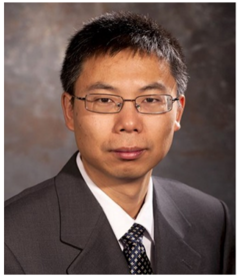
Rethinking Class-Balanced Methods for Long-tailed Visual Recognition from a Domain Adaptation Perspective




M. Abdullah Jamal   *Matthew Brown*   *Ming-Hsuan Yang*   *Liqiang Wang*   *Boqing Gong*

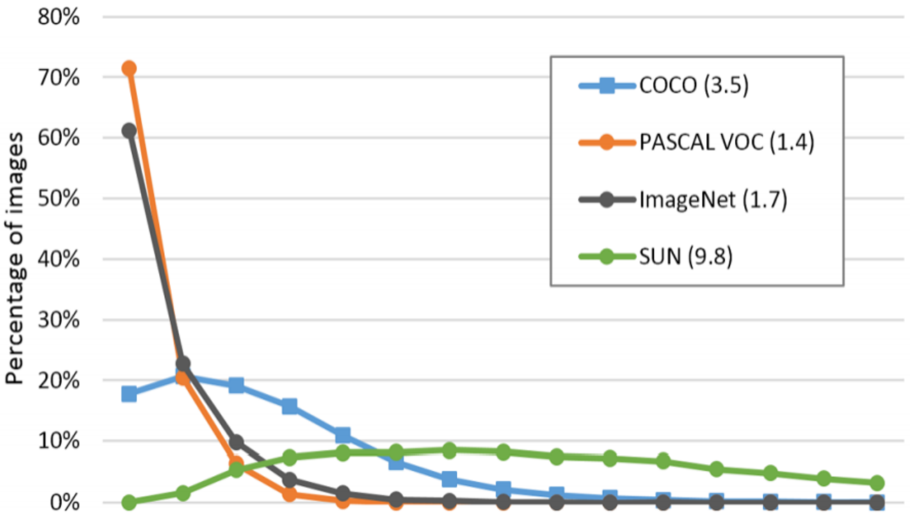
UCF   Google

Long-tailed Problem

Emerging challenge as the datasets grow in scale

Prevalent in fine-grained recognition, detection, etc.

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



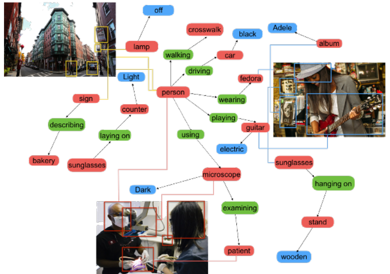
IMAGENET

Microsoft COCO
Common Objects in Context

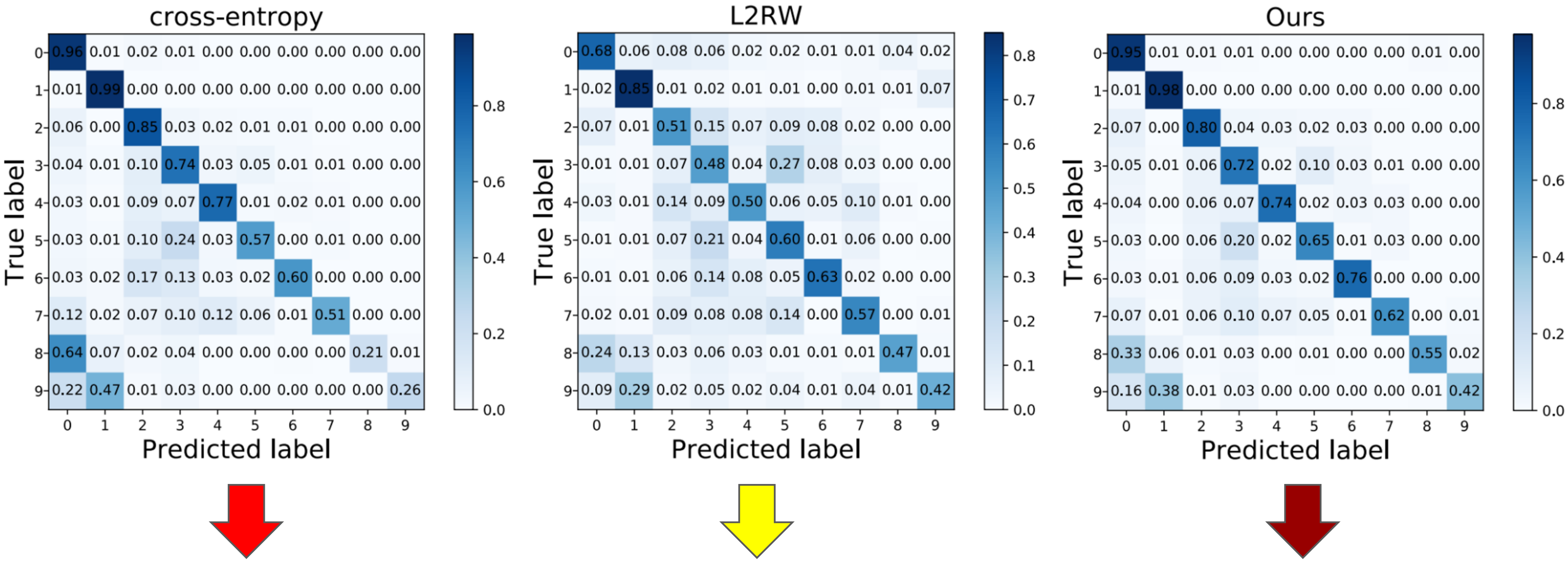
VQA Visual Question Answering

SUN

Visual Genome



Shortcomings of Current Approaches



Accuracy on Head Classes ↑ | Accuracy on Head Classes ↓ | Accuracy on Head Classes ↑

Accuracy on Tail Classes ↓ | Accuracy on Tail Classes ↑ | Accuracy on Tail Classes ↑

New Perspective - Domain Adaptation

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

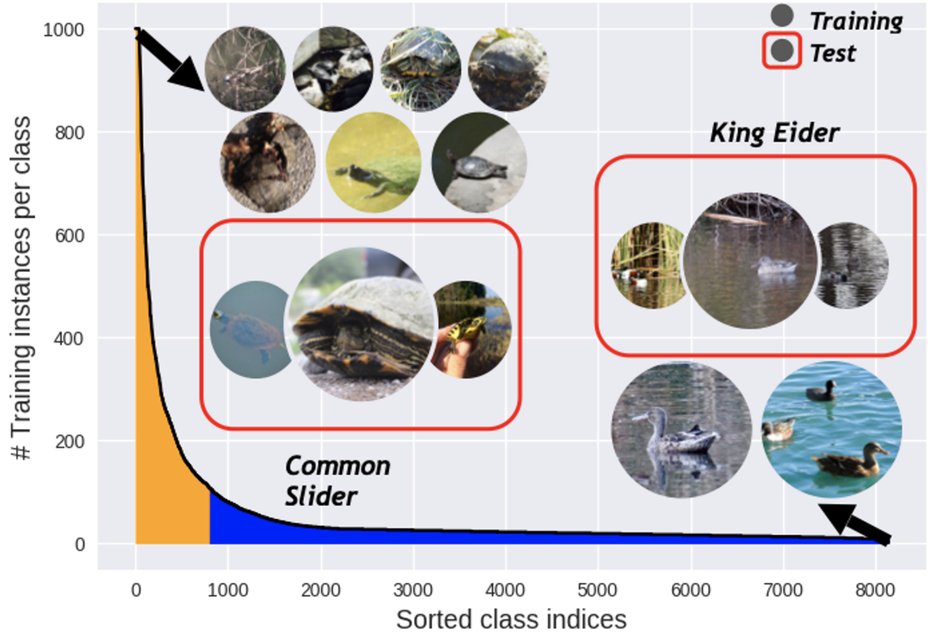
Objective

Learn models to work well on target

[Slide source]()

Existing Works

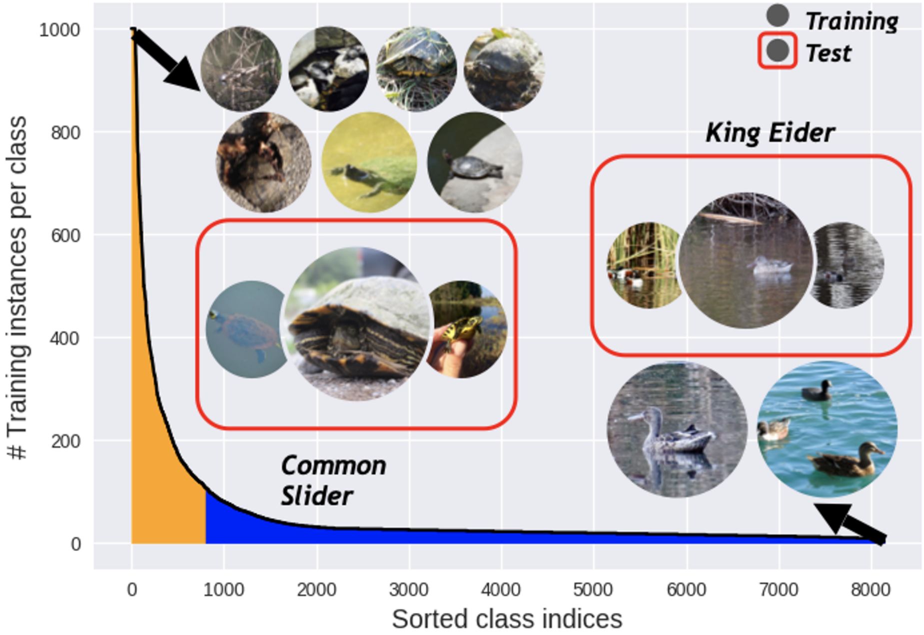
Assume target shift



$P_s(\text{x}|\text{Common Slider}) = P_t(\text{x}|\text{Common Slider})$

$P_s(\text{x}|\text{King Eider}) = P_t(\text{x}|\text{King Eider})$

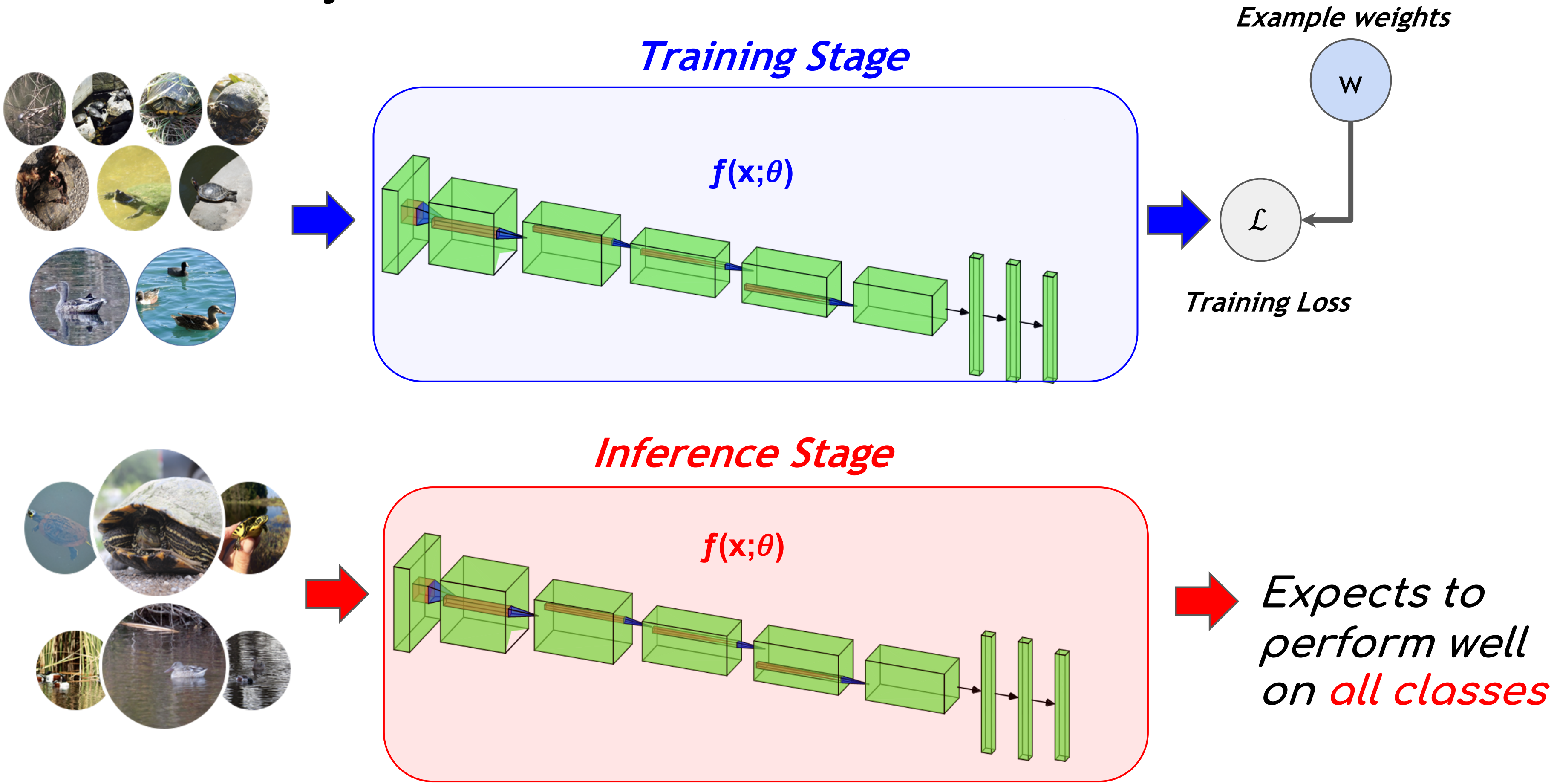
But



$P_s(x|\text{Common Slider}) = P_t(x|\text{Common Slider})$

$P_s(x|\text{King Eider}) \neq P_t(x|\text{King Eider})$

A Bird's Eye View



Two-Component Approach

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

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

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

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

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

$(1 - \beta) / (1 - \beta^n)$

[ICML'18] Learning to reweight examples for robust deep learning

Meta-learning framework

Two-Component Approach

	L2RW	Ours
Pre-training	X	✓
Clip negative ϵ	✓	X
Normalization	✓	X
Free Space of ϵ	reduced	larger

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

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

$(1 - \beta) / (1 - \beta^n)$

[ICML'18] Learning to reweight examples for robust deep learning

Meta-learning framework

Experiments

Six datasets

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

CIFAR-LT-10 - Results

CIFAR-LT-10 - Results

Test top-1 errors (%) of ResNet-32 on CIFAR-LT-10 under different imbalance settings. * indicates results reported in [45].

Imbalance factor	200	100	50	20	10	1
Cross-entropy training	34.32	29.64	25.19	17.77	13.61	7.53/7.11*
Class-balanced cross-entropy loss [7]	31.11	27.63	21.95	15.64	13.23	7.53/7.11*
Class-balanced fine-tuning	33.76	28.66	22.56	16.78	16.83	7.08
Class-balanced fine-tuning*	33.92	28.67	22.58	13.73	13.58	**6.77**
L2RW [41]	33.75	27.77	23.55	18.65	17.88	11.60
L2RW [41]*	33.49	25.84	21.07	16.90	14.81	10.75
Meta-weight net [45]	32.8	26.43	20.9	15.55	12.45	7.19
Ours with cross-entropy loss	**29.34**	**23.59**	**19.49**	**13.54**	**11.15**	7.21

CIFAR-LT-10 - Results

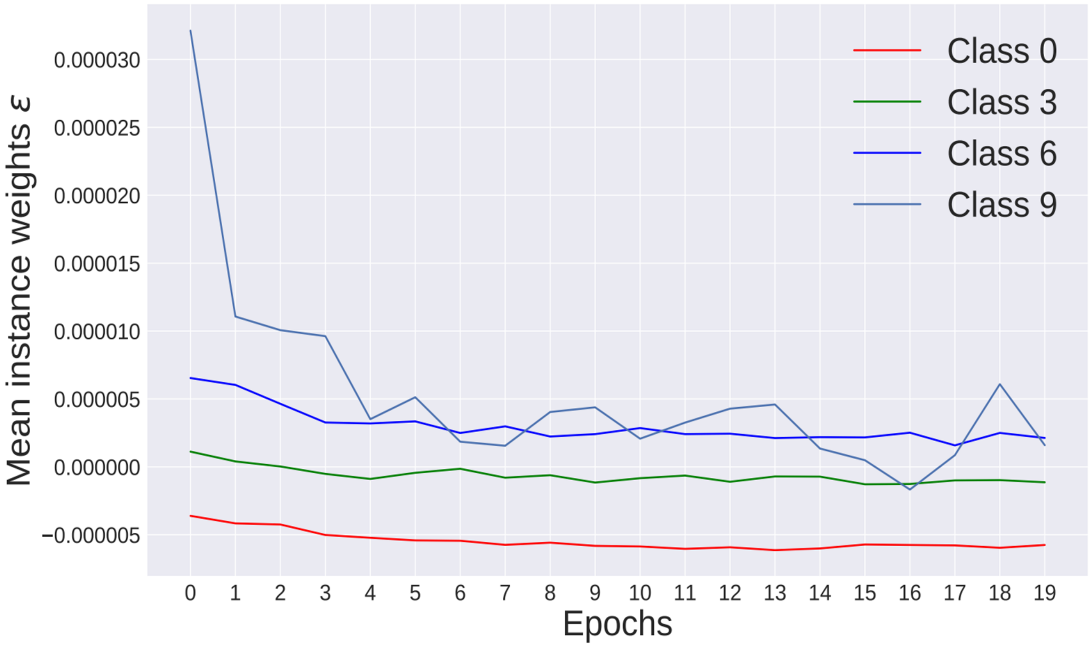
Test top-1 errors (%) of ResNet-32 on CIFAR-LT-10 under different imbalance settings. * indicates results reported in [45].

Imbalance factor	200	100	50	20	10	1
Cross-entropy training	34.32	29.64	25.19	17.77	13.61	7.53/7.11*
Class-balanced cross-entropy loss [7]	31.11	27.63	21.95	15.64	13.23	7.53/7.11*
Class-balanced fine-tuning	33.76	28.66	22.56	16.78	16.83	7.08
Class-balanced fine-tuning*	33.92	28.67	22.58	13.73	13.58	**6.77**
L2RW [41]	33.75	27.77	23.55	18.65	17.88	11.60
L2RW [41]*	33.49	25.84	21.07	16.90	14.81	10.75
Meta-weight net [45]	32.8	26.43	20.9	15.55	12.45	7.19
Ours with cross-entropy loss	**29.34**	**23.59**	**19.49**	**13.54**	**11.15**	7.21

CIFAR-LT-10 - Results

Focal loss [32]	34.71	29.62	23.29	17.24	13.34	6.97
Class-balanced focal Loss [7]	31.85	25.43	20.78	16.22	12.52	6.97
Ours with focal Loss	**25.57**	**21.1**	**17.12**	**13.9**	**11.63**	7.19
LDAM loss [4] (results reported in paper)	-	26.65	-	-	13.04	11.37
LDAM-DRW [4] (results reported in paper)	-	22.97	-	-	11.84	-
Ours with LDAM loss	**22.77**	**20.0**	**17.77**	**15.63**	12.6	**10.29**

What are the learned ϵ



Long-tailed visual recognition

- A new perspective from Domain Adaptation

- A two-component approach

- SOTA results on six datasets

Domain Adaptation

A powerhouse of ideas & techniques

- Domain-invariant representations
- Maximum Mean Discrepancy
- Curriculum Domain Adaptation
- Adversarial adaptation
- Self-supervised adaptation

THANK
YOU
FOR
YOUR
ATTENTION