DA-NET: DEPTH AWARE NETWORK FOR CROWD COUNTING

Van-Su Huynh
National Chung Cheng University, Taiwan

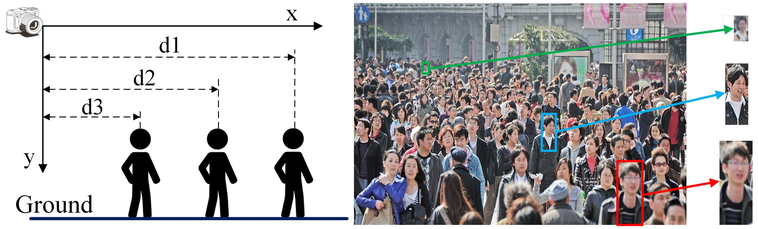
Vu-Hoang Tran
University of Technology and Education, Vietnam

Ching-Chun Huang
National Chiao Tung University, Taiwan

1. INTRODUCTION

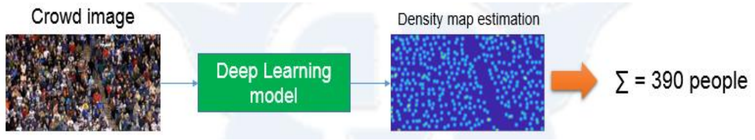
Challenges:

- Scale variation problem is caused by depth changes in crowd images.
- Limitation of crowd dataset (only few hundreds labeled images are available)



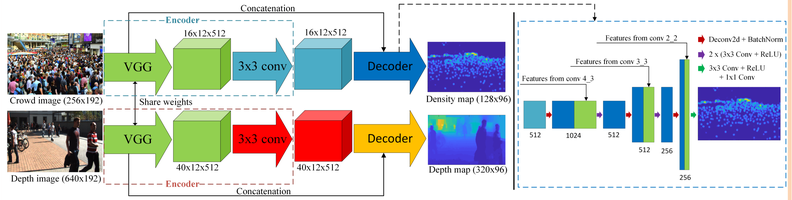
Goal

- Given a crowd image, we build a compact CNN architecture to count number of people and estimate density map that could handle the scale variation problem



2. PROPOSED METHOD

- Multi-task network is trained on separated and individual datasets
- Framework consists of two branches corresponding to two tasks:
  - Crowd density map estimation (**main task**)
  - Depth map estimation (**auxiliary task**)



Encoder:

- Use basic CNNs-based architecture (VGG16)
- Share weights between two tasks

Decoder:

- The decoder is independent for each task
- Adopt U-net architecture to take advantage of both low-level and high-level features for the estimation

Overall loss:

- End-to-end training with overall loss:

$$L = L_{den} + \gamma L_{dep}$$

where L_{den}, L_{dep} are Euclidean distances for density map estimation and depth map estimation tasks, respectively.

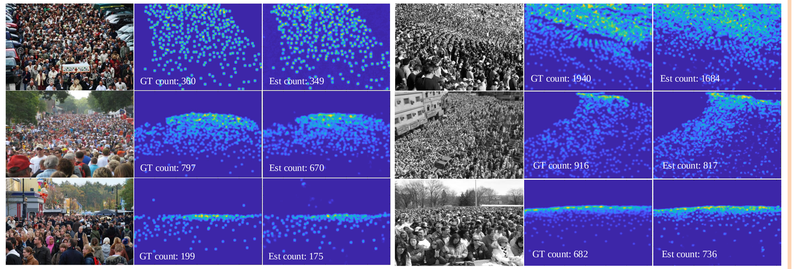
3. EXPERIMENTAL RESULTS

Quantitative results

Method	ShanghaiTech				UCF_CC_50	
	Part_A		Part_B			
	MAE	MSE	MAE	MSE	MAE	MSE
Zhang *et al.*, CVPR 2015	181.8	277.7	32.0	49.8	467.0	498.5
MCNN, CVPR 2016	110.2	173.2	26.4	41.3	377.6	509.1
Switch-CNN, CVPR 2017	90.4	135.0	21.6	33.4	318.1	439.2
Sindagi *et al.*, AVSS 2017	101.3	152.4	20.0	31.1	322.8	397.9
CP-CNN, ICCV 2017	73.6	**106.4**	20.1	30.1	295.8	**320.9**
DecideNet, CVPR 2018	-	-	20.8	29.4	-	-
Liu *et al.*, CVPR 2018	73.6	112.0	13.7	21.4	279.6	388.9
Our DAnet	**71.4**	120.6	**9.1**	**14.7**	**268.3**	373.2

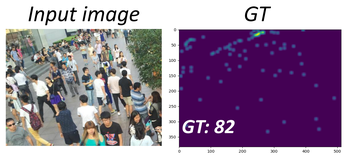
Result Illustrations

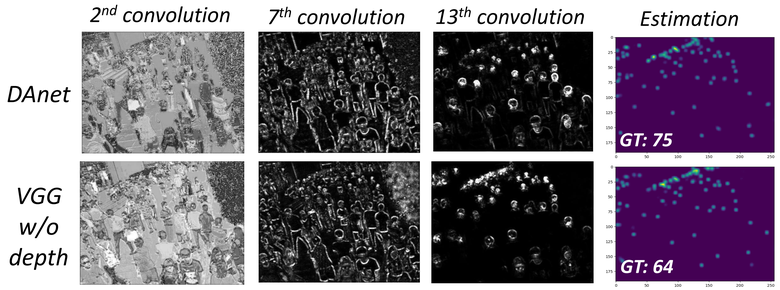
ShanghaiTech dataset *UCF_CC_dataset*



Feature Visualization

- Using **guided backpropagation** to visualize the learning features





4. CONCLUSION

- By leveraging the auxiliary depth estimation dataset, an alternative and novel way was proposed to handle the scale variation problem.
- Our DAnet is trained for two tasks simultaneously: density map estimation and depth map estimation using separated crowd and depth datasets. The experiments demonstrate the efficiency of the proposed method over the state-of-the-art methods.