Abstract #38 | Oral and Poster presentation

# A Novel Voronoi-based Convolutional Neural Network Approach for Crowd Video Analysis and Pushing Person Detection

**A. Alia**[1, 3], M. Maree[2], M. Chraibi[1]

[1] *Juelich Research Center, Institute for Advanced Simulation, Juelich, North Rhine-Westphalia, Germany*
[2] *Arab American University, Department of Information Technology, Jenin, State of Palestine*
[3] *University of Wuppertal, Department of Computer Simulation for Fire Protection and Pedestrian Traffic, Wuppertal, North Rhine-Westphalia, Germany*

At crowded event entrances, some individuals attempt to push others to move quickly and enter the event faster. Such behavior can increase the density over time, which could not only threaten the comfort of pedestrians but also cause life-threatening situations. To prevent such incidents, event organizers and security personnel need to understand the pushing dynamics in crowds. One effective way to achieve this is by detecting pushing individuals from video recordings of crowds. Recently, some automatic approaches have been developed to help researchers identify pushing behavior in crowd videos. However, these approaches only detect the regions where pushing occurs rather than the pushing individuals, limiting their contribution to understanding pushing dynamics in crowds.

To overcome the limitations of previous methods, this work presents a novel Voronoi-based Convolutional Neural Network (CNN) approach for pushing person detection in crowd videos. As depicted in Figure 1, the proposed approach comprises two main phases: feature extraction and labeling. In the first phase, a new Voronoi-based method is developed and utilized to identify the local regions of individuals, employing both the video and the associated trajectory data as inputs. It then uses EfficientNetB0 CNN to extract the deep features of individual behavior from the identified regions. In contrast, the labeling phase utilizes a fully connected layer with a Sigmoid activation function to analyze the extracted deep features and identify the pushing persons. Finally, this phase annotates the pushing persons in the video.

Furthermore, this work produces a novel dataset using five real-world experiments with their associated ground truths, which is utilized for training and evaluating the proposed approach. The resulting dataset consists of 11717 local regions, of which 3067 represent pushing samples, and 8650 represent non-pushing samples. The experimental outcomes demonstrate that the proposed approach attained an accuracy of 83% and a f1-score of 80%.

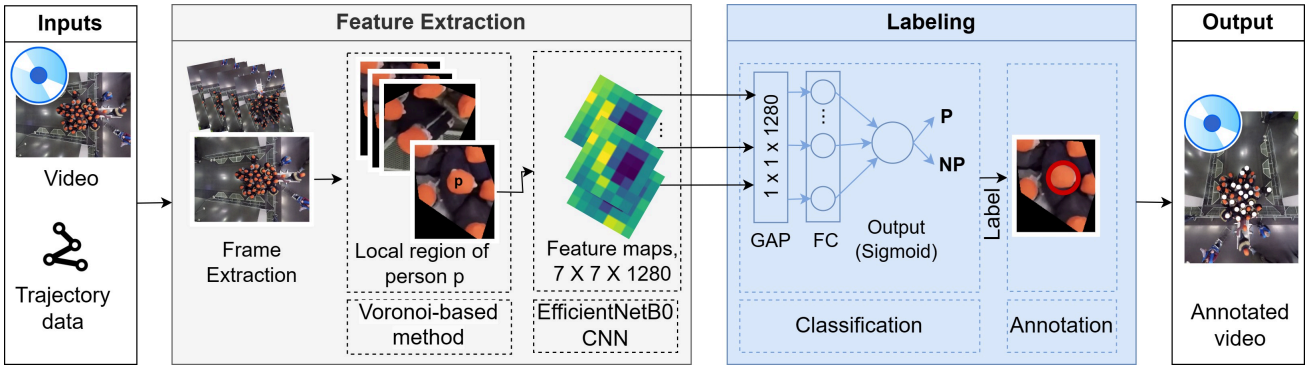


**Figure 1: The proposed approach architecture.**
P and NP terms refer to pushing and non-pushing persons, respectively. FC means fully connected layer. GAP is a global average pooling2D. White circles indicate persons who contribute to pushing.