Artificial Intelligence

Machine Learning and Pattern Recognition
Computer Science

A comparisson between Anime, Manga, Cartoons and Comics

Data Management and Analytics

Author:
Christian Bernabe Cabrera

Academic year 2019/2020

Contents

List of Figures

Chapter 1

Motivation

1.1 Project Goal

For this project I decided to analyse a general pop culture question "What is the difference between a Comic, Cartoon, Manga and Anime?". More specifically, I wish to see if a machine can learn to differentiate between all of this categories.If a machine could accurately classify images, it would mean that there are sufficient features that identify each category of animation.
This report will treat the following aspects:

- Historical difference. I will briefly describe the problem I will analyse
- Dataset
- Neural Network
- SVM
- Results and conclusion.

1.2 Historical difference

The simple way to explain this is that **Comic** and **Cartoon** in the Western World is the same as **Manga** and **Anime** in Japan. Comic and Manga refer to the medium of illustrated stories on physical pages, whereas Cartoon and Anime refer to illustrated stories on TV.
It's pretty easy to recognise Western comic books and manga, because they may look like they're from two different worlds. From the art, to text orientation, themes, and characters.

Instead while the quality of cartoons from America and Japan may not see a significant difference as a whole, there are distinct trends in the popular styles and genres of American cartoons versus anime. American cartoons trend towards comedy, and while they haven't always been aimed at kids, the majority of American animation is family-friendly. Anime has its comedies and its kids/family shows, but the range of genres and demographics is much wider, including realistic dramas that are almost never considered for animation in America.

The differences between the two animation industries can be explained historically. The first American cartoons were theatrical short subjects. When you went to the movies in the 1930s, you'd be paying for a full experience, with newsreels and shorts - both animated and live action - preceding the feature (there was also often a cheaper "B movie" following the main attraction). In that context, animated shorts were most effective as little bursts of comedy. They also served as ways to showcase studios' music libraries - a precursor to the music video - establishing a connection between the cartoon and the musical that doesn't really exist in Japan (but oddly enough, there are lots of stage musicals based on anime - yet hardly any actual anime musicals).

For a long time, feature-length animated films were almost in the exclusive domain of Walt Disney. Disney had experimental ambitions early on in his career, but following the commercial failure of the artier and more adult-oriented Fantasia he retreated to the safety of the fairy tale formula that made him so much money on Snow White. This started the stigma of animation being primarily for kids and families, a stigma which became all the more severe when animation moved to television in the '60s. TV killed the theatrical short subject, and with it, the subversive comedy of the Warner Bros cartoons and the visual experimentation of UPA's shorts. While those theatrical shorts were shown before adult-oriented films and as such often aimed at the adults in the audiences, on television, most animation got relegated to Saturday mornings when most adults were sleeping in. The Flintstones had some success as a primetime sitcom, but as other primetime cartoons flopped, it would be over two decades before The Simpsons made comedy cartoons for adults commercially viable again in the mainstream.

1.3 The visual difference

1.3.1 Manga

Manga has somewhat of a specific style. Features like hair, eyes, and other facial details are delicate yet defined, and action shots are drawn in a more dynamic manner. Additionally, manga tends to incorporate many cultural influences into its artwork, making it unique from any other form of comic books. Most mangas are completely black and white (with the exception of the cover art). They tend to feature two-dimensional drawings, characters with large eyes, and hair of abnormal size and color, and they are about some sort of conflict. Emotions are shown more often by using symbols (such as drops of sweat for worry) than by words.

1.3.2 Comic

When it comes to art styles, Western comics tend to have a bit more variety, especially with indie comics. Superhero comics, like DC and Marvel, tend to have similar styles, but there tends to be more flexibility. Styles can vary from traditional illustration, watercolor, pointillism, and more. Also comics are often colorized.

1.3.3 Anime

Although Manga are often used as the basis for anime, not every anime is from a manga and most manga are never made into anime. But they share most of the body structure features, such as similarities in the way the facial characteristics are exaggerated, as well as having very beautiful environments.
Physical features of characters are, on the whole, closer to reality than cartoons. A few not so realistic characteristics of anime characters are their larger eyes and smaller mouths make for a cuter style.

1.3.4 Cartoon

The body structure of an anime character and a cartoon character is largely different. Whereas a deformed or exaggerated touch is given to a cartoon character to give a comical essence to it, the anime character can be seen in proportionate body structure reflecting an actual human body as we said previously. A huge array of color variants and shades used in animes whilst absent in cartoons is what helps anime to achieve this amazing effect.

Figure 1.1: Cartoons vs Anime[1]

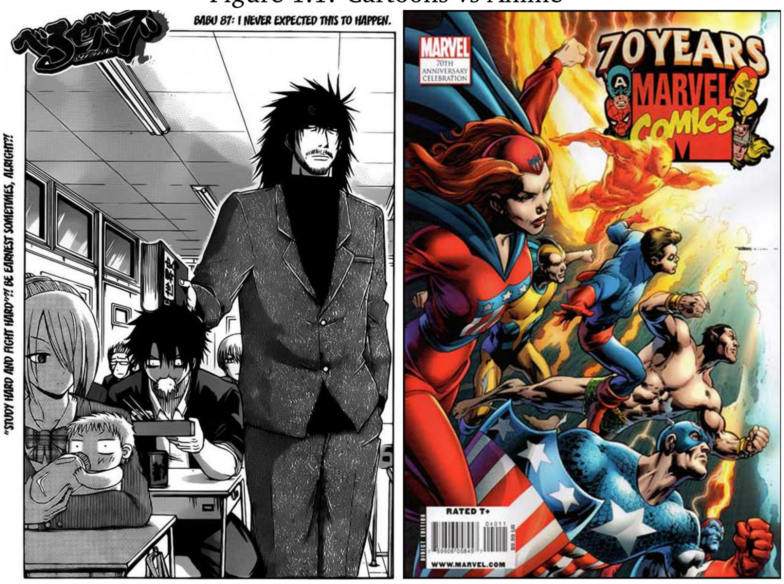

Figure 1.2: Manga vs Comics example[2]

Chapter 2

Introduction and Dataset

I started this project following the footsteps of the project called "Are anime Cartoons?".
Initially I used their dataset which consisted in 3600 images where their convolutional neural network managed to obtain about 92% accuracy. The data includes coloured screencaps of anime and cartoon shows and they were collected through batch image downloads of google image search results. Because of the limited data, they performed some augmentation on the set to decrease bias (zooming, flipping, and shearing).

I tested the dataset provided by the previous project using binary classification methods to predict whether an image was an anime or a cartoon. Having obtained similar results, but not satisfactory ones, I looked out for another dataset. Finally I also searched for other datasets in order to implement multi class classification methods.

To classify anime from cartoons I manually modified the iCartoonFace detection dataset. This is a large-scale dataset established for cartoon face detection, which contains multiple styles. The statistical characteristics of the iCartoonFace detection dataset are listed below:

- Large-scale.
- The iCartoonFace detection dataset consists of the training set which contains 50,000 images with a total of 91,163 faces and the testing set which contains 10,000 images with a total of 18,647 faces.
- High quality. The dataset is manually labeled, it has undergone multiple quality checks to ensure that the error rate is less than 5%. More than 60% of faces have a resolution greater than 50x50 in the training set.

In order to accomplish my desired task, I randomly selected around 20,000 images for each of the two classes from this dataset.

To perform multi class classification I also found a manga dataset and a comic book one. For the manga class I used the Manga109 dataset:

- This data set (hereafter referred to as Manga109) has been compiled by the Aizawa Yamasaki Matsui Laboratory, Department of Information and Communication Engineering, the Graduate School of Information Science and Technology, the University of Tokyo.

- The compilation is intended for use in academic research on the media processing of Japanese manga.
- Manga109 is composed of 109 manga volumes drawn by professional manga artists in Japan. These manga were commercially made available to the public between the 1970s and 2010s

Finally for comic books class I used a comic book images dataset available at kaggle which consists on 52156 rgb images.

Overall I ended up with around:

- 20.807 images for the cartoon class
- 20.886 images for the anime class
- 10.898 images for the comic book class
- 9.609 images for the manga class per manga

Chapter 3

Neural Network binary classification

3.1 Architetecture

The architecture involves stacking convolutional layers with small 3×3 filters and stride equal to 1, followed by batchNormalization layer and a max pooling layer. Together, these layers form a block, and these blocks are repeated where the number of filters in each block is increased with the depth of the network such as 32, 64, 128 features respectively, for the first three blocks of the model.
Through all the layers there is the activation function ReLu and also the Dropout(p) in order to reduce overfitting with p equal to 0.2 for the first two blocks, and 0.50 in the final one for a major consistency in the classifier part of the model.

Since the problem is a binary classification I used the Binary Cross-Entropy as loss function and the Adam criterion as optimizer.
I resized the images in both the training, validation and test set for consistency. Additionally I randomly cropped the images in the training set to introduce some data augmentation to artificially expand the size of a training dataset by creating modified versions of images.

I chose to do this since augmentation techniques can create variations of the images that can improve the ability of the fit models to generalize what they have learned to new images.
The dataset was divided following the proportion training/validation/test sets 70/15/15.

3.2 Results

3.2.1 First model

For the first model includes the first dataset with anime and cartoons. In this case, we can see that the model achieved an accuracy of about 94.43% on the test dataset.
A figure is also created showing a line plot for the loss and another for the accuracy of the model on both the train (blue) and validation (orange) datasets.

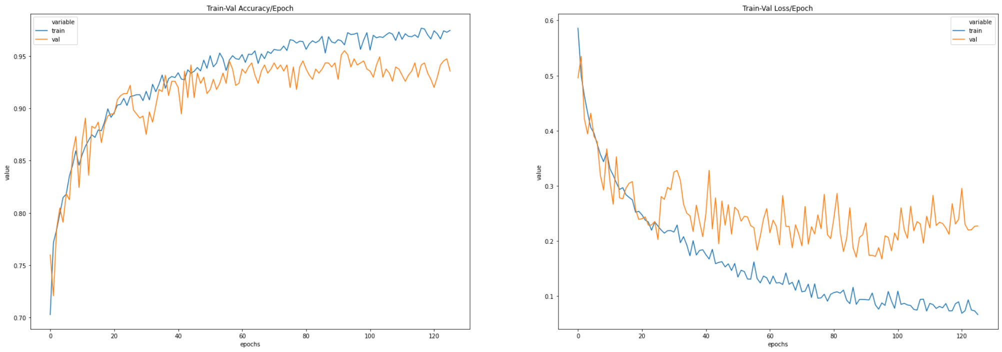

Figure 3.1: Graph Train/Val per epoch[1]

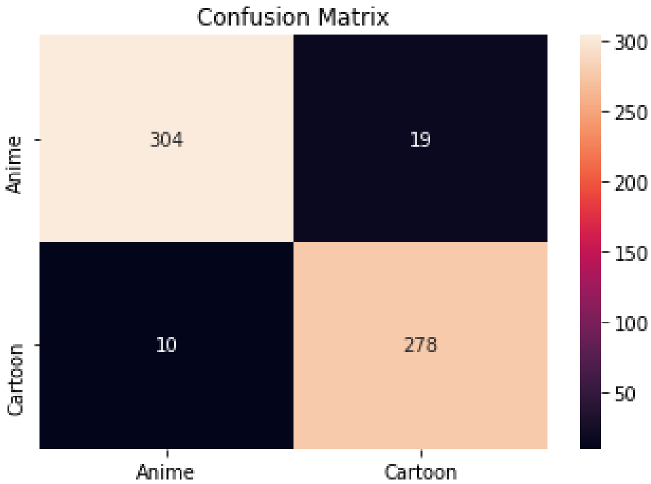

Figure 3.2: Confusion matrix on the Test Set of 611 elements[2]

Reviewing this plot we can see that even though accuracy is satisfactory, the model has overfit the training dataset at about 25 epochs, which is not totally satisfactory.

3.2.2 Second model

For the second model, I used the manually modified iCartoondataset. In this case, we can see that the model achieved a noticeable improvement in performance from about 94.43% to about 96.29% accuracy. Reviewing the plot of the learning curves however, we can see that the over-fitting has been dramatically impacted. Which leads to a improvement of the model overall.

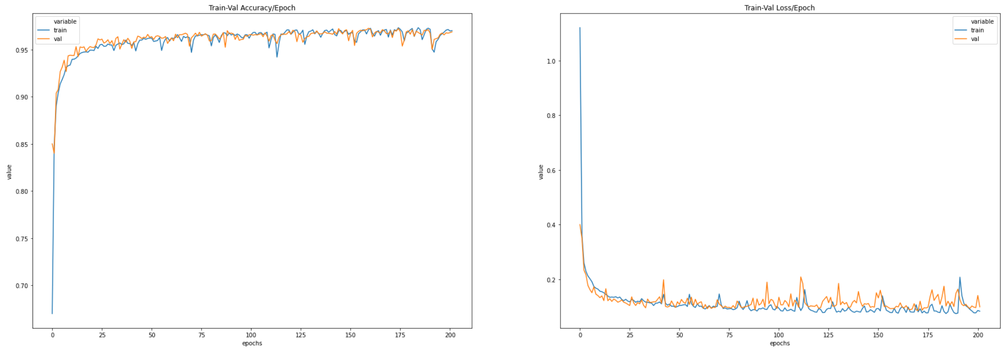

Figure 3.3: Graph Train/Val per epoch[3]

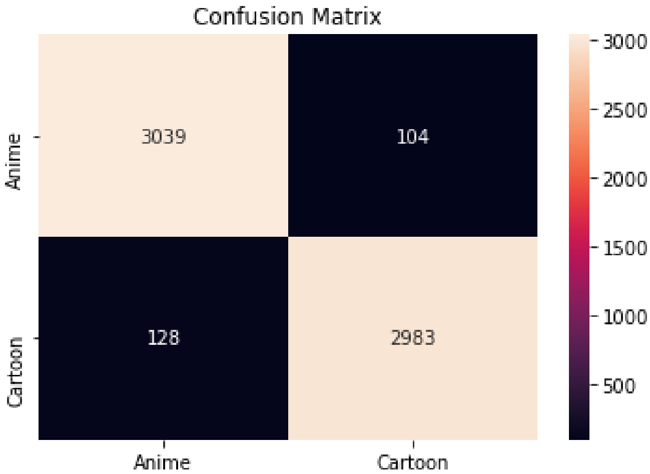

Figure 3.4: Confusion matrix on the Test Set of 6254 elements[4]

3.2.3 Third model

Since the previous model seemed to be stabilized after around 40 epochs, I decided to explore other regularization methods such as early stopping. I decided to be large with the patience and wait at least 30 epochs before stopping the program. In this case, we can see that the model achieved a slightly worse result in performance from about 96.29% to about 96.11% accuracy with early stopping. Reviewing the plot of the learning curves, we can see that again the model behaves very similar as the previous one. However almost the same accuracy has been achieved in half the computational time, specifically 1:45h vs 4:07h.

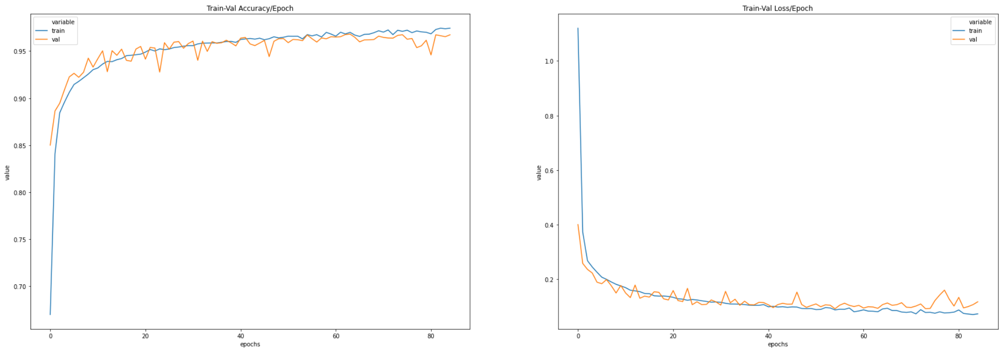

Figure 3.5: Graph Train/Val per epoch[5]

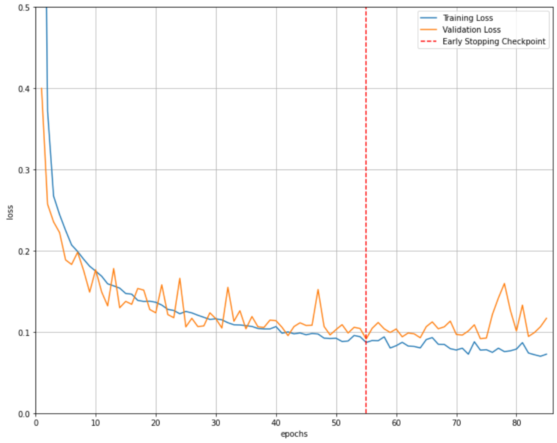

Figure 3.6: Graph Train/Val per epoch with early stopping[6]

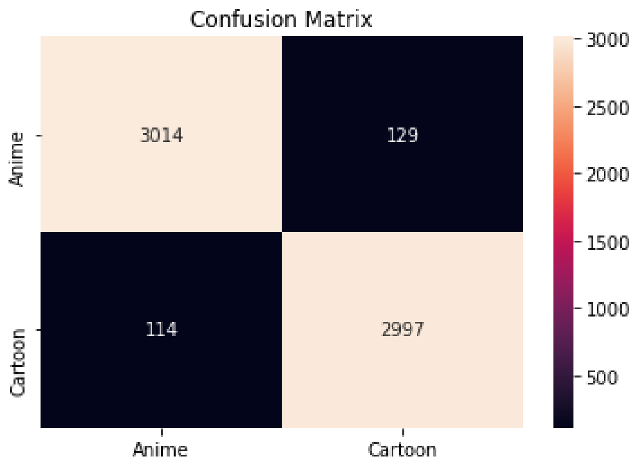

Figure 3.7: Confusion matrix on the Test Set of 6254 elements[7]

Chapter 4

Multiclass classification

4.1 Architecture

The dataset for this model was obtained through the usage of three different datasets:

- Anime and Cartoons
- Comics
- Manga

Even though the initial scope of each dataset was different, nevertheless, they can be used as the basis to develop, evaluate, and use convolutional deep learning neural networks for styles image classification.

The architecture of the CNN remained the same. Since this is a multi-class classification problem, I used a softmax activation function and the categorical cross entropy loss function. As the optimizer I used Adam once again.

4.1.1 Model results

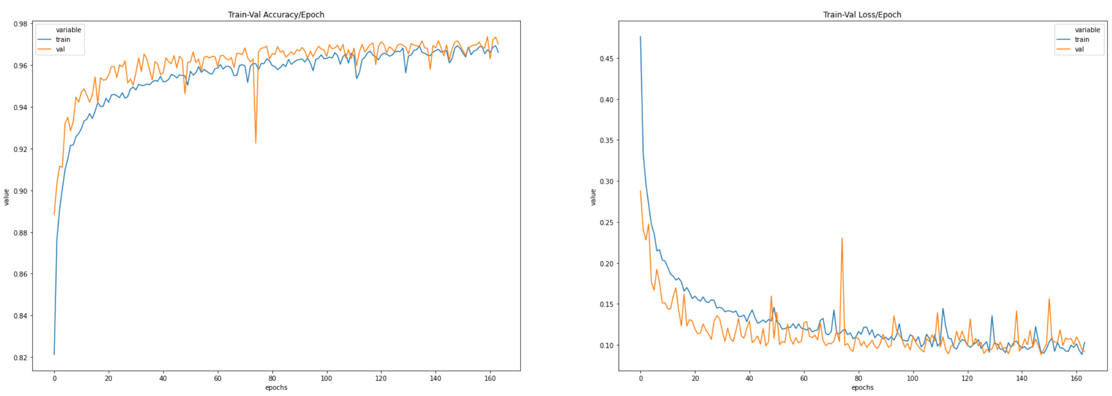

Figure 4.1: Graph Train/Val per epoch[1]

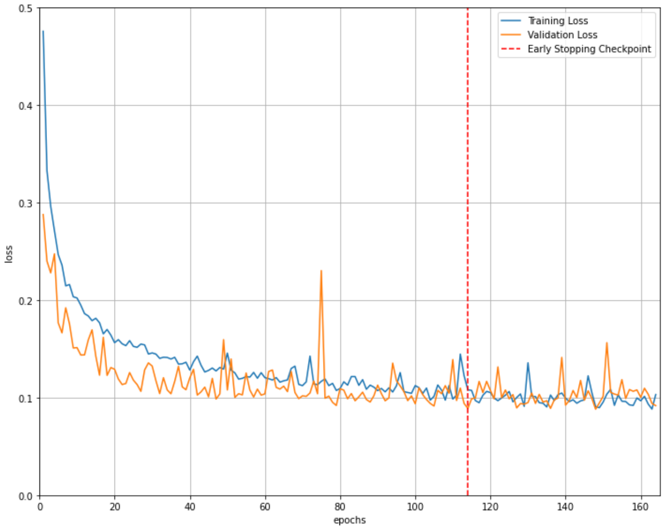

Figure 4.2: Graph Train/Val per epoch with early stopping[2]

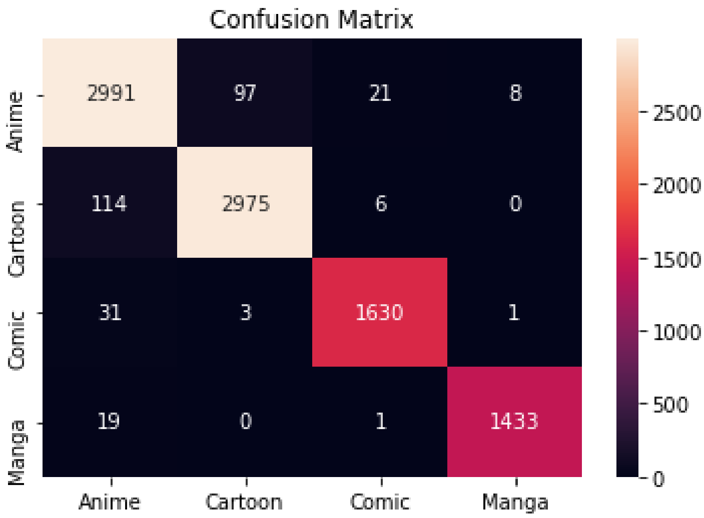

Figure 4.3: Confusion matrix on the Test Set of 9330 elements[3]

For multiclass classification I decided to skip the previous model analysis. Instead i went ahead directly with the usage of early stopping. This time I opted to be even more patient, since the program would take a lot of processing time, so I chose 50 epochs before stopping the program. We can see that the model achieved an overal better result in performance of about 96.77% of accuracy with early stopping. Reviewing the plot of the learning curves, we can see that the overfitting is basically absent, leading this model to be the more satisfactory of all the others.

Chapter 5

SVM

I developed 2 different Support Vector Machines with 2 different kernels:

- Linear Kernel implemented with svm.SVC()
- Radial Basis Function Kernel

I used an "Exhaustive Grid Search" to tune the hyper-parameters, with the following results:

- SVM with Linear Kernel with C = 1000 accuracy = 0.62
- SVM with Radial Basis Function Kernel with C = 1000, gamma = 0.001 accuracy = 0.72

The results seem to indicate that SVM works poorly with this type of classes, which may have some overlapping overall. Due to the nature of SVM, which is not suitable for large datasets because of its high training time and it also takes more time in training I opted to not implement SVM for the upgraded Cartoon vs Anime dataset and the Multi class dataset.

Chapter 6

Discussion

I have explored tow different improvements to the baseline model.
The results can be summarized below, although we must assume some variance in these results given the stochastic nature of the algorithm:

- Old baseline + Dropout Regularization + Data Augmentation: 94.43%
- Old baseline + SVM with Linear Kernel: 62%
- Old baseline + SVM with RBF Kernel: 72%
- Baseline + Dropout Regularization + Data Augmentation: 96.29%
- Baseline + Dropout Regularization + Data Augmentation + Early Stopping: 96.11%
- Multiclass Baseline + Dropout Regularization + Data Augmentation: 96.70%
- Multiclass Baseline + Dropout Regularization + Data Augmentation + Early Stopping:96.77%

The results obtained with the old "cartoon vs anime " dataset and the new one are very similar in terms of accuracy. However the introduction in the new dataset of regularization techniques, as suspected, slows the progression of the learning algorithms and reduces overfitting, resulting in improved performance on the holdout dataset. Overall this dataset achieved excellent results.
The multi class baseline works very well too, recognizing most of the time comics and manga too.
Finally the svm one performed very bad, and would not be taken under further consideration.

Chapter 7

Future

It is likely that the combination of further approaches with maybe a further increase in the number of training epochs will result in further improvements. In addition to tweaks to the regularization methods described, other regularization methods could be explored such as weight decay for example.
It may be worth exploring changes to the learning algorithm, such as changes to the learning rate, use of a learning rate schedule, or different optimizers.
Alternate model architectures may also be worth exploring. The chosen baseline model is expected to offer more capacity than may be required for this problem and a smaller model may faster to train and in turn could result in better performance.

7.1 Future utilizations

As we have seen, there are significant differentiating features between anime, cartoons, comics and manga.
This feature could be utilized to improve for example the "search by image" provided by google or other private sites like "tinEye" , which still are not able to differentiate anime from cartoons, or comics from manga.
It could be further optimized to recognize an artist style in order not only to categorize the class label, but also the artist.
Finally a very interesting feature would be, given an image, give the detection bounding boxes of all cartoon faces in the image and effectuate face recognition. This would actually be very nice and would be an useful implementation to some already existing algorithms like Amazon Prime's character detector in order to have (when the user pauses the video) a summary of the cartoon/anime characters present in the scene, with a brief description for each one.