# Are Anime Cartoons?

## Jia Xun Chai, Haritha Ramesh, Justin Yeo

{jiaxun, haritha, yzxj }@stanford.edu

**References**
- Blessing, A., & Wen, K. (n.d.). Using Machine Learning for Identification of Art Paintings.
- Chollet, F. (2016, June 5). Building powerful image classification models using very little data. Retrieved December 13, 2016, from https://blog.keras.io/building-powerful-image-classification-models-using-very-little-data.html
- Elgammal, A. (2016). Creating Computer Vision and Machine Learning Algorithms That Can Analyze Works of Art. Retrieved December 13, 2016, from https://www.mathworks.com/company/newsletters/articles/creating-computer-vision-and-machine-learning-algorithms-that-can-analyze-works-of-art.html?requestedDomain=www.mathworks.com
- Gatys, L. A., Ecker, A. S., & Bethge, M. (n.d.). Image Style Transfer Using Convolutional Neural Networks.
- Karayev, S., Trentacoste, M., & Han, H. et al. (2014). Recognizing Image Style. Recognizing Image Style. Retrieved December 13, 2016, from https://sergeykarayev.com/files/1311.3715v3.pdf.
- Redi, M., & Merialdo, B. (n.d.). Enhancing Semantic Features with Compositional Analysis for Scene Recognition.
- Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (n.d.). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. Journal of Machine Learning Research.

## Motivation

For this project, we wanted to explore this characteristic of neural networks, and do so in the fun context of pop culture references. Pop culture has many debates, one of which is "Are Anime considered Cartoons?" If a machine could accurately classify images into anime or cartoon, it would mean that there are sufficient features that identify each category of animation - and we would be providing the internet with some proof that anime and cartoons are different. On the other hand, if it fails, we conclude nothing because another method may succeed (and the debate carries on).

## Data

Our data includes coloured screencaps of anime and cartoon shows and they were collected through batch image downloads of google image search results. They are then scaled to 120 by 80 pixels for consistency.

We have about 1800 samples for each of the classes and the split is a 3:1 ratio for training and test sets respectively. Because of the limited data, we performed some augmentation on the set to decrease bias (zooming, flipping, and shearing).

## Models

We tried two models to tackle this problem.

**Multi Layered Perceptron (MLP)**
The first one was a simple MLP neural net which took the raw pixel values as the features. After a few attempts (e.g. grayscale), we realized that this method wouldn't work, and moved on to CNN.
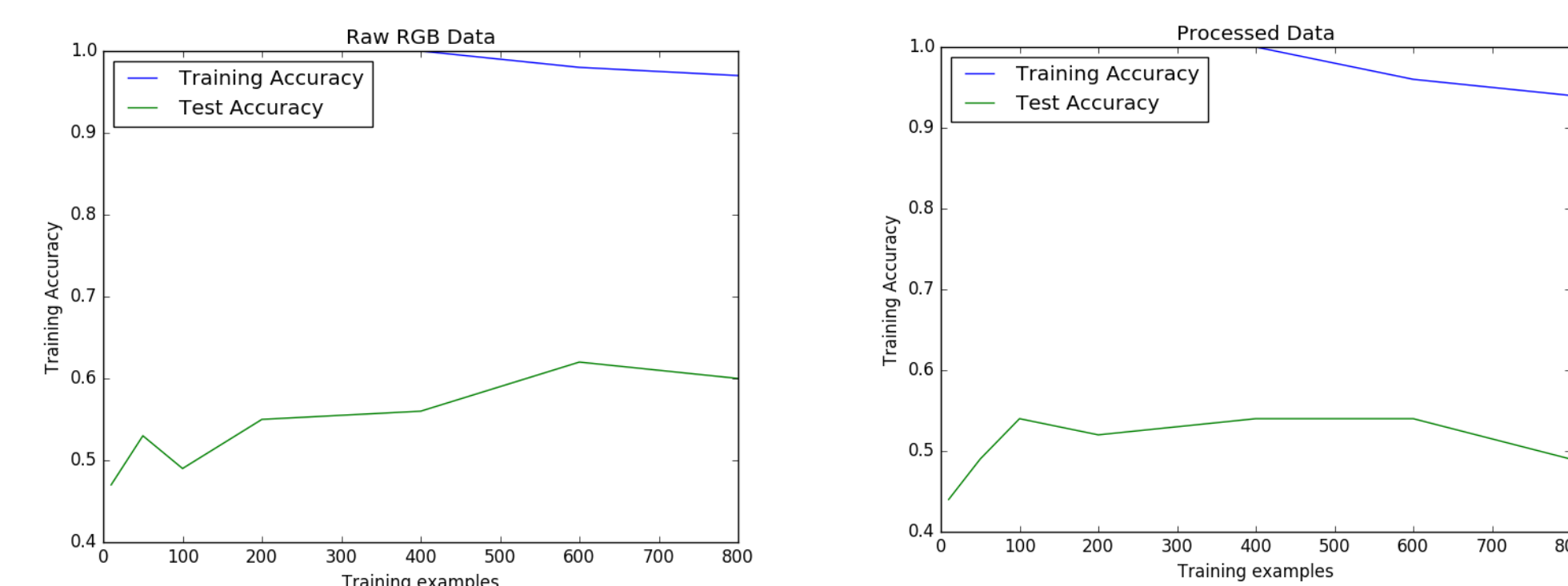
**Convolutional Neural Network (CNN)**
The second one was a CNN with three layers, each with a 3 by 3 filter and a 2 by 2 pool size. The first and second layers had 32 features each and the last layer had 64 features.
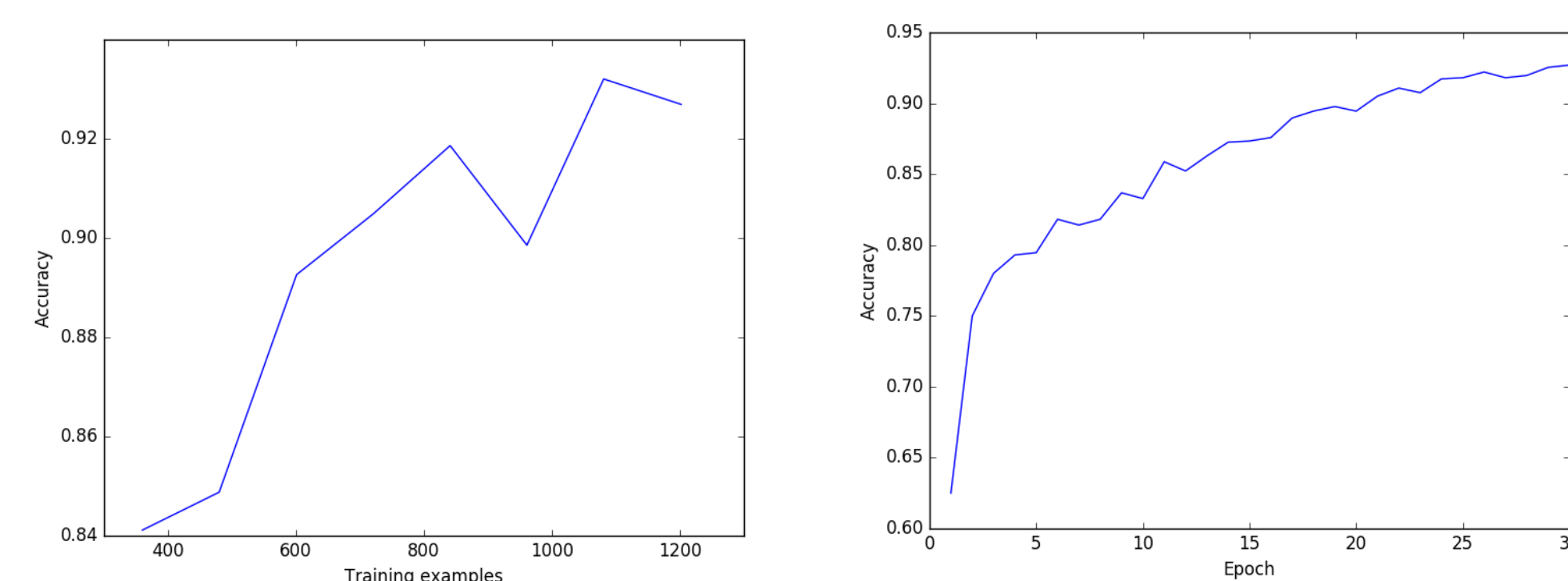
## Results

**Multi Layered Perceptron (MLP)**
This was terrible at predicting the class and had a 50% to 60% accuracy rate, not much more than a random guess. This result held across different permutations of layers and nodes per layer.

**Convolutional Neural Network (CNN)**
The CNN gave much more promising results. With just about 1000 training examples, it gave us a good 90% accuracy on average.

After the initial results, we attempted changing the hyperparameters to see if we could improve the model. The results are shown below.

| Filter Size | Pool Size | Accuracy |
|---|---|---|
| 3 | 2 | 91.74 |
| 5 | 2 | 88.87 |
| 7 | 2 | 82.46 |
| 9 | 2 | 81.80 |
| 3 | 3 | 92.10 |
| 5 | 3 | 92.33 |

## Discussion

Results from CNN were significantly better, for reasons now obvious to us. To begin with, MLP was comparing pixels, rather than features with spatial awareness. The CNN filters abstracted features that can exist across the images. The feature set of our first convolution layer (of filter size 3) is shown below:
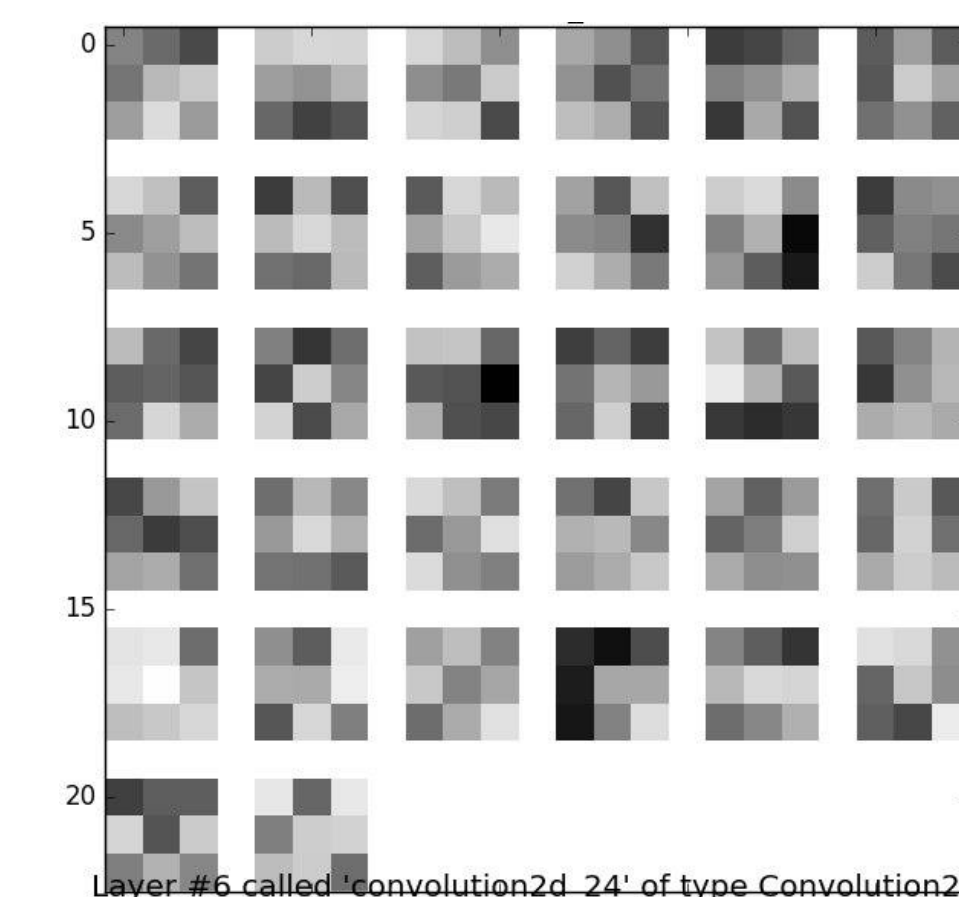
When this feature set is applied (convolved) to an image, we got the following images (of just 4 features, in grayscale for visualization). We observed that the lighter areas trigger the 'activation' layer of the CNN, while the darker areas were muted.

The results we gained in the first round of testing with CNN was around 92%. However, following that, our attempts to modify the hyperparameters and increase training set size had no significant change to the accuracy.

On a side note, data augmentation was observed to help reduce the problem of overfitting due to our limited dataset of 3039 images. It improved our original CNN from 73% accuracy to 91.7% accuracy.

## Ramifications

So, as can be seen from our results, a machine can tell anime from cartoon at about 90% accuracy (hurray).

However, it failed on a popular cartoon series – it classified Avatar: The Last Airbender as an anime, rather than a cartoon. With 20 sets to predict, 85% of the examples were classified as anime. It is clear that there is room for improvement, especially for the more ambiguous cases.

## Future

Anime art styles are quite distinct, and one of the main features can be said to be the way the eyes are drawn. One way we could potentially improve on the model is to focus specifically on eyes as a feature.

Other possible steps we could take would be to train the data sets with new and more relevant features such as HOG, GIST and SIFT.

We also plan on trying different algorithms such as using linear kernels and training with SVM to compare the accuracies amongst the different algorithms of generalisations.