This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2019.2896120, IEEE Internet of Things Journal

Energy-Efficient Deep CNN for Smoke Detection in Foggy IoT Environment

Salman Khan, *Student Member, IEEE*, Khan Muhammad, *Member, IEEE*, Shahid Mumtaz, *Senior Member*, *IEEE*, Sung Wook Baik, *Member*, *IEEE*, Victor Hugo C. de Albuquerque, *Member*, *IEEE*

***Abstract*— Smoke detection in IoT environment is a primary component of early disaster-related event detection in smart cities. Recently, several smoke and fire detection methods are presented with reasonable accuracy and running time for normal IoT environment. However, these methods are unable to detect smoke in foggy IoT environment, which is a challenging task. In this article, we propose an energy-efficient system based on deep convolutional neural networks (CNN) for early smoke detection in both normal and foggy IoT environments. Our method takes advantage of VGG-16 architecture, considering its sensible stability between the accuracy and time efficiency for smoke detection compared to the other computationally expensive networks such as GoogleNet and AlexNet. Experiments performed on benchmark smoke detection datasets and their results in terms of accuracy, false alarms rate and efficiency reveal the better performance of our technique compared to state-of-the-art and verifies its applicability in smart cities for early detection of smoke in normal and foggy IoT environments.**

***Index Terms*— CNNs, Smart Cities, Smoke Detection, Fire Alert, Disaster Management, Image Classification, Surveillance, Foggy IoT Environment**

I. Introduction

The increase of smarter surveillance camera technologies and their advance processing proficiencies gain a lot of advantage in the field of real-time video analysis i.e., objects detection [1, 2], object tracking [3, 4] and action recognition [5-7]. With recent revolution of smart devices and its processing capabilities, they are used for various purposes in Internet of things (IoT) environments in smart cities i.e., [8, 9]. In video-based fire detection systems [10-12] attracted major attention in the current era compared to conventional sensors. Fire can bring massive damages to human lives and economics, thus automatic fire detection system is very important for disaster management to handle the fire on time. Smoke is a primary sign of fire and its early detection provides a convenient way to avoid fire accidents, but it is very difficult to detect smoke using sensors in outdoor IoT environment [13]. As these sensors have a very limited range and outdoor environment needs a wide range of detection system to cover up all areas, which can be made possible through vision sensors. It is a known phenomenon that smoke spreads faster and can be identified from far away using vision sensors. However, its automatic detection using computer vision techniques is still an open challenge for the research community due to various reasons such as the similarity with other natural objects like clouds and shadows, inconsistency of smoke density and motion and complex visual pattern of smoke [14]. The current literature shows that several researchers have presented smoke detection methods based on shape, color, texture and motion features. For instance, [15-17] used color based decision for detection of smoke regions. Chunyu et al. [15] proposed a smoke detection method based on motion and color features by making use of optical flow and back propagation neural network for classification of smoke. Calderara et al. [16] took advantage of image energy and color information for development of smoke detection system. Nguyen et al. [17] developed a smoke detection technique which uses fuzzy C-mean for clustring and back propegation neural network for the classificaiton of smoke based on color features. Besides color and motion, shape features are also exploited for smoke detection. For instance, [18-20]explore different shape features including contours, histograms of edge orientation, spectral, spatial and temporal characteristics of smoke. Similarly, several other smoke detection methods that use motion features, can be found in [21-23]. The motion features used by these methods are, accumulative motion orientation model, approximate median and block processing technique. Finally, [24-26] used textue features including local binary pattern, wavelet and gray level co-occurance matrix. The main issues with all these color, shape, motion and texture features based methods are their low accuracy, high false alarm rate and failure to detect smoke from larger distance or smoke of small size.

To tackle these issues, recently different smoke detection methods are introduced. As, Dimitropoulos et al. [27] proposed a new higher order linear dynamical smoke detection method based on higher order decomposition of multidimensional pictorial data. Their method represents the video subsequences as histograms of high order dynamical system descriptors made by smoke extents in each subsequence and improves the classification accuracy by combining spatiotemporal modeling with multidimensional dynamic texture analysis of smoke using a particle swarm optimization approach. The main limitation of this work is low frame rate due to high computation and low detection rate. Yuan et al. [28] presented a smoke detection technique that uses Haar-like and statistical features with a staircase searching technique and dual threshold AdaBoost classifier. Their algorithm cannot detect some parts of the smoke due to high variation of color and density of smoke and

Manuscript received September 26, 2018; Accepted: XXX, Published: XXXX. This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No.2016R1A2B4011712). This paper was recommended by Associate Editor XYZ. (*Corresponding author: Sung Wook Baik*)

Salman Khan, Khan Muhammad and Sung Wook Baik are with Intelligent Media Laboratory, Digital Contents Research Institute, Sejong University, Seoul 143-747, Republic of Korea (Email: salmank@ieee.org, khan.muhammad@ieee.org, sbaik@sejong.ac.kr)

Shahid Mumtaz is with Campus Universitario de Santiago, Portugal (Email: smumtaz@av.it.pt)

Victor Hugo C. de Albuquerque is with Graduate Program in Applied Informatics at the Universidade de Fortaleza, Fortaleza/CE, Brazil (Email: victor.albuquerque@unifor.br)

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2019.2896120, IEEE Internet of Things Journal

with a very low accuracy in some videos. A more recent fast smoke detection method in video surveillance using compute unified device architecture (CUDA) is proposed in [29]. Their method depends upon shape and color features of the smoke regions and for the fast processing, hybrid approach of utilizing CPU and GPGPU using CUDA are used.

The above-mentioned existing state-of-the-art smoke detection approaches are applicable to normal environment only with low performance in foggy and uncertain environment. Furthermore, computationally expensive methods are not recommended for real-time processing. Moreover, the smoke detection methods still need improvement in terms of false alarm rate and accuracy. To address these problems, in this article we propose an energy-efficient smoke detection framework for early detection of smoke in normal and foggy environment. More precisely, the major contributions of our method are summarized as follows:

1. We investigated different state-of-the-art CNN models for smoke detection in foggy IoT environment and propose an energy-efficient CNN framework for early smoke detection in normal and foggy environment. The reasonable computational complexity of our CNN architecture along with its accuracy and model size, make it a suitable system for smoke detection in IoT-assisted smart cities compared to state-of-the-art.
2. From literature it is evident that pervious methods contain benchmark datasets but these are specially captured in normal environment only. We created our own dataset for detection of smoke in foggy environment for benchmarking purposes. The dataset will be publicly available to researchers for tuning their smoke detection methods for foggy environment.
3. Unlike state-of-the-art methods, that are not appropriate to identify smoke in uncertain or foggy environment, our proposal can detect smoke in certain as well as in uncertain environment, thus fulfills the constraints and requirements of smart cities, increasing its suitability
4. We conducted extensive experimentations on the newly created dataset and the previous benchmark datasets. The incurred results show that the proposed framework overwhelms other state-of-the-art methods in terms of false alarm rate and accuracy.

The rest of this article is arranged as follows. The proposed methodology of our system is disclosed in Section 2. The experimental evaluations are described in Section 3. Lastly, the paper is concluded in Section 4 with its key findings and future directions.

II. The Proposed Framework

Smoke detection using hand-crafted features in smart cites surveillance is a tedious job, particularly when the smoke is in its initial stages or at long distance or in uncertain environment. The existing systems that use the traditional features extraction approach, produce a higher number of false alarm rate with limited smoke detection accuracy. Recently, several smoke detection methods are explored whose performance is reasonable as compared to the traditional methods. However, the major issue with these methods is their limited performance for uncertain and foggy environment. To overcome these issues, we investigated different CNN architectures and proposed an energy-efficient deep CNN based technique for smoke detection in surveillance videos captured in foggy and normal IoT environment. Our framework for smoke detection is overviewed in Fig. 1 whereas description of the proposed model parameters and implementation steps are given in Table 1 and Algorithm 1, respectively.

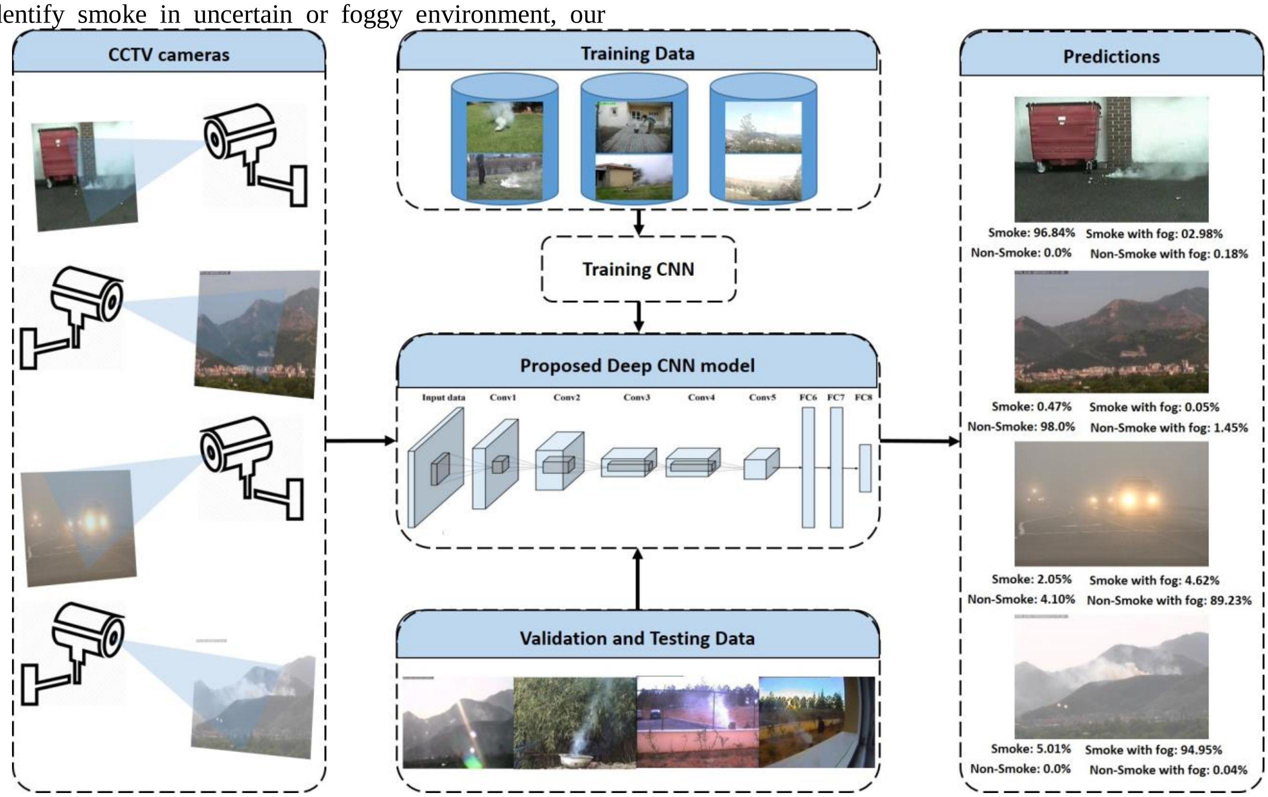

Fig. 1. Energy-efficient deep CNN for smoke detection in certain and uncertain IoT environment.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2019.2896120, IEEE Internet of Things Journal

A. The Proposed CNN Architecture for Smoke Detection

In our proposed method we explored and compared several CNN models with different parameter settings for smoke detection in both normal and foggy environments. After the comparison of different CNNs, we found that VGG-16 [30] is better than other models i.e., AlexNet [31] and GoogleNet [32]. For smoke detection in uncertain IoT environment, we modified the architecture of VGG-16 model according to our problem. Similar to other CNNs, our current VGG-16 is pre-trained on ImageNet [33] dataset for classification of images into 1000 different categories. We fine-tuned this model with our own created dataset by adjusting the last fully connected layer from 1000 to four classes to perform the intended classification of smoke and non-smoke in both normal and foggy IoT environments. Thus, modifying the last layer of the aforementioned architecture enabled our system to detect the smoke patterns effectively. The architecture of our employed CNN model is given in Fig. 2.

The size of input image to the architecture is fixed to 224 × 224 x 3 pixels. Each image is passed through five different convolutions of the architecture. First convolution comprises of two convolutional layers with the input size of 224 × 224 where 64 kernels of size 3 × 3 with stride 1 are applied. The result is then forward propagated to the max pooling layer with 2 × 2 kernel and stride 2 to get the maximum activations from feature maps. The second convolution is same as the previous one, consisting of two convolution layers with the input size of 112 × 112 followed by a max pooling. Third convolution consists of three convolutional layers with 256 kernels of size 3 × 3 with stride 1 to the input of size 56 × 56 followed by a max pooling. The next two convolutions are same as the third one with input size of 28 × 28 and 14 × 14, respectively. A stack of these convolutional layers is followed by three fully-connected layers. The first and second fully-connected layers have 4094 channels apiece, while the third fully-connected layer is modified from 1000 to 4 channels for classification of smoke and non-smoke in certain and uncertain environment. Finally, Softmax classifier is used to predict the probabilities of four target classes (Smoke, Smoke with fog, Non-Smoke and Non-Smoke with fog).

B. Motivations of Using VGG-16 for Smoke Detection

The architecture of VGG-16 is different compared to other state-of-the-art models i.e., GoogleNet and AlexNet. In these models, filter size of initial convolutional layers is 11 × 11 or 7 × 7 with strides of 3 to 5. With the larger size of convolution and wide strides, the convolutional filters can miss the important patterns of smoke area during the training process. The motivation of using VGG-16 is because of its 3 × 3 filter size for all convolutional layers with strides size of 1. Such setup helps the architecture to process and extract features from each pixel of the input image. On the other hand, VGG-16 is slightly different from VGG-19 with less number of convolutions and parameters and almost similar accuracy. The statistical comparison of VGG-16 with other CNN architectures is given in Table II. It can be observed from Table II that VGG-16 is better than other state-of-the-art architectures [34] in terms of top-1 accuracy, top-5 accuracy and top-5 test error rate on large scale ImageNet dataset [35]. This motivated us to employ

Table I
Description of the model parameters in the proposed method.

α	Training data	***FP***	False positive
β	Testing data	***FN***	False negative
γ	Validation data	***A***	Accuracy
Ω	Trained model	***P***	Precision
π	Validation accuracy graph	***R***	Recall
D	Dataset	***F***	F-measure
A	Accuracy		

Table II
Comparative statistics of VGG-16 with other architectures.

Models	Parameters (millions)	Top-1 Accuracy (%)	Top-5 Accuracy (%)	Top-5 test error (%)
GoogleNet	60	69.8	89.3	7.9
AlexNet	**7**	57.1	80.2	16.4
VGG-16	138	**70.5**	**91.0**	**7.0**

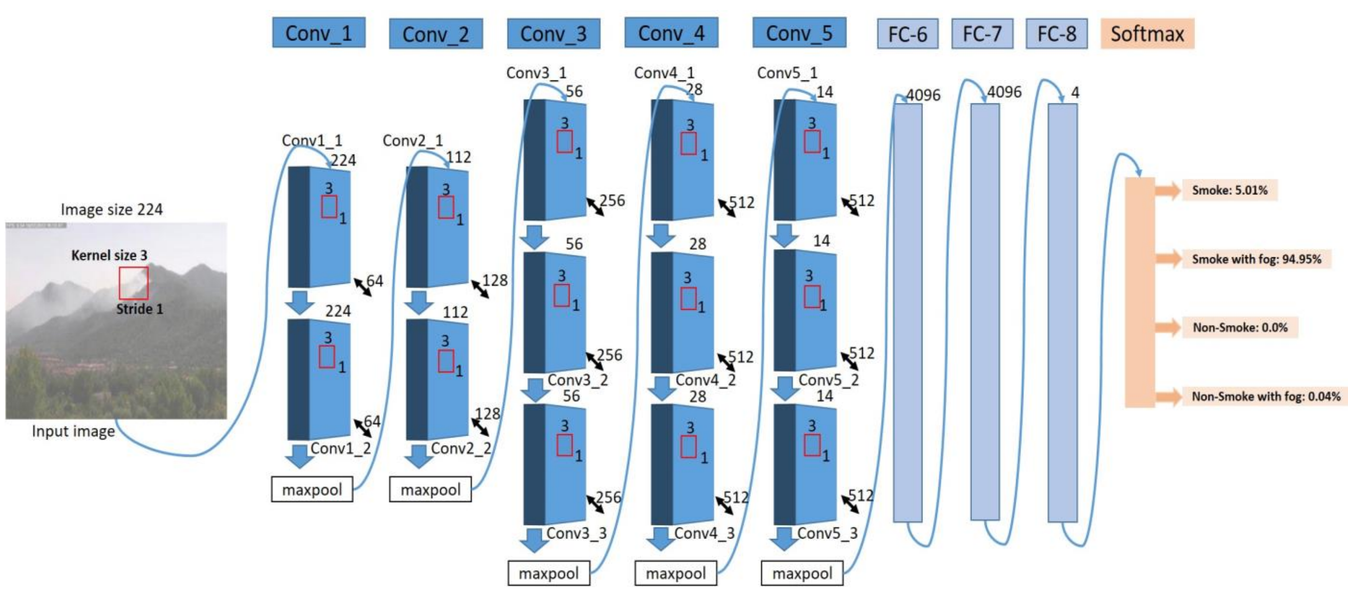

Fig. 2. Overview of the proposed deep CNN architecture for smoke detection in certain and uncertain environment. An input image is passed through five convolutions containing 13 convolutional layers, followed by three fully connected layers. The Softmax classifier provides the final predictions for the intended four classes including "Smoke", "Smoke with fog", "Non-Smoke" and "Non-Smoke with fog".

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2019.2896120, IEEE Internet of Things Journal

a similar architecture for smoke detection in foggy environment, considering the requirements of smart cities in terms of recognition accuracy, false alarms, running time and energy-related constraints.

Algorithm 1. Training and testing of the proposed system.
Input: Dataset ***D***, Train_Val.prototxt, split
Initialization: ***€*** = load VGG-16 pretrained model, split = [0.2, 0.3, 0.5], training_parameters = [epoch = 30, learning rate = 0.001, batch size = 16, validation interval (in epochs) = 1, solver type = SGD (Stochastic Gradient Descent)]
1. [***α, β, γ***] = prepare_data (***D***, split) ($Data_{train}$, $Data_{val}$, $Data_{test}$) = (split[0], split[1], split[2]) ***α*** = random(***D***, $Data_{train}$) ***β*** = random(***D***, $Data_{val}$) ***γ*** = random(***D***, $Data_{test}$) **return** [***α, β, γ***]
2. [***Ω, π***] = ***€*** (***α, β,*** training_parameters)
3. [***FP, FN, A, P, R, F***] = ***Ω*** (***γ***)
4. **Output**: [***FP, FN, A, P, R, F***]

III. Experimental Results and Discussion

In this section, we present the detail about the datasets used in the experimental assessment of the proposed system. Next, we compare the performance of our proposed architecture with other state-of-the-art through different evaluation schemes. Furthermore, we compare the results of our employed method with recent smoke detection methods. Finally, to prove that our system is functional in smart cities we check the adaptability and performance of our system in real-time.

A. Details of the Datasets

The dataset used for experimentation consists of images with four different classes: “Non-Smoke”, “Smoke”, “Non-Smoke with fog” and “Smoke with fog”. In this dataset, the Non-Smoke and Smoke classes belong to three benchmark datasets [36-38] while the other two classes “Non-Smoke with fog” and “Smoke with fog” are synthetically created by adding fog to each image of the first two classes. The dataset comprises of a total of 72012 images, extracted from videos. For the evaluation of our system, we used our recent strategy [11] by using 20% data for training and the rest of 30% and 50% for validation and testing, respectively. Using this distribution of data, our model is trained with 3495 Non-Smoke images, 3706 Smoke images, 3495 Non-Smoke with fog images and 3706 Smoke with fog images. Table III represent the overall statistics of training, validation and testing data while representative images are visualized in Fig. 3. For comparison of the proposed system with state-of-the-art techniques [29] and [28], we also used the seven publicly accessible videos for testing [39, 40]. The complete details, name of each video, duration, frame rate and description are given in Table IV. For simplicity, these videos are termed as V1 to V7, respectively. The representative images of these seven test videos are given in Fig. 4.

B. Comparison with other CNN models

In this section, the comparison of our proposed model with other CNN models is presented using the validation and test set of the overall integrated dataset. The validation accuracy of our proposed model and other state-of-the-art architecture with respect to each epoch is visualized in Fig. 5. To further evaluate, the performance of our method, we have used two types of evaluation matrices for each CNN model. First scheme of evaluation used false-positive (false alarm rate), false negative and accuracy through which our system is compared with AlexNet and GoogleNet and the results are given in Table V.

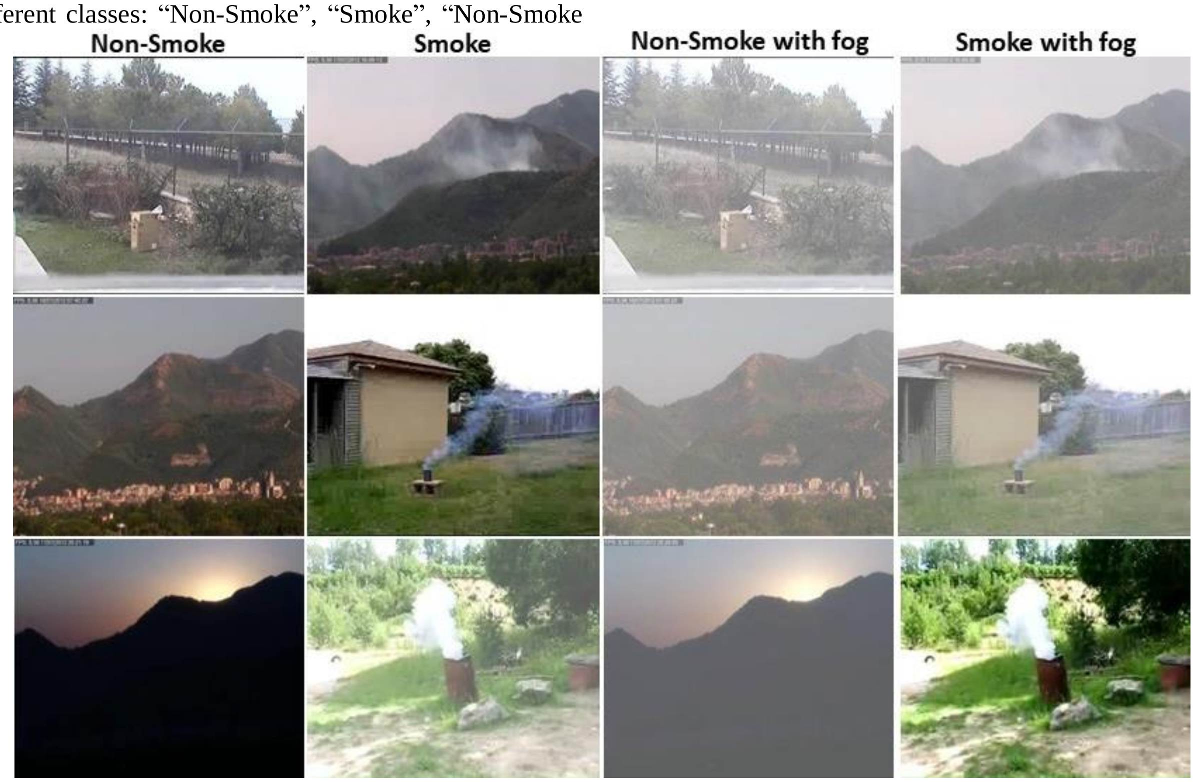

Fig. 3. Representative images of Non-Smoke, Smoke, Non-Smoke with fog and Smoke with fog from the dataset.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2019.2896120, IEEE Internet of Things Journal

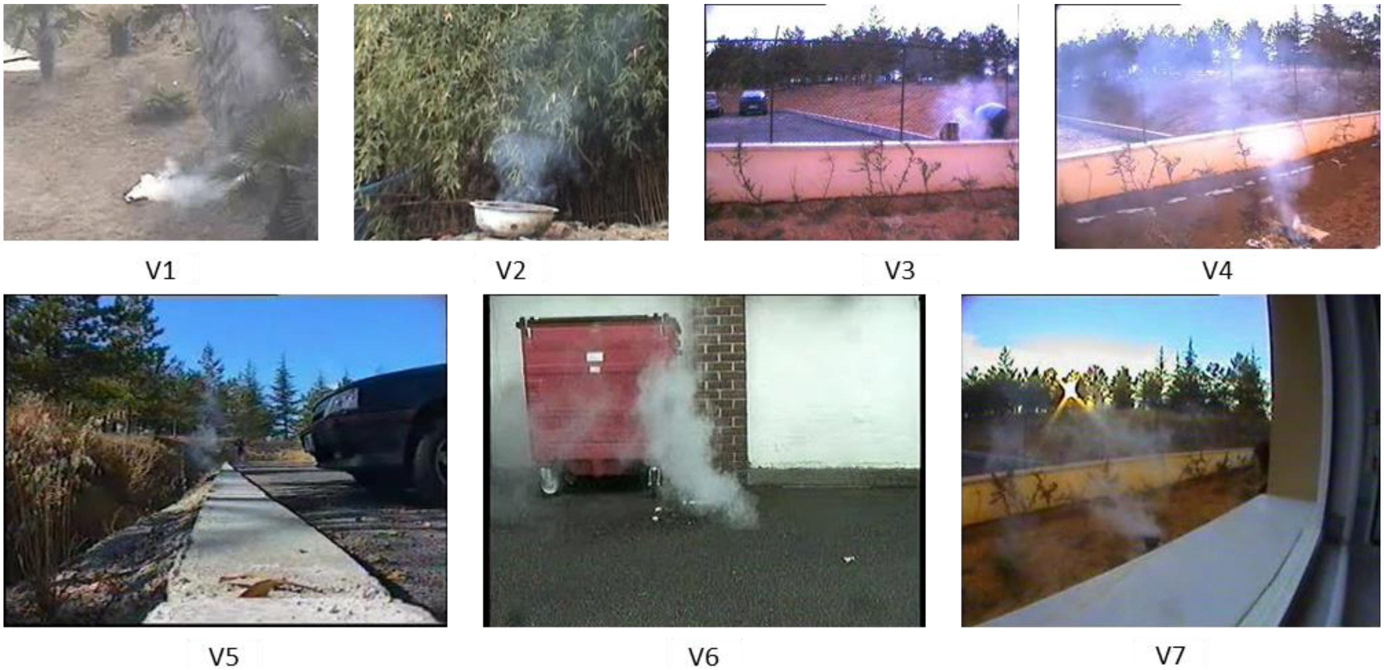

Fig. 4. Representative images of seven smoke videos.

Table III
Overall statistics of training, validation and testing data for the proposed system.

Data distribution	Total images	Data percentage	Non-Smoke	Smoke	Non-Smoke with fog	Smoke with fog
Training Data	14402	20%	3495	3706	3495	3706
Validation Data	21604	30%	5242	5560	5242	5560
Testing Data	36006	50%	8737	9266	8737	9266

TABLE IV
Detail of test videos.

Video Number	Name	Duration (secs)	Frame rate	Description
V1	Cotton_rope_smoke _04.avi	115	25	Smoke from cotton rope with person standing around it
V2	Dry_leaf_smoke_02. avi	48	25	Smoke from dry leaves
V3	sBtFence2.avi	140	10	Smoke at long distance including smoke color background with moving person
V4	sMoky.avi	60	15	Including smoke color background
V5	sParkingLot.avi	69	25	Smoke in a parking lot with moving object and tree shaking
V6	sWasteBasket.avi	90	10	Smoke near a red color waste basket
V7	sWindow.avi	16	15	Smoke in bucket captured from long distance window

Table V
Comparative results using test data on evaluation scheme1.

Model	***FP*** (%)	***FN*** (%)	***A*** (%)
AlexNet	3.39	4.16	95.87
GoogleNet	3.17	2.01	96.11
Proposed	**2.30**	**2.01**	**97.72**

It can be observed from Table V that AlexNet achieved the worst accuracy, false-positive and false negative score as compared to other models. GoogleNet attained better results than AlexNet but its accuracy is still low with high false alarm rate as compared to our proposed method. The proposed method achieves much better results from the previous two models and has minimum false alarm rate of 2.30, minimum false-negatives rate of 2.01 and the highest accuracy of 97.72%. Thus using the first scheme of evaluation, the better performance of our approach can be witnessed compared to other models.

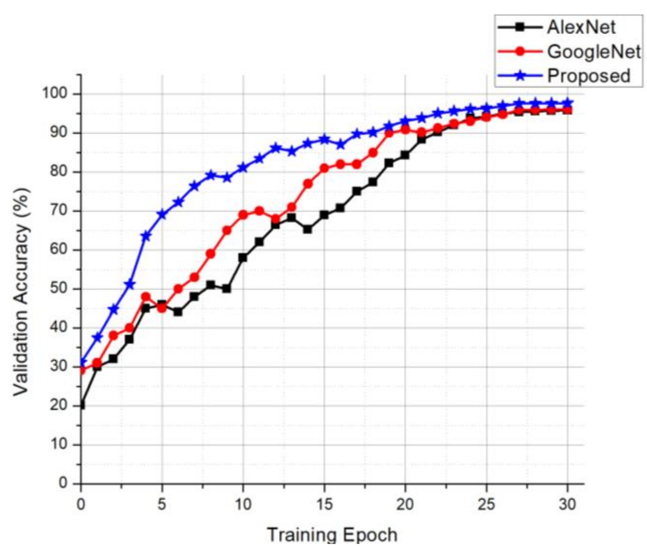

Fig. 5. Validation accuracy of different CNN models for each training epoch.

We also use the second scheme of evaluation to further inspect the performance of our architecture by comparing with other state-of-the-art models. In the second scheme, evaluation

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/JIOT.2019.2896120, IEEE Internet of Things Journal

matrices consist of ***P***, ***R*** and ***F*** where ***P*** is considered as the predicted positive rate calculated for a system and given in Eq. 1. Whereas ***R*** in contrast, refers to sensitivity (true positive rate) of a system and given in Eq. 2. As ***P*** and ***R*** refer to true positive and true negative rate, thus using ***P*** or ***R*** individually for performance evaluation of a system is a biased decision. To evaluate our experimental results more accurately we used ***F*** along with ***P*** and ***R***. The value of ***F*** is calculated by computing the harmonic mean of ***P*** and ***R*** given in Eq. 3 by considering both true positive/negative rate of the system. Experimental results using the second evaluation scheme on a test set are shown is Table VI. From the results, it can be perceived that the ***P*** and ***F*** of AlexNet and GoogleNet are same but in terms of ***R*** GoogleNet is better than AlexNet. Our proposed method dominated both models i.e., GoogleNet and AlexNet using all evaluation matrices of second scheme. Remarkably, our proposed architecture overwhelms all the state-of-the-art models in terms of both evaluation schemes, which shows its effectiveness.

$$\boldsymbol{P} = \frac{\mathrm{TP}}{\mathrm{TP+FP}} \tag{1}$$

$$\boldsymbol{R} = \frac{\mathrm{TP}}{\mathrm{TP+FN}} \tag{2}$$

$$\boldsymbol{F} = 2 \times \left(\frac{\mathrm{P \times R}}{\mathrm{P+R}}\right) \tag{3}$$

Table VI
Comparative results using test data on evaluation scheme2.

Model	***P***	***R***	***F***
AlexNet	0.96	0.95	0.96
GoogleNet	0.96	0.96	0.96
Proposed	**0.98**	**0.97**	**0.98**

C. Comparison of our Method with other State-of-the-Art Smoke Detection Methods

The performance evaluation of our proposed system with respect to other state-of-the-art smoke detection techniques is presented in this section. Proposed results are evaluated using test set of our own created dataset along with seven test videos as mentioned in Section III (A). Our presented technique is first compared with five different state-of-the-art smoke detection methods. The comparison with existing methods in literature is given in Table VII. The evaluation matrices used for comparison with recent methods are accuracy and false alarm rate. Furthermore, the processing time in frame per second (fps) is given with system specifications. From Table VII, we can observe that [41] is worst among all the other methods due to its high false alarm rate, low fps and accuracy. The accuracy of [28] is lower from all the other methods but its false alarm rate and fps are better than [42] and [41]. The method [42] performed average in terms of all the evaluation matrices. The best existing methods are [27] and [29] where [27] outperformed all the existing methods in terms of accuracy and [29] is best in terms of false alarm rate and highest fps against all other methods with lower accuracy. However, all the existing methods have pitfalls of lower accuracy, false alarm rate and processing of low fps. Our proposed system resolved the issue of accuracy and reduce the false alarm rate up to 2.30% but our system is still slower than [29] in terms of fps.

TABLE VII
Comparison with different smoke detection methods.

Method	Accuracy (%)	False positive (%)	fps	System specs
[41]	80.08	15.92	5	-
[42]	90.87	6.63	5.7	System equipped with Core i5 2.4 GHz CPU
[28]	47.71	5.0	25	This system has AMD Phenom (tm) II with X4 955 and 3.22 GHz processor capacity and is equipped with 8 GB of main memory
[27]	94.81	-	5.2	System processor is 2.4-GHz and is Core i5 model.
[29]	84.85	4.29	**61**	This system has Nvidia Geforce GTX with 980M GPU and DDR3 RAM with 16 GB capacity. The system model is Core i7-4720 HQ CPU
Proposed	**97.72**	**2.30**	31.33	Proposed system is equipped with NVidia GeForce GTX TITAN X (Pascal). Our system has Ubuntu operating system installed. Processor model is Core i5 of Intel company and builtin 64 GB of main memory

TABLE VIII
Comparison with two state-of-art techniques using seven test videos.

Video	[28]			[29]			Proposed		
	P	***R***	***F***	***P***	***R***	***F***	***P***	***R***	***F***
V1	0.97	0.99	0.98	1	0.79	0.88	0.98	0.96	0.97
V2	1	0.93	0.96	0.99	0.94	0.97	0.99	0.95	0.97
V3	0.2	0.32	0.3	0.93	0.96	0.95	0.93	0.92	0.92
V4	0.86	0.57	0.69	0.99	0.77	0.87	0.99	0.97	0.98
V5	0	0	-	0.97	0.8	0.88	0.99	0.98	0.99
V6	0.96	0.79	0.43	0.97	0.93	0.95	0.99	1	1
V7	0	0	-	0.85	1	0.92	1	0.91	0.95
average	0.57	0.51	0.48	0.95	0.88	0.91	**0.98**	**0.95**	**0.96**

The proposed system is also compared with two recent methods [29] and [28] using the seven test videos. For further exploration, the second scheme of evaluation is used as discussed in Section III (B). The dominance of the proposed system can be clearly perceived from Table VIII. The ***F*** of [28] shown in the Table is 0.48 and the best score acquired is 0.98 for V1. The minimum value of ***P*** for this method is observed for V7 and the maximum ***P*** value achieved is 1.0 for V2. The average ***F*** value for this method is not convincing to be implemented in smart cities. Similarly, the average ***F*** value calculated of [29] is 0.91 and maximum value of 0.97 is achieved by this method is for V2. The ***P*** value for above-mentioned method lies between 0 and 1 for V7 and V2, respectively. There is 0.43 unit improvement by [29] in accuracy as compared to the previous discussed method. The proposed system in contrast achieves much higher average ***F*** score as compared to both these applied techniques. Our system achieved 0.98, 0.95 and 0.96 ***P***, ***R*** and ***F*** values, respectively. The result comparison from this table assures the effectiveness of the proposed system over the existing state-of-the-art techniques.

D. Running time performance of the system

The experiments of our system are evaluated using a computer with a GPU of NVidia TITAN X (Pascal) having 12 GB onboard memory with a deep learning framework caffe [43] running and Intel Core i5 CPU with Ubuntu OS and 64 GB RAM. It can process 31.33 fps based on this setting and processing of 30 frames are enough to detect smoke in real-time. In addition, the normal camera can capture 25 to 30 fps, therefore our system is faster enough to detect smoke in real-time. The time comparison of the proposed system with state-of-the-art methods is given in Table IX.

Table IX
Average processing time of single frame in millisecond (MS) of the proposed system and state-of-the-art smoke detection methods using seven test videos.

Video	[28]	[29]	Proposed
V1	73.21	16.30	30.29
V2	62.25	17.45	31.5
V3	63.33	21.31	41.28
V4	67.76	27.60	57.11
V5	69.44	18.09	39.76
V6	67.62	17.30	37.77
V7	66.56	22.02	48.08
Average	67.16	**20.01**	40.82

It can be observed from results that the worst method is [28] having a high processing time of average 67.16 millisecond (MS) per frame. The authors in [29] achieved the best performance in terms of processing time and average 20 MS per frame. However, its limited accuracy and false alarm rate (2 times greater than ours) restricts its usefulness. Our proposed system attains the normal reasonable processing time of 40.82 MS per frame with highest accuracy and minimum false alarm rate, which means that our system can process up to 31 fps that is significantly enough for real-time detection of smoke in smart cites in IoT certain and uncertain environment.

TABLE X
Performance comparison of our proposed architecture with other state-of-the-art architectures.

Method Name	Number of layers	Net power consumption (W) / batch size	Memory utilization (MB) / batch size
AlexNet	**8**	13.4	**800**
GoogleNet	22	13.5	2100
VGG-19	19	13.8	1900
Proposed	16	**13**	1850

Table X compares our proposed method with other architectures in terms of number of layers, Net power consumption per batch size and memory utilization per batch size (comparison is based on 16 images per batch). From the table, it can be seen that AlexNet has minimum number of layers and memory utilization per batch also. The limitation of this method is its higher net power consumption and lower accuracy of 95.87%. GoogleNet has 22 layers with memory utilization of 2100 MB which is highest from all the other models and an average power consumption of 13.5 W. VGG-19 has less number of layers and memory utilization than GoogleNet but consumes comparatively more energy. In contrast to these architectures, our CNN model consists of an average number of layers with lower power consumption and minimum memory utilization except AlexNet. Seeing the overall performance evaluation metrics, number of layers, net power consumption per batch and memory utilization per batch, we claim that our proposed system is the better aspirant compared to the other state-of-the-art architectures and techniques for detection of smoke in smart cites normal environment as well as in foggy environment IoT environments.

IV. Conclusion and Future Work

CNNs gained great success recently by addressing problems in various fields and researchers applied it for detection of different abnormal events such as smoke, fire, disasters and other calamities. Fire is one of the most precarious event and it is very essential for disaster management to detect it in its initial stages in smart cities. Detection of smoke before it harms human lives and properties is very essential in its early stages. Unlike the fire flames, smoke can be identified from far away, as it moves in upward direction. Several smoke detection methods are presented by different researches till date. These methods are limited only to normal/certain environment, while in foggy or uncertain environment. Moreover, these methods are computationally expensive and difficult for them to process surveillance video streams in real-time. With these motivations, an energy-efficient CNN based smoke detection in normal and foggy or uncertain IoT environment method is proposed in this work. Our proposed method examines different state-of-the-art CNN models for the detection of smoke in both normal and foggy environments and propose an energy-efficient light-weight CNN architecture that is computationally inexpensive for real-time processing of live surveillance video streams. Secondly, we created our own benchmark dataset for the detection of smoke in uncertain environment. Furthermore, our proposed method dominates other state-of-the-art techniques in terms of false alarm rate and accuracy of the system. Finally, our system is capable to detect smoke in uncertain environment. Thus, it can be concluded that our proposed system is more suitable aspirant for the disaster management system to deploy it for the detection of smoke and fire in its early stages in smart cites normal and foggy or uncertain IoT environment.

This work is mainly focused on early smoke detection in normal and foggy or uncertain IoT environments with a real time processing. As future work, our strategy is to detect and localize the smoke for the extraction of detail information such as area of smoke, growth rate and distance from the camera etc. Moreover, a light-weight trained model can be deployed on embedded systems and board like FPGAs to increase the frame rate and improve decision making for integration with other IoT applications regarding prioritization [44], localization [45, 46], energy efficiency [47], greenery [48], security [49] and healthcare [50].

References

[1] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779-788.

[2] L. Hu and Q. Ni, "IoT-driven automated object detection algorithm for urban surveillance systems in Smart Cities," *IEEE Internet of Things Journal,* vol. 5, pp. 747-754, 2018.

[3] L. Bertinetto, J. Valmadre, J. F. Henriques, A. Vedaldi, and P. H. Torr, "Fully-convolutional siamese networks for object tracking," in *European conference on computer vision*, 2016, pp. 850-865.

[4] S. Ramnath, A. Javali, B. Narang, P. Mishra, and S. K. Routray, "IoT based localization and tracking," in *IoT and Application (ICIOT), 2017 International Conference on*, 2017, pp. 1-4.

[5] A. Ullah, J. Ahmad, K. Muhammad, M. Sajjad, and S. W. Baik, "Action recognition in video sequences using deep Bi-directional LSTM with CNN features," *IEEE Access,* vol. 6, pp. 1155-1166, 2018.

[6] S. Liu, "Class-constrained Transfer LDA for Cross-view Action Recognition in Internet of Things," *IEEE Internet of Things Journal,* 2017.

[7] A. Ullah, K. Muhammad, J. Del Ser, S. Wook Baik, and V. H. C. d. Albuquerque, "Activity Recognition using Temporal Optical Flow Convolutional Features and Multi-Layer LSTM," *IEEE Transactions on Industrial Electronics,* 2018.

[8] M. Mohammadi, A. Al-Fuqaha, M. Guizani, and J.-S. Oh, "Semisupervised deep reinforcement learning in support of IoT and smart city services," *IEEE Internet of Things Journal,* vol. 5, pp. 624-635, 2018.

[9] K. Muhammad, R. Hamza, J. Ahmad, J. Lloret, H. Wang, and S. W. Baik, "Secure Surveillance Framework for IoT Systems Using Probabilistic Image Encryption," *IEEE Transactions on Industrial Informatics,* vol. 14, pp. 3679-3689, 2018.

[10] K. Muhammad, J. Ahmad, and S. W. Baik, "Early fire detection using convolutional neural networks during surveillance for effective disaster management," *Neurocomputing,* vol. 288, pp. 30-42, 2018.

[11] K. Muhammad, J. Ahmad, Z. Lv, P. Bellavista, P. Yang, and S. W. Baik, "Efficient Deep CNN-Based Fire Detection and Localization in Video Surveillance Applications," *IEEE Transactions on Systems, Man, and Cybernetics: Systems,* pp. 1-16, 2018.

[12] K. Muhammad, J. Ahmad, I. Mehmood, S. Rho, and S. W. Baik, "Convolutional Neural Networks Based Fire Detection in Surveillance Videos," *IEEE Access,* vol. 6, pp. 18174-18183, 2018.

[13] T. Song, R. Li, B. Mei, J. Yu, X. Xing, and X. Cheng, "A Privacy Preserving Communication Protocol for IoT Applications in Smart Homes," *IEEE Internet of Things Journal,* vol. 4, pp. 1844-1852, 2017.

[14] C. Brunner, T. Peynot, T. Vidal‐Calleja, and J. Underwood, "Selective combination of visual and thermal imaging for resilient localization in adverse conditions: Day and night, smoke and fire," *Journal of Field Robotics,* vol. 30, pp. 641-666, 2013.

[15] Y. Chunyu, F. Jun, W. Jinjun, and Z. Yongming, "Video fire smoke detection using motion and color features," *Fire technology,* vol. 46, pp. 651-663, 2010.

[16] S. Calderara, P. Piccinini, and R. Cucchiara, "Vision based smoke detection system using image energy and color information," *Machine Vision and Applications,* vol. 22, pp. 705-719, 2011.

[17] T. K. T. Nguyen and J.-M. Kim, "Multistage optical smoke detection approach for smoke alarm systems," *Optical Engineering,* vol. 52, p. 057001, 2013.

[18] B. U. Toreyin, Y. Dedeoglu, and A. E. Cetin, "Contour based smoke detection in video using wavelets," in *European signal processing conference*, 2006.

[19] F. Yuan, "A double mapping framework for extraction of shape-invariant features based on multi-scale partitions with AdaBoost for video smoke detection," *Pattern Recognition,* vol. 45, pp. 4326-4336, 2012.

[20] C.-C. Ho, "Machine vision-based real-time early flame and smoke detection," *Measurement Science and Technology,* vol. 20, p. 045502, 2009.

[21] F. Yuan, "A fast accumulative motion orientation model based on integral image for video smoke detection," *Pattern Recognition Letters,* vol. 29, pp. 925-932, 2008.

[22] T. X. Tung and J.-M. Kim, "An effective four-stage smoke-detection algorithm using video images for early fire-alarm systems," *Fire Safety Journal,* vol. 46, pp. 276-282, 2011.

[23] C.-Y. Lee, C.-T. Lin, C.-T. Hong, and M.-T. Su, "Smoke detection using spatial and temporal analyses," *International Journal of Innovative Computing, Information and Control,* vol. 8, pp. 4749-4770, 2012.

[24] Y. Chunyu, Z. Yongming, F. Jun, and W. Jinjun, "Texture analysis of smoke for real-time fire detection," in *Computer Science and Engineering, 2009. WCSE'09. Second International Workshop on*, 2009, pp. 511-515.

[25] H. Maruta, A. Nakamura, and F. Kurokawa, "A new approach for smoke detection with texture analysis and support vector machine," in *Industrial Electronics (ISIE), 2010 IEEE International Symposium on*, 2010, pp. 1550-1555.

[26] F. Yuan, "Video-based smoke detection with histogram sequence of LBP and LBPV pyramids," *Fire safety journal*, vol. 46, pp. 132-139, 2011.

[27] K. Dimitropoulos, P. Barmpoutis, and N. Grammalidis, "Higher Order Linear Dynamical Systems for Smoke Detection in Video Surveillance Applications," *IEEE Trans. Circuits Syst. Video Techn.,* vol. 27, pp. 1143-1154, 2017.

[28] F. Yuan, Z. Fang, S. Wu, Y. Yang, and Y. Fang, "Real-time image smoke detection using staircase searching-based dual threshold AdaBoost and dynamic analysis," *IET Image Processing,* vol. 9, pp. 849-856, 2015.

[29] A. Filonenko, D. C. Hernández, and K.-H. Jo, "Fast Smoke Detection for Video Surveillance Using CUDA," *IEEE Transactions on Industrial Informatics*, vol. 14, pp. 725-733, 2018.

[30] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556,* 2014.

[31] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097-1105.

[32] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, *et al.*, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1-9.

[33] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "Imagenet: A large-scale hierarchical image database," in *Computer Vision and Pattern Recognition, 2009. CVPR 2009. IEEE Conference on*, 2009, pp. 248-255.

[34] A. Canziani, A. Paszke, and E. Culurciello, "An analysis of deep neural network models for practical applications," *arXiv preprint arXiv:1605.07678,* 2016.

[35] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, *et al.*, "Imagenet large scale visual recognition challenge," *International Journal of Computer Vision,* vol. 115, pp. 211-252, 2015.

[36] P. Foggia, A. Saggese, and M. Vento, "Real-time fire detection for video-surveillance applications using a combination of experts based on color, shape, and motion," *IEEE TRANSACTIONS on circuits and systems for video technology,* vol. 25, pp. 1545-1556, 2015.

[37] K. Dimitropoulos, P. Barmpoutis, and N. Grammalidis, "Spatio-temporal flame modeling and dynamic texture analysis for automatic video-based fire detection," *IEEE transactions on circuits and systems for video technology,* vol. 25, pp. 339-351, 2015.

[38] R. D. Labati, A. Genovese, V. Piuri, and F. Scotti, "Wildfire Smoke Detection Using Computational Intelligence Techniques Enhanced With Synthetic Smoke Plume Generation," *IEEE Transactions on Systems, Man, and Cybernetics: Systems,* vol. 43, pp. 1003-1012, 2013.

[39] "http://staff.ustc.edu.cn/~yfn/vsd.html (cited on 12 August 2018) ".

[40] "http://signal.ee.bilkent.edu.tr/VisiFire/Demo/SmokeClips/ (cited on 12 August 2018)."

[41] K. Avgerinakis, A. Briassouli, and I. Kompatsiaris, "Smoke detection using temporal HOGHOF descriptors and energy colour statistics from video," in *International workshop on multi-sensor systems and networks for fire detection and management*, 2012.

[42] P. Barmpoutis, K. Dimitropoulos, and N. Grammalidis, "Smoke detection using spatio-temporal analysis, motion modeling and dynamic texture recognition," in *Signal Processing Conference (EUSIPCO), 2014 Proceedings of the 22nd European*, 2014, pp. 1078-1082.

[43] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, *et al.*, "Caffe: Convolutional architecture for fast feature

embedding," in *Proceedings of the 22nd ACM international conference on Multimedia*, 2014, pp. 675-678.

[44] K. Muhammad, T. Hussain, and S. W. Baik, "Efficient CNN based summarization of surveillance videos for resource-constrained devices," *Pattern Recognition Letters,* 2018.

[45] G. Han, C. Zhang, L. Shu, and J. J. P. C. Rodrigues, "Impacts of Deployment Strategies on Localization Performance in Underwater Acoustic Sensor Networks," *IEEE Transactions on Industrial Electronics,* vol. 62, pp. 1725-1733, 2015.

[46] G. Han, J. Jiang, C. Zhang, T. Q. Duong, M. Guizani, and G. K. Karagiannidis, "A Survey on Mobile Anchor Node Assisted Localization in Wireless Sensor Networks," *IEEE Communications Surveys and Tutorials,* vol. 18, pp. 2220-2243, 2016.

[47] K. Wang, Y. Wang, Y. Sun, S. Guo, and J. Wu, "Green industrial internet of things architecture: An energy-efficient perspective," *IEEE Communications Magazine,* vol. 54, pp. 48-54, 2016.

[48] C. Zhu, V. C. Leung, L. Shu, and E. C.-H. Ngai, "Green internet of things for smart world," *IEEE Access,* vol. 3, pp. 2151-2162, 2015.

[49] K. Muhammad, M. Sajjad, I. Mehmood, S. Rho, and S. W. Baik, "Image steganography using uncorrelated color space and its application for security of visual contents in online social networks," *Future Generation Computer Systems,* vol. 86, pp. 951-960, 2018.

[50] J. J. Rodrigues, D. B. D. R. Segundo, H. A. Junqueira, M. H. Sabino, R. M. Prince, J. Al-Muhtadi, *et al.*, "Enabling Technologies for the Internet of Health Things," *Ieee Access,* vol. 6, pp. 13129-13141, 2018.

Salman Khan (S'19) received his Bachelor's degree in Computer Science from Islamia College Peshawar, Pakistan in 2017 with research in medical image analysis. Currently, he is pursuing MS degree in software convergence from Sejong University, Seoul, Republic of Korea. He is working as a research assistant at Intelligent Media Laboratory (IM Lab). His research interests include Medical Image Analysis, Computer Vision, Deep Learning, Fire/Smoke Scene Analysis, and Video Surveillance.

Khan Muhammad (S'16–M'18) received the Ph.D degree in Digital Contents from Sejong University, South Korea. He is currently working as a researcher at Intelligent Media Laboratory (IM Lab), Sejong University, Seoul. His research interests include medical image analysis (brain MRI, diagnostic hysteroscopy and wireless capsule endoscopy), information security (steganography, encryption, watermarking and image hashing), video summarization, computer vision, fire/smoke scene analysis, and video surveillance. He has published over 50 papers in peer-reviewed international journals and conferences in these research areas with target venues as IEEE COMMAG, TII, TIE, TSMC-Systems, IoTJ, Access, TSC, Elsevier INS, Neurocomputing, PRL, FGCS, COMCOM, COMIND, JPDC, PMC, BSPC, CAEE, Springer NCAA, MTAP, JOMS, and RTIP, etc. He is also serving as a professional reviewer for over 40 well-reputed journals and conferences.

Shahid Mumtaz (M'13–SM''16) (smumtaz@av.it.pt) received the M.Sc. degree from the Blekinge Institute of Technology, Sweden, and the Ph.D. degree from the University of Aveiro, Portugal. He is currently a senior research engineer with the Instituto de Telecornunicações, Aveiro, where he is involved in EU funded projects. His research interests include MIMO techniques, multihop relaying communication, cooperative techniques, cognitive radios, and game theory.

Sung Wook Baik (M'16) received the B.S degree in computer science from Seoul National University, Seoul, Korea, in 1987, the M.S. degree in computer science from Northern Illinois University, Dekalb, in 1992, and the Ph.D. degree in information technology engineering from George Mason University, Fairfax, VA, in 1999. He worked at Datamat Systems Research Inc. as a senior scientist of the Intelligent Systems Group from 1997 to 2002. In 2002, he joined the faculty of the College of Electronics and Information Engineering, Sejong University, Seoul, Korea, where he is currently a Full Professor and the Chief of Sejong Industry-Academy Cooperation Foundation. He is also the head of Intelligent Media Laboratory (IM Lab) at Sejong University. His research interests include computer vision, multimedia, pattern recognition, machine learning, data mining, virtual reality, and computer games. He is a member of the IEEE.

Victor Hugo C. de Albuquerque (M'17) received the graduation degree in mechatronics technology from the Federal Center of Technological Education of Ceará, Fortaleza, Brazil, in 2006, the M.Sc. degree in tele-informatics engineering from the Federal University of Ceará, Fortaleza, in 2007, and the Ph.D. degree in mechanical engineering with emphasis on materials from the Federal University of Paraíba, João Pessoa, Brazil, in 2010. He is currently an Assistant VI Professor with the Graduate Program in Applied Informatics at the University of Fortaleza, Fortaleza. He has experience in computer systems, mainly in the research fields of applied computing, intelligent systems, visualization and interaction, with specific interest in pattern recognition, artificial intelligence, image processing and analysis, Internet of Things, Internet of Health Things, as well as automation with respect to biological signal/image processing, image segmentation, biomedical circuits, and human/brain–machine interaction, including augmented and virtual reality simulation modeling for animals and humans.