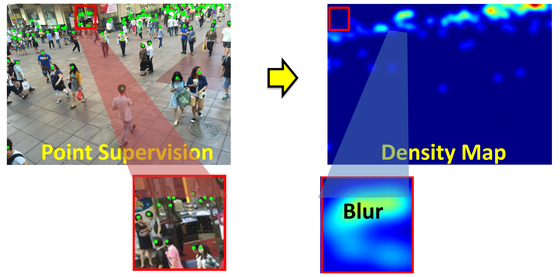
Point in, Box out: Beyond Counting Persons in Crowds

Yuting Liu[1], Miaojing Shi[2], Qijun Zhao[1], Xiaofang Wang[2]
[1]College of Computer Science, Sichuan University [2]Univ Rennes, Inria, CNRS, IRISA

CVPR LONG BEACH CALIFORNIA June 16-20, 2019

Motivation

➢ **Regression-based Counting**



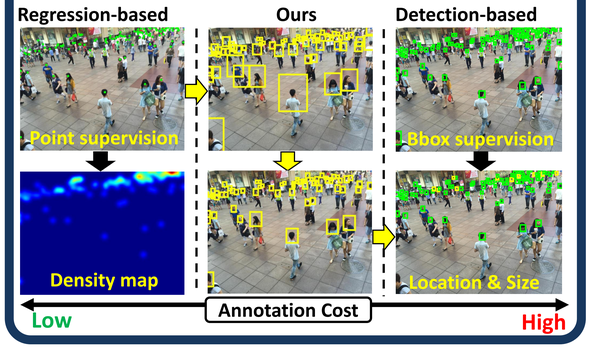
Missing **location & size** information

➢ **Detection-based Counting**

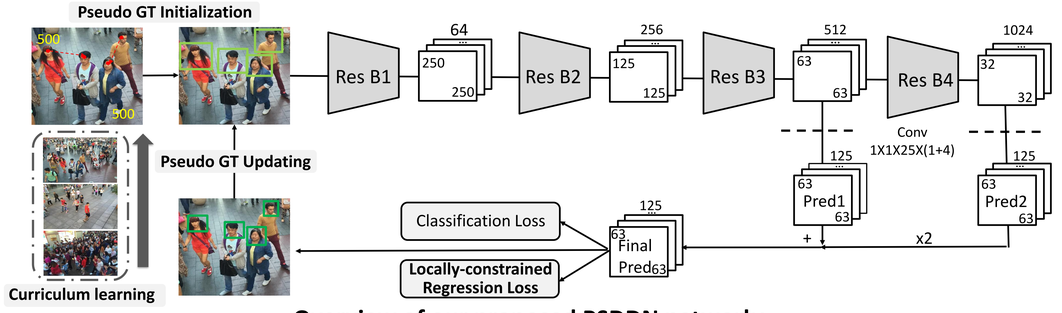
Annotation cost: **high** fine output information

➢ **Ours: Point in, Box out**

Mine latent information from point annotations



Method



Overview of our proposed PSDDN network:
One-shot anchor-based detection network (multi-scale Training & testing scheme)

➢ **Pseudo ground truth initialization**

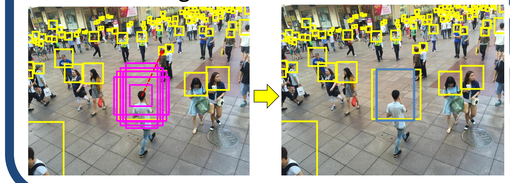
Box initialization : $h(g^0) = w(g^0) = d(g, NN_g)$

➢ **Online pseudo ground truth updating**

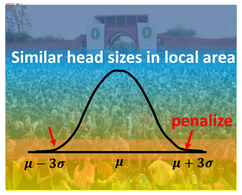
1) Select positive anchors:

$$IoU(pos(g^t)) > 0.7 \;\&\&$$
$$size((pos(g^t)) < d(g, NN_g)$$

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



➢ **Locally-constrained regression loss**



$$\mu w_i = \frac{1}{|G_i|}\sum_{mn\in G_i} gw_{mn}$$

$$\sigma w_i = \sqrt{\frac{1}{|G_i|}\sum_{mn\in G_i}(gw_{mn} - \mu w_i)^2}$$

$$lw_{ij} = \begin{cases} \left(\widehat{gw_{ij}} - (\mu w_i + 3\sigma w_i)\right)^2 & \widehat{gw_{ij}} > \mu w_i + 3\sigma w_i \\ \left((\mu w_i - 3\sigma w_i) - \widehat{gw_{ij}}\right)^2 & \widehat{gw_{ij}} < \mu w_i - 3\sigma w_i \\ 0 & otherwise \end{cases}$$

➢ **Curriculum learning**

Train the model from images of relatively accurate and easy pseudo ground truth first

$$\mathrm{TL} = 1 - \frac{1}{|G_i|}\sum_{g\in G}\Phi(d_g|\mu,\sigma)$$

Training difficulty is defined according to image density

Results

➢ **Counting Performance (MAE & MSE)**

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

Different variants of PSDDN:
Pv0: Training with initialized fixed pseudo Gt;
Pv1: Pv0 + pseudo Gt updating;
Pv2: Pv1 + proposed regression loss;
Pv3: Pv2 + Curriculum learning;
[20]: Csrnet: regression-based method

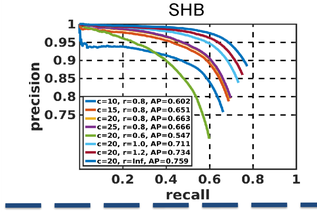
Counting	UCF		
Measures	MAE	MSE	AP
Li et al. [20]	**266.1**	**397.5**	-
Liu et al. [24]	279.6	388.9	-
Sindagi et al. [41]	295.8	320.9	-
Sam et al. [35]	318.1	439.2	-
PSDDN	359.4	514.8	0.536

➢ **Detection Performance (AP)**

True positives: IoU / if no GT BB:

- $d(g, \hat{g}) < c$
- $size(\hat{g}) < r \times d(g, NN_g)$

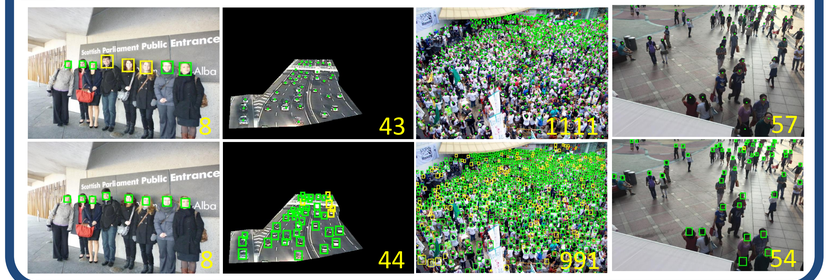
Effects of different c and r



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

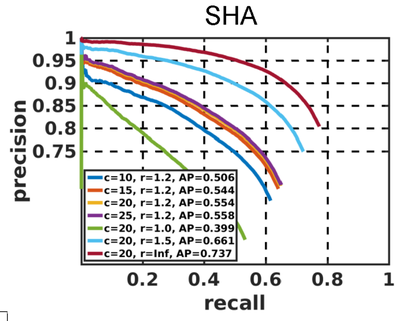
Methods	Annotations	WiderFace		
		easy	medium	hard
Avg. BB	points(test)+ mean size	0.002	0.083	0.059
FR-CNN (ps)	points(train) + mean size	0.008	0.183	0.108
FR-CNN (fs)	bounding boxes (train)	**0.840**	**0.724**	0.347
PSDDN	points(train)	0.605	0.605	**0.396**

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



Counting	UCF		
Measures	MAE	MSE	AP
Li et al. [20]	**266.1**	**397.5**	-
Liu et al. [24]	279.6	388.9	-
Sindagi et al. [41]	295.8	320.9	-
Sam et al. [35]	318.1	439.2	-
PSDDN	359.4	514.8	0.536

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



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**

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

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