Open-world Domain Adaptation and Generalization

Sicheng Zhao
schzhao@tsinghua.edu.cn
BNRist, Tsinghua University
China

Jianhua Tao
jhtao@tsinghua.edu.cn
BNRist, Department of Automation, Tsinghua University
China

Guiguang Ding
dinggg@tsinghua.edu.cn
BNRist, School of Software, Tsinghua University
China

ABSTRACT

Deep learning has achieved unprecedented success in various artificial intelligence areas and tasks. One precondition is that large-scale labeled training data is provided to train a neural network. Although recent self-supervised pre-training can utilize unlabeled data to learn discriminative representations, label information is still required for specific downstream tasks. In real-world applications, such as fine-grained recognition and pixel-level segmentation, human annotation might be prohibitively expensive and time-consuming. One intuitive solution is to utilize other labeled datasets with similar desired tasks to perform cross-domain transfer. Domain adaptation and domain generalization aim to address this issue by enhancing the transferability of the models trained on the labeled source domains so that they can well adapt and generalize to the target domain. In this paper, we summarize recent methods in our group on open-world domain adaptation and generalization.

CCS CONCEPTS

• **Computing methodologies → Transfer learning**.

KEYWORDS

Domain adaptation; domain generalization; transfer learning

ACM Reference Format:
Sicheng Zhao, Jianhua Tao, and Guiguang Ding. 2024. Open-world Domain Adaptation and Generalization. In *ACM Turing Award Celebration Conference 2024 (ACM-TURC '24), July 05–07, 2024, Changsha, China.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3674399.3674462

1 OVERVIEW

Domain adaptation (DA) is typically investigated under unsupervised, homogeneous, single-source, single-target, closed-set, and centralized settings [20], *i.e.*, there are one labeled source domain and one unlabeled target domain, the source and target data is observed in the same data space and can be accessed simultaneously during training, and the label sets of the source and target domains are the same. However, in the open world, the DA settings are more complex [10, 12]. For example, there are multiple source domains, the label sets of the source and target domains are different, and only pre-trained source models are given. When target data is unavailable during training, the task becomes domain generalization (DG). As shown in Figure 1, this paper will briefly introduce our recent efforts on single-source DA, multi-source DA, and DG.

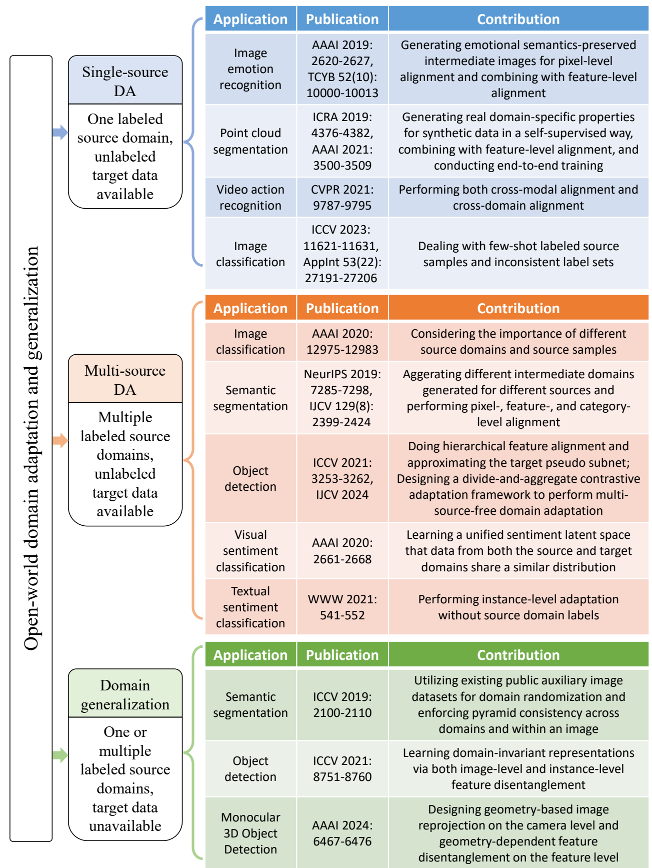

Figure 1: Overview of recent methods in our group on open-world domain adaptation and generalization.

2 SINGLE-SOURCE DOMAIN ADAPTATION

We proposed to study single-source DA for both image emotion recognition [11, 15] and point cloud segmentation [5, 17]. Besides feature-level alignment, we investigated how to generate intermediate domains based on the tasks' characteristics to perform pixel-level alignment. For image emotion recognition, we employed CycleGAN-based methods to generate adapted images with target styles and source emotional semantics [11, 15]. For point cloud segmentation, we rendered the dropout noise for synthetic data based on a rendering network trained on unlabeled real data in a self-supervised manner [17]. The adapted images have similar styles

ACM ISBN 979-8-4007-1011-7/24/07
https://doi.org/10.1145/3674399.3674462

to the target images and the semantics are preserved as the source images. Therefore, the feature-level alignment between the adapted and target domains usually performs better than the alignment between the source and target domains.

For video action recognition, apart from cross-domain alignment, we considered the alignment across modalities by spatio-temporal contrastive learning [4]. For image classification, we addressed the special settings of few-shot and universal DA by confidence-based dispersal learning [6] and consensual contrastive learning [1].

3 MULTI-SOURCE DOMAIN ADAPTATION

On the one hand, we investigated, for the first time, the DA for semantic segmentation, object detection, and visual sentiment classification with multiple sources [3, 8, 13, 14, 19]. For semantic segmentation, we designed adversarial aggregation to aggregate different intermediate domains and aligned the aggregated domain and the target domain on the pixel, feature, and category levels [13, 14]. We demonstrated the models' interpretability through feature transferability, style translation, and attention visualization. For object detection, we designed a hierarchical feature alignment and approximated the target pseudo subset using the weighted combination of source parameters [8]. To preserve the privacy of source data, we also studied multi-source-free domain adaptive object detection. We designed a novel divide-and-aggregate contrastive adaptation framework to efficiently leverage the advantages of multiple source-pretrained models and aggregate their contributions to adaptation in a self-supervised manner [19]. For visual sentiment classification, we learned a unified sentiment latent space where the source and target data share a similar distribution [3]. Image reconstruction and cycle-reconstruction constraints aim to preserve the original information, while the image translation with emotional semantic consistency tries to align different domains.

On the other hand, we designed effective multi-source strategies to improve the performance of traditional domain adaptive image classification and textual sentiment classification [16, 18]. We proposed multi-source distilling DA to consider the importance of different source domains and source samplesn [16]. The source samples that are closer to the target samples are distilled to finetune the source classifiers. The source domains that respectively look more similar to the target domain are assigned higher weights when aggregating different target predictions. To deal with the situation without source domain labels, we further designed curriculum cycle-consistent generative adversarial network to perform instance-level adaptation [18]. Different from images, we generated an intermediate domain for the encoded textual representations of the mixed source. The source samples are assigned weights using novel weighting mechanisms to explore their importance.

4 DOMAIN GENERALIZATION

For semantic segmentation, we proposed to randomize the synthetic images with the visual styles of real images by utilizing public auxiliary datasets and enforcing pyramid consistency to learn domain-invariant and scale-invariant features [9]. This is based on the conjecture that if the network is exposed to sufficient domains in training, it should interpolate well to new real-world target domains. For object detection, we disentangled the representations on both image and instance levels and designed a cross-level reconstruction to preserve informative object representations [2]. For monocular 3D object detection, we designed geometry-guided domain generalization with two main components [7]. Geometry-based image projection mitigates the impact of camera discrepancy by unifying intrinsic parameters, randomizing camera orientations, and unifying the field of view range. Geometry-dependent feature disentanglement overcomes the negative transfer problems by incorporating domain-shared and domain-specific features.

ACKNOWLEDGMENTS

This work is supported by CCF-DiDi GAIA Collaborative Research Funds for Young Scholars and the National Natural Science Foundation of China (Nos. U21B2010, 61925107, 62021002).

REFERENCES

[1] Haojin Liao, Qiang Wang, Sicheng Zhao, Tengfei Xing, and Runbo Hu. 2023. Domain consensual contrastive learning for few-shot universal domain adaptation. *Applied Intelligence* 53, 22 (2023), 27191–27206.
[2] Chuang Lin, Zehuan Yuan, Sicheng Zhao, Peize Sun, Changhu Wang, and Jianfei Cai. 2021. Domain-Invariant Disentangled Network for Generalizable Object Detection. In *ICCV*. 8771–8780.
[3] Chuang Lin, Sicheng Zhao, Lei Meng, and Tat-Seng Chua. 2020. Multi-Source Domain Adaptation for Visual Sentiment Classification. In *AAAI*. 2661–2668.
[4] Xiaolin Song, Sicheng Zhao, Jingyu Yang, et al. 2021. Spatio-temporal Contrastive Domain Adaptation for Action Recognition. In *CVPR*. 9787–9795.
[5] Bichen Wu, Xuanyu Zhou, Sicheng Zhao, Xiangyu Yue, and Kurt Keutzer. 2019. SqueezeSegV2: Improved Model Structure and Unsupervised Domain Adaptation for Road-Object Segmentation from a LiDAR Point Cloud. In *ICRA*. 4376–4382.
[6] Yizhe Xiong, Hui Chen, Zijia Lin, Sicheng Zhao, and Guiguang Ding. 2023. Confidence-based Visual Dispersal for Few-shot Unsupervised Domain Adaptation. In *ICCV*. 11621–11631.
[7] Fan Yang, Hui Chen, Yuwei He, Sicheng Zhao, Chenghao Zhang, Kai Ni, and Guiguang Ding. 2024. Geometry-Guided Domain Generalization for Monocular 3D Object Detection. In *AAAI*. 6467–6476.
[8] Xingxu Yao, Sicheng Zhao, Pengfei Xu, and Jufeng Yang. 2021. Multi-Source Domain Adaptation for Object Detection. In *ICCV*. 3273–3282.
[9] Xiangyu Yue, Yang Zhang, Sicheng Zhao, et al. 2019. Domain Randomization and Pyramid Consistency: Simulation-to-Real Generalization Without Accessing Target Domain Data. In *ICCV*. 2100–2110.
[10] Sicheng Zhao, Hui Chen, Hu Huang, Pengfei Xu, and Guiguang Ding. 2024. More is Better: Deep Domain Adaptation with Multiple Sources. In *IJCAI*.
[11] Sicheng Zhao, Xuanbai Chen, Xiangyu Yue, Chuang Lin, et al. 2022. Emotional Semantics-Preserved and Feature-Aligned CycleGAN for Visual Emotion Adaptation. *IEEE TCYB* 52, 10 (2022), 10000–10013.
[12] Sicheng Zhao, Xiaopeng Hong, Jufeng Yang, Yanyan Zhao, and Guiguang Ding. 2023. Toward Label-Efficient Emotion and Sentiment Analysis. *Proc. IEEE* 111, 10 (2023), 1159–1197.
[13] Sicheng Zhao, Bo Li, Pengfei Xu, Xiangyu Yue, Guiguang Ding, and Kurt Keutzer. 2021. MADAN: Multi-source Adversarial Domain Aggregation Network for Domain Adaptation. *IJCV* 129, 8 (2021), 2399–2424.
[14] Sicheng Zhao, Bo Li, Xiangyu Yue, Yang Gu, Pengfei Xu, Runbo Hu, Hua Chai, and Kurt Keutzer. 2019. Multi-source domain adaptation for semantic segmentation. *NeurIPS* (2019), 7285–7298.
[15] Sicheng Zhao, Chuang Lin, Pengfei Xu, et al. 2019. CycleEmotionGAN: Emotional Semantic Consistency Preserved CycleGAN for Adapting Image Emotions. In *AAAI*. 2620–2627.
[16] Sicheng Zhao, Guangzhi Wang, Shanghang Zhang, et al. 2020. Multi-Source Distilling Domain Adaptation. In *AAAI*. 12975–12983.
[17] Sicheng Zhao, Yezhen Wang, Bo Li, et al. 2021. ePointDA: An End-to-End Simulation-to-Real Domain Adaptation Framework for LiDAR Point Cloud Segmentation. In *AAAI*. 3500–3509.
[18] Sicheng Zhao, Yang Xiao, Jiang Guo, Xiangyu Yue, Jufeng Yang, Ravi Krishna, Pengfei Xu, and Kurt Keutzer. 2021. Curriculum CycleGAN for Textual Sentiment Domain Adaptation with Multiple Sources. In *WWW*. 541–552.
[19] Sicheng Zhao, Huizai Yao, Chuang Lin, Yue Gao, and Guiguang Ding. 2024. Multi-Source-Free Domain Adaptive Object Detection. *IJCV* (2024).
[20] Sicheng Zhao, Xiangyu Yue, Shanghang Zhang, et al. 2022. A Review of Single-Source Deep Unsupervised Visual Domain Adaptation. *IEEE TNNLS* 33, 2 (2022), 473–493.