ASIAN JOURNAL OF ELECTRICAL AND ELECTRONIC ENGINEERING
Vol. 4 No. 2 2024
E-ISSN: 2785-8189

Improving Crowd Counting Performance: A Convolutional Neural Network Approach with Transfer Learning

Marwah M. Ahmeed[1] and Othman O. Khalifa[2]

[1]*Collage Of Electronic Technology, Bani Walid, Libya*
[2]*Libyan Center for Engineering Research and Information Technology, Bani Walid, Libya*

*Corresponding author: ookhalifa@gmail.com

(Received: 16 August 2024; Accepted: 27 August 2024)

Abstract— Precise crowd counting is critical to public safety and smart city planning since it solves the problems associated with the time-consuming manual counting of people in photos and videos. Transfer learning has become a key building block for improving crowd counting techniques, especially when used to Convolutional Neural Networks (CNNs). Because pretrained models already know the pertinent weights and architecture, using them in transfer learning minimizes computational demands and shortens training time. This paper presents a crowd counting method with an emphasis on optimizing the VGG16 model with a mall dataset. The results show that using VGG16 for transfer learning leads to higher performance when compared to more modern methods like AdaCrowd and PSSW models. In addition, the paper highlights how adaptable our proposed method is and how well it can transfer knowledge from one dataset to another.

Keywords: *Crowd Counting, VGG16, Enhanced Performance, Smart Cities, Public Safety, Convolutional Neural Networks (CNNs)*

1. INTRODUCTION

Crowd counting is a challenging task in computer vision that involves estimating the number of people present in each scene or image. It has a wide range of applications, such as traffic monitoring, crowd management, and security surveillance. The computer vision community has recently given crowd counting much attention, and several solutions have been put out to address this issue [1]. Traditional approaches to crowd counting involve manually designing features and using handcrafted algorithms for counting people [2]. However, these methods often suffer from low accuracy and scalability issues when dealing with complex scenes with many people. With the recent advances in deep learning, there has been a shift towards using deep neural networks for crowd counting [3]. The most advanced solution for this task is now deep learning-based, thanks to their impressive performance in crowd counting [4].

Deep neural networks require large amounts of labelled data to achieve high accuracy. However, collecting and annotating a large crowd counting dataset is a challenging and time-consuming task. Furthermore, the diversity and complexity of real-world crowd scenes make it difficult to capture and label all possible scenarios [5]. Therefore, transfer learning, a technique that enables the transfer of knowledge from one task to another, has become a popular approach in deep learning-based crowd counting [6].

Transfer learning can be used to leverage pre-trained deep neural networks that have been trained on large-scale datasets such as ImageNet [7]. The pre-trained models have already learned to recognize high-level features such as edges, shapes, and textures, which are also relevant to crowd counting. By fine-tuning the pre-trained models on a small crowd counting dataset, we can achieve high accuracy with limited labeled data [1]. This approach can significantly reduce the time and effort required for collecting and labeling crowd counting data [5]. In recent years, various pre-trained models have been proposed for transfer learning-based crowd counting. Among them, VGG16, a deep residual network with 16 layers, has shown promising results in several computer vision tasks.[8]. VGG16 has achieved state-of-the-art performance on the ImageNet dataset, and its deep architecture allows it to capture

DOI: https://doi.org/10.69955/ajoeee.2024.v4i2.63

high-level features in complex scenes [4]. Therefore, we propose to use VGG16 for transfer learning-based crowd counting in this research.

2. RELATED WORK

Crowd counting has become a crucial computer vision problem in recent years, with applications found in a variety of fields including public safety, event management, and surveillance. Scholars have investigated a range of approaches to improve the precision and effectiveness of crowd counting models. This section examines seminal research that addresses the problems related to crowd counting by utilizing Convolutional Neural Networks (CNNs) and transfer learning approaches. Early crowd counting research frequently used conventional computer vision methods. But the emergence of deep learning—and CNNs in particular—marked a paradigm change in this field. In a groundbreaking study, [5] presented the application of a multi-column CNN architecture for crowd counting.

This established the groundwork for later research to investigate CNNs' capacity to manage complicated scenarios with a range of pedestrian sizes. Khalifa et al. [15] explored transfer learning for crowd counting using ResNet50 on the Mall dataset. While transfer learning reduced the computational burden and training time, the results showed mediocre Mean Absolute Error (MAE) and Mean Squared Error (MSE) compared to other recent techniques. Further improvements are required to make this approach more beneficial. However, it's worth noting that not all transfer learning approaches yield the same results. Additionally, Feng et al. [9] proposed a new deep learning model called Spatiotemporal Convolutional LSTM (ConvLSTM) for crowd counting in videos. This model captures both spatial and temporal dependencies in crowd counting videos and has shown improved accuracy compared to traditional ConvLSTM models.

Table I. Summary of Related Work

Authors/ Year	**Methodology**	**Strengths**	**Limitations**
Yingying, *et al*, 2016 [18].	Multi-column CNN for single-image crowd counting (MCNN)	Effective for estimating crowd counts from single images	Limited to single images; may not capture temporal dynamics in videos
Deepak, *et al*, 2017 [10].	Switch-CNN	Improved accuracy through fine-tuning	Limited to specific architecture (VGG16)
Vishwana, et al, (2017 [11]	Contextual pyramid CNN for generating crowd density maps(CP-CNN)	High-quality crowd density maps; Improved accuracy	May require large-scale datasets for pre-training
Yuhong, et al, 2018 [12].	CSRNet with dilated CNNs for congested scenes	State-of-the-art accuracy and efficiency	May require extensive computational resources
Feng, et al, 2017 [9].	Spatiotemporal ConvLSTM for crowd counting in videos	Captures spatial and temporal dependencies; Improved accuracy	May be computationally intensive
Mahesh, et al, 2021 [17[.	AdaCrowd	Adapts crowd counting models to new, unlabelled target scenes using adversarial learning; Overcomes lack of labelled target data	Adversarial learning complexity; Network training complexity
Zhen, et al, 2020 [16].	PSSW	Accurate crowd counting with limited labelled data	Framework complexity; May require substantial computation
Khalifa, et al, 2022 [14].	Transfer learning with ResNet50 on Mall dataset	Reduced computational burden; Faster training	Mediocre performance compared to other techniques; Further improvements needed
Lijia Deng Yudong Zhang (2020), [13].	FOCNN	Superior performance in low-density scenarios	Limited applicability to specific crowd scenarios

Furthermore, Zhao et al. [16] proposed an active learning framework for crowd counting that combines several innovative components, including partition-based sample selection, density regression module, domain classification module, and Mix-up regularization. This framework enables accurate crowd counting with very limited labelled data.

Lastly, Reddy et al. [17] introduced AdaCrowd, a method that addresses the challenge of adapting crowd counting models to new, unlabeled target scenes. It leverages unlabeled target data during training and employs teacher-student learning framework combined with an adaptation module based on adversarial learning. These studies [9,14,16,18] provide valuable insights and methodologies for applying transfer learning in crowd counting tasks. They highlight the importance of leveraging pre-trained CNN architectures and fine-tuning them on crowd counting datasets to achieve improved accuracy while reducing training time and computational complexity.

Table 1 shows the summary of the methodology and advantages and disadvantages of published articles that are used as references for this literature review.

3. METHODOLOGY AND PROPOSED SOLUTION

In this work, a methodical strategy utilizing Convolutional Neural Networks (CNNs) and Transfer Learning is used to improve crowd counting performance. Figure 1 shows the proposed Solution.

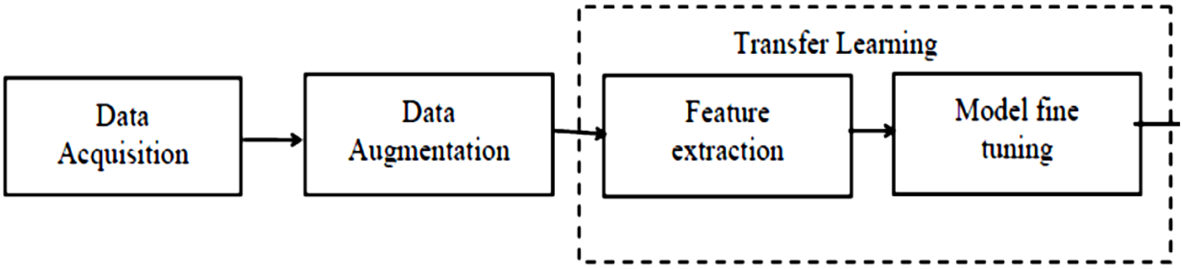

Fig. 1. Proposed Solution

A. Dataset Acquisition

The Mall dataset, a collection of surveillance images captured from a shopping mall, was obtained from https://paperswithcode.com/dataset/mall. The Mall dataset contains 2000 annotated images, and all images have a resolution of 320 x 240 each accompanied by a corresponding crowd count label. These labels accurately depict the number of individuals present within each image, ranging from a minimum of 11 people to a maximum of 53 people. The dataset's comprehensive annotations facilitate supervised learning, allowing us to train and evaluate crowd counting models effectively.

B. Data Augmentation

To enhance the model's generalization capabilities and reduce overfitting, various data augmentation techniques were employed, including Rotation, Scaling, and flipping. These techniques increase the diversity of the training data, enabling the model to learn robust features.

C. Image **Preprocessing**

Preprocessing the images is essential to optimize model performance. The following preprocessing techniques were applied to the images:

Resizing: All images were resized to a fixed dimension (e.g., 224x224 pixels) to ensure uniformity and compatibility with the VGG16 model architecture.

Rescale (Pixel Rescaling): This operation involves rescaling pixel values to a specific range, such as [0, 1]. It is typically expressed by the equation for each pixel.

If X is the original pixel value and X_rescaled is the rescaled pixel value: X_rescaled=X/255

ZCA Whitening: ZCA Whitening is a method that reduces data variance and enhances data quality. Feature-wise Standardization normalizes the distribution of features in the dataset, ensuring that each feature has a mean of zero and a standard deviation of one. Sample-wise Standardization normalizes data on a per-sample basis. It ensures that each sample in the dataset has a mean of zero and a standard deviation of one.

D. Model *Architecture* Design

The selection of an appropriate model architecture significantly impacts the crowd counting system's performance. In this paper, an enhanced VGG16 model, a deep convolutional neural network renowned for its effectiveness in computer vision tasks, was chosen as the backbone architecture. Figure 3 shows the VGG16 model after it was modified.

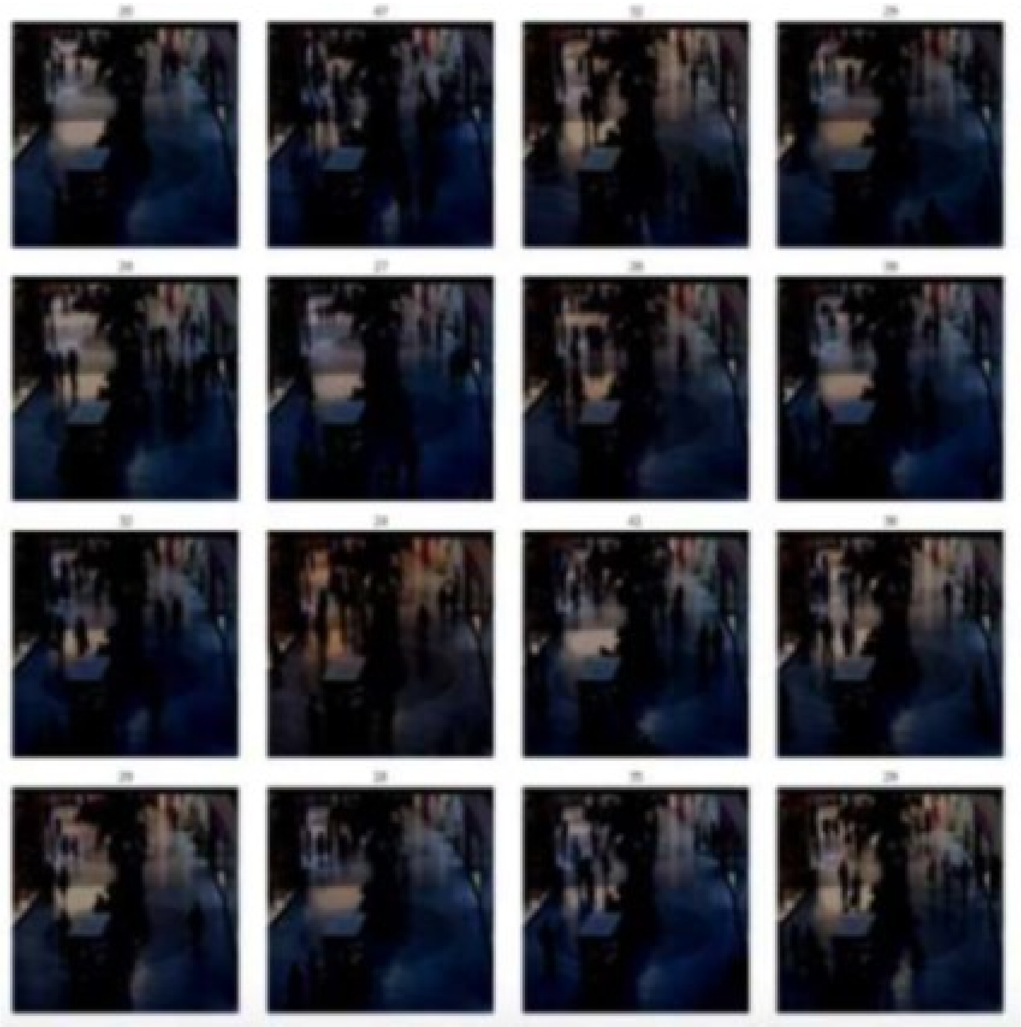

Fig. 2. Image Preprocessing view.

4. RESULTS ANALYSIS

To evaluate the performance of crowd counting models, mean absolute error (MAE) and mean squared error (MSE) are the most used parameters. MAE and MSE are defined as

$$MAE = \frac{1}{N}\sum_{i-1}^{N}|y_i - \hat{y}|$$

$$MSE = \frac{1}{N}\sum_{i-1}^{N}(y_i - \hat{y})^2$$

Where N is the number of test images

y_i is the number of actual people in the image and

y ̂ is the number of people estimated to be in the image

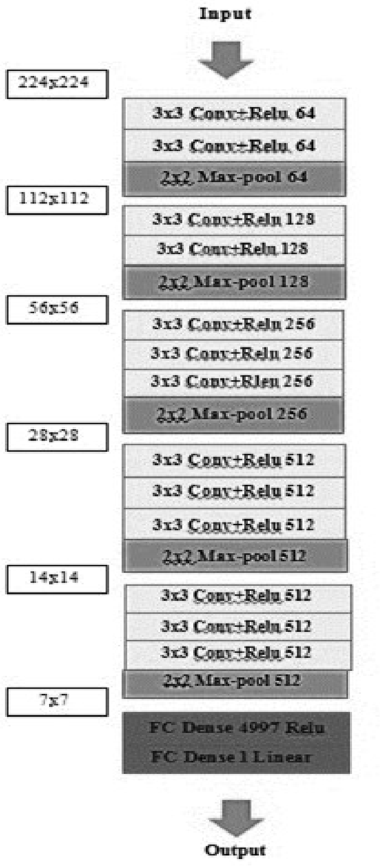

Fig. 3. Modified VGG-16 Architecture

4.1 *Simulation Parameters Setup*

A meticulous enumeration and elucidation of the parameters adopted are imperative. These parameters should be systematically presented in a table, accompanied by a clear elucidation of the type of Convolutional Neural Network (CNN) employed and its specific configurations. Table II shows the Optimized Parameters

Table II. Optimized Parameters

Parameter	**Range**
CNN Type	VGG16
Mini Batch Size	32
Optimizer	ADAM
Initial Learn Rate	0.001
Beta 1	0.8
Beta 2	0.95
Number Of Epochs	50
Number Of FC Layers	2

4.2 Finding and Outcome

The model is trained for 50 epochs and the model achieves 1.6 MAE, 4.1 MSE, 5.3 MAPE, 2.0 RMSE and 0.915 R^2. Figure 4 & 5 shows the deployment of the trained model in Smart City Dataset. However, the predicted count is more than the actual count.

5. COMPARATIVE PERFORMANCE ANALYSIS

The comparative performance analysis is shown in the table. Each method's performance is measured using metrics like Mean Absolute Error (MAE) and Mean Squared Error (MSE). Interestingly, the proposed method works better than the current methods, obtaining lower MSE and MAE values and indicating improved crowd density estimation accuracy. However, this paper offers insightful information about how well the strategy offered performs when compared to cutting-edge techniques from various years of publication.

Figure 6 shows the deployment of the trained model in the Beijing BRT dataset. However, the predicted count is less than the actual count.

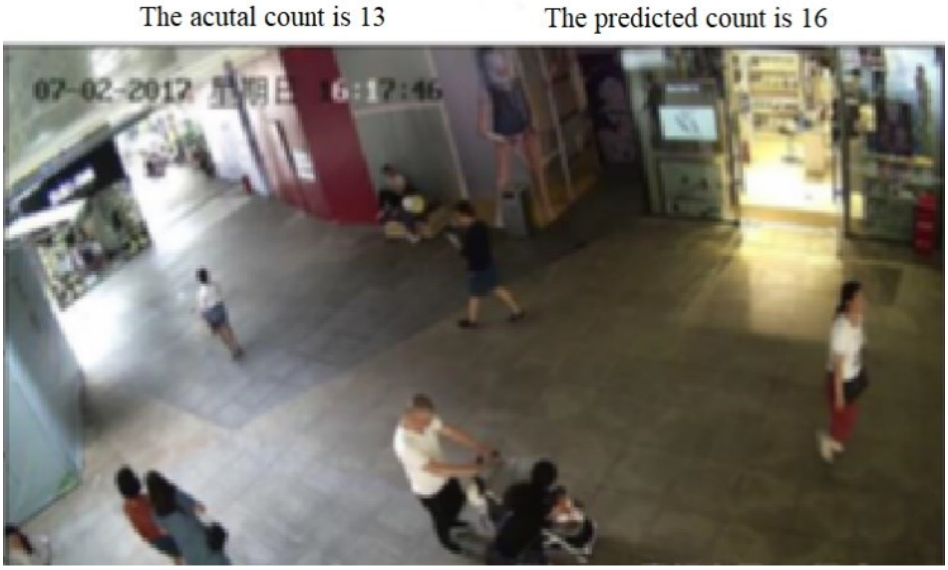

Fig. 4. An image Smart city dataset

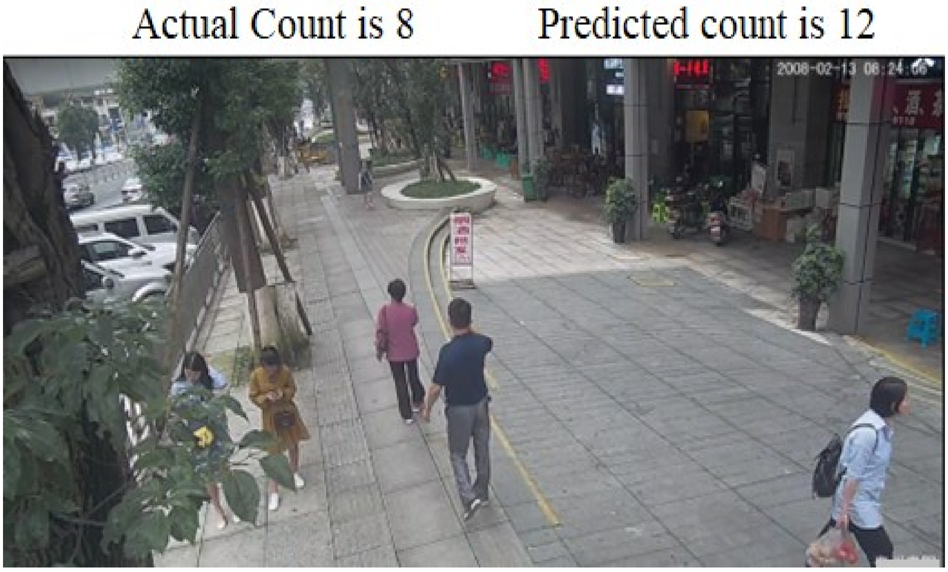

Fig. 5. An image Smart city dataset

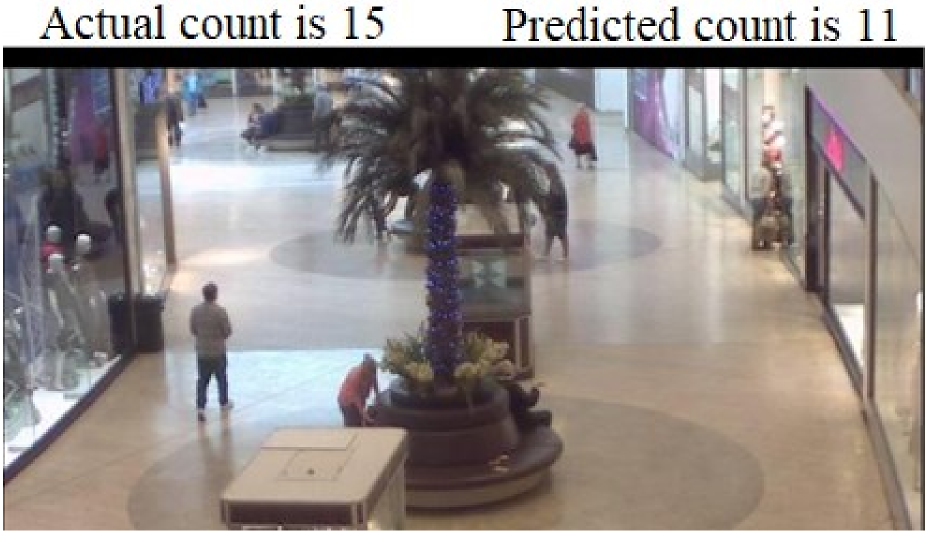

Fig. 6. An image of the Beijing BRT dataset

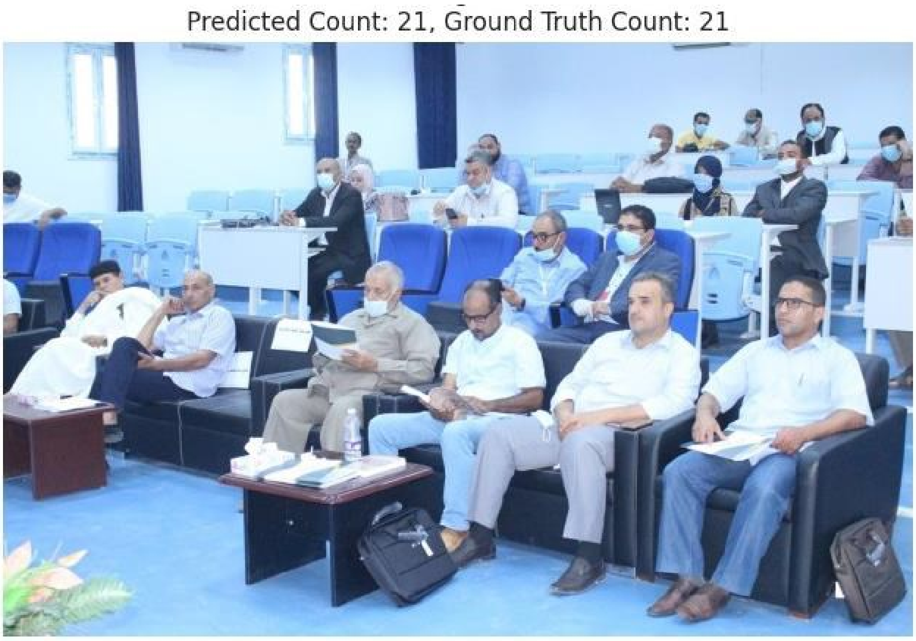

Fig. 7. An image from the College of Electronic Technology Bani Walid, Libya

Table III. Comparative Performance Analysis of Crowd Counting Methods

Methods	**MSE**	**MAE**	**YEAR**
AdaCrowd [17]	5	4	2021
MCCN [18]	8.5	2.24	2016
PSSW [16]	5.4	3.8	2020
Bidirectional ConvLSTM [9]	7.6	2.10	2017
ResNet50 with transfer learning [14]	15.5	3.31	2022
Proposed Method	**4.19**	**1.60**	**2023**

6. CONCLUSION

Crowd counting is a critical component of smart city planning and public safety. The conventional manual counting methods are resource-intensive, prompting the adoption of advanced techniques like transfer learning via Convolutional Neural Networks (CNNs) for crowd counting. This paper contributes to the field by presenting an innovative approach using the VGG16 model fine-tuned on a mall dataset. Transfer learning is known to be effective for problems involving crowd counting, providing significant computational and training time savings. Utilizing a pretrained model with predetermined weights and architecture, like VGG16, offers a strong basis for precise assessment of crowd density. The finding presented in this paper demonstrates the effectiveness of the proposed method and validate its efficacy.

REFERENCES

[1] C. Zhang, H. Li, X. Wang, and X. Yang, "Survey on crowd counting: Methods and datasets," Neurocomputing, vol. 399, pp. 67-89, 2020.

[2] A. B. Chan, Z. Q. Liang, and N. Vasconcelos, "Privacy preserving crowd monitoring: Counting people without people models or tracking," in European Conference on Computer Vision, 2008, pp. 412-425. https://doi.org/10.1109/CVPR.2008.4587569

[3] D. Wang, D. Zhang, Y. Chen, C. Zhang, and F. Yang, "Comprehensive study on convolutional neural network-based crowd counting methods," Neurocomputing, vol. 375, pp. 270-285, 2020.

[4] N. Liu, J. Zhang, K. Huang, and Z. He, "A Review on Deep Learning for Crowd Counting," IEEE Access, vol. 9, pp. 55801-55818, 2021.

[5] C. Zhang et al., "Data-driven crowd understanding: A baseline for a large-scale crowd dataset," IEEE Trans. Multimed., vol. 18, pp. 1048-1061, 2016. https://doi.org/10.1109/TMM.2016.2542585

[6] V. A. Sindagi and V. M. Patel, "A survey of recent advances in CNN-based single image crowd counting and density estimation," Pattern recognition letters, vol. 107, pp. 3-16, 2018. https://doi.org/10.1016/j.patrec.2017.07.007

[7] J. Pan, S. Liu, D. Sun, J. Yang, and C. C. Loy, "Crowd sampling the parameter space of deep neural networks for robustness," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019, pp. 11205-11214.

[8] J. Deng et al., "Imagenet: A large-scale hierarchical image database," in 2009 IEEE conference on computer vision and pattern recognition, 2009, pp. 248-255. https://doi.org/10.1109/CVPR.2009.5206848

[9] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2016, pp. 770-778. https://doi.org/10.1109/CVPR.2016.90

[10] X. Feng, X. Shi, and D. Yeung, "Spatiotemporal modeling for crowd counting in videos," in ICCV, 2017, pp. 5161-5169.

[11] A. Sam, S. Surya, and R. V. Babu, "Switching convolutional neural network for crowd counting," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017. https://doi.org/10.1109/CVPR.2017.429

[12] V. A. Sindagi and V. M. Patel, "Generating high-quality crowd density maps using contextual pyramid CNNs," in Proceedings of the IEEE International Conference on Computer Vision, 2017. https://doi.org/10.1109/ICCV.2017.206

[13] 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. https://doi.org/10.1109/CVPR.2018.00120

[14] L. Deng, S. H. Wang, Y. D. Zhang, "Fully optimized convolutional neural network based on small-scale crowd," presented at the 2020 IEEE International Symposium on Circuits and Systems (ISCAS), 2020. https://doi.org/10.1109/ISCAS45731.2020.9180823

[15] O. O. Khalifa, A. Albagul, A. H. Abdallah Hashim, N. Abdul Malik Hashim and K. N. Sakinahbt Wan Zainuddin, "Transfer Learning for Crowed Counting," 2022 IEEE 2nd International Maghreb Meeting of the Conference on Sciences and Techniques of Automatic Control and Computer Engineering (MI-STA), Sabratha, Libya, 2022, pp. 248-253. https://doi.org/10.1109/MI-STA54861.2022.9837673

[16] X. Feng, X. Shi, and D. Yeung, "Spatiotemporal modeling for crowd counting in videos," in ICCV. IEEE, 2017, pp. 5161-5169.

[17] Z. Zhao et al., "Active crowd counting with limited supervision," presented at the ECCV 2020: 16th European Conference on Computer Vision, 2020. https://doi.org/10.1007/978-3-030-58565-5_34

[18] M. K. Krishna Reddy et al., "AdaCrowd: unlabeled scene adaptation for crowd counting," IEEE Transactions on Multimedia, vol. 24, pp. 1008-1019, 2022. https://doi.org/10.1109/TMM.2021.3062481