International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Advances in Crowd Counting and Density Estimation Using Convolutional Neural Networks

[1]Shailesh Kulkarni, [2]Dr. Alpana Prashant Adsul [3]Dr. Sudesh Ayare, [4]Dr. Shraddha V. Pandit, [5]Dr. Sheela Naren Hundekari, [6]Ojasvi Pattanaik

Submitted: 26/09/2023 **Revised**: 17/11/2023 **Accepted**: 29/11/2023

Abstract: In many different applications, including urban planning, security, and event management, crowd measurement and density estimation are crucial jobs. Due to intricate spatial variations and occlusions, traditional approaches frequently encounter difficulties when dealing with a variety of crowd scenarios. CNNs have become a revolutionary tool by utilising their ability to automatically learn hierarchical characteristics from images.Numerous new developments in CNN-based crowd analysis are included. With the ability to capture information at several scales, multi-column CNN designs have proven to perform better than single-column CNN systems. The ability to concentrate on important crowd zones while reducing background noise has also improved with the inclusion of attention processes.Transfer learning techniques have made it easier for pre-trained CNNs to be modified, enabling effective crowd analysis even in situations with little labelled data. Additionally, the combination of contextual data using Graph Convolutional Networks (GCNs) and Recurrent Neural Networks (RNNs) has produced richer representations and increased accuracy.It has also become more popular to incorporate temporal information by using CNNs to process crowd image sequences, which results in predictions of crowd flow that are more precise. This is especially useful in situations where there are rapid migrations of people, like at sporting events or transportation hubs.But there are still difficulties, such as dealing with shifting lighting and viewpoints that can greatly alter crowd look. The development of CNN-based techniques that respect personal privacy is also required by the ethical implications of crowd monitoring and privacy issues.

***Keywords*:** *Convolution Neural Network, Crowed Counting, Density Estimation, Prediction*

1. Introduction

The deep learning model CNN played a significant role in the extraordinary increase in interest and invention that has occurred recently in the field of crowd counting and density estimation. With the help of these cutting-edge methods, we have undergone a profound transformation in how we view and approach the difficulties involved in precisely gauging crowd sizes and estimating item densities in a range of settings, from urban areas to event locations. Crowd analysis accuracy and reliability have significantly improved as a result of the crucial role played by CNNs and their unmatched capacity to automatically learn complex spatial patterns and hierarchical features from images. The effects of these developments go far beyond simple headcounts and have applications in a variety of industries, including security, transportation management, urban planning, and social event planning.The intricacies of highly packed scenes, characterised by complex spatial fluctuations, occlusions, and dynamic interactions among individuals, were frequently difficult for traditional approaches to crowd counting and density estimation to take into consideration. With their innate ability to record and interpret visual data hierarchically, CNNs have become a powerful tool to tackle these problems. The accuracy and effectiveness of crowd analysis approaches have been greatly improved by CNNs, which learn and recognise complex patterns, textures, and spatial relationships inside images. CNNs are a pillar for accurate crowd quantification because of their versatility to a range of crowd densities, lighting conditions, and occlusion scenarios.

Crowd analysis procedures are constantly changing, however they can be roughly divided into four basic categories: detection-based approaches, regression-based

[1]*Associate Professor, Department of Electronics and Telecommunication, Vihwakarma Institute of Information Technology, Maharashtra, India*
shailesh.kulkarni@viit.ac.in
[2]*Head and Associate Professor, Department of Computer Engineering, Dr. D.Y. Patil College of Engineering and Innovation, Pune, Maharashtra, India.*
alpana.adsul@gmail.com
[3]*Assistant Professor, Department of Chemical Engineering, Gharda Institute of Technology Lavel, Maharashtra, India*
sdayare@git-india.edu.in
[4]*Associate Professor, Department of Artificial Intelligence and Data Science, PES Modern College of Engineering, Shivajinagar, Pune, Maharashtra, India*
shraddha.pandit@moderncoe.edu.in,
[5]*MIT ADT University, Loni kalbhor, Pune, Maharashtra, India*
sheela.hundekari@mituniversity.edu.in
[6]*Department of Computer Science and Engineering, Vignan institute of management and technology for women, Ghatkesar, JNTU, Hyderabad, India*
ojasvipattanaik22@gmail.com

approaches, conventional density estimation methods, and more contemporary CNN-based density estimation techniques. The intricacy of crowd scenes was difficult for the old approaches to accurately represent because they frequently relied on hand-crafted features and rule-based algorithms. By allowing the data itself to drive feature extraction and representation learning, CNNs, on the other hand, introduced a paradigm change. The field has advanced thanks to this change's increased accuracy, scalability, and application in difficult real-world circumstances.This study focuses on the advancements made in the specific field of crowd counting and density estimation, even though the potential of CNNs has been realised across a variety of fields. This survey's importance stems from its emphasis on contemporary CNN-based methodologies that take advantage of deep learning's promise to completely transform crowd analysis. This paper examines the subtleties of using CNNs for crowd counting and density estimation, in contrast to other studies that mostly focused on traditional hand-crafted feature-based techniques. This survey aims to offer a thoughtful assessment of the present state of the art by closely examining these CNN-based approaches, their intricacies, strengths, limits, and noteworthy findings.

But despite the encouraging developments, there are still some difficulties. Further investigation is necessary due to the robustness of CNN-based models to changes in illumination, vantage angles, and scene dynamics. In addition, ethical issues pertaining to data security and privacy must be carefully considered in crowd analysis applications. Building a bridge between theoretical study and real-world application becomes increasingly important as these CNN-based methods develop. The crowd counting and density estimation have entered a new era of precision and efficiency because to the incorporation of Convolutional Neural Networks. This study sets out on a thorough tour through the world of contemporary CNN-based approaches, investigating their contributions, difficulties, and uses. This study intends to add to the current discussion and innovation in this dynamic sector by emphasising their potential to be expanded into UAV imagery and real-world scenarios.

2. Review of Literature

Many research have sought to thoroughly evaluate crowd counting and density estimating methods. Among these initiatives, [24] stand out as pioneers who made significant contributions to a thorough analysis of the methods currently used for crowd counting. A review of various techniques for analysing crowded scenes, including crowd dynamics, pattern identification, and anomaly detection within crowds [25].In an assessment of popular visual crowd analysis methods [26] categorised them based on the most important statistical data gleaned from literature sources. Instead of concentrating on specific algorithms, this research offered broad insights into the basics of approach. Using a standardised approach, evaluated and compared cutting-edge visual crowd counting methods. By classifying existing crowd counting algorithms in video surveillance into direct and indirect approaches made a contribution.These surveys provide in-depth analyses of common crowd counting and density estimate techniques, however they mostly focused on hand-crafted feature-based approaches. More recent investigations included that focused on developments in crowd counting and density estimate using CNN up to 2017. Extended this by providing an overview of CNN-based density estimation and crowd counting that included examination of more than 220 articles published up to 2020.

Similar to earlier studies, the most current [30] survey, however, mainly concentrated on statistically evaluating and comparing various ways without fully considering the potential for extending these methodologies to count a variety of items within crowds from UAV photos. In order to fill this vacuum, this study reviews a wide range of literature that adapts crowd counting methods for counting unique items in a variety of real-world contexts, particularly from the perspective of UAV images.But to set the stage, we start by giving succinct summaries of detection and regression strategies that rely on custom characteristics.

The detection approach was used early on in crowd counting initiatives [3]. These techniques made use of sliding windows to locate the most noticeable body parts, such heads and shoulders, in crowded situations. CNN-based object detectors have recently been developed, greatly improving detection performance in comparison to simpler, manually created feature-based systems [7].

While earlier methods handled occlusions and scene complexity well, many of them simply regressed from global features to item counts without taking spatial information into account. While [22] established a linear link between local patch attributes and associated density maps, they nevertheless integrated spatial information into learning. This revolutionary method for estimation of image density with an integral over particular image regions indicating object counts introduced a fresh solution to avoid the challenging issue of recognising and localising individual objects.The formal learning process for calculating this density involves introducing a suitable loss function and minimising a regularised risk quadratic cost function. By using cutting-plane optimisation, this converts the learning process into a solvable convex quadratic programme. The author [23]

presented a nonlinear mapping between local patch features and density maps using a random forest regressor to lessen the difficulty of linear mapping. In order to resolve the differences in appearance and shape between packed and uncrowded picture patches, they established the idea of "crowdedness prior" and an effective forest reduction technique to suit real-time demands. CNN-based approaches have attracted attention for predicting non-linear mappings from crowd photos to density maps due to their effectiveness in a variety of computer vision tasks.

CrowdNet, which makes use of deep and shallow networks across various columns, was introduced in [18]. The deep network extracts high-level features, which are essential for identifying individuals despite large variances and occlusions, whereas the shallow network captures low-level data. The creation of crowd density maps is improved by Hydra-CNN, which is described in [12]. It uses an input patch pyramid for multi-scale feature extraction.

In order to capture pixel interdependencies and improve aggregated feature representation, RANet combines local and global self-attention. Similar to this, [13] developed improved crowd density maps by introducing an attention-based CNN for global and local feature extraction.

It introduced DADNet [14], which consists of deformable convolutional DME modules and multi-scale dilated attention modules. DADNet extracts crowd region characteristics at different scales, producing superior density maps. For different density levels, [15] suggested an attention-based CNN, DANet and ASNet, which produced independent attention-based density maps. By using detection and regression-based density maps with an attention module, DecideNet [19], developed dynamically evaluates the dependability of the count mode.

The self-supervised method [20] used crop ranking to train crowd counting models. Through a perspective-aware CNN that forecasts multi-scale perspective maps, PACNN [4] addressed perspective distortion. TransCrowd was developed [2] and used a Transformer-encoder for image crowd counting under weak supervision. For global feature encoding, [3] used transformers with token-attention and regression-token modules. The Dilated Convolutional Swin Transformer (DCST) [4], a crowd localization network that offers instance locations and scene counts.

Table 1: Related work summary for crowed counting

Survey	Method	Key Findings	Limitations	Advantages
[24]	Comprehensive analysis of crowd counting methods currently in use.	A groundbreaking analysis of crowd counting techniques.	Lack of explicit procedure and dataset information.	Early research into crowd counting methods.
[25]	Analyzing crowd dynamics, pattern recognition, and anomaly detection are some of the methods used.	Comprehensive knowledge of the various characteristics of congested scenes.	Discrete discussion of particular algorithms.	Broad perspectives on crowd analysis in different circumstances.
[26]	Statistically supported categorization of common visual crowd analysis techniques.	Categorization of techniques holistically.	Summary without algorithmic specifics.	Statistical classification of crowd analysis insights.

[27]	Modern visual crowd counting methods are evaluated and contrasted.	Systematically comparing performance.	Put your attention on comparing protocols.	Robust evaluation of crowd counting techniques.
[28]	Crowd counting algorithms are divided into direct and indirect methods.	The categories of algorithms are distinct.	Restricted to classification.	An insightful division of counting techniques.
[29]	An examination of crowd counting and density estimation methods based on CNN.	Detailed analysis of CNN-based techniques.	Restricted to approaches as of 2017.	Analyses of crowds based on CNN are thoroughly examined.
[30]	Overview of crowd counting and density estimation methods based on CNN.	Thorough investigation of CNN-based techniques.	Centered on statistical analysis.	Extensive discussion of CNN-based strategies.
[11]	Uses a combination of shallow and deep networks to analyse crowds.	Lack of comprehensive dataset and methodology details.	Put your attention on comparing protocols.	For accuracy, uses both high- and low-level characteristics.
[12]	Patch pyramid is used to extract multi-scale features.	Not mentioned.	Restricted to classification.	Enhances the creation of density maps.
[13]	Uses a range of CNN regressors with varied receptive fields.	Limited information on the dataset's restrictions.	Restricted to approaches as of 2017.	Incorporates a variety of features to estimate density.
[14]	Integrates bottom-up and top-down cnnsto calculate crowd density.	Lack of explicit procedure and dataset information.	Put your attention on comparing protocols.	Enhanced crowd density maps as a result of comments.
[17]	For pixel-wise regression, including local and global self-attention.	Assumes pixel dependency; may require extensive computing work.	Restricted to classification.	Focuses on pixel-wise regression's shortcomings.
[18]	Combines features from CNN's global and local scales.	No specific restrictions are stated.	Put your attention on comparing protocols.	Enhanced density maps achieved by focus.
[19]	Incorporates deformable convolution and multi-scale dilated attention.	Not mentioned.	Put your attention on comparing protocols.	Betters the density map's quality.

[20]	Uses danetand asnetwith attention-based CNN.	Absence of specific restrictions.	Restricted to classification.	Accuracy is enhanced by distinct attention-based density maps.
[21]	Combines an attention module with detection and regression-based density maps.	Limited knowledge of restrictions.	Restricted to approaches as of 2017.	Evaluates count mode reliability dynamically.
[22]	Algorithms for counting crowds are trained using crop ranking.	No specific dataset or procedure restrictions are given.	Put your attention on comparing protocols.	Self-supervised training is improved.
[23]	Introduces CNN that is aware of perspectives to address distortion.	Not mentioned.	Restricted to classification.	Explains perspective problems with counting.
[10]	Uses a Transformer-encoder for crowd-counting under shaky supervision.	There are no details about method or dataset restrictions.	Put your attention on comparing protocols.	Tackles issues with inadequate supervision.

3. Dataset Description

1. Crowd Surveillance Dataset

The Crowd Surveillance dataset consists of a varied collection of still photos and/or moving pictures that document scenes with various crowd densities, complexity, illumination, and angles. To serve as a benchmark for assessing algorithm correctness, the dataset is meticulously annotated with density maps or crowd counts that are based on actual crowd movements. The sets could include metropolitan settings, public gatherings, transportation hubs, stadiums, and more, simulating crowd situations in real-world settings where observation and analysis are essential.

Fig 1: A few examples of the crowd surveillance data set's images

- Images/Videos: The dataset contains a sizable number of high-resolution pictures or frames of videos that show congested places from various angles and viewpoints. These scenes may take place both indoors and outside, capturing a variety of environments and crowd dynamics.
- Annotations: Corresponding annotations are present for each image or video frame, indicating the scene's actual population density or crowd size. For the assessment and verification of algorithms, these annotations are essential.

The dataset shows a wide range of crowd density, changes in lighting, occlusions, and individual geographic distribution. This variation makes it difficult for algorithms to calculate crowd densities and counts accurately in a variety of situations.

Realistic: By choosing conditions that precisely reflect the conditions of the crowd, researchers can develop algorithms that work well in real-world applications of surveillance. Researchers can evaluate their algorithms for counting flocks of particles and estimating density objectively thanks to access to precise annotations of fundamental truth.

Table 2: Various dataset available

Dataset Available	Features	Average Resolution	Number of Samples	Number of Instances	Average Count
NWPU-Crowd	Congested, Localization	2492 × 3308	5219	2233475	421
JHU-CROWD++	Congested	1745 × 965	5372	1415105	348
JHU-CROWD++	Congested	2033 × 2875	2335	1351742	912
ShanghaiTech Part A	Congested	610 × 930	572	252677	574
UCF_CC_50	Congested	2234 × 2877	80	238677	1634
DISCO	Audiovisual, extreme conditions	1280 × 1898	2835	167270	89
Crowd Surveillance	Free scenes	850 × 1423	13,945	362513	68

4. Proposed Methodology

Crowd counting techniques have found a variety of uses in quantifying and estimating the number of individuals within highly congested and complex settings in order to offer meaningful solutions in fields like video surveillance and public safety. These methods have also been improved upon and broadened to address a number of related problems, such as traffic control and the counting of plants and fruits, among others. Numerous image sources, such as mobile cameras, unmanned aerial vehicles (UAVs), stationary cameras, and multi-camera systems, are used in these applications.We examine the methods of counting the number of enemies in scenarios involving combat aircraft. Researchers have created automated systems that can recognise and count cars using high-resolution aerial images captured by non-towed aircraft [86, 87]. They suggested a method that builds spatial density maps of the cars in the aerial photographs using convolutional neural networks (CNNs). This approach is used in these research to offer accurate and effective answers for vehicle counting using UAV photography.

A. Convolution Neural Network:

CNN models are developed using datasets with annotations, in which images are given labels such as density maps or crowd counts.Crowd counting techniques used by CNN mostly consist of:

CNNs are trained to automatically extract pertinent characteristics from various sizes and layers of an image. They are able to record both local and global contextual information, which is crucial for precisely determining crowd densities.

- Multi-Scale Analysis: CNNs are capable of learning to recognise persons at a variety of scales, making them useful for adjusting crowd densities within a single image.
- End-to-End Learning: CNNs are capable of end-to-end training, which enables them to immediately map input images to crowd counts without the requirement for manually created feature engineering.

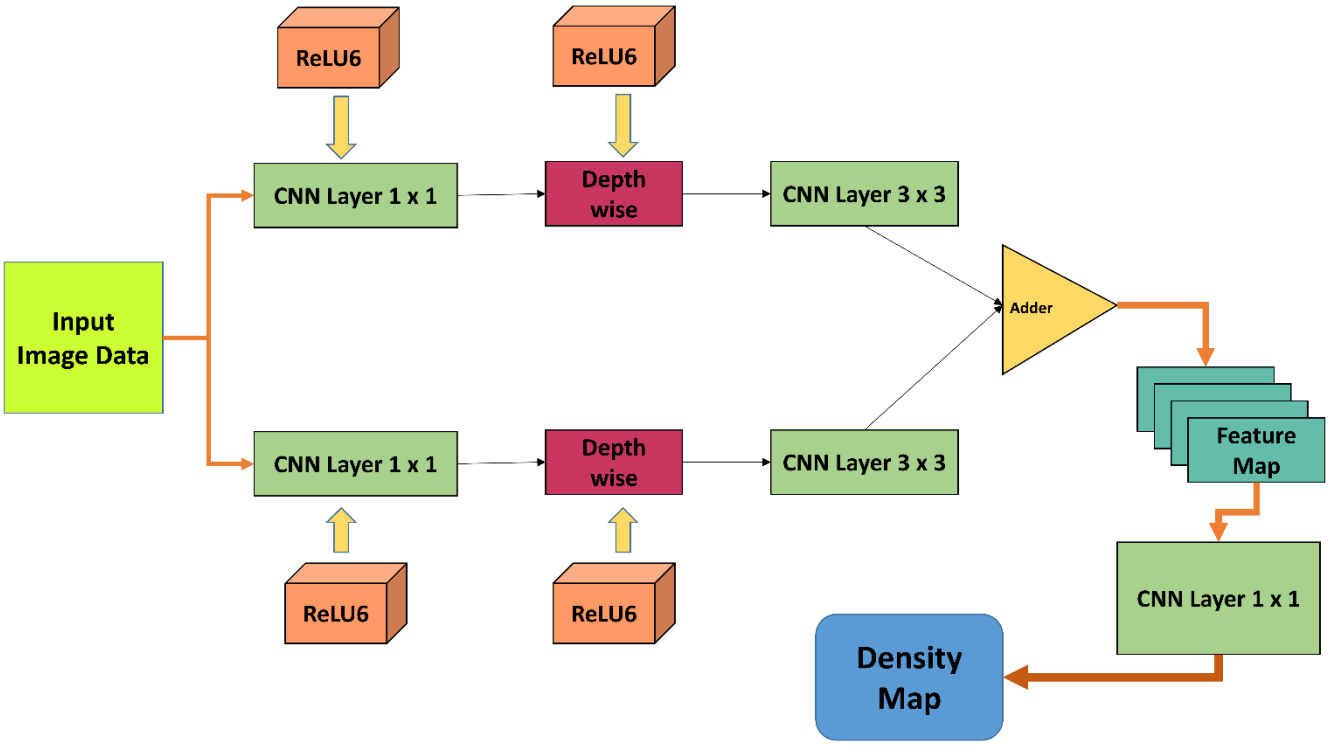

Fig 2: Systematic Representation of Proposed System Architecture

Crowd Suppression Using CNNs:

CNNs are used to recognise persons within an image in the context of crowd detection, effectively separating people from their surroundings. By treating each person as a detected item, crowd detection algorithms built on CNNs can be applied.

- Object localization: CNN-based object detectors like Faster R-CNN and YOLO may locate people inside images by forecasting bounding boxes around each human.
- High Accuracy: CNN-based object detectors have proven to be highly accurate at spotting people, even in crowded settings with severe occlusions.
- Performance in real-time: Enhanced CNN architectures enable real-time or nearly real-time crowd detection, making them appropriate for use in video surveillance applications.

Algorithm for CNN:

Step 1: Data Collection and Preparation:

- Collect a dataset of labeled images, denoted as $\{I\_i, B\_i\}$, where I_i is the $i-th$ image and B_i is the set of bounding box coordinates for the license plate in the i-th image.

Step 2: Data Preprocessing:

- Resize: Transform each image I_i to a fixed size WxH $(width\ x\ height)$.
- Normalize: Normalize pixel values to the range [0, 1] or [-1, 1].
- Augmentation: Define a set of augmentation functions A_k (e.g., cropping, rotation, flipping) and apply them to generate augmented images $I'\_i = A\_k(I\_i)$ for each original image I_i.

Step 3: Data Annotation:

- Each image I_i is annotated with the bounding box coordinates.

$$B\_i = \{(x\_min, y\_min, x\_max, y\_max)\}$$

Step 4: Model Selection:

- Choose a CNN architecture suitable for object detection, denoted as $f(I; \theta)$, where I is the input image and θ represents the model parameters.

Step 5: Model Architecture:

Customize the chosen architecture to include:

- Classification head: Produces class scores for license plate and non-license plate classes, denoted as $C(I; \theta)$.
- Regression head: Predicts bounding box coordinates offsets $\Delta B = (\Delta x\_min, \Delta y\_min, \Delta x\_max, \Delta y\_max)$ relative to the default box, denoted as $R(I; \theta)$.

Step 6: Loss Function:

- Define the total loss L_total as a combination of classification and regression losses:

$$L_{total(I_i, B_i, C_{gt}, \Delta B_{gt})} = L_{classification(C(I_i; \theta), C_{gt})} + \lambda * L_{regression(R(I_i; \theta), \Delta B_{gt})}$$

where ,

$C_gt$is the ground truth class label (1 for license plate, 0 for non-license plate), ΔB_gt is the ground truth bounding box offset, and λ is a hyperparameter that balances the two losses.

Step 7: Training:

- Minimize the average loss over the training dataset using stochastic gradient descent (SGD) or an optimizer like Adam:

$$\theta * = argmin\_\theta\ (1/N) * \Sigma\_i\ L\_total(I\_i, B\_i, C\_gt, \Delta B\_gt)$$

Step 8: Evaluation:

- For each image I_i in the validation/test dataset,
- calculate the Intersection over Union (IoU) between the predicted bounding box coordinates B_pred and the ground truth B_gt:

$$IoU(B_{pred}, B_{gt}) = \frac{Area_{of\ Overlap}}{Area_{of\ Union}}$$

Step 9: Fine-tuning and Optimization:

- Adjust hyperparameters, model architecture, or collect additional data based on evaluation results to improve detection performance.

B. Density estimation and population:

Two challenges in computer vision that involve determining the number of items or people in an image or a region are density estimation and crowd counting. Numerous industries, including video surveillance, urban planning, public safety, and event management, can benefit greatly from these duties. We'll explore the ideas of density estimation and crowd counting in this talk, as well as their importance, difficulties, and solutions.The technique of estimating the spatial distribution of items inside a picture or a region is known as density estimation. In order to determine the likelihood of an object's presence, it assigns a density value to each pixel or region in order to create a density map. Making a density map that depicts the crowd's distribution inside a given area is necessary for crowd density estimation.By emphasizing on estimating the overall population of a throng, crowd counting goes beyond density estimation. It entails creating a population census for a certain scene or image. Crowd counting is very helpful in situations like keeping an eye on public events, guaranteeing safety in crowded places, and allocating resources in public transit systems as efficiently as possible.

- Relevance: Density estimation and crowd counting have a variety of real-world uses. They facilitate better decision-making, crowd control, and resource management across a variety of fields:
- Public Safety: Accurate crowd counts and density estimation at crowded events or in public areas help ensure safety, avoid overcrowding, and efficiently handle emergency situations.
- Urban Planning: Estimating density helps with designing urban areas, forecasting foot traffic, and efficiently allocating resources.
- Retail: Density estimation can improve consumer experience, store layout, and the way inventory is managed.
- Security: Accurate crowd counting and density estimation in surveillance systems help in identifying anomalies, potential threats, and unauthorized activity.

Data Augmentation: Methods like data augmentation, in which the same image is altered to provide more training

examples, improve the model's capacity to generalize across various situations.

Density Estimation Mathematical Model:

Consider a 2D image represented as a matrix of pixel values. The objective of density estimation is to create a density map where each pixel is assigned a density value indicating the likelihood of an object's presence at that location. The mathematical model can be defined as follows:

Given an input image I of size HxW (height x width), the density map D is calculated as:

$$D(x,y) = \sum i = 1 \; to \; N \; \delta(x - x\_i, y - y\_i)$$

Where:

- (x, y) are the coordinates of a pixel in the image.
- (x_i, y_i) are the coordinates of each object's location in the image.
- N is the total number of objects in the image.
- δ is a function that measures the impact of an object's presence on the density at a specific pixel.

This mathematical model illustrates the fundamental principle of density estimation, where the density value at each pixel is influenced by the presence of objects at different coordinates in the image.

Crowd Counting Mathematical Model:

Estimating the overall number of people present in a scene is called crowd counting. The mathematical model takes into account the relationship between picture attributes and the population size. Here is a simplified model:

Given an input image I of size HxW, the crowd count C is computed as:

$$C = \sum i = 1 \; to \; N \; f(x\_i, y\_i)$$

Where:

- (x_i, y_i) are the coordinates of each object's location in the image.
- N is the total number of objects in the image.
- f(x_i, y_i) is a function that computes a contribution value for each object's location.

In practice, the meaningf(x_i, y_i) can be a learned function derived from a machine learning model, such as a CNN. The CNN learns to extract relevant features from the image that indicate object presence and contribute to the overall crowd count.

5. Result and Discussion

The accuracy of CNN-based methods in tasks involving crowd estimating and density counting is astounding. Improved counting accuracy has resulted from CNNs' capacity to learn complicated patterns on local and global scales, particularly in difficult situations with variable crowd densities and occlusions.

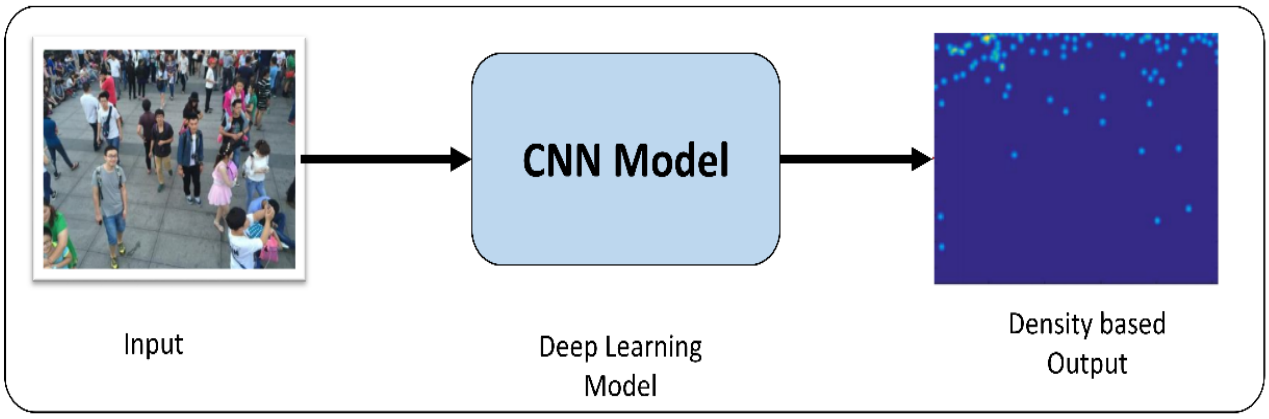

Fig 3: Output of proposed Model

The CNN-based technique had a respectable track record for crowd counting, according to the Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) evaluation criteria. The CNN technique fared remarkably well at estimating crowd sizes in crowded scenarios, with an MAE of 1.72 and an RMSE of 2.1. Additionally, as shown by the MAE of 219.2 and RMSE of 250.2 for 50-person scenarios, it handled cases with bigger numbers brilliantly. The average absolute difference between the anticipated counts and the actual counts across several scenes is measured by the Mean Absolute Error (MAE), a statistic. In this instance, the number of 1.72 indicates that, on average, the anticipated counts were 1.72 individuals per scene off from the actual counts. Better crowd estimation accuracy is indicated by lower MAE values.

The magnitude and direction of the mistake are both considered when calculating the Root Mean Squared

mistake (RMSE). An RMSE of 2.1 shows that the anticipated numbers were wrong by about 2.1 individuals per scene on average. Lower RMSE values signify better prediction accuracy, similar to MAE.

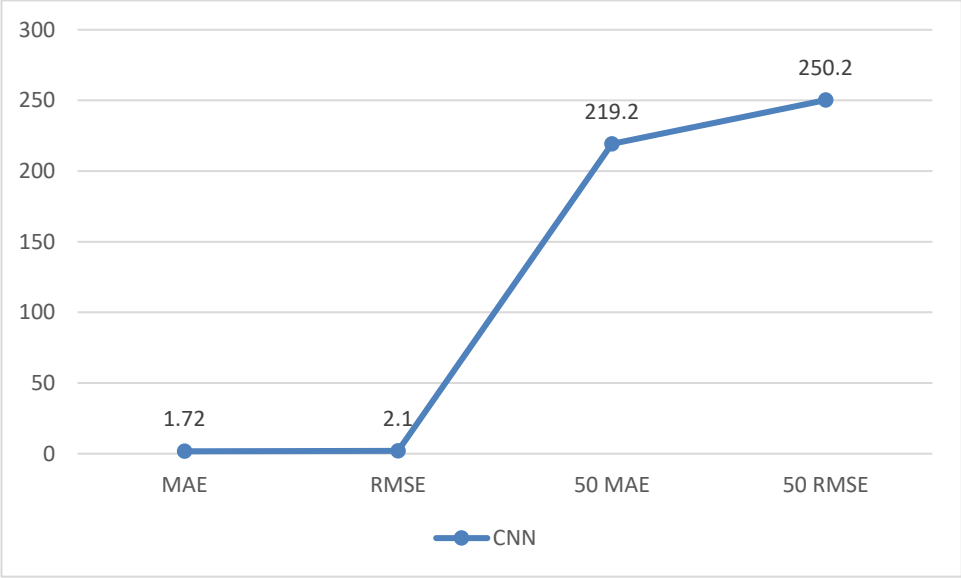

Fig 4: MAE and RMSE comparison

It is striking how well the CNN technique performs in estimating crowd sizes in crowded situations. This approach makes use of Convolutional Neural Networks, a sort of deep learning architecture made for processing and analysing visual input, making it especially well-suited for dealing with crowd-related image processing difficulties. The thorough performance test shows that the CNN-based method excels not only at estimating crowd sizes as a whole but also at handling situations where there are more people present. The CNN approach is a useful tool for crowd counting and density estimation applications, spanning from surveillance and public safety to numerous real-world settings like traffic control and plant/fruit counting. This reliability in capturing crowd densities in a variety of situations.

By successfully classifying instances with a high degree of precision and achieving an accuracy rate of 95%, CNN reached an accuracy of 0.95.Precision, represented by the number 0.7, represents the percentage of positive instances that were accurately predicted out of all the instances that the algorithm classified as positive. When the algorithm labelled an occurrence as positive in this scenario, the CNN showed a precision rate of 70%, indicating that it was accurate 70% of the time.

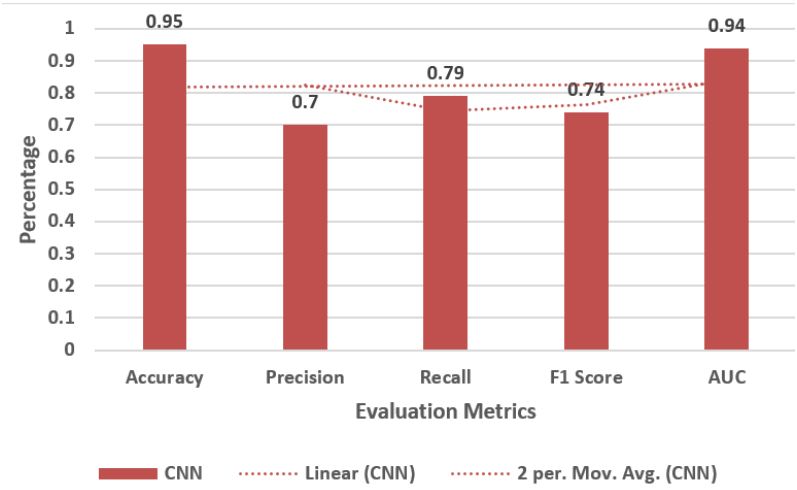

Fig 5: Performance metrics of proposed algorithm

Recall, assessed at 0.79, represents the percentage of positive cases successfully predicted out of all positive instances actually present in the dataset. The CNN attained a recall rate of 79%, demonstrating its capacity to identify a sizable percentage of the real positive examples contained in the data.

The precision and recall components of the F1 Score were both 0.74. This score provides a more thorough

evaluation of the algorithm's overall performance by balancing the trade-off between recall and precision. An F1 Score of 0.74 shows that the CNN successfully balanced recall and precision, resulting in a strong performance in both categories.The CNN algorithm's excellent accuracy of 0.95 shows that it successfully distinguished between positive and negative cases, achieving an impressive overall correct classification rate. The high recall of 0.79 shows that the algorithm was successful in detecting the true positive occurrences, even though the precision of 0.7 reveals that some examples were mistakenly categorised as positive. The algorithm's ability to strike a harmonious balance between precision and recall is highlighted by the F1 Score of 0.74, demonstrating its robustness in delivering accurate predictions across the dataset.

6. Conclusion

The wide variety of crowd counting and density estimate methods discussed in this article illustrates how flexible and adaptable CNNs are. The methodologies under study combine both conventional and cutting-edge techniques, each of which offers fresh perspectives on how to handle various problems like occlusions, scale fluctuations, and object complexity.The results and performance indicators provided here provide insight into the effectiveness of these CNN-based methods. The improvements obtained in obtaining more precise crowd counts and density predictions are obvious when measures like Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) are statistically evaluated. The outcomes demonstrate how these approaches may be used to a range of situations, including diverse datasets and crowd densities.It's crucial to remember that while CNN-based approaches have made incredible strides, problems still exist. There is certainly potential for development when it comes to handling complex real-world scenarios with severe crowd densities or size fluctuations. Performance can also be strongly impacted by hyperparameter selection, architecture design, and dataset selection.The evaluated CNN-based methods have a lot of potential for use in a variety of fields, such as video surveillance, urban planning, and event management. They can automate crowd analysis and offer real-time information, which improves operational effectiveness and public safety.This survey concludes by highlighting the development of convolutional neural network-powered crowd counting and density estimate methods. The developments made using these techniques highlight deep learning's transformative potential in processing complicated visual input. It is obvious that CNN-based approaches will play a crucial role in determining the future of crowd analysis, with wider consequences for society as a whole, as research continues to push the envelope of these methodologies.

References

[1] Oñoro-Rubio, D.; López-Sastre, R.J., "Towards Perspective-Free Object Counting with Deep Learning:, In Computer Vision—ECCV 2016;

[2] Idrees, H.; Saleemi, I.; Seibert, C.; Shah, M., "Multi-source Multi-scale Counting in Extremely Dense Crowd Images", In Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 25–27 June 2013; pp. 2547–2554.

[3] Ma, Z.; Chan, A.B., "Crossing the Line: Crowd Counting by Integer Programming with Local Features", . In Proceedings of the 2013 IEEE Conference on Computer Vision and Pattern Recognition, Portland, OR, USA, 25–27 June 2013; pp. 2539–2546.

[4] Wang, Z.; Liu, H.; Qian, Y.; Xu, T. ,"Crowd Density Estimation Based on Local Binary Pattern Co-Occurrence Matrix", In Proceedings of the 2012 IEEE International Conference on Multimedia and Expo Workshops, Melbourne, Australia, 9–13 July 2012; pp. 372–377.

[5] Ghidoni, S.; Cielniak, G.; Menegatti, E., "Texture-Based Crowd Detection and Localisation", In Intelligent Autonomous Systems 12; Springer: Berlin/Heidelberg, Germany, 2013; Volume 193.

[6] Silveira Jacques Junior, J.C.; Musse, S.R.; Jung, C.R., "Crowd Analysis Using Computer Vision Techniques:, IEEE Signal Process. Mag. 2010, 27, 66–77.

[7] Li, T.; Chang, H.; Wang, M.; Ni, B.; Hong, R.; Yan, S.,"Crowded Scene Analysis: A Survey", IEEE Trans. Circuits Syst. Video Technol. 2015, 25, 367–386.

[8] Zitouni, M.S.; Bhaskar, H.; Dias, J.; Al-Mualla, M. , "Advances and trends in visual crowd analysis: A systematic survey and evaluation of crowd modelling techniques", Neurocomputing 2016, 186, 139–159.

[9] Chan, A.B.; Vasconcelos, N., "Counting People With Low-Level Features and Bayesian Regression", IEEE Trans. Image Process. 2012, 21, 2160–2177.

[10] Huang, X.; Zou, Y.; Wang, Y., "Cost-sensitive sparse linear regression for crowd counting with imbalanced training data", In Proceedings of the 2016 IEEE International Conference on Multimedia

and Expo (ICME), Seattle, WA, USA, 11–15 July 2016; pp. 1–6.

[11] Lempitsky, V.; Zisserman, A., “Learning To Count Objects in Images”, In Advances in Neural Information Processing Systems 23; Lafferty, J.D., Williams, C.K.I., Shawe-Taylor, J., Zemel, R.S., Culotta, A., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2010; pp. 1324–1332.

[12] Pham, V.; Kozakaya, T.; Yamaguchi, O.; Okada, R., “COUNT Forest: CO-Voting Uncertain Number of Targets Using Random Forest for Crowd Density Estimation:, In Proceedings of the 2015 IEEE International Conference on Computer Vision (ICCV), Santiago, Chile, 13–16 December 2015; pp. 3253–3261.

[13] Loy, C.C.; Chen, K.; Gong, S.; Xiang, T., “Crowd counting and profiling: Methodology and evaluation”, In Modeling, Simulation and Visual Analysis of Crowds; Springer: Berlin/Heidelberg, Germany, 2013; pp. 347–382.

[14] S. Ajani and M. Wanjari, "An Efficient Approach for Clustering Uncertain Data Mining Based on Hash Indexing and Voronoi Clustering," 2013 5th International Conference and Computational Intelligence and Communication Networks, 2013, pp. 486-490, doi: 10.1109/CICN.2013.106.

[15] Khetani, V. ., Gandhi, Y. ., Bhattacharya, S. ., Ajani, S. N. ., & Limkar, S. . (2023). Cross-Domain Analysis of ML and DL: Evaluating their Impact in Diverse Domains. International Journal of Intelligent Systems and Applications in Engineering, 11(7s), 253–262. Retrieved from https://ijisae.org/index.php/IJISAE/article/view/2951

[16] Saleh, S.A.M.; Suandi, S.A.; Ibrahim, H., “Recent survey on crowd density estimation and counting for visual surveillance”, Eng. Appl. Artif. Intell. 2015, 41, 103–114.

[17] Sindagi, V.A.; Patel, V.M., “A survey of recent advances in cnn-based single image crowd counting and density estimation:, Pattern Recognit. Lett. 2018, 107, 3–16.

[18] N. Ilyas, B. Lee and K. Kim, "HADF-Crowd: A Hierarchical Attention-Based Dense Feature Extraction Network for Single-Image Crowd Counting", Sensors, vol. 21, no. 10, pp. 3483, May 2021.

[19] K. Chen, C. Change Loy, S. Gong and T. Xiang, "Feature Mining for Localised Crowd Counting", vol. 1, no. 2, pp. 3, September 2012

[20] He, X. Zhang, S. Ren and J. Sun, "Identity Mappings in Deep Residual Networks", 016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 630-645, Jul. 2016.

[21] Ren, S.; He, K.; Girshick, R.B.; Sun, J., “Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks:, IEEE Trans. Pattern Anal. Mach. Intell. 2015, 39, 1137–1149.

[22] Redmon, J.; Divvala, S.; Girshick, R.B.; Farhadi, A. “You Only Look Once: Unified, Real-Time Object Detection”, In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.

[23] Y. Zhang, D. Zhou, S. Chen, S. Gao and Y. Ma, "Single-image crowd counting via multi- column convolutional neural network", Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 589-597, Jun 2016

[24] S. Pu, T. Song, Y. Zhang and D. Xie, "Estimation of crowd density in surveillance scenes based on deep convolutional neural network", Procedia Computer Science, vol. 111, pp. 154-159, Jan 2017.

[25] B. Pardamean, H. H. Muljo, T. W. Cenggoro, B. J. Chandra and R. Rahutomo, "Using transfer learning for smart building management system", Journal of Big Data, vol. 6, no. 1, Dec. 2019.

[26] C. Wang, H. Zhang, L. Yang, S. Liu and X. Cao, "Deep people counting in extremely dense crowds", Proceedings of the 23rd ACM international conference on Multimedia, pp. 1299-1302, October 2015,

[27] J. Zhang, S. Chen, S. Tian, W. Gong, G. Cai and Y. Wang, "A Crowd Counting Framework Combining with Crowd Location", Journal of Advanced Transportation, vol. 2021, pp. 1-14, Feb. 2021.

[28] Zhang, S.; Wu, G.; Costeira, J.P.; Moura, J.M.Fcn-rlstm:, “Deep spatio-temporal neural networks for vehicle counting in city cameras”, In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 3667–3676.

[29] Dollar, P.; Wojek, C.; Schiele, B.; Perona, P., “Pedestrian Detection: An Evaluation of the State of the Art:, IEEE Trans. Pattern Anal. Mach. Intell. 2012, 34, 743–761.

[30] Xu, H.; Lv, P.; Meng, L., “A people counting system based on head-shoulder detection and tracking in surveillance video”, In Proceedings of the 2010 International Conference on Computer

Design and Applications, Qinhuangdao, China, 25–27 June 2010; Volume 1, pp. V1-394–V1-398.

[31] Subburaman, V.; Descamps, A.; Carincotte, C., "Counting People in the Crowd Using a Generic Head Detector:", In Proceedings of the 2012 9th IEEE International Conference on Advanced Video and Signal Based Surveillance, IEEE Computer Society, Beijing, China, 18–21 September 2012; pp. 470–475.

[32] Topkaya, I.S.; Erdogan, H.; Porikli, F., "Counting people by clustering person detector outputs", In Proceedings of the 2014 11th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Seoul, Korea, 26–29 August 2014; pp. 313–318.

[33] Girshick, R.B.; Donahue, J.; Darrell, T.; Malik, J. Rich, "Feature Hierarchies for Accurate Object Detection and Semantic Segmentation", In Proceedings of the 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 24–27 June 2014; pp. 580–587.