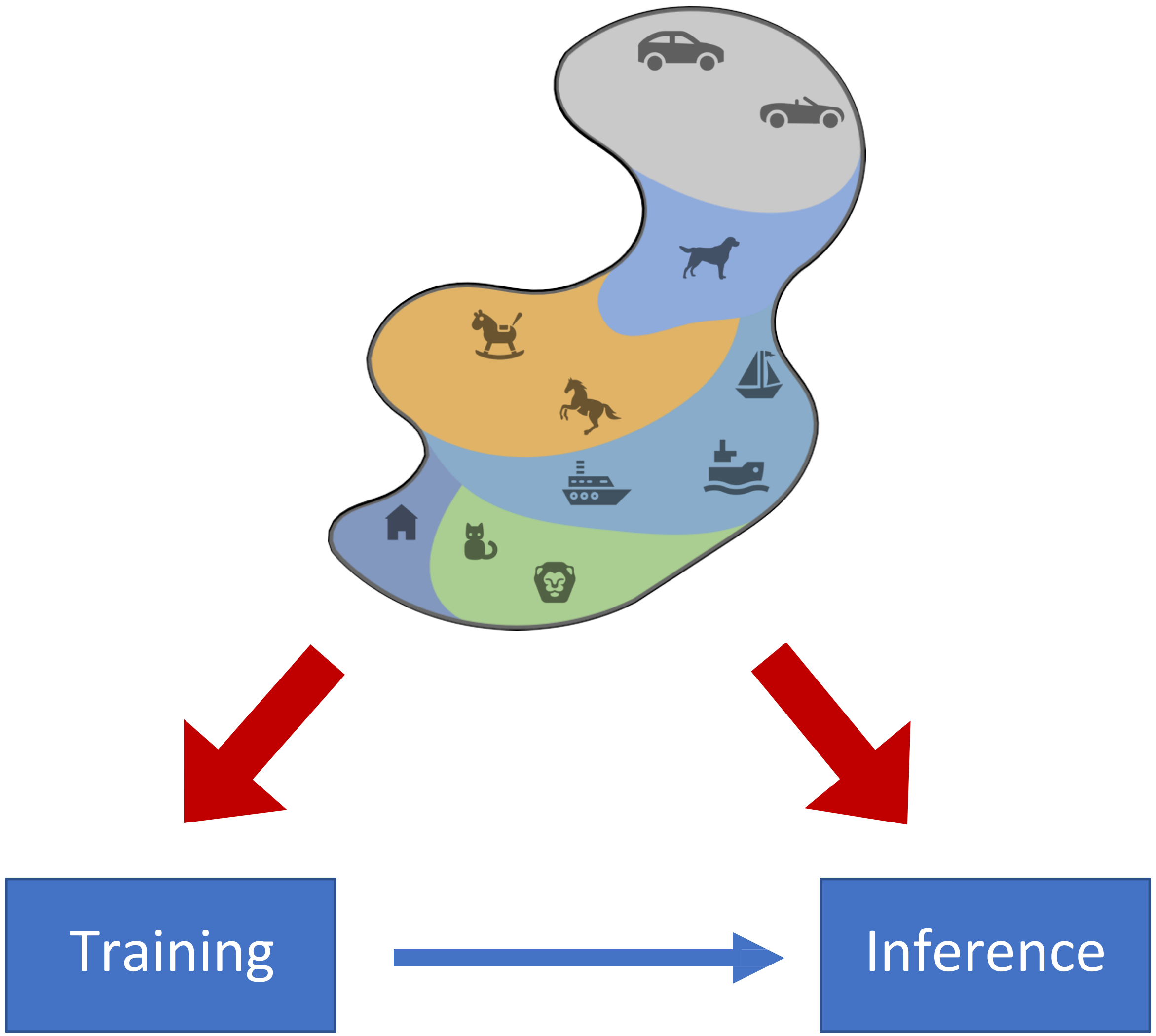
CMSC 475/675 Neural Networks

Lecture 6: Domain Adaptation

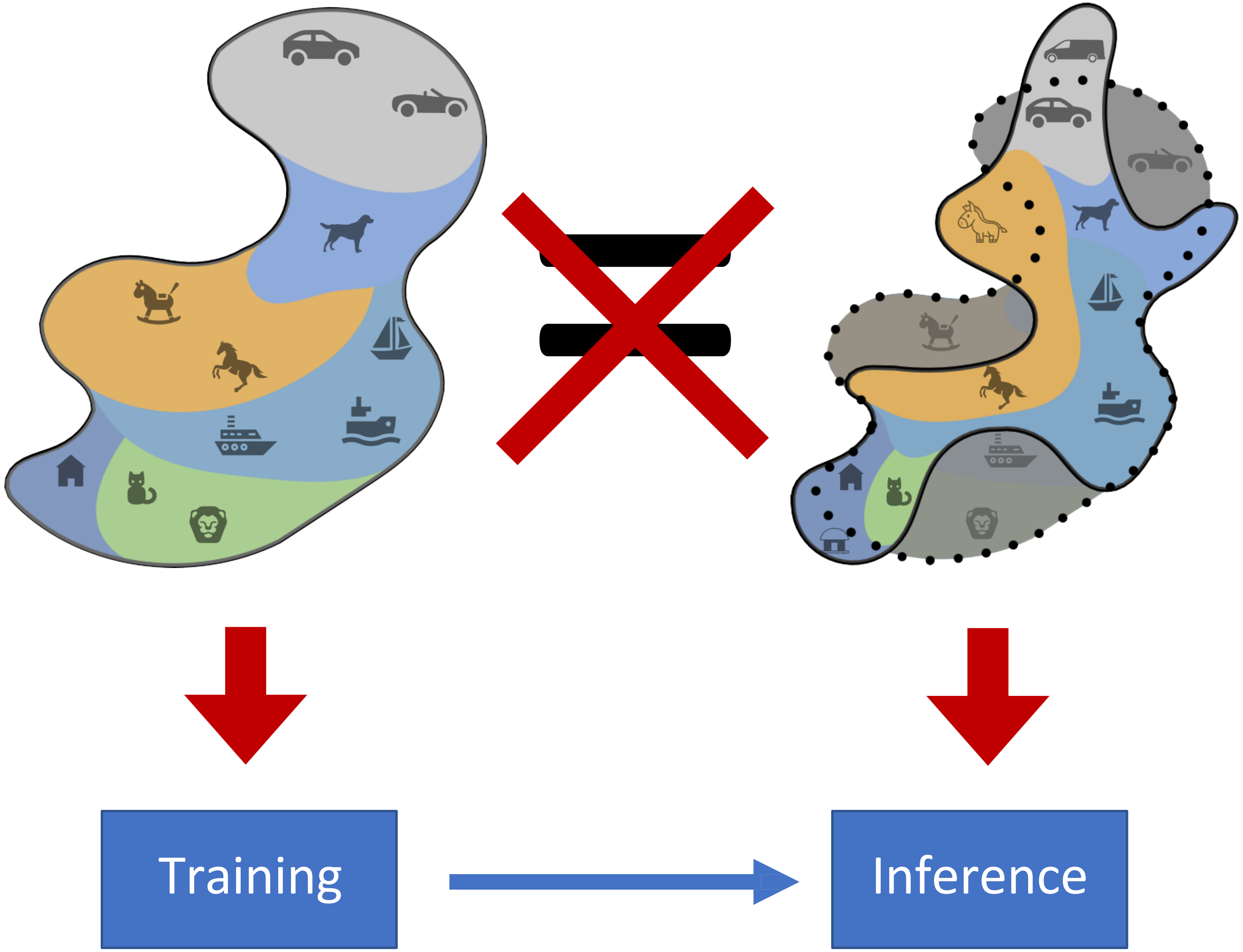
A Limitation of the (Supervised) ML Framework



Measure of performance:
Fraction of mistakes during testing

But: In reality,
the distributions we **use** ML on
are NOT the ones we **train** it on

A Limitation of the (Supervised) ML Framework



Measure of performance:
Fraction of mistakes during testing

But: In reality,
the distributions we **use** ML on
are NOT the ones we **train** it on

What can go wrong?

Standard *i.i.d.* Assumption in Machine Learning

Training Data Distribution = Test Data Distribution

"Independent and Identically Distributed"

Models learn useful patterns

Standard *i.i.d.* Assumption in Machine Learning

Training Data Distribution $\neq$ Test Data Distribution

IID Assumption collapses in real-world “in-the-wild” settings

Model performance deteriorates

Example Scenarios

What your net is trained on

What it's asked to label

"Dataset Bias"
"Domain Shift"
"Domain Adaptation"
"Domain Transfer"

Figure Source: Kate Saenko

Example Scenarios

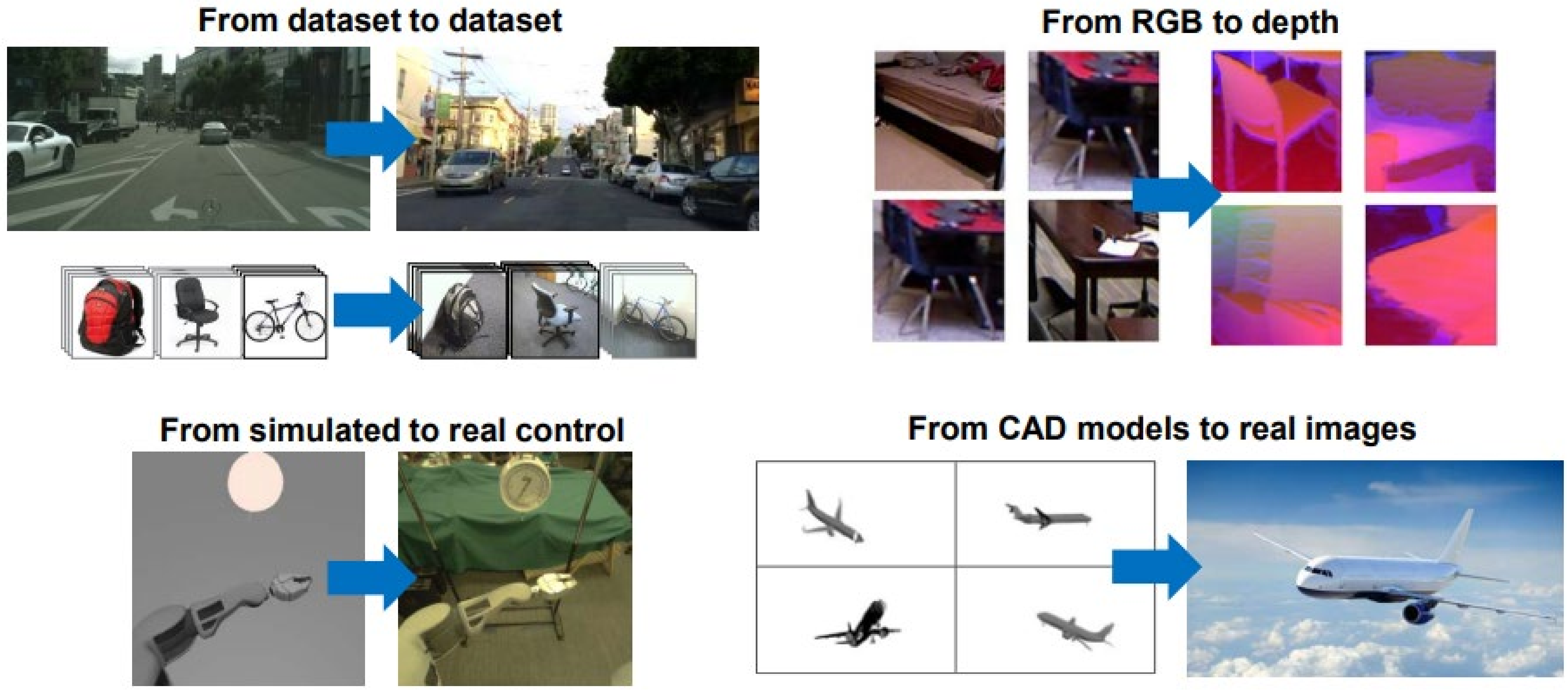
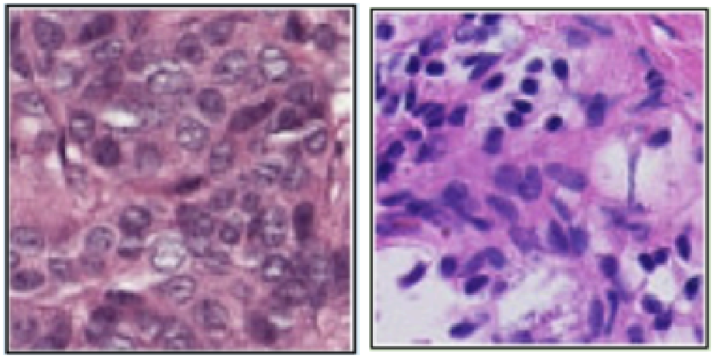


Figure Source: Kate Saenko

Example Scenarios

Tumor detection & classification

Source hospital Target hospital



varying imaging techniques, different demographics

Land use classification

Source region Target region

appearance of buildings, plants; weather conditions, pollution

Text classification, generation

Source corpus Target corpus

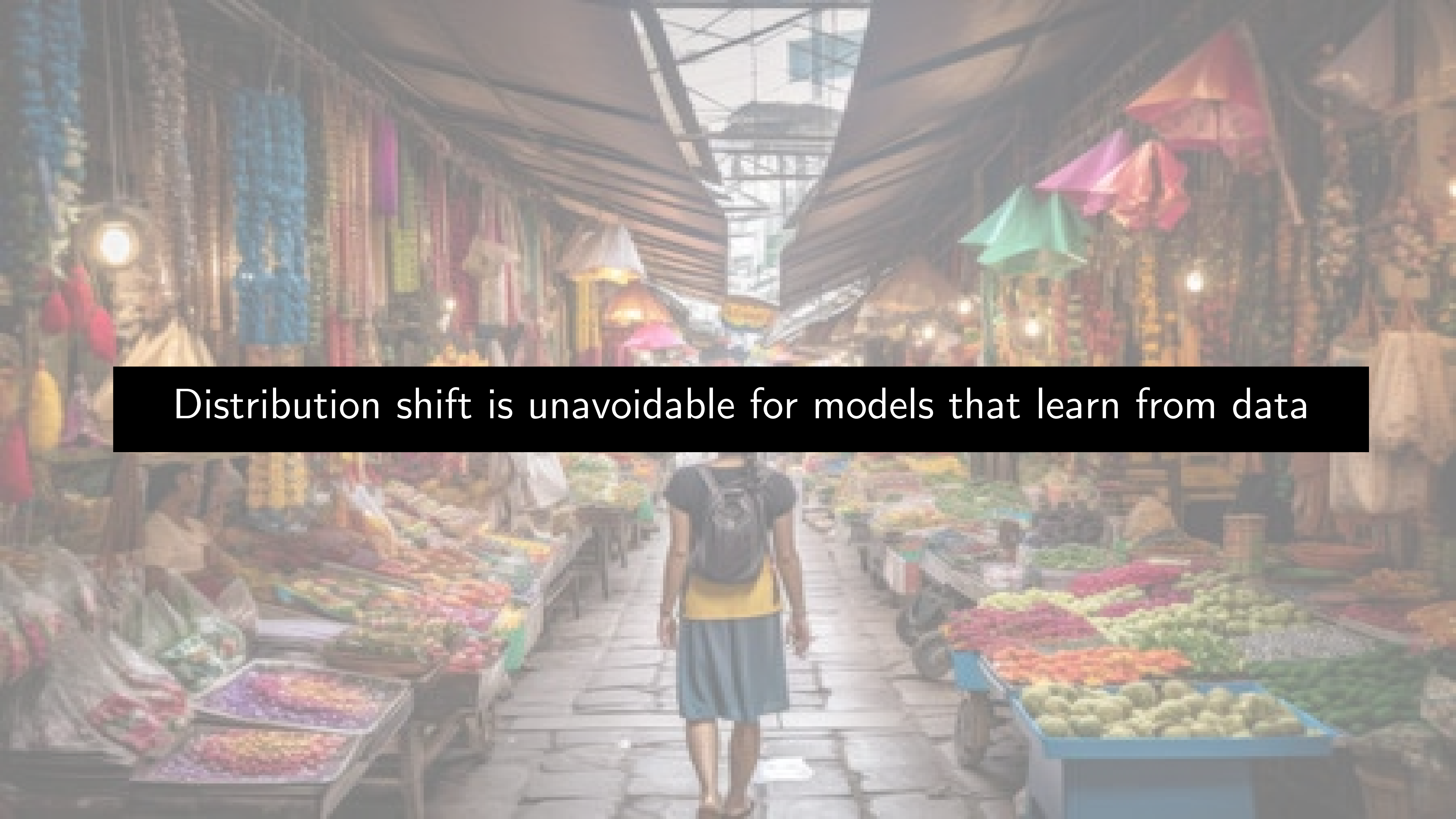
Simple English Wikipedia

arXiv

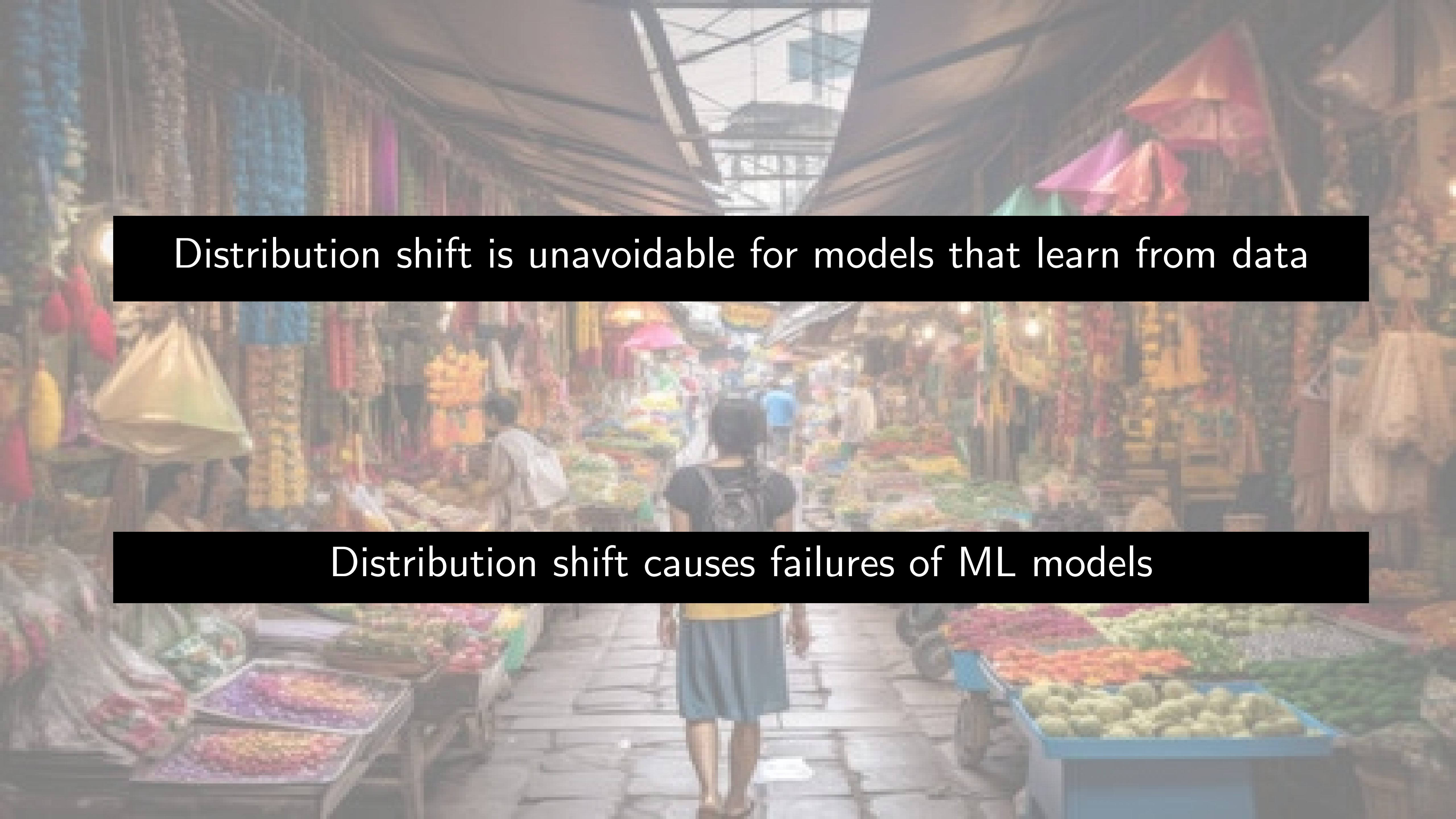
PubMed

differing sentence structure, vocabulary, word use

Figure Source: Chelsea Finn



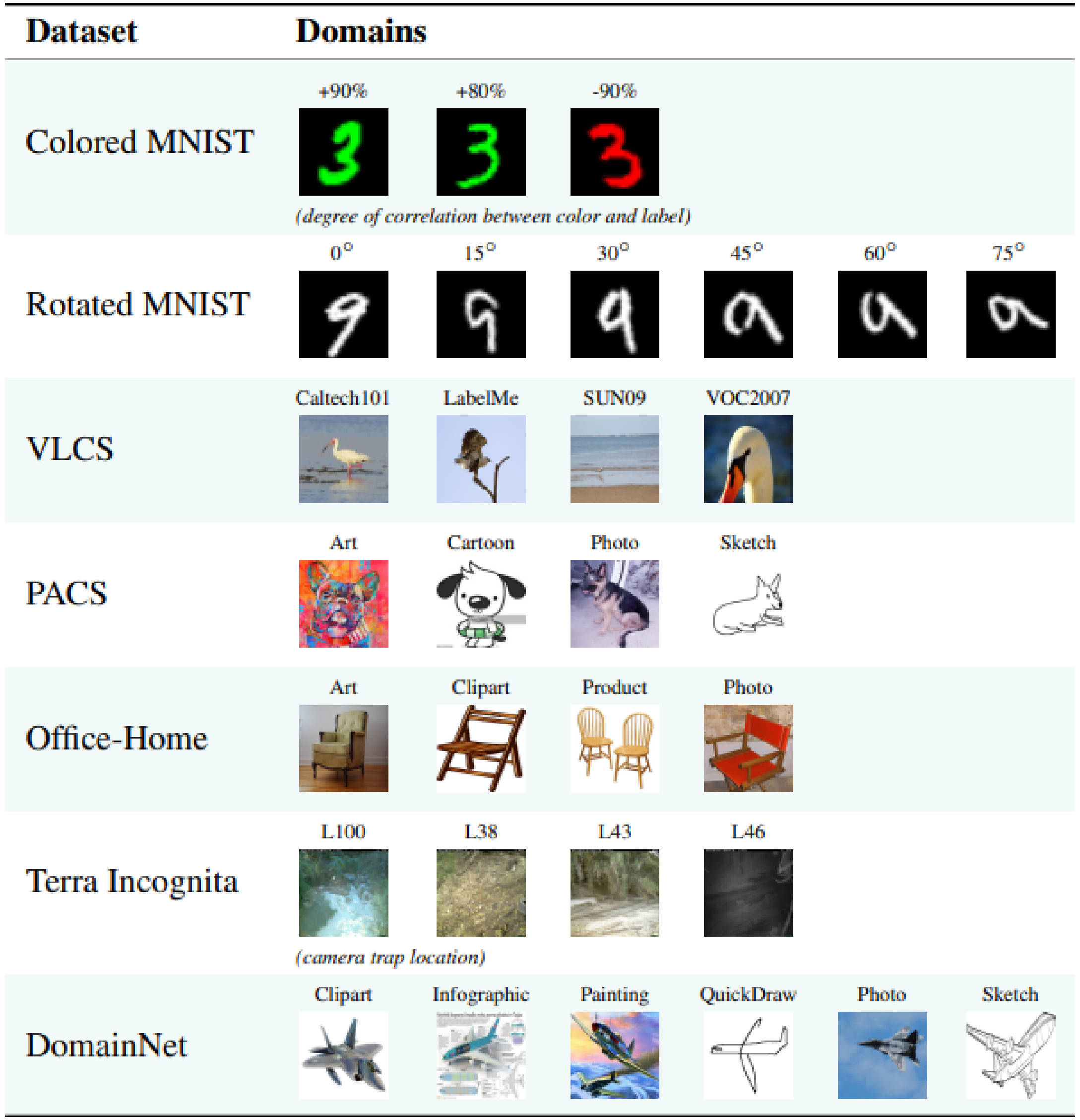
Distribution shift is unavoidable for models that learn from data



Distribution shift is unavoidable for models that learn from data

Distribution shift causes failures of ML models

Benchmarks / Challenge Datasets



Algorithm	CMNIST	RMNIST	VLCS	PACS	OfficeHome	TerraInc	DomainNet	Average
ERM	51.5 ± 0.1	98.0 ± 0.0	77.5 ± 0.4	85.5 ± 0.2	66.5 ± 0.3	46.1 ± 1.8	40.9 ± 0.1	66.6
IRM	52.0 ± 0.1	97.7 ± 0.1	78.5 ± 0.5	83.5 ± 0.8	64.3 ± 2.2	47.6 ± 0.8	33.9 ± 2.8	65.4
GroupDRO	52.1 ± 0.0	98.0 ± 0.0	76.7 ± 0.6	84.4 ± 0.8	66.0 ± 0.7	43.2 ± 1.1	33.3 ± 0.2	64.8
Mixup	52.1 ± 0.2	98.0 ± 0.1	77.4 ± 0.6	84.6 ± 0.6	68.1 ± 0.3	47.9 ± 0.8	39.2 ± 0.1	66.7
MLDG	51.5 ± 0.1	97.9 ± 0.0	77.2 ± 0.4	84.9 ± 1.0	66.8 ± 0.6	47.7 ± 0.9	41.2 ± 0.1	66.7
CORAL	51.5 ± 0.1	98.0 ± 0.1	78.8 ± 0.6	86.2 ± 0.3	68.7 ± 0.3	47.6 ± 1.0	41.5 ± 0.1	67.5
MMD	51.5 ± 0.2	97.9 ± 0.0	77.5 ± 0.9	84.6 ± 0.5	66.3 ± 0.1	42.2 ± 1.6	23.4 ± 9.5	63.3
DANN	51.5 ± 0.3	97.8 ± 0.1	78.6 ± 0.4	83.6 ± 0.4	65.9 ± 0.6	46.7 ± 0.5	38.3 ± 0.1	66.1
CDANN	51.7 ± 0.1	97.9 ± 0.1	77.5 ± 0.1	82.6 ± 0.9	65.8 ± 1.3	45.8 ± 1.6	38.3 ± 0.3	65.6
MTL	51.4 ± 0.1	97.9 ± 0.0	77.2 ± 0.4	84.6 ± 0.5	66.4 ± 0.5	45.6 ± 1.2	40.6 ± 0.1	66.2
SagNet	51.7 ± 0.0	98.0 ± 0.0	77.8 ± 0.5	86.3 ± 0.2	68.1 ± 0.1	48.6 ± 1.0	40.3 ± 0.1	67.2
ARM	56.2 ± 0.2	98.2 ± 0.1	77.6 ± 0.3	85.1 ± 0.4	64.8 ± 0.3	45.5 ± 0.3	35.5 ± 0.2	66.1
VREx	51.8 ± 0.1	97.9 ± 0.1	78.3 ± 0.2	84.9 ± 0.6	66.4 ± 0.6	46.4 ± 0.6	33.6 ± 2.9	65.6
RSC	51.7 ± 0.2	97.6 ± 0.1	77.1 ± 0.5	85.2 ± 0.9	65.5 ± 0.9	46.6 ± 1.0	38.9 ± 0.5	66.1

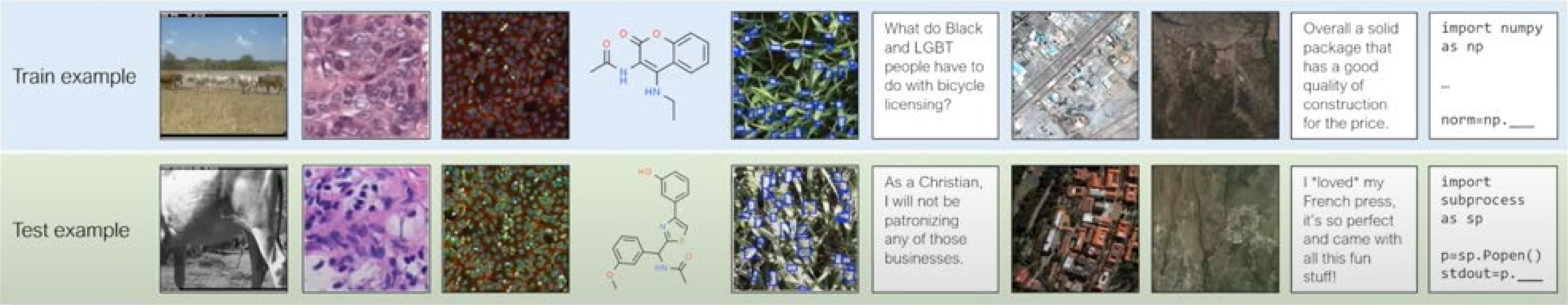
Model selection: training-domain validation set

DomainBed (Gulrajani and Lopez Paz ICLR 2021)

Benchmarks / Challenge Datasets

WILDS

	Domain shift					Subpopulation shift	Domain shift + subpopulation shift			
Dataset	iWildCam	Camelyon17	RxRx1	OGB-MolPCBA	GlobalWheat	CivilComments	FMoW	PovertyMap	Amazon	Py150
Input (x)	photo	tissue slide	cell image	molecular graph	wheat image	online comment	satellite image	satellite image	product review	code
Prediction (y)	animal species	tumor	perturbed gene	bioassays	wheat head bbox	toxicity	land use	asset wealth	sentiment	autocomplete
Domain (d)	camera	hospital	batch	scaffold	location, time	demographic	time, region	location	user	git repository



Adapted from	Beery et al. 2020	Bandi et al. 2018	Taylor et al. 2019	Hu et al. 2020	David et al. 2021	Borkan et al. 2019	Christie et al. 2018	Yeh et al. 2020	Ni et al. 2019	Raychev et al. 2016

"WILDS" (Koh et al. 2021)

Solution?

Domain Adaptation

Domain Adaptation

- Problem Setup (Handwritten Notes)

- Theory
    - A theory of learning from different domains (Ben-David et al. MLJ 2010) https://link.springer.com/content/pdf/10.1007/s10994-009-5152-4.pdf
    - Learning from multiple sources (Crammer et al. JMLR 2008) https://www.jmlr.org/papers/volume9/crammer08a/crammer08a.pdf

Domain Adaptation Scenarios

(adapted from Mathieu Salzmann)

Standard Visual Recognition

Training data

Test data

Train a classifier on the training data and directly apply it to the test data

Domain Shift

Training data

Source domain

Test data

Target domain

A classifier trained on one domain may perform poorly on another domain

Semi-supervised vs Unsupervised

- Semi-supervised: Some labeled target data, but not enough to train from scratch

Source data

Fully-labeled

Target data

A few labels

Semi-supervised vs Unsupervised

- Unsupervised: No labels for the target data

Source data

Fully-labeled

Target data

Single vs Multiple Source Domains

Source domain 1

Source domain 2

Target domain

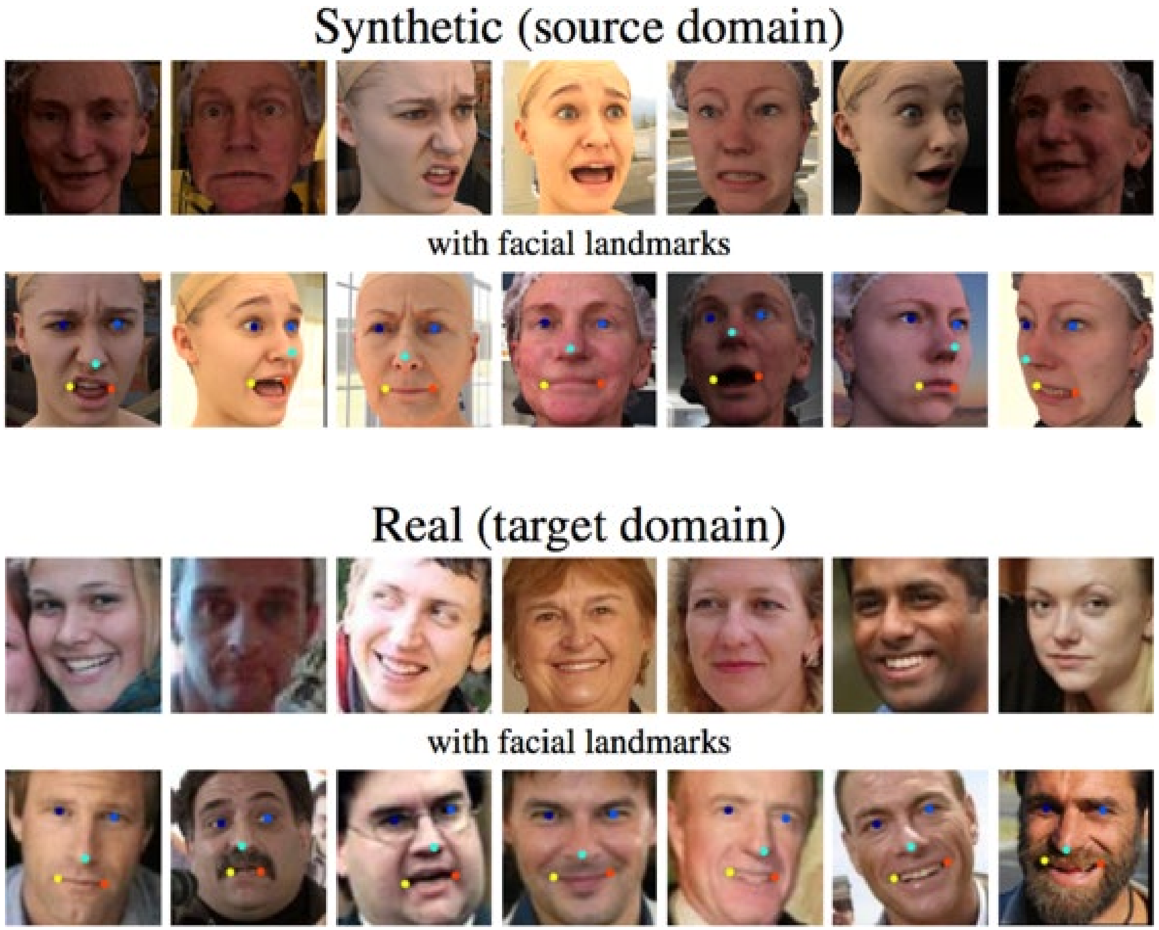
- Moving towards domain generalization

Domain Adaptation: Other Scenarios

Synthetic (source domain)

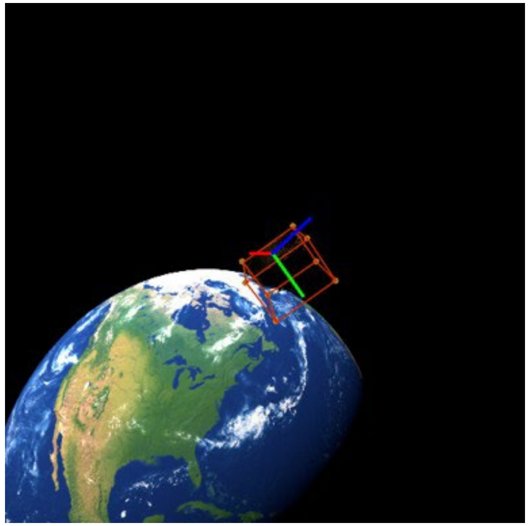
Real (target domain)

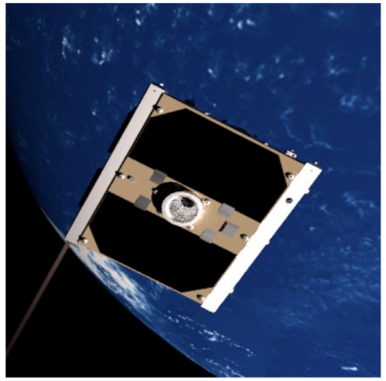
Domain Adaptation: Other Scenarios



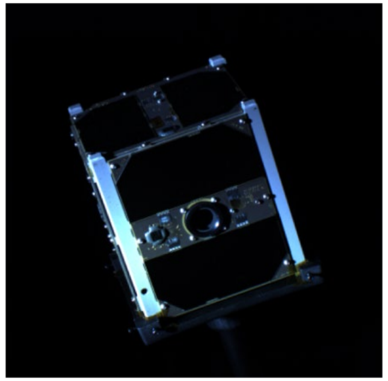
Domain Adaptation: Other Scenarios

Satellite 6D pose estimation





Synthetic (source)



Real (target)

Setup

- Each sample is represented by a feature vector:
  - In the traditional methods, e.g., bag of SURF features
  - More recently, features extracted by a deep backbone network

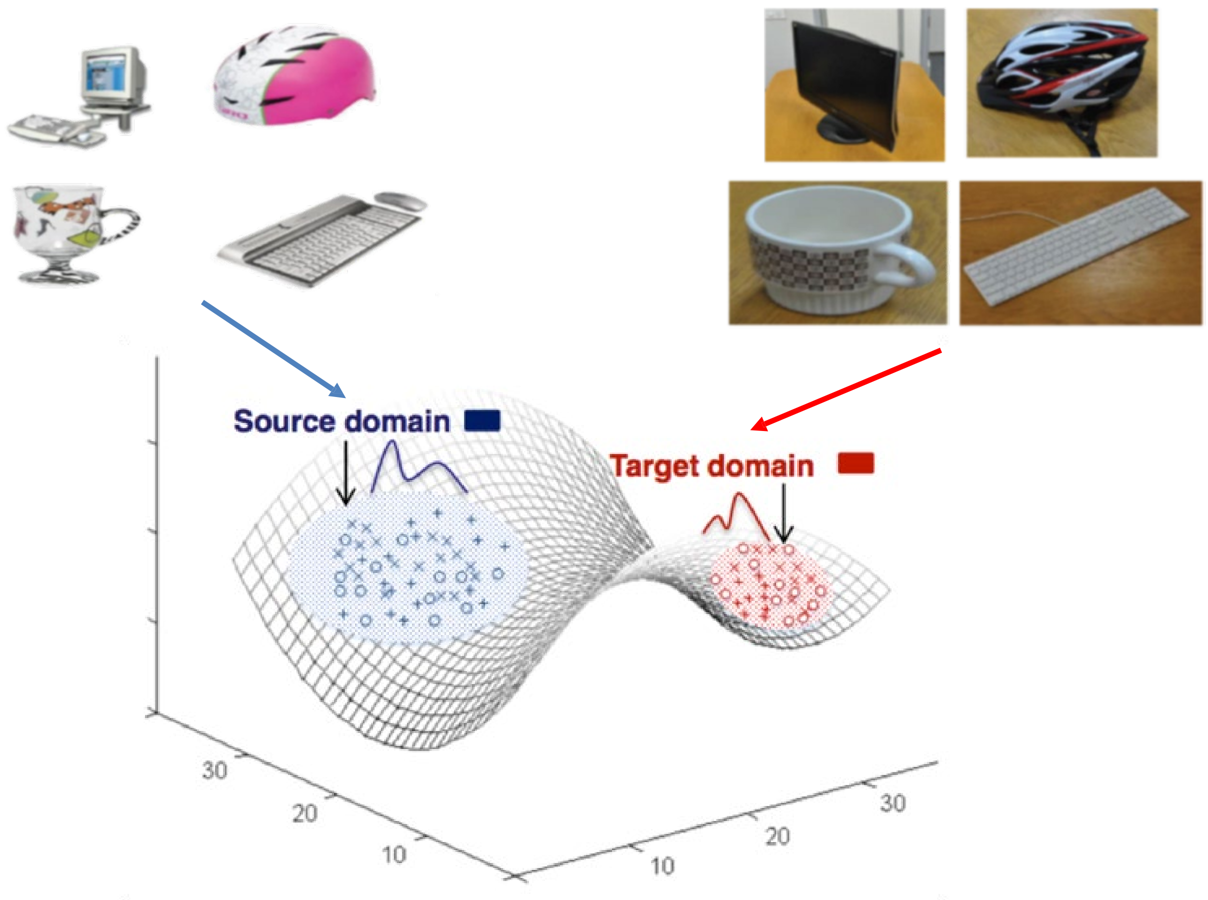
$$\mathbf{X}_s = \{\mathbf{x}_s^i\}_{i=1}^{n}$$

$$\mathbf{X}_t = \{\mathbf{x}_t^j\}_{j=1}^{m}$$

Label: $\{y_s^i\}_{i=1}^{n}$

Domain Shift

- The domain shift is defined as a difference in the distribution of the source and target samples



Domain Shift

- Typically, the literature focuses on the covariate shift case, where

$$p_t(x_t) \neq p_s(x_s)$$

- But

$$p_t(y|x_t) = p_s(y|x_s)$$

- The goal of domain adaptation is then often expressed as that of finding a transformation $T(.)$, such that

$$p_t(T(x_t)) = p_s(T(x_s))$$