Image Processing: Identifying Age and Gender

1st Gloria Abuka
Department of Computer Science
Middle Tennessee State University
Murfreesboro, TN, USA
gea2f@mtmail.mtsu.edu

2nd Rober Makram
Department of Computer Science
Middle Tennessee State University
Murfreesboro, TN, USA
rbm3d@mtmail.mtsu.edu

3rd Kirolous Shihataa
Department of Computer Science
Middle Tennessee State University
Murfreesboro, TN, USA
krs6q@mtmail.mtsu.edu

4th Jessica Wijaya
Department of Computer Science
Middle Tennessee State University
Murfreesboro, TN, USA
jcw8h@mtmail.mtsu.edu

5th Hannah Williams
Department of Computer Science
Middle Tennessee State University
Murfreesboro, TN, USA
hnw2y@mtmail.mtsu.edu

***Abstract*—This document is a model and instructions for LaTeX. This and the IEEEtran.cls file define the components of your paper [title, text, heads, etc.]. *CRITICAL: Do Not Use Symbols, Special Characters, Footnotes, or Math in Paper Title or Abstract.**

***Index Terms*—component, formatting, style, styling, insert**

I. Introduction

Facial recognition has generally gained much popularity over the years for the important role it plays across various industries. Its applications are greatly in use in the security, medical, and target-advertising industries. In 2020, the global pandemic changed the world as we knew it, and this brought about the need to find a way to reduce physical contact while maintaining usual day to day interactions. Facial recognition also plays a vital role in that aspect. The ever-growing need for applications of age and gender classification through facial images has attracted the attention of many researchers over the years.

Many previous studies have been conducted to create neural networks that can obtain information from facial images including a person's age, gender, ethnicity, etc. The goal of our research project is to train two neural nets, one that can classify images of people by their gender and one to determine the age of the individual. The IMDB-WIKI dataset will be used to train the models. This dataset contains over 500,000 face images that were web scraped from IMDb and Wikipedia [1]. Additionally, only the images containing a single person will be used as input and the data will be preprocessed to prepare the images for the neural net. We will be using transfer learning to build upon pretrained models and attempt to improve the performance of these networks.

A. History

Talking about image processing and computer vision, convolutional neural networks (CNN) are recognized as the foundation on which such models thrive. The idea of the CNN started with the discovery of David Hubel and Torsten Wiesel back in 1959. They discovered the idea of simple cells and complex cells in the human cortex, these cells are used in pattern recognition [2].

While a simple cell responds to edges and bars of a particular orientation, a complex cell responds to those edges and bars even when they are shifted in different positions around the scene. This property is known as "spatial invariance". Spatial invariance is achieved by summing the output of several simple cells that prefer the same orientation [2]. This concept forms the basis of convolutional neural networks which is adopted for our age and gender model in this paper.

A core contribution to this field was made in the 1980s by Dr. Kunihiko Fukushima. His discovery was inspired by the work of Hubel and Wiesel. He proposed the concept of the "neocognitron," this model consists of the S-cells and C-cells. The S-cells are like the simple cell while C-cells are like the complex cells. The major idea of this model was to capture the simple-complex and turn it into a computational model for pattern recognition [2].

The first model built on the concept of CNNs was in the 1990s by Yann Lecun who was inspired by the previous research. Lecun used a CNN model trained on the popular MNIST dataset to recognize handwritten digits 0-9. CNNs gained the huge popularity it has today in 2012 when a CNN called AlexNet achieved a great deal of success labelling pictures in the imageNet challenge [2]. Convolutional neural networks have come a long way over the past decades and the future looks even brighter.

II. Background

Of course, many developers have built models to attack this problem using various techniques. A research group at the Open University of Israel [3] created deep-convolutional neural networks to attempt getting an improved accuracy in identifying an individual's age and gender from an image of their face as compared to current models at the time of the article in 2015. Their network was composed of just three moderate sized convolutional layers, two relatively small dense layers, and an output layer. This model was much smaller than other models in the past that were built to solve the same problem. The group's reason for keeping it small is that they did not want to risk overfitting the model to the training data. The group utilized units with ReLU activation functions and also used dropout layers to make sure the network was being trained as evenly as possible. Their final net could identify the gender typically around 86

In a similar study that aimed to improve the accuracy of age group and gender predictions of real-world faces, researchers from the University of Kwazulu-Natal created a convolutional neural net with the goal of handling the many variations present in unfiltered images [4]. The neural net consists of four convolutional layers that are each followed by a dense layer with the ReLU activation function, a batch normalization layer, max-pooling, and a dropout layer. The softmax function is used for the output layer. The model was trained using the IMDb, MORPH-II, and OIU-Adience datasets. To achieve a higher accuracy in making predictions on unfiltered facial images, they implemented an image preprocessing algorithm to prepare the images being fed into the network. Their results yielded some success with an accuracy of about 96

For this project, we originally aimed to create our own models that could achieve a reasonable accuracy in predicting age and gender from facial images. Although the initial goal was to create two networks from scratch for age and gender classification, we ended up using transfer learning to add layers on to the Xception network in order to attain better results.

III. Methods

Like most convolutional neural networks that compute information from images, our workflow requires the input be preprocessed before being fed into the model. The data set we are using consists of thousands of images of people's faces, as well as MATLAB files that contain the meta data relating to each image. The MATLAB files have a lot of information about the person in each image like their gender and their birthdate or birthyear as well as the date of when the photo was taken. We used this information to determine what each individual's age was at the time of the photo. These are the parts of the data that we trained the age predicting portion of our network on.

In order to prepare the data for training, the metadata of the images also had to be processed. The data contains information of an individual's birthdate and the date that the image was taken. From this, we wrote some functions to calculate the age, in years, of the depicted person. Additionally, we extracted the gender, and face locations from the IMDb MATLAB file.

Furthermore, to process the images themselves for training, the image is first cropped using the coordinates that specify the face location. This will eliminate potential noise from the background and make the face the main focus of the image. Then the image is resized to be 200 by 200 pixels. Also, when the images are read in before training, they are checked for corruption and skipped over if corrupted. The images that contained no faces were similarly discarded.

After the preprocessing, we discovered that 129 of the IMDd dataset were bad but the wiki data set had no bad images. We deleted the corrupted images from the data set and moved onto building our models.

Our model for gender predictions begins with the pretrained Xception model. We tried various ways to avoid using a pretrained net such as adding a variety of convolutional, dense, and max pooling layers of various sizes, but ultimately the only thing that greatly increased the accuracy of the output was to use a pretrained net as the base model and add layers on top of it. The top of the Xception model is left off and the output is rerouted into the new layers added on top. After the base, our model consists of a global average pooling layer in order to take the average of each feature map of an image. After that we have a dense layer with 128 units with the ReLU activation function. Lastly, we have another dense layer the size of the output with the softmax activation function to help deal effectively with the classification problem. We then compile our model with the Adam optimizer and categorical cross entropy. We use categorical accuracy as the metric of the model.

The age model is set up in a largely similar way. It is also built on top of the Xception network with a pooling layer, a dense layer with 128 units using the ReLU activation function, and the output layer that uses softmax. Together, these two models would make up what we would use to train the model to predict the age and gender of a person in each image.

IV. Results

**Discuss the results and accuracy of our models.

**Add in the graphs and examples of images with their predicted age and gender vs real. Show examples where our models were successful and ones where it was incorrect.

After a lot of training on the data set, both models were sufficiently trained with decent accuracies. Around 360,000 images were trained on our model, we did this by splitting our datasets into groups of 40,000 images, training in groups, and saving the weights and the history of the accuracy.

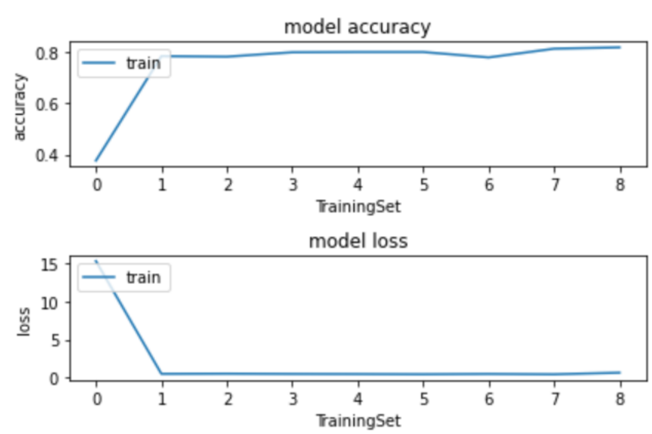

Fig. 1. Loss Graph Of Age Model

Figure 1 shows the loss graph of the model used to determine gender. Our model did well classifying images it did not train on to an accuracy of up to 82%. We started at an accuracy of 37% which we were able to drive up to 82% after training for a while.

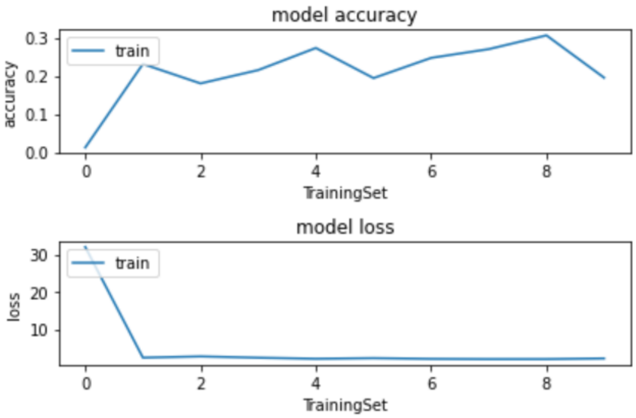

Fig. 2. Loss Graph Of Gender Model

Our age prediction model is designed to predict age groups rather than actual age of an individual from 0-130 in groups of 5 which makes it a total of 26 different age groups (0-4, 5-9, 10-14, 15-19...). We had a harder time training and getting our expected level of accuracy because of the nature of classification it is intended to do. Figure 2 shows the loss graph of our model. We started at an accuracy of 12% and we were up to drive it up to an accuracy of 31% after all the training was done. In the graph though, you see a dip in accuracy at the very end. This is because we trained it on one last data set which decreased the accuracy. Because we were saving the model after each set of 40,000 images were trained, we decided to just use the previous set of weights that were calculated and ignore the training that last part of the data set added.

V. Discussion

**Talk about expected results vs actual results.

**What sort of noise in the data made our models predict incorrectly, etc.

**Discuss what problems we encountered.

Over the course of this project, we encountered some challenges that made us reexamine our approach to creating a neural network that could predict age and gender from an image. The original goal had been to design and train a model from scratch that had some differences from the ones previously created. However, we had difficulties in training these models to reach a reasonable accuracy and decided to instead add layers on top of a pretrained base model. This yielded more accurate results for both the gender and age predictions.

Additionally, the training of these networks took a considerable amount of time with the large number of images present in the dataset. Some of the training also had to be repeated after it was discovered that the dataset contained images that has no faces in them since this would add unintended noise.

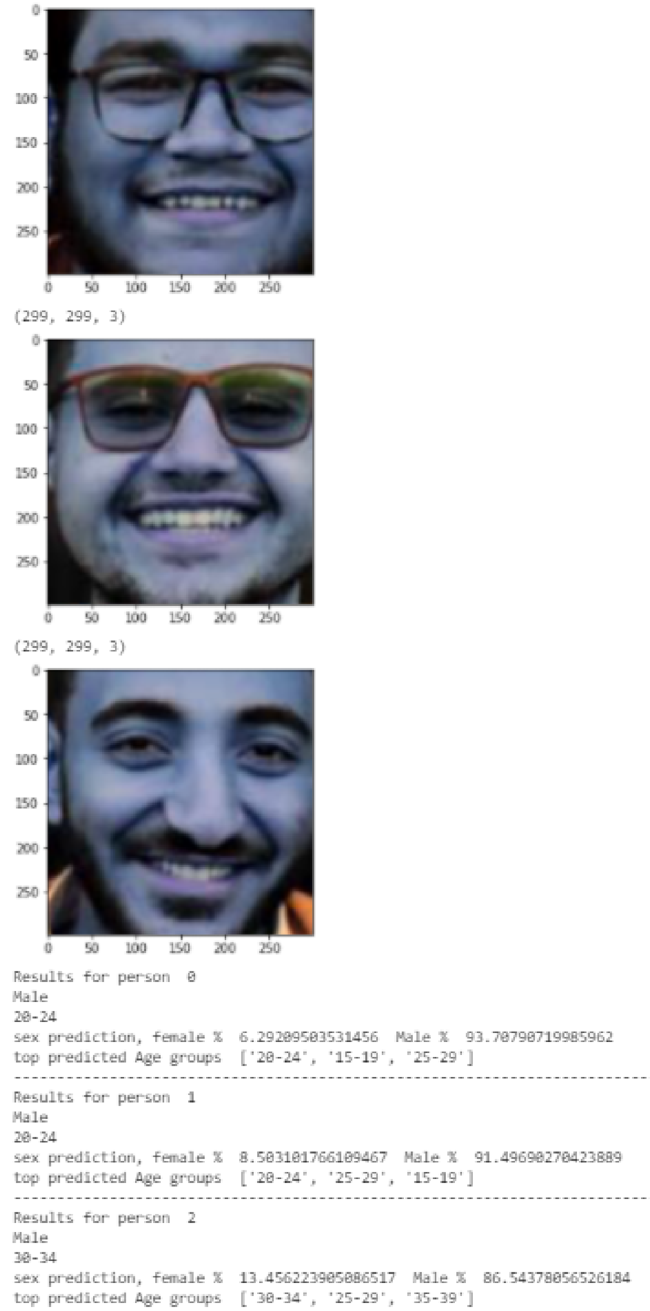

Fig. 3. Some Example Calculations Made By The Models

Write conclusion.

References

[1] R. Rothe, R. Timofte, and L. V. Gool, "Deep expectation of real and apparent age from a single image without facial landmarks," *International Journal of Computer Vision*, vol. 126, no. 2-4, pp. 144–157, 2018.

[2] R. L. B. Draelos, "The history of convolutional neural networks," 2020.

[3] G. Levi and T. Hassner, "Age and gender classification using convolutional neural networks," 2015.

[4] O. Agbo-Ajala and S. Viriri, "Deeply learned classifiers for age and gender predictions of unfiltered faces," *TheScientificWorldJournal*, 2020.

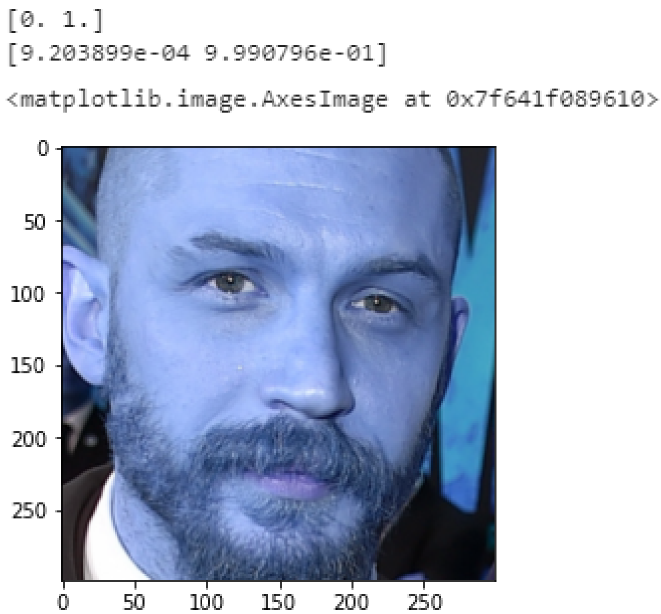

Fig. 4. Correct Classification Made By The Network

Fig. 5. Incorrect Classification Made By The Network