Adaptation Across Extreme Variations using Unlabeled Bridges

Shuyang Dai[1*] Kihyuk Sohn[2] Yi-Hsuan Tsai[3] Lawrence Carin[1] Manmohan Chandraker[3,4]
[1]Duke University [2]Google [3]NEC Labs America [4]UC San Diego

Abstract

We tackle an unsupervised domain adaptation problem for which the domain discrepancy between labeled source and unlabeled target domains is large, due to many factors of inter- and intra-domain variation. We propose to decompose domain discrepancy into multiple but smaller, and thus easier to minimize, discrepancies by introducing unlabeled bridging domains that connect the source and target domains. We realize our proposed approach through an extension of the domain adversarial neural network with multiple discriminators, each of which accounts for reducing discrepancies between unlabeled (bridge, target) domains and a mix of all precedent domains including source.

1. Introduction

With advances in supervised deep learning, many vision problems have realized significant performance improvements [12, 15, 16, 8, 7, 13, 14, 3]. However, this success is strongly dependent on the existence of large-scale labeled data [5], often not available in practice. To address this challenge, unsupervised domain adaptation (UDA) [6, 4, 11, 10, 9, 17] has been proposed to improve the generalization ability of classifiers, using unlabeled data from the target domain. The core idea is to reduce the discrepancy metric [2, 1] between the two domains, measured by the domain discriminator [6] or MMD kernel [18] at certain representation of deep networks. Nevertheless, it could be difficult to model such dynamics when there are many factors of inter- and intra-domain variation applied to transform the source domain into the target domain.

In this work, we aim to solve unsupervised domain adaptation challenges whose domain discrepancy is large due to variation across the source and target domains. Figure 1 provides an illustrative example of adapting from labeled images of cars from the internet to recognize cars for surveillance applications at night. Two dominant factors, the perspective and illumination, make this a difficult adaptation task. As a step towards solving these problems, we introduce *unlabeled domain bridges* whose factors of variation are partially shared with the source domain, while the others are in common with the target domain. As in Figure 1, the domain on the bottom left shares a consistent lighting condition (day) with the source, while viewpoint is similar to that of the target domain. There also could be multiple bridging domains, such as the bottom right one whose lighting intensity is in the midst of the first bridging domain and the target domain.

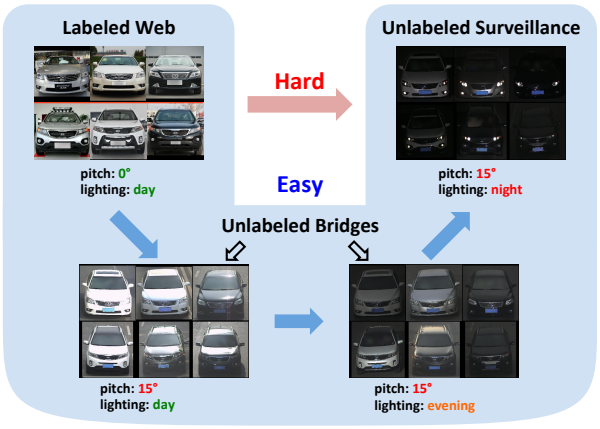

Figure 1. We introduce *bridging domains* composed of unlabeled images with some common factors to the source (e.g., lighting) and the target domain (e.g., viewpoint, image resolution).

To utilize unlabeled bridging domains, we propose to extend the domain adversarial neural network (DANN) [6] using multiple domain discriminators, each of which accounts for learning and reducing the discrepancy between unlabeled (bridging, target) domains and the mix of all precedent domains. We hypothesize that the decomposition of a single, large discrepancy into multiple, small ones leads to a series of easier optimization problems, culminating in better alignment of source and target domains.

2. Method

Our proposed domain adaptation framework is built atop DANN [6], which transfers a classifier learned from the labeled source domain $\mathcal{D}_S$ to the unlabeled target domain $\mathcal{D}_T$ by learning domain-invariant features. It uses a domain discriminator d to control the amount of domain-related information in the extracted feature. First, d needs to be well trained to tell the difference between source and target domains. In comparison, the feature extractor f wants to confuse the discriminator d to remove any domain-specific information. Moreover, to make sure the extracted feature is task-related, f is trained to generate features that can be correctly classified by a classifier C.

*This work is done when S. Dai was an intern at NEC Labs America.

In our proposed framework, we introduce additional sets of unlabeled examples, which we call bridging domains, that reside in the transformation pathway from labeled source to unlabeled target domains to guide adaptation process. Besides $\mathcal{D}_S$ and $\mathcal{D}_T$, we denote $\mathcal{D}_B$ as a bridging domain. Our framework is composed of feature extractor $f(x)$ from an input $x \in \mathcal{D}_S \cup \mathcal{D}_B \cup \mathcal{D}_T$ and classifier $C(f(x))$. Unlike the DANN that directly aligns $\mathcal{D}_S$ and $\mathcal{D}_T$, we decompose the adaptation into two as follows: First, $\mathcal{D}_S$ and $\mathcal{D}_B$ are aligned. This is an easier task than direct adaptation as in DANN, since there are less discriminating factors between $\mathcal{D}_S$ and $\mathcal{D}_B$. Second, we adapt $\mathcal{D}_T$ to the union of $\mathcal{D}_S$ and $\mathcal{D}_B$. Similarly, the task is easier since it needs to discover remaining factors between $\mathcal{D}_T$ and $\mathcal{D}_S$ or $\mathcal{D}_B$, as some factors are already found from the previous step. To accommodate two adaptation steps, we use two binary domain discriminators d_1 for learning discrepancy between $\mathcal{D}_S$ and $\mathcal{D}_B$ and d_2 between $\mathcal{D}_S \cup \mathcal{D}_B$ and $\mathcal{D}_T$. The objectives are:

$$\mathcal{L}_{d_1} = \mathbb{E}_{\mathcal{D}_S} \log d_1(f) + \mathbb{E}_{\mathcal{D}_B} \log(1 - d_1(f)) \quad (1)$$

$$\mathcal{L}_{d_2} = \mathbb{E}_{\mathcal{D}_S \cup \mathcal{D}_B} \log d_2(f) + \mathbb{E}_{\mathcal{D}_T} \log(1 - d_2(f)) \quad (2)$$

Both $\mathcal{L}_{d_1}$ and $\mathcal{L}_{d_2}$ are minimized to update their respective model parameters θ_{d_1} and θ_{d_2}. Once d_1 and d_2 are updated, we update the classifier according to the classification loss:

$$\mathcal{L}_C = \mathbb{E}_{(x,y)\sim \mathcal{D}_S \times \mathcal{Y}}[-y \log C(f(x))] \quad (3)$$

and the feature extractor to confuse discriminators as follows:

$$\min_{\theta_f, \theta_C} \mathcal{L}_C + \lambda_1 \mathcal{L}_{d_1} + \lambda_2 \mathcal{L}_{d_2} \quad (4)$$

where θ_C is the parameter for the classifier, and λ_1 and λ_2 are two hyperparameters to adjust the strengths of adversarial loss. We alternate updates between d_1, d_2 and f, C.

Our framework can be extended to the case for which multiple unlabeled bridging domains exist, desirable to span larger discrepancies between source and target domains. To formalize, we denote $\mathcal{D}_0 = \mathcal{D}_S, \mathcal{D}_{M+1} = \mathcal{D}_T$ as source and target domains, and $\mathcal{D}_m$, $m = 1, ..., M$ as unlabeled bridging domains with $\mathcal{D}_m$ closer to source than $\mathcal{D}_{m+1}$. We introduce $M+1$ domain discriminators $d_1, ..., d_{M+1}$, each of which is trained by maximizing the following objective:

$$\mathcal{L}_{d_m} = \mathbb{E}_{\bigcup_{i=0}^{m-1} \mathcal{D}_i} \log d_m(f) + \mathbb{E}_{\mathcal{D}_m} \log(1 - d_m(f)) \quad (5)$$

and the learning objective for f and C is given as follows:

$$\min_{\theta_f, \theta_C} \mathcal{L}_C + \textstyle\sum_{m=1}^{M+1} \lambda_m \mathcal{L}_{d_m}. \quad (6)$$

3. Experiments

We evaluate our methods on the CompCars dataset [19], which provides two sets of images: 1) the web-nature images are collected from car forums, public websites and search engines, and 2) the surveillance-nature images are collected from surveillance cameras. The dataset is composed of $52,083$ web images across 431 car models and $44,481$ SV images across 181 car models, with these categories of SV set being inclusive of 431 categories from web set. We use an illumination condition as a metric for adaptation difficulty and partition the SV set into SV1–5 based on the illumination condition of each image. SV1 contains the brightest images, thus is the easiest domain for adaptation, whereas SV5 contains the darkest ones, thus is the hardest to adapt. Sample web and SV images can be found in Figure 1.

Model	SV1–3	SV4–5
Web (source only)	72.67	19.87
Web→SV4–5	68.90±1.28	49.83±0.70
Web→SV4→5	74.03±0.71	61.37±0.30
Web→SV1–5	**83.29**±0.14	77.84±0.34
Web→SV1–3→4–5	**82.83**±0.40	**78.78**±0.33

Table 1. Accuracy and standard error over 5 runs on SV test sets for models with and without bridging domain.

Model	SV5
Web→SV5	37.83±0.51
Web→SV4→5	58.40±0.60
Web→SV1→5	69.69±0.99
Web→SV3→4→5	74.01±0.52
Web→SV2→3→4→5	75.15±0.18
Web→SV1→2→3→4→5	75.47±0.20

Table 2. Accuracy and standard error over 5 runs on SV5 test set for models with different bridging domain configurations.

We present two experimental protocols. First, we evaluate on an adaptation task from web to SV night (SV4–5) using SV day (SV1–3) as one domain bridge. We compare the following models in Table 1: baseline model trained on labeled web images, DANN from source to target (Web→SV4–5), from source to mixture of bridge and target (Web→SV1–5), and the proposed model from source to bridge to target (Web→SV1–3→SV4–5).

Second, we adapt to extreme SV domain (SV5) using different combinations of one or multiple bridging domains (SV1–4) and characterize the properties of an effective bridging domain. As in Table 2, DANN fails at adaptation without domain bridge (Web→SV5). While including SV4 as the target domain raises adaptation difficulty, using it as a bridging domain (Web→SV4→5) greatly improves the performance on the SV5 test set. Including SV3 as an additional bridging domain (Web→SV3→4→5) shows additional improvement.

4. Conclusions

This work aims to simplify adaptation problems with extreme domain variations, using unlabeled bridging domains. A novel framework based on DANN is developed by introducing additional discriminators to account for decomposed many, but smaller source-to-target domain discrepancy.

References

[1] Shai Ben-David, John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, and Jennifer Wortman Vaughan. A theory of learning from different domains. *Machine learning*, 2010. 1

[2] Shai Ben-David, John Blitzer, Koby Crammer, and Fernando Pereira. Analysis of representations for domain adaptation. In *NIPS*, 2007. 1

[3] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *TPAMI*, 2018. 1

[4] 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. 1

[5] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *CVPR*, 2009. 1

[6] Yaroslav Ganin, Evgeniya Ustinova, Hana Ajakan, Pascal Germain, Hugo Larochelle, François Laviolette, Mario Marchand, and Victor Lempitsky. Domain-adversarial training of neural networks. *Journal of Machine Learning Research*, 2016. 1

[7] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *CVPR*, 2014. 1

[8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *CVPR*, 2016. 1

[9] Judy Hoffman, Eric Tzeng, Taesung Park, Jun-Yan Zhu, Phillip Isola, Kate Saenko, Alexei A Efros, and Trevor Darrell. Cycada: Cycle-consistent adversarial domain adaptation. In *ICML*, 2018. 1

[10] Judy Hoffman, Dequan Wang, Fisher Yu, and Trevor Darrell. Fcns in the wild: Pixel-level adversarial and constraint-based adaptation. *arXiv preprint arXiv:1612.02649*, 2016. 1

[11] Naoto Inoue, Ryosuke Furuta, Toshihiko Yamasaki, and Kiyoharu Aizawa. Cross-domain weakly-supervised object detection through progressive domain adaptation. In *CVPR*, June 2018. 1

[12] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. In *NIPS*, 2012. 1

[13] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. In *NIPS*, 2015. 1

[14] E. Shelhamer, J. Long, and T. Darrell. Fully convolutional networks for semantic segmentation. *PAMI*, 2017. 1

[15] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. In *ICLR*, 2015. 1

[16] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *CVPR*, 2015. 1

[17] Yi-Hsuan Tsai, Wei-Chih Hung, Samuel Schulter, Kihyuk Sohn, Ming-Hsuan Yang, and Manmohan Chandraker. Learning to adapt structured output space for semantic segmentation. In *CVPR*, June 2018. 1

[18] Eric Tzeng, Judy Hoffman, Ning Zhang, Kate Saenko, and Trevor Darrell. Deep domain confusion: Maximizing for domain invariance. *arXiv preprint arXiv:1412.3474*, 2014. 1

[19] Linjie Yang, Ping Luo, Chen Change Loy, and Xiaoou Tang. A large-scale car dataset for fine-grained categorization and verification. In *CVPR*, 2015. 2