Towards Category and Domain Alignment: Category-Invariant Feature Enhancement for Adversarial Domain Adaptation

Yuan Wu*, Diana Inkpen†, Ahmed El-Roby*
*SCS, Carleton University †EECS, University of Ottawa

2021 ICCV OCTOBER 11-17 VIRTUAL

Background:

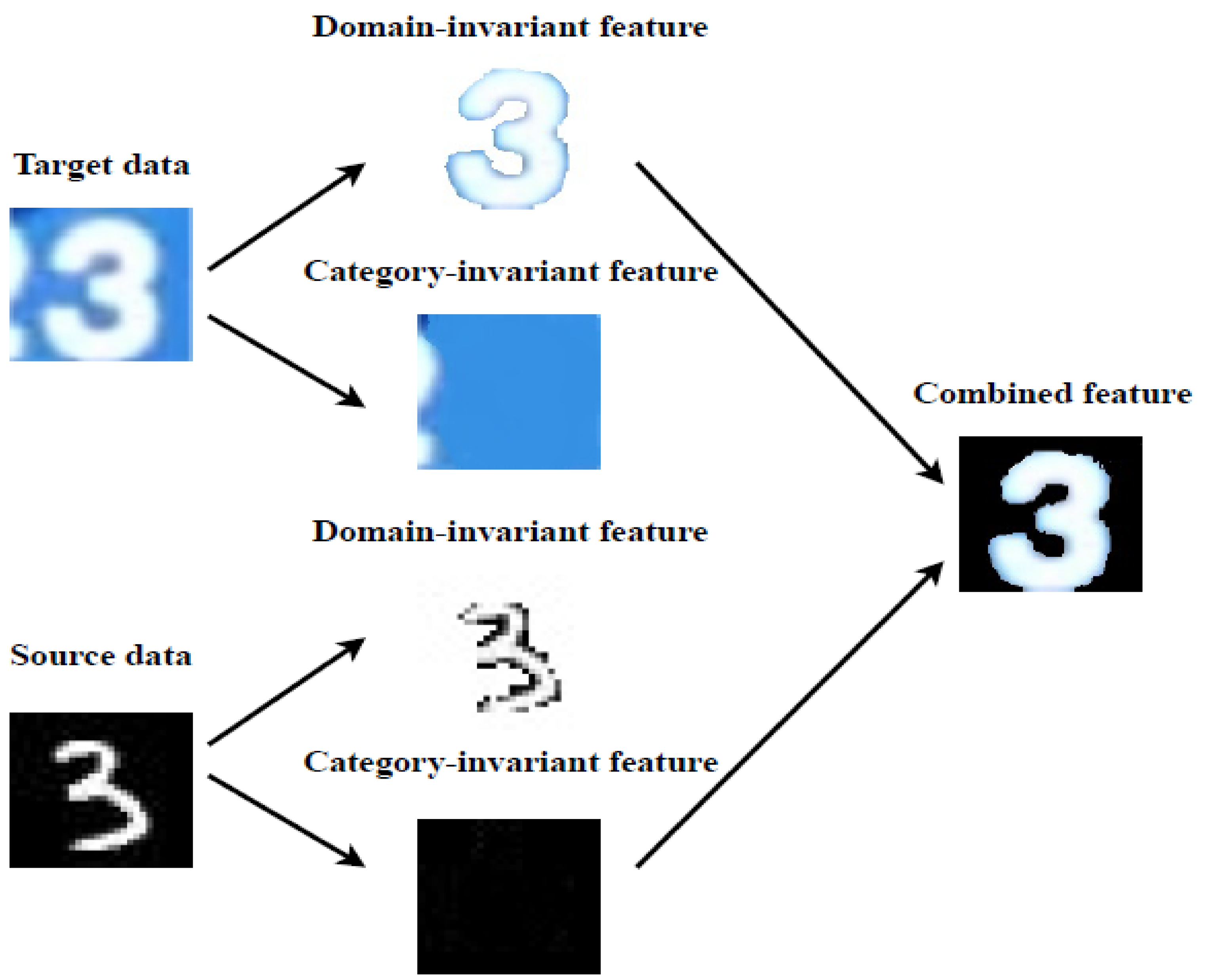
Domain Adaptation:

As collecting label is often expensive and time consuming, it is of great significance to learn knowledge from a label-dense (source) domain and apply the learned knowledge to a label-scarce (target) domain. Domain adaptation is proposed to learn transferable representations across domains such that a model trained on the source domain can also perform well on the target domain.

Adversarial Training:

Adversarial Training was first proposed to generate images, then it was extended to learn domain-invariant features across different domains for domain adaptation. Adversarial training can align different feature distributions in the latent space. When applying adversarial training in domain adaptation, it plays a two player mini-max game between a domain discriminator and a feature extractor: the domain discriminator aims to distinguish source features from target features, while the feature extractor strives to confuse the domain discriminator.

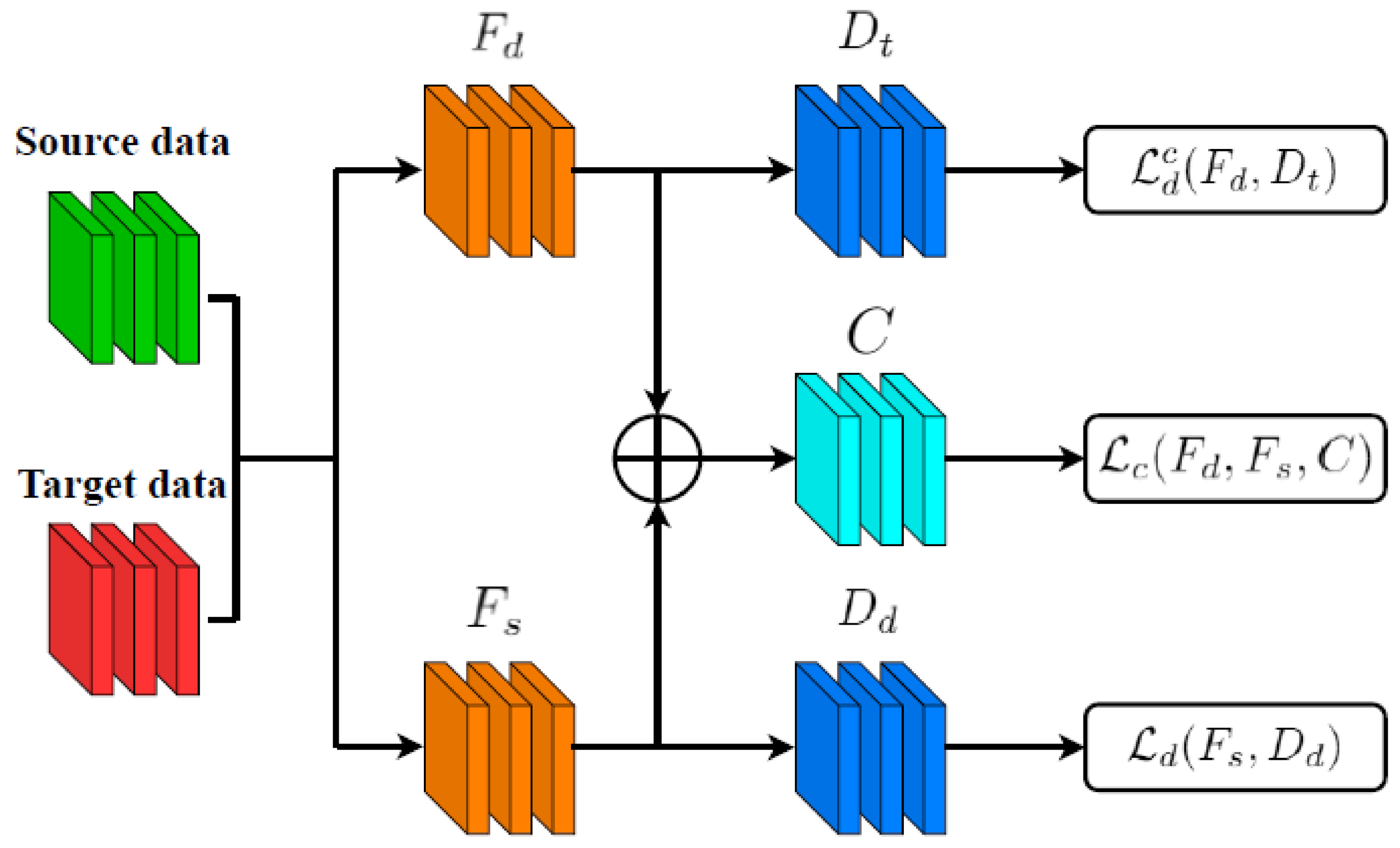
Motivation:



Contribution:

- ❖ We propose a category-invariant feature enhancement (CIFE) mechanism, which enhances the discriminability of the domain-invariant features. The proposed CIFE improves the system performance by optimizing the adaptability, rather than further reducing the domain divergence.
- ❖ To evaluate the efficacy of the CIFE, we embed CIFE into two existing adversarial domain adaptation methods and evaluate them on five benchmarks. Our proposed CIFE significantly improves upon these two methods by yielding state-of-the-art results.
- ❖ Further experiments are conducted to validate the feasibility of advancing domain adaptation by optimizing the adaptability, and explore how the hyperparameter influences the performance of the model.

Methods:



Our model consists of two types of adversarial training: (1) The adversarial training between the domain-specific feature extractor $\mathcal{F}_s$ and the category discriminator $\mathcal{D}_t$, aiming to extract category-invariant features; (2) The adversarial training between the domain-invariant feature extractor $\mathcal{F}_s$ and the domain discriminator $\mathcal{D}_d$, aiming to learn domain-invariant features. The classifier $\mathcal{C}$ uses the concatenation of the domain-invariant feature vector and the category-invariant feature vector as its input and outputs the label probabilities.

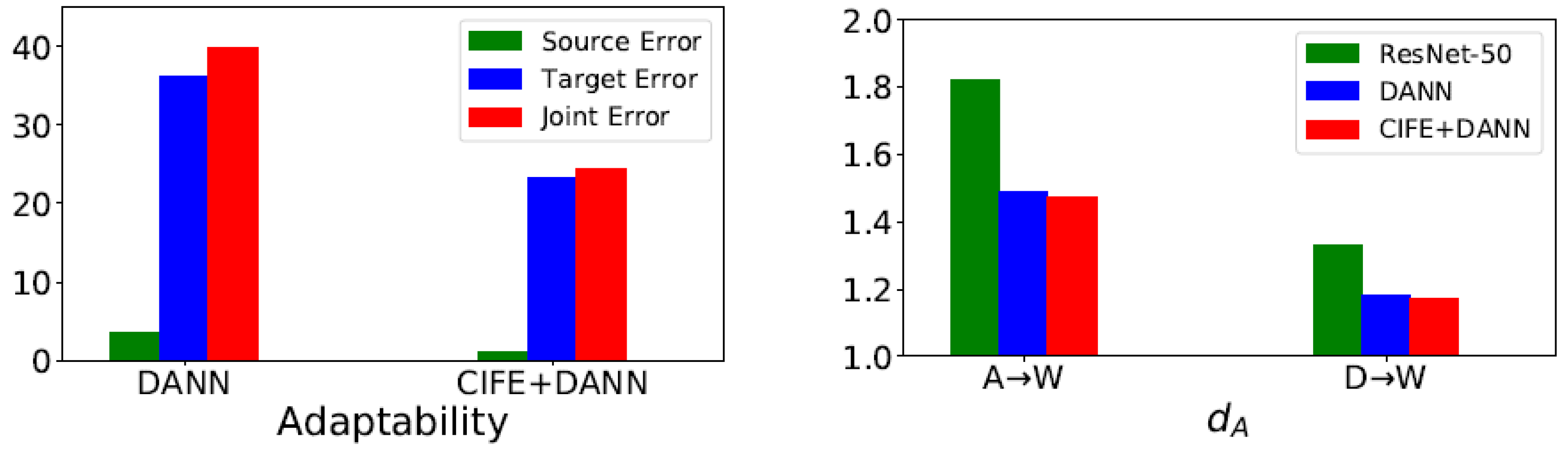
Experiments:

Table 1. Accuracy (%) on Office-31.

Method	A→W	D→W	W→D	A→D	D→A	W→A	Avg
ResNet-50 [11]	68.4±0.2	96.7±0.1	99.3±0.1	68.9±0.2	62.5±0.3	60.7±0.3	76.1
DAN [15]	80.5±0.4	97.1±0.2	99.6±0.1	78.6±0.2	63.6±0.3	62.8±0.2	80.4
DANN [6]	82.0±0.4	96.9±0.2	99.1±0.1	79.7±0.4	68.2±0.4	67.4±0.5	82.2
JAN [17]	85.4±0.3	97.4±0.2	99.8±0.2	84.7±0.3	68.6±0.3	70.0±0.4	84.3
MADA [19]	90.0±0.1	97.4±0.1	99.6±0.1	87.8±0.2	70.3±0.3	66.4±0.3	85.2
CDAN [16]	93.1±0.2	98.2±0.2	**100.0**±0.0	89.8±0.3	70.1±0.4	68.0±0.4	86.6
BSP [4]	93.3±0.2	98.2±0.2	**100.0**±0.0	93.0±0.2	73.6±0.3	72.6±0.3	88.5
ETD [14]	92.1	**100.0**	**100.0**	88.0	71.0	67.8	86.2
A^2LP [25]	87.7	98.1	99.0	87.8	75.8	**75.9**	87.4
BNM [5]	92.8	98.8	**100.0**	92.9	73.5	73.8	88.6
CIFE+DANN	90.7±0.3	99.0±0.1	**100.0**±0.0	90.0±0.5	71.0±0.3	69.9±0.3	86.8
CIFE+CDAN	**94.0**±0.2	99.3±0.1	**100.0**±0.0	**93.4**±0.2	**75.9**±0.2	74.3±0.3	**89.5**

Table 2. Accuracy (%) on ImageCLEF-DA.

Method	I→P	P→I	I→C	C→I	C→P	P→C	Avg
ResNet-50 [11]	74.8±0.3	83.9±0.1	91.5±0.3	78.0±0.2	65.5±0.3	91.2±0.3	80.7
DAN [15]	74.5±0.4	82.2±0.2	92.8±0.2	86.3±0.4	69.2±0.4	89.8±0.4	82.5
DANN [6]	75.0±0.6	86.0±0.3	96.2±0.4	87.0±0.5	74.3±0.5	91.5±0.6	85.0
JAN [17]	76.8±0.4	88.0±0.2	94.7±0.2	89.5±0.3	74.2±0.3	91.7±0.3	85.8
MADA [19]	75.0±0.3	87.9±0.2	96.0±0.3	88.8±0.3	75.2±0.2	92.2±0.3	85.8
CDAN [16]	76.7±0.3	90.6±0.3	97.0±0.4	90.5±0.4	74.5±0.3	93.5±0.4	87.1
DAAN [30]	78.5	91.3	94.4	88.4	74.0	94.3	86.8
ETD [14]	**81.0**	91.7	97.9	93.3	**79.5**	95.0	89.7
A^2LP [31]	79.3	91.8	96.3	91.7	78.1	96.0	88.9
CIFE+DANN	77.0±0.2	91.1±0.2	97.3±0.3	90.8±0.3	74.5±0.5	93.7±0.3	87.4
CIFE+CDAN	79.5±0.3	**93.0**±0.2	**98.2**±0.3	**93.6**±0.3	79.2±0.4	**96.1**±0.4	**90.0**



(a) Adaptability (b) A-distance