ISTITUTO ITALIANO
DI TECNOLOGIA
PATTERN ANALYSIS
AND COMPUTER VISION

UNIVERSITÀ di VERONA
Dipartimento di INFORMATICA

AIDA Course

Domain Adaptation and Generalization

Vittorio Murino, Pietro Morerio

April 8, 2022

Credits

- Tutorial by Pietro Morerio and Massimiliano Mancini

- Some slides are courtesy of Prof. Elisa Ricci and Dr. Riccardo Volpi

- Other material is referred in the corresponding slides

Outline

Session 1 - Introduction (1h)

- What is domain adaptation and why do we need it?
- The domain shift issue in vision
- Domain shift - formal statement
- Common Domain Adaptation scenarios
- Classical methods and benchmarks

Session 2 - Recent Methods (Deep learning) (1h)

- Adversarial DA
- Image translation methods
- Feature alignment/confusion
- Batchnorm-based methods
- Pseudo-labeling (TODO)

Outline

Session 3 - Beyond Domain Adaptation (1h)

- Source Free UDA (TODO)
- Domain Discovery
- Continuous DA
- Predictive DA
- Validation issues in Unsupervised Domain adaptation

Session 4 - Domain generalization (1h)

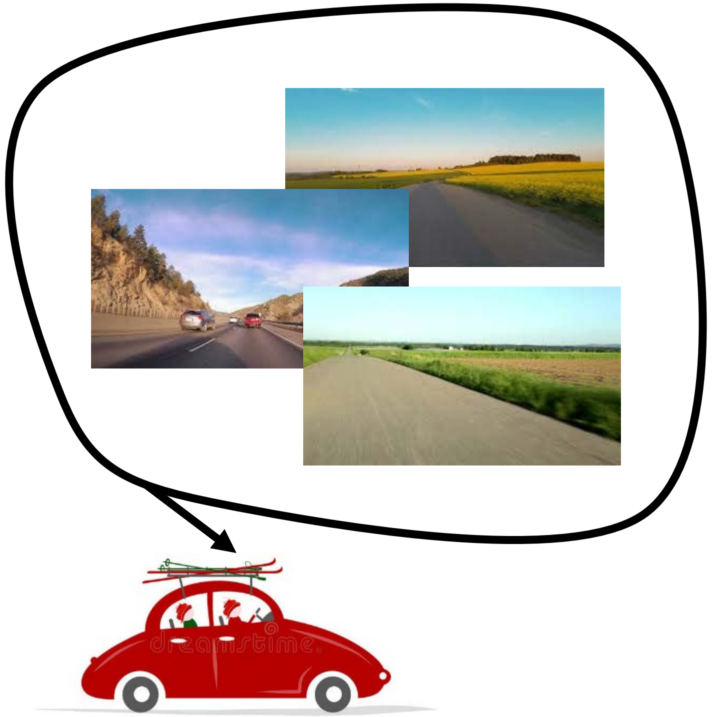
- A more challenging problem
- Single source domain generalization
- Other issues
- Conclusions

Session 4

Domain Generalization

Problem formulation

Each dataset carries its own bias [1], and models trained on it result biased, too.



[1] A. Torralba and A. Efros. Unbiased Look at Dataset Bias. CVPR11.

Domain adaptation

Domain adaptation has been the main strategy to bridge the gap between source and target distributions.

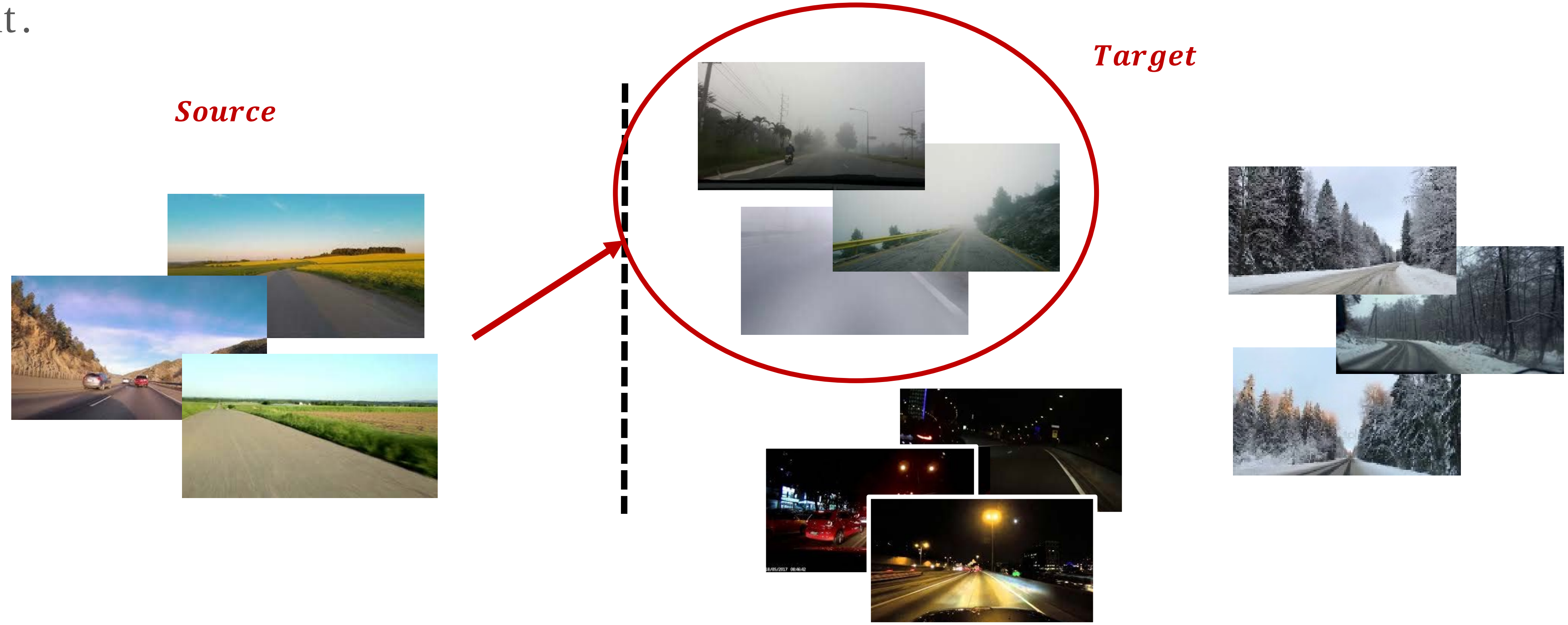
Assumptions: we can fix *a priori* a target distribution and we are able to sample from it.

Source

Domain adaptation

Domain adaptation has been the main strategy to bridge the gap between source and target distributions.

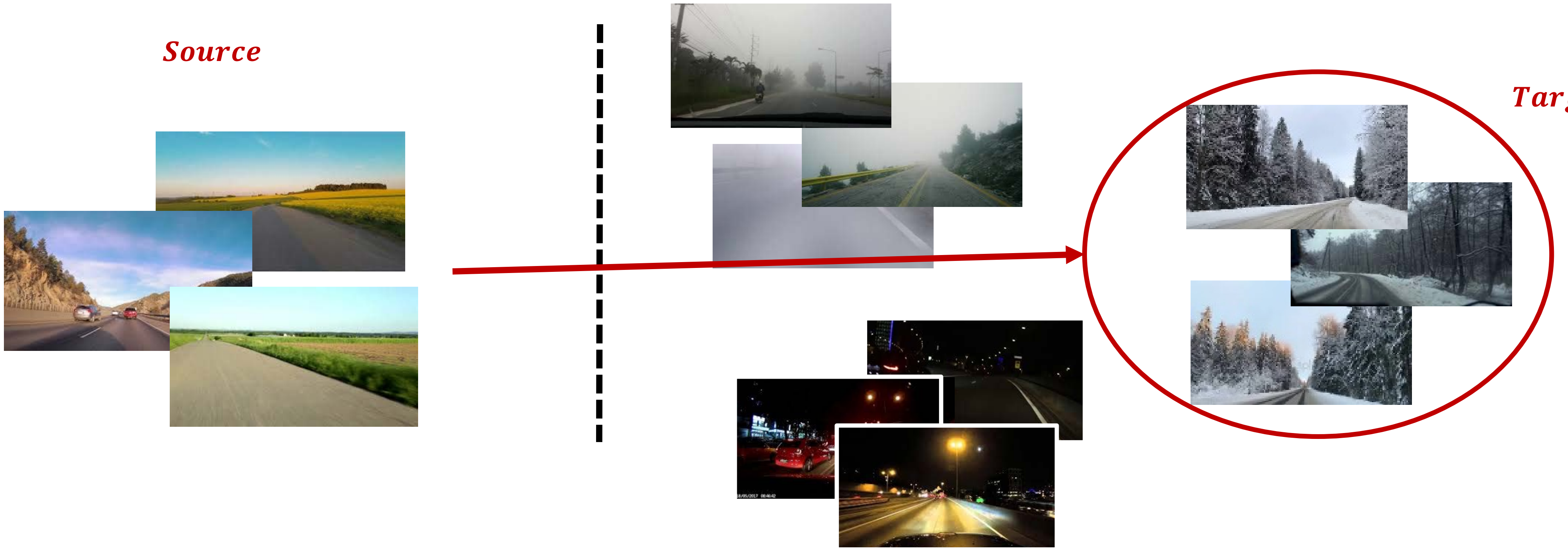
Assumptions: we can fix *a priori* a target distribution and we are able to sample from it.



Domain adaptation

Domain adaptation has been the main strategy to bridge the gap between source and target distributions.

Assumptions: we can fix *a priori* a target distribution and we are able to sample from it.



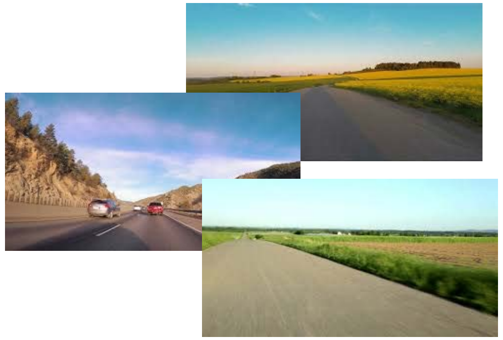
Domain adaptation

Domain adaptation has been the main strategy to bridge the gap between source and target distributions.

Assumptions: we can fix *a priori* a target distribution and we are able to sample from it.

Generalising to unseen domains

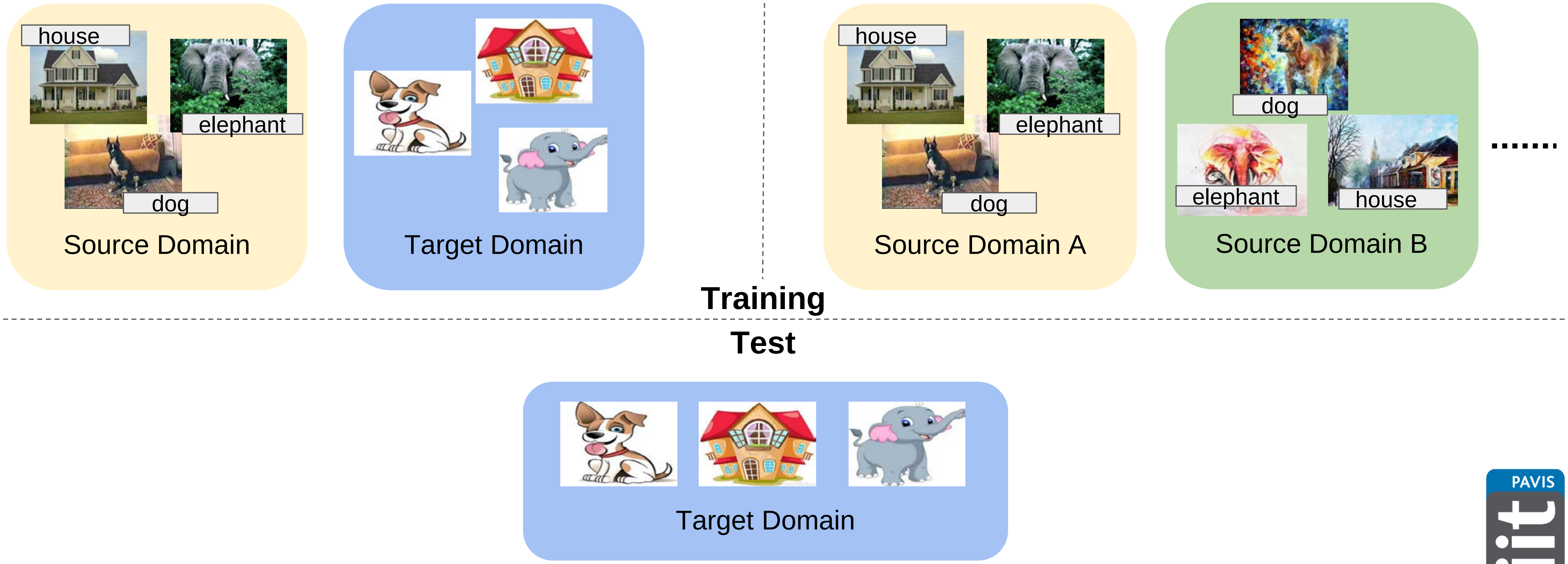
Goal: generalizing to unseen domains using data from a single source.



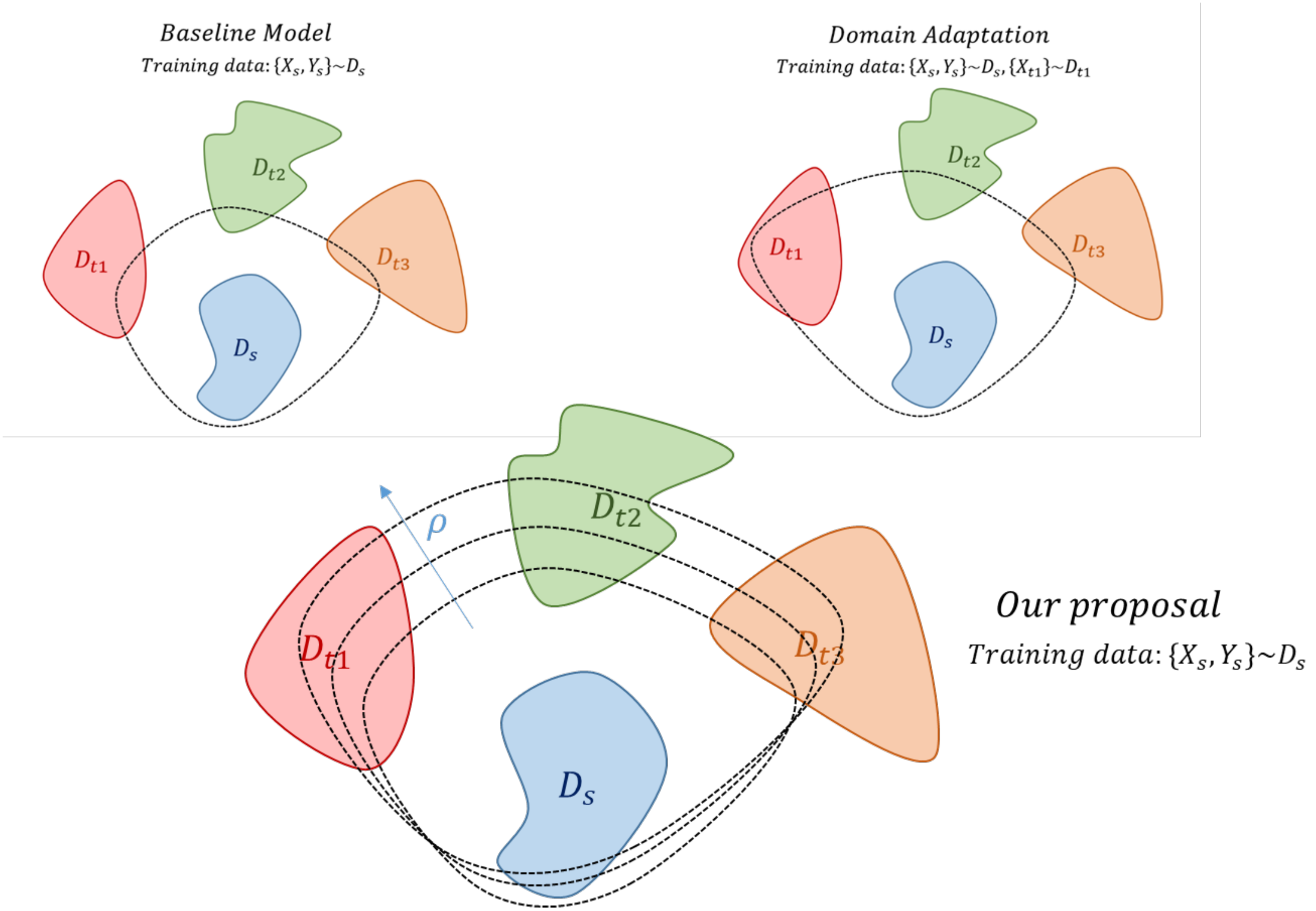
Domain Generalization (DG)

Domain Adaptation:
Given a one or multiple source domains for which we have labeled data, we want to find a model able to generalize to a target domain for which few or no labeled data are available during training.

Domain Generalization:
Given a set of multiple labeled source domains, we want to find a model able to generalize to any target domain for which no data are available during training:



Generalising to unseen domains



Volpi et al., *Generalizing to Unseen Domains via Adversarial Data Augmentation*, NeurIPS 2018

Locating the work …

Robust statistics

$$\min_{w} \quad E_{x,y\sim p'}\left\{l(y, f(x; w))\right\}$$

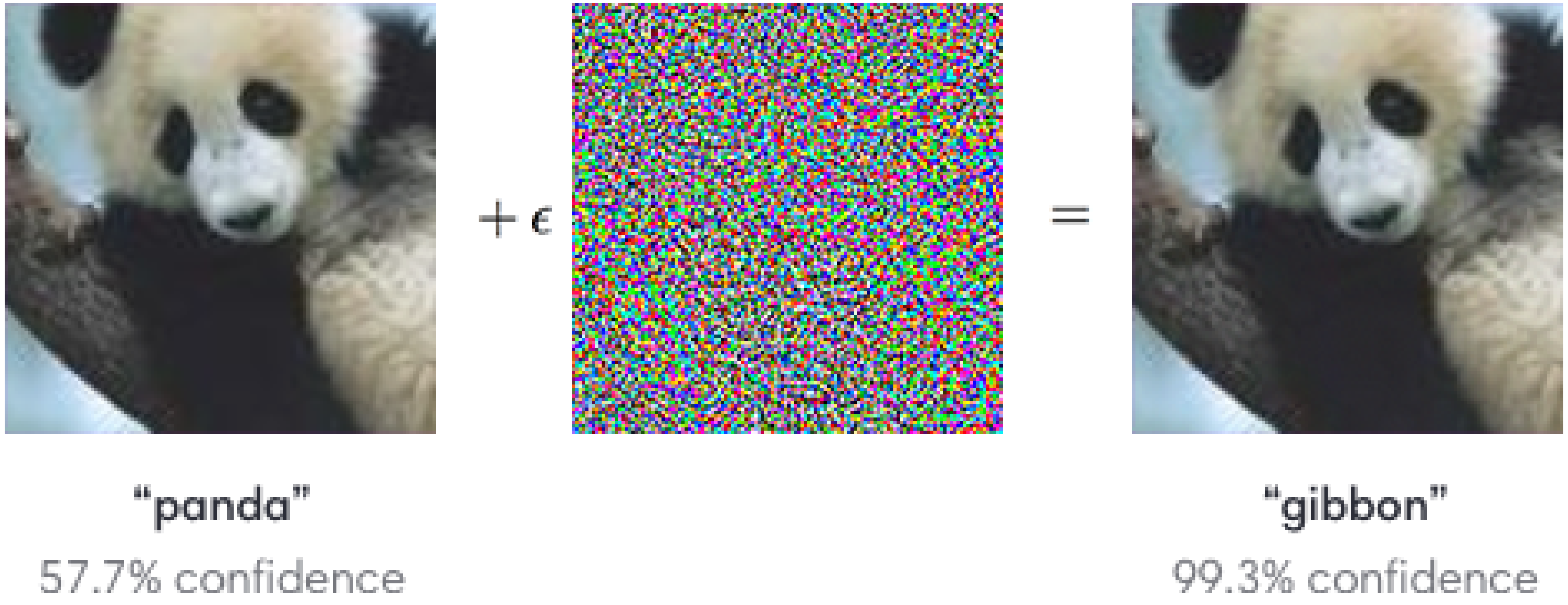
$$\Downarrow$$

$$\min_{w} \max_{p' \text{ st. } \Delta(p', p_{source}) \leq \delta} E_{x,y\sim p'}\left\{l(y, f(x; w))\right\}$$

Volpi et al., *Generalizing to Unseen Domains via Adversarial Data Augmentation*, NeurIPS 2018

Locating the work ...

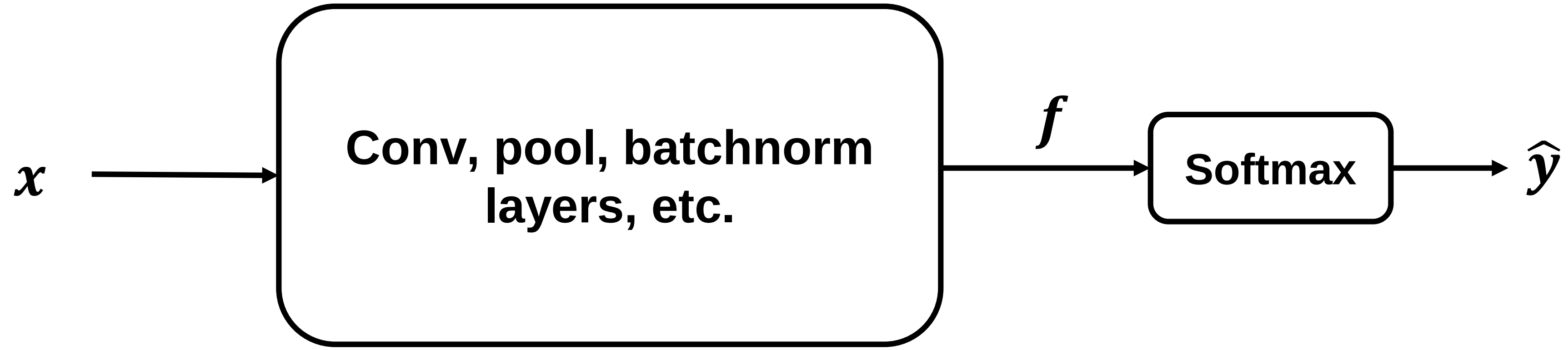
Defense against adversarial samples



Goodfellow et al., *Explaining and harnessing Adversarial Samples*, ICLR15

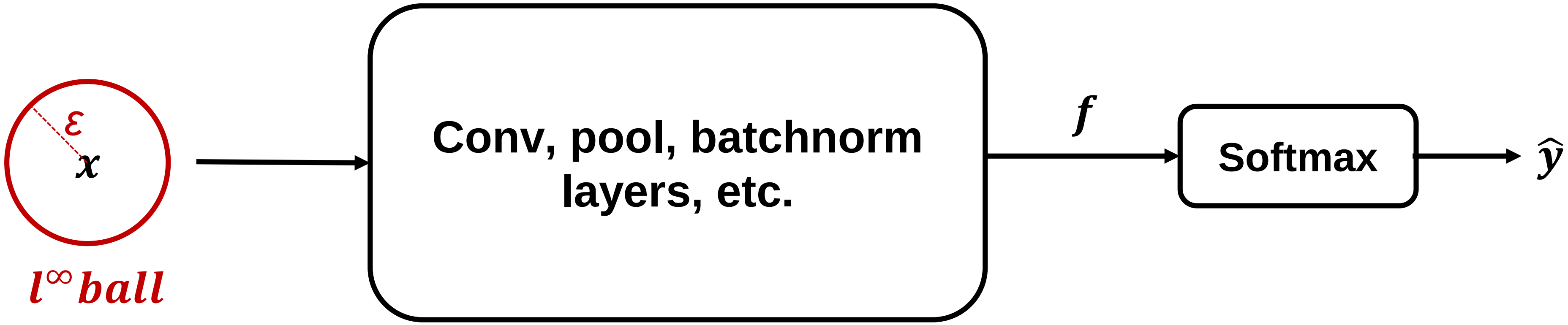
Locating the work …

Defense against adversarial samples



Locating the work …

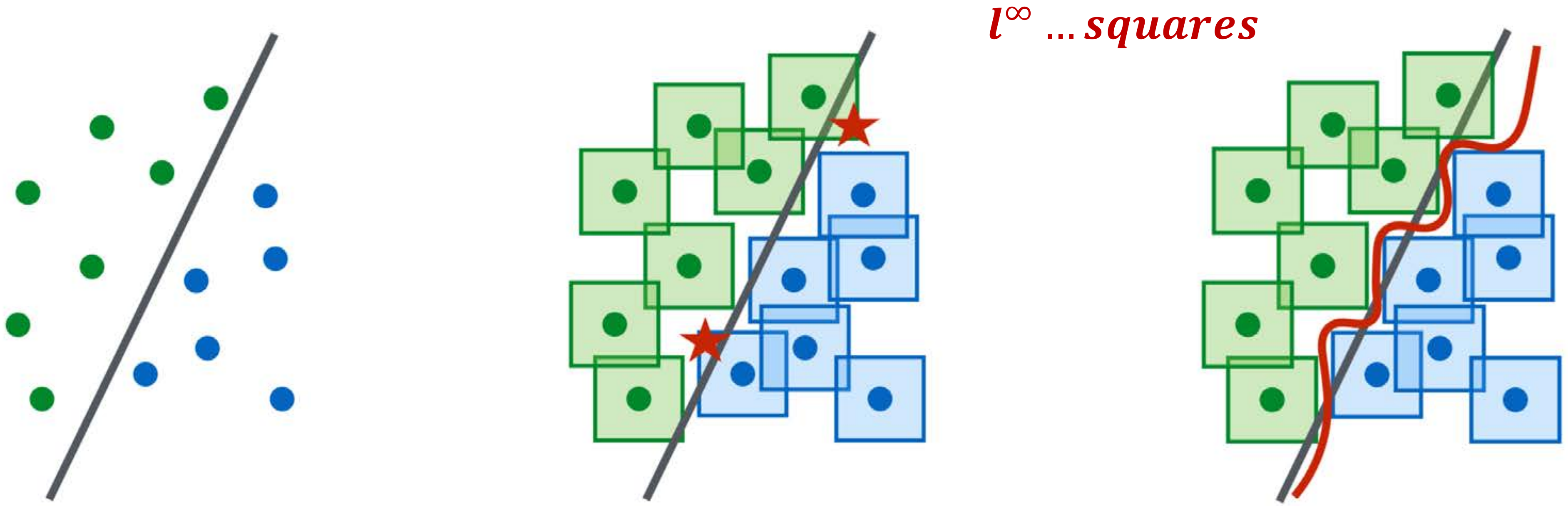
Defense against adversarial samples



Goodfellow et al., *Explaining and harnessing Adversarial Samples*, ICLR15
Madry et al., *Toward Deep Learning Models Resistant to Adversarial Attacks*, ICLR18

Locating the work …

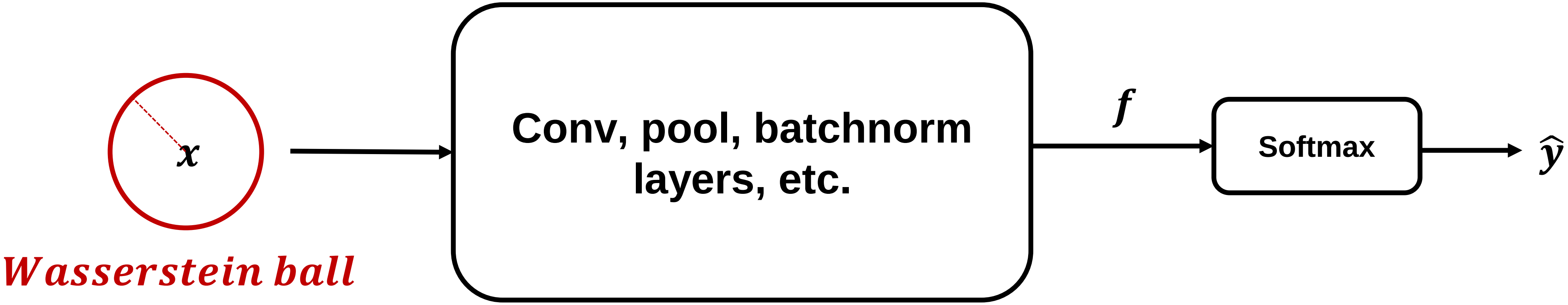
Defense against adversarial samples (pic from Madry et al.)



Goodfellow et al., *Explaining and harnessing Adversarial Samples*, ICLR15
Madry et al., *Toward Deep Learning Models Resistant to Adversarial Attacks*, ICLR18

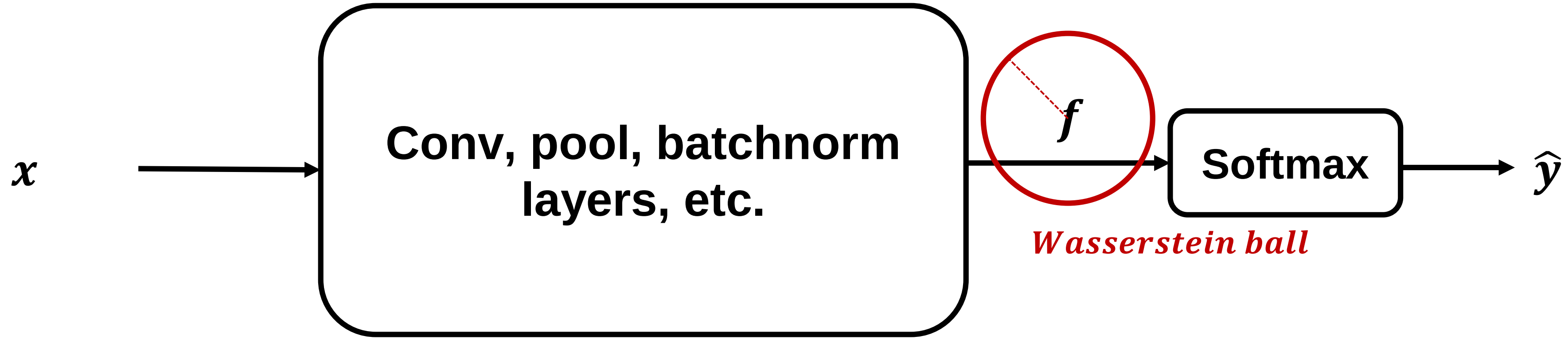
Locating the work ...

Defense against adversarial samples



Locating the work …

Defense against **perturbations in the feature space**, which – in high capacity networks – approximates a semantic space



Johnson et al., *Perceptual Losses for Real-Time Style Transfer and Super-Resolution*, ECCV16
Dosovitskiy A. and Brox T., *Generating Images with Perceptual Similarity Metrics Based on Deep Networks*, NIPS16

Method formulation (from robust statistics)

Distributionally robust optimization

$$\underset{\theta \in \Theta}{\text{minimize}} \; \sup_{P} \{\mathbb{E}_P[\ell(\theta; (X, Y))] : D_\theta(P, P_0) \leq \rho\}$$

We consider the Lagrangian relaxation [17]

$$\underset{\theta \in \Theta}{\text{minimize}} \; \sup_{P} \{\mathbb{E}_P[\ell(\theta; (X, Y))] - \gamma D_\theta(P, P_0)\}$$

Defining the surrogate loss ϕ_γ

We finally have:

$$\nabla_\theta \phi_\gamma(\theta; (x_0, y_0)) = \nabla_\theta \ell(\theta; (x_\gamma^\star, y_0))$$

Volpi et al., *Generalizing to Unseen Domains via Adversarial Data Augmentation*, NeurIPS 2018

PAVIS

Method formulation (from robust statistics)

Distributionally robust optimization

$$\underset{\theta \in \Theta}{\text{minimize}} \ \sup_P \{\mathbb{E}_P[\ell(\theta; (X, Y))] : D_\theta(P, P_0) \leq \rho\}$$

We consider the Lagrangian relaxation [17]

$$\underset{\theta \in \Theta}{\text{minimize}} \ \sup_P \{\mathbb{E}_P[\ell(\theta; (X, Y))] - \gamma D_\theta(P, P_0)\}$$

Defining the surrogate loss ϕ_γ

We finally have:

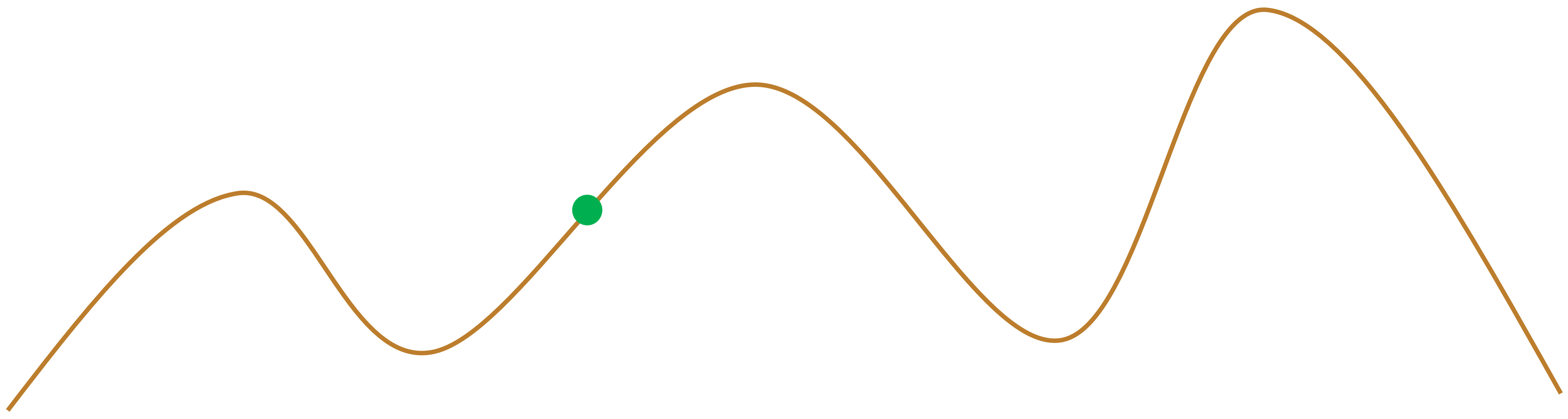
Computed by gradient ascent over the surrogate loss. c is a distance

$$\nabla_\theta \phi_\gamma(\theta; (x_0, y_0)) = \nabla_\theta \ell(\theta; (x_\gamma^\star, y_0))$$

$$x_\gamma^\star = \arg\max_{x \in \mathcal{X}} \{\ell(\theta; (x, y_0)) - \gamma c_\theta((x, y_0), (x_0, y_0))\}$$

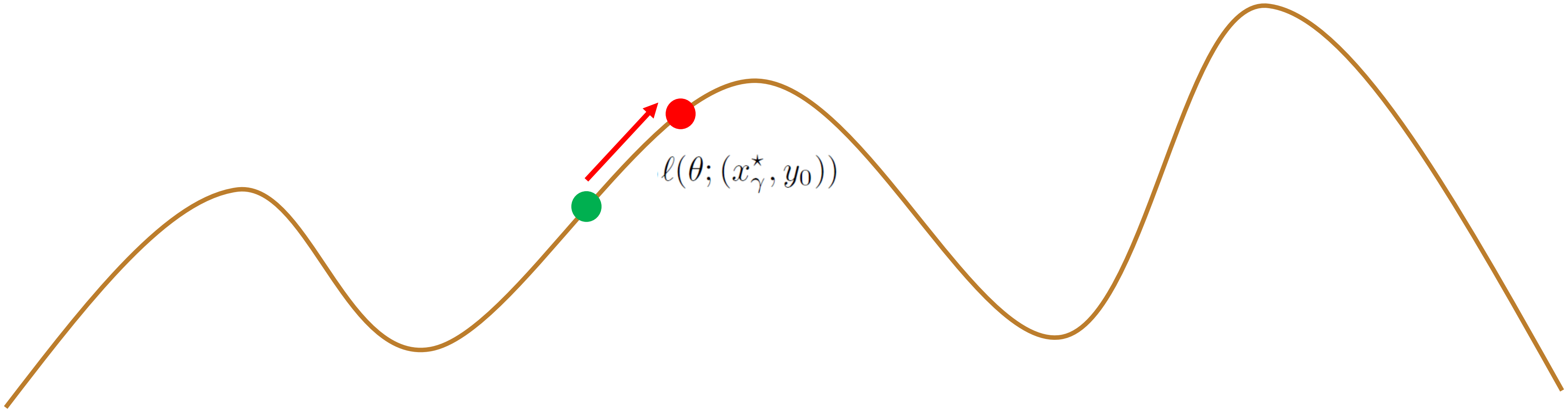
[17] Sinha et al., *Certifying Some Distributional Robustness with Principled Adversarial Training*, ICLR 2018

"Long-story short"



$$\nabla_\theta \phi_\gamma(\theta; (x_0, y_0)) = \nabla_\theta \ell(\theta; (x^\star_\gamma, y_0))$$

“Long-story short”

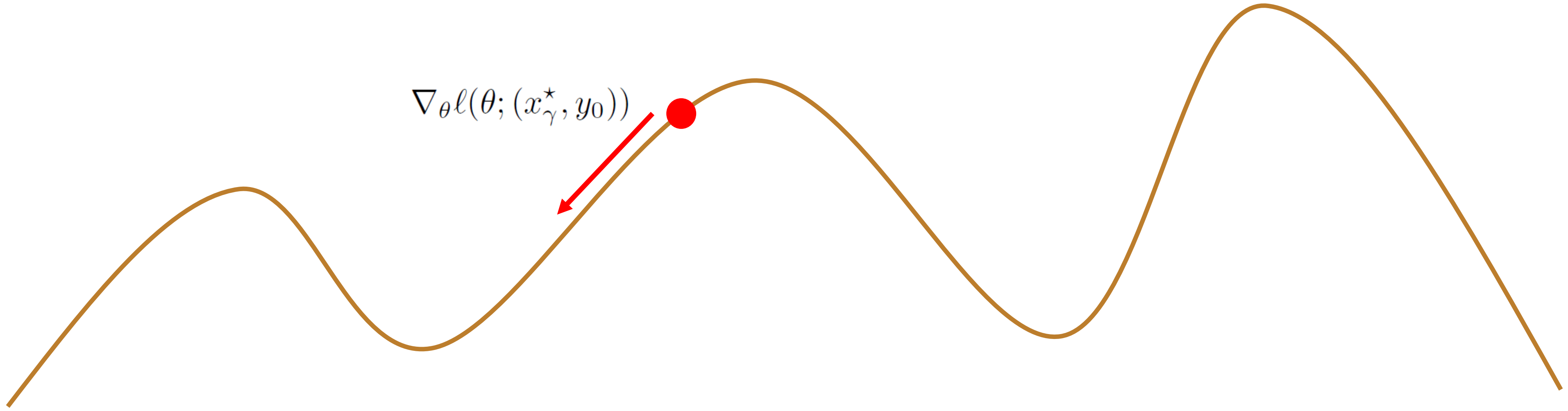


Gradient ascent

$$\nabla_\theta \phi_\gamma(\theta; (x_0, y_0)) = \nabla_\theta \ell(\theta; (x^\star_\gamma, y_0))$$

PAVIS

“Long-story short”



$$\boxed{\nabla_\theta \phi_\gamma(\theta; (x_0, y_0)) = \nabla_\theta \ell(\theta; (x^\star_\gamma, y_0))}$$

Gradient descent

PAVIS

Adversarial Data Augmentation

Algorithm:

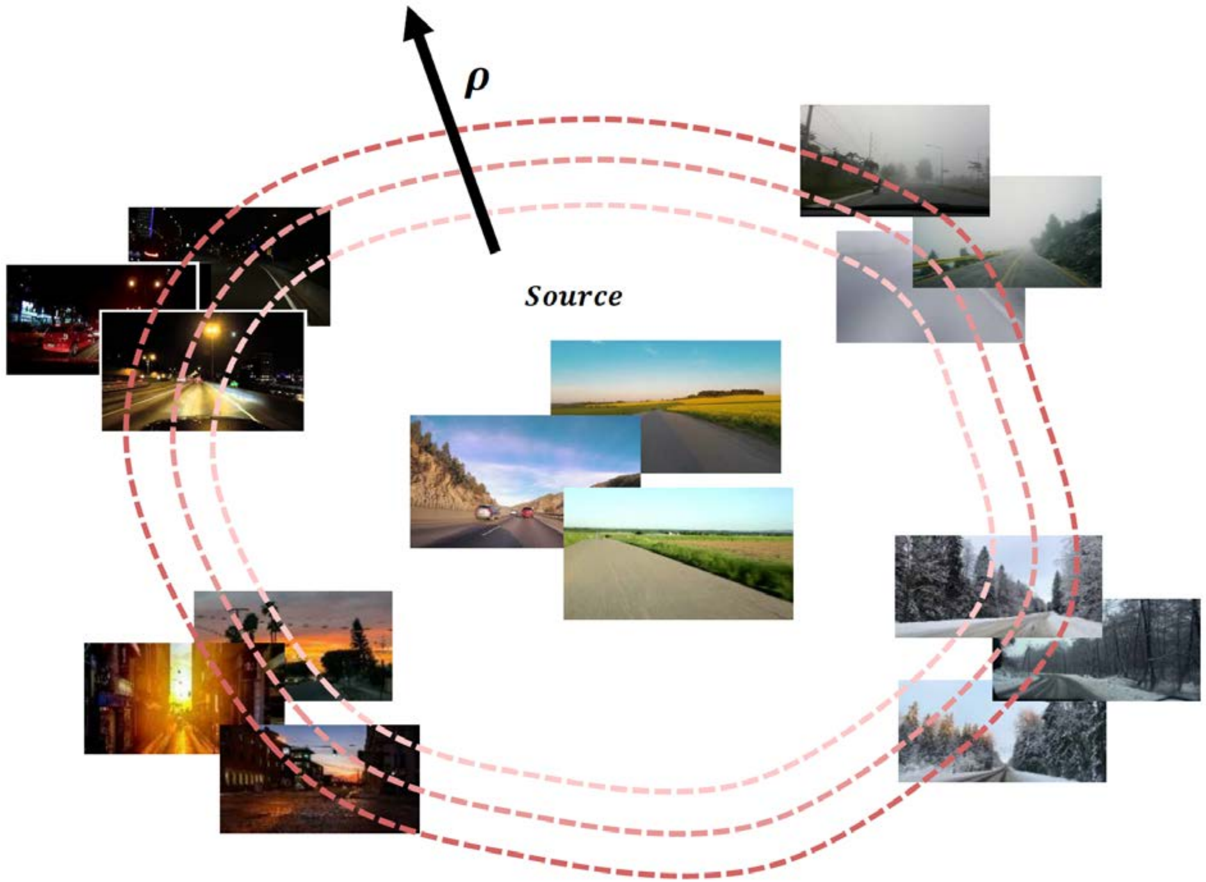
Input: model θ and dataset D

For K iterations:

1. Update θ via stochastic gradient descent
2. Generate perturbed samples and append them to D

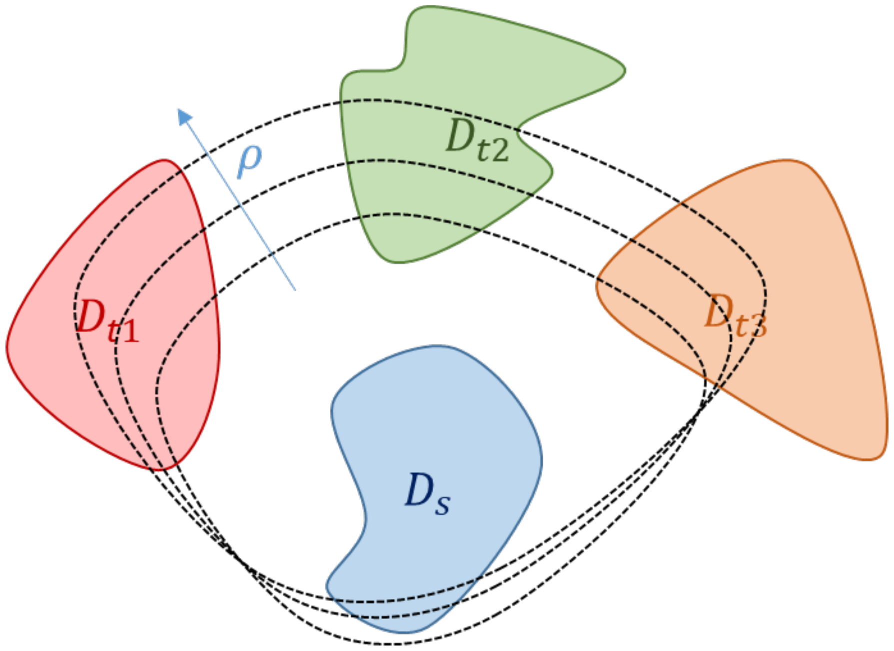
Update θ via stochastic gradient descent until convergence

Adversarial Data Augmentation



The "unknown-domain" problem

We don't know the target domain, thus it is difficult to set ρ



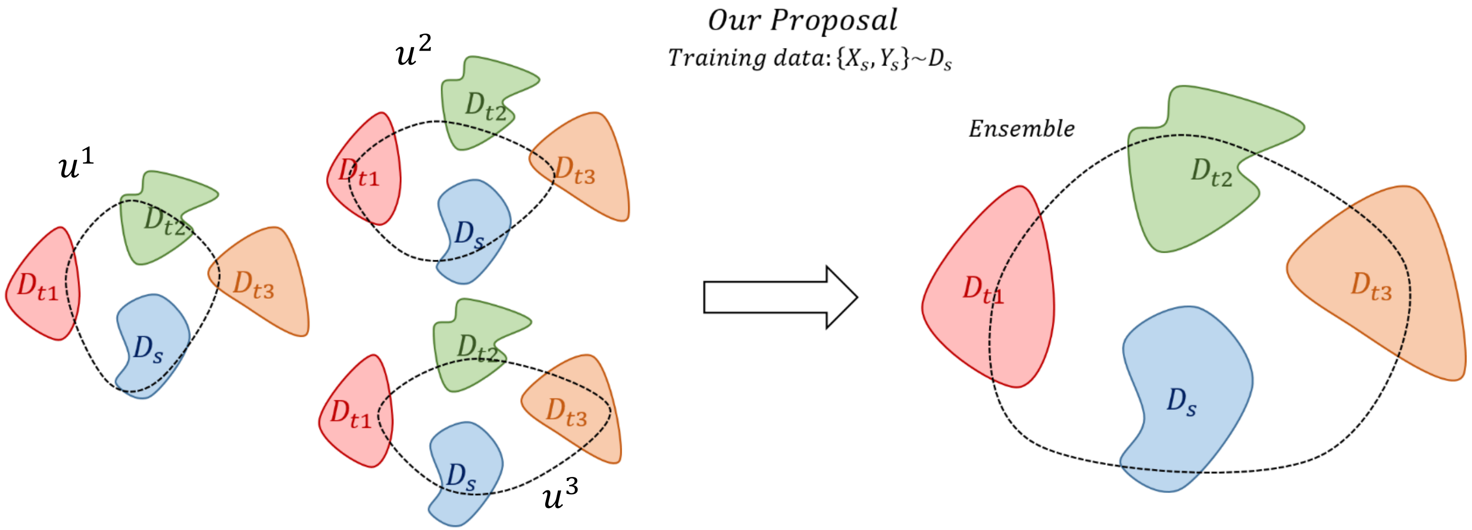
Our proposal

Training data: $\{X_s, Y_s\} \sim D_s$

The “unknown-domain” problem

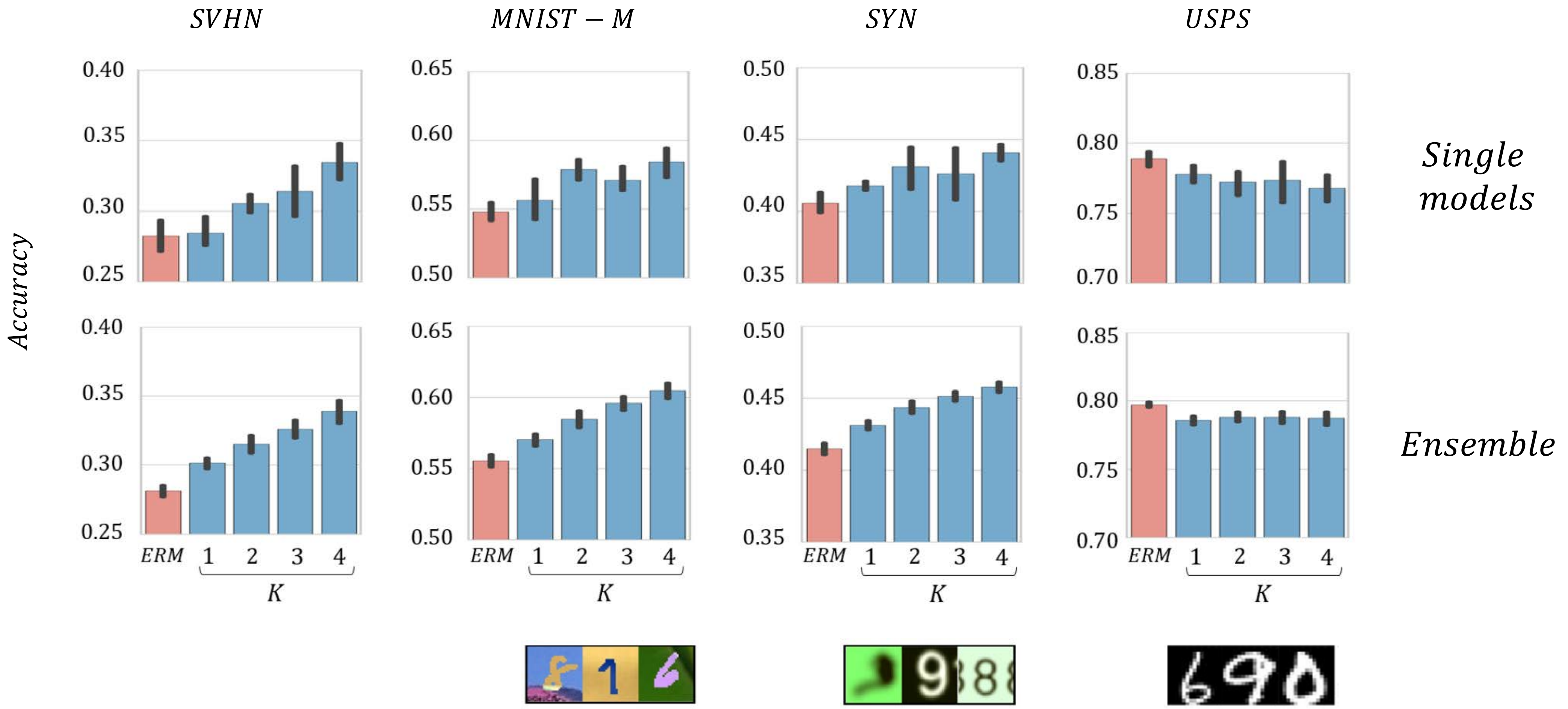
We don’t know the target domain, thus it is difficult to set ρ

→ **ENSEMBLE APPROACH**



$$u^{\star}(x) := \arg\max_{1 \leq u \leq s} \max_{1 \leq j \leq k} \underbrace{\theta_{c,j}^{u\top} g(\theta_f^u; x)}_{\text{softmax}}$$

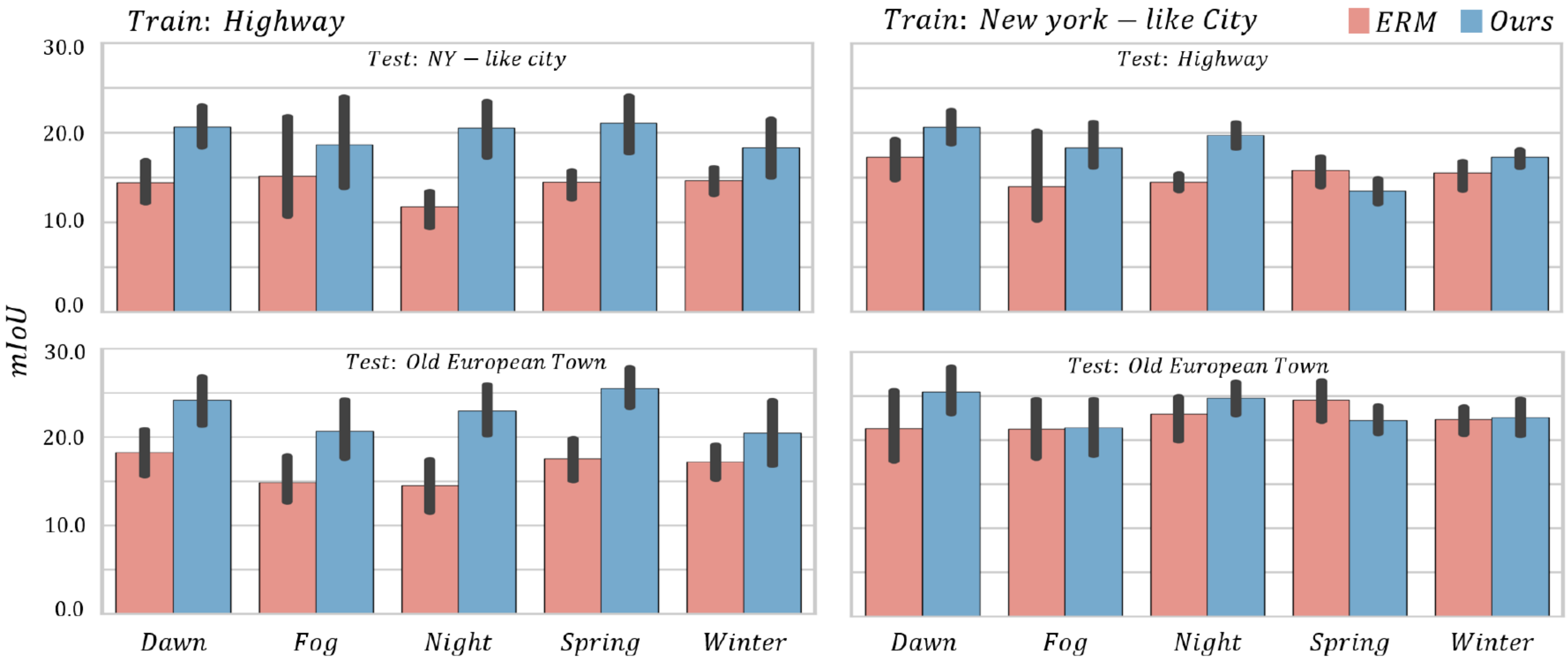
Results – Digits



Results – Digits

Results

SYNTHIA dataset



SPRING

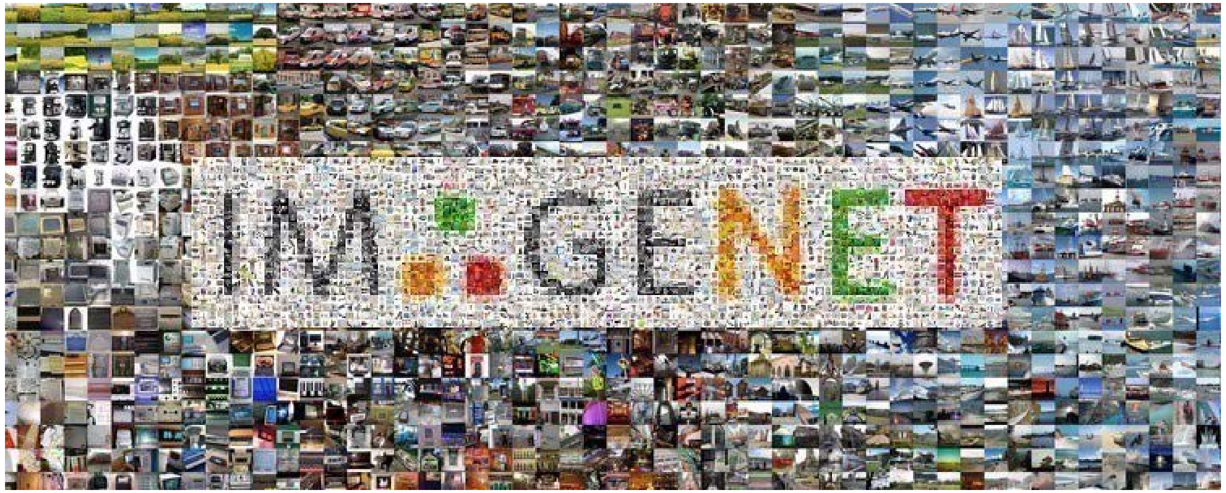
NIGHT

WINTER

Wrapping up ...

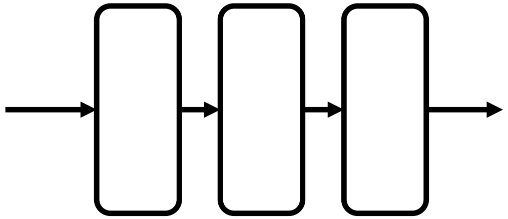
To recap … a standard situation

- Your data(set), your model



+

(your favorite neural net)



PyTorch

TensorFlow

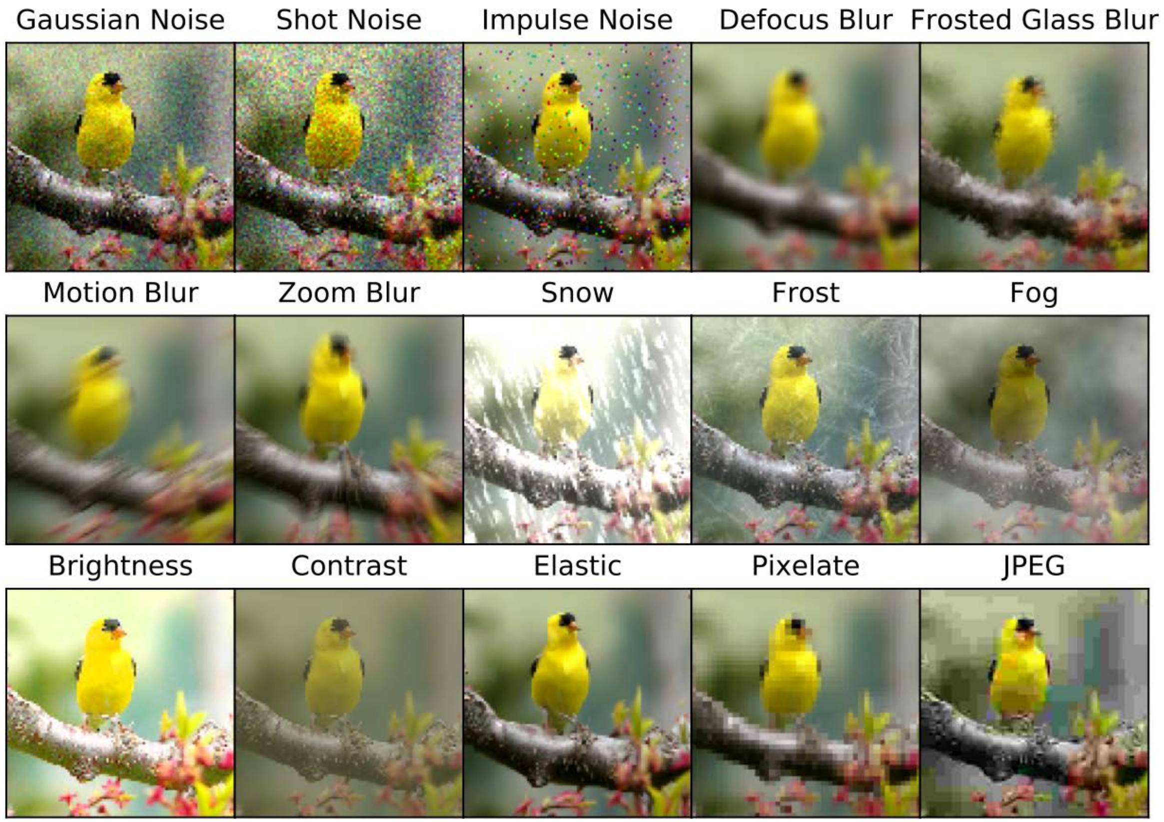
Deng et al., CVPR 2009

But then… corruption (lack of) robustness

BENCHMARKING NEURAL NETWORK ROBUSTNESS TO COMMON CORRUPTIONS AND PERTURBATIONS

Dan Hendrycks
University of California, Berkeley

Thomas Dietterich
Oregon State University



mCE	Clean Error
53.6%	24.2%
56.5%	17.90%
63%	23.9%
64.9%	21.2%
65.3%	22.47%
69.3%	25.41%
74.3%	24.5%
76.7%	23.85%

Hendrycks & Detterich, Benchmarking Neural Networks Robustness to Common Corruptions and Perturbations, ICLR 2019

But then … texture bias in DNNs

IMAGENET-TRAINED CNNS ARE BIASED TOWARDS TEXTURE; INCREASING SHAPE BIAS IMPROVES ACCURACY AND ROBUSTNESS

Robert Geirhos
University of Tübingen & IMPRS-IS

Patricia Rubisch
University of Tübingen & U. of Edinburgh

Claudio Michaelis
University of Tübingen & IMPRS-IS

Matthias Bethge*
University of Tübingen

Felix A. Wichmann*
University of Tübingen

Wieland Brendel*
University of Tübingen

(a) Texture image

81.4%	**Indian elephant**
10.3%	indri
8.2%	black swan

(b) Content image

71.1%	**tabby cat**
17.3%	grey fox
3.3%	Siamese cat

(c) Texture-shape cue conflict

63.9%	**Indian elephant**
26.4%	indri
9.6%	black swan

Geirhos et al., ImageNet-trained CNNs are biased towards texture; increasing shape bias improves accuracy and robustness, ICLR 2019

But then … dataset bias/domain shift

- Each dataset carries its own *bias*, and models trained on it result *biased*, too.

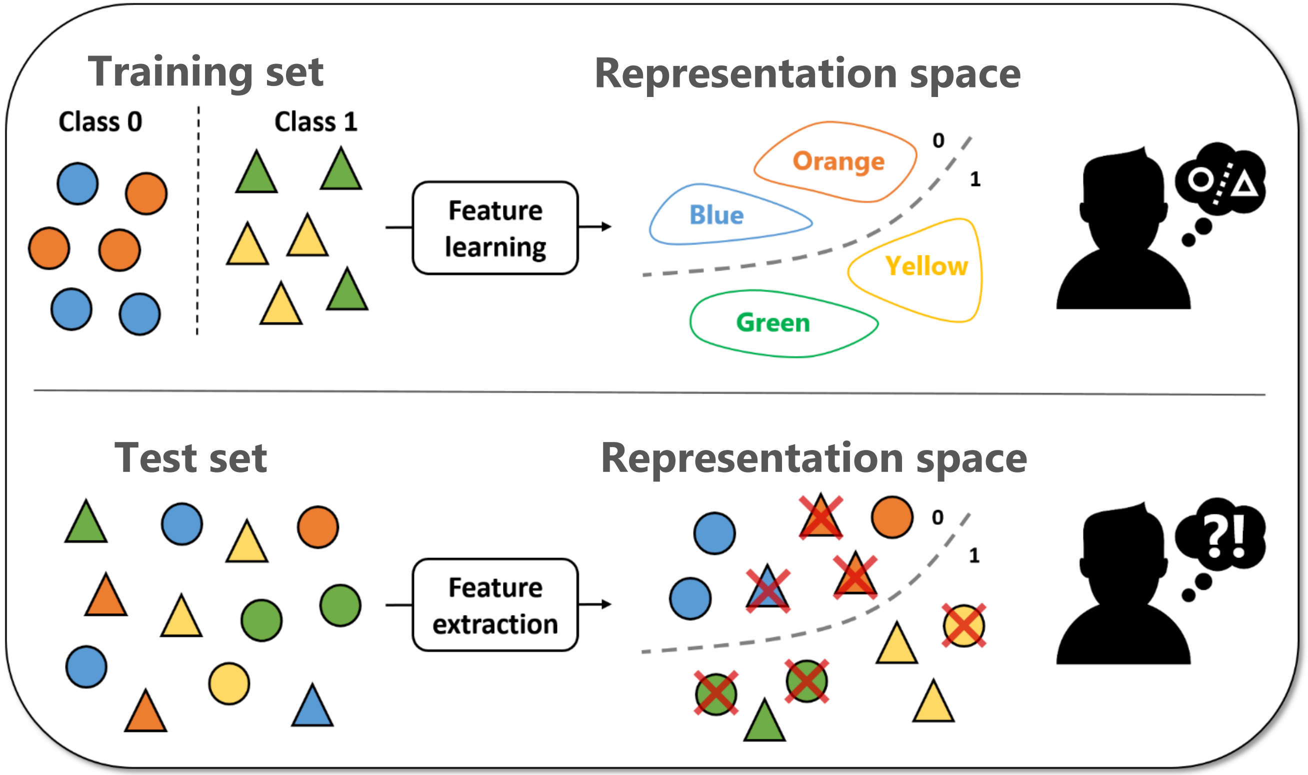
Training set

Test set

Torralba A. and Efros A. A., Unbiased Look at Dataset Bias, CVPR 2011

But then … dataset bias/domain shift

- Each dataset carries its own *bias*, and models trained on it result *biased*, too.



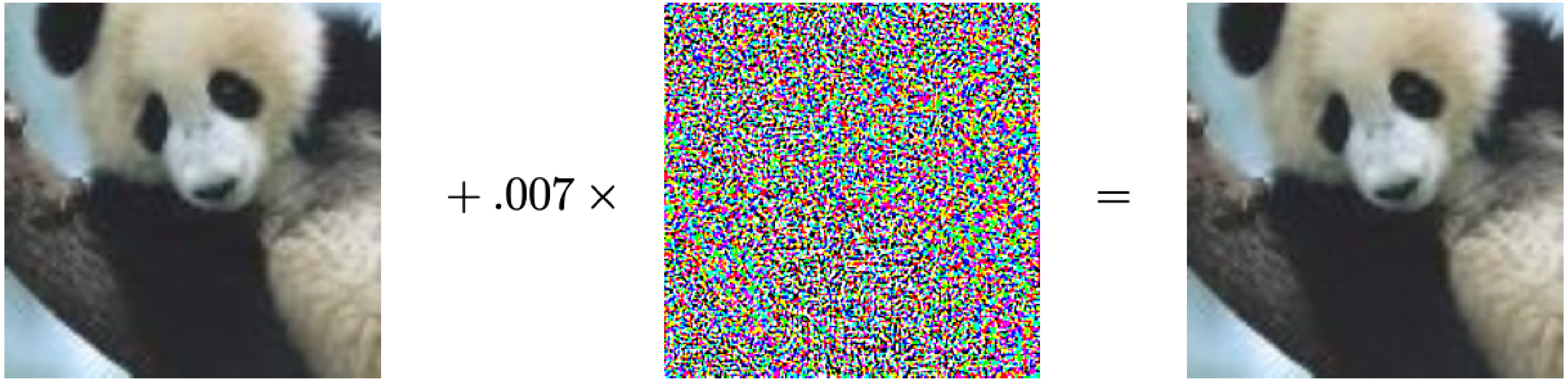
Ragonesi et al., Learning Unbiased Representations via Mutual Information Backpropagation, CVPR Workshops 2021

PAVIS

But then … adversarial samples

Explaining and Harnessing Adversarial Examples

Ian J. Goodfellow, Jonathon Shlens & Christian Szegedy
Google Inc., Mountain View, CA



$+ .007 \times$ ... $=$

$\boldsymbol{x}$	$\text{sign}(\nabla_{\boldsymbol{x}} J(\boldsymbol{\theta}, \boldsymbol{x}, y))$	$\boldsymbol{x} + \epsilon\text{sign}(\nabla_{\boldsymbol{x}} J(\boldsymbol{\theta}, \boldsymbol{x}, y))$
"panda"	"nematode"	"gibbon"
57.7% confidence	8.2% confidence	99.3 % confidence

Modern machine learning models

Something to keep in mind (among many other things)

- Data greedy
- Vulnerabilities against domain shifts
- Dataset bias
- Human bias
- Vulnerabilities against adversarial samples

Problem formulation(s)

Empirical Risk Minimization

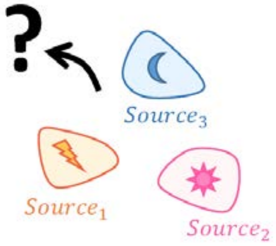
Training data

$\{x, y\} \sim P_{source}$

Multi-source Domain Generalization

Training data

$\{(x, y, d)\} \sim P_{source}$



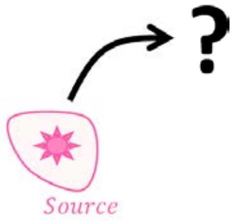
(sim2real)

(corruption robustness)

Single-source Domain Generalization

Training data

$\{(x, y)\} \sim P_{source}$

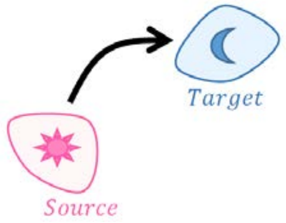


(sim2real)

Unsupervised Domain Adaptation

Training data

$\{(x, y)\} \sim P_{source}, \{x\} \sim P_{target}$



Fair/Unbias representations

Training data

$\{(x, y, s)\} \sim P_{source}$

(s is a sensitive attribute)

(s is a biased attribute)

PAVIS

Thanks for the attention