# A REAL-TIME DEEP NETWORK FOR CROWD COUNTING

*Xiaowen Shi*[2], *Xin Li*[1,2], *Caili Wu*[1,2], *Shuchen Kong*[3], *Jing Yang*[2], *Liang He*[1,2]

[1]Shanghai Key Laboratory of Multidimensional Information Processing,
[2]East China Normal University, Shanghai, China,
[3]Videt Tech Ltd., Shanghai, China

華東師範大學
EAST CHINA NORMAL UNIVERSITY

兑观
VIDET

# Outline

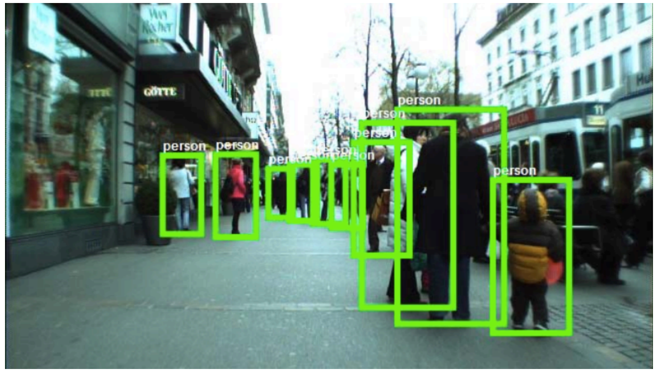
- Introduction and Motivation
- Framework
- Experiments
- Conclusion

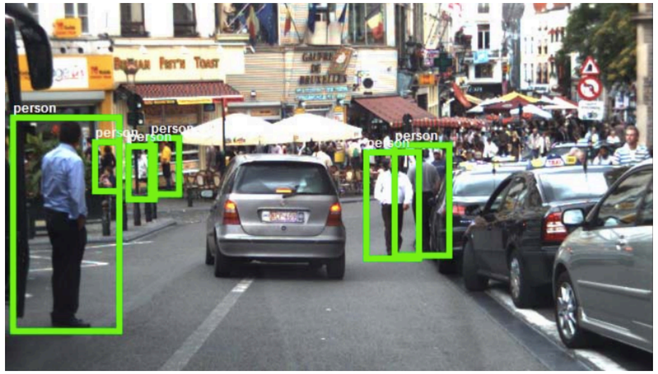
# Outline

- Introduction and Motivation
- Framework
- Experiments
- Conclusion

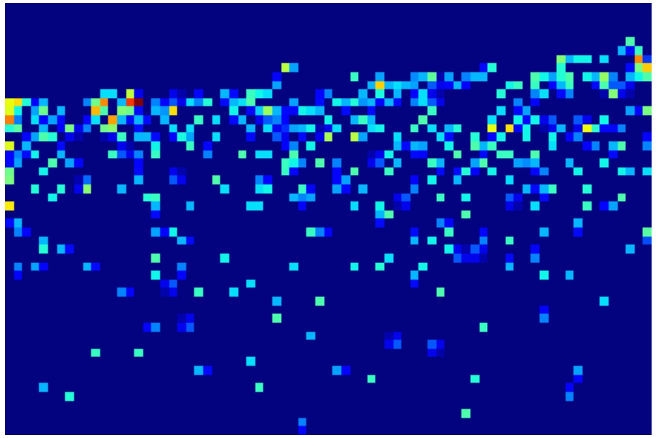
# Background

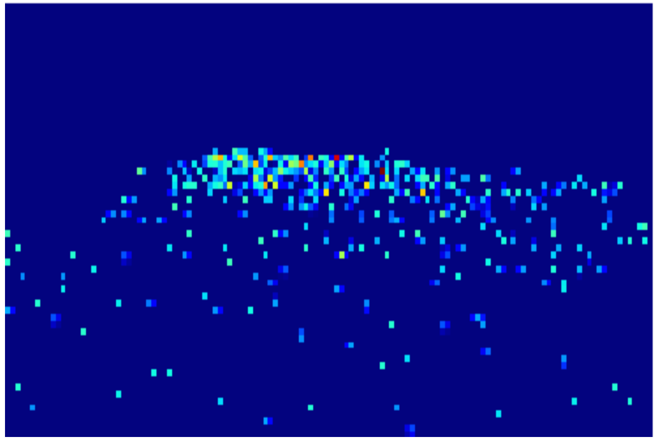
- Crowd counting
  - Count-oriented Approaches




  - Density-oriented Approaches

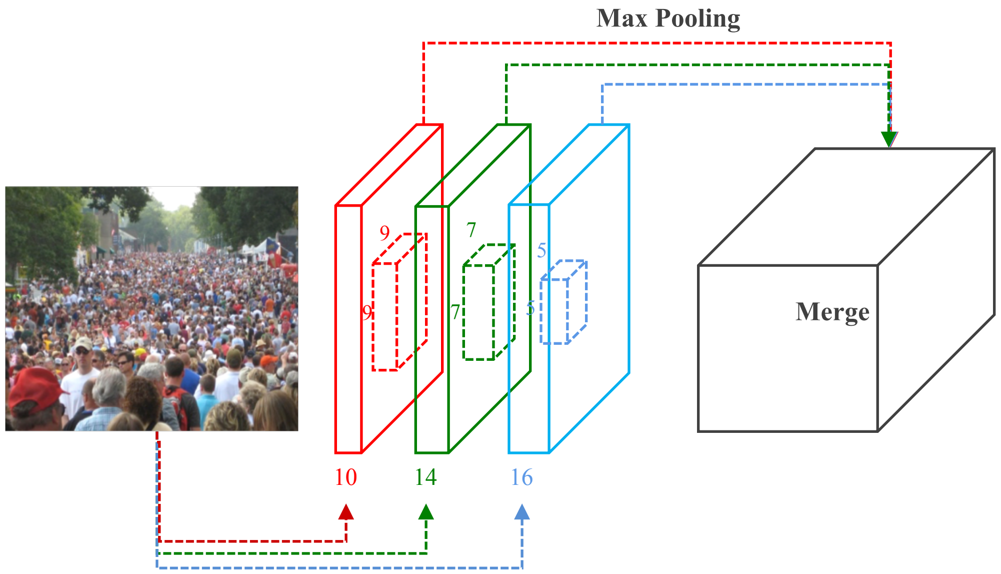




# Challenges

- Occlusions, high clutter, scale and perspective
- limited computing resources in practical applications
- High requirement to the processing speeds

# Motivation




- to deal with the problem as bellow:
  - sub-optimal and time-consuming problem.
  - need to store a large amount of parameters

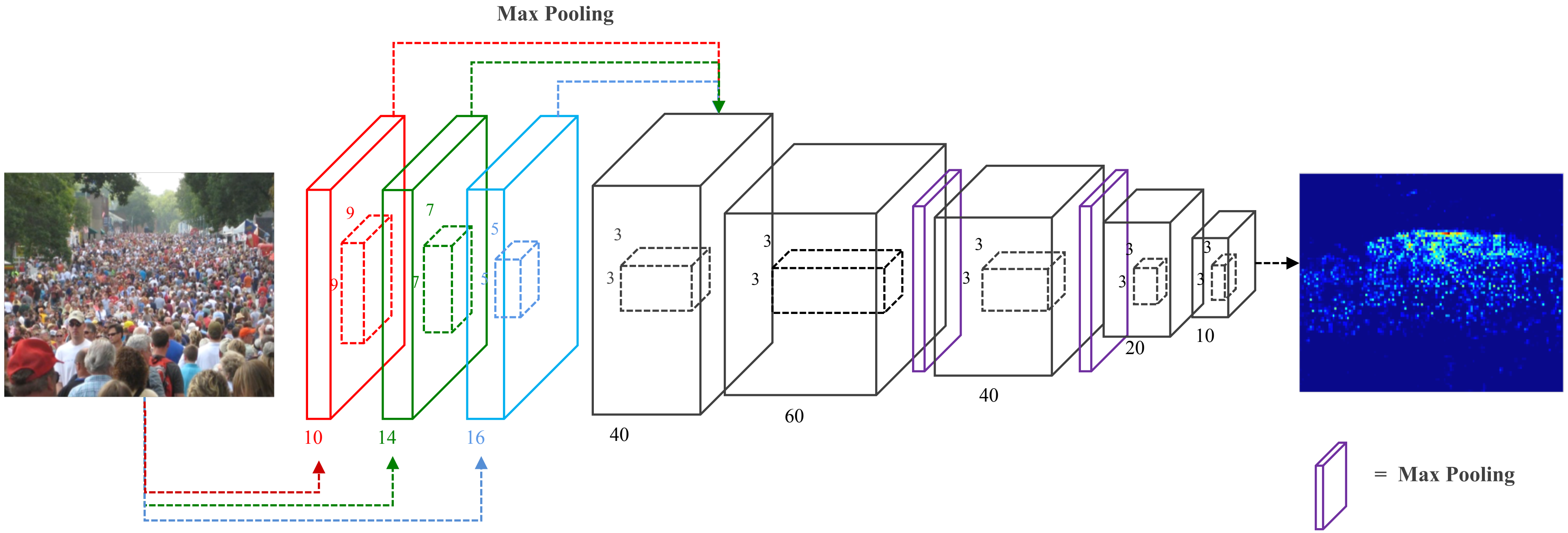

Y. Zhang, D. Zhou, S. Chen, S. Gao, and Y. Ma. Single-image crowd counting via multi-column convolutional neural network. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 589–597, 2016.

# Outline

- Introduction and Motivation
- Framework
- Experiments
- Conclusion

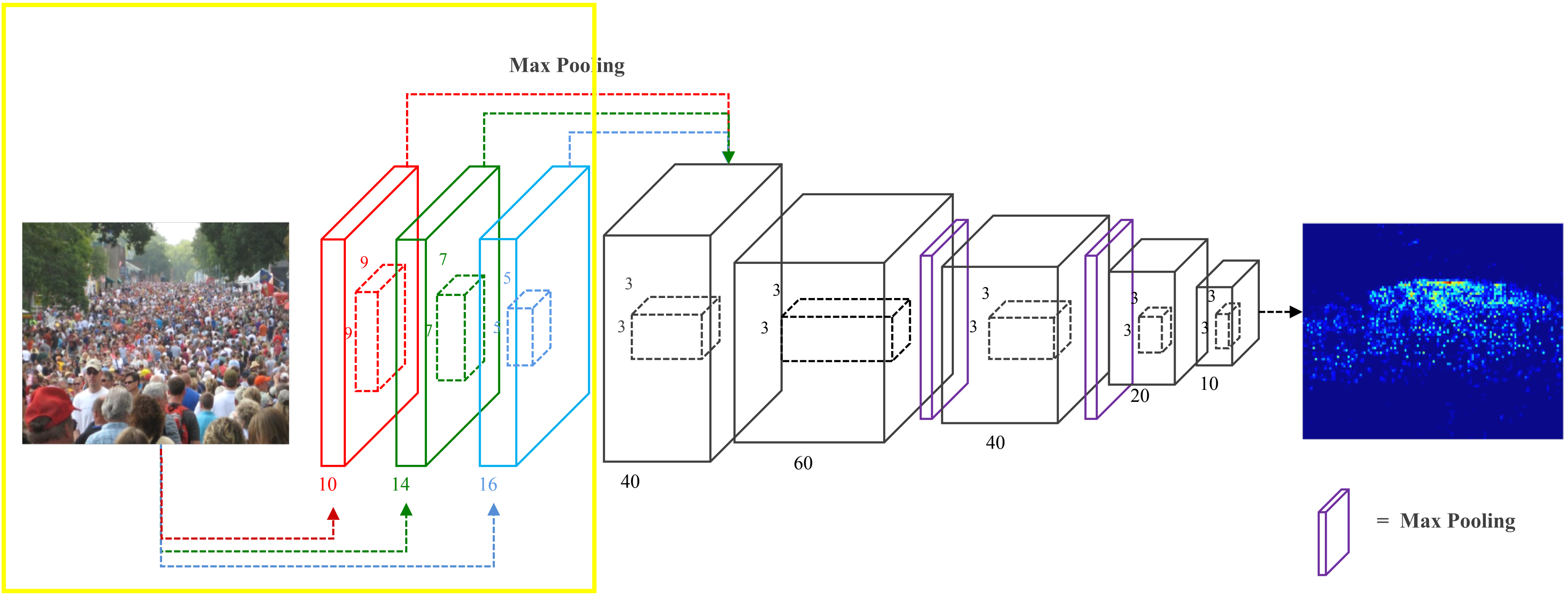
# Framework

- two components：
  - The parallel convolution layer with different kernels
  - The convolution with pooling layers that followed.



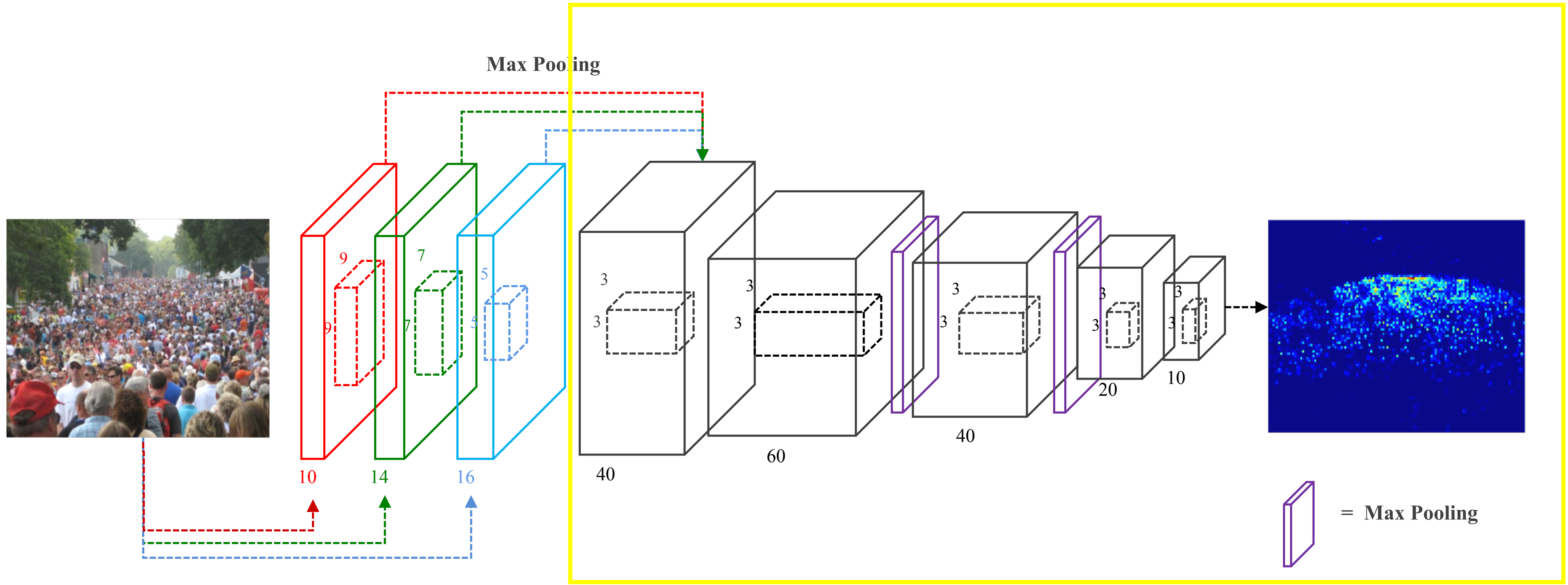
# Framework

- In the front part：
  - using three filters different receptive fields in one layer (red/green/blue).
  - the feature maps are merged directly after receptive fields.



# Framework

- In the latter part：
  - consists of 6 convolutional layers specifically.
  - last convolution layer aggregate the feature maps into a density map.

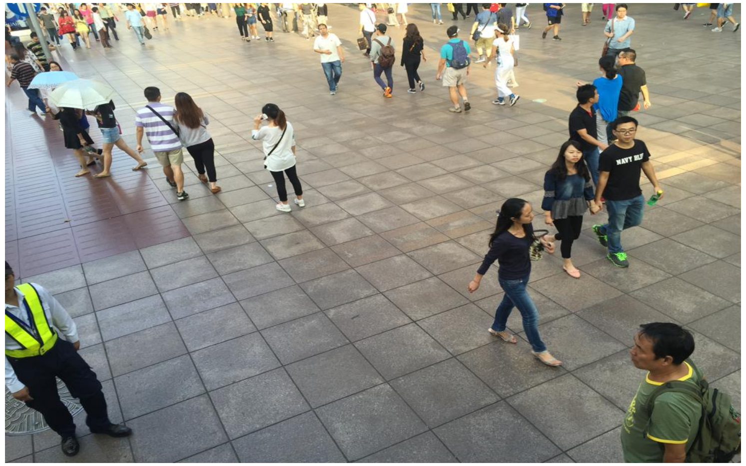


# Outline

- Introduction and Motivation
- Framework
- Experiments
- Conclusion

# Experiment

- Dataset
  - ShanghaiTech dataset
  - The WorldExpo’10 dataset
- Evaluation
  - MAE & MSE



# Experiment

- Result & Comparing

| Method | Part A | | Part B | | Parameter size |
|---|---|---|---|---|---|
| | MAE | MSE | MAE | MSE | |
| CMTL [18] | 101.3 | 152.4 | 20.0 | 31.1 | 2.36M |
| Zhang *et al.* [17] | 181.8 | 277.7 | 32.0 | 49.8 | 0.62M |
| MCNN [7] | 110.2 | 173.2 | 26.4 | 41.3 | 0.15M |
| TDF-CNN [19] | 97.5 | 145.1 | 20.7 | 32.8 | 0.13M |
| C-CNN | **88.1** | **141.7** | **14.9** | **22.1** | **0.07M** |
| ACSCP [20] | 75.7 | **102.7** | 17.2 | 27.4 | 5.10M |
| Switching CNN [1] | 90.4 | 135.0 | 21.6 | 33.4 | 15.30M |
| CSRNet [21] | **68.3** | 115.0 | **10.6** | **16.0** | 16.26M |
| SaCNN [22] | 86.8 | 139.2 | 16.2 | 25.8 | 24.06M |
| CP-CNN [2] | 73.6 | 106.4 | 20.1 | 30.1 | 68.40M |

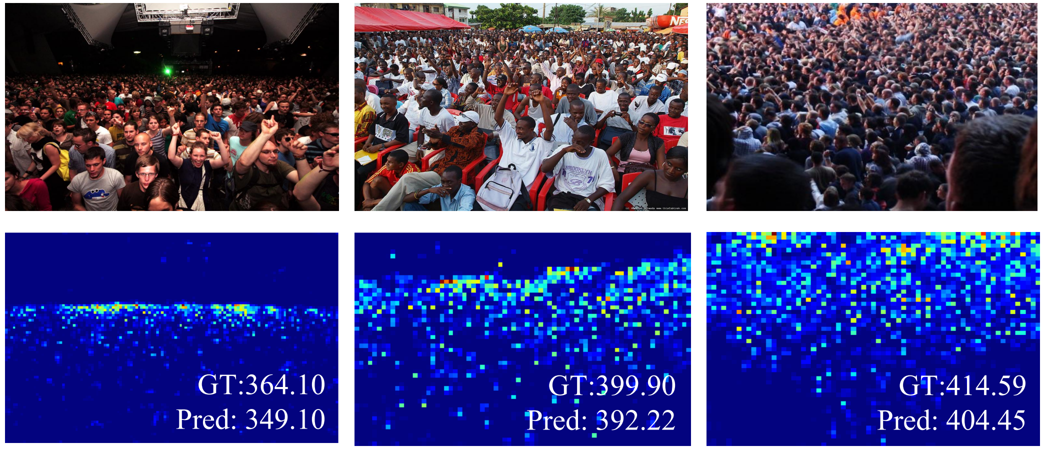
Table1. Comparison on ShanghaiTech dataset

| Method | S1 | S2 | S3 | S4 | S5 | Avg. | Params |
|---|---|---|---|---|---|---|---|
| Zhang *et al.* [17] | 9.8 | **14.1** | 14.3 | 22.2 | 3.7 | 12.9 | 0.62M |
| MCNN [7] | 3.4 | 20.6 | 12.9 | 13.0 | 8.1 | 11.6 | 0.15M |
| TDF-CNN [19] | **2.7** | 23.4 | 10.7 | 17.6 | **3.3** | 11.5 | 0.13M |
| C-CNN(ours) | 3.8 | 20.5 | **8.8** | **8.8** | 7.7 | **9.9** | **0.07M** |
| CSRNet [21] | 2.9 | **11.5** | **8.6** | 16.6 | 3.4 | 8.6 | 16.26M |
| SaCNN [22] | **2.6** | 13.5 | 10.6 | 12.5 | **3.3** | **8.5** | 24.06M |
| CP-CNN [2] | 2.9 | 14.7 | 10.5 | 10.4 | 5.8 | 8.86 | 68.40M |

Table2. Comparison on WorldExpo’10 dataset

# Experiment

- Results demonstration

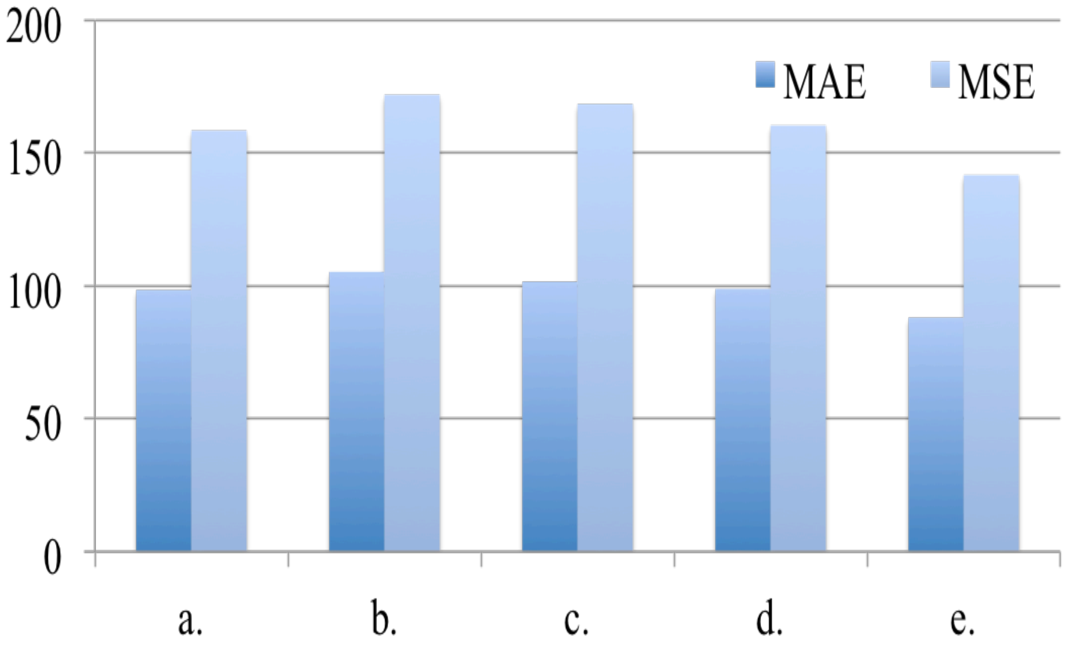



(a) ShanghaiTech Part A


(b) ShanghaiTech Part B

# Experiment

- Ablation Experiments & Speed Comparison




| Method | CMTL [18] | MCNN [7] | C-CNN |
|---|---|---|---|
| FPS | 8.37 | 64.52 | **104.16** |

# Conclusion

- A compact CNN for crowd counting is proposed to deal with the lack of real-time performance of existing methods.
- Utilizes three filters with different sizes of local receptive field in one layer and directly targeting a merged feature map at once.
- Compared with the baseline approaches, the proposed model obtains an improvement significantly.

*Thanks!*