

Channel-wise attention model-based fire and rating level detection in video

ISSN 2468-2322

Received on 22nd March 2019

Revised on 15th April 2019

Accepted on 21st April 2019

doi: 10.1049/trit.2019.0022

www.ietdl.org

Yirui Wu^{1,2}, Yuechao He¹, Palaiahnakote Shivakumara² ✉, Ziming Li¹, Hongxin Guo¹, Tong Lu³

¹College of Computer and Information, Hohai University, 210098, Nanjing, People's Republic of China

²Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur 50603, Malaysia

³National Key Lab for Novel Software Technology, Nanjing University, Nanjing 210093, People's Republic of China

✉ E-mail: hudempsk@yahoo.com

Abstract: Due to natural disaster and global warning, one can expect unexpected fire, which causes panic among people and extent to death. To reduce the impact of fire, the authors propose a new method for predicting and rating fire in video through deep-learning models in this work such that rescue team can save lives of people. The proposed method explores a hybrid deep convolutional neural network, which involves motion detection and maximally stable extremal region for detecting and rating fire in video. Further, the authors propose to use a channel-wise attention mechanism of the deep neural network for detecting rating of fire level. Experimental results on a large dataset show the proposed method outperforms the existing methods for detecting and rating fire in video.

1 Introduction

As one of the most common and largely distributed disasters, fire often happens and brings life and economic damages to human society. Therefore, how to detect fire accurately has been a hot topic and many researchers have proposed methods to perform fire detection. In general, most of the traditional methods work based on smoke or temperature sensors [1]. However, such methods suffer from the following inherent limitations: (1) they will be easily affected by complex environmental factors to provide false alarms; (2) applicable scenario for such methods is highly restricted into indoor scenes, which cannot be applied in the wild; and (3) these methods require additional devices to perform this task.

With the rapid development of computer vision and machine learning technologies, researchers perform fire detection by analysing images directly captured by CCTV cameras [2, 3]. With this idea, people save cost and time since this way does not require expensive sensors. In addition, video and image-based methods are fast and accurate compared to sensor-based methods irrespective of situations. Therefore, the methods can be categorised into two classes, that is, those use flame detection as feature for fire detection [4, 5], and those explore texture, colour and spatial information of fire in video for fire detection [6–8]. However, the formers may not be robust when work well for different situations, while the latter are generally more accurate and may work well for different situations. Therefore, the proposed method falls on the second category.

Here, we explore different combinations of deep convolutional neural networks for accurate fire detection as well as rating detection in video. The motivation for choosing the above method is that deep learning models have strong the discriminative ability for complex situations. It is noted that the considered work here is complex because one cannot predict background and the nature of fire in video. The main contribution of the proposed work is that exploring the channel-wise attention deep-learning model for fire and rating detection in video.

The rest of the paper is organised as follows. Section 2 reviews the related work. Details of fire detection and fire rating are discussed in Section 3. Section 4 presents the experimental results and discussions. Finally, Section 5 concludes the paper.

2 Related work

During burning process, fire produces the unique red and yellow circular structure, which results in a special colour distribution. This observation leads to develop methods for fire detection in literature. For example, Yan *et al.* [4] extract features based on RGB color space and Ohta space, which further help to generate four suspected regions of fire. However, the process involves human intervention to fix thresholds for achieving better results. Celik *et al.* [2] extract features to describe colour characteristics of fire by fusing intensity and brightness information in YCbCr colour space. However, the method is not robust and perfect enough for complex situations. Though colour-based methods are simple, they are not effective for complex situations.

To overcome the above limitation, there are methods that explore motion features of fire in video for fire detection. For example, Verstockt *et al.* [8] propose an image differential method to firstly acquire amplitude information of fire, and then calculate the mean amplitude variance that reflects motion dynamic information of fire and helps to detect accurate fire in video. This method works well for images having a high resolution. Kolesov *et al.* [9] use an optical flow method to construct features for the description of physical properties of fire in two different stages, i.e. smoke and fire process. The method works well when an image contains both fire and smoke. Wang *et al.* [10] explore temporal features in frequency domain for fire detection. The method is computationally expensive due to high operation for converting an input image into a frequency matrix. Borges *et al.* [11] propose to analyse variations of low-level features to detect fire candidate regions in video frames. Then, it combines light frequency and motion features with the help of a Bayesian classifier for fire detection.

In summary, there are methods for fire detection in video as well as still images. It is noted from the review of the above methods is that most of these methods target specific situations for achieving better results but not for different situations. However, as mentioned, the nature of fire and background complexity are unpredictable in reality. Therefore, there is an urgent need for developing a new method that can work well for different situations. To achieve our goal, we explore hybrid deep-learning models, which involve an attention concept in this work.

There are many deep learning-based methods are proposed literature because of advantage of attention models for solving different issues. The method can be classified broadly into two classes, namely, hard attention and soft attention-based methods. Hard attention mechanically chooses parts of the input data as focuses. For example, Mnit *et al.* [12] propose a hard attention model for image recognition, which adaptively selects a sequence of regions and processes the selected regions. He *et al.* [13] present scene text detection by proposing a novel text-attentional convolutional neural network (Text-CNN), which particularly focuses on extracting text-related regions and features from image components. In the same way, soft attention-based methods consider the entire input into account by weighting each part or step dynamically. For example, Yeung *et al.* [14] propose the method based on the fusion of neighbouring frames within a sliding window in RGB video. Liu *et al.* [15] propose a global context-aware attention LSTM for RGB-D action recognition, which recurrently optimises the global contextual information and further utilises it as an informative function to assist accurate action recognition. Song *et al.* [16] achieve the goal of action recognition from skeleton data by selectively focusing on discriminative joints of skeleton within each frame of the inputs and assigning different levels of attention to the outputs of different frames. Chen *et al.* [17] introduce a novel convolutional neural network dubbed SCA-CNN that incorporates Spatial and Channel-wise Attentions in CNN for the task of image captioning. Hu *et al.* [18] propose a novel and relatively independent architectural unit, which is termed as the ‘Squeeze-and-Excitation’(SE) block and adaptively recalibrates channel-wise feature responses by explicitly modelling interdependencies between channels.

It is observed from the review of attention-based deep learning models for solving different issues that attention-based deep learning models have the ability to adapt for complex situation without much changes. Inspired by the special property of attention deep learning models, we explore the same in a new way for fire and rating detection in video in this work. The main contribution of the proposed work is follows. It is noted from the literature review that the channel-wise attention deep learning model is not explored for the application of fire detection and fire rating level detection. Therefore, the way we explore the above concept for fire detection and fire rating level detection is new.

3 Proposed method

The block diagram of the proposed method is shown in Fig. 1, which consists of two stages, i.e. fire detection and fire rating detection. In the first stage, the proposed method detects frames that probably contain fire based on motion features and maximally stable extremal regions. This results in candidate fire regions. The candidate fire regions are then passed to a pre-trained VGG-16 CNN network [19] to extract features. Next, an SVM classifier is

designed for the features given by the VGG-16 model for fire region detection from candidate fire regions. In the second stage, the proposed method considers the output of the first stage as the input for fire level rating detection. To achieve this, the proposed explores a VGG-19 CNN structure involving a channel-wise attention scheme for fire-level rating detection. The fire rating levels considered in this work are non-fire, beginning, small, medium, and big. Since these rating levels are quite common according to fire rescue experts, we consider these levels in this work.

3.1 Fire detection

In case of fire video, it is true that motion feature plays a vital role to study the properties of fire and smoke because fire and smoke move with a certain speed. Motivated by this observation, the proposed method uses Inter frame Differential for detecting frames that contain fire. The algorithmic steps for detecting frames containing fire are as described below.

First, we divide the original video frames into small blocks with size of 32×32 in a non-overlapping way. The proposed method calculates the sum of intensity difference of corresponding blocks between frames as defined in (1).

$$D_t = \sum_{(x,y)} |M_t(x,y) - M_{t-1}(x,y)| \quad (1)$$

where x and y are the row and column index for blocks, and function $M_t(x,y)$ represents the mean intensity value of all the pixels contained by the block located at (x,y) in frame M_t . Then the proposed method detects the frame that contains fire as defined in (2).

$$I_{s,t} = \begin{cases} Motion, & D_t \geq T_B \\ Static, & D_t < T_B \end{cases} \quad (2)$$

where T_B is the threshold to decide the state of the frame. The value of T_B is determined empirically. This step outputs a number of frames that contain fire.

To detect exact fire regions in the frames detected by the previous step, the proposed method explores an improved maximally stable extremal regions (MSER) [20] algorithm. The reason to choose MSER algorithm is that it performs grouping for pixels, which share the same color values and outputs candidate regions. In case of smoke and fire in the proposed work, the pixels share almost the same values. Therefore, an improved MSER algorithm [20] is used for detecting fire and smoke regions as candidate fire regions. The steps for detecting candidate fire regions using MSER are a follows.

$$q(i) = \frac{|Q_{i+\Delta} - Q_{i-\Delta}|}{Q_i} \quad (3)$$

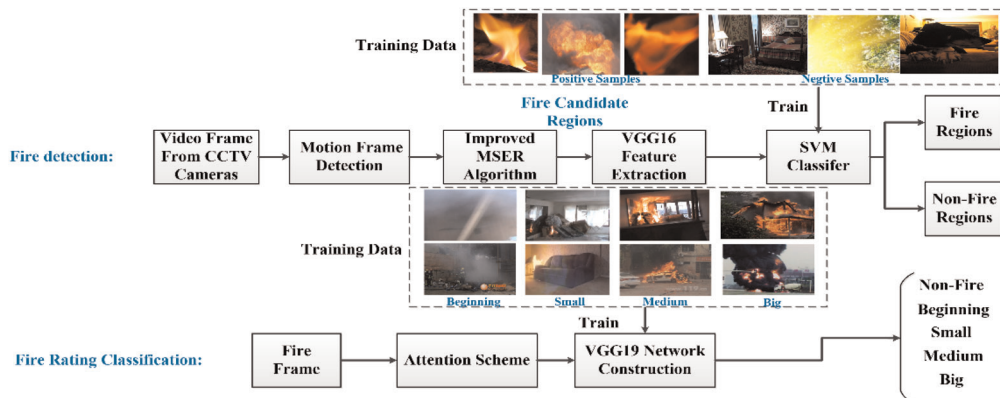


Fig. 1 Block diagram of the proposed method

where Q_i represents ER whose construction threshold is defined as i , Δ represents a small variation of colour intensity, and $q(i)$ is the rate of variation of region Q_i with threshold value i . When $q(i)$ is smaller than a certain threshold, Q_i can be considered as an MSER, namely, a fire candidate region. The effect of candidate region detection by MSER is shown in Fig. 2a, where we can see almost all the fire regions are detected. It is also noted from Fig. 2a that the step detects non-fire regions as candidate fire ones.

To reduce the number of false candidate fire regions, we propose a deep CNN model, i.e. VGG16, to extract features. The features are passed to an SVM classifier with Gaussian kernel for eliminating false candidate fire regions as shown in Fig. 2b, where false candidate fire regions are removed. This results in fire detection.

3.2 Fire rating detection

This section presents the step for fire level rating from fire detection results given by the previous step. To achieve this, we explore a channel-wise attention model. This model works based on lightweight gating mechanism, which defines the relationship between feature maps of multiple layers as the structure of the model in Fig. 3, where it can be seen the whole general architecture of the channel-wise attention model.

Inspired by Mnit *et al.*'s work [12] which considers the attention problem as a sequential decision process for interactions in visual environment, we explore the same by deriving weights assigned to each feature channel. Specifically, let input feature map be $U = [u_i; z_i = 1, \dots, c]$, where u_i represents the i th channel of U and c is the total number of feature channels. Since each of the learned convolutional filters operate with a local receptive field to compute features, which do not exploit contextual information between the region and the neighbour information [18], we propose to utilise

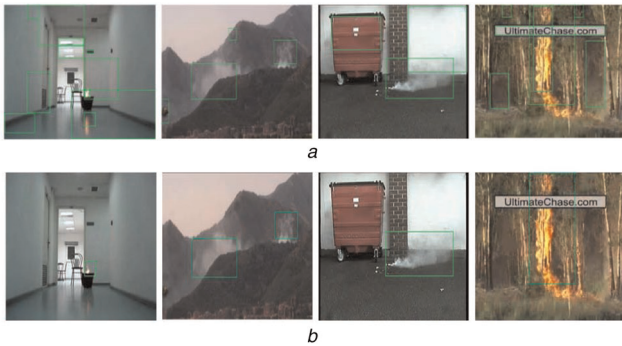


Fig. 2 Candidate fire region detection based on MSER and Deep learning model

(a) Candidate fire regions detection. It is noted that there are many false candidate fire regions, (b) False candidate fire region removal elimination

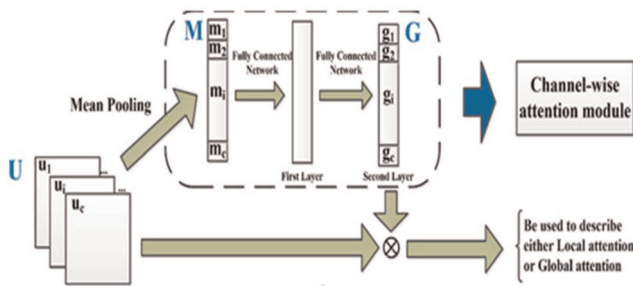


Fig. 3 Structure of the proposed channel-wise attention module for classification

the mean pooling operator to collect contextual information into a channel descriptor, which is defined as in (4)

$$m_i = \frac{1}{W \times H} \sum_{j=1}^W \sum_{k=1}^H u_i(j, k) \quad (4)$$

where W and H are width and height of u_i , and M is defined as the constructed channel descriptor formed by quantity of vectors represented by m_i .

To utilise the advantage of the information of M , the proposed method follows the idea of 'interaction level' to construct the weighting scheme G as shown in Fig. 4, where the proposed channel-wise attention module is designed, which consists of two fully connected layers and two corresponding nonlinear activation functions. It is defined as in (5).

$$g_i = \text{Nor}(\text{sig}(W_2 \eta(W_1 m_i + b_1) + b_2)) \quad (5)$$

where function $\text{sig}()$, $\text{Nor}()$ and $\eta()$ refer to sigmoid, normalisation and ReLU functions, respectively, W_1 and W_2 are the learnable parameter matrices, and b_1 and b_2 are bias vectors. There are two reasons for adapting such a structure to construct the modules. First, the designed structure must be capable of learning a highly nonlinear interaction between channels and second, it allows multiple channels to be emphasised opposed to one-hot activation. The resulting weight vector G is thus defined as $\{G = g_i; i = 1, \dots, c\}$ and leads to the attention on informative channel features as shown in Fig. 4, where it shows how the channel-wise attention module works with

$$\tilde{u}_i = u_i \odot g_i \quad (6)$$

where \odot represents element-wise multiplication.

To implement the above idea, we propose to adapt a popular VGG19 network to modify such that it works well for fire level rating. As mentioned earlier, we use pre-trained VGG16 and VGG19 for the purpose of fire level rating detection using Image net dataset. This gives proposed hierarchically channel-wise attention scheme as shown in Fig. 4, where the convolutional layer is represented as:

$$U = \sigma(X * W + b) \quad (7)$$

where σ is the ReLU activation function, X represents the input image, $*$ represents the convolution operator, W is a convolution kernel, b is the offset, and U is the output feature map.

As shown in Fig. 4, the channel-wise attention weights are built as a function of the l th CNN channel feature $\{u_{k,l}; l = 1, \dots, c\}$ output by the k th vgg layer and then work with it as

$$\tilde{u}_{k,l} = u_{k,l} \times \emptyset(u_{k,l}) \quad (8)$$

where function $\emptyset()$ represents the channel-wise attention module described in the last layer, and \cdot is element-wise multiplication. We build such channel-wise attention modules at the end of several VGG19 layers, which utilises the channel-wise property

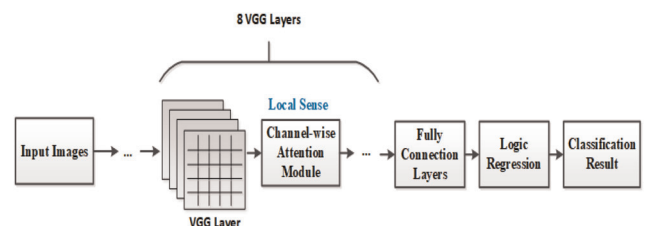


Fig. 4 Proposed CNN model with Channel-wise attention module, which is used in the last 8 convolution layers of VGG19

of CNN features to automatically focus on discriminative and representative features for the classification in a local sense. The last full connection layer outputs a feature vector, which is then be classified by a softmax function into five classes. In this way, the proposed method uses a channel-wise attention model for fire level rating detection in this work.

4 Experimental results

For evaluating the proposed method for fire detection and fire level rating detection, we collect dataset from different resources, such as Internet from searching engines, such as Google, Bing, and Baidu. The dataset includes video 4273 for training and 1436 video for testing. Of training, 2049 for positive sample images and 2224 negative sample images. Positive sample images cover different kinds of environments, such as indoor and outdoor scenes, while negative sample images include common scenes without fire, such as residence, street, and mall. Since the dataset includes images of outdoor and indoor scenes, the considered dataset is complex and good for evaluating the proposed method.

For measuring the performance of the proposed fire detection step, we use standard measures, namely, false positive, false negative, and accuracy. For the measuring the performance of the fire level rating detection step, we use standard measures, namely, recall, precision, and F-measure. To show the effectiveness and usefulness of the proposed method, we implement the following method for comparative study. The method [21] uses co-variance matrix-based features for fire and flame detection in video. The method [22] proposes a convolution neural network for early fire detection during disaster management. The method [23] uses colour, shape, and motion features for fire detection in real time environment. The method [24] explores video analytics for fire detection. These methods are used for comparing with the proposed fire detections step. The reason to choose the above methods for comparative study is that the methods are developed for fire detection as the proposed method. In addition, each method has its own application and specific objective for fire detection. All these comparative methods can show that the developed method with specific objective may not perform well for complex situations.

Similarly, we implement the well-known deep learning tools, namely, AlexNet [25], VGG19 [19], ResNet [26], DenseNet [27], and the method [28] for comparing with the proposed fire level rating detection and a method from [28]. The reason to choose the above models is that those models are state-of-the-art for different applications in literature.

4.1 Evaluating fire detection

Qualitative results of the proposed method for fire detection are shown in Fig. 5, where it is noted that the proposed method works



Fig. 5 Sample results for fire frames detection

well for images of different backgrounds and different situations. Therefore, we can confirm that the proposed fire detection is robust and can be used for other applications, where video contains fire and flame. Quantitative results of the proposed and existing methods are reported in Table 1, where it is observed that the proposed method is the best at false positive, false negative and accuracy compared to the existing methods. The reason for poor results of the existing methods is that the methods are developed for addressing specific situations but not different and complex situations. Therefore, one can conclude the proposed method outperforms the existing methods for fire detection.

4.2 Evaluating fire level rating detection

For fire-level rating detection, we use the following details for training and testing of the proposed channel-wise attention model. We define defining batch size as 32 and adopt ADAM solver to optimise the training procedures. The proposed model is trained within 100 epoches and initialised with a learning rate of 0.0001. The momentum is set as 0.9. Most important is that we define rules based on expert advice for determining the number of rating levels as reported in Table 2.

Then, we rate the fire based on the burning area, where we show the detail of rating rules in Table 2. It is noted that such a rating rule and further rating process on the training or testing dataset are under the guidance of an expert. Sample images for each level are shown in Fig. 6, where we can visualise fire-level rating ranging from 1 to 4

Table 1 Performance of the proposed and existing methods for fire detection

Methods	False positives,%	False negatives,%	Accuracy,%
proposed method	3.02	2.26	97.36
Habibuglu <i>et al.</i> [21]	5.88	14.29	90.32
Muhammad <i>et al.</i> [22]	9.07	2.13	94.39
Foggia <i>et al.</i> [23]	11.67	0	93.55
Di Lascio <i>et al.</i> [24]	13.33	0	92.86

Table 2 Rules and conditions based on expert advice for fire-level rating detection, where we use area with m^2 to define different fire levels

Scenes grade rating	Beginning, m^2	Small, m^2	Medium, m^2	Big, m^2
indoor	smoke	<0.25	0.25-1	>1
outdoor	smoke	<5	5-20	>20

Classification Results



Fig. 6 Sample results for fire-level rating detected frames

Table 3 Performance of the proposed and existing deep learning tools for fire level rating detection

Method	Precision	Recall	F-measure
VGG19 [19]	0.72	0.73	0.72
AlexNet [25]	0.69	0.72	0.70
ResNet [26]	0.76	0.76	0.76
Ma <i>et al.</i> [28]	0.77	0.78	0.77
DenseNet [27]	0.78	0.78	0.78
proposed method	0.79	0.81	0.80

corresponding to small, beginning, medium and big fire. According to the rules defined for fire-level rating, we report results for fire level rating estimated by the proposed and other deep learning models in Table 3. Table 3 shows that the proposed method is the best at recall, precision and F-measure compared to the existing deep learning tools. The main reason for the poor results of the existing deep learning tools is that the tools miss attention-based context information, while the proposed method incorporates the attention-based context. In other words, a convolutional layer usually scans the input image and computes the corresponding 3D feature map, which essentially encodes spatial-visual responses given by a channel filter. As a result, the low-layer channel filters extract low-level visual cues such as edges and corners, while the high-layer extracts high-level semantic patterns such as parts of the objects or objects. However, it is true that all the channel-wise features may not extract high level semantics for all situations. Thus channel attention scheme plays a vital role in extracting the task-specified feature map for fire rating, which exploits cross-channel relationship to achieve the best results. Therefore, the proposed method is the best for fire level rating detection.

Note that for each image, the average processing time of the proposed method is 3.2 s with 3.4 GHz Intel i7 CPU (64 G RAM) and NVIDIA Titan 1080Ti GPU (12 G Memory) system configuration.

5 Conclusion and future work

Here, we have proposed a new method for fire detection and fire-level rating detection by exploring deep learning models. The proposed method uses frame difference based on motion features for detecting frames that contain fire from video. Fire regions in frames are detected as candidate fire regions by applying the MSER method, and then features are extracted from candidate fire regions from a deep-learning model. The extracted features are passed to an SVM classifier for eliminating false candidate fire regions. For frames containing fire regions, the proposed method explores a channel-wise attention-based deep learning model for fire-level rating detection. Experimental results on both fire detection and fire-level rating detection show that the proposed method outperforms the existing methods. The future work would extend the same for implementing an embedding system in a real-life scenario. In addition, we plan to combine the proposed two architectures as a single architecture to make the proposed system more efficient with better results.

6 Acknowledgment

This work was supported by National Key R&D Program of China under Grant no.2018YFC0407901, the Natural Science Foundation of China under Grant Grant no.61702160, Grant 61672273 and Grant no. 61832008, the Science Foundation of Jiangsu under Grant BK20170892, the Science Foundation for Distinguished Young Scholars of Jiangsu under Grant BK20160021, Scientific Foundation of State Grid Corporation of China (Research on Ice-wind Disaster Feature Recognition and Prediction by Few-shot Machine Learning in Transmission Lines), and the open Project of

the National Key Lab for Novel Software Technology in NJU under Grant K-FKT2017B05.

7 References

- [1] Khan, M.J.A., Imam, M.R., Uddin, J., *et al.*: 'Automated fire fighting system with smoke and temperature detection'. Proc. of 7th Int. Conf. on Electrical and Computer Engineering, Dhaka, Bangladesh, 2012, pp. 232–235
- [2] Celik, T., Ma, K.K.: 'Computer vision based fire detection in color images'. Proc. of 2008 IEEE Conf. on Soft Computing in Industrial Applications, Mumbai, India, 2008, pp. 258–263
- [3] Van Hamme, D., Veelaert, P., Philips, W., *et al.*: 'Fire detection in color images using Markov random fields'. Proc. of Int. Conf. on Advanced Concepts for Intelligent Vision Systems, Sydney, Australia, 2010, pp. 88–97
- [4] Yan, Y., Guo, Z., Wang, H.: 'Fire detection based on feature of flame color'. Proc. of 2009 Chinese Conf. on Pattern Recognition, Nanjing, China, 2009, pp. 1–5
- [5] Truong, T.X., Kim, J.M.: 'Fire flame detection in video sequences using multi-stage pattern recognition techniques', *Eng. Appl. Artif. Intell.*, 2012, **25**, (7), pp. 1365–1372
- [6] Frizzi, S., Kaabi, R., Bouchouicha, M., *et al.*: 'Convolutional neural network for video fire and smoke detection'. Proc. of 42nd Annual Conf. of the IEEE Industrial Electronics Society, Lviv, Ukraine, 2016, pp. 877–882
- [7] Xi, Z., Fang, X., Zhen, S., *et al.*: 'Video flame detection algorithm based on multi-feature fusion technique'. Proc. of 24th Chinese Control and Decision Conf. (CCDC), Taiyuan, China, 2012, pp. 4291–4294
- [8] Verstockt, S., Van Hoecke, S., Beji, T., *et al.*: 'A multi-modal video analysis approach for car park fire detection', *Fire Saf. J.*, 2013, **57**, pp. 44–57
- [9] Kolesov, I., Karasev, P., Tannenbaum, A., *et al.*: 'Fire and smoke detection in video with optimal mass transport based optical flow and neural networks'. Proc. of 2010 IEEE Int. Conf. on Image Processing, Hong Kong, China, 2010, pp. 761–764
- [10] Wang, L., Ye, M., Ding, J., *et al.*: 'Hybrid fire detection using hidden Markov model and luminance map', *Comput. Electr. Eng.*, 2011, **37**, (6), pp. 905–915
- [11] Borges, P.V.K., Izquierdo, E.: 'A probabilistic approach for vision-based fire detection in videos', *IEEE Trans. Circuits Syst. Video Technol.*, 2010, **20**, (5), pp. 721–731
- [12] Mnih, V., Heess, N., Graves, A., *et al.*: 'Recurrent models of visual attention'. Proc. of Conf. on Neural Information Processing Systems, Montreal, Canada, 2014, pp. 2204–2212
- [13] He, T., Huang, W., Qiao, Y., *et al.*: 'Text-attentional convolutional neural network for scene text detection', *IEEE Trans. Image Process.*, 2016, **25**, (6), pp. 2529–2541
- [14] Yeung, S., Russakovsky, O., Jin, N., *et al.*: 'Every moment counts: dense detailed labeling of actions in complex videos', *Int. J. Comput. Vis.*, 2018, **126**, (2–4), pp. 375–389
- [15] Liu, J., Wang, G., Hu, P., *et al.*: 'Global context-aware attention LSTM networks for 3D action recognition'. Proc. of IEEE Conf. on Vision and Pattern Recognition, Honolulu, HI, USA, 2017, pp. 3671–3680
- [16] Song, S., Lan, C., Xing, J., *et al.*: 'An end-to-end spatio-temporal attention model for human action recognition from skeleton data'. Proc. of AAAI Conf. on Artificial Intelligence, San Francisco, California, USA, 2017, pp. 4263–4270
- [17] Chen, L., Zhang, H., Xiao, J., *et al.*: 'SCA-CNN: spatial and channel-wise attention in convolutional networks for image captioning', arXiv preprint arXiv:1611.05594, 2016
- [18] Hu, J., Shen, L., Sun, G.: 'Squeeze-and-excitation networks', arXiv preprint arXiv:1709.01507, 2017
- [19] Simonyan, K., Zisserman, A.: 'Very deep convolutional networks for large-scale image recognition', arXiv preprint arXiv:1409.1556, 2014
- [20] Matas, J., Chum, O., Urban, M., *et al.*: 'Robust wide-baseline stereo from maximally stable extremal regions', *Image Vis. Comput.*, 2004, **22**, (10), pp. 761–767
- [21] Habiboğlu, Y.H., Günay, O., Çetin, A.E.: 'Covariance matrix-based fire and flame detection method in video', *Mach. Vis. Appl.*, 2012, **23**, (6), pp. 1103–1113
- [22] Muhammad, K., Ahmad, J., Baik, S.W.: 'Early fire detection using convolutional neural networks during surveillance for effective disaster management', *Neurocomputing*, 2018, **288**, pp. 30–42
- [23] Foggia, P., Saggese, A., Vento, M.: 'Real-time fire detection for video-surveillance applications using a combination of experts based on color, shape, and motion', *IEEE Trans. Circuits Syst. Video Technol.*, 2015, **25**, (9), pp. 1545–1556
- [24] Di Lascio, R., Greco, A., Saggese, A., *et al.*: 'Improving fire detection reliability by a combination of video analytics'. Proc. of Int. Conf. Image Analysis and Recognition, Vilamoura, Portugal, 2014, pp. 477–484
- [25] Krizhevsky, A., Sutskever, I., Hinton, G.E.: 'Imagenet classification with deep convolutional neural networks'. Proc. of Advances in Neural Information Processing Systems, Lake Tahoe, Nevada, 2012, pp. 1097–1105
- [26] He, K., Zhang, X., Ren, S., *et al.*: 'Deep residual learning for image recognition'. Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 2016, pp. 770–778
- [27] Huang, G., Liu, Z., Van Der Maaten, L., *et al.*: 'Densely connected convolutional networks'. Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 2017, pp. 4700–4708
- [28] Ma, J., Li, K., Han, Y., *et al.*: 'Image-based air pollution estimation using hybrid convolutional neural network'. Proc. of 24th Int. Conf. on Pattern Recognition (ICPR), Beijing, China, 2018, pp. 471–476