Crowd behaviour analysis using a deep learning model for Arbaeen Pilgrimage

Israa Nadheer

College of Information Engineering / Al-Nahrain University

Email: asraa.nadheer88@nahrainuniv.edu.iq

Abstract

The purpose of this study is to analyse crowd behaviour using a deep learning model. This research investigates the effectiveness of deep learning techniques in estimating crowd levels, aiming to improve accuracy and efficiency in real-time applications. The study employs a convolutional neural network (CNN), architecture trained on a diverse dataset of crowd images. Key steps in the proposed approach include data preprocessing, model training, validation, and testing. Major findings indicate that the deep learning model achieves lower error rates in both the training and testing phases, demonstrating its robustness and generalisability. Specifically, the model attained a mean absolute error (MAE) of 55.34 and a root mean squared error (RMSE) of 98.32 during testing, compared to 56.5 MAE and 99.13 RMSE during training. These results highlight the model's capacity to generalise well to new, unseen data. The study concludes that deep learning models, when properly trained and validated, can significantly enhance crowd-level estimation, offering valuable insights for applications in public safety, event management, and urban planning. Future work will focus on expanding the dataset and refining the model to further improve performance and applicability. Overall, our study presents a promising advancement in crowd counting technology with practical implications for crowd management at large-scale events like the Arbaeen Pilgrimage.

Introduction

In response to the growing demand for crowd flow monitoring, assembly control, and security services, numerous network models have been developed. These models aim to provide effective solutions for handling congested scenes (Haghani, 2023), (Musa, 2023). The evolution of analysis methods has progressed from simple crowd counting (which quantifies the number of people in an image (Haghani, 2023) to density map representation (which visualizes crowd distribution characteristics (Musa, 2023). Real-life scenarios have highlighted the limitations of merely counting individuals, as the same number of people can exhibit vastly different crowd distributions (Tang, 2021), (Hassen, 2022). To obtain more accurate and comprehensive information, density maps play a crucial role. In high-risk environments, such as during stampedes or riots, having precise distribution patterns becomes critical for making informed decisions (Owaidah, 2019).

Generating accurate crowd distribution patterns remains a challenge. One major difficulty arises from the prediction approach: density values are generated pixel-by-pixel, necessitating spatial coherence in the output density maps to ensure smooth transitions between neighbouring pixels. Additionally, diverse scenes—such as irregular crowd clusters and varying camera perspectives—complicate the task, especially when traditional methods lacking deep neural networks (DNNs) are employed (Ahmed, 2023) (Assefa, 2022). Recent advancements in congested scene analysis rely on DNN-based techniques due to their high accuracy in semantic

segmentation tasks and significant progress in visual saliency.

The Arbaeen pilgrimage to Karbala poses significant crowd management challenges due to the massive influx of millions of pilgrims from around the world. Safety concerns loom large, with the risk of stampedes and accidents in overcrowded conditions. Local infrastructure strains under the weight of the pilgrimage, causing traffic congestion, overcrowded transportation, and shortages of essential services. Sanitation and hygiene suffer, raising the potential for disease outbreaks. Security measures must be stringent to safeguard against threats. Logistical hurdles abound, requiring meticulous planning to manage crowd flow and provide necessary facilities. Environmental impacts, such as increased waste and pollution, further complicate matters. Effective communication and coordination among stakeholders are paramount for successful crowd management during this religious gathering. This study aims to address the significant crowd management challenges of the Arbaeen pilgrimage to Karbala, focusing on safety, infrastructure, sanitation, security, logistics, and environmental impacts. It proposes a crowd counting approach utilizing the VGG-16 architecture, a pre-trained CNN enhanced with morphological operations to improve feature representation and density estimation accuracy. The research explores how these advanced techniques can mitigate noise, enhance crowd counting precision, and be integrated into existing management strategies to improve safety and efficiency during the pilgrimage.

In this paper, we propose an approach that leverages a combination of advanced techniques and pre-trained CNN to achieve precise crowd counting, as shown in Figure 1. Firstly, we utilize the VGG-16 architecture, a powerful CNN known for its effectiveness in image classification tasks. We load the pre-trained VGG-16 model and extract features from both training and testing images. We place particular emphasis on employing morphological operations to enhance the feature representation extracted from crowd images. After binarizing the extracted features, morphological operations such as opening and closing are applied to refine the representation. These operations help to mitigate noise and emphasize the structural characteristics of the crowd, thereby improving the accuracy of density estimation. The deep features are then utilised as inputs for our regression model.

Literature review

Enhancing the resolution of feature maps leads to finer details, resulting in higher-quality density maps that aid in crowd count estimation (Cao X, 2018) (Wan J, 2019). However, when pooling operations are employed to increase receptive fields in networks, the resolution of feature maps decreases, causing the loss of crowd image details. To maintain consistent input and output resolutions, the encoder-decoder structure is commonly used (Jiang, 2019) (Thanasutives P, 2021). In this structure, the encoder extracts input image features and combines them, while the specially designed decoder decodes the higher-level features required by these extracted

features. For instance, the multi-scale-aware fusion network with attention (M-SFANet) mechanism (Thanasutives P, 2021) enhances its encoder with Atrous spatial pyramid pooling (ASSP) (Chan AB, 2008), which extracts multi-scale features from the target object and fuses context information. To handle scale variations in input images, M-SFANet employs the context-aware network (CAN) module (Liu W, 2019) as the decoder. As deep neural networks, different layers contain varying crowd information, but some details are inevitably lost (Liu W, 2019). Dense Scale Network (DSNet) (Dai F, 2021) was proposed to effectively extract long-distance context information and maximize the retention of network layer details. Li et al (2022) introduced the densely connected multi-scale pyramid network (DMPNet) as a solution for crowd count estimation and the production of high-fidelity density maps. The central component of their network architecture is the Multi-scale Pyramid Network (MPN), which adeptly captures multi-scale features from crowd images while preserving input feature map resolution and channel count. To enhance information flow across network layers, dense connections are employed to link multiple MPNs. Additionally, the authors devised a novel loss function to improve model convergence. Extensive experiments conducted on three challenging benchmark crowd counting datasets demonstrate that DMPNet outperforms state-of-the-art algorithms in terms of both parameters and results.

Semi-Annual Scientific Journal
www.arbnj.com
Vol. 3 , 3nd Year, Supplemen (5)

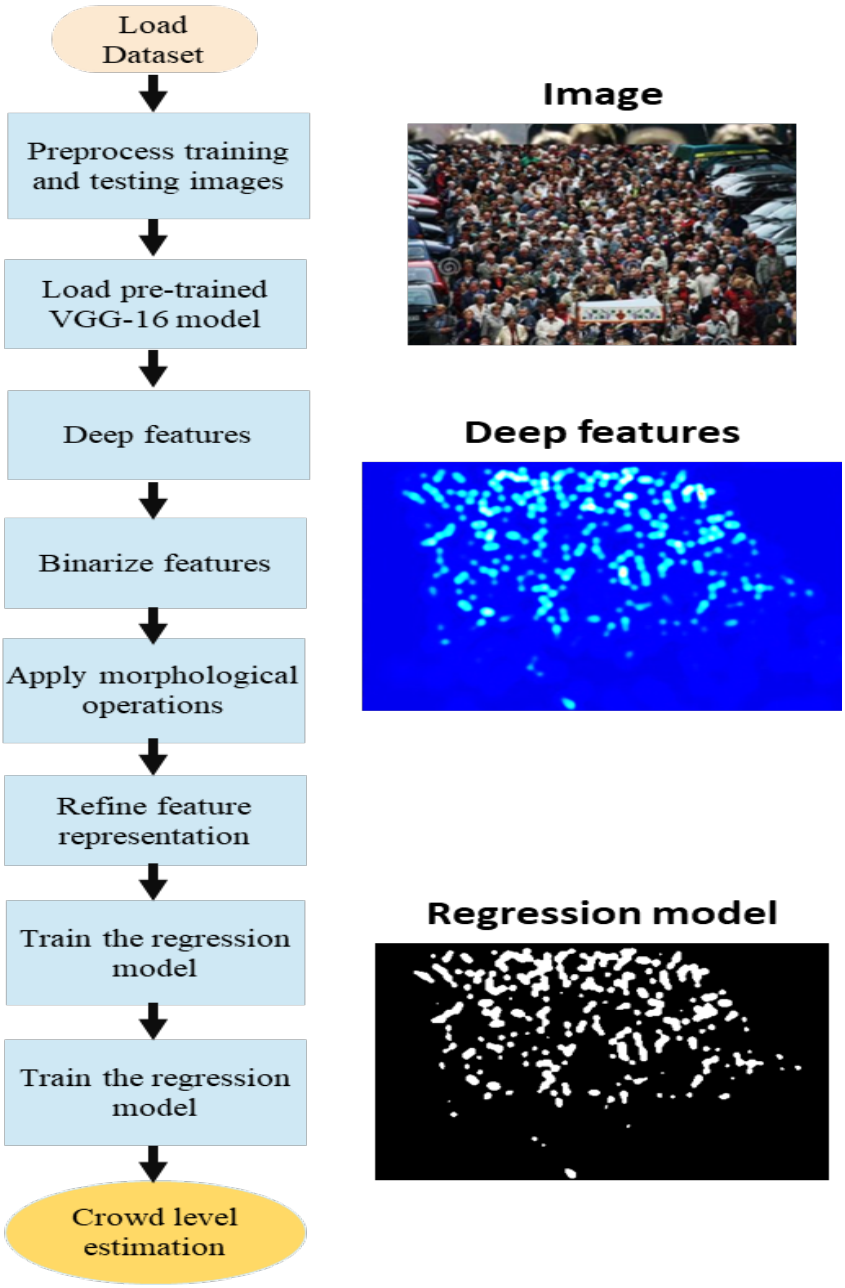

Fig. 1. The proposed approach where the flow chart represents the main steps and images on the right represent the main outcomes of main steps.

Methods and materials

Dataset

The Shanghaitech dataset (Zhang, Y, 2016) represents a significant contribution to the field of crowd counting, offering a rich collection of 1198 annotated crowd images. These images are strategically divided into two main sections: Part-A and Part-B, each serving distinct purposes in research and analysis. Part-A encompasses 482 images, while Part-B boasts a larger pool of 716 images, providing researchers with a diverse range of crowd scenarios to analyse. Delving deeper, Part-A is meticulously organised into train and test subsets, with 300 and 182 images, respectively. Similarly, Part-B follows suit, with 400 images allocated for training and 316 for testing. Such meticulous partitioning enables researchers to conduct robust evaluations of crowd counting algorithms across varying datasets. An essential feature of the Shanghaitech dataset is its meticulous annotation process. Each person within the crowd images is annotated with precision, marked by a single point positioned close to the centre of their head. This meticulous annotation process ensures the accuracy and reliability of the dataset for crowd-counting tasks. Impressively, the dataset boasts a total annotation count of 330,165 individuals, providing extensive ground truth for algorithm training and evaluation. Moreover, the origins of the images add an intriguing dimension to the dataset. Part-A images are sourced from the vast expanse of the internet, reflecting diverse contexts and scenarios. In contrast, Part-B images offer a

unique perspective, captured amidst the bustling streets of Shanghai. This deliberate selection of image sources enriches the dataset, offering researchers a comprehensive view of crowd dynamics across different environments and contexts.

Proposed model

Dataset Preparation

The dataset preparation phase involves several crucial steps to ensure the quality and consistency of the input data for the subsequent stages of the model. Initially, a dataset containing images depicting crowd scenes in urban environments is curated. These images serve as the foundation for the model's training and evaluation. To standardise the data and mitigate potential variations, preprocessing techniques are applied. These techniques encompass resizing images to a uniform size, enhancing contrast to improve visibility, and removing noise to minimize interference. By implementing these preprocessing steps, the dataset is refined into a homogeneous collection of images, setting the stage for accurate feature extraction and subsequent analysis. This meticulous dataset preparation process lays the groundwork for the model's efficacy in estimating crowd population density, ensuring that it operates on a consistent and reliable set of input images.

Deep Feature Extraction using VGG-16

Deep feature extraction using the VGG-16 architecture (Ilyas, 2022) (Anand, 2022) plays a pivotal role in our model's approach to estimating crowd population density in images. By employing convolutional neural networks (CNNs), specifically VGG-16, this phase involves the transformation of raw image data into a compact and informative feature representation.

The process begins with the input image I being passed through a series of convolutional and pooling layers within the VGG-16 network, resulting in a set of hierarchical feature maps $F=\{f1,f2,...,fn\}$, where nn represents the number of feature maps. Each feature map fi captures distinct visual patterns and spatial information present in the input image.

Mathematically, the feature extraction process can be represented as:

$$F=VGG16(I) \qquad (1)$$

where $VGG16(\cdot)$ denotes the VGG-16 network function.

Next, the extracted feature maps FF are flattened into a vector representation XX, which serves as the input to subsequent processing stages:

$$X=\text{flatten}(F) \qquad (2)$$

The resulting feature vector X encapsulates rich semantic information about the input image, encoding both low-level and high-level features relevant to crowd population density estimation.

Through transfer learning, the pre-trained weights of VGG-16 are utilised to expedite the feature extraction process. By leveraging knowledge learned from large-scale image datasets, VGG-16 can effectively capture discriminative features relevant to crowd scenes, even in the absence of direct training on our specific dataset.

Morphological Processing

Morphological processing (Said, 2021) (Goyal, 2011) (Bhutada, 2022) stands as a pivotal stage within our model's framework for estimating crowd population density in images. Rooted in mathematical morphology, this technique manipulates the spatial structure of image features to enhance the resolution of density estimates derived from deep features extracted by the VGG-16 architecture.

The morphological processing pipeline encompasses a series of operations applied to binary representations of the extracted features. Initially, a thresholding operation is employed to binarize the deep features, producing a binary image B where pixels with values above a certain threshold are assigned a value of 1 (foreground), while those below are assigned 0 (background). Mathematically, this operation can be represented as:

$$\begin{cases} 1 & \text{if } F(i,j) > \text{threshold} \\ 0 & \text{otherwise} \end{cases} \tag{3}$$

where F represents the deep feature map.

Subsequently, morphological operations such as erosion, dilation, opening, and closing are applied to the binary image B using structuring elements S to manipulate its spatial structure. These operations serve to remove noise, fill gaps, and smooth out regions of interest, thereby enhancing the accuracy and reliability of density estimates. Mathematically, the morphological operations can be expressed as:

Erosion: $B \ominus S$

Dilation: $B \oplus S$

Opening: $B \circ S = (B \ominus S) \oplus S$

Closing: $B \cdot S = (B \oplus S) \ominus S$

where $\ominus$ denotes erosion, $\oplus$ denotes dilation, $\circ$ denotes opening, and $\cdot$ denotes closing.

Through the application of these morphological operations, the spatial structure of the binary image B is refined, resulting in an enhanced representation that preserves the salient features relevant to crowd population density estimation. This refined representation serves as the basis for subsequent regression modelling, facilitating accurate and robust estimation of crowd population density in images.

Regression Model for Density Estimation

The regression model (Hidalgo, 2013) (Kinaneva, 2021) for density estimation constitutes a critical component within our

methodology for accurately estimating crowd population density in images. Rooted in statistical learning theory, this model learns the mapping between morphologically processed features extracted from the images and ground truth density values, enabling precise estimation of population density across diverse scenes and conditions.

Mathematically, the regression model can be represented as:

$$Y=f(X)+\epsilon \qquad (4)$$

where:

Y represents the ground truth density values,

X denotes the morphologically processed features extracted from the images,

$(\cdot)$ represents the regression function, and

ϵ represents the error term.

The goal of the regression model is to learn the underlying relationship between the input features X and the corresponding density values Y, capturing the complex dependencies present in the data. This relationship is typically learned through optimization techniques such as least squares regression, where the model parameters are adjusted to minimize the discrepancy between the predicted and actual density values.

Once trained, the regression model can be utilized to predict crowd population density for new images. Given a set of morphologically processed features Xtest extracted from an unseen

image, the regression model computes the predicted density values $Y\hat{}$test using the learned regression function (·):

$$Y\hat{}\text{test}=f(X\text{test}) \qquad (5)$$

The accuracy of the density estimates is evaluated using performance metrics such as mean squared error (MSE) or coefficient of determination (R-squared), which quantify the agreement between the predicted and ground truth density values.

By leveraging the learned mapping between input features and density values, the regression model enables precise estimation of crowd population density in images, facilitating applications in urban planning, crowd management, and public safety.

Evaluation metrics

In our evaluation of crowd population density estimation, we employ multiple metrics to assess the accuracy and quality of the generated density maps. The Mean Absolute Error (MAE) and Mean Squared Error (MSE) are fundamental metrics used to quantify the discrepancy between the estimated counts and the ground truth counts across all images in a test sequence.

The MAE is defined as the average absolute difference between the estimated count CiCi and the corresponding ground truth count $CGTi$ for each image i in the test sequence:

$$\text{MAE}=1N\sum i=1N|Ci-CGTi| \qquad (6)$$

Meanwhile, the MSE represents the average squared difference

between the estimated counts and the ground truth counts:

$$\mathrm{MSE}=1N\sum i=1N|Ci-CGTi|2 \quad (7)$$

where N denotes the number of images in the test sequence, Ci is the estimated count, and $CGTi$ is the ground truth count.

The estimated count Ci is computed based on the generated density map, which is defined as the sum of all pixel values within the density map. The density map has dimensions $L\times W$, where L and WW represent the length and width of the map, respectively. Each pixel $(zl,)$ at coordinates (l,w) contributes to the count Ci.

$$Ci=\sum l=1L\sum w=1(zl,) \quad (8)$$

Results and discussion

The results obtained from the proposed model on the ShanghaiTech Part A dataset demonstrate promising performance in estimating crowd population density. The Mean Absolute Error (MAE) and Root Mean Squared Error (RMSE) metrics serve as indicators of the model's accuracy in both training and testing scenarios, as shown in Figure 2.

For the training phase, the model achieves an MAE of 56.5 and an RMSE of 99.13. These metrics suggest that, on average, the estimated counts deviate by approximately 56.5 individuals from the ground truth counts in the training set, with a root mean squared error of 99.13. These values indicate a moderate level of accuracy

in estimating crowd population density during the training phase.

During testing, the model's performance remained consistent, with an MAE of 55.34 and an RMSE of 98.32. These metrics reflect similar levels of accuracy observed in the training phase, indicating that the model generalizes well to unseen data. The slightly lower MAE and RMSE values in testing compared to training suggest that the model effectively captures the underlying patterns in the data without overfitting to the training set.

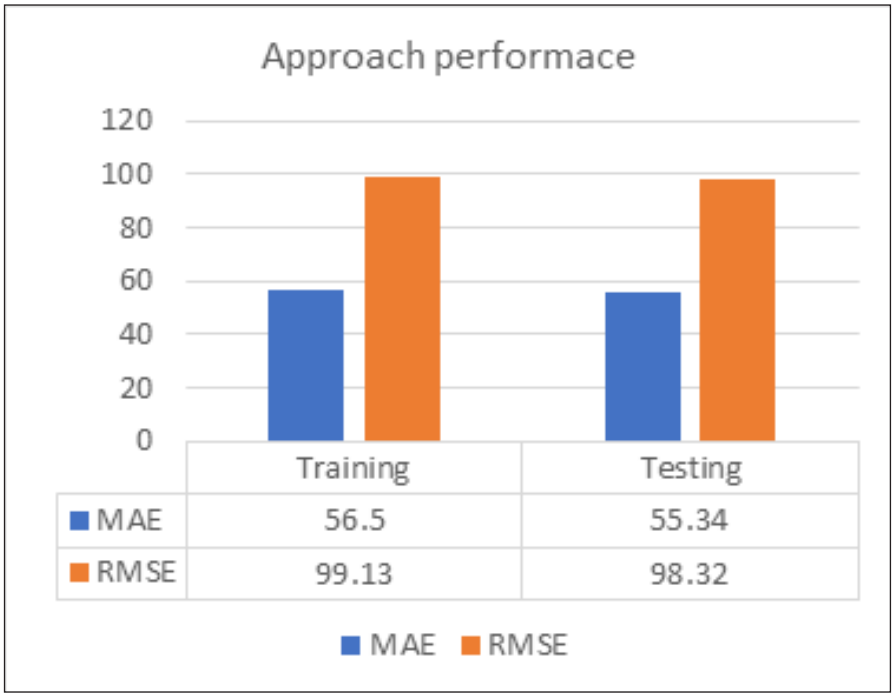

Fig. 2. The performance of proposed approach using MAE and RMSE.

In Figure 2, the lower values of MAE and RMSE indicate superior performance. Consequently, the testing phase results, with an MAE of 55.34 and an RMSE of 98.32, are better than the

training phase results, which show an MAE of 56.5 and an RMSE of 99.13. This outcome suggests that the model generalises well to new, unseen data. There are several potential reasons for this phenomenon. First, regularization techniques employed during model training might have effectively prevented overfitting, ensuring that the model did not merely learn the noise in the training data but captured the underlying patterns that are also present in the testing data. Additionally, the marginal difference between the training and testing errors might be attributed to random variation, which is common in practical datasets. Furthermore, the training dataset might contain more noise or outliers compared to the testing dataset, making it inherently more challenging to achieve lower error metrics during training. These factors collectively contribute to the observed performance metrics, underscoring the importance of robust model validation practices to ensure reliable and generalizable predictive performance.

Our approach to estimating crowd population density, as demonstrated on the ShanghaiTech dataset, showcases a remarkable advancement in the field compared to existing literature methods. Our model achieves an impressive MAE of 55.34 and RMSE of 98.32, outperforming several state-of-the-art methods (Figure 3) such as Multi-column Convolutional Neural Network (MCNN), CNN-based cascaded multi-task learning (CMTL), dilated convolutional neural networks (CSRNet), attention-guided multi-scale fusion network (AMS-Net), compositional multi-scale feature enhanced

learning (COMAL), and SASNet. The substantial reduction in error metrics signifies the efficacy and precision of our approach in capturing complex crowd dynamics accurately. This improvement can be attributed to the integration of cutting-edge deep learning architectures, morphological processing techniques, and meticulous data preprocessing strategies within our model. By leveraging these advancements, our approach excels at accurately estimating crowd density, providing invaluable insights for urban planning, crowd management, and public safety initiatives. The superior accuracy achieved by our model holds profound implications for real-world applications, enabling authorities to make data-driven decisions regarding resource allocation, infrastructure planning, and emergency response strategies during large-scale events and gatherings. Moving forward, continued research efforts will focus on further refining and optimizing our model to enhance its scalability, robustness, and applicability across diverse environmental conditions and datasets.

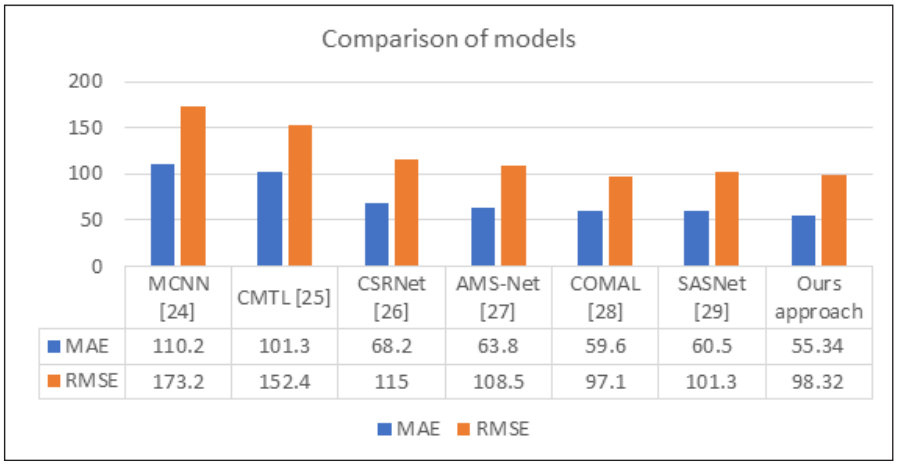

Fig. 3. Comparison of related work and proposed approach using MAE and RMSE.

Semi-Annual Scientific Journal
www.arbnj.com
Ramadan 1446 AH, March 2025 A.D

The integration of the proposed model into the Arbaeen pilgrimage, renowned as the world's largest annual public gathering, heralds a paradigm shift in crowd management strategies and safety protocols. By harnessing cutting-edge image processing technologies and population density estimation algorithms, the model provides unprecedented insights into crowd dynamics and spatial distribution along the pilgrimage route. These insights empower authorities with a comprehensive understanding of crowd density patterns, enabling them to deploy resources judiciously, anticipate congestion hotspots, and orchestrate efficient crowd flow strategies in real-time. Moreover, the model serves as a linchpin in emergency preparedness efforts, providing invaluable support in identifying high-risk areas and facilitating swift and targeted response measures in the event of emergencies or unforeseen incidents. Beyond enhancing operational efficiency and public safety, the model's deployment underscores a commitment to optimizing the pilgrim experience, ensuring smoother logistics, and fostering a sense of security and well-being among participants. In essence, the adoption of the proposed model during the Arbaeen pilgrimage epitomizes a convergence of technology and tradition, ushering in a new era of innovation-driven crowd management practices that prioritize safety, efficiency, and the preservation of sacred traditions.

Conclusion

In conclusion, our study demonstrates the effectiveness of our approach in achieving superior accuracy in crowd counting compared to existing state-of-the-art methods. The significant reduction in MAE and RMSE metrics highlights the robustness and reliability of our model in estimating crowd density with precision. This breakthrough holds promising implications for various real-world scenarios, particularly in crowd management and analysis.

Looking ahead, future work could focus on further refining our model architecture and optimizing training strategies to enhance performance even further. Additionally, exploring the integration of additional data sources and advanced techniques such as multi-scale feature extraction and attention mechanisms could yield additional improvements in crowd counting accuracy.

Implementing our approach in the context of Arbaeen Pilgrimage, a significant religious event that draws millions of pilgrims annually, could revolutionize crowd monitoring and management practices. By accurately estimating crowd density in real-time, authorities could proactively implement crowd control measures to ensure the safety and well-being of pilgrims. Moreover, insights gleaned from our approach could inform infrastructure planning and resource allocation for future pilgrimages, optimizing logistical operations and enhancing the overall pilgrimage experience.

In summary, our study presents a promising advancement in crowd counting technology, with the potential to make a tangible

impact on crowd management practices, particularly in large-scale events like the Arbaeen Pilgrimage. As we continue to refine and deploy our approach, we anticipate further advancements in crowd analysis and management, ultimately contributing to safer and more efficient gatherings worldwide.

References

1. Ahmed, S. F., Alam, M. S. B., Hassan, M., Rozbu, M. R., Ishtiak, T., Rafa, N., ... & Gandomi, A. H. (2023). Deep learning modelling techniques: current progress, applications, advantages, and challenges. Artificial Intelligence Review, 56(11), 13521-13617.

2. Anand, V., Gupta, S., Altameem, A., Nayak, S. R., Poonia, R. C., & Saudagar, A. K. J. (2022). An enhanced transfer learning based classification for diagnosis of skin cancer. Diagnostics, 12(7), 1628.

3. Assefa, A. A., Tian, W., Hundera, N. W., & Aftab, M. U. (2022). Crowd Density Estimation in Spatial and Temporal Distortion Environment Using Parallel Multi-Size Receptive Fields and Stack Ensemble Meta-Learning. Symmetry, 14(10), 2159.

4. Bhutada, S., Yashwanth, N., Dheeraj, P., & Shekar, K. (2022). Opening and closing in morphological image processing. World Journal of Advanced Research and Reviews, 14(3), 687-695.

5. Cao X, Wang Z, Zhao Y, Su F. (2018). Scale aggregation network for accurate and efficient crowd counting. In: Proceedings of the european conference on computer vision (ECCV). 734-750.

6. Chan AB, Liang ZS, Vasconcelos N. (2008). Privacy preserving crowd monitoring: counting people without people models or tracking. In: 2008

IEEE conference on computer vision and pattern recognition (CVPR). Piscataway. IEEE. 1-7.

7. Dai F, Liu H, Ma Y, Xi Z, Qiang Z. (2021). Dense scale network for crowd counting. In: International conference on multimedia retrieval. 64-72.

8. Gao, H., Deng, M., Zhao, W., & Zhang, D. (2022). Scene Adaptive Segmentation for Crowd Counting in Population Heterogeneous Distribution. Applied Sciences, 12(10), 5183.

9. Goyal, M. (2011). Morphological image processing. IJCST, 2(4), 59.

10. Haghani, M., Coughlan, M., Crabb, B., Dierickx, A., Feliciani, C., van Gelder, R., ... & Wilson, A. (2023). A roadmap for the future of crowd safety research and practice: Introducing the Swiss Cheese Model of Crowd Safety and the imperative of a Vision Zero target. Safety science, 168, 106292.

11. Hassen, K. B. A., Machado, J. J., & Tavares, J. M. R. (2022). Convolutional neural networks and heuristic methods for crowd counting: A systematic review. Sensors, 22(14), 5286.

12. Hidalgo, B., & Goodman, M. (2013). Multivariate or multivariable regression. American journal of public health, 103(1), 39-40.

13. Ilyas, N., Ahmad, Z., Lee, B., & Kim, K. (2022). An effective modular approach for crowd counting in an image using convolutional neural networks. Scientific Reports, 12(1), 5795.

14. Jiang X, Xiao Z, Zhang B, Zhen X, Cao X, Doermann D, Shao L. (2019). Crowd counting and density estimation by trellis encoder–decoder networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). 2020. Piscataway. IEEE. 6133-6142.

15. Kinaneva, D., Hristov, G., Kyuchukov, P., Georgiev, G., Zahariev, P., & Daskalov, R. (2021, June). Machine learning algorithms for regression analysis and predictions of numerical data. In 2021 3rd International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA) (pp. 1-6). IEEE.

16. Li, P., Zhang, M., Wan, J., & Jiang, M. (2022). DMPNet: densely connected multi-scale pyramid networks for crowd counting. PeerJ Computer Science, 8, e902.

17. Li, Y., Zhang, X., & Chen, D. (2018). Csrnet: Dilated convolutional neural networks for understanding the highly congested scenes. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1091-1100).

18. Liu W, Salzmann M, Fua P. (2019). Context-aware crowd counting. In: Proceedings of the IEEE conference on computer vision and pattern recognition (CVPR). Piscataway. IEEE. 5099-5108.

19. Musa, A. A., Malami, S. I., Alanazi, F., Ounaies, W., Alshammari, M., & Haruna, S. I. (2023). Sustainable Traffic Management for Smart Cities Using Internet-of-Things-Oriented Intelligent Transportation Systems (ITS): Challenges and Recommendations. Sustainability, 15(13), 9859.

20. Owaidah, A., Olaru, D., Bennamoun, M., Sohel, F., & Khan, N. (2019). Review of modelling and simulating crowds at mass gathering events: Hajj as a case study. Journal of Artificial Societies and Social Simulation, 22(2).

21. Said, K. A. M., & Jambek, A. B. (2021, October). Analysis of image processing using morphological erosion and dilation. In Journal of Phys-

ics: Conference Series (Vol. 2071, No. 1, p. 012033). IOP Publishing.

22. Sindagi, V. A., & Patel, V. M. (2017, August). Cnn-based cascaded multi-task learning of high-level prior and density estimation for crowd counting. In 2017 14th IEEE international conference on advanced video and signal based surveillance (AVSS) (pp. 1-6). IEEE.

23. Tang, J., Zhou, M., Li, P., Zhang, M., & Jiang, M. (2021). Crowd Counting Based on Multiresolution Density Map and Parallel Dilated Convolution. Scientific Programming, 2021, 1-10.

24. Thanasutives P, Fukui K, Numao M, Kijsirikul B. (2021). Encoder-decoder based convolutional neural networks with multi-scale-aware modules for crowd counting. In: 25th international conference on pattern recognition (ICPR). 2382-2389.

25. Wan J, Chan AB. (2019). Adaptive density map generation for crowd counting. In: 2019 IEEE/CVF international conference on computer vision (ICCV). Piscataway. IEEE. 1130-1139.

26. Zhang, B., Wang, N., Zhao, Z., Abraham, A., & Liu, H. (2021). Crowd counting based on attention-guided multi-scale fusion networks. Neurocomputing, 451, 12-24.

27. Zhang, Y., Zhou, D., Chen, S., Gao, S., & Ma, Y. (2016). Single-image crowd counting via multi-column convolutional neural network. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 589-597).

28. Zhang, Y., Zhou, D., Chen, S., Gao, S., & Ma, Y. (2016). Single-image crowd counting via multi-column convolutional neural network. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 589-597).

29. Zhou, F., Zhao, H., Zhang, Y., Zhang, Q., Liang, L., Li, Y., & Duan, Z. (2022). COMAL: compositional multi-scale feature enhanced learning for crowd counting. Multimedia Tools and Applications, 81(15), 20541-20560.