**Academic Journal of Science and Technology**
ISSN: 2771-3032 | Vol. 5, No. 2, 2023

# A Survey of Crowd Counting Algorithm Based on Domain Adaptation

**Yanan Wang[1], Fen Luo[1, *]**

[1] School of Software, Henan Polytechnic University, China

[*] **Corresponding author**: Fen Luo (Email: luofenjsj@hpu.edu.cn)

**Abstract:** Crowd counting, the task of estimating the number of individuals in a crowded scene, has gained increasing attention in computer vision research. However, crowd counting remains a challenging problem due to the complex and diverse nature of crowd scenes. In recent years, domain adaptation has emerged as a promising approach to improve crowd counting performance by adapting a pre-trained model to a target domain with different characteristics. This paper provides a survey of domain adaptation-based crowd counting algorithms, including their methods, datasets, and evaluation metrics. Overall, domain adaptation shows great potential in improving the accuracy and robustness of crowd counting algorithms, and further research in this direction is expected to lead to more effective and practical crowd counting solutions.

**Keywords:** Crowd counting, Domain adaptation.

## 1. Introduction

With the accelerating urbanization and continuous population growth, the problem of crowd counting has become a widely researched area. Traditional crowd counting methods mainly rely on manually designed features and models, which often fail to adapt to the changes and complexities of different scenarios, resulting in low counting accuracy. Therefore, crowd counting methods based on deep learning have been widely applied, achieving excellent results by learning more effective feature representations and building more accurate models.

However, in practical applications, factors such as differences between different scenarios, distribution differences between different datasets, etc., may lead to a decrease in model performance, which requires domain adaptation. Domain adaptation is a method of transferring knowledge learned in the source domain to the target domain, improving the generalization performance of the model by solving domain differences.

Therefore, this paper mainly reviews the research status and development trends of crowd counting algorithms based on domain adaptation. This paper categorizes domain-adaptive crowd counting methods into two types: one based on distribution strategy framework, and the other based on image transformation strategy framework. Firstly, the two methods are briefly introduced, then the commonly used datasets and evaluation indicators in population counting are introduced, and finally, the development direction of population counting based on domain adaptation is reasonably predicted.

## 2. Methods

### 2.1. Methods based on the distribution strategy

The first type of distribution strategy-based framework mainly uses discriminators to distinguish the density maps generated from the source and target domain images to align the data distribution.

Wang et al. [1] proposed an adversarial learning method, namely CODA (adaptation to counting objects through scale-perceived adversarial density). To handle different target scales and density distributions, through adversarial training by adapting multi-scale pyramid pyramid patches from the source and target domain, consistent count objects can be produced for different scales along with the ranking constraints of the pyramid input level.

Yan et al. [2] solves the problem of domain shift by training using synthetic images and their associated labels and unlabeled real images, so that it can be born as a label for the purpose of fine-tuning the network.

Zou et al. [3] proposed an adversarial scoring network (ASNet) that gradually closes the cross-domain gap from coarse-to fine-grained. The coarse-grained phase designs a dual discriminator strategy to adapt the source domain from the global and local feature space perspectives close to the target domain. In the fine-grained stage, the transferability of source features is explored by scoring the similarity of source and target samples from multiple levels based on the generation probability derived from the coarse-grained stage.

### 2.2. Deep learning methods

The second type of image transformation strategy-based framework is mainly based on Cycle GAN[4], converting the synthesized images into images with a similar style to the target domain, and then training the transformed images in the population counting task.

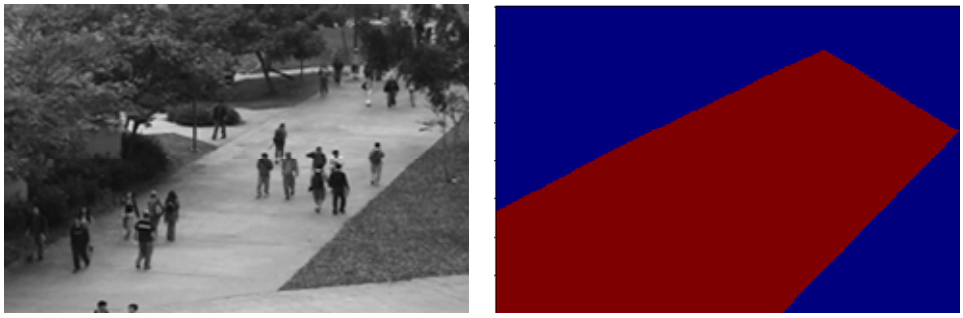
Wang et al. [5] developed a data collector and hashtag that can generate synthetic population scenarios and annotate them without the need for manual labor. Based on this work, a large-scale, diverse synthetic dataset is constructed.

Gao et al. [6] proposed a DACC (Domain Adaptive Crowd Counting) framework composed of high-quality image transformation and density map reconstruction. The former transforms synthetic data into real images to improve conversion quality by separating domain sharing/ independent features and design content perceived consistency loss, while the latter aims to generate pseudo-labels on real scenes, using labels to retrain the final counter to improve count performance.

## 3. Datasets

ShanghaiTech[7]: The dataset consists of two subsets: PartA and PartB. PartA contains 482 images with different resolutions, with 300 in the training set and 182 in the test set. The numbers in each image ranged from 33 to 3139, with an average of 501. The PartA images contain a variety of scenes, including indoor and outdoor scenes, and are mostly crowded images. Similarly, PartB contains 716 images with a resolution size of 768 * 1024, with 400 in the training set and 316 in the test set. The number of people per image ranged from 9 to 578, with an average of 123 people per image. The images of PartB are taken from the city streets of Shanghai.

UCSD[8]: The dataset was collected from a surveillance video recording on campus. Contains 2,000 frames with a resolution of 158 * 238. Number per frame ranged from 11 to 46, with an average of 25 people per frame. This dataset provides a region of interest ROI (Region of Interest, ROI) for each frame that ignores background interference, is trained using the # 601 to # 1400 frames, and the rest are tested. Figure 1 shows the population image and the ROI region of the UCSD.



（a）UCSD dataset image　　（b）ROI regions provided by the dataset

**Figure 1.** Population image and the ROI region of the UCSD.

Mall[9]: Similar to UCSD, the dataset was collected in a surveillance camera in a shopping mall. Mall contains 2000 frame files with a resolution of 240 * 320. Each frame contained numbers ranging from 13 to 53, with an average of 31 per frame. The dataset provides a ROI ignoring background interference for the frame files. Unlike the grayscale frames of the UCSD, the Mall frame file is colored and contains different changes in light conditions. In addition, Mall has perspective distortion and occlusion phenomenon.

QNRF[11]: The dataset is composed of 1,535 images with about 1.25 million annotations. It ranged from 49 to 12865, with an average count of 815.4. Specifically, the training set consisted of 1,201 images, and the test set consisted of 334 images. The images of this dataset species contain a wider range of scenarios and contain the most diverse viewpoint, density, and illumination changes.

## 4. Evaluating Indicator

Mean absolute error (MAE) and root mean square error (RMSE) are two common indicators of population counting algorithm. MAE reflects the accuracy of model prediction, defined as formula (1), RMSE reflects the robustness of algorithm prediction and defined as formula (2): promising results in crowd counting, but further research is needed to develop more efficient and effective techniques for this challenging problem.

$$MAE = \frac{1}{N}\sum_{i=1}^{N} | C_i - C_i^{GT} | \#(1)$$

$$RMSE = \sqrt{\frac{1}{N}\sum_{i=1}^{N} (C_i - C_i^{GT})^2} \#(2)$$

## 5. Conclusion

Although domain adaptation-based crowd counting algorithms have achieved good results, there are still many problems. Firstly, most methods are limited to single-source domain adaptation. However, in practical scenarios, domain shift not only exists between source and target domains, but also between different source domains. Data from different source domains can interfere with each other during mutual learning. Therefore, it is necessary to solve the problem of multi-source domain adaptation in crowd counting. Secondly, current domain adaptation methods are tested on crowd counting datasets in normal environments. However, there may be adverse environments in real-world scenarios, such as low-light, rain, and fog, and the model cannot effectively handle these scenarios. Therefore, the generalization ability of the model in adverse environments needs to be further improved.

## References


[1] Li W, Yongbo L, Xiangyang X. Coda: Counting objects via scale-aware adversarial density adaption[C]//2019 IEEE International Conference on Multimedia and Expo (ICME). IEEE, 2019: 193-198.

[2] Yan Z, Li P, Wang B, et al. Towards learning multi-domain crowd counting[J]. IEEE Trans. Circuits Syst. Video Technol, 2021.Badrinarayanan, V., Kendall, A., Cipolla, R.: 'Segnet: A deep convolutional encoder-decoder architecture for image segmentation', IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39, pp. 2481–2495

[3] Zou Z, Qu X, Zhou P, et al. Coarse to fine: Domain adaptive crowd counting via adversarial scoring network[C]//Proceedings of the 29th ACM International Conference on Multimedia. 2021: 2185-2194.Ghiasi,

[4] Zhu J Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks[C]//Proceedings of the IEEE international conference on computer vision. 2017: 2223-2232.

[5] Wang Q, Gao J, Lin W, et al. Learning from synthetic data for crowd counting in the wild[C]//Proceedings of the IEEE/CVF

conference on computer vision and pattern recognition. 2019: 8198-8207.

[6] Gao J, Han T, Yuan Y, et al. Domain-adaptive crowd counting via high-quality image translation and density reconstruction[J]. IEEE transactions on neural networks and learning systems, 2021.

[7] Zhang Y, Zhou D, Chen S, et al. Single-image crowd counting via multi-column convolutional neural network [C] //Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 589-597.

[8] Chan A B, Liang Z S J, Vasconcelos N. Privacy preserving crowd monitoring: Counting people without people models or tracking[C]//2008 IEEE conference on computer vision and pattern recognition. IEEE, 2008: 1-7ision, 2021, 129, (11), pp. 3051–3068

[9] Zhang Y, Zhou D, Chen S, et al. Single-image crowd counting via multi-column convolutional neural network [C] // Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 589-597.

[10] Caesar, H., Uijlings, J., Ferrari, V.: 'Coco-stuff: Thing and stuff classes in context',Proceedings of the IEEE conference on computer vision and pattern recognition,2018, pp. 1209–1218