Gender Classification via a Convolutional Neural Network

CDT Jack Summers
Advisor: Colonel Thomas Babbitt
Department of Electrical Engineering and Computer Science
United States Military Academy
West Point, New York 10996
Email: jack.summers@westpoint.edu
18 April, 2022

1. Abstract

The project's purpose was to create a state-of-the-art neural network from scratch to demonstrate the capabilities of machine learning (ML). Machine learning is a sub-field of artificial intelligence (AI), and AI attempts to automate intellectual tasks usually performed by humans. The code associated with this paper applies a convolutional neural network (CNN) to an image-classification task to demonstrate how ML can perform intellectual tasks. Specifically, the ML algorithm is a bi-classification model capable of taking a joint photographic experts group file (JPG) of a human face and identifying it as male or female.

2. Data Analysis

The data set used to train the CNN contains just under 58.6K colored JPG images [1]. Each image is a human face and is 92 x 114. The data set has two classes and is approximately half male and half female.

Figure 1. Female and Male JPG

The large data set made it suitable for training the CNN. Additionally, there was a variation in the race, pose, and eyewear. As a result, many photos with a degree of variation provide the CNN with different examples of what a male and female human look like. However, the data set does have limitations. Due to the fact that the images are cropped around the face of each person, the machine learning model produced from training on the current data set may not be able to generalize well to images of people from head to toe.

3. Methodology

Building and training the CNN with a GPU required the use of a Jupiter notebook. The first step involved data processing with training generators. The purpose of the training generators was to connect the directories to the model [2]. The generators automatically take 128 JPG images at a time from the directories containing the data set. The generators also split the data at random into two sets, 80 percent training data and 20 percent validation data. Additionally, the generators resized and labeled each image using the directory's name. Meaning, if an image originated from the "male" directory, then the image was labeled as "male". The generators also provided the program with a secondary purpose. The generators utilized data augmentation to artificially expand the size of the data set. Rotating, shifting, shearing, zooming, and flipping images at random provides the model with images it has already seen, but from a different perspective. Hence, providing the ML model with more examples of what a male or female face looks like. The larger the data set, the more powerful the model. Finally, the generators normalized the data by dividing each pixel value by 255. Each pixel value is between 0 and 255, and dividing by 255 makes each value between 0 and 1. In turn, all of the values are within the same range, and will effect the models training with equal weight.

The second step in creating a state-of-the-art ML model required defining callbacks. An EarlyStopping callback helps to prevent over-fitting during the training process by terminating training if the validation accuracy does not increase by at least 0.001 after 10 epochs.

After creating the callback function, the model was defined. The model contains two Conv2D layers each with a relu activation function and input shape of (92, 114, 3). Relu is the standard function for CNNs [2], and the model accepts an input vector of (92,114,3) because the database images are 92 x 114 with 3 pixel values (red, blue, green). After each conv2D layer is a Maxpooling layer, their purpose is to down sample [2]. Down sampling reduces the number of feature maps to be processed. Finally, the feature maps are fed into a Flatten layer which create a one dimensional vector for the dense layers. The dense layers and their corresponding neurons are what classify each image. The

final dense layer has one neuron with a sigmoid activation function. The final dense layer only has one neuron because the model only produces one value per image. The final value is the probability that the image is female. In turn, a value of 0.90 would mean that an image has a 90 percent probability of being the female class. A sigmoid activation function is optimal for classifying between two classes [2]. The final steps involved compiling and fitting the model.

4. Results

An initial setting of 1000 epochs provided the model with a generous amount of time to optimize. After 42 epochs the gender classification CNN stopped training and achieved a training accuracy of 0.9515 with a loss of 0.1319. The model also achieved a validation accuracy of 0.9458 with a loss of 0.1415. The tight correlation between training and validation accuracy shows that the model did not memorize the training data, hence not over-fitting [2].

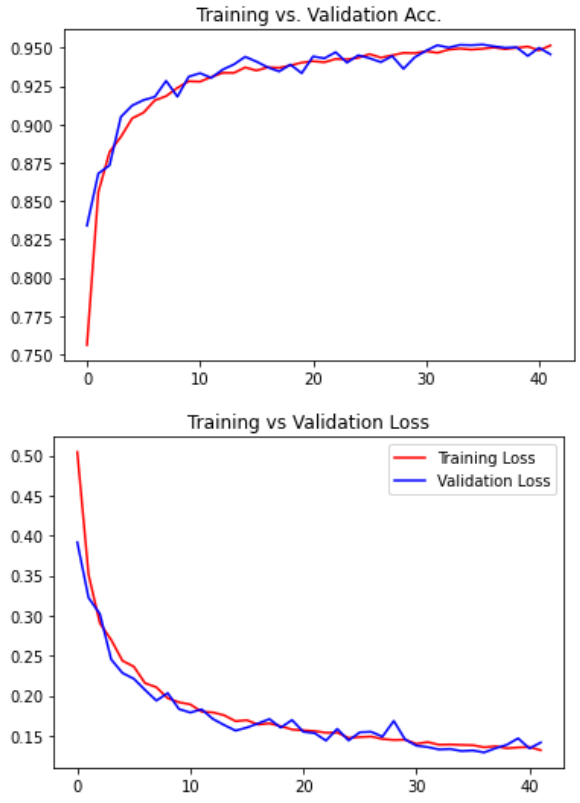

Figure 2. Results

Finally, 5,841 female images and 5,808 male images were set aside for testing purposes. The gender classification ML model did not train with any of the images. The final model was evaluated with the testing data, and the model achieved a state-of-the-art accuracy of 0.9609 with a loss of 0.105.

Conclusion

The gender classification model can serve practical purposes. For example, the model could help profile pedestrians via CCTV security cameras with an ensemble of other ML models. If security personnel were notified a male in a black shirt stole a purse, officers could query the CCTV system for a suspect with ”FIND ALL males AND blackshirts”. A second practical purpose for the gender classification model would be for E-commerce. For example, the model could classify if an unknown user is male or female based on their profile image. In turn, a program could target that individual with advertisements specific to a male or female. However the main purpose of this project was to demonstrate the capabilities of ML algorithms. A powerful model can be made with a small data set, an hour of training, and a small number of neurons. With more data, AI can become extremely powerful. For instance, China has developed a program called ”Skynet” which can identify the face of one of China’s 1.4 Billion citizens within a second [3]. Additionally, Ukrainians are utilizing Clearview AI to identify Russian soldiers, living or dead [4]. In other words the power of ML is revolutionizing all aspects of society from shopping, to war and espionage.

References

[1] A. Chauhan, “Gender classification dataset,” Kaggle.com, https://www.kaggle.com/datasets/cashutosh/gender-classification-dataset (accessed Mar. 7, 2022).

[2] F. Chollet, *Deep Learning with Python*. Manning, 2018.

[3] K. Strittmatter, *We Have Been Harmonized*. HarperCollins, 2018.

[4] K. Hill, “Facial recognition goes to war,” nytimes.com/, https://www.nytimes.com/2022/04/07/technology/facial-recognition-ukraine-clearview.html (accessed April. 8, 2022).