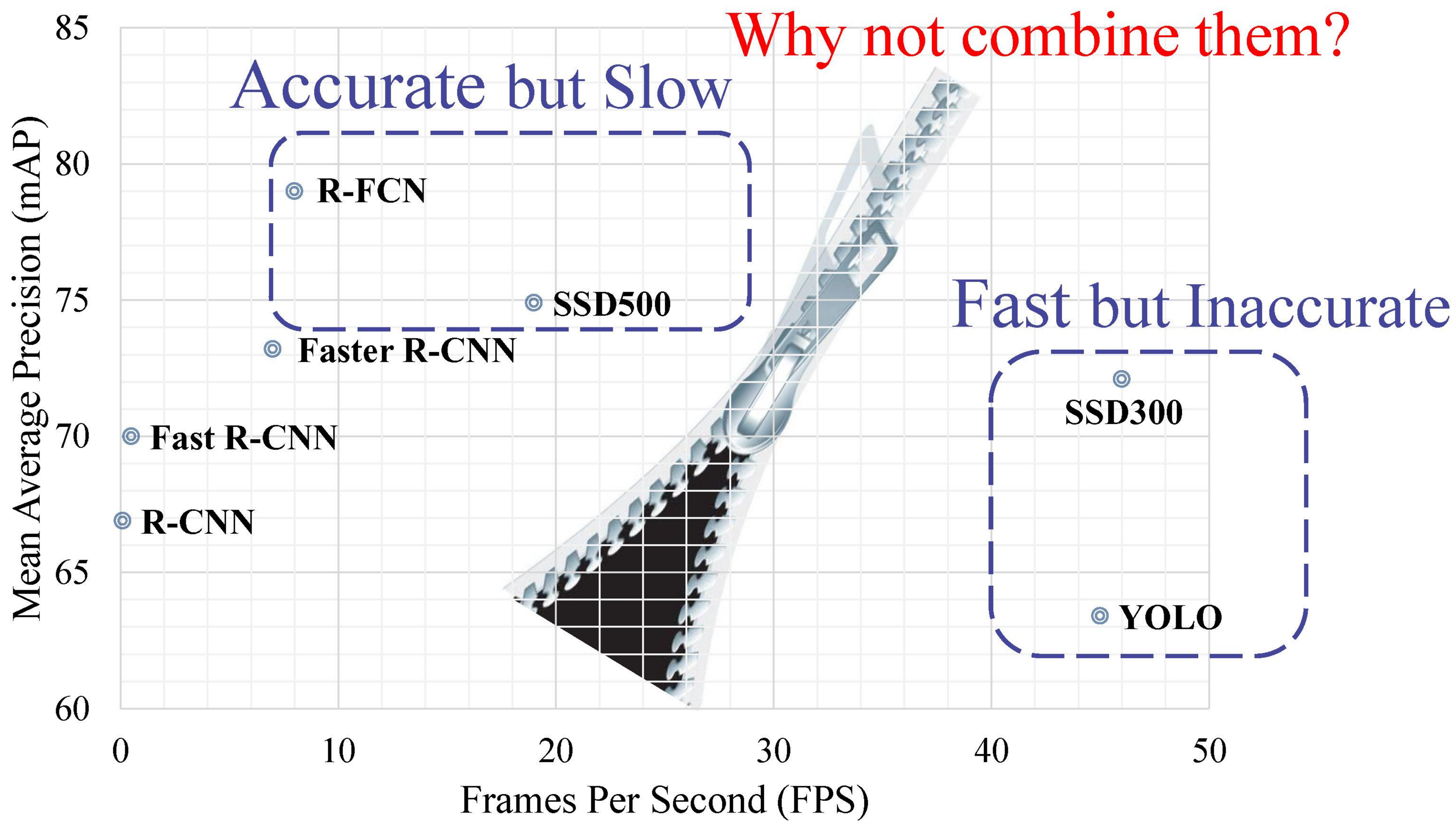
Adaptive Feeding: Achieving Fast and Accurate Detections by Adaptively Combining Object Detectors

Hong-Yu Zhou, Bin-Bin Gao, Jianxin Wu

National Key Laboratory for Novel Software Technology, Nanjing University, China

Learning And Mining from DatA
http://lamda.nju.edu.cn

A Brief Introduction 1



Motivation: Fast AND Accurate 2

✓ In most images, the fast detector is as precise as the accurate one (and in few cases it is even better).

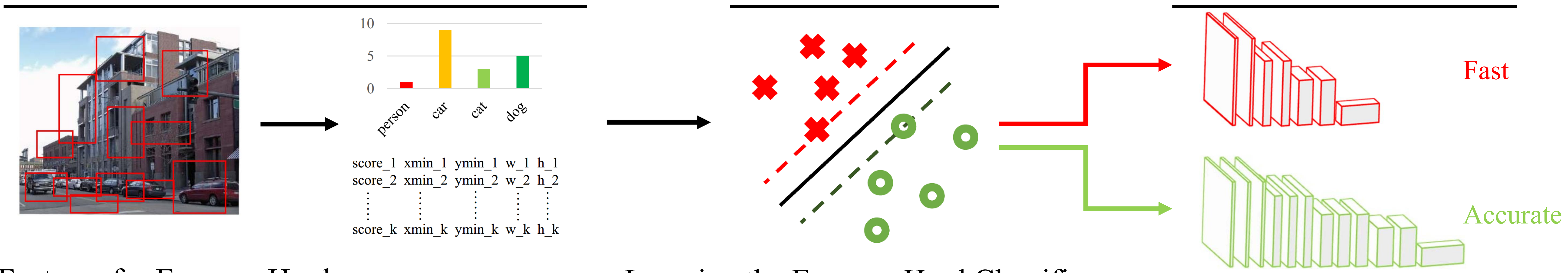
- **Easy:** Fast ≥ Accurate
- **Hard:** Fast < Accurate

✓ Let Fast model deal with **Easy** images while the Accurate model should focus on **Hard** ones. **So we can speed up the detection process while keep accuracy**.

✓ In Pascal VOC datasets, the percentage of Easy examples is large (around 80%).

An Effective Framework 3

Instance Proposals | Easy or Hard ? | Fast or Accurate ?



Features for Easy vs. Hard

- ➢ We take instance proposals as features.
- ➢ Feature generation process should be extremely fast to maintain speed.
- ➢ Features are supposed to be strong enough in order to discriminate Hard images to keep accuracy.
- ✓ Best feature form: **class hist + (conf + coords) × top-k**. **A naïve combination is already stronger than SSD500**.

Learning the Easy vs. Hard Classifier

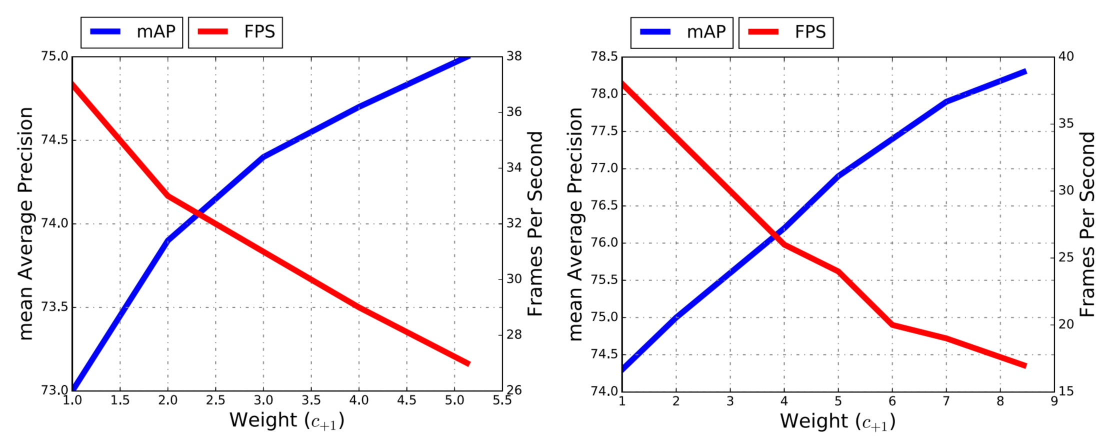
Imbalanced SVM Classification

➢ Since the number of Easy images is much larger than that of Hard ones, the problem is imbalanced.

$$\min_{\vec{w},b} \ \frac{1}{2}\vec{w}^T\vec{w} + C\sum_{i=1}^{n} c_{y_i}\varepsilon_i$$

$$s.t. \quad y_i(\vec{w}^T\vec{x}_i + b) \ \geq \ 1-\varepsilon_i, \varepsilon_i \geq 0, \ 1 \leq i \leq n.$$

$y_i \in \{-1,+1\}$ is image's label (Easy or Hard). A larger c_{y_i} value puts more emphasis on the correct classification of hard images, and hence will in general lead to higher recall.



- Impact of sampling weights on mAP and FPS. The experiments are performed on VOC07 test.

Experimental Results 4

Pascal VOC 2007

SUR: Speed-Up Ratio. **DmAP:** Decreased mAP based on accurate model. **A**: the accurate mode. **F**: the fast mode. **W**: the sampling weight.

Method	W	mAP	FPS	SUR	DmAP
SSD300	-	72.1	46	-	-
SSD400	-	74.0	32	-	-
SSD500	-	74.9	19	-	-
Simple Ensemble	-	73.0	19	-	-
R-FCN	-	79.0	8	-	-
300-500-A	5.13	75.0	27	42%	-0.1
300-500-F	3	74.4	33	74%	0.5
300-R-FCN-A	8.43	78.3	17	113%	0.7
300-R-FCN-F	5	76.9	24	200%	2.1

MSCOCO 2015 test-dev

Method	W	test	FPS	AP	AP^{50}	AP^{75}	AP^S	AP^M	AP^L
SSD300	-	test-dev	46	20.8	38.0	20.5	3.9	18.5	38.7
SSD500	-	test-dev	19	24.4	43.7	24.7	7.2	25.3	40.1
R-FCN	-	test-dev	8	28.6	48.8	30.1	8.8	31.4	44.1
300-500-A	1.16	test-dev	25	23.7	42.7	23.9	6.6	23.7	39.8
300-500-F	2	test-dev	31	23.0	41.6	23.1	6.0	22.2	39.5
300-R-FCN-A	1.07	test-dev	15	27.0	47.4	29.2	7.5	29.9	43.1
300-R-FCN-F	2	test-dev	21	26.2	46.7	27.9	6.3	28.8	41.9

Statistics on COCO minival

- ➢ When IoU increases, the number of Hard images is up.
- ➢ **But our approach still works** with different sampling weights.

Fast	Accurate	Dataset	IoU = 0.5	
			⩾ 0 (Easy)	< 0 (Hard)
SSD300	SSD500	minival2014	53.6%	46.4%
SSD300	R-FCN	minival2014	51.7%	48.3%

SVM Visualization 5

Weights of Each Group in SVM

Method	dataset	class	conf	xmin	ymin	width	height
300-500	voc	0.22	1.46	0.08	0.02	-0.55	-0.23
300-R-FCN	voc	0.25	1.58	0.13	0.05	-0.89	-0.12
300-500	coco	0.31	1.46	-0.02	-0.05	-0.37	-0.33
300-R-FCN	coco	0.36	1.48	-0.05	-0.01	-0.60	-0.18

✓An image with many objects might be hard for a detector.

✓Large proposal hints Easy images.

✓Easy images prefer shorter proposals (w>h) while Hard images like taller instances.

✓Positions of proposals have small impacts.

ICCV 17

Project Page | PDF File