DOI: 10.46299/ISG.2025.MONO.TECH.2.5.3

5.3 Neural networks as a necessary component for modern methods of searching for anomalies in video sequences

To control various processes and objects in order to prevent a wide range of risks and dangers, video surveillance systems are widely used, i.e. complexes of technical equipment and software designed to monitor the course and behavior of these processes and objects. This is a slightly supplemented formulation of the essence of the video surveillance process from work [174].

Such processes and objects can be, for example, cars moving along a street or a country road, road surface during control of its condition and quality, a security system for any object, including the “Smart Home” system, and much more.

Anything that differs from the usual course or behavior of the observed processes or objects is an anomaly in the video sequence of surveillance cameras. An anomaly in the video sequence of surveillance cameras is usually a sign of violations or danger in the observed processes or objects. But, for example, in video surveillance in industry during technological processes, an anomaly can be evidence of violations or safety precautions, or a certain technology, that is, be useful information. Search and registration or detection of anomalies is the main goal of video surveillance.

A review of modern systems for detecting anomalies in video footage of video surveillance cameras has been conducted.

An analysis of methods for detecting anomalies in video footage of video surveillance cameras has been conducted. For this purpose, a classification of traditional modern methods for detecting anomalies in video footage has been developed and the foundations of the theory of neural networks and convolutional neural networks have been considered from the point of view of their application for classification, localization, segmentation, detection, identification and tracking of objects in video footage of surveillance cameras.

The formation of an input data set for conducting experiments on detecting anomalies in the video sequence of surveillance cameras is considered and a methodology for comparative experiments on searching for anomalies in the video sequence of surveillance cameras is developed.

An anomaly of a time series in general, and the video sequence belongs to such series, is, as this very word indicates, something that differs from the norm, that is, an outlier or deviation. The search for anomalies is a scientific direction of enormous scope, which concerns many branches of science and technology.

As for anomalies in the video sequence of surveillance cameras, they can belong to useful information, that is, be evidence of some violations in the normal course of the observed processes. In this case, the registration or detection of anomalies is the main goal of video surveillance. Anomalies in the video series of surveillance cameras can also be caused, for example, by equipment malfunctions or interference in the signal transmission path from video cameras to recording and analytical equipment. Of course, such anomalies are not informative, and the causes of their appearance must be eliminated.

Recently, many different methods of detecting anomalies in time series in general, and in video series of video surveillance cameras in particular, have appeared in the world information space. Many articles devoted to research in this direction have appeared, mostly in English. Although among their authors there are many researchers from Asian countries. These articles are mostly very vague, many are overviews. It is very difficult to obtain really useful practical information from them. Apparently, software developers for intelligent IP video cameras and video information processing systems carefully protect their developments.

On the other hand, some general-purpose software companies, such as Microsoft, publish computational schemes for processing video data, test suites, machine learning programs, etc. In general, these are huge amounts of computational and video information. Researchers and developers of modern video surveillance systems find it difficult to navigate this variety of methods for detecting anomalies.

During video surveillance, information is transmitted from video cameras and television cameras to a certain number of monitors or recording devices. Video surveillance systems are used in places that require constant supervision, such as banks, ATMs, train stations, airports, military facilities, vehicles and transportation systems, private houses and apartments, shopping malls and regular stores, etc.

In industry, surveillance cameras can be used for centralized monitoring of the production process or in the presence of an environment that is dangerous to humans. Video surveillance systems can record continuously or be turned on only after a given event. More advanced surveillance systems using video recorders allow you to create recordings that will be stored for years, with different quality and with additional features (such as motion detection and email notifications) [175].

A video sequence is a product of the action of video cameras of a surveillance system and refers to time series, the study of which, including their anomalies, has been devoted to a lot of work. But the study of anomalies in video sequences themselves is not so much.

Usually, a video sequence is considered as a sequence of frames, and significant changes from one frame to another may indicate the emergence of new situations, and this is considered as a way out of the state of stability, that is, as an anomaly.

We have analyzed a considerable number of scientific works that consider problems related to the study of video sequences. These are not only video cameras used for video surveillance in security systems, but also those video cameras used in many other areas of human activity. These are defense, industry, medical research, agriculture, various scientific research, including space, and many others.

The large volume of video data that has been produced recently and continues to be produced continuously requires automated processing using powerful computers, specialized software, and the latest intelligent information technologies. Currently, the role of humans in analyzing video sequences from the point of view of searching for anomalies is practically minimized. This is explained by both the large volume of video data that requires analysis and the low capabilities of humans compared to computing equipment (speed of operation, vision features, etc.).

The system for detecting anomalies in video sequences or a video surveillance system functions as follows. The video sequence, that is, the useful signal from the video camera, which is the result of video surveillance, enters the video recorder or computer for registration, further processing, and storage.

A video recorder is an electronic device designed to record, store, and play video information [176]. It is similar in structure to a computer or video server and contains an analog-to-digital converter (ADC), processor, hard drive and other components. A specialized operating system is installed on it to control the video recorder. Before recording, digitized video images are usually compressed to reduce the space occupied on the hard drive. Almost all video recorders can work with both monochrome and color video images. Many video recorders have the ability to connect to a computer network to transmit video images to computers of remote users.

In the case of using a video recorder, video processing is carried out in the video processing unit. This process of processing video sequences in order to detect threats or anomalies in conjunction with appropriate software is called video analytics and can be implemented both by separate devices and by computer or video recorder software. The results of video analytics or the video itself can be transmitted to the Internet using an Internet connection unit.

Now autonomous controlled robotic rotating Internet Protocol (IP) video surveillance cameras have appeared on the market, which allow remote or automatic rotation of the lenses horizontally and vertically to increase the volume of the monitored space, for example, Hikvision DS-2DF8C448I5XS-AELW (Fig. 1) [177].

Such a video camera is controlled in an application for smartphones and tablets or in the software for the video surveillance system itself.

Figure 1. Hikvision robotic pan/tilt IP video camera DS-2DF8C448I5XS-AELW with video analytics and Internet connectivity [177]

Such devices are wireless and can transmit video online via the Internet in real time, as well as record an archive to the cloud or to a memory card with a capacity of up to 256 GB. This is a modern self-sufficient tool for video surveillance, which has video analytics software and does not require additional equipment. Surveillance cameras of this type, manufactured by well-known companies, cost 100 thousand hryvnias or more per unit. Therefore, such cameras are not yet very common, but practice proves that in the future such devices will become cheaper and will be widely used in video surveillance systems.

For a robotic IP video camera with video analytics, the functional block diagram of the anomaly detection system in the video sequence looks like this (Fig. 2):

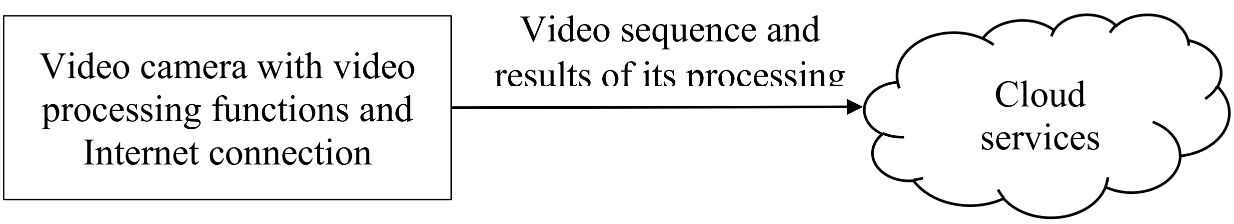

Figure 2. Diagram of the anomaly detection system in the video sequence with combined functional blocks

Such a video surveillance device can independently store a certain amount of video information, can transfer it for storage using cloud technologies. This video camera can also process the received video information using certain algorithms independently (within its capabilities), and can transfer it for cloud computing if more powerful computing resources are required.

The above-mentioned state-of-the-art systems for detecting anomalies in video sequences are still a thing of the future. Therefore, the task of classifying a somewhat simpler traditional technique, which is now predominantly used in anomaly detection systems in video sequences at the present time, continues to be relevant.

It is customary to distinguish several types of classic video surveillance systems depending on the technical equipment used [178, 179].

1) Analog systems. Components: AHD/HD-CVI/HD-TVI video surveillance camera, Digital Video Recorder (DVR).

2) Digital systems. They are based on IP technology. The abbreviation IP (Internet Protocol) indicates a collection of rules by which Internet technology must be used to exchange information between computer networks. This refers to the use of IP video cameras in conjunction with network switches. The composition of components: IP video surveillance camera, Network Video Recorder (NVR) and a network switch for connecting video cameras and video recorders to the Internet.

3) Mixed (hybrid) systems. The principle of their operation is based on receiving video images from analog cameras and digitizing the image for further use.

Analog systems are not very modern, but sometimes the use of this type of system is justified [178].

Digital systems today are more promising compared to analog systems: improved video image quality, flexibility and easy scalability, the possibility of deep analysis and remote configuration, ease of integration into an existing computer network.

Hybrid systems are so far the most common. In them, the quality of the received image is inferior to systems with a digital format, but other parameters of the operation of systems of this type do not differ from similar characteristics of digital systems, and this makes it possible to create reliable and multifunctional systems of a mixed type.

It should be noted that modern video surveillance systems necessarily use the Internet and for this purpose have in their composition either video cameras and video recorders for receiving, recording and saving video images, as well as a router and a control unit for all video cameras, or one network IP video camera for small areas or local tasks. To transmit video information simultaneously from several points, a video surveillance system can consist of several network video cameras.

Let us try to clearly present the classification of anomaly detection systems in the video sequence, based on the above-mentioned classes of these systems (Fig. 3) [180].

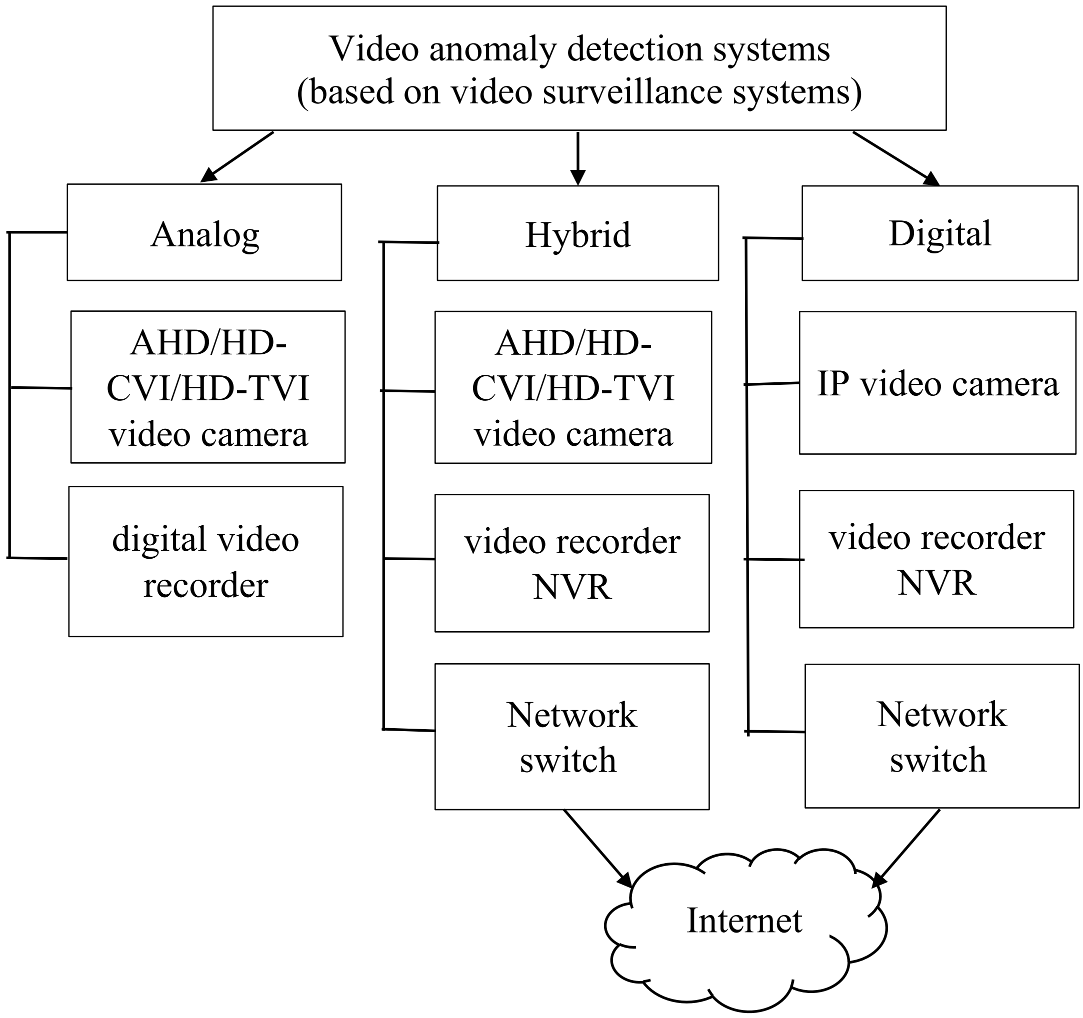

Figure 3. Classification scheme of anomaly detection systems in surveillance camera video sequences

The above analysis shows that for modern video cameras, the review of anomaly detection systems in video sequences is identical to the review of methods for detecting anomalies in video sequences of surveillance cameras. All of the above schemes are functionally the same, but with the continuous development of optics, electronics and computing technology, the hardware composition of anomaly detection systems in video sequences of video surveillance cameras is changing somewhat.

For a better understanding, we can say that now the computing and communication (network) capabilities of IP video cameras are better than those of powerful computers of the 2000s. Those blocks in such systems that required separate housings and many cables for the operation of video surveillance systems have turned into small microprocessor boards that are easily mounted in IP video camera housings. But new methods for processing video capture results (and these are mostly methods associated with the use of neural networks) require significantly greater computing power. This is where the network capabilities of IP video cameras and cloud services come in handy, with the help of which huge amounts of recorded data are stored and processed using new network technologies.

Above, we talked about the fact that with the development of microprocessor technology and cloud technologies for storing and processing information, it is difficult to separate systems for detecting anomalies in video sequences from video surveillance systems. We believe that with the further development of radio electronics, information technologies, computing power of modern microprocessors and mathematical modeling, each video surveillance system will simultaneously be a system for detecting anomalies in video sequences.

As can be seen from Fig. 2, modern intelligent video cameras such as the Hikvision video camera themselves are a system for detecting anomalies in video sequences, and the differences between one system for detecting anomalies in video sequences from another are the differences between the methods for detecting such anomalies.

It can be summarized that intelligent video cameras themselves become anomaly detection systems in video sequences due to the following factors:

- the presence of a module in the video camera for primary processing of video surveillance results;

- the presence of a module in the video camera for Internet communication, i.e. the ability to use cloud technologies and services for storing and processing information.

There are two types of tasks related to the search for anomalies [179] in video sequences.

1) Detection of outliers, which are defined as observations that are significantly different from others. Algorithms for determining such outliers try to find in which section of the video sequence the bulk of the data is located. Anomaly observations interfere with the use of this data and are removed before machine learning. The procedure for detecting such anomalies is called Outlier detection.

2) Detection of anomalous behavior in a situation where there are observations that describe normal states of the system, and when new data appears, it is necessary to determine whether they are anomalous. This approach is called Novelty detection. It is most suitable for analyzing a video sequence and determining whether there are anomalies in a new frame compared to the previous ones.

One of the simplest approaches used in video surveillance is video detection. A video detector compares the next frame pixel by pixel with the previous one or with a group of frames and signals a change in static images. At the same time, it is determined in advance whether to take into account more or less changes in color depth, or to exclude reactions to certain zones in the frame, and the area of changes is determined. The most significant disadvantage of video detection is low noise immunity. If, for example, there are branches swaying in the wind in the video detection zone, then the accuracy is significantly reduced. And this requires continuous manual control in real time or long-term viewing of video archives to track important events. Such noise in a fixed video sequence is a big problem in the video surveillance process.

That is, to the problem of the huge amount of video information that is constantly accumulating in the process of video surveillance, there is added the problem of analyzing this video information. In addition, this analysis needs to be carried out mostly in real time, that is, during observations.

We have come to the point where it is necessary to introduce a new concept for this study. This is video analytics or video content analysis (Video content analysis or video content analytics (VCA)) - the ability to automatically analyze a video sequence to detect and determine temporal and spatial events [181].

Video analytics does not compare static images and pixels as such, but changes in the nature of activity. It works with dynamics, finds a new movement against the background of other movements. When an object enters the frame or starts to move, its nature of activity is analyzed and remembered, which becomes a sign of the identity of this object. The reaction occurs when this pattern of movement changes or when a different nature of movement appears in the frame, and even their combinations when superimposed. One of the most urgent tasks in video surveillance systems is to reduce the amount of useless information, remove unnecessary data, unlike other approaches to video analytics, when information necessary for a person is recognized.

This goal is served by the so-called video semantics - "a brief logical presentation of video information by decomposing it into semantic units (video plots), each of which has its own complete content, which differs from both the previous and the next video segment" [182]. Video semantics does not work every time the twig shakes in the last example, but only once, when the weather changes and becomes windy. Subsequent vibrations will not be registered as unimportant information. At the same time, the “short data” technology shortens everything, including the obstacles themselves. The results panel will report a twig shaking, but once during windy weather, which is thousands of times less than several hours of continuous video recording tied to standard video detection.

One of the first video surveillance systems that worked on the “short data” technology was an automated surveillance technology called Annotation of Critical Evidence (ACE) Surveillance [182]. It is engaged in the extraction and processing of annotated critical sections of video surveillance. ACE technology consists of the ACE Capture module and the ACE Browser module. ACE Capture works as follows: the video signal from the camera is constantly processed on a computer. When a new object or activity is detected, the software triggers an alarm. Critical snapshots of the data are automatically captured and stored. Each such snapshot is provided with text annotations (explanatory captions) that help to understand the data and increase the manageability of archived data.

Annotations to the snapshots include: the location and size of the moving object, the direction from which the object appeared and its speed. The ACE browser is responsible for the preparation, efficient viewing and search of archived data. This system structure is intended to simplify the detection of anomalous events in the huge amount of stored video surveillance data.

Another author [183] gives a different definition of this new scientific and technical direction. Video analytics is a hardware-software or technology that uses computer vision methods for automated data collection based on the analysis of streaming video (video analysis), which relies on image processing and pattern recognition algorithms that allow video analysis without direct human participation.

In principle, these two definitions do not contradict each other. Both state the automatic nature of processing and analyzing video information using computer technology. Recently, the use of video analytics in video surveillance systems has been given great attention.

Let's try to understand how video analytics relates to the search for anomalies in the video sequence of surveillance cameras. In [183], the basic functions of analytics are given, on which all other functions of its work are built.

1) Object detection. Detection of objects in the field of view of the video camera is carried out using video motion detectors, and analysis of several objects is also possible. If movement is not a sufficient sign for localization of an object in the frame, then sometimes templates are used, including Haar features, which are known to us from the problems of computer vision.

2) Object tracking. Tracking algorithms allow you to get the trajectory of an object's movement both in the field of view of one camera and according to the results

of tracking by several cameras. For example, this is how cars moving at high speed are identified.

3) Object classification. Video analytics systems classify objects. For example, a typical object classifier, using shape features and absolute dimensions, divides objects into groups: a person, groups of people, and a vehicle.

4) Object identification. Object identification is the most complex component of video analytics systems. Modern video surveillance systems equipped with a video analytics unit allow you to identify people by biometric features of a person, or vehicles by license plates.

5) Detection or recognition of situations. By following certain algorithms and using the same operations as in points 1-4, video analytics software allows not only to highlight objects from streaming video, but also to recognize alarming situations based on the analysis of the behavior of these objects. Such situational video analytics can automatically detect, for example, crossing a signal line, falling people, prohibited parking, fires, etc.

The above basic video analytics functions for video surveillance systems require various methods for their implementation. In recent years, quite a few such methods have appeared. In the next section, we will try to classify them and analyze the features of their application.

The last point of the above basic video analytics functions almost completely corresponds to the tasks facing systems for detecting anomalies in the video sequence of surveillance cameras. Although it all depends on the purpose of creating or purchasing each specific video surveillance system. Its tasks are discussed with the client or customer of such a system. It is necessary to choose such a method of detecting anomalies in the video sequence of video surveillance cameras or a set of methods so that he or they are able to solve the security tasks set.

Now, having information about the use of software or hardware video analytics blocks in anomaly detection systems in video sequences of video surveillance cameras, we present a modern functional block diagram of an intelligent anomaly detection system in video sequences (Fig. 4) [180].

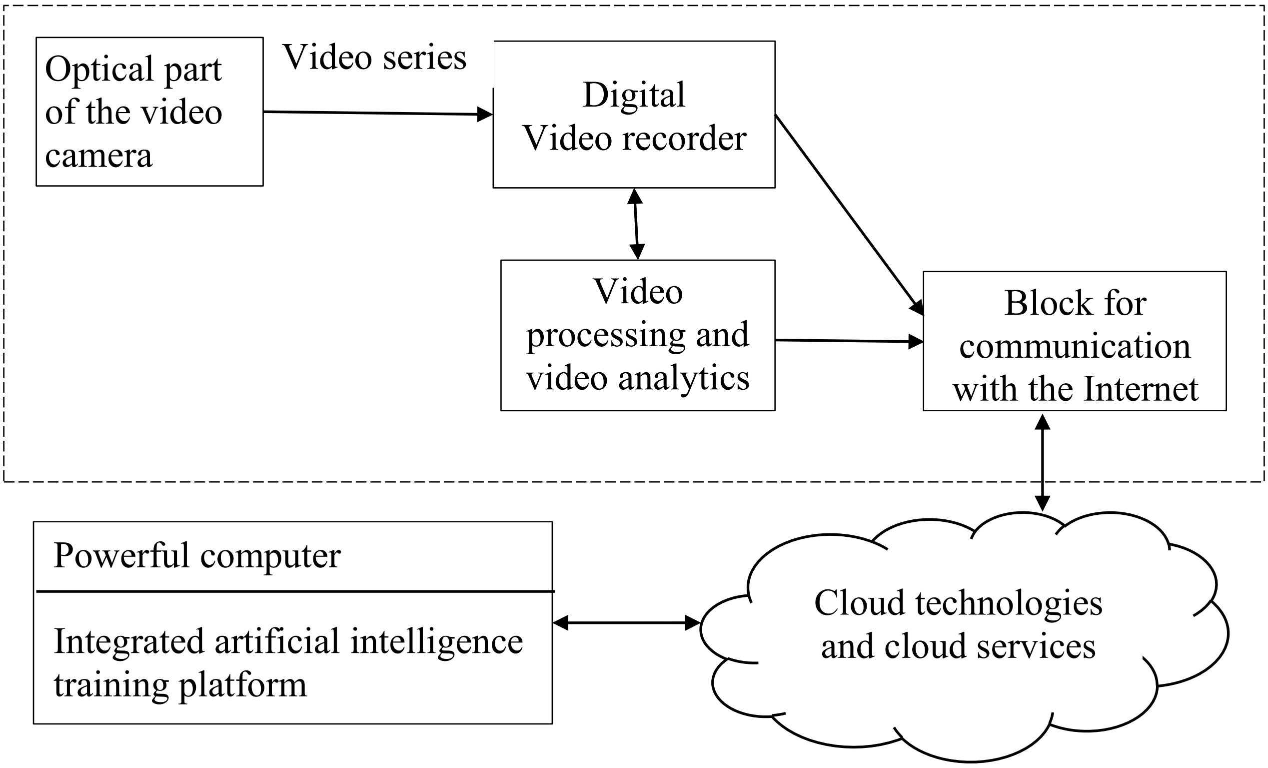

Figure 4. Proposed functional block diagram of an intelligent video anomaly detection system

The functional block diagram of the intelligent system for detecting anomalies in the video sequence of video surveillance cameras, which we have proposed, is shown in Fig. 4, and is universal in nature. It shows the composition of the anomaly detection system, which is based on a conventional video camera. In this case, the signal from the video camera, that is, the video sequence, is transmitted to the video recorder for storage and further operations. After video recording, the signal from the video camera is transmitted to the processing unit or, if necessary, to the Internet for deeper processing using the most modern technologies (neural computing networks and others), for further analysis or storage. Intelligence in this version is provided by modern algorithms and powerful resources of the cloud environment. In the case of using a more expensive IP video camera, the blocks highlighted by dash-dotted lines are present in the camera itself. Depending on the tasks of the system and the availability of certain equipment, in each specific case there may be a different composition of the anomaly detection system in the video sequence.

Modern intelligent video surveillance cameras can also have built-in video analytics functions. The advantage here is that such analytics capabilities in video cameras do not depend on network bandwidth and server response time. Such a solution is beneficial where high efficiency and immediate response are required. Video data in this case is stored on the video cameras themselves and can be retrieved for analysis at any time via a remote client program. In the case where an intelligent IP video camera is used for surveillance, it combines several blocks from Scheme 4 (dotted line). Primary processing can be performed by the camera itself (video processing and video analytics unit). If more powerful computing resources are required, for example, in the case of using neural technologies, cloud resources are used.

In any case, the anomaly detection system in the video sequence of video surveillance cameras shown in Fig. 4 belongs to the Intelligent Video System (IVS) systems. There are different types of intelligent video equipment in the field of IVS: from simple devices that record any movement at all, built into the lower tier of digital video cameras, to expensive advanced systems that classify objects and their behavior, compensating for false alarms and illusory actions. Some video surveillance and anomaly detection systems in the video sequence are intelligent due to the use of an intelligent video camera, and some - due to their deep learning in neural networks.

A universal functional block diagram of the anomaly detection system in the video sequence of video surveillance cameras, which belongs to intelligent video analytics systems, has been developed. Such a system is based on simple digital video cameras of the lower tier to expensive advanced systems that classify objects and their behavior, compensating for false alarms and illusory actions. A video surveillance system and anomaly detection in video footage can be intelligent through the use of expensive intelligent video cameras, or perhaps through its deep learning in neural networks and the use of cloud computing resources.

In general, there are many methods for detecting anomalies in video footage from video surveillance cameras, and their successful use requires appropriate analysis. Such an analysis will be conducted by us in the next section.

In modern systems for detecting anomalies in the video sequence of surveillance cameras, the presence of video analytics is almost mandatory. The existence and development of video analytics are inextricably linked to artificial intelligence.

The concept of artificial intelligence includes Mashing Learning (ML) algorithms [182]. Such algorithms automatically create a mathematical model of a process or object, using huge amounts of sample data to obtain the ability to calculate results without a specially created program. Machine learning algorithms are developed using an iterative process, in which the cycle of collecting training data, labeling it, using labeled data to train the algorithm, and testing the trained algorithm is repeated until the desired level of model quality is achieved. After that, the algorithm is ready for use in an application for processing and analyzing data, which can be implemented on the surveillance object. Training is complete, since the application will not learn anything new. A typical task of AI-based video analytics is to detect and classify objects (people and vehicles) in a video stream. A machine learning algorithm learns a combination of visual characteristics that allows it to identify these objects.

A deep learning algorithm is more sophisticated and can detect much more complex objects. However, it also requires more effort to develop and train, and much more computing resources when used in a fully developed application. The development of algorithms and the expansion of the capabilities of data processing cameras have made it possible to use advanced AI-based video analytics directly in the camera, instead of performing all the calculations on the server. This means that the received data can be analyzed in real time, because applications can immediately use the compressed video material. Thanks to dedicated hardware accelerators, such as the Machine Learning Processing Unit (MLPU) and Deep Learning Processing Unit (DLPU), an embedded camera with video analytics requires less power than using a central or graphics processor.

Before installing an application for intelligent video analytics, you must carefully study and strictly follow the manufacturer's recommendations, based on known prerequisites and limitations. All video surveillance systems are unique, and the effectiveness of the application must be evaluated on a case-by-case basis. If the quality is lower than expected, you need to look for the cause in the complex, and not focus only on the analytical application. The performance of video analytics depends on many factors related to the equipment and configuration of the camera, the quality and dynamic characteristics of the video image, as well as the illumination in the frame. In most cases, knowledge of the impact of these factors and their appropriate optimization can improve the effectiveness of video analytics. As artificial intelligence is increasingly used in video surveillance, in addition to the benefits and new scenarios for its use, it is necessary to consider when and where this technology will be used. Artificial intelligence has been developed and discussed since the invention of the first computers. Although full-fledged artificial intelligence is still a long way off, technologies with its elements are already widely used to perform well-defined tasks in applications and systems such as voice recognition, search engines and virtual assistants. In addition, artificial intelligence is increasingly used in medicine, for example, in X-ray diagnostics and for analyzing the results of retinal scans.

AI-based video analytics is one of the hottest topics in the video surveillance industry, and it holds great promise. Applications have been developed that use AI algorithms to significantly speed up data analysis and automate repetitive tasks.

Machine learning is a subset of artificial intelligence that uses statistical learning algorithms to create systems that can automatically learn and improve without explicit programming.

The difference between traditional programming and machine learning is that computer vision, which allows computers to understand what is happening in a frame by analyzing photo or video data. Video analytics, like finding anomalies in video sequences, are related to computer vision, with some differences. Traditionally, programmed computer vision is based on methods for calculating image features, such as sharp edges and corners. These features must be manually specified by an algorithm developer who knows what to look for in the visual data. The developer then combines these features to create an algorithm that can infer what is detected in the frame.

Machine learning algorithms automatically build a mathematical model using a large amount of training data to be able to make decisions by calculating the results without special programming. The features can be specified, but how to interpret them is learned by the algorithm itself as it processes large amounts of labeled or annotated training data. This technique of using features specified in combinations that the algorithm learns itself is called classical machine learning.

In other words, in the case of machine learning, you have to teach a computer to use the desired program. The data is collected and then annotated by humans, sometimes with the help of a server computer. The result is fed into the system, and this process continues until the program learns enough to detect the desired object, for example, a specific type of car. The trained model becomes the program. Note: once the program is ready, the system will not have learned anything new.

Deep machine learning is an advanced version of machine learning, where the system learns to extract features and how to combine these features in deep rule structures to produce an output [184]. Deep learning is also based on input data. The algorithm automatically determines which features to look for in the training data and learns the deep structures of related feature combinations. The algorithms used in deep learning are based on how neurons work and how the brain uses them to form higher-level knowledge by combining the output neural signals in a deep hierarchy. The brain is a system in which combinations are also formed by neurons, erasing the difference between extracting and combining features, making these processes in some sense the same. Scientists have modeled these structures in so-called artificial neural networks, which are the most popular (and one would think necessary) type of deep learning algorithm. Using deep learning algorithms, it is possible to create sophisticated visual detectors and automatically train them to detect very complex objects that are robust to scaling, rotation, and other variations.

The reason for this flexibility is that deep learning systems are trained on a much larger amount of data (and are also characterized by greater diversity) than classical machine learning systems [184]. They often significantly outperform hand-crafted computer vision algorithms. This makes deep learning ideal for complex tasks where humans have difficulty formulating combinations of features (e.g., image classification, linguistic information processing, object or anomaly detection) (Fig. 5).

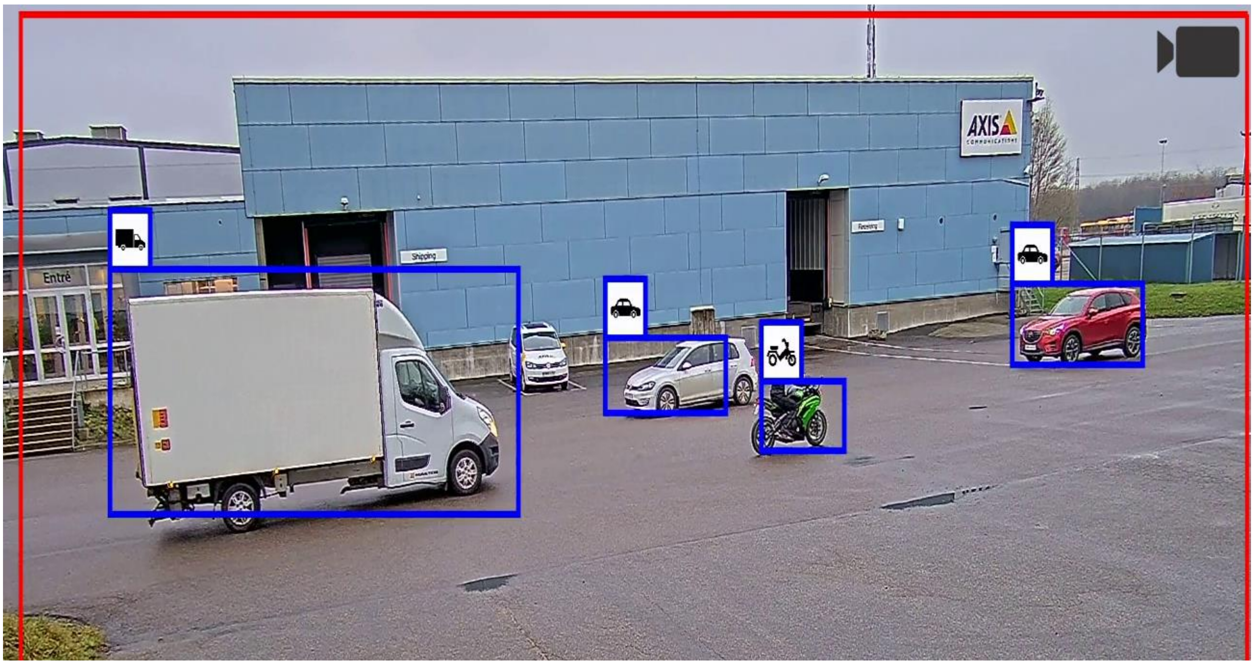

Figure 5. Object detection and classification by type based on deep learning [184]

After training, the model needs to be thoroughly tested. This phase usually includes an automated part and thorough testing in real-world deployment conditions. During the automated part, the program is tested using new data sets that were not used to train the model. If the results of such tests do not meet expectations, the process starts again: new training data is collected, instructions are created or refined, and the model is trained again. Once the desired level of quality is reached, field testing begins: the program is tested on real video sequences. The scope and variability depend on the scope of the program. The smaller it is, the fewer variations need to be tested. Conversely, the broader the scope, the more tests are needed.

The results are again compared and evaluated. This phase can again lead to the process starting all over again. Another possible outcome may be the identification of prior conditions that explain a scenario in which the program is not recommended for use or is recommended for use only partially.

The deployment phase is also called the inference or prediction phase. Inference or prediction is the process of executing a trained model [184]. The algorithm uses what it has learned during the training phase to produce the desired result. In the context of video analytics, the inference stage is the program used in a video surveillance system to monitor objects in real-world conditions. To achieve real-time performance when executing a machine learning algorithm on input audio or video data, special hardware acceleration is usually required. As noted above, the development of optimal algorithms and the increase in the computing power of peripheral devices in recent years have made it possible to use advanced video analytics based on artificial intelligence in the cameras themselves. Embedded video analytics has obvious advantages: analytical applications have almost instant access to compressed video material, which allows applications to be used in real time, avoiding the additional costs and complexities associated with moving data to the cloud for computing. Embedded analytics also requires less equipment and deployment costs, as the surveillance system requires fewer server resources. For some applications, it may be useful to combine two types of data processing: pre-processing on the camera and post-processing on the server. Such a hybrid system can contribute to the economical scaling of analytical applications by working with video streams from multiple cameras.

The information provided on machine learning will be useful for further analysis of methods for detecting anomalies in video sequences

At the beginning, we concluded that the differences between systems for searching for anomalies in video sequences from video surveillance cameras are due not only (and not so much) to the technical composition of such systems. For the most part, this is the influence of the methods that customers and designers, mathematicians, programmers and other specialists choose to process video information.

Currently, various methods for detecting anomalies in video sequences from video surveillance cameras are known. For greater clarity, we will present a classification of such methods in a graphical form (Fig. 6). They are usually divided into three classes: Supervised Anomaly Detection, Unsupervised Anomaly Detection, and Change Point Detection [174].

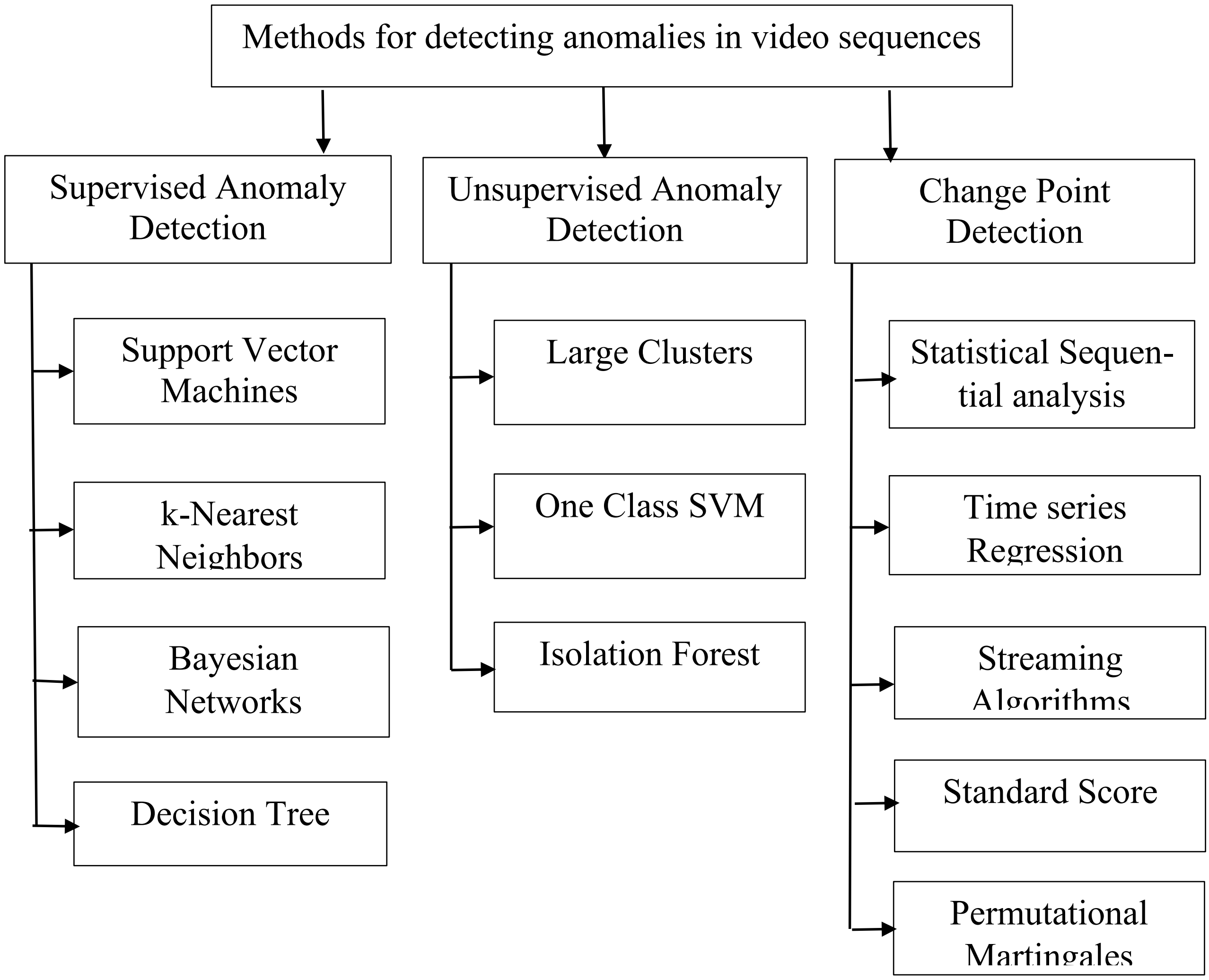

Figure 6. Classification of traditional methods for searching for anomalies in video footage from video surveillance cameras

Judging by the publications, anomaly detection methods are inextricably linked with machine learning. Moreover, methods of the Supervised Anomaly Detection class are related to supervised learning tasks, and methods of the Unsupervised Anomaly Detection class are related to unsupervised learning. Anomaly detection is one of the most common uses of machine learning.

In methods that use supervised anomaly detection, machine learning receives data with labels that determine whether there are anomalies or not. These are methods such as Support Vector Machines (SVM) [174, 185], k-Nearest Neighbors, Bayesian Networks, Decision Tree, and others [186]. However, a big problem for them is the correct installation of labels. As a result, the input video data contains noise, which leads to frequent false positives.

The developers of the SVM method [185] state in the conclusions of their article that machine learning methods have attracted considerable attention among intrusion detection researchers to overcome the shortcomings of knowledge base detection methods. Anomaly detection includes supervised and unsupervised methods. Many algorithms have been used to achieve good results for these methods. This paper provides an overview of machine learning methods for anomaly detection. Experiments have shown that supervised learning methods significantly outperform unsupervised ones if the test data does not contain unknown attacks. Among the supervised methods, nonlinear methods such as SVM, multilayer perceptron (remember this) and rule-based methods show the best performance.

Unsupervised methods such as K-Means, SVM and One-class SVM have achieved better performance compared to other methods, although they differ in their ability to effectively detect all classes of attacks.

Next, we will analyze the k-nearest neighbors method, which is most often used for classification, although it can also be used for estimation and prediction. According to the authors of a large study [186], the k-nearest neighbors method is an example of instance-based learning, in which the training data set is stored, so that the classification for a new unclassified record can be found simply by comparing it with the most similar records in the training set.

Next, we will consider methods with unsupervised anomaly detection. For this class of methods, data is used for which it is not known which data are normal and which are anomalous. This situation is more typical of video surveillance analysis, and in this case, various approaches are used.

The simplest solution is clustering [187]. The authors proceed from the assumption that most of the data (large clusters) are the norm for many tasks, and only a small percentage of the data are anomalous. That is, data belonging to small clusters are the anomalies sought. In the conclusions, the authors state: “In this paper, we investigate the solution of the video anomaly detection problem by comparative representation of the pretext learning problem. We introduce a new instance discrimination task, Cluster Attention Contrast, to obtain reliable normality subcategories as clusters. Experiments on control datasets demonstrate that our approach is effective and interpretable. The Cluster Attention Contrast task can help establish reliable and discriminative feature spaces for video anomaly detection. By exploiting the similarity between subcategory-specific representations and their corresponding centers, the proposed method achieves state-of-the-art performance on reference datasets. In this way, we have established a reliable normal concept without any prior assumptions about reconstruction or prediction errors. In the future, we would like to explore the following directions to improve this work. First, we will investigate the improvement of the performance of detecting small abnormal objects. In addition, we will develop a more reliable method to automatically determine the number of K feature gaps in encoders. Finally, we investigated different distance measures to obtain a more task-appropriate Cluster Attention Contrast method (Fig. 7)",

Other authors use the One Class Support Vector Machines or One Class SVM method. They model only normal data. This algorithm is one of the forms of the classical SVM algorithm [179, 188]. As the name suggests, to train it, it is enough to have only one class – "pure" observations without anomalies. The authors of the study [189] write about this method as follows: "One-class SVMs are used to separate data from one specific class, the target class, from other data. They are trained only on positive examples, that is, on data from the target class. In the feature space, the one-class SVM method separates all data points from the origin by a hyperplane and maximizes the distance of this hyperplane to the origin." After several numerical experiments, the authors state that the corresponding one-class SVM classifier is accepted as the best model.

Figure 7. The bounding box in each subfigure indicates the location of the anomaly [188]

The third class of methods is based on the assumption that anomalies are not only few, but also have attribute values that differ significantly from the values of numerous normal instances. One example of this type is the Isolation Forest method (iForest - isolated forest) [190], which is based on the partitioning of the feature space of the detected video sequence in the form of so-called “isolating trees”. In this case, the anomalous points are significantly distant from the others.

The authors themselves [190] in the conclusions to the article say that this work proposes a fundamentally different model and method that focuses on anomaly isolation. The concept of isolation and its use have not been studied in modern literature. It is shown that the method is very effective in detecting anomalies. Taking advantage of the “few and different” nature of anomalies, iForest isolates anomalies closer to the root of the trees. This unique feature allows iForest to build partial models (as opposed to full models in profiling) and to use only a small portion of the training data to build efficient models. As a result, iForest has linear and low time complexity and low memory requirements, which is very suitable for the case of large data sets.

The authors state that empirical evaluation shows that iForest performs significantly better than some other methods, especially on large data sets. In addition, iForest converges quickly, which allows for high-efficiency anomaly detection. For high-dimensional problems that contain a large number of irrelevant attributes, iForest can achieve high levels of speed and performance. It is also demonstrated that iForest works well even when anomalies are absent from the training set. In essence, Forest isolation is an accurate and efficient anomaly detector, especially for large databases.

A whole family of promising and widespread methods solve the problem of anomaly detection, such as finding points of change of certain properties or characteristics of a video sequence as a time series due to the fact that video frames change over time. The name of this group of methods is appropriate - Change Point Detection (CPD). Several effective methods are based on this idea, allowing to determine the moments in time when significant changes occur in the time series [191]. And this, according to the definition, is a sign of anomalies in the video sequence. The authors of the study [191] say this about their work: "Change point detection is an important part of time series analysis, because the presence of a change point indicates a sharp and significant change in the data generation process. Although many algorithms have been proposed for change point detection, relatively little attention has been paid to evaluating their performance on real time series. Algorithms are usually evaluated on simulated data and a small number of commonly used series with unreliable ground truth.

Clearly, this does not provide sufficient insight into the comparative performance of these algorithms. Therefore, instead of developing yet another change point detection method, we believe it is much more important to properly evaluate existing algorithms on real data. To achieve this, we present a dataset specifically designed for evaluating change point detection algorithms, which consists of 37 time series from different application domains. Each series was annotated by five people to obtain real information about the presence and location of change points.

We analyze the consistency of human annotators and describe evaluation metrics that can be used to measure algorithm performance in the presence of multiple ground truth annotations. We then present a comparative study in which fourteen algorithms are evaluated for each time series in the dataset. Our goal is for this dataset to serve as a testing ground for the development of new change point detection algorithms."

The class of CPD methods, in turn, consists of two groups:

1) Online CPD methods, which are used in real-time video sequence studies, which are often found in modern video surveillance systems.

2) Offline CPD methods, which are used in anomaly detection in video sequences after the shooting process.

To detect moments of changes in the video sequence during video surveillance, the following methods are used (Fig. 1):

1) Statistical sequential analysis - a section of mathematical statistics that studies statistical methods based on a sequential sample formed during a numerical experiment, and at each step it is decided whether more data is needed for evaluation [192].

2) Time series regression [192] - the use of such a model allows you to predict trends, and then compare the predicted and actual indicators; a significant difference indicates a deviation or anomaly.

3) Streaming algorithms – are used to process a data sequence in one or a small number of passes and solve problems in which data is received sequentially and in large volumes [193].

4) Standard score (z-score) – the size of the range of values of measured values [192]. The larger the z-score, i.e. the range of values, the greater the probability that there is a deviation in the stream of values. To detect anomalies, it is necessary to specify the limits of z-scores that are considered normal. Z-scores that fall outside this range are considered anomalous. The method is capable of numerous false positives.

5) Permutational Martingales investigate whether each measurement in the sequence can be determined with equal probability (commutativity property). If there is a commutativity violation in this observed sequence, this indicates the presence of an anomaly [193].

In conclusion, we note that we have quite a lot of methods for searching for anomalies in the video series of surveillance cameras. Of course, neither the limited scope of this work nor the lack of time for relevant research allows us to check all the above methods for searching for anomalies. We called these methods in Figure 6 traditional, because the vast majority of them have been known to the scientific community for a long time.

When you study the achievements of developers of certain methods or devices, it is difficult to expect an objective assessment of them from the developers themselves. You need to critically approach the analysis of such developments, especially in conditions of lack of time to personally verify the validity of the developers' positive conclusions. The criteria can be the conclusions and reviews of other scientists and researchers, as well as the repeated use of these developments in practice. When there is information about the successful use of such developments from third-party researchers, this significantly increases the likelihood that this really happened.

Time moves forward, and new developments, new technical achievements, devices and methods appear. To search for and detect anomalies in the video series of surveillance cameras, researchers and specialists in machine learning are also constantly trying to develop new methods and involve already known developments. Thus, in the twenties, neural networks began to be used for this purpose, although these algorithms were known much earlier. We will talk about this in more detail in the next section.

The article [189] presents an overview of the areas of application of supervised and unsupervised machine learning methods to solve the problem of anomaly detection. The authors used 60 literature sources and conducted complex numerical experiments on various models using a wide range (but not all) of anomaly detection methods known at that time (2013). Below is Table 1 with the results of these studies.

In the conclusions to the article [189], the authors note that neural networks perform well in numerical experiments on anomaly detection in time series. However, some classical machine learning methods have also received positive reviews.

In Table 1, we see that neural networks have serious advantages over other analyzed methods. The disadvantages of this method are also noted. The first of them, the need for training, does not seem to be a disadvantage now. After all, this is a study from 2013. During this time, the use of machine learning has become almost a prerequisite for time series analysis.

Table 1.

Comparative analysis of anomaly detection methods [188]

Method name	Advantages	Disadvantages
1	2	3
K-Nearest Neighbor	1) Easy to understand when there are multiple predictor variables; 2) Useful for building models that include non-standard data types such as text;	1) Has large storage requirements; 2) is sensitive to the choice of similarity function used to compare instances; 3) lacks a principled way to choose k, other than cross-validation or similar; 4) requires powerful computing.
Decision Tree	1) Easy to understand and interpret; 2) Requires little data preparation; 3) Capable of handling both numeric and categorical data; 4) Uses a white-box model; 5) Ability to test the model with tests; 6) Robust; 7) Works well with big data	1) The problem of learning the optimal decision tree; 2) Decision tree scientists create overly complex trees that do not generalize well; 3) There are concepts that are difficult to learn because decision trees cannot express them
Support Vector Machine	1) Finds the optimal separating hyperplane; 2) Can handle very high dimensional data; 3) Some kernels have infinite Vapnyk-Chervonensky dimension, which means they can learn very complex concepts; 4) Usually works very well	1) Requires both positive and negative examples; 2) A good kernel function must be chosen; 3) Requires a lot of memory and CPU time; 4) There are some numerical stability issues
Self-organizing map	1) Simple and straightforward algorithm that works; 2) Unsupervised topological clustering algorithm that works with nonlinear data sets; 3) Visualizes multidimensional data well in 1 or 2-dimensional space. This makes it unique, especially for dimensionality reduction	1) Time algorithm

Continuation of table 1

1	2	3
K-means	1) Low complexity	1) Need to specify k; 2) Sensitive to noise and outliers; 3) Custers are sensitive to the initial assignment of centroids
Expectation - Maximization Meta	1) Can easily modify the model to adapt to different distribution of data sets; 2) The number of parameters does not increase with increasing training data	1) Slow convergence in some cases
Neural Network	1) Can perform tasks that a linear program cannot; 2) When an element of a neural network fails, it can continue to work without any problems; 3) A neural network learns and does not need to be reprogrammed; 4) Can be implemented in any program	1) Neural networks require training to work; 2) Neural network architecture is different from microprocessor architecture, so it needs to be tuned; 3) Requires a lot of processing time for large neural networks

Another disadvantage, such as the need for a lot of processing time for large neural networks, has also almost disappeared over the years due to the continuous development of computing and microprocessor technology, the increase in speed and memory in modern computers. Given this, we can conclude that it is possible and advisable to use neural networks to detect anomalies in video footage from surveillance cameras, which is confirmed by many modern researchers.

Despite the presence of many methods for searching for anomalies in video sequences, which are in the arsenal of researchers, recently the main tool in this field is neural networks. This situation is not only in the direction of searching for anomalies. Analysis of big data in general, which ten years ago was carried out using traditional or classical machine learning algorithms, now uses mainly methods based on neural networks. Our research of literary sources indicates that modern video surveillance systems in large campaigns and in government institutions most often use the so-called "smart" intelligent technologies based on models and methods of neural networks that detect objects in video sequences. Such technologies also allow finding various properties of objects both in each frame separately and in the dynamics of the video stream. Neuronal networks (NN) are mathematical models of the organization of real biological neural networks [194]. But unlike mathematical models of biological neural networks, artificial neural networks do not require an accurate description of all chemical and physical processes, such as the description of the "firing" of an action potential, the work of neurotransmitters, ion channels, secondary mediators, transport proteins, etc. Artificial neural networks are required to resemble the work of real biological neural networks only at the functional level, not at the physical level.

The basic element of a biological neural network is a neuron, shown in Fig. 8.

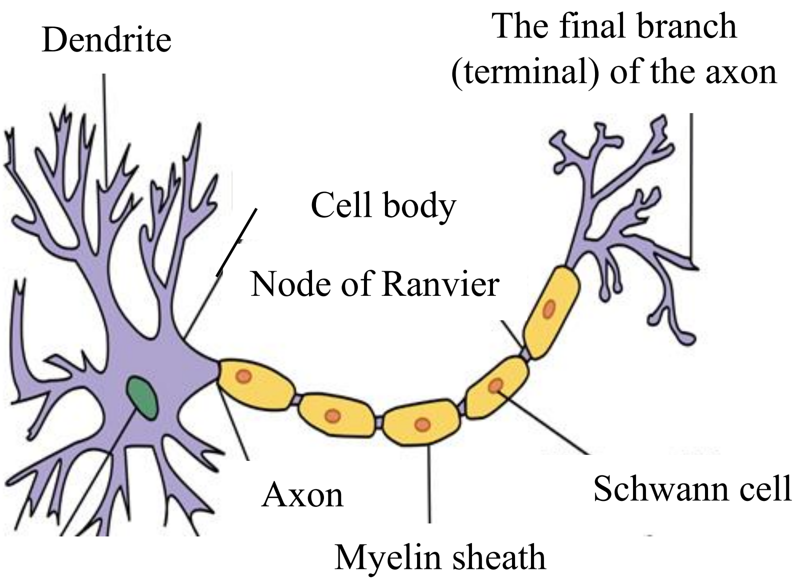

Figure 8. Typical structure of a biological neuron [189]

To understand the operation of an artificial neuron, we will describe the functioning of a biological neuron in general. The structure of a biological neuron can be simplified to the following [194]: dendrites, neuron body and axon. Dendrites are branching processes that collect information from the input to the neuron (this can be external information from receptors, for example, from a cone in the case of color, or internal information from another neuron). If the input information has activated the neuron (in the biological case, the potential has become higher than a certain threshold value), an excitation wave is generated that propagates along the membrane of the neuron body, and then through the axon, by releasing a neurotransmitter, transmits a signal to other nerve cells and tissues. In nervous tissue, the axon connects to the body of the next nerve cell, to dendrites or to another axon. Based on this, Warren McCulloch and Walter Pitts in 1943 proposed a model of a mathematical neuron. And in 1958, Frank Rosenblatt, based on the McCulloch-Pitts neuron, created a computer program, and then a physical device - a perceptron. This is where the history of artificial neural networks began. Now let's consider the structural model of a neuron, which we will deal with further (Fig. 9).

This diagram contains the following components:

1) the input column vector of parameters X. The components of the vector xi are numbers that, for a real neuron, determine the degree of activation of various receptors and which came to the input of the neuron;

2) the column vector of weights W, the numerical values of which wi change during the learning process. Biological learning also exists. In the presence of synaptic plasticity, the neuron learns to respond correctly to signals from its receptors);

3) the adder is a functional block of a neuron that adds up all the input parameters multiplied by their corresponding weights;

4) the neuron activation function f(S) is the dependence of the neuron output value on the value coming from the adder;

5) subsequent neurons, where one of their many own inputs is fed with the value from the output of this neuron (this layer may be absent if this neuron is the last in the computational network).

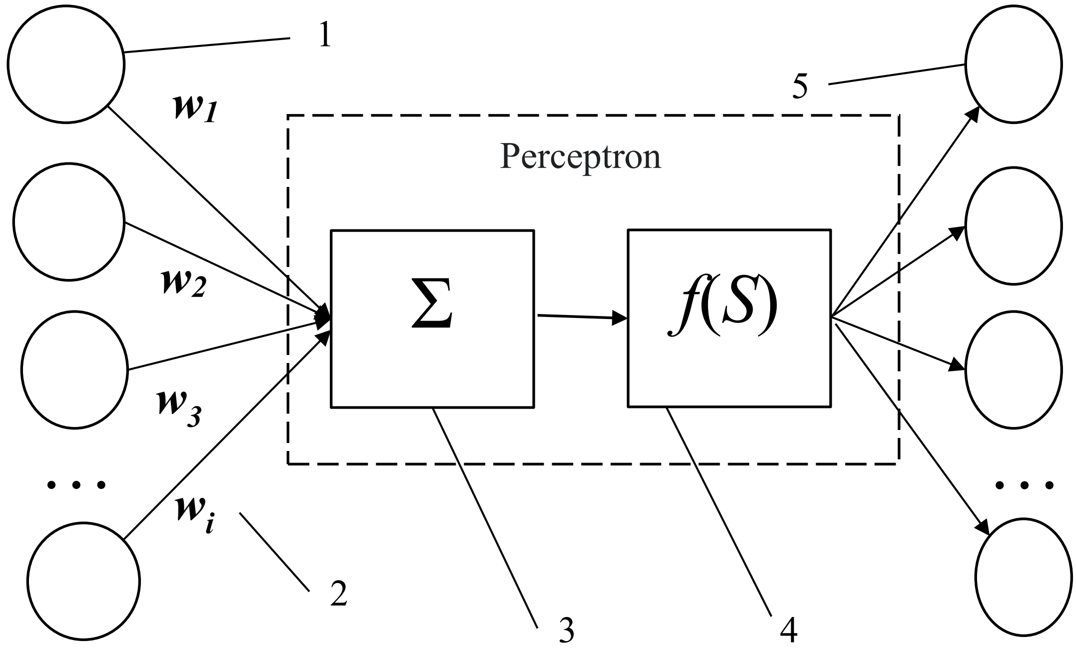

Figure 9. Scheme of an artificial neuron (perceptron): 1 - neurons, the output signals of which are fed to the input of this neuron; 2 - weights of input signals; 3 - adder of input signals; 4 - transfer function calculator; 5 - neurons, the inputs of which are fed with the signal of this neuron [193]

Then, from these minimal structural units, classical artificial neural networks are assembled [194]. The following terminology is adopted:

1) input (receptor) layer - this layer consists of neurons that do not perform computational operations, but simply transmit the received input signal to the output, possibly amplifying or weakening it. The task of the layer is to provide digital information received by receptors from the "external" world. The input layer contains as many elements as there are input parameters (plus a bias-term, necessary to shift the activation threshold) (Fig. 10);

2) associative (hidden) layer - a deep structure capable of memorizing examples, finding complex correlations and nonlinear dependencies, and constructing abstractions and generalizations. It consists of neurons that perform basic computational operations. In the general case, these are several layers between the input and output. We can say that each layer prepares a new (higher-level) feature vector for the next layer. It is this layer that is responsible for the emergence of high-

level abstractions in the learning process. The structure contains as many neurons and layers as you like, or may not exist at all;

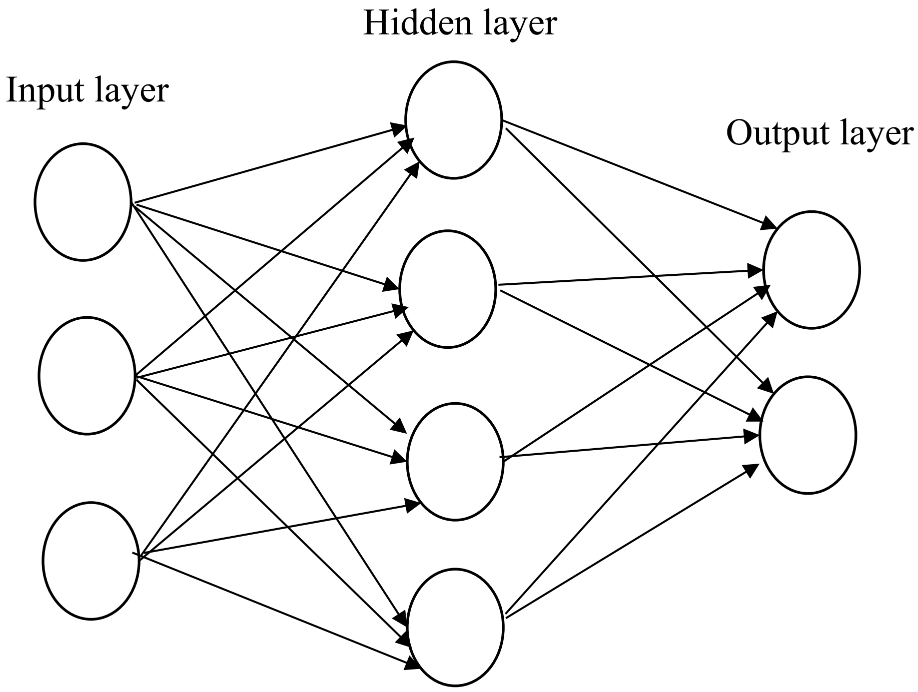

Figure 10. Classical topology of a neural network with input (receptor), output, decision, and associative (hidden) layers [194]

3) the output layer is a layer, each neuron of which is responsible for a specific class. The output of this layer can be interpreted as a probability distribution function of the object belonging to different classes. The layer contains as many neurons as there are classes represented in the training sample. If there are two classes, then you can use two output neurons or limit yourself to just one. In this case, one neuron, as before, is responsible for only one class, but if it gives values close to zero, then the sample element according to its logic must belong to another class.

Due to the presence of hidden associative layers, an artificial neural network is able to build hypotheses based on finding complex dependencies. For example, for convolutional neural networks that recognize images, the input layer will be fed with the brightness values of the image pixels, and the output layer will contain neurons responsible for specific classes (person, car, tree, house, etc.). In the hidden layers, neurons that are excited by straight lines of different angles will begin to appear (specialize) "by themselves", then react to corners, squares, circles, primitive patterns: alternating stripes, geometric mesh ornaments. Closer to the output layers are neurons that react, for example, to an eye, wheel, nose, wing, leaf, face, etc. (Fig. 11).

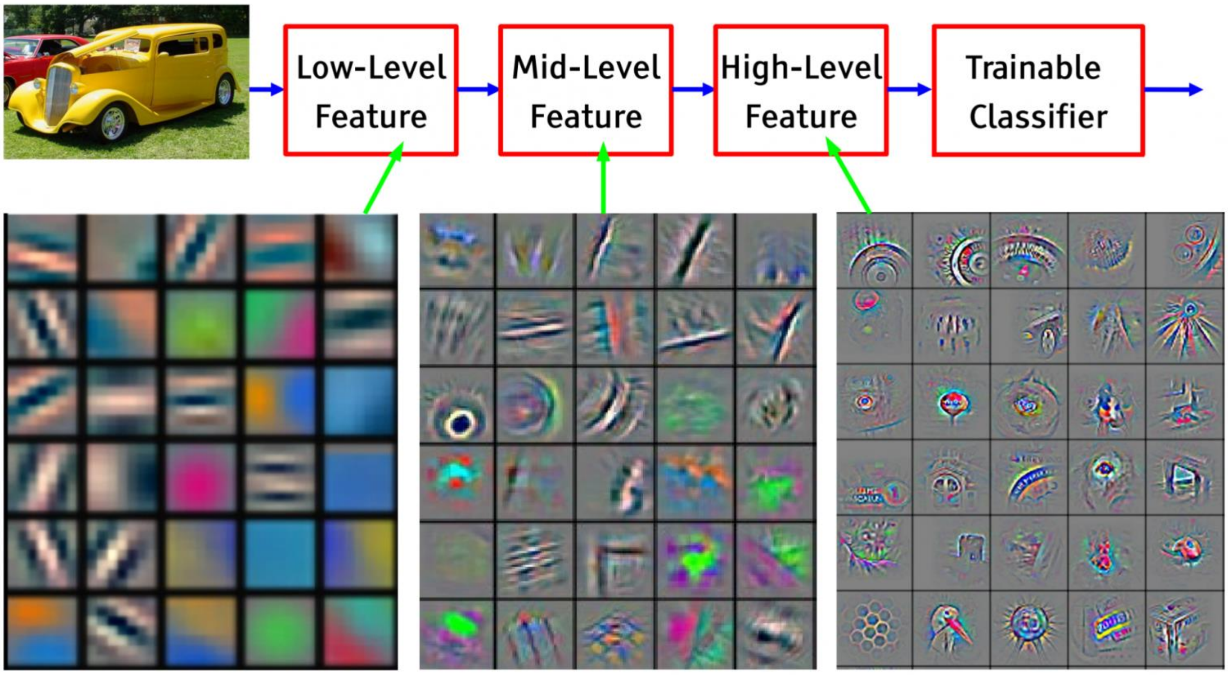

Figure 11. Formation of hierarchical associations in the process of training a neural network [194]

In the simplest case, the entire neural network can be represented by a single neuron. In this case, even a single neuron sometimes copes well with the task, especially when it comes to recognizing an object class in a space in which objects of these classes are linearly separable.

For a mathematical description of the above about artificial neurons, we first write. At the input of the neuron we have a vector of parameters, represented in numerical form [194]:

$$X_i = \left\{x_{i_1}, x_{i_2}, \ldots, x_{i_n}\right\}$$

where n – is the number of parameters;

i – is the number of classes, if the classification task is considered.

In this case, each class is assigned Y_i − a class that characterizes the success of the form 1 or 0. The neural network must find the optimal separating hypersurface in a vector space, the dimension of which corresponds to the number of features. Training a neural network in this case − is finding such values (coefficients) of the weight matrix W for which the neuron responsible for the class will give values close to unity if the corresponding class is defined and values close to zero if the corresponding class is not defined.

The result of the neural network is usually called a hypothesis. The hypothesis is denoted as h(X), showing the dependence of the hypothesis on the input features (parameters) of the object. The value of the hypothesis is calculated by the following formula [195]:

$$h_w(X) = f(\sum_{k=1}^{|w|} w_k \cdot x_k) \equiv \sigma(w \cdot x)$$

where $h_w(X)$ − the value of the hypothesis for the vector of input parameters X;

− the weight of each parameter;

− the k-th input parameter of the vector X.

As can be seen from the formula, the result of the neuron's operation is the activation function (often denoted by) of the sum of the product of the input parameters by the coefficients sought in the training process. Let's figure out what an activation function is.

Any real values can be input to the neural network, and the coefficients of the weight matrix can also be any real numbers [194]. Therefore, the result of adding such numbers can be any real number from minus to plus infinity. Each element of the training sample has a class value for this neuron from zero to one. It is necessary to obtain from the neuron an output value in the same range from zero to one and make a decision about the class depending on what this value is closer to. It is better to interpret this value as the possibility that the element belongs to this class. That is, such a monotonic smooth function is needed that will map elements from the set of real numbers into the range from zero to one.

In real biological neurons, such a continuous activation function has not been implemented by nature. In human cells, there is a resting potential, which is on average -70 mV [194]. If information is supplied to a neuron, the activated receptor opens the ion channels associated with it, which leads to an increase or decrease in the potential in the cell. An analogy can be drawn between the strength of the reaction to receptor activation and one coefficient of the weight matrix obtained during the learning process. As soon as the potential reaches the value of -50 mV, an action potential arises, and a wave of excitation reaches the presynaptic ending along the axon, releasing the neurotransmitter into the intersynaptic medium. That is, real biological activation is stepwise, not smooth: the neuron is either activated or not. This shows how mathematically free we are in building our models. Taking the basic principle of distributed computing and learning from nature, we are able to build a computational graph consisting of elements that have any desired properties. In the example, we want to receive continuous, i.e. continuous, analog, rather than discrete values from a neuron. Although in the general case the activation function can be different.

The conclusion from the above is as follows: "Neural network training (synaptic learning) should be reduced to the optimal selection of weight matrix coefficients in order to minimize the error allowed" [194]. In the case of a single-layer neural network, these coefficients can be interpreted as the contribution of the element parameters to the possibility of belonging to a particular class.

It is necessary that the hypotheses of the neural network correspond as much as possible to reality (to real classes of objects). Actually, this is where the main idea of learning from experience is born. Now we will need a measure that describes the quality of the neural network. This functional is usually called the "loss function". The functional is usually denoted as J(W), showing its dependence on the coefficients of the weight matrix. The smaller the functional, the less often our neural network makes mistakes and the better. It is to minimize this functional that training is reduced. Depending on the coefficients of the weight matrix, the neural network can have different accuracy. The learning process is a movement along the hypersurface of the loss functional, the goal of which is to minimize this functional (Fig. 12) [193].

Usually, the coefficients of the weight matrix are initialized randomly. During the training process, the coefficients change. The graph (Fig. 12) shows two different iterative learning paths for the coefficients w1 and w2 of the neural network weight matrix.

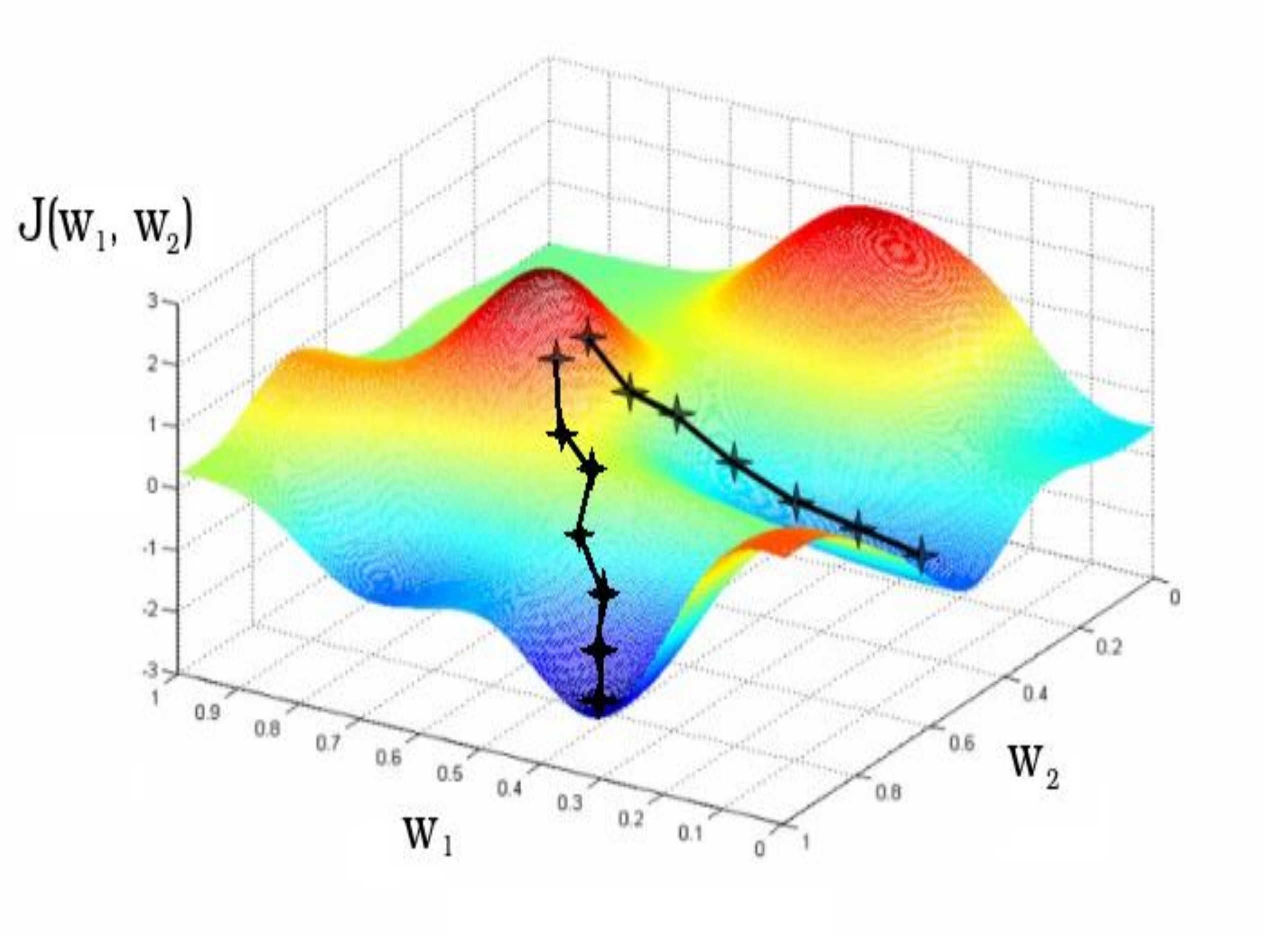

Figure 12. The learning process as gradient descent to a local minimum of the loss functional [194]

Conventional artificial neural networks require enormous computational power from the computers used. For example, in a dataset that is typically used to train computer vision models, images are 32x32 pixels in size and have 3 color channels. This means that one fully connected neuron in the first hidden layer of this neural network will have 32x32x3=3072 weight units. This is already quite a lot, because there are other neurons and other layers. But the image can be larger, for example, 300x300x3=270000 weight units, the training of which requires large computational power.

Such a huge neural network requires a lot of computational resources. Reducing this volume is the task of convolutional neural networks.

In the study [174], the authors state that the modern basic models that are successfully applied to the analysis of visual images are convolutional neural networks (CNN). These networks have two or more hidden layers. The first such successful network, AlexNet [194], which was created in 2012, contained 8 hidden layers. Since then, deep neural networks have been widely developed and distributed in various areas due to the fact that features in such networks are detected automatically, as well as due to their high efficiency in solving many tasks, including when working with images and videos [174]. The limited scope of our study does not allow us to dwell in detail on the history of the emergence and development of convolutional neural networks, which began in the middle of the twentieth century. In the study [195], scientists from Pakistan summarized data on convolutional neural networks known as of 2019. They processed 255 (!) literary sources. One of the results of their work is a detailed chronological and classification scheme for CNNs (Fig. 13).

The general conclusion that the researchers make and which best describes the development of convolutional neural networks for image processing is this: CNNs, when analyzing time series, allow using fewer parameters to obtain more accurate results.

There are many options for training a neural network. For example, the evolutionary (genetic) algorithm and the gradient descent method [194]. Evolutionary algorithms are a direction in artificial intelligence based on modeling natural selection. Previously, this approach was used to train deep layers of a neural network. Evolutionary algorithms have completely given way to gradient descent-based optimizers for the task of training deep neural networks, but they can still be used for educational purposes: evolutionary algorithms are very easy to understand. It is better to start with them for beginners, it is worth knowing them for further comparison of computational complexity and convergence with other (currently popular) methods. The gradient descent and backpropagation method is more complicated, but it is one of the most effective and popular training methods.

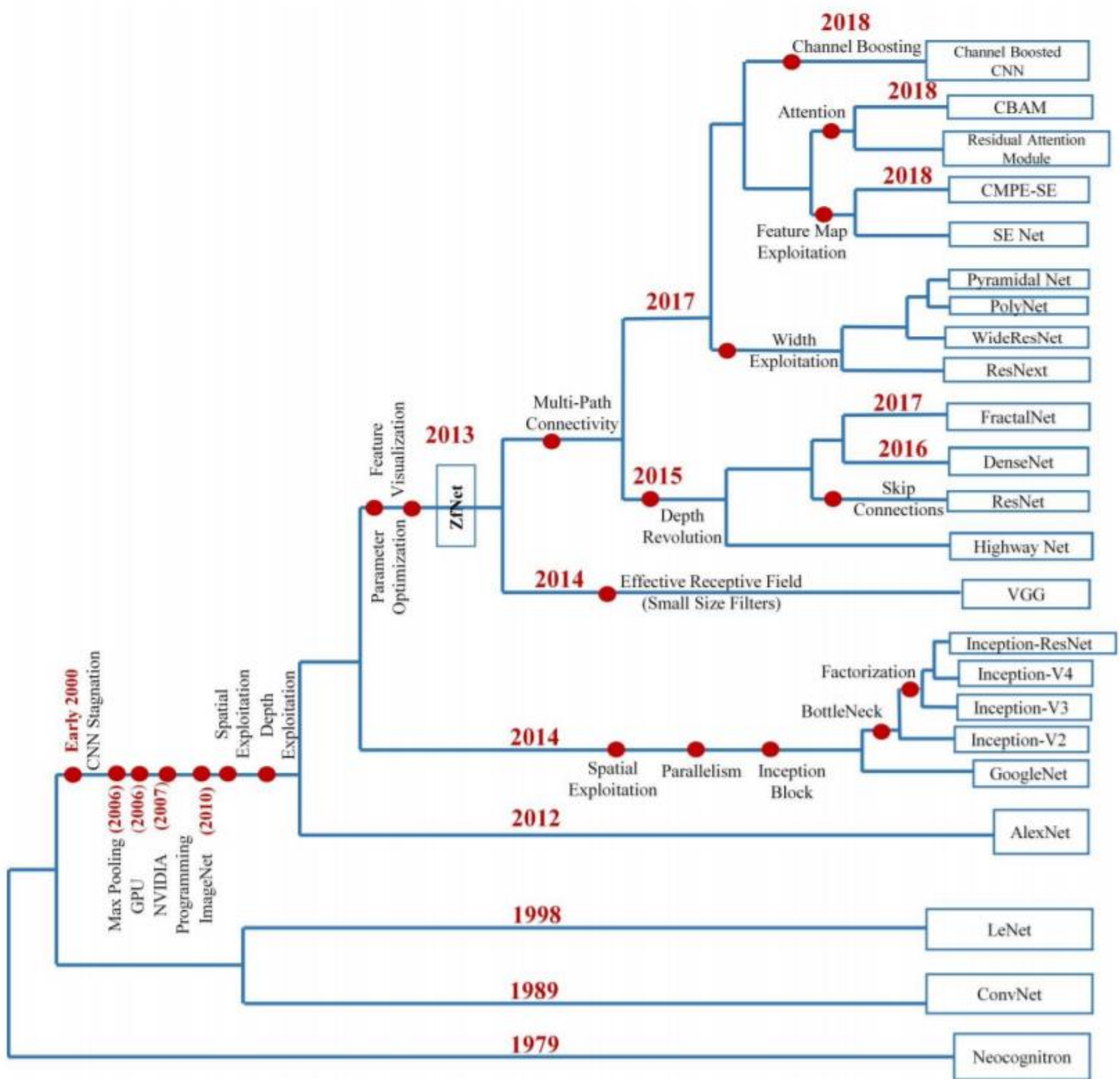

Figure 13. Historical development and hierarchical classification of convolutional neural networks [196]

Unfortunately, the scope of this monograph does not allow us to consider convolutional neural networks in more detail.

The input data set should be test and training video sequences that define various dynamic situations that affect the quality of the neural network as a component of the anomaly detection system. In particular, such situations include the speed of cars, the distance of the object from the camera lens, the lighting conditions of the scene, extraneous objects in the scene, etc. The standard frequency of 24 frames per second was chosen as the shooting frequency. The total number of frames in the sequence is 124 frames, the image format is HD_720 (1280 pixels × 720 pixels), the color system is sRGB; the graphic image format is JPEG.

The training set consists of sequences of ordinary video frames. The model will be trained to restore these sequences. It is necessary to prepare the data for the model by performing the following three steps:

1) divide the training video frames into time sequences, each of size 10, using the sliding window technique;

2) change the size of each frame to 256×256 to ensure the same resolution of the input images;

3) scale the pixel values from 0 to 1, dividing each pixel by 256.

The development of experiments to verify the research conducted in this work and the statements and conclusions made on their basis should begin with the selection of components and components of the video surveillance system. First of all, it is necessary to decide how the video surveillance itself and the analysis of its results will be carried out - only with human participation, with automatic analysis of video recordings under human supervision, or fully automated from changing the video recording parameters (camera tilt, viewing angle, etc.) to intelligent analysis of recorded video information. In addition, it is necessary to decide when this video analysis will be done, simultaneously with video surveillance or after the latter.

The studies presented below have shown that from the point of view of searching for anomalies in the video sequence of surveillance cameras, the main factor is the financial capacity of the customer of the video surveillance system. Even the cheapest video camera can record video of some objects, events or processes, and then, using video analytics, analyze the recorded information using cloud services in terms of detecting anomalies in the video sequence. This option will cost quite a bit - from 3-4 thousand UAH (Fig. 14).

This is the AHD KIT 10800 PRO video surveillance kit, assembled on the basis of AHD format equipment (analog high resolution) [197].

The cylindrical universal AHD KIT 10800 PRO video camera with infrared illumination is made of street ABS plastic and has an IP-66 housing protection rating, which can withstand winter temperatures down to -300. The camera is assembled on the basis of a modern CMOS IR-CUT matrix and is capable of broadcasting video with a resolution of 2MP. A megapixel lens with a focal length of f=3.6 mm forms a high-quality focused image with high clarity and a wide viewing angle (900). The number and power of LEDs provide bright illumination for detailed night shooting with a viewing range of up to 20 meters. The wavelength of the infrared illumination is 850 nm and is invisible to the human eye. The illumination automatically turns on in low light.

Figure 14. Video surveillance system aHD KIT 10800 PRO [197]

The AHD KIT 10800 PRO DVR video recorder supports up to 4 video cameras and is capable of recording video in real time on all four channels. Through the implemented VGA and HDMI outputs, video can be displayed on a monitor or LCD TV. The recorder supports a choice of recording modes based on motion detection, constantly or on a schedule. Using the remote access function via connection via the

Xmeye Cloud cloud service, viewing video, as well as archived recordings, is available both via the global Internet on a computer and via a smartphone based on Android or iOS.

Such a video surveillance system fully corresponds to the functional diagram of the anomaly detection system in the video sequence, shown in Fig. 1. Only the video sequence processing unit will be after the unit for connecting to the Internet.

For a comparative analysis of several methods for searching for anomalies in the video sequence, test video clips may be sufficient. But in many literary sources studied by us, it was said that training models and their further use to search for anomalies in the video sequence is an individual matter. Therefore, it is better to conduct such research work in conditions close to real ones, that is, with video sequences obtained from specific video cameras.

Above, we discussed the simplest version of a modern video surveillance system, to the results of which it would be possible to apply methods for searching for anomalies. Next, it would be advisable to conduct a comparative analysis of different methods for searching for anomalies in the video sequence for video surveillance systems that differ in their capabilities, but also in price. The volume and material base of such experiments is entirely dependent on the appropriate funding.

If research funding allows, then, undoubtedly, it would be necessary to search for anomalies in the video sequence of cameras of such a surveillance system, which is designed to process the results of video surveillance online. Such an automated system is much more expensive, but it is also more in demand.

The automated system will receive broadcast video from video surveillance cameras and independently analyze the situation, which will relieve the operator of the load, thereby reducing the likelihood of error. That is, the proposed system consists of one or more video cameras that broadcast what is happening, as well as software that will process the broadcast to search for anomalies in the video sequence [174].

A valuable feature of video surveillance in such a system will be the ability to perform its functions with an autonomous power source. The video camera must be equipped with a battery and a recorder to continue recording even in the event of a power outage. But this requires economical algorithms for processing the results of surveillance. The video camera must have good recording quality in the dark and be equipped with a motion sensor. Such a set of equipment will provide the most efficient and accurate analysis of information in any situation.

The video surveillance system must be developed as a client-server. The server, together with the main software system, is designed to process the broadcast to search for suspicious situations, focusing the operator's attention on what is happening.

The client software will allow those who have been authorized and can see the broadcast from the video cameras to receive notifications in case of suspicious events. In the event of a loss of connection between the video camera and the server, the client program should receive a message about possible suspicious activity.

As mentioned above, in order to choose a method for searching for anomalies in the video sequence of surveillance cameras, it is necessary to conduct a comparative analysis of several well-known modern methods. Now large software development companies offer their services and software products for using convolutional neural networks to work with images and video sequences.

We have made numerous attempts to use these software products, but the computing capabilities of the computer did not allow this.

Thus, a universal functional block diagram of the anomaly detection system in the video sequence of video surveillance cameras, which belongs to the intelligent video analytics systems, has been developed. The video surveillance and anomaly detection system in the video sequence can be intelligent due to the use of valuable intelligent video cameras, or perhaps due to its deep learning in neural networks and the use of cloud computing resources.

For the successful use of methods for detecting anomalies in the video sequence of video surveillance cameras, their appropriate analysis is required. Such a comparative analysis is carried out in this study. For this purpose, many modern methods for detecting anomalies in the video sequence have been considered and the foundations of the theory of deep neural networks and convolutional neural networks have been considered from the point of view of the possibility of their application for detecting anomalies in the video sequence of video surveillance cameras.

Both traditional statistical methods and modern methods based on neural computing networks have been considered, including the history of their emergence and features of use. A historical and chronological classification of methods based on neural networks is given. A conclusion is made about the need to use convolutional neural networks to search for anomalies in video sequences.

A methodology has been developed for conducting numerical experiments on the comparative analysis of methods for detecting anomalies in video sequences from video surveillance cameras.