



Boulder

Computer Vision; Image Classification; Domain Adaptation

Maziar Raissi

Assistant Professor

Department of Applied Mathematics

University of Colorado Boulder

maziar.raissi@colorado.edu



Boulder

Domain-Adversarial Training of Neural Networks

Domain Adaptation

Data at training and test time come from similar but different distributions!

$X \rightarrow$ input space

$Y = \{0, 1, \dots, L - 1\} \rightarrow$ set of L possible labels

$\mathcal{D}_S \rightarrow$ source domain

$\mathcal{D}_T \rightarrow$ target domain

$\mathcal{D}_S, \mathcal{D}_T \rightarrow$ distributions over $X \times Y$

$S \rightarrow$ labeled source sample drawn i.i.d from \mathcal{D}_S

$T \rightarrow$ unlabeled target sample drawn i.i.d from \mathcal{D}_T^X

$\mathcal{D}_T^X \rightarrow$ marginal distribution of \mathcal{D}_T over X

$S = \{(x_i, y_i)\}_{i=1}^n \sim (\mathcal{D}_S)^n$

$T = \{x_i\}_{i=n+1}^N \sim (\mathcal{D}_T^X)^{n'}$

$N = n + n' \rightarrow$ total number of samples

Build a classifier $\eta : X \rightarrow Y$ with a low target risk:

$$\mathcal{R}_{\mathcal{D}_T}(\eta) = \Pr_{(x,y) \sim \mathcal{D}_T}(\eta(x) \neq y)$$

Theorem

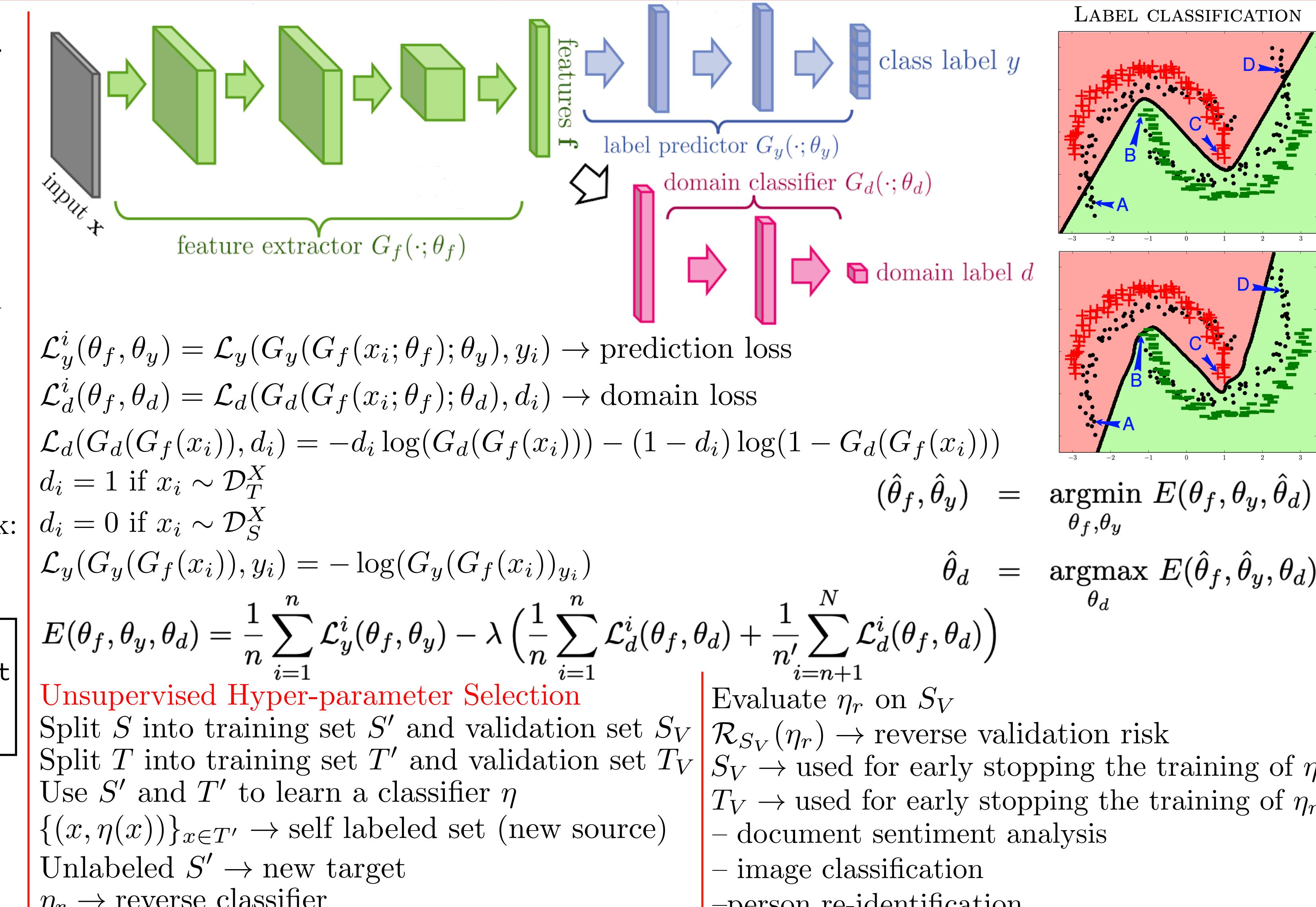
For effective domain transfer to be achieved, predictions must be made based on features that cannot discriminate between the training (source) and test (target) domains.

Domain-Adversarial Neural Network (DANN)

$G_f(\cdot; \theta_f) \rightarrow$ feature extractor

$G_y(\cdot; \theta_y) \rightarrow$ label predictor

$G_d(\cdot; \theta_d) \rightarrow$ domain classifier



Unsupervised Hyper-parameter Selection

Split S into training set S' and validation set S_V

Split T into training set T' and validation set T_V

Use S' and T' to learn a classifier η

$\{(x, \eta(x))\}_{x \in T'} \rightarrow$ self labeled set (new source)

Unlabeled $S' \rightarrow$ new target

$\eta_r \rightarrow$ reverse classifier

Evaluate η_r on S_V

$\mathcal{R}_{S_V}(\eta_r) \rightarrow$ reverse validation risk

$S_V \rightarrow$ used for early stopping the training of η

$T_V \rightarrow$ used for early stopping the training of η_r

- document sentiment analysis

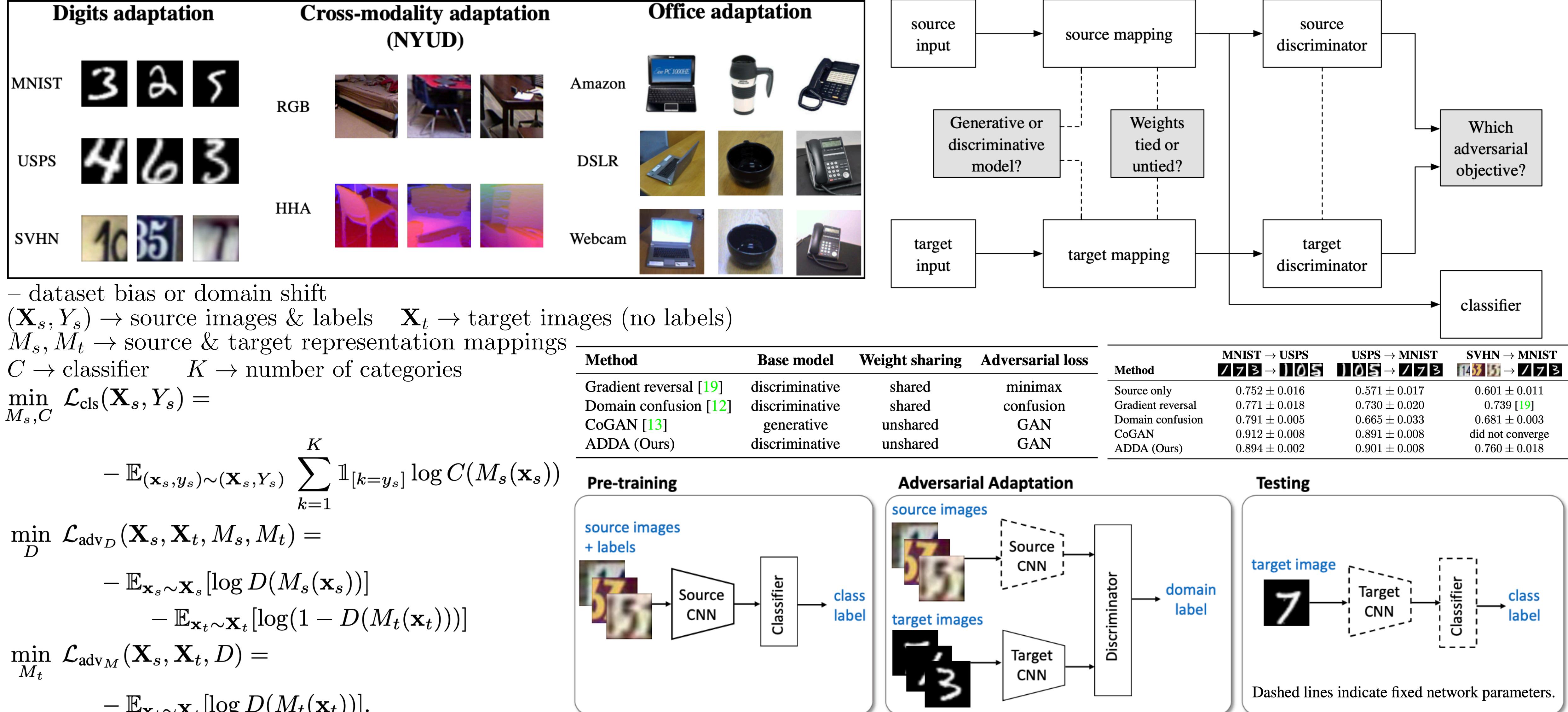
- image classification

- person re-identification



Boulder

Adversarial Discriminative Domain Adaptation





Boulder

Questions?
