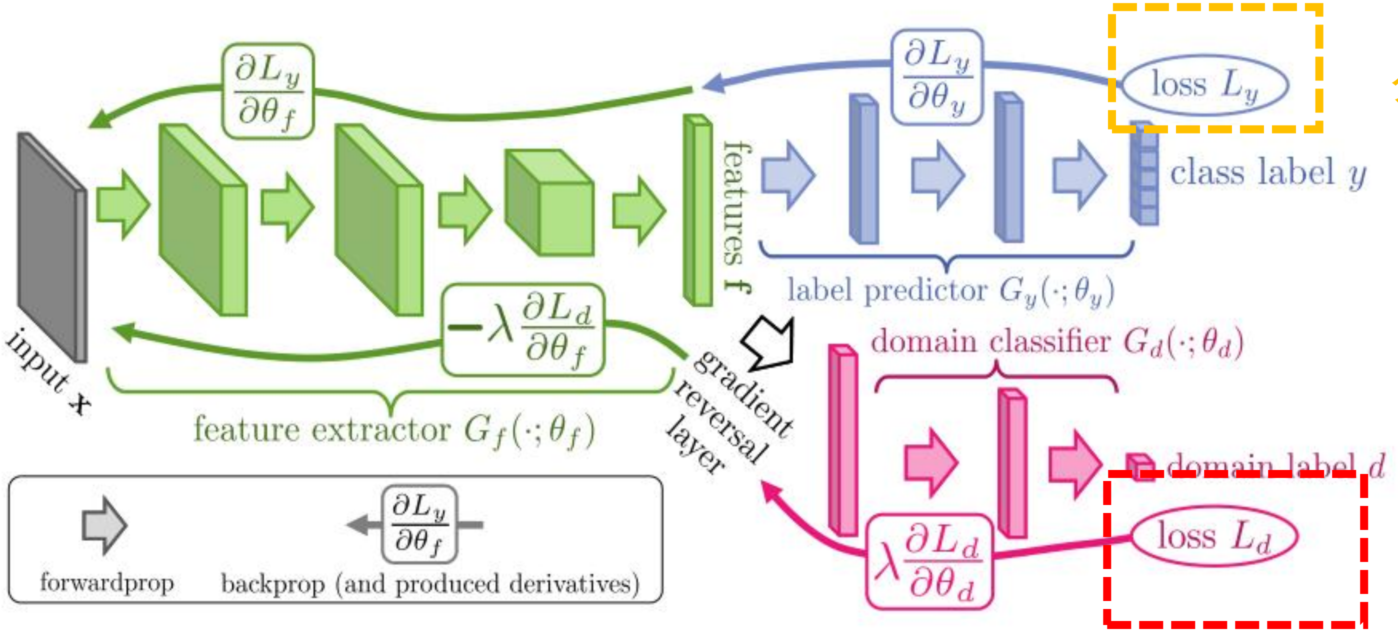
Asymmetric Tri-training for Unsupervised Domain Adaptation

Kuniaki Saito [1] Yoshitaka Ushiku [1] Tatsuya Harada [1 2]

ICML 2017

Background —— Domain-Adversarial Training of Neural Networks

- DANN aimed at obtaining domain-invariant features by minimizing the divergence between domains, as well as a category loss on the source domain.



分类Loss

Domain-invariant features

域判别分类Loss

Domain-Adversarial Training of Neural Networks[1]

$$\epsilon_T(h) \leq \epsilon_S(h) + \frac{1}{2} d_{\mathcal{H}\Delta\mathcal{H}}(\mathcal{D}_S, \mathcal{D}_T) + \lambda$$

$$h* = \underset{h \in \mathcal{H}}{\operatorname{argmin}}\, \epsilon_S(h) + \epsilon_T(h)$$

$$\lambda = \epsilon_S(h^*) + \epsilon_T(h^*)$$

[1] Y aroslav Ganin and Victor Lempitsky. Unsupervised domain adaptation by backpropagation. arXiv preprint arXiv:1409.7495, 2014.

Motivation

- DANN aims to learn domain-invariantly discriminative representations.
- However, if a classifier that works well on both the source and the target domains does not exist, we theoretically cannot expect a discriminative classifier to be applicable to the target domain.
- This methods aims to learn target-discriminative representations for target domain by assigning pseudo-label to the target samples and training the target-specific networks as if they were true labels.

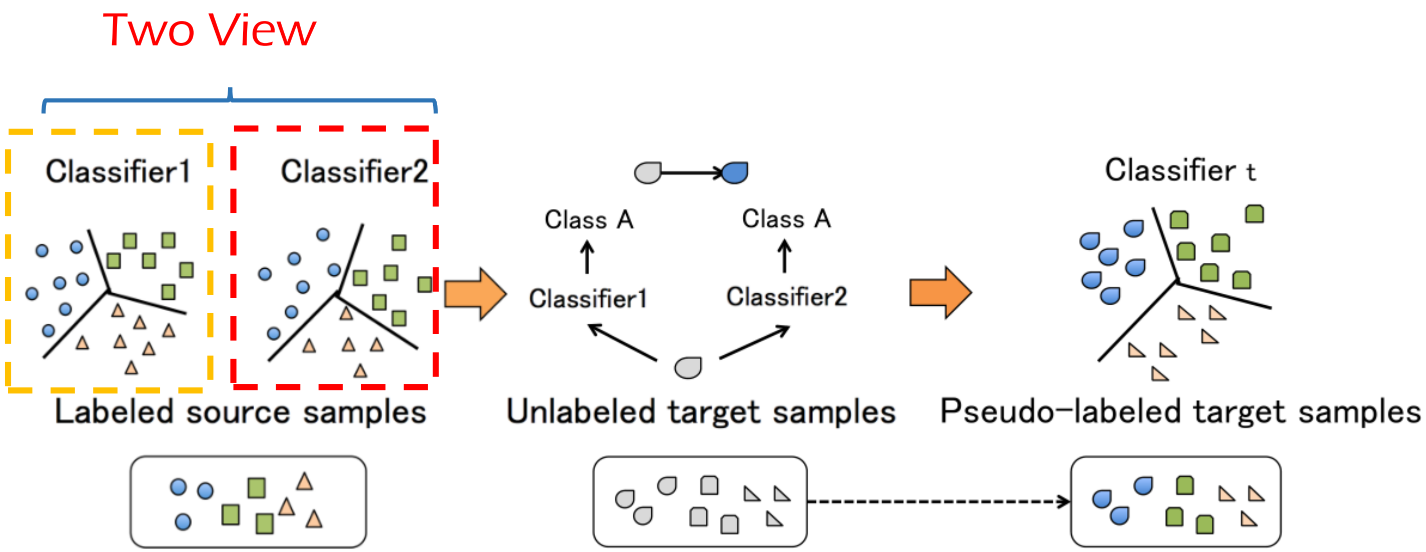
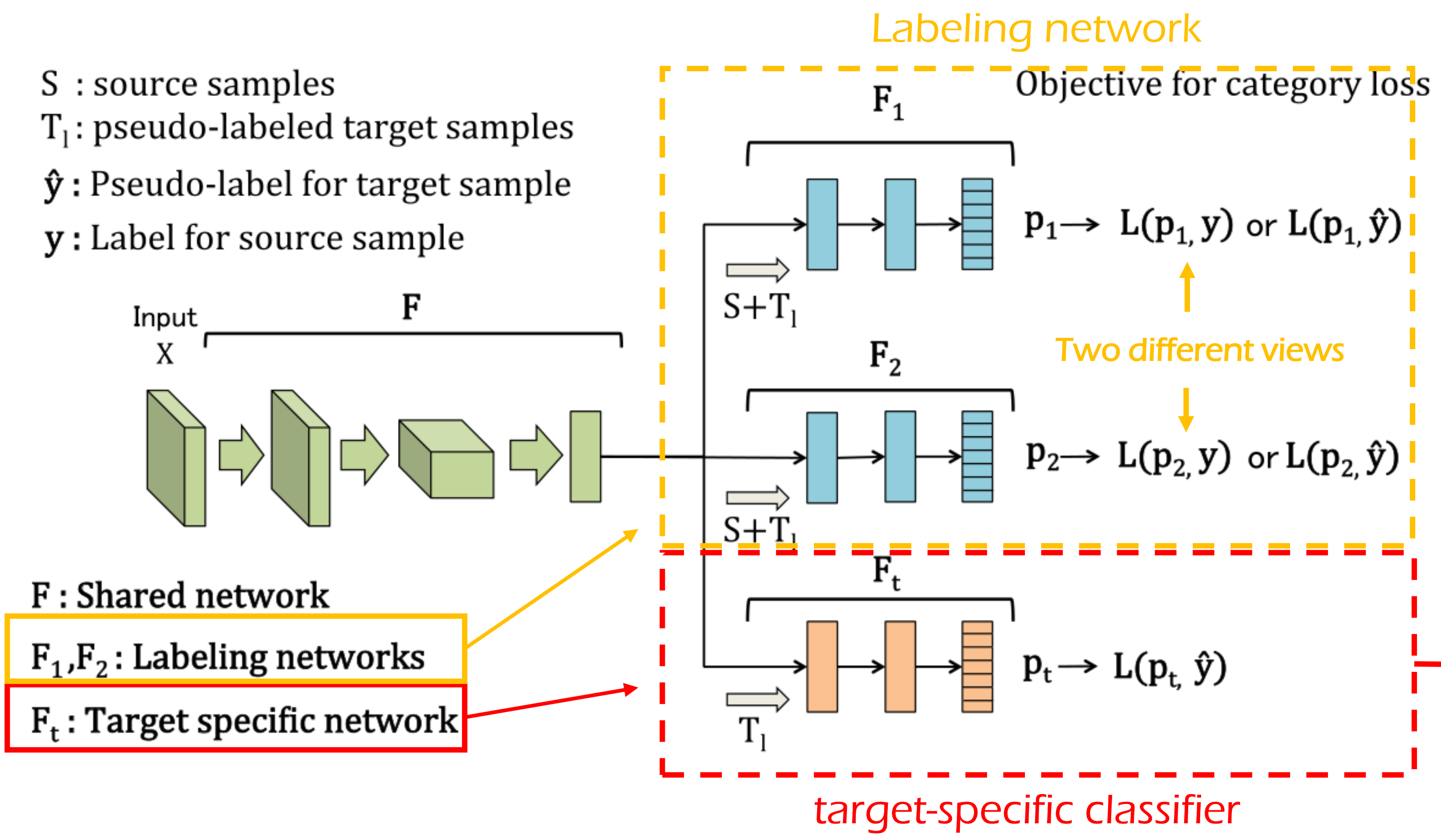


Figure 1. Outline of our model. We assign pseudo-labels to unlabeled target samples based on the predictions from two classifiers trained on the source samples.

Method —— Asymmetric Tri-training

- Asymmetric means that every classifiers has been assigned different roles.



Loss:

确保两个分类器F1，F2从两个不同的视角来为目标域样本分配伪标记。

$$E(\theta_F, \theta_{F_1}, \theta_{F_2}) = \frac{1}{n}\sum_{i=1}^{n}[L_y(F_1 \circ F(x_i)), y_i) + L_y(F_2 \circ (F(x_i)), y_i)] + \boxed{\lambda|{W_1}^T W_2|} \quad (1)$$

To learn Target-discriminative representations

Method —— Algorithm

Algorithm 1 $iter$ denotes the iteration of the training. The function *Labeling* indicates the labeling method. We assign pseudo-labels to samples when the predictions of F_1 and F_2 agree, and at least one of them is confident of their predictions.

Input: data
$\mathbf{X^s} = \{(x_i, t_i)\}_{i=1}^{m}$, $\mathbf{X^t} = \{(x_j)\}_{j=1}^{n}$
$\mathbf{X^t}_l = \emptyset$
for $j = 1$ **to** $iter$ **do**
  Train F, F_1, F_2, F_t with a mini-batch from the training set $\mathcal{S}$
end for
$N_t = N_{init}$ # 5000
$\mathbf{X^t}_l$ = Labeling($F, F_1, F_2, \mathbf{X^t}, N_t$)
$\mathcal{L} = \mathbf{X^s} \cup \mathbf{X^t}_l$
for K steps **do**
  for $j = 1$ **to** $iter$ **do**
    Train F, F_1, F_2 with mini-batch from training set $\mathcal{L}$
    Train F, F_t with mini-batch from training set $\mathbf{X^t}_l$
  end for
  $\mathbf{X^t}_l = \emptyset$, $N_t = K/20 * n$
  $\mathbf{X^t}_l$ = Labeling($F, F_1, F_2, \mathbf{X^t}, N_t$)
  $\mathcal{L} = \mathbf{X^s} \cup \mathbf{X^t}_l$
end for

Labeling() 为目标域数据加入伪标记

Two conditions:
1.两个分类器给出相同的分类标记（两个视图）.
2.分类置信度 > 0.9 or 0.95

N_t: pseudo-labeled candidates

Experiments

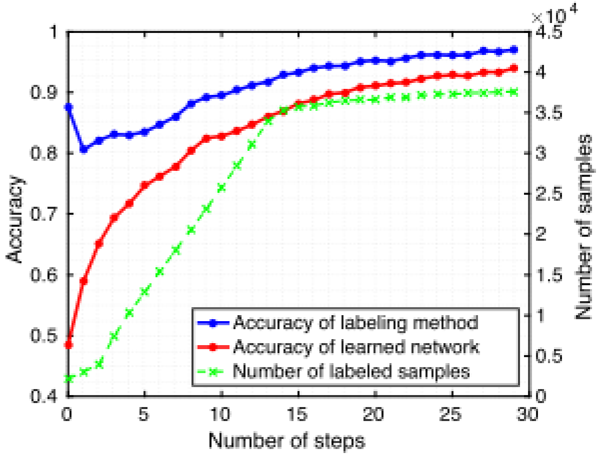
METHOD SOURCE / TARGET	MNIST / MNIST-M	SVHN / MNIST	MNIST / SVHN	SYN DIGITS / SVHN	SYN SIGNS / GTSRB
Source Only w/o BN	59.1(56.6)	68.1(59.2)	37.2(30.5)	84.1(86.7)	79.2(79.0)
Source Only with BN	57.1	70.1	34.9	85.5	75.7
MMD (Long et al., 2015b)	76.9	71.1	-	88.0	91.1
DANN (Ganin & Lempitsky, 2014)	81.5	71.1	35.7	90.3	88.7
DRCN (Ghifary et al., 2016)	-	82.0	40.1	-	-
DSN (Bousmalis et al., 2016)	83.2	82.7	-	91.2	93.1
kNN-Ad (Sener et al., 2016)	86.7	78.8	40.3	-	-
Ours w/o BN	85.3	79.8	39.8	**93.1**	**96.2**
Ours w/o weight constraint ($\lambda = 0$)	**94.2**	**86.0**	49.7	92.4	94.0
Ours	94.0	85.8	**52.8**	92.9	**96.2**

Table 1. Results of the visual domain adaptation experiment on digit and traffic sign datasets. In every setting, our method outperforms other methods by a large margin. In the source-only results, we show the results reported in (Bousmalis et al., 2016) and (Ghifary et al., 2016) in parentheses.

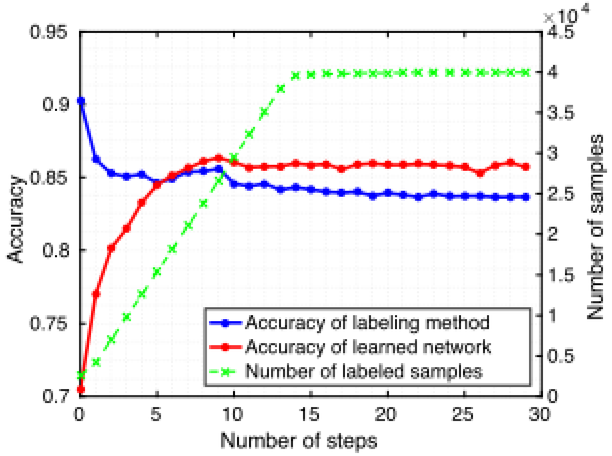
Source→Target	VFAE	DANN	Our method
books→dvd	79.9	78.4	**80.7**
books→electronics	79.2	73.3	**79.8**
books→kitchen	81.6	77.9	**82.5**
dvd→books	**75.5**	72.3	73.2
dvd→electronics	**78.6**	75.4	77.0
dvd→kitchen	82.2	78.3	**82.5**
electronics→books	72.7	71.1	**73.2**
electronics→dvd	**76.5**	73.8	72.9
electronics→kitchen	85.0	85.4	**86.9**
kitchen→books	72.0	70.9	**72.5**
kitchen→dvd	73.3	74.0	**74.9**
kitchen→electronics	83.8	84.3	**84.6**

Table 3. Amazon Reviews experimental results. The accuracy (%) of the proposed method is shown with the result of VFAE (Louizos et al., 2015) and DANN (Ganin et al., 2016).

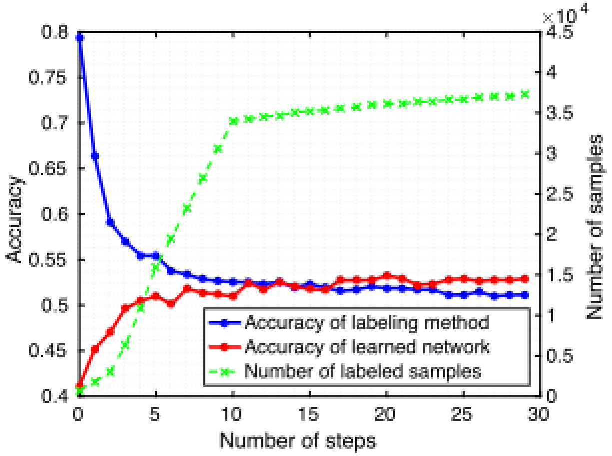
Experiments



(a) MNIST→MNIST-M

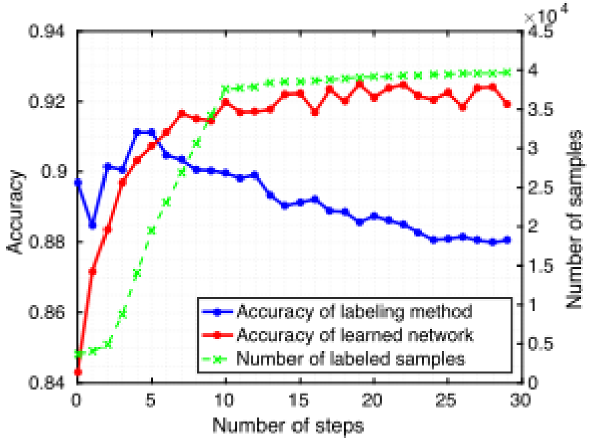


(b) SVHN→MNIST

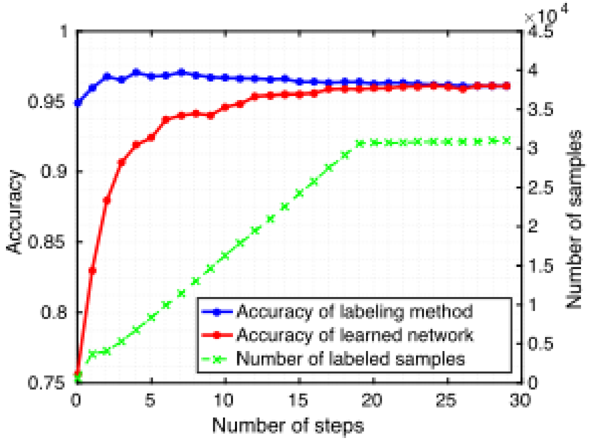


(c) MNIST→SVHN

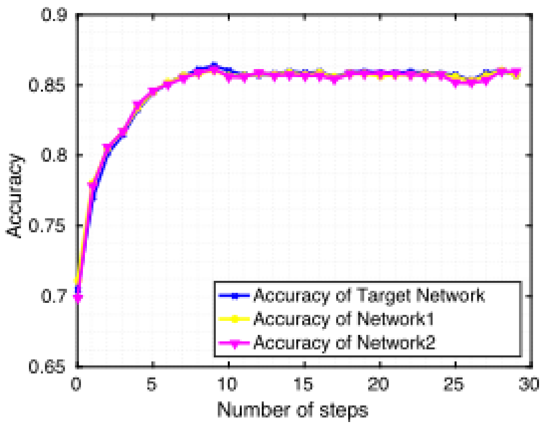
Labeling accuracy : (the number of correctly labeled samples)/(the number of labeled samples)



(d) SYNDIGITS→SVHN

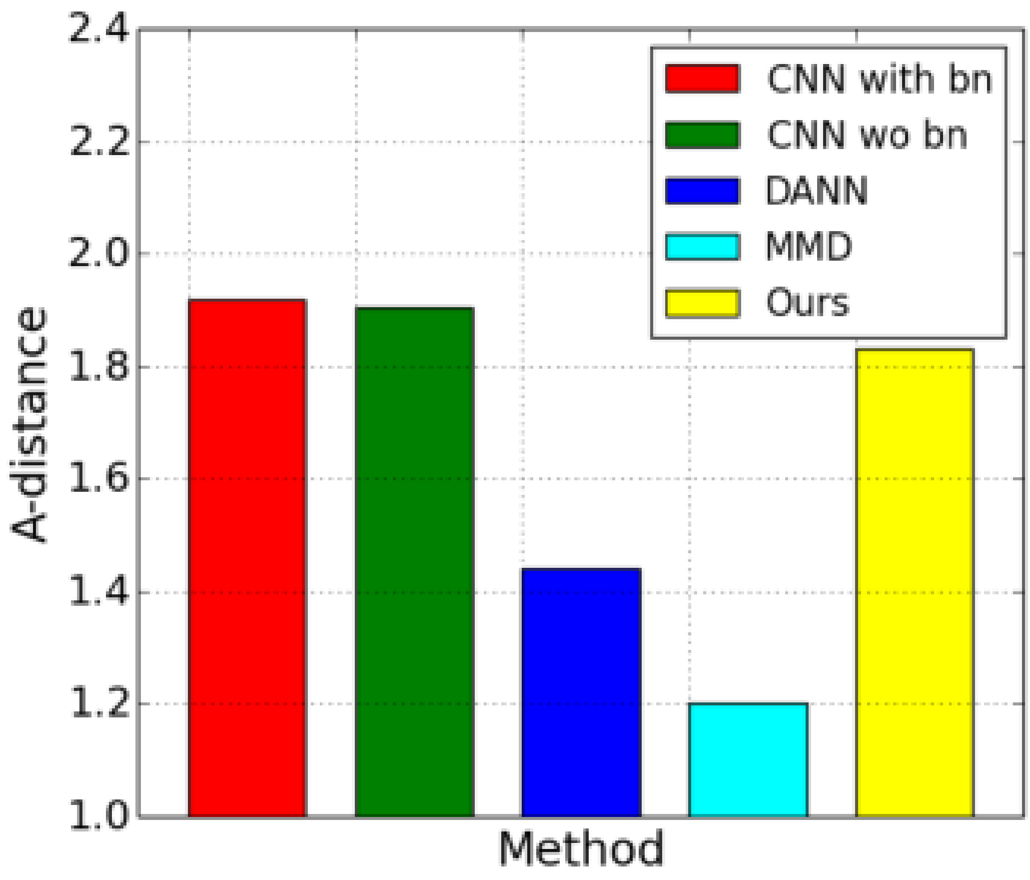


(e) SYNSIGNS→GTSRB



(f) Comparision of accuracy of three networks on SVHN→MNIST

Experiments



(g) $\mathcal{A}$-distance in MNIST→MNIST-M

$$\hat{d}_{\mathcal{A}} = 2(1 - 2\epsilon)$$

ϵ is a generalization error

Thanks