Label-Efficient Learning of Transferable Representations across Domains and Tasks

Zelun Luo[1], Yuliang Zou[2], Judy Hoffman[3], Li Fei-Fei[1]

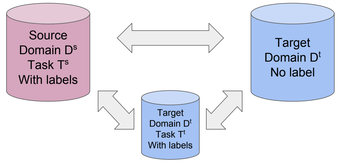
[1] Stanford University [2] Virginia Tech [3] University of California, Berkeley

Conference on Neural Information Processing Systems 2017

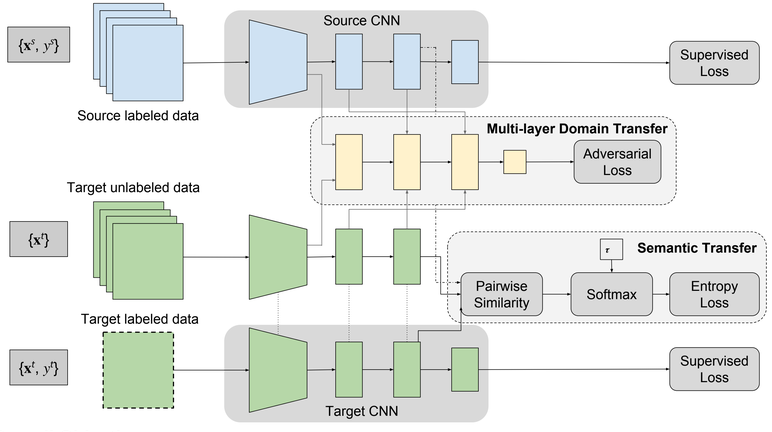
Project Page

Motivation

- Traditional machine learning methods assume that training and test data are drawn from the same feature space and distribution.
- However, this assumption usually does not hold in real-world scenarios.
- Relying on **a large amount of labeled data**, fine-tuning is a baseline method to address these issues:
  - Task transfer
  - Domain shift
- In this work, we propose a **label-efficient** learning approach that learns a transferable representation across domains and tasks.



Overview



Overall Objective

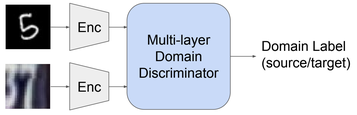
$$\mathcal{L}(\mathcal{X}^{\mathcal{S}}, \mathcal{Y}^{\mathcal{S}}, \mathcal{X}^{\mathcal{T}}, \mathcal{Y}^{\mathcal{T}}, \tilde{\mathcal{X}^{\mathcal{T}}}) = \mathcal{L}_{\text{sup}}(\mathcal{X}^{\mathcal{T}}, \mathcal{Y}^{\mathcal{T}}) + \alpha\mathcal{L}_{DT}(\mathcal{X}^{\mathcal{S}}, \tilde{\mathcal{X}^{\mathcal{T}}}) + \beta\mathcal{L}_{ST}(\mathcal{X}^{\mathcal{S}}, \mathcal{X}^{\mathcal{T}}, \tilde{\mathcal{X}^{\mathcal{T}}})$$

References

[1] Vinyals et al. Matching networks for one shot learning. In *NIPS 2016*
[2] Simonyan and Zisserman. Two-stream convolutional networks for action recognition in videos. In *NIPS 2014*
[3] Ganin et al. Domain-adversarial training of neural networks. In *JMLR 2016*
[4] Tzeng et al. Simultaneous deep transfer across domains and tasks. In *ICCV 2015*
[5] Tzeng et al. Adversarial discriminative domain adaptation. In *CVPR 2017*

Multi-layer Domain Transfer

- **Data**:
  - A large amount of source domain data
  - A large amount of target domain data
- **Objective**: Learn a universal embedding function to map both domain into such a subspace, that the discriminator cannot reliably predict their domain labels
- **Method**: Multi-layer Adversarial Domain Discriminator



Discriminator

$$\mathcal{L}_{DT}^{D} = -\mathbb{E}_{\mathbf{x}^s \sim \mathcal{X}^{\mathcal{S}}}\left[\log \mathbf{d}_l^s\right] - \mathbb{E}_{\mathbf{x}^t \sim \mathcal{X}^{\mathcal{T}}}\left[\log(1 - \mathbf{d}_l^t)\right]$$

Encoder

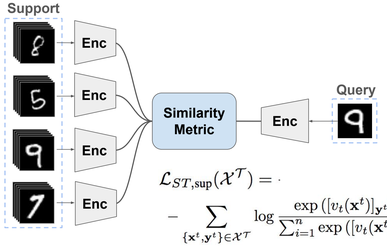
$$\mathcal{L}_{DT}^{E^t} = -\mathbb{E}_{\mathbf{x}^s \sim \mathcal{X}^{\mathcal{S}}}\left[\log(1 - \mathbf{d}_l^s)\right] - \mathbb{E}_{\mathbf{x}^t \sim \mathcal{X}^{\mathcal{T}}}\left[\log \mathbf{d}_l^t\right]$$

Semantic Transfer

- **Data**:
  - A large amount of labeled source domain data + A small amount of labeled target domain data
  - A large amount of unlabeled target domain data + A small amount of labeled target domain data
- **Objective**: Representation learning
- **Method**:
  - Episodic training + metric learning + entropy minimization

$$\mathcal{L}_{ST}(\mathcal{X}^{\mathcal{S}}, \mathcal{X}^{\mathcal{T}}, \tilde{\mathcal{X}^{\mathcal{T}}}) = \mathcal{L}_{ST}(\tilde{\mathcal{X}^{\mathcal{T}}}, \mathcal{X}^{\mathcal{S}}) + \mathcal{L}_{ST,\text{sup}}(\mathcal{X}^{\mathcal{T}}) + \mathcal{L}_{ST,\text{unsup}}(\tilde{\mathcal{X}^{\mathcal{T}}}, \mathcal{X}^{\mathcal{T}})$$

Episodic Metric learning



$$\mathcal{L}_{ST,\text{sup}}(\mathcal{X}^{\mathcal{T}}) = \cdot$$
$$- \sum_{\{\mathbf{x}^t, \mathbf{y}^t\} \in \mathcal{X}^{\mathcal{T}}} \log \frac{\exp\left([v_t(\mathbf{x}^t)]_{\mathbf{y}^t}\right)}{\sum_{i=1}^{n} \exp\left([v_t(\mathbf{x}^t)]_i\right)}$$

Entropy Minimization

Source ⇒ Target

$$\mathcal{L}_{ST}(\tilde{\mathcal{X}^{\mathcal{T}}}, \mathcal{X}^{\mathcal{S}}) = \sum_{\tilde{\mathbf{x}}^t \in \tilde{\mathcal{X}^{\mathcal{T}}}} H(\sigma(v_s(\tilde{\mathbf{x}}^t)/\tau))$$

Target ⇒ Target

$$\mathcal{L}_{ST,\text{unsup}}(\tilde{\mathcal{X}^{\mathcal{T}}}, \mathcal{X}^{\mathcal{T}}) = \sum_{\tilde{\mathbf{x}}^t \in \tilde{\mathcal{X}^{\mathcal{T}}}} H(\sigma(v_t(\tilde{\mathbf{x}}^t)/\tau))$$

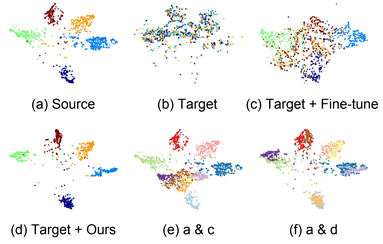
Experiments and Results

- Different tasks in source and target domain
- Domain shift between source and target domain
- Limited labeled examples in target domain (k examples in each class)

SVHN 0-4 ⇒ MNIST 5-9

Methods	k=2	k=3	k=4	k=5
Target only	0.642±0.026	0.771±0.015	0.801±0.010	0.840±0.013
Fine-tune	0.612±0.020	0.779±0.018	0.802±0.016	0.830±0.011
Matching net [1]	0.469±0.019	0.455±0.014	0.566±0.013	0.513±0.023
Fine-tuned matching net	0.645±0.019	0.755±0.024	0.793±0.013	0.827±0.011
Ours: fine-tune + adv.	0.702±0.020	0.800±0.013	0.804±0.014	0.831±0.013
Ours: full model	**0.917±0.007**	**0.936±0.006**	**0.942±0.006**	**0.950±0.004**

t-SNE



Ablation Study:
Unsupervised Domain Adaptation

Methods	Accuracy
Source only	0.601±0.011
Gradient reversal [3]	0.739
Domain confusion [4]	0.681±0.003
ADDA [5]	0.760±0.018
Ours	**0.810±0.003**

Image object recognition ⇒ video action recognition

Methods	k=3	k=5	k=10	All
Target only (img.)	0.098±0.003	0.126±0.022	0.100±0.035	-
Target only (vid.)	0.105±0.003	0.133±0.024	0.106±0.038	-
Fine-tune (img.)	0.380±0.013	0.486±0.012	0.529±0.039	-
Fine-tune (vid.)	0.406±0.05	0.523±0.010	0.568±0.042	-
Two-stream spatial [2]	-	-	-	0.708 - 0.720
Ours (img.)	0.393±0.006	0.459±0.013	0.523±0.002	-
Ours (vid.)	**0.467±0.007**	**0.545±.0014**	**0.620±0.005**	**-**