IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received April 25, 2021, accepted May 7, 2021, date of publication May 17, 2021, date of current version May 24, 2021.

*Digital Object Identifier 10.1109/ACCESS.2021.3081050*

# IA-SSLM: Irregularity-Aware Semi-Supervised Deep Learning Model for Analyzing Unusual Events in Crowds

**ABDULAZIZ SALAMAH ALJALOUD**[1] **AND HABIB ULLAH**[2]
[1]Univeristy of Ha'il, Hail 55211, Saudi Arabia
[2]UiT The Arctic University of Norway, 9037 Tromsø, Norway

Corresponding author: Habib Ullah (engr_habibullahkhan@yahoo.com)

**ABSTRACT** Analyzing unusual events is significantly important for video surveillance to ensure people safety. These events are characterized by irregular patterns that do not conform to the expected behavior in the surveillance scenes. We present a novel irregularity-aware semi-supervised deep learning model (IA-SSLM) for detection of unusual events. While most existing works depend on the availability of large amount of labeled data for training, our proposed method utilizes a semi-supervised deep model to automatically learn feature representations from limited number of labeled data samples. Our method extracts meaningful information from both labeled and unlabeled data during the training stage to improve the performance. For this purpose, we explore the concept of consistency regularization and entropy minimization to output confident predictions on unlabeled data. For experimental analysis, we consider various standard and diverse datasets. The results show that our IA-SSLM method outperforms several reference methods using different performance metrics.

**INDEX TERMS** Unusual event, semi-supervised learning, convolutional neural network, consistency regularization.

## I. INTRODUCTION

The analysis of unusual events is a prominent topic that involves the detection of human behaviours from a surveillance context. Unusual events have various interpretations importantly depending on the context. In our work, unusual events are represented by moving objects revealing motion patterns in the surveillance scenes that do not conform to the expected behavior and may warrant special attention or action. These events show infrequent behavior compared with all other behaviors. Similar clarifications are given by a number of authors in recent years [1]–[3]. Due to the cost reduction of surveillance equipments, it is very common to install surveillance cameras for people safety in public areas such as airports, train stations, and stadiums to name a few. Early detection may reduce the possible dangerous consequences of an unusual event, or may alert a human operator for inspecting more carefully the ongoing situation. However, there are several challenges for human operators observing the monitor screens to identify unusual events efficiently. These challenges are severe occlusions, inconsistent flow of people, and low resolution videos with temporal variations of the scene background. Therefore, with the increased number of surveillance cameras, the demand for accurate event detection increases.

Automatic detection of unusual events is a challenging task due to pattern variations in different scenes. In fact, an unusual event varies with the type of application and scenario. To detect unusual events, hand-crafted features are regarded as one key factor in the existing models, where trajectory [4], flow [5] and vision modeling [6] encode spatio-temporal information of scenes based on color, texture, optical flow, and bag-of-words features [7], [8]. However, finding the informative, discriminative, and independent set of features for unusual event detection is complex. Therefore, various abnormal event detection methods based on deep learning are developed in the past few years. The deep learning based methods can learn better feature representation than hand-crafted feature modeling, if large amount of labeled data is available for training. However, collecting large amount of labeled data is expensive and very time consuming for unusual event detection because it necessarily involves the

The associate editor coordinating the review of this manuscript and approving it for publication was Gulistan Raja.

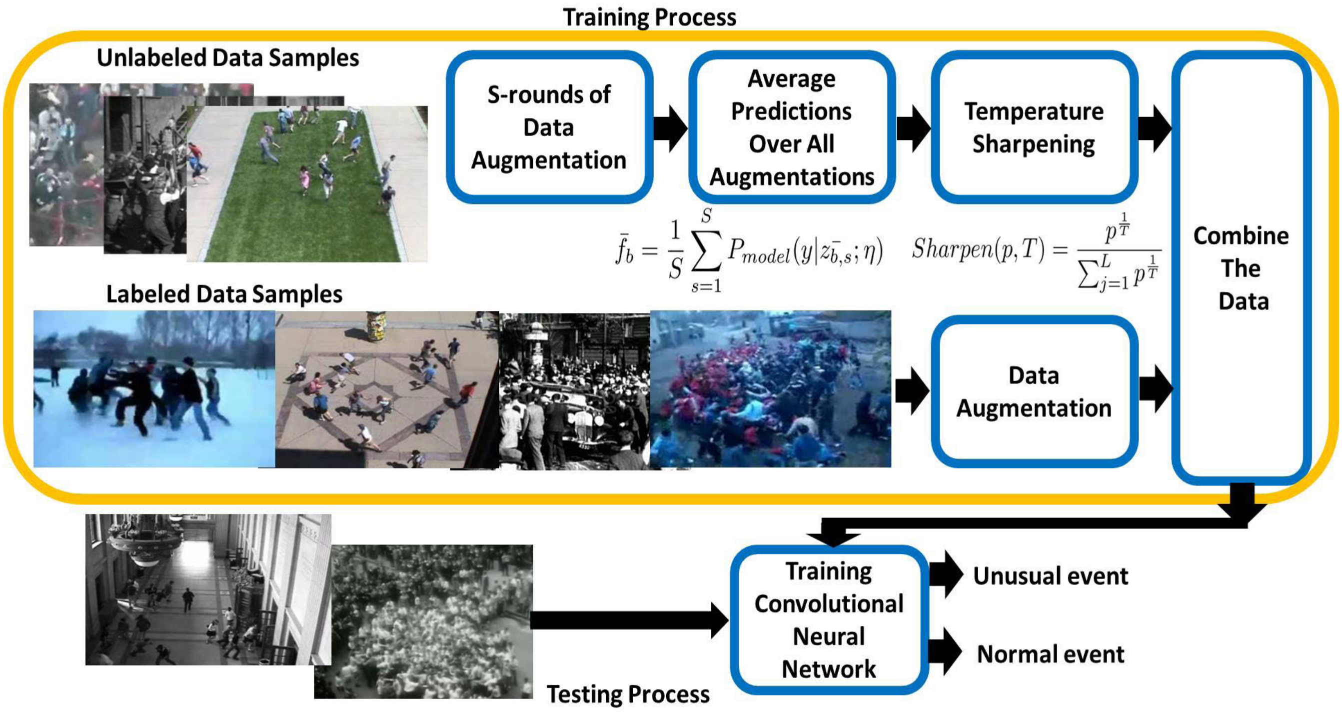


**FIGURE 1.** Our proposed IA-SSLM method performs data augmentation on both labeled and unlabeled data, which are then combined to train the convolutional neural network during the training stage.

compilation of huge archives. Furthermore, data labels may contain private information making it more challenging to label large amount of data.

In many tasks it is much easier or cheaper to obtain unlabeled data. Semi-supervised learning (SSL) [9], [10] is an efficient learning technique for leveraging unlabeled data to mitigate the reliance on large amount of labeled data samples. In fact, semi-supervised learning (SSL) seeks to largely alleviate the need for labeled data by allowing a model to leverage unlabeled data. Therefore, we propose a novel irregularity-aware semi-supervised deep learning method (IA-SSLM) for unusual event detection based on the work of Berthelot *et al.* [11]. In our proposed method, we consider batches of both labeled and unlabeled data during the training of a convolutional neural network. Our method explores the concept of consistency regularization and entropy minimization to encourage the method to output confident predictions on unlabeled data. We present our proposed IA-SSLM method in Fig. 1. Our main motivation is to demonstrate that including the paradigm of semi-supervised learning into the model has a positive impact on classification accuracy.

The main contributions of our paper are:

1) We propose a novel irregularity-aware semi-supervised deep learning method (IA-SSLM). One of the major attractions of the IA-SSLM method is its capability to model unusual events distinctively using very limited number of labeled samples.
2) We extensively evaluate our proposed IA-SSLM method on standard datasets and compared to different reference methods. Our results show that our proposed method outperforms all the reference methods.

We organize the rest of this paper as follows. Section II briefly reviews the related works. Section III describes our proposed method in details. We provide the datasets description, the experimental results, and discussion in Section IV. Finally, we present the conclusion in Section V.

## II. RELATED WORK

We categorize the literature related to the unusual event detection into hand-crafted features based methods and deep learning based method.

In the first category, Chen and Huang *et al.* [12] estimated optical flows to cluster human crowds into groups in unsupervised manner using adjacency-matrix based clustering. The behaviors with attributes, orientation, position, and crowd size of the clusters of human crowds are characterized by a model of force field. Zhou *et al.* [13] proposed a statistical method to detect abnormal behaviors by modeling trajectories of pedestrians. Initially, the trajectories are obtained by Kanade-Lucas-Tomasi feature tracker. Then trajectories are grouped to form representative trajectories, which characterize the underlying motion patterns of different scenes. Subsequently, trajectories are evaluated by multi-observation hidden markov model to find whether frames are normal or abnormal. Loy *et al.* [14] introduced an active learning approach to incorporate human feedback for on-line unusual event detection. Their method is formulated as a stream-based solution to make decision on-the-fly whether to query for labels. Furthermore, the method combined multiple active learning techniques to achieve quick discovery of unknown event classes and refinement of classification boundary. For unusual event detection, Chamle *et al.* [15] used background subtraction to detect the moving object and to mark those objects with a rectangular and elliptical bounding box followed by extracting the features like aspect ratio, fall angle, and silhouette height. Li *et al.* [16] introduced a low-rank and compact coefficient dictionary learning method for unusual event detection. For this purpose, they used the histogram of maximal optical flow projection feature and a joint optimization technique. Tariq *et al.* [17] investigated an anomaly based video surveillance model based on a particle filtering framework for online anomaly detection. They detected video frames with anomalous activities based on the posterior probability of activities in a video sequence. Benetti *et al.* [18] presented a smart ultra-low power vision system targeted to detect unusual events. The sensor embeds a low-level image processing technique that autonomously detects unusual events occurring in the scene, relying on adaptive background subtraction. The resulting binary image is then directly segmented by a field programmable gate array (FPGA), which triggers the higher layer of processing, transferring only aggregate and useful feature information. Amraee *et al.* [19] calculated histograms of oriented gradients - local binary patterns (HOG-LBP) and histogram of optical flow (HOF) in the extracted candidate regions to describe appearance and motion information. To obtain more accurate event detection, the large regions are divided into non-overlapping cells, and the abnormality of each cell is examined separately using two distinct one-class SVM models. Yuan *et al.* [20] exploited structural information within a sparse representation framework to detect unusual events. Then a smoothness regularization is constructed to describe the relationships among video events. The relationship consists of both similarities in the feature space and relative positions in the video sequences. Unusual events are identified as samples which cannot be well represented using the learned dictionary.

In the second category of deep learning based methods, Chu *et al.* [21] presented an unsupervised deep feature learning algorithm for the abnormal event detection problem. To consider the spatio-temporal information of the inputs, they used deep three-dimensional convolutional network (C3D) to perform feature extraction. By jointly optimizing between the sparse coding and the unsupervised feature learning, they modeled feature representations to detect the anomaly score of each testing input. Ye *et al.* [22] used a hybrid modulation method via feature expectation subgraph calibrating classification in video surveillance scenes for event detection. They exploited a convolutional neural network and long short-term memory models to extract the spatio-temporal features of video frame. Then they modeled the feature expectation subgraph for each key frame of every video, which could be used to encode the internal sequential and topological relational characteristics of structured feature vector. Li *et al.* [23] introduced a two-stream deep spatial-temporal auto-encoder, which is composed of a spatial stream and a temporal stream. Firstly, the spatial stream extracted appearance characteristics and the temporal stream extracted the motion patterns, respectively. Al-dhami *et al.* [24] considered the transfer deep learning model based on a pre-trained convolutional neural network to extract descriptive features. Next, the feature vector is passed into binary support vector machine classifier to construct a binary-SVM model for detecting irregular patterns. Traoré and Akhloufi *et al.* [25] proposed the violence-aware deep model for violence detection. The model combined both recurrent neural networks (RNN) and 2-dimensional convolutional neural networks (2D CNN). The CNN extracted spatial characteristics in each frame, and the RNN extracted temporal characteristics. Agrawal and Dash [26] presented the anomaly based convolutional autoencoder to detect unusual activities in different scenes. They used an autoencoder-based deep learning framework to categorize abnormality. They calculated optical flow using motion influence map and fed it to the convolutional autoencoder. Next, the spatio-temporal features collected from the output of the encoder are used for classification. Ullah *et al.* [27] extracted spatio-temporal features from a series of frames by passing each one to a pre-trained convolutional neural network model. They then fed the extracted deep features to multi-layer bi-directional long short-term memory model, which can classify ongoing anomalous/normal events in complex surveillance scenes. Asad *et al.* [28] introduced a two-stage architecture for detecting anomalous behavior in videos. In the first stage, they used a 3D convolutional autoencoder to extract spatio-temporal features from normal event training videos. In the second stage, they combined the 3D spatio-temporal features into different normality clusters, and then removed the sparse clusters to represent a stronger pattern of normality. From these clusters, one-class SVM classifier is used to differentiate between normal and abnormal events based on the normality scores. Deepak *et al.* [29] proposed a residual spatio-temporal autoencoder, which is trainable end-to-end to carry out the anomaly detection task in surveillance videos. Irregularities are detected using the reconstruction loss, where normal frames are reconstructed with a low reconstruction cost, and the converse is identified as abnormal frames.

## III. IRREGULARITY-AWARE SEMI-SUPERVISED LEARNING METHOD

Our proposed IA-SSLM method is based on a convolutional neural network consisting of 6 layers. Each layer is a container of functions. There are four convolutional layers and two fully connected layers. A convolutional layer performs a convolution operation that involves the multiplication of a set of weights with the input. A fully connected layer takes the output of the previous layer and flattens them into a single vector that can be an input for the next stage. During the training stage, our proposed method learns useful information from the training data through the semi-supervised learning process [11]. Our method exploits a loss term computed on unlabeled data to invigorate the model to generalize properly to unseen data. This loss term falls into one of the three classes namely entropy minimization [30], [31], consistency regularization, and generic regularization.

Consistency regularization uplifts the method to engender the same output distribution when its inputs are discomposed. An ordinary regularization method in supervised learning is data augmentation (DA). The DA method transforms inputs without affecting the class semantics. To extend the size of training data in image classification, image augmentation is very common to change the pixel content of an image without changing its label [32], [33]. Consistency regularization uses data augmentation in semi-supervised learning. Considering this concept, a classifier may output the same class distribution for an unlabeled sample even after it has been perturbed. In fact, the consistency regularization handles the idea that an unlabeled sample $d$ should be classified the same as $Perturb(d)$, an augmentation of itself. This is further illustrated in Eq. 1,

$$||P_{model}(y|Perturb1(d);\eta)|| - ||P_{model}(y|Perturb2(d);\eta)||_2^2 \tag{1}$$

The $Perturb1(d)$ and $Perturb2(d)$ represent different non-linear transformations. The method [34] replaces one of the transformations in eq. 1 with the output of the model using an exponential moving average of model parameter values. However, the weakness of this method is that it uses domain-specific data augmentation techniques. In our method, we explore the consistency regularization through the use of standard data augmentation for images (random horizontal flips and crops). We augment both labeled and unlabeled data. For each data sample $d_b$ in the batch of labeled data $D$, we produce an augmented form $\overline{d_b}$ = Perturb($d_b$). For each $z_b$ in the batch of unlabeled data $Z$, we produce $S$ augmentations $\overline{z_{b,s}}$ = Perturb($z_b$), where s $\in$ (1,…,S). We use these individual augmentations to produce a guessed label $f_b$ for each $z_b$. In fact, for each unlabeled sample in $Z$, we produce a guess for the sample's label using the predictions of the model. We calculate the average of the model's predicted class distributions across all the $S$ augmentations of $z_b$ according to Eq. 2,

$$\bar{f_b} = \frac{1}{S}\sum_{s=1}^{S} P_{model}(y|\bar{z_{b,s}};\eta) \tag{2}$$

Entropy minimization uplifts the method to output confident predictions on unlabeled data samples. A general concept in several SSL methods is that the decision boundary of the model should not pass through high-density areas of the marginal data distribution. One technique to implement this is to ensure that the model generates low-entropy predictions on unlabeled data samples. Grandvalet and Bengi [30] implemented it by using a loss term to minimize the entropy of their method for unlabeled data samples. Our proposed method enforces entropy minimization by exploring a sharpening function on the target distribution for unlabeled data samples. Given the average prediction over augmentations $\bar{f_b}$, we use the sharpening function to degrade the entropy of the label distribution. To use the sharpening function, we tune the temperature hyperparameter. The sharpening function is illustrated in Eq. 3,

$$Sharpen(p,T) = \frac{p^{\frac{1}{T}}}{\sum_{j=1}^{L} p^{\frac{1}{T}}} \tag{3}$$

where $p$ is the average class prediction over augmentations $\bar{z_{b,s}}$, and $T$ is a hyperparameter. As $T$ reaches 0, the output of Eq. 3 will approach a Dirac (one-hot) distribution. Because we use $f_b$ = Sharpen($\bar{f_b}$, $T$) as a target for the model's prediction for an augmentation of $z_b$, reducing the temperature helps the method to generate lower entropy predictions.

Generic regularization tunes the method to generalize well and avoid overfitting the training data. In fact, regularization enforces a constraint on the method to avoid memorizing the training data. Therefore, the method generalizes better to unseen data. To take into account the generic regularization, our method exploits the weight decay which penalizes the $L_2 - norm$ of the parameters [35], [36] of the method.

To consider the regularization techniques and the entropy minimization, our method is based on a single loss for both labeled and unlabeled data. To explain this loss, we consider a batch $D$ of labeled samples with one-hot targets (representing one of $G$ possible labels) and an equally-sized batch $Z$ of unlabeled samples. The method generates a processed batch of augmented labeled samples $\bar{D}$ and a batch of augmented unlabeled samples with guessed labels $\bar{Z}$. $\bar{Z}$ and $\bar{D}$ are then used in calculating separate labeled and unlabeled loss terms. The unified loss $\Delta$ is formulated as,

$$\bar{D},\bar{Z} = theMethod(D,Z,T,S) \tag{4}$$

$$\Delta_D = \frac{1}{|\bar{D}|}\sum_{d,p\in\bar{D}} H(p, P_{model}(y|d;\eta)) \tag{5}$$

$$\Delta_Z = \frac{1}{G|\bar{Z}|}\sum_{z,f\in\bar{Z}} ||(f, P_{model}(y|z;\eta))||_2^2 \tag{6}$$

$$\Delta = \Delta_D + \phi\Delta_Z \tag{7}$$

where $H(p,f)$ is the cross-entropy between distributions $p$ and $f$. In these equations, $T$ and $\phi$ are hyperparameters. We present the process of our proposed method in Algorithm 1.

## IV. EXPERIMENTS AND RESULTS

We perform extensive experiments on publicly available standard datasets which are extensively used in the literature. We also compare the performance of the proposed IA-SSLM method with many other reference methods using different performance metrics.

**Algorithm 1** Proposed IA-SSLM Algorithm

**Input :** Input images.
Hyperparameters : $T, K, \phi$.
**Output:** Analyzing unusual events.
Read unlabeled data samples.
Perform $S$ augmentations on unlabeled data samples.
Average predictions across all augmentations.
$\bar{f_b} = \frac{1}{S}\sum_{s=1}^{S} P_{model}(y|\bar{z_{b,s}}; \eta)$
Apply sharpening function over average class prediction.
$Sharpen(p, T) = \frac{p^{\frac{1}{T}}}{\sum_{j=1}^{L} p^{\frac{1}{T}}}$
Perform augmentation on labeled data samples.
Combine labeled and unlabeled loss terms into a unified loss.
$\Delta_D = \frac{1}{|\bar{D}|}\sum_{d,p\in\bar{D}} H(p, P_{model}(y|d; \eta))$
$\Delta_Z = \frac{1}{G|\bar{Z}|}\sum_{z,f\in\bar{Z}} ||(f, P_{model}(y|z; \eta))||_2^2$
$\Delta = \Delta_D + \phi\Delta_Z$
Trained model is used during testing.

**FIGURE 2.** We show three sample images from the three datasets. The first row shows images from the web dataset, the second row shows images from the violent-flow dataset, and the last row shows images from the UMN dataset.

### A. DATASET DESCRIPTION

To evaluate our proposed method using challenging videos, we consider three standard datasets namely the web dataset [37], the violent-flow dataset [38], and the UMN dataset [39]. The web dataset has been collected from the websites like Getty Images and ThoughtEquity.com. This dataset consists of documentary and high quality crowd videos in different urban scenes representing both normal and abnormal motion patterns. There are 12 videos representing normal crowd behaviors such as pedestrian walking, and 8 videos representing abnormal behaviors such as escape panics, protesters clashing, and crowd fighting. The violent-flows dataset consists of 246 real-world videos of crowd violence (123 violence and 123 non-violence), along with standard benchmark protocols designed for violent/non-violent classification. All the videos are downloaded from the web representing videos of uncontrolled and in-the-wild conditions. The shortest clip duration is 1.04 seconds, the longest clip is 6.52 seconds, and the average length of the videos is 3.6 seconds, with shortest/longest 1.04/6.52 seconds. The UMN dataset is collected from the University of Minnesota. The dataset consists of videos of 11 different scenarios of an escape event. The videos are collected considering three different scenes, including both indoor and outdoor. Each video clip starts with an initial part of normal behaviors and ends with sequences of abnormal behaviors. We depict three sample frames from each dataset representing unusual events in Fig. 2. For experimental analysis, we extracted frames from the three datasets and combine them into two classes representing normal and unusual events. We use 80% data for training and 20% data for validation.

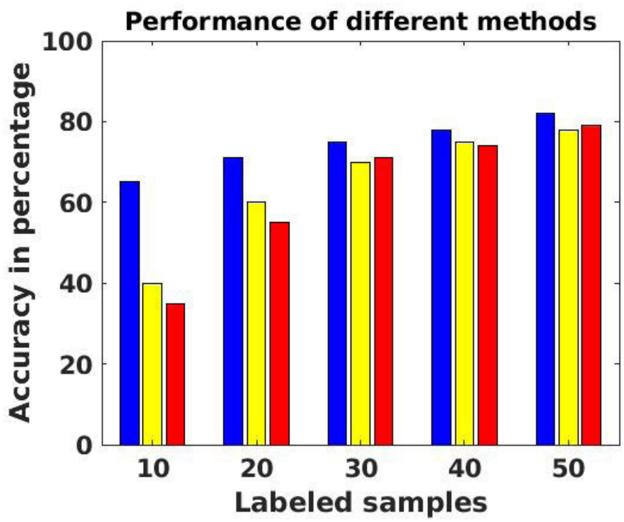


**FIGURE 3.** The accuracies of our proposed method, the CNN method, and the WideResNet method [40] are annotated in blue, yellow, and red, respectively. These accuracies are shown considering 10, 20, 30, 40, and 50 labeled samples.

### B. PERFORMANCE ANALYSIS AND DISCUSSION

For performance evaluation we consider different performance metrics including accuracy, receiver operating characteristics curves (ROC), and equal error rate (EER). We compare our proposed method with the baseline CNN model and the WideResNet model [40]. For this purpose, we use different number of labeled samples in Fig. 3, where blue bar, yellow bar, and red bar represent the accuracies our proposed IA-SSLM method, the baseline CNN model, and the

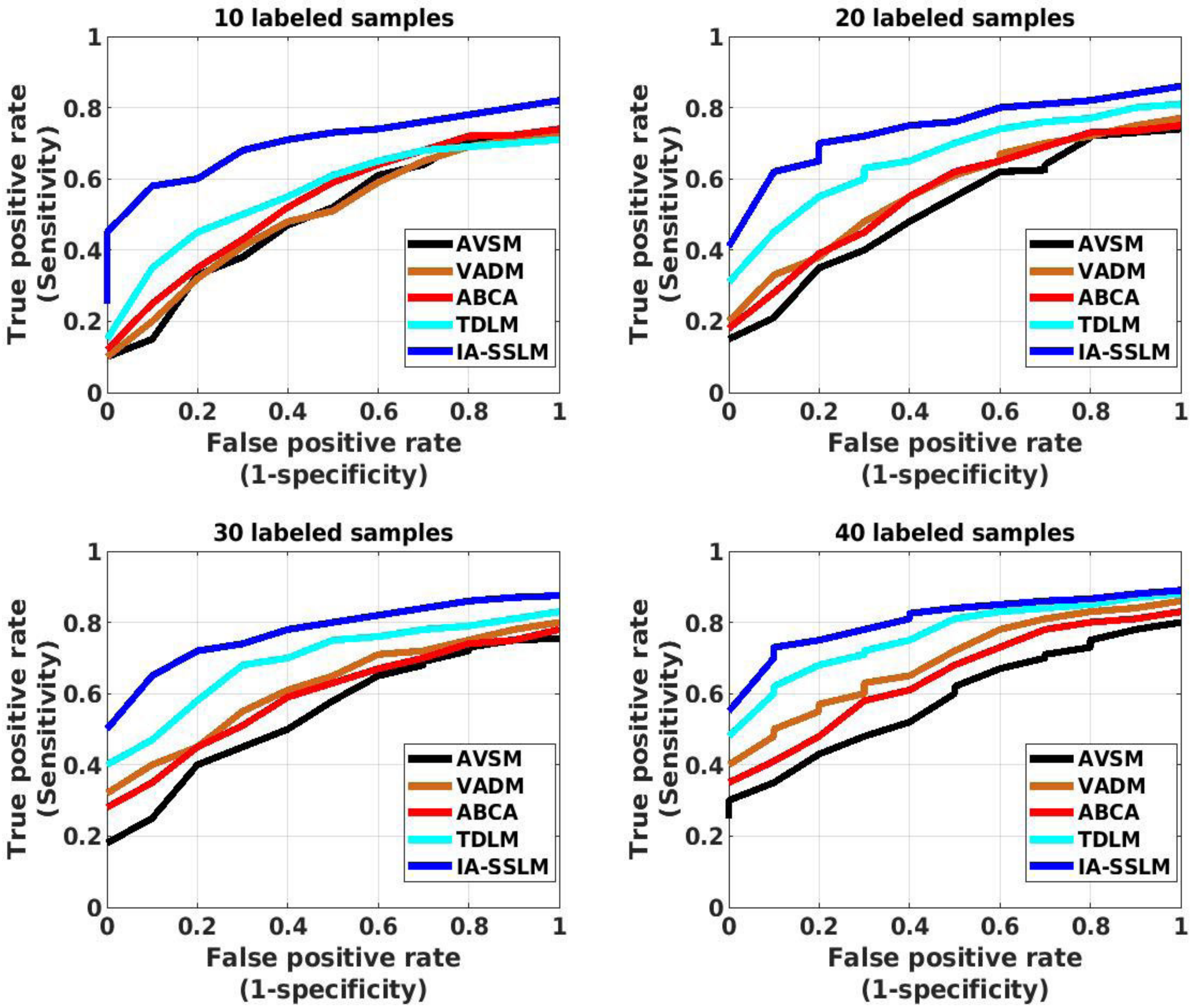


**FIGURE 4.** We show the ROC curves considering different number of labeled samples. The first row shows the ROC curves using 10 and 20 labeled samples, respectively. The bottom row shows the ROC curves using 30 and 40 labeled samples, respectively.

WideResNet model, respectively. We consider 10, 20, 30, 40, and 50 labeled samples from the training data to investigate the impact on performance. In case of our proposed IA-SSLM method, these five different number of labeled samples from the training data is considered and the rest of the training data is treated as unlabeled data. As can be seen from Fig. 3, our proposed method shows higher accuracies considering different number of labeled samples. Our motivation to compare our IA-SSLM method with the other learning methods is to show that our method can achieve better performance even with smaller number of labeled data samples. In fact our semi-supervised learning method also extracts useful information from the unlabeled data to improve the performance in term of percentage accuracy. As the number of labeled samples increases, the difference in the performances of our proposed method, the baseline CNN model, and the WideResNet model decreases. The reason is that the extraction of meaningful information from the unlabeled data decreases by increasing the number of labeled samples. However, our proposed method renders higher performance when we provide very limited number of labeled samples and that is the essence of the semi-supervised deep learning method.

We also compare our proposed IA-SSLM method with the transfer deep learning model (TDLM) [24], the violence-aware deep model (VADM) [25], the anomaly based convolutional autoencoder (ABCA) [26], and the anomaly based video surveillance model (AVSM) [17] considering different number of labeled samples. In this case, we plot the ROC curves using both the true positive rate (sensitivity) and the false positive rate (1-specificity). We present these ROC curves in Fig. 4, where the top row shows the ROC curves considering 10 and 20 labeled samples randomly extracted from the training data. The bottom row shows the ROC curves considering 30 and 40 labeled samples extracted from the

**TABLE 1.** Equal error rate (EER). We present the EER rates for the reference methods and our proposed IA-SSLM method.

| No | Different methods | Equal error rate (EER) |
|---|---|---|
| 1 | AVSM [17] | 34.28 |
| 2 | VADM [25] | 32.01 |
| 3 | ABCA [26] | 31.24 |
| 4 | TDLM [24] | 26.28 |
| **5** | **Prop. IA-SSLM** | **19.34** |

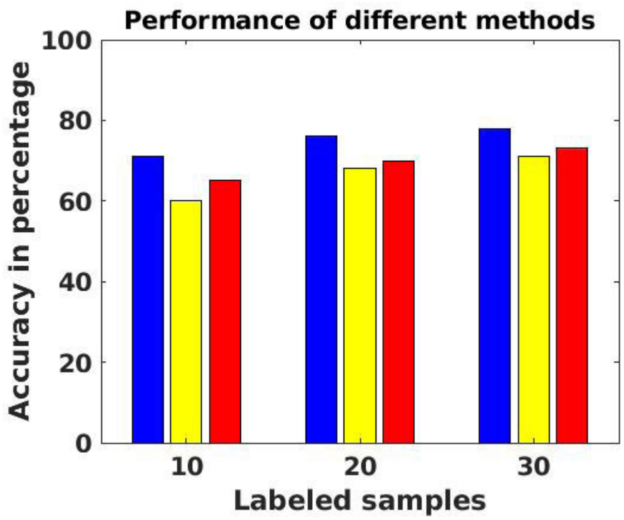


**FIGURE 5.** The accuracies of our proposed method, the trajectory based anomaly detection model (TADM) [42], and the hybrid anomaly model (HAM) [?] are annotated in blue, yellow, and red, respectively. These accuracies are shown considering 10, 20, and 30 labeled samples.

training data. As can be seen, our proposed IA-SSLM method outperforms all the reference methods in all the ROC curves. Therefore, it shows the effectiveness of our proposed method when we have different number of labeled samples.

To further elaborate the robustness of our proposed method, we present the equal error rate (EER) for our proposed method and the reference methods in Table.1. The EER rates of the AVSM model [17], the VADM model [25], the ABCA model [26], and the TDLM model [24] are 34.28, 32.01, 31.24, and 26.28, respectively. Our proposed IA-SSLM method presents the lowest EER rate equal to 19.34. Hence, our method outperforms the reference methods also in case of EER rates.

To present the generalization capability of our proposed method, we perform more experiments by taking into account an additional dataset namely the UCD dataset [41]. This dataset consists of surveillance sequences representing flows of students moving outdoor across two buildings. The unusual events are represented by the groups of students running in different directions. Considering the UCD dataset, we compare the performance of our method with two additional and recent methods namely trajectory based anomaly detection model (TADM) [42] and a hybrid anomaly model (HAM) [43]. In Fig. 5, we present the results in term of accuracy using only 10, 20, and 30 labeled data samples. The accuracies of our proposed method, the trajectory based anomaly detection model (TADM) [42], and the hybrid anomaly model (HAM) [43] are annotated in blue, yellow, and red, respectively. As can be seen, our proposed method outperforms the reference methods in all cases. Hence, our method is driven by the generalization capability to detect any unusual event depending on the training strategy and the available data in the forms of labeled and unlabeled samples.

## V. CONCLUSION

We proposed irregularity-aware semi-supervised deep learning model for detecting unusual events. For this purpose, we consider a convolutional neural network driven by semi-supervised learning process during the training stage. Our proposed method shows higher performance with very limited number of labeled data samples as compared to several reference methods considering different standard datasets. The reason is that our method also learns useful information from unlabeled data during the training stage. Therefore, our method can be extended to several other applications where collecting labeled data is time-consuming and difficult. The improvement in the performance of our method is not significantly high when the available number of labeled data samples increases. In this case, the useful information in the unlabeled data does not contribute substantially due to the utilization of information from increased labeled data.

In our future work, we would like to improve the augmentation techniques. We would also extend our method to encompass several other applications areas.

## REFERENCES

[1] H. Ammar and A. Cherif, ''DeepROD: A deep learning approach for real-time and online detection of a panic behavior in human crowds,'' *Mach. Vis. Appl.*, vol. 32, no. 3, pp. 1–15, May 2021.

[2] M. U. Farooq, M. N. M. Saad, and S. D. Khan, ''Motion-shape-based deep learning approach for divergence behavior detection in high-density crowd,'' *Vis. Comput.*, vol. 99, pp. 1–25, Feb. 2021.

[3] W. Lin, J. Gao, Q. Wang, and X. Li, ''Learning to detect anomaly events in crowd scenes from synthetic data,'' *Neurocomputing*, vol. 436, pp. 248–259, May 2021.

[4] Y. Cai, H. Wang, X. Chen, and H. Jiang, ''Trajectory-based anomalous behaviour detection for intelligent traffic surveillance,'' *IET Intell. Transp. Syst.*, vol. 9, no. 8, pp. 810–816, Oct. 2015.

[5] H. Ullah, A. B. Altamimi, M. Uzair, and M. Ullah, ''Anomalous entities detection and localization in pedestrian flows,'' *Neurocomputing*, vol. 290, pp. 74–86, May 2018.

[6] B. Zhao, L. Fei-Fei, and E. P. Xing, ''Online detection of unusual events in videos via dynamic sparse coding,'' in *Proc. CVPR*, Jun. 2011, pp. 3313–3320.

[7] K.-W. Cheng, Y.-T. Chen, and W.-H. Fang, ''Video anomaly detection and localization using hierarchical feature representation and Gaussian process regression,'' in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 2909–2917.

[8] H. Ullah, M. Ullah, and N. Conci, ''Real-time anomaly detection in dense crowded scenes,'' *Proc. SPIE*, vol. 9026, Mar. 2014, Art. no. 902608.

[9] A. Iscen, G. Tolias, Y. Avrithis, and O. Chum, ''Label propagation for deep semi-supervised learning,'' in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 5070–5079.

[10] Y. Luo, J. Zhu, M. Li, Y. Ren, and B. Zhang, "Smooth neighbors on teacher graphs for semi-supervised learning," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8896–8905.

[11] D. Berthelot, N. Carlini, I. Goodfellow, N. Papernot, A. Oliver, and C. Raffel, "Mixmatch: A holistic approach to semi-supervised learning," in *Proc. Adv. Neural Inf. Process. Syst.*, 2019, pp. 1–14.

[12] D.-Y. Chen and P.-C. Huang, "Motion-based unusual event detection in human crowds," *J. Vis. Commun. Image Represent.*, vol. 22, no. 2, pp. 178–186, Feb. 2011.

[13] S. Zhou, W. Shen, D. Zeng, and Z. Zhang, "Unusual event detection in crowded scenes by trajectory analysis," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Apr. 2015, pp. 1300–1304.

[14] C. C. Loy, T. Xiang, and S. Gong, "Stream-based active unusual event detection," in *Asian Conf. Comput. Vis.* Springer, 2010, pp. 161–175.

[15] M. Chamle, K. G. Gunale, and K. K. Warhade, "Automated unusual event detection in video surveillance," in *Proc. Int. Conf. Inventive Comput. Technol. (ICICT)*, vol. 2, Aug. 2016, pp. 1–4.

[16] A. Li, Z. Miao, Y. Cen, X.-P. Zhang, L. Zhang, and S. Chen, "Abnormal event detection in surveillance videos based on low-rank and compact coefficient dictionary learning," *Pattern Recognit.*, vol. 108, Dec. 2020, Art. no. 107355.

[17] Ata-Ur-Rehman, S. Tariq, H. Farooq, A. Jaleel, and S. M. Wasif, "Anomaly detection with particle filtering for online video surveillance," *IEEE Access*, vol. 9, pp. 19457–19468, 2021.

[18] M. Benetti, M. Gottardi, T. Mayr, and R. Passerone, "A low-power vision system with adaptive background subtraction and image segmentation for unusual event detection," *IEEE Trans. Circuits Syst. I, Reg. Papers*, vol. 65, no. 11, pp. 3842–3853, Nov. 2018.

[19] S. Amraee, A. Vafaei, K. Jamshidi, and P. Adibi, "Abnormal event detection in crowded scenes using one-class SVM," *Signal, Image Video Process.*, vol. 12, no. 6, pp. 1115–1123, Sep. 2018.

[20] Y. Yuan, Y. Feng, and X. Lu, "Structured dictionary learning for abnormal event detection in crowded scenes," *Pattern Recognit.*, vol. 73, pp. 99–110, Jan. 2018.

[21] W. Chu, H. Xue, C. Yao, and D. Cai, "Sparse coding guided spatiotemporal feature learning for abnormal event detection in large videos," *IEEE Trans. Multimedia*, vol. 21, no. 1, pp. 246–255, Jan. 2019.

[22] O. Ye, J. Deng, Z. Yu, T. Liu, and L. Dong, "Abnormal event detection via feature expectation subgraph calibrating classification in video surveillance scenes," *IEEE Access*, vol. 8, pp. 97564–97575, 2020.

[23] T. Li, X. Chen, F. Zhu, Z. Zhang, and H. Yan, "Two-stream deep spatial-temporal auto-encoder for surveillance video abnormal event detection," *Neurocomputing*, vol. 439, pp. 256–270, Jun. 2021.

[24] A. Al-Dhamari, R. Sudirman, and N. H. Mahmood, "Transfer deep learning along with binary support vector machine for abnormal behavior detection," *IEEE Access*, vol. 8, pp. 61085–61095, 2020.

[25] A. Traore and M. A. Akhloufi, "Violence detection in videos using deep recurrent and convolutional neural networks," in *Proc. IEEE Int. Conf. Syst., Man, Cybern. (SMC)*, Oct. 2020, pp. 154–159.

[26] S. Agrawal and R. Dash, "Anomaly detection in crowded scenes using motion influence map and convolutional autoencoder," in *Proceedings of International Conference on Computational Intelligence and Data Engineering*. Springer, 2021, pp. 155–164.

[27] W. Ullah, A. Ullah, I. U. Haq, K. Muhammad, M. Sajjad, and S. W. Baik, "CNN features with bi-directional LSTM for real-time anomaly detection in surveillance networks," *Multimedia Tools Appl.*, vol. 871, pp. 1–17, Aug. 2020.

[28] M. Asad, J. Yang, E. Tu, L. Chen, and X. He, "Anomaly3D: Video anomaly detection based on 3D-normality clusters," *J. Vis. Commun. Image Represent.*, vol. 75, Feb. 2021, Art. no. 103047.

[29] K. Deepak, S. Chandrakala, and C. K. Mohan, "Residual spatiotemporal autoencoder for unsupervised video anomaly detection," *Signal, Image Video Process.*, vol. 15, no. 1, pp. 215–222, Feb. 2021.

[30] Y. Grandvalet and Y. Bengio, "Semi-supervised learning by entropy minimization," in *Proc. CAP*, 2005, pp. 281–296.

[31] D.-H. Lee, "Pseudo-label: The simple and efficient semi-supervised learning method for deep neural networks," in *Proc. Workshop Challenges Represent. Learn.*, vol. 3, 2013, pp. 1–6.

[32] E. D. Cubuk, B. Zoph, D. Mane, V. Vasudevan, and Q. V. Le, "AutoAugment: Learning augmentation policies from data," 2018, *arXiv:1805.09501*. [Online]. Available: https://arxiv.org/abs/1805.09501

[33] D. C. Cireşan, U. Meier, L. M. Gambardella, and J. Schmidhuber, "Deep, big, simple neural nets for handwritten digit recognition," *Neural Comput.*, vol. 22, no. 12, pp. 3207–3220, Dec. 2010.

[34] A. Tarvainen and H. Valpola, "Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results," 2017, *arXiv:1703.01780*. [Online]. Available: https://arxiv.org/abs/1703.01780

[35] I. Loshchilov and F. Hutter, "Fixing weight decay regularization in adam," Tech. Rep., 2018.

[36] G. Zhang, C. Wang, B. Xu, and R. Grosse, "Three mechanisms of weight decay regularization," 2018, *arXiv:1810.12281*. [Online]. Available: https://arxiv.org/abs/1810.12281

[37] R. Mehran, A. Oyama, and M. Shah, "Abnormal crowd behavior detection using social force model," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 935–942.

[38] T. Hassner, Y. Itcher, and O. Kliper-Gross, "Violent flows: Real-time detection of violent crowd behavior," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2012, pp. 1–6.

[39] *Unusual Crowd Activity Dataset*. [Online]. Available: https://mha.cs.umn.edu/Movies/Crowd-Activity-All.avi

[40] S. Zagoruyko and N. Komodakis, "Wide residual networks," 2016, *arXiv:1605.07146*. [Online]. Available: https://arxiv.org/abs/1605.07146

[41] H. Ullah and N. Conci, "Crowd motion segmentation and anomaly detection via multi-label optimization," in *Proc. ICPR Workshop pattern Recognit. Crowd Anal.*, vol. 75, Nov. 2012, pp. 1–7.

[42] S. Mondal, A. Roy, and S. Mandal, "A supervised trajectory anomaly detection using velocity and path deviation," in *Proceedings of International Conference on Frontiers in Computing and Systems*. Springer, 2021, pp. 777–784.

[43] Z. Ilyas, Z. Aziz, T. Qasim, N. Bhatti, and M. F. Hayat, "A hybrid deep network based approach for crowd anomaly detection," *Multimedia Tools Appl.*, vol. 61, pp. 1–15, Mar. 2021.

**ABDULAZIZ SALAMAH ALJALOUD** received the M.Sc. and Ph.D. degrees from the University of New England, Australia, in 2015 and 2018, respectively. He is currently working as an Assistant Professor with the College of Computer Science and Engineering, University of Ha'il, Sauid Arabia. He is also working as the Vice-Dean of the Quality Assurance Unit, University of Ha'il. His research interests include machine learning and signal processing.

**HABIB ULLAH** received the M.Sc. degree in computer engineering from Hanyang University, Seoul, South Korea, in 2009, and the Ph.D. degree from the University of Trento, Italy, in 2015. He is currently working as a Researcher with the UiT The Arctic University of Norway. His research interests include computer vision and machine learning.