DAPC : Domain Adaptation People Counting via Style-level Transfer Learning and Scene-aware Estimation

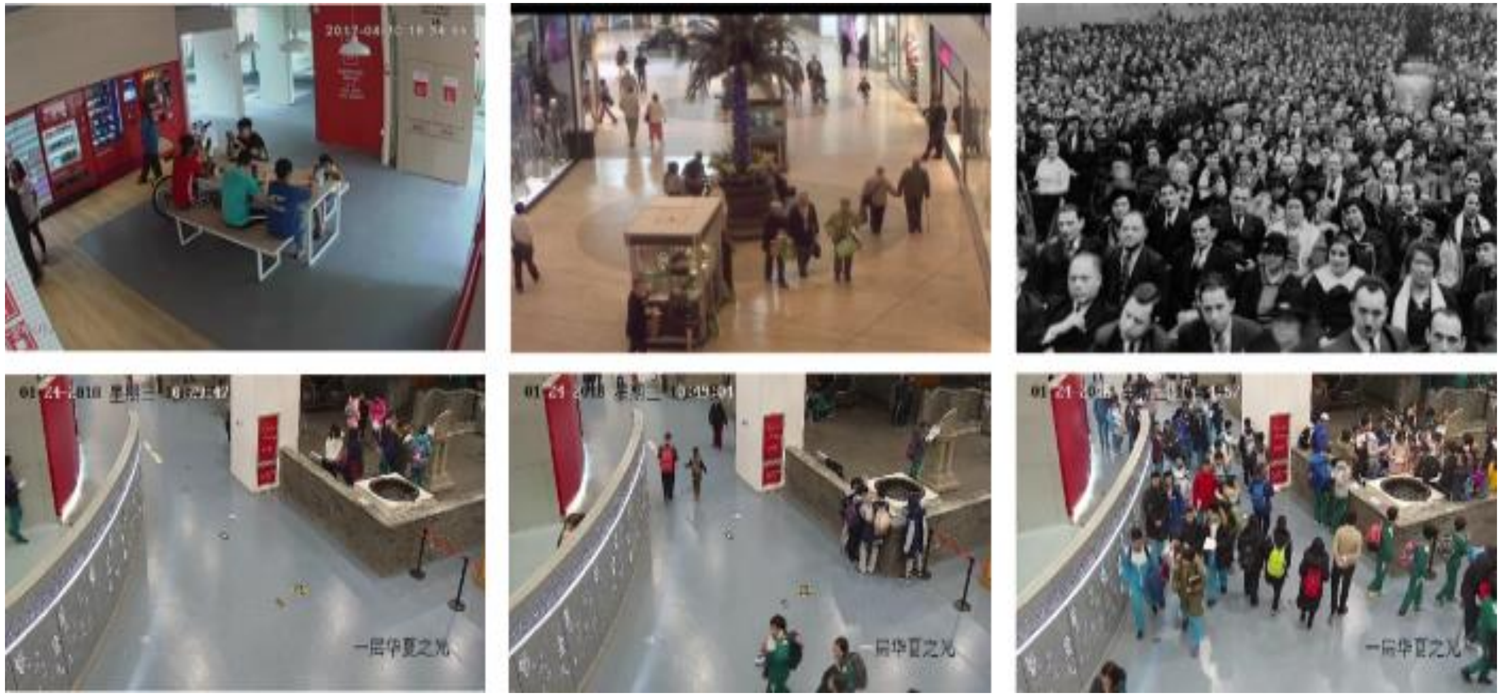
ICPR 2020

Na Jiang , Xingsen Wen , Zhiping Shi

Capital Normal University

What is people counting?

The emphasis is on the statistics of the number of people in the real-world surveillance images !



Challenges and methods

- CNN-based People Counting :

  [1] designed an end-to-end counting framework for extremely dense scenes.
  [2] proposed DecideNet with detection and regression branches.
  [3] designed CSRNet to understand highly congested scenes and estimate high-quality density maps.

- Domain Adaptation in Vision Tasks :

  [4] adopted generative adversarial networks (GANs) to alleviate the shift between source domain and target domain.
  [5] presented a learning framework to translate the labeled images from source to target domain.
  [6] transferred the different style of each camera with CycleGan.

[1] C. Wang, H. Zhang, L. Yang, S. Liu, and X. Cao, "Deep people counting in extremely dense crowds," *in Proceedings of the 23rd ACM international conference on Multimedia*, 2015, pp. 1299–1302.
[2] J. Liu, C. Gao, D. Meng, and A. G. Hauptmann, "Decidenet: Counting varying density crowds through attention guided detection and density estimation," *in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 5197–5206.
[3] Y. Li, X. Zhang, and D. Chen, "Csrnet: Dilated convolutional neural networks for understanding the highly congested scenes," *in Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 1091–1100.
[4] S. Sankaranarayanan, Y. Balaji, A. Jain, S. Nam Lim, and R. Chellappa, "Learning from synthetic data: Addressing domain shift for semantic segmentation," *in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 3752–3761.
[5] W. Deng, L. Zheng, Q. Ye, G. Kang, Y. Yang, and J. Jiao, "Imageimage domain adaptation with preserved self-similarity and domaindissimilarity for person re-identification," *in Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 994–1003.
[6] Z. Zhong, L. Zheng, Z. Zheng, S. Li, and Y. Yang, "Camera style adaptation for person re-identification," *in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp.5157–5166.

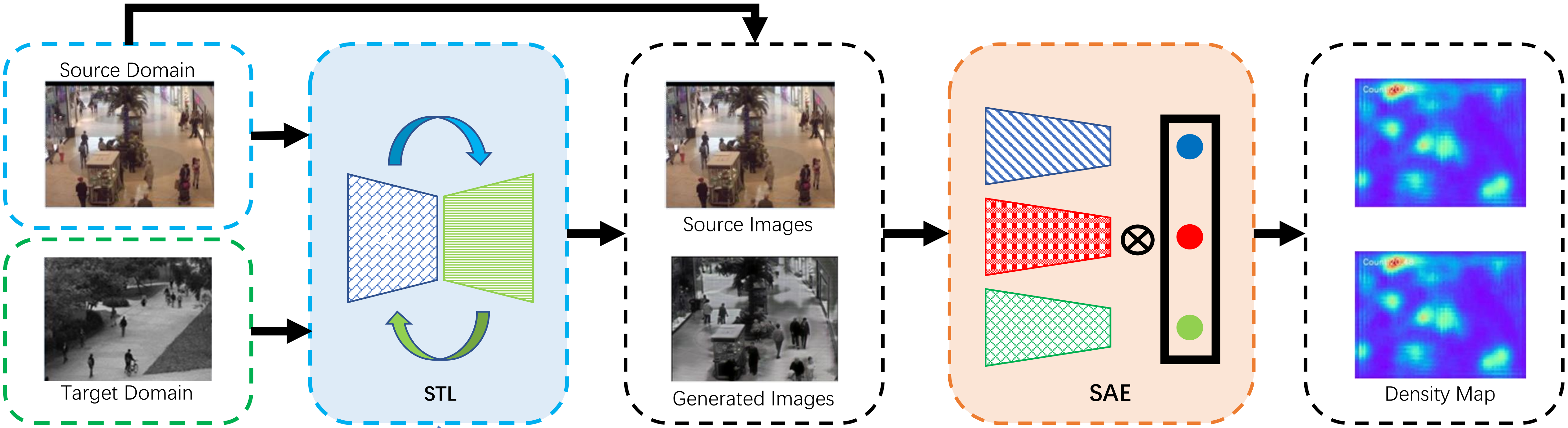
Challenges and methods

- Challenges :
  - Rich variations in scene type
  - Crowd density
  - The result generalization of the closed-set is limited
- We need to do:
  - Improving the current general approach through domain adaptation
  - Alleviates interference of variations in scene type and crowd density on domain adaptation people counting

Contributions

- Propose a style-level transfer learning
- Propose a scene-aware estimation
- Method achieves the state-of-the-art generalization in domain adaptation experiments

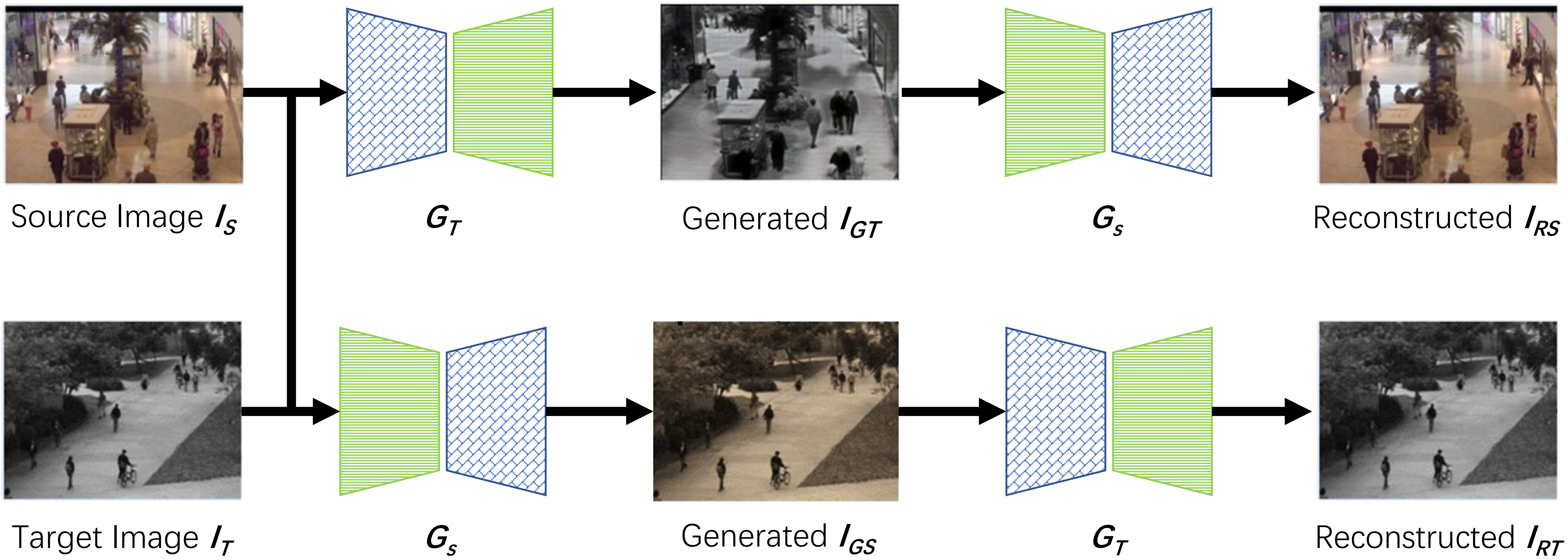
Our Outline



This method can effectively improve the generalization characteristics of the closed-set, enabling it to achieve better cross-domain learning.

This method can achieve partial balance between different scenarios, and can also improve the statistical accuracy of different crowd densities

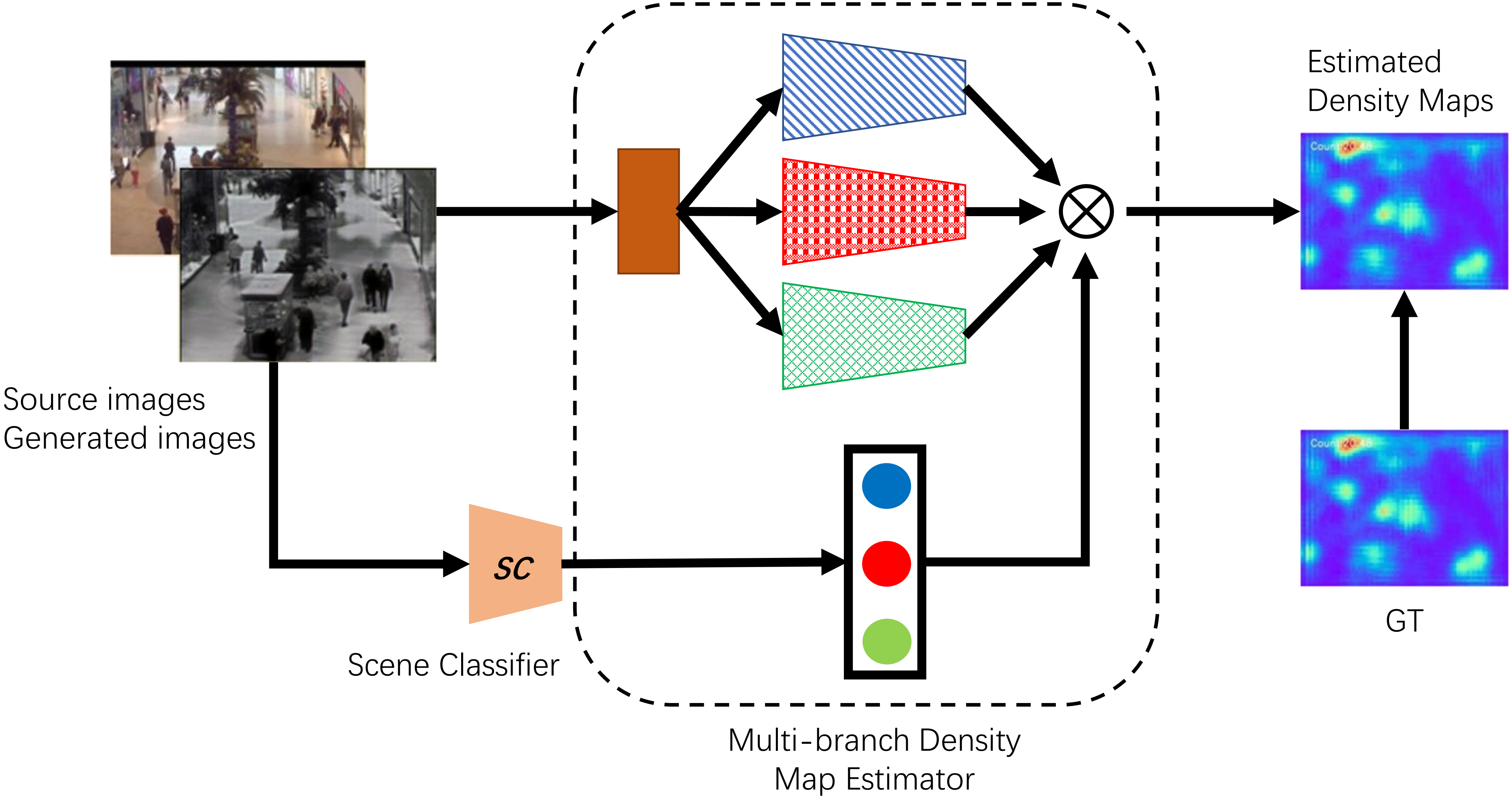
How to design Style-level Transfer Learning (STL) ?



Demonstrates the transfer process of STL

✓ **The generated images by this method and source images will be used in the next stage**

How to design Scene-aware Estimation (SAE) ?



Some of our experimental results

- Method achieves the state-of-the-art generalization in domain adaptation experiments

Table 1 : Ablation results of different cross-domain groups

Configurantion	TL	A:Baidu→Mall			B:Mall→Baidu			C:Mall→UCSD			D:SHA→Mall		
		MAE↓	MSE↓	C_{rate}↑	MAE↓	MSE↓	C_{rate}↑	MAE↓	MSE↓	C_{rate}↑	MAE↓	MSE↓	C_{rate}↑
Baseline	N	8.3	11.9	–	25.3	33.7	–	16.3	22.5	–	9.5	13.4	–
SAE	N	6.6	8.3	0.27	25.0	32.9	0.02	13.1	17.5	0.22	7.9	9.3	0.22
STL+Baseline	Y	4.5	6.3	0.61	24.3	30.7	0.06	5.4	8.3	0.74	5.5	7.2	0.54
STL+SAE	Y	**2.9**	**3.4**	**0.87**	**23.4**	**28.1**	**0.12**	**3.1**	**5.1**	**0.89**	**3.8**	**4.1**	**0.77**
SUP	–	2.1	2.8	–	9.0	14.6	–	1.5	1.7	–	2.1	2.8	–

'**SUP**' represents the results obtained by baseline with supervised training on the target domain.

Some of our experimental results

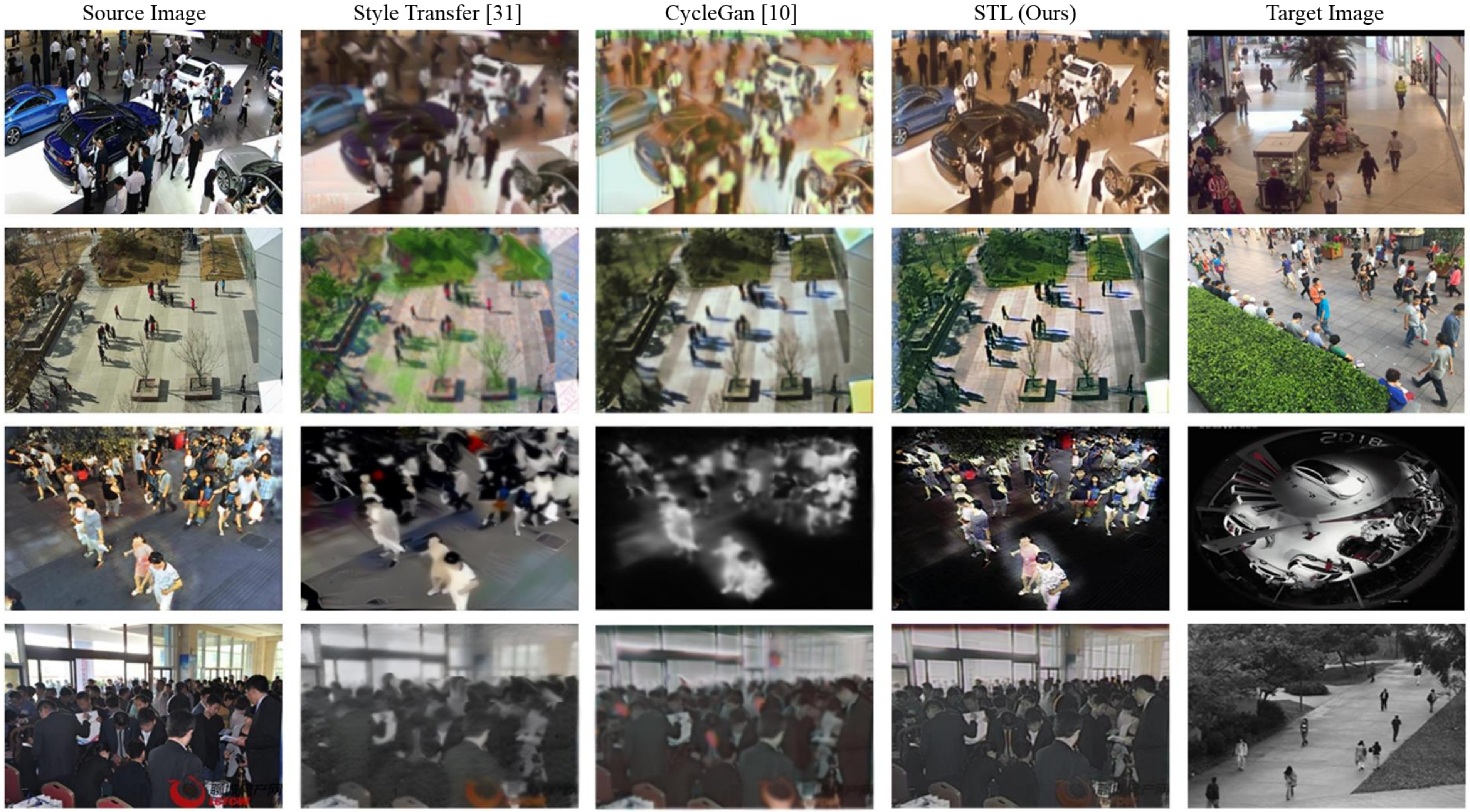
Table 2 : Comparative results of domain adaptation people counting

Method	TL	Baidu→SHA		Baidu→SHB		Baidu→Mall		Baidu→UCSD	
		MAE↓	MSE↓	MAE↓	MSE↓	MAE↓	MSE↓	MAE↓	MSE↓
MCNN[7]	N	246.4	310.4	35.4	45.7	13.5	18.5	20.5	25.3
CSRNET[8]	N	169.3	237.5	24.5	33.1	8.5	12.5	18.5	23.5
MCNN[7]+CLG[10]	Y	210.5	250.4	30.7	39.0	7.9	12.9	15.0	19.6
CSRNET[8]+CLG[10]	Y	149.9	233.6	20.6	30.4	6.5	8.9	8.5	13.5
MCNN[8]+STL	Y	178.4	230.1	25.4	**25.1**	7.5	15.9	8.0	13.9
CSRNET[8]+STL	Y	**140.3**	**202.7**	**19.4**	27.4	**4.3**	**7.1**	**4.9**	**9.1**
DAPC(ours)	Y	**119.3**	**190.1**	**18.3**	**24.0**	**2.9**	**3.4**	**2.6**	**3.0**

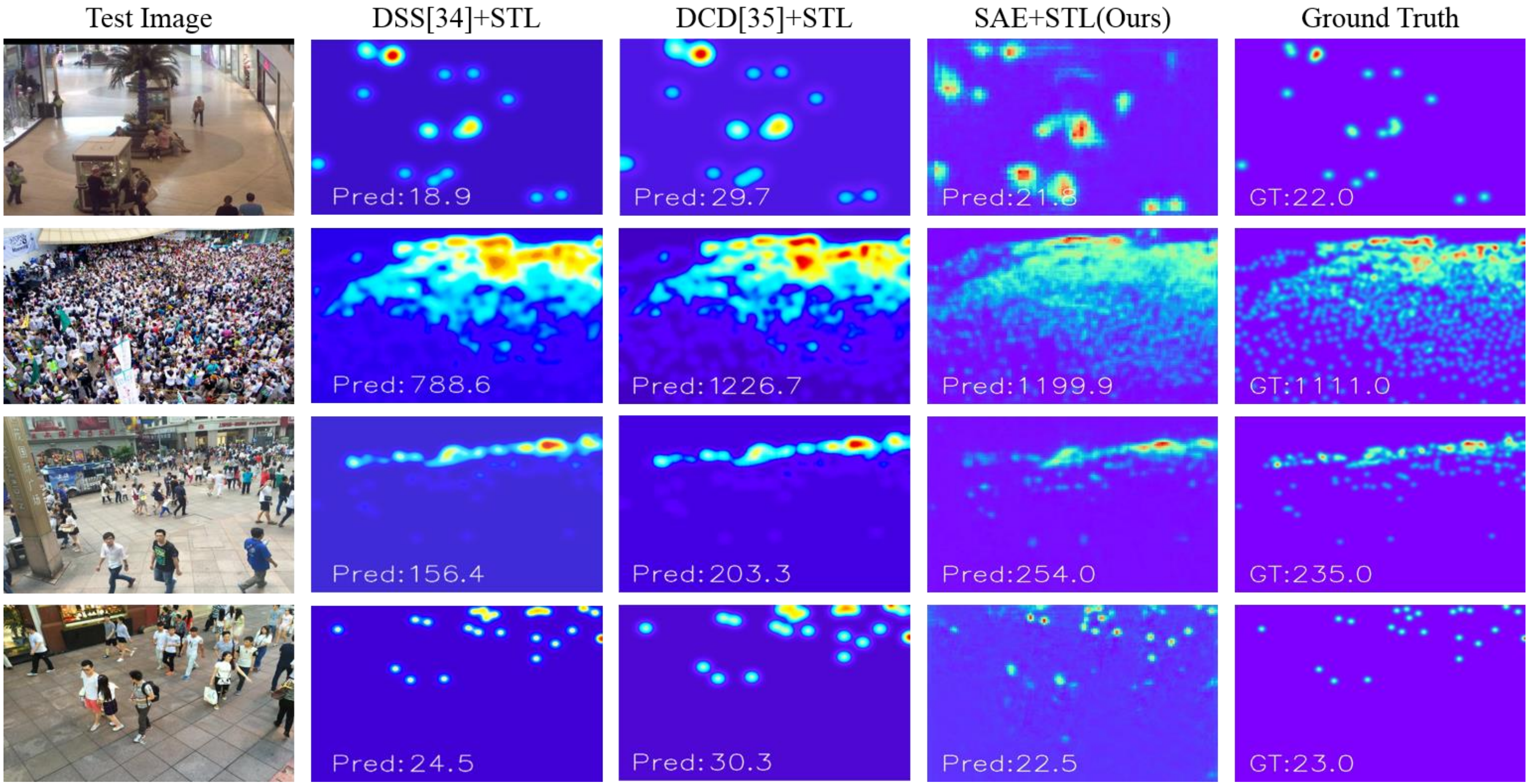
Method	TL	GCC→SHA		GCC→SHB		PE09+UC→Mall		PE09+Ma→UCSD	
		MAE↓	MSE↓	MAE↓	MSE↓	MAE↓	MSE↓	MAE↓	MSE↓
DA-ELM	Y	–	–	–	–	3.58	5.27	2.71	4.86
SeCycleGan	Y	125.7	194.3	19.9	28.3	–	–	–	–
DAPC(ours)	Y	**120.6**	**191.2**	**18.8**	**26.4**	**3.34**	**4.76**	**2.60**	**4.11**

'CLG' means the CycleGan Red and blue values in bold highlight the first and second results

Some of our experimental results



Conclusion: results map



Conclusion:

- **Style-level transfer learning (STL)**

  Is designed to generate fake images and establish the effective knowledge transfer between source and target domains.

- **Scene-aware Estimation (SAE)**

  Introduces scene classification to alleviate the impact of crowd distribution on people counting. These steps together achieve the goal of improving the accuracy and generalization of domain adaptation people counting.

- **DAPC**

  Significantly outperforms the state-of-the-art methods. In the future, we will continue to work on people counting and will pay attention to the optimization of efficiency.

Thank you!

ICPR 2020

Na Jiang , Xingsen Wen , Zhiping Shi

Capital Normal University