IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received March 12, 2020, accepted March 17, 2020, date of publication March 24, 2020, date of current version April 10, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2982906

Transfer Deep Learning Along With Binary Support Vector Machine for Abnormal Behavior Detection

AHLAM AL-DHAMARI[1,2], RUBITA SUDIRMAN[1], AND NASRUL HUMAIMI MAHMOOD[1]
[1]Division of Electronic and Computer Engineering, Universiti Teknologi Malaysia, Johor 81310, Malaysia
[2]Department of Computer Engineering, Hodeidah University, Hodeidah, Yemen

Corresponding author: Ahlam Al-Dhamari (kmaahlam2@live.utm.my)

This work was supported in part by the Universiti Teknologi Malaysia (UTM) and the Ministry of Education (MOE), Malaysia, under the UTM Fundamental Research (UTMFR) under Grant Q.J130000.2551.21H49, and in part by the Fundamental Research Grant Scheme (FRGS) under Grant R.J130000.7857.5F123.

ABSTRACT Today, machine learning and deep learning have paved the way for vital and critical applications such as abnormal detection. Despite the modernity of transfer learning, it has proved to be one of the crucial inventions in the field of deep learning because of its promising results. For the purpose of this study, transfer learning is utilized to extract human motion features from RGB video frames to improve detection accuracy. A convolutional neural network (CNN) based on Visual Geometry Group network 19 (VGGNet-19) pre-trained model is used to extract descriptive features. Next, the feature vector is passed into Binary Support Vector Machine classifier (BSVM) to construct a binary-SVM model. The performance of the proposed framework is evaluated by three parameters: accuracy, area under the curve, and equal error rate. Experiments performed on two different datasets comprising highly different context abnormalities accomplished an accuracy of 97.44% and an area under the curve (AUC) of 0.9795 for University of Minnesota (UMN) dataset and accomplished an accuracy of 86.69% and an AUC of 0.7987 for University of California, San Diego Pedistrain1 (UCSD-PED1) dataset. Moreover, the performance of the pre-trained network VGGNet-19 with handcrafted feature descriptors and with other CNN pre-trained networks, respectively, has been investigated in this study for abnormal behavior detection. The results demonstrated that VGGNet-19 has better performance than histogram of oriented gradients, background subtraction, and optical flow. In addition, the VGGNet-19 shows higher detection accuracy than other pre-trained networks: GoogleNet, ResNet50, AlexNet, and VGGNet-16.

INDEX TERMS Abnormal behavior detection, transfer learning, convolutional neural network, VGGNet-19, handcrafted feature descriptors, pre-trained networks.

I. INTRODUCTION

Abnormal detection points out the issue of detecting the statistically large deflection of data from the anticipated normal data [1], [2]. Those skewed patterns of data are often called abnormalities or outliers. Abnormal detection has been utilized in many vital areas, such as video surveillance, cyber intrusion detection, credit fraud detection, and health care. In general, based on the training dataset of normal patterns or of both normal and abnormal patterns, abnormal detection frameworks make a decision in which the detecting rate is increased, whilst the false alarm rate can be controlled by setting some specified threshold τ.

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

Support vector machine (SVM) is a ubiquitous technique in the machine learning community. In the last decade of the 20th century, it was first proposed by Vapnik [3]. Owing to its importance, SVM is vastly utilized in scientific research in various applications [4], especially for regression analysis, outlier detection, and statistical classification. Furthermore, by utilizing different kernel functions (KFs), SVM has the ability to solve both linear and nonlinear classification problems. SVM classifier's basic idea is that training data are used to construct a classification model. Then, the created classification model is utilized to classify the unknown patterns into two or more different classes of data [5]. SVM demands very few patterns for training, and it has not only good training impact but also good detection accuracy for new patterns with the same characteristics during the testing [6].

Recently, non-parametric methods for abnormal detection have been widely used. These methods suppose few data hypotheses on the data, and they appear to be more stabilized [7]. These methods involve density-based methods [8] and SVM-based methods [9].

The contributions of the proposed framework can be summarized as:

- Using transfer learning to extract motion features from videos. Hand-crafted feature extraction techniques suffer from some limitations such as the limited adaptability (i.e., perform well in one scenario and fail in another), limited design specification (i.e., working for a specific purpose such as appearance and motion to represent the main characteristic of a scene) [10]. Therefore, to tackle these limitations, a VGGNet-19 is used to extract motion features from complex and noisy surveillance scenes.
- Proposing a new computationally efficient framework for detecting abnormalities in video streams utilizing BSVM.
- Presenting a comparison between VGGNet-19 with different hand-crafted descriptors using the BSVM classifier.
- Comparing five CNN pre-trained networks along with the BSVM classifier for the purpose of abnormal detection. The comparison is made using UMN and UCSD-PED1 datasets and in terms of three significant criteria: Accuracy, AUC, and EER.

The rest of the paper is organized as follows. Section II presents a significant theoretical background on transfer learning and SVM. In section III, the most relevant related work is mentioned and highlighted. Section IV describes the proposed VGGNet-19-BSVM framework in detail. Section V provides the experimental results and analytical discussion. Section VI concludes this study.

II. THEORETICAL BACKGROUND

A. TRANSFER LEARNING BACKGROUND

As shown in Fig. 1, four main kinds of deep learning architectures are utilized for abnormalities detection namely: Convolution Neural Network (CNN) [11]–[13], Autoencoder (AE), Generative Neural Network (GAN) [14] and Recurrent Neural Network (RNN) [15]. Training a new deep learning model from scratch demands a considerable amount of data, high computational capabilities, and very long processing time. Thus, aggregation, as well as annotation of such massive data in real systems requires high cost and time-consuming. Therefore, applying deep learning models is quite challenging. In order to overcome this challenge, the concept of transfer learning has been introduced [16], [17], which means that the CNNs networks that trained on a particular dataset or certain mission may be fine-tuned for a new mission even if the scope is different. Because of the rapid increase in the amount of published information or data (e.g., audios and videos), demand for high values on accuracy and computational efficiency is also increased. Owing to these factors, transfer learning has drawn a lot of attention.

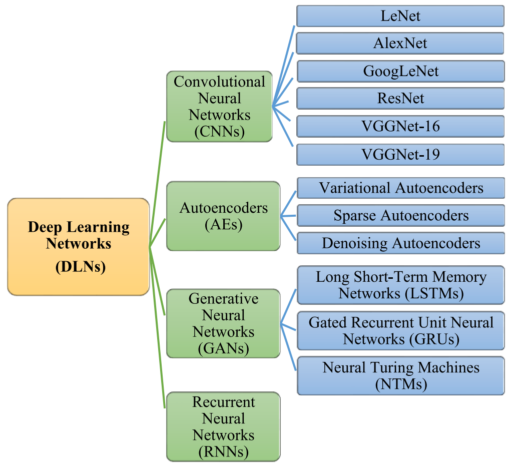


FIGURE 1. DLNs types and variants.

While transfer learning allows doing new enhancements for visual classification, the traditional methods of machine learning have limited ability to improve the mentioned factors above. Fundamentally, transfer learning uses two approaches:

- Maintaining the original pre-trained network and updating the weights based on a new training dataset by continuous backpropagation up to the upper layers.
- Utilizing the pre-trained network for extracting visual features. In this approach, the final fully connected layer, which is basically the classifier layer (softmax classifier for VGGNet), is replaced with a new suitable classifier like SVM for visual classification [18].

The last approach has been implemented effectively to many functions of classification and recognition. The proposed framework also depended on the second approach. In this study, VGGNet-19, a recently proposed benchmark for deep models, is investigated. The following four common CNN deep pre-trained networks are described briefly:

1) ResNet50

ResNet is a CNN architecture 152 layers deep, introduced in 2015 by Microsoft Research Asia. ResNet puts forward residual connections where the output of a convolutional-rectified linear unit-convolutional series is inserted into the original input, and afterward processes via rectified linear unit (ReLU) as seen in Fig. 2. Through that, the data is easily transferred from the former layer to the subsequent layer and the gradient streams during backpropagation due to the addition operations that disseminate the gradient. The ResNet50 is one of the models proposed by the Microsoft research team in deep residual learning for object recognition. ResNet50 consists of 50 layers of identical blocks with shortcut connections. Such connections retain low computation as well as produce rich combination features.

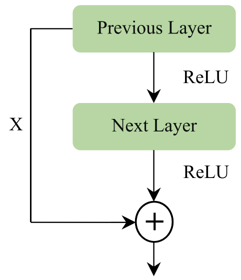


FIGURE 2. Residual connection [18].

2) GoogleNet

GoogleNet is a small network composing three layers of convolution. One important aspect of GoogleNet is that it is developed very deep with 22 layers deep. Another aspect of GoogleNet is that it has a new level of organization named "Inception Module". The essential concept of Inception module is to locate the optimum local construction as well as to repeat it spatially. Also, one of the first valuable features of GoogleNet structure is that it permits for increasing the number of units at each phase without an uncontrolled blow-up regarding the computational complexity. Thus, the network can be structured very deep as well as be effectively trainable [16], [19].

3) AlexNet

AlexNet is deemed as a significant breakthrough in the computer vision area [20]. AlexNet reduced the error rate of classification from 26% to 15%, which is a major improvement [18]. It is 8 layers deep and learns rich feature representations for a wide variety of images. The image input size for this network is 227×227. AlexNet includes five convolutional layers and three fully connected layers as well as three max-pooling layers after the 1st, 2nd and 5th convolutional layers. 96 filters with size 11×11 accompanied by a stride of four pixels as well as padding with two pixels have been included in the first convolutional layer. For other convolutional layers, both the stride and padding are designated by one pixel. The second convolutional layer includes 256 filters of size 5×5. The other remaining convolutional layers have filters with a size of 3×3: 384, 384, and 256 filters for the third, fourth and fifth layers, respectively [21].

4) VGGNet

VGGNet was introduced in 2014. It has uncomplicated and repeated architecture. VGGNet fulfilled an error rate of 7.32% in ImageNet Large-Scale Visual Recognition Challenge (ILSVRC) 2014 [18]. VGGNet resembles AlexNet but with extra convolution layers. VGGNet has been introduced by the Visual Geometry Group (VGG) of Oxford University. VGGNet effectively builds 16-19 layers (convolution layer and fully linked layer) of deep convolution neural networks by stacking 3×3 small convolution cores and 2×2 maximum pool layer repeatedly [22].

B. SUPPORT VECTOR MACHINE

Over the previous two decades, SVM and its applications have developed rapidly. SVM notion was developed from the statistical learning theory [23]. It is employed in a wide range of applications such as image recognition [24], text categorization [25], bioinformatics [26], biometrics [27], and chemoinformatics [28]. The primary aim of the SVM is to project the data points into higher dimensional space where the different classes become linearly separable and to construct a hyper plane (an optimal decision surface) that can distinguish the data points in this space [29], [30]. This usually requires solving quadratic programming (QP) problem utilizing kernel functions [31]. Different kernel functions like linear, polynomials, and radial basis function (RBF) can be utilized with SVM [32]. Since SVM is based on the structural risk minimization, it offers better generalization ability than other classification algorithms that use empirical risk minimization [33]. Suppose that a training dataset of N data points are given: $A = \{a_1, a_2, \ldots a_N\}$, where each pattern $a_i \in R^n$ has d features and it belongs to one of two classes $y_i \in \{\pm 1\}$. If the patterns of training are linearly separable, SVM defines a hyperplane that splits up the data in the feature space, which can be represented as in the following equation, supposing that w is an adjustable weight vector and b is a bias:

$$f(a) = w^T a + b = 0 \quad (1)$$

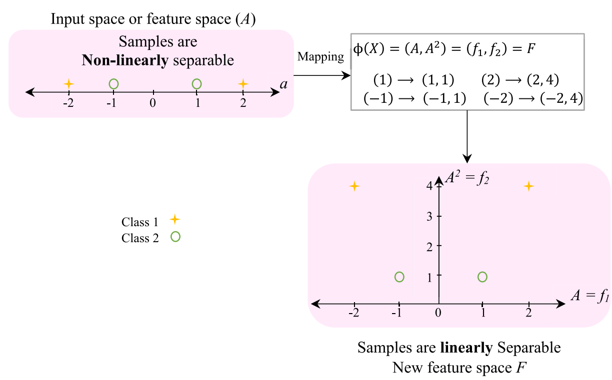


FIGURE 3. An example illustrates how the KF is utilized to transform the input data.

The KFs are used in SVM to transform the data $A \in R^n$ into a higher-dimensional space ($F \in R^d$), as illustrated in Fig. 3, where $d > n$. Let $A \in R^1$ be the input data, which has two classes and each class has two non-separable patterns. The KF is utilized to map A into a new feature space $F \in R^2$. As can be seen, if the data are linearly separable in F space, then the standard SVM can be applied [34]. There are three widely used KFs [35], described in Table 1. The RBF, also called Gaussian kernel function as shown in Table 1 has only one variable σ, which controls the performance of SVM. There are several options to RBF kernel such as Laplacian and Exponential kernels [36].

TABLE 1. Common KFs with SVM.

KF	Equation	Description
Linear kernel	$K(x_i, x_j) = x_i^T x_j$	Two class learning
Polynomial kernel of degree d	$K(x_i, x_j) = (x_i^T x_j + 1)^d$	-
Gaussian or Radial basis function (RBF)	$K(x_i, x_j) = exp(-\|x_i - x_j\| /2\sigma^2)$	One class learning, σ is the width of the kernel

III. RELATED WORK

Many researchers have concentrated their studies on video-based abnormal behavior detection. Comprehensive review papers of this area can be found in [1], [37]. The existing state-of-the-art algorithms for abnormal behavior detection can be generally categorized into dynamic Bayesian network (DBN), Bayesian topic models (BTMs), clustering-based models, artificial neural network (ANN), deep learning-based models and sparse representation-based models [1]. In the following sub-sections, we focus only on two types, which are SVM-based and deep learning based abnormal detection algorithms.

SVM algorithms are widely utilized in abnormal detection [38]. Fu *et al*. [39] have developed a self-evolving framework for detecting abnormalities to improve the reliability of cloud computing platforms. Based on their experiments, it is verified that SVM starts to perform reasonably well on imbalanced datasets once the percentage of anomaly patterns reaches 10%.

In [40], the authors put forward an SVM-based intrusion detection model along with the time-varying chaos particle swarm optimization (TVCPSO) algorithm that is utilized for determining the best parameters for the SVM classifier. In [41], after computing optical flow (OF) between pairs of consecutive frames of the input video stream, individual and interactive behaviors using static and dynamic agents are extracted. Static agents calculate the variation of the OF at the static location over time to observe the individual behavior of a specific location in the crowded view. Dynamic agents calculate interactions between neighbors using models of social force and energy interaction potential. Individual and interactive behaviors observed by the agents are shown as characteristic words. As a consequence, a bag-of-words (BoWs) representation is utilized to represent an input video. Last, SVM is employed to classify the abnormalities.

Kim *et al*. [42] presented an algorithm to detect unusual behaviors that depend on the estimation of body parts utilizing a geodesic graph as well as SVM classifier. The performance of this method was greatly influenced by the detection of human body region, especially for adjacent people. In addition, Shu *et al*. [43] put forward an approach to detect violence in the lift that depends both on visual features and SVM classifier. They incorporated corner kinetic energy based optical flow features and motion characteristics to depict the fierce event. Later, they utilized the SVM classifier to detect violence in the lift. In [44], a simple strategy on a smart camera network to recognize human behavior sequence was presented. The SVM classifier was designed for single event recognition, using a 10-fold cross validation method to train and test the classifier. Liu *et al*. [45] proposed representing the spatial-temporal silhouette in order to characterize motion properties such as everyday activities. They made use of the multi-class SVM to classify different samples with multiple views and motion descriptor scenarios for each action.

The growing success of deep learning algorithms in diverse applications leads to utilize them for detecting abnormality or normality of human behaviors. The deep learning term points to learning a hierarchical set of features employing several layers of hidden nodes in an artificial neural network. Dissimilar from other methods, deep learning algorithms are not required to define a specific set of features to extract from the dataset. Deep learning algorithms learn the handy features directly from the data [46].

Borja *et al*. [47] review deep learning algorithms for abnormal behavior detection. CCN networks contain a stack of convolutional layers with a fully connected layer and a softmax classifier. Zhou *et al*. [48] proposed a method to classify abnormalities utilizing a spatial–temporal 3D CNN model. Authors in [49] proposed to learn regular patterns using autoencoders (AEs) with limited supervision. First, they take advantage of the conventional spatiotemporal local features and learn a fully connected AE. Then, a fully convolutional AE is built to learn both the local features and the classifiers in a single learning framework. Additionally, Li [50] proposed a deep spatiotemporal architecture, which has convolutional and recurrent nature for understanding pedestrian behavior in a crowd. In [11], authors study abnormal behavior detection in different situations such as various background settings and number of subjects using CNN. Zenati *et al.* [14] proposed a new framework based on Bidirectional GAN, which simultaneously learns an encoder E that maps samples x in latent representation z together with a generator G and a discriminator D during training.

However, Gnouma *et al*. [51] put forward a new approach for recognizing human activity based on the history of binary motion image (HBMI). They took a range of silhouettes of human activities as a basis for representing the characteristics. To obtain human silhouette, they utilized background subtraction (BS), which uses a combination of both the uniform motion of magnitude of optical flow (MOF) and Gaussian mixing model (GMM). The stacked sparse autoencoder (SSAE) was used to detect automated human activity. SSAE can capture in an unsupervised way the high-level features of pixel intensity. In [46], the authors formulate the abnormality detection problem as a spatiotemporal sequence outlier detection, and they used a combination

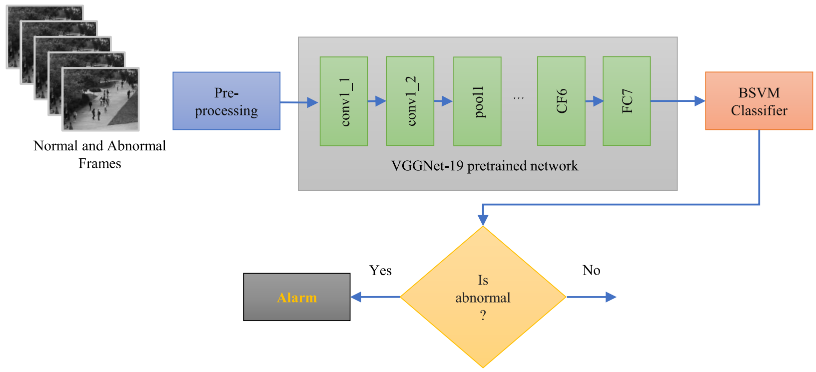


FIGURE 4. **Abnormal detection framework using transfer learning and BSVM.**

of spatial feature extractor and temporal sequencer convolutional long short-term memory (ConvLSTM) model to address the problem.

IV. VGGNet-19-BSVM PROPOSED FRAMEWORK

The proposed framework includes three main stages: pre-processing, feature extraction using VGGNet-19 pretrained network, and BSVM model. These stages are illustrated in the flowchart in Fig. 4. Note that the flowchart describes both the training and the testing procedures. The first stage is to pre-process the raw video. In the second stage, the transfer learning employing deep CNNs is used. This is very useful for model training with a dataset of restricted size as CNNs are susceptible to over-fitting with small size datasets. Over-fitting could be obviated by raising the volume of the training dataset. However, the provision of a large amount of annotated information is very hard and costly. Therefore, transfer learning is useful in this case and solves the issue by utilizing pre-trained profound representation as a source architecture to build a new structure. After that, BSVM model for two classes (normal and abnormal patterns) is built and learned during the training stage, to predict to any class the new features belong to. After the training, to know if a new video frame is abnormal or not, the extracted features from that frame are passed as a vector of features to the BSVM model, which provides a prediction.

A. PRE-PROCESSING FOR VGGNet-19 PRETRAINED NETWORK

Changes in lighting conditions have an important effect on the performance of abnormal detection algorithms [1], [52]. One way to handle disparate illumination is to utilize illumination normalization as a pre-processing step. In the proposed framework, we apply the illumination normalization algorithm using histogram equalization technique to control lighting conditions. After that, we use Gaussian filtering to remove the unwanted small objects. Since VGGNet-19 network can only process an image input size of 224×224 for RGB images, we then resize all frames into $224 \times 224 \times 3$. In addition, because UCSD-PED1 frames are gray images, all the frames are converted into color images.

B. FEATURE EXTRACTION BASED ON VGGNet-19 PRETRAINED NETWORK

Features extraction was performed by passing video frames through the pre-trained VGGNet-19 network [53]. The VGGNet-19 model was trained using the ImageNet dataset and has a depth of 19 layers [54]. The adopted VGGNet-19 network with learnable weights is shown in detail in Fig. 5. The VGGNet-19 network has an image input size of 224×224 for RGB images, and it is always fixed.

The pre-trained VGGNet-19 has five blocks of 16 convolutional layers and 3 fully connected layers. Those convolutional layers employ very small 3×3 kernels and this in general has significantly improved the accuracy of VGGNet compared with other architectures such as ZFNet [18]. It uses a stride of 1 and padding of 1 throughout the whole network to guarantee that every activation map has similar spatial dimensions like the former layer. For spatial down-sampling, five max-pooling layers are used. These max-pooling layers are performed utilizing a 2×2 window with a stride of two and zero padding to guarantee that every spatial dimension of the former layer is halved. All the convolutional layers of VGGNet-19 use the ReLU, which is carried out after every convolution. It used two fully connected layers with 4,096 ReLU activated units and a fully connected softmax layer with 1,000. The five convolutional blocks of VGGNet-19 can be viewed as feature extraction layers. The activation maps created by these layers in our framework are called human motion features. Fig. 6 shows a visualization of training features from 6 randomly selected deep layers.

C. BINARY SUPPORT VECTOR MACHINE CLASSIFIER

As presented in section 4.2, the proposed framework made used of VGGNet-19 network was to extract human motion features from training and testing data without needing hand-crafted techniques to extract features. To perform the binary classification, BSVM with linear kernel was utilized.

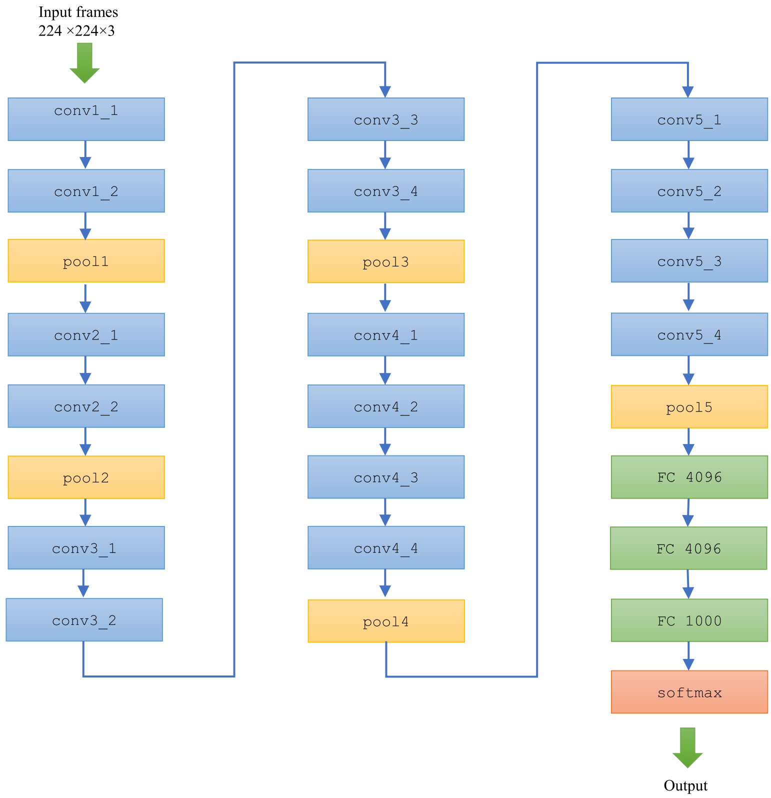


FIGURE 5. Architecture diagram of VGGNet-19.

The BSVM classifier was selected on the basis of the following factors:

- To date, BSVM is the most utilized classifier to solve different classification problems.
- Owing to the small size of the evaluated datasets, using the full VGGNet-19 model would not be recommended since it requires a large set of training samples.
- Only two parameters should be selected to implement BSVM, upperbound and the kernel parameters.
- It produces a good performance of generalization by implementing the principle of structural risk minimization (SRM) [55].

In the binary classification problem, the machine learning approach provides the output labels with a hyperplane separation where $y_i \in \{\pm 1\}$ points out the classification "label" of the input vector x. In case the two classes in the input space are linearly separable, the hyperplane is represented as in Eq. (1).

Given that the input frames and a set of training labels are given as: $\{x_i, y_i\}_{i=1}^{n}, y_i \in \{\pm 1\}$. The BSVM is determined as [56]:

$$y(x) = sign\left[\sum_{i=1}^{n} \alpha_i y_i k(x, x_i) + b\right] \tag{2}$$

where α_i represents the Lagrange multiplier, y_i is the target value of training pattern x_i, $k(x, x_i)$ is the kernel function, and b is the bias, which is the offset of the hyperplane along its normal vector.

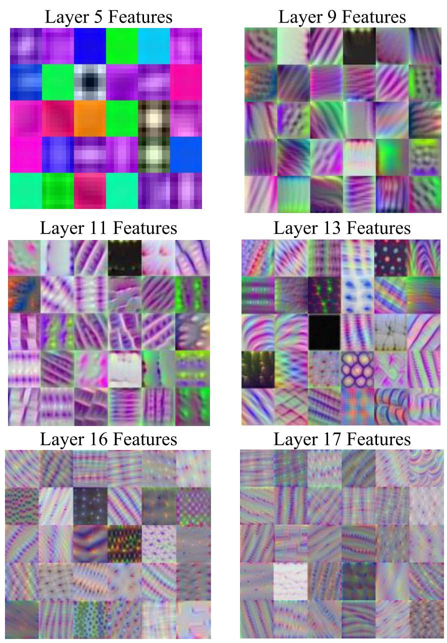


FIGURE 6. Features visualization using Matlab deepDreamImage function.

V. EXPERIMENTAL RESULTS

The experiments on this study were conducted using MATLAB R2017b (9.3.0.713579) x64 on Windows platform

TABLE 2. Confusion matrix for binary classification.

Actual	Detected	
	Normal	Abnormal
Normal	TN	FP
Abnormal	FN	TP

with an Intel Core i7-4600U CPU working at 2.10 GHz with a 4 MB cache and 8 GB RAM, and NVIDIA GTX 750 Ti GPU. We validate the effectiveness of the proposed frameworks by testing their abilities to detect abnormalities on two common datasets, UMN, and UCSD-PED1. To analyze video frames and to detect abnormalities, frame-level based criterion has been adopted to indicate whether abnormalities occur within the frame or not [57]. The frame-level based receiver operating characteristic (ROC) curve and area under the curve (AUC) are utilized in this study as the performance evaluation metrics. Furthermore, we use the confusion matrix to measure the detection accuracy, and equal error rate (EER) values in the experimental datasets. The structure of the confusion matrix for classification of abnormal detection algorithms is represented in Table 2, where the definition for True Negative (TN), False Negative (FN), True Positive (TP), and False Positive (FP) are:

- TN: The number of normal samples that are correctly detected as normal,
- FN: The number of abnormal samples incorrectly detected as normal,
- TP: The number of abnormal samples that are correctly detected as abnormal,
- FP: The number of normal samples that are incorrectly detected as abnormal.

The computation formulas of accuracy (Acc), and equal error rate (EER) measures are as in the following [58]:

$$Acc = \frac{TP + TN}{TP + TN + FP + FN} \tag{3}$$

$$EER = \frac{FP + FN}{TP + TN + FP + FN} \tag{4}$$

A. DATASETS

The performance of the proposed framework is validated on standard benchmark datasets using UMN and UCSD, which are available online for abnormal behavior detection. UMN and UCSD datasets are summarized as follows:

UMN: The UMN [59], an unusual crowd activity dataset, is the most straightforward and widely used dataset. It includes three different scenes, which are Lawn, Indoor, and Plaza. The three scenes have 1453, 4144, and 2144 frames, respectively for both training and testing stages. In particular, we used 3,085 frames for training and 4,656 for testing. However, although the training set seems to be insignificant, BSVM requires only a few training patterns for abnormal detection, and it provides pretty good detection accuracy for new patterns with the same characteristics during the test. The frame resolution for this dataset is 240 × 320 pixels. In this dataset, the unexpected running of people recorded as abnormal.

UCSD: The UCSD [60] consists of two categories: PED1 and PED2. UCSD-PED1 comprises 34 training and 36 testing video clips, with 200 frames each (i.e.,14,000 frames in total for training and testing). The frame resolution for the UCSD-PED1 dataset is 158 × 238 pixels. It is worth mentioning that PED1 is more challenging than PED2 because the angle of camera produces larger perspective distortion. Moreover, anomalous events in PED1 involves not only abnormalities resulted by small carts, bikers and skateboarders etc., but also contextual abnormalities such as a person walking over the grass. Fig. 7 and Fig. 8 illustrate some patterns from UMN and UCSD-PED1 datasets, respectively.

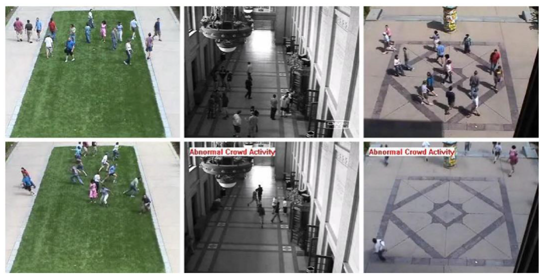


FIGURE 7. Instances from each view in UMN benchmark. The images in the first row represent normal samples and in the second row show abnormal samples.

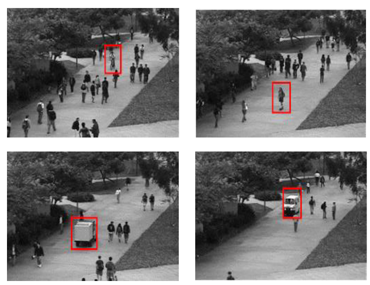


FIGURE 8. Four different abnormal events from UCSD-PED1.

TABLE 3. Detection accuracy applying UMN dataset.

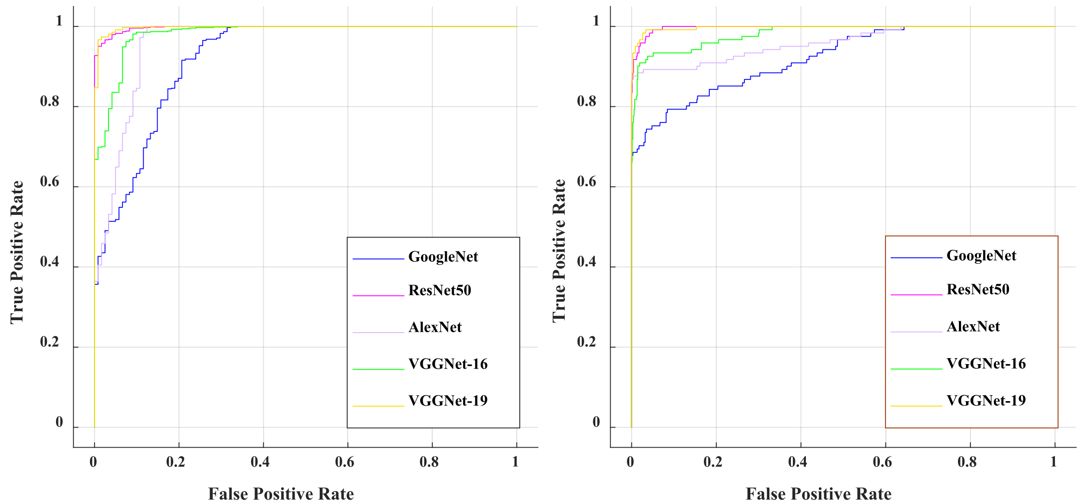
Dataset	Hog (%)	BS (%)	OF (%)	VGGNet-19 (%)
Lawn scene	81.2986	91.5661	78.5939	96.9059
Indoor scene	60.6450	90.7085	51.5145	95.8044
Plaza scene	79.3060	38.0770	90.1385	99.6183
Average:	73.7499	73.4505	73.4156	97.4429

B. DISCUSSION

The main aim of this paper is to detect abnormalities efficiently. To achieve this, transfer learning technique through VGGNet-19 and BSVM are utilized. The VGGNet-19 is trained on more than a million images from the ImageNet database. Also, it is trained to extract more efficient

TABLE 4. Detection accuracy applying UMN dataset using five different transfer learning networks.

Dataset	GoogleNet		ResNet50		AlexNet		VGGNet-16		VGGNet-19	
	Accuracy (%)	EER	Accuracy (%)	EER	Accuracy (%)	EER	Accuracy (%)	EER	Accuracy (%)	EER
Lawn scene	82.0545	0.1795	94.5545	0.0545	97.6485	0.0235	93.9356	0.0606	96.9059	0.0309
Indoor scene	80.2095	0.1979	76.3005	0.2370	80.4512	0.1955	86.9317	0.1307	95.8044	0.0419
Plaza scene	79.6437	0.2036	97.0738	0.0293	97.7099	0.0229	97.7099	0.0229	99.6183	0.0038
Average:	80.6359	0.1937	89.3096	0.1069	91.9365	0.0806	92.8591	0.0714	97.4429	0.0255



(a) Normal class

(b) Abnormal class

FIGURE 9. ROC curve comparisons between BSVM and different pretrained networks.

descriptive features, which leads to increase in the performance. For any input video sequence, the aim is to determine whether that sequence is normal or abnormal. Therefore, the classification problem here is a two-class problem. Moreover, the training and testing sets are prepared to contain both normal and abnormal frames.

Table 3 represents the detection accuracies for BSVM applying different descriptor: histogram of oriented gradients (Hog), background subtraction (BS), optical flow (OF), and VGGNet-19. As shown, the VGGNet-19 provides the best performance by an increase of 24%.

In addition, we performed a comparison between VGGNet-19 and other CNN pre-trained networks: GoogleNet, ResNet50, AlexNet, and VGGNet-16. As shown in Table 4, VGGNet-19 attained the highest classification accuracy and the lowest EER in all the testing patterns, followed by VGGNet-16, AlexNet, ResNet50, and GoogleNet. Generally, VGGNet variants have the best results and they are considered as suitable models for abnormal detection since they pre-trained on similar images such as bicycles, small carts, cars, and pedestrians.

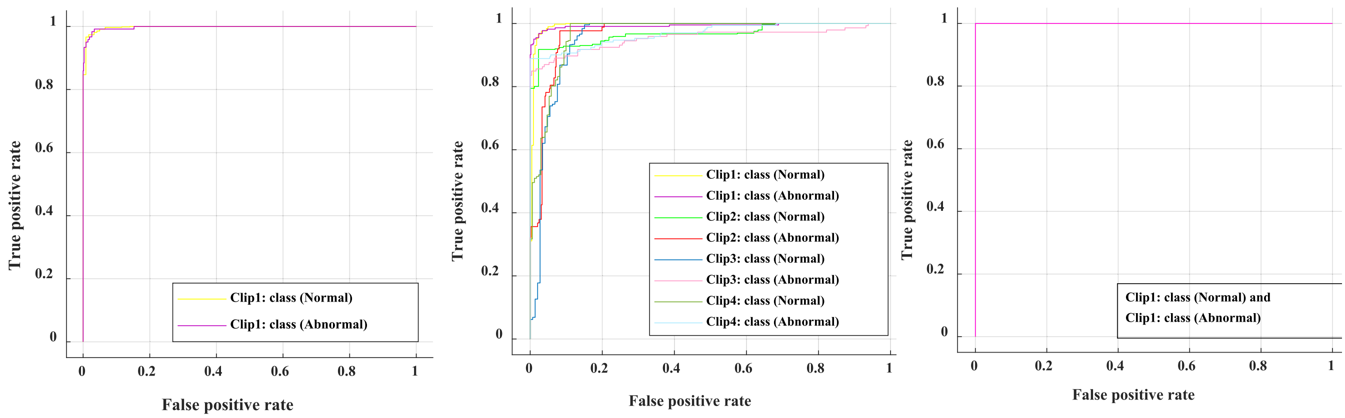
Table 5 presents a comparison between the proposed framework with other contemporary frameworks [17], [61]–[67] using UMN dataset in terms of AUC and accuracy. Note that, the proposed framework provides the best value of AUC and very competitive value of accuracy to Sabokrou *et al.* [67].

TABLE 5. Comparison using UMN with other contemporary frameworks.

Framework	AUC	Accuracy (%)
Hu *et al.* [62]	0.9770	-
Cong *et al.* [61]	0.9730	-
Lu *et al.* [63]	0.7010	-
Biswas and Babu [64]	0.7360	-
Leyva *et al.* [65]	0.8830	-
Pennisi *et al.* [66]	0.9500	-
Sabokrou *et al.* [67]	-	97.50
Bansod and Nandedkar [17] set A	-	96.35
Bansod and Nandedkar [17] set B	-	96.99
Proposed	0.9795	97.44

Similarly, we compare the proposed framework with other contemporary frameworks [61]–[64], [68]–[70] in terms of accuracy and EER in Tables 6 and 7 using UCSD-PED1 dataset. It is noted that the proposed framework presents the lowest EER value and the best accuracy among all other frameworks.

We also report the frame-level ROC curve for UMN dataset using our proposed framework with five pre-trained models. These results are shown in Fig. 9. It can be observed that with the use of VGGNet-19, our framework outperforms all the other pre-trained models. The results of frame-level ROC curves for all UMN scenes are shown in Fig. 10. Therefore, it is clear that all UMN video clips for all scenes keep its ROC levels high.



(a) Lawn scene (b) Indoor scene (c) Plaza scene

FIGURE 10. ROC curves for UMN dataset scenes.

TABLE 6. Comparison using UCSD-PED1 with other contemporary frameworks in terms of EER.

Framework	EER (%)
Roshtkhari and Levine [68]	15.00
Hu *at al.* [62]	18.00
Cheng *et al.* [69]	19.90
Cong *at al.* [61]	23.00
Zhu *at al.* [70]	15.00
Lu *at al.* [63]	15.00
Biswas and Babu [64]	24.66
Proposed	13.31

TABLE 7. Comparison using UCSD-PED1 with other contemporary frameworks in terms of accuracy.

Framework	Accuracy (%)
Zhou *et al.* [48]	76.00
Xu *at al.* [71]	84.00
Revathi and Kumar [72]	85.00
Xingjian *at al.* [73]	84.90
Tran and Hogg [74]	85.20
Bansod and Nandedkar [17]	82.92
Proposed	86.69

VI. CONCLUSION

The novelty of this study is in using deep learning via transfer learning approaches to detect abnormal incidents. In this paper, we propose an abnormal event detection framework in videos based on VGGNet and BSVM, which provides good accuracy in complex motion scenarios. Specifically, we use VGGNet-19 to extract high-level features and then adopt the BSVM algorithm to establish the abnormal event detection model. The conducted experiments on two common datasets show that our framework can realize the automatic detection of abnormal behavior, and has better performance compared with the classical frameworks.

In addition, we conducted a comparison between hand-crafted descriptors and VGGNet-19 as a fixed feature extractor. The experimental results show that VGGNet-19 obtained better accuracy than other hand-crafted descriptors with an average accuracy of 97.44%. Furthermore, we compared VGGNet-19 with other pre-trained CNN networks such as GoogleNet, ResNet50, AlexNet, and VGGNet-16, and the VGGNet-19 shows the best performance, which means that VGGNet-19 can extract very descriptive features that help to detect more abnormal incidents. Results achieved so far strongly urge for future work with other real-world datasets to test and improve the proposed framework. Further, one future work could be to extend the proposed framework to other video monitoring applications such as detecting child abuse.

REFERENCES

[1] A. Al-Dhamari, R. Sudirman, and N. H. Mahmood, "Abnormal behavior detection in automated surveillance videos: A review," *J. Theor. Appl. Inf. Technol.*, vol. 95, no. 19, pp. 5245–5263, 2017.

[2] A. Al-Dhamari, R. Sudirman, N. H. Mahmood, N. H. Khamis, and A. Yahya, "Online video-based abnormal detection using highly motion techniques and statistical measures," *Telkomnika*, vol. 17, no. 4, pp. 2039–2047, 2019.

[3] H. Drucker, D. Wu, and V. N. Vapnik, "Support vector machines for spam categorization," *IEEE Trans. Neural Netw.*, vol. 10, no. 5, pp. 1048–1054, Sep. 1999.

[4] Z. Jia, W. Wu, L. Ren, H. Li, and Z. Wang, "Pipeline abnormal classification based on support vector machine using FBG hoop strain sensor," *Optik*, vol. 170, pp. 328–338, Oct. 2018.

[5] A. Tharwat, A. E. Hassanien, and B. E. Elnaghi, "A BA-based algorithm for parameter optimization of support vector machine," *Pattern Recognit. Lett.*, vol. 93, pp. 13–22, Jul. 2017.

[6] Z. Fang, F. Fei, Y. Fang, C. Lee, N. Xiong, L. Shu, and S. Chen, "Abnormal event detection in crowded scenes based on deep learning," *Multimedia Tools Appl.*, vol. 75, no. 22, pp. 14617–14639, Nov. 2016.

[7] Y. Chen, J. Qian, and V. Saligrama, "A new one-class SVM for anomaly detection," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process.*, Vancouver, BC, Canada, May 2013, pp. 3567–3571.

[8] J. Qian and V. Saligrama, "New statistic in P-value estimation for anomaly detection," in *Proc. IEEE Stat. Signal Process. Workshop (SSP)*, Ann Arbor, MI, USA, Aug. 2012, pp. 393–396.

[9] B. Schölkopf, J. C. Platt, J. Shawe-Taylor, A. J. Smola, and R. C. Williamson, "Estimating the support of a high-dimensional distribution," *Neural Comput.*, vol. 13, no. 7, pp. 1443–1471, Jul. 2001.

[10] X. Hu, Y. Huang, H. Zhang, H. Wu, and S. Hu, "Video anomaly detection using deep incremental slow feature analysis network," *IET Comput. Vis.*, vol. 10, no. 4, pp. 258–267, Jun. 2016.

[11] N. C. Tay, T. Connie, T. S. Ong, K. O. M. Goh, and P. S. Teh, "A robust abnormal behavior detection method using convolutional neural network," in *Proc. Comput. Sci. Technol. Conf.*, Singapore, 2019, pp. 37–47.
[12] K. Simonyan and A. Zisserman, "Two-stream convolutional networks for action recognition in videos," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2014, pp. 568–576.
[13] T.-H. Vu, A. Osokin, and I. Laptev, "Context-aware CNNs for person head detection," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Santiago, Chile, Dec. 2015, pp. 2893–2901.
[14] H. Zenati, C. Sheng Foo, B. Lecouat, G. Manek, and V. Ramaseshan Chandrasekhar, "Efficient GAN-based anomaly detection," 2018, *arXiv:1802.06222*. [Online]. Available: http://arxiv.org/abs/1802.06222
[15] H. Salehinejad, S. Sankar, J. Barfett, E. Colak, and S. Valaee, "Recent advances in recurrent neural networks," 2017, *arXiv:1801.01078*. [Online]. Available: http://arxiv.org/abs/1801.01078
[16] S. P. K. Karri, D. Chakraborty, and J. Chatterjee, "Transfer learning based classification of optical coherence tomography images with diabetic macular edema and dry age-related macular degeneration," *Biomed. Opt. Express*, vol. 8, no. 2, p. 579, Feb. 2017.
[17] S. Bansod and A. Nandedkar, "Transfer learning for video anomaly detection," *J. Intell. Fuzzy Syst.*, vol. 36, no. 3, pp. 1967–1975, Mar. 2019.
[18] M. A. Wani, F. A. Bhat, S. Afzal, and A. I. Khan, *Advanced Machine Learning*, vol. 57. Cham, Switzerland: Springer, 2020.
[19] Z. Zhong, L. Jin, and Z. Xie, "High performance offline handwritten chinese character recognition using GoogLeNet and directional feature maps," in *Proc. 13th Int. Conf. Document Anal. Recognit. (ICDAR)*, Tunis, Tunisia, Aug. 2015, pp. 846–850.
[20] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.
[21] H.-W. Kang and H.-B. Kang, "Prediction of crime occurrence from multi-modal data using deep learning," *PLoS ONE*, vol. 12, no. 4, 2017, Art. no. e0176244.
[22] H. Huang, Y. Chong, C. Nie, and S. Pan, "Hand gesture recognition with skin detection and deep learning method," *J. Phys., Conf. Ser.*, vol. 1213, Jun. 2019, Art. no. 022001.
[23] V. Vapnik, *Statistical Learning Theory*, vol. 3. New York, NY, USA: Wiley, 1998.
[24] I. El-Naqa, Y. Yang, M. N. Wernick, N. P. Galatsanos, and R. M. Nishikawa, "A support vector machine approach for detection of microcalcifications," *IEEE Trans. Med. Imag.*, vol. 21, no. 12, pp. 1552–1563, Dec. 2002.
[25] A. H. Mohammad, T. Alwada'n, and O. Al-Momani, "Arabic text categorization using support vector machine, Naïve Bayes and neural network," *GSTF J. Comput. (JoC)*, vol. 5, no. 1, p. 108, 2018.
[26] M. P. S. Brown, W. N. Grundy, D. Lin, N. Cristianini, C. W. Sugnet, T. S. Furey, M. Ares, and D. Haussler, "Knowledge-based analysis of microarray gene expression data by using support vector machines," *Proc. Nat. Acad. Sci. USA*, vol. 97, no. 1, pp. 262–267, Jan. 2000.
[27] M. Vatsa, R. Singh, and A. Noore, "Improving biometric recognition accuracy and robustness using DWT and SVM watermarking," *IEICE Electron. Express*, vol. 2, no. 12, pp. 362–367, 2005.
[28] J.-P. Doucet, F. Barbault, H. Xia, A. Panaye, and B. Fan, "Nonlinear SVM approaches to QSPR/QSAR studies and drug design," *Current Comput. Aided-Drug Design*, vol. 3, no. 4, pp. 263–289, Dec. 2007.
[29] S. Lekha and M. Suchetha, "A novel 1-D convolution neural network with SVM architecture for real-time detection applications," *IEEE Sensors J.*, vol. 18, no. 2, pp. 724–731, Jan. 2018.
[30] M. K. Das, "Electrocardiogram signal analysis for heartbeat pattern classification," Ph.D. dissertation, Nat. Inst. Technol., Rourkela, India, 2015.
[31] S. Ari, K. Hembram, and G. Saha, "Detection of cardiac abnormality from PCG signal using LMS based least square SVM classifier," *Expert Syst. Appl.*, vol. 37, no. 12, pp. 8019–8026, Dec. 2010.
[32] M. Achirul Nanda, K. B. Seminar, D. Nandika, and A. Maddu, "A comparison study of kernel functions in the support vector machine and its application for termite detection," *Information*, vol. 9, no. 1, p. 5, 2018.
[33] E. Çomak and A. Arslan, "A new training method for support vector machines: Clustering k-NN support vector machines," *Expert Syst. Appl.*, vol. 35, no. 3, pp. 564–568, Oct. 2008.
[34] N. J. Nilsson, *Learning Machines: Foundations of Trainable Pattern-Classifying Systems*. New York, NY, USA: McGraw-Hill, 1965.
[35] J. Friedman, T. Hastie, and R. Tibshirani, *The Elements of Statistical Learning*, vol. 1, no. 10. New York, NY, USA: Springer, 2001,
[36] A. Tharwat, "Parameter investigation of support vector machine classifier with kernel functions," *Knowl. Inf. Syst.*, vol. 61, no. 3, pp. 1269–1302, Dec. 2019.
[37] A. A. Sodemann, M. P. Ross, and B. J. Borghetti, "A review of anomaly detection in automated surveillance," *IEEE Trans. Syst., Man, Cybern. C, Appl. Rev.*, vol. 42, no. 6, pp. 1257–1272, Nov. 2012.
[38] G. Wang, J. Yang, and R. Li, "Imbalanced SVM-based anomaly detection algorithm for imbalanced training datasets," *ETRI J.*, vol. 39, no. 5, pp. 621–631, Oct. 2017.
[39] S. Fu, J. Liu, and H. Pannu, "A Hybrid anomaly detection framework in cloud computing using one-class and two-class support vector machines," in *Proc. Int. Conf. Adv. Data Mining Appl.* Nanjing, China: Springer, 2012, pp. 726–738.
[40] S. M. Hosseini Bamakan, H. Wang, T. Yingjie, and Y. Shi, "An effective intrusion detection framework based on MCLP/SVM optimized by time-varying chaos particle swarm optimization," *Neurocomputing*, vol. 199, pp. 90–102, Jul. 2016.
[41] S.-H. Cho and H.-B. Kang, "Abnormal behavior detection using hybrid agents in crowded scenes," *Pattern Recognit. Lett.*, vol. 44, pp. 64–70, Jul. 2014.
[42] H. Kim, S. Lee, Y. Kim, S. Lee, D. Lee, J. Ju, and H. Myung, "Weighted joint-based human behavior recognition algorithm using only depth information for low-cost intelligent video-surveillance system," *Expert Syst. Appl.*, vol. 45, pp. 131–141, Mar. 2016.
[43] G. Shu, G. Fu, P. Li, and H. Geng, "Violent behavior detection based on SVM in the elevator," *Int. J. Secur. Appl.*, vol. 8, no. 5, pp. 31–40, Sep. 2014.
[44] J.-Y. H. W.-P. Lee, "A smart camera network with SVM classifiers for crowd event recognition," in *Proc. World Congr. Eng.*, vol. 1, 2014, pp. 1–6.
[45] H. Liu, Z. Ju, X. Ji, C. S. Chan, and M. Khoury, "Study of human action recognition based on improved spatio-temporal features," in *Human Motion Sensing and Recognition: A Fuzzy Qualitative Approach*. Berlin, Germany: Springer, 2017, pp. 233–250.
[46] Y. S. Chong and Y. H. Tay, "Abnormal event detection in videos using spatiotemporal autoencoder," in *Proc. Int. Symp. Neural Netw.* Hokkaido, Japan: Springer, 2017, pp. 189–196.
[47] L. F. Borja-Borja, M. Saval-Calvo, and J. Azorin-Lopez, "A short review of deep learning methods for understanding group and crowd activities," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2018, pp. 1–8.
[48] S. Zhou, W. Shen, D. Zeng, M. Fang, Y. Wei, and Z. Zhang, "Spatial–temporal convolutional neural networks for anomaly detection and localization in crowded scenes," *Signal Process., Image Commun.*, vol. 47, pp. 358–368, Sep. 2016.
[49] M. Hasan, J. Choi, J. Neumann, A. K. Roy-Chowdhury, and L. S. Davis, "Learning temporal regularity in video sequences," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 733–742.
[50] Y. Li, "A deep spatiotemporal perspective for understanding crowd behavior," *IEEE Trans. Multimedia*, vol. 20, no. 12, pp. 3289–3297, Dec. 2018.
[51] M. Gnouma, A. Ladjailia, R. Ejbali, and M. Zaied, "Stacked sparse autoencoder and history of binary motion image for human activity recognition," *Multimedia Tools Appl.*, vol. 78, no. 2, pp. 2157–2179, Jan. 2019.
[52] N. Sjarif, S. Shamsuddin, and S. Hashim, "Detection of abnormal behaviors in crowd scene: A review," *Int. J. Advance. Soft Comput. Appl*, vol. 4, no. 1, pp. 1–33, 2012.
[53] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," 2014, *arXiv:1409.1556*. [Online]. Available: http://arxiv.org/abs/1409.1556
[54] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2009, pp. 248–255.
[55] A. M. Deris, A. M. Zain, and R. Sallehuddin, "Overview of support vector machine in modeling machining performances," *Procedia Eng.*, vol. 24, pp. 308–312, Jan. 2011.
[56] J. A. Rohwer, C. T. Abdallah, and C. G. Christodoulou, "Least squares support vector machines for fixed-step and fixed-set CDMA power control," in *Proc. 42nd IEEE Int. Conf. Decis. Control*, Maui, HI, USA , vol. 5, Dec. 2003, pp. 5097–5102.
[57] W. Li, V. Mahadevan, and N. Vasconcelos, "Anomaly detection and localization in crowded scenes," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 1, pp. 18–32, Jan. 2014.

[58] A. Tharwat, "Classification assessment methods," *Appl. Comput. Inform.*, 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S2210832718301546

[59] 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.

[60] V. Mahadevan, W. Li, V. Bhalodia, and N. Vasconcelos, "Anomaly detection in crowded scenes," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, San Francisco, CA, USA, Jun. 2010, pp. 1975–1981.

[61] Y. Cong, J. Yuan, and J. Liu, "Sparse reconstruction cost for abnormal event detection," in *Proc. CVPR*, Providence, RI, USA, Jun. 2011, pp. 3449–3456.

[62] Y. Hu, Y. Zhang, and L. S. Davis, "Unsupervised abnormal crowd activity detection using semiparametric scan statistic," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2013, pp. 767–774.

[63] C. Lu, J. Shi, and J. Jia, "Abnormal event detection at 150 FPS in MATLAB," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 2720–2727.

[64] S. Biswas and R. V. Babu, "Real time anomaly detection in H.264 compressed videos," in *Proc. 4th Nat. Conf. Comput. Vis., Pattern Recognit., Image Process. Graph. (NCVPRIPG)*, Dec. 2013, pp. 1–4.

[65] R. Leyva, V. Sanchez, and C.-T. Li, "Video anomaly detection with compact feature sets for online performance," *IEEE Trans. Image Process.*, vol. 26, no. 7, pp. 3463–3478, Jul. 2017.

[66] A. Pennisi, D. D. Bloisi, and L. Iocchi, "Online real-time crowd behavior detection in video sequences," *Comput. Vis. Image Understand.*, vol. 144, pp. 166–176, Mar. 2016.

[67] M. Sabokrou, M. Fayyaz, M. Fathy, and R. Klette, "Deep-cascade: Cascading 3D deep neural networks for fast anomaly detection and localization in crowded scenes," *IEEE Trans. Image Process.*, vol. 26, no. 4, pp. 1992–2004, Apr. 2017.

[68] M. J. Roshtkhari and M. D. Levine, "An on-line, real-time learning method for detecting anomalies in videos using spatio-temporal compositions," *Comput. Vis. Image Understand.*, vol. 117, no. 10, pp. 1436–1452, Oct. 2013.

[69] K.-W. Cheng, Y.-T. Chen, and W.-H. Fang, "Gaussian process regression-based video anomaly detection and localization with hierarchical feature representation," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 5288–5301, Dec. 2015.

[70] X. Zhu, J. Liu, J. Wang, C. Li, and H. Lu, "Sparse representation for robust abnormality detection in crowded scenes," *Pattern Recognit.*, vol. 47, no. 5, pp. 1791–1799, May 2014.

[71] D. Xu, Y. Yan, E. Ricci, and N. Sebe, "Detecting anomalous events in videos by learning deep representations of appearance and motion," *Comput. Vis. Image Understand.*, vol. 156, pp. 117–127, Mar. 2017.

[72] A. R. Revathi and D. Kumar, "An efficient system for anomaly detection using deep learning classifier," *Signal, Image Video Process.*, vol. 11, no. 2, pp. 291–299, Feb. 2017.

[73] S. Xingjian, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 802–810.

[74] H. Tran and D. Hogg, "Anomaly detection using a convolutional Winner-Take-All autoencoder," in *Proc. Procedings Brit. Mach. Vis. Conf.*, London, U.K., Sep. 2017.

AHLAM AL-DHAMARI received the B.Sc. degree in computer engineering from Hodeidah University, in 2008, and the M.Sc. degree in computer engineering and networks from the University of Jordan, Jordan, in 2015. She is currently pursuing the Ph.D. degree in electronic and computer engineering with Universiti Teknologi Malaysia, Malaysia. Her research interests include image and video processing, computer vision, machine learning, deep learning, and computer architectures.

RUBITA SUDIRMAN received the B.Sc. and M.Sc. degrees from the University of Tulsa and the Ph.D. degree in electrical engineering from Universiti Teknologi Malaysia (UTM). She is currently an Associate Professor and a Certified Professional Engineer serving with the School of Electrical Engineering, UTM. Her current research interests include applications of soft computing in biomedical engineering, particularly in speech processing, EEG & EOG signal analysis, medical electronics, and rehabilitation engineering.

NASRUL HUMAIMI MAHMOOD received the B.Eng. degree in electrical and electronic engineering from Universiti Kebangsaan Malaysia (UKM), the M.Eng. degree in electrical and electronic engineering from Universiti Teknologi Malaysia (UTM), and the Ph.D. degree in engineering from the University of Warwick, U.K. His research interests include three-dimensional (3D) object reconstruction from multiple views, biomedical image processing, medical electronics, biomedical materials, and rehabilitation engineering. Since 1999, he has been serving at the Faculty of Electrical Engineering, UTM. Within the Faculty of Electrical Engineering at the UTM, he teaches a number of courses related to his research and elementary to the electrical and electronics curriculum. He was the Head of the Department for Bio-Instrumentation, Signal Processing Department, and Bio-Technology and Medical Engineering Department at the Faculty of Biomedical Engineering and Health Science and Faculty of Bio-Sciences and Medical Engineering, respectively, from 2011 to 2015. He was also a Postgraduate Manager with the Faculty of Electrical Engineering. He is currently an Associate Professor of electronic and computer engineering (ECE) with the School of Electrical Engineering.