

Video-Based Crowd Analytics Using Deep Learning Methods for Density Estimation

Deepadharshana M¹, Vijayarani S², 

1. *Ph.D Research Scholar, Department Computer Science, Bharathiar University, Coimbatore, IND*

2. *Associate Professor, Department of Computer Science, Bharathiar University, Coimbatore, IND*

Received: April 24, 2026 | Review began: May 12, 2026 | Review ended: May 24, 2026 | Published: May 29, 2026

© Copyright 2026

This is an open access article distributed under the terms of the Creative Commons Attribution License CC-BY 4.0., which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Abstract

Crowd analytics with video-based analytics is a major research area, and crowds are being monitored automatically for surveillance and public safety. The proposed work proposes a frame-based crowd density estimation model that has the ability to work on both image and video inputs with four different machine learning-based density estimation models. The input videos are extracted as frames, and images and frames are pre-processed into grayscale, normalization, and noise removal. The work proposes a crowd density estimator, namely the Gradient-Texture Crowd Density Estimator, Multi-Scale Crowd Density Estimator, Context-Aware Smooth Crowd Density Estimator, and Global-Context Crowd Density Estimator, to compute frame-wise crowd density maps and density scores in the proposed framework. The density values estimated for each frame are aggregated to analyze the progressive crowd behavior. A comparative evaluation is presented based on the metrics such as Mean Absolute Error, Mean Squared Error, density entropy, prediction variance, runtime, as well as density level and density maps. The experimental results underline that the best overall performance is achieved by the Gradient-Texture Crowd Density Estimator, since this method offers stable density estimation with the least computational overhead. The proposed framework provides an efficient, scalable, and real-time performance solution for crowd density estimation in video analytics and smart surveillance systems.

Categories: Big Data and Analytics, Data Mining, Deep Learning

Keywords: crowd analytics, deep learning, density estimation, video analytics, neural network, artificial intelligence

Introduction

Due to the increase in the rate of urbanization and the number of events where a significant number of people come together, the need arises for intelligent crowd surveillance systems. Crowd analytics, a subfield of video analytics, plays a critical role in the safety of the crowd, crowd management, and overall decision-making in areas such as airports, malls, sports stadiums, and cities. Normally, the conventional way of crowd surveillance is inefficient, error-prone, and cannot be applied on a massive scale. This brings into focus the need for automatic and vision-based crowd analytics surveillance systems [1].

It is important to note that crowd density estimation is an imperative dimension of crowd analytics, providing numerical information with respect to the number of individuals as well as their density in the given images. It is essential for the early detection of crowded locations, anomalies, and possible dangerous situations. Although more recent achievements have contributed greatly to the increased accuracy of crowd counting, the existing recent-state solutions prove complex, requiring an enormous number of images from the training database, as well as an enormous number of calculations [2].

How to cite this article:

M D, S V (May 29, 2026) Video-Based Crowd Analytics Using Deep Learning Methods for Density Estimation. Cureus J Comput Sci 3 : es44389-026-00091-1. DOI <https://doi.org/10.7759/s44389-026-00091-1>

This work responds to the issues described by proposing a frame-level crowd density estimation approach using a frame-based crowd density estimation system for the processing of images and videos using simplified density estimation models. Videos are broken down into distinct frames. Further processing of images and frames is conducted using the grayscale and normalization methods for the reduction of complexity. Four different models for estimating crowd density have been proposed in this work and have been used in the current work for estimating the crowd density scores for the videos using the models named the Gradient-Texture Crowd Density Estimator, the Multi-Scale Crowd Density Estimator, the Context-Aware Smooth Crowd Density Estimator, and the Global-Context Crowd Density Estimator [3].

A comparative assessment is carried out in terms of the average density value, density variance, and the computation time in every frame, in order to determine the best algorithm. From the experimental outcome, it is noted that the Gradient-Texture Crowd Density Estimator performs better while consuming low amounts of computation time. A scalable solution for the real-time crowd density estimation technique of the smart surveillance system is presented in [4]. The remaining portion of this work contains a Literature review section, methodology in "Materials and Methods" section, experimental results and future discussion in the "Results and Discussion" section and Conclusions.

Literature review

Gao et al. [1] presented an in-depth review of crowd counting/density estimation via the application of deep learning approaches, dividing the techniques of crowd counting into detection-based approaches, regression-based approaches, and density maps approaches.

Patwal et al. [2] examined recent deep learning techniques being used in crowd counting, focusing on the effect of occlusion, perspective transformation, and dense crowds on accuracy and considering density map-based techniques as more accurate in dense scenarios.

Fan et al. [3] surveyed state-of-the-art crowd counting approaches and identified current trends, noting that many existing methods aim for improved accuracy at the expense of high complexity, thereby making them unsuitable for real-time analysis.

Wang et al. [4] suggested a lightweight convolutional neural network architecture for crowd density estimation on edge devices and thereby demonstrated that the simplicity of a model can significantly improve the efficiency of a process with acceptable accuracy of results.

Wang M et al. [5] discovered crowd density estimation combined with edge intelligence, enhancing the reputation of effective frame-level dispensation and scalability for real-time video-based surveillance classifications.

Deng et al. [6] presented a crowd density estimation prototype optimized for real-time video-based surveillance, strengthening the necessity for harmonizing computational efficiency and accuracy in practical apps.

Khan et al. [7] explored efficient crowd counting via structured knowledge transfer, which enabled small models to excel with reduced computational complexity.

Zhang et al. [8] offered a comprehensive review of convolutional neural network-based techniques for crowd counting and density estimation and provided insights into the evaluation metrics and benchmark datasets that are commonly utilized in the calculations carried out in this study.

How to cite this article:

Materials And Methods

How to cite this article:

M D, S V (May 29, 2026) Video-Based Crowd Analytics Using Deep Learning Methods for Density Estimation. Cureus J Comput Sci 3 : es44389-026-00091-1. DOI <https://doi.org/10.7759/s44389-026-00091-1>

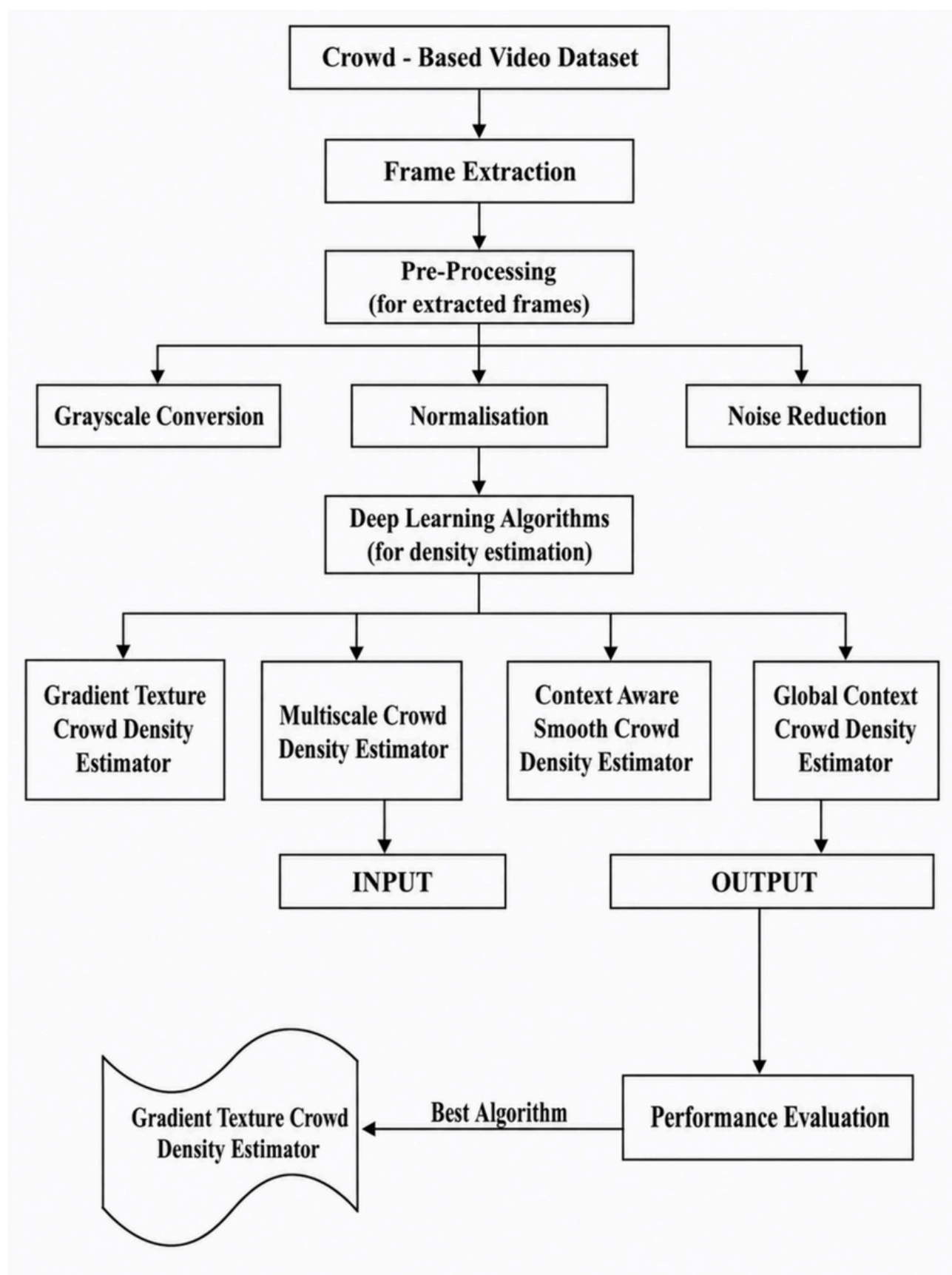


FIGURE 1: Proposed work.

The UCF Crowd Segmentation Dataset comprises 38 high-density crowd videos and has been widely used as a basis for crowd segmentation and flow analysis jobs. The dataset enables realistic crowd behavior and analysis based on motion and was first proposed by Saad Ali and Mubarak Shah at CVPR 2007 and is meant for use with due citations in any research work. The starting point of this work is a crowd-related video dataset, which is essentially the input of this system. The video information has a frame sequence; hence, the initial task of this system is frame extraction, in which a video is fragmented into a set of image frames, further exposed to investigation [9].

Gradient-texture crowd density estimator

This algorithm determines crowd density by examining patterns related to the gradient and textures in image frames. It detects intensity variations in an image, which could be edges or textures generated in denser regions of crowds. The algorithm can work in real-time, meaning there is no significant delay in determining crowd density.

Multi-scale crowd density estimator

The Multi-Scale Crowd Density Estimator calculates crowd density based on the analysis of image features across various scales. This feature helps in dealing with differences in crowd scales and perspectives as it focuses on both fine and coarse scales for crowd density analysis. The drawback of this algorithm is that it consumes more computation and processing time.

Context-aware smooth crowd density estimator

This algorithm focuses on incorporating local contextual information to make the density estimates smooth across neighboring regions. By incorporating the relationship of surrounding pixels, it minimizes noise and sudden fluctuations in the density map. It yields a visually smooth density map but may lower sensitivity to fine-grained changes in density.

Global-context crowd density estimator

Overall scene information is used by this estimator to estimate crowd density. Long-range spatial dependencies and global distribution patterns within the frame are captured. However, it usually results in higher computational cost and may be less responsive to changes in local density.

Once the frames have been extracted, there is a pre-processing stage that enhances the quality of the extracted data and allows for some degree of uniformity. At this stage, there is grayscale conversion, which simplifies complexity by concentrating on intensity and reducing complexity by ignoring color; normalization, which establishes a set intensity range for pixels; and noise reduction, which eliminates unwanted variation in the frames. All these measures improve visual clarity and simplify complexity. At this point, the pre-processed images (Figure 2) are then subject to density estimation based on deep learning algorithms. At this stage, four different algorithms for crowd density estimation are used separately [10]: Gradient-Texture Crowd Density Estimator, Multi-Scale Crowd Density Estimator, Context-Aware Smooth Crowd Density Estimator, and Global Context Crowd Density Estimator on the same frames, estimating density differently for each frame based on the algorithms used in Figure 1.

The input of this stage includes the pre-processed frames, and the output includes a set of density values and density maps for each of the algorithms. The outcomes are then treated based on performance assessment, such as accuracy, stability, scalability, and runtime. After a comparative analysis of performances, the best-suited algorithm is determined. As illustrated in the diagram, based on comparative performance analysis, the Gradient-Texture Crowd Density Estimator has been identified as the best-suited algorithm because of its optimized accuracy, stability, and faster processing speed [11]. As illustrated in the diagram, it represents a complete systematic approach that starts from raw video inputs and progresses until arriving at a suitable algorithm that can perform effective crowd analysis accurately and efficiently.

How to cite this article:

Results And Discussion

Experimental setup

The experiments were conducted using Python and TensorFlow deep learning frameworks on a system equipped with an Intel Core i7 processor, 16 GB RAM, and NVIDIA GPU support. The proposed framework was evaluated using the UCF Crowd Segmentation Dataset containing high-density crowd videos. The input videos were converted into frames and resized before pre-processing operations such as grayscale conversion, normalization, and noise reduction (Figure 2). The extracted frames were processed using four different crowd density estimation approaches: Gradient-Texture Crowd Density Estimator, Multi-Scale Crowd Density Estimator, Context-Aware Smooth Crowd Density Estimator, and Global-Context Crowd Density Estimator. The performance evaluation was carried out using Mean Absolute Error (MAE), Mean Squared Error (MSE), density entropy, prediction variance, and runtime metrics.

To evaluate the performance of the proposed method, several features are measured. MAE in persons per frame defines the accuracy of the prediction, while MSE in the square of persons per frame defines the penalty for higher errors in marker prediction [12]. Entropy of density, measured in bits, defines the crowd density, while the variance of prediction in the square of persons per frame defines the stability of the prediction of density. Runtime in milliseconds per frame defines the efficiency of the algorithm. All the factors collectively give compact feedback about the proposed approach.

Algorithm	MAE	MSE	Density entropy	Prediction variance	Runtime
Gradient-Texture Crowd Density Estimator	73	9,272	18	56,327	271
Multi-Scale Crowd Density Estimator	256	106,241	17	224,682	266
Context-Aware Smooth Crowd Density Estimator	324	188,414	18	557,163	290
Global-Context Crowd Density Estimator	372	387,008	18	710,211	266

TABLE 1: Evaluation metrics.

MAE, mean absolute error; MSE, mean-squared error.

Table 1 presents the performance comparison of four crowd density estimation algorithms in terms of MAE, MSE, density entropy, prediction variance, and runtime. The Gradient-Texture Crowd Density Estimator has the least MAE (73) and the least MSE (9272) among all the algorithms, reflecting that the predicted value of density by this approach is closest to the ground-truth data [13]. In addition, it has lower variance in prediction (56,327), indicating more stability and consistency in the estimated density between frames. Although its runtime is relatively higher (271 ms) than that of the multi-scale method, the improved accuracy and stability make this the most effective overall approach. The Multi-Scale Crowd Density Estimator demonstrates a moderate performance with respect to its higher MAE (256) and MSE (106,241) values

How to cite this article:

compared to the gradient-texture approach. Its low entropy of density (17) reflects reduced variation of density distribution; however, the increased prediction variance indicates reduced consistency in the frame-wise predictions. On the other hand, the Context-Aware Smooth Crowd Density Estimator has a higher MAE (324) and MSE (188,414), indicating its reduced accuracy in estimation. While the incorporated smoothing mechanism helps to keep the entropy values consistent (18), it results in a significant increase in prediction variance (557,163) and runtime [14].

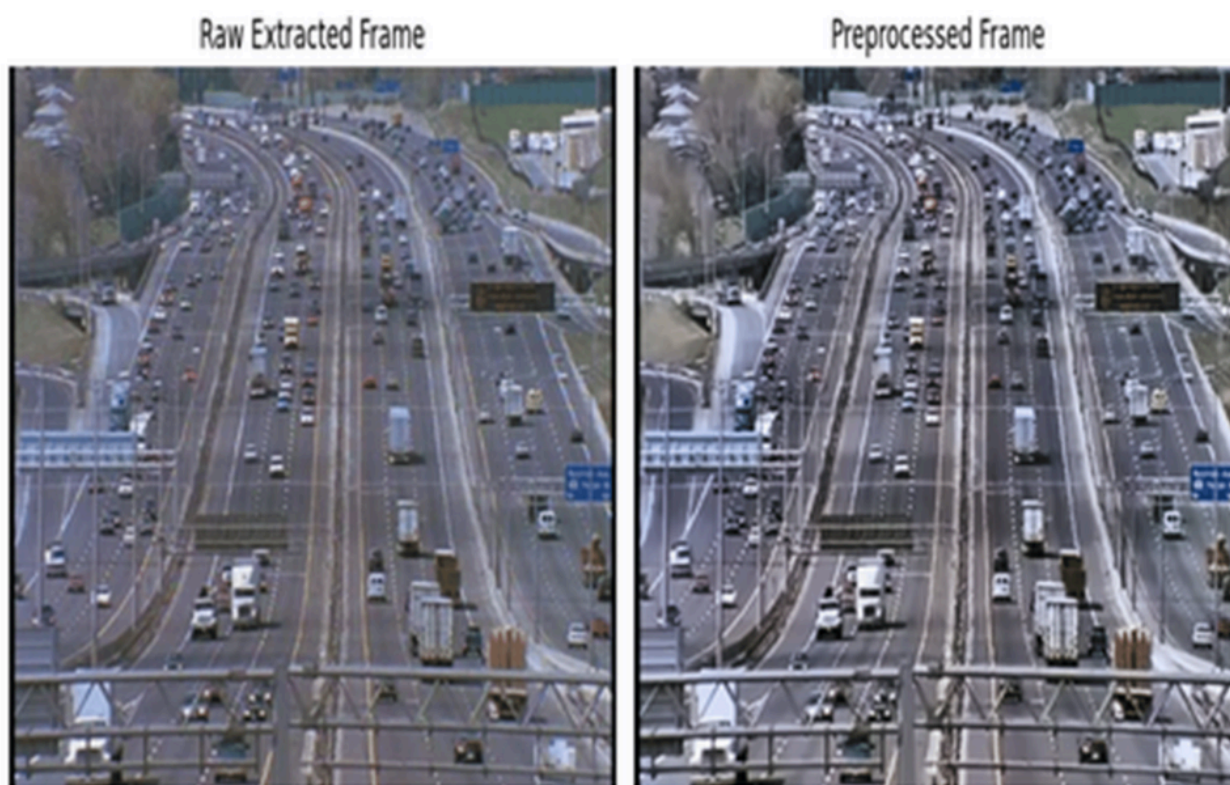


FIGURE 2: Before and after pre-processing.

Figure 3 illustrates a density estimation map and a density estimation level for a given input as an image and a video. For video input, it converts into frames and predicts density estimation for each frame, generated by a deep learning algorithm called Gradient-Texture Crowd Density Estimator [15].

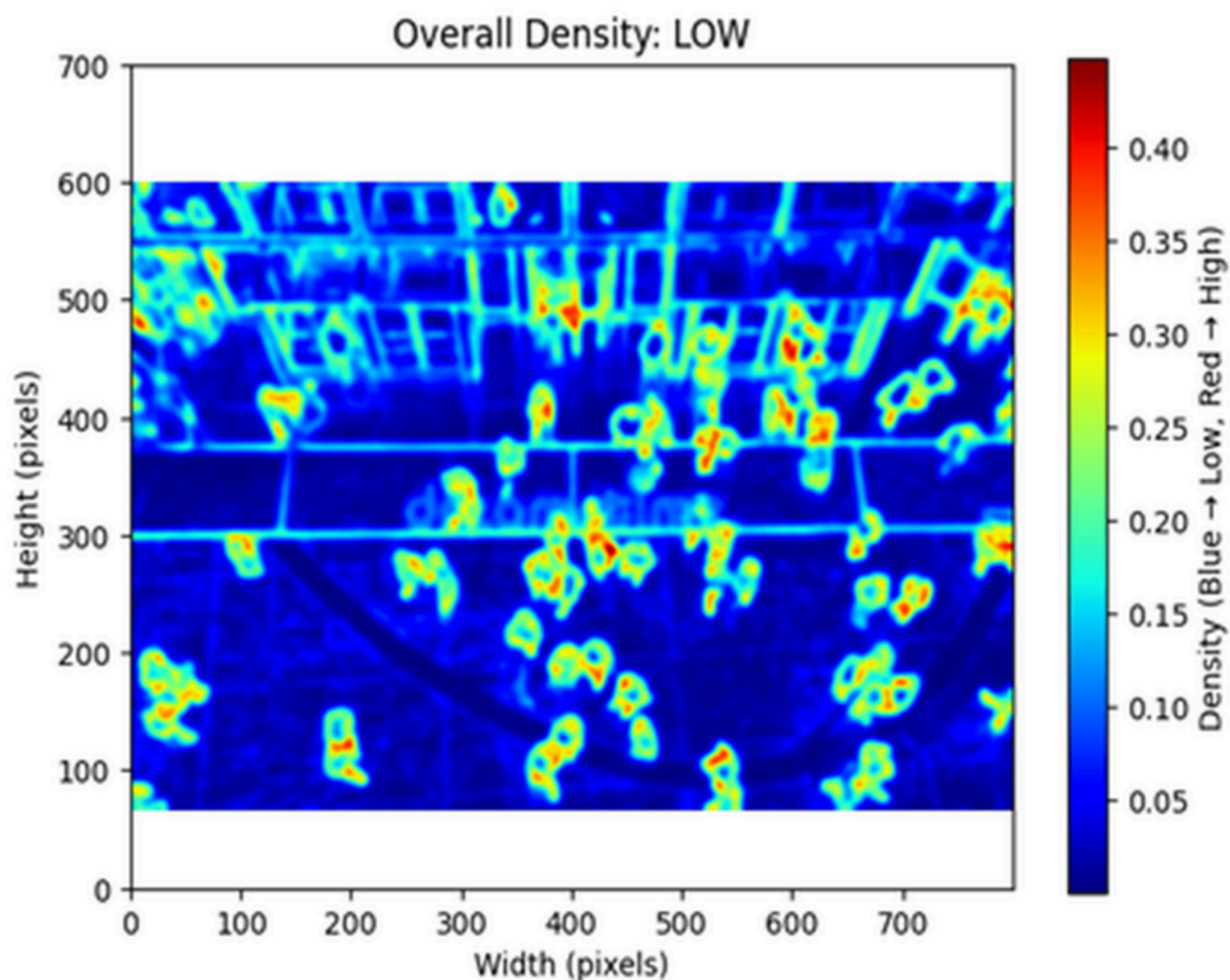


FIGURE 3: Density estimation for gradient-texture crowd density estimator.

Figure 4 illustrates the density estimation map and the density estimation level for a given input as an image and a video. For video input, it converts into frames and predicts density estimation for each frame, generated by a deep learning algorithm called Multi-Scale Crowd Density Estimator [7].

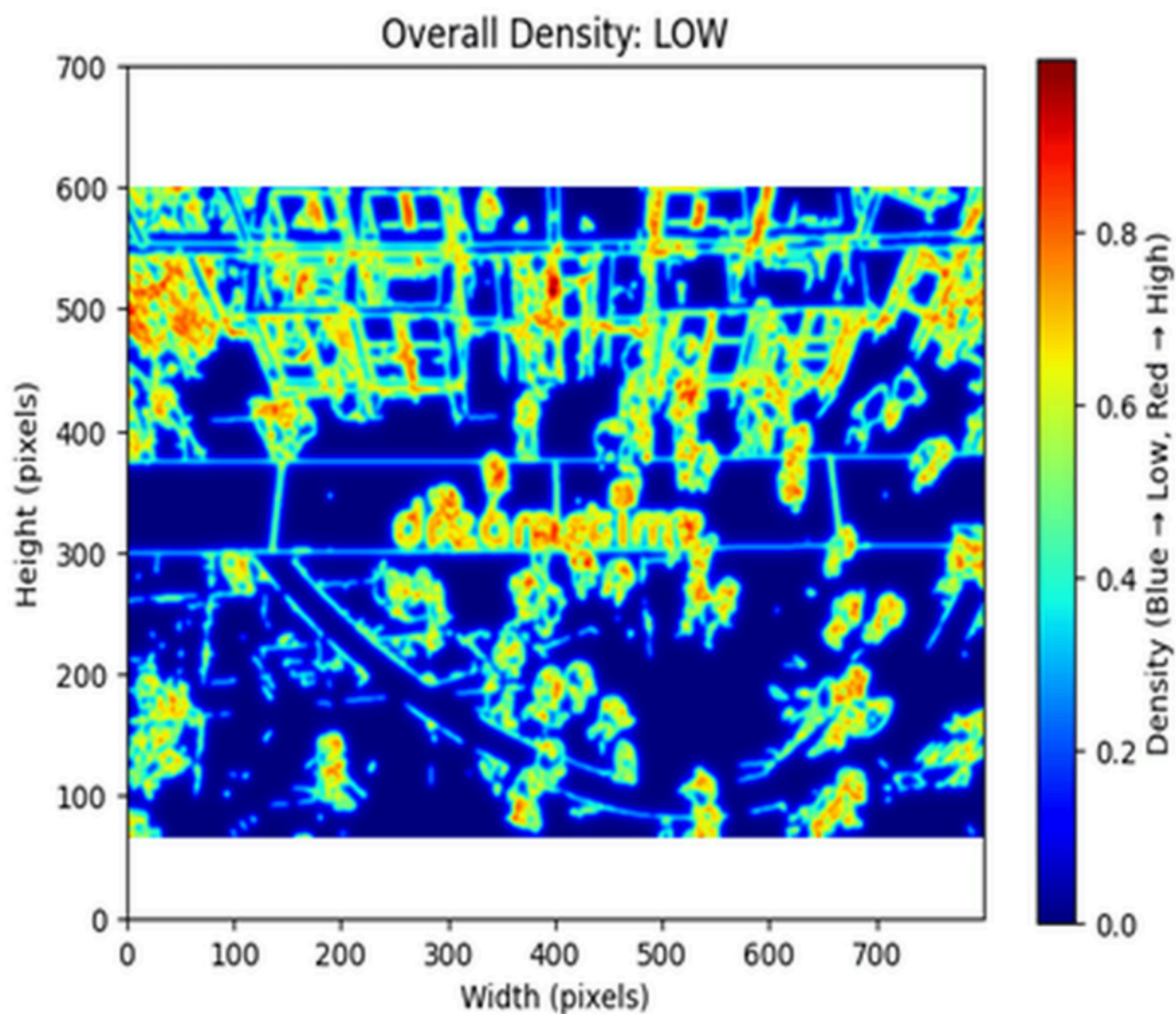


FIGURE 4: Density estimation for multi-scale crowd density estimator.

Figure 5 illustrates the density estimation map and density estimation level for a given input as an image and a video. For video input, it converts into frames and predicts density estimation for each frame, generated by a deep learning algorithm called Context-Aware Smooth Crowd Density Estimator [10].

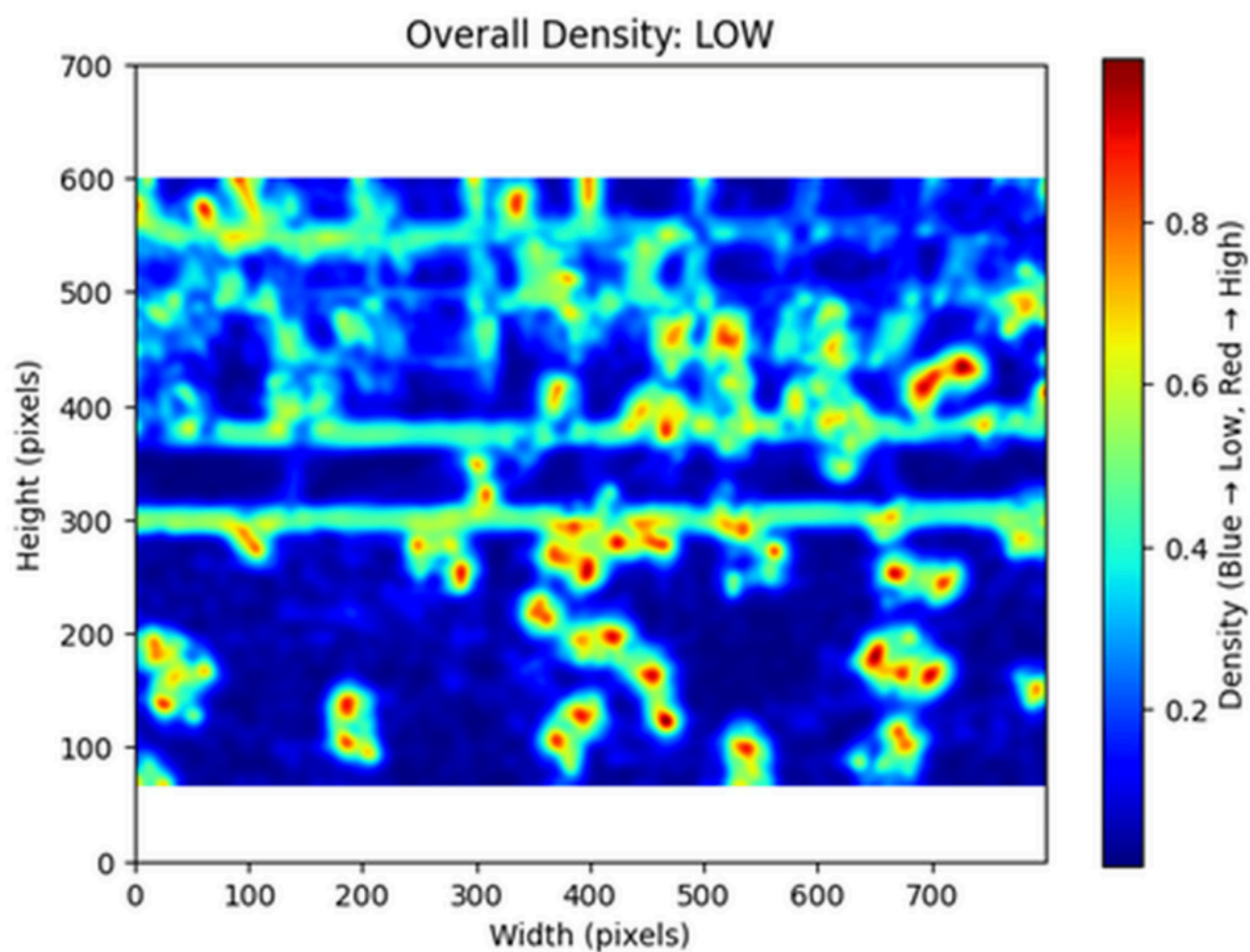
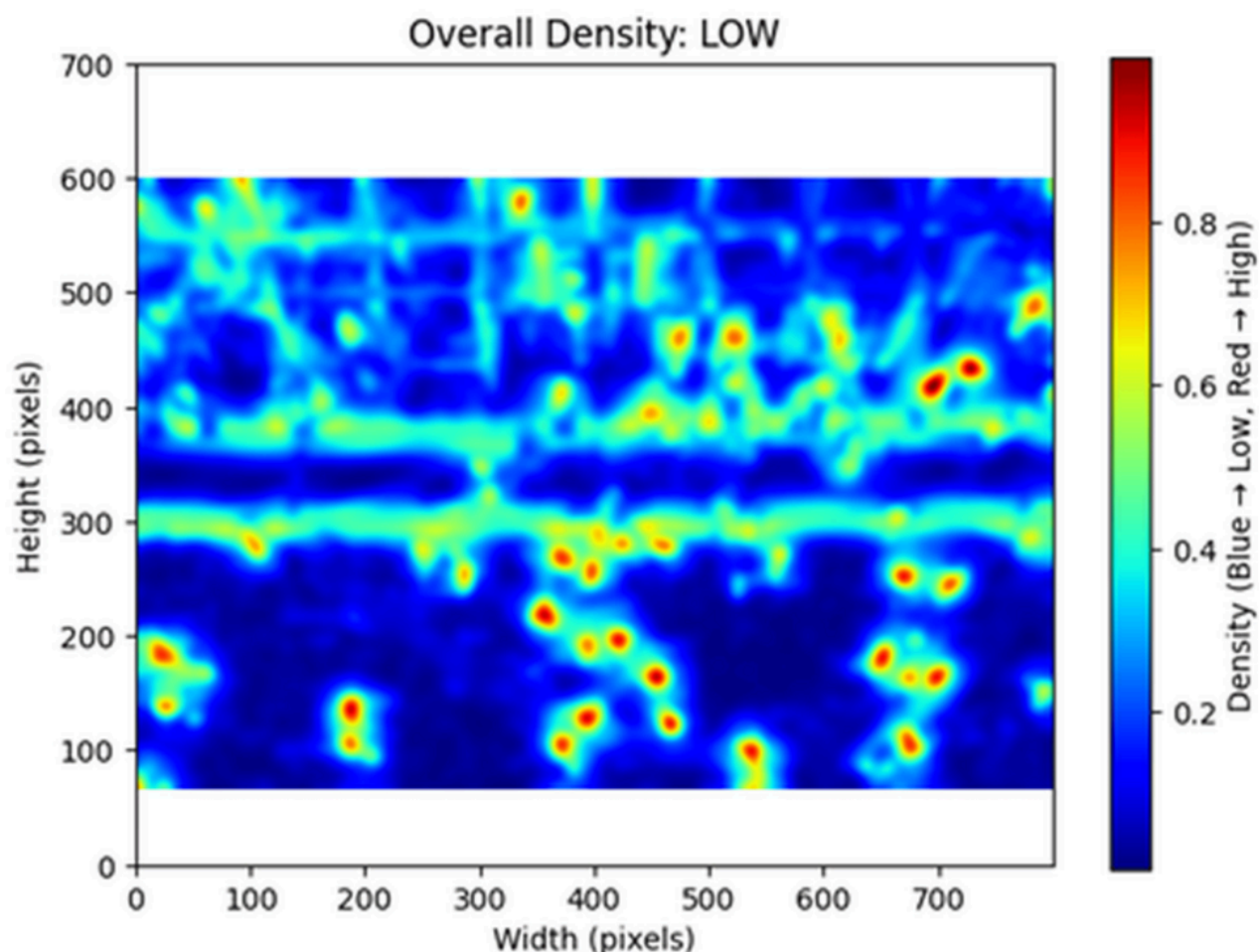


FIGURE 5: Density estimation for context-aware smooth crowd density estimator.

Figure 6 illustrates a density estimation map and density estimation level for a given input as an image and video. For video input, it converts into frames and predicts density estimation for each frame, generated by a deep learning algorithm called Global-Context Crowd Density Estimator [12,15].

Future direction**FIGURE 6: Density estimation for global-context crowd density estimator.**

For future work, integrating state-of-the-art deep learning architecture models will help improve the accuracy of density estimation in extreme crowd situations. The model can be developed further to facilitate real-time execution with edge computing and multi-camera fusion for large-scale surveillance systems. Implementing temporal models and adaptive thresholding can help improve behavior prediction and anomaly identification of crowd behavior. Investigation of greater datasets will support the progress of the generalization and strength of the future models.

Conclusions

In this work, a frame-based density estimation technique has been proposed for video analysis using four different density estimation approaches. From the experimental analysis, it has been observed that the Gradient-Texture Crowd Density Estimator performs better compared to other density estimation approaches in terms of MAE, MSE, density entropy, prediction variance, runtime, as well as density level and density maps. This work provides a scalable, clear and efficient technique for progress with real-time crowds and is used for real-time crowds' analysis or surveillance systems in smart cities.

How to cite this article:

Additional Information

Author Contributions

All authors have reviewed the final version to be published and agreed to be accountable for all aspects of the work.

Concept and design: Deepadharshana M, Vijayarani S

Acquisition, analysis, or interpretation of data: Deepadharshana M, Vijayarani S

Drafting of the manuscript: Deepadharshana M, Vijayarani S

Critical review of the manuscript for important intellectual content: Deepadharshana M, Vijayarani S

Supervision: Deepadharshana M, Vijayarani S

Disclosures

Human subjects: All authors have confirmed that this study did not involve human participants or tissue. **Animal subjects:** All authors have confirmed that this study did not involve animal subjects or tissue. **Conflicts of interest:** In compliance with the ICMJE uniform disclosure form, all authors declare the following: **Payment/services info:** All authors have declared that no financial support was received from any organization for the submitted work. **Financial relationships:** All authors have declared that they have no financial relationships at present or within the previous three years with any organizations that might have an interest in the submitted work. **Other relationships:** All authors have declared that there are no other relationships or activities that could appear to have influenced the submitted work.

Data Availability Statements

The datasets (and/or code) supporting this study are available from the corresponding author upon reasonable request.

Acknowledgements

The paper was submitted in - "International Conference on Applied Intelligence and Innovative Computational Solutions 2026 (IC-AIICS 2026)", February 9–11, 2026, The Gandhigram Rural Institute, Dindigul.

References

1. Gao G, Gao J, Liu Q, Wang Q, Wang Y: [A survey of deep learning methods for density estimation and crowd counting](#). *Vicinagearth*. 2025, 2:1-37. [10.1007/s44336-024-00011-8](#)
2. Patwal A, Diwakar M, Tripathi V, Singh P: [Crowd counting analysis using deep learning: a critical review](#). *Procedia Computer Science*. 2023, 218:2448-2458. [10.1016/j.procs.2023.01.220](#)
3. Fan Z, Zhang H, Zhang Z, Lu G, Zhang Y, Wang Y: [A survey of crowd counting and density estimation based on convolutional neural network](#). *Neurocomputing*. 2022, 472:224-251. [10.1016/j.neucom.2021.02.103](#)
4. Wang S, Pu Z, Li Q, Wang Y: [Estimating crowd density with edge intelligence based on lightweight convolutional neural networks](#). *Expert Systems with Applications*. 2022, 206:117823. [10.1016/j.eswa.2022.117823](#)
5. Wang M, Zhou X, Chen Y: [A comprehensive survey of crowd density estimation and counting](#). *IET Image Processing*. 2025, 19:1-32. [10.1049/ipr2.13328](#)
6. Deng L, Zhou Q, Wang S, Górriz JM, Zhang Y: [Deep learning in crowd counting: a survey](#). *CAAI Transactions on Intelligence Technology*. 2023, 9:1043-1077. [10.1049/cit2.12241](#)
7. Khan K, Albattah W, Khan RU, Qamar AM, Nayab D: [Advances and trends in real time visual crowd analysis](#). *Sensors*. 2020, 20:1-28. [10.3390/s20185073](#)
8. Zhang X, Sun Y, Li Q, Li X, Shi X: [Crowd density estimation and mapping method based on surveillance video and GIS](#). *ISPRS International Journal of Geo-Information*. 2023, 12:1-17. [10.3390/ijgi12020056](#)

How to cite this article:

M D, S V (May 29, 2026) Video-Based Crowd Analytics Using Deep Learning Methods for Density Estimation. *Cureus J Comput Sci* 3 : es44389-026-00091-1. DOI <https://doi.org/10.7759/s44389-026-00091-1>

9. Jiang G, Wu R, Huo Z, Zhao C, Luo J: [LigMSANet: lightweight multi-scale adaptive convolutional neural network for dense crowd counting](#). Expert Systems with Applications. 2022, 197:116662. [10.1016/j.eswa.2022.116662](#)
10. Wang L, Li Y, Peng S, Tang X, Yin B: [Multi-level feature fusion network for crowd counting](#). IET Computer Vision. 2021, 15:60-72. [10.1049/cvi2.12012](#)
11. Sindagi VA, Patel VM: [CNN-based cascaded multi-task learning of high-level prior and density estimation for crowd counting](#). 14th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). 2017, 1-6. [10.1109/AVSS.2017.8078491](#)
12. Idrees H, Tayyab M, Athrey K, Zhang D, Al-Maadeed S, Rajpoot N, Shah M: [Composition loss for counting, density map estimation and localization in dense crowds](#). arXiv. 2018, 1-15. [10.48550/arXiv.1808.01050](#)
13. Liu YB, Jia RS, Liu QM, Zhang XL, Sun HM: [Crowd counting method based on the self-attention residual network](#). Applied Intelligence. 2020, 51:427-440. [10.1007/s10489-020-01842-w](#)
14. Zhang L, Shi M, Chen Q: [Crowd counting via scale-adaptive convolutional neural network](#). 2018, 1113-1121. [10.1109/WACV.2018.00127](#)
15. Hossain M, Hosseinzadeh M, Chanda O, Wang Y: [Crowd counting using scale-aware attention networks](#). IEEE Winter Conference on Applications of Computer Vision (WACV). 2019, 1280-1288. [10.1109/WACV.2019.00141](#)

How to cite this article:

M D, S V (May 29, 2026) Video-Based Crowd Analytics Using Deep Learning Methods for Density Estimation. Cureus J Comput Sci 3 : es44389-026-00091-1. DOI <https://doi.org/10.7759/s44389-026-00091-1>