Multi-Resolution Fusion and Multi-scale Input Priors Based Crowd Counting

Usman Sajid, Wenchi Ma, and Guanghui Wang
{usajid, wenchima, ghwang}@ku.edu

OBJECTIVE & CONTRIBUTIONS

Problem: Crowd counting in low to high density scenarios in given static images.

Contributions:
1) Designed a new multi-resolution feature-level fusion based end-to-end crowd counting approach for still images that effectively deals with significant variations of crowd-density, lighting conditions, and large perspective;
2) Proposed an alternative patch rescaling module by more effectively using the input priors;
3) Outperformed the state-of-the-art methods, including the PRM based schemes, by a large margin with up to **10% improvements**.

EXPERIMENTS

	ShanghaiTech Dataset		UCF-QNRF Dataset	
Method	**MAE**	**RMSE**	**MAE**	**RMSE**
CFF [30]	65.2	109.4	93.8	146.5
RRSP [33]	63.1	96.2	-	-
CAN [21]	**62.3**	100.0	107	183
TEDNet [17]	64.2	109.1	113	188
L2SM[38]	64.2	98.4	104.7	173.6
BL [21]	62.8	101.8	**88.7**	154.8
ZoomCount [27]	66.6	94.5	128	201
PRM-Based [28]	67.8	86.2	94.5	141.9
v5 (ours)	67.1	**81.0**	96.9	**130.1**

AHU-Crowd Dataset		
Method	**MAE**	**RMSE**
DPM [8]	395.4	-
BOW-SVM [7]	218.8	-
Ridge Regression [6]	207.4	-
Hu et. al. [14]	137	-
DSRM [41]	81	129
ZoomCount [27]	88.2	126.1
PRM-Based [28]	74.9	111
v5 (ours)	**60.2**	**91.7**

The Proposed Architecture

The original 128×128 input patch (I_2) is used to produce the new up-scaled (I_1) and down-scaled (I_3) input priors, which go through their respective stems ($stem_1$, $stem_2$, $stem_3$). The resultant initial channels (IC_1, IC_2, IC_3) then pass through the phase-based main network, containing three deep columns/branches with the residual modules (RM). Multi-resolution fusion regularly occurs between these columns, followed by passing through the auxiliary (RH_1, RH_2, RH_3) and the final (RH_{final}) crowd regression heads to yield the respective crowd counts (ccp(1), ccp(2), ccp(3), and cc_{final}). The final crowd count for the input patch (I_2) is the weighted average of these crowd estimates.