Point in, Box out: Beyond Counting Persons in Crowds

Yuting Liu[1], Miaojing Shi[2], Qijun Zhao[1], Xiaofang Wang[2]

[1]Biometrics Research Laboratory, College of Computer Science, Sichuan University

[2]Univ Rennes, Inria, CNRS, IRISA

INVENTEURS DU MONDE NUMÉRIQUE

CVPR 2019, Long Beach, CA

Crowd Counting

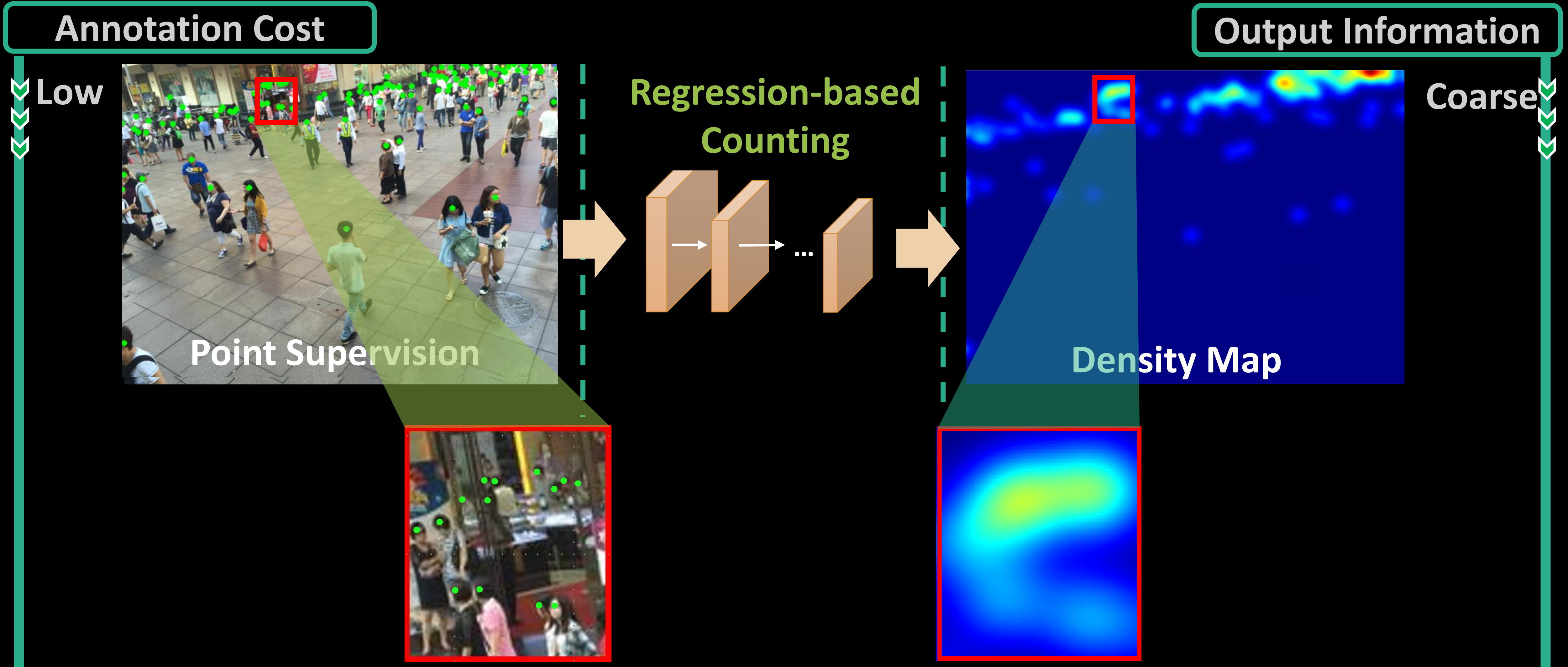
#persons ?

Related Works

Regression-based Counting

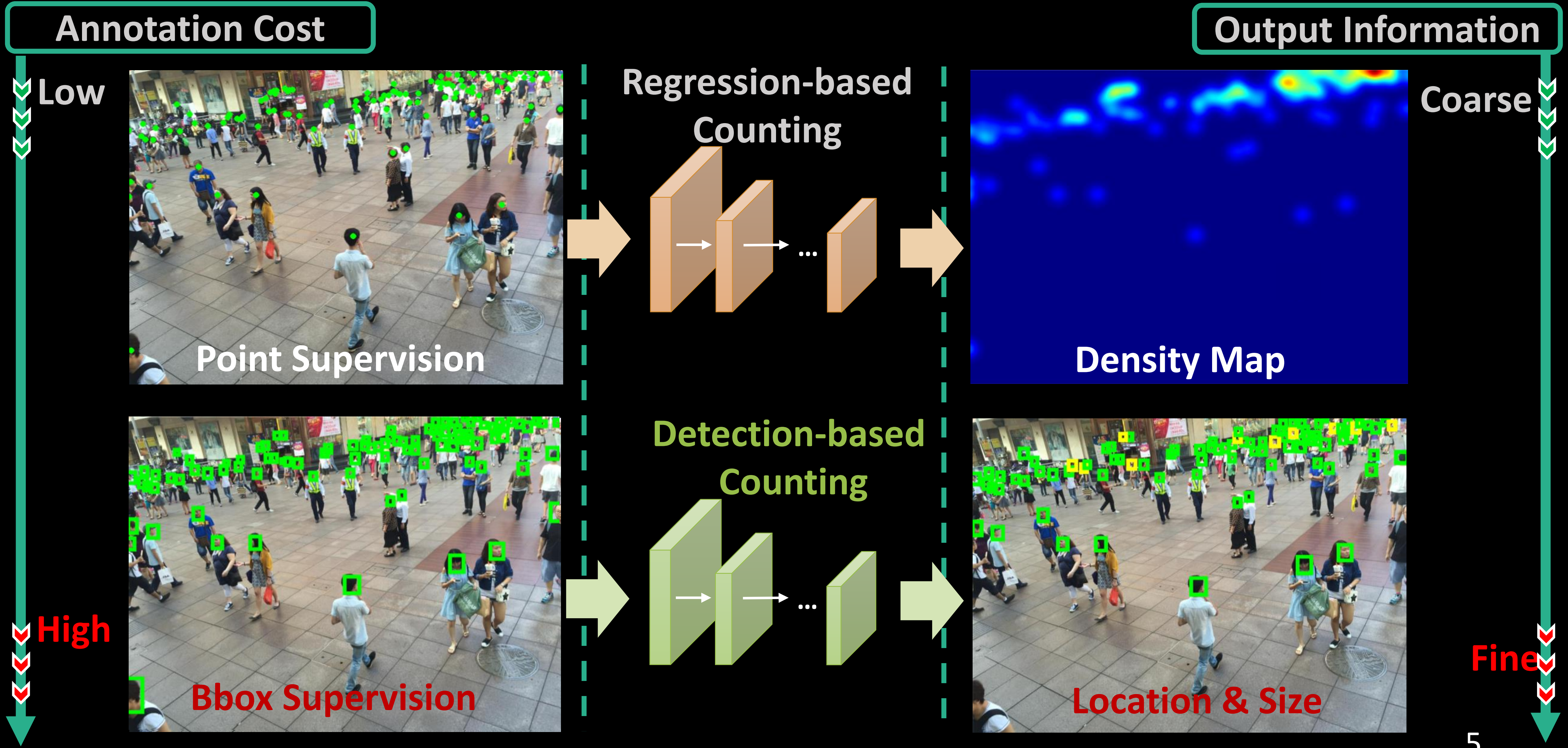
Detection-based Counting

Related Works



Can not detect the persons
(Tracking, Recogniton, ...)

Related Works



Our Goal: Point In, Box Out

Annotation Cost

Low

Our proposed PSDDN:

- Pseudo GT Initialization
- Online Pseudo GT Updating
- Locally-constrained Regression Loss
- Curriculum Learning

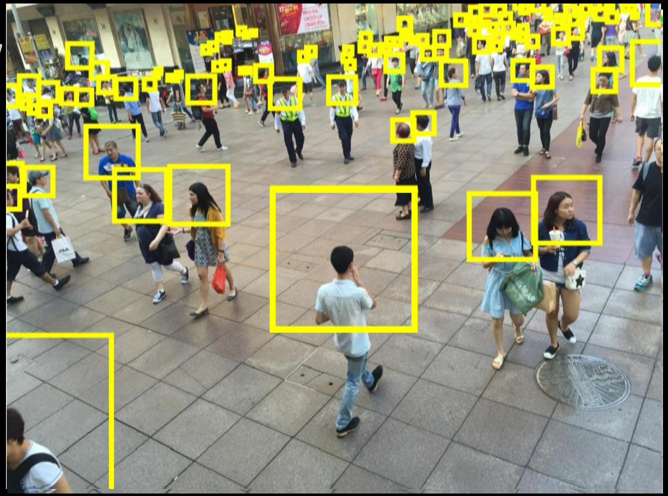
Output Information

Fine

Our Approach: Overview

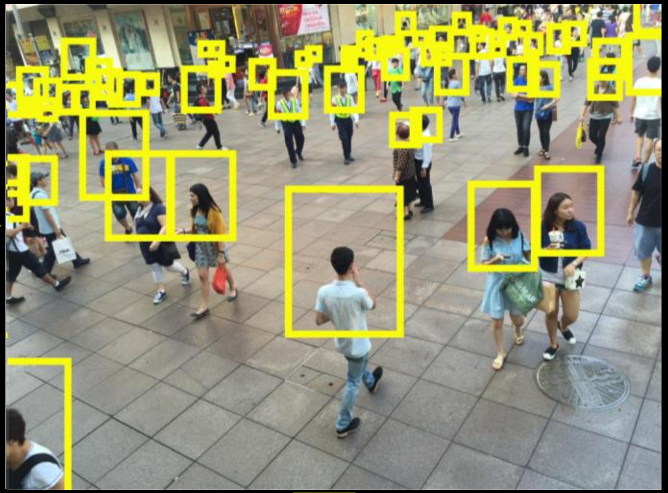
Annotation Cost

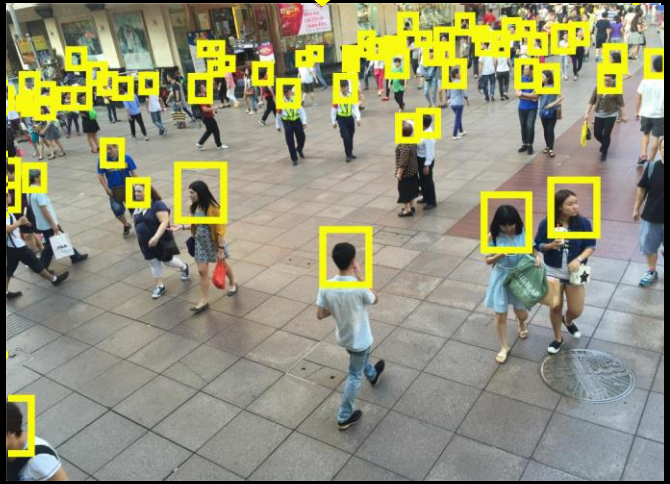
Low



Pseudo GT Initialization

Pseudo GT Updating

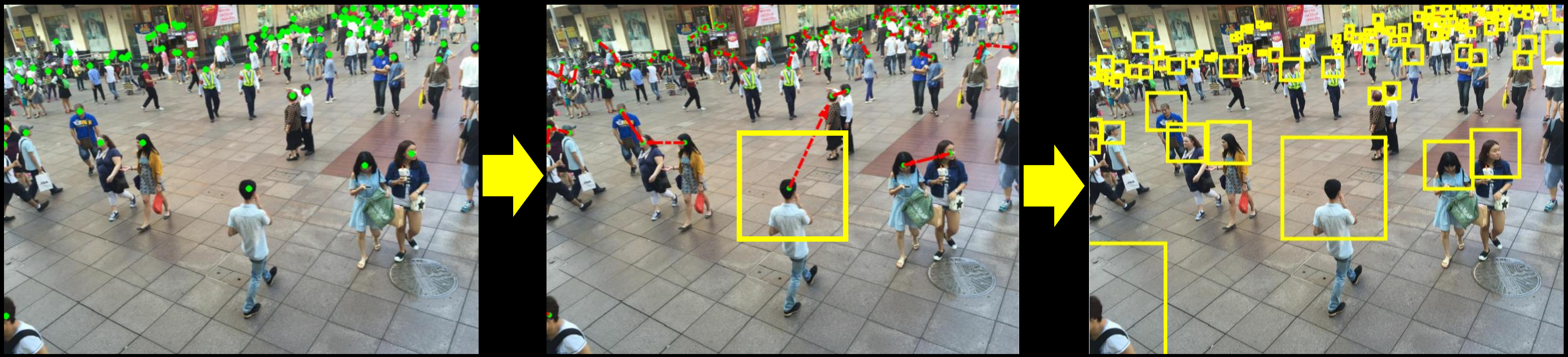




Output Information

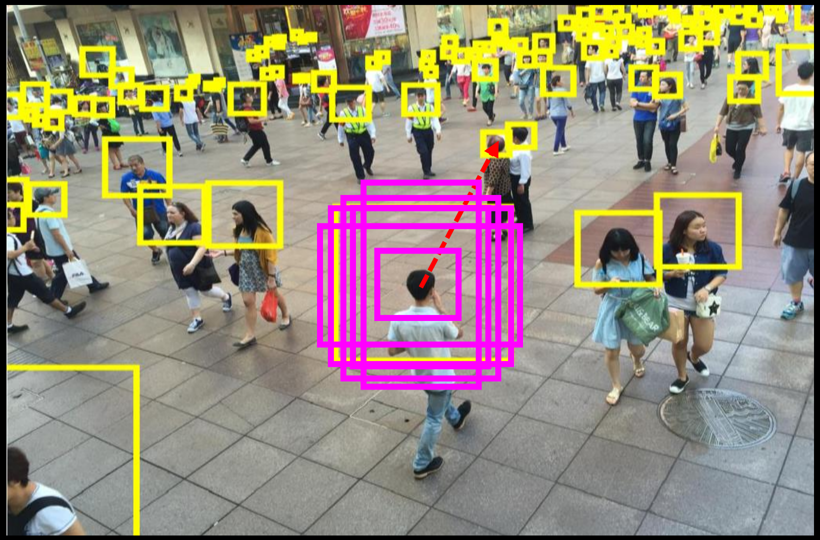
Fine

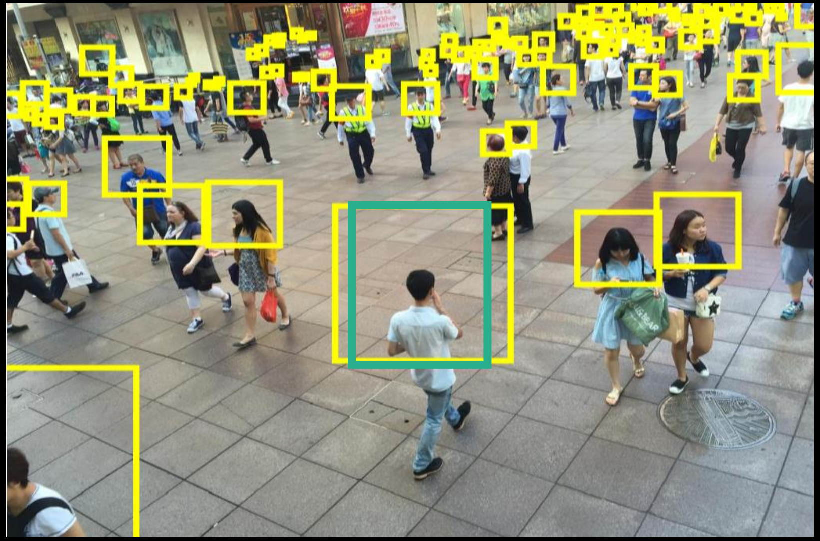
Our Approach: Pseudo GT Initialization



Nearest neighbor distances indeed reflects head **size information**

Our Approach: Online Pseudo GT Updating





1) Select positive anchors:

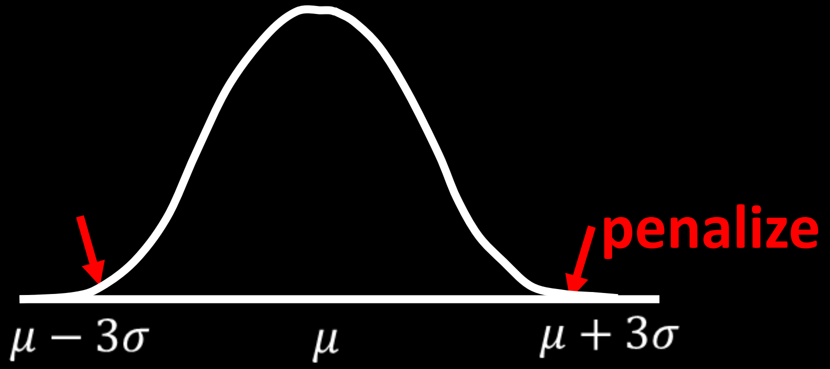
$$\boldsymbol{IOU(pos(g^t), g^t) > 0.7 \;\&\&}$$
$$\boldsymbol{SIZE((pos(g^t)) < d(g, NN_g)}$$

2) $\boldsymbol{g^{t+1}}$ is from those $pos(g^t)$ that has the **highest detection score**

(g^t: Pseudo GT at t^{th} iteration; $pos(g^t)$: positive anchors of g^t; $\mathbf{d(g, NN_g)}$: distance from g to its nearest neighbor head ; $SIZE(\cdot)$: smallest side of height or width)

Our Approach: Locally-constrained Regression Loss

We **penalize** the predicted Bboxes if its size clearly **violate the observation**



Our Approach: Curriculum Learning

Crowd Density:	Sparse	→	Moderate	→	Dense
Initialize Pseudo Gt:	bigger than real	→	Moderate	→	too small
Training Difficulty:	Hard	←	Easy	→	Hard

Our Results

True positives False positives

Our Results

Dataset	SHA		SHB	
Measures	MAE	MSE	MAE	MSE
Pv0	168.6	268.3	69.8	98.1
Pv1	104.7	193.8	41.7	66.6
Pv2	89.8	169.5	19.1	42.4
Pv3(PSDDN)	85.4	159.2	16.1	27.9
PSDDN + [20]	**65.9**	**112.3**	**9.1**	**14.2**
Li et al. [20]	**68.2**	115.0	**10.6**	**16.0**
Ranjan et al. [31]	68.5	116.2	10.7	16.0
Liu et al. [24]	73.6	112.0	13.7	21.4
Liu et al. [22]	-	-	20.7	29.4
DetNet in [22]	-	-	44.9	73.2
Sindagi et al. [41]	73.6	**106.4**	20.1	30.1
Sam et al. [35]	90.4	135.0	21.6	33.4

Table 1: Counting: ablation study of PSDDN on ShanghaiTech dataset (Pv0: trained with initialized pseudo GT; Pv1: Pv0 + pseudo GT updating; Pv2: Pv1 + our regression loss; Pv3: Pv2 + curriculum learning).

Dataset	Pv0	Pv1	Pv2	Pv3 (PSDDN)
SHA	0.308	0.491	0.539	0.554
SHB	0.015	0.241	0.582	0.663

Table 2: Detection: ablation study of PSDDN on ShanghaiTech dataset. AP is reported.

Methods	Annotations	WiderFace		
		easy	medium	hard
Avg. BB	points(test)+ mean size	0.002	0.083	0.059
FR-CNN (ps)	points(train) + mean size	0.008	0.183	0.108
FR-CNN (fs)	bounding boxes (train)	**0.840**	**0.724**	0.347
PSDDN	points(train)	0.605	0.605	**0.396**

Table 3: Detection results on WiderFace. (AP)

Methods	GAME0	GAME1	GAME2	GAME3	AP
Victor et al. [19]	13.76	16.72	20.72	24.36	-
Onoro et al. [27]	10.99	13.75	16.09	19.32	-
Li et al. [20]	**3.56**	5.49	8.57	15.04	-
PSDDN	4.79	**5.43**	**6.68**	**8.40**	**0.669**

Table 4: Counting and detection results on TRANCOS dataset. GAME and AP are reported.

Thank you!

Poster: #32