Poster Abstract: A Fast, Multi-Camera, and Intelligent System for Exact Stampede Detection in Large Crowds

Nitika Nigam and Tanima Dutta
Indian Institute of Technology (BHU) Varanasi, India
{nitikanigam.rs.cse18,tanima.cse}@iitbhu.ac.in

Abstract

With the increasing population, events with large crowds also increased. It often leads to uncontrolled stampede situations, causing several deaths. Deployment of intelligent systems with the quick alert feature may reduce the impact of stampedes. Researchers utilized traditional deep learning models on a centralized server for stampede detection. These models have high time, computational complexity, unaddressed public privacy concerns, and misclassification due to less inter-class variance. We thus propose a low-cost, fast, and intelligent system named **StampSys**, for accurate stampede detection over large crowds in multi-camera environment. To address complexity and privacy issues, we introduce a novel light-weight multi-modal federated learning setup. We include a novel multi-label fuzzy classifier to improve the global decision. We create a new annotated dataset, entitled **CrowdStampede** with 6K images. The experimentation results show that our system accurately classifies stampede situations on our dataset.

CCS Concepts: • **Computing methodologies** → Artificial intelligence .

Keywords: Federated Learning, Fuzzy, Stampede Detection

ACM Reference Format:
Nitika Nigam and Tanima Dutta. 2022. Poster Abstract: A Fast, Multi-Camera, and Intelligent System for Exact Stampede Detection in Large Crowds. In *The 20th ACM Conference on Embedded Networked Sensor Systems (SenSys '22), November 6–9, 2022, Boston, MA, USA.* ACM, Boston, USA, 2 pages. https://doi.org/10.1145/3560905.3568094

1 Introduction

The presence of surveillance cameras are nowadays common in crowded places, like pilgrimage sites and shopping malls. These places are prone to hazardous incidents, i.e., stamping, causing to the death [2]. Deployment of an autonomous intelligent system may generate quick alerts to reduce the impact of the disaster [6]. With the success of deep learning (DL) approaches in familiar computer tasks, researchers extended use of DL models on centralized servers for classifying stampede situations [1, 3]. Implementation of stampede detection using DL models is impacted by three challenges in context of real environment; 1) DL techniques are computational heavy with large amount of training data. 2) Issues of public information leakage are not handled. 3) DL models may lead to misclassification due to high inter-class similarity. Also, a single camera is unable to predict stampede as it captures small area. Aforementioned issues motivated us to devise a low-cost, fast, intelligent stampede detection system, named as StamSys, in multi-camera environment (MCE). StamSys contains a novel multi-modal federated learning (MFL) setup and a multi-label fuzzy classifier (MFC). We summarize our contributions as: • We tackle the problem of stampede detection for large crowds considering multi-camera deployment in real environment. We create a new dataset, named as CrowdStampede. • We introduce MFL to obtain multi-modal features with local decisions and less training overhead. It is light-weight in nature and also address the public privacy concerns. • We also propose a novel MFC that helps to take global decision based on the multi-modal local decisions, which reduces possibility of misclassification.

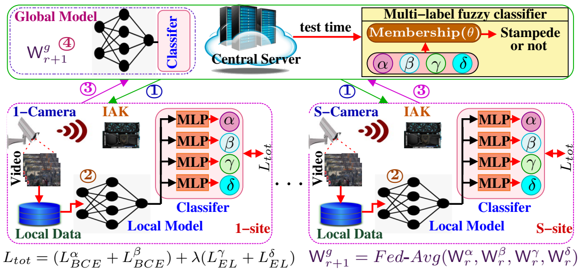

Figure 1. StampSys: Our stampede detection system with multi-modal FL setup and multi-label fuzzy classifier. Steps of Federated Learning is taken from [5].

2 Methodology

The overall framework of StampSys is shown in Figure 1.
• **Multi-modal FL setup:** We propose a novel MFL setup that trains a DL model to extract correlated representations based on multiple sites **S**. MFL follows steps to send and update the model is explained in Figure 1. We first extract p frames from each video $\mathbf{V} = \{\mathbf{F}_1, \cdots, \mathbf{F}_p\}$, where $\mathbf{F}_p$ has a spatial size of $H \times W$ with 3 channels, followed by extracting feature vectors $\mathbf{Q} \in \mathbb{R}^{p \times d}$ using SqueezeNetV1 [4]. Next, we incorporate four classifiers to predict stampede (α), violence

SenSys '22, November 6–9, 2022, Boston, MA, USA
© 2022 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-9886-2/22/11.
https://doi.org/10.1145/3560905.3568094

(β), crowd density (γ), and crowd head count (δ). A classifier consists of MLP layer is followed by an activation function Φ, i.e., $\mathbf{M}^{\theta} = \Phi^{\theta}(\mathbf{MLP}^{\theta}(\mathbf{Q}))$, $\theta = \{\alpha, \beta, \gamma, \delta\}$, where $\mathbf{M}$ is the probability vector. Further, we train the local model at each site with binary cross entropy (BCE) loss to classify stampede and violence and euclidean loss (EL) to estimate crowd density and head count. The overall loss is:

$$L_{tot} = \lambda_1 \, L^{\alpha}_{BCE} + \lambda 2 \, L^{\beta}_{BCE} + \lambda_3 \, L^{\gamma}_{EL} + \lambda_4 \, L^{\delta}_{EL}, \tag{1}$$

where λ_* is scaling factor, determined empirically. We use FedAvg algorithm for server CNN weight updation as follows:

$$\mathrm{W}^{sm}_{r+1} = \frac{1}{|S|} \sum_{s \in \mathbf{S}} (\mathrm{W}^{\alpha}_{r,s}, \mathrm{W}^{\beta}_{r,s}, \mathrm{W}^{\gamma}_{r,s}, \mathrm{W}^{\delta}_{r,s}), \tag{2}$$

where W^{sm} is server model and $\mathrm{W}^{\theta}_{r,s}$ is local model with $r - th$ number of round and different modalities on each site.

• **Multi-label Fuzzy Classifier:** We propose a MFC to produce accurate global decision at test time, despite of getting different local outputs. The fuzzy rules $\mathbf{R}_{i,j}$ on given feature vector $\mathbf{Z}$ obtained from MFL server with N dimensions are:

$$\begin{aligned} \mathbf{R}_{i,j} = & \textbf{IF } \mathbf{z1} \textbf{ IS } A_{i,j} \textbf{ AND } \cdots \textbf{ AND } \mathbf{z}_n \textbf{ IS } A_{n,(i,j)} \\ & \textbf{THEN } \Psi_{i,j} = a_{i,j}(\mathbf{Z}) + b_{i,j}, \end{aligned} \tag{3}$$

where $\mathbf{Z}$ is combination of $\{\alpha, \beta, \gamma, \delta\}$. $i \in [1, \cdots, K_{\alpha} \; or \; K_{\beta}]$ and $j \in [1, \cdots, K_{\gamma} \; or \; K_{\delta}]$. $a_{i,j}$ is the parameter vector of classification and regression coefficients and $b_{i,j}$ is scalar offset. The correct decision obtained based on $\mathbf{Z}$ as follows:

$$\Gamma(\mathbf{Z}) = \sum_{i,j \in \mathbf{R}} \chi_{i,j}(\mathbf{Z}) \Psi_{i,j}(\mathbf{Z}) \text{ and } \chi_{i,j}(\mathbf{Z}) = \frac{\mu_{i,j}(\mathbf{Z})}{\sum_m \sum_n \mu_{m,n}(\mathbf{Z})}, \tag{4}$$

where $\mu_{i,j}(\mathbf{Z})) = exp(-\frac{1}{2}(\mu - c_{i,j})^{\top} \sum_i^{-1} (\mu - c_{i,j})$ is rule activation or membership values of $\mathbf{Z}$ of i-th classification and $j - th$ modality rule. $c_{i,j}$ is the center and $\sum^{-1}$ is inverse covariance matrix. $\chi_{i,j}$ is normalized rule membership degrees in feature vector $\mathbf{f}$, i.e. the membership degrees in each rule is relation to the sum of memberships in all rules. If the value of $\Gamma(\mathbf{Z})$ tends to zero, the overall decision is not stampede.

Table 1. Impact of MFL on CrowdStampede dataset.

D. Distribution	A_{iid}	B_{iid}	C_{iid}	A_{niid}	B_{niid}	C_{niid}
#sites	2	4	8	2	4	8
#samples	4000	2000	1000	4000	2000	1000
#rounds	62	50	40	60	61	50
Acc.	79.0	**81.5**	78.2	75.3	76.4	75.8

3 Experiment

CrowdStampede is a stampede dataset collected and annotated and validated by 15 and 7 participants, respectively. 6353 frames of stampede are present. We evaluate our system in terms of test accuracy in % (**Acc.**). **Training setup:** We train our model using SGD optimizer with 0.001 learning rate and 0.9 momentum. The input frame size is 224 and batch size is 32. **Inference setup:** We provide the feature vector of as an input to fuzzy classifier for the stampede detection.

Ablation 1: Data distribution and #sites. We create three different data distributions with iid setting, i.e., $\{A_{iid}$, B_{iid}, $C_{iid}\}$. Similar for non-iid setting. We conduct experiments on $L = 2^l = \{2, 4, 8\}$ sites. We observe that iid model provides 81.5 test accuracy, as shown in Table 1. We notice that our model performs best with 4 sites in iid setting. **Ablation 2: Multi-modalities in FL setup.** We compare the presence of different multi-modalities and inference time (in seconds) on our dataset in Table 2. We observe that combination of all modalities outperforms others, but has a bit more inference time (in seconds). **Ablation 3: Multi-label Fuzzy Classifier:** We compare the performance of our model for single-label fuzzy classifier (SFC) and MFC in both distribution with 4 sites, as shown in Table 3. We note that accuracy of MFC is higher and inference time is less in iid setting.

Table 2. Effect of multiple modalities in iid setup.

	α	β	γ	δ	$\alpha + \beta + \gamma + \delta$	
Inference	1.2	2.9	3.6	3.2	**0.7**	
Acc.	78.2	78.0	77.5	75.1	**81.5**	
	$\alpha + \beta$	$\alpha + \gamma$	$\alpha + \delta$	$\beta + \gamma$	$\beta + \delta$	$\gamma + \delta$
Inference	1.3	1.6	1.9	2.0	1.2	1.0
Acc.	78.9	80.6	80.1	80.0	78.2	79.1

Table 3. Impact of MFC on CrowdStampede dataset.

	iid		Non-iid	
Fuzzy	**Inference**	**Acc.**	**Inference**	**Acc.**
SFC	**0.5**	78.1	6.1	71.6
MFC	0.7	**81.5**	1.2	75.2

4 Discussion and Future Scope

In this work, we propose a low-cost, fast and intelligent system for stampede prediction in multi-camera environment. It utilizes multi-modal federated learning to preserve the safety of the crowd and reduce high computation cost. To obtain accurate global decision, we incorporate a multi-label fuzzy classifier. In future, we will incorporate different parameters to remove the ambiguity occur through multiple cameras, as our system fails to distinguish between structured or non-structured motion patterns of crowd.

Acknowledgement

We thank SERB MTR/2021/604 for supporting this research.

References

[1] Henry Cruz et al. 2020. Automatic counting of people in crowded scenes, with drones that were applied in internal defense operations on October 20, 2019 in Ecuador. In *Proc. of ICRADS*. 111–121.

[2] Tanu Gupta et al. 2019. CrowdVAS-net: A deep-CNN based framework to detect abnormal crowd-motion behavior in videos for predicting crowd disaster. In *Proc. of SMC*. 2877–2882.

[3] Sabrina Haque et al. 2020. Real-time crowd detection to prevent stampede. In *Proc. of IJCCI*. 665–678.

[4] Iandola et al. 2016. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size. *arXiv* (2016), 1–13.

[5] Yang Liu et al. 2020. Fedvision: An online visual object detection platform powered by federated learning. In *Proc. of AAAI*, Vol. 34. 13172–13179.

[6] Jinghong Wang et al. 2013. Risk of Large-Scale Evacuation Based on the Effectiveness of Rescue Strategies Under Different Crowd Densities. *Risk Anal.* 33, 8 (2013), 1553–1563.