Summer Internship Report

Gan Group Toy Images Recognition Subsystem

IDEA LAB
浙江大学-阿里巴巴联合实验室

Intern: Zhe ZHANG
Advisor: Prof. Lingyun SUN
Supervisor: D. Zejian LI, D. Jingyu WU
Time: 2020.07.22~2020.08.03
Format: Online internship

Plans and Goals

In this internship, I was assigned to the GAN group in the laboratory due to special reasons. And conducted the related work of "Toy Image Recognition Subsystem" under the guidance of D. Zejian LI and D. Jingyu WU. I received the assignment on the afternoon of July 22. Under the plan of professors and seniors, the work I did was mainly concentrated in three stages:

I. Purchase toys for image recognition system; Investigate existing object detection models and try to build
II. Collect and expand the dataset; Perform model training and related verification
III. Further optimization for training model; Generate scene graph

Summary

In the current internship (one week), I have completed the 1st and 2nd stages of the prescribed plan, and successively carried out the 3rd stage of practice and attempt. During the internship, I sorted out the documents of the internship diary as well as core modules every day. I also release my work in: https://github.com/doubleZ0108/IDEA-Lab-Summer-Camp, and the work of “Data Augmentation” is released in： https://github.com/doubleZ0108/Data-Augmentation, hoping some one else could need the related algorithms and experiments.

Table1: Summary of internship work

	Overview	Principal	Concrete Content
Environment Setup	① Configure for server environment ② Install dependencies for deep learning ③ Building yolo network	Zhe ZHANG	① connect ssh and ftp services of the server ② configure opencv environment and write documentation ③ configure the yolo environment for testing and write documentation
Experiment	① Dataset collection ② Data augmentation ③ Data labeling ④ Dataset summary	Zhe ZHANG	① take 93 pictures as dataset ② use 11 methods to expand the dataset and write documentation ③ manually annotate more than 500 pictures, and use scripts to automatically annotate the rest ④ sort out the data left before, and sum up 1551 total pictures
Model Training	① yolov3 ② yolov4 ③ Testing for model	Zhe ZHANG	① use yolov3 network to train for 2000 epochs ② use yolov4 network to train for 2000 epochs ③ test the mAP and actual effect of the two networks ④ use pictures and videos for acceptance testing of the network

Contents

【Environment Setup】

1. **Server**：In this internship, an Alibaba Cloud server applied by seniors was used for deep learning training, guided by the thought of "If you want to be good at work, you must first sharpen your tools". I first checked and compared many remotely connected services. Finally, I chose "Termius" on the macOS side and "Termius" on the iPadOS side for ssh service; "Cyberduck" for ftp service, after the basic application configuration, greatly reduced the complexity of my use of the server, and improved my work efficiency in file upload and processing

2. **opencv**：The server is pre-installed with the MiniConda environment, which has a certain impact on the configuration of the opencv environment, so many problems were encountered when the configuration was started. After continuous exploration attempts and related information inquiries, I completed the configuration of the opencv environment under the server cuda version related issues. Among them, I recorded some representative issues to facilitate future work:

    1. Unable to locate package libjasper-dev.
    2. Unable to download ippicv_linux_20151201.tgz due to network problems.
    3. The following variables are used in this project, but they are set to NOTFOUND.
    4. cuda9 no longer supports 2.0 architectures
    5. Unsupported gpu architecture 'compute_20' .
    6. Package opencv was not found in the pkg-config search path.

3. **Yolo**：According to the official information and related blogs, under the premise that other environments are well configured. Firstly, configure the Yolo environment to check whether the target detection practice can be carried out normally. The main steps are as follows:

    1. Cloning and building darknet
    2. Download pre-trained yolov weights
    3. Run detections with darknet and yolo to verify whether the build is successful

【Dataset】

1. **Dataset Collection**：Since the identification object of this system is a separately purchased toy, there is no suitable sample in a relatively large public data set (such as Google Open Images Dataset). And the previous pictures was a little rough so the mobile phone(iPhone 11) + PTZ(DJI OSMO Mobile3) shooting method was used in the process of data collection, so that the parameters such as iso and exposure can be manually controlled for taking higher quality pictures. I also choose a warmer-toned sofa and a cooler-toned wall for the scene, so that the data set has better diversity. On the other hand, due to the high quality pictures of the mobile phone, it will cause a lot of pressure during uploading and training. Therefore, after the pictures is taken, the script is used to compress the image with high fidelity, and the image size is reduced from 5M to 500K.

2. **Data Augmentation**：On the condition that I was the only one who built the subsystem and couldn't take a lot of images in a limited time. After conducting related research, I used the concept of "data expansion" in the deep learning course of Professor Enda WU to expand the toy dataset. Different from the practice of previous students, the literature emphasizes that "the operation of the data when artificially expanding the training data is best to reflect the changes in the real world", that is, although we are virtual expansion of the data set through simulation, it is supposed to reflect the real changes in the physical world. After several days of practice, I used five categories and a total of 11 methods to expand the data set:

    - Intensity Transform
        - Luminance Fluctuation： lightness darkness

    - Contrast Transform: contrast
- Filtering
    - Image Sharpening: sharpen
    - Gaussian Blur: blur
- Perspective Transform
    - Mirror Flip: flip
    - Image Clipping: crop
    - Image Stretching: deform
    - Lens Distortion: distortion
- Injected Noise
    - Salt ane Pepper Noise: noise
    - Vignetting: vignetting
- Others
    - Random Cutout: cutout

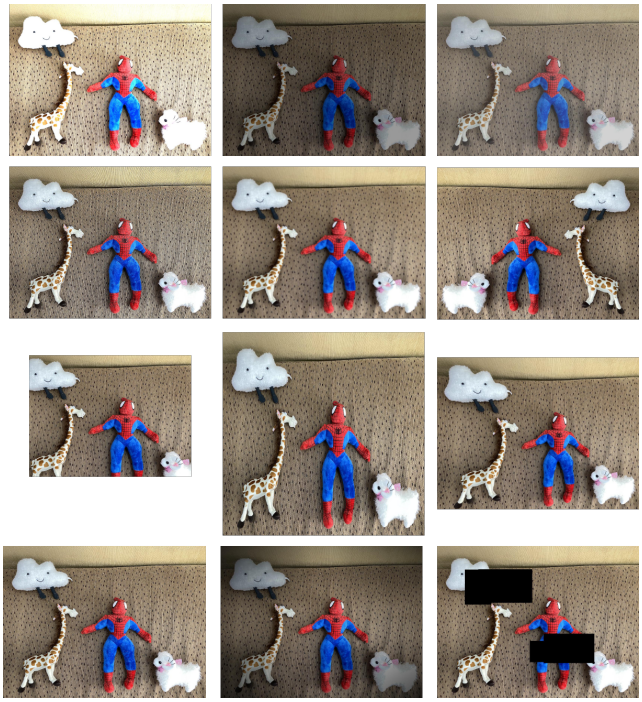

Figure 1: Thumbnail for Dataset Augmentation

3. **Data Labeling**: With enough data, the "labelImg" tool is used to mark images and get the ground truth of each bounding box. At the same time, the data augmentation through the dataset in the second step can be further classified. Most of the images are filtered and processed on the pixel, which will not affect the labeling result, so this part of the image can avoid repeated labeling. Some images, such as lens distortion and cropping, are difficult to realize due to poor adaptive effect, so manual annotation is adopted.

4. **Dataset Summary**: After synthesizing the above three steps, a total of 1551 annotated image data were obtained for subsequent model training
    - previous: Filter the data left by previous students and select 342 images that meet the specifications
    - main: 93 images taken with a mobile phone without any processing
    - man_labled: 279 images that need to be manually labeled after crop, deform, and distortion of the image in main
    - auto_generated: 837 images with main processed pictures and automatically relabeled

【Model Training】

After the construction of the training environment and the acquisition of the dataset in the 1st stage, use yolo to train the custom data set through the relevant configuration files. The most important is the three-part configuration file:

1. **cfg configuration file**: the cfg file is related to the architecture of the neural network and the parameters of the target detector. The core parameters are as follows:
    - bash = 64(train), bash = 1(test): number of batches, images sent to the neural network in each iteration
    - subdivision = 32(train), subdivision = 1(test): divide bash into the neural network to reduce the use of video memory, but it is limited by server performance
    - classes = 4: the type of classification to be performed on the network is changed to four toys in the three YOLO layers
    - filters = (classes + 5) * 3: number of convolution kernels in the convolutional layer before the three YOLO layers
    - max_batches = (# of classes) * 2000: maximum number of iterations
2. **obj.data & obj.name configuration file**: the obj file is related to the basic configuration information and the category name to be classified
    - classes = 4: number of categories to be classified
    - train = data/train.txt: path for training list
    - valid = data/valid.txt: path for validation list

- names = data/obj.name: name of each category to be classified
- backup = backup: oath for gradually store the weights obtained during training

3. **train、valid、test list files**：Each piece of data is a piece of image data used for training, verification, and testing. In the process of project construction, different scripts are written to automatically generate these list files, and the total list is divided according to the ratio of 8, 1, and 1 to simplify manual labor, and improve the uniform division of data.

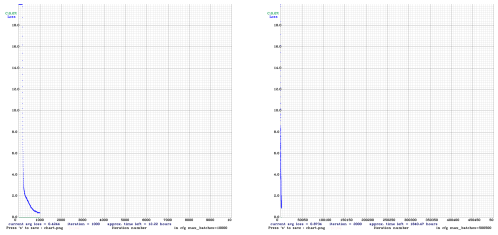

Figure 2: loss curve of yolov3 and yolov4 during training

After getting all the configuration files, use yolov3 and yolov4 to train the custom dataset. The loss curve of yolov3 training 1000 epochs, yolov4 training 2000 epochs, and the recognition effect of different situations after training 2000 epochs is shown in the figure：

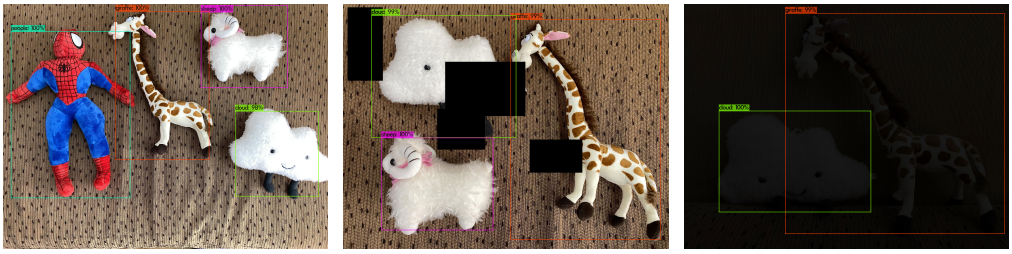

Figure3: recognition effect of yolov3 and yolov4

Not only the image, but also the trained model is used to recognize the video taken in real time. Yolov3 can reach 41.8FPS and yolov4 can reach 31.4FPS. These frame rates can already meet the needs in actual use, and the accuracy of recognition in the actual acceptance is also very high. Specifically, we can also observe from mAP:

```
 detections_count = 346, unique_truth_count = 290
class_id = 0, name = sheep, ap = 100.00%          (TP = 88, FP = 0)
class_id = 1, name = giraffe, ap = 100.00%        (TP = 78, FP = 0)
class_id = 2, name = cloud, ap = 100.00%          (TP = 77, FP = 0)
class_id = 3, name = people, ap = 100.00%         (TP = 47, FP = 0)

 for conf_thresh = 0.25, precision = 1.00, recall = 1.00, F1-score = 1.00
 for conf_thresh = 0.25, TP = 290, FP = 0, FN = 0, average IoU = 87.04 %

 IoU threshold = 50 %, used Area-Under-Curve for each unique Recall
 mean average precision (mAP@0.50) = 1.000000, or 100.00 %
Total Detection Time: 17 Seconds
```

```
 detections_count = 392, unique_truth_count = 290
class_id = 0, name = sheep, ap = 99.97%           (TP = 88, FP = 1)
class_id = 1, name = giraffe, ap = 99.98%         (TP = 78, FP = 1)
class_id = 2, name = cloud, ap = 100.00%          (TP = 77, FP = 0)
class_id = 3, name = people, ap = 100.00%         (TP = 47, FP = 0)

 for conf_thresh = 0.25, precision = 0.99, recall = 1.00, F1-score = 1.00
 for conf_thresh = 0.25, TP = 290, FP = 2, FN = 0, average IoU = 88.30 %

 IoU threshold = 50 %, used Area-Under-Curve for each unique Recall
 mean average precision (mAP@0.50) = 0.999896, or 99.99 %
Total Detection Time: 19 Seconds
```

Figure4: mAP of yolov3 and yolov4

Experience and Inspiration

2020 is a unique year, under such a social environment, I am really honored to be able to join the Zhejiang University inLab internship through remote summer camps. Although only a short period of more than a week, I do think I have seen and learned so much here. First of all, I would like to express my sincere respect to the teacher's patient listening, the senior's hard work these days, and all the teachers, seniors and sisters responsible for the normal operation of this summer camp. At such a special moment, it is you who supported my hope.

During this week's internship, I was individually assigned to the laboratory GAN team to undertake the task of implementing a subsystem. During this time, I fully studied the whole process of project research, literature research, environment construction, data collection, data augmentation, model training, and model verification, allowing me to contact and try a field of work from scratch. I do think this is not only a test for me, but also a very important experience for my life. Although I have done a small amount of scientific research work in the undergraduate stage, more often it is completely carried out according to the senior's steps and according to the plan, but this time the senior only gave me a global idea, and all the specific planning and

implementation are planned by myself. This is a very precious exercise for me, and it will definitely become a highlight in my life.

I develop corresponding best practices and solutions in accordance with the three main stages of the internship program: At the beginning of the project, the environment preparation work was carried out, and some existing tools on the market were investigated, adhering to the idea of "If you want to be good at work, you must first sharpen your tools", lays a foundation for continuous and efficient promotion in the following practice. In the part of target recognition, I first conducted a survey of common neural networks and tried to build an operating environment. Then I customized the model to lay a good foundation for the training of customized data sets. After the acquisition of the dataset is completed, a series of data processing such as dataset collection (shooting), augmentation and annotation will be conducted respectively. This not only enables me to have a deeper understanding of dataset, but also enables me to better master and apply the courses related to "digital image processing" and "computer vision" that I learned in my undergraduate study. In the part of model training, I made several attempts with different methods and data, recorded necessary experimental data, and finally verified the model and improved the functions of the sub-system.

Although I only had a short week, I started with a classic problem. From discovering problems, researching materials and solving problems, I trained my basic qualities as a researcher and laid a good foundation for my future study and research career. At the same time, this internship experience also let me see some shortcomings of my own, such as the speed of literature reading and the ability of writing article need to be improved. I think if I have the honor to enter the laboratory and engage in research, I will make great efforts in this aspect. Undergraduate education is more of a general education, which enables us to get in touch with different fields, find out our interests and have a vague understanding of the subject. Different from master's and doctor's degree, it is more important to study in a certain field and add our own innovative ideas, hoping to make a difference in a certain field. I cherish the valuable internship experience and the opportunity to join our lab very much. I believe that under the guidance of the teacher and with my efforts, I will surely achieve good results. I also believe that the computer and design related industries will have more vigorous vitality in the future.

Finally, I would like to thank the teachers and seniors for your patience reading again. Thank you very much for your help during this time.

Appendix

【Study Notes】

1. opencv environment configuration under ubuntu
2. Usage of labelImg Tool
3. Data Augmentation

【Daily Report】

4. Intern Daily Report 7-23
5. Intern Daily Report 7-24
6. Intern Daily Report 7-25
7. Intern Daily Report 7-26