Computer Vision Final Report
Cross Domain Object Detection

NTNU CSIE 61147050S 林詠閎

NTNU CSIE 61147053S 何承祐

1. Introduction

With the rapid development of computer vision, object detection has been a fundamental and challenging task. The goal of object detection is to identify and localize specific objects in images or videos, providing a foundation for subsequent analysis and applications. However, traditional object detection methods often perform poorly when faced with variations across different domains, limiting their wide-ranging applicability in real-world scenarios.

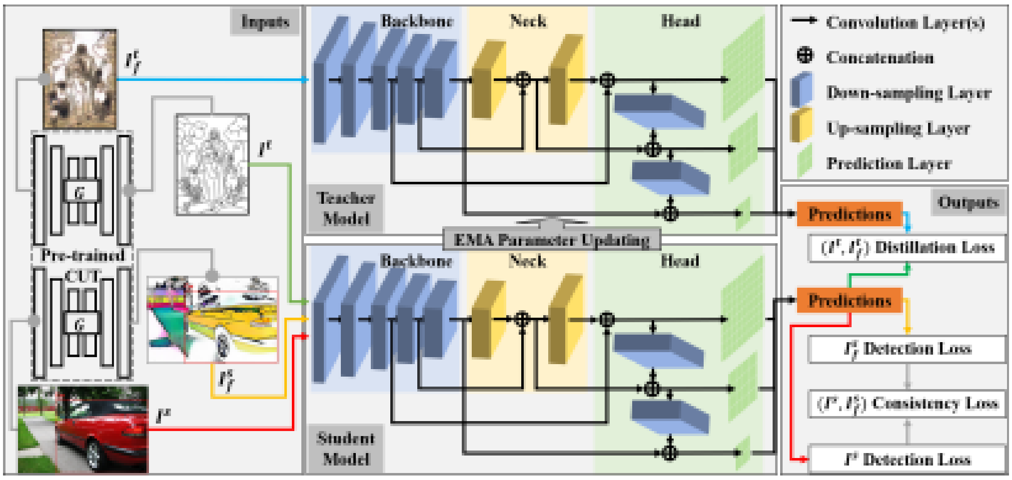
In recent years, domain adaptation in object detection has emerged as a significant research area. Images and videos from different domains exhibit variations in lighting conditions, backgrounds, and object appearances. Therefore, designing object detectors that can adapt to different domains has become a critical challenge. Previous research efforts have primarily focused on object detection within a single domain, with limited exploration of domain adaptation in object detection.

In this paper, we address the problem of domain adaptation in object detection and propose a novel method called SSDA-YOLO (Source-free Semi-Supervised Domain Adaptation YOLO). Our approach leverages the power of transfer learning and domain adaptation techniques to enable object detectors to achieve robust performance across diverse domains. We introduce a teacher-student framework, where the teacher model generates pseudo-labels for the target domain, and the student model is trained using these pseudo-labels. Furthermore, we introduce a consistency loss to encourage consistency between the teacher and student predictions. Through extensive experiments on benchmark datasets, we demonstrate the effectiveness of our approach in improving the domain adaptation performance of object detectors.

Overall, our work contributes to the field of domain adaptation in object detection by proposing a novel method that addresses the challenges of adapting object detectors to diverse domains. We believe that our research opens up new possibilities for deploying object detection models in real-world scenarios with domain shifts.

2. Method

In this section, we present the proposed method, SSDA-YOLO (Source-free Semi-Supervised Domain Adaptation YOLO), for domain adaptation in object detection. Our method leverages transfer learning and domain adaptation techniques to enable object detectors to achieve robust performance across diverse domains.



2.1.Contrastive Unsupervised Transfer learning

CUT's primary function is to learn a shared feature space that aligns the source and target domains. It does this by maximizing the similarity between images from the same class while minimizing the similarity between images from different classes. By training the object detector in this shared feature space, it becomes robust to domain shifts and can effectively detect objects in the target domain.

2.2.Teacher-Student Framework

Our approach adopts a teacher-student framework to facilitate domain adaptation in object detection. The teacher model is pretrained on a large-scale labeled dataset from the source domain, while the student model is trained on the target domain using pseudo-labels generated by the teacher model. This framework allows the student model to benefit from the knowledge transferred by the teacher model, improving its performance in the target domain.

2.3.Pseudo-Label Generation

To generate pseudo-labels for the target domain, the teacher model applies object detection inference on unlabeled target images. The resulting detections are treated as pseudo-labels and used to train the student model. This strategy allows the student model to learn from unlabeled target data and adapt to the target domain.

2.4. Consistency Loss

To encourage consistency between the teacher and student predictions, we introduce a consistency loss term in the training process. The consistency loss penalizes discrepancies between the predictions of the teacher and student models, encouraging them to produce similar outputs. This promotes alignment between the source and target domains and enhances the generalization ability of the student model.

2.5. Training Procedure

During training, we alternate between two steps: teacher model updating and student model updating. In the teacher model updating step, we optimize the teacher model parameters using the labeled source domain data. In the student model updating step, we optimize the student model parameters using the pseudo-labeled target domain data. This alternating training procedure allows the student model to gradually adapt to the target domain while leveraging the knowledge transferred by the teacher model.

2.6. Domain Adaptation Evaluation

We evaluate the domain adaptation performance of our method on benchmark datasets by measuring the detection accuracy and localization precision. We compare the performance of our SSDA-YOLO with baselines and state-of-the-art domain adaptation methods to demonstrate its effectiveness in improving object detection performance across diverse domains.

3.Experiments

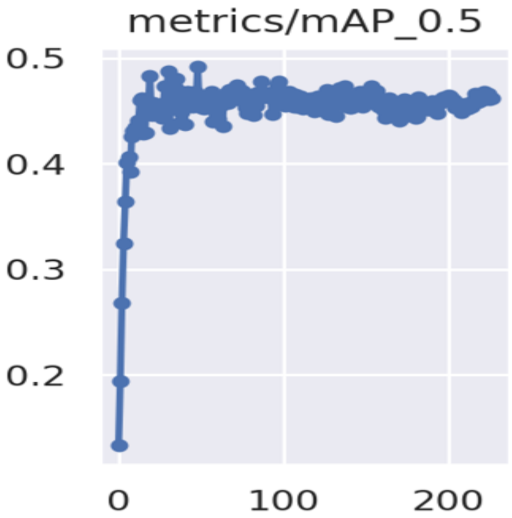
3.1. Dataset

We used the original->foggy data on CityScapes and CityScapes Foggy Dataset. The original training data consists of 2575 images, and the validation data includes 300 images. The foggy data follows the same setup as mentioned earlier.

3.2. Training result on source data

First, We train the yolo v5 medium size model on source data, evaluate the model on source data, the table show the mAP.

	map	**map50**	**map75**	**map_s**	**map_m**	**map_l**
Yolov5m	0.2839	0.4499	0.2880	0.0246	0.1845	0.5053



3.3. Evaluate the model for foggy data

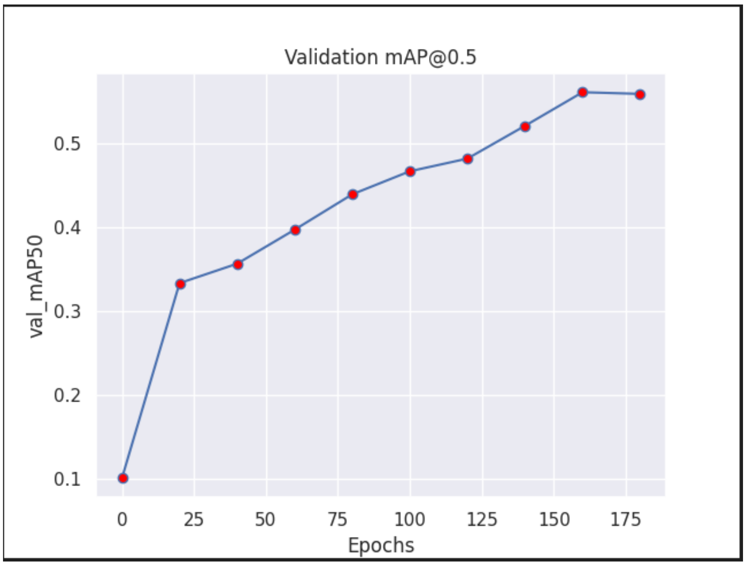
We finish the original training, then we evaluate it on foggy data, the results indicate that the model trained on the training set is not able to perform well in cross-domain scenarios.

	map	map_50	map_75	map_s	map_m	map_l
Yolov5m	0.1850	0.2866	0.1911	0.0034	0.0932	0.4064

3.4. Our Cross Domain Model

The final results are as follows: our model not only achieved cross-domain capability but also surpassed the original performance.

	map	map_50	map_75	map_s	map_m	map_l
ssda-yolov5m	0.3763	0.5860	0.3872	0.0267	0.2273	0.6610



4.Discussion and Conclusion

Due to hardware limitations, we were constrained to use only the yolov5m version for selecting a pre-trained model. Our approach utilized the mean teacher framework and showed potential for further improvement by exploring a semi-supervised architecture. We employed a generative style transfer model for pretraining, which allowed our model to effectively leverage cross-domain information. As a result, we observed significant benefits in the obtained outcomes.

5.Reference

[1] Zhou, Huayi and Jiang, Fei and Lu, Hongtao, "Semi-supervised Domain Adaptive YOLO for Cross-Domain Object Detection", arXiv preprint arXiv:2211.02213, 2022

[2] Taesung Park, Alexei A Efros, Richard Zhang, and JunYan Zhu, "Contrastive learning for unpaired image-toimage translation," in ECCV. Springer, 2020

*Subtask:
林詠閎:report,ppt,ssda-yolo training
何承祐:presentation,ppt,CUT training