Gender and Age detection with Emotion recognition

Geethika Lingamaneni
geethika.lingamaneni@mavs.uta.edu

Roshini Tadi
roshini.tadi@mavs.uta.edu

Snigdha Kethireddy
snigdha.kethireddy@mavs.uta.edu

Department of Computer Science and Engineering
The University of Texas at Arlington

2. ABSTRACT

Human facial image processing has been an active and interesting research topic over years. Since human faces provide a lot of information, many topics have drawn lots of attention. Predicting gender and age as well as emotions from the images in unconstrainted conditions is one of the challenging tasks in many real-world applications. Gender-age and emotion detection plays a key role in consumer applications. Humans are capable of producing thousands of facial actions during communication that varies in complexity, intensity, and meaning.

KEYWORDS: *Gender, Age, Emotion, Detection, Recognition, Coco, Deep learning, CNN, Tensorflow, Keras*

3. INTRODUCTION

Face recognition is one of the wonders that AI research has brought into the world. It is a subject of curiosity for many people. Biometrics are used by a facial recognition system for mapping facial features from an image. Also can be used for various categories from surveillance to marketing, even under changing light and changes occurred by age or with any obstructions. It is a process that comprises of four steps. They are detection, alignment, feature extraction, and a recognition task. These models and methods can support very large datasets of faces and learn great and compact representations of faces. Age and gender play important roles in social interactions. It's classification became applicable to an increasing amount of applications from the increase in social platforms. So, extracting gender-age related attributes from facial images has received great attention in recent years and lots of methods are proposed. Previous methods for age and gender estimation are based on calculating ratios between various measurements of facial features. In these methods a small set of near-frontal face images are trained using a neural network for gender-age classification. For gender-age estimation , CNN gathered an outsized, labeled image training set from social image repositories. Humans are generally used to taking in non-verbal cues from facial emotions. Computers are too getting better in reading emotions. Emotion recognition helps to identify human emotion. People vary widely in their accuracy at recognizing the emotions of others. The use of Technology to help people with emotion recognition is an interesting area of research. The most work has been conducted on automating the recognition of facial expressions from video, audio, text.

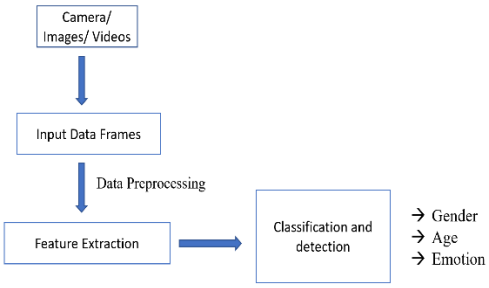
In deep learning, CNN(Convolutional neural network or ConvNet) could be a class of deep neural networks, most ordinarily applied to analyzing visual imagery. CNNs use little pre-processing when compared to other image classification algorithms. This might be an advantage of CNN as it is independence from previous knowledge and human effort in feature design. They have applications in image, video recognition, recommender systems, image classification, brain-computer interfaces, linguistic communication processing, and financial statistics. TensorFlow is capable of training and also to run deep neural networks for sequence-to-sequence models for artificial intelligence, handwritten digit classification, image recognition, recurrent neural networks, linguistic communication processing, and partial equation based simulations. Keras is an open-source library. It provides a Python interface for artificial neural networks. It consists of various implementations of commonly used neural-network building necessary for writing deep neural network code. OpenCV is short for Open Source Computer Vision. It is an open-source Machine Learning library and Computer Vision. This library is not only capable of processing real-time images and videos but also

have analytical capabilities. It assists Deep Learning frameworks, TensorFlow, Caffe, and PyTorch.

3.1 PROPOSED METHOD

Our proposed method is to detect the gender and age of the people along with their emotions for the images, videos and live inputs through webcam. For this we are considering a dataset and preprocessing the data to separate the class of persons from the entire dataset and then using it for building the model and running the output.

3.2 System Architecture:



4. DATASET DESCRIPTION:

We used both UTKFace and Adience datasets for training. COCO dataset and the webcam for testing.

UTKFace dataset is a large-scale face dataset that consists of a long age span range from 0 to 116 years old. It has over 20,000 face images with annotations of age, gender. This dataset could be used for different purposes.

Adience dataset is available in the public domain. This dataset acts as a benchmark for photos with faces and has various real-world imaging conditions like , appearance, noise, lighting, pose. It consists of 26,580 photos of 2,284 subjects in eight age ranges and the size is about1GB.

COCO (Common Objects in Context) dataset is a large-scale image recognition/classification, object detection, segmentation, and captioning dataset. It has a volume of 330K images where more than 200K are annotated and has more than 2M instances in 80 object categories, 91 stuff categories, with 5 captions per image, and 250,000 people with key points.

4.1 Data Preparation:

We have preprocessed the dataset by labelling the classes and separated the persons category from the entire dataset. We need only the person class or category from the coco dataset our project to detect age, gender and emotions. We have then categorized the person class from all other classes and stored it in another folder for our further use.
This classified folder of only persons is used in our project for further process. From these images we are training and testing the model to detect the age, gender and emotion of the person.

5. PROJECT DESCRIPTION:

5.1 Description:

For our proposed model we need to initially detect the faces from the considered dataset. Here we have used UTKData set in order to train our model and tested using the coco images and open cv. In UTK dataset we just have images with the labels in format of "age"_"gender", we denoted age with numbers, and "0" and "1" for gender. So we basically extract the information of the image from the label. Next, we have split the dataset into 75% training and 25% testing in order to generate our model. In the process of model generation we used CNN, where we underwent four convolutions with initial kernel size as (3,3) and filters from 64,128,256. Finally we applied the activation functions like the relu and sigmoid and constructed the model. With the help of the saved model by above process, we generated Epoch for every batch of the data and got the accuracy, mean-loss as shown in output screenshot. In emotion detection, we have used CNN to train our model. We have taken coco dataset to train and test the model, 70% for training , 30% for testing. We have used 2D convolution and used resnet layers to filter our test set. We have divided the test set into batch size of 128 and 30 epochs and got accuracy, mean-loss for every batch. We classified the emotions as angry, disgust, fear, happy, neutral, sad, surprise. We have added activations like relu in the convolution process in order to smoothen the data. We have written a block of code inorder to detect and highlight the face using opencv or the images. We have then captured the video from the camera and obtained the faceboxes and iterated through it and passed the same to the model.

We have a .pb file for face detection. This file consists of the trained weights and the graph definition of the model. Using this we can run the trained model. A .pb file is in binary format and the .pbtxt extension is in text format. For age and gender, .caffemodel file gives the internal states and the .prototxt file explains the network configuration. We use the argparse library to form an argument parser so we are able to get the image argument from the prompt. We make the parser to parse the argument which is having the path to the image which needs to be classified. The readNet() method is used to load the networks. When we capture through the webcam, we read the stream and store content into the names. If it's not a video, it must wait, and then we call waitKey() from the cv2 then end it with a break. When we call the hl_Face() we will be storing the resultants as resultImg and faceBoxes. And when ever we get 0 faceBoxes, it means there is no face to detect and also the list is empty. Then the input is set and a forward pass to the network is made. When we find the confidence threshold lesser than the confidence mentioned, we then get x1, x2, y1 and y2 coordinates and thus it is then appended as list to faceBoxes. Then, we plot rectangles on these images for list of coordinates and return the list of faceBoxes and the shallow copy. But if there are faceBoxes, we define the face and create a 4D blob from the image. With all this process being done we will be able to resize it, scale it and pass the mean values. We give the network a pass forward to attain the confidence after feeding the input. At last we add the gender and age labels to the resulting image and display it with imshow().

5.2 Main References used:

[1]https://towardsdatascience.com/facial-data-based-deep-learning-emotion-age-and-gender-prediction-47f2cc1edda7

[2] https://github.com/smahesh29/Gender-and-Age-Detection

[3] https://www.kaggle.com/krsna540/toon-emotion-detection/notebook

[4] https://www.kaggle.com/rahuldshetty/age-and-gender-estimation-using-cnn

5.3. DIFFERENCE IN APPROACH BETWEEN OUR PROJECT AND REFERENCES

Our entire project is to build a model to detect gender and age with emotions of a person. In the references we used, there is no model which does both the detections (Emotions and Gender-Age) together and the models are built on different datasets. We couldn't find even one model which uses COCO dataset for detecting either the emotions or the gender-age. A model to detect Emotions and Gender-Age together is our enhancement. We have also implemented the model for the images, video and through web cam as well. We were able to detect multiple faces with age, gender, emotion detection.

5.4. DIFFERENCE IN ACCURACY BETWEEN OUR PROJECT AND REFERENCES

We have attached the outputs below for the gender-age and emotions.

- Gender Accuracy: 90%-95%
- Age Accuracy: 40%-80%
- Emotion Accuracy: 90%-98%

5.5. CONTRIBUTION

1. Implementation from scratch -CNN – 4 Layers of convolution with filter sizes 64,128,256. Initial kernel size is (3,3) and later (2,2)

2. We have used Keras model from Tensorflow to load the saved model.

3. Bounding boxes are plotted on faces by using 0.7 Threshold.

4. We almost got the similar accuracy w.r.t the original reference.

- Code: Snigdha, Geethika, Roshini
- Documentation: Snigdha, Geethika, Roshini
- Presentation: Snigdha, Geethika, Roshini

OUTPUTS

CNN Table:

conv2d_5 (Conv2D)	(None, 24, 24, 64)	18496	max_pooling2d_4[0][0]
dropout_7 (Dropout)	(None, 24, 24, 64)	0	conv2d_5[0][0]
activation_5 (Activation)	(None, 24, 24, 64)	0	dropout_7[0][0]
max_pooling2d_5 (MaxPooling2D)	(None, 12, 12, 64)	0	activation_5[0][0]
conv2d_6 (Conv2D)	(None, 12, 12, 128)	73856	max_pooling2d_5[0][0]
dropout_8 (Dropout)	(None, 12, 12, 128)	0	conv2d_6[0][0]
activation_6 (Activation)	(None, 12, 12, 128)	0	dropout_8[0][0]
max_pooling2d_6 (MaxPooling2D)	(None, 6, 6, 128)	0	activation_6[0][0]
conv2d_7 (Conv2D)	(None, 6, 6, 256)	295168	max_pooling2d_6[0][0]
dropout_9 (Dropout)	(None, 6, 6, 256)	0	conv2d_7[0][0]
activation_7 (Activation)	(None, 6, 6, 256)	0	dropout_9[0][0]
max_pooling2d_7 (MaxPooling2D)	(None, 3, 3, 256)	0	activation_7[0][0]
flatten_1 (Flatten)	(None, 2304)	0	max_pooling2d_7[0][0]
dense_2 (Dense)	(None, 64)	147520	flatten_1[0][0]
dense_3 (Dense)	(None, 64)	147520	flatten_1[0][0]
dropout_10 (Dropout)	(None, 64)	0	dense_2[0][0]
dropout_11 (Dropout)	(None, 64)	0	dense_3[0][0]
sex_out (Dense)	(None, 1)	65	dropout_10[0][0]

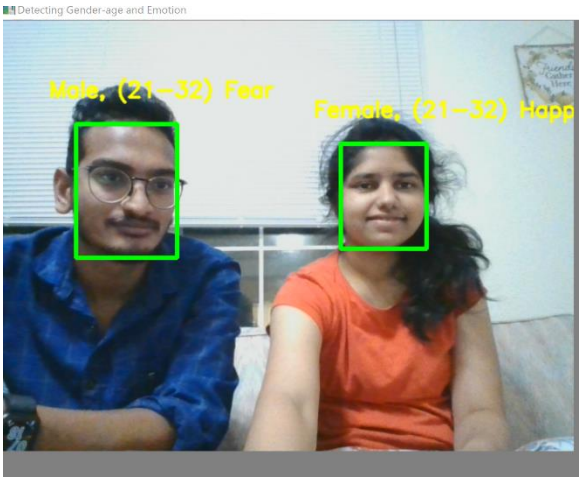
Gender and age Process

```
sex_out_accuracy: 0.9194 - age_out_accuracy: 0.0441
Epoch 00109: val_loss did not improve from 7.63451
278/278 [==============================] - 89s 319ms/step - loss: 5.5287 - sex_out_loss: 0.1935 - age_out_loss:
4.3028 - sex_out_accuracy: 0.9194 - age_out_accuracy: 0.0441 - val_loss: 8.7586 - val_sex_out_loss: 0.2539 -
val_age_out_loss: 7.4711 - val_sex_out_accuracy: 0.8957 - val_age_out_accuracy: 0.0457
Epoch 110/500
278/278 [==============================] - ETA: 0s - loss: 5.6489 - sex_out_loss: 0.1928 - age_out_loss: 4.4215 -
sex_out_accuracy: 0.9180 - age_out_accuracy: 0.0430
Epoch 00110: val_loss did not improve from 7.63451
278/278 [==============================] - 93s 334ms/step - loss: 5.6489 - sex_out_loss: 0.1928 - age_out_loss:
4.4215 - sex_out_accuracy: 0.9180 - age_out_accuracy: 0.0430 - val_loss: 8.0251 - val_sex_out_loss: 0.2483 -
val_age_out_loss: 6.7422 - val_sex_out_accuracy: 0.8998 - val_age_out_accuracy: 0.0457
Epoch 111/500
278/278 [==============================] - ETA: 0s - loss: 5.5779 - sex_out_loss: 0.1869 - age_out_loss: 4.3563 -
sex_out_accuracy: 0.9220 - age_out_accuracy: 0.0438
Epoch 00111: val_loss did not improve from 7.63451
278/278 [==============================] - 90s 325ms/step - loss: 5.5779 - sex_out_loss: 0.1869 - age_out_loss:
4.3563 - sex_out_accuracy: 0.9220 - age_out_accuracy: 0.0438 - val_loss: 8.3509 - val_sex_out_loss: 0.2381 -
val_age_out_loss: 7.0770 - val_sex_out_accuracy: 0.8998 - val_age_out_accuracy: 0.0457
Epoch 112/500
278/278 [==============================] - ETA: 0s - loss: 5.6440 - sex_out_loss: 0.1932 - age_out_loss: 4.4147 -
sex_out_accuracy: 0.9193 - age_out_accuracy: 0.0442
Epoch 00112: val_loss did not improve from 7.63451
278/278 [==============================] - 91s 328ms/step - loss: 5.6440 - sex_out_loss: 0.1932 - age_out_loss:
4.4147 - sex_out_accuracy: 0.9193 - age_out_accuracy: 0.0442 - val_loss: 8.4844 - val_sex_out_loss: 0.2345 -
val_age_out_loss: 7.2135 - val_sex_out_accuracy: 0.9015 - val_age_out_accuracy: 0.0457
Epoch 113/500
278/278 [==============================] - ETA: 0s - loss: 5.5664 - sex_out_loss: 0.1913 - age_out_loss: 4.3382 -
sex_out_accuracy: 0.9217 - age_out_accuracy: 0.0429
Epoch 00113: val_loss did not improve from 7.63451
278/278 [==============================] - 93s 335ms/step - loss: 5.5664 - sex_out_loss: 0.1913 - age_out_loss:
4.3382 - sex_out_accuracy: 0.9217 - age_out_accuracy: 0.0429 - val_loss: 8.4440 - val_sex_out_loss: 0.2471 -
val_age_out_loss: 7.1595 - val_sex_out_accuracy: 0.8969 - val_age_out_accuracy: 0.0457
186/186 [==============================] - 7s 38ms/step - loss: 7.6345 - sex_out_loss: 0.2340 - age_out_loss:
6.6531 - sex_out_accuracy: 0.8922 - age_out_accuracy: 0.0457
Predicted Age: 89
Predicted Sex: Male
```

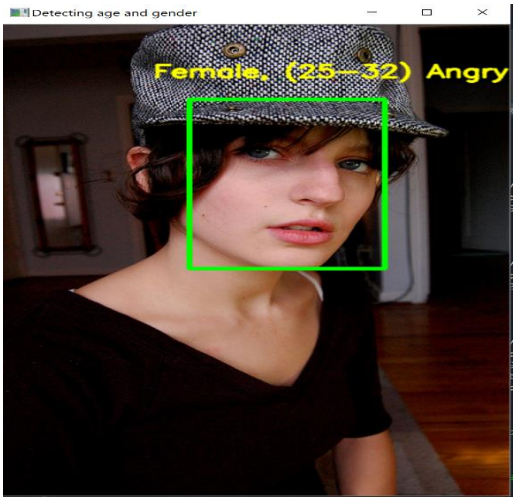
Emotion Process:

```
==
Total params: 284,042
Trainable params: 282,666
Non-trainable params: 1,376
_________________________________________________________________
Model_Net20v1
Rate of Learning:  0.001
Epoch 1/30
44/44 [==============================] - ETA: 0s - loss: 0.2571 - accuracy: 0.9736
```

Testing Using Open CV:



Testing using COCO images:



6. ANALYSIS

We worked on combining gender-age detection with emotions on COCO dataset as there is no model that does both. We could have worked more on the accuracy part. We also extended our project for multiple faces using open cv.

7. CONCLUSION

We have implemented a CNN to predict both the gender-age and emotion from facial images in this project. We have detected faces, then classified them into male/female with age ranges and emotions and then the results are displayed on the image.

8. REFERENCES

[1]https://towardsdatascience.com/facial-data-based-deep-learning-emotion-age-and-gender-prediction-47f2cc1edda7

[2]https://github.com/smahesh29/Gender-and-Age-Detection

[3] S. Kumar, S. Singh and J. Kumar, "A study on face recognition techniques with age and gender classification," 2017 International Conference on Computing, Communication and Automation (ICCCA), Greater Noida, 2017, pp. 1001-1006, doi: 10.1109/CCAA.2017.8229960.

[4]Safak, Emre & Barişçi, Necaattin. (2018). Age and Gender Prediction Using Convolutional Neural Networks. 1-7. 10.1109/ISMSIT.2018.8567066.

[5]Khaung Tin, Dr.Hlaing Htake. (2011). Gender and Age Estimation Based on Facial Images. International Journal : ACTA TECJNICA NAPOCENSIS Electronics and Telecommunications.

[6]A. Varghese, J. P. Cherian and J. J. Kizhakkethottam, "Overview on emotion recognition system," 2015 International Conference on Soft-Computing and Networks Security (ICSNS), Coimbatore, 2015, pp. 1-5, doi: 10.1109/ICSNS.2015.7292443.

[7]https://www.kaggle.com/krsna540/toon-emotion-detection/notebook

[8]https://github.com/awslabs/keras-apache-mxnet/blob/master/examples/cifar10_resnet_multi_gpu.py

[9] https://arxiv.org/pdf/1506.02203.pdf

[10]http://people.vision.caltech.edu/~mronchi/projects/Cocoa/

[11]https://www.kaggle.com/rahuldshetty/age-and-gender-estimation-using-cnn

[12] https://www.kaggle.com/jangedoo/utkface-new?

[13]https://www.kaggle.com/ttungl/adience-benchmark-gender-and-age-classification

[14]https://data-flair.training/blogs/python-project-gender-age-detection/

9. ACKNOWLEDGEMENTS

The system discussed in this article was implemented as a project for the FALL 2020 CSE 6363-Machine Learning class at The University of Texas at Arlington.
This course was lectured by **Dr Jesus Antonio Gonzalez Bernal**.
Thank you for reading this article carefully. You are welcome to submit any feed back or suggestions to our respective email addresses:
geethika.lingamaneni@mavs.uta.edu
roshini.tadi@mavs.uta.edu
snigdha.kethireddy@mavs.uta.edu