Open Images V5 Detection Challenge: 5th Place Solution without External Data

Xi Yin, Jianfeng Wang, Lei Zhang

{xiyin1, jianfw, leizhang}@microsoft.com

Microsoft Cloud and AI

Microsoft

ICCV 2019 Seoul, Korea

Highlights

(1) Challenges in Open Images Detection

- Imbalanced class distribution.
- Incomplete annotations with non-exclusive image-level negative labels.
- Large-scale hierarchical structured classes.

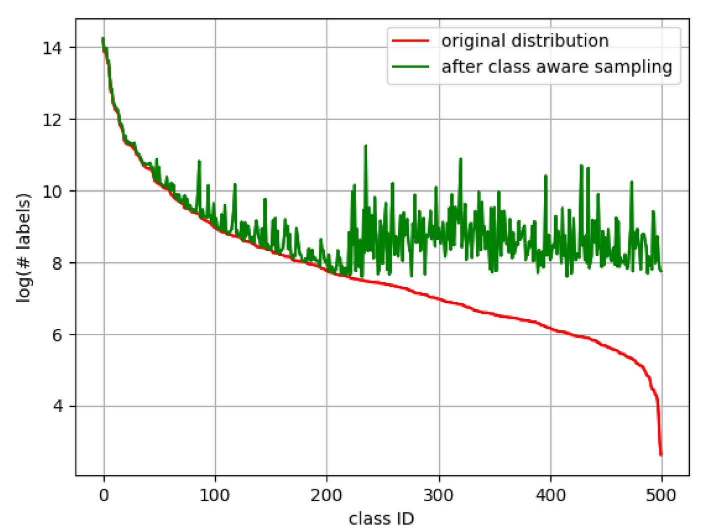
(2) Insights and Contributions

- Class-aware sampling can solve the imbalanced distribution problem to a large extent.
- We propose a novel soft label propagation to alleviate the incomplete annotation issue.
- Score aggregation during inference to take the hierarchy structure into consideration.
- Score and box voting for model ensemble to boost the performance significantly.

Open Images Detection Challenge (OID-C)

(1) Imbalanced Distribution

CAS (class-aware sampling) [1]: duplicate the images with rare classes such that each class has a minimum number of 2,000 bounding boxes. (~ 5 points gain in mAP)



OID-C:

1.7M images

12.2M bounding boxes

500 classes

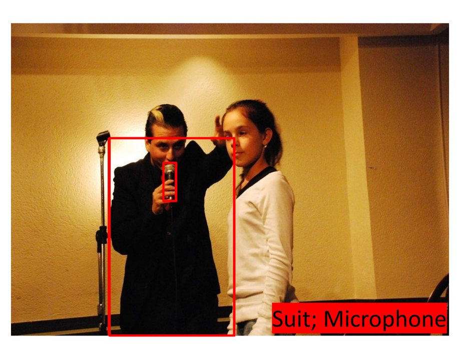
Imbalanced:

"Man": 1.4M bounding boxes

"Pressure cooker": 14 bounding boxes

(2) Incomplete Annotations (non-exhaustive image-level labeling)

2.3 positive classes and 1.1 negative classes per image, 7.3 bounding boxes per image.



(3) Other Properties: Group-of box, Hierarchy labels

655K (5.4%) Group-of bounding boxes.

68 parent classes.

Soft Label Propagation

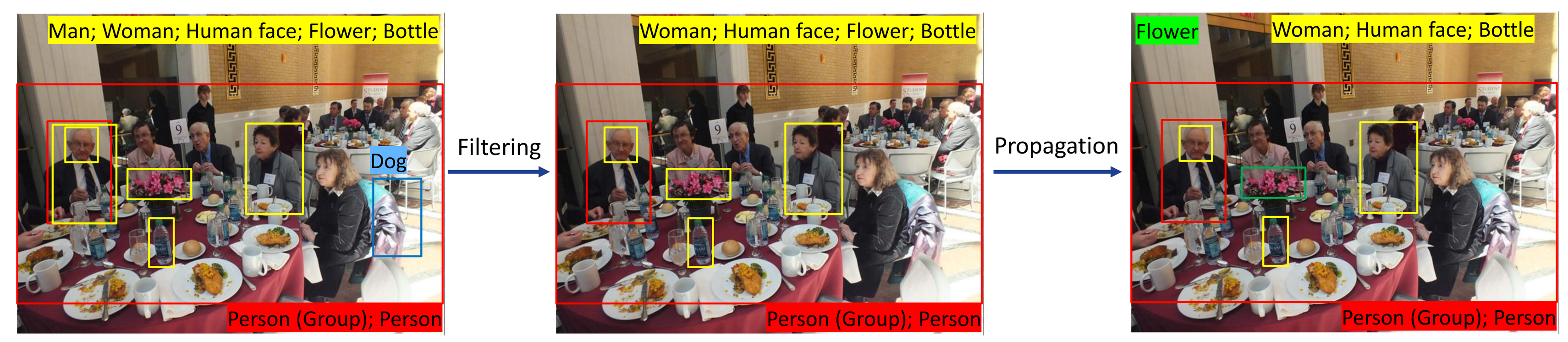
(1) Key Ideas

Use a pre-trained model to perform detection on the training images. Based on the ground-truth annotations and priors to generate new labels: positive label and neutral label.

Positive label: a bounding box with class and confidence, confidence is used as the loss weight.

Neutral label: a bounding box without class label, ignored during RoI matching.

(2) Label Propagation Procedure

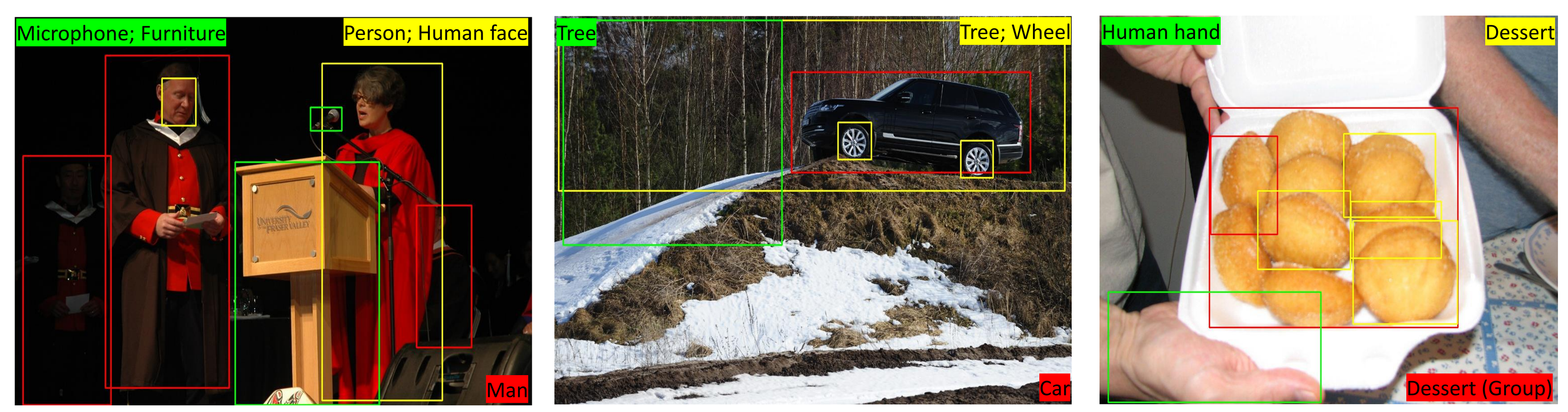


Filtering: remove detections of image-level negative labels, or that matched to a ground-truth box.

Propagation: propagate based on Group-of box, object-part prior [2], and confidence scores.

(3) Label Propagation Results

Re-train the model using original labels and propagated labels. (~ 1.2 points gain in mAP)



Single Model Results

(1) Modeling

Dataset: class-aware sampling and soft label propagation.

Framework: Faster R-CNN with ResNeXt-152-FPN backbone.

Loss: softmax cross-entropy loss for classification and smoothed-L1 loss for regression.

(2) Hierarchy Structure

Label expansion (after NMS): duplicate a child box to its ancestors and perform NMS on parent classes.

Score aggregation (before NMS): aggregate the scores of all children to its parent class before NMS.

(3) Single Model Results

Training: batch_size = 32, max_iter = 800K (step at 400K, 750K), LR = 0.02, weight_decay = 0.00005.

Testing: NO test time augmentation, min_size_input = 800, score = 0.0001, max_detections = 600.

Single model results on the validation set.

	X152	X152-SE	X152-DCN	X152-cascade
Label expansion	69.47	69.62	70.21	69.42
Score aggregation	69.59	69.66	70.38	69.50

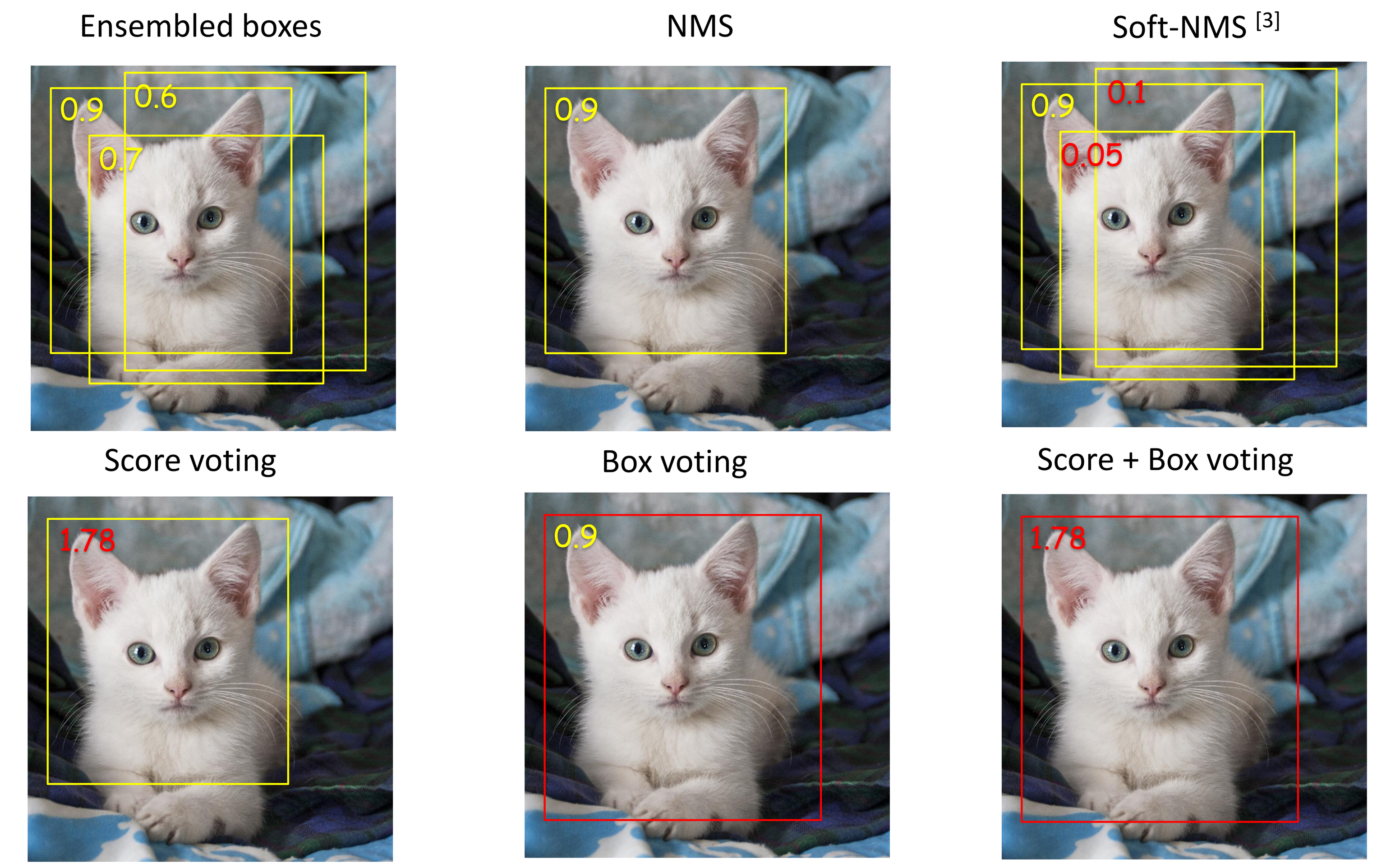
Model Ensemble

(1) Single Models + Expert Models

Single models: X152, X152-SE, X152-DCN, X152-cascade.

Expert models: 219 classes with low AP (below average). Split into 20 groups based on the label hierarchical structure so that similar classes and sibling classes are in the same group. All parent classes that are not included will be used as **neutral** labels.

(2) Score and Box Voting



Box A: s1 = 0.9, Box B: s2 = 0.7, Box C: s3 = 0.6. IoU (A, B) = 0.75, IoU(A, C) = 0.6.

Score voting (IoU-based voting): s = 0.9 + 0.7 * 0.75 + 0.6 * 0.6 = 1.78.

Box voting (score-based voting): b = (b1 * 0.9 + b2 * 0.7 + b3 * 0.6) / (0.9 + 0.7 + 0.6).

(3) Model Ensemble Results

Model ensemble results on the validation set.

	NMS	Score Voting	Box Voting	Soft-NMS	mAP@Val
single models	0.5	NO	NO	NO	71.95
single models	NO	NO	NO	0.2	72.10
single models	0.6	NO	NO	0.2	72.27
single models	NO	0.5	NO	NO	73.61
single models	NO	0.6	NO	0.2	74.02
single models	NO	0.6	0.6	0.2	74.22
single + expert models	NO	0.6	0.6	0.2	75.22

References

[1] Yuan Gao, Xingyuan Bu, Yang Hu, Hui Shen, Ti Bai, Xubin Li, Shilei Wen. "Solution for Large-Scale Hierarchical Object Detection Datasets with Incomplete Annotation and Data Imbalance." In *ECCV Workshop*, 2018.

[2] Yusuke Niitani, Takuya Akiba, Tommi Kerola, Toru Ogawa, Shotaro Sano, Shuji Suzuki. "Sampling Techniques for Large-Scale Object Detection From Sparsely Annotated Objects." In *CVPR*, 2019.

[3] Navaneeth Bodla, Bharat Singh, Rama Chellappa, Larry S. Davis. "Soft-NMS – Improving Object Detection With One Line of Code." In *ICCV*, 2017.

[4] Francisco Massa, Ross Girshick. "maskrcnn-benchmark: Fast, modular reference implementation of Instance Segmentation and Object Detection algorithms in PyTorch." https://github.com/facebookresearch/maskrcnn-benchmark.