Multi-Source Domain Adaptation and Semi-Supervised Domain Adaptation with Focus on Visual Domain Adaptation Challenge 2019

Yingwei Pan, Yehao Li, Qi Cai, Yang Chen, and Ting Yao
JD AI Reseach, Beijing, China

{panyw.ustc, tingyao.ustc}@gmail.com

Abstract

This notebook paper presents an overview and comparative analysis of our systems designed for the following two tasks in Visual Domain Adaptation Challenge (VisDA-2019): multi-source domain adaptation and semi-supervised domain adaptation.

Multi-Source Domain Adaptation: We investigate both pixel-level and feature-level adaptation for multi-source domain adaptation task, i.e., directly hallucinating labeled target sample via CycleGAN and learning domain-invariant feature representations through self-learning. Moreover, the mechanism of fusing features from different backbones is further studied to facilitate the learning of domain-invariant classifiers. Source code and pre-trained models are available at https://github.com/Panda-Peter/visda2019-multisource.

Semi-Supervised Domain Adaptation: For this task, we adopt a standard self-learning framework to construct a classifier based on the labeled source and target data, and generate the pseudo labels for unlabeled target data. These target data with pseudo labels are then exploited to re-training the classifier in a following iteration. Furthermore, a prototype-based classification module is additionally utilized to strengthen the predictions. Source code and pre-trained models are available at https://github.com/Panda-Peter/visda2019-semisupervised.

1. Introduction

Generalizing a model learnt from a source domain to target domain, is a challenging task in computer vision field. The difficulty originates from the domain gap that may adversely affect the performance especially when the source and target data distributions are very different. An appealing way to address this challenge would be unsupervised domain adaptation (UDA) [1, 11, 15], which aims to utilize labeled examples in source domain and the large number of
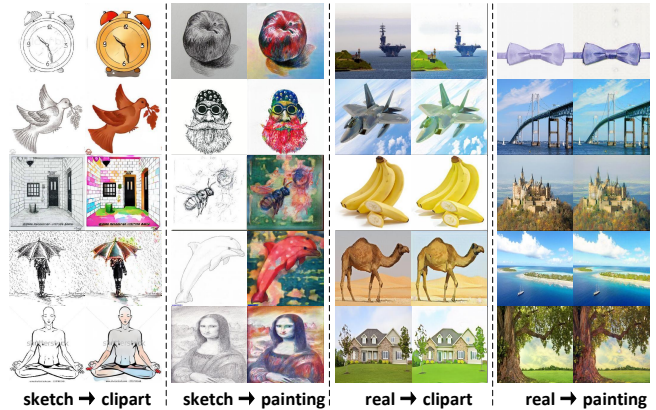

Figure 1. Examples of Pixel-level adaptation between source domains (sketch and real) and target domain (clipart/painting) via CycleGAN in multi-source domain adaptation task.

unlabeled examples in the target domain to generalize a target model. Compared to UDA which commonly recycles knowledge from single source domain, a more difficult but practical task (i.e., multi-source domain adaptation) is proposed in [10] to transfer knowledge from multiple source domains to one unlabeled target domain. In this work, we aim at exploiting both pixel-level and feature-level domain adaptation techniques to tackle this challenge problem. In addition, another task of semi-supervised domain adaptation [4, 14] is explored here when very few labeled data available in the target domain.

2. Multi-Source Domain Adaptation

Inspired from unsupervised image/video translation [3, 17], we utilize CycleGAN [17] to perform unsupervised pixel-level adaptation between source domains (sketch and real) and target domain (clipart/painting), respectively. Thus, each unlabeled training image in sketch or real domains is translated into an image in target domain via the generator of CycleGAN (named as sketch* and real* domains). Figure 1 shows several examples of such pixel-level adaptation from source domains (sketch and real) to target domain (clipart/painting). Next, we
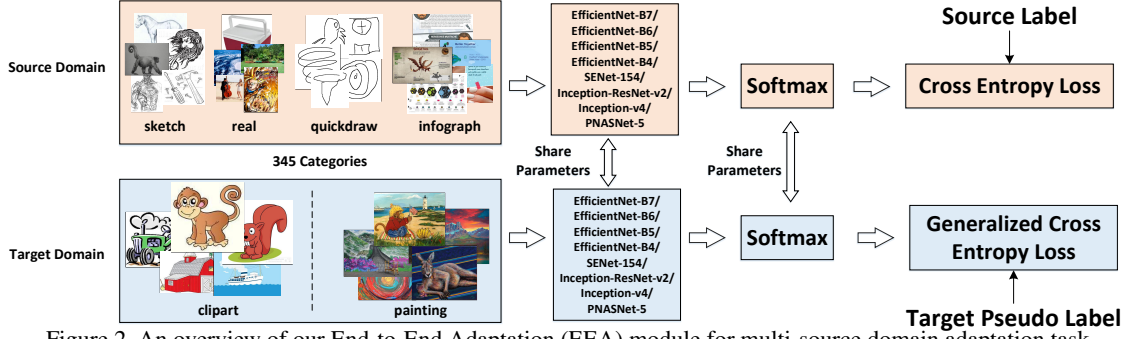
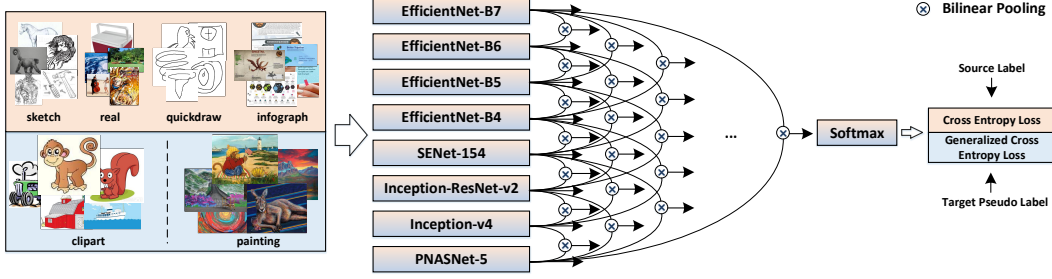Figure 2. An overview of our End-to-End Adaptation (EEA) module for multi-source domain adaptation task.

Figure 3. An overview of our Feature Fusion based Adaptation (FFA) module for multi-source domain adaptation task.

combine all the six source domains (sketch, real, quick-draw, infograph, sketch*, and real*) and train eight source-only models in different backbones (EfficientNet-B7 [13], EfficientNet-B6 [13], EfficientNet-B5 [13], EfficientNet-B4 [13], SENet-154 [5], Inception-ResNet-v2 [12], Inception-v4 [12], PNASNet-5 [8]). All backbones are pre-trained on ImageNet and we can achieve the initial pseudo label for each unlabeled target sample by averaging the predictions of eight source-only models. Furthermore, a hybrid system with two kinds of adaptation models (End-to-end adaptation module and Feature fusion based adaptation module) are utilized to fully exploit pseudo labels for this task. We alternate the two adaptation models in four times for enhancing pseudo labels.

End-to-End Adaptation Module (EEA). This module performs domain adaptation by fine-tuning source-only models with updated pseudo labels in an end-to-end fashion. Figure 2 depicts its detailed architecture. In particular, for unlabeled target data, generalized cross entropy loss [16] is adopted for training with pseudo labels. After training, we update pseudo labels of unlabeled target samples by averaging the predictions of eight adaptation models in different backbones.

Feature Fusion based Adaptation Module (FFA). This module directly extracts features from each backbone in the former module and fuses features from every two backbones via Bilinear Pooling. Next, for each kind of fused feature for input source/target sample, we take it as input and train a classifier from scratch. Each classifier is equipped with cross entropy loss (for labeled source sample) and generalized cross entropy loss (for unlabeled target sample). We illustrate this module in Figure 3. After training the
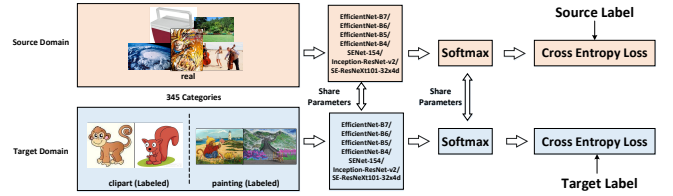

Figure 4. An overview of classifier pre-training for semi-supervised domain adaptation task.
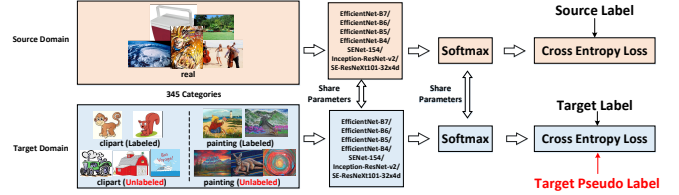

Figure 5. An overview of our End-to-End Adaptation (EEA) module for semi-supervised domain adaptation task.

36 classifiers (28 classifiers with input fused feature and 8 classifiers with input single feature), we update pseudo labels of unlabeled target sample by averaging the predictions of 36 classifiers. At inference, we take the averaged output from 36 classifiers (learnt in Feature fusion based adaptation module at the last time) as the final prediction.

3. Semi-Supervised Domain Adaptation

For semi-supervised domain adaptation task, we over-sample the labeled target samples ($\times 10$) and combine them with labeled source samples for training classifier in a supervised setting. Figure 4 depicts the detailed architecture for classifier pre-training. Note that here we train seven kinds of classifiers in different backbones (EfficientNet-B7,

Table 1. Comparison of different sources and backbones in source-only model for multi-source domain adaptation task on Validation Set.

Method	Source	Target	Backbone	mean_acc_all	mean_acc_classes
Source-only	real	sketch	SE-ResNeXt101_32x4d	40.24%	39.59%
Source-only	real, quickdraw	sketch	SE-ResNeXt101_32x4d	43.09%	41.76%
Source-only	real, quickdraw, infograph	sketch	SE-ResNeXt101_32x4d	48.22%	46.95%
Source-only	real, quickdraw, infograph, real*	sketch	SE-ResNeXt101_32x4d	**50.27**%	**48.59**%
Source-only	real, quickdraw, infograph, real*	sketch	Inception-v4	51.08%	49.22%
Source-only	real, quickdraw, infograph, real*	sketch	Inception-ResNet-v2	52.50%	50.94%
Source-only	real, quickdraw, infograph, real*	sketch	PNASNet-5	51.64%	49.52%
Source-only	real, quickdraw, infograph, real*	sketch	SENet-154	52.40%	50.46%
Source-only	real, quickdraw, infograph, real*	sketch	EfficientNet-B4	53.30%	51.82%
Source-only	real, quickdraw, infograph, real*	sketch	EfficientNet-B6	53.85%	51.98%
Source-only	real, quickdraw, infograph, real*	sketch	EfficientNet-B7	**54.72**%	**52.92**%

Table 2. Comparison of different methods for multi-source domain adaptation task on Validation Set.

Method	Source	Target	Backbone	mean_acc_all	mean_acc_classes
Source-only	real, quickdraw, infograph	sketch	ResNet-101	43.53%	42.73%
SWD [7]	real, quickdraw, infograph	sketch	ResNet-101	44.36%	43.74%
MCD [11]	real, quickdraw, infograph	sketch	ResNet-101	45.01%	44.03%
Source-only	real, quickdraw, infograph	sketch	SE-ResNeXt101_32x4d	48.22%	46.95%
BSP+CDAN [2]	real, quickdraw, infograph	sketch	SE-ResNeXt101_32x4d	53.01%	51.36%
CAN [6]	real, quickdraw, infograph	sketch	SE-ResNeXt101_32x4d	54.74%	52.89%
CAN [6] +TPN [9]	real, quickdraw, infograph	sketch	SE-ResNeXt101_32x4d	56.49%	54.43%
End-to-End Adaptation (Cross Entropy)	real, quickdraw, infograph	sketch	SE-ResNeXt101_32x4d	54.42%	53.18%
End-to-End Adaptation (Generalized Cross Entropy)	real, quickdraw, infograph	sketch	SE-ResNeXt101_32x4d	**58.09**%	**56.15**%

EfficientNet-B6, EfficientNet-B5, EfficientNet-B4, SENet-154, Inception-ResNet-v2, SE-ResNeXt101-32x4d). All backbones are pre-trained on ImageNet and we can achieve the initial pseudo label for each unlabeled target sample by averaging the predictions of the seven classifiers.

End-to-End Adaptation Module (EEA). Next, an end-to-end adaptation module is utilized to incorporate pseudo labels for training classifiers (in the backbones pre-trained on ImageNet), which further bridges the domain gap between source and target domain. Figure 5 illustrates this module. After training, we update pseudo labels of unlabeled target samples by averaging the predictions of seven classifiers in different backbones. The updated pseudo labels will be utilized to train the end-to-end adaptation module again. We repeat such procedure for three times.

Prototype-based Classification Module (PC). Taking the inspiration from Prototype-based adaptation [9], we construct an additional non-parametric classifier to strengthen the predictions from the previous EEA module. Specifically, under each backbone, we define the prototype of each class as the average of all labeled target samples in that class (according to the given labels and pseudo labels). Therefore, the prototype-based classification for each target sample is performed by measuring the distances to prototypes of each class. At inference stage, we take the averaged output from 1) seven classifiers learnt in end-to-end adaptation module at the last time and 2) seven prototype-based classifiers as the final prediction.

4. Experiments

4.1. Multi-Source Domain Adaptation

Effect of pixel-level adaptation in source-only model. Compared to traditional UDA, the key difference in multi-source domain adaptation task is the existence of multiple sources. To fully explore the effect of multiple source do-

Table 3. Comparison of different backbones in our End-to-End Adaptation (EEA) module and Feature Fusion based Adaptation (FFA) module for multi-source domain adaptation on Validation Set (Source: real, quickdraw, infograph, real*; Target: sketch).

Method	Backbone	mean_acc_all	mean_acc_classes
EEA	SE-ResNeXt101_32x4d	59.07%	57.05%
EEA	Inception-v4	59.93%	57.42%
EEA	Inception-ResNet-v2	60.58%	58.32%
EEA	PNASNet-5	60.07%	57.84%
EEA	SENet-154	60.88%	58.29%
EEA	EfficientNet-B4	60.41%	58.30%
EEA	EfficientNet-B6	61.12%	58.80%
EEA	EfficientNet-B7	63.01%	60.33%
FFA	Ensemble	**67.58**%	**64.54**%

mains and the synthetic domain via pixel-level adaptation, we show the performances of source-only model on validation set by injecting one more source domain in Table 1. The results across different metrics consistently indicate the advantage of transferring knowledge from multiple source domains. The performance is further improved by incorporating synthetic domain (real*) via pixel-level adaptation. Table 1 additionally shows the performances of source-only model under different backbones and the best performance is observed when we construct source-only model under EfficientNet-B7.

Effect of End-to-End Adaptation (EEA). We evaluate our End-to-End Adaptation module on Validation Set and compare the results to recent state-of-the-art UDA techniques (e.g., SWD [7], MCD [11], BSP+CDAN [2], CAN [6], and TPN [9]). Results are presented in Table 2. Overall, our adopted EEA with Generalized Cross Entropy exhibits better performance than other runs, which demonstrates the merit of self-learning for multi-source domain adaptation. Note that here we include one variant of our EEA by replacing Generalized Cross Entropy with traditional Cross Entropy, which results in inferior performance. The results verify the advantage of optimizing classifier with Generalized Cross Entropy for unlabeled target samples in self-learning paradigm.

Table 4. Comparison of different components in our system for multi-source domain adaptation on Testing Set (Source: sketch, real, quickdraw, infograph, sketch*, real*; Target: clipart/painting).

Method	Backbone	mean_acc_all (clipart)	mean_acc_all (painting)	mean_acc_all
Source-only	Inception-ResNet-v2	67.77%	59.24%	62.59%
EEA+FFA	Ensemble	78.16%	67.56%	71.73%
$(EEA+FFA)^2$	Ensemble	79.66%	69.51%	73.50%
$(EEA+FFA)^2$, Higher resolution	Ensemble	81.25%	71.65%	75.42%
$(EEA+FFA)^4$, Higher resolution	Ensemble	**81.61**%	**72.31**%	**75.96**%

Table 5. Comparison of different components in our system for semi-supervised domain adaptation on Testing Set (Source: real; Target: clipart/painting).

Method	Backbone	mean_acc_all
Source-only	Ensemble	64.3%
EEA	Ensemble	68.8%
EEA^2	Ensemble	70.5%
EEA^3	Ensemble	71.35%
EEA^3+PC	Ensemble	**71.41**%

Effect of Feature Fusion based Adaptation (FFA). One of the important design in our system is feature fusion based adaptation (FFA) which facilitate the learning of domain-invariant classifier with fused features from different backbones. As shown in Table 3, by fusing the features from every two backbones in EEA via Bilinear Pooling, our FFA leads to a large performance improvement.

Performance on Testing Set. Table 4 illustrates the final performances of our submitted systems with different settings on Testing Set. The basic component in our submitted systems is the hybrid system consisting of two adaptation modules (EEA and FFA), which will be alternated in several times. For simplicity, we denote the system which alternates (EEA+FFA) in N times as $(EEA+FFA)^N$. Note that we also try to enlarge the input resolution of each backbone (+ 64 pixels in both width and hight) in the submitted systems and such processing is named as "Higher resolution." As shown in Table 4, our system with more alternation times and Higher resolution achieves the best performance on Testing Set.

4.2. Semi-Supervised Domain Adaptation

The performance comparisons between our submitted systems for semi-supervised domain adaptation task on Testing Set are summarized in Table 5. Note that here we denote the setting which alternates End-to-End Adaptation (EEA) module in N times as EEA^N. In general, our system with more alternation times obtains higher performance. In addition, by fusing the predictions from both EEA and Prototype-based Classification (PC), our system boosts up the performance.

References

[1] Qi Cai, Yingwei Pan, Chong-Wah Ngo, Xinmei Tian, Lingyu Duan, and Ting Yao. Exploring object relation in mean teacher for cross-domain detection. In *CVPR*, 2019.

[2] Xinyang Chen, Sinan Wang, Mingsheng Long, and Jianmin Wang. Transferability vs. discriminability: Batch spectral penalization for adversarial domain adaptation. In *ICML*, 2019.

[3] Yang Chen, Yingwei Pan, Ting Yao, Xinmei Tian, and Tao Mei. Mocycle-gan: Unpaired video-to-video translation. In *ACMMM*, 2019.

[4] Hal Daumé III, Abhishek Kumar, and Avishek Saha. Frustratingly easy semi-supervised domain adaptation. In *Proceedings of the 2010 Workshop on Domain Adaptation for Natural Language Processing*, 2010.

[5] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *CVPR*, 2018.

[6] Guoliang Kang, Lu Jiang, Yi Yang, and Alexander G Hauptmann. Contrastive adaptation network for unsupervised domain adaptation. In *CVPR*, 2019.

[7] Chen-Yu Lee, Tanmay Batra, Mohammad Haris Baig, and Daniel Ulbricht. Sliced wasserstein discrepancy for unsupervised domain adaptation. In *CVPR*, 2019.

[8] Chenxi Liu, Barret Zoph, Maxim Neumann, Jonathon Shlens, Wei Hua, Li-Jia Li, Li Fei-Fei, Alan Yuille, Jonathan Huang, and Kevin Murphy. Progressive neural architecture search. In *ECCV*, 2018.

[9] Yingwei Pan, Ting Yao, Yehao Li, Yu Wang, Chong-Wah Ngo, and Tao Mei. Transferrable prototypical networks for unsupervised domain adaptation. In *CVPR*, 2019.

[10] Xingchao Peng, Qinxun Bai, Xide Xia, Zijun Huang, Kate Saenko, and Bo Wang. Moment matching for multi-source domain adaptation. *arXiv preprint arXiv:1812.01754*, 2018.

[11] Kuniaki Saito, Kohei Watanabe, Yoshitaka Ushiku, and Tatsuya Harada. Maximum classifier discrepancy for unsupervised domain adaptation. In *CVPR*, 2018.

[12] Christian Szegedy, Sergey Ioffe, Vincent Vanhoucke, and Alexander A Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. In *AAAI*, 2017.

[13] Mingxing Tan and Quoc V Le. Efficientnet: Rethinking model scaling for convolutional neural networks. *arXiv preprint arXiv:1905.11946*, 2019.

[14] Ting Yao, Yingwei Pan, Chong-Wah Ngo, Houqiang Li, and Tao Mei. Semi-supervised domain adaptation with subspace learning for visual recognition. In *CVPR*, 2015.

[15] Yiheng Zhang, Zhaofan Qiu, Ting Yao, Dong Liu, and Tao Mei. Fully convolutional adaptation networks for semantic segmentation. In *CVPR*, 2018.

[16] Zhilu Zhang and Mert Sabuncu. Generalized cross entropy loss for training deep neural networks with noisy labels. In *NIPS*, 2018.

[17] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *ICCV*, 2017.