Gundam Classification with CNN

HE, Jianle
The Hong Kong University of Science and Technology
jhebg@ust.hk

Ho, Chun To
The Hong Kong University of Science and Technology
cthoab@ust.hk

1. Introduction

Consider the images below in Figure 1. It is very hard for people to recognize all the models of "Mobile Suit" in the whole series of the "Mobile Suit Gundam" animations. Only die heart fan of this series would have the ability to classify all of the machines. Is it possible to train a classification model having the same ability compare to those fans? There are two approaches for tackling this problem. Our first approach is to build a convolutional neural network with custom architecture for the image classification and the second approach is to do some transfer learning with ImageNet model to evaluate the predict result. Numbers of image processing algorithms will also be implemented, including histogram of color (HoS), gradient and orientation histogram, and some data augmentation algorithms.

Figure 1

2. Problem statement

This research topic aims to help people to recognize different types of "Mobile Suit" inside the animation of "Mobile Suit Gundam" by using the classifier to identify different types of "Mobile Suits".

To achieve the research goal, a large dataset of Gundam images and their name are needed for the image classification. However, there are no pre-processed dataset of Gundam available online, it is decided to collect the images of Gundam from the Internet by web scraping. In this state, there are two different data sources, mahq.net and fan-made wiki. The two sorts of dataset are in huge different. Images from mahq.net have a clean background with 400x400 dimension, but the quantity is small. For the images from wiki, some of them are toys' product shot, some of them are from manga or non-official drawing, some of them crossover with character from other brands, large amount of data with low quality and are polluted. In order to increase the quantity and quality of sample for better training, some data augmentation skills and data cleaning skills would be applied to the images, for instance, applying flip, shift, rotate, zooming in/out method to images, can increase the number of data to be fed into the network. Another way to increase the sample size is grabbing the images from google image, with searching the model number (e.g. RX-78-2 is the model number of a titular mobile suit) as keyword, to obtain the images with less pollution.

For the result evaluation, confusion matrix will be used for checking whether the classification is correct. It is a combination of actual and predicted classes. The overlapped grid of the matrix is representing how the model understands the classes. This is a good way to study which classes are easily confused. Class prediction error is another analysis that extends the confusion matrix. It visualizes the misclassified classes as a stacked bar. Each bar is a composite of predicted classes. However, there are nearly a thousand of classes, directly plotting a confusion matrix maybe messy, so that grouping the classes will be considerate.

3. Technical Approach

In order to solve the problem, there are two different approaches in this state. The first one is to train our own model by the convolutional neural network. With extracting the feature from images, such as the color histogram, object context. Applying the small filter layers and pooling layers to down-sampling the size of the image for easier managing the representation of the data. Before finetuning the best hyperparameters of the model, our approach is to build a naïve model for testing and adjust the model by observing the learning and loss curve, accuracy and loss curve. The architecture should be:

```
[[Conv-ReLU]*3]-BN-MaxPool]*N
- [FC-Dropout]*N - FC - output
```

The second one is transfer learning. There are many image classification models on the Internet, fine tune the layers of these models and see how they perform. Some pre-trained models from CIFAR-100 and ImageNet should be able to provide some inspiring ideas about image classification problems.

4. Intermediate/Preliminary Results

The first attempt is using the "raw" data mentioned above, the result is poor. The possible reasons for the result are as following, huge different among images, the hyperparameters are setup randomly, the sample size is not large enough for each individual class. In the coming state, all the potential problems will be tackled to increase the performance of the model. In case the improvement is small, ensemble learning will also be considerate. Combining all the algorithms and techniques mentioned above to evaluate how the performance change.

Figure 2 (Mobile Suit Gundam dataset)