Selective Feature Learning with Filtering Out Noisy Objects in Background Images

Soonyong Song, Heechul Bae, Hyonyoung Han and Youngsung Son
Smart ICT Convergence Research Department
Electronics and Telecommunications Research Institute (ETRI)
Daejeon, South Korea
{soony, hessed, hyonyoung.han, ysson}@etri.re.kr

***Abstract*—In this competition, we propose a selective feature learning method to eliminate irrelevant objects in target images. We applied a Single Shot multibox Detection (SSD) algorithm to select desired objects. The SSD algorithm alleviates performance degradation by noisy objects. We trained SSD weights with annotated images in task 1. The refined dataset is fed into a traditional MobileNet classification network. We summarize our next research points through this competition in future works.**

***Index Terms*—Selective Feature Learning, Single Shot multibox Detection**

I. Introduction

Cameras installed in robots record images around them continuously. Vision processing techniques based on deep neural networks extract meaningful information from the recorded data. The extracted information is utilized to recognize contextual objects in arbitrary environments. Naturally, most of the environments around the robots change their states. Images recorded by the cameras might vary dynamically while the robots are operated even if shapes of target objects did not change. Thus vision processing units with deep neural networks should update their weights against those data variations. The updated weights keep object recognition performances in response to new training dataset. In case of sequential training, however, overall recognition performances of vision processing units probably decrease since old memories in the weights are overwritten by brand-new memories [1]. To protect old memory from the new training dataset, recent frontiers who research phenomena for catastrophic forgetting in deep neural networks were proposed weight regularization methods [2] and self-expandable networks [3] - [5]. Those of researches tried to transfer object features during continual learning.

They analyze final dataset to design object recognition software, and find that target objects in the dataset coexist with unlabeled objects, such as bags, clocks, milk cartons and paper clips. Therefore, they propose a selective feature learning method by eliminating irrelevant features in training dataset. The selective learning procedure is as follows: 1) extracting target objects from training dataset by an object detection algorithm, 2) feeding the refined dataset into a deep neural network to predict labels. In their software, they applied to a SSD as the object detection algorithm due to convenience of flexible feature network design and proper detection performances. We trained the SSD model with human-annotated dataset in task 1, and converted the SSD model to a frozen graph to infer object location in feeding images. The refined dataset was resized by 224x224 resolution since our prediction neural network requires a specified input image size.

II. Dataset Analysis

Before designing software, we analyzed characteristics of the provided final dataset [8], [9] in order to figure out proper object prediction mechanism. Basically, the provided data set of each task were taken in different environment conditions, such as illumination, occlusion, pixel shifting and clutter. Also each 69 objects had different sizes and backgrounds. Details of object region and position are as follows.

- Region of objects (relative scale): Median relative size is 0.142 and relative size difference is 4.14.
- Position of objects (relative scale): Center of object is greater than 0.2 and smaller than 0.8.

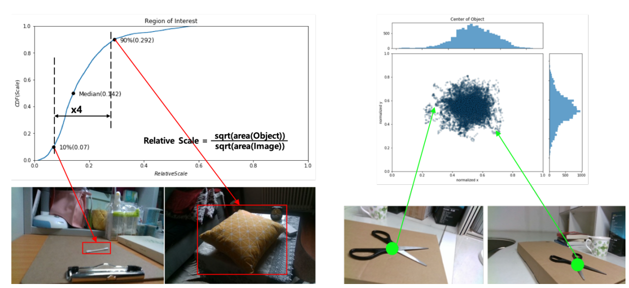

Fig. 1. Region of interest analysis for the dataset

Environments located in the labeled objects have such characteristics, 1) coexistence of interferential objects around the labeled objects, 2) diminishing native feature for the labeled objects in source images by variation of lights and masking portion. In this competition, we focused on the first one. Thus we setted up our goal as reducing to transfer noisy features from other unlabeled objects. We modified preprocessing step to obtain correct objectives from the provided images at the next section.

III. Software Design

Based on the previous analysis, we noticed that target objects coexist other unlabeled objects in the training images.

Therefore, to train correct target object selectively, a proposed software applied an SSD algorithm that is one of object detection methods. Reasons selecting the SSD algorithm are as follows.

- Customizing feature extraction networks
- Excellent object detection performance for its processing time
- Single pipeline of deep neural network

The SSD algorithm was developed by modifying examples in tensorflow open source [7]. We applied MobileNet version 2 as a backbone network of the feature extraction network in SSD algorithm. The backbone network model was pretrained with coco dataset. We retrained the backbone network model using our handcrafted annotation data.

The extracted object images were resized by 224x224 resolution, and were supplied to an image classification network. The MobileNet version 1 was used as the image classification network since its validation loss was converged faster than other networks in our experiments with a dataset opened at the first phase. After finishing sequential training for task 12, trained model and weight were saved files with json and h5 extension respectively. We programmed our software using python 3.6 interpreter with tensorflow version 2.0.x and keras version 2.3.x.

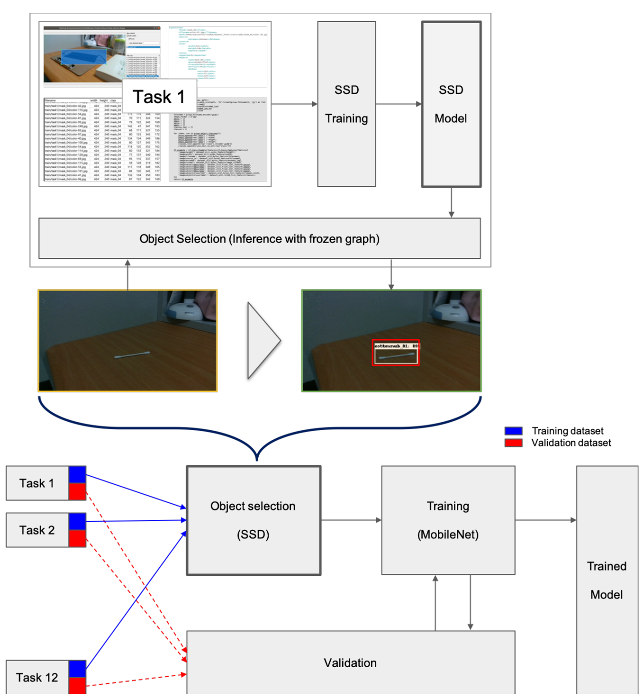

Fig. 2. Software architecture for selective feature learning

IV. Future Works

Our submitted software is a draft version of lifelong object recognition. Thus this project needs complementary works.

- Lifelong learning of object selection inference graph: Our SSD model is only trained with task 1. This model has insufficient features since its weights do not have any information for different directions. Currently, our software is implemented to infer object locations with frozen graph. In the future, the software should be modified to adopt sequential training procedures to trace environment variations.
- Lifelong learning of feature extraction network: In the sequential training, weights in a feature extraction network memorize additional information. In some case, additional neurons and connections might be needed in order to save new features. Current SSD models use feature extraction networks with fixed size. Therefore dynamic expanded networks should be devised to train SSD model sequentially.
- Single pipeline for object recognition: Basically the SSD model includes image classification network. Thus, the object detection and training block in Fig. 2 could be integrated to single pipeline.

Acknowledgment

This work was supported by Electronics and Telecommunications Research Institute (ETRI) grant funded by the Korean government. [19ZH1100, Distributed Intelligence Core Technology of Hyper-Connected Space]

References

[1] German I. Parisi, Ronald Kemker, Jose L. Part, Christopher Kanan, Stefan Wermter, "Continual lifelong learning with neural networks: A review", Neural Networks, vol. 113, May 2019, pp.54–71.

[2] Kirkpatrick J., Pascanu R., Rabinowitz N., Veness J., Desjardins G., Rusu A.A., et al. "Overcoming catastrophic forgetting in neural networks", Proceedings of the National Academy of Sciences, vol. 114, no. 13, 2017, pp. 3521–3526

[3] Jaehong Yoon, Eunho Yang, Jeongtae Lee, Sung Ju Hwang, "Lifelong Learning with Dynamically Expandable Networks", arXiv preprint arXiv:1708.01547, 2017.

[4] J Xu, Z Zhu, "Reinforced Continual Learning", Advances in Neural Information Processing Systems 31 (NeurIPS), 2018, pp. 899–908.

[5] Xu, Ju, Jin Ma, and Zhanxing Zhu. "Bayesian Optimized Continual Learning with Attention Mechanism." arXiv preprint arXiv:1905.03980, 2019.

[6] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, Alexander C. Berg, "SSD: Single Shot MultiBox Detector", European Conference on Computer Vision (ECCV), September 2016, pp 21–37.

[7] https://github.com/tensorflow/models/

[8] Q. She, F. Feng, X. Hao, Q. Yang, C. Lan, V. Lomonaco, X. Shi, Z. Wang, Y. Guo, Y. Zhang, F. Qiao, and R. H. M. Chan, "Openloris-object: A robotic vision dataset and benchmark for lifelong deep learning," 2019.

[9] F. Feng, R. H. M. Chan, X. Shi, Y. Zhang, and Q. She, "Challenges in task incremental learning for assistive robotics," *IEEE Access*, vol. 8, pp. 3434–3441, 2020.