De-centralized Self-Training with Class-Wise Weight Aggregation

Ali Haroon Turk, Odilbek Urmonov, Zafar Aziz, HyungWon Kim*
Chungbuk National University.

Abstract

Aggregation strategies in distributed training are key in training well-generalized object detectors. The available aggregation strategies aggregate the weights for all classes even though each detector is trained in a different domain. Aggregating weights from non-identical domains introduces bias in weights. Biased aggregation of these weights leads to poor or incorrect features of the minority class samples in the data. Hence, the resulting global detector suffers from a confirmation bias problem and performs poorly on local data. We propose a simple class-wise weight selection strategy before aggregation. Our proposed strategy aims to improve the performance of the detector for minority classes. The resultant weights learn better features of the minority classes and generalize well.

Ⅰ. Introduction

Object detection has been broadly explored for various computer vision applications, including autonomous driving[1], visual information for scene understanding[2], and industrial inspection[3].

Object detectors are mainly trained in supervised training methodology[4] requiring labeled training data. Labeled training data is scarce due to the expensive and time-consuming data annotation step. In contrast, unlabeled data is easily and abundantly available. Therefore, the research community is more focused on semi-supervised[5] training methods.

In conventional semi-supervised training, detectors are trained in a centralized setup. In centralized learning, detectors have access to sensitive training data. Distributed training[6] addresses the data privacy problem by accessing only the weights of the clients. In this work, we propose to combine distributed training with semi-supervised learning.

Most clients are often trained under the I.I.D assumption, which assumes that the train and test data are independently and identically distributed[7]. This assumption is not always true. In distributed training, some clients are trained on data collected from rural areas, some are trained on data collected from crowded areas and some are trained on data collected from highways. Based on this, some detectors are good at detecting some object classes due to the availability of more samples belonging to those classes in the training data. Aggregation based on conventional methods introduces a bias to the global model weights[8]. Biased aggregation results in misleading representation of minority class features. Hence the aggregated global model underperforms on local data.

In this study, our key idea is to extend the existing aggregation schemes[6] by combining them with a simple class-wise weight selection strategy. We evaluate the weight parameters of each client for various validation sets available on the server. Based on this evaluation, the best weights are prioritized for each class for aggregation.

Ⅱ. Proposed Method

We propose a simple method, where the server evaluates the weights of each client on every validation dataset available on the server side. Based on these results, model weights are prioritized accordingly before aggregation.

The proposed methodology is shown in Fig. 1. We provide details of the proposed method for class-wise aggregation in the following sections.

A. Training on Client Side

In the first step, unlabeled data is collected from the current client domain. In the next step, pseudo-labels for the unlabeled data are generated using the server global model. Finally, each client is trained using labeled and pseudo-labeled data following the Mean Teacher framework[9].

B. Uploading Weight Parameters

Once the training on the client side is complete, the client sends the best weights to the server for aggregation. The server waits for all the clients to send their weights, once server receives all the weights, server proceeds to the next step.

C. Class-wise Weight Selection and Aggregation

Before running the federated averaging algorithm, the server will evaluate the weights of each client on the available validation dataset on the server side. Based on this evaluation, the best weights for each class are

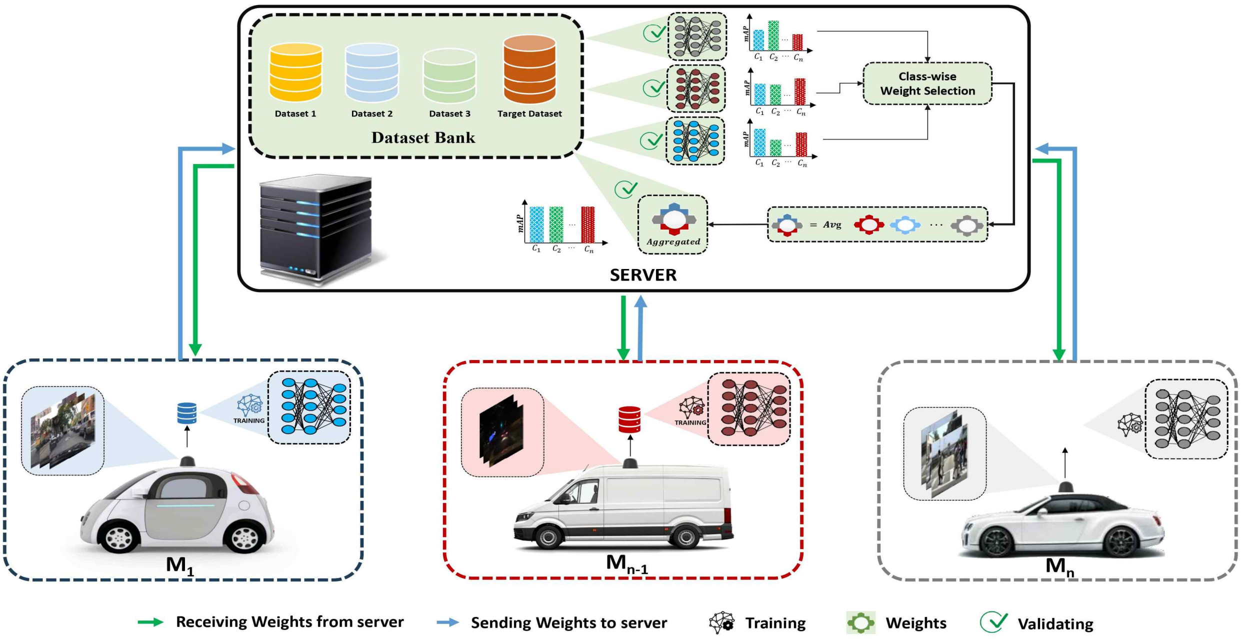

Fig. 1: Proposed methodology

prioritized for aggregation. The best weights for each class are aggregated using class-wise weighted aggregation. These aggregated weights are then validated on the target dataset, which is difficult as compared to habitual datasets.

After the aggregation process is complete, each client receives updated weights. The client will use these weights to generate new pseudo-labels and resume the training process.

Ⅲ. Conclusion

In this study, we propose decentralized training combined with semi-supervised learning and a class-wise weight selection mechanism. In our method, aggregated weights are prioritized for respective classes. The resulting aggregated global weights result in better features of the minority classes without affecting the performance of the majority classes.

ACKNOWLEDGMENT

This work was supported by Regional Leading Research Center (RLRC) of the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) (No. 2022R1A5A8026986) and supported by Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No.2020-0-01304, Development of Self-Learnable Mobile Recursive Neural Network Processor Technology). It was also supported by the MSIT (Ministry of Science and ICT), Korea, under the Grand Information Communication Technology Research Center support program (IITP-2024-2020-0-01462) supervised by the IITP (Institute or Information & communications Technology Planning & Evaluation).

References

[1] A. H. Khan, M. S. Nawaz, and A. Dengel, "Localized semantic feature mixers for efficient pedestrian detection in autonomous driving," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 5476–5485.

[2] Z. Zhang and M. Hoai, "Object detection with self-supervised scene adaptation," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 21589–21599.

[3] Z. Liu, Y. Zhou, Y. Xu, and Z. Wang, "Simplenet: A simple network for image anomaly detection and localization," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2023, pp. 20402–20411.

[4] Z. Shen, Z. Liu, J. Li, Y.-G. Jiang, Y. Chen, and X. Xue, "Dsod: Learning deeply supervised object detectors from scratch," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 1919–1927.

[5] W. Chen, L. Lin, S. Yang, D. Xie, S. Pu, and Y. Zhuang, "Self-supervised noisy label learning for source-free unsupervised domain adaptation," in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, IEEE, 2022, pp. 10185–10192. Accessed:

[6] O. Urmonov, S. Sajid, Z. Aziz, and H. Kim, "Federated Object Detection Scenarios for Intelligent Vehicles: Review, Case Studies, Experiments and Discussions," *IEEE Trans. Intell. Veh.*, 2024

[7] Q. Liu, L. Lin, Z. Shen, and Z. Yang, "Periodically exchange teacher-student for source-free object detection," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2023, pp. 6414–6424.

[8] S. P. Karimireddy, S. Kale, M. Mohri, S. Reddi, S. Stich, and A. T. Suresh, "Scaffold: Stochastic controlled averaging for federated learning," in *International conference on machine learning*, PMLR, 2020, pp. 5132–5143.

[9] A. Tarvainen and H. Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," *Adv. Neural Inf. Process. Syst.*, vol. 30, 2017