

Visualization System for Fire Detection in the Video Sequences

N.V. Laptev¹, V.V. Laptev², O. M. Gerget³, A.A. Kravchenko⁴, D.Yu. Kolpashchikov⁵

National Research Tomsk Polytechnic University

¹ ORCID: 0000-0003-0709-9974, nikitalaptev77@gmail.com

² ORCID: 0000-0001-8639-8889, vvl39@tpu.ru

³ ORCID: 0000-0002-6242-9502, olgagerget@mail.ru

⁴ ORCID: 0000-0001-6828-3279, aak224@tpu.ru

⁵ ORCID: 0000-0001-8915-0918, Dyk1@tpu.ru

Abstract

The paper deals with the analysis of the visual images obtained from fire detection systems. We review the existing approaches to the analysis of video surveillance data and propose a tool for data labeling and visualization. The proposed solution for visual image analysis is based on a neural network (object detection technology). Recognition of hazard locations was carried out using the EfficientDet-D1 model. Video pre- and post-processing algorithms were implemented to improve visual image classification. The pre-processing was used to generate a frame preserving the features of objects that dynamically change over time. The post-processing combines the results of sequential detection of characteristic features on each frame, in particular, features of a smoke cloud. The results of the system operation are presented: visual image classification accuracy was 81%, while localization accuracy was 87%.

Keywords: computer vision, neural network, object detection, video analysis, image visualization, machine learning, algorithm.

1. Introduction

Advanced analysis of visual images obtained from fire detection systems, accurate localization and performance measurements play a crucial role in preventing environmental disasters and minimizing environmental damage. The importance of developing real time visual systems for accurate fire detection and forest fire location is beyond doubt.

The most of existing forest fire detection systems are based on mathematical models. Among them the model of a forest fire as a source of infrared radiation [1] was used to highlight forest fire contours based on infrared emission data. Another model processes the transparency of the atmosphere [2] based on two factors: the brightness of objects and the presence of suspended particles in the atmosphere. However, building reliable mathematical models is a complicated task that is often impossible due to a large amount of dynamic information in surveillance videos, hidden patterns between input data and the complexity of identifying characteristic features.

However, the described limitations appear non-essential for a visual system based on technical vision. Such systems can analyze camera images and timely identify fire sources, which makes them suitable for early warning of fire. Fire detection systems based on technical vision are cost effective and can be used to analyze massive amounts of visual images.

One solution based on technical vision was described in [3] and is of great interest for fire detection. This method was based on combining HSV and YCbCr color models. In contrast to traditional methods, this approach allowed additional transformations of the color space, which improved the quality of recognition. However, this system was focused on the static characteristics of the flame, which can negatively affect the ultimate detection accuracy.

Another solution was presented in [4]. The authors analyse temporal and spatial factors of fire, while using the Gaussian model to describe the HSV color model. Despite many positive aspects, the Gaussian model requires immense computation time when working with a large amount of data, which is unacceptable for real-time solutions as the analysis occurs fuzzy.

In [5], the authors propose to use a combination of SVM and GLCM methods to detect and segment a hard-to-find object. This approach can also be applied to the current task in order to improve the detection accuracy.

Despite several existing systems and approaches, the problem of visual image analysis remains relevant and understudied, including the task of precise fire detection and localization. Meanwhile, machine learning has been successfully used to achieve high-quality object detection [6], [7], e.g. based on object detection technology. Therefore, the use of neural networks for automated surveillance and accurate forest fire detection was investigated in this study.

2. Initial data

The initial data for the study was obtained from several sources, in particular, from the open online resources of Nevada Seismological Laboratory (University of Nevada) [8], [9], Center for Wildfire Research (sponsored by the University of Split) [10], and Perm forestry [11]. For the purpose of direct training, all data were labeled using the SuperVisely web service [12]. This service allows not only visualizing the labeled data, but also operating the data in a semi-automatic mode. The total number of collected video recordings was 1000, including 766 videos that contained fire and 234 videos without fire. A sequence of 7 frames was extracted from each video for training. This number of frames was obtained experimentally. The statistics of the test frames with and without fire, as well as the number of annotated areas are given in Table 1.

Table 1. Initial training and test data

Quantity, pcs	Total number of frames, 7000			Number of frames in the training set, 6300			Number of frames in the test set, 700		
	300	2100	4600	100	1850	4350	200	250	250
Resolution, pixels	400	1280	1920	400	1280	1920	400	1280	1920
	x	x	x	x	x	x	x	x	x
	400	1024	1080	400	1024	1080	400	1024	1080

3. Dynamic feature extraction algorithm

One of the informative discriminative features in surveillance video is the shape of smoke, which constantly changes due to the fire dynamics. Some traditional methods of fire detection take this condition into account, including continuous frame change [13], background subtraction [14], frame difference and background modeling using Gaussian mixture model [15]. Among modern methods [16] there is a neural network approach to image processing based on the use of generative adversarial networks.

In this work, the emphasis was placed on the investigation and implementation of the frame difference method. This method demonstrates the advantage of insensitivity to scene changes (e.g., to lighting) and the ability to adapt to various dynamic environments with good stability. However, it fails to extract the complete area of an object. This work proposes an improved frame difference method.

The designed algorithm for extracting dynamic features of an image based on the frame difference method included the following steps:

Step 1. A video stream was converted into a sequence of frames.

Step 2. Frames were extracted from three RGB channels at a certain interval and converted into one channel (transition to grayscale) to save computation time during the following steps.

Step 3. An averaged frame was calculated according to Formula (1). This operation reduced camera noise and increased the stability of results.

Further frame pre-processing was neglected, since adding, for example, the Gaussian blur using a 5x5 kernel lead to an accuracy drop by the average of 15-20% in the object detection step as demonstrated in the test results.

Step 4. A dark frame was created based on the differences between the original and average frames. The difference was calculated according to Formula (2).

Step 5. Noise was reduced according to Formula (3). This operation allowed highlighting the objects with greater dynamics, while removing extraneous noise.

$$F_c(x, y) = \frac{1}{N} * \sum_{i=1}^N F_i(x, y), \quad (1)$$

$$F_{d_i}(x, y) = |F_c(x, y) - F_i(x, y)|, i = 1, 2, \dots, N, \quad (2)$$

$$F_{r_i}(x, y) = \{maxval, \text{ if } F_{d_i}(x, y) > thresh, 0, \text{ otherwise, } \}, \quad (3)$$

where $F_c(x, y)$ – averaged frame, N – the total number of processed frames, $F_i(x, y)$ – the current frame of the sequence, $F_{d_i}(x, y)$ – the difference between the current frame of the sequence and the averaged frame. $F_{r_i}(x, y)$ – the resulting frame after noise reduction.

Since the air flow and combustion properties cause a constant change in the flame pixels [17], pixel images without fire can be removed by comparing two consecutive images.

It should be noted that ten-second recordings from static cameras were considered in the experiments. Each video sequence was divided into frames, where seven frames were extracted at equal time intervals. It was expected to receive four processed frames (number 1, 3, 5, 7) at the output for high-accuracy fire detection.

4. Object detection

After pre-processing, each received frame was sequentially processed using the EfficientDet-D1 object recognition model. The general architecture of EfficientDet [18] largely corresponds to the paradigm of one-stage detectors. It is based on the EfficientNet model previously trained on the ImageNet dataset. A distinctive feature of the EfficientDet-D1 model [19]–[22] is an additional weighted bi-directional feature pyramid network (BiFPN) followed by class and block networks used to generate predictions of object classes and bounding boxes (boxes), respectively. A box had four parameters: two coordinates (x, y) for the upper left corner and two coordinates for the lower right corner. The network was trained using frames labeled with boxes indicating the class.

The other object detection models considered for the experiments include: EfficientDet-Do, EfficientDet-D1, SSD ResNet50 v1, SSD MobileNet v2, Faster R-CNN ResNet50 V1, Faster R-CNN Inception ResNet. All models were trained under the same conditions. The quality of visual image analysis based on the neural network models was assessed according to three criteria: Mean Average Precision (MAP), Accuracy (classification accuracy), and Speed (time of processing one frame). The results are shown in Table 2 and Table 3.

Table 2. Model performance on the test set

Model	Input size	Weight, mb	Accuracy	MAP	Speed, s
EfficientDet-Do	512x512	18.6	0.6	0.336	0.03
EfficientDet-D1	640x640	24.9	0.69	0.514	0.11
SSD ResNet50 v1	640x640	10.1	0.39	0.64	0.08
SSD MobileNet v2	640x640	7.215	0.64	0.46	0.04
Faster R-CNN ResNet50 V1	640x640	4.597	0.45	0.32	0.26
Faster R-CNN Inception ResNet V2	640x640	18.2	0.55	0.12	0.58

Table 3. Dependence of model performance on image resolution

Model	Accuracy		
	400x400	1280x1024	1920x1080
EfficientDet-D0	0.57	0.53	0.64
EfficientDet-D1	0.51	0.71	0.63
SSD_Resnet50_v1	0.49	0.3	0.37
SSD MobileNet v2	0.50	0.75	0.59
Faster R-CNN ResNet50 V1	0.53	0.45	0.39
Faster R-CNN Inception ResNet V2	0.41	0.62	0.54

As illustrated in Tables 2-3, EfficientDet-D1 showed the highest efficiency. It is also worth noting the efficiency of the SSD ResNet50 model in object localization. On the contrary, Faster R-CNN Inception ResNet V2 demonstrated the lowest efficiency.

5. Post-processing

The post-processing algorithm is laid out in Figure 1. It includes the following sequence of actions: four frames of the same perspective but spaced in time arrived at the input of the neural network. Since the smoke has a very unstable structure (density, variability of shape, direction of movement), the shape of a smoke cloud was different in all frames. Thus, in most cases, the algorithm selected the most discriminative areas of smoke at a given time, which can be clearly seen in Figure 1. After this step the resulting frame had up to 20 bounding boxes, which highlighted the same object with different probabilities. Then two or more boxes that overlapped more than 25% were merged. This approach made it possible not only to detect fire with great confidence, but also to localize it with increased accuracy. Figure 2 shows the impact of the algorithm on system performance: the resulting image before post-processing (a, the detection probability 59 and 17 %) and after post-processing (b, the detection probability 65%). Figure 3 shows a histogram, which reflects the dependence of the fire detection accuracy on the area of box overlapping at the stage of merging.

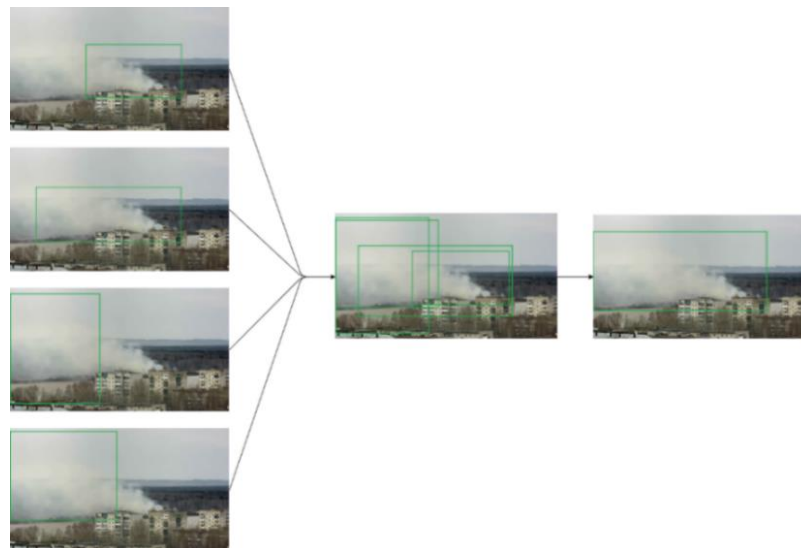


Fig. 1. The principle of the post-processing algorithm

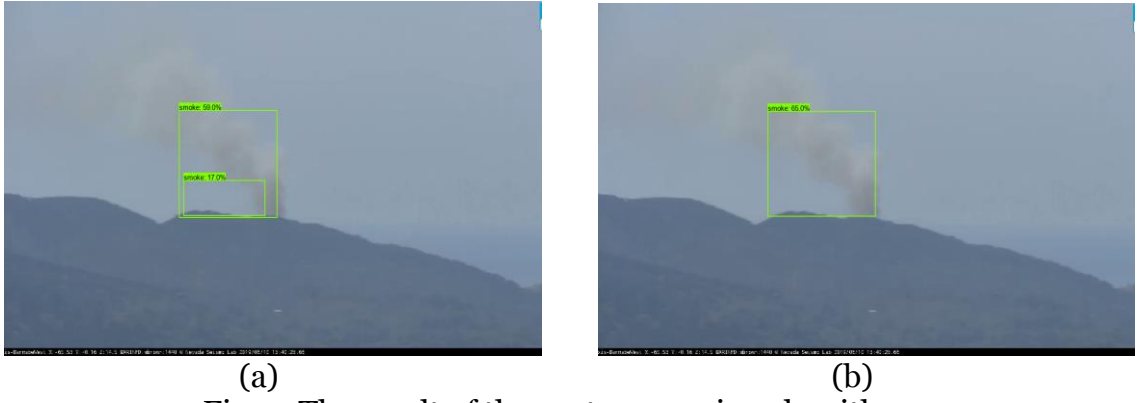


Fig. 2. The result of the post-processing algorithm

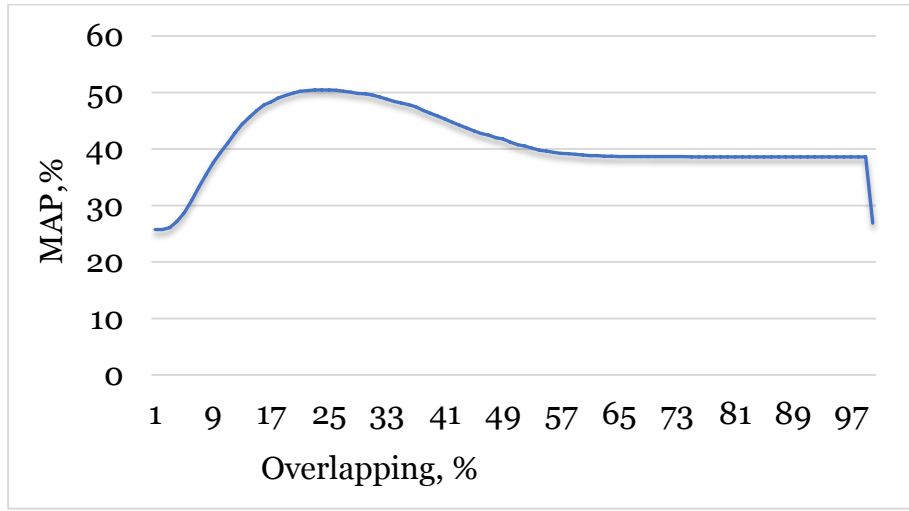


Figure 3. The dependence of the detection accuracy on the area of box overlapping

It should be noted that only those bounding rectangles (boxes) were applied to the analyzed frames, which had the probability of more than 15%. The percentage of overlapping was calculated according to the IOU (Intersection over Union) metric using Formula (5). The final box area was calculated according to Formula (6). The overall system result is presented in Table 4. The probability of detecting and merging bounding boxes was determined empirically. The aim of the experiments was to maximize the number of fire detections and minimize the number of false alarms.

$$IOU = \frac{Area\ of\ Overlap}{Area\ of\ Union}, \quad (5)$$

$$C_{total} = \frac{C_1^2 + C_1^2 + \dots + C_n^2}{\sum C_n}, \quad (6)$$

where Area of Overlap – the area of overlapping between the predicted areas 1 and 2, Area of Union – the total predicted area, C_i – the area of the i -th box.

6. Results

The sequence of steps in the video processing algorithm is visualized in Figure 4, where the original image is marked (a), the conversion of a color frame to grayscale is marked (b), the result of subtracting the averaged frame from the sequence frame according to Formula (2) is marked (c), and the resulting image after noise reduction according to Formula (3) is marked (d). The resulting frame has a sharp outline of smoke and a minimal number of objects, which remained in the image due to various types of noise.

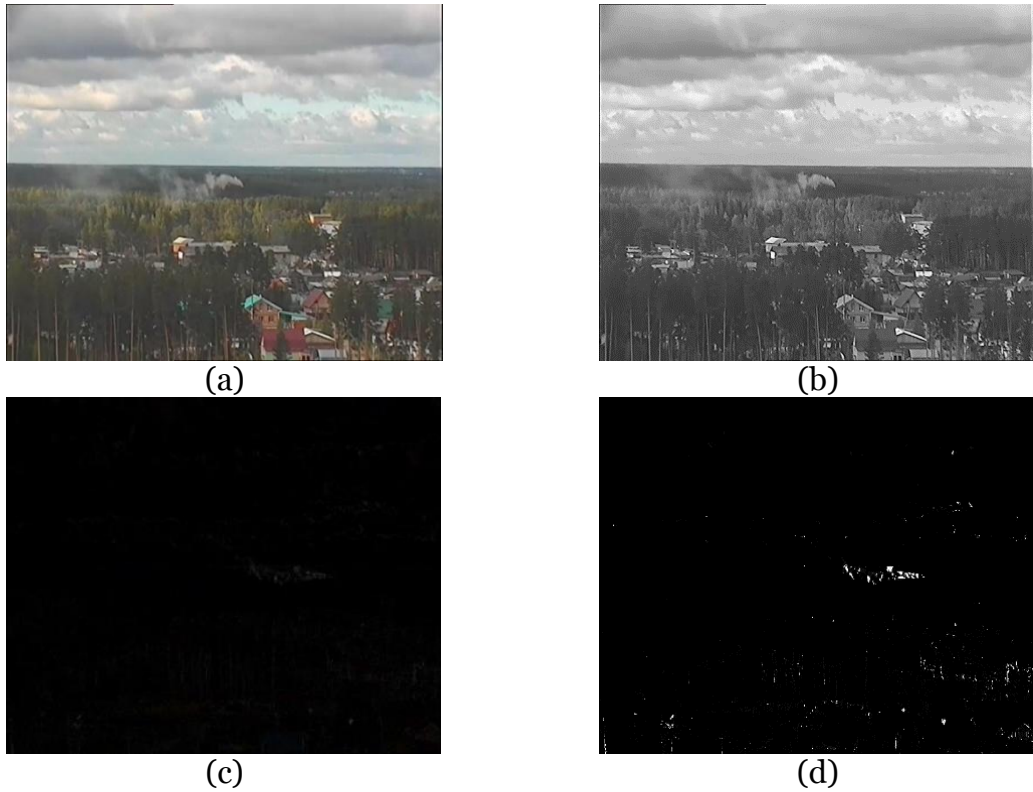


Figure 4. Visual interpretation of steps in the video processing algorithm

Figure 5 illustrates the results of detecting fire hazards. These are output frames. An example of correct detection is marked (a). In this frame the entire area of smoke was enclosed in a bounding box. Image (b) gives an example of detecting objects that fit into the class of smoke, but do not belong to the class of fire hazards. An example of partially correct operation of the system is marked (c). Here a fire hazard was detected, but an object that should not have been classified as fire hazards was also detected with less probability. System malfunction is marked (d). The detected object has discriminative features that are similar to a smoke cloud, but is not a fire hazard.

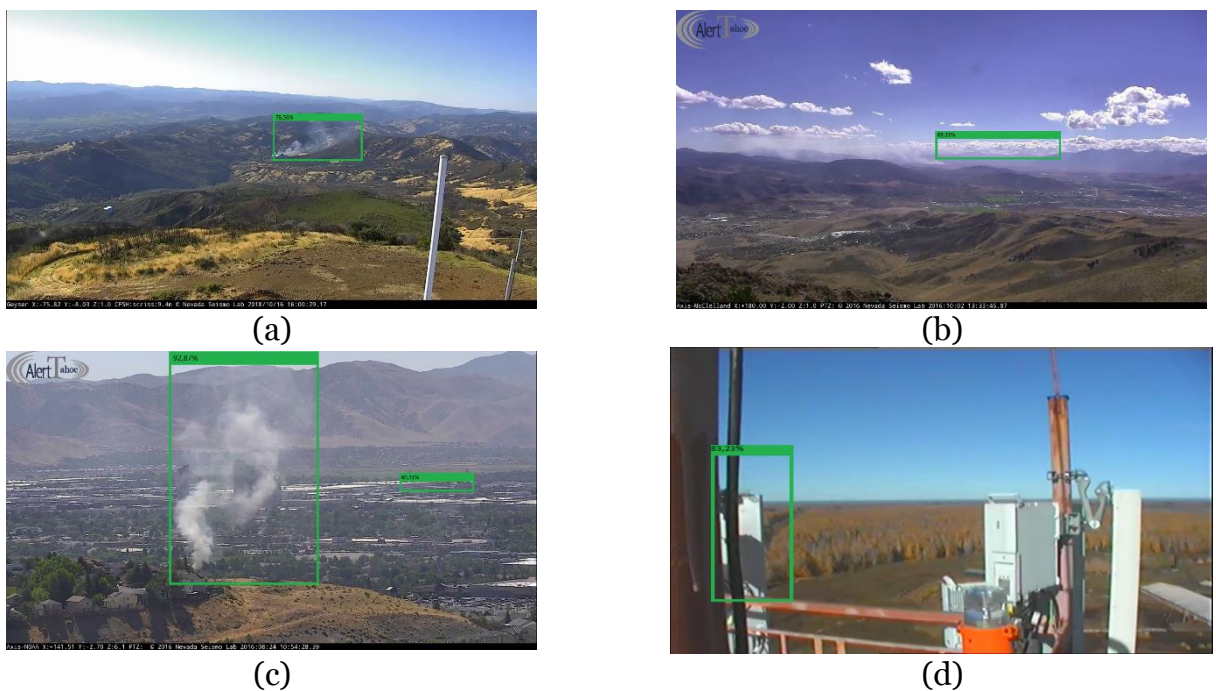


Figure 5. Object detection results

Table 4. System performance

Model	Classification					Localization
	Accuracy	N	FN	P	FP	MAP
EfficientDet-Do	0.67	234	124	466	107	0.72
EfficientDet-D1	0.81		88		45	0.87
SSD ResNet50 v1	0.61		197		76	0.79
SSD MobileNet v2	0.69		94		123	0.70
Faster R-CNN ResNet50 V1	0.58		175		119	0.70
Faster R-CNN Inception Res-Net V2	0.62		128		111	0.74

In Table 4, N is the number of negative frames (frames with no fire), FN is a false negative detection result (fire was detected in the frames with no fire), P is the number of positive frames (frames with fire), FP is a false positive result (no fire was detected in the frames with fire).

Therefore, the EfficientDet-D1 architecture has shown the greatest efficiency in visual image analysis and smoke detection. Its classification accuracy was 69%, and fire localization accuracy exceeded 51%. Further use of the post-processing algorithm makes it possible to reduce the fire detection error in frames with no smoke, which can increase classification accuracy up to 81% and localization accuracy up to 87%. Localization accuracy can be raised up to 92%, but this will drop classification accuracy due to numerous system operations on frames with no fire.

7. Conclusion

The proposed in this work technology has been used to solve the problem of analyzing visual images obtained from fire detection systems. The approach outlined in the article was capable of removing noise and highlighting dynamic features when visualizing images. Thus, only the objects with the most pronounced features remained in analyzed frames, which had a positive effect on the detection and localization accuracy.

An important characteristic of the developed system is the image pre-processing stage. Besides, the use of the dynamic feature extraction algorithm makes it possible to process frames of different resolutions. This expands the application of the technology, as it can be used for cameras with different resolutions.

The post-processing algorithm also occupies an important place in the operation of the visual system. During the experiments, it increased the localization accuracy by 36% and the classification accuracy by 12%, which is a significant improvement for the detection problem. The final demonstrated classification accuracy was 81% and localization accuracy was 87%.

The next important attribute of the system is visualization. The fire detection results were visualized using the capabilities of the Tensorflow library. High-quality image visualization and analysis is important when an operator makes a decision.

At the moment, the authors continue solving the problem of additional classification for various negative visual images others than a smoke cloud, which should significantly increase the efficiency of the presented technology.

Acknowledgement

The study has been funded with support from the Russian Foundation for Basic Research, projects No. 20-37-90055 and No. 19-07-00351.

References

- [1] A. Vasiliev and A. Krasnyashchikh, "Mathematical model of a forest fire as a source of infrared radiation," *cyberleninka.ru*, Accessed: Apr. 29, 2021. [Online]. Available: <https://cyberleninka.ru/article/n/16469818>.
- [2] V. M. Gusyatina and A. P. Ostroushko, "Mathematical model and algorithm for processing weather conditions for visualization systems," 2000. Accessed: Apr. 20, 2021. [Online]. Available: https://openarchive.nure.ua/bitstream/document/7548/1/ASU_2000-111-9-14.pdf.
- [3] J. Seebamrungsat, ... S. P.-2014 T. I., and undefined 2014, "Fire detection in the buildings using image processing," *ieeexplore.ieee.org*, Accessed: Dec. 02, 2020. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/6923226/?casa_token=3CXE3OSK2J8AAAAA:3v-HZxPnQh9mmSkL_J8VwTb9UwiKEQkRjrlcZxi6OAKV8SDwMm6yjCMajde3xs4unPrXPkohOuOe.
- [4] L.-H. Chen and W.-C. Huang, *Fire Detection Using Spatial-temporal Analysis*.
- [5] V. V. Danilov, I. P. Skirnevskiy, R. A. Manakov, D. Y. Kolpashchikov, O. M. Gerget, and F. Melgani, "Catheter detection and segmentation in volumetric ultrasound using SVM and GLCM," *Sci. Vis.*, vol. 10, no. 4, pp. 30–39, 2018, doi: 10.26583/sv.10.4.03.
- [6] D. Alexandrov, E. Pertseva, ... I. B.-... 24th C. of, and undefined 2019, "Analysis of machine learning methods for wildfire security monitoring with an unmanned aerial vehicles," *ieeexplore.ieee.org*, Accessed: Dec. 02, 2020. [Online]. Available: https://ieeexplore.ieee.org/abstract/document/8711917/?casa_token=8gaW9VszHEgAAAA:RoouzYoDduzDBHluM-pA_bYPfBd4FW9-rgIgg9EieK6ZJAYJN8yw9ai3Ffa1FfkxJnXaoXTsC_WUO.
- [7] V. Danilov, "Comparative Study of Deep Learning Models for Automatic Coronary Stenosis Detection in X-ray Angiography," 2020. Accessed: Feb. 08, 2021. [Online]. Available: <http://ceur-ws.org/Vol-2744/paper75.pdf>.
- [8] "Nevada Seismological Laboratory." <https://www.youtube.com/user/nvseismolab/about> (accessed Dec. 02, 2020).
- [9] "Nevada Seismological Laboratory, University of Nevada." <http://www.seismo.unr.edu/> (accessed Dec. 02, 2020).
- [10] "Wildfire Observers and Smoke Recognition." <http://wildfire.fesb.hr> (accessed Nov. 19, 2020).
- [11] "Perm forest fire center." https://www.youtube.com/channel/UCsKn1hQgGh5n7NGoqLNoh_Q/videos (accessed Nov. 19, 2020).
- [12] "Supervisely - Web platform for computer vision. Annotation, training and deploy." <https://supervise.ly/> (accessed Dec. 02, 2020).
- [13] T. Song *et al.*, "IEEE TRANSACTIONS ON NANOBIOSCIENCE Spiking Neural P Systems with Learning Functions." Accessed: Dec. 02, 2020. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/8632967/>.
- [14] A. Aggarwal, S. Biswas, S. Singh, S. Sural, and A. K. Majumdar, "Object tracking using background subtraction and motion estimation in MPEG videos," in *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, 2006, vol. 3852 LNCS, pp. 121–130, doi: 10.1007/11612704_13.
- [15] T. Song, X. Zeng, P. Zheng, ... M. J.-I. transactions on, and undefined 2018, "A parallel workflow pattern modeling using spiking neural P systems with colored spikes," *ieeexplore.ieee.org*, Accessed: Dec. 16, 2020. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/8478391/>.
- [16] V. V. Kniaz, M. I. Kozyrev, A. N. Bordodymov, A. V. Papazian, and A. V. Yakhanov, "Segmentation and visualization of obstacles for the enhanced vision system using gen-

- erative adversarial networks,” *Sci. Vis.*, vol. 11, no. 4, pp. 43–52, 2019, doi: 10.26583/sv.11.4.04.
- [17] W. Yang, M. Mörtberg, and W. Blasiak, “Influences of flame configurations on flame properties and NO emissions in combustion with high-temperature air,” *Scand. J. Metall.*, vol. 34, no. 1, pp. 7–15, Feb. 2005, doi: 10.1111/j.1600-0692.2005.00710.x.
- [18] M. Tan, R. Pang, and Q. V. Le, “EfficientDet: Scalable and efficient object detection,” in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2020, pp. 10778–10787, doi: 10.1109/CVPR42600.2020.01079.
- [19] S. Liu, L. Qi, H. Qin, J. Shi, and J. Jia, “Path Aggregation Network for Instance Segmentation.” Accessed: Dec. 02, 2020. [Online]. Available: http://openaccess.thecvf.com/content_cvpr_2018/html/Liu_Path_Aggregation_Network_CVPR_2018_paper.html.
- [20] J. Redmon and A. Farhadi, “YOLO9000: Better, faster, stronger,” in *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, Nov. 2017, vol. 2017-January, pp. 6517–6525, doi: 10.1109/CVPR.2017.690.
- [21] T.-Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, “Feature Pyramid Networks for Object Detection.” Accessed: Dec. 02, 2020. [Online]. Available: http://openaccess.thecvf.com/content_cvpr_2017/html/Lin_Feature_Pyramid_Networks_CVPR_2017_paper.html.
- [22] T. Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, “Focal Loss for Dense Object Detection,” *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 42, no. 2, pp. 318–327, Feb. 2020, doi: 10.1109/TPAMI.2018.2858826.