ICPR 2020
25th INTERNATIONAL CONFERENCE ON PATTERN RECOGNITION
Milan, Italy 10 | 15 January 2021

Learning from Web Data: Improving Crowd Counting via Semi-Supervised Learning

Tao Peng, Rong Li, Shang Li, Pengfei Zhu

Tianjin University, Tianjin, China
63961 Troops of P.L.A, China
Innovation & Research Institute of HIWING Technology Academy, China

Motivation

- Dataset Problem:

  Small dataset size or low average density

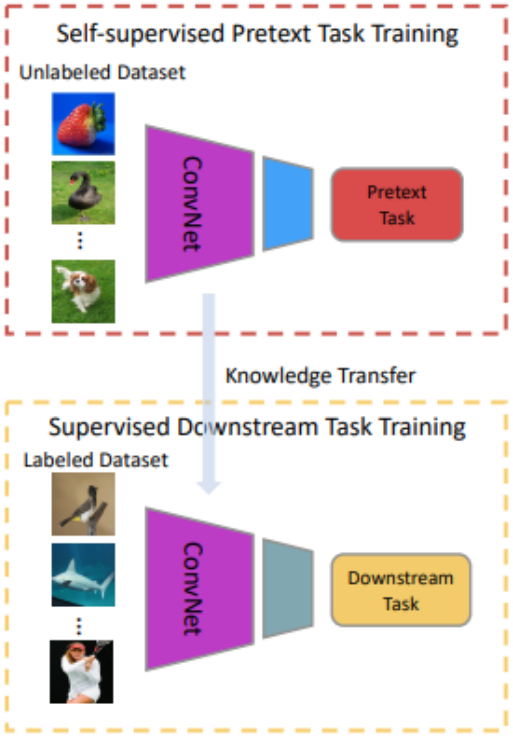
  difficult and expansive data annotation

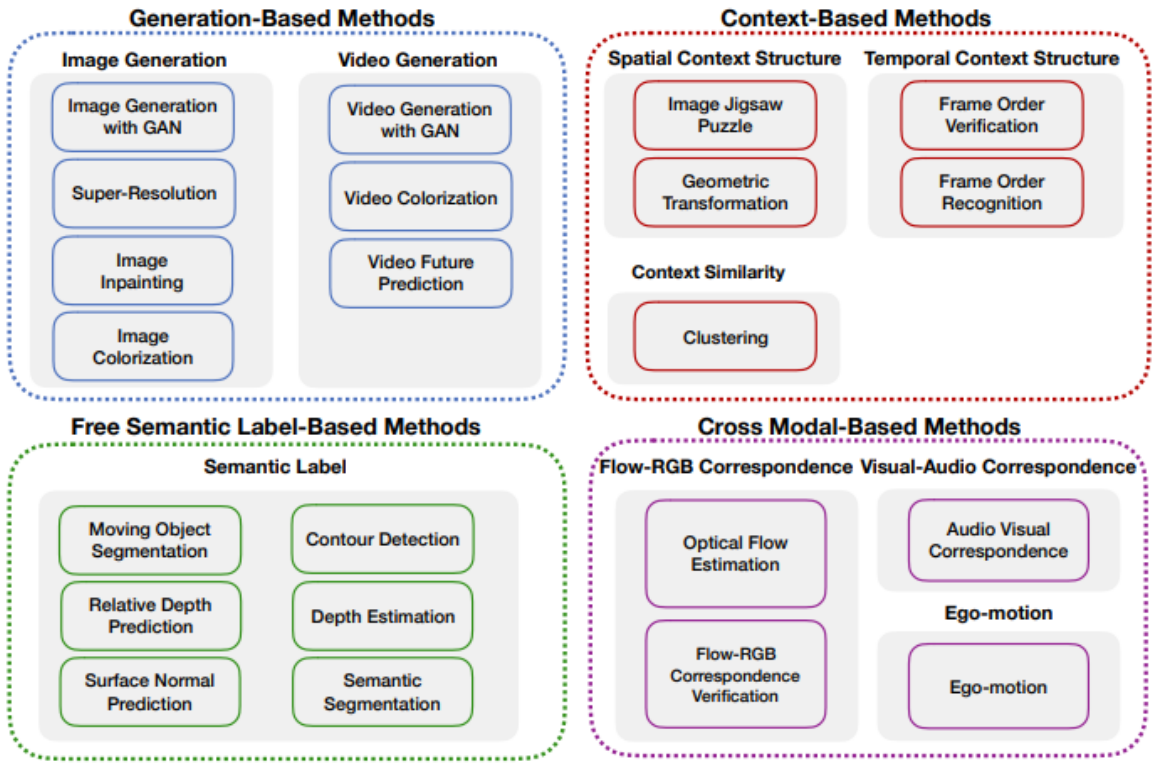
Dataset	Frames	Total	Year
UCSD [3]	2,000	49,885	2008
Mall [13]	2,000	62,315	2013
UCF_CC_50 [8]	50	63,974	2013
WorldExpo2010 [24]	3,980	199,923	2015
Shanghaitech A [9]	482	241,677	2016
Shanghaitech B [9]	716	88,488	2016
UCF-QNRF [6]	1,535	1,251,642	2018
FDST [12]	15,000	394,081	2018
Crowd Surveillance [11]	13,945	386,513	2019
GCC [14]	15,211	7,625,843	2019
NWPU-Crowd [10]	5,109	2,133,375	2020

Our strategy : Auxiliary tasks learning + Unlabeled data

Motivation

- How to use web data to boost the crowd counting models?

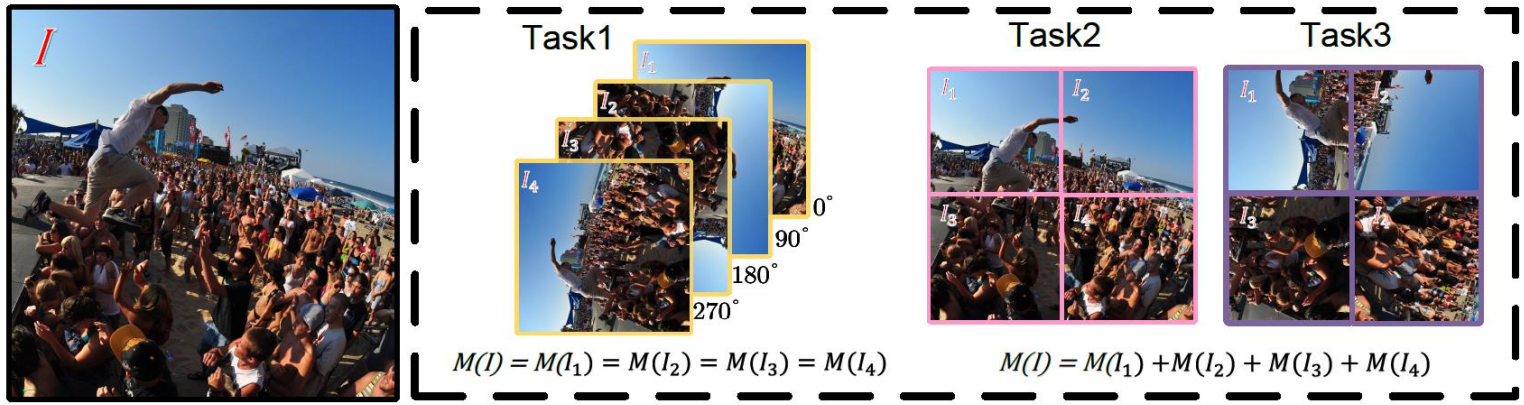




Jing L, Tian Y. Self-supervised visual feature learning with deep neural networks: A survey[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2020.

Challenge

☐ How to design a appropriate auxiliary tasks ?



Task 1:

$$L_s(\theta, w_s; D_b) = Smooth_{L_1}\left(\sum_{r,p\in R'} M\left(x_p^r\right) - M(x)\right)$$

Task 2:

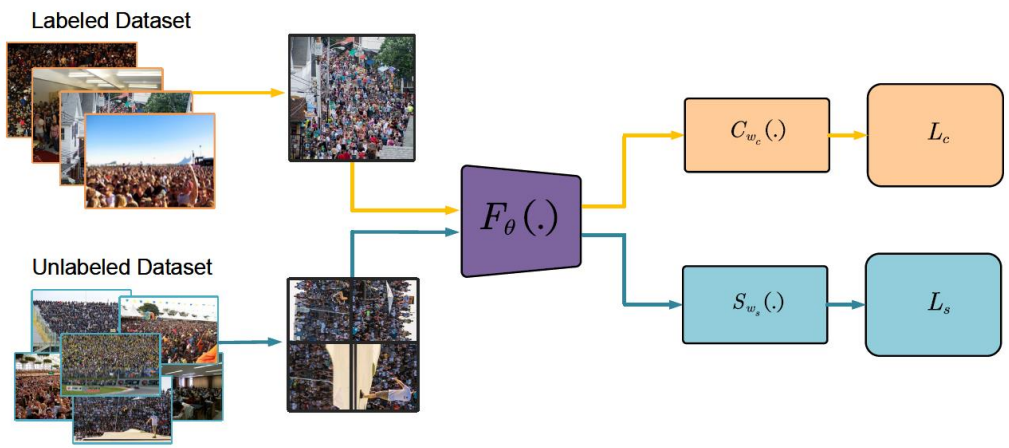
$$L_s(\theta, w_s; D_b) = Smooth_{L_1}\left(\sum_{p=1}^{4} M(x_p) - M(x)\right)$$

Task 3:

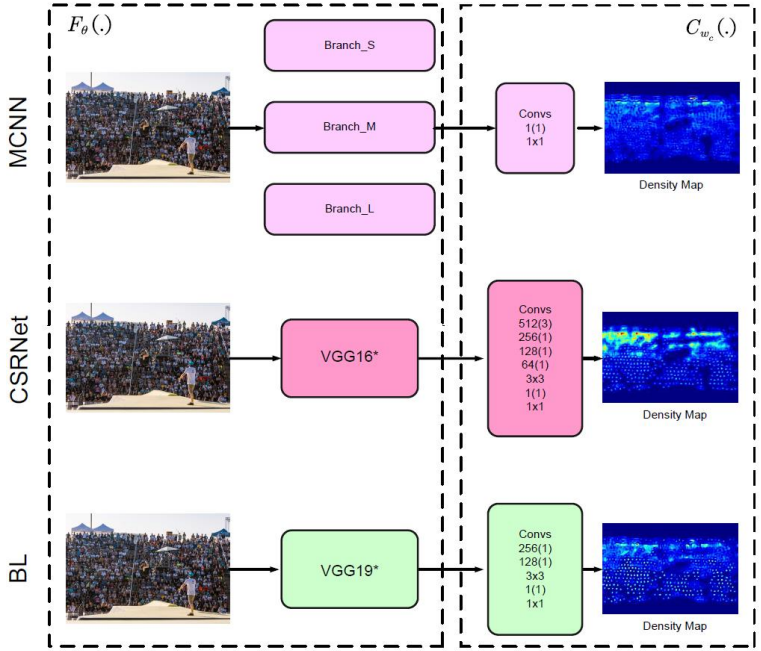
$$L_s(\theta, w_s; D_b) = \sum_{p=1}^{4}\sum_{r_i,r_j\in R} Smooth_{L_1}\left(M\left(x_p^{r_i}\right) - M\left(x_p^{r_j}\right)\right)$$
$$+\omega \sum_{r_i,r_j\in R} Smooth_{L_1}\left(\sum_{p=1}^{4} M\left(x_p^{r_i}\right) - M\left(x_p^{r_j}\right)\right)$$

Framework

What is semi-supervised multi-task framework ?



The semi-supervised multi-task framework combining both counting task and pretext task.



The division results of three baselines.

Experiment

Experiment with different ratios of unlabeled data

Method	Ratio	Learning Paradigm	MAE	MSE
BL	1:1	supervised	88.7	154.8
BL^{self}	1:0	self-supervised	84.1	151.3
BL_1^{semi}	1:1	semi-supervised	84.6	150.0
BL_2^{semi}	1:2	semi-supervised	83.2	145.8
BL_3^{semi}	1:4	semi-supervised	87.7	159.6
BL_4^{semi}	1:8	semi-supervised	102.3	168.5

We experimented with the impact of different ratios of unlabeled data on the overall model. We use BL model for baseline and Task3. All experiments were performed on UCF-QNRF dataset.

Performance of ablation experiments for different pretext tasks

Method	Baseline		$Baseline_1^{semi}$		$Baseline_2^{semi}$		$Baseline_3^{semi}$	
	MAE	MSE	MAE	MSE	MAE	MSE	MAE	MSE
MCNN [9]	277.0	426.0	283.6	519.7	250.8	435.1	236.2	405.8
CSRNet [31]	120.3	208.5	114.3	192.7	117.3	200.9	110.6	184.4
BL [33]	88.7	154.8	85.2	150.0	84.6	147.7	83.2	145.8

Here we evaluate the influence of different pretext tasks.

Experiment

Method	UCF-QNRF		ShanghaiTechA		ShanghaiTechB		Journal/Venue & Year
	MAE	RMSE	MAE	RMSE	MAE	RMSE	
MCNN [9]	277.0	426.0	110.2	173.2	26.4	41.3	2016–CVPR
CMTL [29]	252.0	514.0	101.3	152.4	20.0	31.1	2017–AVSS
Switching CNN [30]	228.0	445.0	90.4	135.0	21.6	33.4	2017–CVPR
L2R [23]	–	–	72.0	106.6	13.7	21.4	2018–CVPR
CSRNet [31]	120.3	208.5	68.2	115.0	10.6	16.0	2019–CVPR
TEDnet [40]	113.0	188.0	64.2	109.1	8.2	12.8	2019–CVPR
CAN [41]	107.0	183.0	62.3	100.0	7.8	12.2	2020–AAAI
DUBNet [42]	105.6	180.5	64.6	106.8	7.7	12.5	2019–ICCV
S-DCNet [43]	104.4	176.1	**58.3**	95.0	**6.7**	10.7	2019–CVPR
SFCN [14]	102.0	171.4	64.8	107.5	7.6	13	2019–ICCV
DSSINet [32]	99.1	159.2	60.6	96.1	6.9	**10.3**	2019–ICCV
BL [33]	88.7	154.8	62.8	101.8	7.7	12.7	2019–ICCV
Our*	**83.2**	**145.8**	60.7	**94.6**	7.1	10.9	2020

We made a review on the public datasets. We select BL as the baseline, use Task3 as pretext task, and set the Ratio to 1:2. We evaluate our method compared with 12 state-of-the-art methods

Thanks!

To contact:

Email : wspt@tju.edu.com