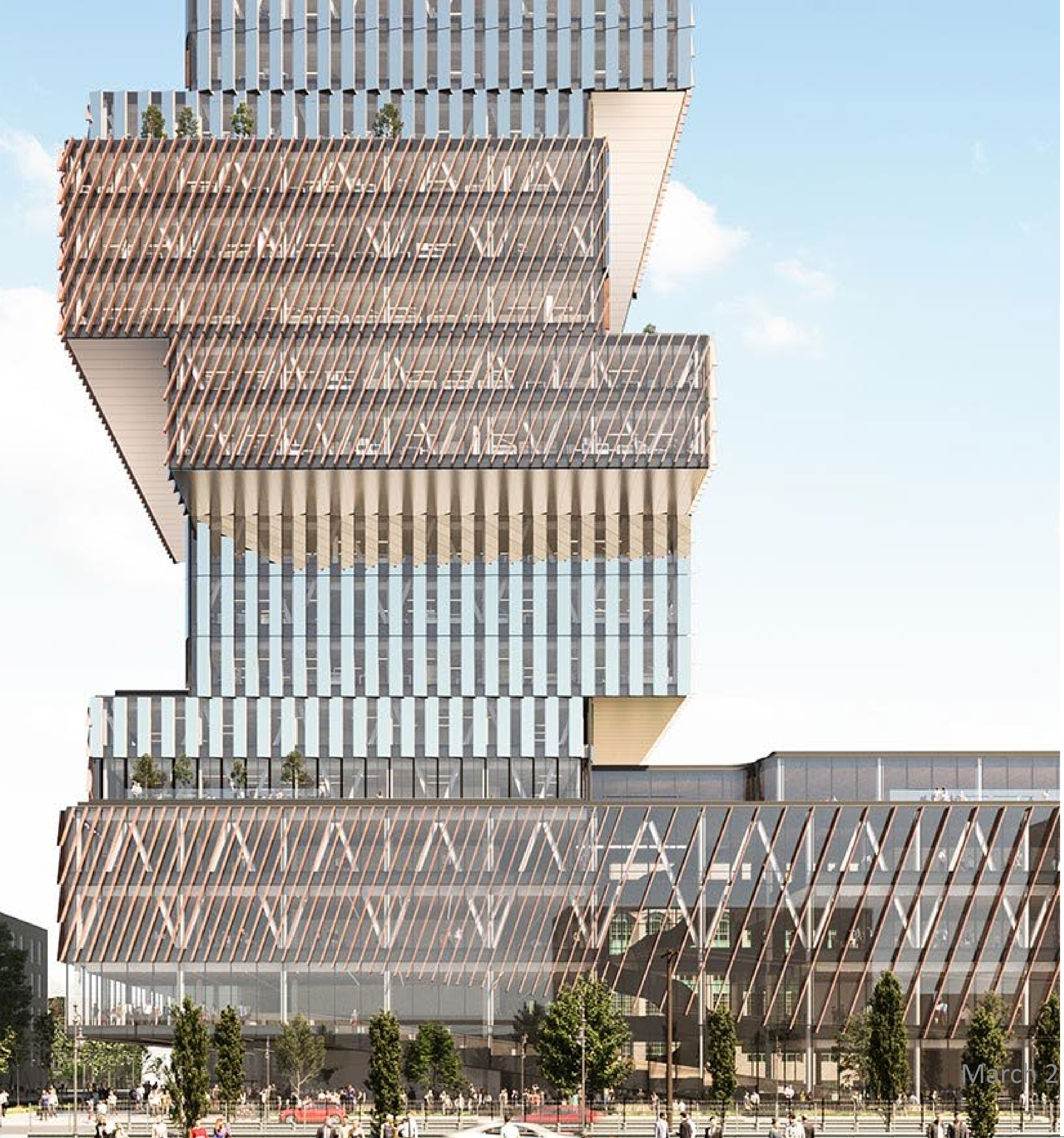


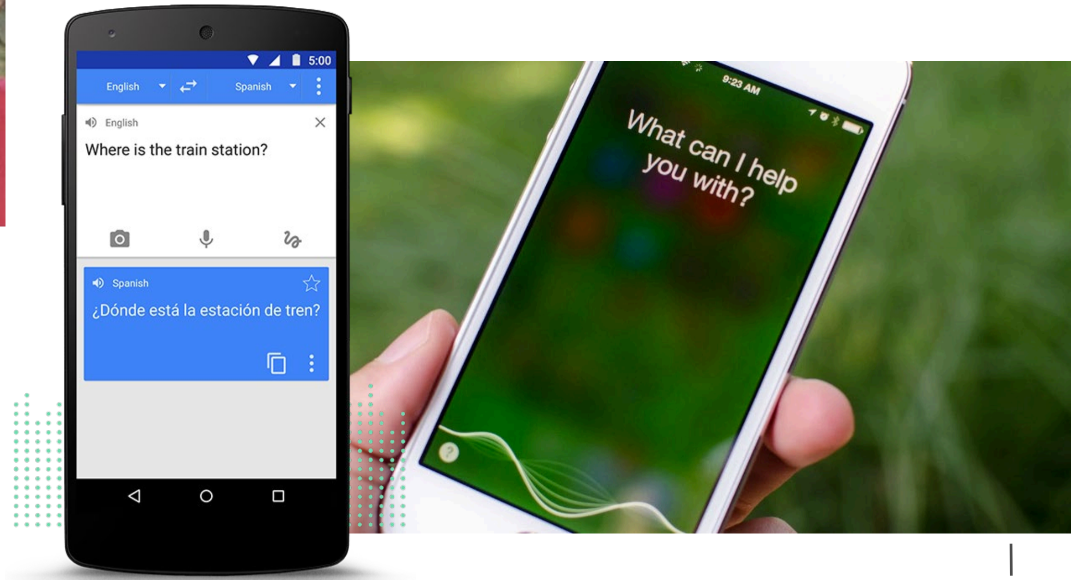
Learning Generalizable Visual Representations

Kate Saenko

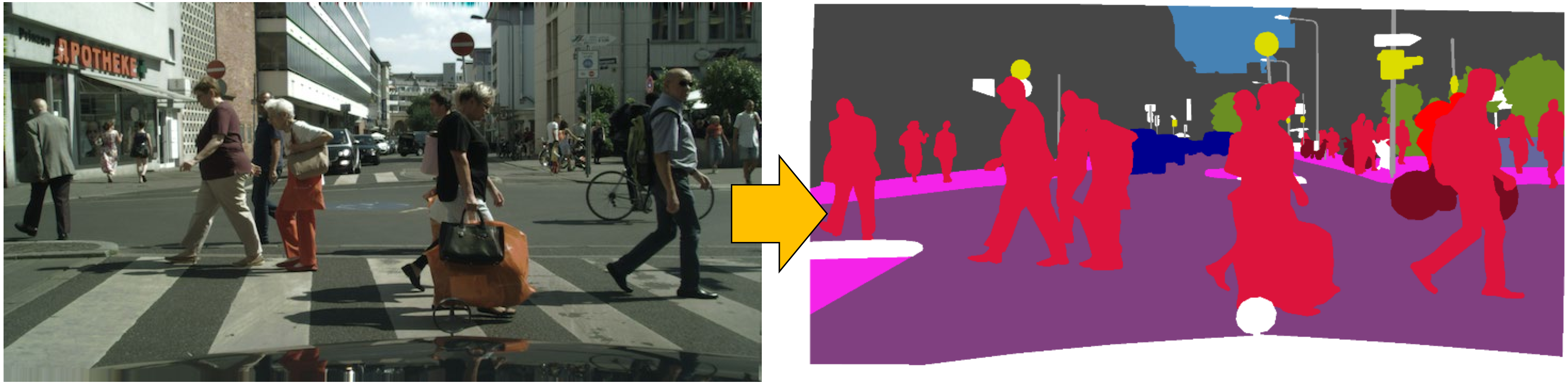
Boston University & MIT-IBM Watson AI

Successes of AI

2015 2020



Automate tasks previously only done by humans



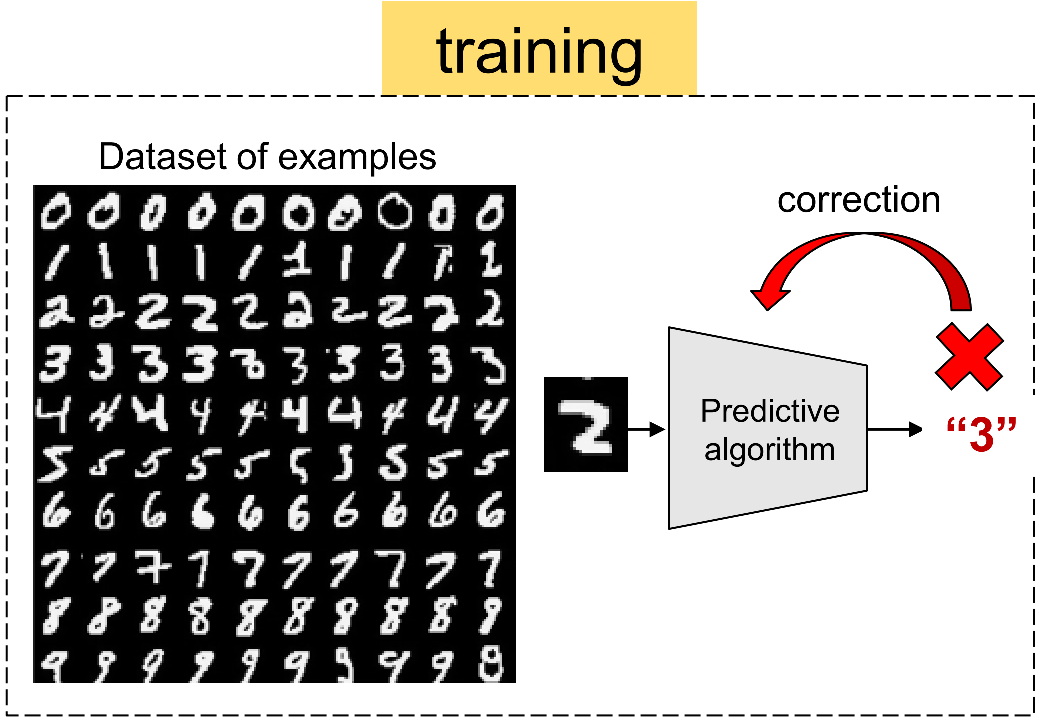
Driving scene segmentation Project at BU

ZeroWaste Project at BU

What is Machine Learning?

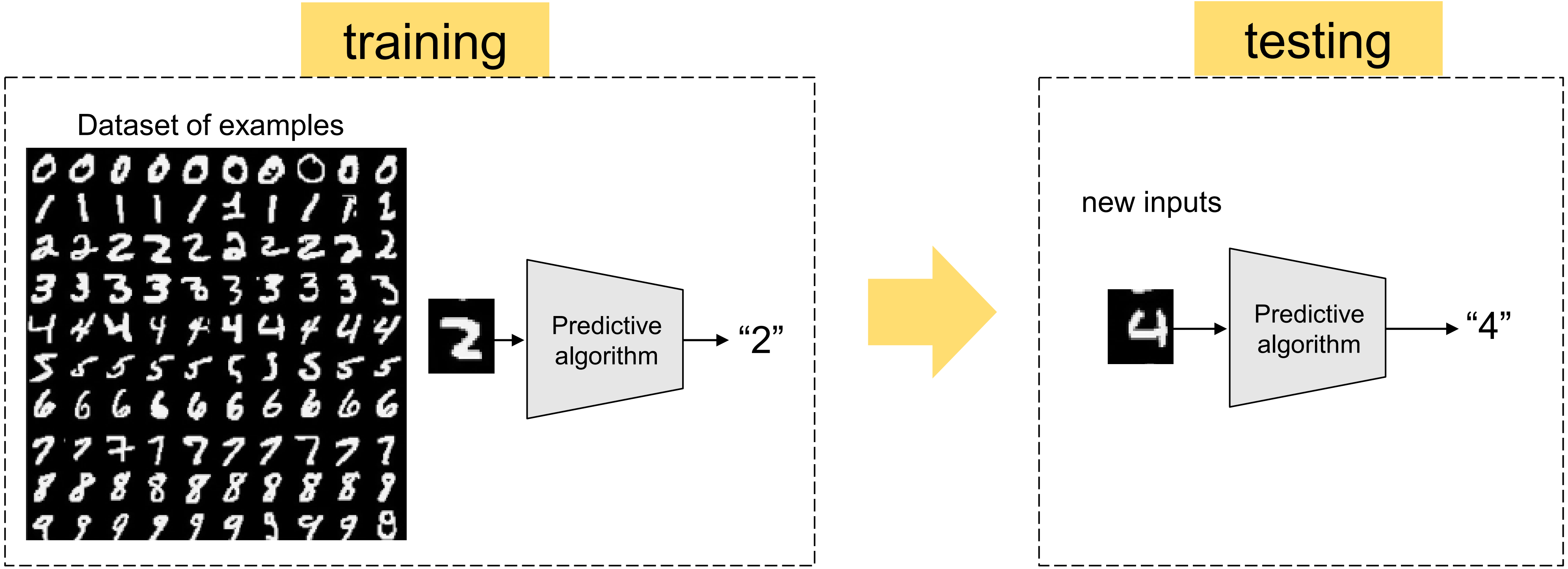
What is Machine Learning?

- Machine learning powers Artificial Intelligence
- Software algorithm that “learns” to make decisions from examples



What is Machine Learning?

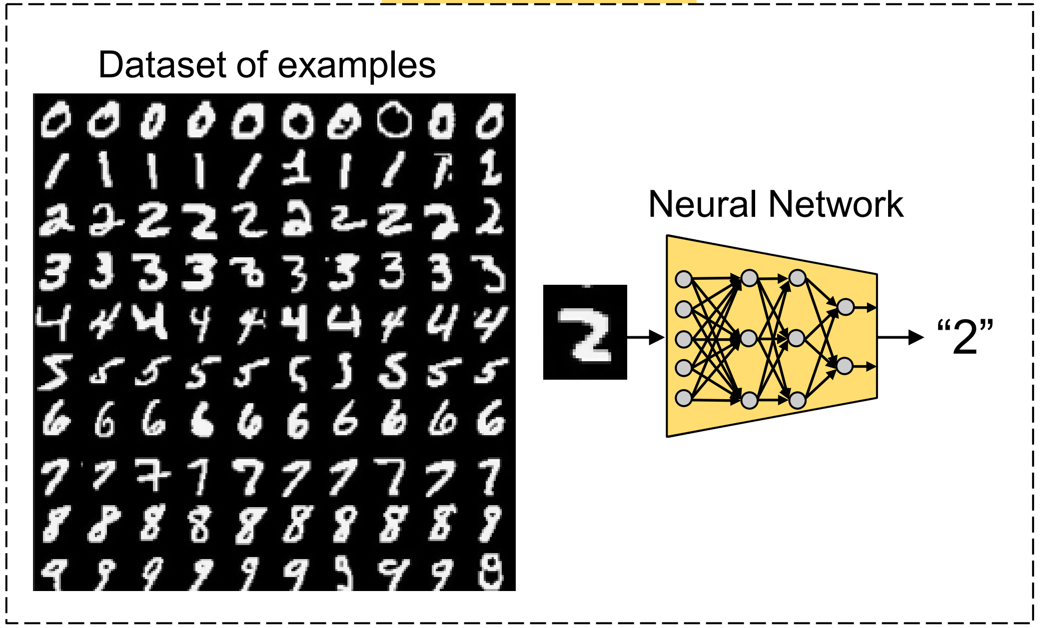
- Machine learning powers Artificial Intelligence
- Software algorithm that “learns” to make decisions from examples



“Deep” Learning == Neural Networks

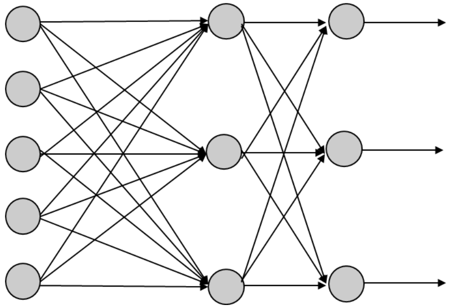
- a popular algorithm is a deep Neural Network
- It learns features (patterns) from data – does not rely on hand-coded ones

training



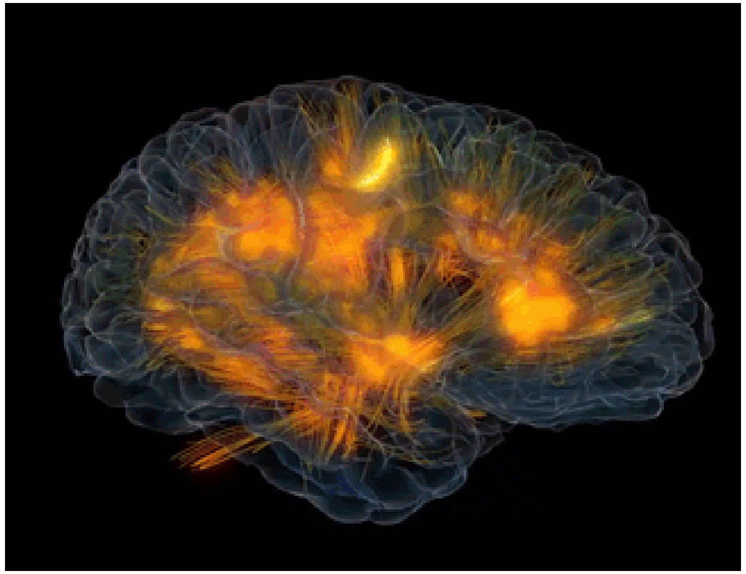
Neurons in the brain…

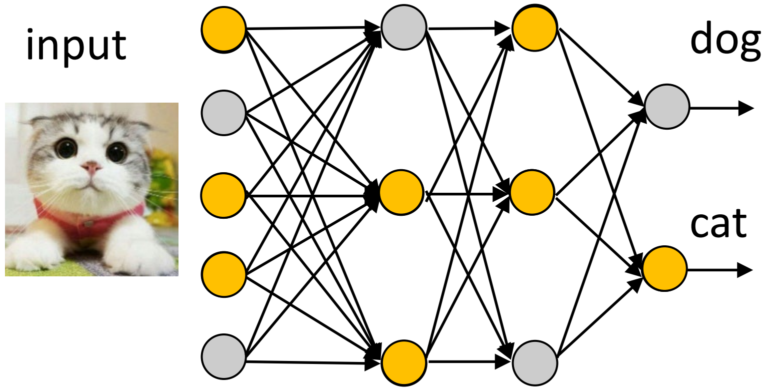
…inspired "Artificial Neural Networks"



Neurons are cells that process chemical and electrical signals and transmit these signals to neurons and other types of cells

Real vs Artificial Neural Network





- **Artificial neural networks:** consist of many inter-connected neurons organized in layers
- **Neurons**: each neuron receives inputs from neurons in previous layer, passes its output to next layer
- **Activation**: neuron's output between 1 (excited) and 0 (not excited)

Neural Networks learn features (patterns)

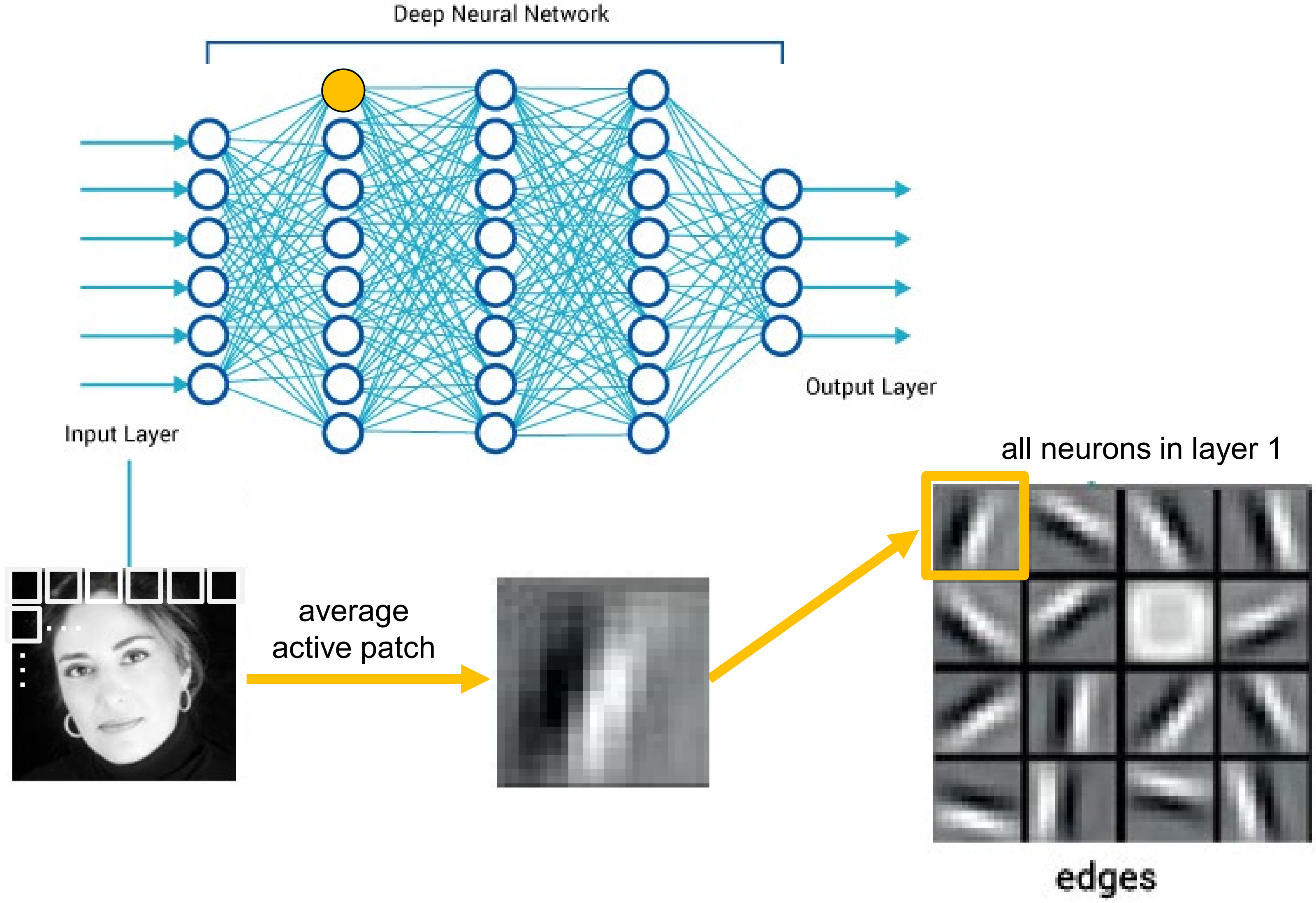


Image: http://www.amax.com/blog/wp-content/uploads/2015/12/blog_deeplearning3.jpg

Neural Networks learn **hierarchical** features

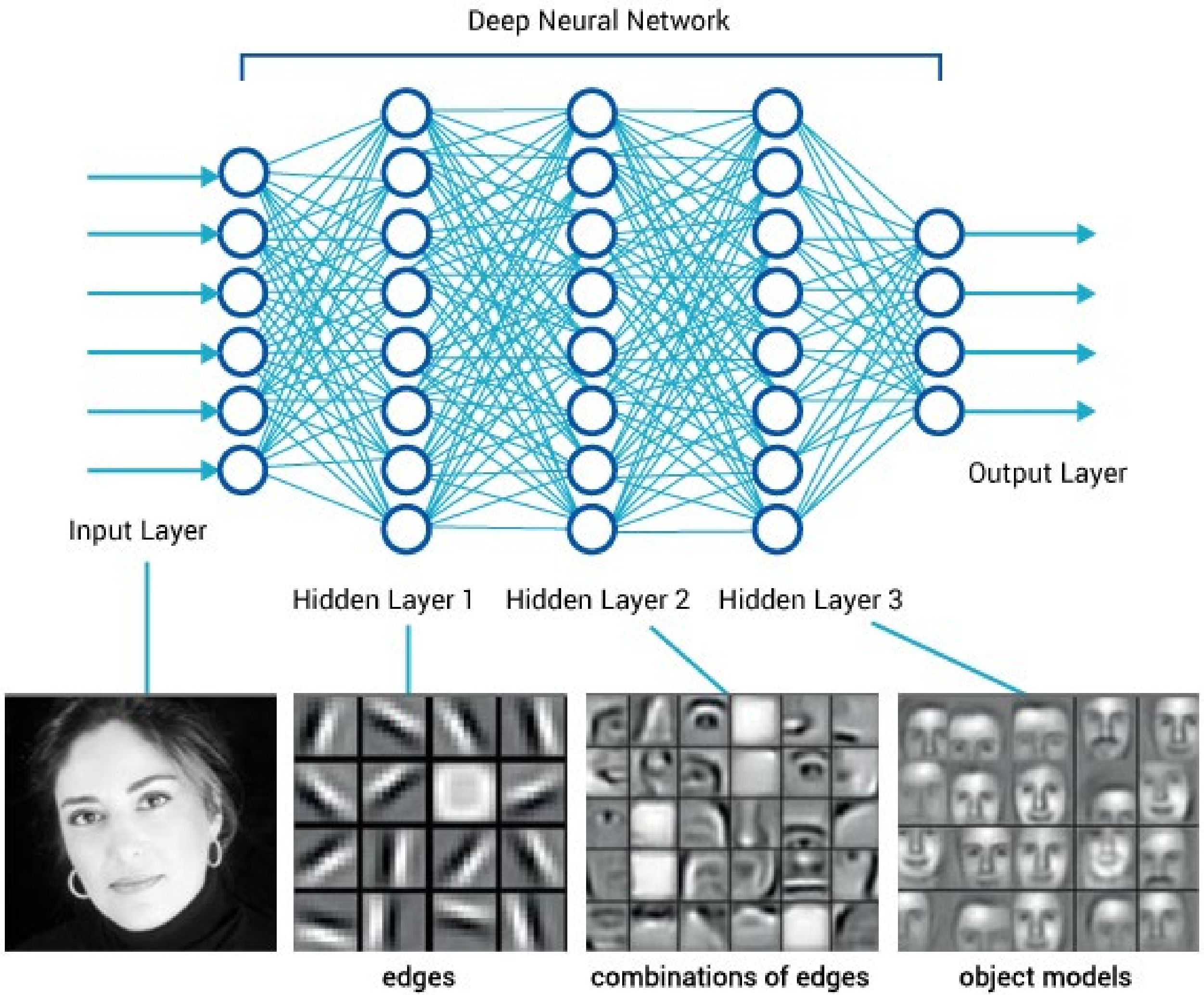
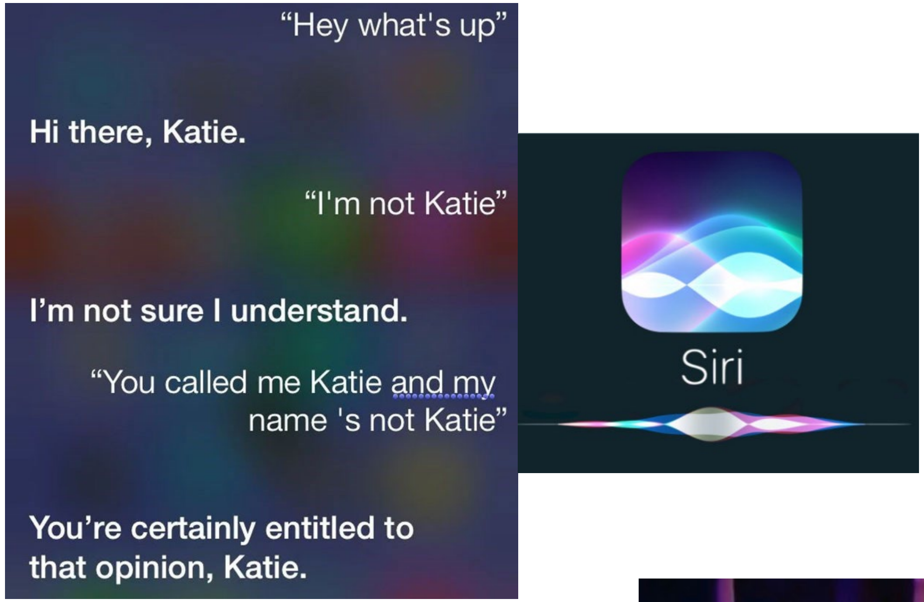


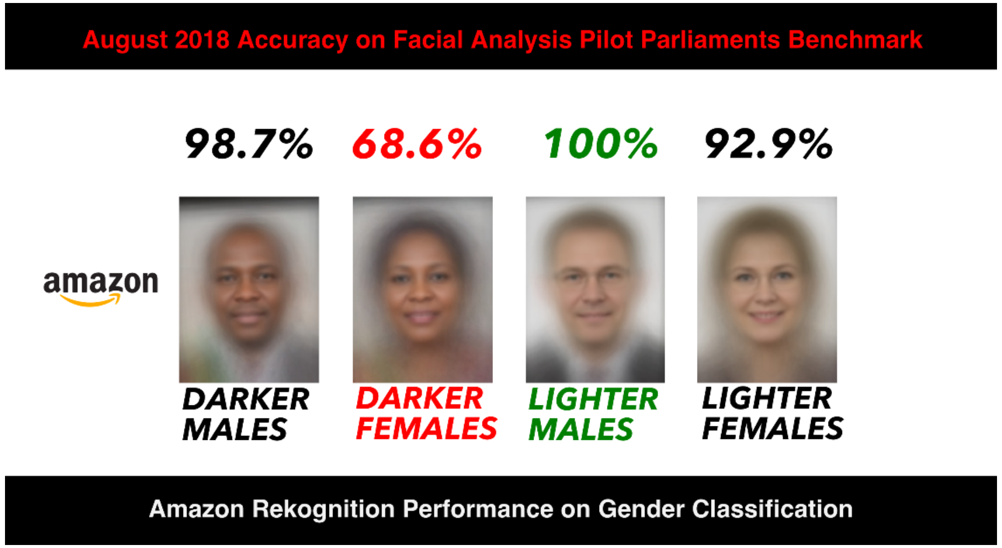
Image: http://www.amax.com/blog/wp-content/uploads/2015/12/blog_deeplearning3.jpg

Talk outline

- What are AI and machine learning?
- Dataset bias
- Solutions to dataset bias
- Trash
- Language to the rescue

AI makes mistakes…





Dataset Bias

Problem: dataset bias

What your AI is trained on

What it's asked to label

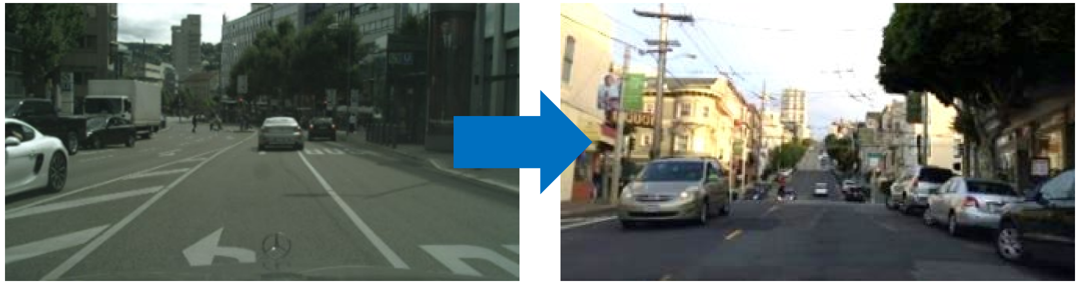
"Dataset Bias"
"Domain Shift"

vs

Adapting visual category models to new domains K Saenko, B Kulis, M Fritz, T Darrell, ECCV, 2010

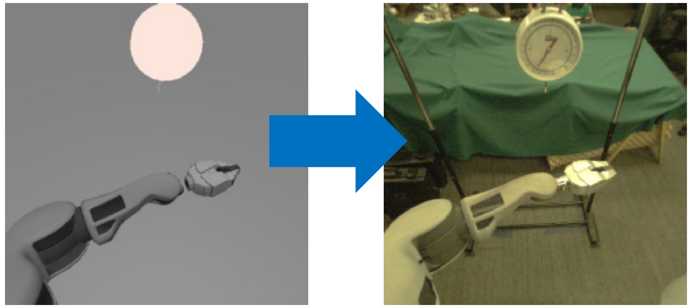
When does dataset bias happen?

From one city to another



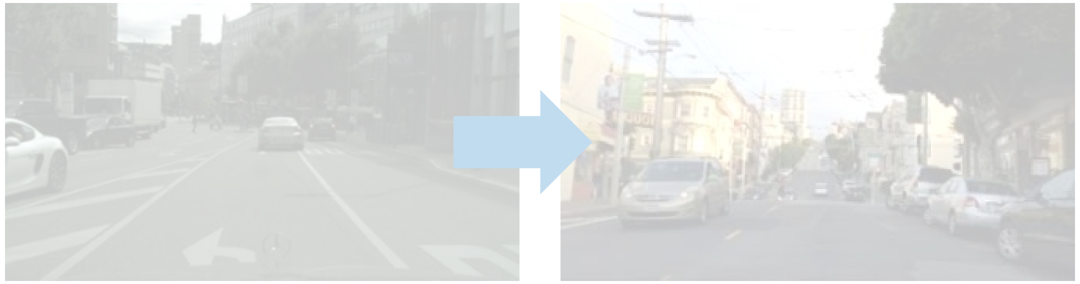
From web to robot

From simulated to real control

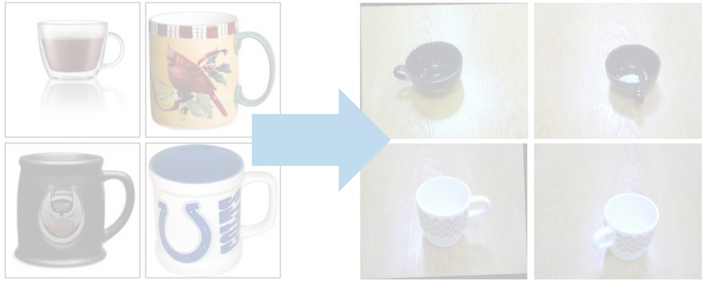


When does dataset bias happen?

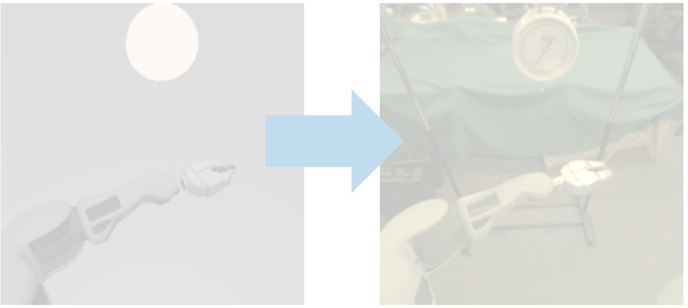
From one city to another



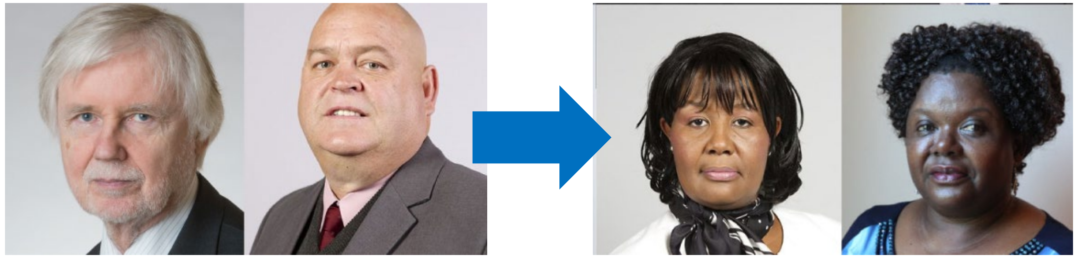
From web to robot



From simulated to real control

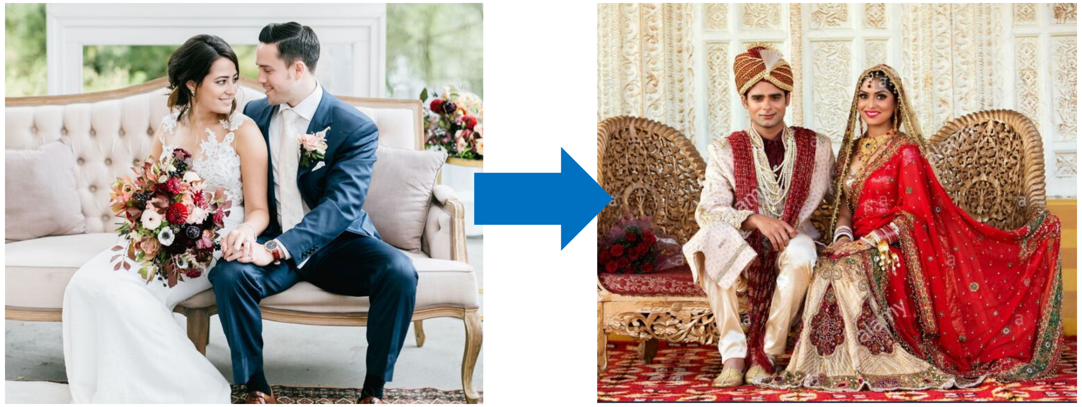


From one demographic to another



https://www.nytimes.com/2018/02/09/technology/facial-recognition-race-artificial-intelligence.html

From one culture to another



Dataset bias reduces accuracy

90%
accuracy

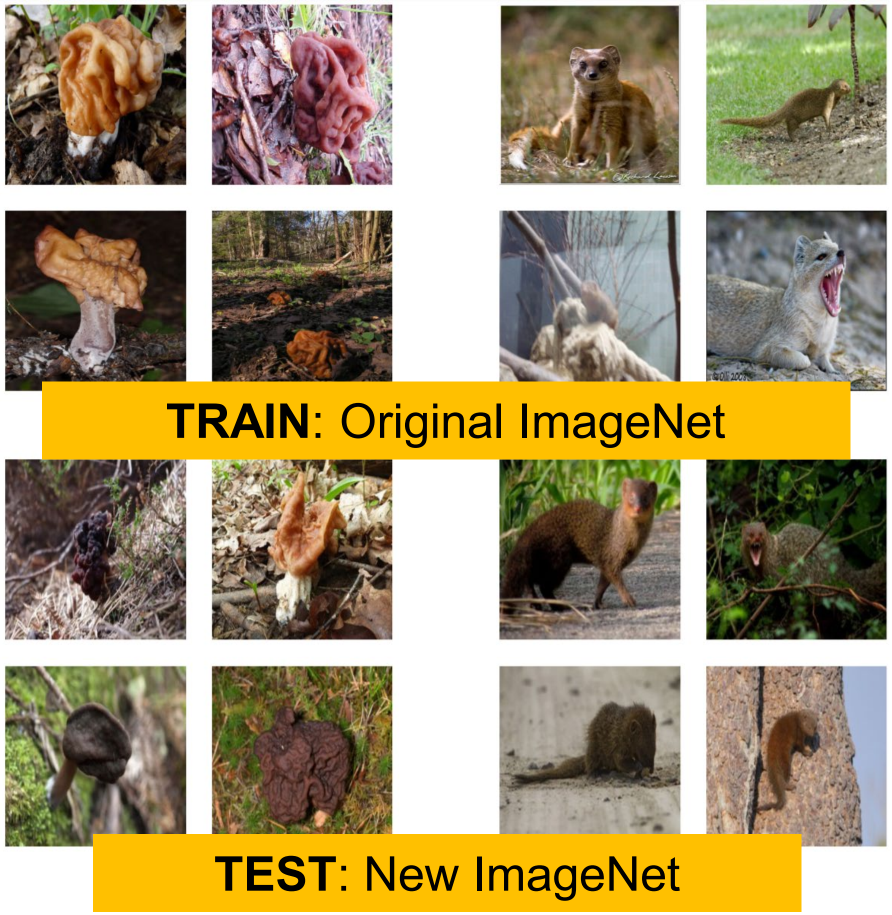
70% accuracy

target domain

Dataset bias reduces accuracy

Dataset bias reduces accuracy

11% – 14% accuracy drop!

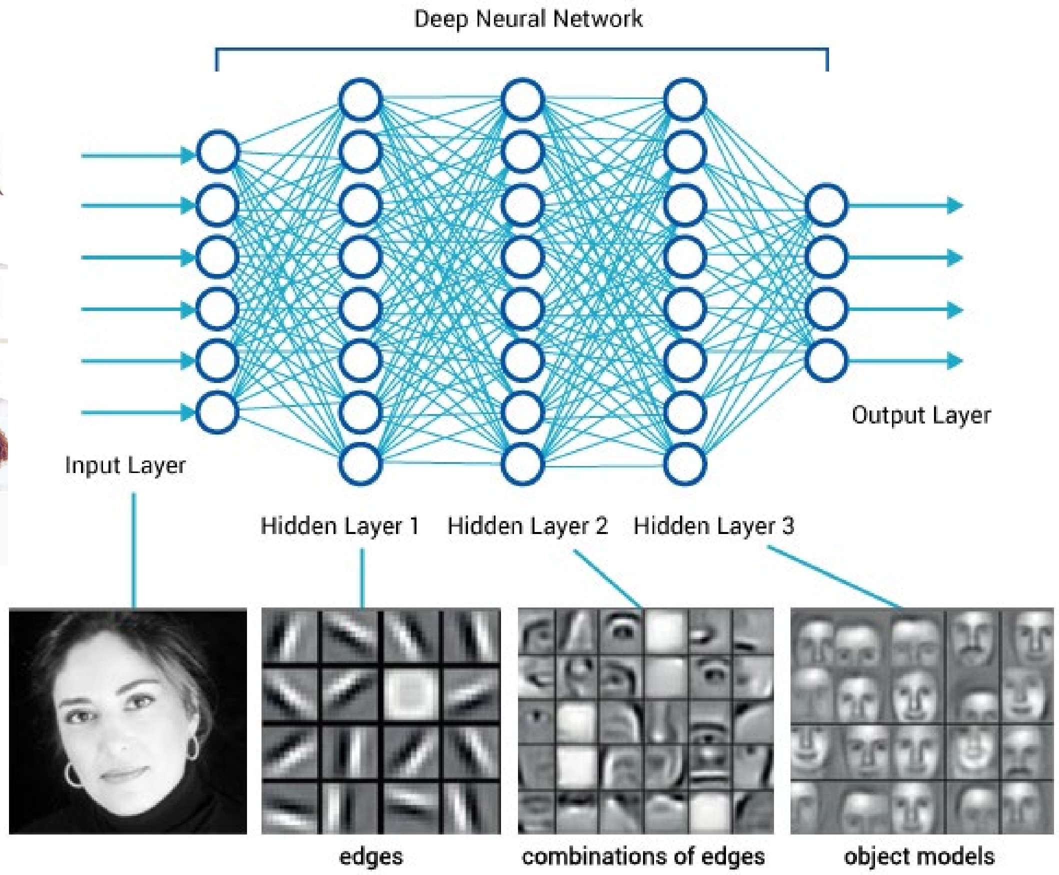


Do ImageNet Classifiers Generalize to ImageNet? [Recht et al. 2019]

Why does dataset bias affect machine learning?

Training: biased dataset

e.g., adult faces



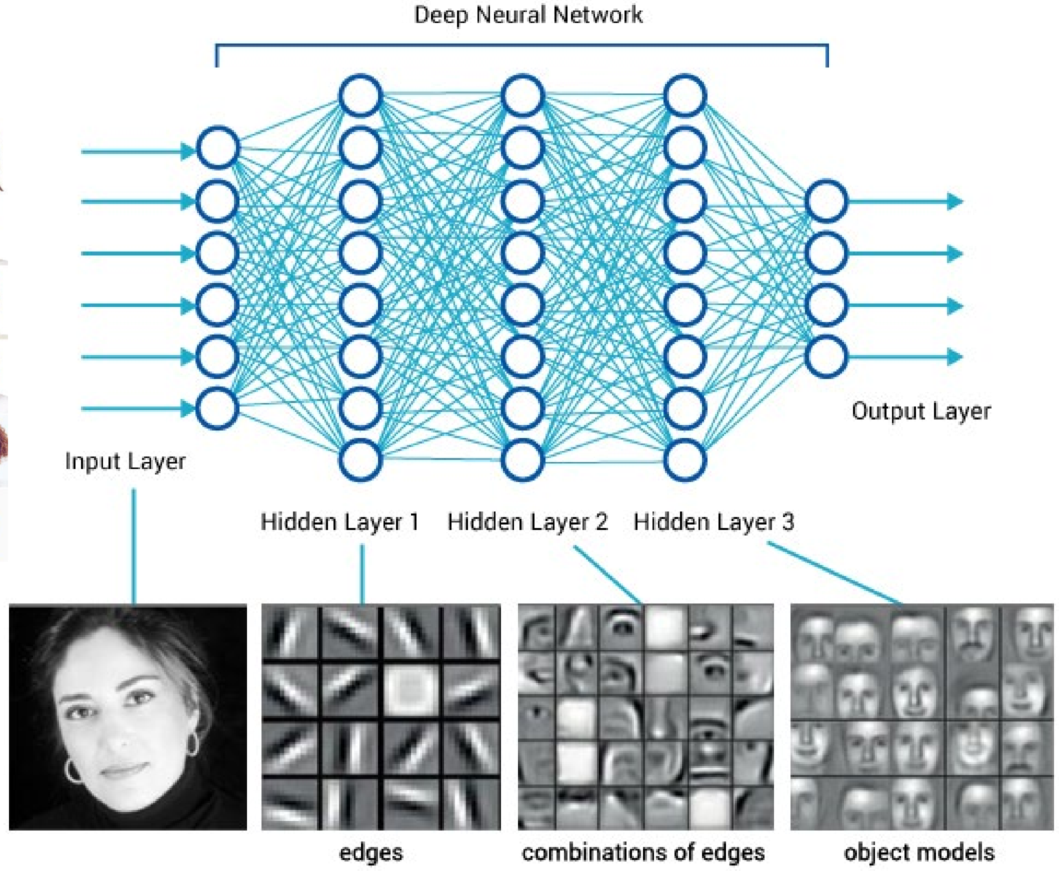
test

Poor accuracy

Problem: dataset bias

Training: biased dataset

e.g., mostly white faces



test

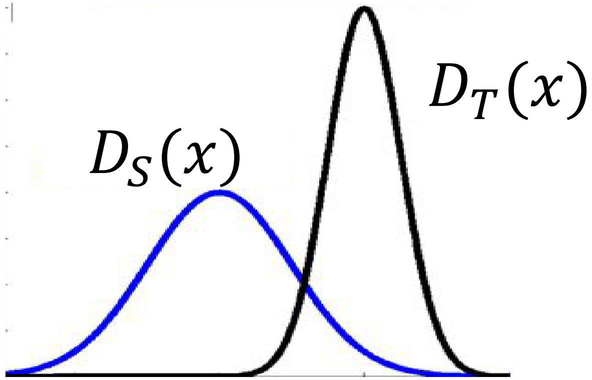
Poor accuracy

Solutions to dataset bias

- Collect and label more representative data
- Cons: $$$, time
- Long-tailed distribution of labels

A learning solution to dataset bias

Problem: dataset bias



Source Domain $\neq$ Target Domain

lots of labeled data

unlabeled data

$$D_S = \{(\mathbf{x}_i, y_i), \forall i \in \{1, \ldots, N\}\}$$

$$D_T = \{(\mathbf{z}_j, ?), \forall j \in \{1, \ldots, M\}\}$$

Goal: learn a classifier h that achieves low expected loss under distribution D_T

Theorem: target error is bounded by the source error, the difference between labeling functions f_S and f_T, and the divergence between the distributions D_S and D_T

$$\epsilon_T(h) \leq \epsilon_S(h) + d_1(\mathcal{D}_S, \mathcal{D}_T) + \min\left\{\mathrm{E}_{\mathcal{D}_S}\left[\left|f_S(\mathbf{x}) - f_T(\mathbf{x})\right|\right], \mathrm{E}_{\mathcal{D}_T}\left[\left|f_S(\mathbf{x}) - f_T(\mathbf{x})\right|\right]\right\}$$

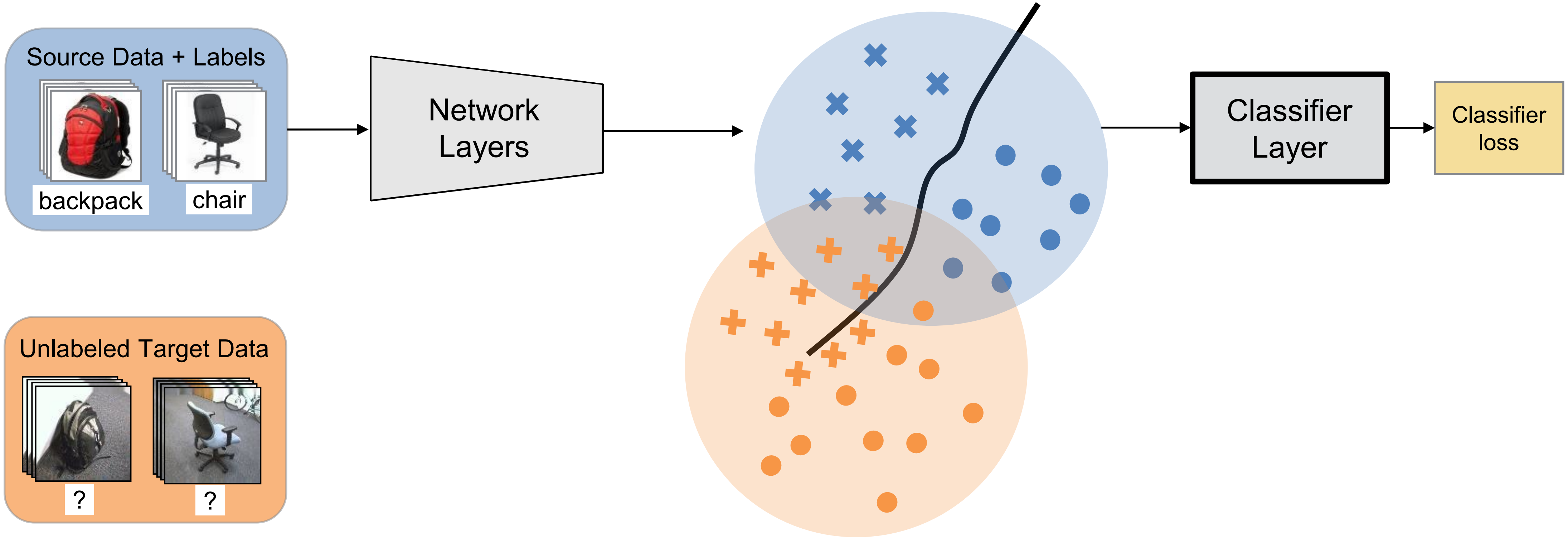
We expect the first and third terms to be small, the problem is the second

Divergence between two distributions; B is the set of measurable subsets under D and D’

$$d_1(\mathcal{D}, \mathcal{D}') = 2 \sup_{B \in \mathcal{B}} \left|\Pr_{\mathcal{D}}[B] - \Pr_{\mathcal{D}'}[B]\right|$$

A theory of learning from different domains, Shai Ben-David et al. 2009

What causes poor performance?



Problem: features are different

ADDA: Tzeng, Eric, et al. "Adversarial discriminative domain adaptation." CVPR 2017.

What causes poor performance?

- Train and test data distributions are different
- Model lacks discriminative features

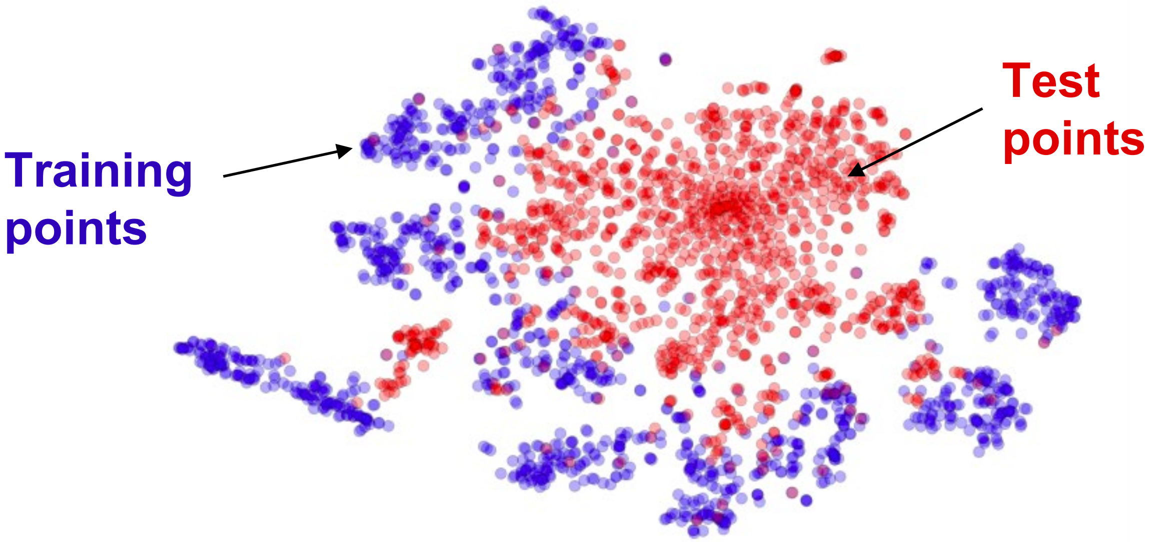
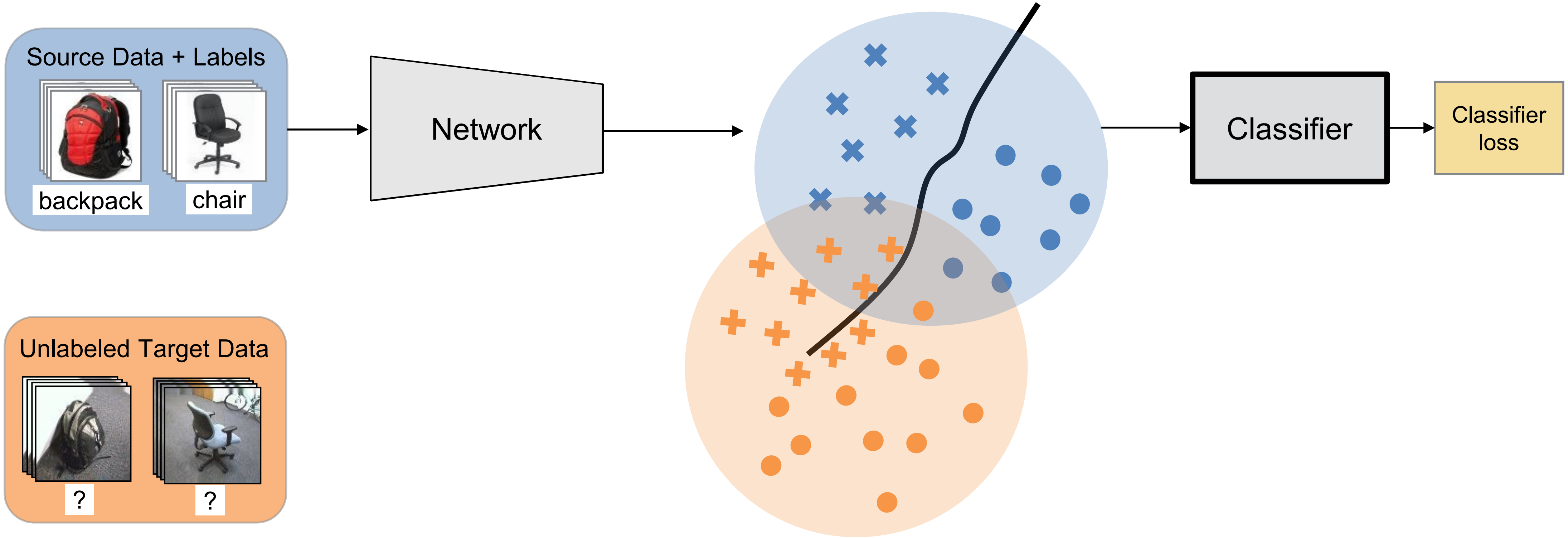


Figure from Ganin and Lempitsky. "Unsupervised domain adaptation by backpropagation." ICML 2015

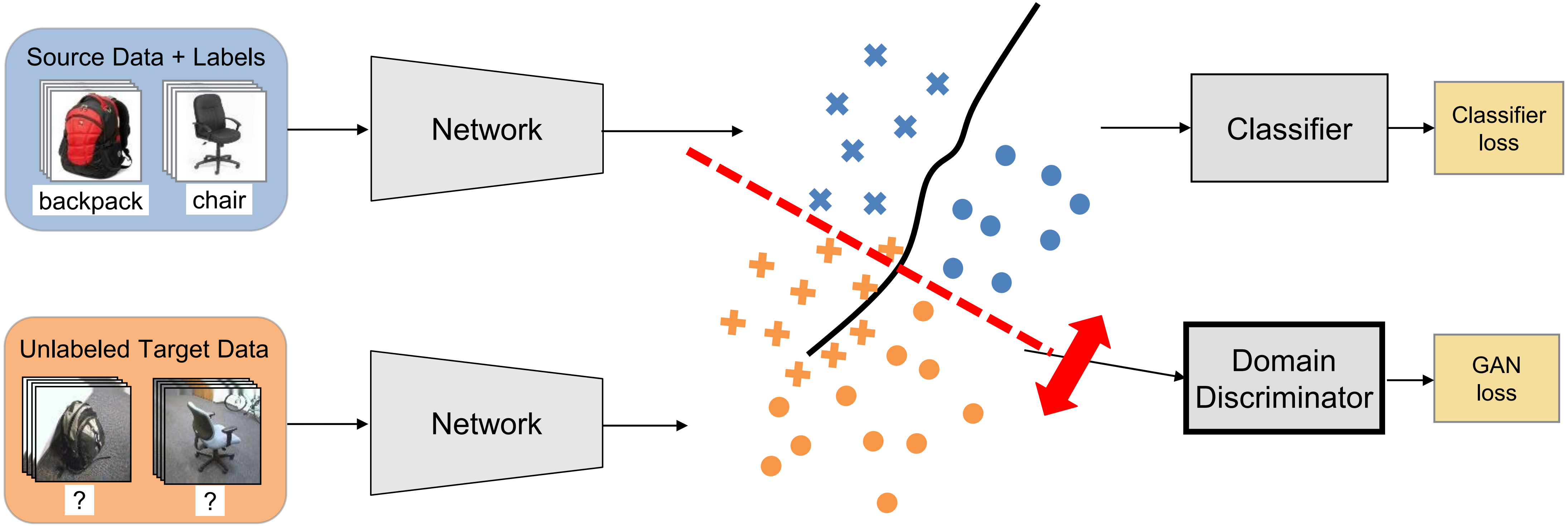
IDEA 1: Adversarial domain alignment



Goal: align distributions

ADDA: Tzeng, Eric, et al. "Adversarial discriminative domain adaptation." CVPR 2017.

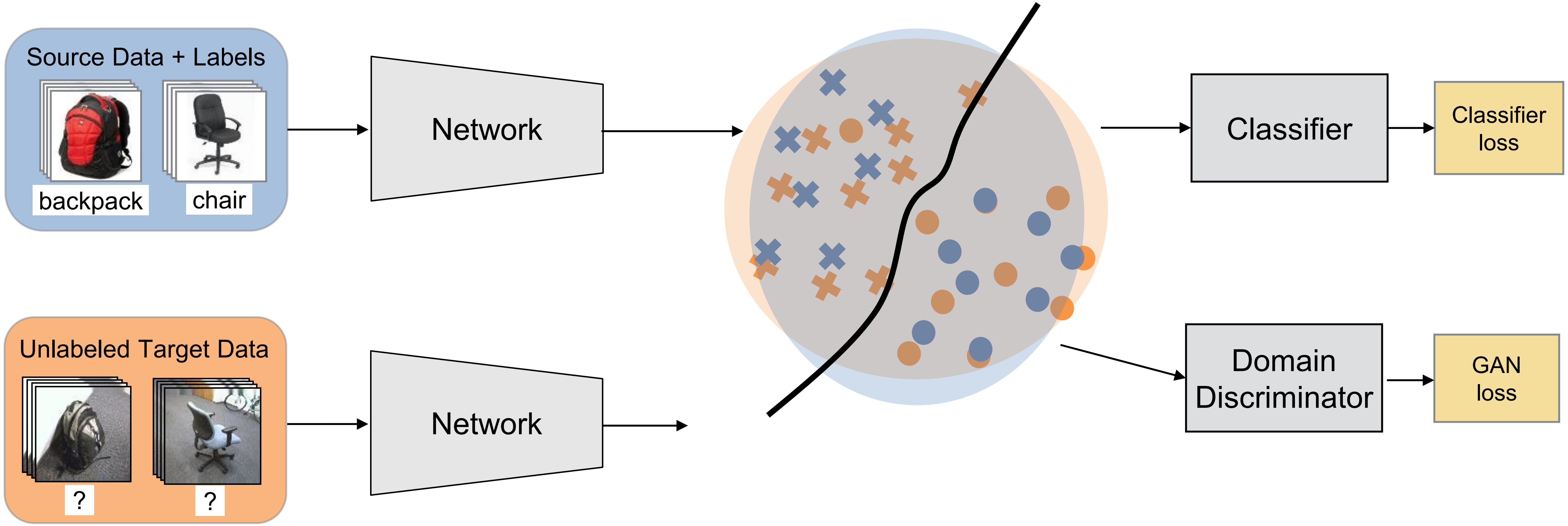
IDEA 1: Adversarial domain alignment



Goal: align distributions

ADDA: Tzeng, Eric, et al. "Adversarial discriminative domain adaptation." CVPR 2017.

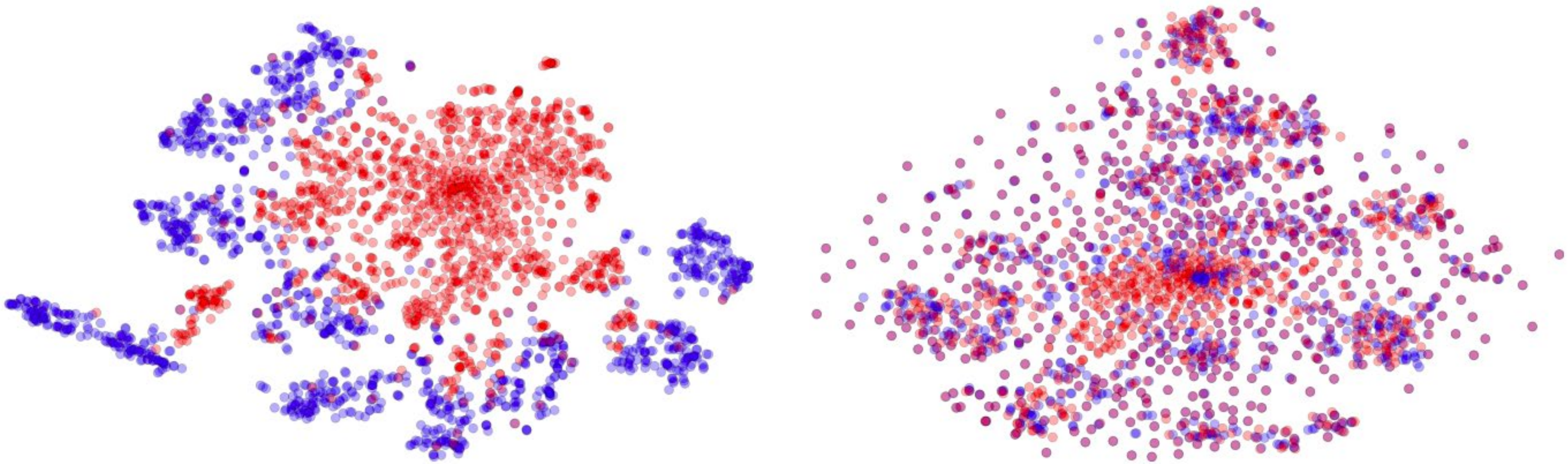
IDEA 1: Adversarial domain alignment



Goal: align distributions

ADDA: Tzeng, Eric, et al. "Adversarial discriminative domain adaptation." CVPR 2017.

Domain alignment: feature visualization on digits

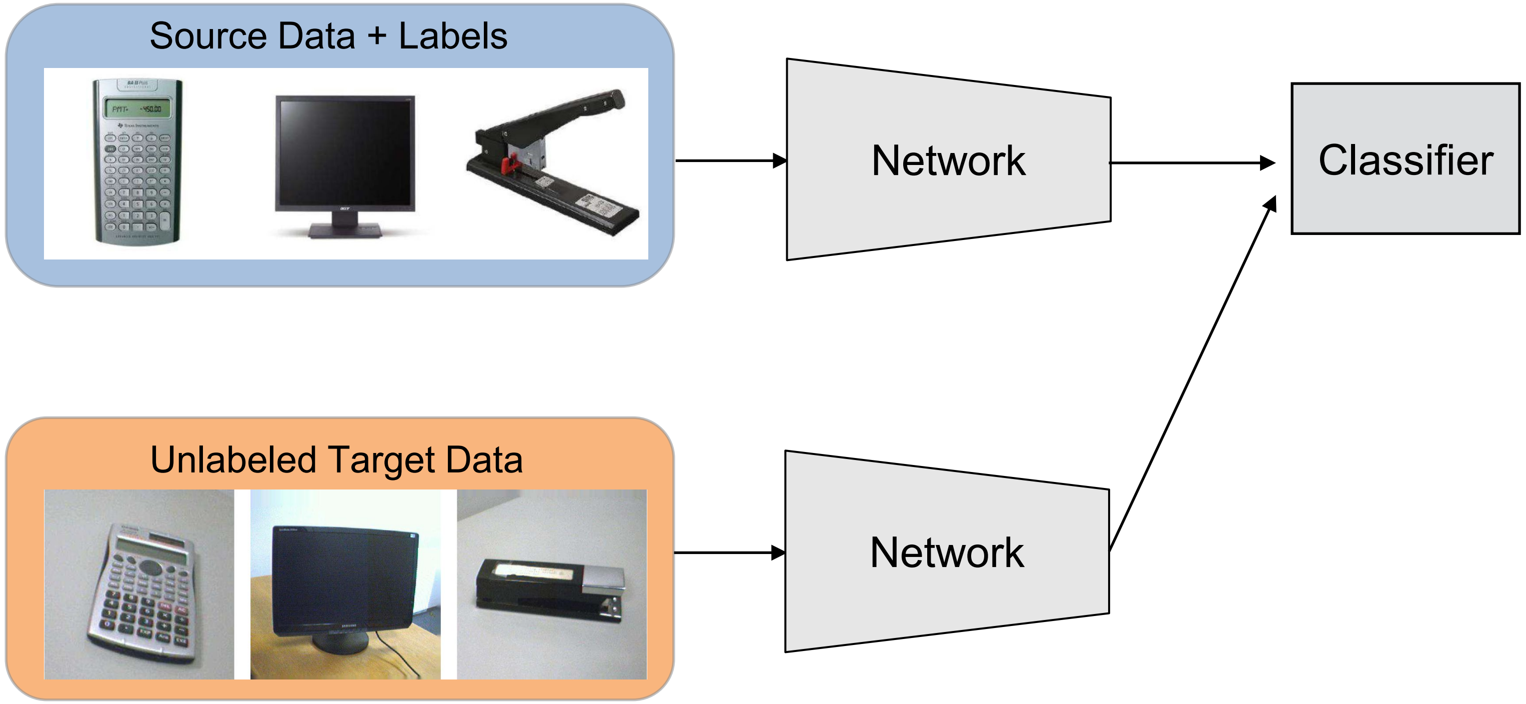


(a) Non-adapted (b) Adapted

Effect of adaptation on features in MNIST → MNIST-M shift
(top feature extractor layer)

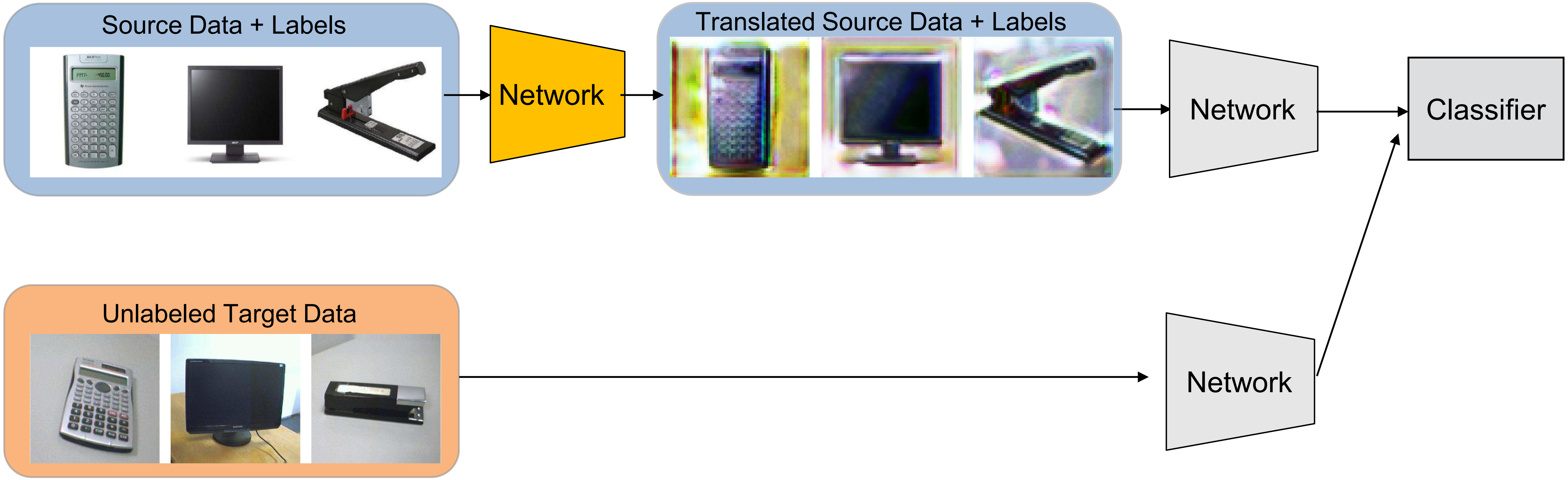
Figure from Ganin, Yaroslav, and Victor Lempitsky. "Unsupervised domain adaptation by backpropagation." ICML 2015

IDEA 2: Pixel-space domain alignment



CyCADA, **Hoffman**, Tzeng, Park, Zhu, Isola, Saenko, Efros, Darrell, ICML 2018
From source to target and back: Symmetric Bi-Directional Adaptive GAN, Russo et al. CVPR 2018

IDEA 2: Pixel-space domain alignment

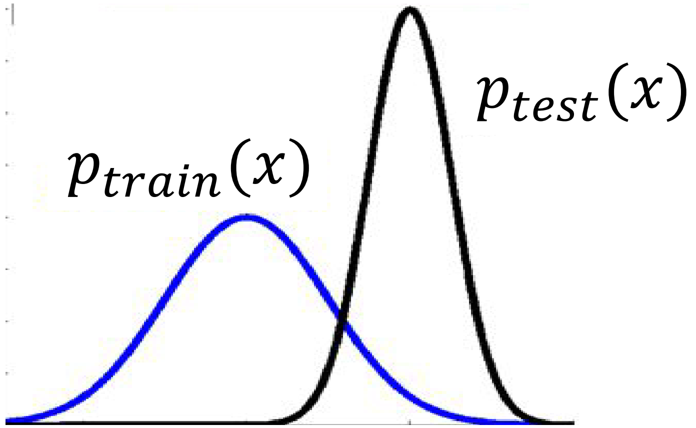


CyCADA, Hoffman, Tzeng, Park, Zhu, Isola, Saenko, Efros, Darrell, ICML 2018
From source to target and back: Symmetric Bi-Directional Adaptive GAN, Russo et al. CVPR 2018

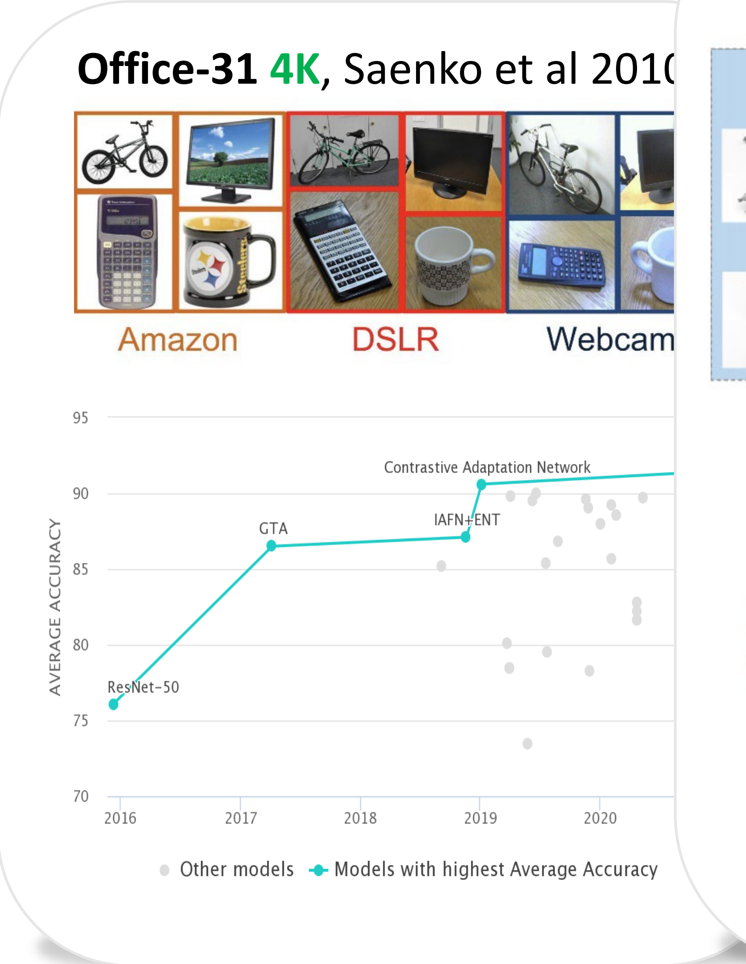
Does domain alignment improve o.o.d accuracy?

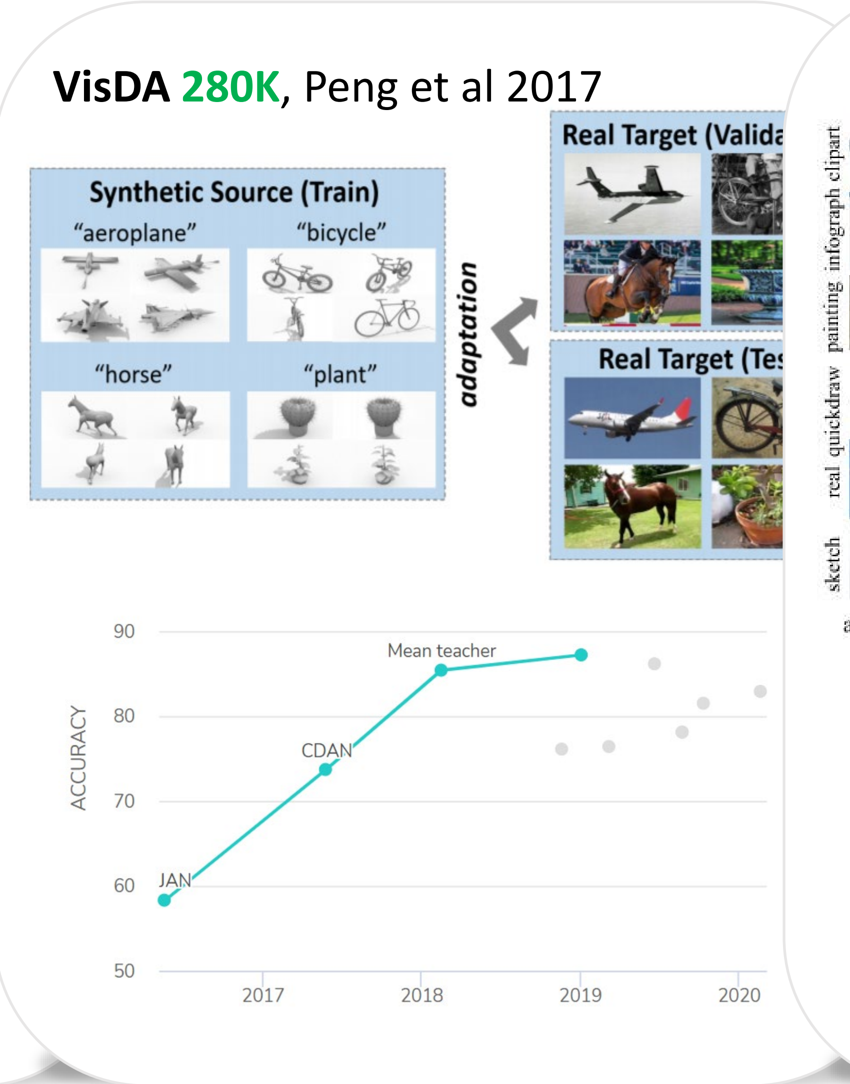
i.i.d testing overestimates generalization

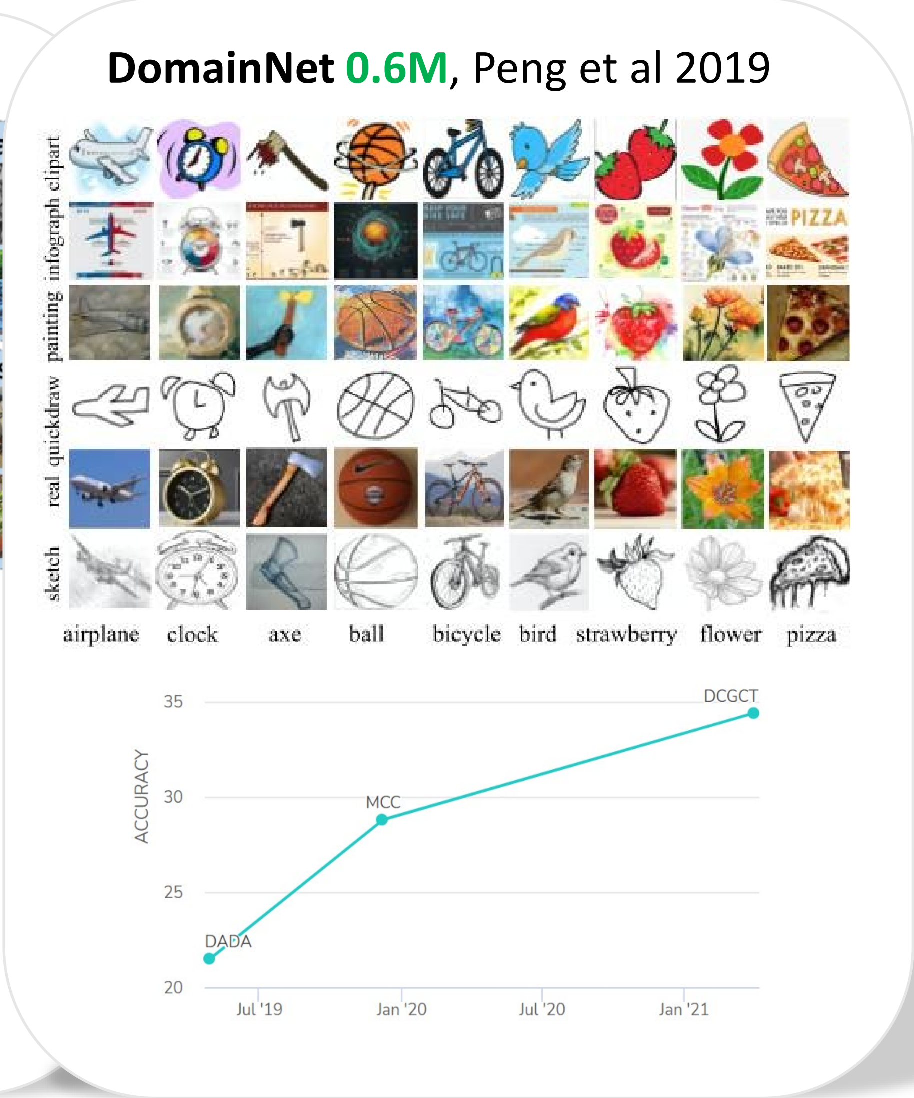
- The vast majority of datasets have i.i.d test sets
- This over-estimates generalization of models
- Solution: evaluate on out-of-distribution (o.o.d) data



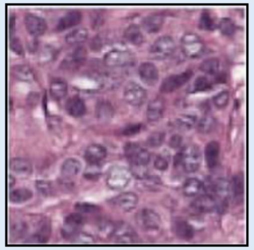
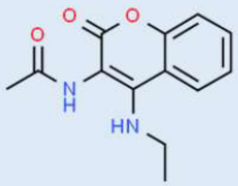
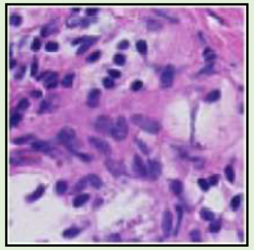
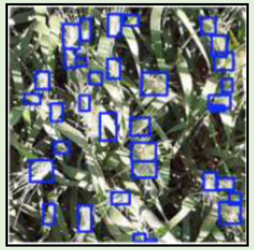
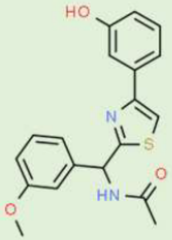
O.O.D datasets





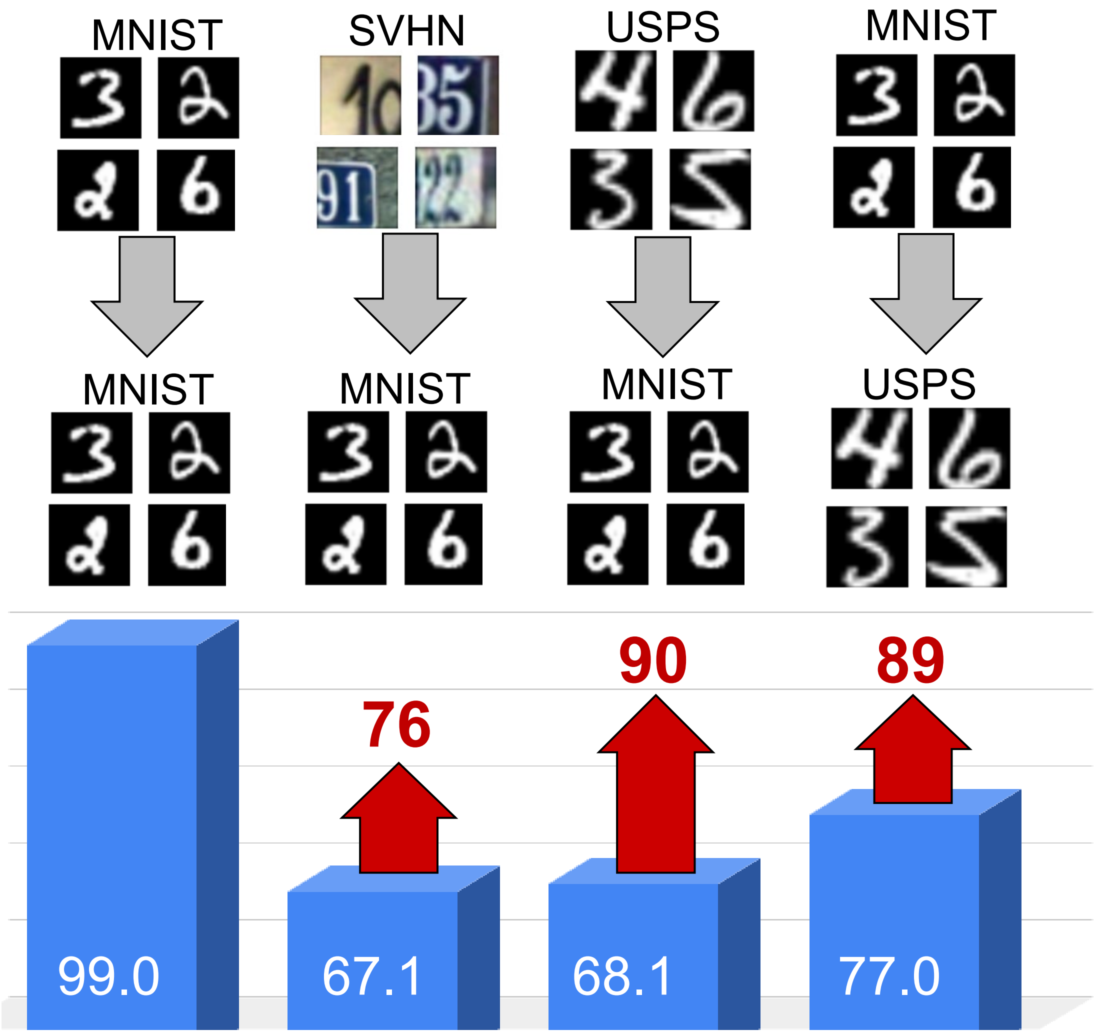


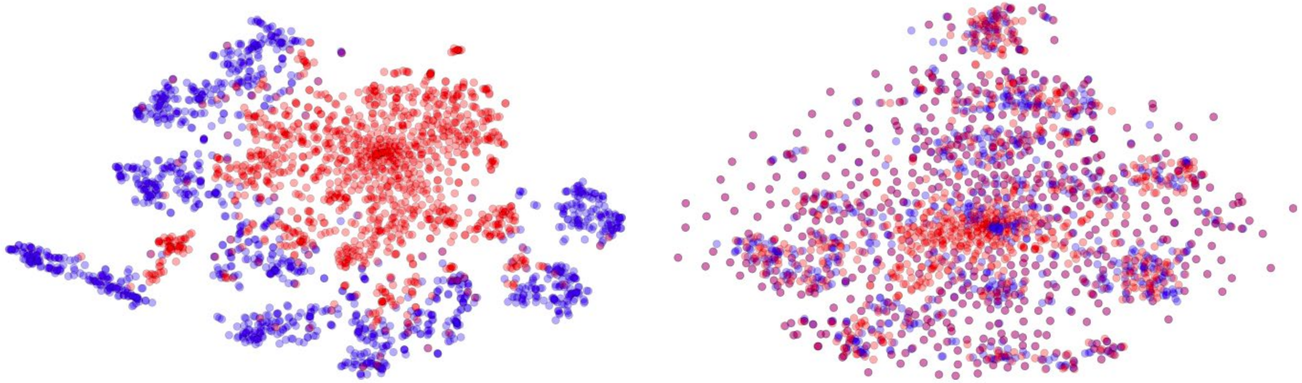
WILDS Dataset (Koh et al., 2021)

Dataset	iWildCam	Camelyon17	FMoW	PovertyMap	GlobalWheat	OGB-MolPCBA	CivilComments	Amazon
Input (x)	camera trap photo	tissue slide	satellite image	satellite image	wheat image	molecular graph	online comment	product review
Prediction (y)	animal species	tumor	land use	asset wealth	wheat head bbox	bioassays	toxicity	sentiment
Domain (d)	camera	hospital	time, region	country, ru/ur	location, time	scaffold	demographic	user
Source example							What do Black and LGBT people have to do with bicycle licensing?	Overall a solid package that has a good quality of construction for the price.
Target example							As a Christian, I will not be patronizing any of those businesses.	I *loved* my French press, it's so perfect and came with all this fun stuff!

Wilds dataset (Koh et al., 2021) contains labeled data from the source domains (for training), validation domains (for hyperparameter selection), and target domains (for held-out evaluation). In Wilds 2.0, we extend these datasets with unlabeled data

Results on image classification





(a) Non-adapted

(b) Adapted

[Ganin, ICML 2015]

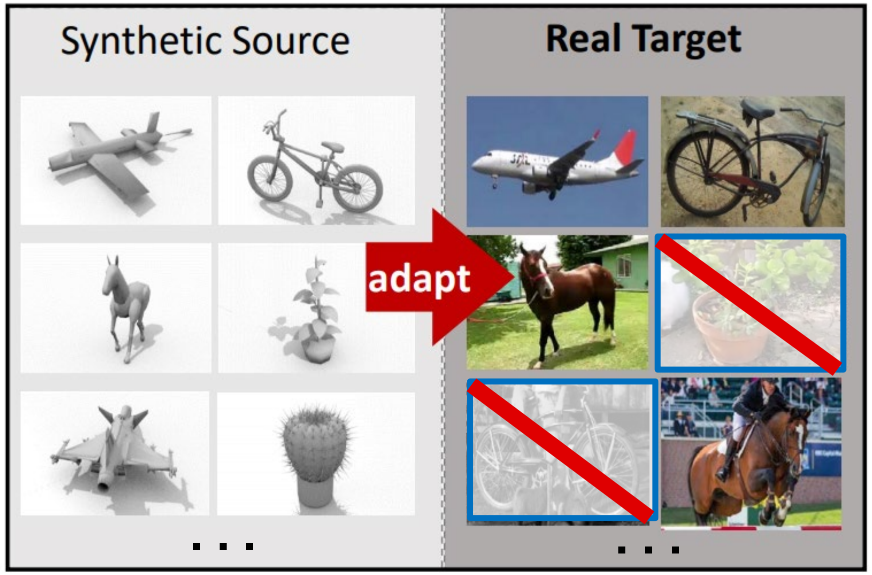
Takeaway:

Domain adaptation can improve accuracy on target data without any labels; “unsupervised fine-tuning”

ADDA: Tzeng, Eric, et al. "Adversarial discriminative domain adaptation." CVPR 2017.

Results on image classification

- Example: syn2real object recognition
- 6 categories missing in target ("partial" shift)
- **improves accuracy compared to SOTA**



Source Only	49.9
DANN (Ganin et al., 2016)	38.7
ETN (Zhangjie Cao, 2019)	59.8
STA (Liu et al., 2019)	48.2
UAN (You et al., 2019)	39.7
DANCE (ours)	**73.7**

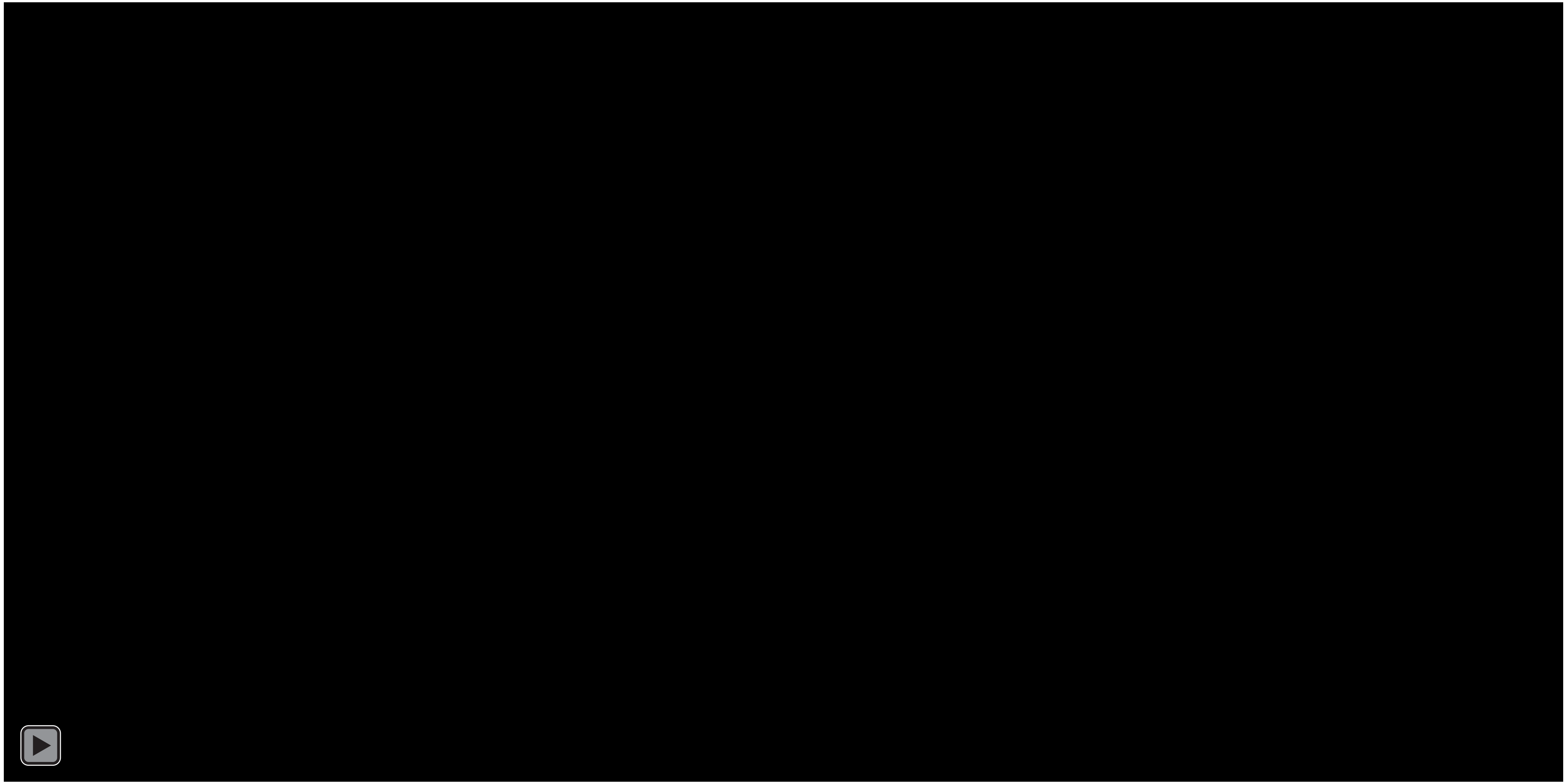
+14%

Universal Domain Adaptation through Self Supervision, Saito, Kim, Sclaroff, Saenko, NeurIPS, 2020

Application: Image segmentation for driving

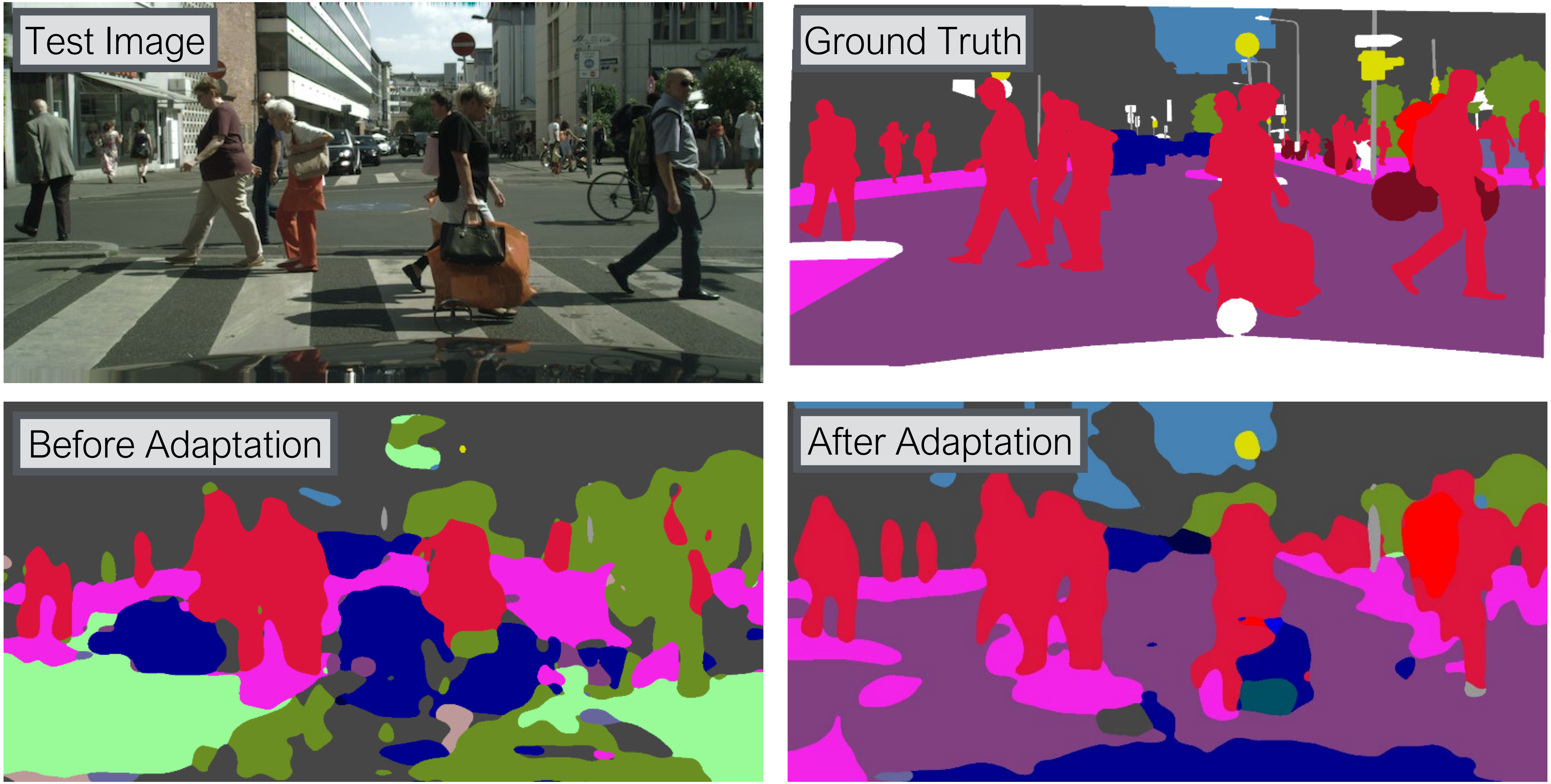
- Labeled data is very expensive
- Learn from simulation?

Alignment: make training data look like test data



CycleGAN. Zhu*, Park*, Isola, Efros. ICCV 2017.

Segmentation Results: Train on GTA game, test on real city



Hoffman, Tzeng, Park, Zhu, Isola, Saenko, Efros, Darrell, ICML18.

Pixel-level alignment with CyCADA [Hoffman'18]

Source image (GTA5) | Adapted source image (**Ours**) | Target image (CityScapes)

Pixel accuracy on target
Source-only: 54.0%
Adapted (**ours**): **83.6%**

Takeaway:

Unsupervised image-to-image translation can change the image style to match the target domain

CyCADA, Hoffman, Tzeng, Park, Zhu, Isola, Saenko, Efros, Darrell, ICML 2018

A Real-world Small Data Problem

Classifying recyclables

Only 30% of waste is recycled

- The EPA estimates that 75% of the American waste stream is recyclable, but we only recycle about **30%** of it.

- Dull, dirty, dangerous job

ZeroWaste Dataset: Towards Deformable Object Segmentation in Cluttered Scenes

Dina Bashkirova, Mohamed Abdelfattah, Ziliang Zhu,
James Akl, Fadi Alladkani, Ping Hu,
Vitaly Ablavsky, Berk Calli, Sarah Adel Bargal and Kate Saenko

CVPR 2022

https://github.com/dbash/zerowaste

TACO dataset

Labeled Waste in the Wild dataset

ReSortIT dataset

Real MRF facility (our setup)

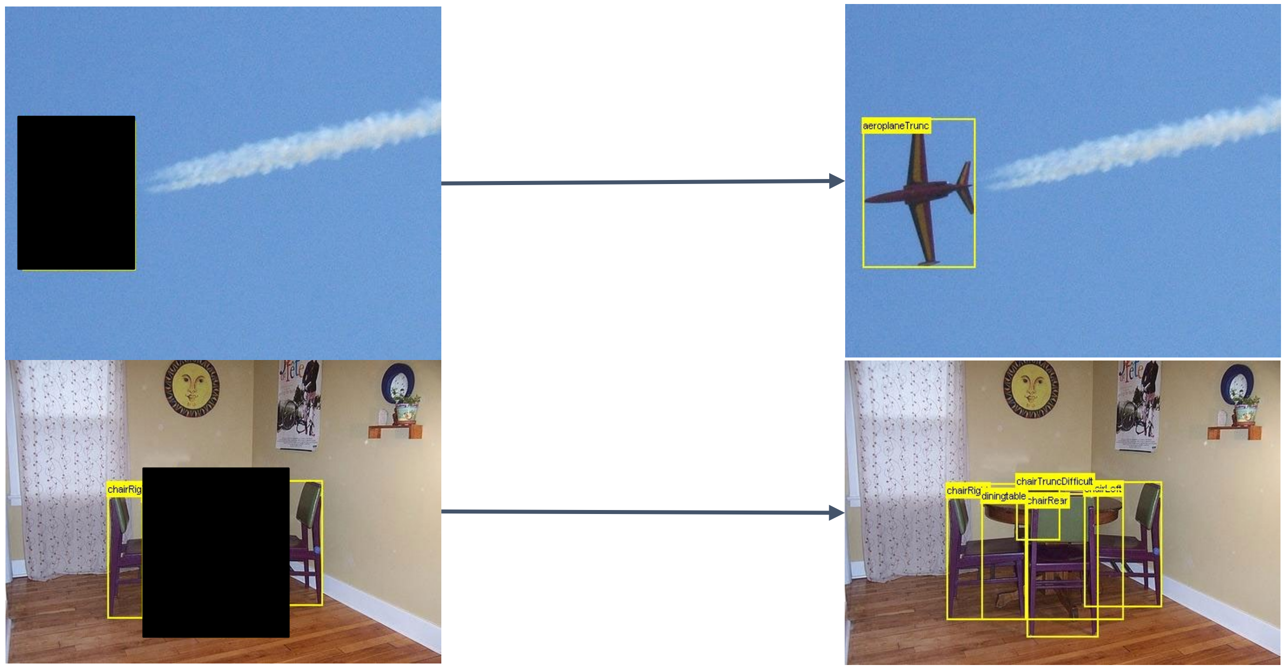
our ZeroWaste dataset

ZeroWaste: Overview

- Task: remove non-paper objects from the conveyor belt;
- Annotated objects of four classes: cardboard, soft plastic, rigid plastic and metal;
- 4503 fully annotated frames (ZeroWaste-f)
- 6212 unlabeled frames (ZeroWaste-s);
- 1410 frames before and after collection (ZeroWaste-w)

ZeroWaste: key differences

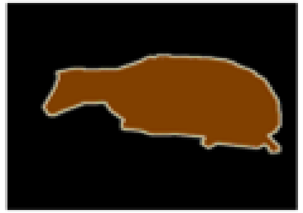
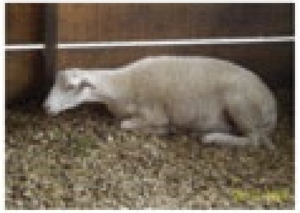
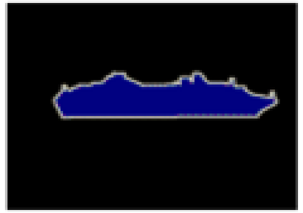
- A lot of background clutter and occlusions
- Objects out of context



ZeroWaste: key differences

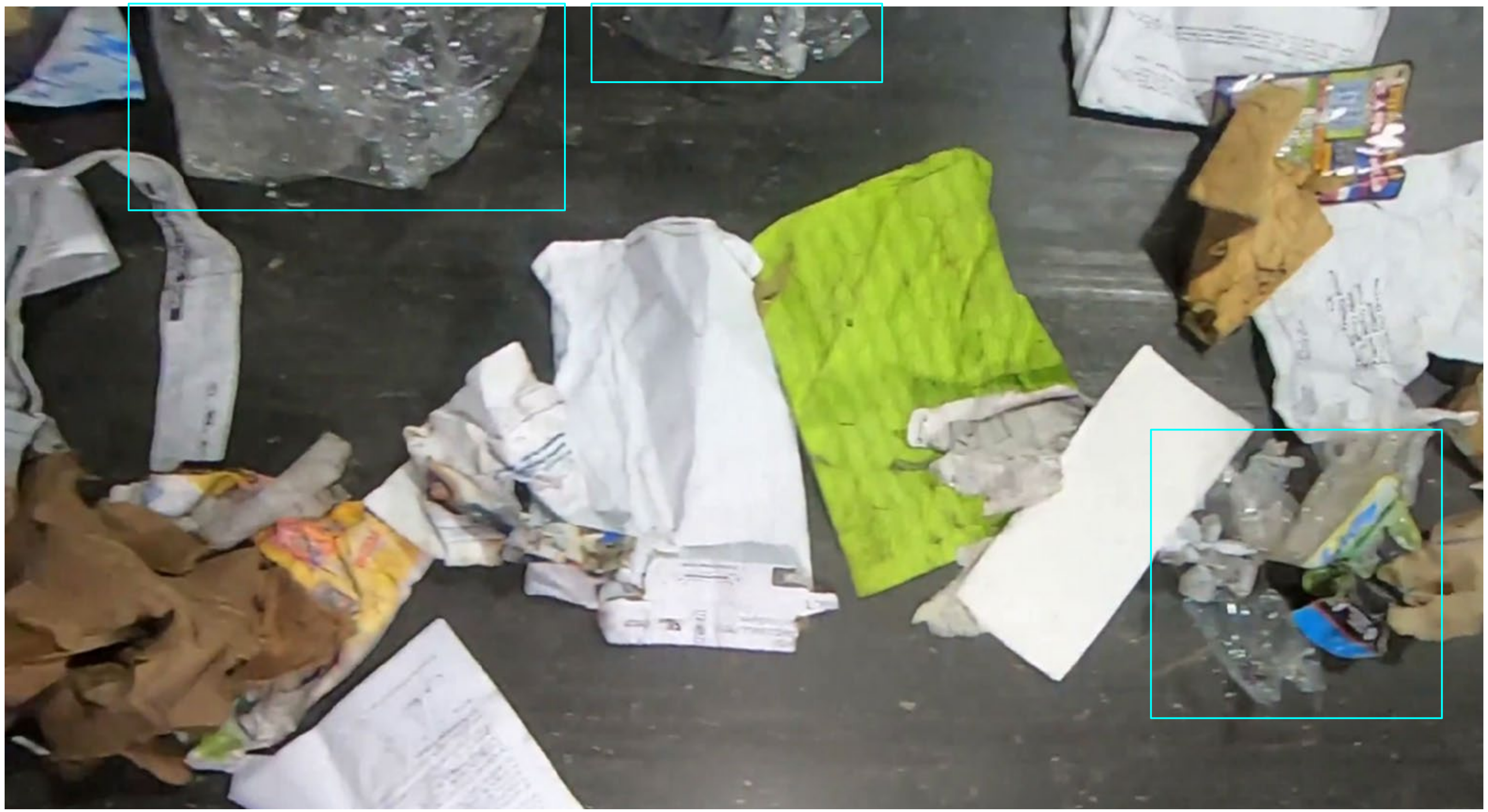
- A lot of background clutter and occlusions
- Objects out of context
- **Highly deformable objects**




ZeroWaste: key difference

- A lot of background clutter and occlusions
- Objects out of context
- Highly deformable objects
- **Translucent objects**



Summary: Learning Generalizable Visual Representations

- O.o.d testing is more realistic than i.i.d
- Can learn domain invariance using self supervision, “bridge” domain
- Real-world small data problem: recycling dataset
- Learning from language helps