International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 09 September 2025
Page No: 3784-3792
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0549

Three-Layer Single Convolutional Neural Network Model for Estimating Crowd Density

Mr. Danayakanakere Umesha[1], Dr. Sapna.B. Kulkarni[2]
[1]Mtech 2nd Sem, Department of CSE, RYM Engineering college-RYMEC, Ballari, VTU Belagavi, India.
[2]Professor, Department of CSE, RYM Engineering College, VTU Belagavi, Karnataka, India.
Emails: *umeshad614@gmail.com[1], sapnabkulkarni@gmail.com[2]*

Abstract

The estimation of crowd density is a crucial area of study in computer vision because of its extensive use in intelligence collection, urban planning, and surveillance. The crowd density prediction that comes from a thorough analysis considers several factors, including inter-blocking in dense crowds, background elements, and individual appearance similarity. To monitor populous areas and avoid congestion, we are interested in using machine learning for crowd control in this research. We suggest using a Single Convolutional Neural Network with Three Layers (S-CNN3) model to estimate the crowd size and count the number of individuals in a scene. Next, a comparison study for density counting determines how well the suggested model performs in comparison to switched convolutional neural networks (SCNN) and convolutional neural networks with four layers (single CNN4). This study makes use of the Shanghai Tech dataset, which is regarded as the biggest database for crowd counting. With an average test accuracy of 99.88% and an average validation loss of 0.02 for crowd density estimate, the suggested model demonstrates remarkable efficacy and efficiency. These outcomes outperform the state-of-the-art models already in use.

Keywords: *Transportation and traffic management, public safety and surveillance, urban planning and smart cities, retail foot traffic, and consumer analytics.*

1. Introduction

In recent years, more people have opted to live in cities, where the benefits of this phenomenon include enhancing cultural life and effectively utilizing easily accessible urban infrastructure, drawing a diverse range of individuals to participate in a variety of activities. Both domestic and international events draw sizable crowds, both indoors and out. These events usually entail at least one activity that requires attendees to participate simultaneously, like watching a display, an outdoor performance, going through checkpoints, or entering areas with capacity restrictions. Nevertheless, these kinds of gatherings are prone to crowding because there is no quick and effective way to get a sense of the entire space and interact with people in other areas, which frequently leads to accidents and crowding. On the other hand, a centralized monitoring system that can simultaneously estimate the density of people in multiple locations is a better choice for making informed decisions that will protect attendees' safety while letting them continue to enjoy the event. There are several seasons for the Hajj, Umrah, and Ramadan in the Kingdom of Saudi Arabia (KSA), particularly on the 27th of Ramadan, when people from all over the world swarm Mecca to participate in religious ceremonies during a constrained time of year. Various nations and the security authorities are mobilized in various ways to commemorate their founding, national, or independent days. Locations to avoid the detrimental effects of crowding. In this regard, the creation of an artificial intelligence (AI)-based technique aids in the counting of people in various locations, offers the capability of managing them to prevent mishaps, and stops the spread of infectious diseases like COVID-19. A centralized, automated system that combines machine learning and crowd density estimation can be used to count, monitor, and control overcrowding issues wherever

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 09 September 2025
Page No: 3784-3792
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0549

they occur. To control large crowds in various locations, research on automatic detection, counting, and density estimation in large crowds is crucial to security and management. It takes a lot of work to oversee big events with sizable crowds to guarantee the attendees' safety and the provision of quality services. To develop an effective monitoring system, numerous studies have been carried out. The ability to estimate crowd density will help management officials plan and schedule their movements between various locations during transitions. Several CNN models have been proposed considering recent developments in machine learning to improve estimation problems and take advantage of both classification accuracy and crowd density estimation. Nevertheless, crowd size was employed as a discriminator for crowd density in those models. The structure of this paper is as follows. Background information on convolutional neural networks and crowd density estimation is provided in Section II. The pertinent state-of-the-art research on convolutional neural networks for crowd counting is reviewed in Section III. A thorough explanation of our suggested CNN model is provided in Section IV. The simulation environment and the specifics of the experiment are covered in Section V. A comparison of the outcomes between the suggested model, and a few other models, is shown in Section VI. The work contributions are displayed in Section VII. The paper is finally concluded in Section VIII, which also lists some upcoming projects [1].

2. Background

Volunteers are positioned throughout the venue to assist event organizers in keeping an eye on crowds and guiding guests. Nevertheless, without a comprehensive and current picture of the event, each person must decide for themselves when and where to lead the crowd. Uncontrolled and haphazard crowd movement can result in traffic jams, crowding, and in more severe situations, injuries from stampedes. But each person in charge of the event only keeps an eye on and plans the movements of a single, small group in a single location. For this reason, a lot of research has investigated crowd density estimation and classification as components of crowd management solutions. Crowd density counting, a branch of artificial intelligence (AI) that uses algorithms to give computers the capacity to recognize patterns in large amounts of data to generate predictions, has made use of deep learning techniques. With this approach to learning, computers can carry out certain tasks on their own. A few of the methods for estimating crowd density rely on various technologies, including neural network, computer vision, and sensor technologies [2].

2.1. Overall, Method for Determining Crowd Density

Crowd counting includes counting the number of people in a scene (image) or estimating the crowd density. Crowd Scientists used to manually count the crowd by counting how many people were in specific regions of an image and then extrapolating those numbers for estimation. This approach is characterized by a high likelihood of error and a waste of time and effort. Crowds are typical at political, religious, sporting, and festival events because they draw and assemble large numbers of people in one location. The most popular technique for counting people at rallies, protests, and religious ceremonial gatherings is the Jacobs Method. Jacobs' method involves breaking up a crowd into smaller areas, figuring out how many people are on average in each sector, and then multiplying that figure by the total number of sections. Estimating crowd density aids in the creation of management strategies like creating secure public areas and an emergency evacuation strategy. Establishing correlations between image parameters from different image processing techniques and the actual crowd densities at an investigation site is the first step in the crowd density estimation process. The two primary approaches are the direct approach and the indirect approach. If individuals are properly segmented, the direct approach simultaneously counts and tracks them. The indirect approach uses pixel-based analysis, texture-based analysis, and corner point-based analysis to perform counting and estimation processes by relating a set of measurement features to crowd learning algorithms [3].

2.2. CNN, or Convolutional Neural Network

Convolutional neural networks, also known as CNNs or Conv Nets, are a subset of artificial neural

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 09 September 2025
Page No: 3784-3792
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0549

networks that use image pixels as input to carry out tasks like face recognition, object detection, image identification, and classification. Using image pixels as input, CNN image classifications can count the number of objects in an image and output a class, such as cat, dog, or bird. An input layer, several hidden layers, and an output layer make up a CNN. A sequence of convolutional, ReLU, pooling, and fully connected layers typically make up the hidden layers. It is necessary to learn the set of filters to filter parameters in the convolutional layer. ReLU layers speed up the training process, which enhances neural networks. The pooling layer of the neural network is used to reduce processing and parameter usage. The final classification decision is made by the Fully Connected (FC) layer. CNN's structure is shown in (Figure 1).

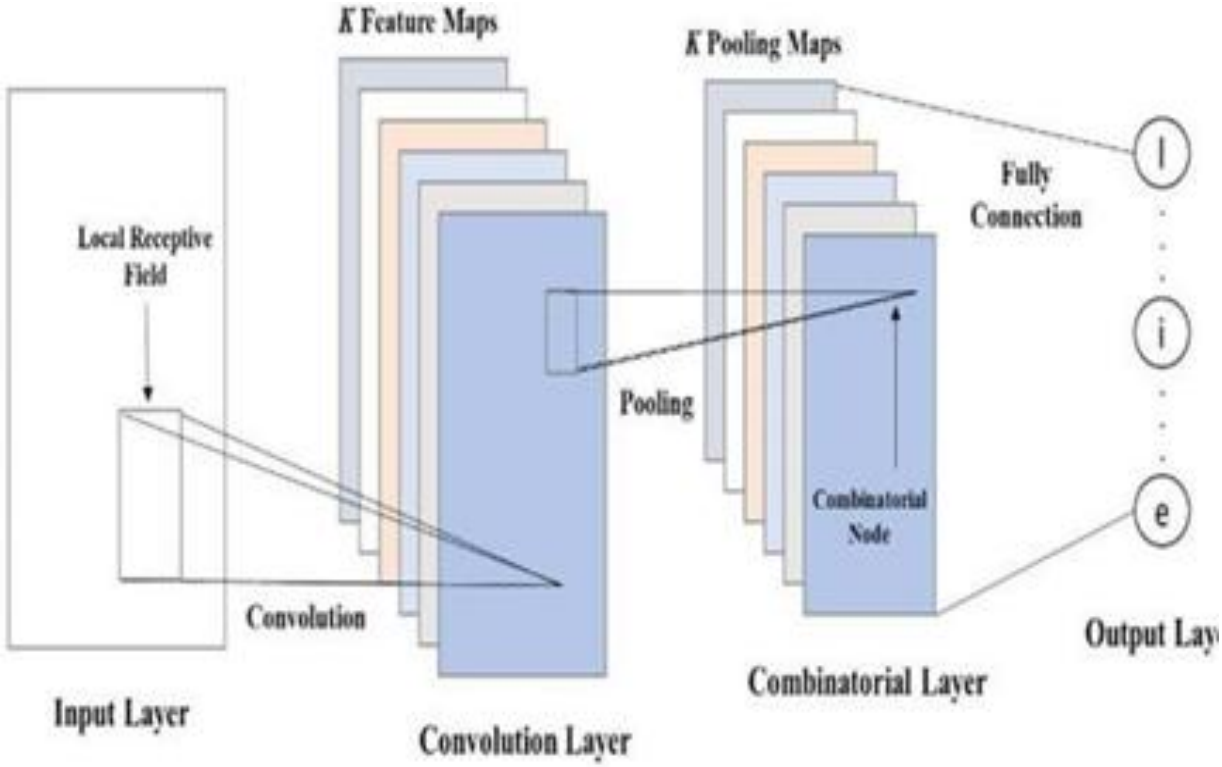

Figure 1 Shows the Convolutional Neural Networks (CNN) Structure

2.3. Additional Methods for Crowd Counting.

2.3.1. Solutions Based on Sensors.

The issue of crowd estimation has been the subject of numerous studies employing sensor technologies like radio frequency identification (RFID) and wireless sensor networks (WSN). The premise behind sensor technology approaches is that people in the crowd have network devices (such as smart phones, RFID wristbands, or sensor tags) that can send wireless signals. By counting the Unique Identifiers (UIDs) of the signals detected or read in the region of interest, some sensor-based technologies calculate the density of the crowd. Other approaches linked to the Link Quality Indicator (LQI), Channel State Information (CSI), and Received Signal Strength Indicator (RSSI) [4].

2.3.2. Solutions For Computer Vision

On the other hand, computer vision solutions don't rely on crowd members' cooperation and require fewer devices. All that is required to take pictures of the crowded scene is a camera that has been placed there. Following processing of those photos to extract relevant data about the degree of crowd density.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 09 September 2025
Page No: 3784-3792
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0549

3. Review Of Literature

Table 1 Deep Learning Methods for Crowd Counting: Key Insight

Ref No	Authors	Title	Key Learning
1	I. Zhang, L. Zhao, D.Tao	Using Convolutional Neural Networks to Count Crowds	Explains CNNs for density map estimation and emphasizes how crucial end-to-end learning is for managing occlusions and crowd fluctuations.
2	S. Idrees, J. Zhang, Y. Zhu, M. Shah	Deep Crowd: A Deep Learning Method for Counting Crowded Scenes	The Deep Crowd, a deep CNN that focuses on learning spatial interactions and produces high-resolution density maps for crowd counts.
3	J. Li, L. Zhang, H. Zhang, X. Li	Dilated Convolutional Neural Networks for Scenes with High Congestion (CSRNet)	Presents CSRNet, the state-of-the-art in crowd benchmarking, which uses dilated convolutions to record wide receptive fields for crowded surroundings.
4	J.Zhang. S. Zheng, X. Xu, S. Yang	MCNN: Convolutional Neural Networks with Multiple Columns for Crowd Counting	Presents MCNN, a multi-column CNN that can handle various crowd sizes and performs well in scenes with variable densities
5	X. Zhang, X. Zhao, H. Zhang	CNN with Density Awareness for Counting Crowds	SFCN: An Easy and Quick Crowd Counting System handles significant density changes and occlusions by introducing a density-aware CNN that is suited to different crowd densities.
6	Y.Chen, J. Yang, X. Zhang	Using Deep Convolutional Neural Networks for Crowd Counting	Focuses on multi-scale characteristics and data augmentation for better generalization when discussing the use of deep CNNs for crowd counting.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 09 September 2025
Page No: 3784-3792
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0549

4. Proposed Approach

This work's primary objective is to create a model that can learn and analyze crowd features using deep learning, after which it will be able to predict the crowd density class. In a busy setting, this work enables appropriate decision-making for incident prevention and efficient management. The goal of this study is to count crowds automatically without the need for human intervention so that we can handle them more quickly and professionally. To do this, we used a dataset of photos with varying numbers of people. The mod counts the heads in a particular area where people are congested. Since avoiding crowds and keeping a safe distance are crucial right now to stop the spread of the Covid-19 virus, this aids security officials in breaking up the crowd to prevent accidents [5].

4.1. Preprocessing

During the preprocessing stage, new images are added to the dataset using the segmentation method and the available data, thereby expanding its size. The image is segmented into nine non-overlapping sub- images, and the density maps for each are then determined. On the one hand, the density map's sum indicates how many heads are in that map. On the other hand, the distribution of the total count across the image segments must be understood. Consequently, the entire image is first used to create the dot-map, which is subsequently segmented. The headcount and head locations in the corresponding image segment are then represented by the sum and placement of 1s in a dot-map segment. Labeling also requires a complete dot- map. Figure 2 illustrates the segmentation technique [6].

Figure 2 Shows an Illustration of Image Segmentation.

4.2. Extraction Of Features and Labeling

This stage involves extracting the image's primary feature, which is the quantity of heads. The number of heads in each image is then used to label them all. After that, these pictures are split up into 20 and 33 labels, or classes, which are then used in two distinct experiments, respectively.

4.3. Details of the Proposed Model Architecture:

To count the number of people in a scene, we suggest using a Single Convolution Neural Network with three convolution layers (Single-CNN3). First, multiple classes with varying crowd count ranges are developed to shift the problem solving from counting to density estimation. These ranges are chosen to define a particular density level. These levels denote different indicators, with high levels signifying the possibility of traffic jams Shown in Figure 3 [7].

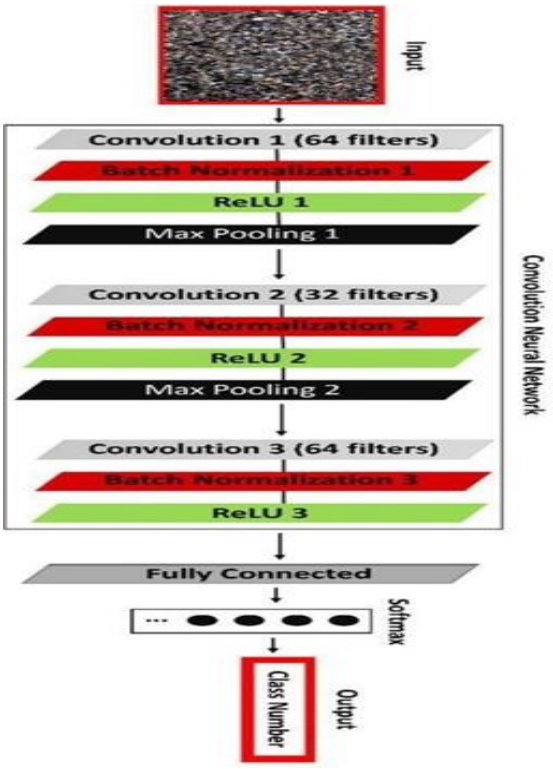

Figure 3 Illustrations from the Shanghai Tech Collection

To solve the crowd density estimation problem as a classification problem rather than a regression problem, we secondly constructed the Single-CNN3

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 09 September 2025
Page No: 3784-3792
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0549

model with a straightforward structure. It is composed of a single CNN regressor with three filter size convolution layers. While the second layer uses 32 filters, the first and third layers use 64 filters. The suggested architecture additionally comprises an output layer of n neurons representing the n density classes, two max- pooling layers, Shown in Figure 4.

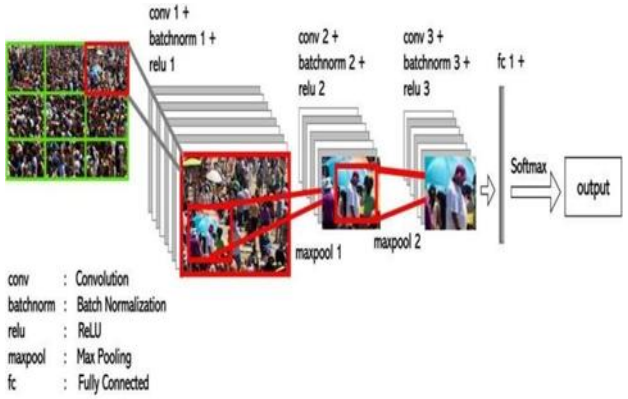

Figure 4 The Suggested Model's Architecture

Three batch normalization layers, three rely on layers, one fully connected layer, and one SoftMax layer in addition to the input and output layers. A Gaussian distribution with a learning rate of 0.0001 is used to initialize the model Shown in [8].

Figure 5 Three-Layer Single Convolutional Neural Network (Single-CNN3) Model

5. Implementation

5.1. Equipment and Tools

A MATLAB program was used to run this suggested model with training parameters. Our training parameters are listed in Table 1. An MSI laptop with a Core (TM) i7- 9750H CPU running at 2.60 GHz, 16 GB of RAM, and an NVIDIA GeForce RTX 2060 is used as the simulation tool Shown in Figure 6 & 7.

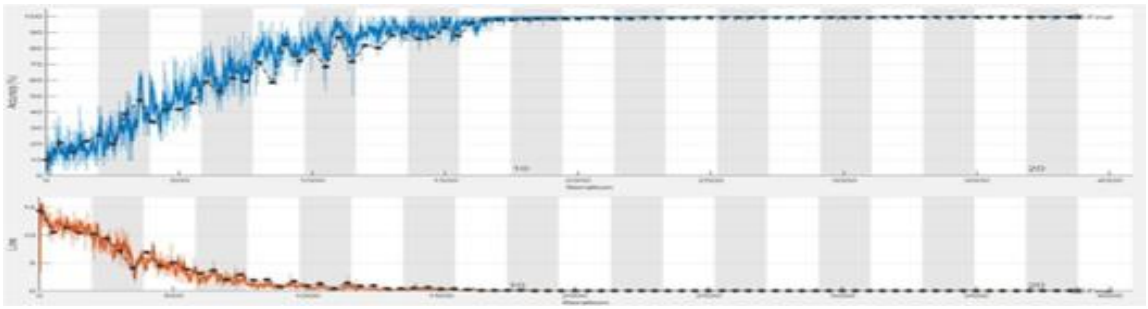

Figure 6 Shows Accuracy and Loss Progress Over the First Phase of Training

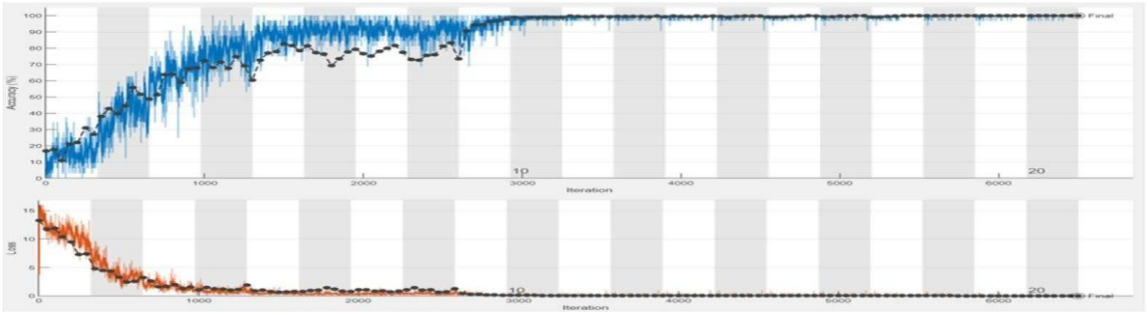

Figure 7 Shows the Accuracy and Loss Progress Over the Second Phase of Training

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 09 September 2025
Page No: 3784-3792
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0549

5.2. The Dataset of Shanghai Tech

Shanghai Tech is a brand-new, extensive dataset for crowd density estimation that contains 1198 annotated photos of 330,165 people with annotations on the center of their heads. In terms of the total number of annotated individuals, this dataset is the biggest. There are two sections to the dataset. The 482 photos in the first section were gathered at random from the Internet, Arabian nations, and Makkah. 716 photos from Shanghai's bustling city streets are included in the second section. Both sections are separated into groups for testing and training. Compared to most existing datasets, accurate crowd estimation is more difficult because of the significant differences in crowd density between the two subsets. Samples of the dataset are displayed in Figure 5 [9].

5.3. Metrics Of Performance

Our suggested model's performance is assessed using several metrics that are computed over several experiments. In this work, three performance metrics are used.

5.3.1. Accuracy Of Classification

The accuracy can be defined as the proportion of accurate predictions to all inputs entered. Equation 1 provides it.

$$\text{Precision} = \frac{\text{Quantity of accurate forecasts}}{\text{Total number of forecasts}}$$

5.3.2. Loss Rate

The model's prediction accuracy on a single example is gauged by the loss rate metric. The loss is zero if the model's prediction is accurate; if not, it is larger. Finding a set of weights and biases that, on average, have low loss across all examples is the aim of model training. In contrast to accuracy, the loss is determined by the model's performance on training and validation sets. The loss is not a percentage as a result. Rather, it is a total of the mistakes made for every instance in the training or validation sets.

5.3.3. Matrix of Confusion

To assess, visualize, and summarize a classification model's performance, we employed a confusion matrix. Additionally, being aware of our model's actual and predicted classifications.

6. Performance Evaluation and Experiments

6.1. Results And Experiment

The Shanghai Tech dataset is subjected to Single-CNN3. This dataset, which includes labels used by the neural network to train a model and provide context, is regarded as a crowd density estimation dataset. It includes a lot of pictures with different crowd densities, which makes it possible to learn about the variety of congestion situations effectively. To expedite the training process, multiple GPUs are used in parallel to train the model. The datasets A and B are combined and divided into train and test subsets at random by 80% and 20%, respectively. A specific range of head counts is included in the number of classes created in this work. A particular degree of crowdedness is represented by each class. For a total of 20 and 33 classes, numerous experiments are conducted. It goes without saying that many classes significantly increase the classification problem. A specific crowd level is represented by the range of head counts included in each class. As a result, this step shifts the focus of our problem solving from regression to classification with a large class size [10].

6.2. Evaluation Of Performance

Using the Shanghai Tech dataset, the performance evaluation compares the outcomes and other features of our suggested Single-CNN3 model to those of the Single-CNN4 and Switch-CNN models. The outcomes of the three models are displayed in Table 4 along with the number of classes, validation accuracy, validation loss, and test accuracy. Using 20 labels, the single-CNN3 model produced the best results in terms of test and validation accuracy. In terms of validation accuracy, the Single-CNN4 model outperformed the Single-CNN3 model with 33 labels, while the Single-CNN3 model produced the best test accuracy results.

7. Contributions to Work

The following is a summary of the research's most significant contributions.

- It helps create a system that can automatically estimate the number of people in record time during certain events, like religious ceremonies or pilgrimages. In addition, it saves money, time, and effort when

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 09 September 2025
Page No: 3784-3792
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0549

compared to conventional counting and crowd control methods.

- By defining distinct classes that estimate different density levels, the density estimation challenge has been transformed from a regression problem to a classification problem.
- The most well-known Shanghai Tech dataset that reflects crowded events has been used to evaluate performance.

This system's intended use includes assisting security personnel in managing and arranging crowds in tourist and leisure areas as well as scientific locations like stadiums, universities, commercial districts, and beaches. Secure Covid-19 vaccine supply supervision in the available centers can be aided by the system [11].

Conclusion

In this work, we examined methods based on convolutional neural networks that are intended to precisely estimate the degree of crowd density in various settings. Deep learning has recently caught the attention of the industry and research community for a variety of image classification and speech recognition applications. The Shanghai Tech dataset, which is a sizable dataset in terms of the annotated heads for crowd counting, uses our suggested CNN with three layers. The method's outcomes have demonstrated a low loss rate and high accuracy up to 100%. The accuracy and loss metrics of our suggested model and the switched convolutional neural networks have then been compared in a comparative analysis. This model is developed on three layers CNN, using very large and recognized crowd counting dataset and evaluated compared to the existing state-of-the-art models. To sum up, our method helps create a system that can automatically estimate the number of people during pilgrimages and religious ceremonies. This system counts people more quickly and efficiently and assists security in dispersing and dismantling crowds for everyone's safety. Additionally, in the centers that are available, the system helps with Secure Covid-19 Vaccine Supply Supervision. Our method can be used in a broader context, encompassing different databases, as future research. Additionally, the model can be used for crowd density estimation and control by the relevant authorities in both the public and private sectors.

References

[1]. Saleh, S. A. M., Suandi, S. A., & Ibrahim, H. (2015). Recent survey on crowd density estimation and counting for visual surveillance. Engineering Applications of Artificial Intelligence, 41, 103-114.

[2]. Addanki, S. C., & Venkataraman, H. (2017). Greening the economy: A review of urban sustainability measures for developing new cities. Sustainable cities and society, 32, 1- 8.

[3]. Conway, D. G. (2014). The Event Manager's Bible 3rd Edition: The Complete Guide to Planning and Organizing a Voluntary or Public Event. Hachette UK.

[4]. Li, B., Huang, H., Zhang, A., Liu, P., & Liu, C. (2021). Approaches on crowd counting and density estimation: a review. Pattern Analysis and Applications, 24, 853- 874.

[5]. Alashban, A., Alsadan, A., Alhussainan, N. F., & Ouni, R. (2022). Single convolutional neural network with three layers model for crowd density estimation. IEEE Access, 10, 63823-63833.

[6]. Bhardwaj, S., Dwivedi, A., Pandey, A., Perwej, D. Y., & Khan, P. R. (2023). Machine Learning-Based Crowd Behavior Analysis and Forecasting. International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN, 2456-3307.

[7]. Lin, Y., Lv, F., Zhu, S., Yang, M., Cour, T., Yu, K., ... & Huang, T. (2011, June). Large-scale image classification: Fast feature extraction and SVM training. In CVPR 2011 (pp. 1689-1696). IEEE.

[8]. Alashban, A., Alsadan, A., Alhussainan, N. F., & Ouni, R. (2022). Single convolutional neural network with three layers model for crowd density estimation. IEEE Access, 10, 63823-63833.

[9]. Gleckler, P. J., Taylor, K. E., & Doutriaux, C. (2008). Performance metrics for climate models. Journal of Geophysical Research:

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 09 September 2025
Page No: 3784-3792
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0549

Atmospheres, 113(D6).

[10]. Hernández-Orallo, J., Flach, P., & Ferri, C. (2012). A unified view of performance metrics: Translating threshold choice into expected classification loss. The Journal of Machine Learning Research, 13(1), 2813-2869.

[11]. Fan, Z., Zhang, H., Zhang, Z., Lu, G., Zhang, Y., & Wang, Y. (2022). A survey of crowd counting and density estimation based on convolutional neural network. Neurocomputing, 472, 224-251.