



# **PFDet: 3rd Place Solution to Instance Segmentation Track**

Yusuke Niitani\*, Toru Ogawa\*, Shuji Suzuki\*,  
Takuya Akiba, Tommi Kerola, Kohei Ozaki, Shotaro Sano

\*: Equal Contribution

Preferred Networks, Inc., Japan

<https://arxiv.org/abs/1910.11534>

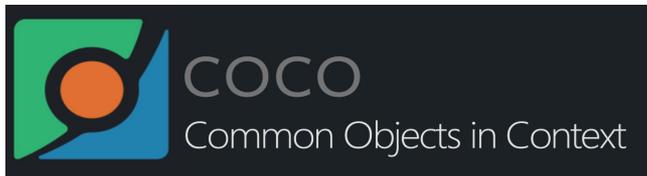
# Outline

1. Fast R-CNN
2. Learning with Federated Dataset
3. Expert Models
4. Results
5. Lessons Learned

# Large Number of Images

	MS COCO	Open Images
# of classes	80	500
# of images	0.12M	1.7M

Increase by more than  
**x14**



# Fast R-CNN is Efficient

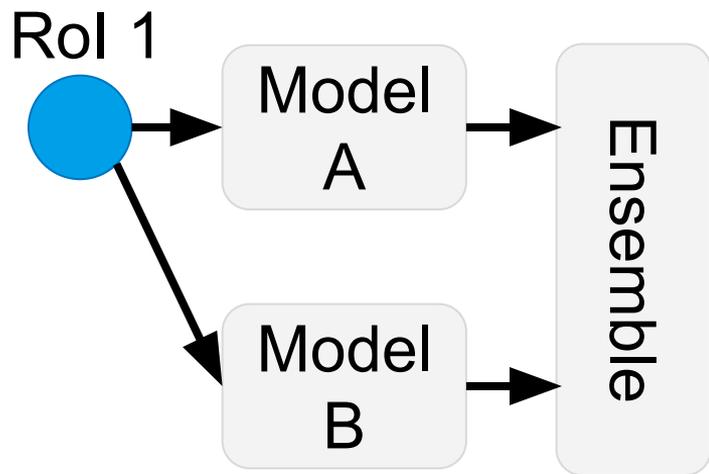
## Faster R-CNN

1. Forward backbone
2. RPN
3. GT assignment
4. Loss

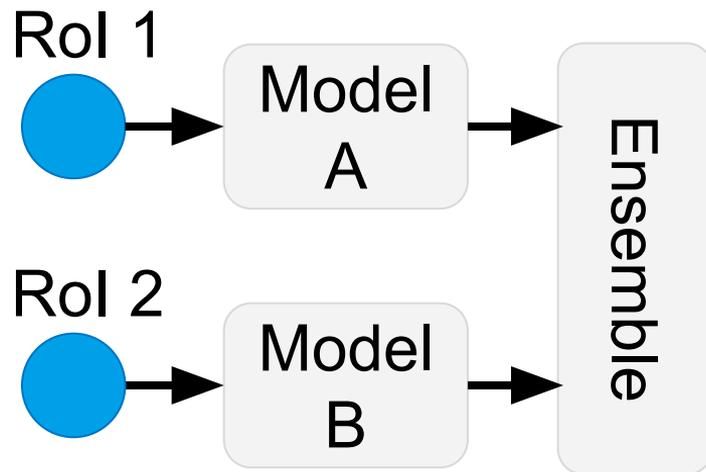
## Fast R-CNN

1. GT assignment  
(in parallel)
2. Forward backbone
3. Loss

# Variation of Rols is Important



One Rol, two models:  
71.0mAP

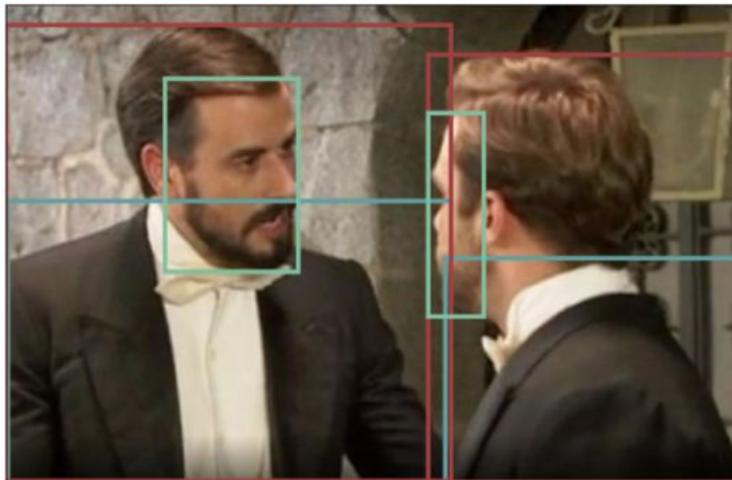


Two Rol, two models:  
71.8mAP (+0.8)

# Federated Annotations

Categories are grouped into three groups

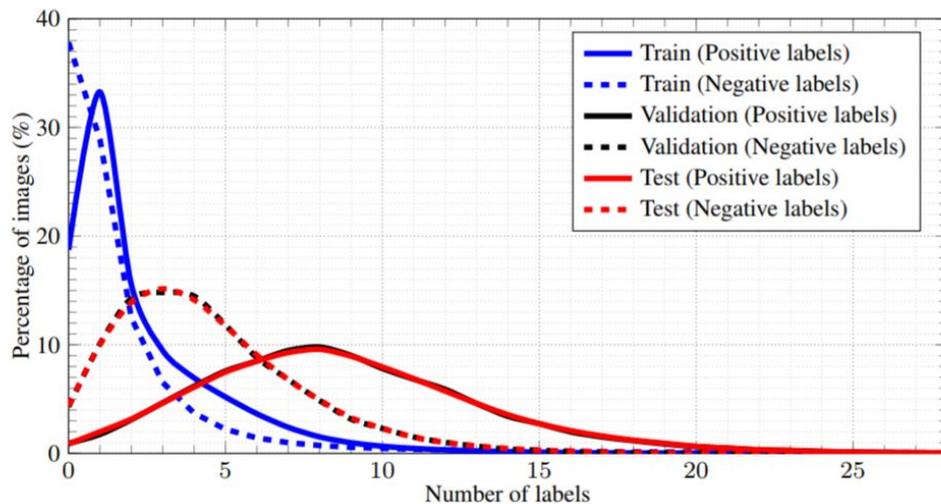
1. Positively Verified:  
always annotated
2. Negatively Verified:  
never exist
3. Unverified



Man  
Suit  
Human face

Dress  
Tie  
Footwear  
Woman

# Federated Annotations Statistics



**Fig. 12 Human-verified image-level labels:** Histogram of number of labels per image.

- On average <10 categories are verified per image
- Most of the categories are unverified

Image from

A. Kuznetsova et. al., The Open Images Dataset V4 Unified image classification, object detection, and visual relationship detection at scale

# Loss for Federated Dataset

Binary Cross Entropy Loss

$$\mathcal{L}_{cls} = - \sum_i \sum_c w_{ic} \log p_{ic}$$

Assigned category:  $w_{ic} = +1$

Unassigned verified:  $w_{ic} = -1$



Man  
Suit  
Human face

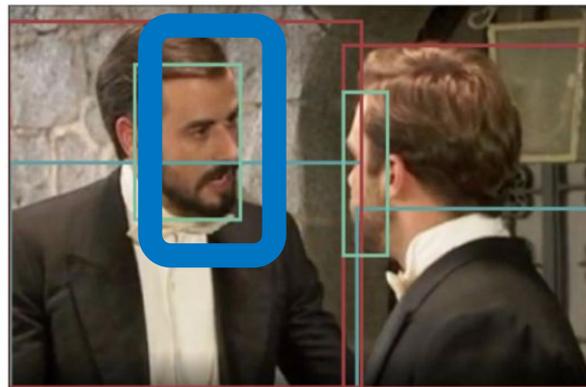
Dress  
Tie  
Footwear  
Woman

# Loss for Unverified Categories

Binary Cross Entropy Loss

$$\mathcal{L}_{cls} = - \sum_i \sum_c w_{ic} \log p_{ic}$$

- Set  $w_{ic} = -1$ :
  - Treat unverified as negatives
  - Same as ignoring negative verification
- Set  $w_{ic} = 0$  (better):
  - Ignore unverified categories



Man  
Suit  
Human face

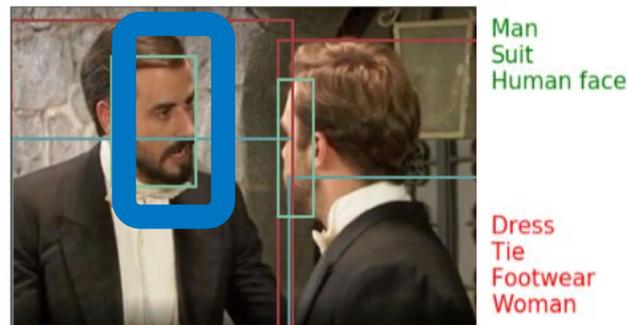
Dress  
Tie  
Footwear  
Woman

# Loss for Unverified Categories

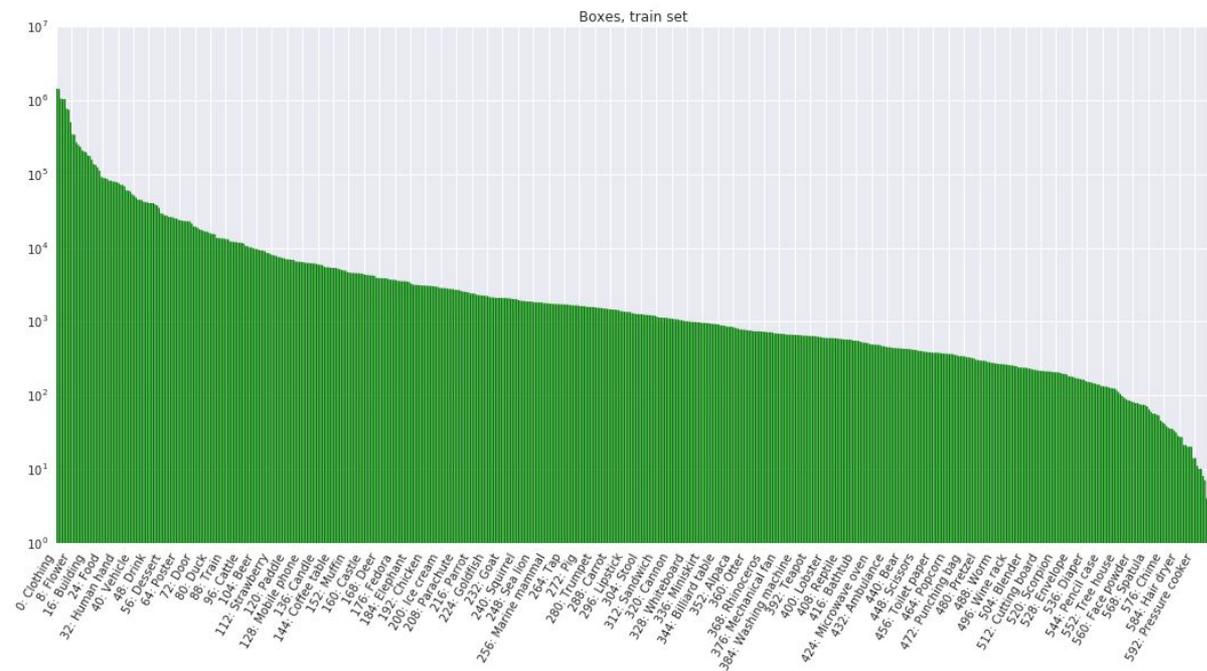
Binary Cross Entropy Loss

$$\mathcal{L}_{cls} = - \sum_i \sum_c w_{ic} \log p_{ic}$$

- Set  $w_{ic}$  based on pseudo label [1]
  - The previous work generally shows an improvement from the baseline
  - Many parameters to tune



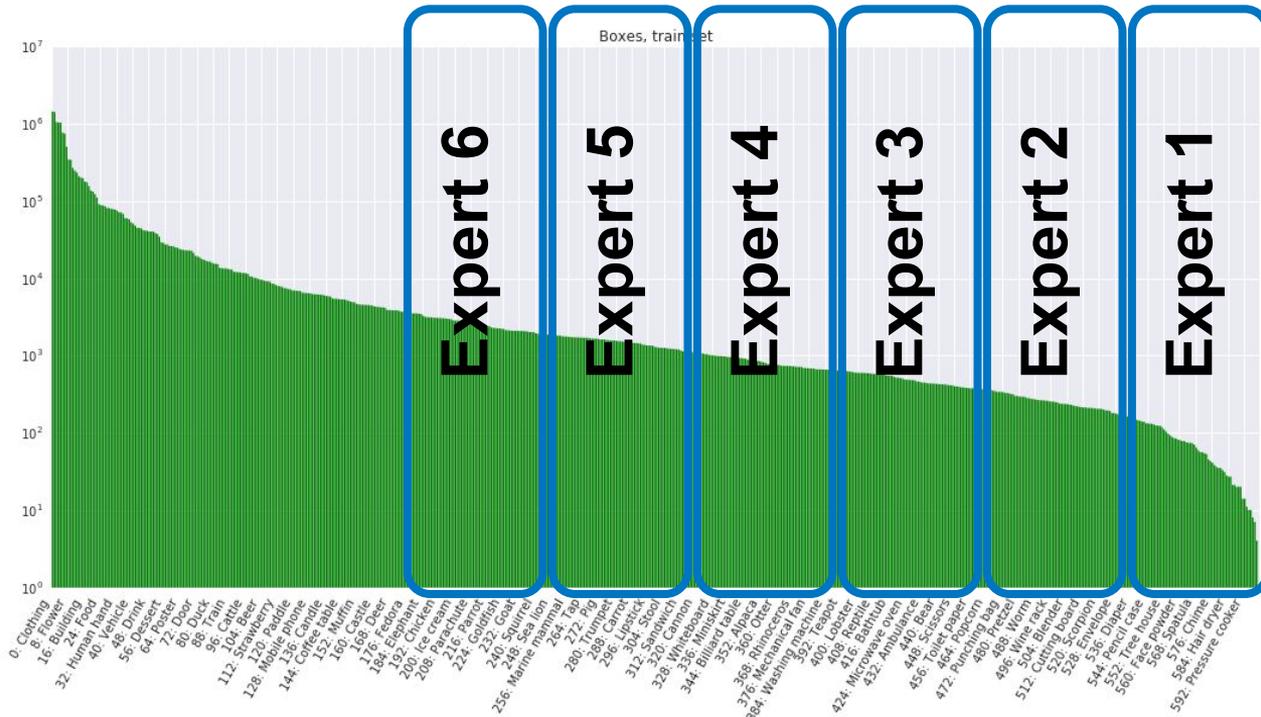
# Huge Class Imbalance



**Pressure Cooker: 17 images**  
**Person: 800k images**

**238 classes appear in <1000 images**

# Expert Models by Occurrence

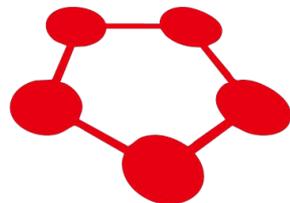


# SW Setup



CuPy

GPU Array Library



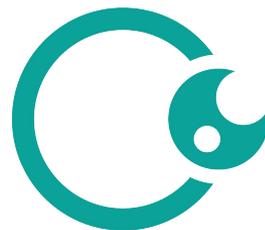
Chainer

Define-by-run DL framework



Chainer MN

Multi-Node Distributed DL



Chainer CV

CV with DL made easy

# Details

- FPN with SENet154 as backbone
- Cosine learning rate schedule with warmup
- Momentum SGD with momentum 0.9
- Batchsize was 240 with 120 GPUs. Training finishes in one and half days.
- NMS with IoU threshold set to 0.5
- AutoAugment Policy v0 was used.
- Multi-scale training. The length of the shorter edge was between [650, 1056]

# Final Submission

- Two Fast R-CNN models trained on 500 detection categories. One of them is trained for 16 epochs and another is trained for 24 epochs.
- 47 Fast R-CNN expert models. On average each expert predicts 43 categories.
- Faster R-CNN models. We used predictions from Faster R-CNN models trained in the preliminary experiments. Some of them are from the last year's submission [11, 2].

# Results

	val	public test	private test
Full (16 epochs)	70.62	51.33	46.33
Full (24 epochs)	71.02	51.80	47.17
Ensemble of above two	71.61	52.67	47.32
+ Expert Models	75.74	54.55	50.55
+ Remove small masks		54.83	50.76
+ Faster R-CNN models		<b>55.33</b>	<b>51.10</b>

# Conclusion

- We won 3rd place in the competition
- OID is a challenging, but powerful dataset

We are hiring!!!

<https://www.preferred-networks.jp/en/jobs>



Team members  
with Chainer T-shirts

