A Fast Hybrid Deep Neural Network Model for pushing behavior detection in human crowds

Ahmed Alia[1,2,3], Mohammed Maree[4], Mohcine Chraibi[1]

[1]*Institute for Advanced Simulation, Forschungszentrum Jülich, 52425 Jülich, Germany*

[2]*Department of Computer Simulation for Fire Protection and Pedestrian Traffic, University of Wuppertal, 42285 Wuppertal, Germany*

[3]*Department of Management Information Systems, An-Najah National University, Nablus, Palestine*

[4]*Department of Information Technology, Arab American University, Jenin, Palestine*

a.alia@fz-juelich.de, mohammed.maree@aaup.edu, m.chraibi@fz-juelich.de

***Keywords*— Artificial Intelligence, Deep Neural Network, EfficientNetB1, Convolutional Neural Network, Crowd Behavior Analysis, Pushing Forward Motion Detection.**

ABSTRACT

Deep learning technology is regarded as one of the latest advances in data science and analytics due to its learning abilities from the data [1]. As a result, deep learning is widely applied in the human crowd analysis domain [2]. Although it has achieved remarkable success in this area, a fast and robust model for pushing behavior detection in the human crowd is unavailable. This paper proposes a model that allows crowd-monitoring systems to detect pushing behavior early, helping organizers make timely decisions before dangerous situations appear. This particularly becomes more challenging when applied to real-time video streams of crowded events, which the proposed model accomplishes with reasonable time latency. To achieve this, the model employs a hybrid deep neural network.

A. *Related works*

In crowded events, particularly at entrances, pedestrians may obey the social norm of queuing or imposing some pushing behavior to access events faster [3]. Helena et al. [4] developed a manual rating system to understand when, where and why pushing appears in video recordings of crowded entrance areas. Although this system is manual, it emphasizes the need for computer scientists to develop automatic approaches for pushing detection in crowds. The system has clearly defined pushing behavior and provided ground truth data for pushing behavior. In this context, this behavior involves pushing others and moving forward quickly by using one's arms, shoulders, elbows, or upper body, as well as using gaps among crowds to overtake and gain faster access.

However, developing automatic pushing detection is difficult due to the dense crowds, the diversity of pushing behavior, and the fact that the relevant features for pushing behavior representation are not well understood [5]. In 2022, Alia et al. [5] proposed the first automatic approach for pushing forward motion detection in video recordings at entrances of crowded events. The authors combined the EfficientNetB0-based classifier with CPU-based optical flow and false reduction algorithm to tackle the challenges of pushing detection. However, the used model in this approach is slow, and its accuracy decreases in complex scenarios of pushing behavior.

B. *Proposed Deep Neural Network model*

To address the above issues, we proposed a new model for localizing pushing patches from top-view video streams. It is important to note that the duration of each input is two seconds of streams to meet the duration of samples in the dataset that will train our model. The main goal of the model is to detect the pushing patches with a reasonable time delay that allows organizers to act. As shown in figure 1, our model consists mainly of feature extraction and classification components. Feature extraction employs three methods to extract the relevant features from the input: pre-trained deep optical flow model [6], wheel color [7], and EfficientNetB1 [8]. The pre-trained deep optical flow model is based on Convolutional Neural Network (CNN) and recurrent neural network architectures. This composition makes it an efficient approach for dense crowds because it reduces the effect of occlusions on optical flow estimation. EfficientNetB1 is one of the most efficient and simple CNN architectures. Firstly, the pre-trained model estimates the optical flow vectors in the input streams. The color wheel method then calculates the speed and direction of each pixel from the optical flow vectors. Then, it visualizes the calculated information to generate a Motion Information Map (MIM)-patches. Every patch represents the visual motion information of a specific region of the crowd at a particular time. After that, EfficientNetB1 extracts the feature maps from the generated MIM-patches. Next, the classification component labels each MIM-patch as pushing or non-pushing using a fully connected layer with one neuron and a Sigmoid activation function. Finally, the input is annotated.

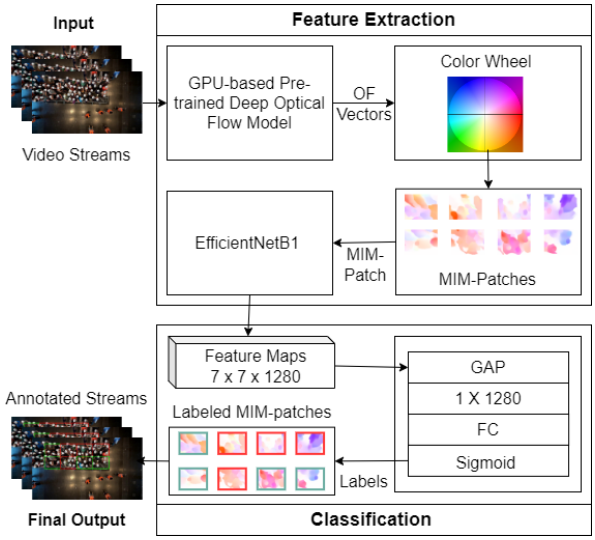

Fig. 1 The proposed model architecture

C. *Dataset Preparation*

To train and evaluate EfficientnetB1 with the fully connected layer, we prepared a new labeled dataset (training, validation, and test sets) containing several pushing scenarios. The samples in the dataset are pushing and non-pushing MIM-Patches. The data sources used to prepare the dataset are: 1) five video experiments, including their trajectory data, were

selected from the data archive hosted by Forschungszentrum Jülich under CC Attribution 4.0 International license [9, 10]. 2) the ground truth data of pushing generated by the manual rating system for experiments [4]. To create the dataset from the data sources, firstly, we utilized the pre-trained deep optical flow model and the color wheel method to generate MIM-patches. After that, we used the trajectory and ground truth data to label the patches as pushing and non-pushing. Then, the labeled dataset was randomly divided into three sets: 70 % for training, 15 % for validation, and 15 % for testing. Finally, we obtained 1585 pushing samples with 1182 non-pushing samples for the training set. In contrast, each of the validation and test set contains 336 and 251 pushing and non-pushing MIM-patches, respectively.

D. Preliminary Evaluation and Results

In order to evaluate the performance of the proposed model, we used the pushing detection model developed by Alia et al. [5] as the baseline of the proposed model. Moreover, overall accuracy, macro F1-score, and computational time metrics were employed in this evaluation process. For a fair comparison, the implementations, training processes, and all experiments for both models were conducted over the same dataset (our labeled sets) and the environment (Google Collaboratory Pro) using Python 3 with Keras library.

The results in Table 1 show the proposed model obtained 86% for both accuracy and F1-score, whereas the baseline model achieved 83% accuracy and F1-score. The main reason for this result is that the EfficientNetB1 is more efficient on datasets containing complex pushing scenarios than the EfficientNetB0.

TABLE I
PERFORMANCE COMPARISON OF THE PROPOSED MODEL WITH THE EXISTING MODEL (BASELINE MODEL) ON OUR DATASET

Model	**Accuracy (%)**	**F1-score (%)**	**Computational time (s)**
Baseline model	83	83	17.4
Our model	86	86	1.3

To measure the computational time of the models, we ran every model on 20 input streams, where the duration of each stream is two seconds with a 1920 x 1080 pixels resolution and eight patches. Then, we calculated the average computational time of all runs of every model. The last column in Table 1 displays a significant difference in the average computational time between the two models. Our model's computational time is less than the baseline model by 13 times for annotating one input of streams. It took only 1.3 seconds; in contrast, the baseline model needed 17.4 seconds for the same task. This result can be explained by using the GPU-pre-trained deep optical flow model, which is significantly faster than the deep optical flow model used in the baseline model.

E. Conclusion

In this paper, we proposed a fast and robust model for detecting pushing behavior in video streams of crowded events. In particular, the proposed model, based on a hybrid deep neural network, identifies pushing patches from top-view video streams in real-time. It combines a handcrafted crowd motion descriptor with EfficientNetB1 to extract the relevant features from the input streams. A fully connected layer with a neuron and Sigmoid activation function then identifies pushing patches based on the extracted features. Moreover, we introduced a new dataset for pushing behavior containing varied scenarios of entrance areas to train and evaluate the proposed and baseline models. The results showed that our model identifies pushing patches with 86% accuracy and 1.3 seconds delay time. On the other hand, the baseline model achieved 83% accuracy with 17.4 seconds delay time.

F. ACKNOWLEDGEMENTS

The authors are thankful to Armin Seyfried for the many helpful and constructive discussions. This work was funded by the German Federal Ministry of Education and Research (BMBF: funding number 01DH16027) within the Palestinian-German Science Bridge project framework.

References

[1] Sarker, Iqbal H. "Deep learning: a comprehensive overview on techniques, taxonomy, applications and research directions." SN Computer Science 2.6 (2021): 1-20

[2] Sánchez, Francisco Luque, et al. "Revisiting crowd behavior analysis through deep learning: Taxonomy, anomaly detection, crowd emotions, datasets, opportunities and prospects." Information Fusion 64 (2020): 318-335.

[3] Adrian, Juliane, Armin Seyfried, and Anna Sieben. "Crowds in front of bottlenecks at entrances from the perspective of physics and social psychology." Journal of the Royal Society Interface 17.165 (2020): 20190871.

[4] Üsten, Ezel, Helena Lügering, and Anna Sieben. "Pushing and Non-pushing Forward Motion in Crowds: A Systematic Psychological Observation Method for Rating Individual Behavior in Pedestrian Dynamics." Collective dynamics 7 (2022): 1-16.

[5] Alia, Ahmed, Mohammed Maree, and Mohcine Chraibi. "A Hybrid Deep Learning and Visualization Framework for Pushing Behavior Detection in Pedestrian Dynamics." Sensors 22.11 (2022): 4040.

[6] Teed, Zachary, and Jia Deng. "Raft: Recurrent all-pairs field transforms for optical flow." European conference on computer vision. Springer, Cham, 2020.R. Zeng, W. Dietzel, R. Zettler, J. Chen, and K. U. Kainer, "Microstructure evolution and tensile properties of friction-stir-welded AM50 magnesium alloy," Trans. Nonferrous Met. Soc. China, Vol. 18, Pp. s76–s80, Dec. 2008.

[7] Baker, Simon, et al. "A database and evaluation methodology for optical flow." International journal of computer vision 92.1 (2011): 1-31.

[8] Tan, Mingxing, and Quoc Le. "Efficientnet: Rethinking model scaling for convolutional neural networks." International conference on machine learning. PMLR, 2019.

[9] http://doi.org/10.34735/ped.2018.1

[10] http://doi.org/10.34735/ped.2013.1