

USER GUIDANCE

1. FILE SPECIFICATION

As you can see there are some files in our handout file. And here we explain the content of each file.

A. Bonus Documents

This file includes our bonus document.

B. Experiment Data

i. CNN_model file*:

This file includes the final CNN model of our project. The svhn.caffemodel is our final training weight, while the mydeploy.prototxt tells caffe how to use this weight document.

ii. Dataset file:

a. SVHN_Dataset:

This file includes the svhn dataset and the preprocessed data.

b. Our_Captcha_Dataset*:

This file includes several captcha datasets generated by ourselves.

iii. Weight file:

The weight file includes our training weight snapshot. And this file is not so important, and you can just use the weight in CNN model file.

iv. Extra pictures*:

There are several pictures in this file. And they illustrate some details of our CNN model, it can help you to understand our model.

C. Final Presentation ppt

- i. This file includes our final presentation ppt and our lecture notes, which might help you to remember details of our presentation.

D. Report

This file includes our final report and some extra materials.

The SE_report CaptchaHunter document is our final report.

But we also noticed that as we combine all the software design, requirements and user case together into the final report, it seems to be kind of disordered.

So, in order to let reader to catch our main idea of projects clearly, we also put SE_report CaptchaHunter concise document in the extra file.

And we believe this file might be more easily for readers to follow than the complete version.

More detailed things, you could read **README.md in report file***

E. Student Contribution

This file includes the contribution of each student.

F. Source Code

i. dataset generating:

a. Developed_by_Zhenchao_Zhou:

In this sub-file, there is a method developed by Zhenchao Zhou which crawls the captcha images from website. And it generates part of our dataset. This code should be run under the python2.7 environment.

b. Developed_by_Ruiji_Wang:

In this sub-file, there is a method developed by Ruiji Wang which generates part of our dataset. And also some primitive partitioning methods of our project. You can read the README document in it to see details.

ii. CaptchaHunter_2_0*:

This file is about our final GUI and the combination of our final captcha hunting project's partitioning and recognizing part.

Why we call it _2_0 is because we develop several versions of our GUI and final code. And this is the newest version.

This code is mainly written by Lingkun Kong, under Ubuntu14.04 + python3.5 + several python libraries.

The detailed usage manual is in **the README.md document***.

But in short:

Cd into the captcha_hunter_2_0;

Python final.py;

iii. Ipython notebook*:

It includes specific details of how to train our CNN network and how our CNN network works based on SVHN dataset.

The detailed usage manual is in **the README.md document*** in it.

iv. Caffe config:

This file includes the configuration details of caffe.

2. SOME IMPORTANT THINGS

If you want to just know how our project works, just open the source code file and then open the captcha_hunter_2_0 file, read the README.md carefully, then you will learn anything.

Warning: you must use python3.5, and we strongly recommend you to install anaconda3.5.

But if you want learn more about our CNN model, there are two approaches:

1. *Open the svhn.ipynb document by python notebook, and you can get graphic ideas.*
Ubuntu14.04 + python2.7 + anaconda2.7 + matlab + caffe + openCV3
2. *Open the captcha.learn.py in captchahunter_2_0, read the function one by one. But we don't recommend you to use this method as the former one can show visualization details vividly and is much more friendly to readers.*

Warning: caffe configuration is really annoying. If you meet any trouble, contact me at kenny_kong@foxmail.com