Stampede Event Prediction Model Based on Sociology and Intelligent Algorithms

Jiayi CHANG
The High School Affiliated to Renmin University of China
Beijing, China
e-mail: jiayichang0203@163.com

***Abstract*—Stampedes in crowded environments often lead to substantial casualties, posing significant disruptions to societal order and personal safety. It becomes imperative, therefore, to devise a more precise and efficient system for predicting crowd stampedes, thereby preemptively preventing such accidents in vast spaces. This paper introduces a method utilizing a dynamic convolutional network model, grounded in a social behavior prediction model. This method extracts sociological variables such as group dynamics, activity schedules, emotional contagion, and panic levels. Utilizing multi-head attention, it captures nodes and multi-modal events associated with stampede incidents. Furthermore, it employs the spatio-temporal relationships of adjacent nodes to fuse and encode multi-modal sensing information, ultimately deriving a multi-modal data mode for the stampede event. Experimental results demonstrate that our proposed method, grounded in social behavior, yields superior results in terms of Mean Absolute Error (MAE) and other indices. Particularly applicable in settings where crowds are prone to stampedes—such as large gatherings—this method can be leveraged to predict the likelihood of stampedes.**

Keywords—Stampedes, sociological variables, spatio-temporal, multi-modal, Mean Absolute Error

I. INTRODUCTION

According to news reports, on October 30, 2022, a large crowd celebrating Halloween in Seoul, South Korea poured into an alley in a nightlife area, killing more than a hundred people, most of them teenagers in their 20s. Stampede incidents frequently occur in crowded places such as celebrations, concerts, marathons, etc. Such incidents usually result in multiple casualties and incalculable losses to society and families. A more accurate and efficient crowd prediction system needs to be developed to prevent congestion, stampede and other accidents in areas such as stations, airports, large venues, exhibitions and tourist attractions by estimating crowd flow in public places.

The paper [1] proposes a new IoT architecture, called IoT AI. A density-adaptive Gaussian kernel that varies with the local density and the distance of annotation points is proposed and used to improve the quality of the density map for detecting crowd density during the training process, and the divided density map is constructed accordingly to provide Global object counts and spatial distribution of objects, and use conditional random field-based modules to fuse and reconstruct features at various scales, and use block segmentation to extract context-aware information. Based on the network-controlled cloud platform, it can control video surveillance equipment to actively track crowded areas.

Huo et al. presents a cellular automata model considering the pedestrian trampling phenomenon to test the influences of pedestrian density, pedestrian distribution mode, exit setting, and obstacle at the exit on the stampede accidents and evacuation result. Four basic assumptions in this model are: Stampede accidents only occur near the exit. The causes of pedestrian falls only involve the factors such as crowding force and support force and pedestrian density. The fallen pedestrian occupies two cells and takes a recumbent state on the ground. External factors like fire, smoke, and falling objects are ignored. The pedestrian movement is constant [2]. The movement mode of pedestrian is determined by the extended Moore neighborhood, the probability for a pedestrian to step across the fallen pedestrians is based on the degree of panic. The crowding force in this model is depicted by the social force model proposed by Helbing, and there are significant differences in the possibility of falling under the crowding force and the state after falling to the ground according to different population type. Fallen pedestrians will be either tramped to death, or their physical strength will return to normal and will continue to move toward the exit. Taking the stage change into consideration makes this model more consistent with actual situation.

Fatai et al. proposed a model using context awareness and machine learning to predict stampede events, which is characterized for the application of personal activity recognition and monitoring using built-in sensors of smartphones [3]. The authors highlight the potential for context-aware applications and conduct context-aware and pattern-oriented machine learning frameworks. The model utilizes the k-means clustering algorithm to identify clustered areas in the crowd, and uses personal activity recognition with the help of geolocation system = and accelerometer sensor data to sense the physical environment and respond accordingly, by identifying current conditions that may cause stampedes Abnormal behavior, to mitigate the disaster. Two questions are raised in the paper: How to judge the crowd in an area? How do I identify when a stampede is likely to occur in an area? The authors used a decision tree algorithm in their proposed model, which showed improved performance on the test metrics compared to the baseline in terms of specificity, accuracy, and error rate.

Stampede is a phenomenon that current knowledge cannot fully reveal, and it is difficult to transfer the analysis and experience from major stampede events to the prevention of other stampede events. There are still some problems in the existing predictive stampede models, such as the system performance of accurately and quickly judging the state of pedestrian gathering is still not ideal, and real-time performance is crucial to prevent stampede events.

This paper proposes a dynamic convolutional network model based on the Socio-Behavioral Predictive Model. By analyzing big data from social media and public transportation, sociological variables such as group dynamics, activity schedules, emotional contagion, and panic levels are extracted, and These variables together with the image data form a multi-modal training data set, using multi-head attention to capture the spatio-temporal relationship between the node associated with the stampede event and the multi-order adjacent nodes, and obtain the multi-modality before the stampede event after training Data patterns, such as sudden dense information in social media or changes in crowd movement patterns, can be used to predict the likelihood of a stampede in a scene where crowd stampedes are prone to occur, such as large gatherings.

II. Related Work

Regarding the importance of maintaining safety in crowded situations, Sabrina et al. discussed how to develop a real-time crowd detection system in the paper [4]. The paper proposes a method to process live images captured from a Raspberry Pi camera module from an observation area and count people. If the crowd exceeds a certain limit, it will use wide area network technology to remind and warn and take appropriate measures to deal with it. The model not only counts people in large areas, but also helps managers carry out evacuations by using real-time images. The authors blur each head annotation using a Gaussian filtering technique and use two methods to estimate crowd density: the first uses a shallow convolutional network to identify people far from the camera, which are mostly captured as low-level head blobs, which Makes the training task easier as there is no need to locate the exact point of the head annotation. The second is to use convolutional neural networks to detect crowds, providing data on the contribution of crowds in different regions, enabling CNN to accurately predict crowd density and crowd size. Each detected place may have a different safety threshold of crowd density.

The paper [5] proposed a new automatic algorithm for the Detection of abnormal behavior in Dynamic Crowded Gatherings. The architecture of the system for the DADCG starts with Reading video Frames, then Temporal Features Extraction is achieved through Motion History Image (HMI), and the spatial Features is obtained by calculating the Optical Flow vector for each MHI image using Lucas-Kanade method. The optical flow image is segmented into four equal-sized blocks, and based on a congestion detected by comparing the mean value of the histogram of each segmented optical flow image, an alarm will be generated.

A crowd recognition algorithm based on preference distribution model combines crowd abnormality and intelligent monitoring video acquisition principle was proposed in [6]. The labeling of common attribute labels is achieved through KL distance similarity, and then the classification of abnormal behavior is achieved through the usage of preference distribution model. The two main characteristics of the abnormal population is represented as abnormal crowd density and abnormal crowd movement patterns. The latter one can be further analyzed through the speed, which include kinetic energy, motion entropy, etc, and direction, including the direction of motion histogram and the direction of probability distribution. A multi-tag ranking method is being used to mark the longer behavior sequence, through multi-tag classification and multi-tag sorting. The former is aimed at divide the target sequence into two categories based on whether it is behavior related. To solve the subjective preferences in recognition result, a common behavioral preference distribution is generated, and a preference distribution model learning is conducted.

Hu et al. [7] proposed a convolutional model of crowd flow prediction around dynamic graphs. This model extracts the spatial-temporal dependence of urban crowd flow data, interest point information, and holiday information. Its main structure includes extracting data spatiotemporal features The dynamic graph convolutional layer of , and add a convolutional layer to scale the features and create an output with the desired dimensionality.

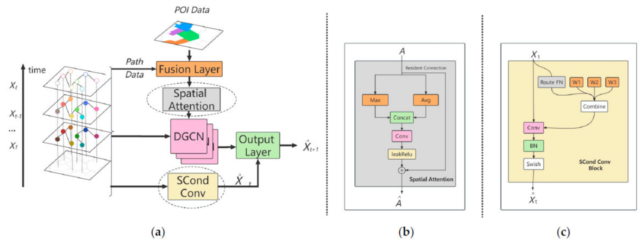

Fig. 1. The dynamic graph convolutional network model[5].

III. Multi-head Attention Multi-modal Training Structure

The paper [8] proposed a method with the usage of sensor-based technique in stampede prediction. The data was gathered through the usage of LoRa communication protocol, and the sensors used in this study include the inertial measurement unit sensor, heartbeat sensor, noise sensor, humidity sensor, and temperature sensor. After the data collected by the LoRa end nodes, the sensory data will be send to LoRa Gateway, where it will then be further send to the server for processing. The monitoring and prediction is achieved by a support vector machine algorithm. The multi-class SVM solves both numerical and categorical problems, and also effectively avoids the problem of over fitting and under fitting. The SVM model aims at increasing the margin between the hyperplane, so the support vector can quickly differentiate between different class. With the SVM model, the multiclass problems are broken down into several binary classification problems. The case studies were conducted in religious places, railway stations, bus stands, and collage campus, result shows that SVM model gives the highest accuracy among other machine learning algorithm including GBN, DT, and KNN.

An general analysis on the requirements of public places crowd stampede early-warning simulation system was proposed in [9]. The function of the system should include initial value setting, image extraction, and risk estimation. The extraction of moving target feature is mainly carried on through video surveillance images modeling to isolate object's characteristics consistently. Methods here include background subtraction, frame difference, and single Gauss background modeling. The tracking moving objects is achieved by the cam-shift algorithm, which can be operates on continuous multi-frame images with mean-shift algorithm. After the establishment of crowd behavior library, the optical flow state of sample extraction can be conducted through Bayesian networks of machine learning and training. The characteristics of head and shoulder are used to count the number of pedestrian and calculate the degree of congestion, the search boxes will calculate the density and mark result box, after comparing with the density threshold, a risk prediction will be generated.

The study in [10] proposes an intelligent model named StampSys for stampede detection in a multi-camera environment, and creates a new annotated data set named CrowdStampede with 6K images. A novel multi-modal federated learning setup (MFL) and a novel multi-label fuzzy classifier (MFC) are integrated in this model to address the privacy issue and improve the global decision respectively. This model may be more effective than models based on deep learning since DL technologies are computational consuming, there are privacy and moral concerns, and DL models may have misclassification due to lack of inter-class variance. MFL is light-weighted and is aimed to obtain multi-modal features with local decisions. This is by first extract p frames from each video and extracting feature through the usage of SqueezeNetV1. Four classifiers are incorporated to predict stampede, violence, crowd density, and crowd head count. The crowd density and crowd head count is estimated through the classification of stampede and violence and euclidean loss gained from the training with binary cross entropy loss.

The study in [11] proposes a real-time system using artificial intelligence to track people and object of interest and to detect unusual crowd density and behavior in able to give early alarm for stampede. The supervision procedure is divided into five parts: entry in camera zone, detection, analysis, warning/notification, and using information. The system mainly consists 2 modules, one is Crowd behavior analysis, and the other is Locating object-of-interest. The latter is achieved by inputting the features of the object, and the output will narrow down the possible locations from more than a million of frames to a very few number of frames. Tools used in this system include Open CV and Optical Flow. In Open CV, Meanshift and Camshift are used. The Meanshift algorithm moves the window to the new location with maximum density. Once the Meanshift has been applied and converged, Camshift updates the size of the window and calculates the orientation of the best fitting ellipse to it. The process continues until the required accuracy is met. The system architecture begins with the video feed from camera, where is then stored into the video database in the AWS cloud.

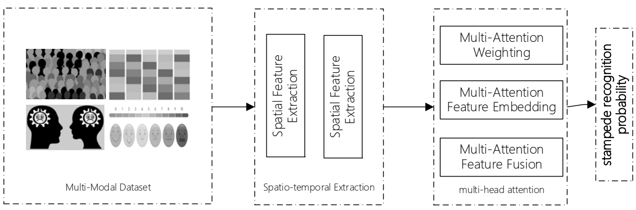

Fig. 2. The proposed multi-head attention multi-modal training structure.

We propose a dynamic convolutional network model based on a social behavior prediction model, as shown in Figure 2. By introducing sociological variables such as group dynamics, activity schedules, emotional contagion, and panic levels, multi-head attention-based perceptual information fusion coding is used to construct the spatio-temporal relationship between nodes related to stampede events and multi-order adjacent nodes, and to identify sociological variables. Sudden dense information or changes are fused with crowd movement patterns, and a recognition model for crowd stampede events is trained based on multi-modal data patterns.

IV. Experiments and Results

To examine the performance of this method, we tested it in UCSD Anomaly Detection Dataset [12], also known as UCSD Ped2 dataset, was acquired with a stationary camera mounted at an elevated position, overlooking pedestrian walkways. The crowd density in the walkways varied, ranging from sparse to very crowded. The normal setting of the video contains only pedestrians. The abnormal events are either due to the circulation of non-pedestrian entities in the walkways or anomalous pedestrian motion patterns. Commonly occurring anomalies include bikers, skaters, small carts, and people walking across a walkway or in the grass that surrounds it. A few instances of people in wheelchairs were also recorded. All abnormalities are naturally occurring, that is, they were not staged for the purposes of assembling the dataset. The data was split into 2 subsets, each corresponding to a different scene. The video footage recorded from each scene was split into various clips of around 200 frames.

The UCF-CC-50 crowd counting dataset, characterized by images of extraordinarily dense crowds, is primarily sourced from FLICKR. These images are disseminated exclusively for research purposes. It is crucial to thoroughly understand and abide by the specific terms and conditions governing the use of these FLICKR-sourced images [13]. To assess the effectiveness of the suggested approach, we analyze metrics in the spatio-temporal context. We employ two primary evaluation metrics: Mean Absolute Error (MAE) and Root Mean Square Error (RMSE). However, these measures do not adequately assess localized performance. To address this, we supplement our evaluation framework with Patch Mean Absolute Error (PMAE) and Patch Root Mean Square Error (PRMSE) [14].

$$\mathrm{MAE} = \frac{1}{M}\sum_{i=1}^{M}\left|C_{X_i} - C_{X_i}^{\mathrm{GT}}\right| \tag{1}$$

$$\mathrm{RMSE} = \sqrt{\frac{1}{M}\sum_{i=1}^{M}\left(C_{X_i} - C_{X_i}^{\mathrm{GT}}\right)^2} \tag{2}$$

$$\mathrm{PMAE} = \frac{1}{n \times M}\sum_{i=1}^{n \times M}\left|C_{X_i} - C_{X_i}^{\mathrm{GT}}\right| \tag{3}$$

$$\mathrm{PRMSE} = \sqrt{\frac{1}{n \times M}\sum_{i=1}^{n \times M}\left(C_{X_i} - C_{X_i}^{\mathrm{GT}}\right)^2} \tag{4}$$

We've classified the methodologies that utilize variables such as group dynamics, activity schedule, emotional contagion, and panic level in conjunction with image data as Proposed_1 through Proposed_4. Furthermore, the method that entails multimodal training, integrating all these variables with image data, has been designated as Proposed_5.

Fig.3 depicts the Mean Absolute Error (MAE) and Root Mean Square Error (RMSE) test results for MCNN, SANet, CP-CNN, and Proposed_1 through Proposed_5 on the UCSD dataset. The figure reveals that the MCNN has the highest average MAE and RMSE. Meanwhile, the results from the Proposed series are comparatively similar, with Proposed_5 demonstrating slightly superior outcomes.Fig.4 presents the test results for the same set of methods applied to the UCF-CC-50 dataset. It can be inferred from the figure that the average MAE and RMSE of MCNN are noticeably greater than the others. When contrasted with Figure 3, the performance disparity between MCNN and the Proposed series becomes more evident. Again, the results of the Proposed series remain relatively consistent, with Proposed_5 yielding marginally better outcomes. Finally, Fig.5 displays the Patch Mean Absolute Error (PMAE) and Patch Root Mean Square Error (PRMSE) results for MCNN, CSRNet, padNet and Proposed_1 through Proposed_5, again applied to the UCSD dataset. As can be discerned from the figure, the PMAE and PRMSE outcomes for the Proposed series are commendable.

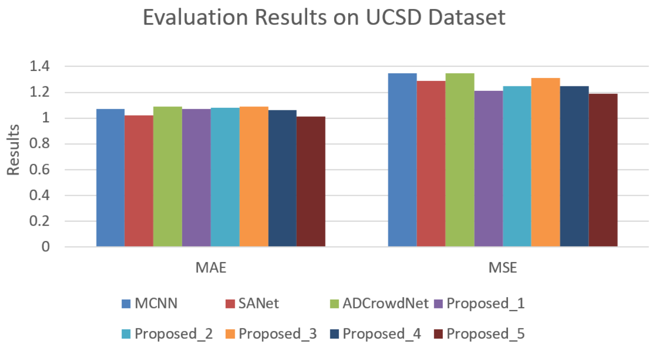

Fig. 3. The MAE/MSE evaluation results on UCSD dataset.

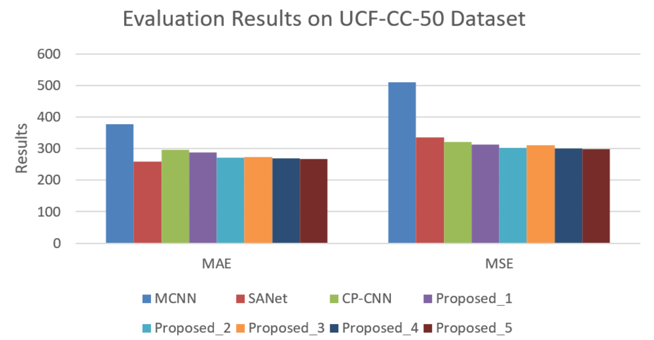

Fig. 4. The MAE/MSE evaluation results on UCF-CC-50 dataset.

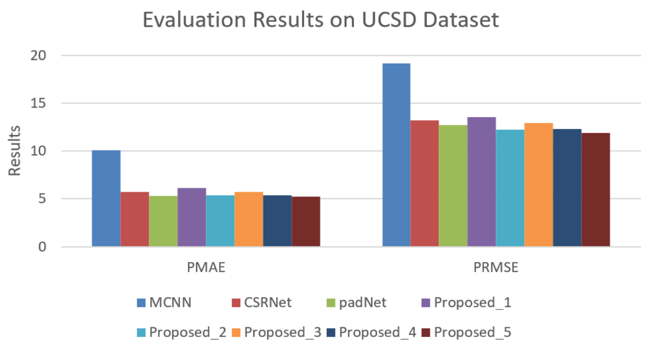

Fig. 5. The PMAE/PRMSE evaluation results on UCSD dataset.

V. Conclusions

Crowd stampede events are very harmful to society and human life, and it is difficult to predict through traditional image processing methods. This paper proposes a multi-head attention dynamic convolutional network structure based on sociological features, by perceiving multi-modal sociological information, such as group dynamics, activity schedules, emotional contagion and panic levels, etc., and fused with real-time images and encoding to construct the spatio-temporal relationship between nodes related to stampede events and multi-order adjacent nodes. Experimental results based on existing data preliminarily show that this method based on multi-head attention dynamic convolutional network has a certain ability to identify and predict crowd stampedes.

References

[1] Zhang T, Zhao Y, Jia W, et al. Collaborative algorithms that combine AI with IoT towards monitoring and control system[J]. Future Generation Computer Systems, 2021, 125: 677-686.

[2] Huo F, Li Y, Li C, et al. An extended model describing pedestrian evacuation considering pedestrian crowding and stampede behavior[J]. Physica A: Statistical Mechanics and its Applications, 2022, 604: 127907.

[3] Sadiq F I, Selamat A, Ibrahim R. Improved stampede prediction model on context-awareness framework using machine learning techniques[C]//Computational Intelligence in Information Systems: Proceedings of the Computational Intelligence in Information Systems Conference (CIIS 2016). Springer International Publishing, 2017: 39-51.

[4] Haque S, Sadi M S, Rafi M E H, et al. Real-time crowd detection to prevent stampede[C]//Proceedings of International Joint Conference on Computational Intelligence: IJCCI 2018. Springer Singapore, 2020: 665-678.

[5] Alqaysi H H, Sasi S. Detection of abnormal behavior in dynamic crowded gatherings[C]//2013 IEEE Applied Imagery Pattern Recognition Workshop (AIPR). IEEE, 2013: 1-6.

[6] Zhai B. Identification of Abnormal Human Behavior in Intelligent Video Surveillance System[C]//4th Workshop on Advanced Research and Technology in Industry (WARTIA 2018). Atlantis Press, 2018: 318-324

[7] Hu C, Liu X, Wu S, Yu F, Song Y, Zhang J. Dynamic Graph Convolutional Crowd Flow Prediction Model Based on Residual Network Structure. Applied Sciences. 2023; 13(12):7271.

[8] Kaur K, Kumar R, Kumar P. Early Prediction of Stampede In Assemblage[C]//2023 6th International Conference on Information Systems and Computer Networks (ISCON). IEEE, 2023: 1-6.

[9] Liu S, Zhu Z, Cheng Q, et al. Analysis and design of public places crowd stampede early-warning simulating system[C]//2016 International Conference on Industrial Informatics-Computing Technology, Intelligent Technology, Industrial Information Integration (ICIICII). IEEE, 2016: 210-213.

[10] Nigam N, Dutta T. A Fast, Multi-Camera, and Intelligent System for Exact Stampede Detection in Large Crowds[C]//Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems. 2022: 792-793.

[11] Suganya P, Gaurav T, Rampuriya P, et al. Locating object-of-interest & preventing Stampede using crowd management[C]//2019 3rd International Conference on Computing Methodologies and Communication (ICCMC). IEEE, 2019: 243-247.

[12] https://paperswithcode.com/dataset/ucsd

[13] Idrees H, Saleemi I, Seibert C, et al. Multi-source multi-scale counting in extremely dense crowd images[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2013: 2547-2554.

[14] Tian Y, Lei Y, Zhang J, et al. Padnet: Pan-density crowd counting[J]. IEEE Transactions on Image Processing, 2019, 29: 2714-2727. .