UNIVERSAL DOMAIN ADAPTATION THROUGH SELF-SUPERVISION

Review Paper - 2

GNR 650

By
Aditya Prakash (210260004)
K.S.Saketh (210260025)

1 Introduction

The paper addresses the challenge of Domain Adaptation (DA) in deep learning for image recognition. DA aims to transfer knowledge from a labeled source domain to an unlabeled target domain, even when the target domain may have different categories or a mix of known and unknown categories compared to the source. Existing methods struggle with this task due to their reliance on prior knowledge about the category shift. The proposed method, called DANCE (Domain Adaptive Neighborhood Clustering via Entropy optimization), introduces a novel approach with two self-supervision based components: neighborhood clustering and entropy separation which can handle arbitrary category shift. It combines self-supervised clustering to learn target domain features and an entropy separation loss to align source and target distributions. This allows the model to handle various category shifts without prior knowledge. DANCE outperforms other methods in all evaluated settings, demonstrating its effectiveness in adapting to new and unknown categories.

2 Universal Domain Adaption (UniDA)

There are various cases and challenges in domain adaptation (DA) including Closed-set Domain Adaptation (CDA), Partial Domain Adaptation (PDA), and Open-set Domain Adaptation (ODA). This introduces the concept of Universal Domain Adaptation (UniDA) which aims to address all these scenarios collectively. The authors propose a method designed to handle CDA, ODA, PDA, and a combination known as OPDA within the UniDA framework. They emphasize the effectiveness of entropy minimization for unlabeled samples in CDA, especially when combined with batch-normalization based domain alignment. Additionally, pseudo-labeling is mentioned as an effective approach in CDA when combined with domain-specific batch-normalization. The proposed method, which incorporates an entropy separation loss, is distinctive as it has the capability to decrease confidence to reject "unknown" classes, providing a more comprehensive solution for UniDA.

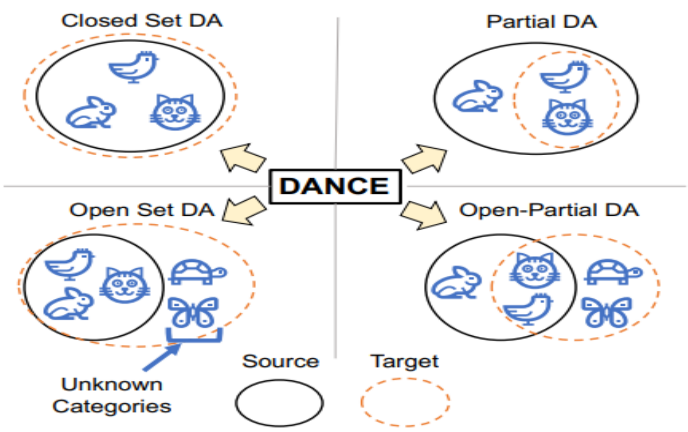


Figure 1. Unlike previous methods tailored to each setting, our proposed approach, called DANCE, performs well on all sub-settings of universal domain adaptation.

3 Use of Self-Supervised Learning

Self-supervised learning is key in Universal Domain Adaptation (UniDA). It helps extract meaningful features from unlabeled images. In UniDA, models are trained on tasks like puzzle solving or instance discrimination with unlabeled target data. This allows the model to learn target-specific representations. Self-supervised techniques, like neighborhood clustering, handle the challenge of unknown cluster numbers in the target domain, making them valuable for UniDA. Incorporating entropy minimization improves the model's class distinction and adaptability to diverse domain shifts. This approach leverages the target domain's structure for more effective adaptation.

4 DANCE (Domain Adaptive Neighborhood Clustering via Entropy optimization)

The task at hand is Universal Domain Adaptation (UniDA), where we have labeled data from a source domain (with known categories) and unlabeled data from a target domain containing potentially unknown categories. The goal is to label target samples into either known source categories or an "unknown" label. The approach involves a prototype-based classifier that aligns samples with their true class centroids. The proposed method employs two key techniques:

Neighborhood Clustering (NC): This involves using self-supervision in the target domain to cluster target samples. Each target point is aligned either to a known class prototype in the source or to its neighbor in the target. This allows for learning a metric that maps a point to its semantically close match, regardless of its category.

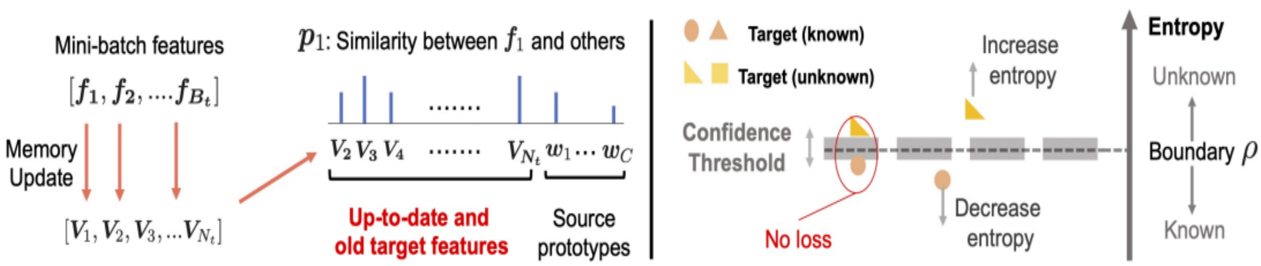
Entropy Separation (ES) Loss: This loss function is applied to the entropy of the known category classifier's output. It forces the output to be either low (indicating the sample belongs to a known class) or high (indicating the sample should be far from any known class). This helps in aligning the target points with source prototypes or rejecting them as unknown.

Additionally, domain-specific batch normalization is used to eliminate domain style information as a form of weak domain alignment. This approach aims to handle category shifts without prior knowledge of them, avoiding catastrophic misalignment by focusing on well-clustered target features while allowing relaxed alignment to the source classes and potential rejection of unknown points.

Figure 1: We propose *DANCE*, which combines a self-supervised clustering loss (red) to cluster neighboring target examples and an entropy separation loss (gray) to consider alignment with source (best viewed in color).

Architecture:

The adopted architecture, based on (Saito et al., 2019), includes an L2 normalization layer before the final linear layer. The weight vectors in this last linear layer serve as prototype features for each class. This architecture is well-suited for clustering both target features and source prototypes. The feature extraction network takes an input and produces a feature vector. The classification network consists of a single linear layer without bias, with weight vectors representing the classes. This network takes L2 normalized features and outputs logits for K classes.



5 Experiments and Results

The goal of the experiments is to compare DANCE with the baselines across all sub-cases of Universal DA (i.e., CDA, PDA, ODA, and OPDA) under the four object classification datasets and four settings for each dataset.

Datasets:

The study utilizes several benchmark datasets for universal domain adaptation. The Office dataset comprises three domains (Amazon, DSLR, Webcam) with 31 classes. OfficeHome dataset contains four domains and 65 classes. The VisDA dataset involves 12 classes across two domains: synthetic and real images. For experiments involving varying class numbers, Caltech and ImageNet datasets are employed due to their extensive class sets.

Evaluation: In CDA and PDA, accuracy is computed across all target samples. In ODA and OPDA, accuracy is averaged over classes, which includes the "unknown" category. For instance, in the Office ODA setting, the average is taken over 11 classes. Each experiment is executed three times, and the reported result is the average outcome across these runs.

Table 1: Summary of the Universal comparisons. Each dataset (Office, OC, OH, VisDA) has multiple domains and adaptation scenarios and we provide the average accuracy over all scenarios. Our DANCE method substantially improves performance compared to the source-only model in all settings and the average rank of DANCE is significantly higher than all other baselines.

Method	Closed DA			Partial DA			Open set DA			Open-Partial DA			Avg	
	Office	OH	VD	OC	OH	VD	Office	OH	VD	Office	OH	VD	Acc	Rank
Source Only	76.5	54.6	46.3	75.9	57.0	49.9	89.1	69.6	43.2	86.4	71.0	38.8	61.7	4.8± 1.2
DANN [12]	**85.9**	62.7	69.1	42.2	40.9	38.7	88.7	72.8	48.2	88.7	76.7	50.6	65.7	3.5± 1.7
ETN [3]	85.2	64.0	64.1	**92.8**	69.4	59.8	88.2	71.9	51.7	88.3	72.6	66.6	70.5	2.9± 1.6
STA [23]	73.6	44.7	48.1	69.8	47.9	48.2	89.9	69.3	48.8	89.8	72.6	47.4	61.2	4.5± 1.3
UAN [43]	84.4	58.8	66.4	52.9	34.2	39.7	91.0	74.6	50.0	84.1	75.0	47.3	62.0	4.1± 1.3
DANCE (ours)	85.5	**69.1**	**70.2**	84.7	**71.1**	**73.7**	**94.1**	**78.1**	**65.3**	**93.9**	**80.4**	**69.2**	**77.3**	**1.2**± 0.4

DANCE demonstrates consistent performance improvement over the source-only model (SO) across all settings. Notably, DANCE excels in open set and open-partial adaptation across all scenarios, as well as in partial and closed domain adaptation for OfficeHome and VisDA. The average performance of DANCE outperforms other baselines in terms of both accuracy and rank.