MAIS 202 - Project Deliverable 2

by Fardin Abdullah, Nicholas Foisy, Japmann Sarin

5 November 2021

1. The goal of this model is to use a CNN to classify drawings of 19 major constellations. The list of such constellations is in Appendix A. The user will be able to draw using black lines on a white background images of constellations (See Figure 3 for an example) on a web-app and a prediction of what is drawn will be returned to the user.

2. As images will be drawn using a controlled Canvas element on a webpage, the model will only be trained with the same black-and-white line drawings. As no dataset currently exists with data of this nature, this data has been created by ourselves.

   The creation of this data has been facilitated by creating and using a Canvas drawing box on a personal website at the following URL: https://foisy.co/, with the page layout shown in Figure 4. The website has been designed so that upon selecting a constellation to draw from the dropdown menu, an image of the constellation selected is displayed, as reference for the user. Then, using the Canvas element, the user can draw an image. When the drawing is complete, the user can click on the 'Download image' to save the drawing locally to their disk, with the file name configured to be the constellation name and a random large integer. This naming convention is used to avoid having files with the same filename.

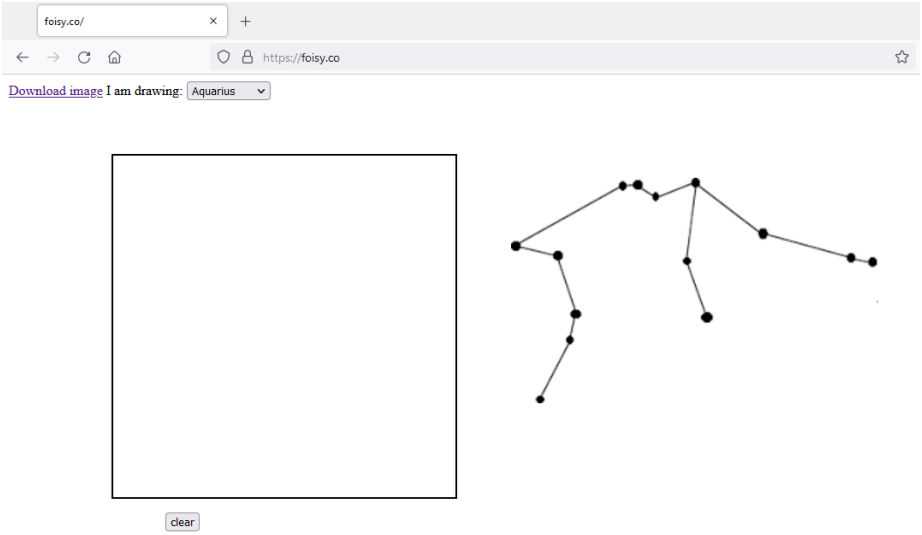


Figure 1: Landing page of the website used to manually draw images for the dataset.

Due to the time-consuming nature of drawing this data, we have decided to limit ourselves to 9 classes for the scope of this deliverable. For future project deliverables, the

full 19 constellations will be supported by the model. In total, there are 873 samples of data. See the chart in Figure 2 for a breakdown of the number of samples per class for the training dataset. The discrepancy in the number of samples, especially for the Cancer constellation, is due to having received a significant number of samples for few classes from one eager contributor. However, this minimally unbalanced dataset is not an issue as the model does not use the frequency of data per class in the dataset as a factor in predictions.

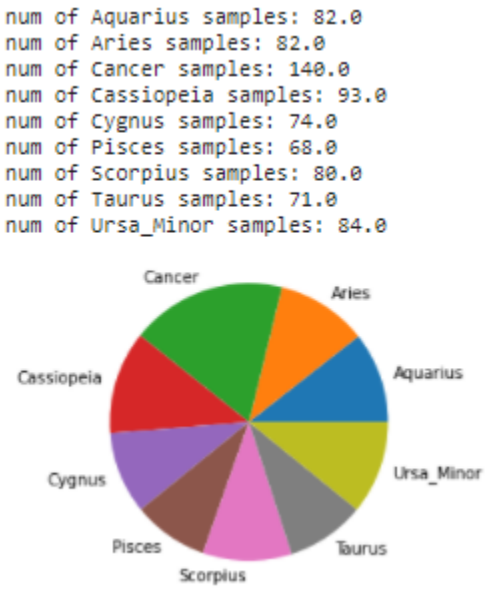


Figure 2: Distribution of of training samples of data for 9 classes of constellations

There are also 99 additional images not shown in the figure above, used for testing purposes.

As data will be augmented using the Keras package in Python, we have drawn a majority of images without substantial variances in the orientation of the drawings. Using the aforementioned package, the data can be flipped and rotated, making up for this lack of orientational diversity in the training data. However, the training data has great variability in terms of the lengths and angles of the lines consisting of the constellations. See Figure 3 below for an example of the drawn data. Nonetheless, these features will be further augmented using shearing and stretching.

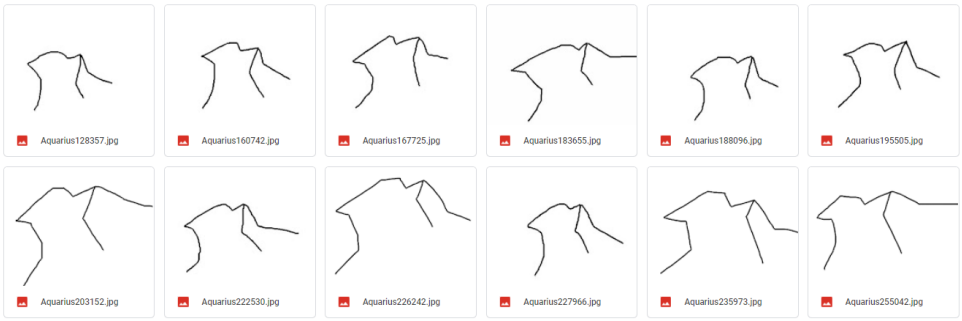


Figure 3: A subset of the data manually drawn for the Aquarius constellation class

The data is currently labelled with numbers ranging from 0 to 8, with 0 representing the Aquarius class, and 8 representing the Ursa_Minor class. When the project is complete, labels will range from 0 to 18 instead, representing each constellation listed in Appendix A.

3. a) The model used is a CNN with only one dense layer and 'softmax' is used as a parameter to normalise the probabilities of each class. Two convolutional layers are used and two pooling layers are used in the model. The images are resized into much smaller dimensions and converted into a numpy array containing the RGB value of each pixel / 255. 50 epochs are used in training the model. Increasing the pixel dimension increased the accuracy of the model at the expense of training time.

b) To evaluate the model, sklearn's accuracy score is used. The model is slightly overfitting because the training accuracies are higher than that of the test accuracy. The test accuracy received is 80%. The model can be modified to under or overfit by changing the number of epochs.

c) A difficulty that was encountered was related to the size of the input images. The input images were originally scaled down to 32 pixels by 32 pixels from 400 pixels by 400 pixels to decrease training time. However, scaling down the images to this small size reduced the amount of uniquely identifiable details in the image, and the model could no longer differentiate reliably between the Cancer and Taurus constellations, which both appear as Y-shapes with only subtle differences, which are largely lost at the size of 32 by 32 pixels. Therefore, the image size was increased to 128 pixels by 128 pixels. This increased the amount of detail in the images for the model to analyze, and predictions improved, with images of Taurus and Cancer being differentiable.

4. The preliminary results of the model were decent, given the small testing set and 90 images were tested and the outputs had an accuracy of 80% as measured by the sklearn's accuracy score. Furthermore, we have also produced a confusion matrix, which gives an insight about which classes were wrongly classified and how often that happened. The confusion matrix is given below:

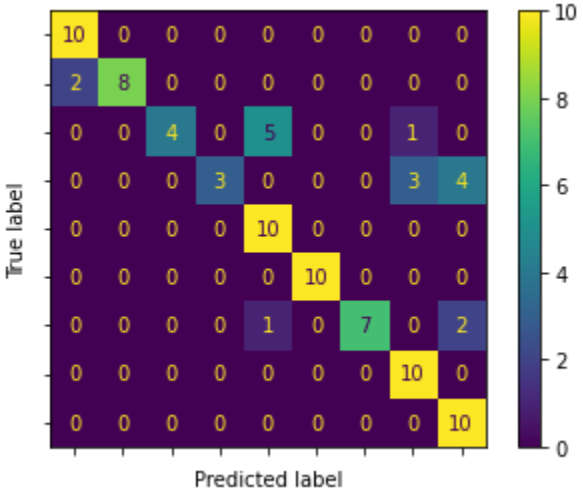


Figure 4: Confusion matrix of model

5. Currently, the model is working very well and no major change appears to be necessary, though hyperparameters will be fine-tuned as the model is tested more.

Thus, the next step of this project is to obtain images for the remaining 10 constellations to be recognized by this model. We will also continue to collect more data for the other 9 constellations currently supported by the model. The model will continue to be tested by processing images with the drawings in further various sizes and angles.

After this, the webapp will be improved by making the layout more visually appealing, and ensuring compatibility with mobile platforms. Finally, the CNN model will be implemented into the website and pictures drawn on the canvas will be fed into it, and a prediction will be returned to the user.

Works Cited

[1]https://www.artofmanliness.com/skills/manly-know-how/15-constellations-every-man-should-know/

[2]https://github.com/codebasics/deep-learning-keras-tf-tutorial/blob/master/16_cnn_cifar10_small_image_classification/cnn_cifar10_dataset.ipynb (tutorial used to create model)

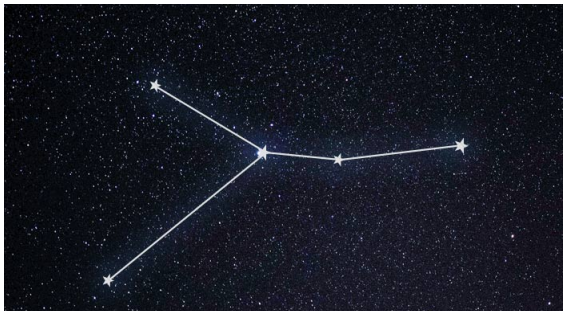
Appendix A

Constellations

1. Aquarius [1]

2. Aries [1]

3. Cancer (https://star-name-registry.com/constellations/cancer)



4. Canis Major [1]

5. Capricornus (https://en.wikipedia.org/wiki/Capricornus)

6. Cassiopeia [1]

7. Cygnus [1]

8. Gemini [1]

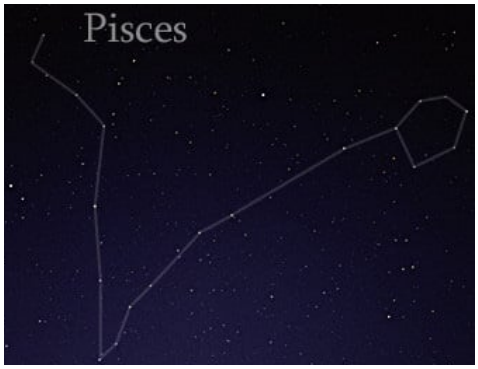
9. Leo [1]

10. Libra (https://en.wikipedia.org/wiki/Libra_(constellation))

11. Lyra [1]

12. Orion (https://www.planetguide.net/orion-constellation/)

13. Pisces [1]

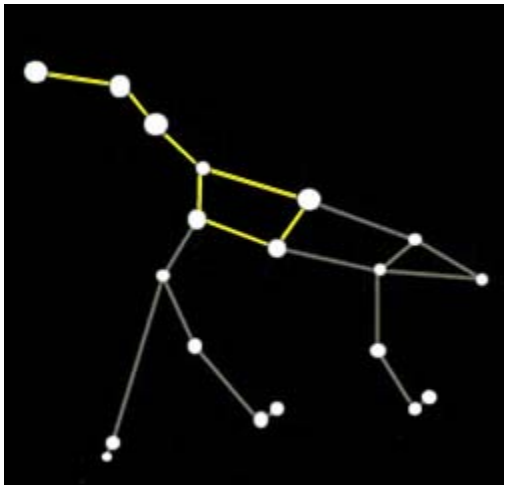


14. Sagittarius [1]

15. Scorpius (https://star-name-registry.com/constellations/scorpius)

16. Taurus [1]

17. Ursa Major (https://www.solarsystemquick.com/universe/ursa-major-constellation)



18. Ursa Minor [1]

19. Virgo (https://en.wikipedia.org/wiki/Virgo_(constellation))