Tutorial: Best Practices of ConvNet Application

Yuan-Hong Andrew Liao

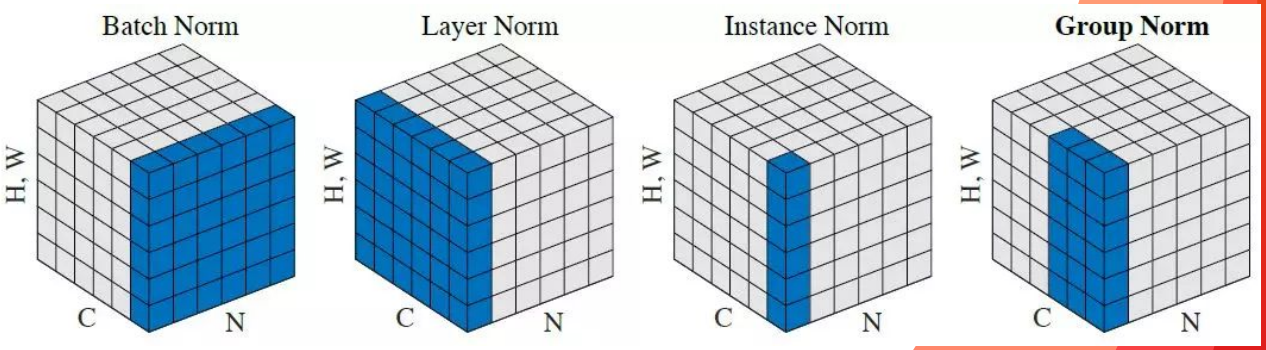
Feb. 25, 2020

Overview

- Normalization
- Transfer Learning
- Label Imbalance
- Model the Long Tail

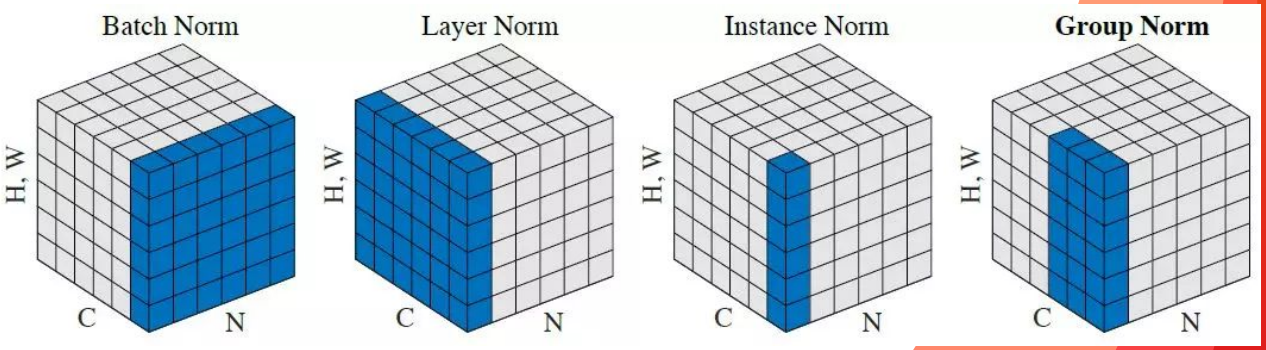
Normalization: BatchNorm

- Avoid Covariance Shift
- Compute batch statistic during training
  - Require large batchsize



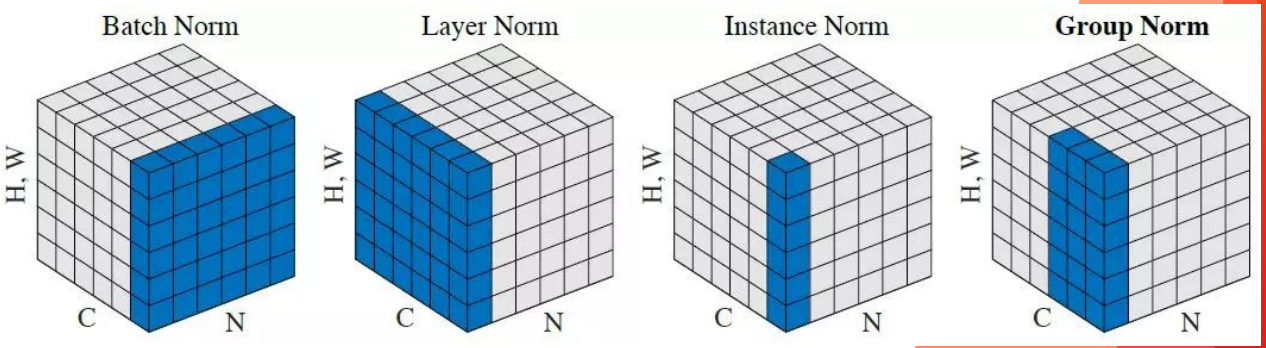
Normalization: LayerNorm

- How about Recurrent Neural Network?
  - LayerNorm

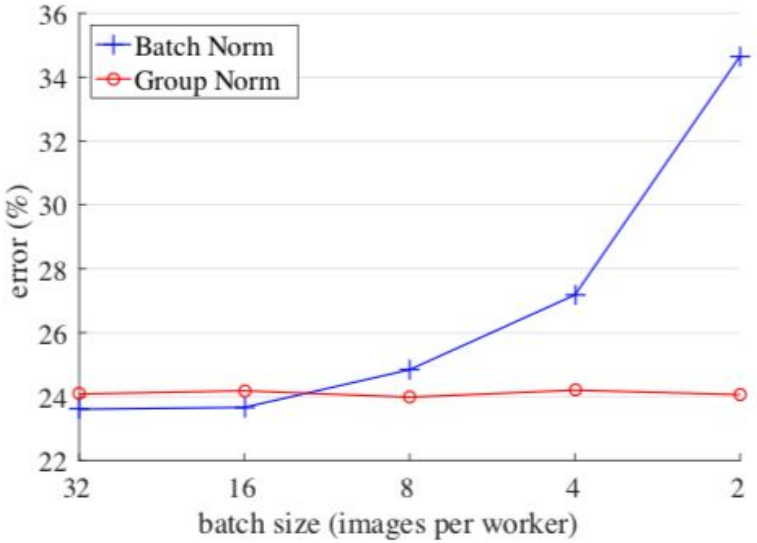


Normalization: GroupNorm

- Large Feed-Forward network
  - Sometimes batch size is small due to computational constraints
- How to adjust?
  - GroupNorm



Normalization: GroupNorm



Normalization: GroupNorm

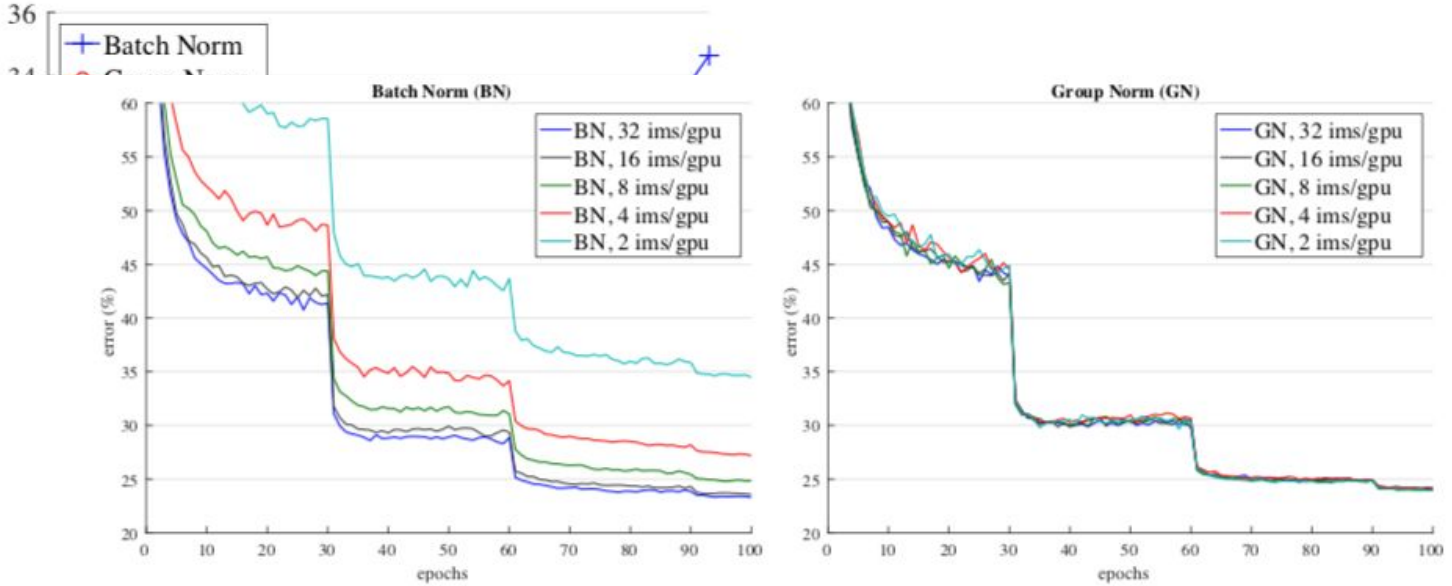


Figure 5. **Sensitivity to batch sizes**: ResNet-50's validation error of BN (left) and GN (right) trained with 32, 16, 8, 4, and 2 images/GPU.

Normalization: SyncBatchNorm

- Large Feed-Forward network
  - Sometimes batch size is small due to computational constraints
- Split large batch into several and distribute them many GPUs
  - Collect the batch statistics from all devices

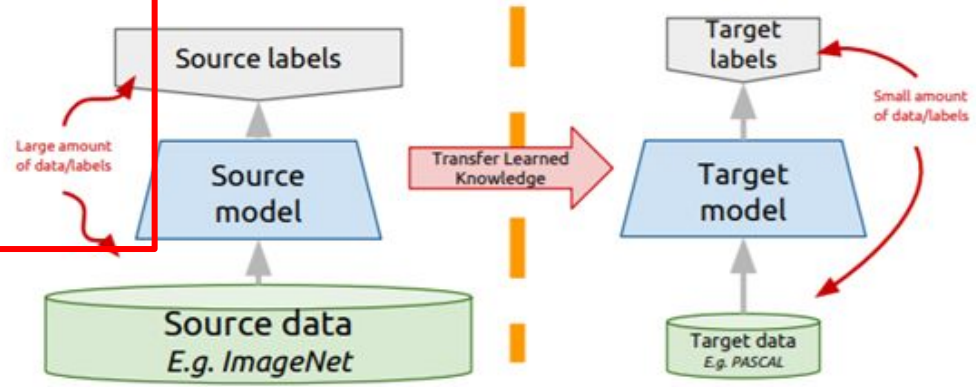
Transfer learning: idea

Instead of training a deep network from scratch for your task:

- Take a network trained on a different domain for a different **source task**
- Adapt it for your domain and your **target task**

Variations:

- Same domain, different task
- Different domain, same task



https://towardsdatascience.com/a-comprehensive-hands-on-guide-to-transfer-learning-with-real-world-applications-in-deep-learning-212bf3b2f27a

Freeze or fine-tune?

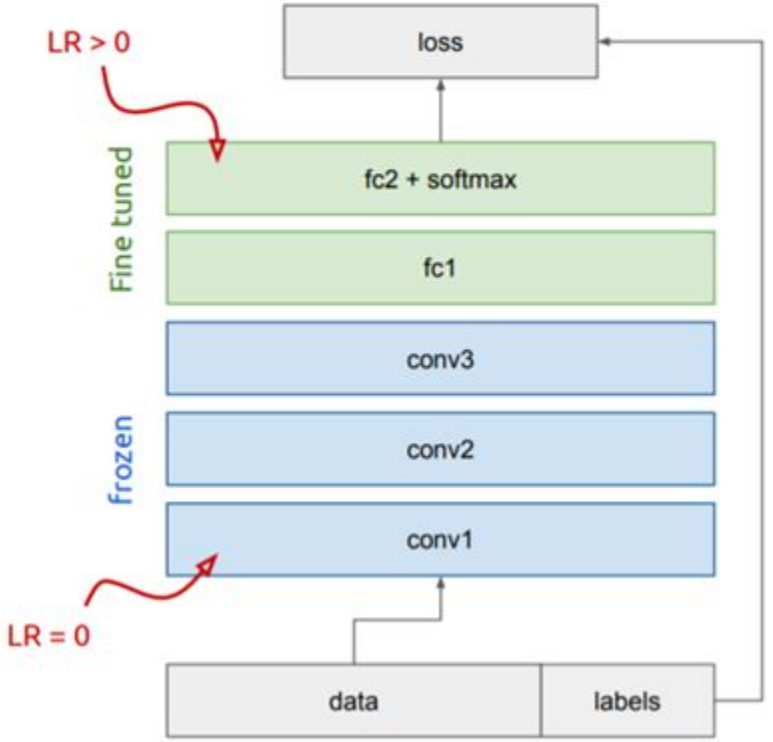
Bottom *n* layers can be frozen or fine tuned.

- **Frozen**: not updated during backprop
- **Fine-tuned**: updated during backprop

Which to do depends on target task:

- **Freeze**: target task labels are scarce, and we want to avoid overfitting
- **Fine-tune**: target task labels are more plentiful

In general, we can set learning rates to be different for each layer to find a tradeoff between freezing and fine tuning



https://towardsdatascience.com/a-comprehensive-hands-on-guide-to-transfer-learning-with-real-world-applications-in-deep-learning-212bf3b2f27a

Transfer Learning: Rule of thumb

	|New Dataset| is small	|New Dataset| is large
Similar to Source dataset	**Freeze**	**Fine-tune all**
Dissimilar to Source dataset	**Try SVM first from low-level features**	**Train from scratch**

http://cs231n.github.io/transfer-learning/#tf

Transfer Learning

- Additional advice:
  - smaller learning rate for ConvNet weights

http://cs231n.github.io/transfer-learning/#tf

Task Transfer Learning

- Same domain, different tasks
- Computer Vision Taskonomy: http://taskonomy.stanford.edu
- What is the relation between *3d keypoint detection* and *depth estimation?*

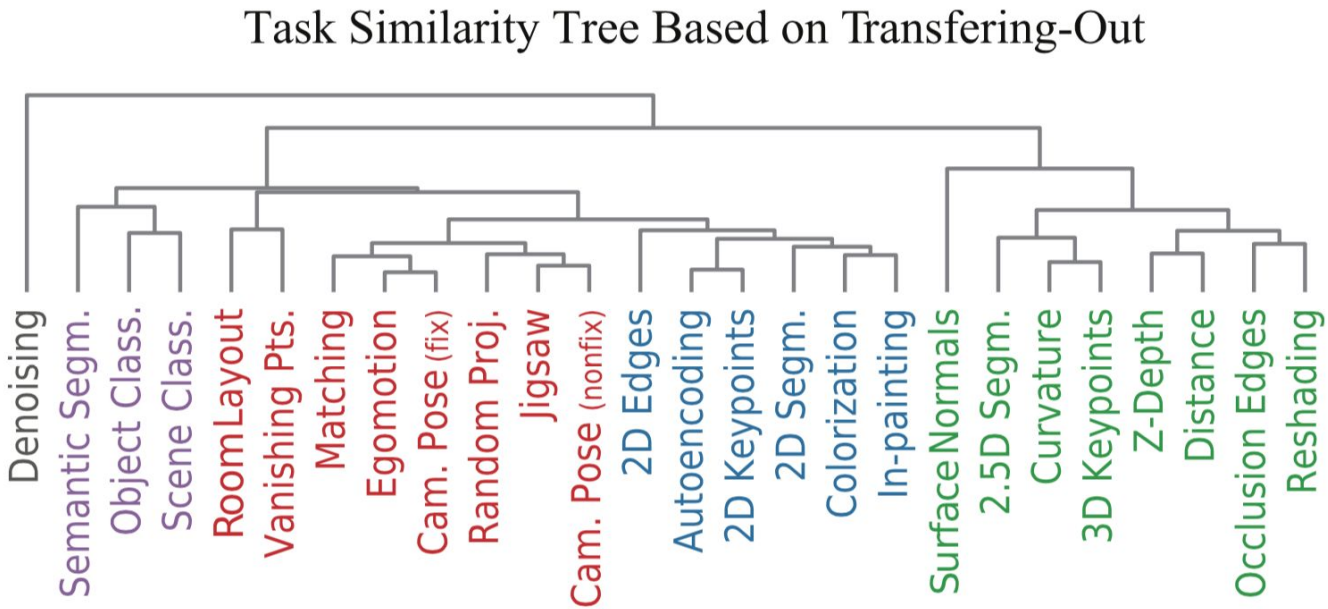
Taskonomy: Disentangling Task Transfer Learning, Amir et al, 2018

Task Transfer Learning

- Same domain, different tasks
- Computer Vision Taskonomy:http://taskonomy.stanford.edu
- What is the relation between *3d keypoint detection* and *depth estimation?*
- Is it able to structurally represented?

Taskonomy: Disentangling Task Transfer Learning, Amir et al, 2018

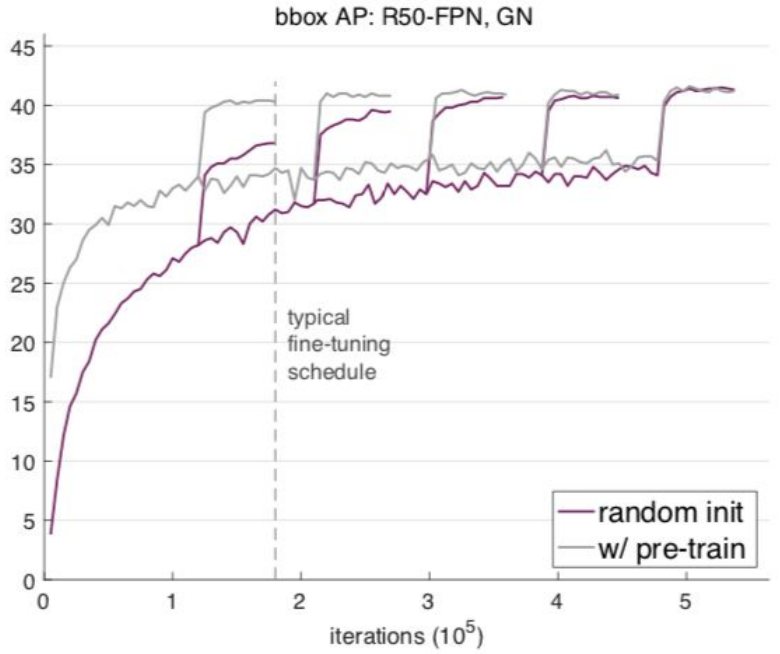
Task Transfer Learning



Taskonomy: Disentangling Task Transfer Learning, Amir et al, 2018

Transfer Learning from ImageNet?

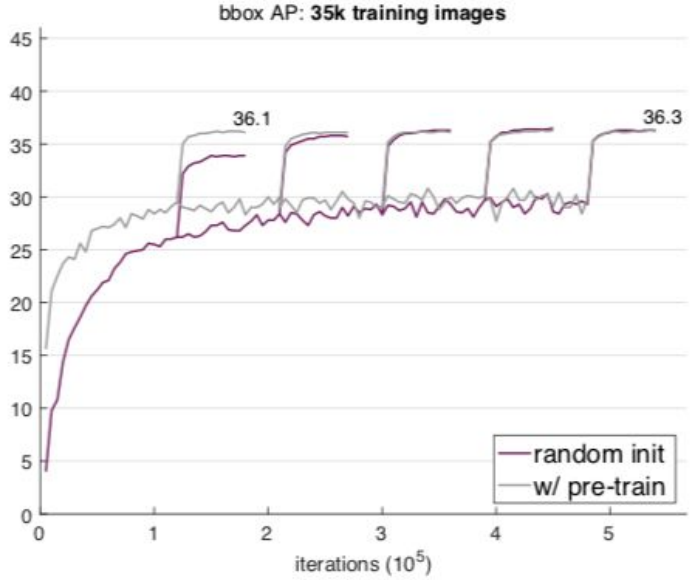
- Always better?
- ImageNet: 130M
- COCO: 8.6M

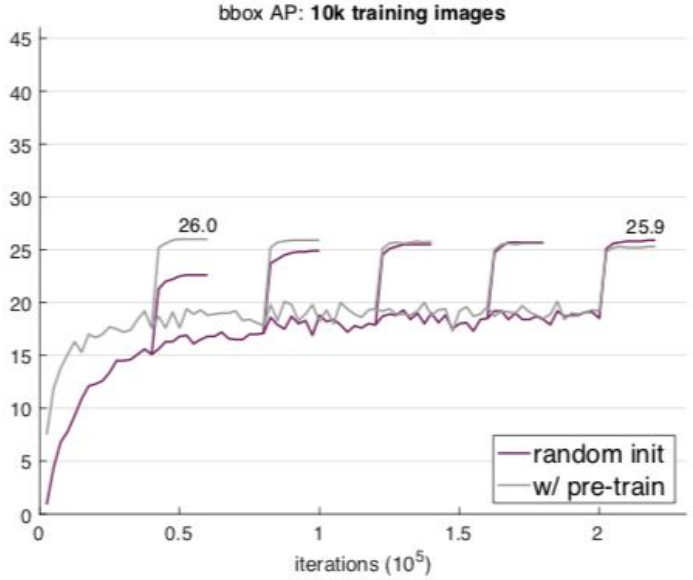


Rethinking ImageNet Pre-training, Kaiming et al, 2019

Transfer Learning from ImageNet?

- Always better?
  ImageNet: 120M

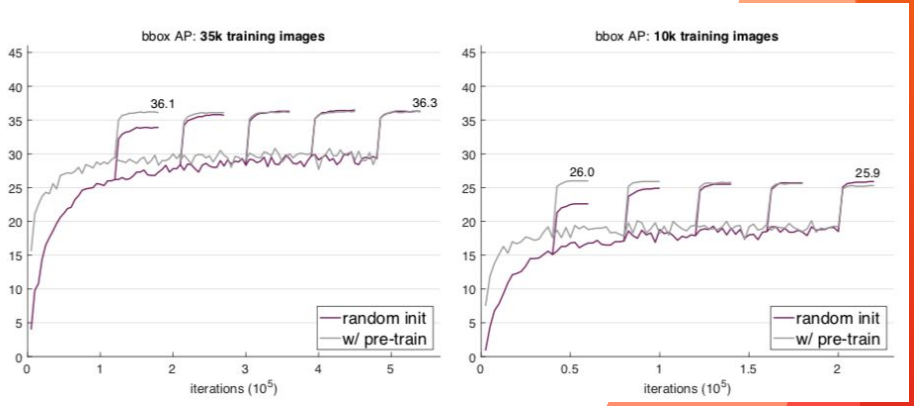




ming et al, 2019

Transfer Learning from ImageNet?

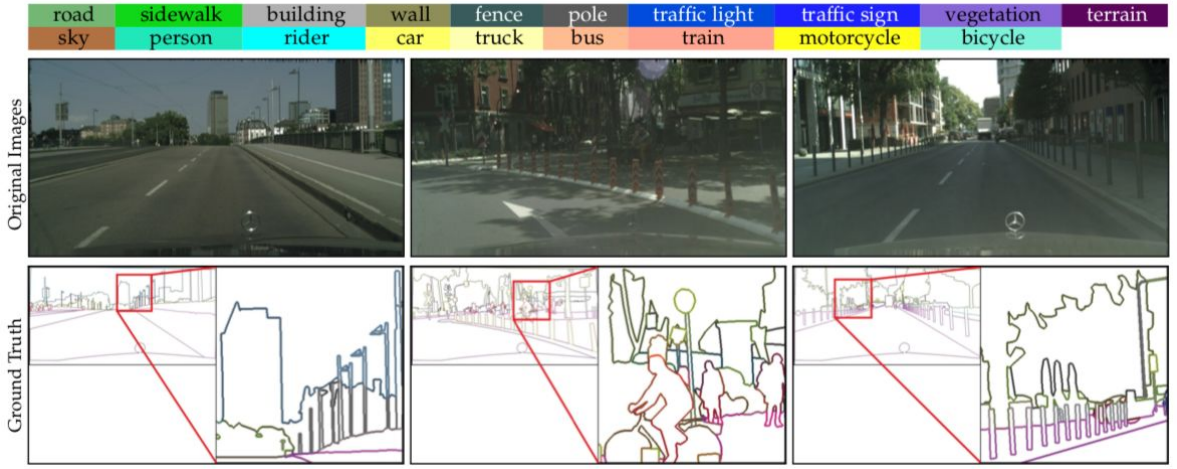
- With only 1k training image:
  - w/ pretrain: 9.9 AP
  - Random init: 3.5 AP



Rethinking ImageNet Pre-training, Kaiming et al, 2019

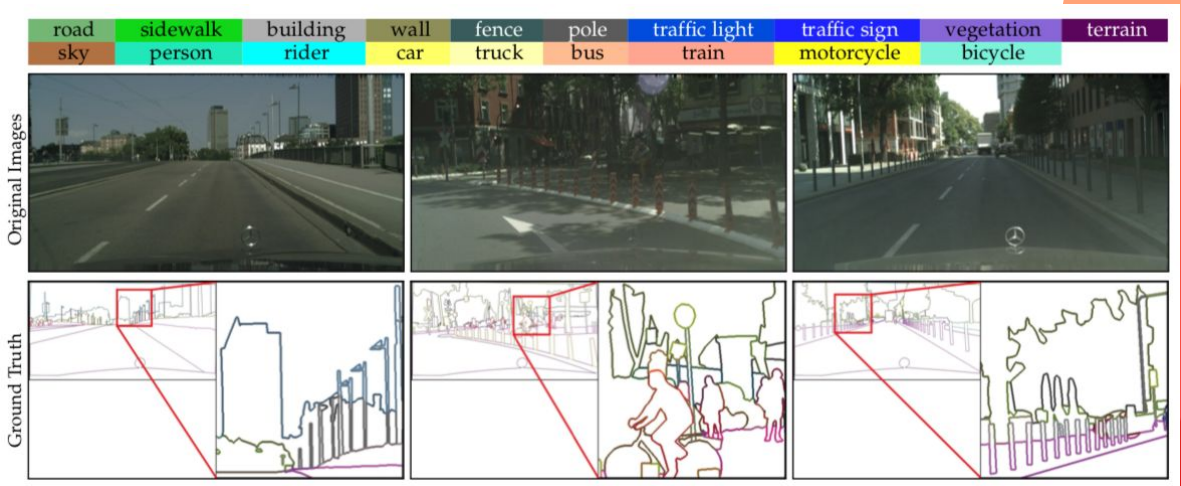
Label Imbalance

- Semantic Segmentation
- Image Segmentation
- Contour Detection

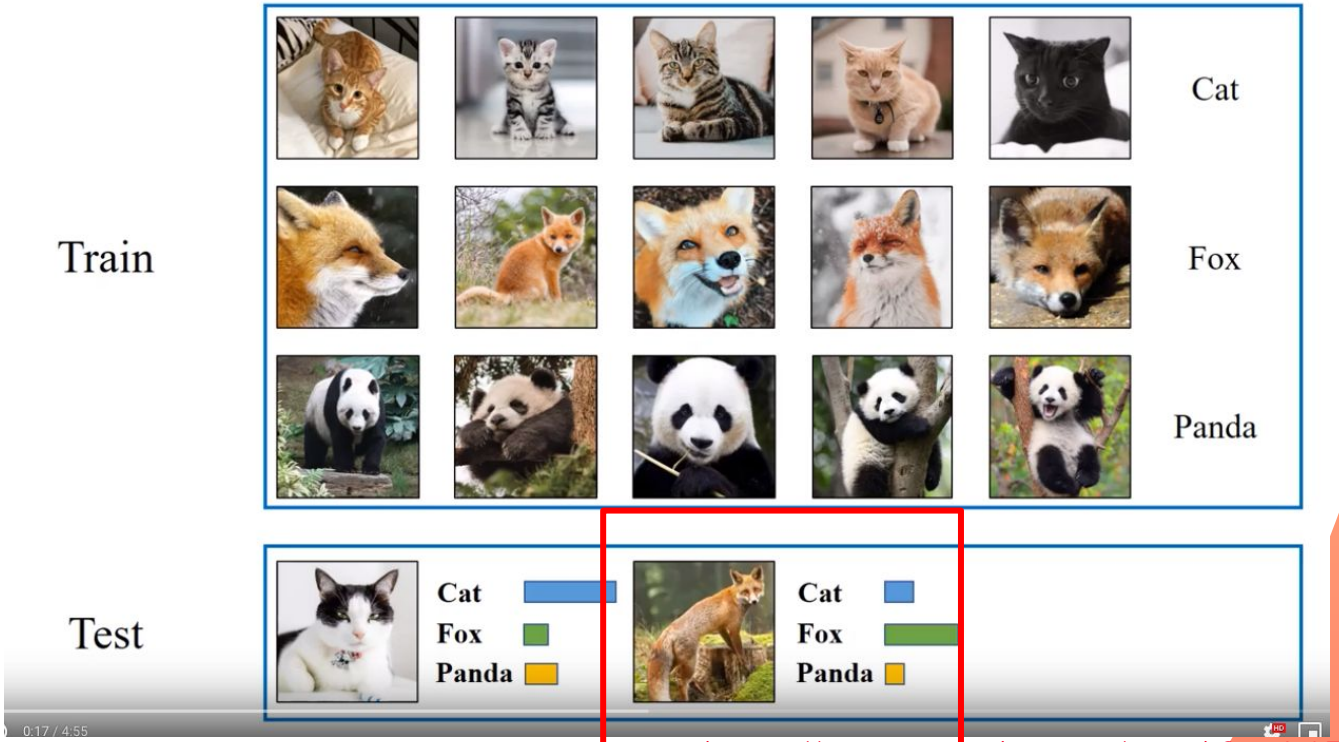


Label Imbalance

- Reweight the loss by class ratio
- Data Resampling

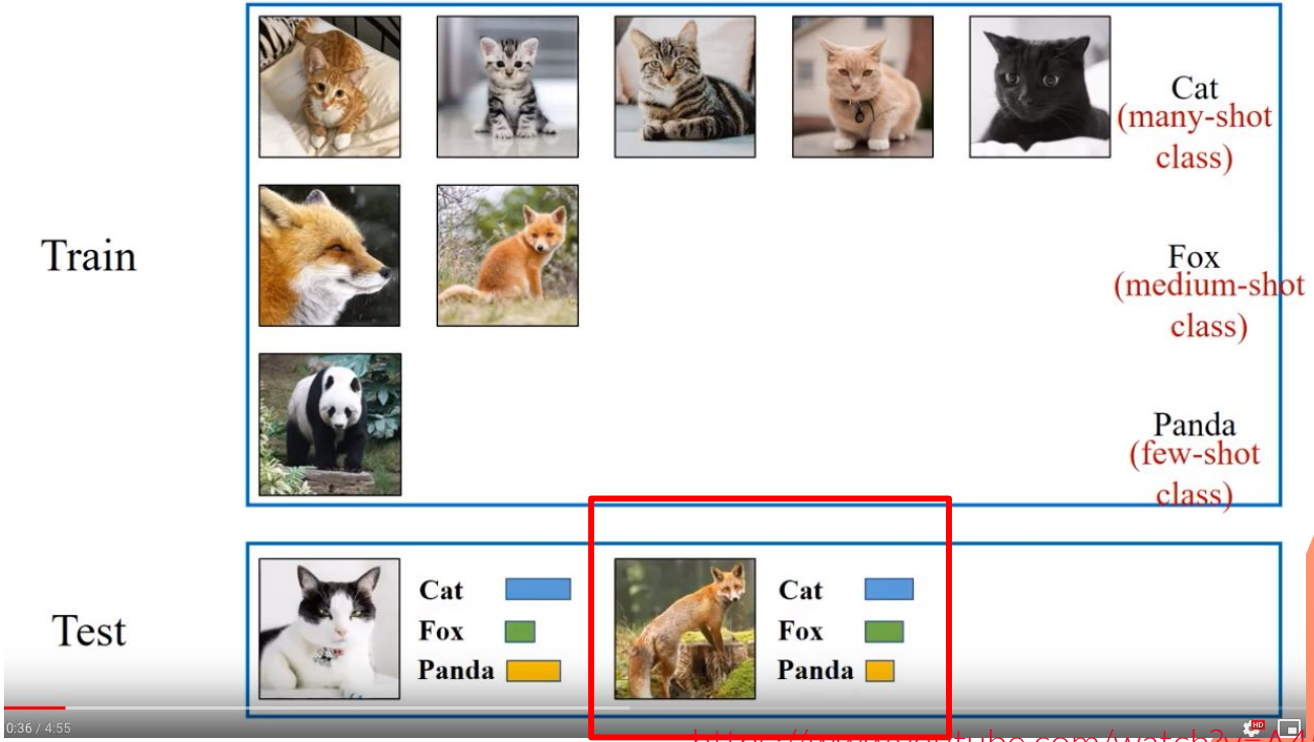


Model the Long Tail



https://www.youtube.com/watch?v=A45wrs1g8VA

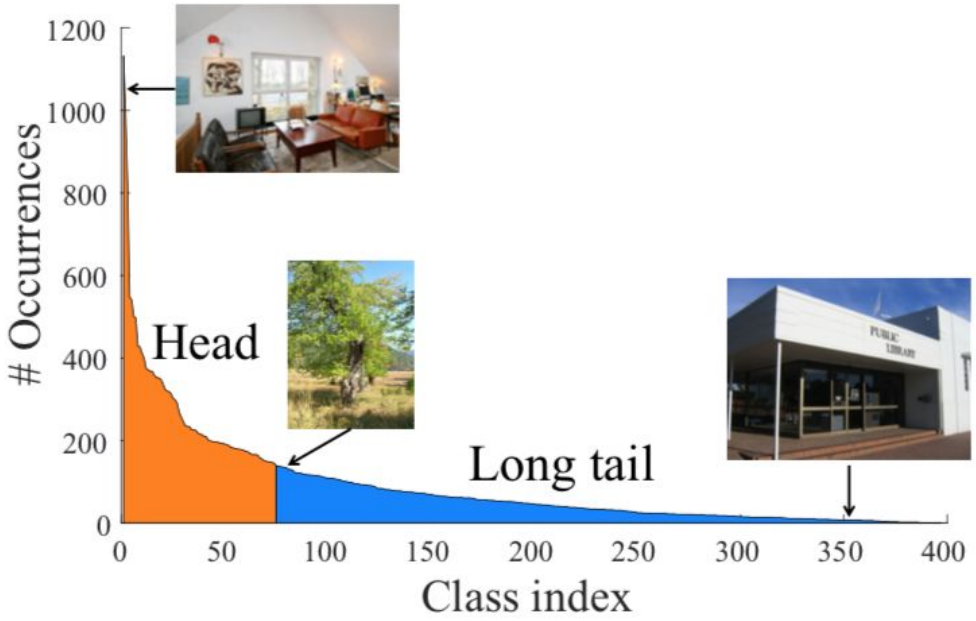
Model the Long Tail



https://www.youtube.com/watch?v=A45wrs1g8VA

Model the Long Tail

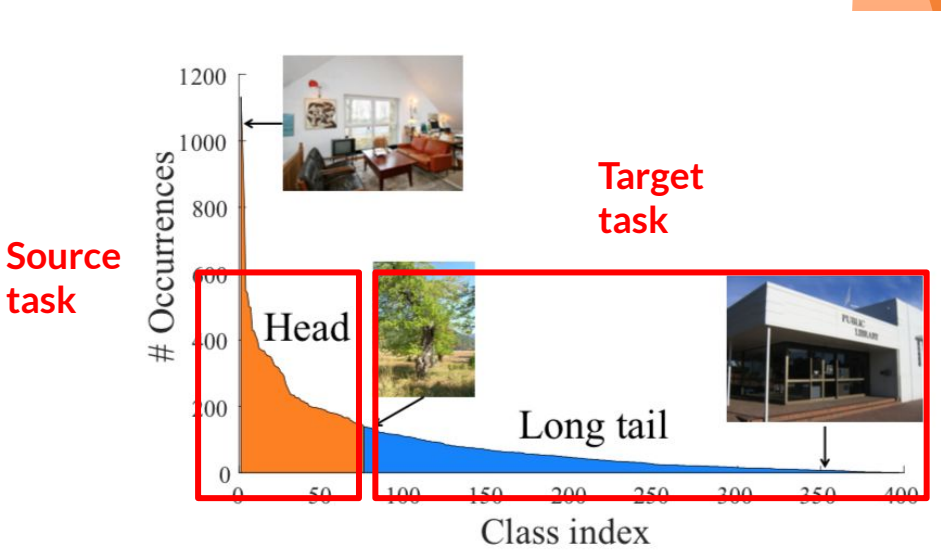
- Tail distribution
- Still an Open question



(a) Long-tail distribution on the SUN-397 dataset.

Learning to Model the Tail, 2017

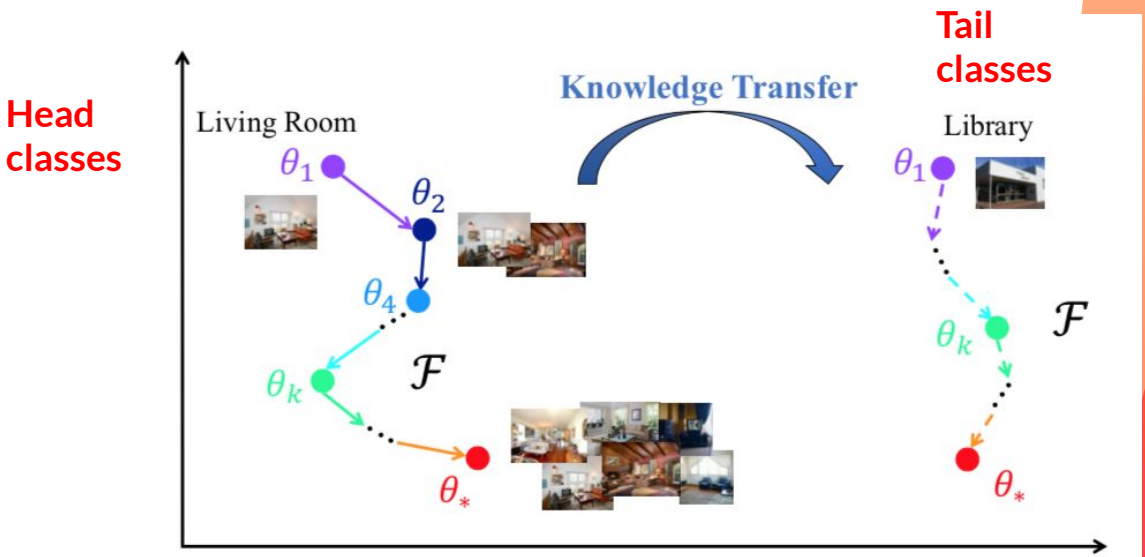
- Transfer learning



(a) Long-tail distribution on the SUN-397 dataset.

Learning to Model the Tail, 2017

- Transfer learning
- Meta-learner to learn the “model dynamics”



Reference

- http://cs231n.github.io/understanding-cnn/