



Switching Convolutional Neural Network for Crowd Counting

[<http://val.cds.iisc.ac.in/crowdcnn/>]

Deepak Babu Sam*, Shiv Surya*, R. Venkatesh Babu (*Equal contribution)

Video Analytics Lab, Indian Institute of Science, Bangalore, INDIA



TL;DR : WANT TO IMPROVE YOUR OBJECT COUNTING MODEL?

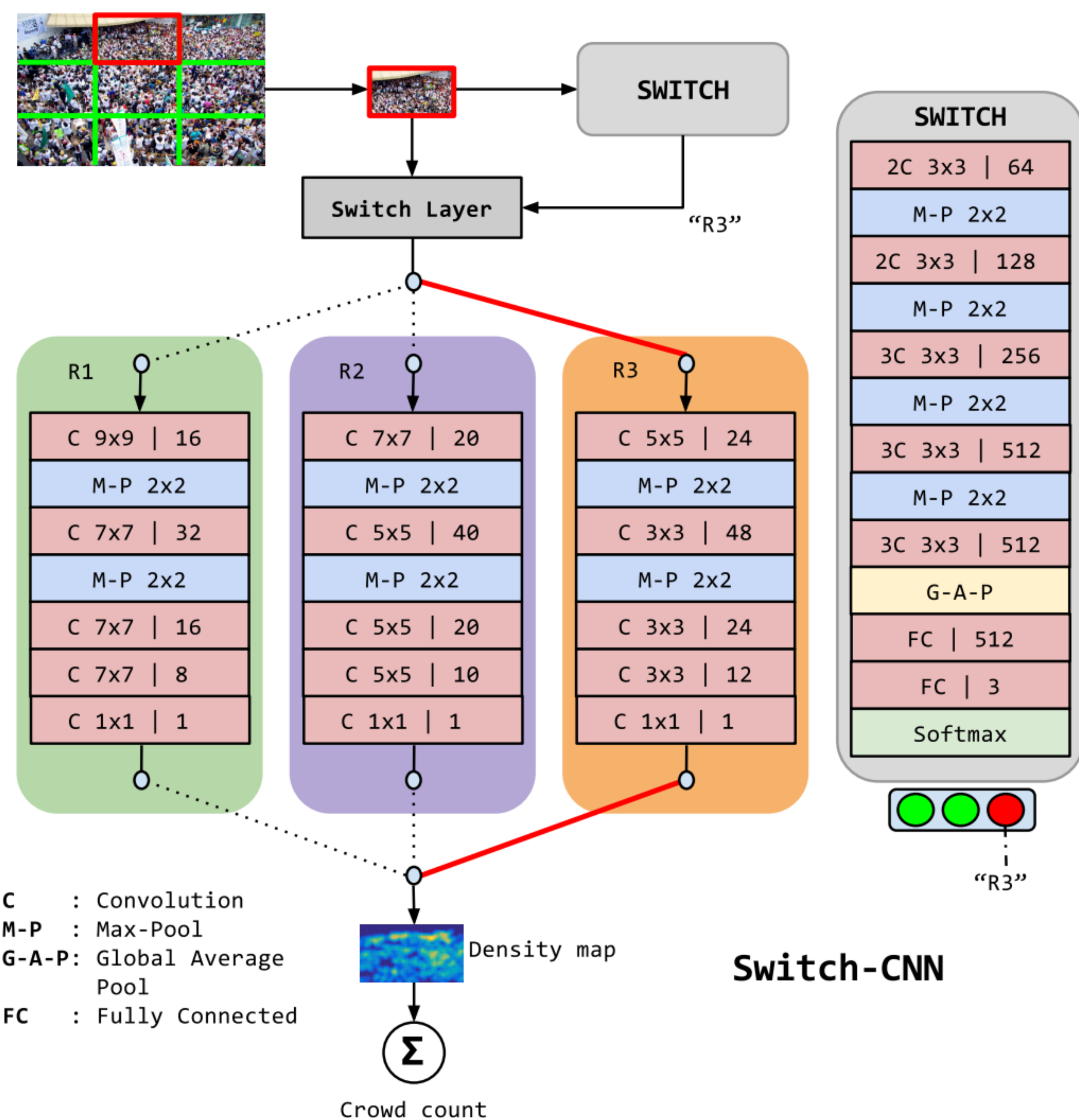
USE AN EXPERT CNN THAT CHOOSES THE BEST OBJECT DENSITY REGRESSOR FOR EVERY SCENE PATCH !

CURRENT COUNTING PARADIGM



CHALLENGES	Global density and scale variations	Varied Illumination and pose	Intra-scene density and scale variation
MODEL			
CNN	X	●	X
Multi-scale CNNs [3]	●	●	●
Multi-column CNNs [1][2]	●	●	●
Switch-CNN (Ours)	●	●	●

OUR APPROACH : SWITCHING CONVOLUTIONAL NEURAL NETWORK



SCNN: Switching CNN

- Novel switching mechanism for handling intra-scene scale variation.
- SWITCH** classifier serves as a expert choosing the best regressor for a part of the crowd scene.
- Regressors **R1-3** and **SWITCH** are trained in an alternating fashion with batch stochastic gradient descent.
- Code available at <https://github.com/val-iisc/crowd-counting-scnn>

ALGORITHM

Initialization and Pre-training

Initialize R_k with weights from Gaussian distribution and pretrain R_k ;
Initialize switch classifier with VGG-16 pretrained on Imagenet;

Differential Training

```

while not converged do
  for every patch in train set do
    Evaluate the patch with all regressors;
    Backpropagate the regressor with best count estimate;
  end
end

```

Coupled Training

```

while not converged do
  for every patch in train set do
    Determine regressor with best count estimate for patch;
    Train switch classifier to predict this label;
  end
end

```

Switched Differential Training

```

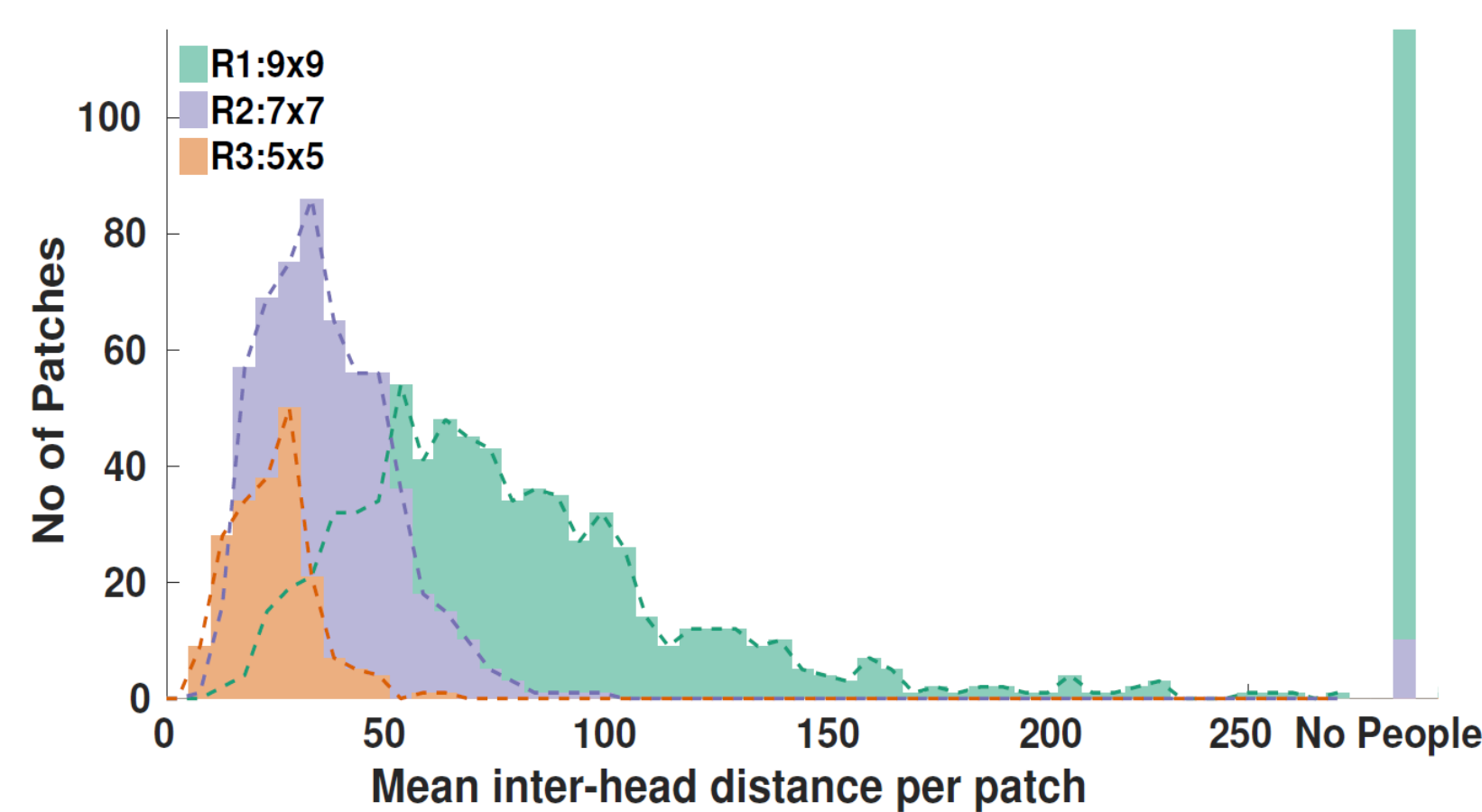
for every patch in train set do
  Infer choice of regressor from switch classifier;
  Backpropagate this chosen regressor;
end

```

MODEL INVESTIGATION



Sample crowd patches preferred by different regressors.



Histogram of average inter-head distance for crowd patches from ShanghaiTech dataset.

Method	MAE
Cluster by mean inter-head distance	94.9
Mixture of Experts	111.6
Late fusion of regressors (MCNN)	110.2
Switch-CNN	90.4

ShanghaiTech dataset Part A

Method	MAE
R1	157.6
R2	178.8
R3	178.1
Switch-CNN (R1, R3)	98.8
Switch-CNN (R1, R2)	110.8
Switch-CNN (R2, R3)	126.6
Switch-CNN	90.4

ShanghaiTech dataset Part A

REFERENCES

- [1] Zhang et al. "Single-image crowd counting via multi-column convolutional neural network.", CVPR 2016.
- [2] Boominathan et al. "Crowdnet: a deep convolutional network for dense crowd counting.", ACM MM 2016.
- [3] Onoro-Rubio et al. "Towards perspective-free object counting with deep learning.", ECCV 2016.
- [4] Zhang et al. "Cross-scene crowd counting via deep convolutional neural networks ", CVPR 2015.

RESULTS

	Part A		Part B	
Method	MAE	MSE	MAE	MSE
Zhang et al. [4]	181.8	277.7	32.0	49.8
MCNN [1]	110.2	173.2	26.4	41.3
Switch-CNN	90.4	135.0	21.6	33.4

ShanghaiTech dataset

Method	S1	S2	S3	S4	S5	Avg. MAE
Zhang et al. [4]	9.8	14.1	14.3	22.2	3.7	12.9
MCNN [1]	3.4	20.6	12.9	13.0	8.1	11.6
Switch-CNN	4.4	15.7	10.0	11.0	5.9	9.4

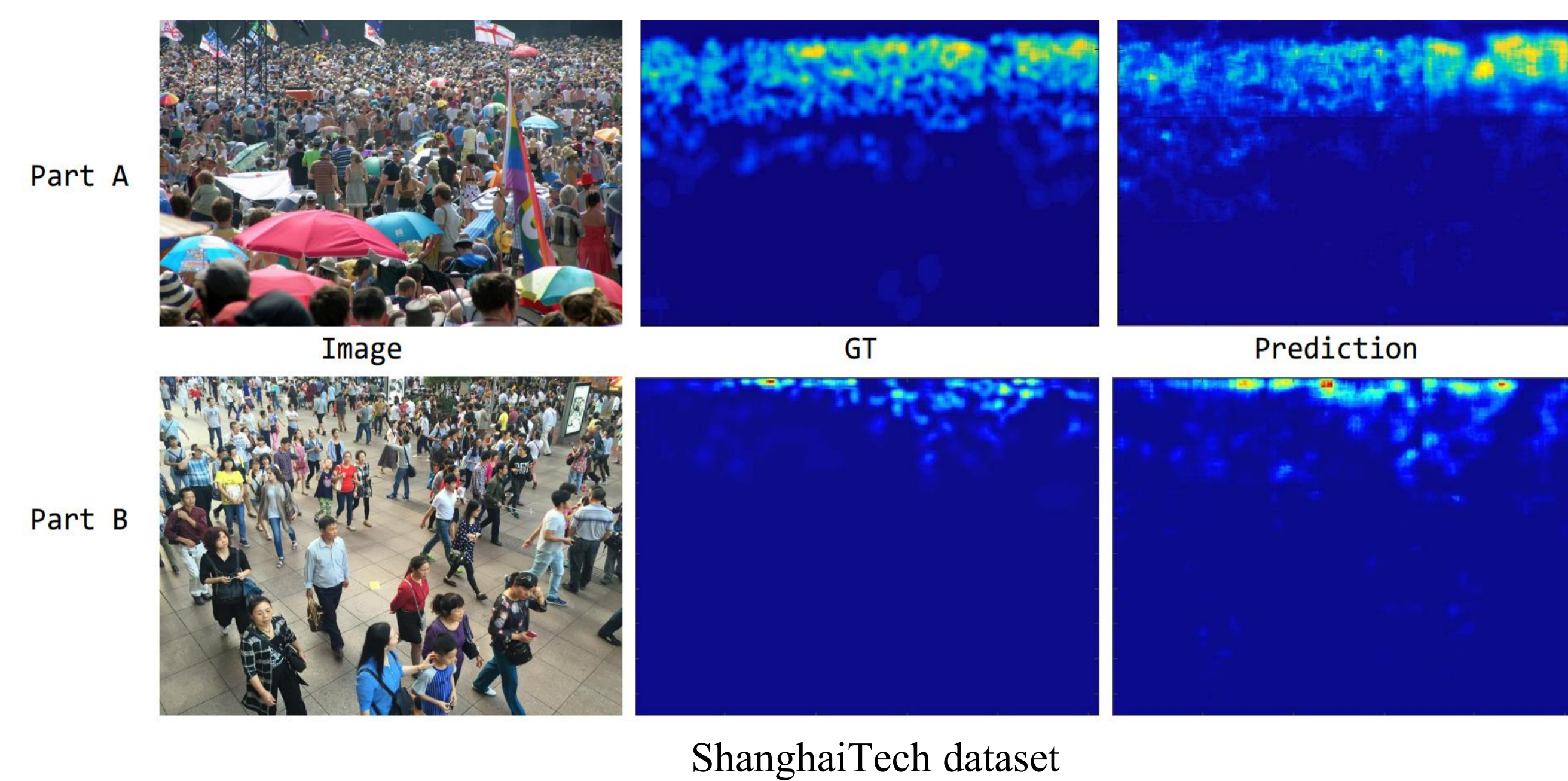
World Expo'10 dataset

Method	MAE	MSE
MCNN [1]	377.6	509.1
Hydra2s [3]	333.7	425.3
Switch-CNN	318.1	439.2

UCF_CC_50 dataset

Method	MAE	MSE
Zhang et al. [4]	1.60	3.31
MCNN [1]	1.07	1.35
CCNN [3]	1.51	—
Switch-CNN	1.62	2.10

UCSD Crowd Counting dataset



ACKNOWLEDGEMENT

This work was supported by SERB, Department of Science and Technology (DST), Government of India (Proj No. SB/S3/EECE/0127/2015). We wish to thank Pratiksha Trust, ACM-India and IEEE Signal Processing Society, Bangalore Chapter for their generous support.