Towards Visual Recognition *in the Wild*: *Long-Tailed* Sources & Open *Compound* Targets

Boqing Gong
Google

Learning To Detect Unseen Object Classes by Between-Class Attribute Transfer

Christoph H. Lampert Hannes Nickisch Stefan Harmeling
Max Planck Institute for Biological Cybernetics, Tübingen, Germany
{firstname.lastname}@tuebingen.mpg.de

Abstract

We study the problem of object classification when training and test classes are disjoint, i.e. *no training examples of the target classes are available. This setup has hardly been studied in computer vision research, but it is the rule rather than the exception, because the world contains tens of thousands of different object classes and for only a very few of them image, collections have been formed and annotated with suitable class labels.*

In this paper, we tackle the problem by introducing attribute-based classification. *It performs object detection*

otter

black:	yes
white:	no
brown:	yes
stripes:	no
water:	yes
eats fish:	yes

polar bear

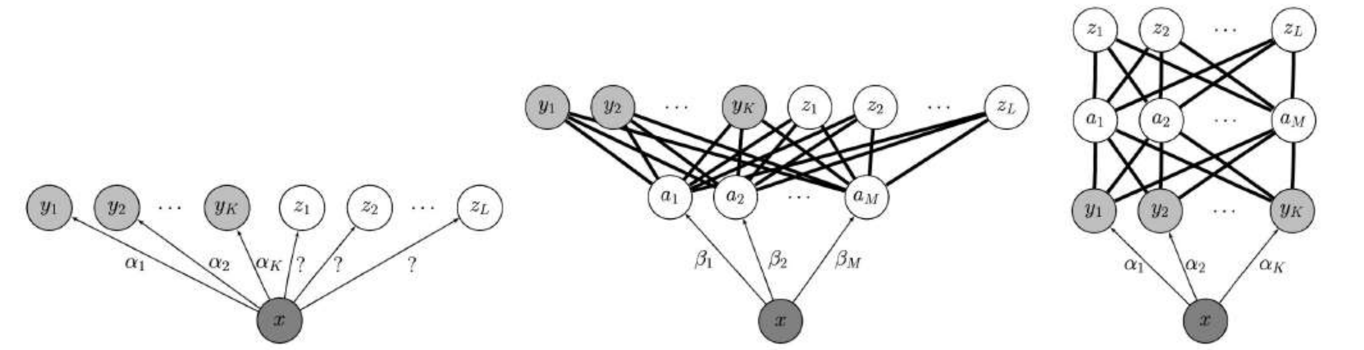
black:	no
white:	yes
brown:	no
stripes:	no
water:	yes
eats fish:	yes

zebra

CVPR 2009

50 classes

85 attributes



(a) Flat multi-class classification (b) Direct attribute prediction (DAP) (c) Indirect attribute prediction (IAP)

Abstract form: *unsupervised* domain adaptation (DA)

Setup

Source domain (with labeled data)

$$D_S = \{(x_m, y_m)\}_{m=1}^{M} \sim \boxed{P_S(X, Y)}$$

Target domain (no labels for training)

$$D_T = \{(x_n, \textbf{?})\}_{n=1}^{N} \sim \boxed{P_T(X, Y)}$$

Different distributions

Objective

Learn models to work well on target

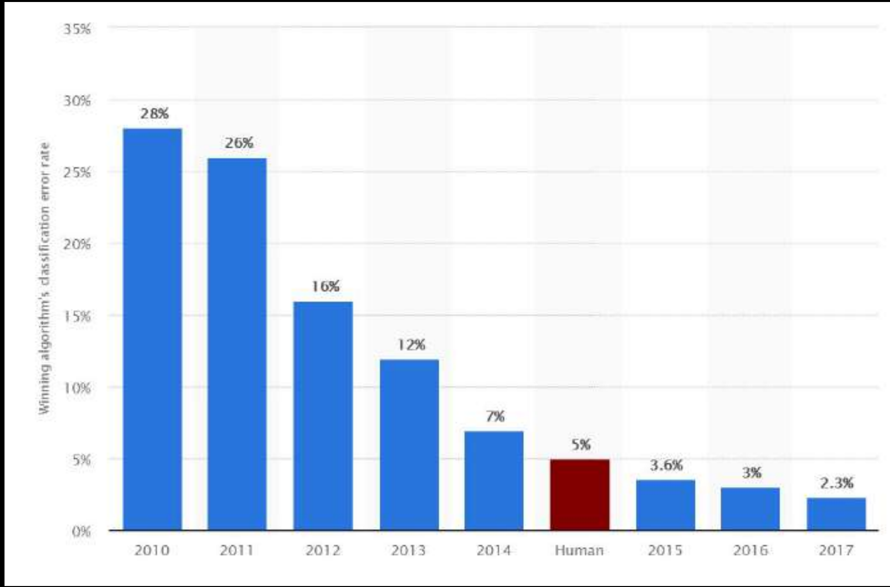
Kernel Methods for Unsupervised Domain Adaptation

10~100 classes

2011-2015

ILSVRC
2010-2017

~1000 classes



Bottom image credit:
http://www.thegreenmedium.com/blog/2019/5/24/why-robots-will-help-you-rather-than-try-to-take-over-the-world-a-brief-history-of-ai

DeCAF: A Deep Convolutional Activation Feature for Generic Visual Recognition

Jeff Donahue*, Yangqing Jia*, Oriol Vinyals, Judy Hoffman, Ning Zhang, Eric Tzeng, Trevor Darrell
{JDONAHUE,JIAYQ,VINYALS,JHOFFMAN,NZHANG,ETZENG,TREVOR}@EECS.BERKELEY.EDU
UC Berkeley & ICSI, Berkeley, CA, USA

Abstract

We evaluate whether features extracted from the activation of a deep convolutional network trained in a fully supervised fashion on a large, fixed set of object recognition tasks can be re-purposed to novel generic tasks. Our generic tasks may differ significantly from the originally trained tasks and there may be insufficient labeled or unlabeled data to conventionally train or

pects of a given domain through discovery of salient clusters, parts, mid-level features, and/or hidden units (Hinton & Salakhutdinov, 2006; Fidler & Leonardis, 2007; Zhu et al., 2007; Singh et al., 2012; Krizhevsky et al., 2012). Such models have been able to perform better than traditional hand-engineered representations in many domains, especially those where good features have not already been engineered (Le et al., 2011). Recent results have shown that moderately deep unsupervised models outperform the

ICML 2014

Deep features!

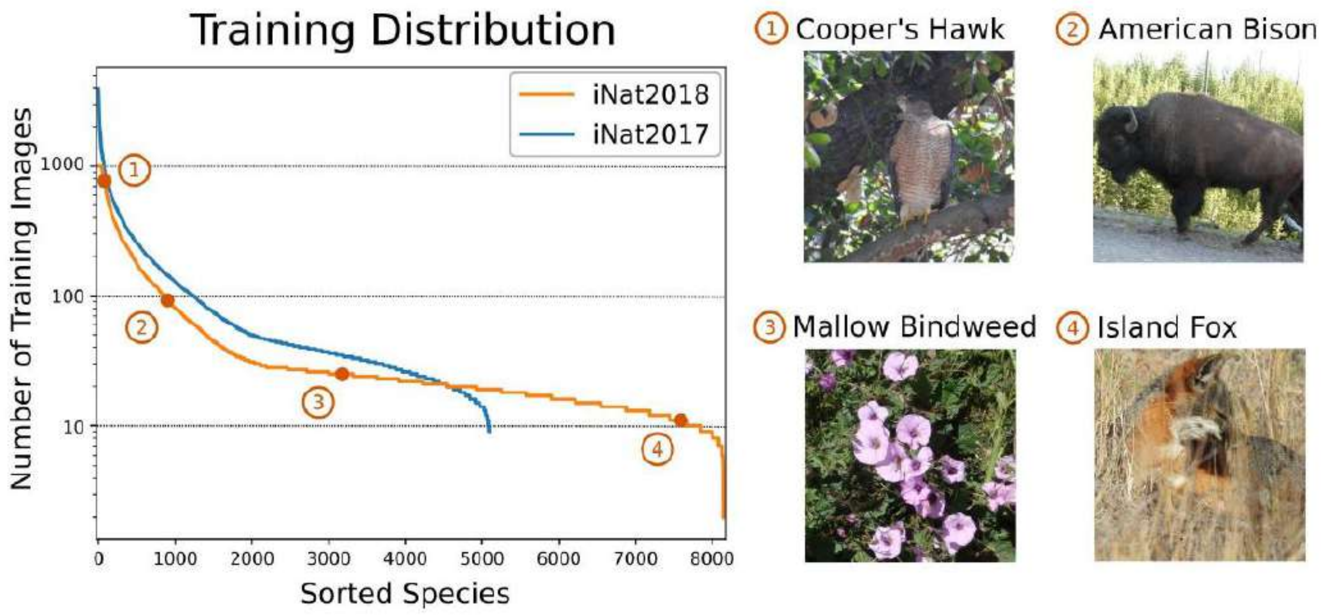
COCO
Common Objects in Context

places
THE SCENE RECOGNITION DATABASE

ACTIVITYNET

Kinetics

Training Image Distribution



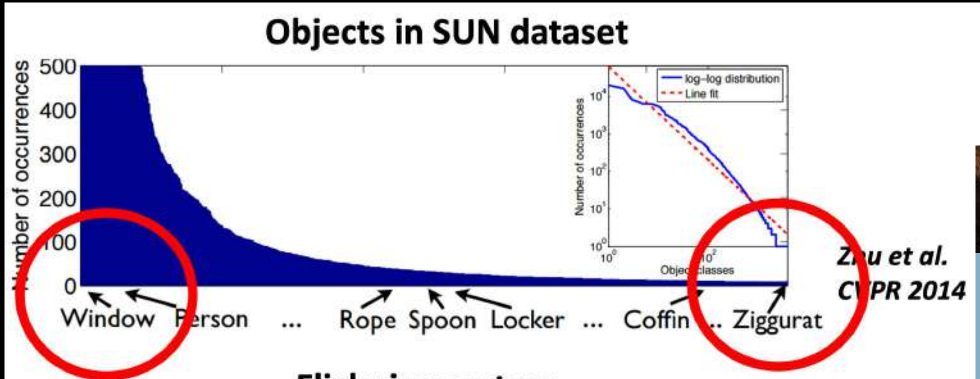
The iNaturalist Species Classification and Detection Dataset

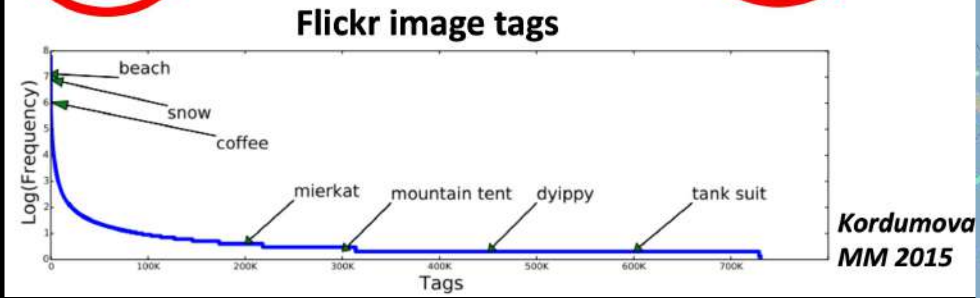
Grant Van Horn[1] Oisin Mac Aodha[1] Yang Song[2] Yin Cui[3] Chen Sun[2]
Alex Shepard[4] Hartwig Adam[2] Pietro Perona[1] Serge Belongie[3]

Object recognition *in the wild*

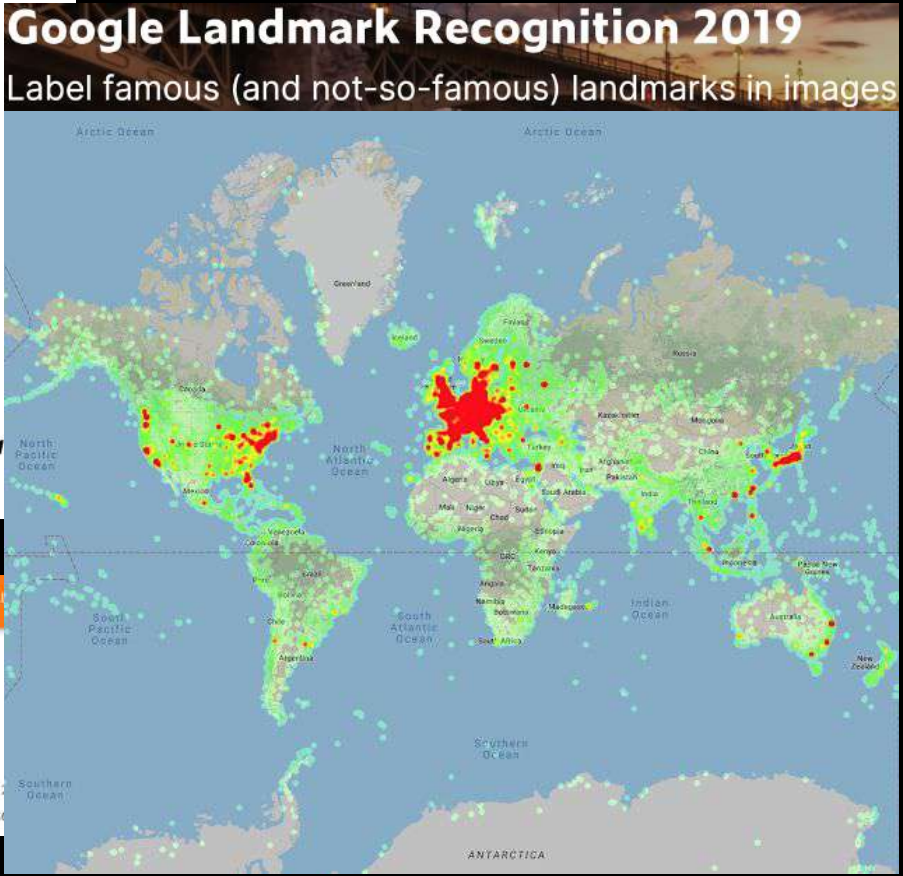
5k~8k classes

in the wild


Objects in SUN dataset
Number of occurrences
500
400
300
200
100
0
log-log distribution
Line fit
Number of occurrences
Object classes
Window
Person
...
Rope
Spoon
Locker
...
Coffin
...
Ziggurat
Zhu et al.
CVPR 2014


Flickr image tags
Log(Frequency)
beach
snow
coffee
mierkat
mountain tent
dyippy
tank suit
Tags
Kordumova
MM 2015


LVIS
1200+ Categories
Found by data-driven object discovery in 164k images.
Long Tail
Category discovery naturally reveals a large number of rare categories.


Google Landmark Recognition 2019
Label famous (and not-so-famous) landmarks in images

in the wild

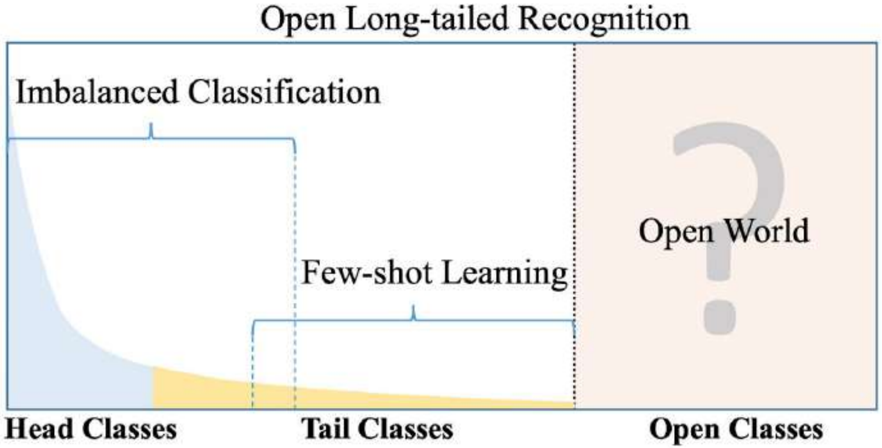
Right image credit: https://natureneedsmore.org/the-elephant-in-the-room/

Large-Scale Long-Tailed Recognition in an Open World

Ziwei Liu[1,2*] Zhongqi Miao[2*] Xiaohang Zhan[1] Jiayun Wang[2] Boqing Gong[2†] Stella X. Yu[2]
[1] The Chinese University of Hong Kong [2] UC Berkeley / ICSI
{zwliu,zx017}@ie.cuhk.edu.hk, {zhongqi.miao,peterwg,stellayu}@berkeley.edu, bgong@outlook.com

Abstract

Real world data often have a long-tailed and open-ended distribution. A practical recognition system must classify among majority and minority classes, generalize from a few known instances, and acknowledge novelty upon a never seen instance. We define Open Long-Tailed Recognition (OLTR) as learning from such naturally distributed data and optimizing the classification accuracy over a balanced test set which include head, tail, and open classes.



CVPR 2019 (oral), improving neural architectures

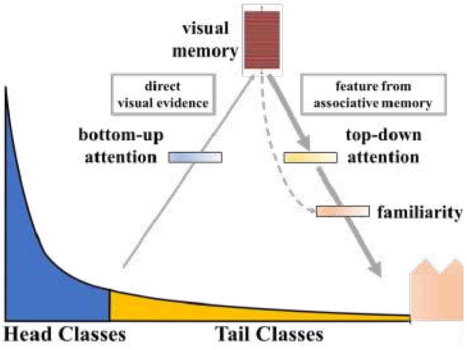
Large-Scale Long-Tailed Recognition in an Open World

Ziwei Liu[1,2*] Zhongqi Miao[2*] Xiaohang Zhan[1] Jiayun Wang[2] Boqing Gong[2†] Stella X. Yu[2]
[1] The Chinese University of Hong Kong [2] UC Berkeley / ICSI
{zwliu,zx017}@ie.cuhk.edu.hk, {zhongqi.miao,peterwg,stellayu}@berkeley.edu, bgong@outlook.com

Long-tailed ImageNet (1000 classes)

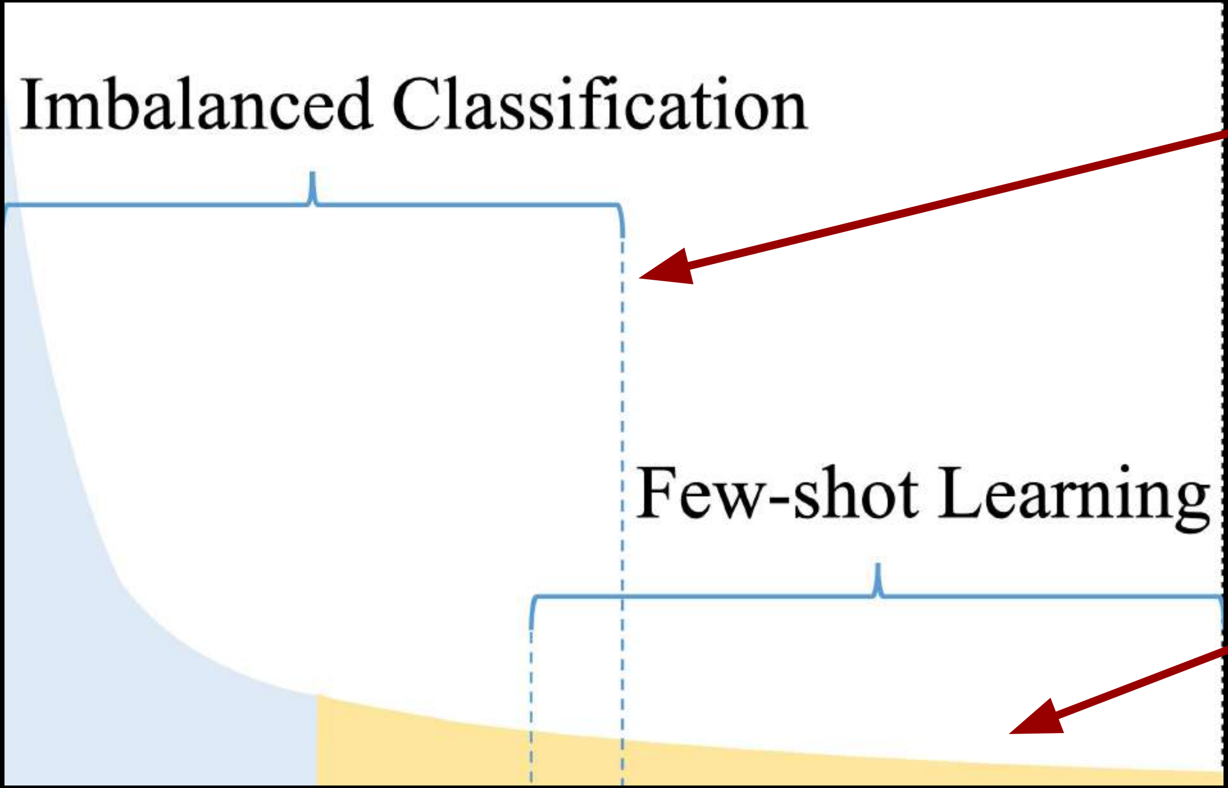
Long-tailed Places-365

Long-tailed MS1M ArcFace (74.5k ids)



A memory bank to enhance tail classes

CVPR 2019 (oral), improving neural architectures



An old AI problem

A new AI problem (meta-learning, transfer learning, zero-shot learning)

Acknowledgement: Matthew Brown @Google

Existing work

Class-wise weighting, over/under-sampling, etc.

[CVPR’18] Large Scale Fine-Grained Categorization and Domain-Specific Transfer Learning

[CVPR’19] Class-Balanced Loss Based on Effective Number of Samples

[NeurIPS’19] Learning Imbalanced Datasets with Label-Distribution-Aware Margin Loss

[ICLR’20] Decoupling Representation and Classifier for Long-Tailed Recognition

DECOUPLING REPRESENTATION AND CLASSIFIER FOR LONG-TAILED RECOGNITION

Bingyi Kang[1,2], Saining Xie[1], Marcus Rohrbach[1], Zhicheng Yan[1], Albert Gordo[1], Jiashi Feng[2], Yannis Kalantidis[1]

Abstract form: *unsupervised* domain adaptation (DA)

Setup

Source domain (with labeled data)

$$D_{\mathcal{S}} = \{(x_m, y_m)\}_{m=1}^{\mathsf{M}} \sim P_{\mathcal{S}}(X, Y)$$

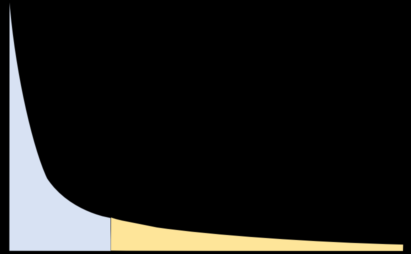
Target domain (no labels for training)

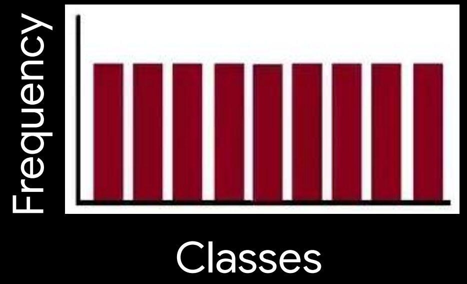
$$D_{\mathcal{T}} = \{(x_n, \textbf{?}\,)\}_{n=1}^{\mathsf{N}} \sim P_{\mathcal{T}}(X, Y)$$

Different distributions

Objective

Learn models to work well on target





Existing work

Class-wise weighting, over/under-sampling, etc.

[CVPR'18] Large Scale Fine-Grained Categorization and Domain-Specific Transfer Learning

[CVPR'19] Class-Balanced Loss Based on Effective Number of Samples

[NeurIPS'19] Learning Imbalanced Datasets with Label-Distribution-Aware Margin Loss

[ICLR'20] Decoupling Representation and Classifier for Long-Tailed Recognition

… as domain adaptation

Target

$$\text{error} = \mathbb{E}_{P_t(x,y)} L(f(x;\theta), y),$$

$$= \mathbb{E}_{P_s(x,y)} L(f(x;\theta), y) P_t(x,y) / P_s(x,y)$$

Source

$$= \mathbb{E}_{P_s(x,y)} L(f(x;\theta), y) \frac{P_t(y) P_t(x|y)}{P_s(y) P_s(x|y)}$$

$$= \mathbb{E}_{P_s(x,y)} L(f(x;\theta), y) w_y (1 + \tilde{\epsilon}_{x,y}),$$

Existing work assumes ϵ=0

Head vs. tail

Many training images in a **head** class: ϵ=0

$$\text{Training cats} \sim P_t(x|\text{cat})$$

Few-shot training images in a **tail** class: ϵ≠0

$$\text{Training tacs} \nsim P_t(x|\text{tac})$$

... as domain adaptation

$$= \mathbb{E}_{P_s(x,y)} L(f(x;\theta), y) \frac{P_t(y) P_t(x|y)}{P_s(y) P_s(x|y)}$$
$$:= \mathbb{E}_{P_s(x,y)} L(f(x;\theta), y) w_y (1 + \tilde{\epsilon}_{x,y}),$$

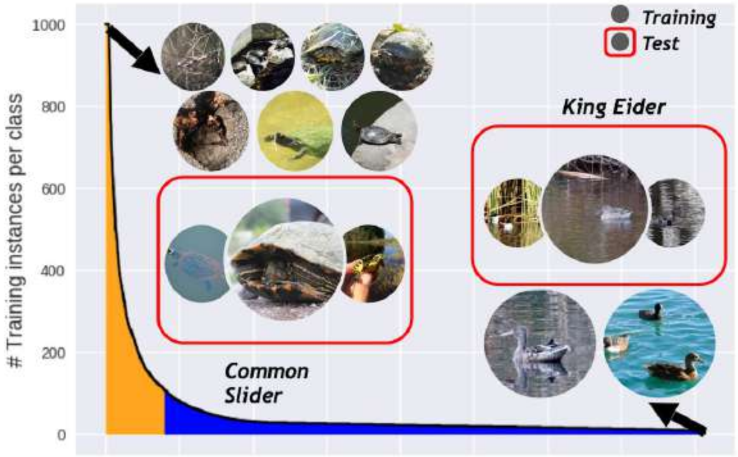
Existing work assumes ϵ=0

Rethinking Class-Balanced Methods for Long-Tailed Visual Recognition from a Domain Adaptation Perspective

Muhammad Abdullah Jamal[1*] Matthew Brown[3] Ming-Hsuan Yang[2,3] Liqiang Wang[1] Boqing Gong[3]
[1]University of Central Florida [2]University of California at Merced [3]Google

Abstract

Object frequency in the real world often follows a power law, leading to a mismatch between datasets with long-tailed class distributions seen by a machine learning model and our expectation of the model to perform well on all classes. We analyze this mismatch from a domain adaptation point of view. First of all, we connect existing class-balanced methods for long-tailed classification to target shift, a well-studied scenario in domain adaptation. The connection reveals that these methods implicitly assume



CVPR 2020 (oral), long-tailed recognition ≅ domain adaptation

Our approach

Estimating both w_y & $\tilde{\epsilon}_{x,y}$

by unifying [CVPR'19] & an improved meta-learning method

SOTA on six datasets

- CIFAR-LT-10
- CIFAR-LT-100
- ImageNet-LT
- Places-LT
- **iNaturalist 2017**
- **iNaturalist 2018**

Long-tailed visual recognition (LTVR)

Emerging challenge as the datasets grow in scale

Timely topic

Datasets: iNaturalist, LVIS, ImageNet, COCO, etc.

Tasks: almost all

… as domain adaptation

New perspective to LTVR

New powerhouse of methods

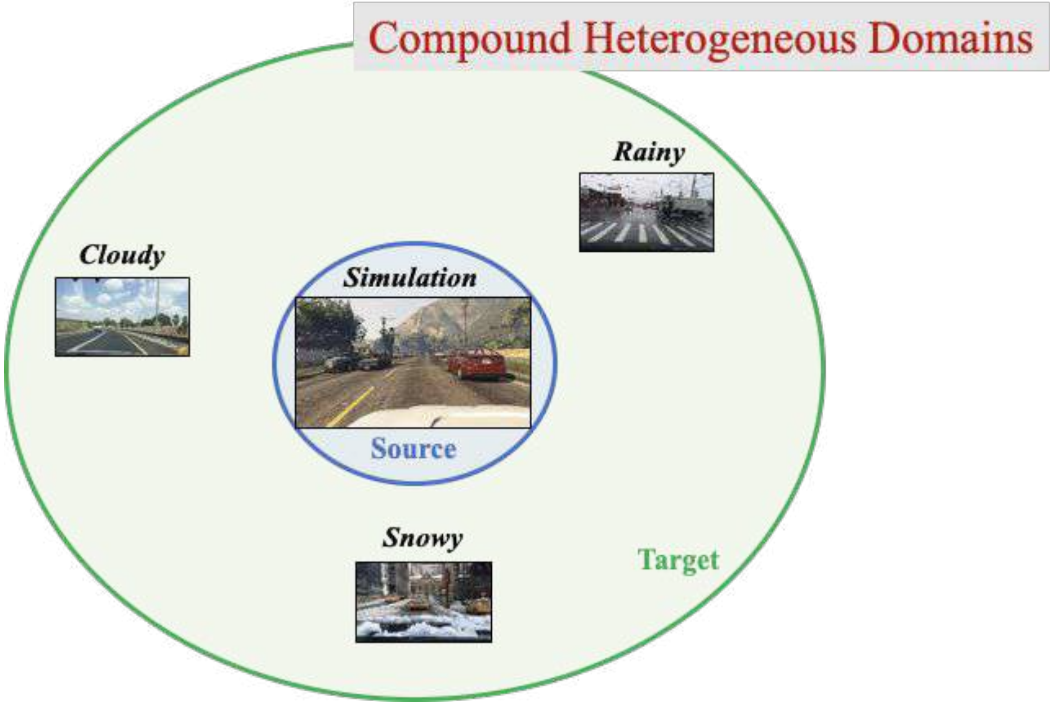
- Domain-invariant representation learning
- Curriculum domain adaptation
- Adversarial learning
- Classifier discrepancy
- Data augmentation & synthesis, etc.

Diff: no access to target data

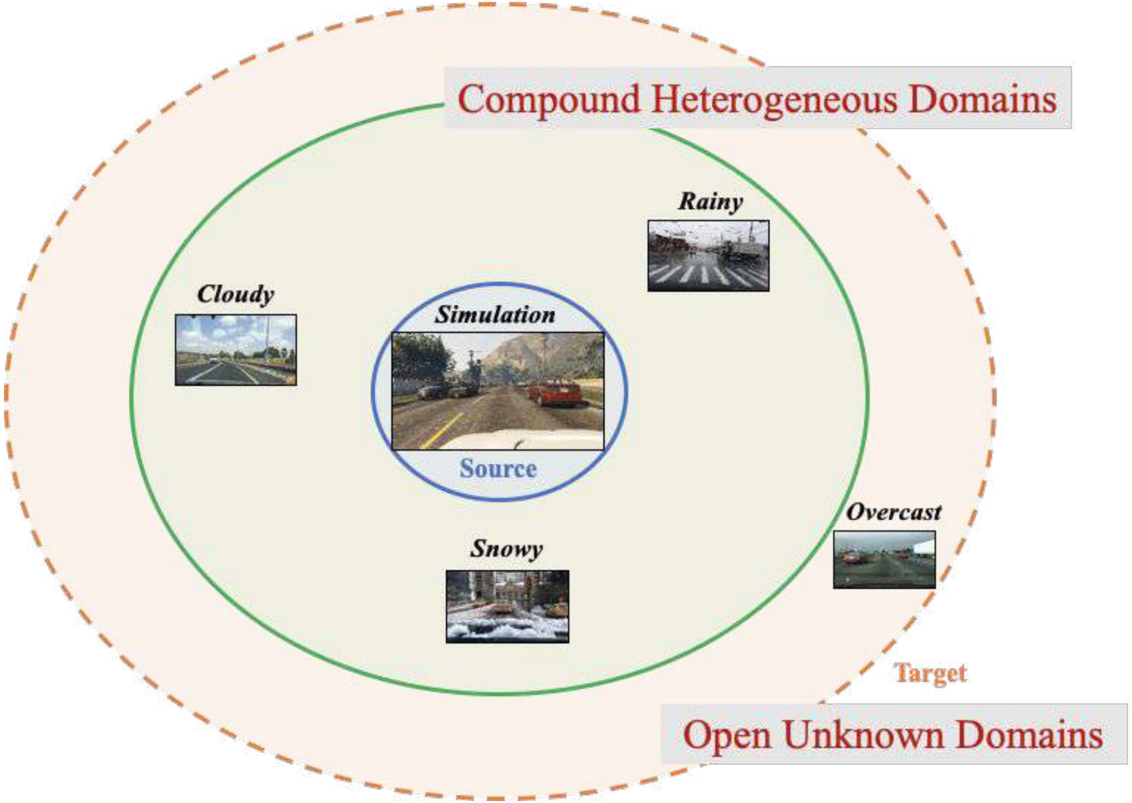
in the wild


Winnipeg
Vancouver
Victoria
Google
WASHINGTON
MONTANA
NORTH DAKOTA
MINNESOTA
Minneapolis
WISCONSIN
MICHIGAN
Toronto
Quebec
Montreal
Ottawa
VERMONT
NEW HAMPSHI
NEW YORK
MASSACHUS
CT
Portland
OREGON
IDAHO
WYOMING
SOUTH DAKOTA
IOWA
NEBRASKA
42 h
2,867 miles
PENNSYLVANIA
Google
ILLINOIS
INDIANA
OHIO
NJ
MARYLAND
DE
NEVADA
UTAH
United States
Kansas City
COLORADO
KANSAS
MISSOURI
44 h
2,959 miles
WEST VIRGINIA
VIRGINIA
KENTUCKY
Sacramento
San Francisco
San Jose
CALIFORNIA
Las Vegas
Nashville
TENNESSEE
NORTH CAROLINA
OKLAHOMA
ARKANSAS
Charlotte
Los Angeles
ARIZONA
NEW MEXICO
MISSISSIPPI
Atlanta
SOUTH CAROLINA
San Diego
Phoenix
Dallas
ALABAMA
GEORGIA
Tucson
El Paso
TEXAS
BAJA CALIFORNIA
Austin
Houston
LOUISIANA
Jacksonville
SONORA
CHIHUAHUA
New Orleans
San Antonio
Orlando
COAHUILA
Tampa
FLORIDA
NUEVO LEON
Miami
Monterrey

Open compound test cases (target)



Open compound test cases (target)



Open compound domain adaptation

Training:

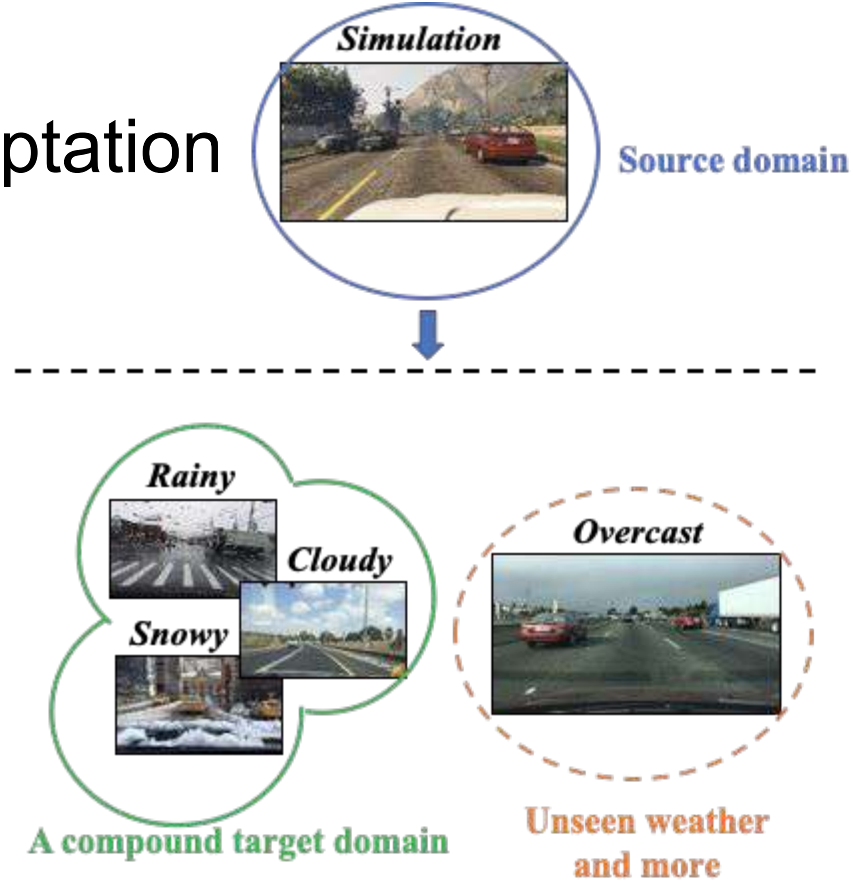
Labeled source domain data

Unlabeled data of the compound target

Testing:

in the compound target domain and

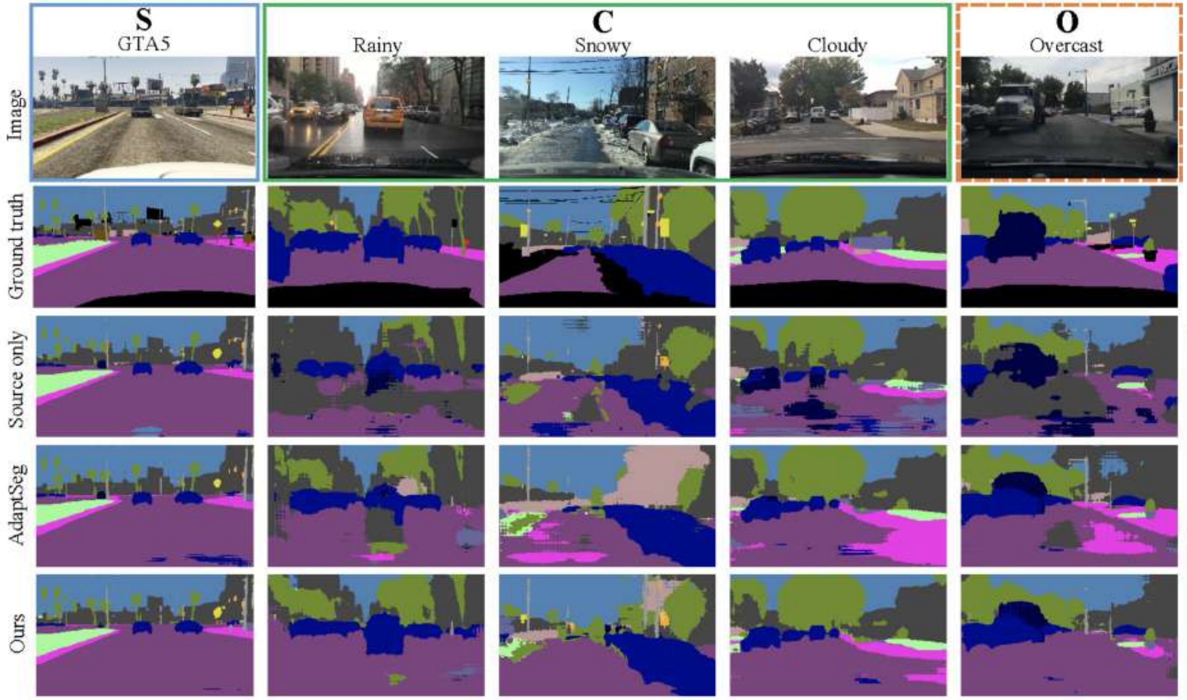
in previously unseen domains

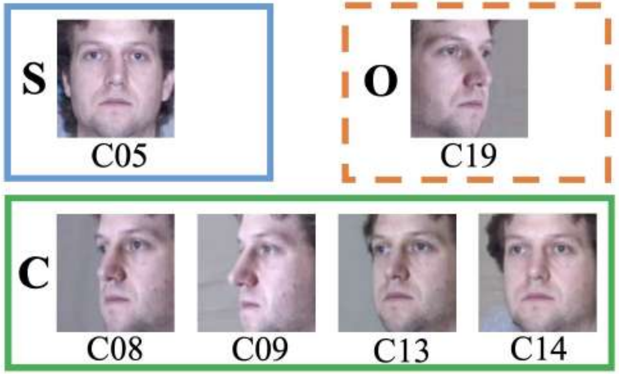


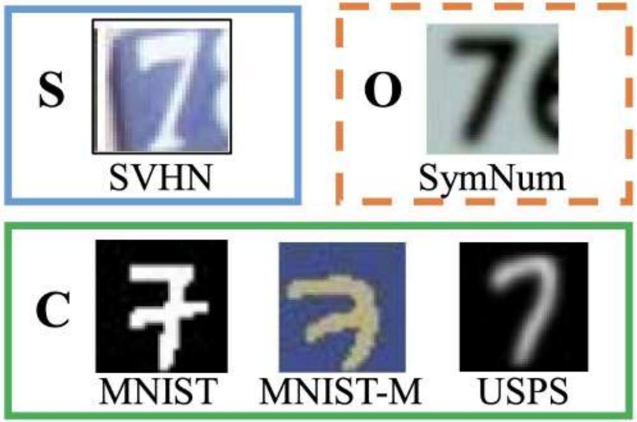
Open Compound Domain Adaptation

Liu, Ziwei, Zhongqi Miao, Xingang Pan, Xiaohang Zhan, Stella X. Yu, Dahua Lin, and Boqing Gong. "Compound domain adaptation in an open world." *CVPR 2020.* (oral)

Experiments

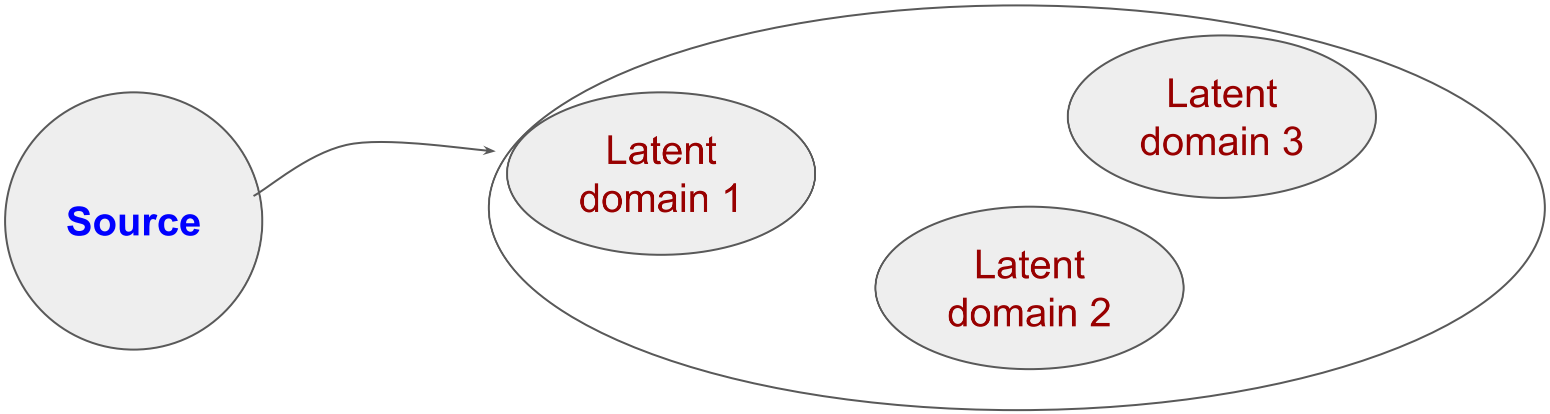






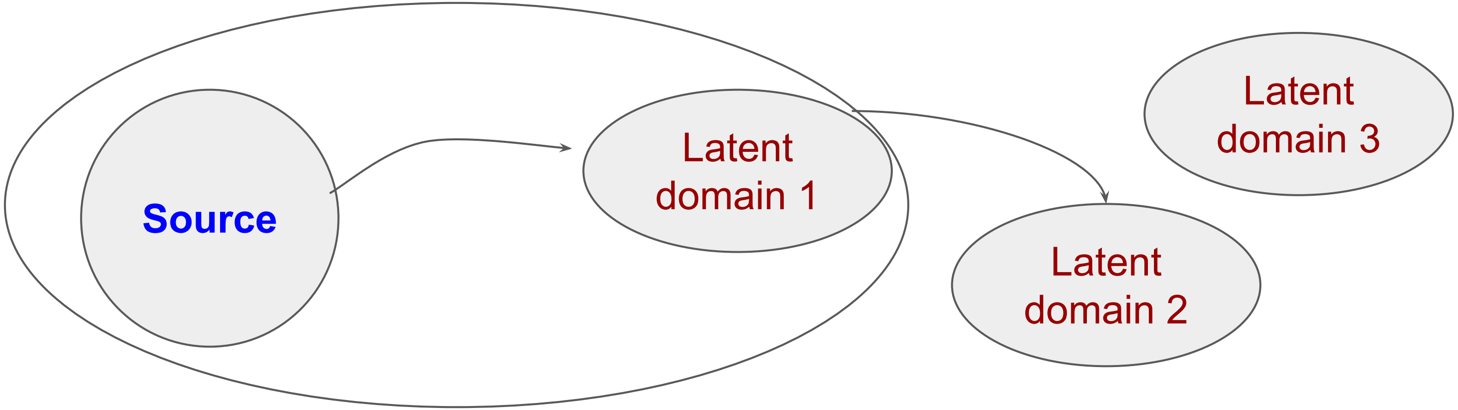
Our approach to break the compound target domain

into a series of bi-domain adaptation problems by “domain distances” between the source and latent domains in the target (curriculum training)



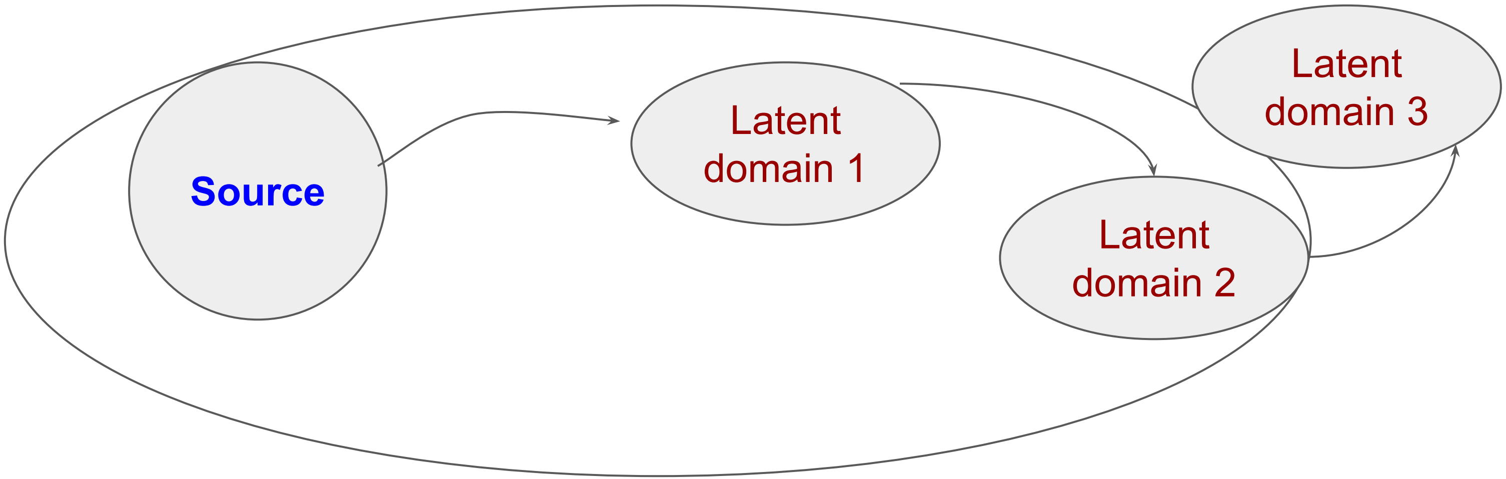
Our approach to break the compound target domain

into a series of bi-domain adaptation problems by “domain distances” between the source and latent domains in the target (curriculum training)



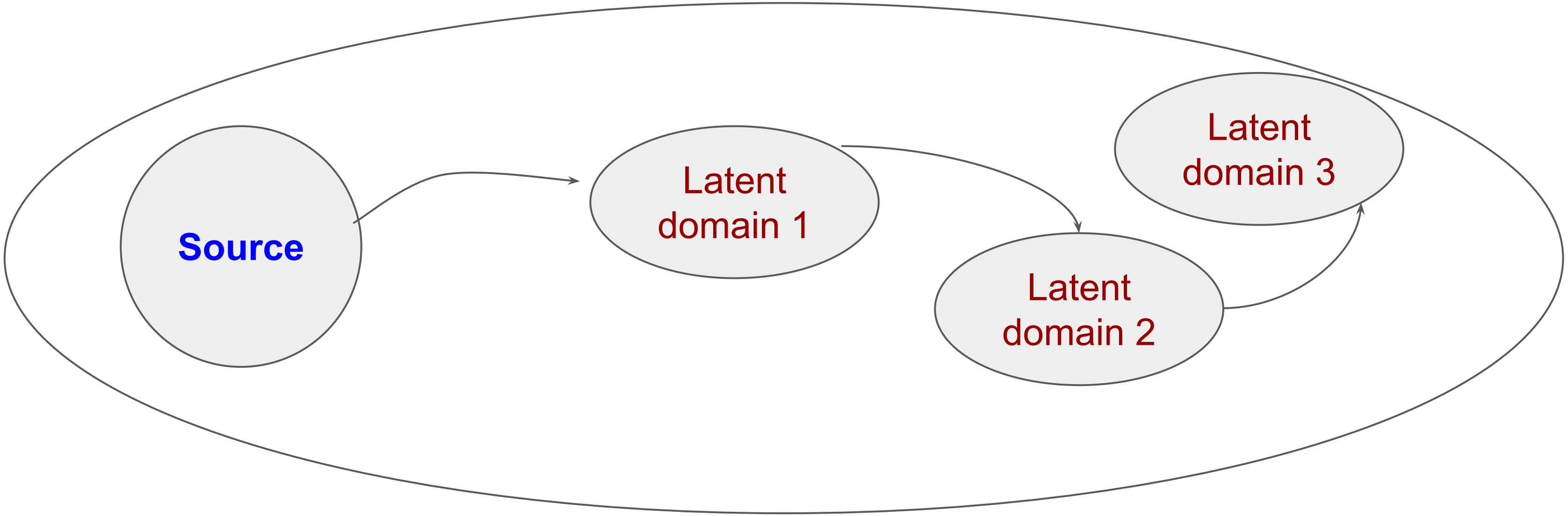
Our approach to break the compound target domain

into a series of bi-domain adaptation problems by “domain distances” between the source and latent domains in the target (curriculum training)



Our approach to break the compound target domain

into a series of bi-domain adaptation problems by “domain distances” between the source and latent domains in the target (curriculum training)



Pushing the boundary of visual recognition

Long-tailed source domains

- The elephant in the room as we scale up classes / study the wild data
- Memory bank to enhance tail classes (CVPR’19, oral)
- Domain adaptation: a new powerhouse of techniques (CVPR’20, oral)
- *Improved meta-learning for long-tailed recognition (undergoing)*

Open compound target domains (CVPR’20, oral)

- *Learning from unlabeled, noisy data in the wild (undergoing)*