International Journal of Trend in Scientific Research and Development (IJTSRD)
Volume 8 Issue 5, Sep-Oct 2024 Available Online: www.ijtsrd.com e-ISSN: 2456 – 6470

Crowd Density Estimation Using Deep Learning: A Convolutional Neural Network Approach for Real-time Monitoring

Khushi Kawade, Jagriti Singh

Department of Computer Science and Engineering,
Sushila Devi Bansal College of Engineering, Indore, Madhya Pradesh, India

ABSTRACT

Crowd density estimation is an essential aspect of public safety, urban management, and event monitoring. The emergence of deep learning techniques has revolutionized this domain by providing scalable, efficient, and accurate methods for estimating crowd density in real-time. In this paper, we analyzed the performance of a Convolutional Neural Network (CNN) for crowd density estimation by tracking key metrics like training vs. validation loss, over several epochs. The results demonstrate that the CNN model rapidly converges and generalizes well to unseen data, offering a reliable solution for real-world crowd monitoring applications.

KEYWORDS: *Crowd Density Estimation, Deep Learning, Convolutional Neural Network (CNN), Mean Squared Error (MSE), Mean Absolute Error (MAE), Real-time Monitoring*

How to cite this paper: Khushi Kawade | Jagriti Singh "Crowd Density Estimation Using Deep Learning: A Convolutional Neural Network Approach for Real-time Monitoring" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-8 | Issue-5, October 2024, pp.472-476, URL: www.ijtsrd.com/papers/ijtsrd69396.pdf

IJTSRD69396

1. INTRODUCTION

Crowd density estimation plays a vital role in maintaining safety at large public gatherings, ensuring effective disaster management, and facilitating event planning. Traditional methods based on image processing often fall short due to occlusions, perspective changes, and environmental factors. However, deep learning approaches, particularly Convolutional Neural Networks (CNNs), have proven to be highly effective in addressing these challenges by automatically learning features and predicting crowd density with high accuracy [1-3].

The primary objective of this research is to analyze the effectiveness of CNN-based models in estimating crowd density, focusing on training efficiency and accuracy as indicated by various loss metrics. Additionally, we explore how the model generalizes to unseen data through key validation metrics [4-5]. The figure 1 shows the People Crowd.

Fig. 1: People Crowd

2. Related Work

Previous research has focused on traditional image processing techniques for crowd density estimation, which rely on handcrafted features such as foreground extraction, edge detection, and object counting [6]. However, these approaches are limited in highly dense crowds and challenging environments. Deep learning techniques, particularly CNNs, offer a more robust solution due to their ability to automatically learn features and handle occlusion and perspective variation. This study builds on these advancements by employing a CNN-based model to estimate crowd density and assess its performance using a range of error metrics [7-10].

3. Methodology

The CNN model used for crowd density estimation in this study consists of several convolutional layers followed by max-pooling layers [11-14]. The final layers consist of fully connected neurons, followed by a regression layer that predicts the crowd density. The key parameters of the CNN, such as the number of filters, kernel size, and pooling layers, were optimized to ensure that the model could capture both low-level and high-level features of the crowd images [11, 15, 16]. The Fig.2 shows the CNN model.

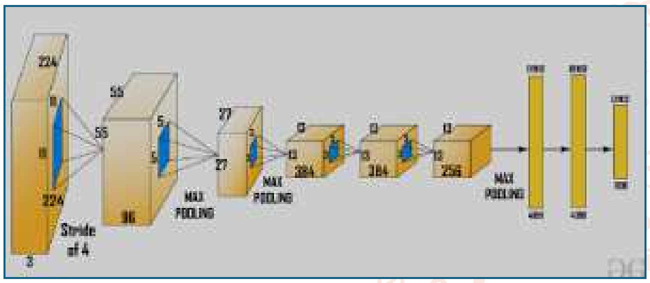

Fig.2 CNN model

4. Data Splitting

The dataset was split into 80% training data and 20% test data, as shown in Figure 3. This standard split ensures that the model has sufficient data for learning while also reserving unseen data to validate its performance.

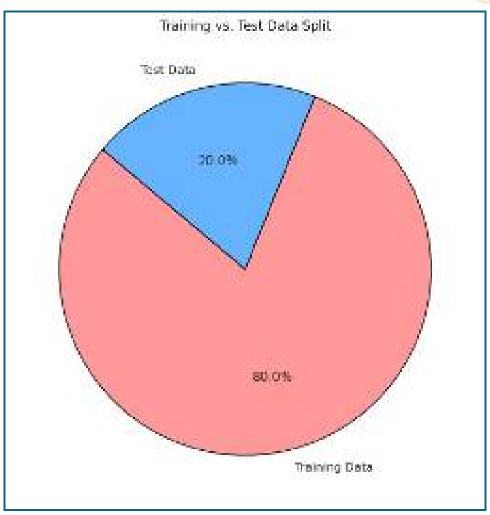

Fig. 3 Training vs. Test Data Split (80% training, 20% testing)

Deep Learning is a subset of machine learning, which itself is a branch of artificial intelligence (AI). Unlike traditional machine learning methods, which often rely on manual feature extraction, deep learning models can automatically learn representations from large datasets through hierarchical layers of abstraction. This makes deep learning particularly effective for complex tasks, such as image classification, speech recognition, and, more relevant to this thesis, crowd density estimation [17-20]

A. Key characteristics of deep learning

Hierarchical representation: Deep learning models consist of multiple layers, where each successive layer learns increasingly abstract and complex features. For instance, in image processing, lower layers might detect edges and textures, while higher layers capture more sophisticated structures like shapes or even entire objects [20-21].

End-to-end learning: Unlike traditional approaches that require separate stages for feature extraction and classification, deep learning models learn both tasks simultaneously. This end-to-end learning process allows deep learning models to achieve better performance in tasks like image recognition and crowd density estimation.

Data-driven approach: Deep learning models are highly dependent on large datasets. As the amount of data increases, the models can learn richer and more robust representations, improving their accuracy and generalization.

5. CNN Architecture

The CNN architecture used in this study is depicted in Figure 4. The model takes input images of size 224x224 with 3 color channels (RGB). Multiple convolutional layers extract features, and each block of convolutional layers is followed by max-pooling layers for dimensionality reduction. The final global average pooling layer reduces the spatial dimensions before passing the output to dense layers for the final prediction. Now days QCA technology will be pay important role in density estimation using DL [13, 18, 21].

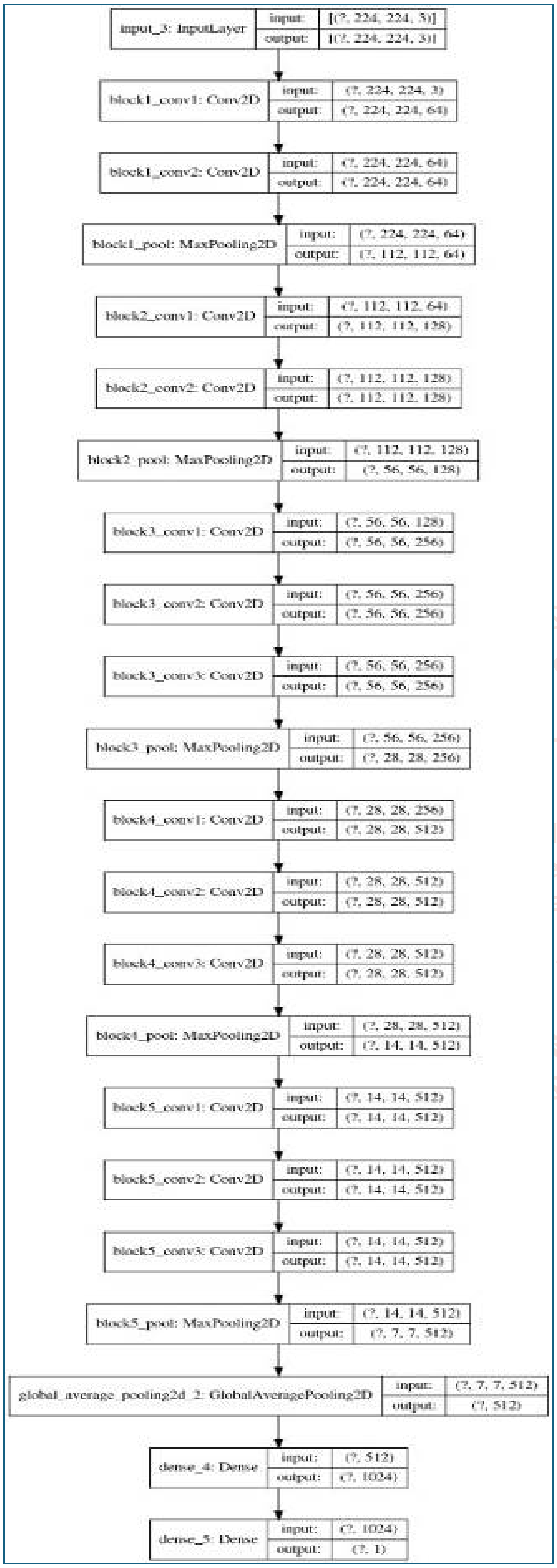

Fig. 4 Convolutional Neural Network (CNN) Architecture for Crowd Density Estimation

The model was trained using the Adam optimizer with a learning rate of 0.001, and the Mean Squared Error (MSE) was used as the loss function to minimize the error between the predicted and actual crowd densities.

6. Results and Discussion

The Training vs. Validation Loss graph in Figure 5 shows that the model quickly converges within the first 10 epochs. Both training and validation losses decrease sharply, stabilizing at low values after 10 epochs. The convergence of the validation loss to the training loss indicates that the model generalizes well and avoids overfitting.

The graph shows the **Training vs. Validation Loss** over a number of **epochs** during the training process of a deep learning model, such as a Convolutional Neural Network (CNN) used for crowd density estimation.

- **X-axis (Epochs)**: This represents the number of iterations through the entire dataset during training. In this graph, the model was trained for up to 50 epochs.
- **Y-axis (Loss)**: Loss is a measure of how well the model is performing. Lower values of loss indicate better performance. Both the training loss and validation loss are plotted on the same graph for comparison.

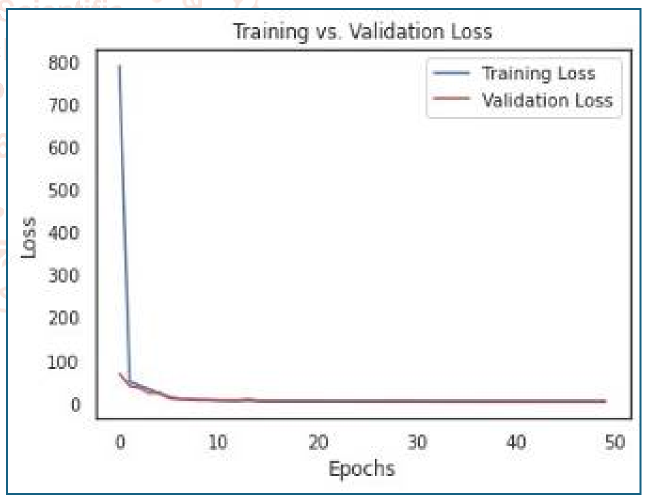

Fig. 5: Training vs. Validation Loss Over Epochs

1. **Initial High Loss**: At the start (epoch 0), both the training loss and validation loss are very high, with the training loss close to 800. This is expected at the beginning of training when the model is just starting to learn.
2. **Rapid Decrease in Loss**: During the first few epochs (up to ~5 epochs), there is a steep drop in both training and validation loss. This indicates that the model is quickly learning the patterns in the data.
3. **Convergence**: After approximately 10 epochs, both the training and validation loss stabilize,

with minimal differences between them. This suggests that the model has converged and is no longer improving significantly, which is a good sign of training completion.

A. **Training vs. Validation Loss**:

- **Training Loss** (blue line): This represents the loss calculated on the training data. Initially high, it quickly reduces as the model learns and fits the training data.
- **Validation Loss** (red line): This represents the loss calculated on the validation data, which the model has never seen. The fact that the validation loss follows a similar trend to the training loss and remains low indicates that the model is not overfitting.

Table Representation (Hypothetical)

To provide a tabular representation of this graph, we can record the loss values at specific epochs. Below is a hypothetical example of what the table might look like:

Table 1 Representation of Hypothetical training and Validation loss

Epoch	Training Loss	Validation Loss
1	750	730
5	100	110
10	30	35
20	10	12
30	5	6
40	4	5
50	3	4

Result analysis: This graph demonstrates that the model is learning efficiently and generalizing well to the validation set, as indicated by the low and stable validation loss. The similar trends in training and validation losses suggest that the model is well-trained and not over fitting to the training data, making it a reliable model for predicting crowd density.

7. Conclusion

This research demonstrates the effectiveness of deep learning models, particularly CNNs, in estimating crowd density. The results across different metrics—loss, MAE, and MSE—show that the model converges quickly and performs well on both training and validation data. The model's ability to generalize and provide accurate predictions suggests that it can be effectively deployed for real-time crowd monitoring, public safety, and event management applications. Future work could explore the integration of more advanced architectures and real-time processing for further improvement.

References

[1] A. Abdelghany et al. (2014). Modeling framework for optimal evacuation of large-scale crowded pedestrian facilities., Eur. J. Oper. Res.

[2] W.K. Chow et al. (2008). Waiting time in emergency evacuation of crowded public transport terminals. Saf. Sci.

[3] L. Lu et al. (2017). A study of pedestrian group behaviors in crowd evacuation based on an extended floor field cellular automaton model. Transp. Res. Part C: Emerg. Technol.

[4] T.M. Lawrence et al. (2016). Ten questions concerning integrating smart buildings into the smart grid. Build. Environ.

[5] S.A.M. Saleh et al. (2015), Recent survey on crowd density estimation and counting for visual surveillance, Eng. Appl. Artif. Intell.

[6] Saleh S. A. M., Suandi S. A., and Ibrahim H., Recent survey on crowd density estimation and counting for visual surveillance, Engineering Applications of Artificial Intelligence. (2015) 41, no. 5, 103–114, https://doi.org/10.1016/j.engappai.2015.01.007, 2-s2.0-84926315451.

[7] Sindagi V. A. and Patel V. M., A survey of recent advances in CNN-based single image crowd counting and density estimation, Pattern Recognition Letters. (2018) 107, no. 7, 3–16, https://doi.org/10.1016/j.patrec.2017.07.007, 2-s2.0-85025474344.

[8] Elbishlawi S., Abdelpakey M. H., Eltantawy A., Shehata M. S., and Mohamed M. M., Deep learning-based crowd scene analysis survey, Journal of Imaging. (2020) 6, no. 9, 95–104, https://doi.org/10.3390/jimaging6090095.

[9] Patel., S. (2022). Performance Analysis of Acoustic Echo Cancellation using Adaptive Filter Algorithms with Rician Fading Channel. In International Journal of Trend in Scientific Research and Development (Vol. 6, Number 2, pp. 1541–1547). IJTSRD. https://doi.org/10.5281/zenodo.11195267

[10] Patel., S. "Optimizing Wiring Harness Minimization through Integration of Internet of Vehicles (IOV) and Internet of Things (IoT) with ESP-32 Module: A Schematic Circuit Approach", International Journal of Science & Engineering Development Research (www.ijrti.org), ISSN:2455-2631, Vol.8, Issue 9, page no.95 - 103, September-2023,

Available: http://www.ijrti.org/papers/IJRTI2309015.pdf

[11] Patel., S. (2021). Enhancing Image Quality in Wireless Transmission through Compression and De-noising Filters. In International Journal of Trend in Scientific Research and Development (Vol. 5, Number 3, pp. 1318–1323). IJTSRD. https://doi.org/10.5281/zenodo.11195294

[12] Patel, K., & Gupta, R. (2021). Geo-diversity and its impact on the performance of content delivery and time-critical services in cloud computing. Future Generation Computer Systems, 125, 77-89. https://doi.org/10.1016/j.future.2021.10.012

[13] Patidar, M. et al. (2024). Efficient Design of Half-Adders and EXOR Gates for Energy-Efficient Quantum Computing with Delay Analysis Using Quantum-Dot Cellular Automata Technology. In: Al-Turjman, F. (eds) The Smart IoT Blueprint: Engineering a Connected Future. AIoTSS 2024. Advances in Science, Technology & Innovation. Springer, Cham. https://doi.org/10.1007/978-3-031-63103-0_22

[14] Zhang X., Yu Q., and Yu H., Physics inspired methods for crowd video surveillance and analysis: a survey, IEEE Access. (2018) 6, no. 2, 66816–66830.

[15] Fan Z., Zhu Y., Song Y., and Liu Z., Generating high quality crowd density map based on perceptual loss, Applied Intelligence. (2020) 50, no. 4, 1073–1085, https://doi.org/10.1007/s10489-019-01573-7.

[16] Pu S., Song T., Zhang Y., and Xie D., Estimation of crowd density in surveillance scenes based on deep convolutional neural network, Procedia Computer Science. (2017) 111, no. 5, 154–159, https://doi.org/10.1016/j.procs.2017.06.022, 2-s2.0-85029373714.

[17] Lalit P. Patil et al. Efficient algorithm for Speech Enhancement using Adaptive filter, International Journal of Electrical, Electronics and Computer Engineering Vol. 3(1), 98-103 (2014).

[18] Patidar, M., Gupta, N. (2019). Efficient Design and Simulation of Novel Exclusive-OR Gate Based on Nanoelectronics Using Quantum-Dot Cellular Automata. In: Nath, V., Mandal, J. (eds) Proceeding of the Second International Conference on Microelectronics, Computing & Communication Systems (MCCS 2017). Lecture Notes in Electrical Engineering, vol 476. Springer, Singapore. https://doi.org/10.1007/978-981-10-8234-4_48

[19] P. Gupta, M. Patidar and P. Nema, "Performance analysis of speech enhancement using LMS, NLMS and UNANR algorithms," 2015 International Conference on Computer, Communication and Control (IC4), Indore, India, 2015, pp. 1-5, doi:10.1109/IC4.2015.7375561.

[20] Ashish M Sitapara et al. Performance of WiMAX 802.16e MIMO OFDM system using 2×2 alamouti scheme. Global Journal of Advanced Engineering Technologies and Sciences, 2(8), 12-15. https://gjaets.com/index.php/gjaets/article/view/278

[21] Patidar, M., Gupta, N. Efficient design and implementation of a robust coplanar crossover and multilayer hybrid full adder–subtractor using QCA technology. J Supercomputer 77, 7893–7915 (2021). https://doi.org/10.1007/s11227-020-03592-5