DIAL: Domain-Invariant Adversarial Learning on Digits Domains

Daniel Zeng

BERKELEY ARTIFICIAL INTELLIGENCE RESEARCH

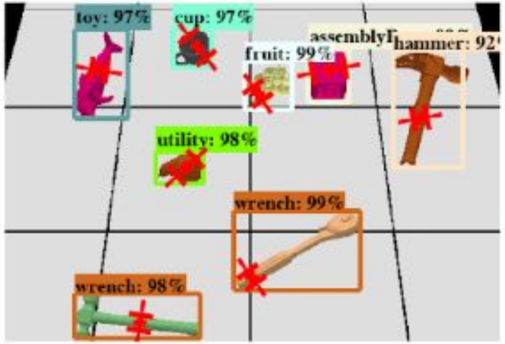
Agenda

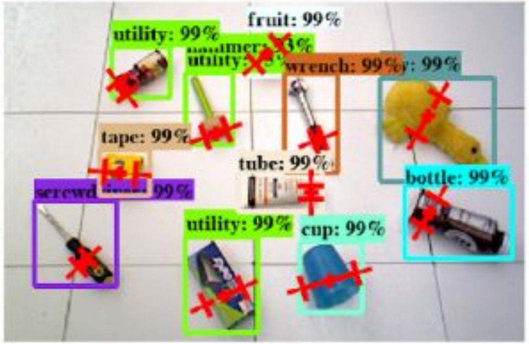
- Recap
- DIAL Approach
- Benchmarking on Digits Domains

Sim-to-real transfer

- Object recognition/grasp planning in decluttering by sim-to-real transfer


- Deformable object manipulation





Domain Adaptation Problem Formulation

-- Simulator or source domain: $\langle D_S, \pi_S \rangle \quad \{(\mathbf{x_i^S}, \mathbf{u_i^S})\}_{\mathbf{i=1}}^{\mathbf{N_S}}$

-- Real or target domain: $\langle D_T, \pi_T \rangle \quad \{(\mathbf{x_i^T}, \mathbf{u_i^T})\}_{\mathbf{i=1}}^{\mathbf{N_T}}$

-- Learn policy

$$\pi : \mathcal{X} \rightarrow \mathbb{R}^{|\mathcal{U}|} \qquad \{0, 1 \ldots . K\} \text{ or } \mathbb{R}^K$$

such that the loss function on target domain is minimized

$$\mathcal{L}_{D_T}(\pi, \pi^*) = \mathbb{E}_{\mathbf{x} \sim \mathbf{D_T}} \left[\mathbb{I}(\pi(\mathbf{x}) \neq \pi^*(\mathbf{x})) \right]$$

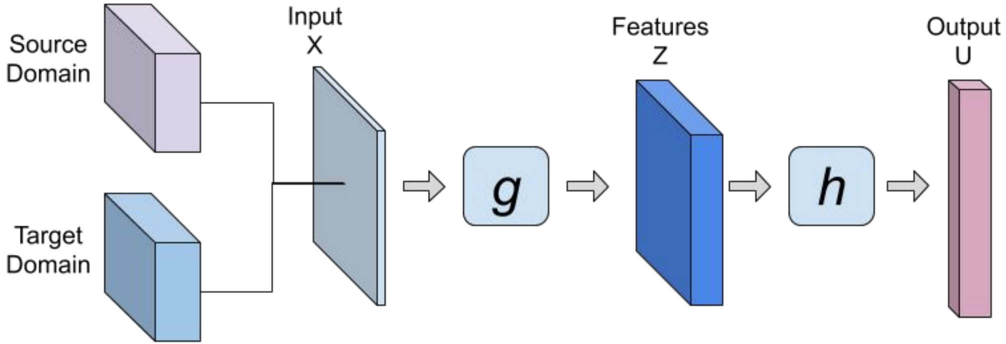
Usually *few or no labels* for the target domain: thus need to learn how to adapt from source to target

DIAL Approach

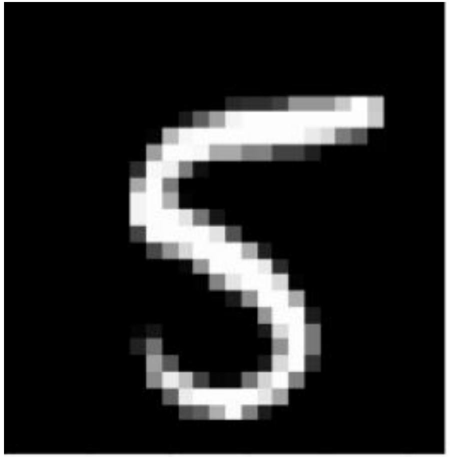
- ***Metric Learning:*** Deep embedding space that pulls together samples from same classes, while pushes away samples from other classes

- ***Domain Alignment:*** Find a *g* such that features *Z* look similar across source and target domains

- ***Class Alignment:*** Find *g* and *h* such that features Z look similar across class categories



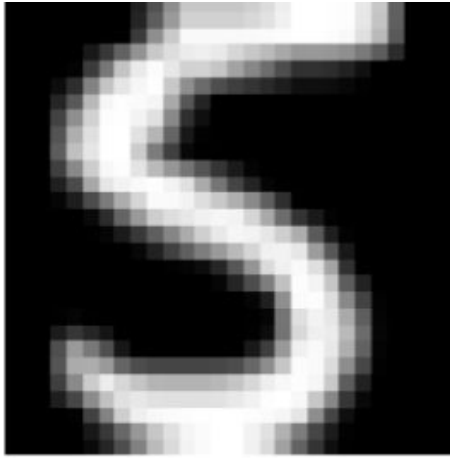
Datasets on Digits Benchmark:



MNIST

MNISTM

SVHN



USPS

- MNIST: Modified National Institute of Standards and Technology database
- MNISTM: MNIST Modified
- SVHN: Street View House Numbers
- USPS: US Post Office Zip Code Data

Unsupervised Baselines

- Source Only
- Domain adversarial neural networks (DANN) [1]
- Reconstruction based adaptation
- Associative adaptation [2]
- Maximum classifier discrepancy (MCD) [3]

[1] Ganin et. al, Domain-Adversarial Training of Neural Networks. 2015
[2] Haeusser et. al, Associative Domain Adaptation. 2017
[3] Saito et. al, Maximum Classifier Discrepancy for Unsupervised Domain Adaptation. 2018

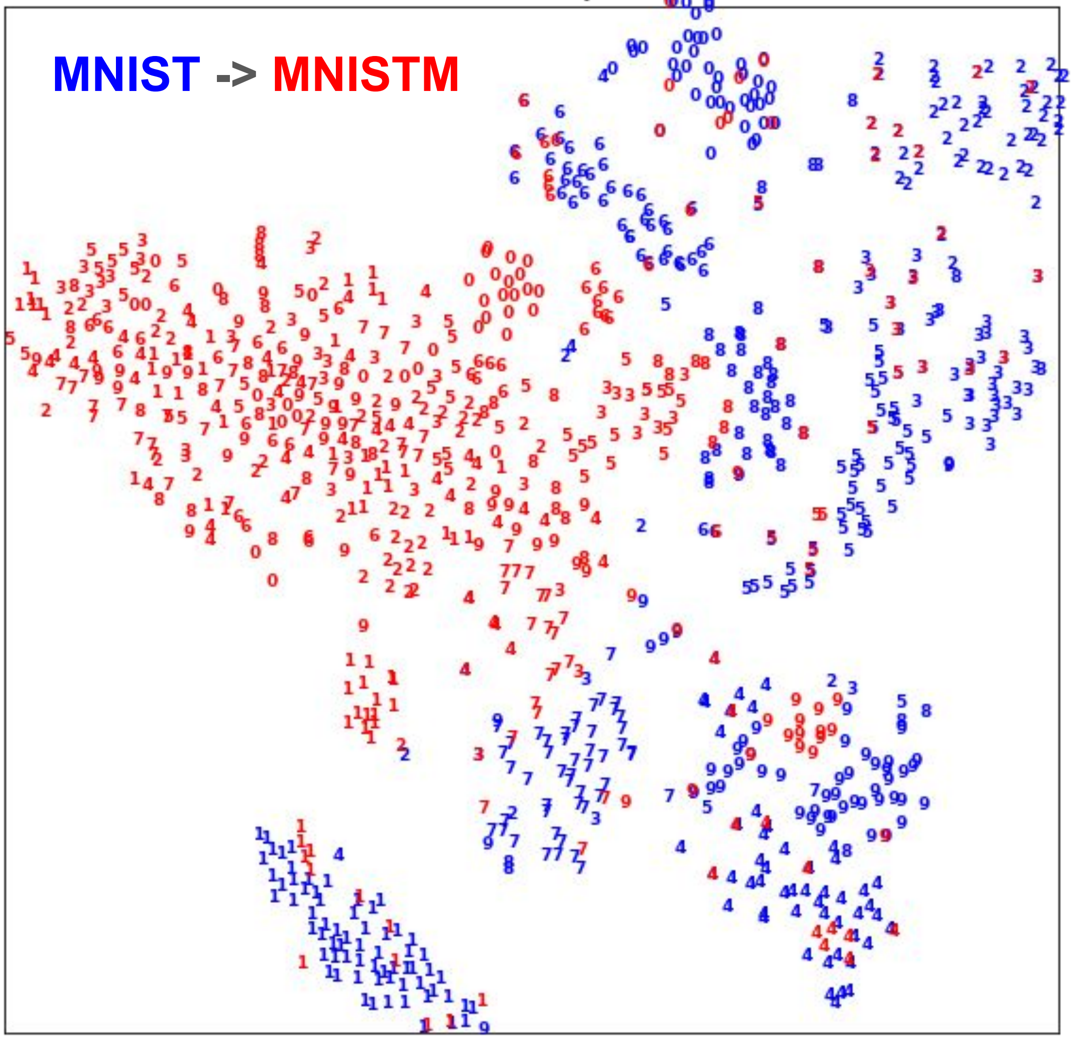
Baselines: Source

Source only classifier

Only trained on MNIST, no alignment

Thus, red exist in different location than source, though some red are in blue clusters

Accuracy = 56.2%



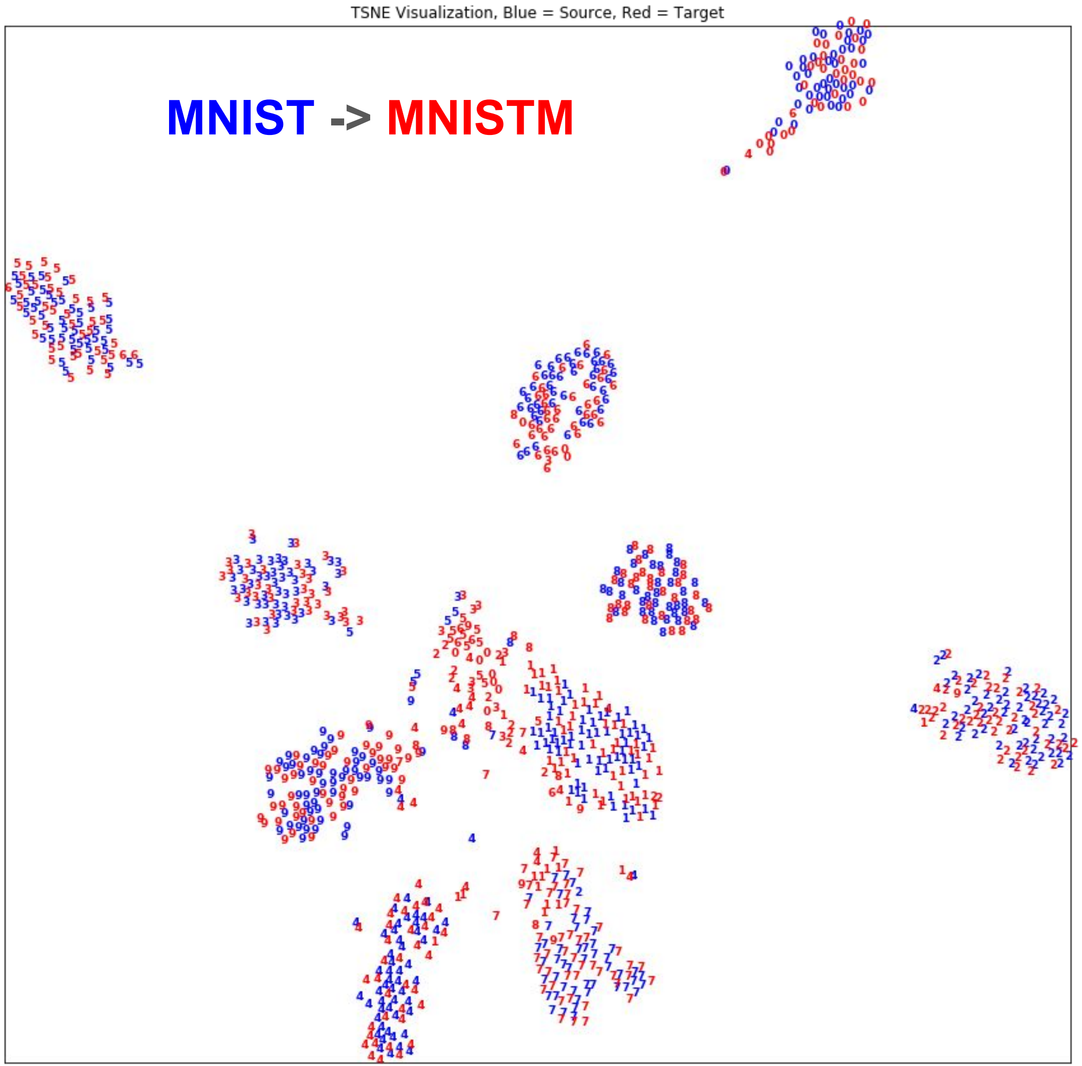
Baselines: DANN

DANN + Source classifier:

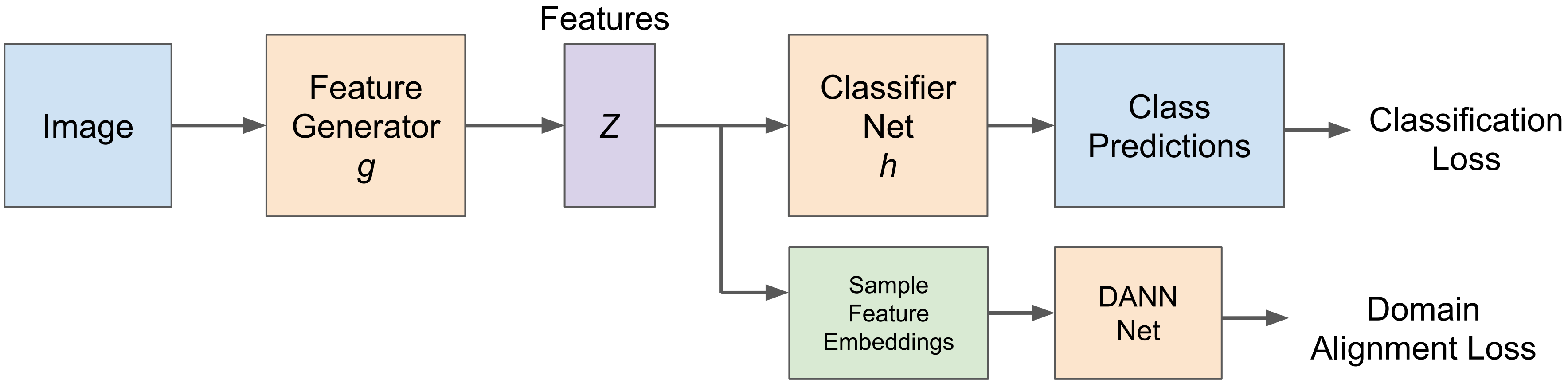
Domain adversarial neural networks

Domain level alignment with DANN, but weaker Class alignment

Accuracy = 79.8%



DANN Architecture



Baselines: Reconstruction

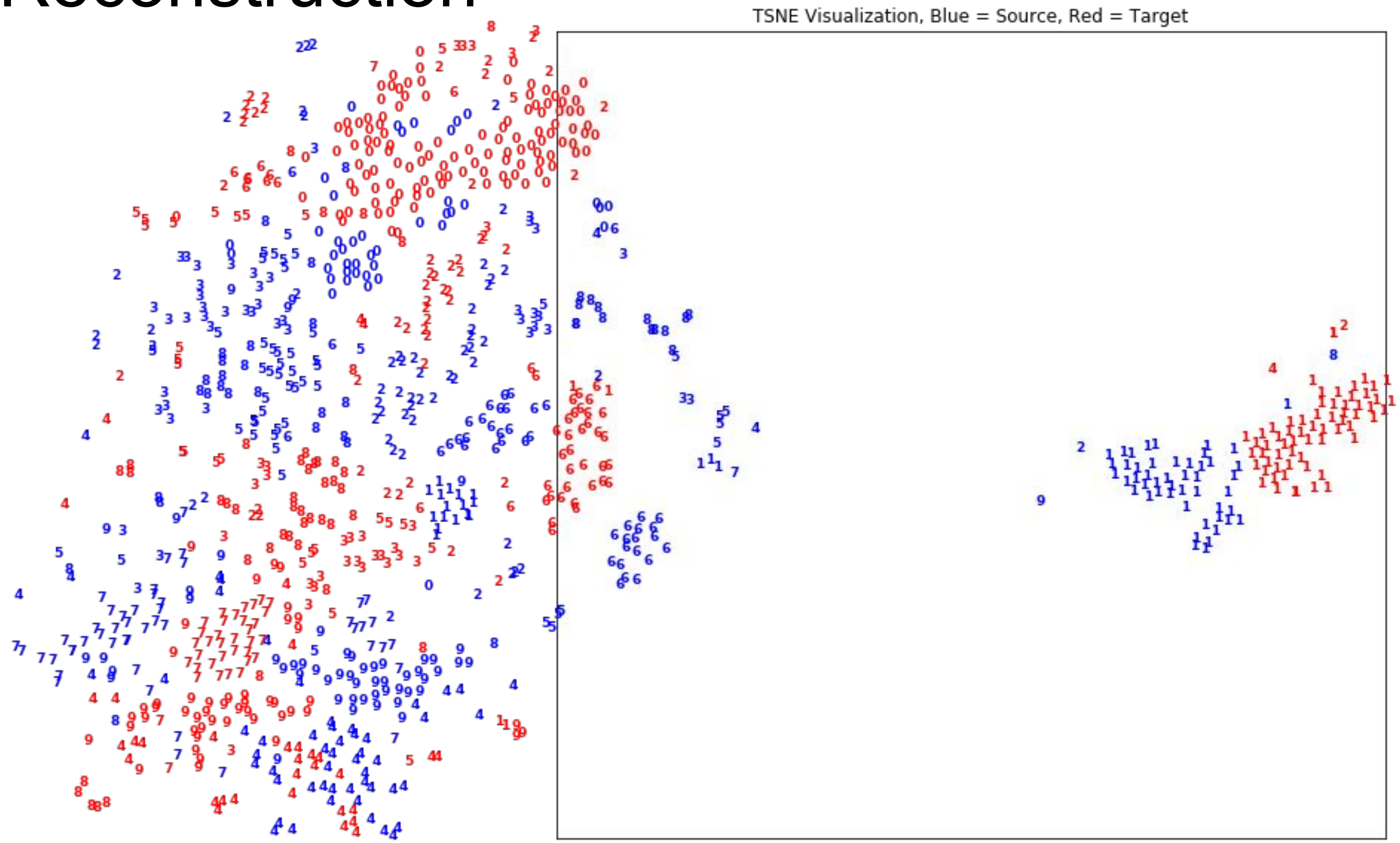
Reconstruction based adaptation:

DANN + Source classifier + Recon.

Reconstruction does not improve

Accuracy = 53.4%

MNIST -> MNISTM

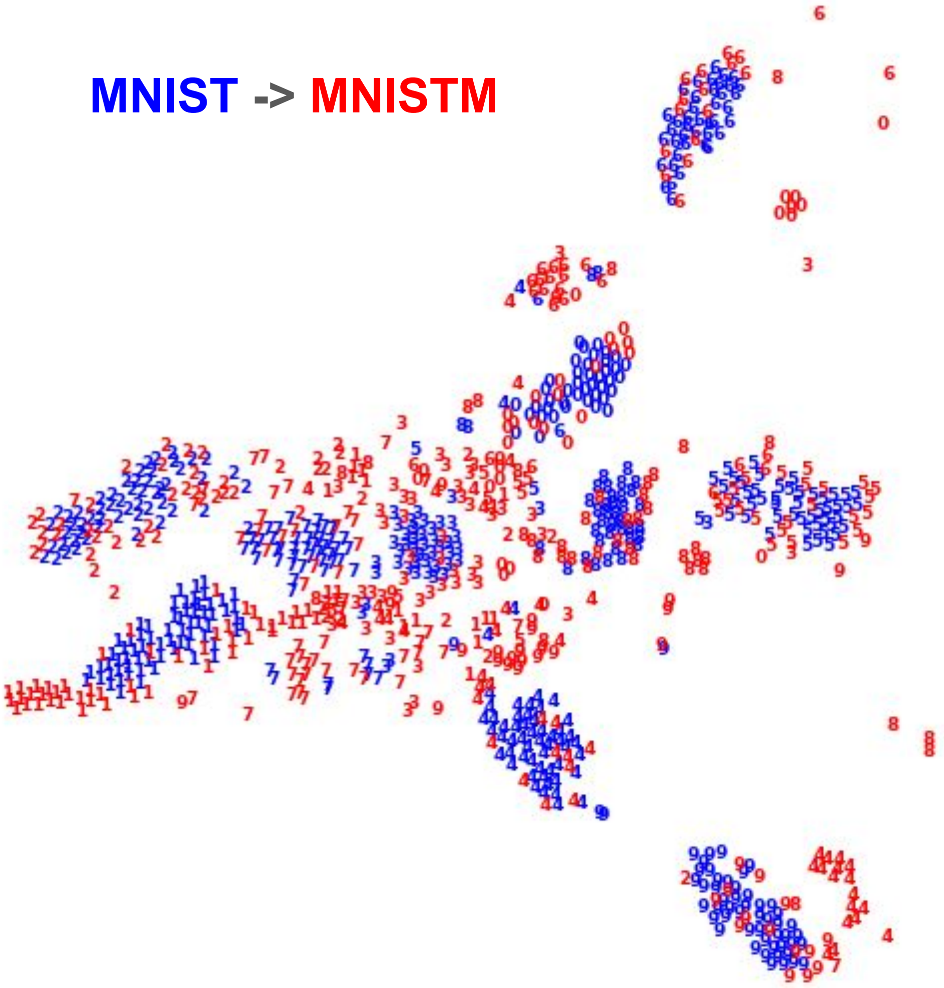


Baselines: MCD

Maximum classifier discrepancy:

Trains 2 different classifiers, and moves target to fit within support region of them

Accuracy = 77%



Results: Unsupervised baselines

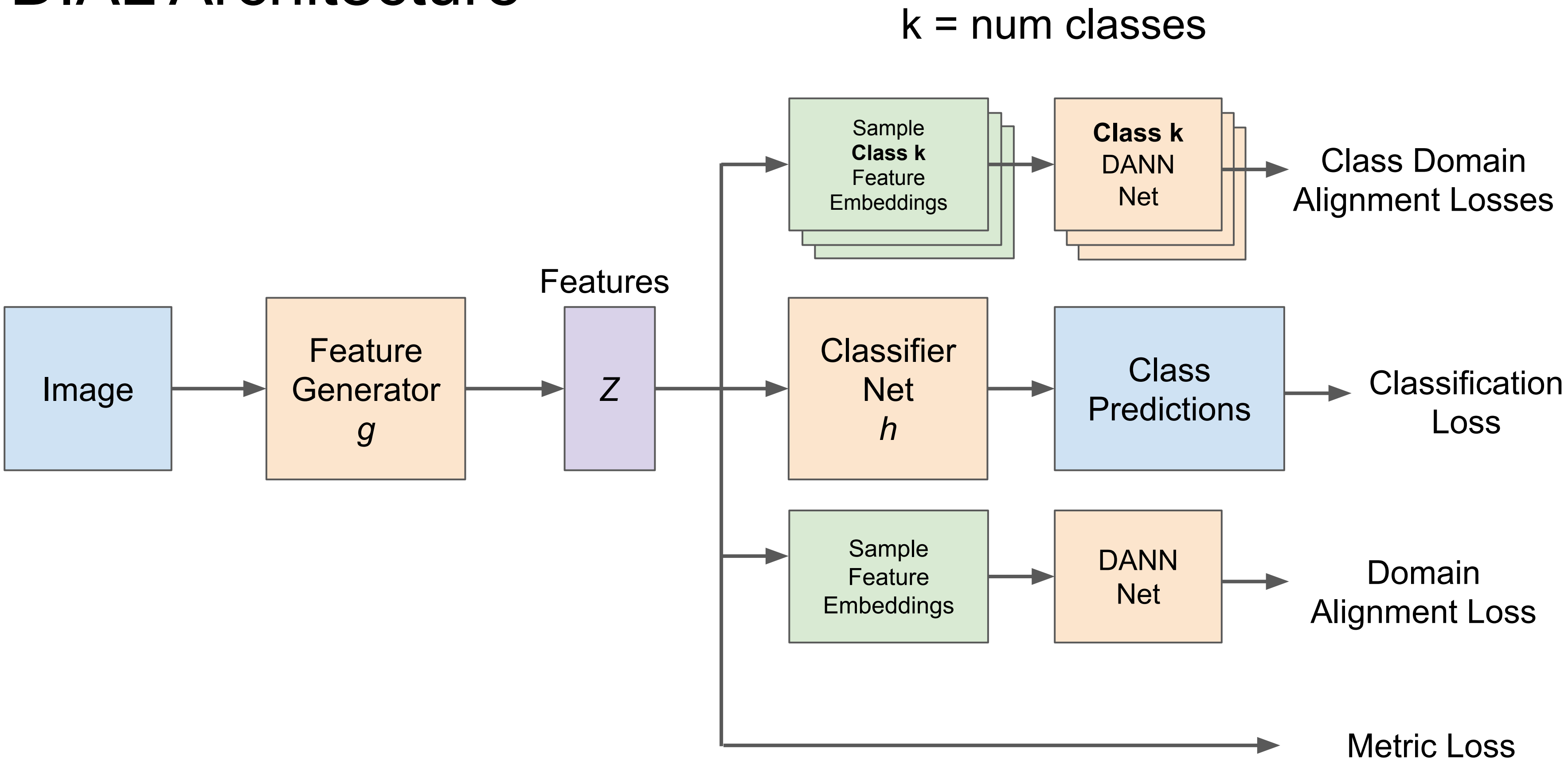
X -> Y: means X is source domain, Y is target domain

Test Accuracy	MNIST-> MNISTM	MNIST-> USPS	SVHN-> MNIST	USPS-> SVHN	USPS-> MNIST
Source Classifier + DANN	0.798	0.873	0.674	0.153	0.751
Triplet + DANN	0.704	0.886	0.805	0.180	0.832
Associative DA	**0.895**	0.212	0.960	0.256	0.570
MCD	0.770	**0.941**	**0.978**	**0.288**	**0.932**
Recon. + DANN	0.534	0.823	0.627	0.155	0.715

Semi-Supervised scenario

- Why Semi-Supervised?
- Typical, we do have at least *few* labels for target dataset domain
- 1, 5, or 10 labels per class can greatly increase target accuracy, as demonstrated in results
- Semi-Supervised is needed when unsupervised fails to perform as well
- Difficult domain adaptation pairs: USPS -> SVHN

DIAL Architecture

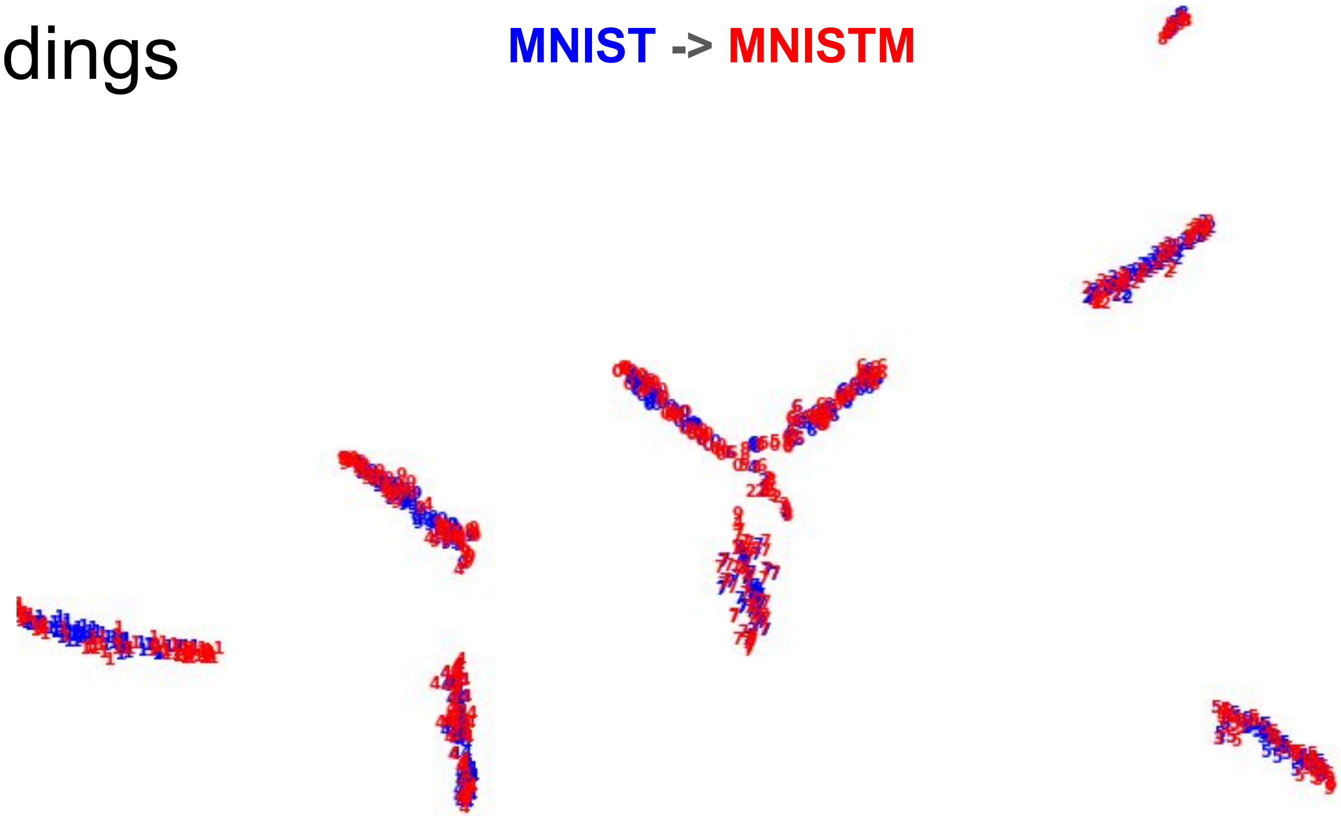


DIAL Embeddings

Class-DANN + Metric learning + DANN + Source classifier

Accuracy with m = 5: 94.1% (Semi-supervised)

Stronger domain and class level alignment



Results: Semi-supervised, *m = 1,* labeled target per class

Test Accuracy	MNIST-> MNISTM	MNIST-> USPS	SVHN-> MNIST	USPS-> SVHN	USPS-> MNIST
DANN Baseline	0.758	0.872	0.764	0.151	0.777
DIAL: Source Classifier + DANN + Class DANN	0.779	**0.906**	0.765	0.351	0.893
DIAL: Source Classifier + Triplet + DANN + Class DANN	**0.786**	0.894	**0.773**	**0.400**	**0.949**

Results: Semi-supervised, *m = 5,* labeled target per class

Test Accuracy	MNIST-> MNISTM	MNIST-> USPS	SVHN-> MNIST	USPS-> SVHN	USPS-> MNIST
DANN Baseline	0.763	0.920	0.793	0.252	0.85
DIAL: Source Classifier + DANN + Class DANN	**0.941**	**0.946**	0.864	0.610	0.945
DIAL: Source Classifier + Triplet + DANN + Class DANN	0.795	0.934	**0.868**	**0.682**	**0.967**

Results: Semi-supervised, *m = 10,* labeled target per class

Test Accuracy	MNIST-> MNISTM	MNIST-> USPS	SVHN-> MNIST	USPS-> SVHN	USPS-> MNIST
DANN Baseline	0.796	0.923	0.806	0.354	0.914
DIAL: Source Classifier + DANN + Class DANN	**0.948**	**0.951**	0.867	0.725	0.950
DIAL: Source Classifier + Triplet + DANN + Class DANN	0.903	0.946	**0.903**	**0.802**	**0.962**

Results: Comparison to another Semi-supervised method

Test Accuracy	MNIST-> MNISTM	MNIST-> USPS	SVHN-> MNIST	USPS-> SVHN	USPS-> MNIST
DIAL (best results**)	m = 1: 0.779 m = 5: 0.941 m = 10: 0.948	**m = 1: 0.906** **m = 5: 0.946** m = 10: 0.951	**m = 1: 0.773** **m = 5: 0.868** m = 10: 0.903	**m = 1: 0.400** **m = 5: 0.682** m = 10: 0.802	**m = 1: 0.949** **m = 5: 0.967** m = 10: 0.962
FADA [4]: Few-Shot Adversarial Domain Adaptation	-	m = 1: 0.891 m = 5: 0.934 m = 7: 0.944	m = 1: 0.728 m = 5: 0.861 m = 7: 0.872	m = 1: 0.275 m = 5: 0.379 m = 7: 0.429	m = 1: 0.811 m = 5: 0.911 m = 7: 0.915

[4] Mottian et. al, Few-Shot Adversarial Domain Adaptation. 2017