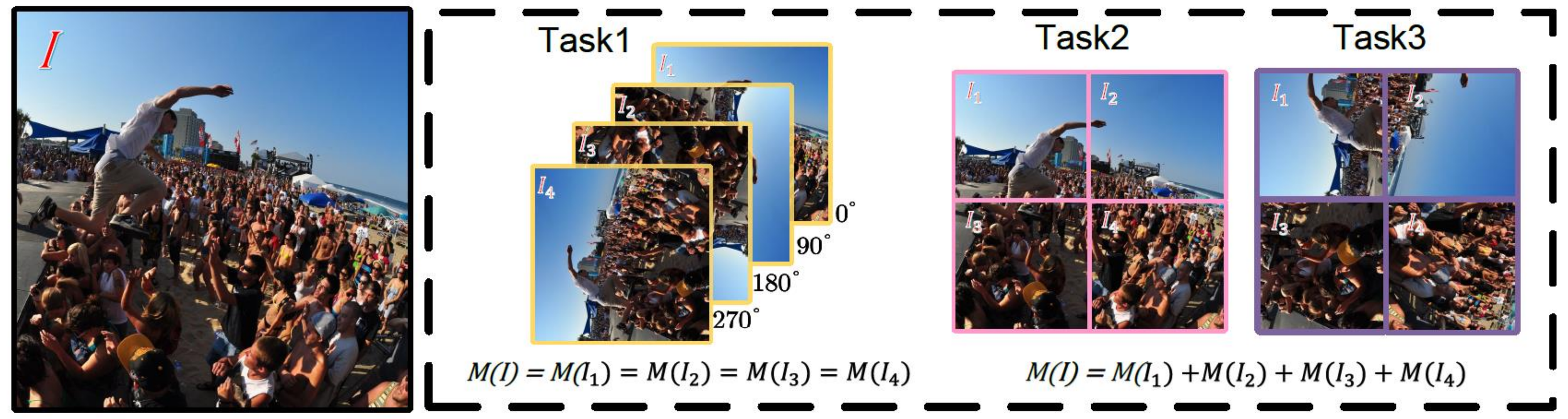
Learning from Web Data: Improving Crowd Counting via Semi-Supervised Learning

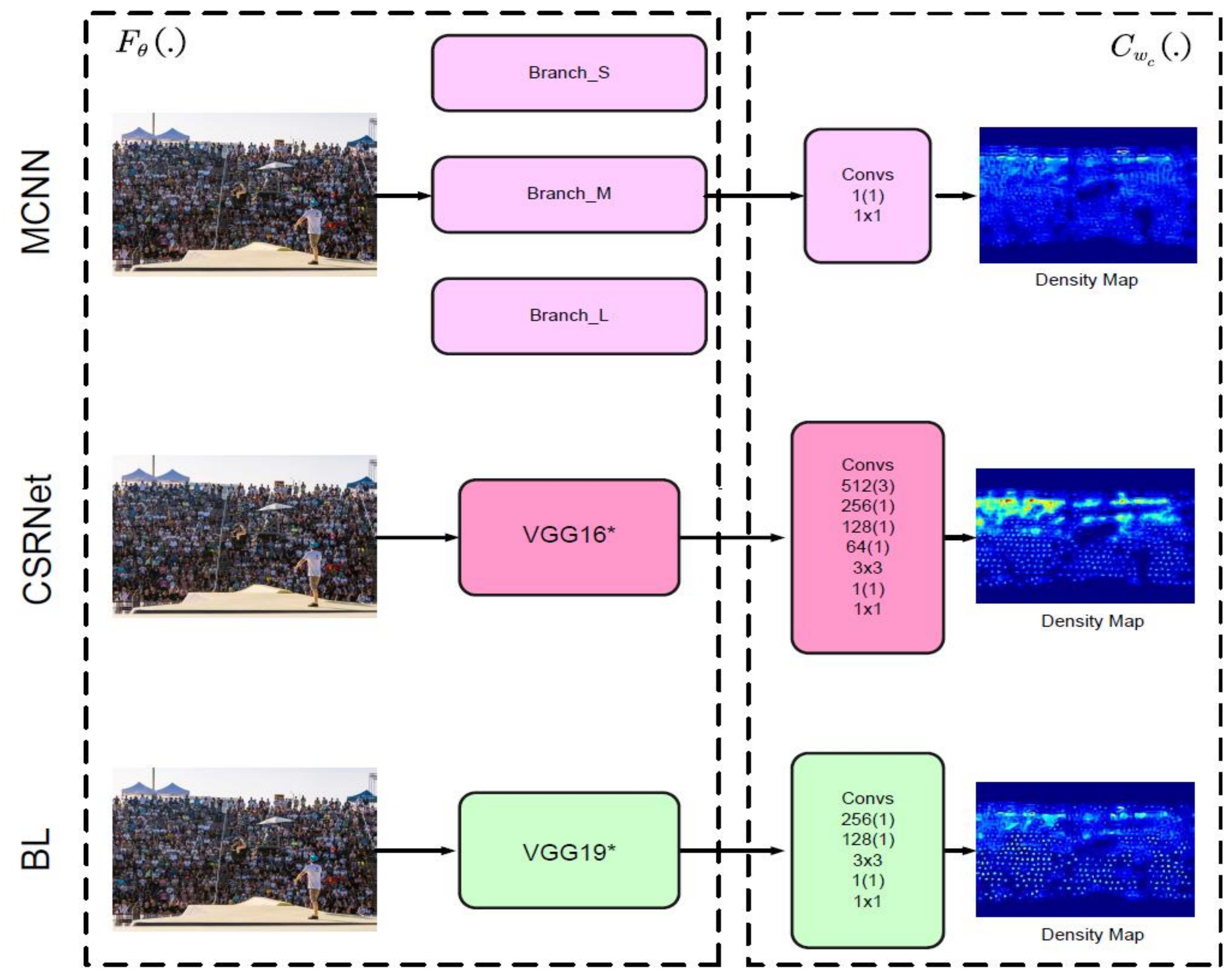
Tao Peng, Pengfei Zhu
Tianjin University, Tianjin, China

Auxiliary Task



Task 1: Equalizing The Rotated Image
Task 2: Summing Sub-images
Task 3: Summing Rotated Sub-images

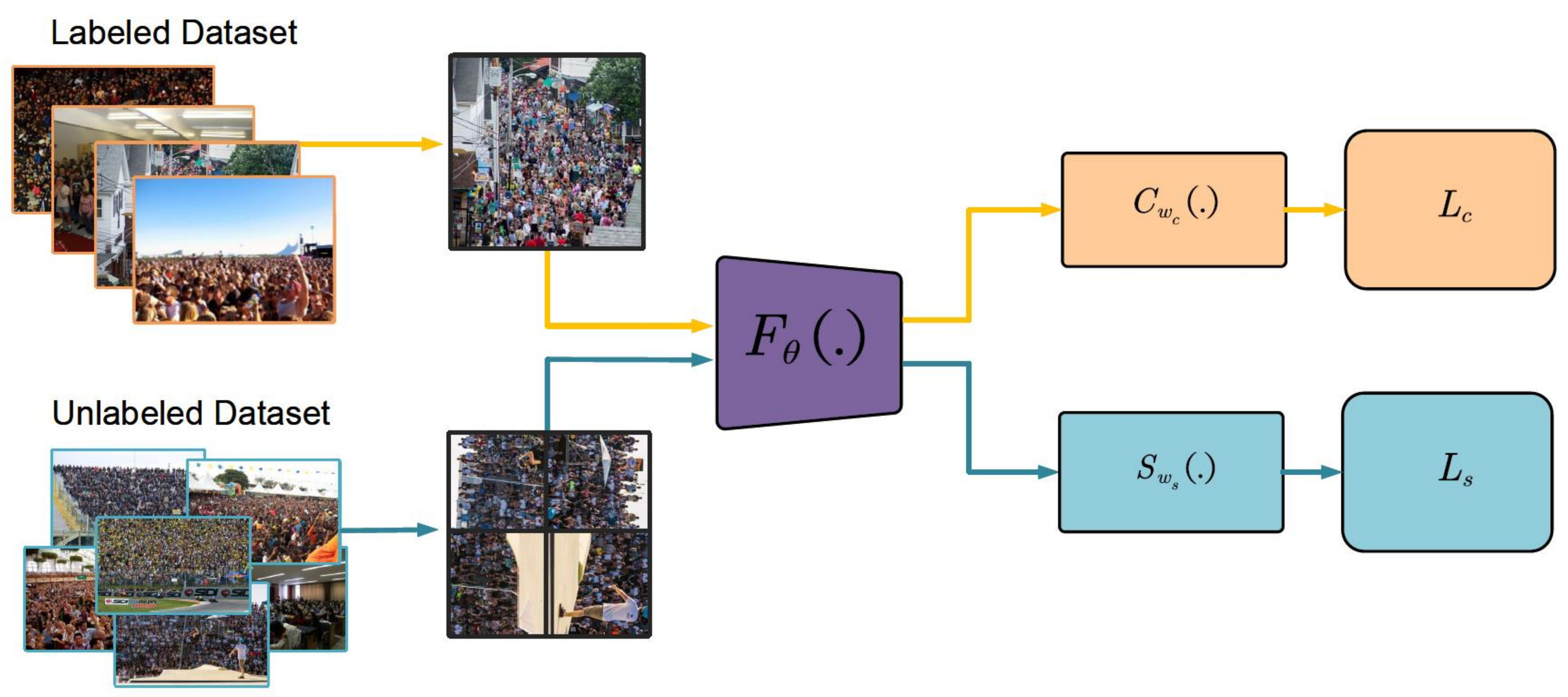
Baseline: MCNN, CSRNet, BL



Main Idea

Deep neural networks need large-scale dataset for better training and evaluation. However collecting and annotating large-scale crowd counting dataset is expensive and challenging. In this work, we exploit unlabeled web crowd image and propose an multi-task framework for boosting crowd counting baseline method through semi-supervision. Based on the observation that the rotation and splitting operations will not change the image crowd counting number, we designed three auxiliary tasks to improve the quality of feature extractors and our framework can be easily extended to other crowd counting baselines. Experiments shows that our semi-supervised learning framework outperforms previous baselines on UCF-QNRF dataset and ShanghaiTech dataset.

Framework



For self-supervised approach, unlabeled data is a subset of original dataset, but our unlabeled data is collected from Internet and does not intersect with original dataset. This is a more flexible and realistic setting, so our approach can be called semi-supervised auxiliary task learning. Finally in our case, we design a multi-task framework where we improve the crowd counting network using joint supervision from the supervised counting task and an semi-supervised pretext task,

Quantitative Results

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

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 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.

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

0726