LEARNING AN EFFICIENT NETWORK FOR LARGE-SCALE HIERARCHICAL OBJECT DETECTION WITH DATA IMBALANCE

3rd place solution to Open Images Detection Challenge 2019

X Bu, J peng, C Wang, C Yu, G Cao

Outline

- Features of Open-Images dataset(Object Detection)
- Single-model solution
- Ensemble

Features of Open-Images Dataset

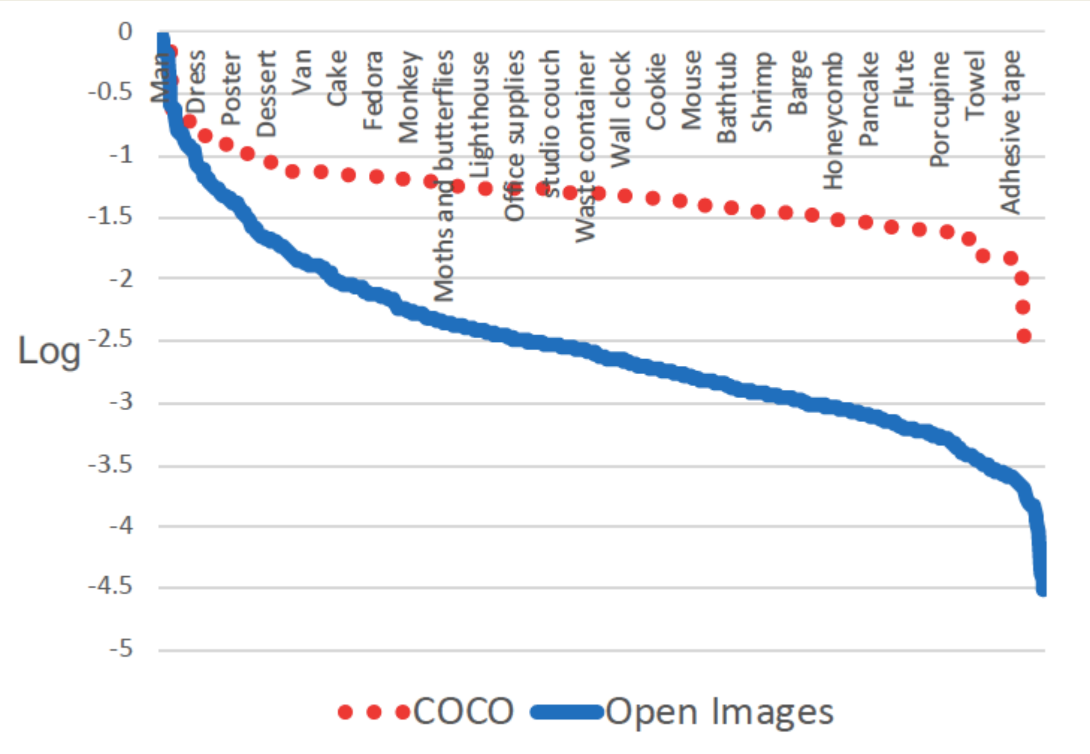
- Large scale of data with partial annotations
- Extremely imbalanced data
- Hierarchical label
- Confused annotations

Large scale of data with partial annotations

- Huge number of training images and instances.
- A great amount of instances are not annotated in training set.

Dataset	Pascal VOC	COCO	Objects365	Open Images
Categories	20	80	365	500
Images	11,540	123,287	638,630	1,784,662
Boxes	27,450	886,287	10,101,056	12,421,955
Boxes/image	2.4	7.2	15.8	7.0

Extremely imbalanced data



- Category like person has 1.4M instances annotated.
- Category like pressure cooker has only 14 instances annotated.

Hierarchical Label

Strawberry => fruit

Sometimes labeled as parent class.

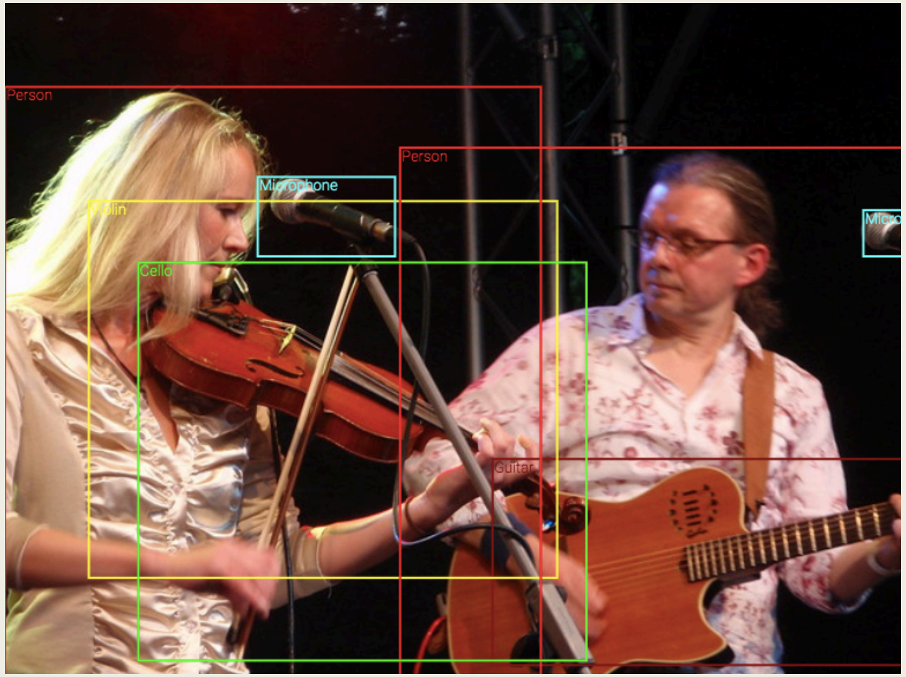
Strawberry => strawberry

Sometimes labeled as leaf class.

Confusing Annotations

An instance is labeled as *flashlight* and *torch* at the same time.

A *cello* is wrongly annotated as *violin* while keeping the label of *cello*.



Single-Model Solution

- Backbone
- Loss function
- Data balance sampling
- Data Augmentation
- Classifier

Baseline

- Config
  - *Backbone: ResNeXt152-32x4d*
  - *Training from scratch(20epochs)*
  - *Sync BN*
  - *More FPN stages*
  - *More anchors*
  - *Multi-scale train*
  - *Multi-scale test*
- Performance
  - *AP50: 53.88*

Method	Public Leader board
Baseline (FPN with ResNeXt-152)	53.88
+EfficientNet-B7	55.59
+Distributed Softmax Loss	56.43
+Class-aware Sampling	61.09
+Auto Augmentation	61.84
+Classifier	62.29
+Ensemble	67.17

EfficientNet(Initial Trial)

- We set 600pixels as initial size.
- Search B0 to B1 under constraint
$$d, w, r = \arg\max_{\alpha,\beta,\gamma}(mAP(model(\alpha, \beta, \gamma))),$$
$$\text{s.t.} \quad \alpha \cdot \beta^2 \cdot \gamma^2 \approx 2$$
- Scale up B1 to B7(1.2/1.15/1.05)
- Single scale improvement: 1.8
- However, multi-scale test brings no improvement.

Two Reasons:

1. FLOPs are equally spread on each stage during scaling-up.
2. Input size of B7 is 1200.

Good for single-scale test, bad for multi-scale test.

EfficientNet(Final Trial)

- Config:
  - *Fix short side to 800.*
  - *Shrink width.*
  - *Make it deeper and concentrate most FLOPs on later stages.*
- Performance:
  - *Single-scale test 1.4*
  - *Multi-scale test 1.7*

Method	Public Leader board
Baseline (FPN with ResNeXt-152)	53.88
+EfficientNet-B7	55.59
+Distributed Softmax Loss	56.43
+Class-aware Sampling	61.09
+Auto Augmentation	61.84
+Classifier	62.29
+Ensemble	67.17

Distributed Softmax Loss

- Vanilla Softmax:

$$\mathcal{L}_{cls} = \sum_{c=1}^{C} \mathbb{1}_{y_c=1} log(\frac{e^{x_c}}{\sum_{i=1}^{C} e^{x_i}})$$

- Distributed Softmax:

$$\mathcal{L}_{cls} = \sum_{c=1}^{C} y_c log(\frac{e^{x_c}}{\sum_{i=1}^{C} e^{x_i}})$$

y_c means (1/k), where k means number of multiple labels

Method	Public Leader board
Baseline (FPN with ResNeXt-152)	53.88
+EfficientNet-B7	55.59
+Distributed Softmax Loss	56.43
+Class-aware Sampling	61.09
+Auto Augmentation	61.84
+Classifier	62.29
+Ensemble	67.17

Data-Balance Sampling

- Equally sample images of each class in an epoch.
- NOTE1: Models with data-balance sampling COULD NOT be trained from scratch.
- NOTE2: Data-balance does not work fine with many tricks, including DCN, Nas-FPN, Heavier head.

Method	Public Leader board
Baseline (FPN with ResNeXt-152)	53.88
+EfficientNet-B7	55.59
+Distributed Softmax Loss	56.43
$\boxed{\text{+Class-aware Sampling}}$	$\boxed{61.09}$
+Auto Augmentation	61.84
+Classifier	62.29
+Ensemble	67.17

Auto Data Augmentation

- We used the searched auto-augmentation strategy.
- Apply it on only rare classes when using data-balance.

Method	Public Leader board
Baseline (FPN with ResNeXt-152)	53.88
+EfficientNet-B7	55.59
+Distributed Softmax Loss	56.43
+Class-aware Sampling	61.09
+Auto Augmentation	61.84
+Classifier	62.29
+Ensemble	67.17

Classifier

- We train several image classifiers of 500 classes and filter the detection result with an extremely low threshold.

- The classifiers are also trained with distributed softmax loss.

Method	Public Leader board
Baseline (FPN with ResNeXt-152)	53.88
+EfficientNet-B7	55.59
+Distributed Softmax Loss	56.43
+Class-aware Sampling	61.09
+Auto Augmentation	61.84
+Classifier	62.29
+Ensemble	67.17

Ensemble

- 1. We trained 20 models in total.
- 2. Randomly split them into 4 groups.
- 3. Rescore boxes, merge and apply NMS(Including box voting).
- 4. Repeat 3 on the four ensemble results.

Method	Public Leader board
Baseline (FPN with ResNeXt-152)	53.88
+EfficientNet-B7	55.59
+Distributed Softmax Loss	56.43
+Class-aware Sampling	61.09
+Auto Augmentation	61.84
+Classifier	62.29
+Ensemble	67.17

Conclusion

- Data-balance are extremely important.
- The ensemble strategy is very important.
- The variety of model pool for ensemble is important.

Thank you!