International Journal of Engineering Applied Sciences and Technology, 2021
Vol. 6, Issue 8, ISSN No. 2455-2143, Pages 147-153
Published Online December 2021 in IJEAST (http://www.ijeast.com)

LITERATURE REVIEW ON CROWD COUNTING AND CROWD DENSITY MAPPING METHODOLOGIES

Bhat Sirish Mahadeva
Student, Department of ISE,
Canara Engineering College, Karnataka, India

Chandan C Rao
Student, Department of ISE,
Canara Engineering College, Karnataka, India

Thushar S Kulal
Student, Department of ISE,
Canara Engineering College, Karnataka, India

Udbhav U
Student, Department of ISE,
Canara Engineering College, Karnataka, India

Mr. Ranganatha K
Asst. Prof, Department of ISE,
Canara Engineering College, Karnataka, India

Abstract - **Purpose of Review: Artificial Intelligence and Machine Learning technologies have enabled the analysis of the crowd which helps monitor and manage the crowd effectively. With further advancements in the field of AI and ML, the quality and the accuracy of the analysis have improved considerably. This review attempts to provide an overview of the currently available technologies enabling crowd analysis operations like crowd counting, crowd density mapping and weigh the benefits and drawbacks of the available methods also take into consideration the situation they are suitable for.**
Recent Finding: We have reviewed 10 articles related to crowd counting and crowd density mapping that used Multi-Column CNN, Deep Convolutional Neural Networks, Real-Time Deep network, Switching CNN, andEncoder-Decoder Based CNN with Multi-Scale-Aware Modules to estimate the crowd density and count for the given crowd scene image. In summary, these studies showed a high level of accuracy and provided excellent examples of the potential of AI in crowd counting. The methodologies and algorithms adopted to handle the problem of crowd counting and density mapping varied considerably from paper to paper. Few Built upon the findings of older methodologies by adding enhancement layers and others approached the problem with a new perspective. The methodologies we studied, had their advantages and disadvantages and were suitable for certain situations.
Summary: With advancements in Computer Vision and Artificial neural network techniques like Convolution Neural Networks, crowd counting, and density mapping which traditionally were considered a difficult task and take up time to train the models has become easier to implement and time for training the models are also reduced considerably reduced. In this paper, we have surveyed journals and articles from a variety of sources via google scholar published from 2015 to 2021 related to crowd counting and crowd density mapping. We have reviewed the methodologies introduced in this paper along with their advantages and disadvantages.

Keywords: **Crowd-counting, Density mapping, neural networks, Convolution neural networks.**

I. INTRODUCTION

Recent years witnessed a significant rise in crowd stampede accidents (CSA). CSA is a life-threatening situation caused during a mass gathering, poor crowd management, and rush

International Journal of Engineering Applied Sciences and Technology, 2021
Vol. 6, Issue 8, ISSN No. 2455-2143, Pages 147-153
Published Online December 2021 in IJEAST (http://www.ijeast.com)

for aid or seemingly without any causes. As pedestrian or crowd movement during these situations has no proper control CSA has been the cause of many preventable deaths [1]. Using crowd counting and effective crowd management protocols, situations like CSA can be prevented thus providing a greater sense of public security. The problem of crowd counting is traditionally approached in two ways, crowd-oriented approach and density map oriented approach. Crowd oriented approach involves counting the number of people in the frame by using an object detector that implements the sliding window technique. This approach struggles to give accurate results when the density of the crowd is extreme. To tackle this shortcoming density-map-oriented approach is used. A density map includes the spatial information which can be effectively utilized to indicate the total number of individuals in the image [10]. Researchers have approached this problem as a computer vision problem and have developed various methods to count the crowd accurately [7].
The need for crowd counting has become greater as it has a broad array of applications, in terms of public safety, surveillance monitoring, and crowd traffic control. With crowd counting an organization can mitigate tragic situations like CSA or just take better decisions in crowd management to get an efficient flow of crowd traffic.

II. LITERATURE SURVEY

2.1 Zhang, Y. et al [4]
The paper introduces a method to estimate the number of people in a crowd from an image with varying crowd density and perspective. For this, the paper introduced a Multi-column Convolutional Neural Network (MCNN) architecture which maps the image to its equivalent crowd density map. The paper uses MAE and MSE as their evaluation metrics. The model was trained and tested on 4 different datasets one of which was introduced by this paper to adequately cover all the challenging situations.

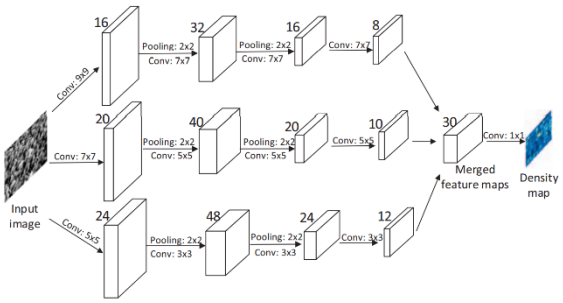

Fig: The above image summarizes in detail the architecture of the MCNN model used in the paper.

The authors of the paper validated the proposed model trained with datasets from ShanghaiTech, UCF_CC_50, UCSD, WorldExpo' 10 and discovered that it performed better than the traditional models.
The paper also goes through the evaluation on transfer learning wherein the model was trained in two settings

1. The model was trained with no samples from the domain of interest.
2. The model was trained on a few samples from the domain of interest.

The model was only able to give the best MAE and MSE only when the last two layers of the MCNN model were fine-tuned. The model also gave a comparable MAE and MSE when the entire model was fine-tuned.

2.2 Sindagi, V. A., et al [8]
The authors tried to solve this problem using a new CNN architecture comprising of end-to-end cascaded network in-order to learn crowd count classification and density map estimation at the same time. The method proposed incorporates a high-level prior into the density estimate network by classifying the crowd count into distinct categories is equivalent to roughly estimating the entire count in the image. This allows the network's layers to acquire globally relevant discriminative features that help estimate density maps that are more refined and have a lower count error. The combined training is done from beginning to conclusion. Extensive tests on very challenging publically accessible data-sets show that the suggested method produces lower count error and higher-quality density maps than recent state-of-the-art techniques.

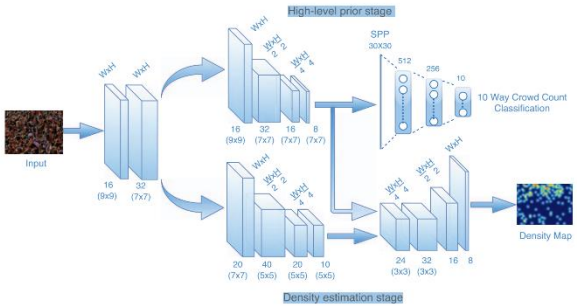

Fig: The above image is an illustration of the cascaded architecture for jointly learning density estimation and high-level prior proposed by the authors of this paper.

The authors have compared the proposed model trained with datasets from ShanghaiTech and UCF_CC_50 with MCNN, Single-stage CNN, and Zhang et al.[6] and discovered that the MAE and the MSE were consistently better than other models.

International Journal of Engineering Applied Sciences and Technology, 2021
Vol. 6, Issue 8, ISSN No. 2455-2143, Pages 147-153
Published Online December 2021 in IJEAST (http://www.ijeast.com)

2.3 Sam, D. B. et al [7]

The authors approach the crowd counting problem by proposing a crowd counting model that converts a crowd scene from an input image to its density. They have addressed the common problems and obstacles faced during crowd analysis like inter-occlusion between people caused by overcrowding, indistinguishability between people and the background scene or objects, high variability of camera point-of-views, etc by introducing a switching convolution neural network-based model that makes use of the image's wide variation in crowd density to increase the final estimated crowd count's accuracy and localization. Based on CNN's crowd count prediction quality acquired during training, patches from a grid within a crowd scene are sent to different CNN regressors. The different receptive fields of the independent CNN regressors are created, and a switch classifier is trained to send the crowd scene patch to the most optimal CNN regressor. The authors run thorough tests on all major crowd counting datasets and find that the proposed method outperforms the current state-of-the-art techniques. Interpretable representations represent the multichotomy of space of crowd scene patches inferred from the switch. An observation is made that the switch relays an image patch to a certain CNN column dependent on crowd density.

The authors evaluated the performance of their model by training it with 4 different data sets, namely ShanghaiTech, UCF_CC_50, UCSD, and The WorldExpo' 10 datasets. During testing, the image patches were sent to a switch classifier that relayed the patch to the most optimal CNN regressor which predicts an optimal crowd density map for the patch relayed to it. The created density maps are then combined into an image to produce the final image, which contains the full scene's density map. The models trained with ShanghaiTech and WorldExpo'10 outperform Zhang et al. [4], MCNN [6] giving lower MAE and MSE, while models trained with UCSD and UCF_CC_50 gave slightly less accurate results than the other model as these datasets had limited training samples compared to the other two.

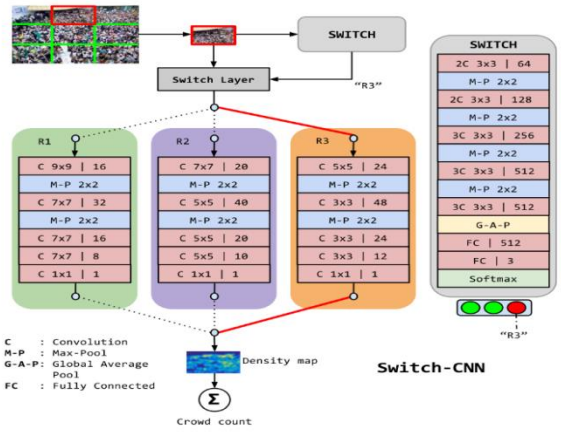

Fig: The above image contains an overview of the proposed Switch-CNN technique.

2.4 Thanasutives, P. et al [3]

The authors of this paper have a different approach to Crowd counting. They suggest a method based on two modified neural networks, SFANet and SegNet, which are dual path multi-scale fusion networks. They named these two networks or models M-SFANet and M-SegNet. The encoder for SFANet is connected with atrous pyramid pooling (ASPP) that contains parallel atrous convolution layers with varying sampling rates as a result of which it can extract multi-scale features of the target object and incorporate that into a larger context. To deal with scale variation in the input image further, to adaptively encode the scales of the contextual information, the authors use the context-aware-module (CAN), which is also coupled to M-SFANet. As a result, the model developed is useful for counting in both dense and sparse crowd scenarios. M-SFANet's decoder, which is based on the SFANet decoder structure has dual paths, one for density path and the other for attention map generation. M-SegNet is the other model, and it is created by substituting the bilinear upsampling in SFANet with the max un-pooling employed in SegNet. This improvement results in a quicker model with comparable counting performance. M-SegNet is designed for high-speed surveillance applications and does not include a multi-scale-aware module to reduce complexity. Both models are encoder-decoder architectures that can be trained from start to finish.

International Journal of Engineering Applied Sciences and Technology, 2021
Vol. 6, Issue 8, ISSN No. 2455-2143, Pages 147-153
Published Online December 2021 in IJEAST (http://www.ijeast.com)

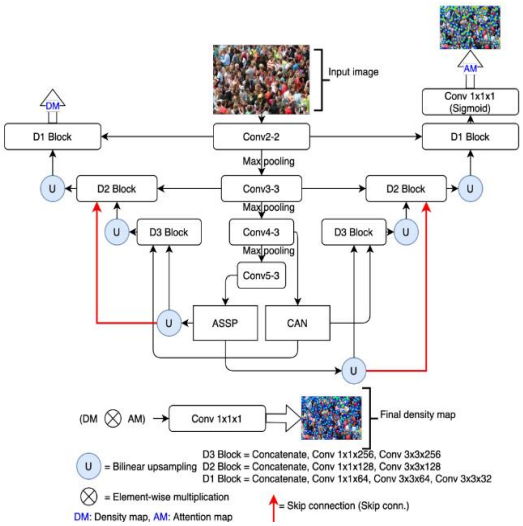

Fig: The above image is an overview of the architecture of M-SFANet proposed by this paper.

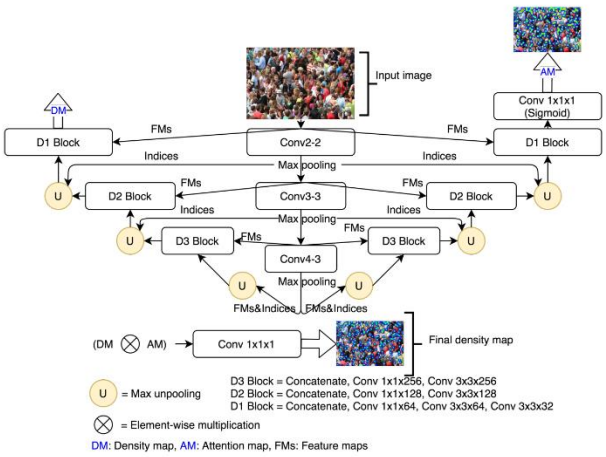

Fig: The above image is an overview of the architecture of M-SegNet proposed by this paper

The authors ran thorough tests on five different datasets related to crowd counting and one dataset related to vehicle counting to show that these changes result in algorithms that can improve current crowd counting approaches. The model reduces the downsides present in the cutting edge strategies and in this way shows prevalent execution on both group and vehicle counting. Furthermore, the M-SFANet decoder is trained to have more remaining associations overall, ensuring that the learnt multi-scale highlights of important level semantic data will influence how the model relapses for the last thickness map. Notwithstanding, the examining paces of the scale aware modules are not learnable and the quantity of these rates is fixed before preparing. This could prompt restricted execution in specific inconspicuous scenes. Consequently, a versatile execution of the modules wherein the testing rates or expanded rates are movable is considered as conceivable future work. For MSegNet, the authors changed the up-testing calculation from bilinear to max unpooling utilizing the retained records utilized in SegNet. This yields a less expensive calculation model while giving cutthroat counting execution material to real-world applications.

2.5 Cong Zhang. et al [6]

The authors of this paper, address one of the most prominent drawbacks of the other models or methods i.e. crowd counting on a different scene or datasets that vary considerably from the dataset the models were trained upon as the performance of already existing crowd counting techniques drops significantly. The method proposed, Cross-scene crowd counting is one of the most challenging tasks as it does not require arduous data annotation for counting the number of individuals in new unseen target surveillance crowd scenes. For crowd counting, the authors employ Deep-CNN and train it with two related learning objectives: crowd density and crowd count. With this switchable learning approach, the authors were able to obtain a better local optimum for objectives. The authors provided a data-driven strategy for fine-tuning the trained CNN model for the target scene to handle unobserved target crowd scenarios. This paper also introduces a new dataset that includes 108 crowd scenes with nearly 200,000 head annotations. This dataset helped in better evaluation of cross-scene crowd counting accuracy.

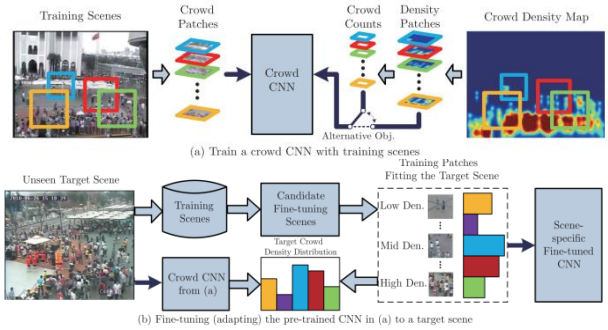

Fig: The above image illustrates the cross-scene crowd counting model.

The authors evaluated the model trained with datasets from WorldExpo10', UCD_CC_50, UCSD and discovered that the model consistently performed better than the existing models providing less MAE and MSE.

2.6 Shi, X, et al. [10]

The authors of this paper have tried to address the real-world problem that is faced by the existing model, i.e. Real-time performance of the crowd counting operation. Computer vision researchers have often given their attention

International Journal of Engineering Applied Sciences and Technology, 2021
Vol. 6, Issue 8, ISSN No. 2455-2143, Pages 147-153
Published Online December 2021 in IJEAST (http://www.ijeast.com)

to solving problems like automatic analysis of densely crowded scenes and as a result of which significant improvements have been witnessed in that area. The authors introduce a method that makes use of compact convolution neural networks for crowd counting which can learn efficiently with a limited number of parameters. The proposed method would achieve near-real-time speed and conserve computer resources by using three parallel filters to perform the convolutional method on the input picture at the front of the network. The experiments conducted by the authors on the two benchmarks show that the proposed model cannot only balance between efficiency and performance, but is also perform suitably for actual real-world scenes, and it is superior in terms of speed to the existing lightweight models.

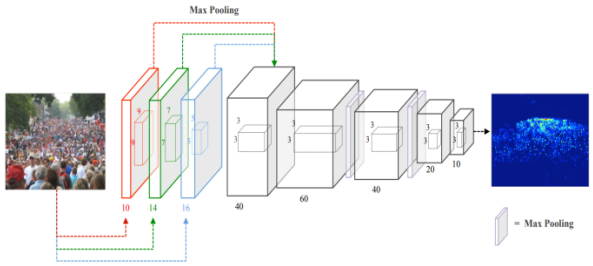

Fig: The above image is an illustration of Compact-CNN architecture.

The authors have conducted an in-depth study of the model trained with the datasets from ShanghaiTech and WorldExpo10'. According to the study with the dataset from ShanghaiTech, the model which had the smallest parameter size can give low MAE and MSE, although deeper models with greater parameter size can achieve better performance on the same. The study using WorldExpo10' revealed that the proposed model gave better accuracy in general compared with the existing models considering the small parameter size. The outcome of this study indicates that the Compact-CNN model is light enough without sacrificing accuracy.

III. COMPARISION OF THE METHODS

3.1 Comparison with ShanghaiTech Dataset

This dataset is a large-scale crowd counting dataset introduced in the paper Zhang et el. [4]. It is a dataset containing 1198 annotated images containing around 330,165 heads of the people annotated. This dataset is divided into two sub datasets, they are Part_A with 482 images collected from the internet randomly, and Part_B containing 716 images collected from the metropolitan streets of Shanghai.

Method	Part_A		Part_B	
	MAE	**MSE**	**MAE**	**MSE**
Zhang, Y. et el. [4]	181.8	277.7	32.0	49.8
Sindagiet el. [8]	101.3	152.4	20.0	31.1
Sam, D. B. et el. [7]	90.4	135.0	21.6	33.4
M-SFANet + M-SegNet [3]	57.55	94.48	6.32	10.6
Shi, X, et el. [10]	88.1	141.7	14.9	22.1

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

3.2 Comparision with UCF_CC_50 dataset

UCF_CC_50 is a dataset consisting of an annotated collection of 50 crowd scene images. The dataset contains scenes with large variance in crowd scenes varying between 94 and 4543

Method	**MAE**	**MSE**
Zhang, Y. et el. [4]	377.6	509.1
Sindagiet el. [8]	322.8	397.9
Sam, D. B. et el. [7]	318.1	439.2
M-SFANet + M-SegNet [3]	167.51	256.26
Cong Zhang et el. [6]	467.0	498.5

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

International Journal of Engineering Applied Sciences and Technology, 2021
Vol. 6, Issue 8, ISSN No. 2455-2143, Pages 147-153
Published Online December 2021 in IJEAST (http://www.ijeast.com)

3.2 Comparision with UCSD dataset
UCSD dataset contains around 2000 frames collected from a single surveillance camera located on the UCSD campus. There are on average 25 people in each frame. The frame size is 158x238 and is recorded at 10 frames per second.

Method	**MAE**	**MSE**
Zhang, Y. et el. [4]	1.07	1.35
Sam, D. B. et el. [7]	1.62	2.10
Cong Zhang et el. [6]	1.60	3.31

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

3.4 Comparision with WorldExpo'10 dataset
WorldExpo10' is a large-scale cross-scene crowd counting dataset introduced in a paper by C, Zhang et al. [6]. It is a large dataset containing 1132 annotated video sequences captured by around 108 surveillance cameras during Shanghai 2010 WorldExpo.

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

From the review, we conducted it is evident that in the past few years remarkable progress can be seen in the area of crowd counting and crowd density analysis. In this paper, we reviewed different methodologies used for crowd counting and crowd density analysis mainly based on CNN architecture. The methods introduced by different authors were subjected to testing on datasets from ShanghaiTech, UCF_CC_50, UCSD, WorldExpo10'etc, and revealed that the methods under review performed differently based on the dataset and the situation the method was used in. The performance is mainly dependent on how the preprocessing of the dataset was conducted and how the processed data was given to the model. From the review, it is evident that each method has its advantages and disadvantages, and based on the situation a particular method can outperform the other models. The authors of these methods have conducted thorough research and have been able to fine-tune their model to give consistent performance efficiently.

V. REFERENCES

[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). Early-warning 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 Multi-Column 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 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).