Adversarial Multiple Source Domain Adaptation

Han Zhao[†], Shanghang Zhang[†], Guanhang Wu[†], João Costeira[♭], José Moura[†] & Geoffrey J. Gordon[†]

[†]Carnegie Mellon University, [♭]Instituto Superior Técnico

{hzhao1,shanghaz,guanhanw,moura,ggordon}@andrew.cmu.edu, jpc@isr.ist.utl.pt

Summary

Unsupervised Domain adaptation: Source $\neq$ Target

Target camera: without label

Source cameras: with label

- We theoretically analyze the multiple source domain adaptation problem under both classification and regression settings.
- We propose two models using adversarial neural networks for multiple source domain adaptation.
- We conduct extensive experiments on sentiment analysis, digit recognition and vehicle counting problems, and we achieve superior adaptation performances on all the tasks.

Preliminary

Given hypothesis class $\mathcal{H}$ and $\mathcal{A}_\mathcal{H} := \{h^{-1}(1) \mid h \in \mathcal{H}\}$,
$\mathcal{H}$-divergence is: $d_\mathcal{H}(\mathcal{D}, \mathcal{D}') := 2\sup_{A\in\mathcal{A}_\mathcal{H}} |\Pr_\mathcal{D}(A) - \Pr_{\mathcal{D}'}(A)|$.
Generalization bound for single-source-single-target binary classification (Blitzer et al. NIPS' 08), using m instances, with probability $\geq 1-\delta$, $\forall h \in \mathcal{H}$:

$$\varepsilon_T(h) \leq \hat{\varepsilon}_S(h) + \frac{1}{2}d_{\mathcal{H}\Delta\mathcal{H}}(\widehat{\mathcal{D}}_S, \widehat{\mathcal{D}}_T) + \lambda + O\left(\sqrt{\frac{d\log(m/d)+\log(1/\delta)}{m}}\right)$$

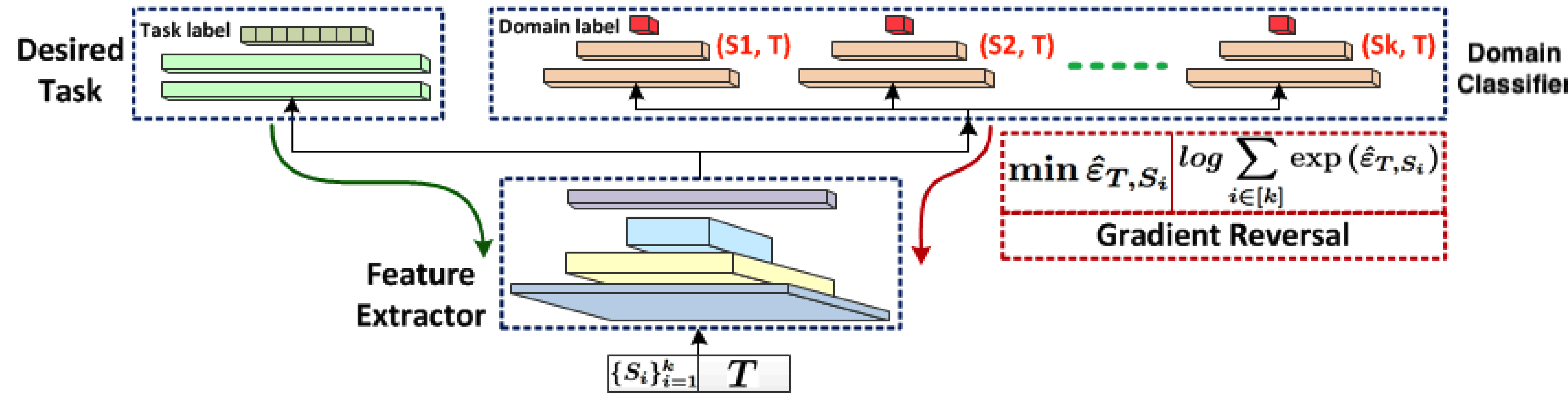
- $\hat{\varepsilon}_S(h)/\varepsilon_T(h)$: empirical/population source/target binary classification error.
- $d_{\mathcal{H}\Delta\mathcal{H}}(\widehat{\mathcal{D}}_S, \widehat{\mathcal{D}}_T)$: empirical $\mathcal{H}\Delta\mathcal{H}$-divergence between source and target domains.
- $\lambda := \min_{h'\in\mathcal{H}} \varepsilon_S(h') + \varepsilon_T(h')$.

A naive extension to k source domains with union bound:

$$\varepsilon_T(h) \leq \max_{i\in[k]}\left\{\hat{\varepsilon}_{S_i}(h) + \frac{1}{2}d_{\mathcal{H}\Delta\mathcal{H}}(\widehat{\mathcal{D}}_T; \widehat{\mathcal{D}}_{S_i}) + \lambda_i\right\}$$
$$+O\left(\sqrt{\frac{1}{m}\left(\log\frac{k}{\delta} + d\log\frac{m}{d}\right)}\right)$$

Carnegie Mellon University

Models and Algorithms



Theorem (informal): $\mathcal{H}$ is a hypothesis class and $VC\dim(\mathcal{H}) = d/P\dim(\mathcal{H}) = d$. $\widehat{\mathcal{D}}_T$ and $\{\widehat{\mathcal{D}}_{S_i}\}_{i=1}^k$ are m samples generated from each domain. Define $\tilde{\mathcal{H}} := \{\mathbb{I}_{|h(x)-h'(x)|>t} : h, h' \in \mathcal{H}, 0 \leq t \leq 1\}$ to be the set of threshold functions. Then for $\forall\alpha \in \Delta^k$, for $\delta \in (0,1)$, $\forall h \in \mathcal{H}$, w.p. $\geq 1-\delta$:

Classification:

$$\varepsilon_T(h) \leq \sum_{i\in[k]}\alpha_i\left(\hat{\varepsilon}_{S_i}(h) + \frac{1}{2}d_{\mathcal{H}\Delta\mathcal{H}}(\widehat{\mathcal{D}}_T; \widehat{\mathcal{D}}_{S_i})\right) + \lambda_\alpha + \widetilde{O}\left(\sqrt{\frac{d\log(1/\delta)}{km}}\right),$$

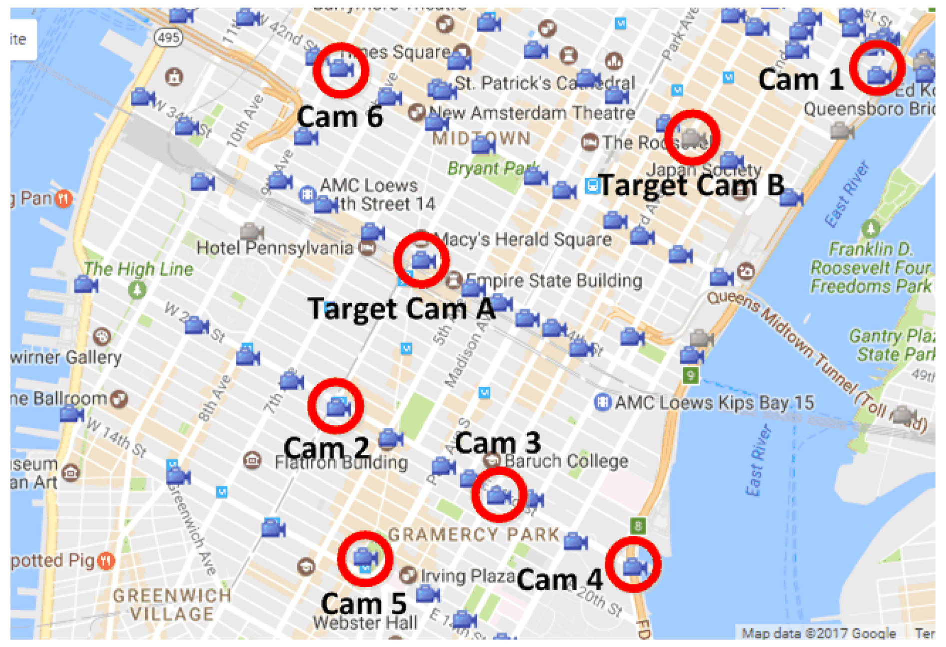
and

Regression:

$$\varepsilon_T(h) \leq \sum_{i\in[k]}\alpha_i\left(\hat{\varepsilon}_{S_i}(h) + \frac{1}{2}d_{\tilde{\mathcal{H}}}(\widehat{\mathcal{D}}_T; \widehat{\mathcal{D}}_{S_i})\right) + \lambda_\alpha + \widetilde{O}\left(\sqrt{\frac{d\log(1/\delta)}{km}}\right)$$

Experiments

Datasets:



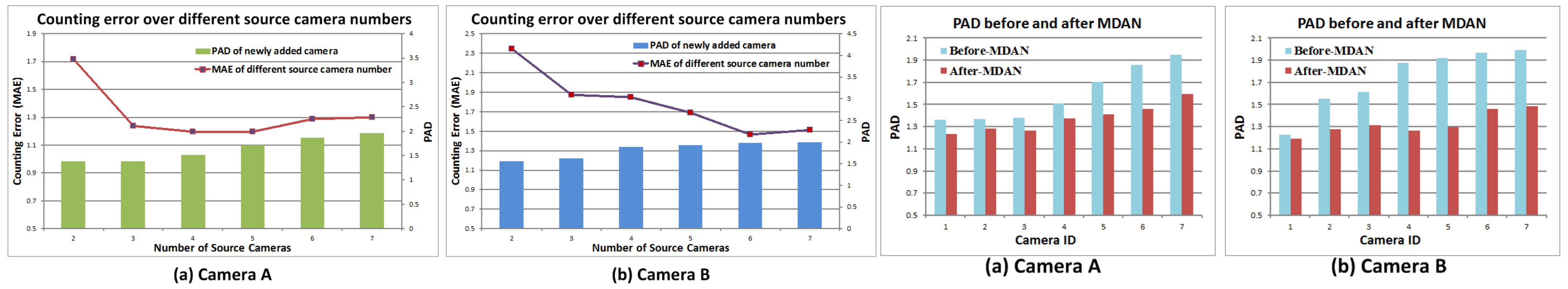
- WebCamT (Zhang et al, CVPR' 17), public dataset for vehicle counting. Image resolution: 352×240.
- 8 cameras. 6 as sources and each of the rest two as target. 2,000 images for each domain.

Methods:

- FCN: Fully-convolutional NN, without domain adaptation.
- DANN: Combine all sources into one, with adversarial learning.

Table: Counting error statistics. S is the number of source cameras; T is the target camera id.

S	T	Ours		DANN	FCN	T	Ours		DANN	FCN
		Hard-Max	Soft-Max				Hard-Max	Soft-Max		
2	A	1.8101	**1.7140**	1.9490	1.9094	B	2.5059	**2.3438**	2.5218	2.6528
3	A	1.3276	**1.2363**	1.3683	1.5545	B	1.9092	**1.8680**	2.0122	2.4319
4	A	1.3868	**1.1965**	1.5520	1.5499	B	**1.7375**	1.8487	2.1856	2.2351
5	A	1.4021	**1.1942**	1.4156	1.7925	B	1.7758	**1.6016**	1.7228	2.0504
6	A	1.4359	**1.2877**	2.0298	1.7505	B	1.5912	**1.4644**	1.5484	2.2832
7	A	1.4381	**1.2984**	1.5426	1.7646	B	1.5989	**1.5126**	1.5397	1.7324



Reference

- Blitzer et al., *Learning bounds for domain adaptation*, NIPS 2010.
- Zhang et al., *Understanding traffic density from large-scale web camera data*, CVPR 2017.