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...
Abstract form: *unsupervised domain adaptation* (DA)

**Setup**

- **Source** domain (with labeled data)
  
  \[ D_S = \{(x_m, y_m)\}_{m=1}^{M} \sim P_S(X, Y) \]

- **Target** domain (no labels for training)
  
  \[ D_T = \{(x_n, ?)\}_{n=1}^{N} \sim P_T(X, Y) \]

**Objective**

Learn models to work well on **target**

Different distributions

Kernel Methods for Unsupervised Domain Adaptation

10~100 classes
ILSVRC 2010-2017

~1000 classes

Bottom image credit:
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}@ECE.S.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 repurposed 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 retrain a model from scratch. In this work, we seek to extend the utility of deep features across domains by mining the representations learned by a deep network for their discriminative and generic aspects 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 state-of-the-art on many tasks. As features increase in representational power, it becomes more difficult to train models by hand-engineering features in a separate manner.
Object recognition in the wild
5k~8k classes

The iNaturalist Species Classification and Detection Dataset

Grant Van Horn\textsuperscript{1}  Oisin Mac Aodha\textsuperscript{1}  Yang Song\textsuperscript{2}  Yin Cui\textsuperscript{3}  Chen Sun\textsuperscript{2}

Alex Shepard\textsuperscript{4}  Hartwig Adam\textsuperscript{2}  Pietro Perona\textsuperscript{1}  Serge Belongie\textsuperscript{3}
in the wild

Objects in SUN dataset

Flickr image tags

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

LVIS
1200+ Categories
Long Tail

Found by data-driven object discovery in 164k images.
Category discovery naturally reveals a large number of rare categories.
in the wild

Right image credit: https://natureneedsmore.org/the-elephant-in-the-room/
Large-Scale Long-Tailed Recognition in an Open World

Ziwei Liu1,2* Zhongqi Miao2* Xiaohang Zhan1 Jiayun Wang2 Boqing Gong2† Stella X. Yu2
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 Liu¹,²* Zhongqi Miao²* Xiaohang Zhan¹ Jiayun Wang² Boqing Gong²† Stella X. Yu²
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)

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
Abstract form: *unsupervised domain adaptation (DA)*

Setup

**Source** domain (with labeled data)

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

**Target** domain (no labels for training)

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

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

\[
\text{error} = \mathbb{E}_{P_t(x,y)} L(f(x; \theta), y),
\]
\[
= \mathbb{E}_{P_s(x,y)} L(f(x; \theta), y) \frac{P_t(x, y)}{P_s(x, y)}
\]
\[
= \mathbb{E}_{P_s(x,y)} L(f(x; \theta), y) \frac{P_t(x) P_t(x|y)}{P_s(y) P_s(x|y)}
\]
\[
= \mathbb{E}_{P_s(x,y)} L(f(x; \theta), y) w_y (1 + \bar{\varepsilon}_{x,y}),
\]

Existing work assumes $\varepsilon=0$
Head vs. tail

Many training images in a head class: $\epsilon=0$

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

Few-shot training images in a tail class: $\epsilon \neq 0$

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

... 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 $\epsilon=0$
Rethinking Class-Balanced Methods for Long-Tailed Visual Recognition from a Domain Adaptation Perspective

Muhammad Abdullah Jamal\textsuperscript{1}\textsuperscript{*} Matthew Brown\textsuperscript{3} Ming-Hsuan Yang\textsuperscript{2,3} Liqiang Wang\textsuperscript{1} Boqing Gong\textsuperscript{3}
\textsuperscript{1}University of Central Florida \textsuperscript{2}University of California at Merced \textsuperscript{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 $\equiv$ 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
Open compound test cases (**target**)

**Compound Heterogeneous Domains**

- Cloudy
- Rainy
- Simulation
- Source
- Snowy
- 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

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)*