Adversarially Constrained Interpolation for Unsupervised Domain Adaptation

Mohamed Azzam, Aurele Tohokantche Gnanha and Hau-San Wong

Department of Computer Science
City University of Hong Kong

Si Wu

School of Comp. Sc. and Eng.
South China University of Technology

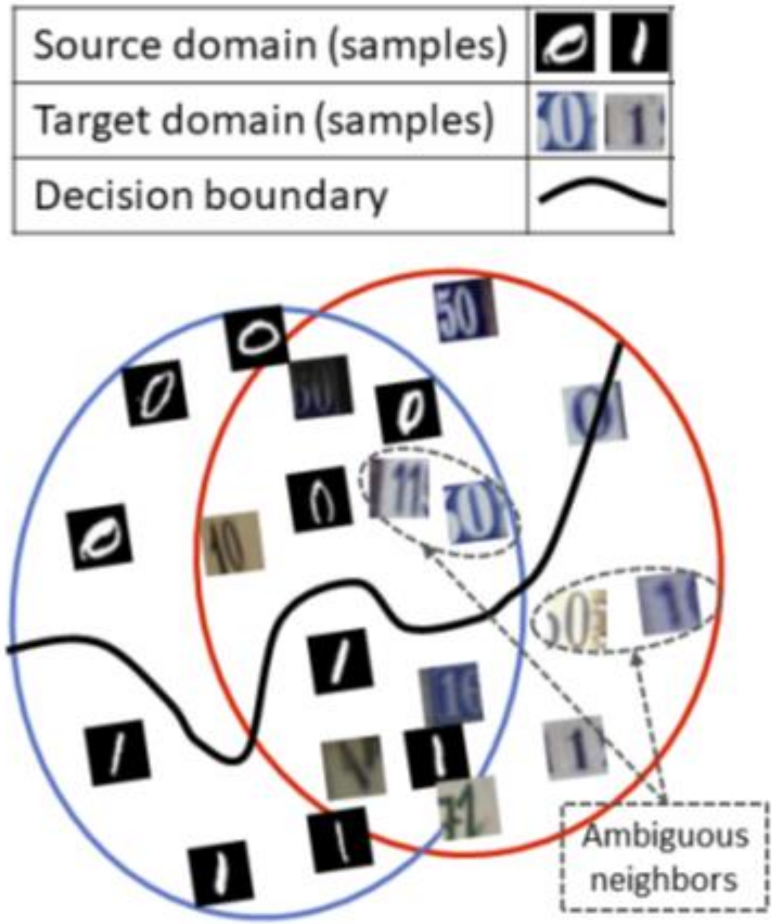
Unsupervised Domain Adaptation (UDA)

- UDA aims at adapting models trained on labeled data from the source domain to a completely unlabeled data from the target domain.
- Possible solution: learning a domain-invariant representation (Domain adversarial training).

MNIST → SVHN

Challenges of Domain Adversarial Learning Methods

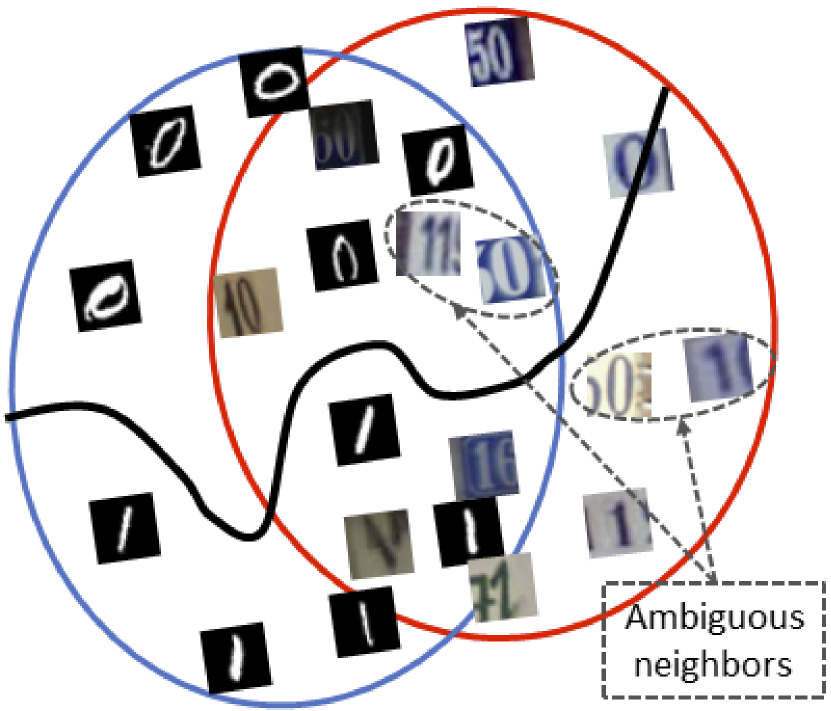
- Data deficiency in both domains
- Target-domain samples of different classes may become neighbors in the feature space



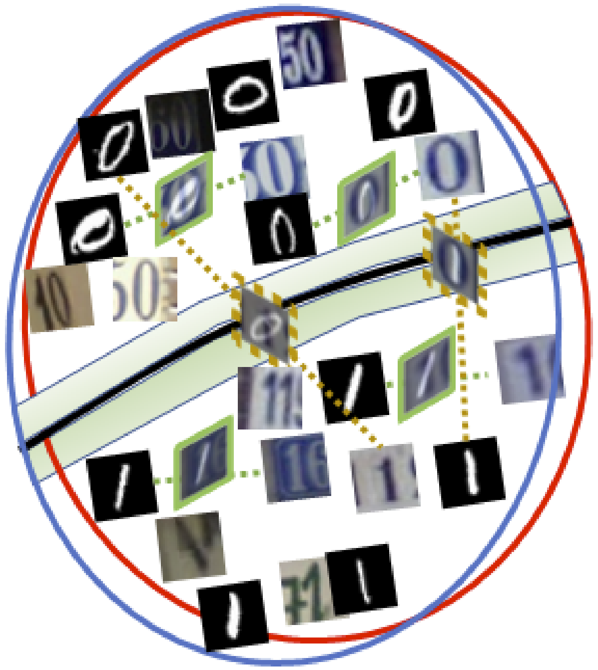
Previous Domain Adaptation Methods

Proposed Method

Source domain (samples)		Dom. Mixup samples (same class)	
Target domain (samples)		Dom. Mixup sample (diff. class)	
Decision boundary		Smooth region	



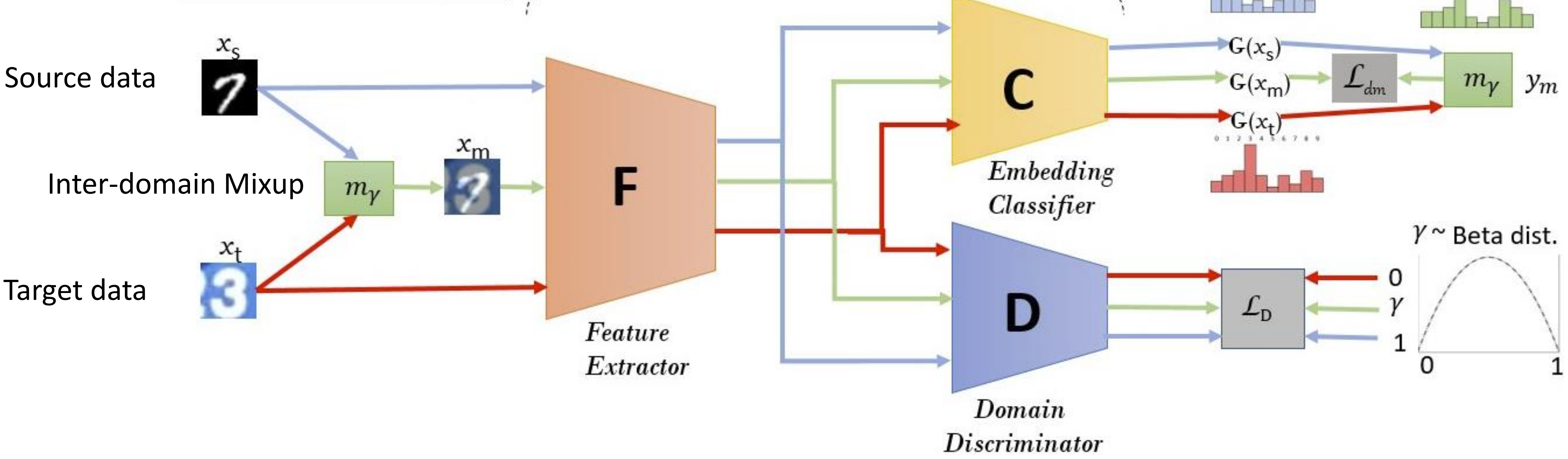
Previous Domain Adaptation Methods



Proposed Method

Proposed Framework

Source input:	x_s	Mixup:	m_γ
Target input:	x_t	Mixup ratio:	γ
Domain mixup input:	x_m		
Domain mixup logit:	y_m		



Inter-domain mixup

Classifier Training:

$$\mathcal{L}_{dm} = -E_{(x,x')\sim(\chi_s,\chi_t)}\left[y_m^T \ln\left(G_\sigma(x_m)\right)\right]$$

Where

$$x_n = m_\gamma(x, x')$$
$$y_m = \sigma\left(m_\gamma\left(G(x), G(x')\right)\right)$$
$$m_\gamma(x_1, x_2) = \gamma x_1 + (1-\gamma)x_2$$

Domain Discriminator

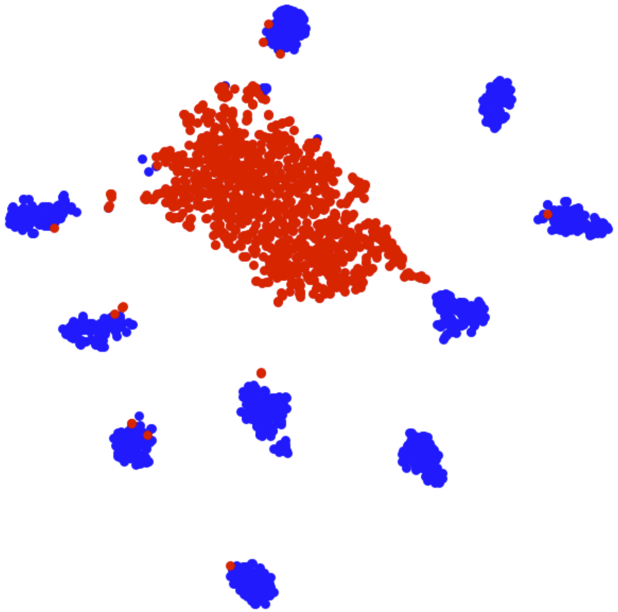
$$\mathcal{L}_D = \mathcal{L}'_D - E_{(x,x')\sim(\chi_s,\chi_t)}\left[\gamma \ln\left(D\left(F(x_m)\right)\right) + (1-\gamma)\ln\left(1 - D\left(F(x_m)\right)\right)\right]$$

Experimental Results

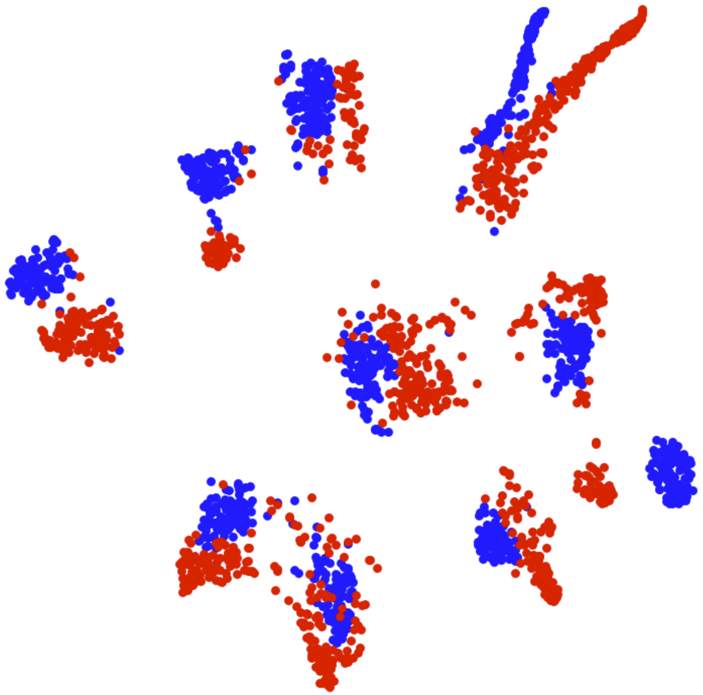
Source data Target data	MNIST SVHN	SVHN MNIST	MNIST MNIST-M	CIFAR-10 STL-10	STL-10 CIFAR-10
Source-only	40.9	82.4	59.9	76.3	63.6
DANN	35.7	73.9	77.4	-	-
VADA	73.3	97.9	95.7	80.0	73.5
Co-DA	81.7	98.8	98.0	81.4	76.4
VMT	85.2	98.9	98.0	82.0	78.5
IIMT	-	97.3	**99.5**	83.1	**81.6**
Ours	**88.7**	**99.0**	98.1	**83.7**	79.7
VADA + DIRT-T	76.5	99.4	98.7	-	75.3
Ours + DIRT-T	**95.9**	**99.6**	**98.9**	-	**82.9**

T-SNE Plot: MNIST → SVHN

- Using the output of the last hidden layer in C



Source-only



Ours

Source data	
Target data	

Ablative Factors in Our Model

Source data Target data	MNIST SVHN	SVHN MNIST	MNIST MNIST-M	CIFAR-10 STL-10	STL-10 CIFAR-10
Source-only	40.9	82.4	59.9	76.3	63.6
Ours W/O Mix Fact. Pred.	83.8	98.3	97.6	83.7	79.7
Ours W/O Dom. Mixup	85.8	98.6	97.8	81.9	78.6
Ours	**88.7**	**99.0**	**98.1**	**83.7**	**79.7**

Conclusion

- Inter-domain mixup encourages binding samples of the same class regardless of their domain

- Enforcing the domain discriminator to predict the mixup ratio is vital for smoothing the feature manifold and facilitate training on inter-domain mixup samples.

Thank You