COMPUTER SCIENCES AND INFORMATION TECHNOLOGIES

UDC 004.93(045)
DOI:10.18372/1990-5548.74.17290

[1]V. M. Sineglazov,
[2]A. A. Komapov

INTELLIGENT ON-BOARD FOREST FIRE SEARCH SYSTEM

Aviation Computer-Integrated Complexes Department, Faculty of Air Navigation Electronics and Telecommunications, National Aviation University, Kyiv, Ukraine
E-mails: [1]svm@nau.edu.ua ORCID 0000-0002-3297-9060, [2]nazarkomarov0410@gmail.com

Abstract—*The paper analyzes the situation with forest fires in Ukraine. It is shown that the situation is deteriorating every year. For forest fire monitoring it is substantiated the need of the integrated use of data from satellites and unmanned aerial vehicles. It has been shown that early detection of a fire before it becomes a disaster is critical to preventing catastrophic fires and saving lives and property. A fire detection approach based on the use of computer vision methods that can work with a non-stationary camera installed on board the unmanned aerial vehicle is substantiated. An approach for detecting a "spot" of fire using convolutional neural networks is proposed. In our task of detecting a forest fire using an unmanned aerial vehicle, tracking based on detection is chosen as the model initialization method, when objects are first detected using the detection method and then linked into tracks (association). The Yolov4-tiny architecture was chosen as the architecture of the neural network detector, which provides high accuracy and speed of binary classification.*

Index Terms—Fire detection; convolutional neural networks; unmanned aerial vehicles; YOLO; R-CNN; single shot MultiBox detector; classifire.

I. INTRODUCTION

For Ukraine, the forest is of great importance, since the forest fund occupies 10.4 million hectares, which is 17.2% of its territory, of which 9.6 million hectares are covered with forest vegetation (the forest cover of the territory of Ukraine, therefore, is 15.9%). The highest forest cover is in the Ukrainian Carpathians (32%).

More than 40,000 hectares of forest fund territories are annually covered by forest fires in the country. Forest fires occur mainly as a result of careless handling of fire. The share of fires of natural origin in comparison with fires caused by anthropogenic activities is small. Fires are easier to prevent than to eliminate, however, in our country, the forest monitoring system is not sufficiently developed, which leads to untimely notification of a fire.

Detecting a fire at an early stage before it turns into a catastrophic event is critical to neutralizing fires and saving lives and property. Due to the rapid spread of fire, it is important to detect it at the stage of the first manifestations of fire.

Despite the importance of this problem, it remains unsolved at the moment. Detectors, the basic principle of which is to detect smoke, work well when a fire has been going on for a certain time and a sufficient amount of smoke has been formed to trigger the alarm. These devices cannot be used on a large scale outdoors, such as in the forest. Satellite systems, fixed remote cameras, and manned aircraft are common technologies for remote fire monitoring, but these technologies have a number of limitations in terms of cost, temporal, and spatial resolution.

Satellite systems, unmanned aerial vehicles are more focused on solving the given task, but require modern means of information processing, which can be artificial intelligence systems using neural networks.

II. FIRE MONITORING PROBLEMS

Although the coverage of the territory when using the satellite monitoring system is quite large, the accuracy and timeliness of detecting a fire or its previous state only by processing data from satellites is currently insufficient.

On the other hand, the use of unmanned aerial vehicles (UAVs) for monitoring purposes requires a fairly large number of them due to a small viewing area, although with a higher resolution than conventional weather satellites.

The joint use of data from satellites and UAVs will allow, on the one hand, to obtain an approximate reference to a more or less localized problem area of the territory where to direct the UAV, which will allow positioning and identify the problem at the required level of accuracy.

At the same time, the use of UAVs alone without satellite monitoring data is not a very efficient process of random walk and, accordingly, random detection of a forest fire or other emergency situation.

In the simplest version of efficiency assessment, it is natural to take as such the average time required to detect a forest fire source, taking into account the probability of an erroneous alarm and/or missing an event. In general, an automated system for aerospace monitoring of forest fires can be represented as follows (Fig. 1).

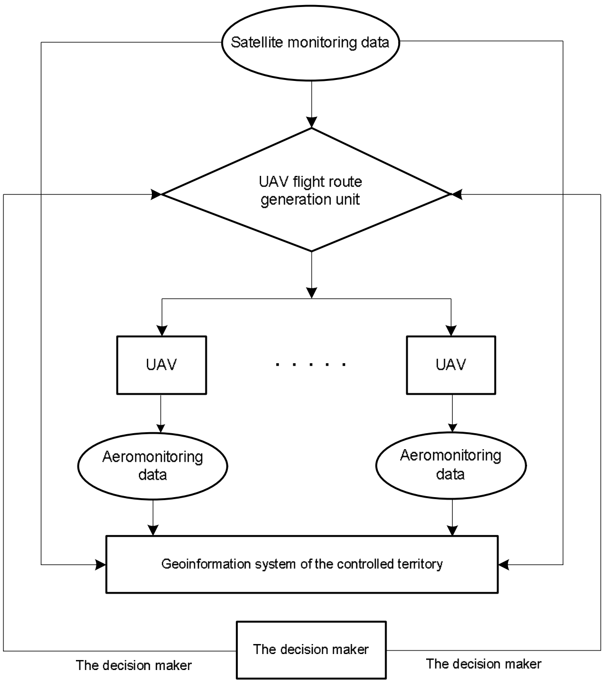

Fig. 1. Structure of the automated system for aerospace monitoring of forest fires

III. JUSTIFICATION OF THE USE OF ARTIFICIAL INTELLIGENCE METHODS FOR THE SEARCH AND LOCATION OF FOREST FIRES

Detecting a fire early, before it becomes a disaster, is critical to preventing catastrophic fires and saving lives and property. While fire and smoke detectors are widely installed indoors, they usually require the fire to burn for a while to generate a large amount of smoke and then trigger the alarm. In addition, these devices cannot be placed in large-scale outdoor environments, such as in the forest, in the wild. In contrast, the vision-based fire detection system captures images from cameras and immediately detects a fire, which is suitable for early fire detection. Such a system is also cheap and easy to install. Therefore, it is advisable to use a fire detection approach based on the use of computer vision methods that can work with a non-stationary camera. A fire detection system can be installed on unmanned aerial vehicles (UAVs) to detect large-scale forest fires.

Several methods of visual fire detection are known. Color model movement, spatial and temporal characteristics are mainly used because fire has very specific characteristics compared to another object. Nearly all of the proposed methods follow a similar detection pipeline, i.e. first find moving pixels using background subtraction and then apply a color model to find areas of fire color. These areas are further analyzed spatially and temporally to detect irregular and flickering fire characteristics. Since motion is the dominant feature, these methods only work with fixed cameras, i.e. in surveillance scenarios. To obtain reliable results, it is advisable to use deep neural networks (convolutional) to extract significant features from the data and train discriminatory classifiers to detect fires. Since fire is a non-rigid object with dynamic shapes, most approaches use motion and color features [1], [2] or spatiotemporal features [3], while background subtraction is widely used to improve accuracy and reliability [4]. In addition, most of the work involves fire detection in the observation scenario, i.e. the camera is installed in a fixed location to detect flames and/or smoke, which is unacceptable when using a UAV. Therefore, in this paper, we consider the structural-parametric synthesis of hybrid convolution neural networks for the analysis of dynamic visual scenes, for example, when the camera is mounted on a UAV, which, in turn, makes background subtraction and motion analysis difficult.

The proposed approach is to perform the following steps.

1) Obtaining moving pixels and areas [4], [5].

2) Extraction of possible areas of flame and/or smoke using a color model, such as HIS, which is used in [4], [5].

3) Further analysis of candidate regions, e.g. foreground area analysis [5], dynamic fire behavior analysis [2], [6].

Instead of following the traditional approach of fire detection based on computer vision methods, we will train the CNN on the selected features of the fire seat and fire classifiers. Deep convolution neural networks have shown great potential for solving many computer vision problems, such as object recognition [7], detection [8], semantic segmentation [9]. In article [10], to detect a fire seat when using non-stationary cameras, a covariance descriptor based on color, spatial and temporal information was proposed, and SVM training as a classifier was proposed. The SVM classifier is also used in [3]. Unlike known methods [3], [7] – [10] that require carefully designed hand-crafted features, [11] uses a CNN to learn the feature representation from raw pixels, so no color model or spatiotemporal information is needed, though such information may supplement the method.

The following criteria are used as an estimate of the fire classifier both at the image level and at the patch level (the patch determines the number of objects in one iteration, the number of iterations determines one training epoch) the following criteria are used:

$$\text{accuracy} = \frac{TP + TN}{N_{\text{pos}} + N_{\text{neg}}},$$
$$\text{detection rate} = \frac{TP}{N_{\text{pos}}}, \qquad (1)$$
$$\text{false alarm rate} = \frac{FN}{N_{\text{neg}}},$$

where TP is the number of true positives, i.e. the number of images with fires that are classified as fire; TN is the number of true negatives, i.e. the number of non-flammable patches that are classified as non-flammable; FN is the number of false negatives, i.e. the number of non-flammable patches that are classified as fire. N_{pos} and N_{neg} are the number of positive and negative results in the ground truth, respectively.

The main advantage of this approach is the ability to detect fire at the spot level. In this case, the classifier solves the problem of binary classification with the appropriate formation of the training sample. There are two options for implementing this approach:

1) train a binary classifier using complete images;

2) firstly train the fire classifier on full images and then apply the fine-grained fragment classifier if the image is classified as containing fire.

Option 1

Training a binary classifier using full image patches is not difficult. Efficient training of such a classifier is quite difficult due to the imbalance of the sample in two classes: there is an area with fire in the full image or it is not. Creating a balanced sample requires the use of generative models. To solve the problem posed, the problem of structural parametric synthesis of CNNs was solved using the approach described in [12].

Option 2

One of the disadvantages of the above patch classifier is that the classifier is trained on patch-level fire examples that do not contain global information.

In addition, the amount of computation by these spot (foci) classifiers is significant due to the fact that they must test several patches for each frame.

To solve the problem of lack of global information, it is necessary to train a complete image classifier based on "positive" (containing at least one torch) images and "negative" (not containing torches) images. Because we have a very small training dataset, training a complete image classifier from scratch can suffer from overfitting. One effective way to train a CNN classifier on a small dataset is to train it in two steps. At the first stage, it is pre-trained on full images (global information). At the second stage, the classifier is retrained on a short sample.

After training a complete image classifier, we can classify whether an image contains fire areas or not.

If fires exist, we apply a fine-grained fires classifier to find the exact location. One way is to directly apply a foci (spot) classifier, however, simply applying a foci classifier trained from scratch is not optimal. The reason is that, firstly, the global (full image) and local (patch) classifiers are independent, there is no interaction or information exchange between them; secondly, the calculation is the sum of both the full image and the local patch classifier, so it is redundant. The best solution is to split the functions between full image classifiers and local focus classifiers.

It is proposed to train the patch classifier directly from the full image CNN, that is, to extract the patch features from the full image CNN.

Unfortunately, this approach is rather complicated in terms of computation and is not reliable for searching for a fire in a forest. The complexity of this approach is due to the use of only one UAV to solve the task and its insufficient technical equipment.

As mentioned above, for the detection and recognition of a forest fire using a UAV, it is necessary to analyze dynamic visual scenes, so let's consider the formulation of this problem.

Let be $K = \{k_n\}$ are multiple frames (dynamic visual scene) $n = 1, \ldots, N$.

$k_n : O_{k_n} = \left\{o_{m_n}^{(k_n)}\right\}$ are set of objects present on the frame k_n, $m_n = 1, \ldots, M_n$, $n = 1, \ldots, N$.

$TR = \{tr_z\}$ are many tracks $z = 1, ..., Z$.

Each of the tracks tr_z represented by a sequential display of a single object $o_z^{(k_i)}$ on a sequence of frames k_i:

$$tr_z : O_{k_n} = \left\{o_z^{(k_i)}\right\},\ i = i_{\text{in}},\ i_{\text{in}+1}, \ldots, i_{\text{out}} \in K,\ i_{\text{in}} \le i_{\text{out}},$$

moreover, the same object is not included in different tracks.

The task of dynamic visual scene analysis is to build object tracks on the input sequence of frames. Thus, the problem of tracking a set of objects on a set of frames is solved.

Classification of multiple object tracking models (MOT, Multiple Object Tracking) can be performed according to three criteria [13]: how the model is initialized, how the data is processed in the model, what is the output. These classification features are shown in Fig. 2.

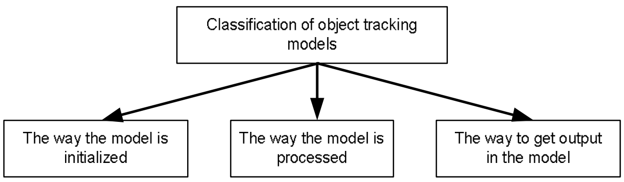

Fig. 2. Classification features of object tracking models

According to the method of initialization of the model, there are: tracking based on detection (Detection-Based Tracking, tracking-by-detection) and tracking without detection (Detection-Free Tracking).

In our task of detecting a forest fire using an UAV, tracking based on detection is chosen as the model initialization method, when objects are first detected using the detection method and then linked into tracks (association).

According to the method of data processing in the model, online and offline methods are distinguished. In our task, we will use online methods [14] – [16], which use only information from previous frames and the frame that is currently being processed for processing.

Such methods are the basis for real-time systems, where it is important to respond immediately after a new frame enters the system.

According to the method of obtaining the output, probabilistic and deterministic models are distinguished. We will use probabilistic models that use an approach based on the concept of a state space.

As detection methods, methods for recognizing objects in static visual scenes are considered.

IV. METHODS FOR DETECTING OBJECTS IN STATIC VISUAL SCENES

Neural network architectures for object detection and recognition can be divided into two large groups:

1) Architectures that process regions in an image (R-CNN).

2) Architectures that process the incoming image as a whole (YOLO, SSD).

A. YOLO (You Only Look Once) Architecture

The works [17] – [19] present the architecture of the neural network object detector called YOLO and its modifications. The YOLO architecture was originally developed for real-time tasks. In the YOLO algorithm, an image is divided into cells using a grid. For each grid cell, the probability of the presence of an object in general is estimated, then several most probable positions of the object are constructed in the form of rectangles centered in this cell, after which, for each resulting rectangle, the probability of the presence of objects of each considered class in it is evaluated.

In the YOLO method, the detection results are presented as a $7\times7\times1024$ tensor. The estimate of the probability of finding an object of a particular class in the current bounding box is the product of the estimate of the probability of finding an object in a cell and the estimate of the probability for a particular class.

In the case of YOLOv3 [19], a convolution neural network is used to extract features, which consists of 53 layers, 3x3 and 1x1 convolutions and Residual blocks are used as filters, which add values from the output of the previous layer to the output of the current layer.

It is also worth noting that in YOLOv3 object detection is performed at three scales, which made it possible to increase the quality of detection of small objects. The network scales the input image until it reaches the first detection level, at which point the filter step is 32. On subsequent convolution layers, the filter step is equal to 2. At each detection scale, the cell predicts three framing rectangles, that is, given the scale, each cell corresponds to 9 framing rectangles.

At the next step, the rectangles are filtered by the probability of finding objects in them. Then, just like in the solid-state drive (SSD) architecture, the rectangles are filtered using the false maximum suppression algorithm.

It is known that most recognition algorithms assume that output labels are mutually exclusive. In the YOLOv1 [17] and YOLOv2 [18] architectures, the softmax function [20] is used to convert estimates into class probabilities, the summation of which over all classes gives one. YOLOv3 uses a multi-label classification.

For example, the output labels could be "Pedestrian" and "Child", which are not mutually exclusive, and the sum of the outputs could be greater than "1". In YOLOv3, the softmax activation function is replaced with independent logistic classifiers to calculate the probability of an output belonging to a particular label. Instead of using the mean squared error when calculating the classification loss, YOLOv3 uses a binary cross-entropy loss function computed for each class. Using this technique also reduces the amount of computation required.

B. Faster R-CNN (Faster Region-based Convolution Neural Network) architecture

To solve the object detection problem Faster R-CNN [21] is currently one of the frequently used architectures based on deep learning. The predecessors of this architecture are R-CNN [22] and Fast R-CNN [23].

The work of R-CNN consists of three main stages.

1) The original image is divided into regions in which objects can be located. For this purpose, the Selective Search algorithm [24] is used, which generates 2000 different areas that are most likely to contain objects.

2) Each region is fed to the input of the corresponding trained convolution neural network, which extracts the feature vector for its region.

3) Feature vectors are fed to the input of a set of SVMs that perform the classification function. Each SVM is trained to define one class of objects. In addition, linear regression is used to refine the parameters of the rectangle enclosing the object.

An additional step can be considered non-maximum suppression (algorithm non-maximum suppression) to eliminate an excessive number of rectangles covering the same object.

The R-CNN architecture showed high rates of object detection accuracy, but such shortcomings as high memory and time costs for training and image processing were noted. Therefore, architecture modifications have been proposed, leading to the creation of Fast R-CNN.

1) The entire original image is fed to the input of one convolution neural network that performs feature extraction, and based on the full-size feature map, candidate regions are selected.

2) The SVM set that performs the classification function has been replaced with a softmax layer.

Thus, the convolution neural network is used once for the entire image instead of processing 2000 intersecting regions, it is also enough to train one network with a softmax layer without additional training of many SVMs.

In terms of speed, the Fast R-CNN method has a significant advantage over R-CNN, but another drawback was the algorithm for selecting candidate regions (Selective Search). Modification of this stage led to the creation of Faster R-CNN.

The Selective Search algorithm was replaced by a region proposal network (RPN). The input of this network is an $n \times n$ area taken from a full-sized feature map, the result is transmitted to two fully connected layers: box-regression and box-classification. Candidate regions obtained using RPN are represented by the coordinates of the circumscribing rectangle and the probability of finding an object in this region, calculated using the softmax function.

The Faster R-CNN architecture currently achieves high object detection accuracy and is considered to be relatively fast. At the same time, the main idea of the original R-CNN architecture is preserved: the selection of regions in the image in which objects may be located, and the classification of the contents of these regions.

C. SSD Architecture (Single Shot MultiBox Detector)

The SSD architecture [25] provides a significant increase in processing speed compared to Faster R-CNN. While the latter performs candidate region selection and region classification in two separate steps, the SSD performs these steps simultaneously while processing the entire image. The operation of an SSD can be described as follows.

1) The original image goes through a series of convolutional layers, resulting in a set of feature maps for different scales (e.g. 19x19, 10x10, 5x5, etc.).

2) At each point of each feature map, a 3x3 convolution filter is applied to obtain a set of circumscribing boxes.

3) For each rectangle, the spatial displacement and the probability of finding the object are simultaneously evaluated.

4) In the learning process, the true rectangles describing the object are compared with the predicted ones to eliminate false detections.

Unlike R-CNN, where there is at least a minimal probability of finding an object in candidate regions, there is no region filtering step in SSD.

As a result, a much larger number of circumscribing boxes are generated at different scales compared to R-CNN, and most of them do not contain an object. In order to solve this problem, SSD firstly uses non-maximum suppression to combine rectangles that are similar to each other into one. Secondly, the technique of hard negative mining [26] is used, according to which only a part of negative examples is used at each iteration of training; in SSD, the ratio of the number of negative examples to positive ones is 3 to 1.

The selection of candidate regions and classification are performed simultaneously: for a given number of classes C, each circumscribing rectangle is associated with a $(4 + C)$-dimensional vector that contains four coordinates and probabilities for all classes. The last step uses the softmax function to classify objects.

As shown above, the main approaches for detecting forest fires are the use of satellite information and information received from UAVs. The use of satellite information makes it possible to develop a UAV route and thereby significantly increase the efficiency of its use. At the same time, the UAV is the main means of detecting a forest fire and the coordinates of its location. Bad weather conditions make it difficult to accurately localize the fire and require additional processing of the received video information. In this paper, the problem of eliminating "blind" noise is not considered; the solution to this problem is given in such works of the authors [27], [28].

V. PROPOSED METHODOLOGY

In this paper, in connection with the use of UAVs as the architecture of the neural network detector, the Yolov4-tiny architecture was chosen. The Yolov4-tiny method was developed based on the Yolov4 method in order to increase the speed of object detection. The object detection speed for Yolov4-tiny can reach 371fps when using 1080Ti GPU, with the precision required by the real application.

This greatly increases the possibility of deploying the object detection method in embedded systems or mobile devices located, for example, on board UAVs [29]. The YOLOv4 model is divided into two layers: a feature extraction layer and a processing layer. The feature extraction layer is a combination of DarkNet and ResNet similar to a functional pyramidal network that has a convolutional layer, a batch normalization layer, and a ReLU layer. The overfitting problem is eliminated with batch normalization.

As in the previous approach [11], the first priority in solving the problem when processing the image stream is the detection of a fire. At the hardware level, this is solved through the presence of IR sensors, a camera for image magnification and an onboard processor, and algorithmically through the use of spot detection methods (Blob detection) aimed at detecting areas of a digital image that differ in properties, such as brightness or color, compared to surrounding areas. Informally, a Blob is a region of an image in which some properties are constant or approximately constant. Given some property of interest, expressed as a function of position in the image, there are two main classes of spot detectors: (i) differential methods based on derivatives of a function with respect to position, and (ii) methods based on local extrema based on finding local maxima and minima functions. In this case, the second approach is used, when using high-precision IR sensors, the intensity of forest fire pixels is measured and the area (there are three different regions depending on the temperature of the fire) in which they fall is analyzed. The pixel brightness is converted into a graph, and local maxima are considered to be an area of high intensity [30].

Histogram-based pixel segmentation is performed using the Otsu method, which allows you to separate the pixels of two classes (“useful” and “background”), calculating such a threshold that the intraclass variance is minimal.

VI. RESULTS

The NASA Space Apps Challenge dataset with and without fire photographs was chosen to build, train, and validate the convolutional neural network and the Grad-CAM algorithm.

In general, the sample includes:

- 700 images of fires, some of which contain heavy smoke;
- 200 images of nature without fires.

The data will be divided into three different categories: training, validation and testing. The training data will be used to train a deep learning CNN model and its parameters will be fine-tuned using the validation data. Finally, data performance will be evaluated using test data

The accuracy graph of the network with localization is shown in Fig. 3. The network loss function graph with localization is shown in Fig. 4.

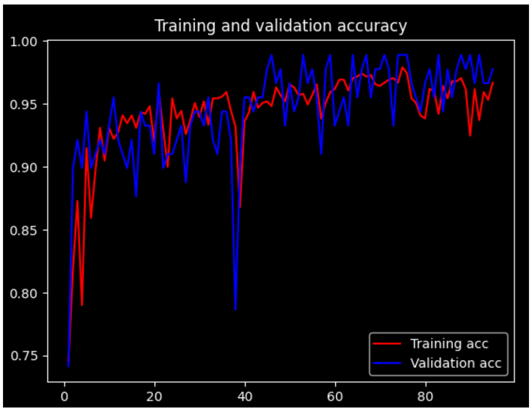

Fig. 3. Accuracy of the network

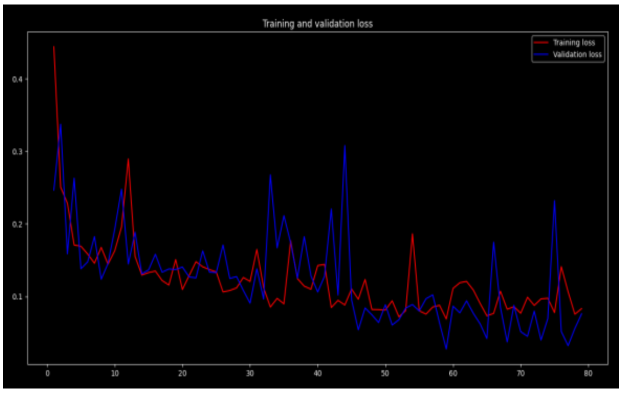

Fig. 4. Network loss function

It should be noted that the obtained results show a fairly high accuracy of the proposed approach. However, the final decision on its application can be made only after the experiment in real conditions.

VII. CONCLUSIONS

The problem of detecting forest fires is considered. The necessity of detecting fires using satellite information and information received from the UAV from infrared sensors and a video camera is shown. The results of information processing with the help of the YOLO neural network detector are presented.

REFERENCES

[1] H. O. H. U. M. Celik, and T. Demirel, "Fire detection in video sequences using statistical color model," in IEEE International Conference on Acoustics, Speech and Signal Processing, 2006.

[2] I. K. Martin Mueller, Peter Karasev and A. Tannenbaum,"Optical flow estimation for flame detection in videos," IEEE Trans. on Image Processing, vol. 22, no. 7, 2013. https://doi.org/10.1109/TIP.2013.2258353

[3] N. A. Che-Bin Liu, "Vision based fire detection," in Int. Conf. in Pattern Recognition, 2004. https://doi.org/10.1109/ICPR.2004.1333722

[4] P. Gomes, P. Santana, and J. Barata, "A vision-based approach to fire detection," International Journal of Advanced Robotic Systems, 2014. https://doi.org/10.5772/58821

[5] X. Z. Chunyu Yu, Zhibin Mei, "A real-time video fire flame and smoke detection algorithm," in Asia-Oceania Symposium on Fire Science and Technology, 2013.

[6] A. E. C. B. Ugur Toreyin, "Online detection of fire in video," in IEEE Conf. on Computer Vision and Pattern Recognition, 2007.

[7] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks." In Advances in Neural Information Processing Systems, Lake Tahoe, Nevada, USA, 2012.

[8] R. B. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in IEEE Conf. on Computer Vision and Pattern Recognition, Columbus, Ohio, USA, 2014. https://doi.org/10.1109/CVPR.2014.81

[9] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in IEEE Conf. on Computer Vision and Pattern Recognition, 2015. https://doi.org/10.1109/CVPR.2015.7298965

[10] Yusuf Hakan Habiboğlu, Osman Günay, and A. Enis Çetin, "Covariance matrix-based fire and flame detection method in video," *Machine Vision and Applications*, vol. 23, no. 6, pp. 1103–1113, 2012. https://doi.org/10.1007/s00138-011-0369-1

[11] Qingjie Zhang, Jiaolong Xu, Liang Xu and Haifeng Guo, "Deep Convolutional Neural Networks for Forest Fire Detection," *International Forum on Management, Education and Information Technology Application* (IFMEITA 2016), pp. 568–575. https://doi.org/10.2991/ifmeita-16.2016.105

[12] Michael Z. Zgurovsky, Viktor M. Sineglazov, Olena I. Chumachenko Artificial Intelligence Systems Based on Hybrid Neural Networks, Springer, 2020, https://link.springer.com/book/10.1007/978-3-030-48453-8. Customer can order it via https://www.springer.com/gp/book/9783030484521

[13] W. Luo and J. Xing, "Multiple Object Tracking: A Literature Review," arXiv.org [Electronic resource]. 2017. – URL: https://arxiv.org/abs/1409.7618 (date of the application: 14.03.2018).

[14] A. Bewley and Z. Ge, "Simple online and realtime tracking," arXiv.org [Electronic resource]. 2016. – URL: https://arxiv.org/abs/1602.00763 (date of the application: 14.03.2018)

[15] A. Sadeghian, A. Alahi, and S. Savarese, "Tracking the Untrackable: Learning To Track Multiple Cues with Long-Term Dependencies," arXiv.org [Electronic resource]. 2017. – URL: https://arxiv.org/abs/1701.01909 (date of the application: 24.02.2018). https://doi.org/10.1109/ICCV.2017.41

[16] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," arXiv.org[Electronic resource]. 2017. – URL: https://arxiv.org/abs/1703.07402 (date of the application: 24.02.2018).

[17] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 779–788. https://doi.org/10.1109/CVPR.2016.91

[18] J. Redmon and A. Farhadi, "YOLO9000: better, faster, stronger," arXiv preprint, arXiv:1612.08242, 2016, 9 p. https://doi.org/10.1109/CVPR.2017.690

[19] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," *Tech report*, arXiv:1804.02767, 2018, 6 p.

[20] C. M. Bishop, "Pattern Recognition and Machine Learning," Springer-Verlag, New York, 2006, 738 p.

[21] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *Extended tech report*, arXiv:1506.01497. 2016, 14 p.

[22] R. B. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014, 21 p. https://doi.org/10.1109/CVPR.2014.81

[23] R. Girshick, "Fast R-CNN," *IEEE International Conference on Computer Vision (ICCV)*, 2015, 9 p. https://doi.org/10.1109/ICCV.2015.169

[24] J. R. R. Uijlings, K. E. A van de Sande, T. Gevers, and A. W. M. Smeulders, "Selective Search for Object Recognition," *International Journal of Computer Vision*, vol. 104, pp. 154–171, 2013. https://doi.org/10.1007/s11263-013-0620-5

[25] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, Ch.-Y. Fu, and A. C. Berg, "SSD: Single Shot MultiBox Detector," "European Conference on Computer Vision (ECCV)," Springer, Cham., 2016, vol. 9905, pp. 21–37. https://doi.org/10.1007/978-3-319-46448-0_2

[26] S. Wan, Z. Chen, T. Zhang, B. Zhang, and K. Wong, "Bootstrapping Face Detection with Hard Negative Examples," arXiv:1608.02236, 2016, 7 p.

[27] V. M. Sineglazov, M. G. Lutsky, and V. S. Ishchenko, "Suppression of noise in visual navigation systems," *Proceedings of 2021 IEEE 6th International Conference on Actual Problems of Unmanned Aerial Vehicles Development (APUAVD)*, Kyiv, Ukraine, October 19-21, 2021, pp. 7–10.

[28] V. M. Sineglazov, "Hybrid Neural Networks for Noise Reductions of Integrated Navigation Complexes," Artificial Intelligence, no. 1, 2022, pp. 168–180. https://doi.org/10.15407/jai2022.01.288.

[29] Zicong Jiang, Liquan Zhao, Shuaiyang Li and Yanfei Jia, "Real-time object detection method for embedded devices," 2021, 11 p.

[30] Z. Jiao, Y. Zhang, X. Xin et al., "A deep learning based forest fire detection approach using UAV and YOLOv3," *in Proceedings of the 2019 1st International Conference on Industrial Artificial Intelligence (IAI)*, pp. 1–5, Shenyang, China, July, 2019. https://doi.org/10.1109/ICIAI.2019.8850815

Received October 24, 2022

Sineglazov Victor. ORCID 0000-0002-3297-9060.
Doctor of Engineering Science. Professor. Head of the Department.
Aviation Computer-Integrated Complexes Department, Faculty of Air Navigation Electronics and Telecommunications, National Aviation University, Kyiv, Ukraine.
Education: Kyiv Polytechnic Institute, Kyiv, Ukraine, (1973).
Research area: Air Navigation, Air Traffic Control, Identification of Complex Systems, Wind/Solar power plant, artificial intelligence.
Publications: more than 700 papers.
E-mail: svm@nau.edu.ua

Komapov Anatoliy. Bachelor.
Aviation Computer-Integrated Complexes Department, Faculty of Air Navigation, Electronics and Telecommunications, National Aviation University, Kyiv, Ukraine.
Education: National Aviation University, Kyiv, Ukraine, (2022).
Research interests: artificial neural networks, programming.
E-mail: nazarkomarov0410@gmail.com

В. М. Синєглазов, А. А. Комаров. Інтелектуальна бортова система пошуку лісових пожеж

У роботі аналізується ситуація з лісовими пожежами в Україні. Показано, що з кожним роком ситуація погіршується. Для моніторингу лісових пожеж обґрунтовано необхідність комплексного використання даних із супутників та безпілотних літальних апаратів. Доведено, що раннє виявлення пожежі, перш ніж вона стане лихом, має вирішальне значення для запобігання катастрофічним пожежам і порятунку життів і майна. Обґрунтовано підхід до виявлення пожеж на основі використання методів комп'ютерного зору, які можуть працювати з нестаціонарною камерою, встановленою на борту безпілотного літального апарату. Запропоновано підхід для виявлення «плями» пожежі за допомогою згорткових нейронних мереж. У нашій задачі виявлення лісової пожежі безпілотним літальним апаратом в якості методу ініціалізації моделі обрано відстеження на основі виявлення, коли об'єкти спочатку виявляються методом виявлення, а потім зв'язуються в треки (асоціації). В якості архітектури нейромережевого детектора обрано архітектуру Yolov4-tiny, яка забезпечує високу точність і швидкість бінарної класифікації.

Ключові слова: виявлення пожежі; згорткові нейронні мережі; безпілотні літальні апарати; YOLO; R-CNN; одиночний детектор MultiBox; класифікатор.

Синєглазов Віктор Михайлович. ORCID 0000-0002-3297-9060.
Доктор технічних наук. Професор. Завідувач кафедрою.
Кафедра авіаційних комп'ютерно-інтегрованих комплексів, Факультет аеронавігації електроніки і телекомунікацій, Національний авіаційний університет, Київ, Україна.
Освіта: Київський політехнічний інститут, Київ, Україна, (1973).
Напрям наукової діяльності: аеронавігація, управління повітряним рухом, ідентифікація складних систем, вітроенергетичні установки, штучний інтелект.
Кількість публікацій: більше 700 наукових робіт.
E-mail: svm@nau.edu.ua

Комаров Анатолій Анатолійович. Бакалавр.
Кафедра авіаційних комп'ютерно-інтегрованих комплексів, Факультет аеронавігації, електроніки та телекомунікацій, Національний авіаційний університет, Київ, Україна.
Освіта: Національний авіаційний університет, Київ, Україна, (2022).
Напрям наукової діяльності: штучні нейронні мережі, програмування.
E-mail: nazarkomarov0410@gmail.com