

PFDet: 2nd Place Solution to OIC Object Detection Track



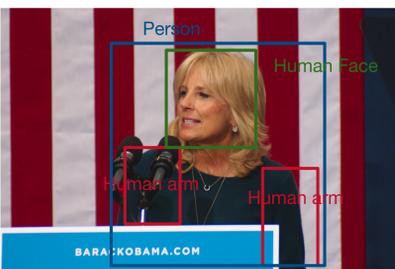
Takuya Akiba[†], Tommi Kerola[†], Yusuke Niitani[†], Toru Ogawa[†],
Shotaro Sanot[†], Shuji Suzuki[†]
(Preferred Networks, Inc.) [†]equal contribution

Summary

- Massive Parallelism:** We train an object detector with batchsize of 512
- Co-occurrence Loss:** We propose a loss that uses class-wise statistics to learn without densely annotated classes
- Expert Models:** To tackle huge class imbalance, we propose to use expert models

Unverified Labels

Instances of classes that are not verified are not annotated



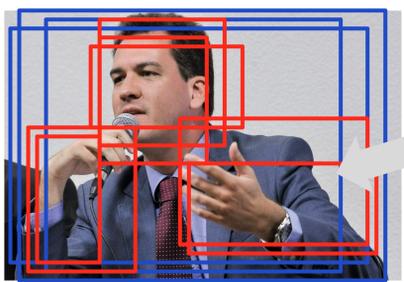
Some human parts are annotated



No human parts are annotated

Co-occurrence Loss

Ignore classification loss for a *part class* if a proposal is inside the *subject class*



Proposals used during training-time.
Blue: positive, Red: negative

Ignore: Arm, Face
Negative: car, etc...

Overall improvement of **+9.1AP**
For *human part classes*, **+22.7AP**

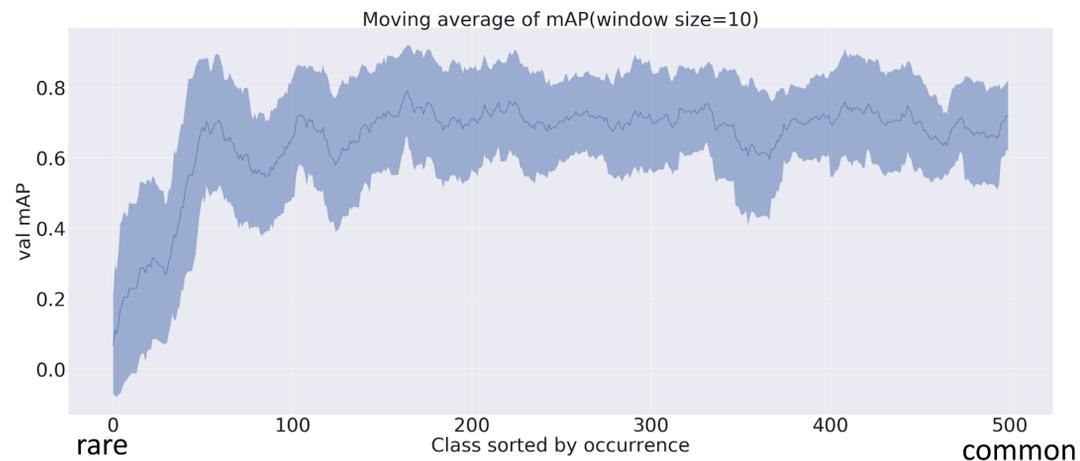
Table 4: Ablative study of co-occurrence loss on classes that can be ignored by the loss. The scores are AP calculated on the validation set of the dataset.

	Arm	Ear	Nose	Mouth	Hair	Eye	Beard	Face	Head	Foot	Leg	Hand	Glove	Hat	Dress	Fedora
Baseline	40.9	17.5	34.7	21.4	63.8	27.3	55.5	82.7	55.1	50.7	41.6	32.3	63.4	64.9	70.6	67.0
Co-occurrence	55.2	62.6	69.6	55.2	74.7	64.0	76.8	91.4	78.9	59.5	54.4	53.6	60.8	69.0	73.9	70.3

	Footwe.	Sandal	Boot	Sports.	Coat	Sock	Glasse.	Belt	Helmet	Jeans	High h.	Scarf	Swimwe.	Earrin.	Bicycl.	Shorts
Baseline	61.9	53.6	61.6	52.9	58.0	70.6	74.9	66.8	80.2	62.7	76.6	71.6	63.4	82.0	75.1	69.7
Co-occurrence	68.5	58.9	57.9	61.2	73.3	67.1	85.4	61.9	82.4	77.6	78.8	75.8	63.4	86.1	75.8	75.4

	Baseba.	Minisk.	Cowboy.	Goggles	Jacket	Shirt	Sun ha.	Suit	Trouse.	Brassi.	Tie	Licens.	Wheel	Tire	Handle	Average
Baseline	67.2	62.5	65.0	79.3	69.5	70.9	61.3	83.7	62.5	82.6	84.7	72.1	48.3	49.4	41.1	61.1
Co-occurrence	62.2	58.7	73.3	86.7	74.3	81.6	66.4	87.0	69.8	74.5	91.5	74.6	66.4	69.6	46.2	70.3

Expert Models for Rare Classes

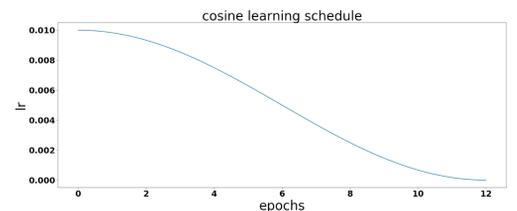


We train a set of expert models for a small subset of classes exclusively

	Index 11-100	Index 101-250	Index 251-350
Full	51.9	70.5	70.9
Class10 experts	65.6	73.1	66.3
Class40 experts	61.0	66.3	50.9

Basic Architecture

- FPN
- Sigmoid Loss
- Multi-node BN
- NMW
- Cosine LR Schedule
- Global Context



- Additional FPN Block
- PSP
- Context Head

$$\eta = \eta_0 \frac{\cos(\% \text{ of progress} \times \pi) + 1}{2}$$

Setup



With 512 V100 (32GB) GPUs, achieved **83%** scalability efficiency

Results

	validation mAP	val mAP	Public LB	Private LB
Baseline (FPN with SE-ResNeXt-101)	60.0			
+ multi-scale training	60.3 (+0.3)			
+ PSP and add BN to head	60.4 (+0.1)			
+ Cosine Annealing	63.4 (+3.0)			
+ Add FPN scale	64.5 (+1.1)			
+ Co-occurrence loss	65.2 (+0.7)			
+ 16 epochs	65.8 (+0.6)			
+ Context head	66.0 (+0.2)			
+ SENet-154 and additional anchors	67.5 (+1.5)			
Single best model		69.95	55.81	53.43
+ class20 experts		71.73	59.34	55.87
+ class10 experts		72.33	60.19	56.61
+ All the others except COCO		73.98	61.83	57.97
+ COCO		74.07	62.34	58.48
+ class-weight ensemble			62.88	58.63
Competition winner			61.71	58.66

