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*—The importance of facial recognition specifically in age and gender classification in today's world cannot be overemphasized as the demand for the applications that make use of these systems have been on the rise lately. In this paper, we examine a few existing models built of the convolutional neural network architecture, their level of success and our attempt to reach a comparable accuracy by building our own models using transfer learning based on a pre-trained Xception model with rich datasets for both age and gender classifications to better enhance the performance. The performance of our gender model was satisfactory with an accuracy of 82% while the age model reached an accuracy of 31% but was still shown to be able to output the correct age group classification within the top few predictions. Although we were not able to achieve a higher accuracy than the networks created in previous research, we found that using transfer learning with the Xception network was able to produce acceptable results in gender and age group predictions.**

I. Introduction

Facial recognition has 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]. 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.

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 in the Xception model through the concept of transfer learning 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 the problem of age and gender prediction using various techniques. A research group at the Open University of Israel created deep-convolutional neural networks to attempt to get 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% of the time and the age within one year around 84% of the time from an image the net had not trained on. The net typically had a hard time identifying the gender of young children because there would tend to be less identifying gender features present because of their age [3].

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. The neural net contains 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 gender predictions and roughly 83% for age group classifications. The images that were incorrectly classified typically had issues such as low resolution and poor lighting [4].

In light of this previous research, we aimed to create our own models that could achieve a comparable 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. With this we intend to see if we can achieve similar performance while utilizing a pretrained model with a few more layers added on for predicting age group and gender.

III. Methods

Like most convolutional neural networks that compute information from images, our workflow requires for the input to 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 dataset includes some variation such as images with multiple people, images with no people, and some noise from mislabeled data. 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. Additionally, the metadata includes the coordinates for each face located within an image. These are the parts of the metadata that we processed and utilized for the training of the networks.

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 in the meta data that specify the face location. This will eliminate potential noise from the background and make the face centered in 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 around 62,000 of the IMDb images were bad, and the WIKI data set had around 18,000 corrupt 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 as a base. The top of the Xception model was left off and the output was 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 that is 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 used 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 make up what we used to train the model to predict the age and gender of a person in each image.

Around 360,000 images were used to train our model. To make this easier, the images were divided into groups of 40,000. The convolution net was trained with each group and the weights and history was saved periodically to keep track of the progress and accuracy. Each group of 40,000 images was used to train the model in batches of 40 with a validation split of 0.2.

When it came to putting our models to the test, we used some images of our own to see what the model would predict. Outside images, of course, do not come with the comprehensive metadata that was included with the IMDb-WIKI dataset. We used a python facial recognition library

called face-recognition to detect a face or set of faces in an image [5]. If no face could be found by the software, we would not use that image. Once the face location or locations were returned, it would generate an image or set of images that are cropped down to the individual's faces. This would eliminate any outside noise from the background of the images that may affect the predictions. These modified images would then be run through the models. It then could output the top decisions of age group and gender as well as the percentage likelihood that the person in the photo was a male or a female and the top three age group predictions made by the net.

IV. Results

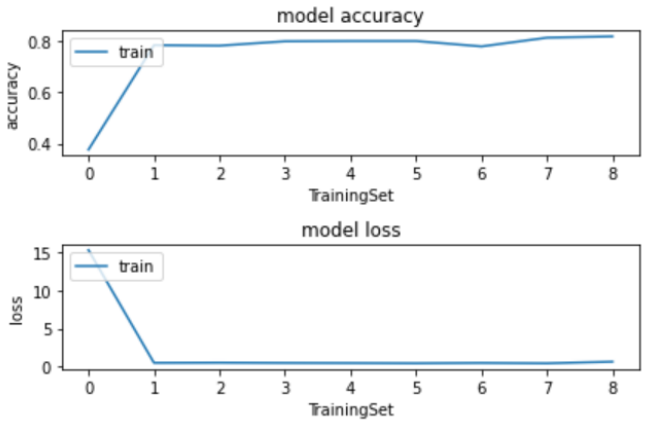

Fig. 1. Accuracy And Loss Graph Of The Gender Model

After extensive training, both models were able to perform acceptably. Figure 1 shows the loss graph of the model used to determine gender. Our model did well in classifying images it had never seen before with an accuracy of up to 82%. We started at an accuracy of 37% which we were able to drive up to 82% after training.

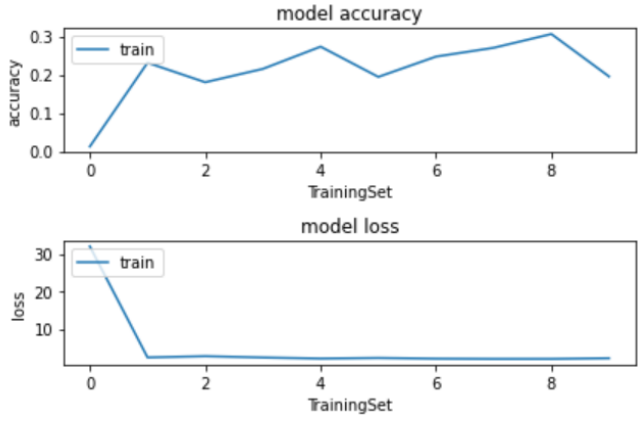

Fig. 2. Accuracy And Loss Graph Of The Age Model

Our age prediction model is designed to predict the age group rather than the specific age of an individual. The range of ages goes from 0 to 130 years in groups of 5 which makes it a total of 26 different age groups (0-4, 5-9, 10-14, 15-19...). We encountered some difficulties in making the accuracy increase. We began at an accuracy of 12% and increased it to 31% after all the training was done. In the graph however, you can observe a dip in accuracy at the very end. This could be either because the last set of data was particularly noisy or a sign of possible overfitting. Because we were saving the model after training it with each set of 40,000 images, we decided to do early stopping and simply use the previous set of weights that were calculated before it was trained on the last group of images.

Figure 3 shows what the network outputs when classifying an image that it has never seen before. It identifies a face or set of faces using the imported facial recognition tools. Here you can see a group of three individuals. Each person's face is cropped off to their own individual picture and each photo is run through the networks. Then the results of the age group and gender prediction is outputted as well as the percentage likelihood of the male and female classifications and the top three guesses for age group.

As an example, Figure 4 shows an individual (actor Tom Hardy) whose gender was predicted correctly as male by the gender identifying network.

Figure 5 shows the same individual (actor Tom Hardy) whose gender was predicted incorrectly as female by the gender identifying net. Any number of things could have led to this wrong decision made by the network such as amount of facial hair, bone structure, the size of certain parts of the face, the angle the image was taken at, the lighting, and so on.

V. Discussion and Conclusion

Our models sometimes misclassified images, most likely due to factors like poor quality, low brightness, an odd camera angle, etc. An image that is grainy or low quality may not capture all the details that the model needs in order to do an accurate prediction. Some of the misclassifications also were due to the incorrect class labels present in the metadata.

The gender classification model performed at 82% accuracy which was within our expectations for it. In some cases, like the one in Figure 5, some factors like lighting and camera angle could make the model classify incorrectly. Other features like facial structure and facial hair most likely played a role in the gender predictions.

The highest accuracy that we were able to reach for the age prediction model was 31% which was an unexpectedly low percentage. However, it is also worth noting that an age is only considered correct if it falls perfectly within the age range group that it is classified in, and everything else is considered incorrect. The accuracy percentage may come off a bit harsher than it should because of this reason. For example, say a 14-year-old gets placed in the 15-19 category by the network. This is a wrong classification made by the network, but let's say the network classified the same 14-year-old in the 75-79 category. That is marked wrong as well, but obviously one

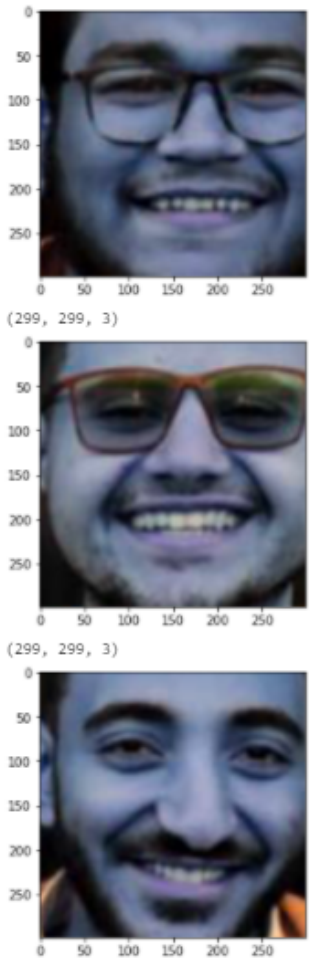

Fig. 3. Some Example Calculations Made By The Models

of those classifications is much closer than the other. The accuracy metric does not take that into account, and marks both incorrect classifications as equally wrong. If the output of an age prediction from our model is examined, the correct age group is usually present within the top three predictions listed. This shows that the model is still producing reasonable results even if the top prediction is off by an age group.

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

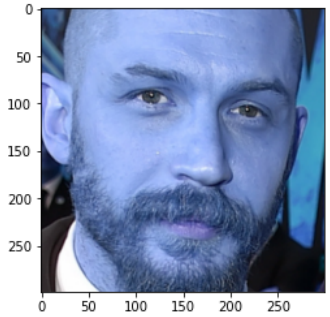

Fig. 4. Correct Classification Made By The Network

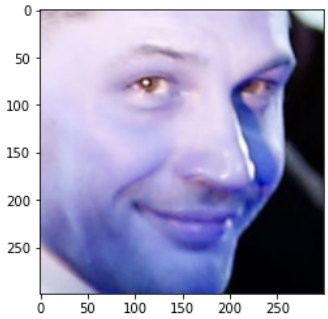

Fig. 5. Incorrect Classification Made By The Network

model from scratch that had some differences from the ones previously created and discussed in the background section. However, after trying a variety of different architectures, we still had difficulties in training these models to reach a reasonable accuracy. We particularly spent a lot of time trying to create a wide convolution network from the ground up to perform the computations, but that ultimately didn't give us the results we wanted. We instead decided to instead add layers on top of a deep 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 had no faces in them since this would add unintended noise and lower the accuracy of the model.

One lesson learned from this project that could be useful for future projects is that transfer learning can be an effective way to build upon previous work with new ideas. Using the deep convolutional architecture of the Xception network as a base model has bolstered the accuracy for our own model when other traditional methods did not. Overall, this makes our project more manageable and reproducible.

In conclusion, the use of convolution nets and transfer learning was successful in creating models that could predict age group and gender from facial images. There have been several other research papers published on this topic with models that were able to achieve a much higher accuracy than

our own. Despite not having results comparable to the ones mentioned above, we reached our goal of making a gender classification model with a reasonably high accuracy while the age prediction model was able to yield acceptable results.

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.

[5] A. Geitgey, "face-recognition," 2020. [Online]. Available: https://pypi.org/project/face-recognition/