Domain specific network for Multi-source domain adaptation

Phat Do Dinh[O], Junhwan Han, Wonjun Hwang
Department of Artificial Intelligence, Ajou University
{phatai, ase2015, wjhwang}@ajou.ac.kr

Summary

In this paper, the method gets knowledge from a multi-source domain and adapts to work well at the target domain. Our network includes three subnets that correspond to two source domains and a target domain. Each source domain subnet is aligned to adapt with target and other source domains by a shared pixel discriminator, nonshared high discriminator. The network has a shared classification and regression head to predict the label and location of each object in image. By using separative features extractor to get information from each image, the network learns more knowledge when trained. Then the network approaches a good accuracy at cross time adaptation. My code is available at https://github.com/EdwardDo69/DSN_GITHUB.

1. Introduction

Computer vision achieves many great approaches by deep learning in image classification, object detection, segmentation, etc. However, if we want deep learning methods to learn the knowledge from images, we must collect a lot of data and annotate them. This work is very expensive in terms of time and money, so unsupervised domain adaptation (UDA) [1,2,3] is receiving much attention. UDA learns knowledge from a labeled dataset (source domain) and transfers it to work well at an unlabeled dataset (target domain). In UDA, domain adaptive object detection (DAOD) [5,1,6] is more challenging. Compared with classification work, DAOD not only predicts labels of objects but also locates the object in images. It is more difficult because there is a domain shift between source and target domain.

Multi-source domain adaptation (MSDA) is more practical for real world scenarios because in the real world, there are many labeled sources. MSDA deals with the gap between source domain and target domain and domain shift among source domains. The first paper MSDA for object detection is DMSN [4]. DMSN shares the first part of the feature extractor and divides the second part into three subnets. Each subnet corresponds with source domains and a target domain. Each source subnet is trained with each source data and the target subnet is made by assembling source subnets. However, the paper has a major limitation: the method has only one first part of the feature extractor. But the extractor has to extract features from many sources (domain sources and target source). When the number of domains increases, the efficiency of the extractor decreases.

To solve the issue, we propose the method Domain specific network for multi-source domain adaptation (DSN). Each network corresponds with one domain. Source subnets are trained with source domains and the target subnet is assembled from source subnets parameters. A subnets model like Faster R-CNN [7] is a good method for both regression and classification: an extractor to get features from images, a region proposal network (RPN), etc. The network has a shared pixel level discriminator and classification, regression prediction head. They reduce domain shift between source domains. In source subnets, the high discriminator classifies the domain that the image is from, which makes each domain closer to the target domain. Our contribution of this study are threefold:

- We propose a new network for multi-source DAOD, which uses specific subnets for each domain to reduce the domain gaps between sources and target domain and between source domains.
- We propose a new method which uses a shared pixel discriminator to align separate networks.
- Our method approaches a better accuracy than the previous work at cross time adaptation.

2. Proposed Method

The architecture includes three subnets: Source Subnet i, Source Subnet j and Target Subnet G^{S_i}, G^{S_j},G^T corresponding to three domains: source domain S_i, S_j, target domain T as shown in Figure 1. Each subnet has two feature extractor $\{G_1, G_2\} \subset G$. We train the network with corresponding source domain and adapt the network to target domain by pixel discriminator and high discriminator.

We train the network 2 steps:

Step 1. We train Source Subnets with corresponding source domain and target domain like SW-DA[10]. After the images are extracted by G_1^S extractor, the feature is fed to GRL (gradient reversal layer) and Pixel Dis domain discriminator to implement adversarial learning. The output of Pixel Dis is a 3D matrix with width, height are the size of the input feature and depth is the number of domains. G_2^S input the output feature of G_1^S and feed the output to High Dis discriminator and RPN to predict classification and regression of objects in image.

$$L_{Dis} = L_{Pixel\ Dis} + L_{High\ Dis}$$

Where $L_{Pixel\ Dis}$, $L_{High\ Dis}$ are Pixel Dis, High Dis discriminator loss and L_{Dis} is the sum of all discriminator losses.

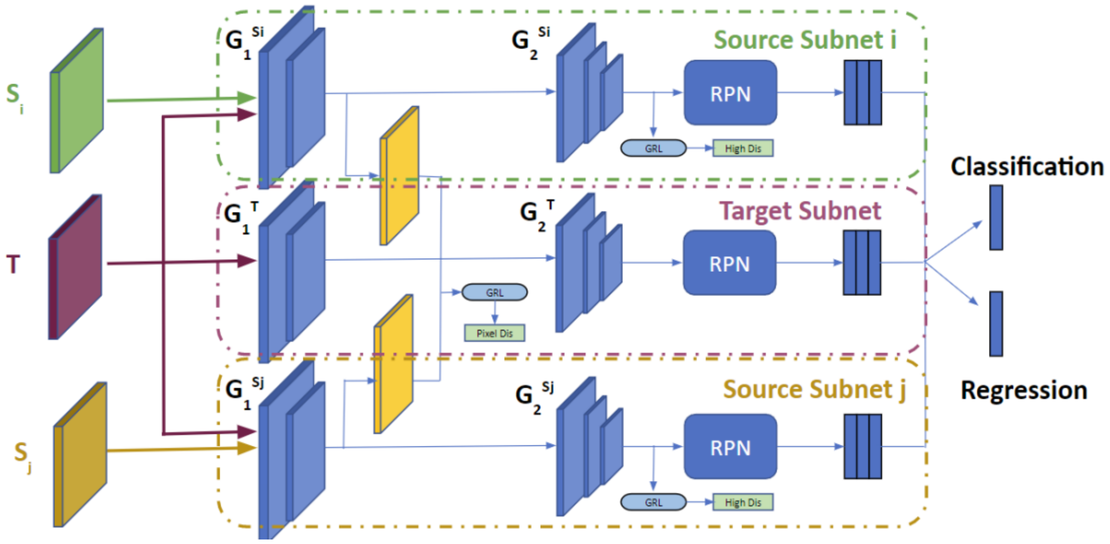

Figure 1: Framework overview of the DSN method.

Step 2. We assemble Target Subnet by sum parameters of Source subnets multiplied by corresponding relative score of this source and target. Relative score is the output value of High Dis discriminator. We also use the Consistency loss is the difference of RPN output between Source Subnets and Target Subnet to lead all sources closer.

$$P_t^T = \alpha P_{t-1}^T + (1-\alpha)(Sc_t^{S_i} P_t^{S_i} + Sc_t^{S_j} P_t^{S_j})$$

where *P* is parameters of subnet, Sc is relative score with target domain, training step *t* and smoothing coefficient parameter α.

3. Experiment

3.1 Experiment Setting

We experiment our method at cross time and cross domain scenarios. Cross time scenario is training with BDD100k dataset: *daytime* and *night* are the sources and *dawn/dusk* is the target domain. We train with Cityscapes and KITTI and test at BDD100k(*daytime*) as the cross camera scenario.

3.2 Cross time adaptation

Table 1. Results on cross time adaptation.

Methods	mAP
MDAN [8]	27.6
MSDA [9]	26.5
DMSN [5]	35.0
DAN (Ours)	**35.9**

As shown in Table 1, by dealing with the problem of previous work, our method gets more knowledge by non-shared feature extractor and approaches the best accuracy 35.9%. The experiment result shows that our method works at the cross time adaptation scenario.

3.3 Cross camera adaptation

Table 2. Results on cross camera adaptation.

Methods	AP on car
MDAN [8]	43.2
MSDA [9]	44.1
DMSN [5]	**49.2**
DAN (Ours)	<u>45.7</u>

In Table 2, our method approaches second best accuracy (underlined number) with 45.7%. We are aware that when increasing the received knowledge the network easily gets overfitting. Because cross camera adaptation only trains the network to predict car objects. When the network gets a lot of information about a car object the accuracy may get worse.

4. Conclusion

In this paper, we present a novel multi-source domain adaptation approach for object detection. To deal with the lack of knowledge when training at multi-source domains, we propose a domain specific network that includes individual networks for each domain. Through experiment, we verified the method works well at multi-source domain adaptation scenario.

Acknowledgement

This work was partially supported by the BK21 FOUR program of the National Research Foundation of Korea funded by the Ministry of Education (NRF5199991014 091) and NRF-2022R1A2C1091402.

References

[1] Yuhua Chen, Wen Li, Christos Sakaridis, Dengxin Dai, and Luc Van Gool. Domain adaptive faster r-cnn for object detection in the wild. In CVPR, 2018.

[2] Taekyung Kim, Minki Jeong, Seunghyeon Kim, Seokeon Choi, and Changick Kim. Diversify and match: A domain adaptive representation learning paradigm for object detection. In CVPR, 2019.

[3] Wanyi Li, Fuyu Li, Yongkang Luo, and Peng Wang. Deep domain adaptive object detection: a survey. arXiv preprint arXiv:2002.06797, 2020.

[4] Xingxu Yao, Sicheng Zhao, Pengfei Xu, and Jufeng Yang. Multi-source domain adaptation for object detection. In ICCV, 2021.

[5] Jinhong Deng, Wen Li, Yuhua Chen, and Lixin Duan. Unbiased mean teacher for cross-domain object detection. In CVPR, pages 4091–4101, 2021

[6] Yangtao Zheng, Di Huang, Songtao Liu, and Yunhong Wang. Cross-domain object detection through coarse-to-fine feature adaptation. In CVPR, pages 13763–13772, 2020.

[7] S. Ren, K. He, R. Girshick, and J. Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In NIPS, 2015.

[8] Han Zhao, Shanghang Zhang, Guanhang Wu, Jose M. F. ´ Moura, Joao Paulo Costeira, and Geoffrey J. Gordon. Adversarial multiple source domain adaptation. In NeurIPS, pages 8568–8579, 2018.

[9] Xingchao Peng, Qinxun Bai, Xide Xia, Zijun Huang, Kate Saenko, and Bo Wang. Moment matching for multi-source domain adaptation. In ICCV, pages 1406–1415, 2019.

[10] [37] Kuniaki Saito, Yoshitaka Ushiku, Tatsuya Harada, and Kate Saenko. Strong-weak distribution alignment for adaptive object detection. In CVPR, 2019.