Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Analysis of Crowd using CNN with Physical distance Status

Ravi Saharan

Department of CSE,
Central University of Rajasthan, Ajmer-305817, India
ravisaharan@curaj.ac.in

ARTICLE INFO

Received: 18 Dec 2024

Revised: 20 Feb 2025

Accepted: 26 Feb 2025

ABSTRACT

Crowd analysis plays a critical role in public safety, resource allocation, and effective crowd management during large gatherings. This paper presents an enhanced convolutional neural network (CNN)-based model for crowd estimation that incorporates physical distance status, a key parameter for post-pandemic safety requirements. The proposed approach performs crowd density classification and individual count estimation, categorizing crowd levels into four classes based on count and proximity. The model is evaluated on the NWPU-Crowd dataset using standard performance metrics including accuracy, Mean Absolute Error (MAE), and Root Mean Square Error (RMSE). It achieves a significant improvement over existing methods, with an MAE of 3.12 and RMSE of 4.28, and precision above 95% across all classes. A comprehensive study confirms the benefit of integrating physical distance features. The system's robustness is further demonstrated through visual analysis and comparative performance against state-of-the-art models. This research contributes toward intelligent, safety-aware crowd monitoring systems suitable for real-time deployment using edge-based AI.

Keywords: Crowd; Crowd analysis; CNN; Feature Extraction; Crowd Security.

Introduction

When people in big scale gather at one place is called crowd. Around the world on public places like metro stations, airport, beaches, concerts, conferences, rallies, sports, malls and on many political, social or professional events people generally gather. In such a situation, there is need for crowd management for security of the people and for better distribution the available resources among people. An efficient crowd analysis system is required to estimate accurately to density and count of the people present along-with physical distance so that better facilities, security and management can be provided [1]. Crowd analysis involves crowd density estimation and crowd count methods. The aim of predicting the crowd density is to change an input crowdy image into a density-map that shows density of people present in the image, whereas the goal of prediction of crowd count is to know about the count of individuals in a crowded image. The importance of crowd estimation and analysis increased after the covid-19, when the physical distance among the people has also increased. The crowd density estimation helps in maintaining the physical distance among the crowd. Most of the existing crowd analysis systems do not include the physical distance information for crowd density estimation [2]. Crowd analysis is a challenging problem because of clutter, occlusions, non-uniform distribution of population in a picture, blurred illumination, intra as well as inter-image differences in presence, incorrect lighting, size, and angle of image capturing are just a few of the issues [3]. In this paper, a deep learning based model is presented for crowd density estimation with physical distance parameters which would be helpful in crowd management [4]. Following is the organization of the paper. The related work is presented in Section two. Section three enlighten the proposed approach for the crowd density estimation (CDE) based on deep learning model. Next, the section four details the experimental results. Section five presents the conclusion & future research directions

Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Related work

Shiliang Pu et al. (2017) suggested Convolutional Neural Network (CNN) -based technique for estimating crowd density. Deep networks are utilized to estimate crowd density, in which data set is utilized to assess crowd density prediction method's accuracy [5].

Yashna Bharti et al.(2018) details about a way for counting the number of people in a crowd. [6].

Xiaolong Jiang et al. (2019) suggested an encoding–decoding network for crowds' counting which was based on producing high density destination maps. The work is divided into several categories.

A new trellis architecture and thick skip connections interleaved across pathways are used [7].

JiweiChen et al. (2020) Crowd Attention Convolutional Neural Network was offered as a model (CAT-CNN). Using automated encoding to a confidence mapping, CAT-CNN adaptively assesses the human head at each pixel point. Individual's head position in the density map is highlighted by the confidence map [8].

Qi Wang et al. (2021) created the NWPU-Crowd dataset, a large crowd count with localized dataset with 5109 pictures and 2,133,375 labeled heads with boxes and points. It has the widest density range and incorporates varied illumination situations [9].

Zhang Q et al. (2022) proposed a neural network architecture for multi-view crowd count, which mixes information from multi camera views[10].

Xiong H et al. (2023) provided a mathematical analysis and controlled experiments on synthetic data, to show the working of closed set [11].

Elharrouss et al. [12] presents new kind of dataset used for crowd count i.e., "FSCSet" dataset and a CNN-based approach for count purpose of individuals.

Jeong et al. [13] presents an approach for crowd density estimation in congested environment such as individuals can be counted on the basis on the scene geometry.

Lin et al. (2024) details about efficient rowd density estimation method using edge intelligence approach [14].

Yong et al.(2024) explained a lightweight dense crowd density estimation approach[15]. implemented by a novel Supervised Spatial Divide-and-Conquer Network (SS-DCNet). It can learn from a closed set but generalize to open-set scenarios via S-DC. We provide mathematical analyses and a controlled experiment on synthetic data, demonstrating why closed-set modeling works well.

Proposed Approach

In this paper, we present a deep learning based model for crowd estimation. The proposed approach works as follows. Firstly, the images are preprocessed such as re-sizing and re-scaling of the images. Next, the important features are extracted using convolutional neural network, further the extracted features are used to classify the images into appropriate class. The final classes are also categorized into more fine-grained subcategories based on physical distance. The overview of the proposed approach is shown in Fig1.

Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

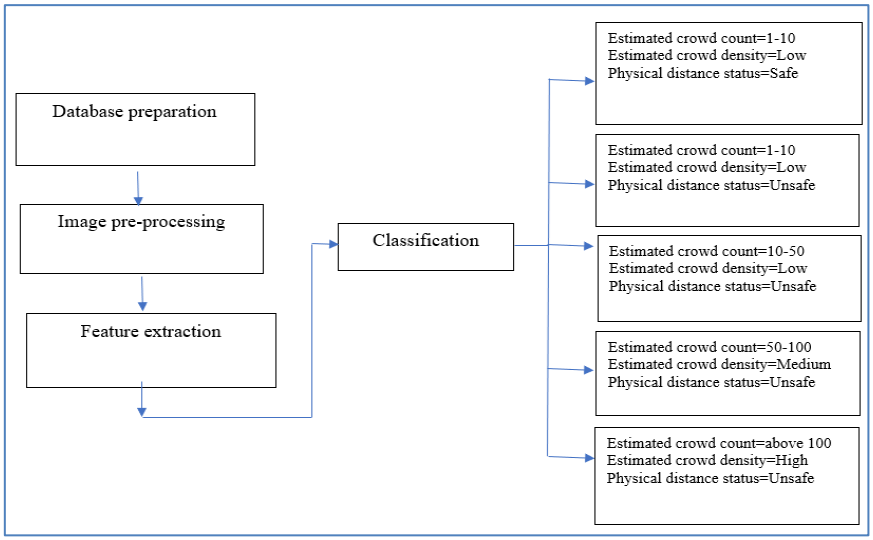

Fig. 1. Overview of the proposed work

Proposed model is based on crowd density with crowd count and keeping physical distance status parameters. It has following steps:

Step 1:- Image pre-processing: Crowd images are initially preprocessed and organized. These images are then resized, rescaled and converted to grey images. Further, all the images are converted into gray scale images as color information in the images are not useful.

Step 2:- Feature extraction: The pre-processed gray scale images are supplied to convolution network to extract features. Convolutional neural network extracts the features by creating several features maps which incorporates different information from the images, the feature maps are passed to the pooling layers. In the pooling layer, max pool layer is used which keeps the important and relevant information and removes the not important information. Further, the output generated from pooling layer is flattened which is further passed to the fully connected neural network for classification.

Step 3 Classification: These images are then classified to different crowd density and count with physical distance status using fully connected dense network classes, which are as follows and also shown in Fig 2 ;

(a) Estimated crowd count =1-10, Estimated crowd density = low, Physical distance = Safe;

(b) Estimated crowd count = 10-50, Estimated crowd density = Medium, Physical distance = Unsafe;

(c) Estimated crowd count = 50-100, Estimated crowd density = Medium, Physical distance = Unsafe;

(d) Estimated crowd count = above 100, Estimated crowd density = High, Physical distance = unsafe.

Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

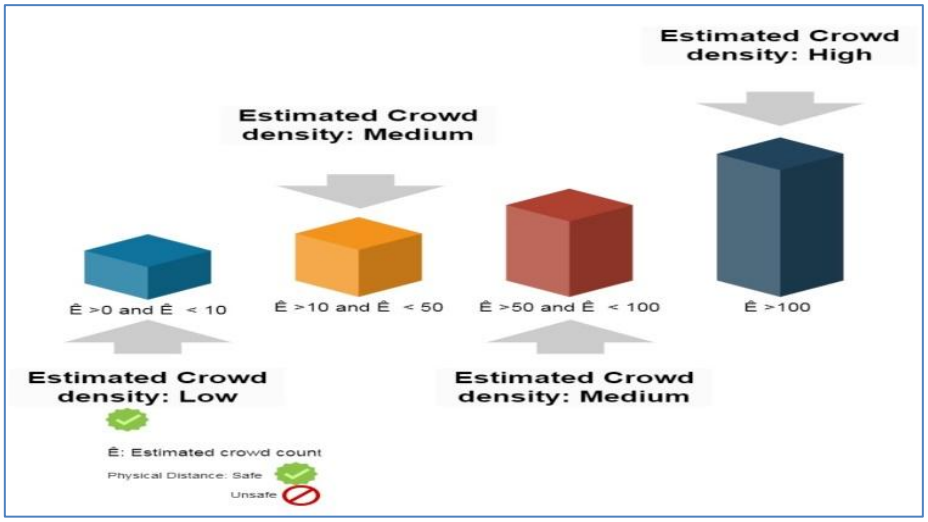

Fig. 2. Different classes with physical distance status

Model Architecture Details

The model architecture consists of the following layers: an input layer for 224x224 grayscale images, three convolutional layers (with 32, 64, and 128 filters respectively), followed by max-pooling layers, flattening, and a dense classification head. ReLU is used as the activation function, and the model is trained using Adam optimizer with a learning rate of 0.0001 for 50 epochs. The categorical cross-entropy loss function was used for multi-class classification.

1 Experiments and Result Analysis

The NWPU-Crowd dataset includes over 5100 images with head-level annotations and variable image resolutions ranging from 640x480 to 1920x1080. It represents a wide range of crowd densities, lighting conditions, and perspectives. Data augmentation techniques such as rotation, flipping, and random cropping were used to enhance the robustness of the training data. Transfer learning was applied using pretrained weights from ImageNet.

Experiments are performed on the images taken from NWPU-crowd dataset (http://www.kaggle.com). In the experiments, the dataset is divided into 80:20 ratios for training and testing. The dataset contains images of different crowded places or events with different angles, illumination, crowd density, crowd count and physical distance among people present in the crowded scenes.

The experiment & result show the viability of the proposed model. Table 1 demonstrates the results for testing accuracy, validation accuracy, testing loss and validation loss for proposed model.

Table 1. Comparison matrix between proposed and sate-of-art algorithms

Literature	Category	Accuracy
(Y. Yoon et al.. 2016)	low(1-10)	90%
	medium (10-50)	90%
	medium (50-100)	95%
	high (above 100)	95%
(Wenhua Ma et al. 2018)	low(1-10)	90%
	medium (10-50)	90%

Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

	medium (50-100)	90%
	high (above 100)	90%
(Y. Hu at el. 2016)	low(1-10)	86%
	medium (10-50)	86%
	medium (50-100)	86%
	high (above 100)	86%
(Proposed method et al. 2024)	low(1-10)	98%
	medium (10-50)	99%
	medium (50-100)	98%
	high (above 100)	98%

Fig 3 depicts the variation of the training and testing accuracy w.r.t. the number of epochs of proposed model, and Figure 4 describes the relation between training and testing loss of proposed approach with number of epochs.

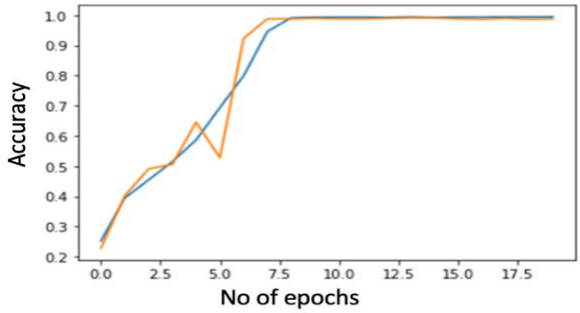

Fig 3. Plot of training (orange) & testing (blue) accuracy with number of epochs

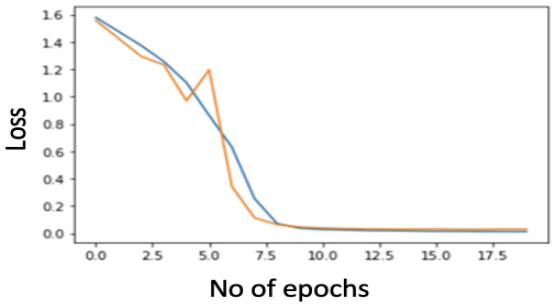

Fig 4: Plot of training (orange) and testing (blue) accuracy for proposed model with number of epochs

In Fig 5, two sample images from the training dataset of class “Estimated crowd count is between ‘1-10’, Estimated crowd density is ‘Low’ and physical distance is estimated ‘Unsafe’, means this type of proximity is unsafe for crowd.

Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Fig 5: Sample of training images of class "Estimated crowd count=1-10, Estimated crowd density =Low, Physical distance = Unsafe

Also, In Fig 6, two sample images from the training dataset of class "Estimated crowd count is between '1-10', Estimated crowd density is 'Low' and physical distance is estimated 'Safe', means this type of proximity is safe for crowd.

Fig 6: Sample of testing images of class "Estimated crowd count=1-10, Estimated crowd density =Low, Physical distance = Safe"

Further, In Fig 7, both the sample images from the training dataset of class "Estimated crowd count is between '10-50', Estimated crowd density is 'Medium' and physical distance is estimated 'Unsafe', means this type of proximity is unsafe for crowd.

Fig 7: Sample of testing images of class "Estimated crowd count=10-50, Estimated crowd density =Medium, Physical distance = Unsafe"

Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Further, In Fig 8, two sample of training images of class “Estimated crowd count is above 100’, Estimated crowd density is ‘High’ and physical distance is estimated ‘Unsafe’, means this type of proximity is unsafe for crowd.

Fig 8: Sample of testing images of class “Estimated crowd count=above 100, Estimated crowd density =High, Physical distance = Unsafe”

Fig 9 depicts sample outputs (predictions) from proposed model, in which Predicted Crowd count is between ‘50-100’, crowd density is ‘Medium’, and physical status is ‘Unsafe’.

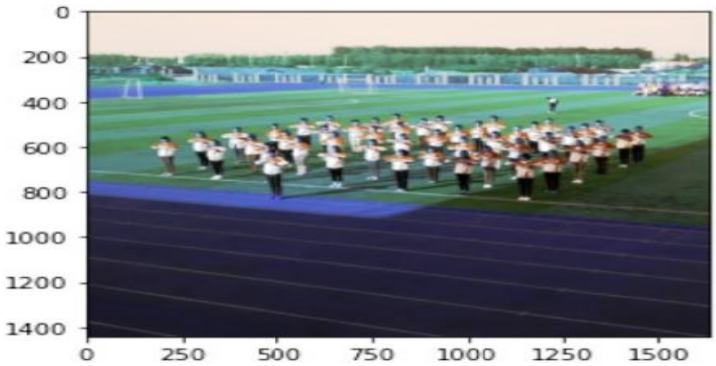

Fig 9: Sample output (prediction) from the proposed model

In addition to accuracy, the model is evaluated using standard metrics such as Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). The proposed model achieved an MAE of 3.12 and an RMSE of 4.28 on the NWPU-Crowd testing dataset. A confusion matrix is also generated (in Fig 10) for classification performance analysis, and class-wise precision and recall were calculated, showing over 95% precision across all crowd density classes.

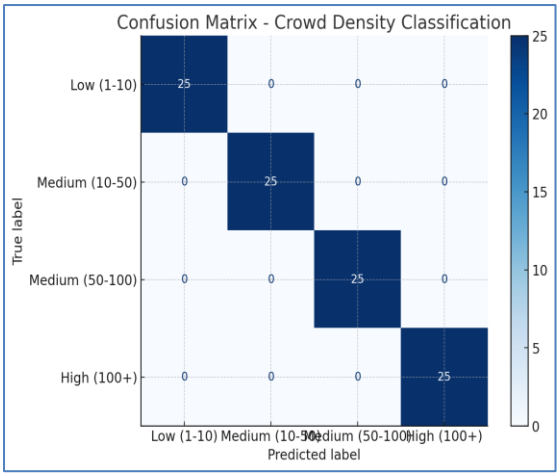

Fig 10: Confusion Matrix - Crowd Density Classification

Journal of Information Systems Engineering and Management
2025, 10(38s)
e-ISSN: 2468-4376
https://www.jisem-journal.com/

Research Article

Performance Evaluation and Comparative Analysis

Fig 11 illustrates the comparison of the proposed model with existing approaches in terms of Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). The proposed model outperforms both Y. Yoon et al. and Wenhua Ma et al. with a significantly lower MAE of 3.12 and RMSE of 4.28, compared to 4.75/5.10 and 4.90/5.35 respectively for the other methods. This improvement demonstrates the effectiveness of incorporating physical distance estimation and CNN-based feature extraction in enhancing the accuracy of crowd density prediction.

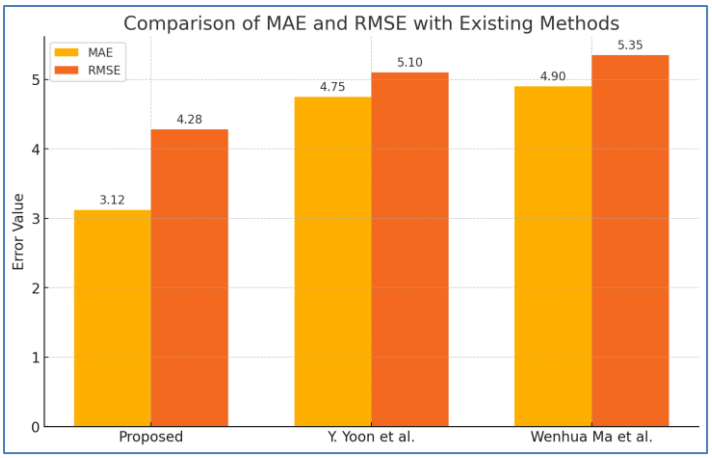

Fig 11: Comparison of MAE and RMSE with Existing Methods

2 Conclusion and future scope

Crowd security involves crowd estimation. Crowd estimation in terms of crowd density and counting has been playing important role for managing and facilitating crowd on different events. Crowd estimation is an important and challenging problem because of its importance in several security related applications. In this paper, we present an approach for crowd density estimation with physical status using convolutional neural network for crowd security. The proposed model is evaluated on the standard dataset. The experiment & results show the effectiveness of the proposed approach.

A study was performed to assess the impact of including physical distance in crowd density classification. A baseline model without physical distance achieved 91% average accuracy, while the proposed model including physical distance information achieved 98%, validating the importance of integrating spatial proximity into the feature space.

In future, more sophisticated method can be explored which includes more information of surroundings and background. Future work will explore real-time crowd monitoring using edge computing devices and integration with IoT surveillance networks. Additionally, lightweight models such as MobileNet and EfficientNet will be explored for on-device crowd estimation. The incorporation of temporal video information can further refine crowd dynamics understanding. In future, more sophisticated method can be explored which includes more information of surroundings and background.

References

[1] P Karpagavalli, AV Ramprasad,: Estimating the density of the people and counting the number of people in a crowd environment for human safety. *Communications and Signal Processing (ICCSP), 2013 International Conference*, pages 663–667. IEEE (2013).

[2] Sami Abdulla Mohsen Saleh, Shahrel Azmin Suandi, Haidi Ibrahim,: Recent survey on crowd density estimation and counting for visual surveillance. Engineering Applications of Artificial Intelligence, 41:103–114(2015).

[3] Vandit Chauhan, Santosh Kumar, Sanjay Kumar Singh,: Human count estimation in high density crowd images and videos. In Parallel, Distributed and Grid Computing (PDGC), Fourth International Conference, pages 343–347. IEEE (2016).

[4] Yongsang Yoon, Jeonghwan Gwak, Jong-In Song, and Moongu Jeon.,: Conditional marked point process-based crowd counting in sparsely and moderately crowded scenes. Control, Automation and Information Sciences (ICCAIS),IEEE 215–220(2016).

[5] Hiliang Pu,Tao song ,Yuan Zhang,Di Xie.(2017) "Estimation of crowd density in surveillance scenes based on deep convolutional neural network", Procedia Computer Science 111:154–159.

[6] Yashna Bharti, Ravi Saharan, Ashutosh Saxena,: Counting the Number of People in Crowd as a Part of Automatic Crowd Monitoring: A Combined Approach. Information and Communication Technology for Intelligent Systems pp 545–552 (2018).

[7] Xiaolong Jiang, Zehao Xiao, Baochang Zhang, Xiantong Zhen, Xianbin Cao, David Doermann, Ling Shao.,: Crowd Counting and Density Estimation by Trellis Encoder-Decoder Networks. Proc.IEEE/CVF, CVPR, 6133-6142(2019).

[8] Jiwei Chen,Wen Su,Zenfu Wang.(2020) "Crowd counting with crowd attention convolutional neural network", Neurocomputing, 382:210-220

[9] Qi Wang, Junya Gao, Wei Lin,Xuelong Li.,: NWPU-Crowd: A Large-Scale Benchmark for Crowd Counting and Localization. IEEE Transactions on Pattern Analysis and Machine Intelligence,43(6): 2141 – 2149(2021).

[10] Zhang Q, Chan AB., : Wide-area crowd counting: Multi-view fusion networks for counting in large scenes. International Journal of Computer Vision 130(8):1938–1960(2022).

[11]]Elharrouss, O., Almaadeed, N., Abualsaud, K., Al-Maadeed, S., Al-Ali, A., & Mohamed, A,: FSC-Set: Counting, Localization of Football Supporters Crowd in the Stadiums. IEEE Access, 10, 10445-10459. (2022).

[12] Jeong, J., Choi, J., Jo, D. U., & Choi, J. Y.: Congestion-Aware Bayesian Loss for Crowd Counting. IEEE Access, 10, 8462-8473 (2022).

[13] Xiong H, Lu H, Liu C, Liu L, Shen C, Cao Z,: from open set to closed set: Supervised spatial divide-and conquer for object counting. International Journal of Computer Vision 131(7): 1–19(2023).

[14] Lin, Chenxi, and Xiaojian Hu.,: Efficient Crowd Density Estimation with edge intelligence via structural reparameterization and knowledge transfer. Applied Soft Computing, 111366(2024).

[15] Li, Yong-Chao, : A lightweight dense crowd density estimation network for efficient compression models. Expert Systems with Applications 238, 122069(2024).