# Anime Character Detection

Yu Hong Zhou
*Computer Science and Information Engineering*
*National Tsing Hua University*
Changhua, Republic of China
chouyuhung1@gmail.com

Jhih Yu Jian
*Computer Science and Information Engineering*
*National Tsing Hua University*
Tainan, Republic of China
e0117e0117@gmail.com

Bo Xin Zheng
*Computer Science and Information Engineering*
*National Tsing Hua University*
Pingtung, Republic of China
deron0506@gmail.com

Yan Zhang Li
*Computer Science and Information Engineering*
*National Tsing Hua University*
New Taipei City, Republic of China
960610vito@gmail.com

***Abstract*—Anime is a youth culture of entertainment stemming from Japan in recent years. Combining elements in sort of fields like workplace culture, school life, or even imaginative outside-world, genres of anime can be wide-ranging, at the same time unfolding touching, fantastic, depressing stories.**

**With large varieties of the style, anime are always interpreted in many ways, one of which is to present the scene with multiple types of illustration style. For example, use heavy tone of color for a sorrowful tale.**

**Our study aims at identifying the characters from different anime based on their illustration, which may be capable of helping anime-lovers in searching unknown anime.**

***Keywords—Anime, Character detection***

## I. INTRODUCTION

The convolutional neural network is a class of artificial neural network which can be used mainly for image classification with one or more convolutional layer.

The convolutional neural network has shown accurate predictions of its outcomes and better performance in image classification. The pre-processing required in the convolutional neural network is much lower as compared to other classification algorithms.

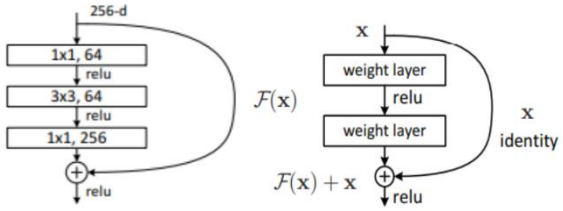
Image classification is the process of taking inputs and outputting a class or a probability that the inputs are a particular class.

## II. METHODS

For the model training part, we had tried three different model architecture: pure CNN model, VGG16, and ResNet50, in which ResNet50 is the prominent one. We will put the part of VGG16 training process in the DISCUSSION part, for now we focus on ResNet50.

For the verification of performance of ResNet50, we refer to the content of the paper which is listed in the REFERENCES part.

In short, ResNet50 makes use of shortcut connection and bottleneck block (1*1 conv) to implement the Residual mapping. With this design, excellent performance can be shown in deeper network training.




Without this way to reinforce the data, CNN model cannot learn and identify the pictures well, due to the depth.

To begin our work, collecting some data is what we need to do. The following talks about how we get the dataset.

First, we download some images in bulk of a particular animation through python website crawler from google. But, we soon encounter a problem that the images on google are not only cluttered but also few in number. We should do it another way since our model requires a large number of training set.

The alternative is to grab images directly from anime episodes. Let's elaborate it step by step. First, we download the anime videos, taking some screenshots automatically by python.[4] Then use a package named cv2 in python to filter the images according to their features like anime character's eyes or faces.[3] Finally, filters again manually since the package may not be smart at all.

The above process is somehow time-consuming, and end up five categories.

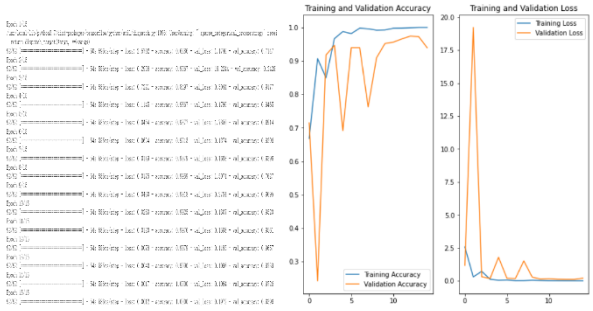
For the UI part, a GUI platform is provided using Python and Tkinter[2] package so that users can commodiously interact with our model. On the interface, user can upload image of anime character, and click the Run button to get the predicted animation name, description, and corresponding anime theme song. It's a efficient way for us to check the correctness of the result.


## III. RESULTS

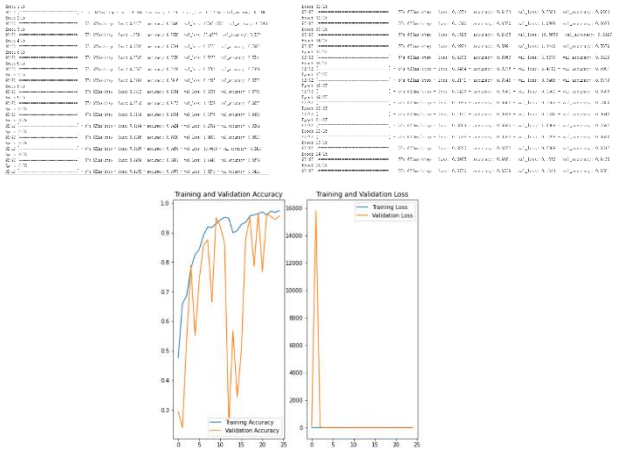
Using ResNet50 model and stochastic gradient decent as optimizer with 0.01 of learning rate, 0 momentum, and softmax classifier activation function, we get a nearly perfect validation set accuracy. However, for the one with data

augmentation, it may have overfitted for we found that the accuracy on testing data (including pictures from Internet or our dataset) cannot reach higher than 60%. For the other one without data augmentation, the accuracy on testing data was great enough for us to say that the model has no fear of testing.

Here is our result without data augmentation.




And here is our result with data augmentation.




## IV. Discussion

We use VGG16 model to predict images previously. But VGG16 model in our anime character detector does not have better performance than ResNet50 model. Next, we discuss the difference and the performance between VGG16 model and ResNet50 in our anime character detector.

VGG16 use multiple 3x3 kernel-sized filters one after another. The images are passed through a stack of convolutional layers, where the filters were used with a very small receptive field. Multiple non-linear layers can increase depth of neural network to guarantee more complicated learning model and it contains fewer parameters.

ResNet50 designs a type of residual building block to train more deep neural network. Training problem of deep learning is called degradation. Then, ResNet50 can solve degradation. ResNet50 use global average pooling layers and residual building block that can train 152 layers of residual network. The accuracy is better than VGG16 and the efficiency is also better than VGG16.

As below are how we can improve the model.First, increase the dataset, which we used about only 3600 in the project at present.Second, we have tried our best to make the performance better, but it can still be improved by using different model and hyperparameter.
Finally, the UI interface can be designed more close to users.

## V. Conclusion

To summarize our project, anime character detection, the main idea of it is to figure out which anime correspond to the given anime character's image. We have both try our best on data preprocessing and model training, and finally find that ResNet50 has the best performance. Although our machine may not always precisely predict the correct answer, its error rate is not high at all. It will only make a mistake when the anime characters look alike or the anime style are similar. Sparing no effort on the final project, we think we make a great work!

## Author Contribution Statements

Jhih Yu Jian (25%): topic discussion, model training and designing, chief of question answering, report.

Yu Hong Zhou (25%): topic discussion, UI interface, report.

Yen Chang Li(25%): topic discussion, data collection and preprocess, final presentation, report.

Bo Xin Zheng (25%): topic discussion, model training and design, report.

## References


[1] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, "Deep Residual Learning for Image Recognition", pp. 770–778, Dec 2015. *(references)*

[2] "tkinter —Python interface to Tcl/Tk" [Online] . Available: https://docs.python.org/3/library/tkinter.html. *(references)*

[3] "Anime face detection with OpenCV" [Online] Available: https://linuxtut.com/en/fda37ac4f9ddab2d7f45/?fbclid=IwAR30H1i1V33ko7yL0DtncJ8f7wcUGLLZL6ZrkmQm5RstJtAfF5D6g45pll4

[4] "Python: take screenshot from video" [Online] Available: https://stackoverflow.com/questions/57791203/python-take-screenshot-from-video?fbclid=IwAR3e-gZTE2ODrt0XlCdC9Aho6bqnfz5OhXHhPqd4kSkt51r6XLc6GwfpzRs

[5] Jianxin Wu, "Introduction to Convolutional Neural Networks", pp.2–4, May 1 2017. *(references)*

[6] Hyungtae Lee, "Going Deeper with Contextual CNN for Hyperspectral Image Classification", pp.1–2, May 9 2017. *(references)*

[7] Tensorflow keras reference
https://www.tensorflow.org/api_docs/python/tf/keras/applications/resnet_v2/ResNet50V2?fbclid=IwAR1NfGOMSOUZSvHIqJ_VNFwpLfWb78l-2RxluqmpwC5kwsrttwYO5YaORjY