Efficient Data Selection with YOLOv5 Deep Learning Model

Hosik Choi* and DaeEun Kim†

Abstract – The state of the art deep learning models such as YOLOv5 have been applied to a wide research field including object detection. The YOLOv5 trains a relatively large size of neural networks and it needs much training time and lots of datasets. In this paper, we propose a strategy of efficient data selection to reduce efforts of data acquisition. The data selection method can be a useful tool for deep learning on a large amount of datasets. If there is no information of data characteristics for training data, all types of datasets should be acquired to train neural networks. We tested YOLOv5 deep learning for a variety of datasets including six different types of custom image datasets, and investigated how a portion of image dataset for a given type can affect the performance of object detection for another type. From the results, we can choose a proper set of training data with a relatively small loss of performance.

Keywords: Deep Learning, Object Detection, Data Selection

1. Introduction

Deep learning models are widely used in applications detecting labeled objects. YOLOv5, state-of-the-art deep learning model as an object detection model, provides high detection performance [1,2,3], and MobileNet with SSD reduces much computing cost, when compared with a large size of neural models [4,5]. These deep learning models consist of a lot of layers to extract feature maps, and need a vast amount of training data. However, acquiring those training data could face a resource problem of time and financial fund. In fact, there have been issues for efficient data selection in machine learning [6,7,8].

In this paper, we provide a strategy of efficient training data selection while the performance of object detection in image database is maintained. Also, we can build the relation graph among a collection of datasets with test performances.

2. Proposed Methods

This paper aims to establish an efficient strategy of training data selection for object detection over image data to reduce computing time or regulate the amount of input data needed for training. The problem is that a large amount of data is generally required when various environments are tested or when we wish to learn image data of distorted images or test images under varying illumination conditions.

In the experiment, we use eight types of datasets shown in Fig. 1; (a)-(b) two preset datasets (open-access public data) and (c)-(h) six custom image datasets (a collection of image datasets acquired manually). We define object classes with person, bicycle and car (PBC) for deep learning model; they are designed for the safety of pedestrians on the road.

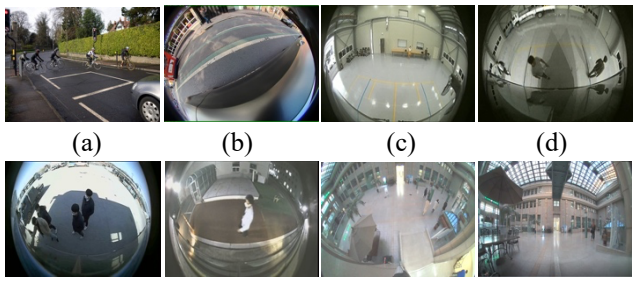

Fig. 1. Examples of preset and custom datasets (a) COCO (b) Woodscape (c) Indoor Day (d) Indoor Night (e) Outdoor Day (f) Outdoor Night (g) High-Angle (h) Low-Angle

The COCO dataset has 71,103 images of PBC and the Woodscape dataset 11,249 distorted images. The custom datasets are acquired with fish-eye cameras with a wide angle view. The pictures were taken in the indoor or outdoor environments during the day or the night. Also, another set of images were collected at two different height positions of a fish-eye camera, called High-Angle view and Low-Angle view. The datasets (c) – (h) have 845, 391, 2515, 2565, 1000 and 1000 images, respectively. Each dataset has its own environmental characteristics of illumination, obstacles, camera altitude and distortion.

Neural network models tend to depend on what sort of training data are given. All training sets include the preset

† Corresponding Author : Dept. of Electrical and Electronic Engineering, Yonsei Univerity, Korea (daeeun@yonsei.ac.kr)

* Dept. of Electrical and Electronic Engineering, Yonsei University, Korea (hosikchoi@yonsei.ac.kr)

dataset of COCO and Woodscape as a base set. The preset dataset has a large number of images, compared to the custom datasets. The features of the datasets can be analyzed by evaluating the test performances

Each of six custom datasets has its own training set and test set, respectively. We allow a set of options for training a custom dataset. A choice for training is given 20%, 40% 60%, 80% and 100% of the whole training set for a custom dataset. For every model-training experiment, the preset data consisting of Coco and Woodscape are always included with the above choice. We try to find what percentage of data would be sufficient to model a given dataset. From the result, efficient data selection can be achieved.

3. Experimental Results

You Only Look Once (YOLO) is a representative single-step object detection algorithm. The advantage of the model is to see the entire image once. The latest model YOLOv5 allows object detection in real time with high performance. We use YOLOv5s in the Pytorch framework for low computing cost, which is a small version of YOLOv5 model. The performance is measured with Mean Average Precision (mAP) of 0.5 Intersection over Union (IoU) threshold, which is often used in deep learning models. The network training was conducted using AMD Rysen 9 3900X CPU and two RTX 2080TI SLI GPUs.

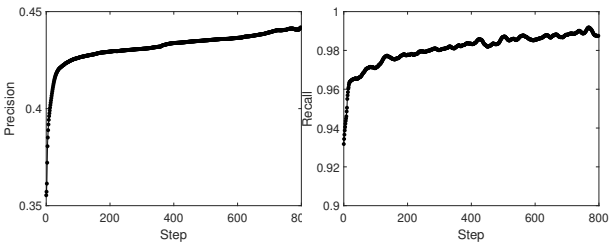

Fig. 2. Precision and recall performance for the preset dataset with 800 iterations of epoch

To determine the parameter of train iterations of epoch, we monitored the transition of performance as the iteration increased. For the purpose, neural networks learned the preset data of COCO and Woodscape. The recall performance was closely saturated with about 50-100 iterations as shown in Fig. 2. From the curve, the train epoch is set to 100 iterations for the upcoming experiments.

Possibly the point with the minimum value at the growth acceleration curve may be a right selection in transition curve, which is close to the saturation level of growth. The idea can be applied to the curve of test performances.

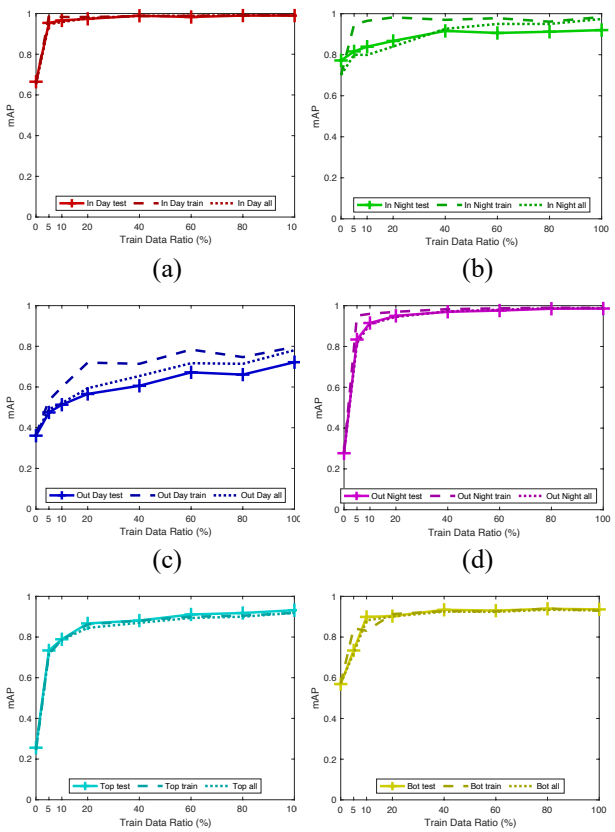

Fig. 3. Performance mAP with varying portions of the full training set (a) Indoor Day (b) Indoor Night (c) Outdoor Day (d) Outdoor Night (e) High-Angle (f) Low-Angle

Fig. 3 shows the performance results with mAP for each custom dataset, where varying portions of the dataset were used for training. From the curves, about 10% of the full training set seems to be sufficient to learn the characteristic or trend of pattern of the dataset except Outdoor-Day dataset, while more than 10% of the set has similar performances. A relatively small portion of the dataset can catch the trend of the dataset. It is believed that those datasets have a homogeneous property to keep consistency in the data patterns. The feature patterns of images within a custom dataset may be analogous, or the training model catches easily the fundamentals with the small portion.

In contrast, it seems that the Outdoor-Day dataset includes the data that the neural networks take more time to learn. Probably, a heterogeneous type of data could be abundant with the Outdoor-Day set. The pictures were taken from the camera of a driving bus outside. The set includes images distorted by fish-eye lenses and dynamic background images from the moving frame as well as a crowd of pedestrians hidden by neighbor objects and distorted images. It provides a hint that training dynamic environmental datasets becomes more difficult than clean stable image datasets. For the environment of indoor datasets (see Fig. 3(a)-(b)), a constant level of illuminations

are observed within a given dataset and thus objects are well distinguished. These types of image datasets seem to be prevalent in the preset dataset.

After learning the full training set for each custom dataset, the neural networks were tested on the preset data. Table 1 shows the performances and there is no much difference among them. Learning the custom dataset does not influence much the preset learning. We evaluated the object detection performances on the test set of the custom datasets after learning each choice of percentage portion (20%, 40%, 60%, 80% and 100%) of the full training set for each custom dataset. Fig. 4 shows the increments of mAP performance based on each choice of the training set from mAP performance based on the preset learning. For example, when 20% of the full set of Indoor-Day dataset plus the preset was trained, the learned neural networks were evaluated on the test set of Indoor-Day dataset. We obtained 30.9% increase in the test performance from the performance result after learning only the preset dataset.

Table 1. Test performances on the preset dataset after learning the full training set for each custom datatset

Train Data	Preset Only	In Day	In Night	Out Day	Out Night	High-Ang	Low-Ang
mAP	0.79	0.77	0.79	0.79	0.78	0.79	0.77
Precision	0.43	0.43	0.43	0.43	0.43	0.43	0.42
Recall	0.85	0.83	0.84	0.85	0.84	0.85	0.83

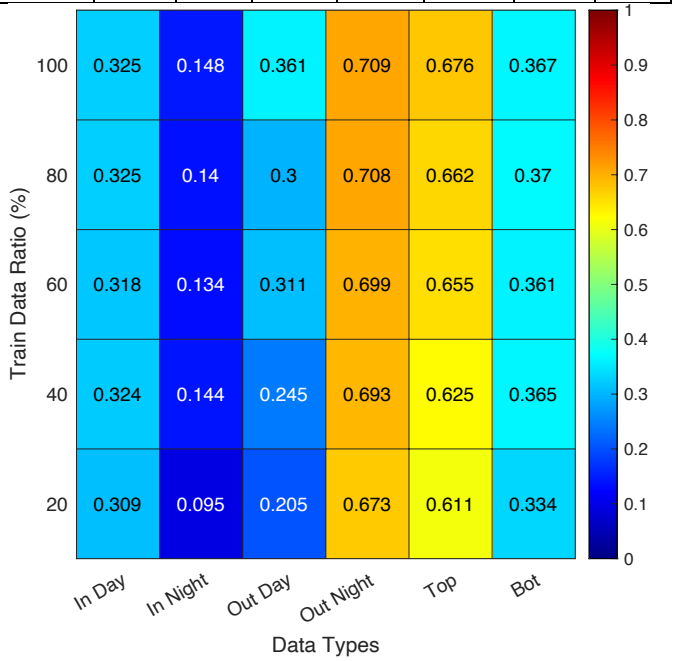

Fig. 4. mAP increment heatmap of the test performances (the performance after each choice of learning is compared to that after learning only preset data)

The Outdoor-Night dataset and High-Angle dataset has large increments and it implies that features in the two datasets were easily learned. The environment for Outdoor-Night dataset has poor illuminations and objects are not distinguished from dim background. However, only a small portion of the dataset can be used to support learning. The suggested approach is effective on these kinds of image datasets, whose images are not easily available in the preset data. Low-Angle dataset includes warped images from fish-eye lens, but has a view similar to the camera view common in the preset data. It is notable that warped or distorted images are easily trained with a small portion of training data. Fig. 5 shows the relationship among the custom datasets. We obtained the learning effects from the average of the performance on the test set of other custom datasets, after learning each choice of percentage portions of the training set of custom datasets. The percentage shows the change of performance compared to that after learning only preset data. It shows that there is relationship depending on the characteristic of the dataset.

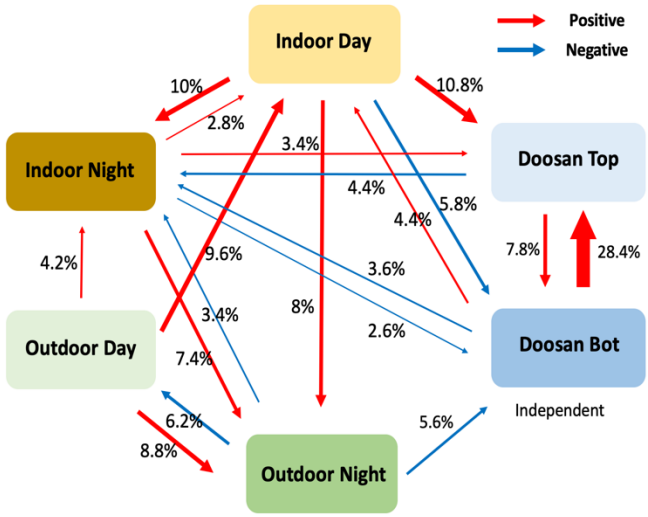

Fig. 5. Relationship of learning effects among the custom datasets (each learned custom dataset is tested on the other datasets and there is no arrow in the independent relationship)

4. Conclusion

In this paper, we investigated the training effects of preset and six different custom datasets by taking a portion of the full training set for learning. The preset neural network model was not sufficient to reflect a new environmental dataset. However, according to the experimental results, a small portion of target dataset, only 5% or 10% can help learning a new environmental

condition, a distortion of images or a different camera view.

Clean images or stable camera setting can be easily adapted only with the preset dataset including COCO and Woodscape data. For object detection with a new type of image dataset, we suggest a strategy of using open-access public image dataset plus a small fraction of image dataset acquired newly, if the image set has homogeneous illumination patterns with a fixed camera view. Based on this idea, dynamic environmental appearances may be decomposed into a subset of images under homogeneous view patterns. We need further study of how various patterns of images can be factorized into easy learning sets.

We can test how a portion of image dataset for a given type can affect the performance of object detection for another type. From the relation, we can build a network graph among a collection of image types and we can infer how much one type of data is close to another type. Thus, redundant datasets can be excluded for training, and efficient data selection organizes the training set to save the time and efforts.

Acknowledgements

This work was supported by the National Research Foundation of Korea through the Korean Government (MSIT) under Grant (NRF-2020R1A2B5B01002395).

References

[1] REDMON, Joseph, et al. You only look once: Unified, real-time object detection. In: Proceedings of the IEEE conference on computer vision and pattern recognition. 2016. p. 779-788.

[2] HUANG, Rachel; PEDOEEM, Jonathan; CHEN, Cuixian. YOLO-LITE: a real-time object detection algorithm optimized for non-GPU computers. In: 2018 IEEE International Conference on Big Data (Big Data). IEEE, 2018. p. 2503-2510.

[3] SHAFIEE, Mohammad Javad, et al. Fast YOLO: A fast you only look once system for real-time embedded object detection in video. arXiv preprint arXiv:1709.05943, 2017.

[4] LIU, Wei, et al. Ssd: Single shot multibox detector. In: European conference on computer vision. Springer, Cham, 2016. p. 21-37.

[5] CHIU, Yu-Chen, et al. Mobilenet-SSDv2: an improved object detection model for embedded systems. In: 2020 International conference on system science and engineering (ICSSE). IEEE, 2020. p. 1-5.

[6] ZHU, Xiangxin, et al. Do we need more training data?. International Journal of Computer Vision, 2016, 119.1: 76-92.

[7] MOORE, Robert C.; LEWIS, William. Intelligent selection of language model training data. In: Proceedings of the ACL 2010 conference short papers. 2010. p. 220-224.

[8] MEHRYARY, Farrokh, et al. Deep learning with minimal training data: TurkuNLP entry in the BioNLP shared task 2016. In: Proceedings of the 4th BioNLP shared task workshop. 2016. p. 73-81.