International Journal of Computer Applications (0975 - 8887)
Volume 185 - No.37, October 2023

Machine Learning Techniques for Crowd Counting: A Survey

Dakshi Chavan
Department of Computer Engineering
Shri G. S. Institute of Technology and Science
Indore, M.P., India

Anuradha Purohit
Department of Computer Engineering
Shri G. S. Institute of Technology and Science
Indore, M.P., India

ABSTRACT

Crowd Counting process estimates the number of people in an image or video. It is a significant area of computer vision research that has numerous applications in crowd management, class student attendance management, temple crowd management, event planning, urban development, security and retail analytic and many more. Due to the increasing interest to provide efficient crowd management and public safety, several researchers have proposed methods based on detection, regression and density. This has made it feasible for both machine learning (ML) and deep learning (DL) approaches to deal with challenges to get accurate crowd counts. Machine learning and deep learning based models identify complex patterns, obtained increased accuracy and adjust to changing environmental conditions efficiently. In this paper, a survey on work done in crowd counting using machine learning techniques has been presented. The advantages and disadvantages of each approach has been discussed in detail.

Keywords

Crowd counting, Crowd management, Machine learning (ML), Deep learning (DL), Detection

1. INTRODUCTION

Crowd counting is a method of counting each individual person in an image or area. Turnstiles can be used to manually estimate crowd by counting them as they pass through. However, with the development of technology, computer vision techniques can also be used to count crowds. One such method is to estimate or count the number of persons in an image using deep learning. The purpose is to implement algorithms and models to automatically estimate the number of people or objects in a scene. It is an important field of study with real world applications in crowd management, event planning public safety, and urban monitoring. These techniques continue to get increase in accuracy and reliability due to developments in computer vision, particularly deep learning techniques, making them vital tools in various industries for dealing with crowd-related difficulties effectively.
Crowd Counting is a challenging problem due to the complexity of crowded scenes, which often contain people of different sizes, shapes, and densities. With the use of computer vision and machine learning, researchers have considerably enhanced crowd counting methods over time. Approaches for early crowd estimating and counting depend mostly on the detection of pedestrians [29, 12, 22]. Crowd counting has numerous uses and has a big impact on a variety of industries. Many different areas found use of crowd counting, in crowd management, safety, urban planning, and more. Accurate crowd counting is essential for safely and effectively managing massive crowds at public events, concerts, and festivals in order to avoid congestion and ensure safety[10]. Crowd counting helps in analysing foot traffic in metropolitan areas, transportation hubs and malls for urban planning and management of infrastructure, resulting in better infrastructure design and allocation of resources[42]. Crowd counting helps in estimating attendance, managing facilities, and guaranteeing an easy event experience for event planners and organisational logistics[24]. Crowd counting is another tool utilised for traffic monitoring and analysis that helps determine the number of vehicles on the road and examine traffic patterns in order to enhance traffic planning and management[15]. Additionally, crowd monitoring and counting helps to ensure public safety by identifying significant dangers to security and unusual crowd behaviour[27]. The importance of crowd counting also extends to retail and business intelligence, where it is utilised to analyse customer preferences, prime shopping hours, and product popularity for marketing and business planning. Traditional computer vision techniques and intuitive methods become a base of crowd counting techniques. The haar cascades can be used to count people in a crowd by identifying and counting them in the environment[33]. In dynamic situations, crowd counting can be done using background subtraction techniques to find moving objects in video streams[28]. While these techniques have several drawbacks in comparison to machine learning-based approaches, they are still helpful in some limited resources scenarios or when real-time performance is not the major issue.
Several approaches, including detection-based counting, have been tried in the recent years to handle the problem of crowd estimation and density counting.
In this paper, a survey of work done in the use of deep learning and machine learning techniques to crowd counting in images and videos has been presented. Increased accuracy, scalability, and adaptability are provided by these techniques, which have effects on event planning, security management, and more. The potential to transform the understanding of crowd dynamics and introduce

novel techniques for crowd management and counting grows more and more real.
The remaining sections of the paper have been organised as follows. Background of crowd counting is presented in section-II and review of crowd counting work using Machine Learning is presented in Section III. A discussion on previous work in crowd counting using Deep learning is presented in Section IV with conclusion in Section V.

2. BACKGROUND

Using methodologies like clustering based counting, detection based counting and regression based counting, researchers have been studying the problem of crowd counting and estimation of density for the past few years[18].
The issue of crowd counting in images has been handled by number of approaches presented by researchers[42] [13, 5, 17, 41].

(1) Detection based approaches
(2) Regression based approaches
(3) Density estimation based approaches

2.1 Detection based approaches

In detection-based techniques, Convolutional Neural Networks (CNN) are used to train a classifier that can identify each individual that is present in the images. The detection type approach, which uses a sliding window detector to count the number of humans, was the focus of the majority of early study[19].
Monolithic and parts-based detection are the two commonly used types of detection. Traditional "monolithic detection approaches" [6, 16, 31, 7] for recognising pedestrians train a classifier using body-wide characteristics such as Haar wavelets[34], histogram-oriented gradients[6], edgelets[36], and shapelets[23]. Several techniques have been applied to different levels of effectiveness, including Support Vector Machines, boosting[32], and random forests [9].
Part-based detection techniques have been used by researchers to try to solve the problem[8, 26, 37], where one generates boosted classifiers for certain body parts, such as the head and shoulder, to estimate individuals in a specific area[19].

2.2 Regression based approaches

Regression methods try to precisely predict the count as a continuous variable using the image's extracted attributes. For this, machine learning techniques like neural networks, support vector regression, and linear regression can be applied. Since the mentioned problems are not relevant in the extremely crowded setting, detection-based approaches are not appropriate. Because of this, regression-based crowd counting was developed to address these problems[2][13]. This method consists of two phases, low-level feature extraction and regression modelling. Regression-based techniques are used to calculate the crowd count using image features and crowd size[42], [13][20, 35, 38, 39]. The following attributes can be extracted from the image patches: texture, edge, gradient, and foreground attributes. These attributes are used to produce low-level data.

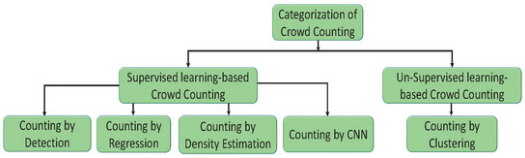

Fig. 1. A taxonomy of crowd counting techniques.

2.3 Density estimation based approaches

These methods involve creating a density map that assigns a density value to each pixel in the image. While earlier approaches were successful in handling obstruction and clutter problems, the majority of them neglected important spatial information in favour of a global count. The density map is generated by placing Gaussian kernels centered around the positions of individuals or objects. Although the occlusion and congestion problems were successfully addressed by the earlier techniques, most of them ignored important spatial information because they were regressing on the overall count [17], however proposed learning a linear mapping between specific attributes and related item density maps to incorporate spatial information into the process of learning. The Multi-Column Neural Network (MCNN) suggests extracting features from multiple columns with different kernel sizes to adjust for scale fluctuation [42]. In addition to the count, the density estimate gives an idea of the location of persons. The method works better than the linear method and uses little memory to store the trees. This method has the drawback of extracting low-level information using conventional features, which makes it difficult to capture low-level data correctly with a high-quality density map.

3. REVIEW OF WORK DONE IN CROWD COUNTING USING MACHINE LEARNING

The use of complicated crowd scenarios with variation densities and occlusions can be challenging using traditional methods without any learning-based approaches. However, several studies have investigated crowd counting methods that do not depend on machine learning or deep learning and instead use patterns or rules.

Lempitsky et al. [17] proposed a method for counting people in a crowd without using deep learning or neural networks. For counting person in overcrowded locations they utilised a traditional machine learning strategy with customised features. The method has the advantage of executing tried-and-true machine learning techniques, which makes it computationally faster and simpler to use than deep learning techniques. Additionally, the approach enables the inclusion of features and domain-specific knowledge that are specific to the crowd counting problem. This method mainly relies on customised features, which might not fully represent the hidden trends and complexities inherent in crowded situations. This could limit the method's capacity to accurately manage complex and varied crowd situations.

Chan et al. [3] presented an idea on behind crowd counting to estimate how many individuals are travelling in various directions while protecting their privacy. Based on the crowd's movement patterns, the system uses a segmentation approach to separate it into two sub-components. It is demonstrated that crowd size estimation is possible utilising low-level data retrieved from each crowd segment, with basic features like the segmentation area displaying an approximate linear correlation with the crowd size. The system

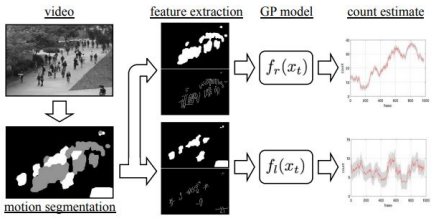

Fig. 2. Architecture of "Privacy Preserving Crowd Monitoring"

for counting people in a crowd includes multiple steps. The video footage is first divided into areas demonstrating crowds moving in various directions. For this, a variety of dynamic textures are combined. Then, different features are taken from every segment of the crowd. Using a perspective map, each image location is weighted according to its approximate size in the real scene to correct for perspective distortion. The obtained characteristics and weighted image data are then used in Gaussian process regression to get the estimated population per segment. By using this method, a system can determine crowd sizes without requiring for individual tracking or particular people models, preserving privacy in crowded scenes. For a variety of applications, accurate and privacy-aware crowd counting becomes possible by combining the techniques of segmentation, feature extraction, perspective correction, and Gaussian process regression. An outline of the crowd counting system presented in Fig. 2.

—Advantage: The main advantage of this approach is its significant focus on privacy protection. The approach makes sure that crowd members' identities are kept anonymous by keeping far from making use of people models or individual tracking. It allows the moral deployment of crowd surveillance devices in public areas. Because the system respects people's rights and privacy by not using individual identification or tracking, it reduces the possibility that the public will be opposed to or express concerns about invasive surveillance techniques.

—Disadvantage: There may be accuracy restrictions in crowd counting, depending on the specific methodologies used in the approach. It may be difficult for privacy-preserving techniques to manage complex situations including overlapping people or high crowd densities, which could result in under counting or over counting errors. The method could be criticised for sacrificing crowd analysis is granularity. The method might not be able to offer in-depth understandings into individual behaviour due to the lack of individual-specific information, which includes crowd dynamics, interactions, or crowd movement patterns.

Idrees et al. [13] proposed approach uses information from several sources and crowd analysis at various scales to precisely count people in densely populated situations. The suggested approach overcomes the issues of occlusions and restrictions related to single-camera viewpoints in scenarios involving extremely dense crowds. In order to estimate the number of people in crowded situations accurately, an approach combines data fusion, multi-scale analysis, feature extraction, perspective correction, density estimation, and density-based regression. The method overcomes the difficulties of crowded situations and offers a reliable solution for crowd counting in densely populated areas by using information obtained from several sources and analysing the number of people at various scales.

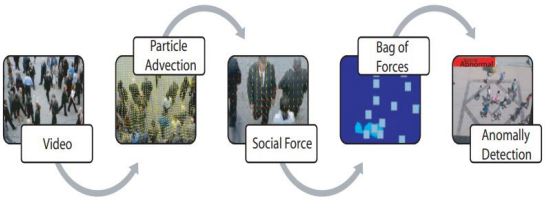

Fig. 3. Architecture of "Abnormal Crowd Behavior Detection using Social Force Model"

—Advantage: The multi-source multi-scale technique takes advantage of information gathered from various camera views or image frames to produce a variety of viewpoints of the crowd scenario. As a result, the crowd appears in a more accurate manner, which leads to increased crowd count estimation accuracy, particularly in difficult and congested conditions. The method has less impact by occlusions that could occur in crowded scenes by combining data from various sources. Accurate counting might be challenging due to occlusions that obstruct the vision of individuals. Utilising multiple sources increases the effectiveness of occlusion avoidance and results in a more accurate crowd count.

—Disadvantage: Keeping the data properly synchronised while using data from many sources might be difficult. The accuracy of crowd count estimation may be affected by misalignment or timing differences between the data sources, requiring further data synchronisation efforts.

Mehran et al. [21] provide a novel approach for the purpose of identifying unexpected behaviours in crowd pictures, the Social Force concept is used. In this method, optical flow is averaged over both time and space to produce a representation after the image is divided into a grid of parts. The social force model, which considers each moving parts as an individual, is used to measure the interactions between them. Then, to compute Force Flow for each pixel in each frame, this interaction force is projected onto the image plane. Force Flow's spatio-temporal volumes are selected at random to imitate typical crowd behaviour. Using the bag of words method, frames are classified as normal or abnormal. Utilising interaction forces, anomalies in the anomalous frames are detected. The results of the studies demonstrate that the suggested approach successfully captures the dynamics of crowd behaviour. Fig. 3. summarizes the main steps of the abnormal crowd behavior detection using social force model.

—Advantage: As the method takes into consideration interactions between individuals and physical forces among people, the Social Force Model can be used to create realistic crowd simulations. Because of this, the method is better able to capture the dynamics and behaviours of a crowd, which makes it useful for identifying abnormal crowd behaviour.

—Disadvantage: The Social Force Model can be computationally expensive to implement and simulate, particularly in large-scale crowd scenarios. A lot of processing effort and resources may be needed to simulate crowd behaviour using physics-based models, which could have an impact on applications that run in real-time or very near it.

Chan et al. [4] approach suggested a Poisson regression model created using the observed properties (such as density, texture, etc.) as predictor variables and the number of persons in a

certain area or image patch as the response variable. The Poisson distribution's parameters and regression coefficients are estimated using Bayesian inference, which enables the model to adjust to changing scene complexity and crowd densities. The ability of Bayesian Poisson regression to handle over dispersion a problem that frequently arises in crowd counting scenarios is one of its key features.

—Advantage: The Poisson distribution is a good option for estimating counts, which is frequently used in crowd counting activities. The method can successfully represent the variation in crowd size by assuming that the crowd counts follow a Poisson distribution.

—Disadvantage: The Poisson distribution may not be able to adequately predict over dispersion on its own when the variance of the count data exceeds the mean. Over dispersion happens when this occurs.

4. REVIEW OF WORK DONE IN CROWD COUNTING USING DEEP LEARNING

The goal of crowd recognition in computer vision is to identify individuals in crowded spaces or in images. Over the years, work has been carried out on a variety of techniques and algorithms to handle this challenging issue.

Zhang et al. [40] approach solves the problem of adapting crowd counting algorithms to novel and unknown domains without any annotated data from the target region. To align the feature distributions of the source and target domains, the technique provides use of mechanisms for cross-domain attention. Even in the absence of labelled data, the model could generalise to the target domain successfully by learning domain-invariant representations. In addition, the cross-domain attention network allows the model to efficiently transfer knowledge from the source domain, resulting in more accurate crowd counting in real-world scenarios with domain shifts, such as changing camera angles, lighting conditions, and crowd densities. An architecture of the crowd counting system of "Cross-Domain Attention Network for Unsupervised" is shown in Fig. 4.

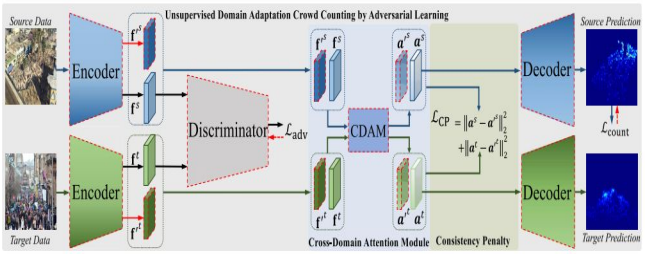

Fig. 4. Architecture of "Cross-Domain Attention Network for Unsupervised"

—Advantage: The primary advantage of the "Cross-Domain Attention Network for Unsupervised Domain Adaptation Crowd Counting" approach is its ability to adapt crowd counting models from a labeled source domain to an unlabeled target domain without requiring labeled target domain data. This reduces the need for extensive manual annotation efforts in the target domain. The use of domain adaptation modules and attention mechanisms allows the model to effectively adapt to domain shifts. The architecture leverages differences between domains to better align features and improve the model's adaptation capabilities.

—Disadvantage: While the architecture is designed to adapt to domain shifts, there could be scenarios where the differences between the source and target domains are too significant for effective adaptation. In cases of extreme domain dissimilarity, the approach may not yield satisfactory results.

Hossain et al. [11] have proposed a novel method to overcome the difficulties of crowd counting in diverse and complicated conditions. The technique makes use of attention networks that are scale-aware in order to adaptive capture elements at different scales within the crowd images. Scale-aware attention techniques allow the model to effectively focus on various areas of the image and collect both local and global data. Even in circumstances with fluctuating crowd densities and scale changes, this provides more precise and reliable crowd counting results. The benefit of this method is that it is capable of handling with the particular variety of crowd scenes, including occlusions, point of view modifications, and density variations. An architecture of the crowd counting system of Scale-Aware Attention Networks is shown in Fig. 5.

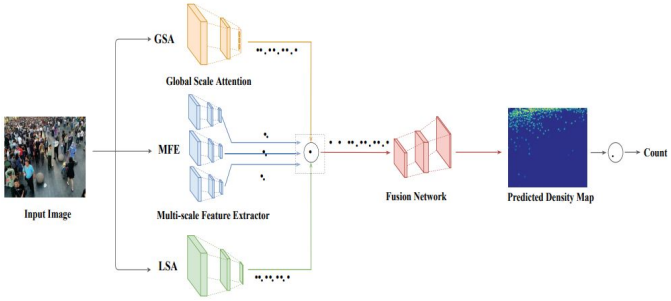

Fig. 5. Architecture of "Scale-Aware Attention Networks"

—Advantage: The "Crowd Counting Using Scale-Aware Attention Networks" method uses scale-aware attention mechanisms to solve the problem of accurate crowd counting. As a result, the algorithm can concentrate on various scales within the crowd scene and estimate crowd sizes more accurately. The model can manage crowd occlusions, where people may be partially or entirely obscured from view. This is made possible by the attention processes. The model can still estimate precise crowd counts even if some of the crowd is blocked off by obstructions since it pays attention to many scales.

—Disadvantage: The model architecture may become more complex as a result of the implementation and training of scale-aware attention networks. Additional computational resources and skill in model building and optimisation may be needed to include attention processes and scale-aware components.

Jiang et al. [14] proposed technique provides a novel approach that makes use of multi-task learning to enhance crowd counting accuracy. Crowd counting and density map estimate are two tasks that the model simultaneously learns. The model improves its performance in counting by better understanding crowd densities

and distributions as a result of jointly optimising both of the tasks. This method has a benefit in that it can capture fine-grained crowd data as well as density variations, which are important in densely populated environments with a variety of product sizes and occlusions. The model may exchange features and information between tasks because of its multi-task learning framework, which increases its effectiveness and efficiency when handling a variety of crowd scenarios. Overall, the Density-Aware Multi-Task Learning technique offers a feasible choice for precise and accurate crowd counting in challenging situations in the real world. An architecture of the Density-Aware Multi-Task Learning is shown in Fig. 6.

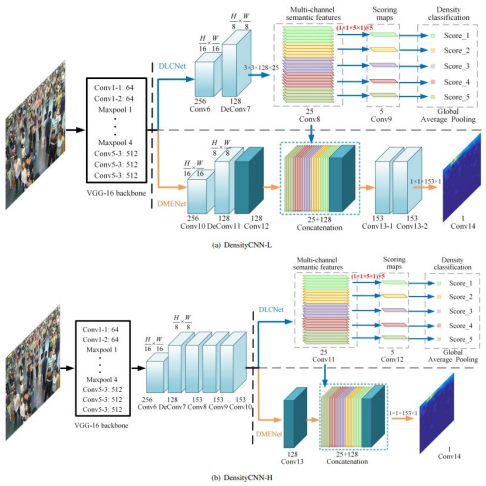

Fig. 6. Architecture of "Density-Aware Multi-Task Learning"

—Advantage: With this method, the learning process explicitly takes density awareness into consideration. When crowd density is taken into consideration as a task, the model is better able to handle various kinds of crowd densities and can adjust its predictions to different densities within an image.

—Disadvantage: Multi-task learning typically requires a more complex training process, involving the optimization of multiple objectives simultaneously. This could increase training time and make the training process more computationally demanding. The use of supervised crowd counting techniques in novel scenarios is strongly constrained by this. It is possible to fail to recognise the domain features that distinguish real-world surveillance applications.

Shami et al. [25] provide a novel approach for counting persons in extremely dense crowd photos. The aspect of persons in a crowded environment is their head, which is the core of this approach. In view of this, a head detector can be utilised to determine the spatially adjustable head size, which is the main component of this head counting method. For the purpose of sparse head detection in dense crowds, the most advanced convolutional neural network has been used. The image is divided into rectangular patches, and after each patch is labelled as either a crowd or not, all not-crowd patches are eliminated using a SURF feature-based SVM binary classifier. Then, regression is used to each patch of the crowd to determine the average head size. By dividing the patch area by the estimated head size, the estimated number of people of each patch is determined. When no heads are found in a crowd patch, the counts are calculated using a distance-based weighted average of the counts from nearby patches. Finally, the total count is calculated by adding the individual patch counts.The architecture of this method is given in fig.7.

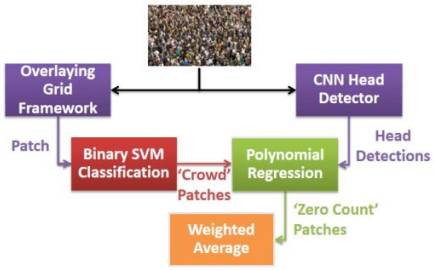

Fig. 7. Architecture of "People Counting in Dense Crowd Images using sparse"

—Advantage: The method decreases the complexity of computers while ensuring counting accuracy by focusing on head detections. The benefit of this method is that it can achieve high accuracy in difficult crowd situations when conventional full-body detection techniques may struggle. The model's enhanced counting performance is a result of the sparse head detections, which allow it to recognise particular situations in dense crowds.

—Disadvantage: Crowd counting and related tasks, such as density estimation, must be correlated for multi-task learning to be effective. The method might not offer significant advantages if the tasks are not properly related.

Tian et al. [30] methodology conveys an in-depth crowd counting technique that solves the issues of fluctuating crowd densities in various parts of an image. By using a pan-density map to estimate crowd counts and densities at various scales, the PaDNet model is able to precisely count both sparse and dense areas inside the crowd scene. The benefit of this method is that it can manage situations when crowd densities are not unified, where other methods can find it difficult to estimate counts accurately. PaDNet uses the pan-density map to represent the crowd distribution in a more accurate and informative way, improving counting precision in challenging and congested locations. An architecture of the "PaDNet: Pan Density Crowd Counting" is shown in Fig. 8.

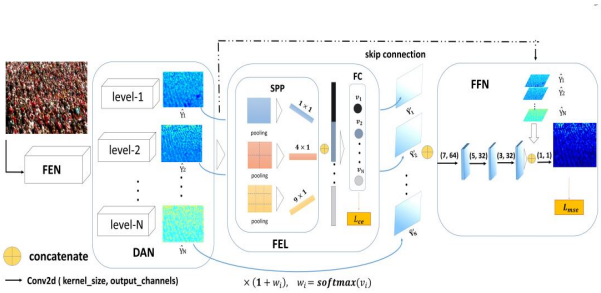

Fig. 8. Architecture of "PaDNet: Pan Density Crowd Counting"

Table 1. Results of Deep Learning Techniques for Crowd Counting

References	Method	Dataset	MAE	MSE
Zhang et al. [40]	Cross Domain Attention Network(CDANet)	Shanghai Tech Part B	14.7	20.9
		UCF-QNRF	218.9	381.3
		WorldExpo	13.2	23.2
Hossain et al. [11]	Scale-Aware Attention Networks	ShanghaiTech Part B	16.86	28.4
		UCF CC 50	271.60	391.00
		Mall	1.28	1.68
Jiang et al. [14]	DensityCNN	ShanghaiTech Part B	11.09	20.81
		UCF CC 50	265.06	384.60
		UCF-QNRF	116.03	203.60
		WorldExpo'10	14.3	8.2
Shami et al. [25]	CNN+SVM Binary Classifier	UCF CC 50 (With Weighted Average)	1085.6	1427.8
		ShanghaiTech Part A (With Weighted Average)	241.6	369.6
		ShanghaiTech Part B (With Weighted Average)	63.4	61.4
		AHU-Crowd (With Weighted Average)	198.2	0.580
Tian et al. [30]	PaDNet	ShanghaiTech Part A	59.2	98.1
		ShanghaiTech Part B	8.1	12.2
		UCF CC 50	185.8	278.3
		UCSD	0.85	1.06
		UCF-QNRF	96.5	170.2
Sam et al. [1]	LSC-CNN	ShanghaiTech Part A	66.4	117.0
		ShanghaiTech Part B	8.1	12.7
		UCF CC 50	185.8	278.3
		UCSD	0.85	1.06
		UCF-QNRF	120.5	218.2

—Advantage: The "PaDNet: Pan-Density Crowd Counting" method aims to count crowds with great accuracy. The method increases the accuracy of crowd count estimation by combining pan-density modelling, which captures extremely fine crowd density changes. The method is scale-invariant due to the pan-density modelling, which allows it to estimate crowd counts with accuracy regardless of the size or scale of persons in the crowd.

—Disadvantage: In comparison to simpler crowd counting techniques, the Pan-Density Crowd Counting approach might require a more complicated model design. Increased processing demands and potential difficulties with model implementation could result from this complexity. Correct parameter adjustment to maximise PaDNet's performance may be necessary for its success. Experimentation and model design knowledge may be needed to find the optimal hyper parameter settings.

Sam et al. [1] provides an innovative technique for precisely recognising, measuring, and counting individuals in dense crowd photos. To locate possible people in the crowd, the design uses a strong object detection backbone, such as Faster R-CNN or YOLO. Candidate object regions are provided by the initial detection stage. Advanced localization techniques are used to fine-tune the boundaries of detected objects. By accurately locating each person, this stage tries to increase accuracy while lowering false positives. The method calculates the bounding box dimensions for each detected individual or makes use of size-related characteristics to estimate each individual's size. A more accurate estimate of the crowd's size results in better individual representation. The architecture counts the identified persons in the scene using accurate localization and size estimations. Generalized architecture of the method is shown in Fig. 9. The model achieves great counting accuracy especially with dense crowds, when traditional counting techniques may struggle due to crowd density changes and occlusions.

—Advantage: The method has been designed to handle situations with a lot of people, where typical counting techniques could have trouble because of occlusions and overlaps. Its robustness in such difficult situations comes from the combination of its detection, localisation, and counting algorithms. It is adaptable in determining the most effective strategy based on the particular task and environment due to the architecture's adaptability to various object detection backbones and methodologies.

—Disadvantage: The method includes a number of processes, such as object detection, localization improvement, size estimation, and counting. This complexity may result in higher computational demands, potentially requiring additional memory and processor resources.

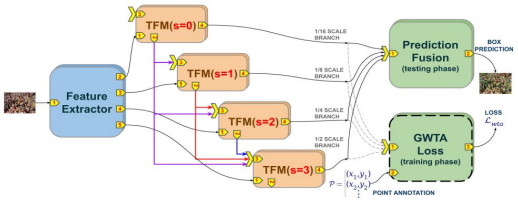

Fig. 9. Architecture of "Locate, Size, and Count Accurately Resolving"

Crowd counting has experienced a fundamental change primarily due to the results of the use of machine learning (ML) and deep learning (DL), which offer several benefits and solve a number of previous issues with traditional approaches. Comparing machine learning and deep learning methods to manual counting or rule-based approaches in crowd counting, it has been observed that the machine learning approaches performed well. Particularly in crowded environments, deep learning (DL) models are able to scoop up on variations and detailed patterns, allowing for accurate

crowd density prediction and reducing counting errors. Without lengthy coding or human modifications, machine learning and deep learning models can adjust to shifting population dynamics and situations.

Real-time video stream processing features of models using machine learning and deep learning enable them to estimate crowd densities instantly. In situations like crowd control, security, and emergency response, this ability is essential for rapid decision-making. By automating crowd counting using ML and DL, the labour and time-intensive manual counting process is eliminated. As a result, operational expenses and human resource costs are reduced.

Results obtained in terms of Mean Square Error (MSE) and Mean Absolute Error (MAE) for various datasets after applying deep learning techniques for crowd counting is presented in Table 1.

5. CONCLUSION

In this paper, a survey on work done by researchers in crowd counting using machine learning techniques along with the advantages and disadvantages of each technique has been presented. Crowd counting has experienced a revolution due to machine learning and deep learning which has enabled previously unobtainable levels of accuracy, scalability, and adaptability. Their fundamental impact on urban design, safety and population control demonstrates how crowded areas will be managed in the future. These methods have the potential to revolutionise our understanding of crowd dynamics and bring in more effective methods for managing and counting crowds as they develop. The development of deep learning (DL) approaches particularly Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs) has led to the true success. These deep learning models have proven their capacity to recognise the complex spatial and temporal patterns present in crowd scenes, resulting in estimation of crowd density with unequalled accuracy. The accuracy and durability of crowd counting systems have reached new levels due to the automatic feature extraction and hierarchical learning capabilities of deep learning architectures.

6. REFERENCES

[1] Deepak Babu Sam, Skand Vishwanath Peri, Mukuntha Narayanan Sundararaman, Amogh Kamath, and Venkatesh Babu Radhakrishnan. Locate, Size and Count: Accurately Resolving People in Dense Crowds via Detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pages 1–1, 2020.

[2] Saleh Basalamah, Sultan Daud Khan, and Habib Ullah. Scale Driven Convolutional Neural Network Model for People Counting and Localization in Crowd Scenes. *IEEE Access*, 7:71576–71584, 2019.

[3] Antoni B. Chan, Zhang-Sheng John Liang, and Nuno Vasconcelos. Privacy preserving crowd monitoring: Counting people without people models or tracking. In *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–7, June 2008. ISSN: 1063-6919.

[4] Antoni B. Chan and Nuno Vasconcelos. Bayesian Poisson regression for crowd counting. In *2009 IEEE 12th International Conference on Computer Vision*, pages 545–551, September 2009. ISSN: 2380-7504.

[5] Ke Chen, Shaogang Gong, Tao Xiang, and Chen Change Loy. Cumulative Attribute Space for Age and Crowd Density Estimation. In *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2467–2474, June 2013. ISSN: 1063-6919.

[6] N. Dalal and B. Triggs. Histograms of Oriented Gradients for Human Detection. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1, pages 886–893, San Diego, CA, USA, 2005. IEEE.

[7] M. Enzweiler and D.M. Gavrila. Monocular Pedestrian Detection: Survey and Experiments. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(12):2179–2195, December 2009.

[8] P F Felzenszwalb, R B Girshick, D McAllester, and D Ramanan. Object Detection with Discriminatively Trained Part-Based Models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(9):1627–1645, September 2010.

[9] J. Gall, A. Yao, N. Razavi, L. Van Gool, and V. Lempitsky. Hough Forests for Object Detection, Tracking, and Action Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 33(11):2188–2202, November 2011.

[10] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep Residual Learning for Image Recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 770–778, Las Vegas, NV, USA, June 2016. IEEE.

[11] Mohammad Hossain, Mehrdad Hosseinzadeh, Omit Chanda, and Yang Wang. Crowd Counting Using Scale-Aware Attention Networks. In *2019 IEEE Winter Conference on Applications of Computer Vision (WACV)*, pages 1280–1288, Waikoloa Village, HI, USA, January 2019. IEEE.

[12] Ya-Li Hou and Grantham K. H. Pang. People Counting and Human Detection in a Challenging Situation. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 41(1):24–33, January 2011.

[13] Haroon Idrees, Imran Saleemi, Cody Seibert, and Mubarak Shah. Multi-source Multi-scale Counting in Extremely Dense Crowd Images. In *2013 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2547–2554, June 2013. ISSN: 1063-6919.

[14] Xiaoheng Jiang, Li Zhang, Tianzhu Zhang, Pei Lv, Bing Zhou, Yanwei Pang, Mingliang Xu, and Changsheng Xu. Density-Aware Multi-Task Learning for Crowd Counting. *IEEE Transactions on Multimedia*, 23:443–453, 2021.

[15] Xiaolong Jiang, Zehao Xiao, Baochang Zhang, Xiantong Zhen, Xianbin Cao, David Doermann, and Ling Shao. Crowd Counting and Density Estimation by Trellis Encoder-Decoder Networks. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 6126–6135, Long Beach, CA, USA, June 2019. IEEE.

[16] B. Leibe, E. Seemann, and B. Schiele. Pedestrian Detection in Crowded Scenes. In *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, volume 1, pages 878–885, San Diego, CA, USA, 2005. IEEE.

[17] Victor Lempitsky and Andrew Zisserman. Learning To Count Objects in Images. In *Advances in Neural Information Processing Systems*, volume 23. Curran Associates, Inc., 2010.

[18] Bo Li, Hongbo Huang, Ang Zhang, Peiwen Liu, and Cheng Liu. Approaches on crowd counting and density estimation:

a review. *Pattern Analysis and Applications*, 24(3):853–874, August 2021.

[19] Min Li, Zhaoxiang Zhang, Kaiqi Huang, and Tieniu Tan. Estimating the number of people in crowded scenes by MID based foreground segmentation and head-shoulder detection. In *2008 19th International Conference on Pattern Recognition*, pages 1–4, December 2008. ISSN: 1051-4651.

[20] Zhiheng Ma, Xing Wei, Xiaopeng Hong, and Yihong Gong. Bayesian Loss for Crowd Count Estimation With Point Supervision. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 6141–6150, Seoul, Korea (South), October 2019. IEEE.

[21] Ramin Mehran, Alexis Oyama, and Mubarak Shah. Abnormal crowd behavior detection using social force model. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 935–942, Miami, FL, June 2009. IEEE.

[22] David Ryan, Simon Denman, Clinton Fookes, and Sridha Sridharan. Crowd Counting Using Multiple Local Features. In *2009 Digital Image Computing: Techniques and Applications*, pages 81–88, December 2009.

[23] Payam Sabzmeydani and Greg Mori. Detecting Pedestrians by Learning Shapelet Features. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, June 2007. ISSN: 1063-6919.

[24] Deepak Babu Sam, Shiv Surya, and R. Venkatesh Babu. Switching Convolutional Neural Network for Crowd Counting. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4031–4039, Honolulu, HI, July 2017. IEEE.

[25] Mamoona Birkhez Shami, Salman Maqbool, Hasan Sajid, Yasar Ayaz, and Sen-Ching Samson Cheung. People Counting in Dense Crowd Images Using Sparse Head Detections. *IEEE Transactions on Circuits and Systems for Video Technology*, 29(9):2627–2636, September 2019.

[26] Sheng-Fuu Lin, Jaw-Yeh Chen, and Hung-Xin Chao. Estimation of number of people in crowded scenes using perspective transformation. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, 31(6):645–654, November 2001.

[27] Vishwanath A. Sindagi and Vishal M. Patel. A survey of recent advances in CNN-based single image crowd counting and density estimation. *Pattern Recognition Letters*, 107:3–16, May 2018.

[28] C. Stauffer and W.E.L. Grimson. Adaptive background mixture models for real-time tracking. In *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149)*, pages 246–252, Fort Collins, CO, USA, 1999. IEEE Comput. Soc.

[29] Venkatesh Bala Subburaman, Adrien Descamps, and Cyril Carincotte. Counting People in the Crowd Using a Generic Head Detector. In *2012 IEEE Ninth International Conference on Advanced Video and Signal-Based Surveillance*, pages 470–475, September 2012.

[30] Yukun Tian, Yiming Lei, Junping Zhang, and James Z. Wang. PaDNet: Pan-Density Crowd Counting. *IEEE Transactions on Image Processing*, 29:2714–2727, 2020.

[31] O. Tuzel, F. Porikli, and P. Meer. Pedestrian Detection via Classification on Riemannian Manifolds. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 30(10):1713–1727, October 2008.

[32] Viola, Jones, and Snow. Detecting pedestrians using patterns of motion and appearance. In *Proceedings Ninth IEEE International Conference on Computer Vision*, pages 734–741 vol.2, October 2003.

[33] P. Viola and M. Jones. Rapid object detection using a boosted cascade of simple features. In *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001*, volume 1, pages I–511–I–518, Kauai, HI, USA, 2001. IEEE Comput. Soc.

[34] Paul Viola and Michael J. Jones. Robust Real-Time Face Detection. *International Journal of Computer Vision*, 57(2):137–154, May 2004.

[35] Boyu Wang, Huidong Liu, Dimitris Samaras, and Minh Hoai Nguyen. Distribution Matching for Crowd Counting. In *Advances in Neural Information Processing Systems*, volume 33, pages 1595–1607. Curran Associates, Inc., 2020.

[36] Bo Wu and R. Nevatia. Detection of multiple, partially occluded humans in a single image by Bayesian combination of edgelet part detectors. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, volume 1, pages 90–97 Vol. 1, October 2005. ISSN: 2380-7504.

[37] Bo Wu and Ram Nevatia. Detection and Tracking of Multiple, Partially Occluded Humans by Bayesian Combination of Edgelet based Part Detectors. *International Journal of Computer Vision*, 75(2):247–266, August 2007.

[38] Zhaoyi Yan, Yuchen Yuan, Wangmeng Zuo, Xiao Tan, Yezhen Wang, Shilei Wen, and Errui Ding. Perspective-Guided Convolution Networks for Crowd Counting. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 952–961, Seoul, Korea (South), October 2019. IEEE.

[39] Yifan Yang, Guorong Li, Zhe Wu, Li Su, Qingming Huang, and Nicu Sebe. Reverse Perspective Network for Perspective-Aware Object Counting. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 4373–4382, Seattle, WA, USA, June 2020. IEEE.

[40] Anran Zhang, Jun Xu, Xiaoyan Luo, Xianbin Cao, and Xiantong Zhen. Cross-Domain Attention Network for Unsupervised Domain Adaptation Crowd Counting. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(10):6686–6699, October 2022.

[41] Cong Zhang, Hongsheng Li, Xiaogang Wang, and Xiaokang Yang. Cross-Scene Crowd Counting via Deep Convolutional Neural Networks. pages 833–841, 2015.

[42] Yingying Zhang, Desen Zhou, Siqin Chen, Shenghua Gao, and Yi Ma. Single-Image Crowd Counting via Multi-Column Convolutional Neural Network. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 589–597, Las Vegas, NV, USA, June 2016. IEEE.