled specifically by the system. Images are initially rescaled to 256 × 256 and a product of 227×227 is bolstered to the system. The three ensuing convolutional layers are then characterized as takes after.

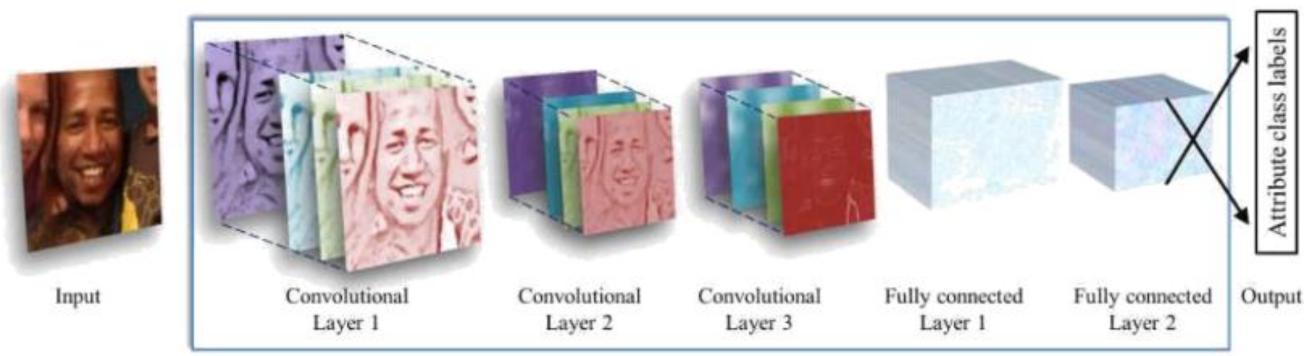


Fig: Illustration of CNN Architecture

B.Testing and Training Initialization:

1. 96 channels of size 3×7×7 pixels are connected to the information in the primary convolutional layer, trailed by an amended straight administrator (ReLU), a maximum pooling layer taking the maximal estimation of 3 × 3 areas with two-pixel strides and a nearby reaction standardization layer

2. The 96 × 28 × 28 yield of the past layer is then handled by the second convolutional layer, containing 256 channels of size 96 × 5 × 5 pixels. Once more, this is trailed by ReLU, a maximum pooling layer and a local reaction standardization layer with the same hyper parameters as some time recently.

3. Finally, the third and keep going convolutional layer works on the 256 × 14 × 14 blob by applying an arrangement of 384 channels of size 256 × 3 × 3 pixels, trailed by ReLU and a maximum pooling layer.

4. A first completely associated layer that gets the yield of the third convolutionallayer and contains 512 neurons, trailed by a ReLU and a dropout layer.

5. A second completely associated layer that gets the 512-dimensional yield of the main completely associated layer and again contains 512 neurons, trailed by a ReLU and a dropout layer.

6. A third, completely associated layer which maps to the last classes for age or gender classification. At long last, the yieldof the last completely associated layer is encouraged to a delicate max layer that doles out likelihood for every class. The forecast itself is made by bringing the class with the maximal likelihood for the given test image. The weights in all layers are instated with irregular qualities from a zero mean Gaussian with standard deviation of 0.01. Tostretch this, we don't utilize pre-prepared models for instating the system; the system is prepared, starting with no outside help, without utilizing any information outside of the images and the makes accessible by the benchmark. This, once more, ought to be contrasted and CNN executions utilized for face acknowledgment, where countless images are utilized for preparing.

Network Training:

Beside our utilization of incline system design, we apply two extra strategies as far as possible the danger of over fitting.To start with we apply dropout learning (i.e. randomly setting the output value of network neurons to zero). The system incorporates two dropout layers with a dropout proportion of 0.5 (half risk of setting a neuron's yield worth to zero). Second, we utilize information growth by taking an arbitrary product of 227 × 227 pixels from the 256 × 256 image data and arbitrarily reflect it in each forward-backward training pass. This, likewise to the different yield and reflect varieties utilized.Prediction:We tried different things with two techniques for utilizing the system as a part of request to create age and gender predictions for novel countenances:

• Center Crop: Feeding the system with the face image, edited to 227 × 227 around the face focus.

•Over-Sampling: We separate five 227 × 227 pixel crop districts, four from the sides of the 256 × 256 face image, and an extra yield area from the focal point of the face. The system is given every one of the five images, alongside their flat reflections. Its lastforecast is taken to be the normal expectation esteem over every one of these varieties. We have found that little misalignments in the Adience images, brought on by the numerous difficulties of these images (impediments, movement obscure, and so forth.) can noticeably affect the nature of our outcomes. This second, over-testing strategy is intended to adjust for these misalignments, bypassing the requirement for enhancing arrangement quality, yet rather specifically bolstering the system with different interpreted adaptations of the same face.

Artificial Nueral Networks

Artificial neural networks (ANN) or connectionist systems are computing systems vaguely inspired by the biological nueral networksthat constitute animal brains. Such systems "learn" to perform tasks by considering examples, generally without being programmed with task-specific rules. For example, in image recognition, they might learn to identify images that contain cats by analyzing example images that have been manually labelled as "cat" or "no cat" and using the results to identify cats in other images. They do this without any prior knowledge of cats, for example, that they have fur, tails, whiskers and cat-like faces. Instead, they automatically generate identifying characteristics from the examples that they process. An ANN is based on a collection of connected units or nodes calledartificial nuerons, which loosely model the neurons in a biological brain. Each connection, like the synapses in a biological brain, can transmit a signal to other neurons. An artificial neuron that receives a signal then processes it and can signal neurons connected to it.

In ANN implementations, the "signal" at a connection is areal number , and the output of each neuron is computed by some non-linear function of the sum of its inputs. The connections are called *edges*. Neurons and edges typically have a that adjusts as learning proceeds. The weight increases or decreases the strength of the signal at a connection. Neurons may have a threshold such that a signal is sent only if the aggregate signal crosses that threshold. Typically, neurons are aggregated into layers. Different layers may perform different transformations on their inputs. Signals travel from the first layer (the input layer), to the last layer (the output layer), possibly after traversing the layers multiple times.

The original goal of the ANN approach was to solve problems in the same way that a human brain would. But over time, attention moved to performing specific tasks, leading to deviations from biology._ANNs have been used on a variety of tasks, including computer vision, speech recognition, machine translation, social network filtering, playing board and video games, medical diagnosis, and even in activities that have traditionally been considered as reserved to humans, like painting.

Image Extraction

There are many methods have been proposed in the literature for the age estimation and gender classification. However, all of them have still disadvantage such as not complete reflection about face structure, face texture. We classified the gender and age based on the association of two methods: geometric feature based method and Principal Component Analysis (PCA) method for improving the efficiency of facial feature extraction stage. The face database contains the 13 individual groups. Within a given database, all weight vectors of the persons within the same age group are averaged together. Experimental results show that better gender classification and age estimation. Gender classification is important visual tasks for human beings, such as many social interactions critically depend on the correct gender perception. As visual surveillance and human-computer interaction technologies evolve, computer vision systems for gender classification will play an increasing important role in our lives. Age prediction is concerned with the use of a training set to train a model that can estimate the age of the facial images. Amount once paid is not refundable or adjustable under any circumstances in future. This project contains full non editable files and database images that we have used.

The Principal Component Analysis (PCA) can do prediction, redundancy removal, feature extraction, data compression, etc. Because PCA is a classical technique which can do something in the linear domain, applications having linear models are suitable. Let us consider the PCA procedure in a training set of M face images. Let a face image be represented as a two dimensional N by N array of intensity values, or a vector of dimension N2. Then PCA tends to find a M-dimensional subspace whose basis vectors correspond to the maximum variance direction in the original image space . New basis vectors define a subspace of face images called face space. All images of known faces are projected onto the face space to find sets of weights that describe the contribution of each vector. By comparing a set of weights for the unknown face to sets of weights of known faces, the face can be identified. PCA basis vectors are defined as eigenvectors of the scatter matrix S defined as:

$$S=\sum_{i=1}^{M}(xi-\mu)(xi-\mu)'$$

Where μ is the mean of all images in the training set and xi is the ith face image represented as a vector i. The eigenvector associated with the largest eigenvalue is one that reflects the greatest variance in the image. That is, the smallest eigenvalue is associated with the eigenvector that finds the least variance.

Nearest Neighbour Classification

One of the most popular non-parametric techniques is the Nearest Neighbor classification (NNC). NNC asymptotic or infinite sample size error is less than twice of the Bayes error . NNC gives a trade-off between the distributions of the training data with a priori probability of the classes involved. KNN (Kth nearest neighbor classifier)classifier is easy to compute and very efficient. KNN is very compatible and obtain less memory storage. So it has good discriminative power. Also, KNN is very robust to image distortions (e.g. rotation, illumination). So this paper can produce good result by combining (PCA and KNN). Euclidian distance determines whether the input face is near a known face. The problem of automatic face recognition is a composite task that involves detection and location of faces in a cluttered background, normalization, recognition and verification.

Feature Extraction

The gender classification procedure is described in this section. Features extraction-deals with extracting features that are basic for differentiating one class of object from another. First, the fast and accurate facial features extraction algorithm is developed. The training positions of the specific face region are applied. The extracted features of each face in database can be expressed in column matrix show in figure

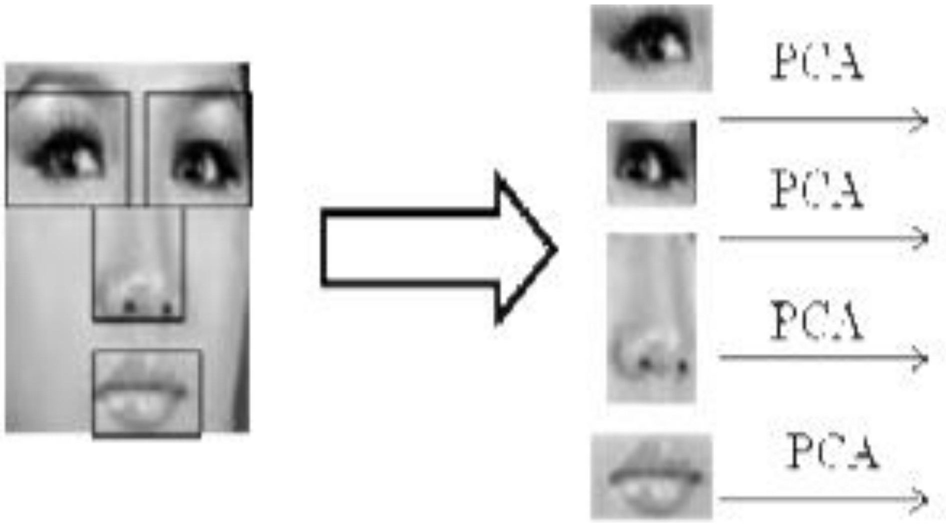


Fig : Facial Images

$$\left\{ \begin{bmatrix} \cdot \\ F_1 \\ \cdot \\ \cdot \end{bmatrix} \begin{bmatrix} \cdot \\ F_2 \\ \cdot \\ \cdot \end{bmatrix} \cdots \begin{bmatrix} \cdot \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} \right\}_{N^2 \times M}$$

Fig: Feature Extraction

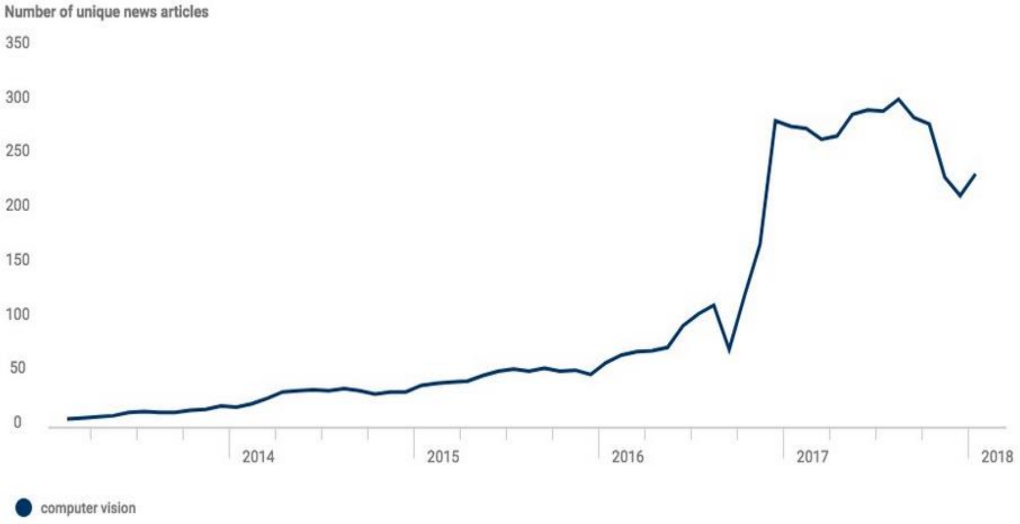
And find the average face for same age group of face images. The mean face feature for the M face images of each age group can be described as

$$A = \left\{ \begin{bmatrix} \cdot \\ \Psi\text{-}F \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} \begin{bmatrix} \cdot \\ \Psi\text{-}F_2 \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} \cdots \begin{bmatrix} \cdot \\ \Psi\text{-}F_M \\ \cdot \\ \cdot \\ \cdot \end{bmatrix} \right\}_{N^2 \times M}$$

The face space is computed from the Euclidean distance of feature points of two faces. The fundamental matrix A is constructed by the difference face space among the input and each face. Then, the matrix Ω can be formed by the average face features of the thirteen age groups. Calculate the Covariance Matrix Cov = ΩΩT. And then built Matrix L= ΩΩT to reduce dimension. Find the eigenvector of Cov. Eigenvector represent the variation in faces. Finally, age is determined through the minimize face space.

Computer Vision

Computer vision is an interdisciplinary field that deals with how computers can be made to gain high-level understanding from digital images or videos. From the perspective of engineering, it seeks to automate tasks that the human visual system can do. "Computer vision is concerned with the automatic extraction, analysis and understanding of useful information from a single image or a sequence of images. It involves the development of a theoretical and algorithmic basis to achieve automatic visual understanding." As a scientific discipline, computer vision is concerned with the theory behind artificial systems that extract information from images. The image data can take many forms, such as video sequences, views from multiple cameras, or multi-dimensional data from a medical scanner. As a technological discipline, computer vision seeks to apply its theories and models for the construction of computer vision systems.



Network Architecture

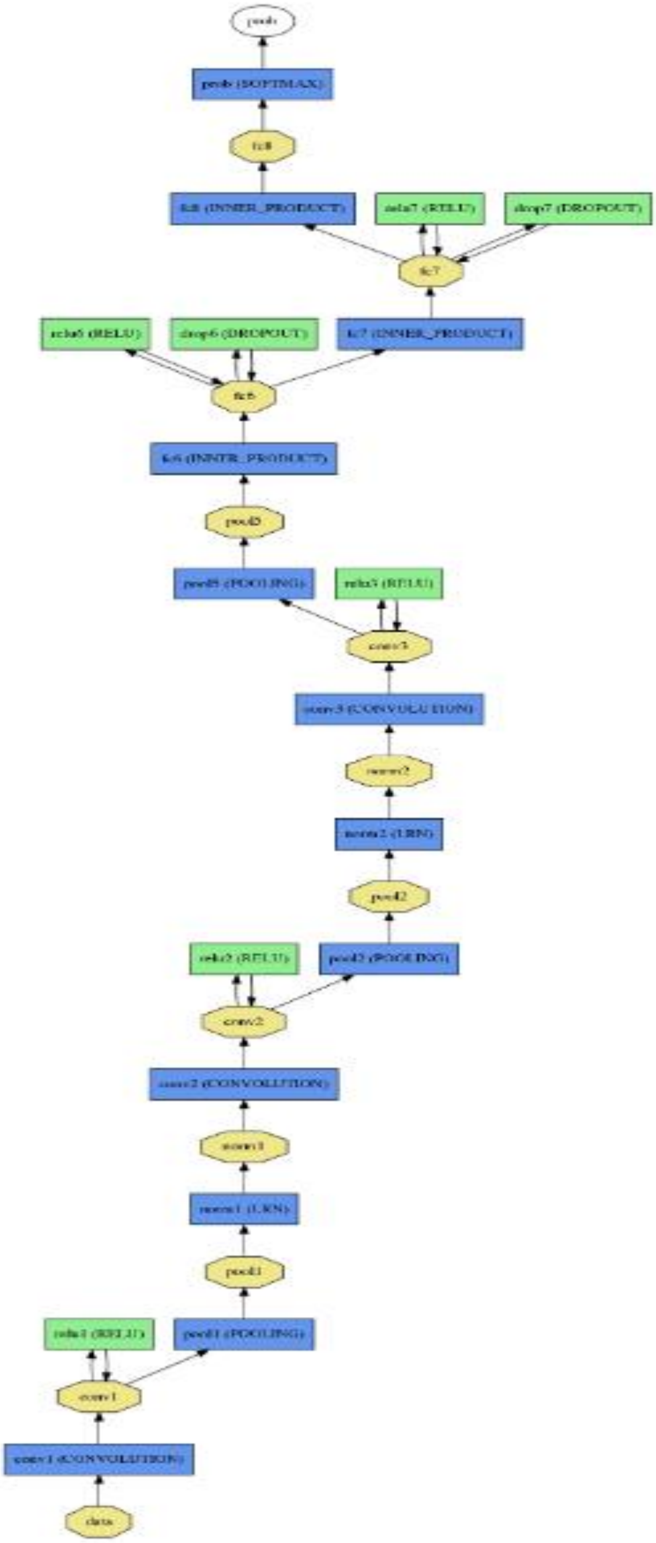
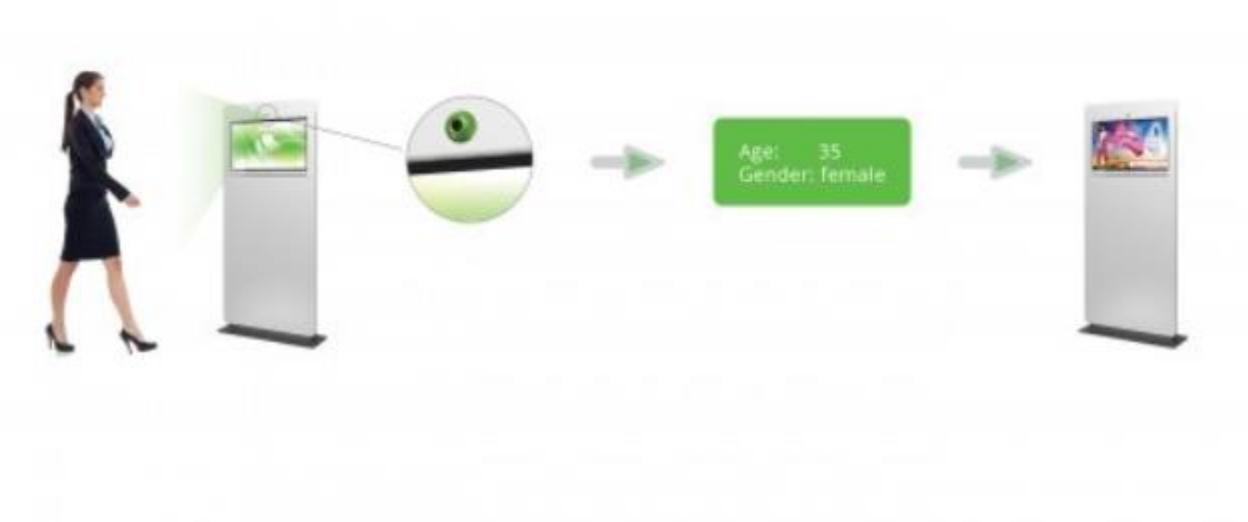


Fig : Full Schematic Diagram of our Network Architecture

Overall Description



Real world Use-Case

Recently I came across Quividi which is an AI software application which is used to detect age and gender of users who passes by based on online face analyses and automatically starts playing advertisements based on the targeted audience. Another example could be **AgeBot** which is an Android App that determines your age from your photos using facial recognition. It can guess your age and gender along with that can also find multiple faces in a picture and estimate the age for each face.

Inspired by the use cases we are going to build a simple Age and Gender detection model in this detailed article. So let's start with our use-case:

Use-case — we will be doing some face recognition, face detection stuff and furthermore, we will be using CNN (Convolutional Neural Networks) for age and gender predictions from a youtube video, you don’t need to download the video just the video URL is fine. The interesting part will be the usage of CNN for age and gender predictions on video URLs.

Requirements :

Pip install OpenCV

Python

numpy

pip install pafy

pip install youtube_dl (to know more about youtube_dl)

pafy : Pafy library is used to retrieve YouTube content and metadata(such as Title, rating, viewcount, duration, rating, author, thumbnail, keywords etc).

Purpose

The detection is the technique in which various factors are recognized on the basis of input and according to requirements. The age and gender detection is the issue which take consideration of researchers from last fewyears. In the topic on age and gender detection various techniques has been proposed to analysis features of the input image and on the basis of image features gender and approximation of age is defined. In this work, novel technique is proposed which is based on CNN for age and gender detection. This technique will scan the input image and detect key features. The simulation is performed in CNN and it is been analyzed that proposed technique performs well in terms of fault detection rate and images that are received from the camera sources, from satellites, aircrafts and the pictures captured in day-to-day lives is called image processing. The images have been processed through many different techniques and calculations have been made on the basis and analysis of the studies. There is a need of analyzing and studying the digitally formed images. There are two main and very common steps followed for image processing which is based upon

CNN that is a deep nueral network(DNN). The improvement of an image such that the resulted image is of greater quality and can be used by other programs, is called image enhancement. The other technique is the most sought after technique used for extraction of information from an image. There is a division of the image into certain number of parts or objects so that the problem is solved. This process is called segmentation. A neural network consists of many simple and similar compressing elements. It is a system with inputs and outputs. There are a number of internal parameters called weights in Artificial Neural networks.

Motivation and Scope

Automatic age and gender classification has become relevant to an increasing amount of applications, particularly since the rise of social platforms and social media. Nevertheless, performance of existing methods on real-world images is still significantly lacking, especially when compared to the tremendous leaps in performance recently reported for the related task of face recognition.

A Convulational Neural work is a deep neural network (DNN) widely used for the purposes of image recognition and processing and NLP. Also known as a ConvNet, a CNN has input and output layers, and multiple hidden layers, many of which are convolutional. In a way, CNNs are regularized multilayer perceptron.

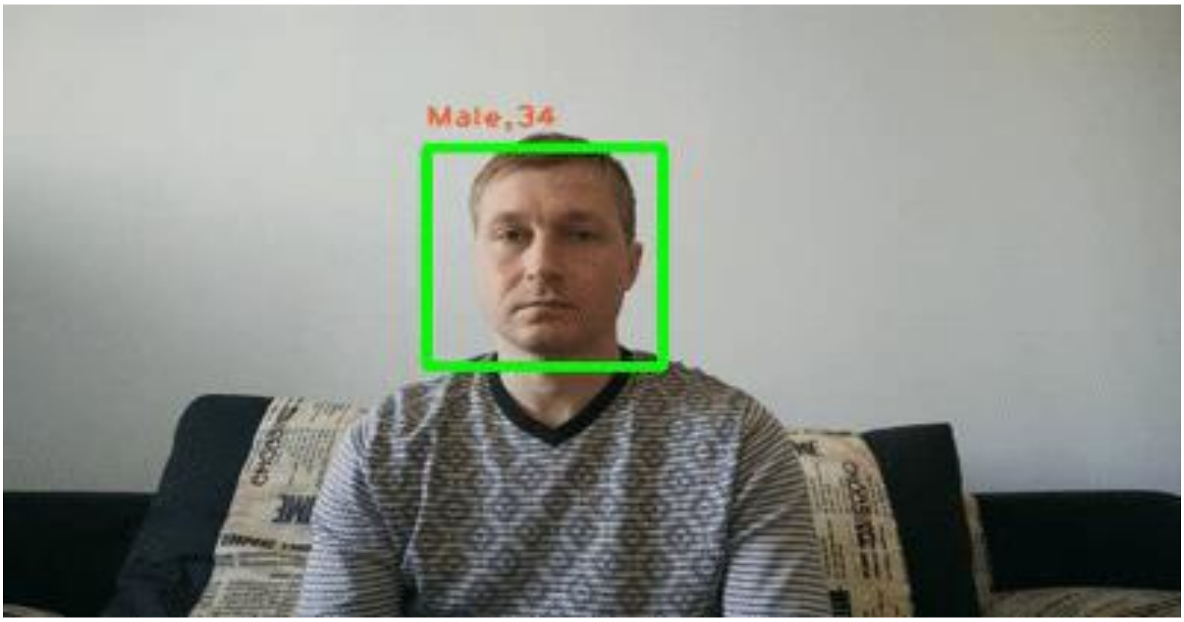
It is very difficult to accurately guess an exact age from a single image because of factors like makeup, lighting, obstructions, and facial expressions. And so, we make this a classification problem instead of making it one of regression.

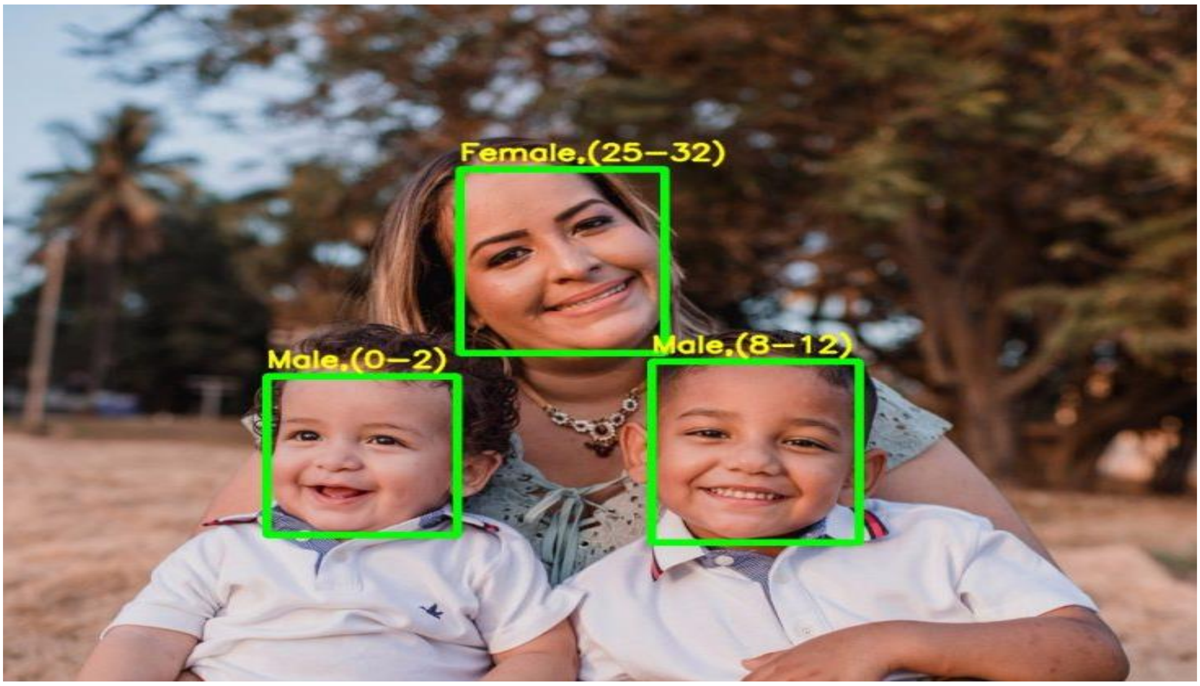
The major applications are

- Detect faces
- Classify into Male/Female
- Classify into one of the 8 age ranges
- Put the results on the image and display it

This will help us in many fields ranging from employee identification to human identification, defense security and CCTV footage identification. It can be used to identify people in somewhat blurred images.

As for example:





Literature Survey

Yunjo Lee, et.al proposed that the fMRI method is used to study upon age detection methods. The study involves a proper recording of the variations of people on the basis of their changes according to age, gender, identity and other features. The brain activation tasks related to face matching are performed and tested outside the scanner. There was a same result in face processing in older as well as young adults. The performance results high in both the cases having same facial viewpoints. The aging of the elders is not based on any one factor. It is combination of various factors that result in accountancy of such results. The results need to be kept a track on which are based on all credentials kept in certain environments.

R. Begg et.al explained the automatic recognition of walking changes because of aging through the artificial neural networks is the aim of the article. The balance control of the locomotors system is disturbed due to the gait factors which are caused through walking patterns which change according to the age. There are many advantages of such techniques. The standard back propagation, scaled conjugate gradient and the back propagation with Bayesian regularization were the three methods involved. The three networks came out with better results but the Bayesian regularization method was the one with greatest results in some fields. The neural networks thus are a great help for the age identification purposes.

Hang Qi et.al, proposed that various techniques have been arising for the detection of faces which can also identify the age of the person. Here, an automated system has been proposed which can classify the age and help distinguishing kids face from that of an adults face. There are three parts that the system encompasses. They are face detection, face alignment and normalization, and age classification. Face samples are created by the normal face detection and alignment methods. ICA is used for the extraction of the local facial components that are present in the images. This system has been proved to be much faster and the results are efficient. So this system can be used in future as a prototype.

Kensuke Mitsukura, et.al that on the basis of the color information the threshold value in multi-value images is considered. There is a lack of versatility when there is no change in the threshold of an image. Whenever there is an influence of any light conditions, the information of the color varies.It becomes prominent to decide the face. It is difficult to determine the face division standard. This is done for providing information to the Genetic Algorithm used in the method. Also a face decision method is proposed further which determines whether it is a decision method face or not. The identification of an individual is also very important. There is a use of the color maps for the differentiation of the detected faces. The features that are missed result in false identifications as well as the poor results.

Chao Yin et.al, the Conditional Probability Neural Network (CPNN) is a distribution learning algorithm used for the age estimation using facial expressions. It follows the three-layer neural network system in which the target values and the conditional feature vectors are used as an input. This can help it in learning the real ages. The relationship between the face image and the related label distribution through the neural network is used as the learning method for this system. The earlier method used proposed that the relationship is to be used according to the maximum entropy model. CPNN has proved to be providing better results than all the previously made methods. Through this method the results provided were very easy, there was less computational involved and the outcomes very efficient. Due to all such advantages it was preferred more than the others.

Sarah N. Kohail et.al proposed that the age estimation is now the current challenge being faced. Here, the article puts forward the approach of neural networks to estimate the age of humans. The main change that has been made in this method is the fine tuning of the age ranges. To learn the multi-layer perception neural networks (MLP) the facial features of the new images were extracted and recorded. The inputs were provided to the layer . The results have shown the MLP method as a good method with minimum errors in the results. These results can be used in many of the applications like age-based access control applications and also in the age adaptive human machine interaction.

Recently various learning machines for pattern classification have been proposed.

For instance, Jiang et al. developed a perturbation-resampling procedure to obtain the confidence interval estimates centred at k-fold cross-validated point for the prediction error and apply them to model evaluation and feature selection.

Liu investigated the effects of confidence transformation in combining multiple classifiers using various combination rules, where classifier outputs are transformed to confidence measures,

Feng et al. proposed a scaled SVM, which is to employ not only the support vectors but also the means of the classes to reduce the mean of the generalization error.

Graf et al. presented a method for combining human psychophysics and machine learning, in which human classification is introduced.

Gender classification is important visual tasks for human beings, such as many social interactions critically depend on the correct gender perception. As visual surveillance and human-computer interaction technologies evolve, computer vision systems for gender classification will play an increasing important role in our lives . Age prediction is concerned with the use of a training set to train a model that can estimate the age of the facial images. Among the first to research age prediction were, Kwon and Vitoria Lobo who proposed a method to classify input face images into one of the following three age groups: babies, young adults and senior adults . Their study was based on geometric ratios and skin wrinkle analysis. Their method was tested on a database of only 47 high resolution face images containing babies, young and middle aged adults. They reported 100% classification accuracy on these data. Hayashi focused their study on facial wrinkles for the estimation of age and gender. Gender classification is arguably one of the more important visual tasks for an extremely social animal like us humans many social interactions critically depend on the correct gender perception of the parties involved. Arguably, visual information from human faces provides one of the more important sources of information for gender classification. Not surprisingly, thus, that a very large number of psychophysical studies has investigated gender classification from face perception in humans . Face aging simulation and prediction is an interesting task with many applications in digital entertainment. A problem of personal verification and identification is an actively growing area of research. Face, voice, lip, movements, hand geometry, odor, gait, iris, retina, fingerprint are the most commonly used authentication methods.

Proposed model

Technologies used :

1. OpenCV : Provide the Haar-Cascades and OpenCV library
2. Computer Vision
3. Image processing
4. CNN(convulational Nueral Networks)

Steps to follow :

1. Get the video URL from YouTube or save pictures.
2. Face detection with Haar-cascades
3. Gender Recognition with CNN
4. Age Recognition with CNN

1.Get a video URL from YouTube:

Get the Youtube video URL and try to get the attributes of the video using pafy as explained above.

2. Face detection with Haar cascades :

This is a part most of us at least have heard of. OpenCV provide direct methods to import Haar-cascades and use them to detect faces. I will not be explaining this part in deep. You guys can refer to my previous article to know more about face detection using OpenCV

3. Gender Recognition with CNN:

Gender recognition using OpenCV's fisherfaces implementation is quite popular and some of you may have tried or read about it also. But, in this example, I will be using a different approach to recognize gender. This method was introduced by two Israel researchers, Gil Levi and Tal Hassner in 2015. I have used the CNN models trained by them in this example. We are going to use the OpenCV's dnn package which stands for "Deep Neural Networks".

In the dnn package, OpenCV has provided a class called Net which can be used to populate a neural network. Furthermore, these packages support importing neural network models from well known deep learning frameworks like caffe, tensorflow and torch. The researchers I had mentioned above have published their CNN models as caffe models. Therefore, we will be using the CaffeImporter import that model into our application.

4. Age Recognition with CNN

This is almost similar to the gender detection part except that the corresponding prototxt file and the caffe model file are "deploy_agenet.prototxt" and "age_net.caffemodel". Furthermore, the CNN's output layer (probability layer) in this CNN consists of 8 values for 8 age classes ("0–2", "4–6", "8–13", "15–20", "25–32", "38–43", "48–53" and "60-")

A caffe model has 2 associated files,

1 .prototxt—The definition of CNN goes in here. This file defines the layers in the neural network, each layer's inputs, outputs and functionality.

2 .caffemodel—This contains the information of the trained neural network (trained model)

The CNN Architecture

The convolutional neural network for this python project has 3 convolutional layers:

- Convolutional layer; 96 nodes, kernel size 7
- Convolutional layer; 256 nodes, kernel size 5
- Convolutional layer; 384 nodes, kernel size 3

Steps for practicing gender and age detection

1.The contents of this zip are:

- opencv_face_detector.pbtxt
- opencv_face_detector_uint8.pb
- age_deploy.prototxt
- age_net.caffemodel
- gender_deploy.prototxt
- gender_net.caffemodel
- a few pictures to try the project on

For face detection, we have a .pb file- this is a protobuf file (protocol buffer); it holds the graph definition and the trained weights of the model. We can use this to run the trained model. And while a .pb file holds the protobuf in binary format, one with the .pbtxt extension holds it in text format. These are TensorFlow files. For age and gender, the .prototxt files describe the network configuration and the .caffemodel file defines the internal states of the parameters of the layers.

2. We use the argparse library to create an argument parser so we can get the image argument from the command prompt. We make it parse the argument holding the path to the image to classify gender and age for.

3. For face, age, and gender, initialize protocol buffer and model.

4. Initialize the mean values for the model and the lists of age ranges and genders to classify from.

5. Now, use the readNet() method to load the networks. The first parameter holds trained weights and the second carries network configuration.

6. Let's capture video stream in case you'd like to classify on a webcam's stream. Set padding to 20.

7. Now until any key is pressed, we read the stream and store the content into the names hasFrame and frame. If it isn't a video, it must wait, and so we call up waitKey() from cv2, then break.

8. Let's make a call to the highlightFace() function with the faceNet and frame parameters, and what this returns, we will store in the names resultImg and faceBoxes. And if we got 0 faceBoxes, it means there was no face to detect..

Following Codes:

age_solver.prorotxt

```
net: "/home/ubuntu/AdienceFaces/age/train_val.prototxt"

test_iter: 1000

test_interval: 1000

base_lr: 0.001

lr_policy: "step"

gamma: 0.1

stepsize: 10000

display: 20

max_iter: 50000
```

```
momentum: 0.9
weight_decay: 0.0005
snapshot: 1000
snapshot_prefix: "caffenet_train"
solver_mode: GPU
```

age_train_val.prototxt

```
name: "CaffeNet"
layers {
name: "data"
type: DATA
top: "data"
top: "label"
data_param {
source: "/home/ubuntu/AdienceFaces/lmdb/age_train_lmdb"
backend: LMDB
batch_size: 50
}
transform_param {
crop_size: 227
mean_file: "/home/ubuntu/AdienceFaces/mean_image/mean.binaryproto"
mirror: true
}
include: { phase: TRAIN }
}
```

```
layers {
name: "data"
type: DATA
top: "data"
top: "label"
data_param {
source: "/home/ubuntu/AdienceFaces/lmdb/age_val_lmdb"
backend: LMDB
batch_size: 50
}
transform_param {
crop_size: 227
mean_file: "/home/ubuntu/AdienceFaces/mean_image/mean.binaryproto"
mirror: false
}
include: { phase: TEST }
}
layers {
name: "conv1"
type: CONVOLUTION
bottom: "data"
top: "conv1"
blobs_lr: 1
blobs_lr: 2
weight_decay: 1
```

```
weight_decay: 0
convolution_param {
num_output: 96
kernel_size: 7
stride: 4
weight_filler {
type: "gaussian"
std: 0.01
}
bias_filler {
type: "constant"
value: 0
}
}
}
layers {
name: "relu1"
type: RELU
bottom: "conv1"
top: "conv1"
}
layers {
name: "pool1"
type: POOLING
bottom: "conv1"
```

```
top: "pool1"
pooling_param {
pool: MAX
kernel_size: 3
stride: 2
}
}
layers {
name: "norm1"
type: LRN
bottom: "pool1"
top: "norm1"
lrn_param {
local_size: 5
alpha: 0.0001
beta: 0.75
}
}
layers {
name: "conv2"
type: CONVOLUTION
bottom: "norm1"
top: "conv2"
blobs_lr: 1
blobs_lr: 2
```

```
weight_decay: 1
weight_decay: 0
convolution_param {
num_output: 256
pad: 2
kernel_size: 5
weight_filler {
type: "gaussian"
std: 0.01
}
bias_filler {
type: "constant"
value: 1
}
}
}
layers {
name: "relu2"
type: RELU
bottom: "conv2"
top: "conv2"
}
layers {
name: "pool2"
type: POOLING
```

```
bottom: "conv2"
top: "pool2"
pooling_param {
pool: MAX
kernel_size: 3
stride: 2
}
}
layers {
name: "norm2"
type: LRN
bottom: "pool2"
top: "norm2"
lrn_param {
local_size: 5
alpha: 0.0001
beta: 0.75
}
}
layers {
name: "conv3"
type: CONVOLUTION
bottom: "norm2"
top: "conv3"
blobs_lr: 1
```

```
blobs_lr: 2

weight_decay: 1

weight_decay: 0

convolution_param {

num_output: 384

pad: 1

kernel_size: 3

weight_filler {

type: "gaussian"

std: 0.01

}

bias_filler {

type: "constant"

value: 0

}

}

}

layers{

name: "relu3"

type: RELU

bottom: "conv3"

top: "conv3"

}

layers {

name: "pool5"
```

```
type: POOLING
bottom: "conv3"
top: "pool5"
pooling_param {
pool: MAX
kernel_size: 3
stride: 2
}
}
layers {
name: "fc6"
type: INNER_PRODUCT
bottom: "pool5"
top: "fc6"
blobs_lr: 1
blobs_lr: 2
weight_decay: 1
weight_decay: 0
inner_product_param {
num_output: 512
weight_filler {
type: "gaussian"
std: 0.005
}
bias_filler {
```

```
type: "constant"
value: 1
}
}
}
layers {
name: "relu6"
type: RELU
bottom: "fc6"
top: "fc6"
}
layers {
name: "drop6"
type: DROPOUT
bottom: "fc6"
top: "fc6"
dropout_param {
dropout_ratio: 0.5
}
}
layers {
name: "fc7"
type: INNER_PRODUCT
bottom: "fc6"
top: "fc7"
```

```
blobs_lr: 1

blobs_lr: 2

weight_decay: 1

weight_decay: 0

inner_product_param {

num_output: 512

weight_filler {

type: "gaussian"

std: 0.005

}

bias_filler {

type: "constant"

value: 1

}

}

}

layers {

name: "relu7"

type: RELU

bottom: "fc7"

top: "fc7"

}

layers {

name: "drop7"

type: DROPOUT
```

```
bottom: "fc7"
top: "fc7"
dropout_param {
dropout_ratio: 0.5
}
}
layers {
name: "fc8"
type: INNER_PRODUCT
bottom: "fc7"
top: "fc8"
blobs_lr: 10
blobs_lr: 20
weight_decay: 1
weight_decay: 0
inner_product_param {
num_output: 8
weight_filler {
type: "gaussian"
std: 0.01
}
bias_filler {
type: "constant"
value: 0
}
```

```
}
}
layers {
name: "accuracy"
type: ACCURACY
bottom: "fc8"
bottom: "label"
top: "accuracy"
include: { phase: TEST }
}
layers {
name: "loss"
type: SOFTMAX_LOSS
bottom: "fc8"
bottom: "label"
top: "loss"
}
```

deploy_age.prototxt

```
name: "CaffeNet"
input: "data"
input_dim: 1
input_dim: 3
input_dim: 227
```

```
input_dim: 227
layers {
name: "conv1"
type: CONVOLUTION
bottom: "data"
top: "conv1"
convolution_param {
num_output: 96
kernel_size: 7
stride: 4
}
}
layers {
name: "relu1"
type: RELU
bottom: "conv1"
top: "conv1"
}
layers {
name: "pool1"
type: POOLING
bottom: "conv1"
top: "pool1"
pooling_param {
pool: MAX
```

```
kernel_size: 3

stride: 2

}

}

layers {

name: "norm1"

type: LRN

bottom: "pool1"

top: "norm1"

lrn_param {

local_size: 5

alpha: 0.0001

beta: 0.75

}

}

layers {

name: "conv2"

type: CONVOLUTION

bottom: "norm1"

top: "conv2"

convolution_param {

num_output: 256

pad: 2

kernel_size: 5

}
```

```
}
layers {
name: "relu2"
type: RELU
bottom: "conv2"
top: "conv2"
}
layers {
name: "pool2"
type: POOLING
bottom: "conv2"
top: "pool2"
pooling_param {
pool: MAX
kernel_size: 3
stride: 2
}
}
layers {
name: "norm2"
type: LRN
bottom: "pool2"
top: "norm2"
lrn_param {
local_size: 5
```

```
alpha: 0.0001

beta: 0.75

}

}

layers {

name: "conv3"

type: CONVOLUTION

bottom: "norm2"

top: "conv3"

convolution_param {

num_output: 384

pad: 1

kernel_size: 3

}

}

layers{

name: "relu3"

type: RELU

bottom: "conv3"

top: "conv3"

}

layers {

name: "pool5"

type: POOLING

bottom: "conv3"
```

```
top: "pool5"
pooling_param {
pool: MAX
kernel_size: 3
stride: 2
}
}
layers {
name: "fc6"
type: INNER_PRODUCT
bottom: "pool5"
top: "fc6"
inner_product_param {
num_output: 512
}
}
layers {
name: "relu6"
type: RELU
bottom: "fc6"
top: "fc6"
}
layers {
name: "drop6"
type: DROPOUT
```

```
bottom: "fc6"

top: "fc6"

dropout_param {

dropout_ratio: 0.5

}

}

layers {

name: "fc7"

type: INNER_PRODUCT

bottom: "fc6"

top: "fc7"

inner_product_param {

num_output: 512

}

}

layers {

name: "relu7"

type: RELU

bottom: "fc7"

top: "fc7"

}

layers {

name: "drop7"

type: DROPOUT

bottom: "fc7"
```

```
top: "fc7"
dropout_param {
dropout_ratio: 0.5
}
}
layers {
name: "fc8"
type: INNER_PRODUCT
bottom: "fc7"
top: "fc8"
inner_product_param {
num_output: 8
}
}
layers {
name: "prob"
type: SOFTMAX
bottom: "fc8"
top: "prob"
}
```

deploy_gender.prototxt

```
name: "CaffeNet"
input: "data"
```

```
input_dim: 1

input_dim: 3

input_dim: 227

input_dim: 227

layers {

name: "conv1"

type: CONVOLUTION

bottom: "data"

top: "conv1"

convolution_param {

num_output: 96

kernel_size: 7

stride: 4

}

}

layers {

name: "relu1"

type: RELU

bottom: "conv1"

top: "conv1"

}

layers {

name: "pool1"

type: POOLING

bottom: "conv1"
```

```
top: "pool1"
pooling_param {
pool: MAX
kernel_size: 3
stride: 2
}
}
layers {
name: "norm1"
type: LRN
bottom: "pool1"
top: "norm1"
lrn_param {
local_size: 5
alpha: 0.0001
beta: 0.75
}
}
layers {
name: "conv2"
type: CONVOLUTION
bottom: "norm1"
top: "conv2"
convolution_param {
num_output: 256
```

```
pad: 2
kernel_size: 5
}
}
layers {
name: "relu2"
type: RELU
bottom: "conv2"
top: "conv2"
}
layers {
name: "pool2"
type: POOLING
bottom: "conv2"
top: "pool2"
pooling_param {
pool: MAX
kernel_size: 3
stride: 2
}
}
layers {
name: "norm2"
type: LRN
bottom: "pool2"
```

```
top: "norm2"

lrn_param {

local_size: 5

alpha: 0.0001

beta: 0.75

}

}

layers {

name: "conv3"

type: CONVOLUTION

bottom: "norm2"

top: "conv3"

convolution_param {

num_output: 384

pad: 1

kernel_size: 3

}

}

layers{

name: "relu3"

type: RELU

bottom: "conv3"

top: "conv3"

}

layers {
```

```
name: "pool5"
type: POOLING
bottom: "conv3"
top: "pool5"
pooling_param {
pool: MAX
kernel_size: 3
stride: 2
}
}
layers {
name: "fc6"
type: INNER_PRODUCT
bottom: "pool5"
top: "fc6"
inner_product_param {
num_output: 512
}
}
layers {
name: "relu6"
type: RELU
bottom: "fc6"
top: "fc6"
}
```

```
layers {
name: "drop6"
type: DROPOUT
bottom: "fc6"
top: "fc6"
dropout_param {
dropout_ratio: 0.5
}
}
layers {
name: "fc7"
type: INNER_PRODUCT
bottom: "fc6"
top: "fc7"
inner_product_param {
num_output: 512
}
}
layers {
name: "relu7"
type: RELU
bottom: "fc7"
top: "fc7"
}
layers {
```

```
name: "drop7"
type: DROPOUT
bottom: "fc7"
top: "fc7"
dropout_param {
dropout_ratio: 0.5
}
}
layers {
name: "fc8"
type: INNER_PRODUCT
bottom: "fc7"
top: "fc8"
inner_product_param {
num_output: 2
}
}
layers {
name: "prob"
type: SOFTMAX
bottom: "fc8"
top: "prob"
}
```

gender_solver.prorotxt

```
net: "/home/ubuntu/AdienceFaces/gender/train_val.prototxt"
```

```
test_iter: 1000

test_interval: 1000

base_lr: 0.001

lr_policy: "step"

gamma: 0.1

stepsize: 10000

display: 20

max_iter: 50000

momentum: 0.9

weight_decay: 0.0005

snapshot: 1000

snapshot_prefix: "caffenet_train"

solver_mode: GPU
```

gender_train_val.prototxt

```
name: "CaffeNet"

layers {

name: "data"

type: DATA

top: "data"

top: "label"

data_param {

source: "/home/ubuntu/AdienceFaces/lmdb/gender_train_lmdb"

backend: LMDB

batch_size: 50

}
```

```
transform_param {
crop_size: 227
mean_file: "/home/ubuntu/AdienceFaces/mean_image/mean.binaryproto"
mirror: true
}
include: { phase: TRAIN }
}
layers {
name: "data"
type: DATA
top: "data"
top: "label"
data_param {
source: "/home/ubuntu/AdienceFaces/lmdb/gender_val_lmdb"
backend: LMDB
batch_size: 50
}
transform_param {
crop_size: 227
mean_file: "/home/ubuntu/AdienceFaces/mean_image/mean.binaryproto"
mirror: false
}
include: { phase: TEST }
}
layers {
```

```
name: "conv1"
type: CONVOLUTION
bottom: "data"
top: "conv1"
blobs_lr: 1
blobs_lr: 2
weight_decay: 1
weight_decay: 0
convolution_param {
num_output: 96
kernel_size: 7
stride: 4
weight_filler {
type: "gaussian"
std: 0.01
}
bias_filler {
type: "constant"
value: 0
}
}
}
layers {
name: "relu1"
type: RELU
```

```
bottom: "conv1"

top: "conv1"

}

layers {

name: "pool1"

type: POOLING

bottom: "conv1"

top: "pool1"

pooling_param {

pool: MAX

kernel_size: 3

stride: 2

}

}

layers {

name: "norm1"

type: LRN

bottom: "pool1"

top: "norm1"

lrn_param {

local_size: 5

alpha: 0.0001

beta: 0.75

}

}
```

```
layers {
name: "conv2"
type: CONVOLUTION
bottom: "norm1"
top: "conv2"
blobs_lr: 1
blobs_lr: 2
weight_decay: 1
weight_decay: 0
convolution_param {
num_output: 256
pad: 2
kernel_size: 5
weight_filler {
type: "gaussian"
std: 0.01
}
bias_filler {
type: "constant"
value: 1
}
}
}
layers {
name: "relu2"
```

```
type: RELU
bottom: "conv2"
top: "conv2"
}
layers {
name: "pool2"
type: POOLING
bottom: "conv2"
top: "pool2"
pooling_param {
pool: MAX
kernel_size: 3
stride: 2
}
}
layers {
name: "norm2"
type: LRN
bottom: "pool2"
top: "norm2"
lrn_param {
local_size: 5
alpha: 0.0001
beta: 0.75
}
```

```
}
layers {
name: "conv3"
type: CONVOLUTION
bottom: "norm2"
top: "conv3"
blobs_lr: 1
blobs_lr: 2
weight_decay: 1
weight_decay: 0
convolution_param {
num_output: 384
pad: 1
kernel_size: 3
weight_filler {
type: "gaussian"
std: 0.01
}
bias_filler {
type: "constant"
value: 0
}
}
}
layers{
```

```
name: "relu3"
type: RELU
bottom: "conv3"
top: "conv3"
}
layers {
name: "pool5"
type: POOLING
bottom: "conv3"
top: "pool5"
pooling_param {
pool: MAX
kernel_size: 3
stride: 2
}
}
layers {
name: "fc6"
type: INNER_PRODUCT
bottom: "pool5"
top: "fc6"
blobs_lr: 1
blobs_lr: 2
weight_decay: 1
weight_decay: 0
```

```
inner_product_param {
num_output: 512
weight_filler {
type: "gaussian"
std: 0.005
}
bias_filler {
type: "constant"
value: 1
}
}
}
layers {
name: "relu6"
type: RELU
bottom: "fc6"
top: "fc6"
}
layers {
name: "drop6"
type: DROPOUT
bottom: "fc6"
top: "fc6"
dropout_param {
dropout_ratio: 0.5
```

```
}
}
layers {
name: "fc7"
type: INNER_PRODUCT
bottom: "fc6"
top: "fc7"
blobs_lr: 1
blobs_lr: 2
weight_decay: 1
weight_decay: 0
inner_product_param {
num_output: 512
weight_filler {
type: "gaussian"
std: 0.005
}
bias_filler {
type: "constant"
value: 1
}
}
}
layers {
name: "relu7"
```

```
type: RELU
bottom: "fc7"
top: "fc7"
}
layers {
name: "drop7"
type: DROPOUT
bottom: "fc7"
top: "fc7"
dropout_param {
dropout_ratio: 0.5
}
}
layers {
name: "fc8"
type: INNER_PRODUCT
bottom: "fc7"
top: "fc8"
blobs_lr: 10
blobs_lr: 20
weight_decay: 1
weight_decay: 0
inner_product_param {
num_output: 2
weight_filler {
```

```
type: "gaussian"
std: 0.01
}
bias_filler {
type: "constant"
value: 0
}
}
}
layers {
name: "accuracy"
type: ACCURACY
bottom: "fc8"
bottom: "label"
top: "accuracy"
include: { phase: TEST }
}
layers {
name: "loss"
type: SOFTMAX_LOSS
bottom: "fc8"
bottom: "label"
top: "loss"
}
```

Training And Testing Data

The database is divided in the CNN release layer (possible layer) on CNN contains 8 values for 8-year courses ("0-2", "4-6", "8--13", "15 - 20", "25– 32 "," 38-43 "," 48-55 "and" 60- ").

Training Data:

A training dataset is a set of examples used to train the model i.e. equations and parameters. Most of the methods used to train the samples tend to skip if the database is not mounted and used in a variety of ways.

Validation Data:

The validation data is also called the 'development dataset' or 'dev set' and is used to fit the hyper parameters of the classifier. You are required to have validation data as well as training and assessment data because it helps to avoid excesses. The ultimate goal is to select the network that performs best on the raw data which is why we use an independent validation database in the training dataset.

Testing Data:

Test data does not depend on training manual or validation data. If the model is suitable for both the training data and the experimental data it can be said that an excessive bias has occurred. Test data is data used only to evaluate the performance of a classifier or model. An evaluation dataset was used to look at performance characteristics such as accuracy, loss, sensitivity, etc.

Fig: Training And Validation Accuracy Graph

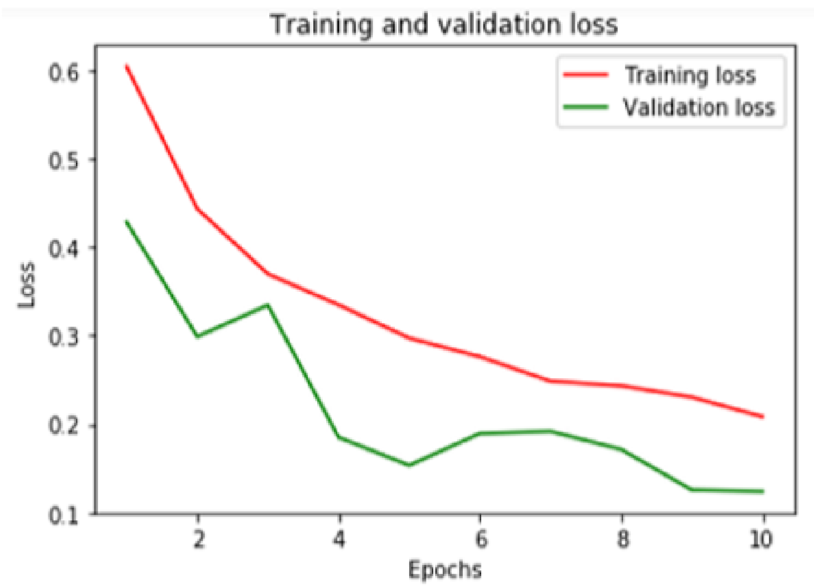


Fig: Training And Validation loss Graph.

The pre-processing unit analyzes image features based on algorithms. Preparing machine learning data is a technique used to convert raw data into a clean dataset i.e. whenever data is collected from different sources, it is collected in a raw format and the raw format is not suitable for analysis. After pre-processing the data, the model is trained using this clean data. Inserting an unknown image to guess the age and sex of an unknown image. The output now will have all the information regarding the age and gender of the unknown image which has been choosen for the input image.

Limitations And Challenges

Most of the applications such as computer, human crowd surveillance, face processing, artificial intelligent, content based image retrieval and video surveillance etc. require face detection for identification and verification the enrolled users. The skin color segmentation is the major problem for face identification. Facial segmentation accuracy depends on the pose, noise, lighting conditions and distance between the object and the camera. The various types of challenges coming in the picture during detection are described below:

1. Pose: The most challenging situation is that the human face varies with respect to the relative camera face pose (45 degrees, profile, frontal and upside down).

2. Facial Expression: The facial expression such as anger, fear, disgust, happiness, sadness and surprise is most influential temperaments for human beings to communicate their feelings.

3. Illumination: Illumination is a major challenge during the detection process. This factor is related to the lightening and angle of the light.

4. Occlusion: Occlusion is the main challenge during gender detection because sometimes the face is partially covered and occluded by others objects.

5. Imaging Condition: During the face image capture some factors such as different lightening conditions and camera characteristic (lenses, sensor response) affect the face recognition accuracy.

6. Different Facial Features: Different type of facial features such as glasses, beard, hair moustache, scars, moles, tattoos, skin colors and makeup affect the face recognition accuracy.

7. Face Size: This factor is also a major challenge because face size can vary a lot person to person. Not only different people have different sized faces but the face closer to the camera and far away from the camera also pose a challenge.

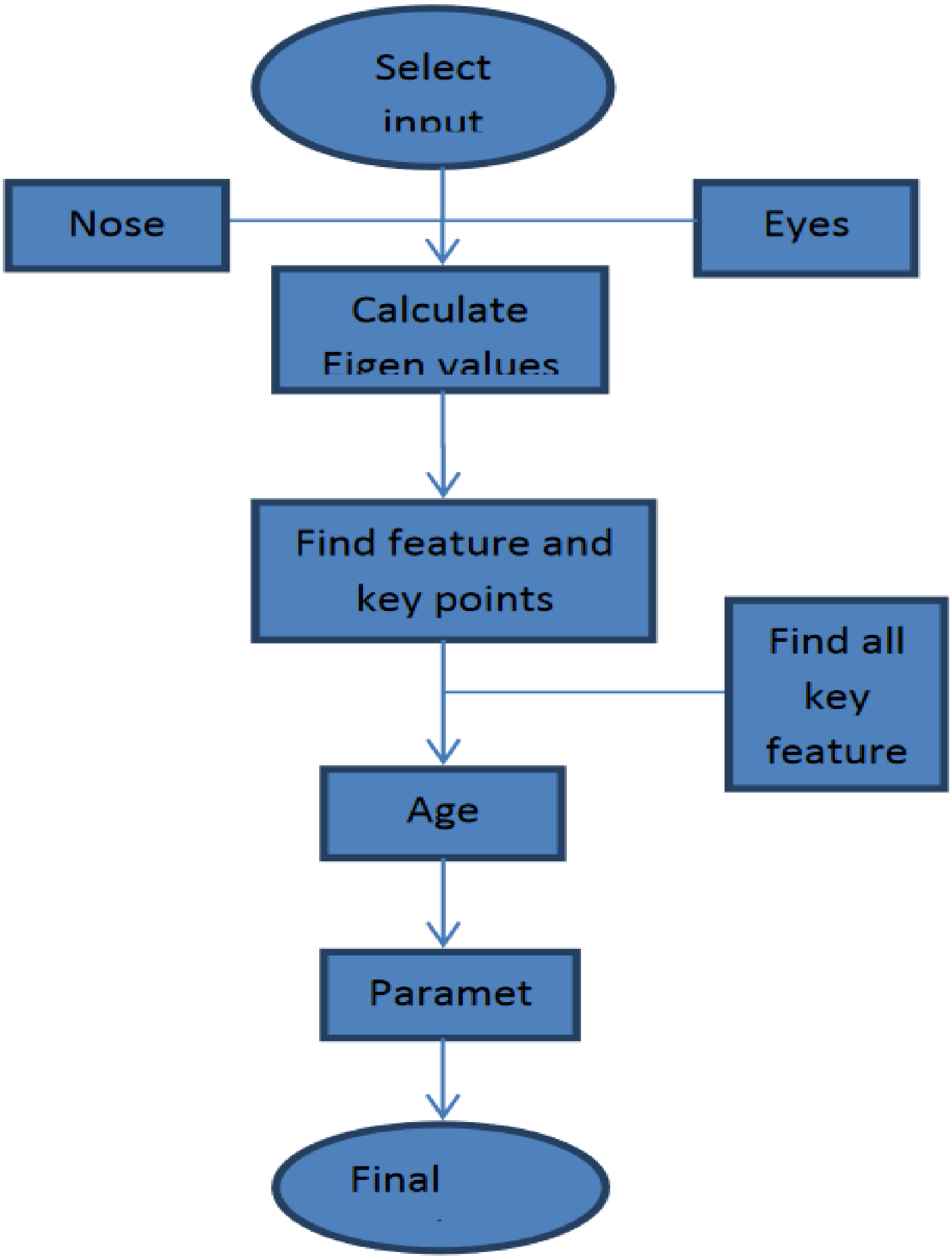
8.Age: It is difficult to gather the information among the small aged ones.

Applications

Some of the Applications of the estimation of Age and Gender Detecction will be as follows :

1. Forensic Department in the medical field as to gather information of the dead bodies.
2. In the banking sector to detect the information about the individual just by the images by age and gender detection.
3. Classify details of the individuals in the ADHAAR database.
4. Criminal Investigation Department to gather the information about the suspects by the age and gender detection.
5. Surveillance Monitoring.

Flow Chart of Proposed Technique



Mathematical Formulations

Local Response Normalization

After the first 2 layers of placement, there are normal response areas (LRN). LRN is a technique first used as far to help the generalization of deep CNNs. The idea behind it is to introduce sequential blocking between various convolution by making them s"competing" for maximum performance over a certain portion of their input. Effectively this prevents the repeated recording of the same information by different alternatives between different pins that point to the same input point and instead a few, more prominent, stimuli to perform other tasks in a specific location. If $a^i_{x,y}$ is a function of a neuron by using the kernel i in the area (x, y), then its local response is normalized to $b^i_{x,y}$ is given by

$$b^i_{x,y} = a^i_{x,y} / \left(k + \alpha \sum_{j=max(0,i-n/2)}^{min(N-1,i+n/2)} (a^j_{x,y})^2 \right)^{\beta}$$

where k, n, α, and all are all hyper parameters. The parameter n is the number of "closest" kernel maps (eflters) in which the LRN is active, and N is the total number of kernels in that given layer.

Softmax

At the top of the proposed structure sits a softmax layer, consisting of improved lost time during training and class opportunities during organization. While other layers of loss such as multiclass SVM loss manage the output of a completely connected layer such as classroom scores, softmax (also known as multinomial logistic regression) treats these schools as unofficial log statistics. That is, if we have the z_i

grade assigned to the class i after the fully connected layer, then the function of the softmax

$$f_j(z) = \frac{e^{z_j}}{\sum_k e^{z_k}}$$

To maximize the log of the class so the minimize the negative log likelihood the formulation will be

$$L_i = -log(\frac{e^{f_{y_i}}}{\sum_j e^{f_j}})$$

Because the softmax function takes the actual output from f and makes it normalized to their specified value, it ensures that the sum of all softmax effects is 1, thus allowing it to be interpreted as a real phase opportunity. It should be noted that softmax loss is actually some form of cross loss. Specifically, the cross-section between the original p distribution and the corresponding distribution q is given as

$$H(p,q) = -\sum_x p(x)logq(x)$$

From this it can be seen that the softmax classi- fier actually reduces the error between the estimated phase distributions and the actual distribution, which may appear as 1 predicted in the real phase and 0 predicted in the rest.

Stochastic Gradient Descent

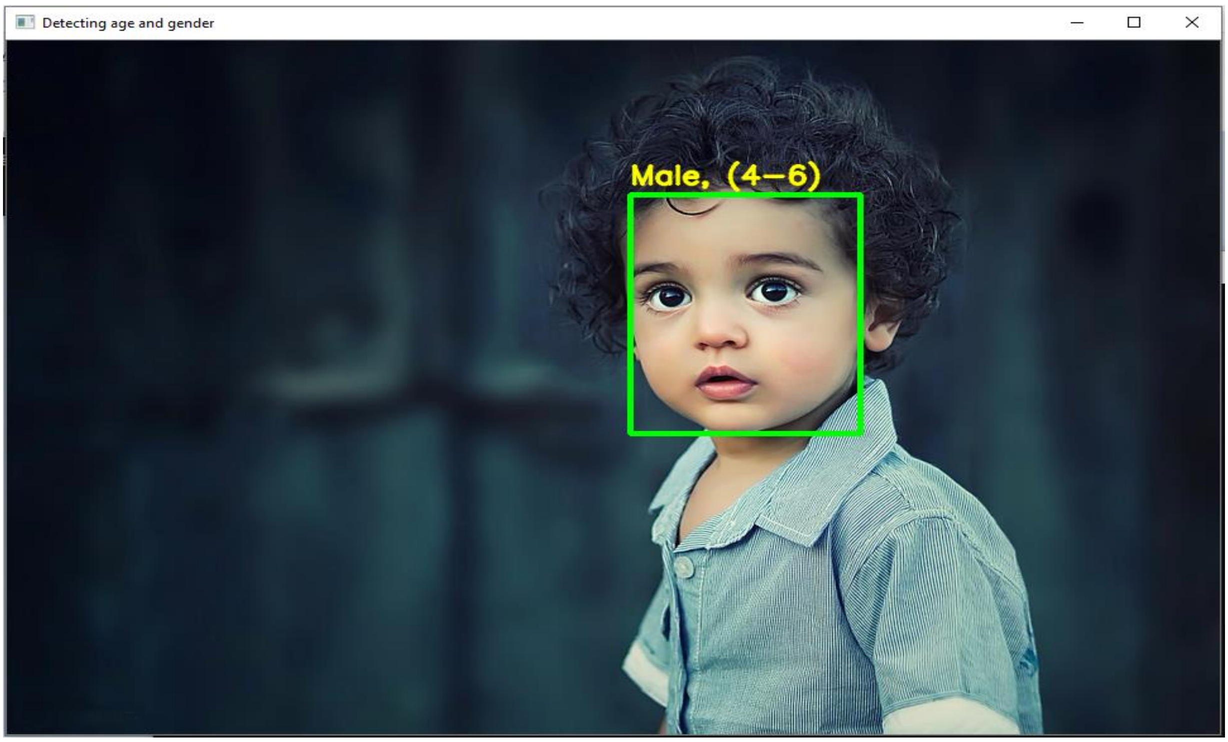
Now that we can calculate losses, we need to know how to reduce them in order to train the right team. The type of material used in this study is a traditional style of style. In order to explain this, first I will elaborate on a very good traditional environment. The task gradient is actually just based on it, and as a result, is the direction of the maximum increase (or decrease if you step back from it). So if we apply the gradient of the loss function with respect to all system transformers / components (on CNNs with billions of these), we will have a direction in which we can proceed to our very small loss immediately following the negative gradient. Each time we calculate a gradient we take a small step (which is controlled by a hyper parameter) on the other side, and we also evaluate the loss, we also measure the gradient, again. The hope (and indeed the truth) is that by repeating this process we will reduce our breakthrough performance, which in turn is a better model for its integrative function. Mathematically, we can write this as

$$\mathbf{w} = \mathbf{w} - \eta \nabla_{\mathbf{w}} L$$

where η is the reading rate, and is sometimes called the step size and $\nabla_W L$ is the time line of loss in terms of mass w.

While this is theoretically great, the truth is that computing the gradient across the entire training set in order to make an incremental update to the weights is prohibitively computationally expensive. Therefore making this form of mini-batch Gradient Descent.

Some examples of Output:



Conclusion

In this work, it is concluded that age and gender research has been the focus of the last few years. Despite the fact that many of the strategies of the past focused on issues of age and sexuality, not so long ago, this work certainly focuses on the compelling images taken in laboratory settings. Such settings do not adequately reflect the general appearance types of current reality photos on social networking sites and online archives. Web images, anytime, are not just about how complex they are: they are equally saturated. Easy access to great collections of high quality video readings of a learning machine with ongoing preparation information. CNN can be used to provide effects of age and age order, not by looking at the smallest size of the uneducated image of age and sexuality, Finally, I hope that more training material will be found with work age and gender cohesion that will allow effective techniques from other forms of big data sets to be used this place. We hope you found this paper well read and useful in your quest. Taking illustration from the related issue of face acknowledgment, we investigate how well profound CNN perform on these assignments utilizing Internet information. We provide results with an incline profound learning architecture designed to keep away from over fitting because of the impediment of constrained marked information. Our system is "shallow" contrasted with a portion of the late system designs, along these lines diminishing the quantity of its parameters and the chance for over fitting. We advance swell the extent of the preparation information by falsely including trimmed variants of the images in our preparation set. The subsequent framework was tried on the Adience benchmark of unfiltered images and appeared to fundamentally beat late cutting edge. Two critical conclusions can be produced using our experimental outcomes. In the first place, CNN can be utilized to give enhanced age and gender arrangement results, notwithstanding considering the much little size of contemporary unconstrained image sets named for age and gender classification. Second, the straight forwardness of our model suggests that more involved frameworks utilizing all the more preparing information might well be able to do significantly enhancing results beyond the one reported here.

Future Works

When changing a dataset, the same model can be trained to predict the feelings of race etc. Age and gender classifications can be used to predict age and gender in uncontrolled real-time situations such as train stations, banks, buses, airports, etc. For example, depending on the number of male and female passengers by the age on the train station, toilets and restrooms can be built to facilitate transportation.

References

1. A. Kumar and F. Shaik," Importance of Image Processing", 2016.

2. Chenjing Yan, Congyan Lang, Tao Wang, Xuetao Du, and Chen Zhang," Age Estimation Based on Convolutional Neural Network", 2014 Springer International Publishing Switzerland.

3. Hang Qi and Liqing Zhang," Age Classification System with ICA Based Local Facial Features", 2009 Springer-Verlag Berlin Heidelberg.

4. Eran Eidinger, Roee Enbar, and Tal Hassner," Age and Gender Estimation of Unfiltered Faces", 2014 IEEE.

5. Aditya K. Saxena, Shweta Sharma and Vijay K. Chaurasiya," Neural Network based Human Age-group Estimation in Curvelet Domain", 2015 The Authors Published by Elsevier.

6. J. Chen, S. Shan, C. He, G. Zhao, M. Pietikainen, X. Chen, and W. Gao. Wld: "A robust local image descriptor", Trans.Pattern Anal. Mach. Intell.,

7. S. E. Choi, Y. J. Lee, S. J. Lee, K. R. Park, and J. Kim. "Age estimation using a hierarchical classifier based on global and local facial features", Pattern Recognition, 44(6):1262–1281, 2011

8. T. F. Cootes, G. J. Edwards, and C. J. Taylor. "Active appearance models", In European Conf. Comput. Vision, pages 484–498. Springer, 1998.

9. C. Cortes and V. Vapnik. "Support-vector networks", Machinelearning, 20(3):273–297, 1995.

10. E. Eidinger, R. Enbar, and T. Hassner. "Age and gender estimation of unfiltered faces", Trans. on Inform.Forensics andSecurity, 9(12), 2014.

11. Y. Fu, G. Guo, and T. S. Huang. "Age synthesis and estimation via faces: A survey", Trans. Pattern Anal. Mach. Intell., 32(11):1955–1976, 2010.

12. Y. Fu and T. S. Huang. “Human age estimation with regression on discriminative aging manifold”, Int. Conf. Multimedia,10(4):578–584, 2008.

13. K. Fukunaga. “Introduction to statistical pattern recognition”, Academic press, 1991.

14. A. C. Gallagher and T. Chen. “Understanding images of groups of people”, In Proc. Conf. Comput. Vision PatternRecognition, pages 256–263. IEEE, 2009.

15. F. Gao and H. Ai. “Face age classification on consumer images with gabor feature and fuzzy LDA method”, In Advancesin biometrics, pages 132–141. Springer, 2009.

16. X. Geng, Z.-H. Zhou, and K. Smith-Miles. “Automatic age estimation based on facial aging patterns”, Trans. PatternAnal. Mach. Intell., 29(12):2234–2240, 2007.

17. B. A. Golomb, D. T. Lawrence, and T. J. Sejnowski. Sexnet: “A neural network identifies sex from human faces”, In NeuralInform. Process. Syst., pages 572–579, 1990.

18. A. Graves, A.-R. Mohamed, and G. Hinton. “Speech recognition with deep recurrent neural networks”, In Acoustics,Speech and Signal Processing (ICASSP), 2013 IEEE Inter-national Conference on, pages 6645–6649. IEEE, 2013

19. G. Guo, Y. Fu, C. R. Dyer, and T. S. “Huang. Image-based human age estimation by manifold learning and locally adjusted robust regression”, Trans. Image Processing, 17(7):1178–1188, 2008. 2

20. G. Guo, G. Mu, Y. Fu, C. Dyer, and T. Huang. “A study on automatic age estimation using a large database”, In Proc. Int.Conf. Comput. Vision, pages 1986–1991. IEEE, 2009.