Deep Convolutional Generative Adversarial Networks for Flame Detection in Video

Süleyman Aslan[1], Uğur Güdükbay[1], B. Uğur Töreyin[2(✉)], and A. Enis Çetin[3]

[1] Department of Computer Engineering, Bilkent University, Ankara, Turkey
suleyman.aslan@bilkent.edu.tr, gudukbay@cs.bilkent.edu.tr
[2] Informatics Institute, Istanbul Technical University, Istanbul, Turkey
toreyin@itu.edu.tr
[3] Department of Electrical and Computer Engineering, University of Illinois at Chicago, Chicago, IL, USA
aecyy@uic.edu
http://cs.bilkent.edu.tr/∼gudukbay, https://spacing.itu.edu.tr/, https://ece.uic.edu/profiles/ahmet-enis-cetin-phd/

Abstract. Real-time flame detection is crucial in video-based surveillance systems. We propose a vision-based method to detect flames using Deep Convolutional Generative Adversarial Neural Networks (DCGANs). Many existing supervised learning approaches using convolutional neural networks do not take temporal information into account and require a substantial amount of labeled data. To have a robust representation of sequences with and without flame, we propose a two-stage training of a DCGAN exploiting spatio-temporal flame evolution. Our training framework includes the regular training of a DCGAN with real spatio-temporal images, namely, temporal slice images, and noise vectors, and training the discriminator separately using the temporal flame images without the generator. Experimental results show that the proposed method effectively detects flame in video with negligible false-positive rates in real-time.

Keywords: Fire detection · Flame detection · Deep Convolutional Generative Adversarial Neural Network

1 Introduction

Fires pose a great danger in open and large spaces. Flames may spread fast and cause substantial damages to properties and human life. Hence, immediate and accurate flame detection plays an instrumental role in fighting fires.

A. Enis Çetin's research is partially funded by NSF with grant number 1739396 and NVIDIA Corporation. B. Uğur Töreyin's research is partially funded by TÜBİTAK 114E426, İTÜ BAP MGA-2017-40964 and MOA-2019-42321.

N. T. Nguyen et al. (Eds.): ICCCI 2020, LNAI 12496, pp. 807–815, 2020.
https://doi.org/10.1007/978-3-030-63007-2_63

Among different approaches, the use of the visible-range video captured by surveillance cameras is particularly convenient for fire detection, as they can be deployed and operated in a cost-effective manner [3]. One of the main challenges is to provide a robust vision-based detection system with negligible false positive rates while securing rapid response. If the flames are visible, this may be achieved by analyzing the motion and color clues of a video in the wavelet domain [5, 21]. Similarly, wavelet-based contour analysis [20] can be used for the detection of possible smoke regions. Modeling various spatio-temporal features such as color and flickering, and dynamic texture analysis [6] can detect fire, as well. In the literature, there are several computer vision algorithms for smoke and flame detection using wavelets, support vector machines, Markov models, region covariance, and co-difference matrices [4]. An important number of fire detection algorithms in the literature not only employ spatial information, but also use the temporal information [4,11,19].

Deep convolutional neural networks (DCNN) achieve successful recognition results on a wide range of computer vision problems [8,15]. Deep neural network-based fire detection algorithms using regular cameras have been developed by many researchers in recent years [9,10,22]. As opposed to earlier computer vision-based fire detection algorithms, in all of the existing DCNN-based methods, the temporal nature of flames is not utilized. Instead, flames are recognized from image frames. In this paper, we utilize the temporal behavior of flames to recognize uncontrolled fires. Uncontrolled flames flicker randomly. The bandwidth of the spectrum of flame flicker can be as high as 10 Hz [7]. To detect such behavior, we group the video frames and obtain temporal slice images. We process the temporal slices using deep convolutional networks.

Radford et al. [17] demonstrate that a class of convolutional neural networks, namely, Deep Convolutional Generative Adversarial Networks (DCGANs), can learn general image representations on various image datasets. In our earlier work, we utilize DCGANs for detecting wildfire smoke based on motion-based geometric image transformation [2]. We utilize a two-stage training approach to have a robust representation of sequences with and without smoke. We first train the network with real images and noise vectors and then train the discriminator using the smoke images without the generator. In a pre-processing stage before training, we integrate the temporal evolution of smoke with a motion-based transformation of images.

We propose utilizing the discriminator of a DCGAN to effectively distinguish ordinary image sequences without flame from those with flame. Additionally, we define a two-stage training approach for the DCGAN such that the discriminator is trained adversarially and to detect flame at the same time. Our main contribution is training the discriminator in such a way that the discriminator acts as a classifier that identifies the images with flame. We develop a DCGAN to utilize adversarial learning in a classification task.

The rest of the paper is organized as follows. Section 2 describes the proposed method that effectively detects flame. Section 3 presents the experimental results of the approach. Finally, Sect. 4 concludes the paper.

2 Method

We describe the proposed flame detection method in this section. In our method, we group the video frames in order to obtain temporal slice images. Then, we process the temporal slices using a DCGAN structure accepting input with size $64 \times 128 \times 384$ pixels. The generator network consists of a fully-connected layer followed by five transposed convolutional layers and the discriminator network consists of five convolutional layers with a fully-connected layer. Figure 1 depicts the neural network architectures and the training framework of the DCGAN.

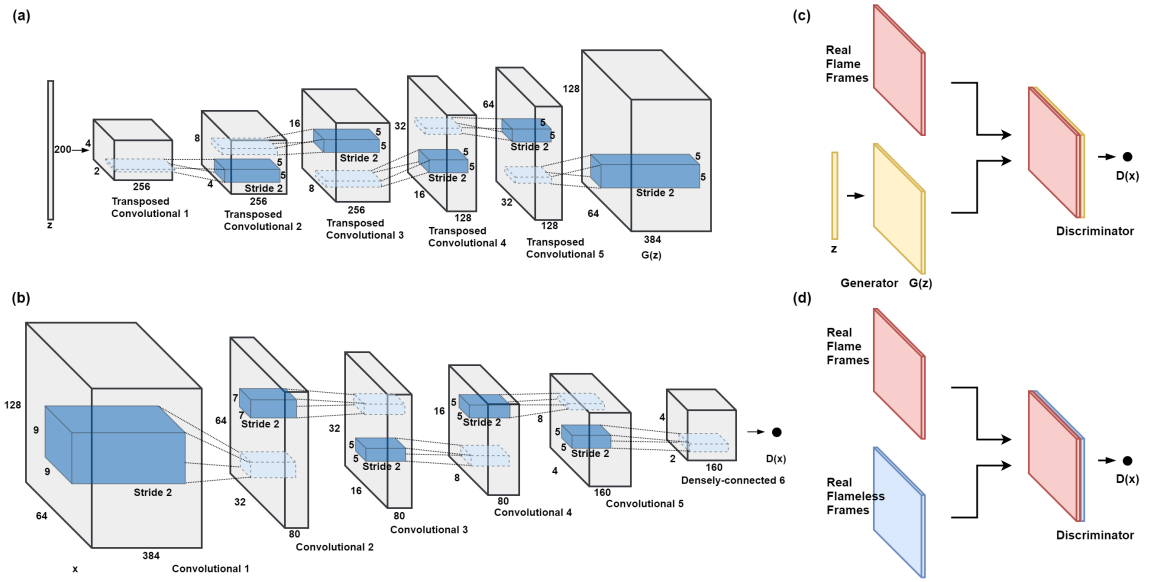

Fig. 1. The block diagram of the DCGAN: (a) the generator of DCGAN, (b) the discriminator of DCGAN, (c) the adversarial learning stage of training, and (d) the second stage of training

We first train the neural networks using a noise distribution z and images that contain flame. The discriminator of the DCGAN is trained to learn a representation of the temporal nature of flames and distinguishes non-flame videos because of the adversarial training. Then, we improve the classification performance by refining and retraining the discriminator network without a generator, where actual non-flame video images constitute the "generated" training data and regular flame images correspond to "real" data as usual. Compared to a generic CNN structure, adversarial training of the DCGAN using the noise vector z and flame data, in addition to the training with the actual non-flame data makes the recognition system more robust.

In our model, we use batch normalization [13] layers and Rectified Linear Unit (ReLU) activation function [16] after each transposed convolutional layer in the generator network, except the last layer, which has a hyperbolic tangent (tanh) activation. Similarly, for the discriminator network, we use batch normalization layers and Leaky ReLU activation function after each convolutional layer, except

the last layer, which has a sigmoid activation. During training, we randomly drop out [18] the outputs of the layers in the discriminator network to prevent overfitting and add Gaussian noise to the inputs to increase the generalization capabilities of the network. We initialize the weights of the convolutional and transposed convolutional layers according to the "MSRA" initialization [12]. Finally, we use Adam optimizer for the stochastic optimization [14]. We use the TensorFlow machine learning framework to train the discriminator and generator networks [1].

2.1 Temporal Slice Images

Exploiting the evolution of flames in time, we obtain slice images from video frames. We first split the videos into blocks containing 64 consecutive frames with size 128×128 pixels. Then, for each column, we extract the pixels along the time dimension, resulting in 128 different 128×64 pixel images (see Fig. 2).

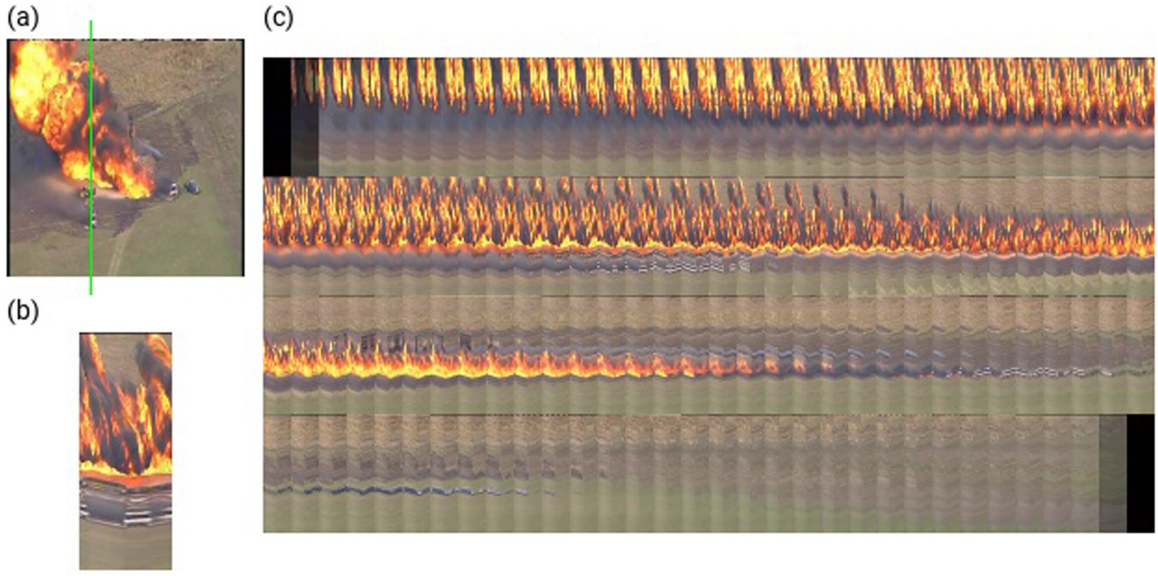

Fig. 2. (a) An example frame from the input video. (b) Temporal slice image of column corresponding to the green line in (a), where the leftmost column contains pixels from the initial frame, namely, the frame at time index $t = 1$, and the rightmost column contains pixels from the final frame, namely, the frame at time index $t = 64$ of the block. (c) Visualization of all 128 slice images. (Color figure online)

To feed the slice image data to the DCGAN model, we stack all 128 slices on top of each other. Thus, we obtain an RGB image cube of size $64 \times 128 \times 384$ pixels because the slice images have three channels each. Figure 3 shows an example of an image cube.

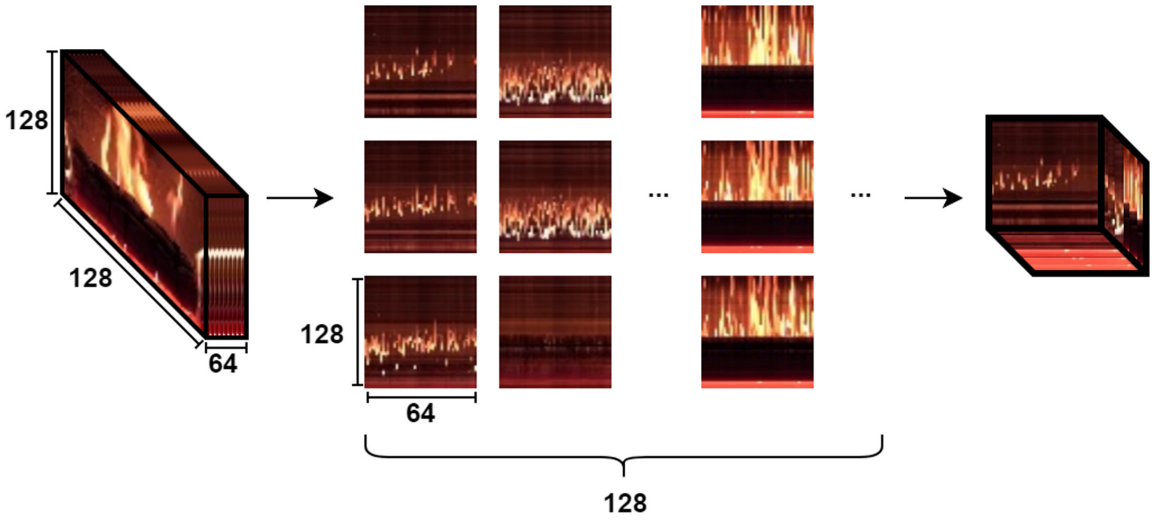

Fig. 3. An example image cube obtained from the input video.

2.2 Proposed GAN-type Discriminator Network

Flame, by its nature, have no particular shape or specific feature as some other objects such as faces, buildings, or cars. Hence, we focus on the temporal behavior of flame instead of spatial information. We utilize the discriminator network of the GAN to distinguish regular camera views from flame videos. This DCGAN structure produces above 0.5 probability value for real temporal flame slices and below 0.5 for slices that do not contain flame, because non-flame slices are not in the initial training set.

In the standard GAN training, the discriminator network that outputs a probability value, D, is updated using the stochastic gradient (Eq. 1)

$$SG_1 = \nabla_{\theta_d} \frac{1}{M} \sum_{i=1}^{M} (\log D(x_i) + \log(1 - D(G(z_i)))), \tag{1}$$

where z_i and x_i are the input noise vector and i^{th} temporal slice, respectively, and G represents the generator network of the GAN which outputs a "fake slice" based on the noise vector z_i; the vector θ_d includes the parameters of D. In this stage, we train the generator network adversarially as in [8]. During this first stage of training, we do not take slices from flame-less videos into account. This GAN can detect flame because discriminator is trained to distinguish flame from any other input. To improve the detection performance, we perform a second stage of training by fine-tuning the discriminator using the stochastic gradient given in Eq. 2.

$$SG_2 = \nabla_{\theta_d} \frac{1}{L} \sum_{i=1}^{L} (\log D(x_i) + \log(1 - D(y_i)), \tag{2}$$

y_i represents the i^{th} slice obtained from regular camera views. The number of non-flame slices, L, is smaller than the size of the slice samples of flame videos

that form the initial training set, M. We do not need to generate any artificial slices during the refinement stage, hence, we do not update the generator network at this stage of training.

3 Experimental Results

In this section, we present the experiments that we carry out for the proposed method. We use 112 video clips containing flame frames and 72 video clips without any flame frames in our experiments.

Throughout the experiments, we first obtain the temporal slice images for both flame and non-flame videos. To this end, we sample 10 frames every second, to be included in a block, i.e., a temporal slice. Because blocks contain 64 frames, they capture the motion for almost six and a half seconds. We partition the video clips into non-overlapping temporal slices. Each video clip has a duration of one minute. Consequently, the dataset is composed of over 210 thousand slices from over 1600 blocks in total.

After this procedure, we split the data into training, validation, and test sets. The training set consists of 60% of the videos and the validation and test sets consist of 20% of videos each. We fine-tune the hyperparameters of the neural networks based on the classification performance on the validation set, then report the final results achieved on the test set.

We evaluate the proposed method, namely, DCGAN with Temporal Slices, in terms of frame-based results. Because all the other deep learning methods are essentially based on CNNs, we compare CNN with Temporal Slices, DCGAN with Video Frames (no temporal information), and DCGAN without refinement stage-based approaches to our CNN implementation. It should be also noted that researchers use different fire datasets, therefore the recognition results are not comparable.

In our approach, we aim to reduce the false positive rate while keeping the hit-rate as high as possible. Experimental results show that the proposed method achieves the best results on the test set (cf. Table 1), where a false-positive rate of 3.91% is obtained corresponding to a hit-rate of 92.19%. We show that the adversarial training in DCGAN structure yields more robust results when compared to a CNN (same architecture as the discriminator). As for the utilization of temporal slices to exploit flame evolution, it can be seen that utilizing the temporal information of flames results in much lower false-positive rates. Figure 4 shows some examples of false negative and false positive temporal slices.

Table 1. The true negative rate (TNR) and true positive rate (TPR) values obtained on the test set for frame-based evaluation.

Method	TNR (%)	TPR (%)
DCGAN with Temporal Slices (Our method)	**96.09**	92.19
CNN with Temporal Slices	87.39	93.23
DCGAN with Video Frames (no temporal information)	92.55	92.39
DCGAN without refinement stage	86.61	90.10

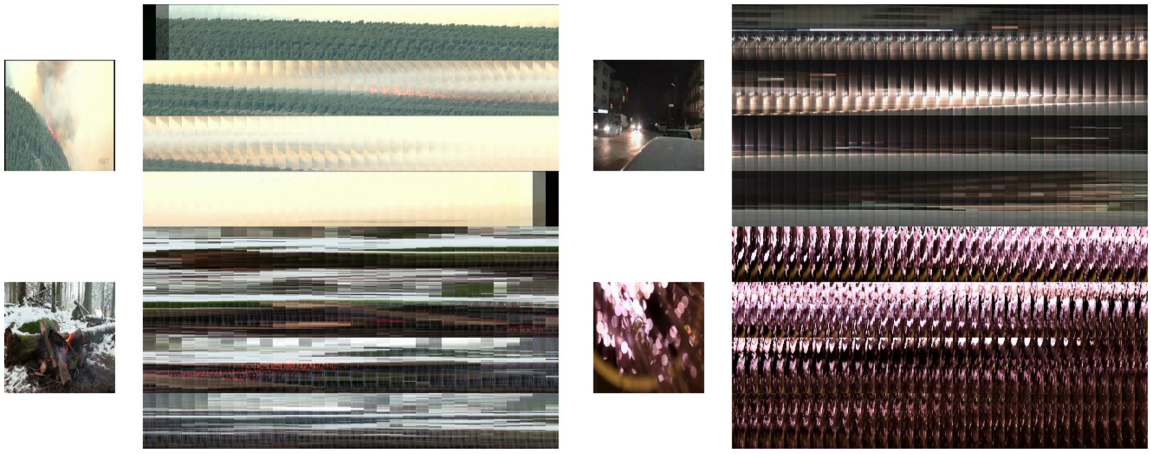

Fig. 4. Examples of false negative temporal slices on the left and false positive temporal slices on the right.

4 Conclusion

We propose a DCGAN-based flame detection method in video exploiting the spatio-temporal evolution of fires and employing an unsupervised pre-training stage. We develop a two-stage DCGAN training approach to represent and classify image sequences with and without flames. To reveal the spatio-temporal dynamics of flame regions, we acquire temporal slice images obtained from consecutive frames.

The main contribution of the proposed method is to utilize not only spatial information but also temporal characteristics of flame regions and the unsupervised representation learning capabilities of the DCGAN-based approach. The results indicate that the proposed method achieves significantly lower false alarm rates, compared to CNNs with temporal slices, while keeping the detection rates high.

References

1. Abadi, M., et al.: TensorFlow: a system for large-scale machine learning. In: Proceedings of the 12th USENIX Conference on Operating Systems Design and Implementation, OSDI 2016, pp. 265–283 (2016)

2. Aslan, S., Güdükbay, U., Töreyin, B.U., Çetin, A.E.: Early wildfire smoke detection based on motion-based geometric image transformation and deep convolutional generative adversarial networks. In: Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, ICASSP 2019, pp. 8315–8319. IEEE, Brighton (2019)
3. Çetin, A.E., et al.: Video fire detection-review. Digit. Signal Proc. **23**(6), 1827–1843 (2013)
4. Çetin, A.E., Merci, B., Günay, O., Töreyin, B.U., Verstockt, S.: Methods and Techniques for Fire Detection. Academic Press, Oxford (2016)
5. Dedeoğlu, Y., Toreyin, B.U., Güdükbay, U., Cetin, A.E.: Real-time fire and flame detection in video. In: Proceedings of the International Conference on Acoustics, Speech, and Signal Processing. ICASSO 2005, vol. 2, pp. ii–669. IEEE (2005)
6. Dimitropoulos, K., Barmpoutis, P., Grammalidis, N.: Spatio-temporal flame modeling and dynamic texture analysis for automatic video-based fire detection. IEEE Trans. Circuits Syst. Video Technol. **25**(2), 339–351 (2015)
7. Erden, F., et al.: Wavelet based flickering flame detector using differential PIR sensors. Fire Saf. J. **53**, 13–18 (2012)
8. Goodfellow, I., et al.: Generative adversarial nets. In: Advances in Neural Information Processing Systems, pp. 2672–2680 (2014)
9. Günay, O., Töreyin, B.U., Köse, K., Çetin, A.E.: Entropy-functional-based online adaptive decision fusion framework with application to wildfire detection in video. IEEE Trans. Image Process. **21**(5), 2853–2865 (2012)
10. Günay, O., Çetin, A.E.: Real-time dynamic texture recognition using random sampling and dimension reduction. In: Proceedings of the International Conference on Image Processing, ICIP 2015, pp. 3087–3091. IEEE (2015)
11. Habiboğlu, Y.H., Günay, O., Çetin, A.E.: Covariance matrix-based fire and flame detection method in video. Mach. Vis. Appl. **23**(6), 1103–1113 (2012)
12. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: surpassing human-level performance on ImageNet classification. In: Proceedings of the International Conference on Computer Vision, ICCV 2015, pp. 1026–1034. IEEE (2015)
13. Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift. CoRR abs/1502.03167 (2015). http://arxiv.org/abs/1502.03167
14. Kingma, D.P., Ba, J.: Adam: a method for stochastic optimization. CoRR abs/1412.6980 (2014). http://arxiv.org/abs/1412.6980
15. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**(1), 436–444 (2015)
16. Nair, V., Hinton, G.E.: Rectified linear units improve restricted Boltzmann machines. In: Proceedings of the 27th International Conference on Machine Learning, ICML 2010, pp. 807–814. Omnipress, Madison (2010)
17. Radford, A., Metz, L., Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks. CoRR abs/1511.06434 (2015). http://arxiv.org/abs/1511.06434
18. Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Dropout: a simple way to prevent neural networks from overfitting. J. Mach. Learn. Res. **15**(1), 1929–1958 (2014)
19. Toreyin, B.U., Cetin, A.E.: Online detection of fire in video. In: Proceedings of the Conference on Computer Vision and Pattern Recognition, CVPR 2007, pp. 1–5. IEEE (2007)
20. Toreyin, B.U., Dedeoğlu, Y., Cetin, A.E.: Contour based smoke detection in video using wavelets. In: Proceedings of the European Signal Processing Conference, EUSIPCO 2006 (2006)

21. Töreyin, B.U., Dedeoğlu, Y., Güdükbay, U., Cetin, A.E.: Computer vision based method for real-time fire and flame detection. Pattern Recogn. Lett. **27**(1), 49–58 (2006)
22. Zhao, Y., Ma, J., Li, X., Zhang, J.: Saliency detection and deep learning-based wildfire identification in UAV imagery. Sensors **18**(3), 712 (2012)