Classifying Galaxies with Machine Learning

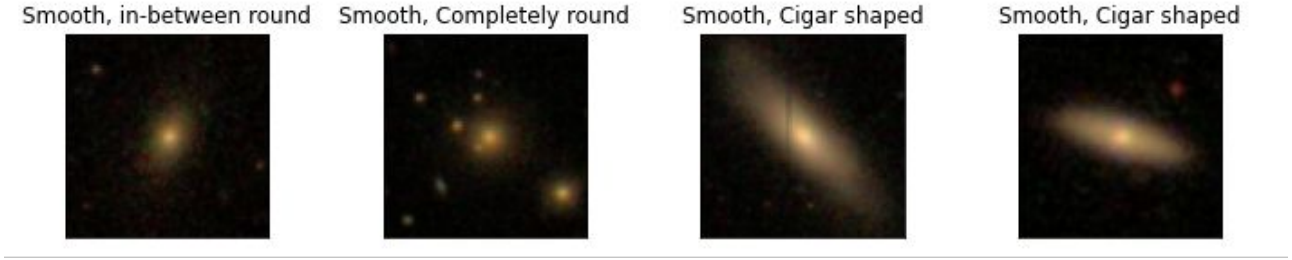
Kaylie Rose Mendoza

Introduction

- How accurately could my model classify galaxy images in comparison to classification by hand
- Galaxy characteristics differ by galaxy shape
  - Elliptical (smooth) galaxies
    - Older
    - Dimmer
    - Redder
    - Little gas and dust
  - Spiral (disk) galaxies
    - Younger
    - Brighter
    - Bluer
    - More gas and dust- lots of star formation

Galaxy 10 Dataset

- ~ 18,000 galaxy images from the Sloan Digital Sky Survey (SDSS)- 2016
- Categorized by hand into 10 different categories
  - Elliptical and spiral galaxies, but no irregular
- Data available in astroNN
  - Python package meant for building neural networks with astronomical data
- Classified by
  - Type (smooth, disk)
  - Shape (spiral/bulge type, round/cigar)
  - Orientation (disk galaxies: edge/face on)



Methods

- Split the data into 80% training and 20% testing (Scipy)
- Set up the neural network using Keras and Tensorflow
  - Sequential model
  - Layers
    - 4 convolutional (conv2D) layers (up to 256 3x3 filters)
    - 4 pooling (MaxPooling2D) layers
    - Flattening layer
    - 2 dense layers
- Compiling and training
  - Adams optimizer
    - Adaptive- the learning rate changes over time
  - EarlyStopping
    - Allows the model to run as many epochs as possible (until overfitting occurs)

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d (Conv2D)             (None, 69, 69, 32)        896

 max_pooling2d (MaxPooling2D  (None, 34, 34, 32)       0
 )

 conv2d_1 (Conv2D)           (None, 32, 32, 64)        18496

 max_pooling2d_1 (MaxPooling  (None, 16, 16, 64)       0
 2D)

 conv2d_2 (Conv2D)           (None, 14, 14, 128)       73856

 max_pooling2d_2 (MaxPooling  (None, 7, 7, 128)        0
 2D)

 conv2d_3 (Conv2D)           (None, 5, 5, 256)         295168

 max_pooling2d_3 (MaxPooling  (None, 2, 2, 256)        0
 2D)

 flatten (Flatten)           (None, 1024)              0

 dense (Dense)               (None, 256)               262400

 dense_1 (Dense)             (None, 10)                2570

=================================================================
Total params: 653,386
Trainable params: 653,386
Non-trainable params: 0
_________________________________________________________________
```

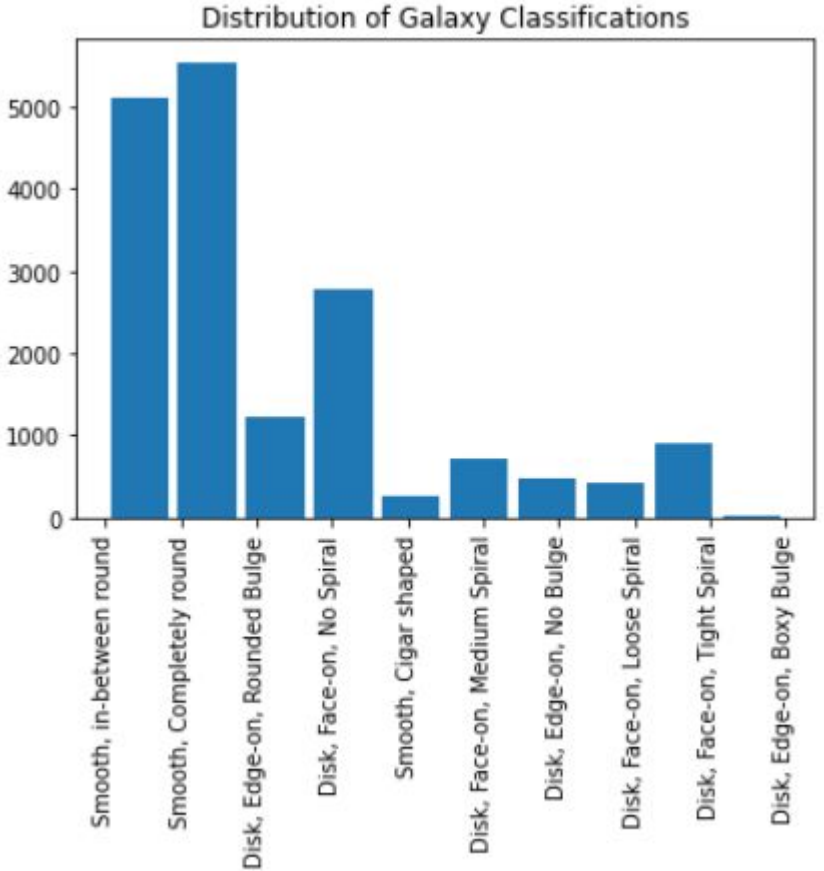
Results

- The model ran for 10 epochs, with 78% accuracy overall when used on the testing images set
- Very mixed results in category accuracy
  - The second category contained the most images, and had an accuracy of 93%
  - The sixth category had the least amount of images, and had an accuracy of 0%

Disk, Face-on, No Spiral:	46% (322/696)
Smooth, Completely round:	93% (1252/1335)
Smooth, in-between round:	87% (1128/1284)
Smooth, Cigar shaped:	47% (36/76)
Disk, Edge-on, Rounded Bulge:	89% (300/337)
Disk, Edge-on, Boxy Bulge:	0% (0/ 3)
Disk, Edge-on, No Bulge:	75% (90/119)
Disk, Face-on, Tight Spiral:	63% (140/221)
Disk, Face-on, Medium Spiral:	50% (91/182)
Disk, Face-on, Loose Spiral:	47% (49/104)

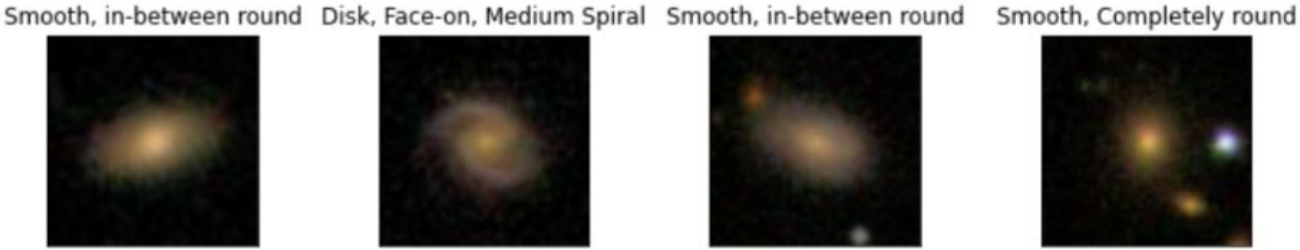
Results

- Overall, categories with more images had more accuracy
  - Smooth, Completely round: 93%
  - Smooth, in-between round: 87%
- Categories with less images had lower accuracy
  - Disk, Edge-on, Boxy Bulge: 0%
    - There were only 17 images in this category in the dataset
  - Smooth, Cigar shaped: 47%



Conclusion

- A more complicated model could create better results
- Could be improved by using a better dataset
  - Fewer, broader categories with more unique parameters
  - Very uneven number of images per category
    - The smallest category contained 17 images, while the largest category had over 5,000
    - Much more images of elliptical than spiral galaxies
      - Overall, the model was less accurate when classifying spiral galaxies
        - Are spiral galaxies harder to classify or were there not enough images?
  - Clearer images



References

- Chen, B. “Early Stopping in Practice: an example with Keras and TensorFlow 2.0: A step to step tutorial to add and customize Early Stopping”. Towards Data Science. 28 July 2020. https://towardsdatascience.com/a-practical-introduction-to-early-stopping-in-machine-learning-550ac88bc8fd
- Howell, Elizabeth. “What Are Elliptical Galaxies?”. Future US, Inc. 9 January 2019. https://www.space.com/22395-elliptical-galaxies.html
- Leung,Henry. Bovy, Jo. “Welcome to astroNN’s documentation!”. 2017-2022 Henry Leung. 22 April 2022. https://astronn.readthedocs.io/en/latest/
- Rosebrock, Adrian. “Keras Conv2D and Convolutional Layers”. 2022 PyImageSearch. 31 December 2018. https://pyimagesearch.com/2018/12/31/keras-conv2d-and-convolutional-layers/#:~:text=The%20final%20Conv2D%20layer%20learns%20128%20filters.