Neural networks application to detect the facts of smoking in video surveillance systems

P V Danilchenko and N S Romanov
Perm National Research Polytechnic University, Department of Applied Mathematics and Mechanics, 614000, Russia, Perm, Komsomolsky prospect 29
E-mail: pavel.danilchenko96@gmail.com

Abstract. The article is devoted to the application of convolutional neural networks and cascade classifiers for tracking smoking facts in relation to video surveillance systems. Two methods of smoking detection are proposed. The first method involves detecting a cigarette against the background of a human face. The second method involves detecting a hand with a cigarette in it. On the basis of the proposed methods, software for a video surveillance system is being developed, which makes it possible to determine the facts of smoking in real time, and thereby prematurely prevent possible undesirable consequences.

1. Introduction
One of the fastest growing areas of software development is the field of computer vision, and the rapid development of existing hardware only spurs the creation of more and more complex methods for the image analysis. Computer vision methods are widely adopted in modern video surveillance systems. They are used to classify and determine the position of objects, track their movement, detection of crossing borders, detection and identification of faces, search and recognition of license plates or some marks, etc. At the moment, when solving such problems, developers apply deep learning neural networks and convolutional neural networks in particular [1-3]. In addition, a computer vision can be used to improve the safety of residential buildings, shopping centers, warehouses, industrial enterprises, etc. by detecting a possible fire or smoke from the video image [4]. Usually timely localization and elimination of a fire with its early detection can manage small losses in material and financial terms, without casualties and injuries, but there are situations when only a potential source of a possible ignition (for instance, a spark or a smoldering cigarette) poses a threat to the existence of the object and the lives of the people there. For example, a spark from a smoldering cigarette in a warehouse with explosive fertilizers, at a natural gas liquefaction plant, or at oil and gas production enterprises can lead to a powerful explosion, death of people, destruction of both the enterprise itself and the nearby infrastructure and buildings.

There are facts when despite the risks employees still decide to smoke on the territory of the facility. Sometimes such cases come into view of video cameras, but go unnoticed. Due to the ineffectiveness of smoke detectors and other existing solutions in open space and in ventilated areas, there is a need to automatically detect the facts of smoking on the video stream. This would help to minimize the number of such incidents and to bring safety violations to justice. In addition, the fact that there are such detectors as a part of a video surveillance system would further motivate employees to comply with safety regulations.
2. Solution
Existing solutions based on one neural network that detect the facts of smoking have very low accuracy and low response rate [5]. The fact of smoking can be detected both by the presence of a cigarette and by the exhaled smoke, but in the second case, under some environmental conditions, for example, at low temperatures, exhaled humid air can be taken as smoke, therefore, the detection of exhaled smoke is not considered. Thus, the fact of smoking can be noticed only when looking for a cigarette in the frame. To solve this problem, it was decided to divide it into two steps: the search for persons with a cigarette in their mouths and the search for hands holding a cigarette. The search for faces with cigarettes is implemented in 2 stages: first, the cascade classifier finds faces in the image, and then the found parts of the image with faces are classified by a neural network.

2.1. Faces with cigarettes
To localize faces in images, it was decided to use cascade classifiers, since they are relatively lightweight and fast, compared to even simple convolutional neural networks that could be used for the same purpose. The development of an intelligent video surveillance system is carried out by using the free computer vision library OpenCV, therefore the adoption of this decision was largely influenced by the availability of free cascading classifiers for detecting faces in a profile and a full face, which can be launched by using OpenCV. In addition, these cascading classifiers facilitate the formation of training samples.

After the face is localized by the cascade classifier, the parts of the images containing the faces are transferred to the input of the convolutional neural network to determine their belonging to one of three classes: false triggering of the cascade classifier, a face without a cigarette, and a face with a cigarette. A significant difference between faces in profile and full face implies the presence of not only two cascading classifiers, but also two different neural networks that should be trained on the corresponding training samples, as shown in figures 1 and 2. For these networks, several variants of structures were tested, and for each of the networks, we selected those that showed the best training results. For example, figure 3 shows the structure of the neural network for faces in profile.

![Figure 1](image1.png)

**Figure 1.** Examples of labelled images of the training set for detecting a cigarette in the mouth in profile: a) images from false triggering of the cascade classifier, b) face in profile, c) face in profile with a cigarette.

Neural networks were trained by using Caffe. The training and test samples for faces in profile included 6000 and 600 images, respectively, and 5000 and 1000 images for full-face faces. As a result of training networks, a classification accuracy of 90% was achieved for faces in profile on a test sample. The neural network sometimes incorrectly determined the class with false positives of the cascade classifier. To eliminate this problem, it was decided to add to the training set the most similar images that were determined incorrectly. Ultimately, it was possible to increase the accuracy of the neural network to 92% and to reduce the number of false positives of the network when the cascade classifier is falsely triggered by a factor of 5. The accuracy of the full face classification was 92.5%.
2.2. Hands with cigarettes

Smokers do not constantly hold a cigarette in a mouth, they hold a cigarette in hands most of the time, therefore, to increase the probability of detecting the facts of smoking, it makes sense to implement a detector of cigarette in hand. However, locating the cigarette in hand is more difficult. This is due to the fact that faces are usually visible in the correct, upright position, and only two existing cascade classifiers are enough, but the hand, unlike the head, is asymmetrical, can be observed from different sides and at different angles while holding a cigarette in hand in several ways. The existing solutions for hand detection and gesture recognition do not suit this task due to their excessive complexity, and the creation of a separate cascade classifier for each case is not an acceptable solution, therefore, it becomes necessary to create a neural network that would allow to localize the hands. There also raises the question whether it is advisable to separate neural networks for localizing the hands and determining the presence of a cigarette or to combine them into one. The solution of this problem will be worked out in the future.
3. Conclusion
Thus, we have proposed a new approach that allows to detect the facts of smoking. We have also trained two convolutional neural networks with an accuracy of more than 90%. The proposed algorithms are implemented in the smoking detector of Domination intelligent surveillance system of Vipax company and the detection of faces with cigarettes has already been implemented in the system and runs on a test server.

References
[1] Goodfellow I, Bengio Y and Courville A 2016 Deep Learning (Cambridge: MIT Press)
[2] Chollet F 2018 Deep Learning with Python (New York: Manning Publications)
[3] Nikolenko S, Kadurin A and Arhangelskaya E 2018 Deep Learning (Saint Petersburg: Piter Publishing House)
[4] Gladkiy S L 2019 Experience in applying differential equations and neural networks for recognizing physical phenomena in video images Artificial intelligence in solving urgent social and economic problems of the XXI century 4 36-43
[5] Complex for smoking detection by photo or video based on Intel NUC URL: https://habr.com/ru/company/intel/blog/493726/