e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:03/March-2024 Impact Factor- 7.868 www.irjmets.com

CROWD COUNTING METHOD BASED ON CONVOLUTION NEURAL NETWORK (CNN)

W. Nirmala[*1], B. Santhosh[*2], Chand Fathima[*3], CH. Vaishnavi[*4]

[*1]Associate Professor Department Of Computer Science And Engineering Malla Reddy College Of Engineering & Technology Hyderabad, India.

[*2,3,4]Final Year Student Department Of Computer Science And Engineering Malla Reddy College Of Engineering & Technology Hyderabad, India.

ABSTRACT

Crowd counting remains a critical research domain in computer vision, where the accuracy of existing methods, particularly multi-column convolution neural networks (MCNN), encounters challenges in scenarios with uneven crowd distributions. Addressing this, our paper integrates crowd global density features into the MCNN framework using cascaded learning. Additionally, we propose an enhanced MCNN architecture featuring max-ave pooling and deconvolutional layers to preserve detailed features lost during down-sampling. Experimental validations on UCF_CC_50 and ShanghaiTech datasets demonstrate superior accuracy and stability of our approach.

Furthermore, we review the evolution of crowd-counting techniques, emphasizing the transition from traditional handcrafted methods to intelligent machine-learning-based approaches, particularly convolutional neural networks (CNNs). Despite facing challenges like occlusion and clutter, CNNs offer promising solutions for intelligent crowd counting and analysis, facilitating adaptive monitoring and management of dynamic crowd gatherings

Our study also discusses the current challenges faced by crowd counting systems, especially those employing density estimation, such as perspective distortion and ineffective population density determination methods. To address these issues, we propose leveraging CNNs for crowd counting through detection, clustering, and regression techniques, promising more robust results by leveraging CNN's advanced image processing capabilities

Given the increasing importance of video surveillance for safety and traffic control, our findings underscore the significance of advancing crowd counting methodologies, particularly leveraging CNNs, to meet evolving surveillance needs and ensure accurate monitoring in dynamic environments.

I. INTRODUCTION

Crowd counting, the process of estimating the number of individuals in images or video frames, plays a crucial role in various domains such as surveillance, traffic management, and event monitoring. Over the years, crowd counting methods have evolved, broadly categorized into three approaches: direct counting via target detection, indirect methods relying on feature regression, and more recently, crowd counting based on deep learning. In the realm of target detection-based approaches, researchers like Line et al. and Kowalak et al. have proposed methods utilizing Haar wavelet transform and shape contours for crowd detection and density estimation. While effective for low-density crowds, these methods encounter accuracy challenges in high-density scenarios. Feature regression-based techniques, championed by Chan et al. and Idrees et al., establish regression relationships between image features and crowd counts, enhancing robustness and adaptability. The introduction of datasets like UCF_CC_50 has facilitated advancements in this area. With the advent of deep learning and big data, crowd counting methodologies have witnessed a paradigm shift towards deep learning-based approaches. Zhang et al. introduced cross-scene crowd counting models, Zhang et al. leveraged multi-column convolution neural networks (MCNN) for scale-adaptive crowd counting, and Boominathan et al. combined shallow and deep CNN features to enhance spatial resolution. Further innovations include Switch-CNN by Sam et al., aggregating multiscale features proposed by Shi et al., leveraging LSTM for contextual information extraction by Fu et al., introducing attention modules for adaptive counting modes by Liu et al., and MMCNN by Yang et al. for robust counting, considering location, detail, and scale variation. However, despite

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:03/March-2024 Impact Factor- 7.868 www.irjmets.com

these advancements, existing methods falter when faced with uneven crowd distributions. To address this challenge, our paper proposes extracting global density features and integrating them into a convolutional neural network framework using multi-task network cascades (MNCs). Additionally, we incorporate max-ave pooling layers to preserve image features and deconvolutional layers to restore lost details during down-sampling, thus enhancing density map accuracy and overall crowd counting precision.

II. LITERATURE REVIEW

In recent years, some important progress has been made in the field of crowd counting and crowd density estimation, but there are still many challenging problems. With the increase of crowd density, there will be serious occlusion in the human body. In addition, the complexity of the background environment of the image, such as noise, uneven light and human body deformation, will affect the final detection results. To solve these problems, researchers have proposed many estimation methods for different problems, e.g., crowd counting based on detection method crowd counting based on regression method , and density estimation method the detection-based method uses bounding boxes to accurately locate each person in the image, and the regression-based method directly outputs the number of people in an image or the density map of the corresponding image. However, in the complex environment or larger crowded scenes, traditional methods are often unable to achieve accurate estimation. In recent years, CNN has been widely used in a great number of applications (image recognition, object detection, image segmentation, etc. Due to CNN's strong learning ability, it can learn more feature information, which makes it perform well in the field of crowd counting. Different from the traditional methods, the convolution neural network method can extract high-level features conveniently and efficiently, and consequently obtain better performance than the traditional methods. For the dense crowd area in the image, the CNN method is used to get better prediction results. At present, the crowd counting method based on CNN has become more popular than before.

In previous works, literatures have studied and analyzed the traditional crowd counting methods. Loy et al. classified the traditional crowd counting methods into three categories, i.e., the detection-based, regression-based, and density estimation-based approaches. Moreover, they evaluated the methods based on image and video, and analyzed the advantages and disadvantages of each method. Ryan et al focused on evaluating regression models with different features such as the local, global, and histogram features, and compared the effects of different features on regression results. Zitouni et al. investigated and analyzed general aspects of techniques of crowd counting instead of specific algorithm. Li et al. and Saleh et al reviewed crowd counting and density estimation methods under visual monitoring. Kang et al analyzed the CNN-based density estimation methods for crowd counting, detection and tracking. Sindagi et al. summarized the existing methods for crowd counting and density estimation in a single image, but they did not analyze detection-based methods ,the methods based on video and methods based on detection and regression In our work, we systematically and comprehensively analyzed the crowd counting and density estimation methods based on CNN, and also discussed the traditional methods that were not analyzed before.

Year	Author	Title	Method
2015	Pawel gard zinki Krzyzt of kowalak	Crow density estimation hased on Voxel model in multi view Surveillance systems.	Voxel modelling
2016	Yingying zhang Yi ma Siqin chan	Single-Image Crowd Counting Via multi Column CNN	Multi column CNN
2017	Deepak Babu Sam Shiv Surya R.Venkatesh Babu	Switching convolutional neural network for Crowd Counting	CNN Regressors
2018	Jingnan Fu, Hongbo young,	A CNN-RNN neural network join long-short term memory for crowd Counting and	CNN Regressions wing MCNN and LSTM

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:03/March-2024 Impact Factor- 7.868 www.irjmets.com

	ping Liu Yuzhun Hu	density estimation	

III. METHODOLOGY

FRCNN ALGORITHM

Here we use FRCNN algorithm as we know neural network is about all the the division of layers of network. CNN is about 3 layers , it can have multiple layers and then feature extraction happens then it produces the output of counting the crowd with best accuracy. So we use faster R-CNN (Region-based Convolutional Neural Network) here.

Faster R-CNN is a deep convolutional network used for object detection, that appears to the user as a single, end-to-end, unified network. The network can accurately and quickly predict the locations of different objects. In order to truly understand Faster R-CNN, we must also do a quick overview of the networks that it evolved from, namely R-CNN and Fast R-CNN.

The article starts by quickly reviewing the region-based CNN (R-CNN), which is the first trial towards building an object detection model that extracts features using a pre-trained CNN. Next, Fast R-CNN is quickly reviewed, which is faster than the R-CNN but unfortunately neglects how the region proposals are generated. This is later solved by the Faster R-CNN, which builds a region-proposal network that can generate region proposals that are fed to the detection model (Fast R-CNN) to inspect for objects.

CNN with global density features:-

Global density features typically refer to some statistical or density-based information about the entire input image rather than local features extracted by individual filters. CNN base network : Design your CNN architecture for the main image features. This typically involves convolutional layers for extracting local features followed by fully connected layers for classification or regression. Global density based features: Compute global density features from the entire image. This could involve statistical measures like mean, standard deviation, or other density-related features. You might also consider using global pooling layers to aggregate information from the entire image. Combine features: Concatenate or merge the features extracted by the CNN with the global density features. This can be done before the final fully connected layers. It is density oriented, as if a person enter slowly then identifies and makes count of it.

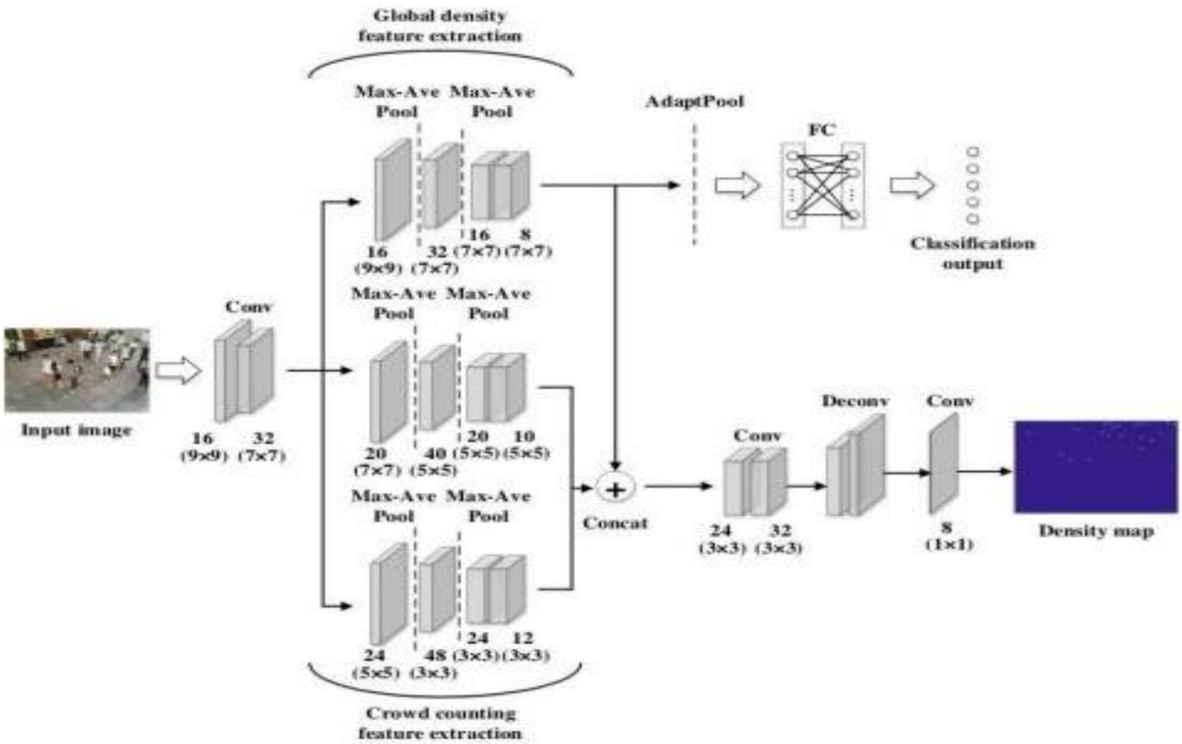
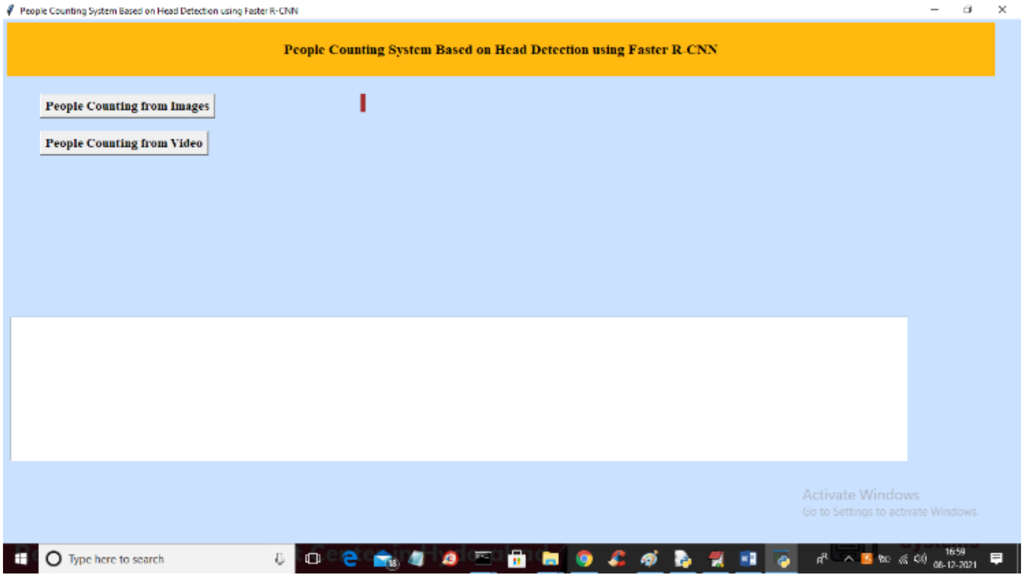
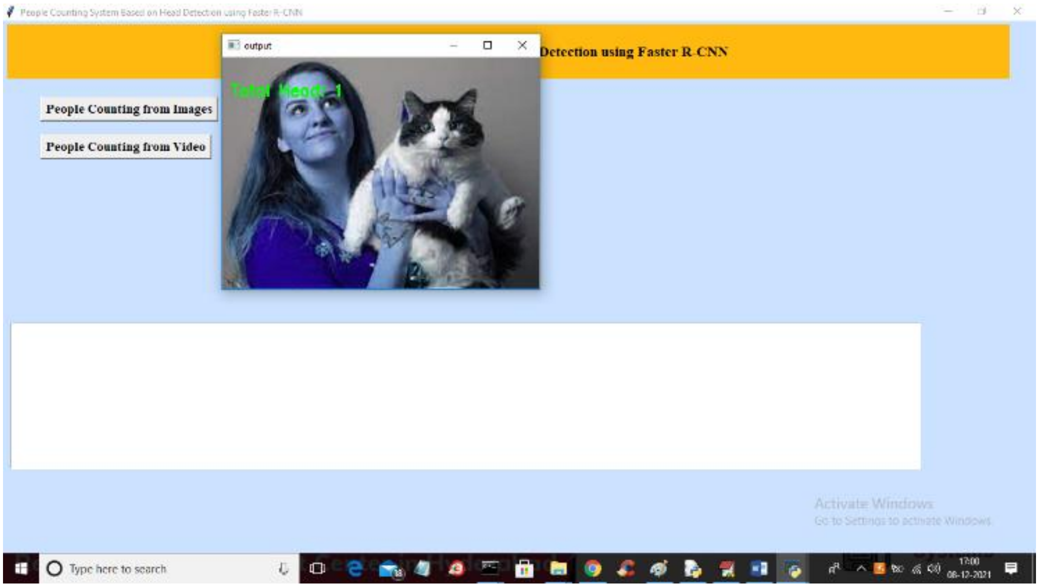


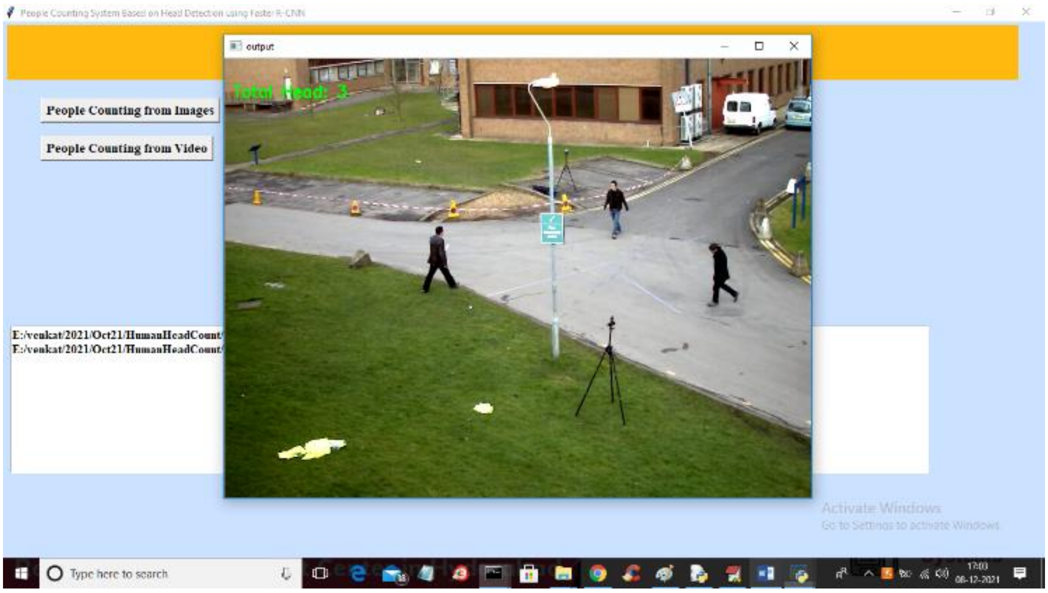
Fig: System Architecture

e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:03/March-2024 Impact Factor- 7.868 www.irjmets.com

IV. RESULT AND ANALYSIS







e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:06/Issue:03/March-2024 Impact Factor- 7.868 www.irjmets.com

V. CONCLUSION

In this paper, an improved convolutional neural network combined with global density feature is proposed. It is different from existing crowd counting methods. The proposed method focuses on uneven crowd distribution. Moreover, the max-ave pooling and de convolutional layers are used to generate a more comprehensive density map. The experimental results show that the proposed method achieves competitive performance on different crowd datasets. Due to the high density crowd, some backgrounds will be taken as people by mistakes. It will bring about noise in the estimated density map and influence the counting results. For the future work, we will focus on reducing the noise in the estimated density map and improving the accuracy of counting.

VI. REFERENCES

[1] S. F. Lin, J. Y. Chen, and H. X. Chao, "Estimation of number of people in crowded scenes using Perspective transformation," IEEE Trans. Syst., Man Cybern. A, Syst., Humans, vol. 31, no.6, pp. 645-654, Nov. 2001.

[2] G. Paweł, K. Kowalak, K. Łukasz, and S. Maćkowiak, "Crowd density estimation based on voxel model in multi-view surveillance systems," in Proc. IEEE Int. Conf. Syst., Signals, Image Process., Sep. 2015, pp. 216- 219.

[3] L. Zhang, M. Shi, and Q. Chen, "Crowd Counting via Scale-Adaptive Convolutional Neural Network," in Proc. IEEE Winter Conf. Appl.Comput. Vis., Mar. 2018, pp. 1113-1121.

[4] T. Zhao, R. Nevatia, and B. Wu, "Segmentation and tracking of multiple humans in crowded environments," IEEE Trans. Pattern Anal. Mach. Intell., vol. 30, no.7, pp. 1198-1211, Jul. 2008.

[5] X. Kong, X. Song, F. Xia, H. Guo, J. Wang and A. Tolba, "LoTAD: long-term traffic anomaly detection based on crowdsourced bus trajectory data," World Wide Web, vol. 21, no. 3, pp. 825–847, May. 2018.

[6] A. B. Chan, Z. S. J. Liang, and N. Vasconcelos, "Privacy preserving crowd monitoring: Counting people without people models or tracking," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2008, pp. 1-7.

[7] A. B. Chan, and N. Vasconcelos, "Counting people with low-level features and Bayesian regression," IEEE Trans. Image Process., vol. 21, no. 4, pp. 2160-2177, Apr. 2012.

[8] H. Idrees, I. Saleem, C. Seibert, and S. Mubarak, "Multi-source multiscale counting in extremely dense crowd images," in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2013, pp. 2547-2554.

[9] W. Ma, L. Huang, and C. Liu, "Crowd density analysis using cooccurrence texture features," in Proc. IEEE Int. Conf. Comput. Sci., Convergence Informat. Techn., Nov. 2010, pp. 170-175.

[10] D. Kong, D. Gray, and H. Tao, "A viewpoint invariant approach for crowd counting," in Proc. Int. Conf. Pattern Recognit., Aug. 2006, pp. 1187-1190.