10. Lysytska I. V. Comparing on effectiveness of superboxes for some modern cipher // Radioelectronics, computer science, management. 2012. Issue 1. P. 37–44.
11. Dolgov V. I., Kuznetsov A. A., Isaev S. A. Differential properties of block symmetric ciphers submitted to the Ukrainian competition // Electronic simulation. 2011. Vol. 33, Issue 6. P. 81–99.
12. Ruzhentsev V. I. The probabilities of two-rounds differentials for Rijndael-like ciphers with random substitutions // Applied Radio Electronics. 2014. Vol. 13, Issue 3. P. 235–238.
13. Practical and Provable Security against Differential and Linear Cryptanalysis for Substitution-Permutation Networks / Kang J.-S. K., Hong S. H., Lee S. L., Yi O. Y., Park C. P., Lim J. L. // ETRI Journal. 2001. Vol. 23, Issue 4. P. 158–167. doi: https://doi.org/10.4218/etrij.01.0101.0402

Запропоновано модель детектора об'єктів і критерій ефективності навчання моделі. Модель містить 7 перших модулів згорткової мережі Squeezenet, два згорткові різномасштабні шари, та інформаційно-екстремальний класифікатор. Як критерій ефективності навчання моделі детектора розглядається мультиплікативна згортка частинних критеріїв, що враховує ефективність виявлення об'єктів на зображенні та точність класифікаційного аналізу. При цьому додаткове використання алгоритму ортогонального узгодженого кодування при обчисленні високорівневих ознак дозволяє збільшити точність моделі на 4 %.

Розроблено алгоритм навчання детектора об'єктів за умов малого обсягу розмічених навчальних зразків та обмежених обчислювальних ресурсів, доступних на борту малогабаритного безпілотного апарату. Суть алгоритму полягає в адаптації верхніх шарів моделі до доменної області використання на основі алгоритмів зростаючого розріджено кодуючого нейронного газу та симуляції відпалу. Навчання верхніх шарів без вчителя дозволяє ефективно використати нерозмічені дані з доменної області та визначити необхідну кількість нейронів. Показано, що за відсутності тонкої настройки згорткових шарів забезпечується 69 % виявлених об'єктів на зображеннях тестової вибірки Inria Aerial Image. При цьому після тонкої настройки на основі алгоритму симуляції відпалу забезпечується 95 % виявлених об'єктів на тестових зображеннях.

Показано, що використання попереднього навчання без вчителя дозволяє підвищити узагальнюючу здатність вирішальних правил та прискорити ітераційний процес знаходження глобального максимуму при навчанні з учителем на вибірці обмеженого обсягу. При цьому усунення ефекту перенавчання здійснюється шляхом оптимального вибору значення гіперпараметру, що характеризує ступінь покриття вхідних даних нейронами мережі

Ключові слова: зростаючий нейронний газ, детектор об'єктів, інформаційний критерій, алгоритм симуляція відпалу

UDC 004.891.032.26:629.7.01.066
DOI: 10.15587/1729-4061.2018.139923

IMPROVING THE EFFECTIVENESS OF TRAINING THE ON-BOARD OBJECT DETECTION SYSTEM FOR A COMPACT UNMANNED AERIAL VEHICLE

V. Moskalenko
PhD, Associate Professor*
E-mail: systemscoders@gmail.com

A. Dovbysh
Doctor of Technical Sciences, Professor,
Head of Department*
E-mail: a.dovbysh@cs.sumdu.edu.ua

I. Naumenko
PhD, Senior Researcher, Colonel
Research Center for Missile Troops and Artillery
Gerasima Kondratyeva str., 165, Sumy, Ukraine, 40021
E-mail: 790895@ukr.net

A. Moskalenko
PhD, Assistant*
E-mail: a.moskalenko@cs.sumdu.edu.ua

A. Korobov
Postgraduate student*
E-mail: artemkorr@gmail.com

*Department of Computer Science
Sumy State University
Rimskoho-Korsakova str., 2, Sumy, Ukraine, 40007

1. Introduction

Unmanned aviation is widely used in the tasks of inspection of technological and residential facilities, protection and reconnaissance activities, as well as in the sphere of transportation of small size loads. One of the ways to increase the functional efficiency of the unmanned aerial vehicle (UAV) is to introduce technologies of artificial intelligence to analyze sensor information. Since most informative sensor system is surveillance cameras, the development of visual detectors of objects of interest is a promising direction. However, the limited computing resources and the weight of the UAV do not make it possible to implement in it the models of analysis of visual data, adapted to a full range of possible observation conditions and a variety of modifications to objects of interest. This causes the need for the development of computationally effective models and algorithms of adaptation to new conditions of functioning, inherent in the specific domain of application area.

In terms of computational complexity and generalizing ability, the leader among the models for visual information analysis is convolutional neural networks. However, training and retraining of convolutional networks requires the selection of the optimal network configuration and significant computing resources. It is possible to reduce computational load at relearning of a neural network for adaptation to the new conditions of functioning due to the use of the transfer learning principle by copying the first layers of the network of ImageNet, trained on the ImageNet dataset [1, 2]. However, the layers of high-level feature representation of observations require learning from scratch. In this case, it is difficult to estimate in advance the required number of convolutional filters in each convolutional layer, which is why the promising approach to learning convolutional filters is to use the principles of growing neural gas, which makes it possible to determine automatically the necessary number of neurons. In this case, the layers of decision rules for the detector and feature extraction layers require fine-tuning, which is typically implemented based on the modifications of the error backpropagation algorithm [2, 3]. However, this algorithm is characterized by a low convergence rate and getting stuck in local minima of loss function. There are alternative metaheuristic search optimization algorithms, however, effectiveness of using these algorithms in problems of networks fine tuning is scantily explored [3].

That is why the research, aimed at enhancing the effectiveness of learning the detector of objects under conditions of limited computing resources and learning data, is relevant.

2. Literature review and problem statement

Papers [4, 5] examined the models of the feature description of visual observations, which are based on the use of Haar-like features, histograms of oriented gradients, local binary templates, histograms of visual words and other low-level features. These models use low-level local information and ignore high-level contextual information, which complicates the implementation of effective objects detectors under conditions of limited volumes of unlabeled datasets. In addition, non-hierarchy feature description models are characterized by high labor consuming computing under conditions of a large variation of observations.

In the image analysis problems, numerous models of hierarchical description of observations based on convolutional neural networks are becoming increasingly popular. The most popular of them include VGG-16, VGG-19, ResNet-50, GoogleNet, MobileNet, SqueezeNet and others [3, 4]. These networks differ in the number of layers, existence of residual connections and multiscale filters in each of the layers. In this case, it was shown in papers [5, 7] that the models trained on dataset ImageNet accumulate in themselves important information regarding the analysis of visual images. However, the more a target domain area differs from the ImageNet images, the fewer layers of trained networks are possible to be reused. In addition, narrowing a domain area, for example, by reducing the number of recognition classes, makes it possible to reduce the need for computational resources for functioning of the objects detector on terrain.

It was proposed in papers [8, 9] to carry out fine tuning of the detector of objects based on the convolutional network VGG-16 using the algorithm of mini-batch error backpropagation. However, this would require a significant volume of learning datasets and a few hours of working on the graphic processor of a personal computer. In article [10], it was proposed to perform scanning of the normalized high-level feature map by a sliding window, in each position of which classification analysis was carried out. In research [11], it was proposed to use a classification analysis of a high-level feature representation within the so-called information-extreme technology (IEI technology). This technology is based on the adaptive binary encoding and construction of the optimal in the information sense radial-basis decision rules in the Hamming binary space. However, these approaches are not very effective, since high-level feature maps are not sufficiently adapted to the domain application area, and effective techniques of fine tuning of the network were not proposed. In addition, network VGG-16 is computationally sophisticated and that is why its use does not ensure decision making in real time under conditions of limited resources of the autonomous compact device.

Implementation of training high-level layers of neural network for adaptation to a domain area through prior unsupervised learning and subsequent fine tuning based on supervised learning is common scheme in learning deep networks. Papers [11, 12] propose unsupervised learning of convolutional networks based on autoencoder or the restricted Boltzmann machine that require a large size of training set and a long time of learning for obtaining an acceptable result. In articles [11, 14], it is proposed to combine the principles of neural gas and sparse coding for learning convolutional filters for unlabeled datasets. This approach has a soft competitive learning scheme, which increases the probability of the algorithm convergence to the optimal distribution of neurons on learning datasets. In this case, embedding the sparse coding algorithms makes it possible to improve the noise immunity and generalization ability of feature representation. However, the number of neurons is not known in advance and is assigned at the discretion of a developer or optimized, which leads to an increase in the number of learning iterations.

It is difficult to estimate the required number of convolutional filters in the high-level convolutional layers in advance, which is why a promising approach to learning convolutional filters is to use the principles of growing neural gas that is able to automatically determine the required number of neurons [15]. However, the mechanism of the insertion of new neurons into the neural gas algorithm, based on setting the insertion period, often leads to distortion of the formed structures and instability of the learning process. However, it was shown in article [16] that it is possible to ensure stability of learning by setting the "accessibility radius" of neurons. It involves the replacement of the neurons insertion period with the threshold of a maximum distance of a neuron from every point of the learning dataset, referred to it. However, the mechanisms of neurons updating and assessment of the distance of the input space points to the neurons for the purpose of adaptation of the learning process to the procedure of sparse coding observations have not been revised yet.

The problem of objects detection based on feature maps of convolutional network is solved based on architectures Yolo (You only look once), Faster R-CNN, DetectNet and SSD (Single Shot MultiBox Detector) [7]. Among them, SSD architecture has gained the widest application in mobile devices, because it has the least computational complexity in the detection mode. However, learning such layers under conditions of a limited size of training datasets and computing resources based on the error backpropagation and the stochastic gradient descent is ineffective. One of the promising ways of training neural networks and fine tuning is the application of metaheuristic algorithms because they are characterized by better convergence and less probability of getting stuck in the "bad" local optimum [17]. Among them, it is worth highlighting the simulated annealing algorithm, since its use made it possible to exceed the results of the gradient descent algorithms in optimization of neural network classifiers [4]. However, its use in the problems of fine tuning of convolutional filters and training the detector of objects remains insufficiently studied.

3. The aim and objectives of the research

The aim of this research is to improve the effectiveness of training a detector of objects under conditions of limited computation resources of on-board systems of an UAV and labeled training datasets of the assigned domain application area. To achieve the set aim, it is proposed to solve the following problems:

– to develop the model of data analysis for detection of objects of interest on aerial images under conditions of limited computing resources and the size of the labeled training datasets from the domain application area;

– to develop the algorithm of training the detector of objects with the use of the ideas and principles of growing neural gas and sparse coding for unsupervised pre-training and metaheuristic optimization for fine tuning;

– to explore the dependence of training efficiency of object detector on the parameters of the algorithms of unsupervised learning of the high-level feature representation of observations.

4. Model and algorithms for training the object detector

4. 1. Model of the object detector

To solve the problem of the development of the model for data analysis under conditions of limited computing resources of a compact UAV, it is necessary to maximize the use of all available a priori information. The use of the available a priori information for the synthesis of the architecture of the convolutional network for analysis of visual information under conditions of limited computing resources of a compact UAV can enhance the effectiveness of training, retraining and inference. Transfer learning technique is one of the examples of using a priori information, accumulated in the trained network for reusing. In this case, the closer the domain area, where the neural network was trained, to the new domain area, where accumulated knowledge will be used, the more layers can be borrowed.

High-level layers of a neural network, as a rule, require training from scratch for maximum adaptation to a new domain area. For this, it is proposed to use the additional convolutional layers, where the filters from the kernels of 1×1 and 3×3 have unsupervised learning. In this case, to ensure the noise immunity and informativeness of feature representation, it is proposed to calculate the activation of each feature map pixel based of the algorithm is orthogonal matching pursuit (OMP) with the ReLU function [12, 14].

For many tasks of terrain monitoring, there is no need for accurate localization of objects, but only information about their location in a particular area is enough. That is why when developing the layers of detecting objects on the aerial images of terrain, the architecture of Shot Single is taken as a basis, but regression to refining the objects' boundaries is not used [7]. In addition, under condition of the limited volume of the labeled learning datasets, reduction of the problem of detection to the problem of classification analysis is appropriate, because the regression analysis may not be enough for acceptable results. It is proposed to carry out this classification analysis of the feature map in the framework of the so-called IEÌ technology, because it makes it possible to synthesize the classifier with low computational complexity and relatively high reliability under conditions of limited volume of training datasets [11]. Fig. 1 schematically shows the architecture of the proposed detector.

A training dataset for the classification analysis is formed by the results of calculation of the measure of intersection of the object bounding box, made by an expert with the projection of the feature map pixel onto the input image (receptive pixel field). If the Jaccard intersection measure (is the same as Intersection-over-Union, IoU) is greater than the threshold value *Th*, then we refer the pixel of the feature map to a positive training dataset, and the rest – to a negative dataset.

Information-extreme classifier that evaluates belonging of an image patch to one of the *R* classes performs discretization of the feature description in the training matrix $\{x_{r,i}^{(j)} \mid i=\overline{1,N};\, j=\overline{1,n_r};\, r=\overline{1,R}\}$ by comparing the value of the *i*-th feature with the corresponding boundaries of the *l*-th one-dimensional receptive field. That is, the formation of binary training matrix $\{b_{r,i}^{(j)} \mid i=\overline{1,N\cdot L};\quad j=\overline{1,n_r};\quad r=\overline{1,R}\}$ is carried out by rule

$$b_{r,\,l\cdot N+i}^{(j)}=\begin{cases}1, \ \mathit{if} \quad x_{i,\max}\left[1-\delta_{i,l}/\delta_{\max}\right]\le E_{r,i}^{(j)}\le x_{i,\max};\\ 0, \ \text{otherwise.}\end{cases}$$

To compute coordinates of binary support vector x, relative to which the class containers in the radial basis is constructed, uses rule

$$b_{r,l\cdot N+i}=\begin{cases}1, \quad \text{if} \quad \dfrac{1}{n_r}\sum_{j=1}^{n_r} b_{r,l\cdot N+i}^{(j)}>\dfrac{1}{R}\sum_{A=1}^{R}\dfrac{1}{n_A}\sum_{j=1}^{n_A} b_{A,l\cdot N+i}^{(j)};\\ 0, \ \text{otherwise.}\end{cases}$$

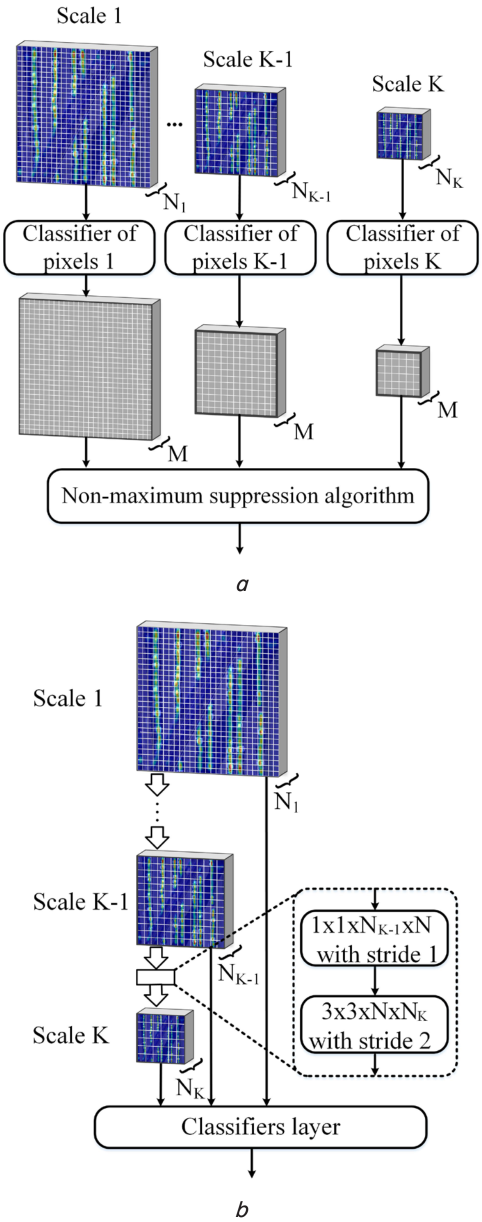

Fig. 1. Generalized architecture of the detector: *a* — the scheme of the object detector from the set of M classes; *b* — the scheme of formation of multi-scale feature maps, each pixel of which is described by $N_1,..., N_{k+1}$ features

In the framework of IEI technology, to increase learning efficiency, it is common to reduce the problem of the multi-class classification to the series of the two-class one by the principle "one-against-all". In this case, to avoid the problems of class imbalance, due to the majority of negative background examples in datasets, a synthetic class is an alternative to the *r*-class. The synthetic class is represented by n_r samples of the remaining classes, which are the closest to support vector b_r, where n_r is the volume of the training dataset of the *r*-class.

Training efficiency criterion of classifier is considered the normed modification of the S. Kullback's information measure [11]:

$$E_r = \frac{1-(\alpha_r+\beta_r)}{\log_2(2+\varsigma)-\log_2\varsigma}\cdot\log_2\left[\frac{2-(\alpha_r+\beta_r)+\varsigma}{(\alpha_r+\beta_r)+\varsigma}\right], \qquad (1)$$

where α_r, β_r are the false-positive and false-negative rates of classification decisions regarding belonging of input vectors to the *r*-th class; ς is any small non-negative number, which is introduced to avoid uncertainty at dividing by zero.

Membership function of binary representation *b* of input vector *x* to *r*-th class, the optimal container of which has support vector b_r^* and radius d_r^*, is derived from formula:

$$\mu_r(b)=\exp\left(-\sum_{i=1}^{N\cdot L} b_i\oplus b_{r,i}^*/d_r^*\right).$$

To optimize the vector of the parameters of receptive fields of classifier $\{\delta_{i,l}\,|\,i=\overline{1,N};\,l=\overline{1,L}\}$, of scaling filters of the detector and correction of the convolutional filters of the last feature extraction layer, it is proposed to use the metaheuristic simulated annealing algorithm [4].

The complex criterion *J* is used as the optimization criterion for training of a scaling filter and the detector classifiers and fine tuning of the last feature extractor layer. This criterion takes into account both effectiveness J_{Loc} of the object detection on the image, and effectiveness of classification analysis J_{Cls}.

$$J=J_{Loc}\cdot J_{Cls}. \qquad (2)$$

It is proposed to calculate the criterion of evaluation of effectiveness of detection of objects of interest in the image from formula

$$J_{Loc}=\frac{1}{n_l}\sum_{i=1}^{n_g}\sum_{j=1}^{n_d}x_{i,j}\cdot\left(\max_r\{\mu_{j,r}\}-\mu_{j,0}\right),$$

where n_g is the number of bounding boxes of objects of interest on the labeled training images of a domain area; n_d is the number of pixels of multi-scaled feature maps of an objects detector; $x_{i,j}$ is the indicator of matching the *j*-th pixel of the feature map of the detection layer to the *i*-th bounding box of the object of interest on labeled images

$$x_{i,j}=\begin{cases}1, \text{if } IoU_{i,j}>Th;\\ 0, \text{otherwise},\end{cases}$$

where $IoU_{i,j}$ is the Jaccard coefficient, which measure of intersection of the *i*-th bounding box of the object of interest on labeled images with the projection onto the input image of the *j*-the pixel of the feature map of the detection layer; n_1 is the number of pixels of the detection feature map, in which the measure of intersection of the projection onto the input image with the bounding box of the object of interest is more than threshold value *Th*; $\mu_{j,r}$ is the function of the *j*-th pixel of the feature map to the *r*-th class, where $r=0$ corresponds to the background class.

It is proposed to calculate the criterion of object classification effectiveness from formula

$$J_{Cls}=\frac{1}{S}\frac{1}{R}\sum_{s=1}^{S}\sum_{r=1}^{R}E_{r,s},$$

where *S* is the number of multiscale feature maps of detection layers; $E_{m,s}$ is the information criterion of effectiveness of machine learning of the *s*-th classifier to recognize the *r*-th class.

The non-maximum suppression algorithm and its modifications are used for filtering unnecessary actions of the detector to one and the same image object. It is proposed to apply the most widely used greedy non-maximum suppression algorithm. The algorithm iteratively finds detection with the locally highest value of membership function to an object of interest and rejects all adjacent detections, overlapping it with more than the threshold value.

Thus, the proposed model of the objects detector on terrain is intended not for accurate localization of objects in the area, but for the most accurate determining the existence of an object of interest for a given area.

4. 2. Algorithm for training the objects detector

To solve the problem of the development of the training algorithm for object detection, we will state the main training stages of the detector and consider the algorithms of their implementation. At the first stage, it is necessary to select a previously trained deep neural network and the number of its low-level layers that will be borrowed according to the transfer learning technique. A network and the number of its layers can be selected by brute force or based on experience.

At the second stage of the algorithm, it is proposed to carry out training of the high-level layers of a neural network using a growing sparse coding neural gas, based on the principles of neural gas and sparse coding. In this case, the dataset for training high-level filters of the convolutional network is formed by partitioning input images or feature maps into the 3D patches that coincide with the dimensions of the filters of a given layer. These patches reshape to 1D vectors, arriving at the input of the algorithm of growing sparse coding neural gas, the basic stages of which are given below:

1) initialization of the counter of learning vectors t:=0;

2) two initial nodes (neurons) w_a and w_b are assigned by random choice out of the training dataset. Nodes w_a and w_b are connected by the edge, the age of which is zero. These nodes are considered non-fixed;

3) the following vector x, which is set to the unit length (L2–normalization), is selected;

4) set each basis vector $w_k, k=\overline{1,M}$ to unit length (L2–normalization);

5) calculation of the measure of similarity of input vector x to basis vectors $w_{s_k} \in W$ for sorting

$$-(w_{s_0}^T x)^2 \leq \ldots \leq -(w_{s_k}^T x)^2 \leq \ldots \leq -(w_{s_{M-1}}^T x)^2;$$

6) the nearest node w_{s0} and the second by proximity node w_{s1};

7) to increase by unity the age of all incidents w_{s0};

8) if w_{s0} is fixed, we proceed to step 9, otherwise – to step 10;

9) if $(w_{s0}^T x)^2 \geq v$, we proceed to step 12. Otherwise, we add new non-fixed neuron w_r to the point that coincides with input vector w_r=x, besides, a new edge that connects w_r and w_{s0} is added, then we proceed to step 13;

10) node w_{s0} and its topological neighbors (the nodes, connected with it by the edge) are displaced in the direction to input vector x according to the Oja's rule [14] by formulas

$$\vartriangle w_{s0} = \varepsilon_b \eta_t y_0 (x - y_0 w_{s0}), \quad y_0 := w_{s0}^T x,$$

$$\vartriangle w_{sn} = \varepsilon_n \eta_t y_n (x - y_n w_{sn}), \quad y_n := w_{sn}^T x,$$

$$0 < \varepsilon_b \ll 1, \quad 0 < \varepsilon_n \ll \varepsilon_b,$$

$$\eta_t := \eta_0 (\eta_{final} / \eta_0)^{t/t_{\max}},$$

where Δw_{s0}, Δw_{sn} are the vector of correction of weight coefficients of the neuron-winner and its topological neighbors, respectively; ε_b, ε_n are the constants of the update force of weight coefficients of the neuron-winner and its topological neighbors, respectively; η_0, η_t, η_{final} are the initial, current and final value of learning rate, respectively;

11) if $(w_{s_0}^T x)^2 \geq v$, we label neuron w_{s0} as fixed;

12) if w_{s0} and w_{s1} are connected by the edge, its age is nulled, otherwise, a new edge with the zero age is formed between w_{s0} and w_{s1};

13) all edges in the graph with the age of more than $a_{\max}$ are removed. In the case when some nodes do not have any incident edges (become isolated), these nodes are also removed;

14) if t<$t_{\max}$, proceed to step 15, otherwise– increment of the step counter t:=t+1 and proceed to step 3;

15) if all neurons are fixed, the algorithm implementation stops, otherwise, proceed to step 3 and a new learning epoch begins (repetition of training dataset).

At stage 3, it is necessary to find optimal parameters of functioning of the detection system. The optimization algorithm is particularly important for the procedure of fine tuning of the feature extraction layer, which adapts to the domain area since unsupervised learning does not take into account non-balanced patches, which match the background and objects of interest.

It is proposed to choose simulated annealing among the metaheuristic search optimization algorithms. The efficiency of the simulated annealing algorithm depends on the implementation of the create_neighbor_solution procedure, forming a new solution s_i on the i-th iteration of the algorithm. Fig. 2 shows a pseudocode of the simulated annealing algorithm, which is implemented by the epochs_max iterations, on each of which function f() is calculated by passing a labeled training dataset t0hrough the model of the system of detection and calculation of a complex criterion (2).

```
$s_{current} \leftarrow create\_initial\_solution()$
$s_{best} \leftarrow s_{current}$
$T \leftarrow T_0$
$A \leftarrow \varepsilon,\ 0 < \varepsilon < 1$
$for(i = 1\ to\ epochs\_\max)$
    $s_i \leftarrow create\_neighbor\_solution(s_{current})$
    $if\ f(s_i) \geq f(s_{current})$
        $s_{current} \leftarrow s_i$
        $if\ f(s_i) \geq f(s_{best})$
            $s_{best} \leftarrow s_i$
        $end\ if$
    $elseif\ \exp\left(\frac{f(s_{current}) - f(s_i)}{T}\right) > uniform\_random(0,1)$
        $s_{current} \leftarrow s_i$
    $end\ if$
    $T \leftarrow c \times T$
$end\ for$
$return(s_{best})$
```

Fig. 2. Pseudocode of metaheuristic simulated annealing algorithm

An analysis of the pseudocode in Fig. 2 shows that current solution $s_{current}$, in relation to which new best solutions s_{best} are sought for, is updated in case of providing a new solution of the criterion increase (2), or randomly from the Gibbs distribution [16]. In this case, an initial search point that is formed by the create_initial_solution procedure can be either randomly generated or a result of the preliminarily training by another algorithm. To generate new solutions in the create_neighbor_solution procedure, it is proposed to use the simplest non-adaptive algorithm, which can be represented as formula:

$$s_{\text{current}} = s_{\text{current}} + \text{uniform\_random}(-1,1)\cdot \text{step\_size},$$

where uniform_random is the function of generation of a random number from the uniformed distribution from the assigned range; step_size is the size of a range of the search for new solutions, neighboring to s_{current}.

Thus, the proposed detector training algorithm lies in adaptation of the upper layers of the model to the domain application area based on the algorithms of growing sparse coding neural gas and simulated annealing.

5. Results of machine learning of the system for detection of objects on aerial video monitoring images

To train the detector of objects in the field of view of an UAV, nearly 180 images from dataset of Inria Aerial Image Labeling Dataset were used [18]. Each image has the resolution of 5,000×5,000 pixels. 2,000 unlabeled images of the size of 224×224 pixels were generated through random crop with rotation, as well as 200 labeled images of size of 224×224 pixels, which were multiplied up to 800 labeled images by putting noise, a contrast change and rotation.

A large number of vehicles in the urban area are presented in the images of the dataset Inria Aerial Image Labeling Dataset. That is why it is proposed to select the means of transport as the objects of interest and to consider the urban area as a domain application area. In this case, the class recognition alphabet equal to R=3, where the first class corresponds to cars, the second class corresponds to freight cars, and the third one – to the background. The size of objects in pixels in random images varies in the range of [16×16,..., 32×32].

It is proposed to carry out transfer learning technique by copying the layers from a pretrained convolutional neural network Squeezenet, which has a high computational efficiency and is popular in mobile systems with limited resources [5, 6]. As a result, each input image is encoded into the feature map at the output of module fire7 of network Squeezenet. In this case, the size of the feature map is 13×13×384. The next layer is trained unsupervised on the unlabeled datasets of the domain area and consists of filters with the kernels of 1x1 and 3×3 with the scan step stride=1. The next layer is intended to increase the scale of the feature map with minimal information losses and has the structure shown in Fig. 1, *a*. The last layer is also pretrained unsupervised. Correction of the layers trained unsupervised and training classifiers are carried out based on the labeled training datasets.

It is proposed first to train the detector using the unsupervised pretraining feature extractor via growing sparse coding neural gas without fine tuning of the network. In this case, a fixed value of the training dataset reconstruction ν=0.8 is used during training. In addition, The OMP algorithm is not used when calculating the activation of the feature map pixels. In the information-extreme classifier of the feature map pixel, the number of thresholds for discretization of feature description is equal to L=3. In the simulated annealing algorithm for training classifiers, the following parameters were used: c=0.98, T_0=10, epochs_max=500, step_size=0.001. Table 1 shows the results of machine learning of the objects detectors at different values of threshold *Th* and the measure of intersection of projection of the feature map pixel with the bounding box from training image markup for inclusion of a pixel to the training dataset of objects of interest.

An analysis of Table 1 shows that the optimal threshold value is equal to Th^{*}=0.4, which provides detection of 69 % of the objects on the test images. In this case, a small threshold value *Th* leads to a decrease in detection efficiency as a result of an increase in false positive rate. Large values of threshold *Th* lead to a decrease in the number of training datasets of each class and to ignoring the objects that are at the boundaries of some feature map pixels projections.

In order to improve the results of machine learning of the detector, informativeness of feature description is increased by fine tuning of the unsupervised trained convolutional layers. In this case, for high-level layers of the feature description formation, we consider the case of both the absence and the presence of OMP algorithm with the ReLU function for calculation of each feature map pixel activation. Table 2 shows the results of machine learning when using fine tuning of the extraction layers, based on simulated annealing.

Table 1

Results of machine learning of the detector at different values of threshold the of overlapping pixel projections with object's bounding box from labeled image in order to refer a pixel to object dataset

Th	J_{Cls}	J_{Loc}	J	Proportion of detected objects on the test dataset
0.1	0.210	0.090	0.0200	0.610
0.2	0.250	0.120	0.0300	0.630
0.3	0.250	0.150	0.0360	0.660
0.4	0.301	0.150	0.0450	0.690
0.5	0.200	0.150	0.0300	0.630
0.6	0.130	0.170	0.0220	0.630
0.7	0.090	0.100	0.0090	0.500
0.8	0.087	0.100	0.0087	0.420
0.9	0.050	0.100	0.0050	0.330

Table 2

Results of machine learning of detector with fine tuning of the feature extractor based on simulated annealing

Use of OMP	J_{Cls}	J_{Loc}	J	Proportion of detected objects on the test dataset
–	0.41	0.360	0.1476	0.91
+	0.52	0.450	0.2340	0.95

An analysis of Table 2 shows that fine tuning of the feature extractor led to an increase in accuracy of object detection on the images of terrain. In this case, the use of the OMP algorithm for calculation of activation of the feature map pixels has also a noticeable effect, since it adds 4 % accuracy of object detecting on the test images.

Thus, the developed algorithm of training decision rules of the object detector makes it possible to obtain the results, acceptable for practical use in the domain area, without the need for learning models from scratch. In this case, fine tuning of high-level layers based on the simulated annealing algorithm makes it possible to improve the outcomes of leaning the detector.

6. Discussion of the results of machine learning of the objects detector

Analysis of the results of machine learning indicates that prior unsupervised learning on the unlabeled datasets, as a rule, simplifies the further process of supervised machine learning. That is why it is worth considering the effect of the parameters of the algorithm of growing sparse coding neural gas, which is used for prior unsupervised learning on the results of supervised machine learning. Table 3 shows dependence of results of machine learning and the number of N_c generated convolutional filters (neurons) on the value of parameter ν, which characterizes the accuracy of coverage of the training dataset with convolutional filters.

Table 3

Results of machine learning of the detector at different values of parameter of training dataset coverage at unsupervised learning

ν	N_c	J_{Cls}	J_{Loc}	J	Proportion of detected objects on the test dataset
0.4	196	0.0912	0.100	0.00912	0.67
0.5	337	0.1502	0.101	0.01517	0.73
0.6	589	0.2508	0.401	0.10057	0.87
0.7	889	0.4203	0.430	0.18073	0.93
0.8	1519	0.5201	0.450	0.23401	0.95
0.9	4058	0.5203	0.450	0.23416	0.92

Analysis of Table 3 shows that at an increase in the value of parameter ν, the number of neurons and the value of a particular and a complex optimization criterion increases (2). However, at ν≤0.8, accuracy of the model for test dataset grows with an increase in parameter ν, and a subsequent increase in the parameter leads to a deterioration of results due to the overfitting effect. That is why parameter ν can be considered the regularization parameter of high-level layers of the feature extractor.

To understand the benefits of using prior unsupervised learning, we will consider the results of machine learning using simulated annealing before and after learning by the algorithm of growing sparse coding neural gas. Fig. 3 shows the dependence of the efficiency criterion of training the detector (2) by the simulated annealing algorithm with parameters c=0.998, T_0=10, epochs_max=5,000, step_size=0.001.

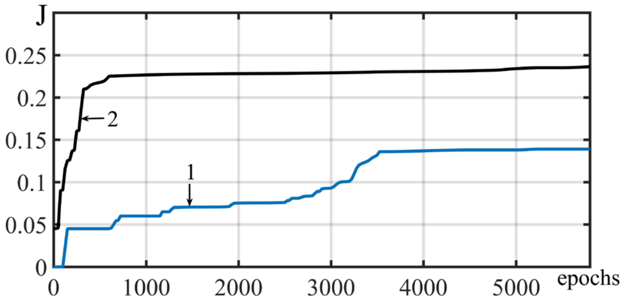

Fig. 3. Diagrams of dependence of the optimization criterion (2) on the number of learning epochs:
1 — before application of prior unsupervised learning;
2 — after application of prior unsupervised learning

Analysis of Fig. 3 shows that prior unsupervised learning based on growing sparse coding neural gas makes it possible to improve the final result of the supervised learning using the simulated annealing algorithm. In this case, if there was unsupervised pretraining, a global maximum of criterion (2) is achieved by more than 10 times faster. In addition, validation the obtained models on the test dataset shows that the use of unsupervised pretraining makes it possible to decrease the overfitting effect under conditions of limited volume of labeled training dataset. When using unsupervised pretraining, criterion (2), calculated for training dataset is equal to J_{train}=0.234, and for the test dataset – J_{test}=0.228. Without using pretraining, corresponding criteria significantly differ from each other, J_{train}=0.1321 and J_{test}=0.081.

Thus, the proposed algorithm of unsupervised pretraining of high-level layers allows increasing the value of the learning criterion and the proportion of detected objects on the test images. In addition, we managed to decrease the overfitting effect and to increase the rate of finding a global maximum at supervised training on the labeled training dataset of limited size. However, the mechanism of refining the boundaries of detected objects were not considered in the framework of this study, neither was dependence of the criterion of learning effectiveness on the parameters of the simulated annealing algorithm. That is why the following studies will be focused on the improvement of the model of detector and the development of algorithms of setting parameters of search optimization algorithm in the course of machine learning.

7. Conclusions

1. The model of the object detector and the training effectiveness criterion of the model were proposed. The model contains 7 first modules of the computationally effective convolutional Squeezenet network, two convolutional layers of different scales, and the information-extreme classifier. The multiplicative convolution of the particular criteria that takes into account the effectiveness of object detection on images and accuracy of the classification analysis was considered as complex criterion of learning effectiveness of the model. In this case, additional use of the orthogonal matching pursuit algorithm in calculating high-level feature maps makes it possible to increase the accuracy of the model by 4 %. This model provides acceptable for practical use reliability of detection of the objects, the size of which in pixels is by 7...14 times smaller than the size of the smallest side of the aerial image.

2. The three-stage training algorithm of the object detector under conditions of a small size of labeled training datasets and limited computing resources available on board of a compact UAV was proposed. At the first stage, it is proposed to select the number of the low-level convolutional layers of the deep neural network that is pretrained on the ImageNet dataset. The second stage involves unsupervised learning of the high-level layers of the neural network on unlabeled dataset. In this case, the algorithm of growing sparse coding neural gas for unsupervised training of convolutional filters was developed. Its application makes it possible to utilize the unlabeled training datasets for the adaptation of the high-level feature description to the domain application area and to determine the number of neurons. At the third stage, there is training of the information-extreme classifier

of pixels of the high-level feature maps and fine tuning of the upper convolutional layers by using the metaheuristic simulated annealing algorithm. The results of physical simulation on the Inria Aerial Image Labeling dataset proved the effectiveness of the developed algorithm of unsupervised learning of convolutional filters. In this case, in the absence of fine tuning of convolutional filters, 69 % detection of objects on the images of the test dataset is provided. After fine tuning of high-level convolutional filters based on the simulated annealing algorithm, 95 % detection of objects in the test dataset is provided.

3. It was shown that the use of prior unsupervised learning makes it possible to decrease the overfitting effect and to increase by more than 10 times the rate of finding the global maximum at supervised training on dataset of limited size. In this case, on the test dataset, it was possible virtually to eliminate the overfitting effect by means of the optimal selection of the value of hyperparameter, which characterizes the measure of coverage of training datasets by the network neurons.

Acknowledgements

The research was performed at the laboratory of intelligent systems of the Department of Computer Sciences at Sumy State University with financial support from MES of Ukraine within the framework of the state-funded scientific research work DR No. 0117U003934.

References

1. Patricia N., Caputo B. Learning to Learn, from Transfer Learning to Domain Adaptation: A Unifying Perspective // 2014 IEEE Conference on Computer Vision and Pattern Recognition. 2014. doi: https://doi.org/10.1109/cvpr.2014.187
2. Nguyen A., Yosinski J., Clune J. Deep neural networks are easily fooled: High confidence predictions for unrecognizable images // 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). 2015. doi: https://doi.org/10.1109/cvpr.2015.7298640
3. Optimization of convolutional neural network using microcanonical annealing algorithm / Ayumi V., Rere L. M. R., Fanany M. I., Arymurthy A. M. // 2016 International Conference on Advanced Computer Science and Information Systems (ICACSIS). 2016. doi: https://doi.org/10.1109/icacsis.2016.7872787
4. Learned vs. Hand-Crafted Features for Pedestrian Gender Recognition / Antipov G., Berrani S.-A., Ruchaud N., Dugelay J.-L. // Proceedings of the 23rd ACM international conference on Multimedia – MM '15. 2015. doi: https://doi.org/10.1145/2733373.2806332
5. A Review of Deep Learning Methods and Applications for Unmanned Aerial Vehicles / Carrio A., Sampedro C., Rodriguez-Ramos A., Campoy P. // Journal of Sensors. 2017. Vol. 2017. P. 1–13. doi: https://doi.org/10.1155/2017/3296874
6. Scaling for edge inference of deep neural networks / Xu X., Ding Y., Hu S. X., Niemier M., Cong J., Hu Y., Shi Y. // Nature Electronics. 2018. Vol. 1, Issue 4. P. 216–222. doi: https://doi.org/10.1038/s41928-018-0059-3
7. DroNet: Learning to Fly by Driving / Loquercio A., Maqueda A. I., del-Blanco C. R., Scaramuzza D. // IEEE Robotics and Automation Letters. 2018. Vol. 3, Issue 2. P. 1088–1095. doi: https://doi.org/10.1109/lra.2018.2795643
8. An Improved Transfer learning Approach for Intrusion Detection / Mathew A., Mathew J., Govind M., Mooppan A. // Procedia Computer Science. 2017. Vol. 115. P. 251–257. doi: https://doi.org/10.1016/j.procs.2017.09.132
9. Qassim H., Verma A., Feinzimer D. Compressed residual-VGG16 CNN model for big data places image recognition // 2018 IEEE 8th Annual Computing and Communication Workshop and Conference (CCWC). 2018. doi: https://doi.org/10.1109/ccwc.2018.8301729
10. Nakahara H., Yonekawa H., Sato S. An object detector based on multiscale sliding window search using a fully pipelined binarized CNN on an FPGA // 2017 International Conference on Field Programmable Technology (ICFPT). 2017. doi: https://doi.org/10.1109/fpt.2017.8280135
11. Development of the method of features learning and training decision rules for the prediction of violation of service level agreement in a cloud-based environment / Moskalenko V., Moskalenko A., Pimonenko S., Korobov A. // Eastern-European Journal of Enterprise Technologies. 2017. Vol. 5, Issue 2 (89). P. 26–33. doi: https://doi.org/10.15587/1729-4061.2017.110073
12. Feng Q., Chen C. L. P., Chen L. Compressed auto-encoder building block for deep learning network // 2016 3rd International Conference on Informative and Cybernetics for Computational Social Systems (ICCSS). 2016. doi: https://doi.org/10.1109/iccss.2016.7586437
13. Aircraft Detection by Deep Convolutional Neural Networks / Chen X., Xiang S., Liu C.-L., Pan C.-H. // IPSJ Transactions on Computer Vision and Applications. 2014. Vol. 7. P. 10–17. doi: https://doi.org/10.2197/ipsjtcva.7.10
14. Labusch K., Barth E., Martinetz T. Sparse Coding Neural Gas: Learning of overcomplete data representations // Neurocomputing. 2009. Vol. 72, Issue 7-9. P. 1547–1555. doi: https://doi.org/10.1016/j.neucom.2008.11.027
15. Mrazova I., Kukacka M. Image Classification with Growing Neural Networks // International Journal of Computer Theory and Engineering. 2013. P. 422–427. doi: https://doi.org/10.7763/ijcte.2013.v5.722
16. Palomo E. J., Lopez-Rubio E. The Growing Hierarchical Neural Gas Self-Organizing Neural Network // IEEE Transactions on Neural Networks and Learning Systems. 2016. P. 1–10. doi: https://doi.org/10.1109/tnnls.2016.2570124
17. Rere L. M. R., Fanany M. I., Arymurthy A. M. Metaheuristic Algorithms for Convolution Neural Network // Computational Intelligence and Neuroscience. 2016. Vol. 2016. P. 1–13. doi: https://doi.org/10.1155/2016/1537325
18. High-Resolution Aerial Image Labeling With Convolutional Neural Networks / Maggiori E., Tarabalka Y., Charpiat G., Alliez P. // IEEE Transactions on Geoscience and Remote Sensing. 2017. Vol. 55, Issue 12. P. 7092–7103. doi: https://doi.org/10.1109/tgrs.2017.2740362