A Human Face Detector for Big Data Analysis of Pilgrim Flow Rates in Hajj and Umrah

Riad Alharbey
Department of Information Systems and Technology, College of Computer Science and Engineering, University of Jeddah, Saudi Arabia
ralharbi@uj.edu.sa

Ameen Banjar
Department of Information Systems and Technology, College of Computer Science and Engineering, University of Jeddah, Saudi Arabia
abanjar@uj.edu.sa

Yahia Said
Department of Electrical Engineering, College of Engineering, Northern Border University, Saudi Arabia
said.yahia1@gmail.com (corresponding author)

Mohamed Atri
College of Computer Sciences, King Khalid University, Saudi Arabia
matri@kku.edu.sa

Mohamed Abid
CES Laboratory, ENIS, University of Sfax, Tunisia
med.abid@enis.tn

Received: 24 November 2023 | Revised: 7 December 2023 | Accepted: 8 December 2023

Licensed under a CC-BY 4.0 license | Copyright (c) by the authors | DOI: https://doi.org/10.48084/etasr.6668

ABSTRACT

In today's digital world, some crowded venues still rely on outdated methods, such as counting people using counters or sensors at the entrance. These techniques generally fail in areas where people move randomly. Crowd management is an important challenge for ensuring human safety. This paper focuses on developing a crowd management system for Hajj and Umrah duty. Motivated by the recent artificial intelligence techniques and the availability of large-scale data, a crowd management system was established and is presented in this paper. Utilizing the most recent Deep Learning techniques, the proposed crowd management system will be charged with detecting human faces, face identification, tracking, and human face counting tasks. Face counting and detection will be achieved by computing the number of people in a given area. Face detection and tracking will be carried out for person identification, flow rate estimation, and security. The suggested crowd management system is composed of three key components: (1) face detection, (2) assignment of a specific identifier (ID) to each detected face, (3) each detected face will be compared to the stored faces in the dataset. If the detected face is identified, it will be assigned to its ID, or a new ID will be assigned. The crowd management system has been developed to improve the Cross-Stage Partial Network (CSPNet) with attention module integration. An attention module was employed to address object location challenges and a channel-wise attention module for determining the objects of focus. Extensive experiments on the WIDER FACE dataset proved the robustness of the proposed face detection module, which allows for building reliable crowd management and flow rate estimation systems through detecting, tracking, and counting human faces. The reported results demonstrated the power of the proposed method while achieving high detection performance in terms of processing speed and detection accuracy.

Keywords-crowd management; face counting; face tracking; deep learning; attention module; channel-wise module; Hajj; Umrah

I. INTRODUCTION

Hajj and Umrah are considered the pillars that each Muslim hopes to perform at least once in life. All Muslims travel to Mecca and Madinah to perform the Hajj and Umrah rituals, which form some of the most populous human gatherings as millions of persons share the same place simultaneously. The Saudi government works to guarantee greater safety and to develop new crowd management techniques to maintain security for pilgrims when they perform their religious duties. Hajj and Umrah present a significant concern for the kingdom as it expects to receive more than 30 million Hajj and Umrah performers by 2030. An efficient crowd management system is essential to control place access and the departure from them and to reduce crowds from congested spots. In this way, the system certifies higher safety and security for the attendants when performing their religious duties. As a very high number of persons per-place are expected, an efficient crowd management system is essential for fast management and flow rate estimation. A reliable crowd management system will be designed using Artificial Intelligence (AI) systems for face detection, tracking, and counting. Human face detection and counting are employed in a crowd management system to quantify the number of people in a given area, whereas face recognition is used for security and health concerns. The main purpose of building a new effective crowd management system is to ensure a safer and more comfortable way for persons to move around overcrowded places. Generally, the system will be charged with the control of people's entry and exit to manipulate crowd flow. The developed system ensures the detection of dangerous situations and can call for specific teams' intervention to address these problems.

In Hajj and Umrah, people move continuously, affecting crowd management accuracy. To address this issue, an additional face detection part was introduced for the crowd management system to identify indivinduals by personal identifying IDs and to track them in overcrowded places. By adding a face recognition framework, we ensure the entry of authorized people to the crowded area and avoid multiple counts of persons contributing to more precise statistics. The two critical components of the suggested crowd control system are data processing and the identification storage framework. First, data collected from cameras are processed for human face detection and identification. In the data storage module, the identified faces will be stored in the dataset for further face recognition and to avoid multiple counting of the same faces.

Deep Learning (DL)-based techniques present a wide range of applications, including indoor object detection and recognition [1-4], wayfinding assistance navigation [5], license plate detection [6], traffic sign detection [7], logo recognition [8], face recognition [9], and pedestrian detection [10].

Channel-wise attention and spatial attention mechanisms are utilized in DL architectures, particularly in Convolutional Neural Networks (CNNs), to enhance the ability to focus on important features and regions within an input. Channel-wise attention allows the network to assign different weights to each channel of the input feature map. This is beneficial when certain channels capture more relevant information for a given task. The model can learn to emphasize important channels while suppressing less informative ones. Channel-wise attention helps the network capture complex relationships between channels, facilitating the learning of non-linear combinations of features. This can be particularly useful in tasks where understanding inter-channel dependencies is crucial. Spatial attention mechanisms enable the model to focus on specific spatial regions within the input feature map. This is valuable when certain parts of the input contain more relevant information for the task at hand. The network can learn to attend discriminative regions. Such spatial attention can reduce the computational cost by focusing on informative regions, allowing the model to allocate resources more efficiently. This is especially important in scenarios where computational resources are limited, such as in real-time applications. Motivated by the power of DL models and the benefits of attention mechanisms, the present work proposes a crowd management system with a face detection framework based on the Cross-Stage Partial Network CSPNet [11]. Enhanced by attention modules, this system ensures an optimized module with high detection accuracy and processing speed. We used attention modules to enable the network to focus better on the target. CSPNet is proposed to reduce the computational cost and the network memory storage without decreasing the network accuracy. By employing the CSPNet, we ensure real-time processing with low computational complexity on low-end devices. To improve accuracy and reduce false-positive predictions, channel-wise and spatial attention modules were applied to the CSPNet at the backbone and prediction stages. The WIDER FACE [12] dataset was used to train and test the suggested model. This dataset provides various challenging conditions, including occlusion, pose variation, varying light conditions, and geometric deformations. The most significant part of the dataset has been collected from overcrowded areas. This makes our work very efficient and effective as it is trained in challenging conditions.

II. RELATED WORK

Various works have been proposed in the literature to address the problem of crowd management and control. Authors in [13] proposed a new simulation approach for real-life crowd management based on DL algorithms. The system can identify the crowd level into five main categories. The developed system comprises two main parts: one for crowd-level classification and the second for constituting the color of warning. Crowd monitoring presents an extensively researched topic as it is primarily related to public safety. This topic is highly challenging as it addresses problems such as occlusion, pose estimation, and density variation. In [14], the authors reviewed the crowd management works by presenting the latest developments and performances obtained in crowd management and control using machine learning techniques.

Most crowd management systems still rely on old methods, such as engaging persons to count at places' entrance. However, these methods fail in crowded places where the movement is random and variable. In [15], the authors introduce the use of Deep Convolutional Neural Networks (DCNNs). The system implementation has been performed employing NVIDIA Graphic Processing Unit (GPU) to parallelize the process and obtain near-real-time results. Authors in [16] proposed a crowd management system based on image classification and an alarm module. The image classification module relies on using CNNs. In [17], a new crowd management system was suggested dealing with high-level architecture with decision support. Authors in [18] review and summarize the strengths and applications of crow management systems in different crowd scenarios. Authors in [19] introduced a new efficient DL model called deep Uran Event. The model is built utilizing Recurrent Neural Networks (RNNs).

The world urbanization prospects in its 2018 revision estimate that 68% of the world population will have been living in urban spaces by 2050. Such an increase in the urbanization flow poses significant pressure on the city and space infrastructures. This fact poses serious problems and a crowd management system would be beneficial in tackling some of those issues. Using many sensors and cameras with limited human intervention to analyze videos, results in developing new effective crowd-monitoring systems. In [20], the authors discuss building an intelligent crowd management system applying condition capabilities. For intelligent systems, detecting and predicting abnormal crowd behavior has emerged as a challenging task in building efficient crowd control systems. In [21], the authors propose a new approach that exploits the psychological state and cognition to detect nine crowd behaviors. The developed approach is based on two cognitive DL-based frameworks with a psychological fuzzy computation framework.

Real-time crowd control and analysis present an active research area within the computer vision and AI community. Various methods for crowd management have been proposed to be implemented for different applications, such as people counting [22], disaster management [23], and event management [24]. Although various works have been suggested to address the problem of crowd management and control, few of them fit real-time conditions. As it is still a challenging matter to be solved, authors in [25] reviewed the currently available crowd management analysis techniques.

III. THE PROPOSED APPROACH

The proposed pipeline followed in this work to build the crowd management system is presented in Figure 1. A camera will capture and collect data from the surrounding environment as a system input. The system will then be charged by conducting face detection, and it will later focus solely on the face cropping in the input data. The saved data will be then compared with each cropped face. If matched faces exist, they will be assigned to their identifier (ID). If not, a new identifier will be assigned to the cropped face. Finally, the cropped face identifiers will be stored in the dataset for further face identification and tracking. The face detection task is highly challenging due to different factors, including lighting, small face sizes, occlusion, different viewpoints, geometrical deformations, etc. So, to overcome all these issues, it is critical to design a robust crowd management system. We propose to build a crowd management system based on a Cross-Stage Partial Network (CSPNet) and two powerful attention blocks. A spatial attention module to address object location challenges and a channel-wise attention module for determining the objects of focus were employed. We note that the applied techniques enhanced the system detection performance and processing speed. Generally, DCNNs require substantial computational resources and high-end devices for application implementations. CSPNet was introduced to resolve this problem as it provides a lightweight architecture that can contribute to real-time applications using mobile devices.

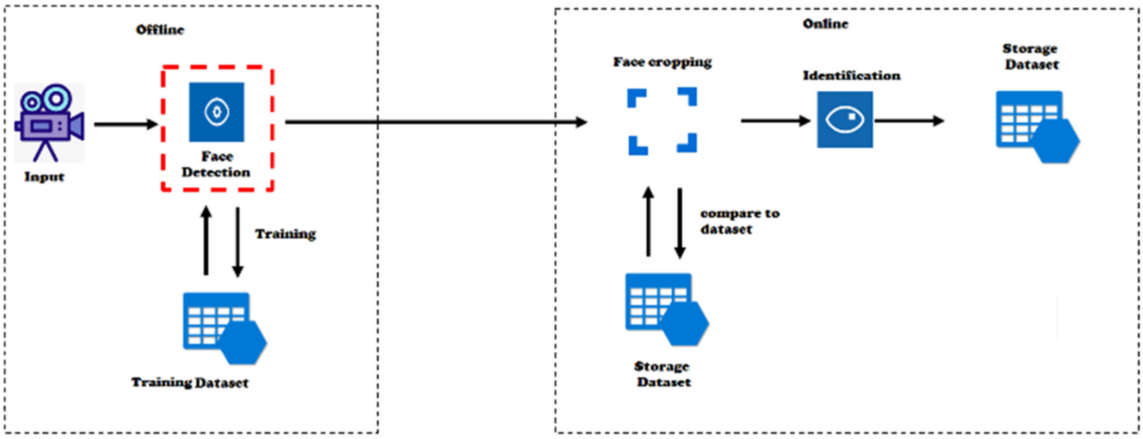

Fig. 1. Proposed crowd management system

CSPNet introduces an optimized version of DenseNet [26]. The basic architecture of DenseNet is based on stages. Every stage contains a dense block containing K dense layers. The DenseNet mechanism is described by:

$$x_1 = W_1 * x_0$$

$$x_2 = W_2 * [x_0, x_1]$$

$$\vdots$$

$$x_K = W_K * [x_0, x_1,\ldots,x_{K-1}] \qquad (1)$$

where * means a convolution operator, $[x_0, x_1,\ldots,x_{K-1]}$ means a concatenation of x_0, x_1, $\ldots x_{K-1}$, W_i is the weight of each layer, and x_i is the output of the i^{th} dense layer. To reduce the error between the network output and the target output, the DCNN uses backpropagation techniques to update the network weights. Weight updating in the DenseNet architecture is computed as:

$$W_1' = f\,(w_1, g_0)$$

$$W_2' = f\,(w_2, g_0, g_1)$$

$$W_3' = f\,(w_3, g0, g1, g2)$$

$$\vdots$$

$$W_K' = f\,(w_K, g_0, g_1, \ldots, g_{K-1}) \qquad (2)$$

where f is the function of weight updating, g_i is the gradient propagation to the i^th^ dense layer.

This approach will produce various dense layers that will repeatedly learn duplicated gradient information. This task presents an issue as it increases the network computational cost. A cross-stage partial network is introduced to limit the computational requirements. CSPNet ensures a richer gradient combination while keeping low computational costs. It achieves this by dividing the gradient to follow different network paths. CSPNet employs cross-stage partial DenseNet as a network backbone. As mentioned in Figure 2, the base layer feature map is split into two paths: [x_0', x_{00}'']. A deep concatenation of x_0', x_{00}''results in the x_0 original feature map. The transition layer output x_T will be a function of the dense layer [x_{00}'', x_1', …, x_K]. x_T will be processed with x_0 and undergo another transition layer to result in x_U. The feed-forward pass can be computed with (3) and the weight updating of the cross-stage partial DenseNet is computed in (4).

$$x_k = W_K * [x_0", x_1, \ldots, x_{K-1}]$$

$$x_T = W_T * [x_0", x_1, \ldots, x_K]$$

$$x_U = W_U * [x_0", x_T] \qquad (3)$$

$$W_K' = f(W_K, g_0", g_1,g_2,\ldots,g_{K-1}]$$

$$W_T' = f(W_T, g_0", g_1,g_2,\ldots,g_K$$

$$W_U' = f(W_u, g_0', g_T] \qquad (4)$$

The CSPDenseNet architecture is presented in Figure 2. It generally reuses the original DenseNet characteristics while avoiding multiple gradient information reuses by dividing the gradient into two paths. The main advantages of splitting the feature map are the following:

- It doubles the gradient paths during backpropagation, which makes it deal with the gradient vanishing problem.
- Deeper depth of the feature map at the base layer whereas the division of the feature map reduces considerably the computation cost.

In our work, the CSPDenseNet architecture fusion was employed first: the merge is performed first, then the transition. As a network for object detection, CSPNet uses an Exact Fusion Model (EFM), as shown in Figure 3.

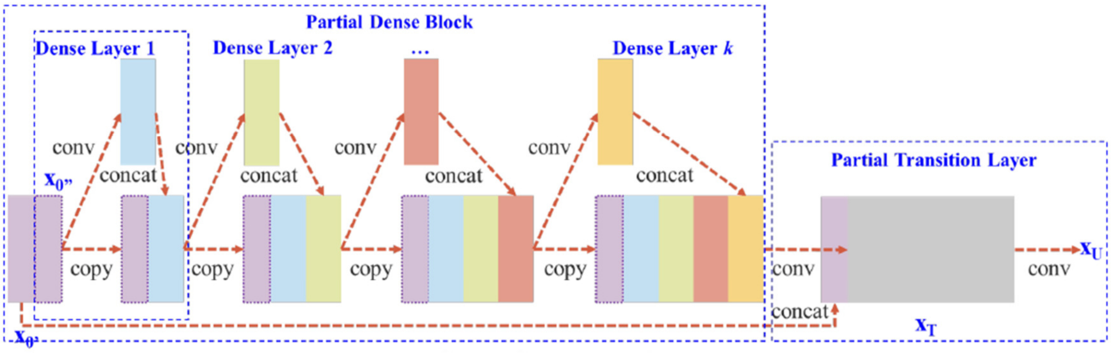

Fig. 2. CSPDenseNet Architecture.

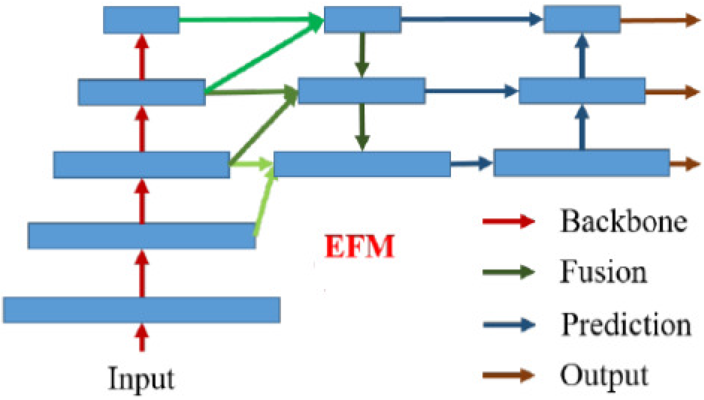

Fig. 3. EFM architecture for object detection. Blue rectangles represent pyramid feature maps. The green arrows represent the fusion of feature maps at different levels.

To detect and localize objects, CSPNet architecture introduces the use of EFM, which captures the most relevant components in a Field of View (FoV) for each anchor. This method is especially employed for aggregating the feature pyramid. This technique is based on assigning each ground truth object with just one prior bounding box. Only one anchor box that surpasses the IOU threshold refers to a ground truth object. Generally, the anchor box is the same size as a grid cell's FoV view. The corresponding bounding box for the i^{th} scale is lower by the $(i-1)^{th}$ scale and upper bounded by the $(i+1)^{th}$ scale. So, the EFM mind is to assemble the most relevant features from three different scales. To minimize the amount of computation complexity and memory cost, CSPNet introduces the max-out techniques. These techniques are utilized to compress the feature maps and address these problems. An attention module was proposed to enhance the model's focus on human faces and ensure greater effectiveness for the detection process. The channel-wise module is presented to address the issue of what to focus on, while the spatial attention module is utilized to address the issue of where to focus. In the proposed work, a combination of the two modules was used for better performance. Attention modules are designed using pooling layers. However, we propose designing an attention module based on fully convolution layers in this work. We note that by replacing pooling by convolution layers [27], we ensure better learning capabilities of the model. Figure 4 depicts the architecture of the spatial attention module.

The 1×1 convolution layer has been used as a squeezing unit. Also, it prevents the backpropagation influence directly on the backbone. The 3×3 convolution layer is utilized as a 3×3 excitation unit followed by an 1×1 convolution. These are the main components of the spatial attention mechanism. The spatial attention map's relative location and receptive field are the same as the backbone output. In the proposed work, the spatial attention module was integrated in a plug-in manner into the backbone architecture. The second attention module applied in our work is the channel-wise attention module. Figure 5 illustrates the designed channel-wise attention module.

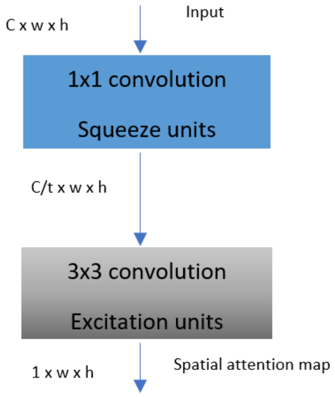

Fig. 4. Spatial attention module architecture.

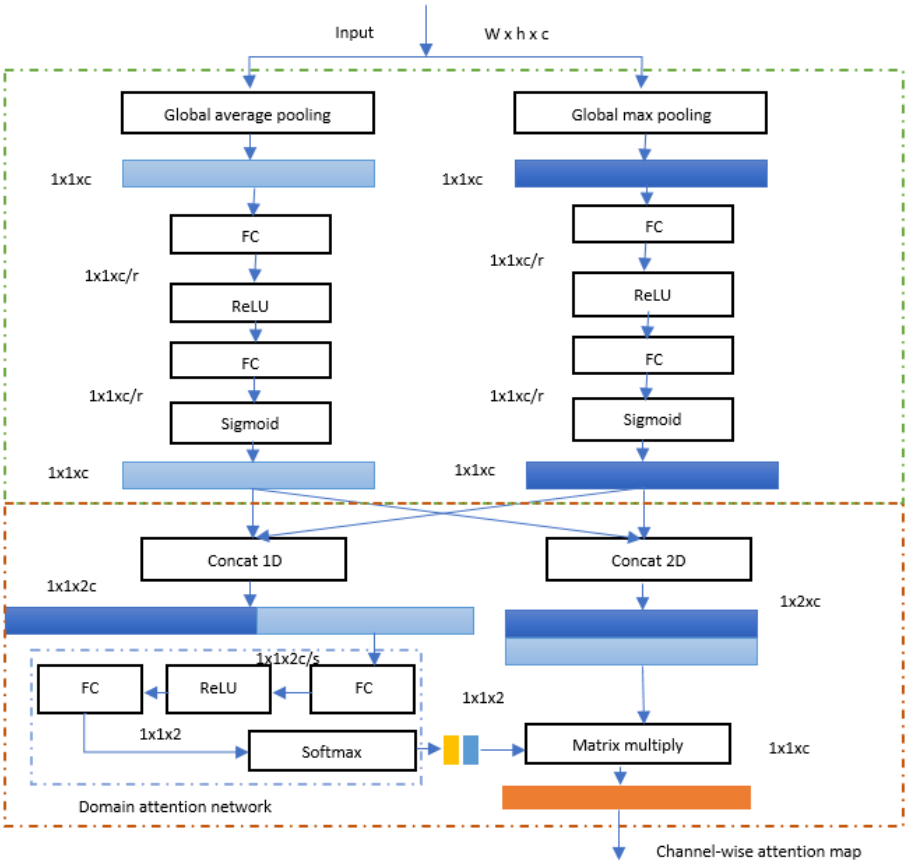

Fig. 5. Channel-wise attention module.

The utilization of distinct pooling layers is critical for the attention process. Using a global average pooling layer allows for object expansion, but using a max-pooling layer assures object location identification. Also, we provide improved identification of small items by employing global max pooling rather than max pooling.

In the current work, human face detection was solved at different aspect ratios to achieve more efficient results. However, only mixed max pooling and global average pooling were adopted for building channel-wise attention. The main novelty in the proposed work is the design of an additional domain attention network to the pooling structure. Three main rules must be respected in the design of the attention domain network: The network must be data-driven, possibly adapting the intermediate feature map and outputs to the input. Also, the network should be powerful to weigh the raw vectors. Finally, the network must be lightweight to reduce the computation complexity. A hidden layer and three fully connected layers are the main components of the proposed domain attention network. We should mention that the suggested attention module was included to keep the network structure and benefit from the pre-trained weights. The spatial module was integrated into the dense layers, while the channel-wise module was integrated into the dense layers after the spatial module application. As the top layers present high semantic information with low spatial features, applying the channel-wise module at this stage was very critical.

IV. EXPERIMENTAL RESULTS

A. Experimental Environment and Evaluation

In this work, we develop a robust crowd management system highly recommended in overcrowded areas, especially during Hajj and Umrah duty. Training and testing experiments have been performed using the WIDER FACE dataset [12], which provides various challenging conditions. All the data were manually annotated and collected from overcrowded events, which match the studied task. All the experiments conducted in this work were carried out on a desktop with an i7 CPU, 32 GB of RAM, and an NVIDIA GTX 960 GPU. The DL framework Pytorch with CUDA support and the CUDNN library for processing acceleration were used to perform the proposed experiments. The Open Computer Vision (OpenCV) library was employed for image display and manipulation. The DCNN was trained for 40 epochs. We utilized 0.001 as an initial learning rate and ADAM as the network optimizer. The algorithm is used to optimize the gradient descent variant and accelerate the convergence process. Training and testing image sizes were fixed to 320×320 pixels to achieve real-time performance. The batch size was 4. Loss optimization provides information on how well the training is. If training goes well, the loss will converge to a worthy minimum. The loss optimization curve of the proposed model for classification and localization is presented in Figure 6. To evaluate the proposed work, mean Average Precision (mAP), processing speed, and model size were considered. To study the effectiveness of the integrated improvements applied to the original CSPNet, the original and the improved version of CSPNet were evaluated on the same dataset. Table I provides the obtained results.

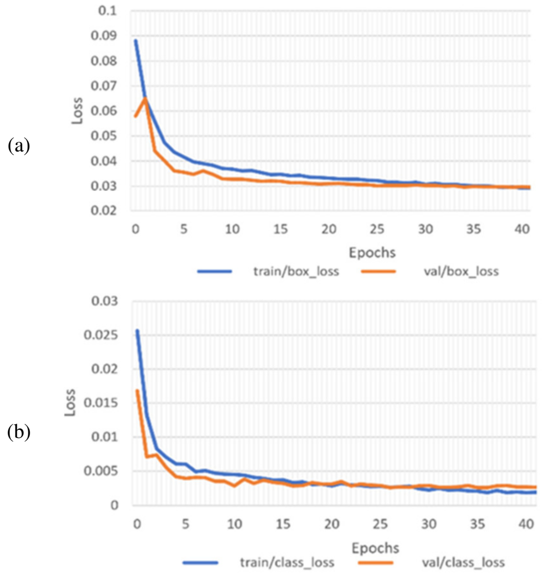

Fig. 6. Loss optimization of the proposed model for classification (class_loss) and localization (box_loss).

TABLE I. OBTAINED PERFORMANCE OF THE PROPOSED AND THE ORIGINAL CSPNET VERSION

Model	mAP (%)	Speed (FPS)	Model size (MB)
CSPNet (original)	91. 25	19	216.7
CSPNet (ours)	93.36	18	227.3

We note that by applying the integration of the attention module, the network detection accuracy was improved. To further study the efficiency and robustness of the proposed work, a comparison with the state-of-the-art works in terms of detection accuracy and speed was proposed. Based on the obtained results, our work outperforms previous works and achieves a satisfactory trade-off between detection precision and speed. Table II provides a comparison with the state-of-the-art techniques.

TABLE II. COMPARISON WITH STATE-OF-THE-ART

Model	mAP (%)	Speed (FPS)
Faster RCNN [28]	88.7	4
[29]	89.1	10
HOANG [30]	75.4	12
CSPNet (ours)	93.36	18

As mentioned in Table II, the proposed crowd management system achieved reliable detection results which outperform those obtained by the other methods regarding both detection and processing time. The output of the face detection algorithm is illustrated in Figure 7. The proposed algorithm presents a high detection rate even for tiny and occluded faces.

B. Ablation Study

An ablation study was conducted to show the impact of different attention modules. The spatial and channel-wise attention modules were adopted to enhance the model's performance in detecting small targets. Table III summarizes the achieved results with different configurations compared to the original model. The spatial attention module greatly improves the detection accuracy with only 0.69% and does not affect the inference speed. The channel-wise attention provides a favorable improvement of 1.32% in detection accuracy, while a slight decrease in inference speed was presented. In the case of adopting both attention modules, an enhancement of 2.11% was achieved with a decrease of only one FPS in inference speed. Both attention modules require 10.6 MB of additional memory. The achieved improvement proved the efficiency of the proposed attention module for enhancing the detection accuracy without degrading the inference speed and without a considerable increase in computation complexity.

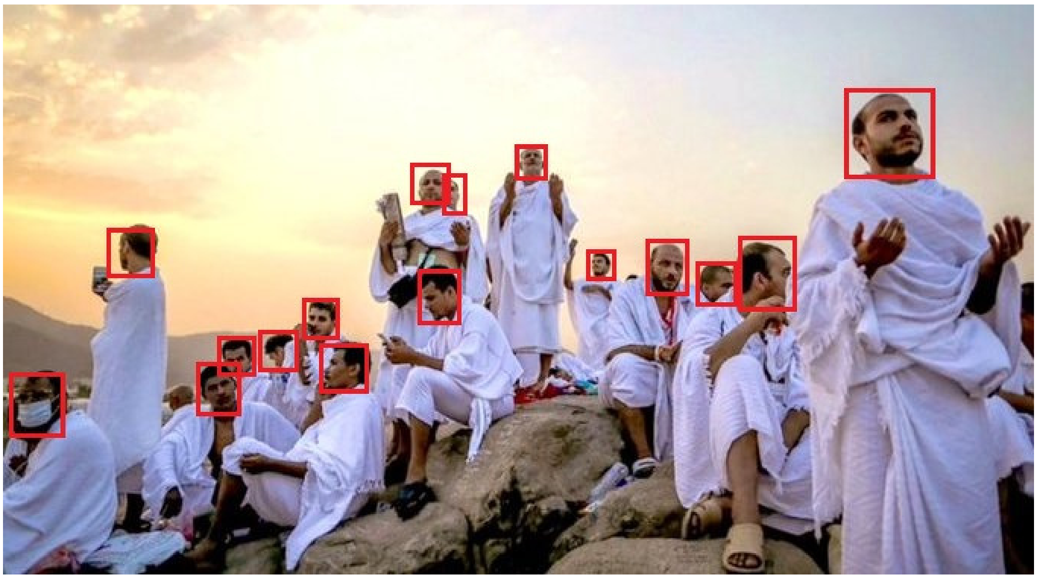

Fig. 7. Demo of the face detection algorithm.

TABLE III. ABLATION STUDY ON THE ATTENTION MODULES

Model	mAP (%)	Speed (FPS)	Model size (MB)
CSPNet (original)	91. 25	19	216.7
CSPNet with spatial attention	91.94	18.7	220.5
CSPNet with channel-wise attention	92.57	18.4	223.2
CSPNet with both attentions	93.36	18	227.3

V. CONCLUSION

Managing large crowds during Hajj and Umrah remains a formidable challenge. The most effective solution for handling crowded spaces involves the development of crowd management systems utilizing human face detection, as opposed to the entire body, due to its efficiency and time-saving benefits. This study introduces an innovative crowd management system that leverages CSPNet with a DenseNet backbone. To enhance the system, attention modules were incorporated, including a spatial attention module to address object location challenges and a channel-wise attention module for determining the objects of focus. Extensive experiments have showcased the system's efficiency, demonstrating superior results in terms of both detection accuracy and speed when compared to state-of-the-art approaches. The significance of using CSPNet with attention mechanisms extends to its practical implementation in real-world crowd management applications. The robustness, adaptability, and high performance demonstrated by the system make it a promising candidate for deployment in settings such as large events, public spaces, and religious gatherings. It is important to note that the proposed system requires high-performance devices for optimal functionality. In future endeavors, the proposed approach will be implemented in a real-world application and evaluated under genuine conditions.

ACKNOWLEDGMENT

This work was funded by the University of Jeddah, Jeddah, Saudi Arabia, under grant No. (UJ-21-ICL-4). The authors thank the University of Jeddah for the provided technical and financial support.

REFERENCES

[1] M. Afif, R. Ayachi, E. Pissaloux, Y. Said, and M. Atri, "Indoor objects detection and recognition for an ICT mobility assistance of visually impaired people," *Multimedia Tools and Applications*, vol. 79, no. 41, pp. 31645–31662, Nov. 2020, https://doi.org/10.1007/s11042-020-09662-3.

[2] J. Niu, Q. Hu, Y. Niu, T. Zhang, and S. Kumar Jha, "Real-Time Recognition and Location of Indoor Objects," *Computers, Materials & Continua*, vol. 68, no. 2, pp. 2221–2229, 2021, https://doi.org/10.32604/cmc.2021.017073.

[3] A. C. Hernandez, C. Gomez, J. Crespo, and R. Barber, "Object Detection Applied to Indoor Environments for Mobile Robot Navigation," *Sensors*, vol. 16, no. 8, Aug. 2016, Art. no. 1180, https://doi.org/10.3390/s16081180.

[4] L. Jiang, W. Nie, J. Zhu, X. Gao, and B. Lei, "Lightweight object detection network model suitable for indoor mobile robots," *Journal of Mechanical Science and Technology*, vol. 36, no. 2, pp. 907–920, Feb. 2022, https://doi.org/10.1007/s12206-022-0138-2.

[5] C. Prandi, B. R. Barricelli, S. Mirri, and D. Fogli, "Accessible wayfinding and navigation: a systematic mapping study," *Universal*

Access in the Information Society, vol. 22, no. 1, pp. 185–212, Mar. 2023, https://doi.org/10.1007/s10209-021-00843-x.

[6] M. Salemdeeb and S. Erturk, "Multi-national and Multi-language License Plate Detection using Convolutional Neural Networks," *Engineering, Technology & Applied Science Research*, vol. 10, no. 4, pp. 5979–5985, Aug. 2020, https://doi.org/10.48084/etasr.3573.

[7] J. Cao, J. Zhang, and X. Jin, "A Traffic-Sign Detection Algorithm Based on Improved Sparse R-cnn," *IEEE Access*, vol. 9, pp. 122774–122788, 2021, https://doi.org/10.1109/ACCESS.2021.3109606.

[8] A. Alsheikhy, Y. Said, and M. Barr, "Logo Recognition with the Use of Deep Convolutional Neural Networks," *Engineering, Technology & Applied Science Research*, vol. 10, no. 5, pp. 6191–6194, Oct. 2020, https://doi.org/10.48084/etasr.3734.

[9] Y. Said, M. Barr, and H. E. Ahmed, "Design of a Face Recognition System based on Convolutional Neural Network (CNN)," *Engineering, Technology & Applied Science Research*, vol. 10, no. 3, pp. 5608–5612, Jun. 2020, https://doi.org/10.48084/etasr.3490.

[10] R. Ayachi, M. Afif, Y. Said, and A. B. Abdelaali, "Pedestrian detection for advanced driving assisting system: a transfer learning approach," in *5th International Conference on Advanced Technologies for Signal and Image Processing*, Sousse, Tunisia, Sep. 2020, pp. 1–5, https://doi.org/10.1109/ATSIP49331.2020.9231559.

[11] C.-Y. Wang, H.-Y. Mark Liao, Y.-H. Wu, P.-Y. Chen, J.-W. Hsieh, and I.-H. Yeh, "CSPNet: A New Backbone that can Enhance Learning Capability of CNN," in *IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, Seattle, WA, USA, Jun. 2020, pp. 1571–1580, https://doi.org/10.1109/CVPRW50498.2020.00203.

[12] S. Yang, P. Luo, C. C. Loy, and X. Tang, "WIDER FACE: A Face Detection Benchmark," in *IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, Jun. 2016, pp. 5525–5533, https://doi.org/10.1109/CVPR.2016.596.

[13] I. T. Nafea, "Simulation of crowd management using deep learning algorithm," *International Journal of Web Information Systems*, vol. 17, no. 4, pp. 321–332, Jan. 2021, https://doi.org/10.1108/IJWIS-04-2021-0045.

[14] A. Khan, J. Ali Shah, K. Kadir, W. Albattah, and F. Khan, "Crowd Monitoring and Localization Using Deep Convolutional Neural Network: A Review," *Applied Sciences*, vol. 10, no. 14, Jan. 2020, Art. no. 4781, https://doi.org/10.3390/app10144781.

[15] U. Bhangale, S. Patil, V. Vishwanath, P. Thakker, A. Bansode, and D. Navandhar, "Near Real-time Crowd Counting using Deep Learning Approach," *Procedia Computer Science*, vol. 171, pp. 770–779, Jan. 2020, https://doi.org/10.1016/j.procs.2020.04.084.

[16] W. Albattah, M. H. Kakakhel, S. Habib, M. Islam, S. Khan, and K. Kadir, "Hajj Crowd Management Using CNN-Based Approach," *Computers, Materials & Continua*, vol. 66, no. 2, pp. 2183–2197, Jan. 2021, https://doi.org/10.32604/cmc.2020.014227.

[17] N. Wijermans, C. Conrado, M. van Steen, C. Martella, and J. Li, "A landscape of crowd-management support: An integrative approach," *Safety Science*, vol. 86, pp. 142–164, Jul. 2016, https://doi.org/10.1016/j.ssci.2016.02.027.

[18] S. Lamba and N. Nain, "Crowd Monitoring and Classification: A Survey," in *Advances in Computer and Computational Sciences*, S. K. Bhatia, K. K. Mishra, S. Tiwari, and V. K. Singh, Eds. New York, NY, USA: Springer, 2017, pp. 21–31.

[19] R. Jiang *et al.*, "DeepUrbanEvent: A System for Predicting Citywide Crowd Dynamics at Big Events," in *25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, Anchorage, AK, USA, Aug. 2019, pp. 2114–2122, https://doi.org/10.1145/3292500.3330654.

[20] E. B. Varghese and S. M. Thampi, "Application of Cognitive Computing for Smart Crowd Management," *IT Professional*, vol. 22, no. 4, pp. 43–50, Jul. 2020, https://doi.org/10.1109/MITP.2020.2985974.

[21] E. B. Varghese, S. M. Thampi, and S. Berretti, "A Psychologically Inspired Fuzzy Cognitive Deep Learning Framework to Predict Crowd Behavior," *IEEE Transactions on Affective Computing*, vol. 13, no. 2, pp. 1005–1022, Apr. 2022, https://doi.org/10.1109/TAFFC.2020.2987021.

[22] C. Wang, H. Zhang, L. Yang, S. Liu, and X. Cao, "Deep People Counting in Extremely Dense Crowds," in *23rd ACM international conference on Multimedia*, Brisbane, Australia, Oct. 2015, pp. 1299–1302, https://doi.org/10.1145/2733373.2806337.

[23] M. Poblet, E. Garcia-Cuesta, and P. Casanovas, "Crowdsourcing Tools for Disaster Management: A Review of Platforms and Methods," in *International Workshop on AI Approaches to the Complexity of Legal Systems*, Bologna, Italy, Dec. 2013, pp. 261–274.

[24] J. L. Abbott and M. W. Geddie, "Event and Venue Management: Minimizing Liability Through Effective Crowd Management Techniques," *Event Management*, vol. 6, no. 4, pp. 259–270, Apr. 2000, https://doi.org/10.3727/152599500108751417.

[25] K. Khan, W. Albattah, R. U. Khan, A. M. Qamar, and D. Nayab, "Advances and Trends in Real Time Visual Crowd Analysis," *Sensors*, vol. 20, no. 18, Jan. 2020, Art. no. 5073, https://doi.org/10.3390/s20185073.

[26] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in *IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, Jul. 2017, pp. 2261–2269, https://doi.org/10.1109/CVPR.2017.243.

[27] R. Ayachi, M. Afif, Y. Said, and M. Atri, "Strided Convolution Instead of Max Pooling for Memory Efficiency of Convolutional Neural Networks," in *8th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications*, Maghreb, Tunisia, Dec. 2018, pp. 234–243, https://doi.org/10.1007/978-3-030-21005-2_23.

[28] X. Zeng, X. Peng, Y. Wang, and Y. Qiao, "Finding hard faces with better proposals and classifier," *Machine Vision and Applications*, vol. 31, no. 7, Sep. 2020, Art. no. 61, https://doi.org/10.1007/s00138-020-01110-4.

[29] Z. Zhang, W. Shen, S. Qiao, Y. Wang, B. Wang, and A. Yuille, "Robust Face Detection via Learning Small Faces on Hard Images," in *IEEE Winter Conference on Applications of Computer Vision*, Snowmass, CO, USA, Mar. 2020, pp. 1350–1359, https://doi.org/10.1109/WACV45572.2020.9093445.

[30] T. M. Hoang, G. P. Nam, J. Cho, and I.-J. Kim, "DEFace: Deep Efficient Face Network for Small Scale Variations," *IEEE Access*, vol. 8, pp. 142423–142433, 2020, https://doi.org/10.1109/ACCESS.2020.3012660.