**Gender Recognition using Hand Images**

Team Members

1. Achyuth Sankineni
2. Paul Marinan
3. Sourav Sethia
4. Tianhao Zhao

# Executive Summary

In our project we decided to gather hand images with the intent to classify the gender of the person to be either male or female. At first we had more ideas on what to classify based on these images, such as race and age, but based on the distribution and other characteristics of our dataset we decided to just go with gender. We will go more in depth about our dataset later but just to give a background on the dataset, we will describe it briefly. We gathered images of hands both palmar and dorsal side up and with open and closed fingers from both the right and left hands. Some hands had accessories or nail polish and there were varying ages and race. Along with these images we obtained the corresponding gender, race, accessories, nail polish, etc., for each image but for the purposes of our project and models all we needed was the gender information. We developed and ran convolutional neural networks on this dataset with the hope that some of the following questions could be answered:

1. Can the gender of someone be determined by a picture of their hand?
2. What features and characteristics of the hands is the model identifying?
3. Do the other characteristics such as race, age, and accessories help or hurt the models ability to classify gender?

We were obviously hoping that our model could accurately predict the gender of a person with an image of their hand and were very curious how the model would do so. Would it identify fingernail length and structure? Would it identify bone structure? Are palm lines different for males and females? Would the accessories and nail polish be the only way the model could make a difference? Since we were only using gender to validate our results some of the answers to these questions would have to come from our evaluation metrics. For instance, if nail polish was the only way the model could make a difference between a male or a female then our model would have very low accuracy. We set forth with two goals, try to classify gender based on hand images and visualize what the model was learning.

# Introduction

In order to build a model for biometric identification we decided to choose the dataset of hand images and use it to classify the gender. The classification of gender is done so as to correctly identify the person and thereafter different features can be used to correctly identify a person. The dataset was gathered from google and we divided the dataset into a training set, validation set and test set. The training set has 6645 images, the validation set has 2215 images while the testing set has 2216 images respectively. The dataset has a mix of dorsal and palmer images of both male and female. Please see [Exhibit 8](#u6agqq6rep8) to see the statistics of the different features of the dataset. We have built convolutional neural networks using a combination of python libraries like tensorflow, opencv, pandas, numpy. We used a different combination of Conv2d and max pooling layers to extract most important features and downsample the image respectively. The two activation functions that we used in our model are relu and sigmoid. To make our model learn better and reduce loss we used optimizers like Adam and RMSprop. The model is run for 500 epochs with a batch size of 25.

# 

# 

# 

# Data Description

Our dataset contained 11,076 hand images with the corresponding descriptive data. Each image had a blank white background, which is important since we did not want the background to play a role in our models. The images were a mix of male and female, different skin tones, right and left hands, and palmar and dorsal side up. Some hands also had accessories on them and some had nail polish as well. See [exhibit\_2](#6p2cifcsl5rh) for image examples. Based off of this dataset, we decided to just classify gender and leave out age and race. The distribution of ages for the hand images was very skewed and we believed it would prohibit the model to predict the age. See [exhibit\_1](#bdreoukkb9ks) for the distribution. We were also fairly certain that the model would be able to identify race based on the hand images and in order to spend more time on classifying gender we decided to leave that out as well. The only data that we needed to go along with the hand images was the gender for each hand. See [exhibit\_3](#9r00dvailwqd) for a dataset example. Once we gathered all the images and the corresponding data we split the data into three sets: training, validation, and test. There were 6645 images and data in our training set, 2215 images and data in our validation set, and 2216 images and data in our test set, or about 60%, 20%, 20% of images and data respectively. We realized that there was a problem with our dataset split so we had to resample our data. When we split it the first time our male and female proportion was way off causing our model to obtain worse results. After we fixed that and our datasets were on par with each other we were ready to run our models.

# Model Description

After a lot of tinkering with our model architecture we arrived with our best model. First we had to figure out how many convolutional and pooling layers were ideal and then how many dense layers were ideal. Along with the number of layers we had to find the right filter sizes for our convolutional layers and then how big our dense layers would be. While we were messing with the architecture we changed the dropout values for our layers to see how that would affect our model results. It seemed that the model performed better with a lower dropout. We eventually obtained a model with great results and we could not improve those results (there wasn’t much room to improve) so that was our final model. Our final model has four convolutional and pooling layers with filter sizes of 32, 64, 128, and 256 respectively with a kernel size of 2 for all and our pooling size remained steady at 2. We did change our kernel size to 3 to see if that would make a difference but it made little difference in the wrong direction. Our final model had 3 dense layers, the last one being the output node. The first dense layer has a size of 256 and the second layer has a size of 64. Both of these layers have a dropout value of 10%, again, we discovered that the less dropout we had the better the model performed. This model generated a ton of trainable parameters, about 5.5 million, which is what we wanted. Our four convolutional layers had 416, 8256, 32,896, and 131,328 trainable parameters with our pooling layers obviously having none. Of course, we then flattened them out and passed them into our dense layers which had 5,308,672, 16,448, and 65 trainable parameters respectively. Please see [exhibit\_4](#8nrnlqemmnjt) for a summary. This model was able to predict gender very accurately and it was very interesting to see what the model was identifying and learning from the hand images which we will discuss further later on.

# Results

After we determined our best model, we evaluated the results on our validation and test sets. At first, we obtained about 90% and 80% accuracy respectively. Although these results weren’t bad we realized that we could improve upon our results by fixing and resampling our dataset as the number of males and females in our training, validation, and test sets differed by a lot. After we corrected this and reorganized our datasets we were able to obtain about 99% accuracy for both our validation and test sets. We were also able to obtain about 99% precision and recall for both sets as well. Please see [exhibit\_5](#4v4yq3ibk6e5) for classification reports. These are excellent results and it seems that our model can very accurately classify gender based on hand images. We ran our model with 500 epochs and a batch size of 25 but we also included early stopping with a patience of 15 in order to avoid overfitting. By doing this we realized that the model should be run for 60 epochs in order to obtain the best results. Please see [exhibit\_6](#zatpgfz84sgt) for the training and validation loss curves. Now that we answered our first question on whether or not the model could classify gender based on hand images we then moved on to what our model was learning and identifying. Going into this we thought that the model would identify things such as nail length, nail polish, palm lines, and hand/finger size in order to determine gender. Once we were finished we realized that the model focused heavily on the fingers and more specifically the edges and outline of them. This leads us to believe that the structure, length, and size of the fingers is a big determinant of gender. Contrary to our beliefs beforehand, the palms, accessories, and nails played a very small role, if one at all, in the model’s ability to classify gender. This also told us that it did not matter if the hand was palmar or dorsal side up which is good since we could still classify gender no matter what side of the hand is showing. Please see [exhibit\_7](#aaqhkwnf5vez) for some images on the features the model was learning. These results make a lot of sense considering what our accuracy was for our model. If the model relied heavily on things such as accessories and nail polish then its ability to classify gender would not be good and hands without those attributes would be hard to classify causing our accuracy to be low. Since the model used the structure of the fingers to classify, all the images have hands with fingers, that would lead us to believe that if fingers tell us something about gender the model will be able to classify with high accuracy and it did. Our results were very satisfactory considering what we were hoping for.

# Discussion

Although our model performed great we have thought about ways to make it better. Although our images had other characteristics such as nail polish and abnormalities we could add or increase some of these features in order to test our model further. What if some hands were missing fingers or fingertips? Could our model still classify with such high accuracy? We realized that we could add more layers so that the model learns smaller more specific attributes. That being said we would like to know what really gives the model trouble. Based off of our results if the images were missing fingertips then our model would perform poorly. Knowing what would “break” our model would help us learn the limitations of it. Also, going forward, we could try to classify more than just gender such as age and race. We would have needed a more normally distributed dataset to classify age and a more robust dataset to classify race, depending on how specific we wanted to get. Classifying things such as these would take a lot of data gathering but could be very interesting to see. What if we could classify gender, race (down to the country they originate from), and age of a person based on an image of their hand? That would really be quite something. “This person is a 32 year old female originating from Russia,” just based off of an image of their hand (certainty depending on how well the model performs). Not only could this have an impact on many different industries such as genetics or even law enforcement, but we could learn a lot about people based on their hands. For this to work on our existing model then there would have to be a distinction in finger structure for people of different ages and race. Another thing to think about is how specific we could classify based on hand images. Think about facial recognition and how we talked about determining whether or not a specific person was in a photograph. Could we do the same with their hands? If we had enough pictures of Dave Wanik’s hands could we train a model to determine whether or not a hand image is his hand or not? In order for something like this to work, hands would have to be different enough for every person for the model to make a difference. There might not be enough features in the hand for something like this but it could be interesting. As of right now, we do know that we can classify gender based on hand images, it would be interesting to see what else.

# Conclusions

On evaluating the different combinations of metrics and finding the best training model, We tested it on validation and test dataset and found the accuracy to be 98.5 % and 98.1 % respectively. We evaluated the channels on all the 8 layers that we used in the model and our findings were as below

1. The first layer acts as a collection of all information of hand. There are lots of purple hands and fewer yellow in the specific area. At that stage, the activations are still retaining almost all of the information present in the initial layer.
2. As the layers go deeper, there are more and more yellow heatmaps in the hand's specific area and the image becomes abstract, which means it contains less information and more information related to the class of image.
3. On evaluating the last layer we can see the majority of yellow area is located on the five fingers of the hand, and especially located on the top of each finger. However, we only can learn from the last layer, there are still lots of activations and filters, and still a large area is yellow, which means we have lots of information. So I think we can put more layers to allow the model to learn until it learns only a small specific area.

# Reference

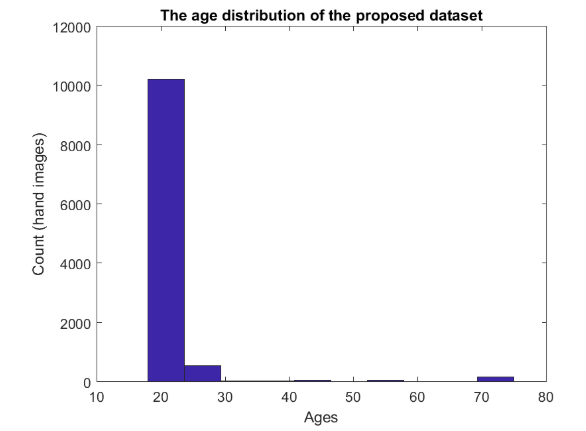
1. [**https://sites.google.com/view/11khands**](https://sites.google.com/view/11khands)**.**
2. **Grokking Deep Learning by Andrew W Trask.**
3. **Deep Learning with Python by Francois Chollet.**

# 

# 

# Appendix

Exhibit\_1



Exhibit\_2



Exhibit\_3



Exhibit\_4

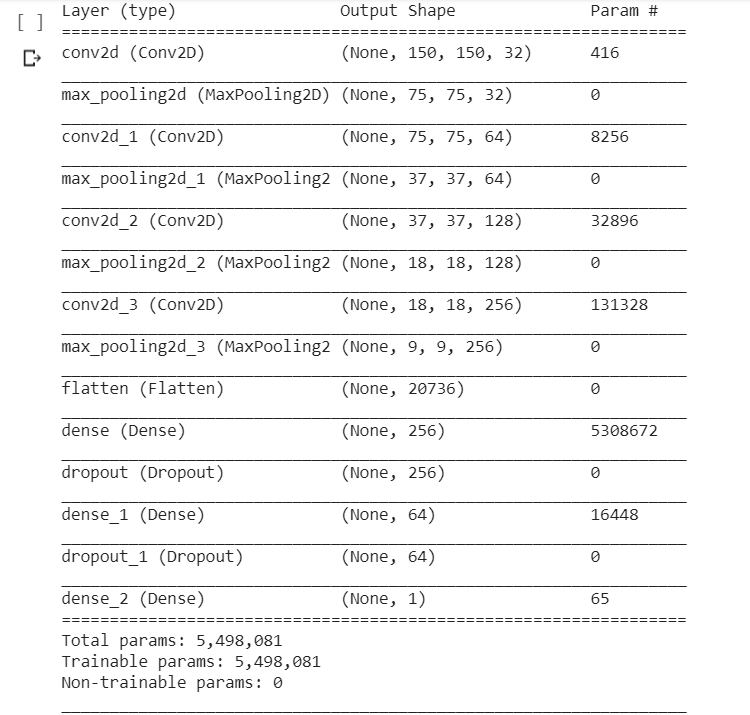
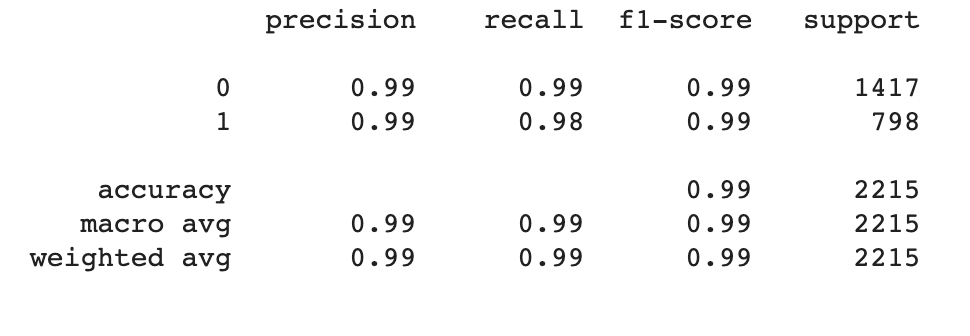


Exhibit 5

Classification Report on Validation Set and Test Set



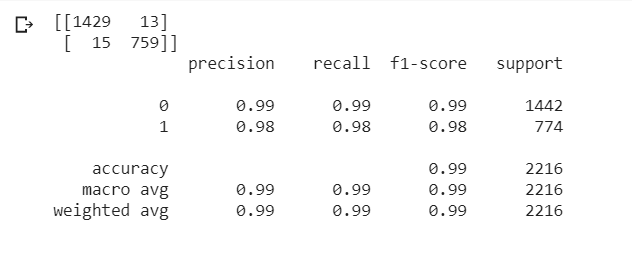


Exhibit 6

Loss Curves

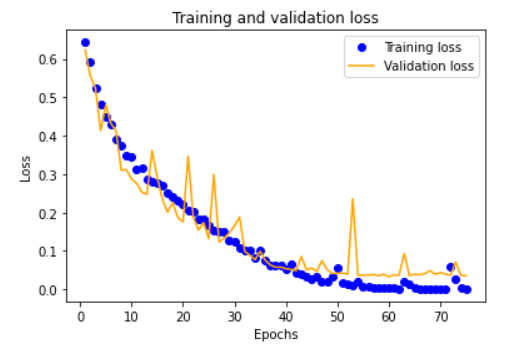


Exhibit 7

Convolution and Pooling

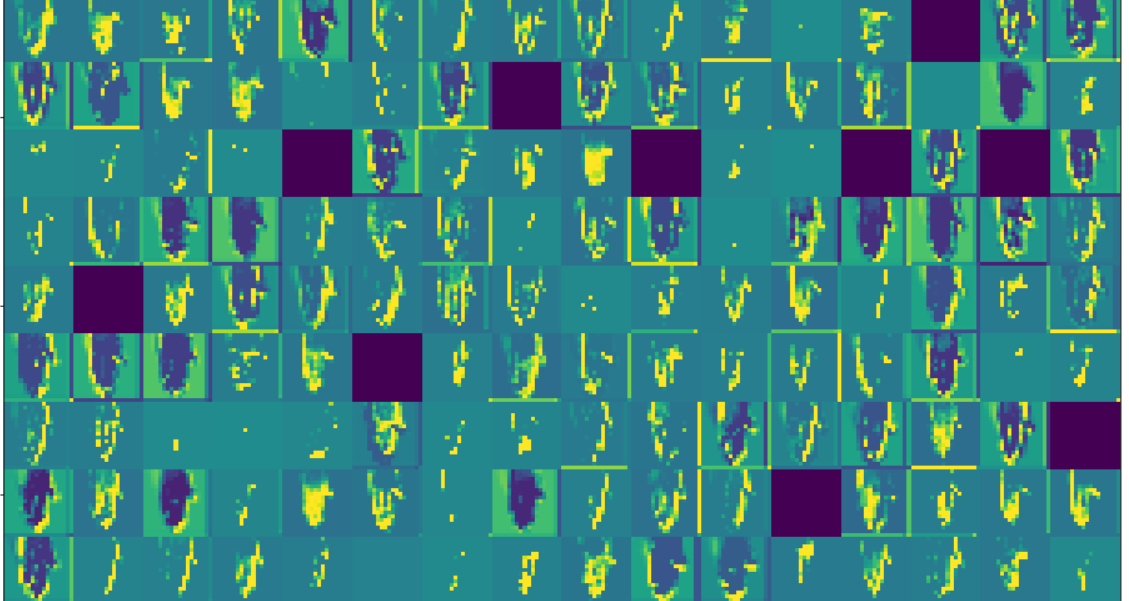


Exhibit 8

