ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN2454-9940
www.ijasem.org
Vol 18, Issue.2 April 2024

CROWD COUNTING METHOD BASED ON CONVOLUTIONAL NEURAL NETWORK WITH GLOBAL DENSITY FEATURE

[1]Dr.REHMAN PASHA, [2]AITLA NARESH,[3]B NAGA DEVI PRIYA,[4]BHANDARU SAI KARTHIK,[5]ANKALA VENKAT

[1]Professor,Department Of CSE,Malla Reddy Institute Of Engineering And Technology(autonomous),Dhulapally,Secundrabad, Telangana, India, Rehaman17@gmail.com

[2,3,4,5]UG Student,Department Of CSE,Malla Reddy Institute Of Engineering And Technology(autonomous),Dhulapally,Secundrabad, Telangana, India.

ABSTRACT:

An important focus within computer vision research lies in crowd counting methodologies. Among these, the multi-column convolutional neural network (MCNN) has shown competitive performance. However, the accuracy of crowd counting with MCNN requires enhancement, especially when dealing with uneven crowd distributions. This study addresses this challenge by integrating crowd global density features into the MCNN framework. Employing cascaded learning techniques, global density features are extracted and incorporated into the MCNN architecture. To preserve detailed features often lost during downsampling, we propose an enhanced MCNN structure. Max pooling is replaced with max-ave pooling to retain more intricate details, while deconvolutional layers aid in recovering lost information during the downsampling process. Experimental results on datasets such as UCF CC 50 and ShanghaiTech demonstrate that our proposed technique yields improved accuracy and stability, offering promising advancements in crowd counting accuracy under varying crowd distributions.

I INTRODUCTION

To determine how many individuals are present in each frame of a picture or video, crowd counting is utilised. Three types of crowd counting techniques can be distinguished: direct counting based on target detection, indirect counting based on feature regression, and deep learning-based crowd counting. Lin et al. [1] suggested using the Haar wavelet transform to extract the feature area of the head like contour and develop the SVM classifier to classify the feature area in the pertinent investigations based on target detection [1]-[5]. Kowalak et al. [2] suggested using body shape and contour to identify crowds and estimate their densities. All of these techniques

ISSN2454-9940
www.ijasem.org
Vol 18, Issue.2 April 2024

work well in situations with low crowd densities, but in situations with high densities, the detection accuracy will suffer. The regression correlations between picture attributes and the population size are developed for crowd counting in the pertinent feature regression studies [6] through [10]. Low-level characteristics and Bayesian regression were suggested by Chan et al. [7] as a way to increase the resilience and adaptability of the regression model. Idrees et alproposal .'s to combine data from many sources to estimate the population of a single image was made in their paper [8], which also introduced the UCF CC 50 dataset. Deep learning-based solutions for crowd counting are steadily being presented in recent times due to the rapid growth of deep learning and large data. An approach for cross-scene crowd counting was suggested by Zhang et al. Density map and global number were the two learning objectives that were alternatively used to train it. This algorithm's implementation is based on CNN with a single column. But, given the increase in crowd size, it is inappropriate. For crowd counting, Zhang et al. suggested using the MCNN with three branch networks. Each branch network employed a different set of receptive fields, allowing the upgraded MCNN to adjust to changes in the crowd's size. They also unveiled a brand-new dataset for crowd counting called ShanghaiTech. To increase spatial resolution, Boominathan et al. suggested combining the advantages of shallow and deep convolutional neural networks. A multi-task network that incorporated the high-level prior and density estimation was proposed by Sindagi et al. Switch-CNN was suggested for crowd counting by Sam et al. In this network, a classifier was trained, and suitable input patches were chosen for the regressor. In their proposal, Shi et al. suggested condensing multiscale features into a small single vector and using a deep supervised technique to add more supervision signal. Fu et al. suggested using the LSTM structure to retrieve the crowd region's contextual information. In order to adaptively choose the counting mode utilised at various positions on the image, Liu et al. suggested adding an attention module. The MMCNN was suggested by Yang et al. for accurate crowd counting. In order to increase the robustness of the crowd counting method, in this work the location, specific information, and scale variation were taken into account to build density map. In general, these algorithms perform well while counting

crowds, however when the crowd distribution is unequal, their results are ineffective.

A branch of artificial intelligence called computer vision enables computers to recognise and comprehend the visual world, or the world of pictures and movies. Computer vision involves classifying and identifying images captured by cameras, then drawing conclusions from them or using images to analyse events and determine the best course of action. The goal of crowd counting is to count or estimate the population of an image. There are two types of crowds: sparse crowds and dense crowds. Dense crowd counting involves counting the number of individuals in densely populated areas, whereas sparse crowd counting involves counting the number of people in spatially dispersed areas.

II.LITERATURE REVIEW

Yang Zhang et al. [4] The article presents a technique for estimating the crowd size from an image with varied crowd density and perspective. In order to do this, the article created an architecture known as the Multi-column Convolutional Neural Network (MCNN), which converts a picture into a map of crowd density. The paper's evaluation measures are MAE and MSE. In order to appropriately cover all the tough cases, the model was trained and evaluated on 4 different datasets, one of which was introduced by this research.

V. A. Sindagi and others [8] In order to learn crowd count classification and density map estimate simultaneously, the authors attempted to solve this challenge using a new CNN architecture consisting of end-to-end cascaded networks. By categorising the crowd count into different groups, the method suggested incorporates a high-level prior into the density estimate network, which is similar to roughly estimating the total count in the image. This makes it possible for the layers of the network to accumulate globally applicable discriminative characteristics that aid in estimating density maps that are more accurate and have a smaller count error. From beginning to end, the integrated training is completed. Thorough testing on extremely difficult publicly available data sets demonstrates that the suggested method creates denser maps with lower count errors and higher quality than current state-of-the-art methods.

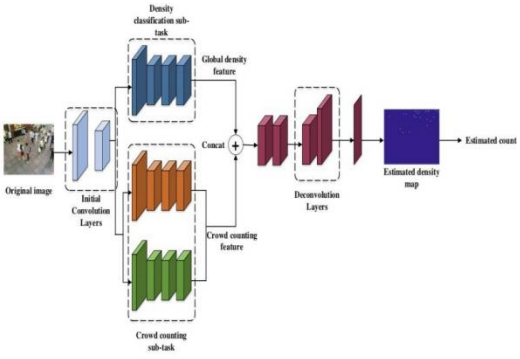
D. B. Sam et al. [7] The authors propose a crowd counting model that transforms a crowd scene from an input image to its density as a solution to the crowd

counting problem. They have introduced a switching convolution neural network-based model to address the common issues and challenges encountered during crowd analysis, such as inter-occlusion between people caused by crowding, indistinguishability between people and the background scene or objects, high variability of camera point-of-views, etc. by utilising the image's wide variation in crowd density to increase the final estimated crowd count's accuracy and localization.

P. Thanasutives et al. This paper's authors take a different tack when it comes to crowd counts. They propose a technique based on dual path multi-scale fusion networks (SFANet and SegNet), two modified neural networks. These two networks or models were given the names M-SFANet and M-SegNet. With the help of atrous pyramid pooling (ASPP), which consists of parallel atrous convolution layers with different sampling rates, the SFANet encoder is able to extract multi-scale features from the target object and integrate those features into a wider context. The context-aware-module (CAN), which is also related to MSFANet, is used by the authors to deal with scale variation in the input image further and to adaptively encode the scales of the contextual information. As a result, the developed model may be used to count people in both scenarios of dense and sparse crowds.

III.ALGORITHM FRAMEWORK

The primary framework of the suggested strategy can be separated into three sections in this paper: First, the sub-task for density classification yields the global density characteristics. These are combined with information gleaned from the crowd counting task. Then, in order to maintain more features, max-ave pooling is introduced. Deconvolutional layers are also utilised to recover lost features during the downsampling process. Lastly, a feature map with a global density feature is used to construct the estimated density map. Estimated counting is obtained using the estimated density map. The algorithm's framework.



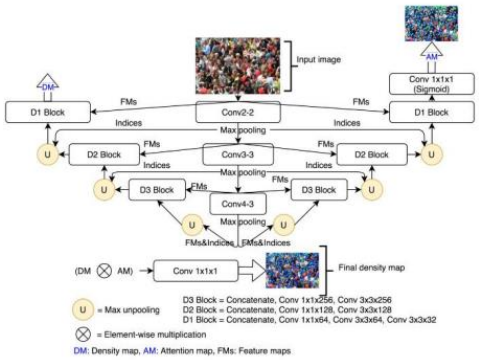
CONVOLUTIONAL NEURAL NETWORK WITH GLOBAL DENSITY FEATURE

ISSN2454-9940
www.ijasem.org
Vol 18, Issue.2 April 2024

The MCNN-based method of crowd counting has thus far shown accurate counting results. The methods for counting crowds that are currently in use, however, do not account for uneven crowd distributions. The global density characteristic is taken into consideration in order to address the issue of unequal crowd dispersion. In this study, the network is built using two methods: extracting feature maps with global density features, and creating a denser map overall.

Working:

Convolution and deconvolution use different computation techniques. You can think of deconvolution as the convolution computation done backwards. In essence, it is an up sampling procedure. The high-dimensional characteristics will be down sampled to the low-dimensional features throughout the calculation procedure in the pooling layers. The feature map's resolution will be decreased. Each pooling region's location of the maximum in the low-dimensional feature matrix will be noted in a position set. The resolution of the feature map will be restored during the up sampling procedure in the deconvolutional layers. The maximum inside each pooling zone will be roughly restored in the high-dimensional feature matrix in accordance with the position selected after the high-dimensional feature matrix has been reconstructed. The remaining values in the high-dimensional feature matrix will be 0.



The process of convolution down-sampling and deconvolution up-sampling is shown in Fig.3. In Fig.3, when the down-sampling operation in the first convolutional layer is finished, the resolution of output feature map in the pool1 becomes half of that in the input image. When the downsampling operation of the second convolutional layer is finished, the resolution of output feature map in the pool2 becomes half of that in the pool1. That means it only keeps 1/4 resolution of the input image. Through adding two deconvolutional layers and four times up-sampling.

ISSN2454-9940
www.ijasem.org
Vol 18, Issue.2 April 2024

Method	Avg. MAE	Avg. MSE
Zhang, Y. et el. [4]	11.6	-
Sam, D. B. et el. [7]	9.4	-
Cong Zhang et el. [6]	10.7	15.0

Table: Comparision of the models reviewed in this paper on UCF_CC_50 dataset

IV.CONCLUSION

According to the assessment we conducted, there has been a notable advancement in the field of crowd counting and crowd density analysis during the last few years. In this study, we explored various approaches to crowd density analysis and crowd counting, primarily based on CNN architecture. Testing the approaches proposed by various authors using datasets from ShanghaiTech, UCF CC 50, UCSD, WorldExpo10, etc. showed that the performance of the methods under evaluation varied depending on the dataset and the context in which they were applied. The manner in which the dataset was preprocessed and the manner in which the model received the processed data both have a significant impact on performance. It is clear from the review that each technique has pros and cons, and depending on the circumstances, a single way may perform better than the other models. These approaches' creators have done extensive research and have been able to improve their model to deliver reliable results.

V.REFERANCES

[1] Kefan, X., Song, Y., Liu, S., & Liu, J. (2018). Analysis of crowd stampede risk mechanism. Kybernetes. doi:10.1108/k-11-2017-0415

[2] Xie, K., Mei, Y., Gui, P., & Liu, Y. (2018). Earlywarning analysis of crowd stampede in metro station commercial area based on internet of things. Multimedia Tools and Applications. doi:10.1007/s11042-018-6982-5

[3] Thanasutives, P., Fukui, K., Numao, M., &Kijsirikul, B. (2021). Encoder-Decoder Based Convolutional Neural Networks with Multi-Scale-Aware Modules for Crowd Counting. 2020 25th International Conference on Pattern Recognition (ICPR). doi:10.1109/icpr48806.2021.9413286

[4] Zhang, Y., Zhou, D., Chen, S., Gao, S., & Ma, Y. (2016). Single-Image Crowd Counting via MultiColumn Convolutional Neural Network. 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi:10.1109/cvpr.2016.70

[5] Ma, Z., Wei, X., Hong, X., & Gong, Y. (2019). Bayesian Loss for Crowd Count Estimation With Point Supervision. 2019 IEEE/CVF

ISSN2454-9940
www.ijasem.org
Vol 18, Issue.2 April 2024

International Conference on Computer Vision (ICCV). doi:10.1109/iccv.2019.00624

[6] Cong Zhang, Hongsheng Li, Wang, X., &Xiaokang Yang. (2015). Cross-scene crowds counting via deep convolutional neural networks. 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi:10.1109/cvpr.2015.7298684

[7] Sam, D. B., Surya, S., &Babu, R. V. (2017). Switching Convolutional Neural Network for Crowd Counting. 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). doi:10.1109/cvpr.2017.429

[8] Sindagi, V. A., & Patel, V. M. (2017). CNN-Based cascaded multi-task learning of high-level prior and density estimation for crowd counting. 2017 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). doi:10.1109/avss.2017.8078491

[9] Battogtokh, B., Mojirsheibani, M., &Malley, J. (2017). The optimal crowd learning machine. BioData Mining, 10(1). doi:10.1186/s13040-017- 0135-7

[10] Shi, X., Li, X., Wu, C., Kong, S., Yang, J., & He, L. (2020). A Real-Time Deep Network for Crowd Counting. ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).