DOI: https://doi.org/10.15276/aait.03.2020.2
UDC 004.931

INTELLIGENT SYSTEM BASED ON A CONVOLUTIONAL NEURAL NETWORK FOR IDENTIFYING PEOPLE WITHOUT BREATHING MASKS

Oleksii I. Sheremet[1)]
ORCID: https://orcid.org/0000-0003-1298-3617, sheremet-oleksii@ukr.net
Oleksandr Ye. Korobov[2)]
ORCID: https://orcid.org/0000-0003-4530-7535, korobov.alex@gmail.com
Oleksandr V. Sadovoi[3)]
ORCID: https://orcid.org/0000-0001-9739-3661, sadovoyav@ukr.net
Yuliia V. Sokhina[3)]
ORCID: https://orcid.org/0000-0002-4329-5182, jvsokhina@gmail.com
[1)]Donbas State Engineering Academy, 39, Mashinobudivnykiv Blvd. Kramatorsk, 84313, Ukraine
[2)]AI-labs, 2, Maidan Nezalezhnosti Sq. Kyiv, 01001,Ukraine
[3)]Dnipro State Technical University, 2, Dniprobudivska Str, Kamyanske, 51918, Ukraine

ABSTRACT

The COVID-19 pandemic is having a huge impact on people and communities. Many organizations face significant disruptions and issues that require immediate response and resolution. Social distancing, breathing masks and eye protection as preventive measures against the spread of COVID-19 in the absence of an effective antiviral vaccine play an important role. Banning unmasked shopping in supermarkets and shopping malls is mandatory in most countries. However, with a large number of buyers, the security is not able to check the presence of breathing masks on everyone. It is necessary to introduce intelligent automation tools to help the work of security. In this regard, the paper proposes an up-to-date solution – an intelligent system for identifying people without breathing masks. The proposed intelligent system works in conjunction with a video surveillance system. A video surveillance system has a structure that includes video cameras, recorders (hard disk drives) and monitors. Video cameras shoot sales areas and transmit the video image to recording devices, which, in turn, record what is happening and display the video from the cameras directly on the monitor. The main idea of the proposed solution is the use of an intelligent system for classifying images periodically received from cameras of a video surveillance system. The developed classifier divides the image stream into two classes. The first class is “a person in a breathing mask” and the second is “a person without a breathing mask”. When an image of the second class appears, that is, a person who has removed a breathing mask or entered a supermarket without a breathing mask, the security service will immediately receive a message indicating the problem area. The intelligent system for image classification is based on a convolution neural network VGG-16. In practice, this architecture shows good results in the classification of images with great similarity. To train the neural network model, the Google Colab cloud service was used – this is a free service based on Jupyter Notebook. The trained model is based on an open source machine learning platform TensorFlow. The effectiveness of the proposed solution is confirmed by the correct processing of the practically obtained dataset. The classification accuracy is up to 90 %.

Keywords: Intelligent System; VGG-16; Convolution Neural Network; Tensorflow; Image Classification; Accuracy; Loss Function; Machine Learning

For citation: **Oleksii I. Sheremet, Oleksandr Ye. Korobov, Oleksandr V. Sadovoi, Yuliia V. Sokhina. Intelligent System Based on a Convolutional Neural Network for Identifying People without Breathing Masks. *Applied Aspects of Information Technology.* 2020; Vol.3 No.3: 133–134. DOI: https://doi.org/10.15276/aait.03.2020.2**

INTRODUCTION

The current situation with COVID-19 requires the implementation of specific technical solutions in the security field. One of such solutions is the video surveillance systems intellectualization.

New requirements are being introduced to such systems related to identifying signs of quarantine violations. Thus, violation of the mask regime in public places should be suppressed by security services. In this regard, video surveillance systems in large supermarkets and shopping malls in the leading countries of the world are beginning to be modernized. The intelligent systems for identifying people without masks can significantly help security guards who monitor trading floors. A video story the board or a photos stream incoming in real time at intelligent system input is classified automatically.

All input images are divided into two classes. The first class is “a person in a breathing mask” and the second is “a person without a breathing mask”. When an image of the second class appears, that is, a person who has removed a breathing mask or entered a supermarket without a breathing mask, the security service will immediately receive a message indicating the problem area. Thus, the task of an intelligent identification system is reduced to the image classification task in real time.

In the context of machine learning, classification can be referred to the supervised learning. This type of training implies that the data sent to the system inputs is already labeled, that is, the images are already divided into separate categories or classes. In image classification problem, Convolutional Neural Networks (CNNs) [1–3] have achieved particular success, which have repeatedly won the ImageNet Large Scale Visual Classification Challenge (ILSVRC) [4–5]. In addition to classification task, CNNs are used for pattern recognition [6–7], object detection and tracking [8–9], semantic segmentation [10–11] and other tasks [12–14].

The main features of the modern convolutional neural networks architecture were laid in the famous convolutional neural network – LeNet-5 [15]. According to the classical approach, an image of 32×32×1 pixels is fed to the network input (*Fig. 1*). Naturally, modern CNNs process color images with three layers (usually RGB: red, green and blue).

In CNNs, convolution and subsampling layers are made up of multiple neurons layers called feature maps or channels. Convolutional layers work on the kernels basis or filters that deal with the certain image features recognition. The filter moves along the image and determines whether some desired feature is present in a particular part of it. To obtain an answer of this kind, a convolution operation is performed, which is the sum of the products of the filter elements and the input signals matrix (*Fig. 2*). Subsampling is performed to speed up the learning process and reduce the computing resources consumption. Most often, subsampling is performed using the max pooling operation. The max pooling means moving the window along the matrix with data. From the pixels falling into its field of view, the maximum is selected and moved to the resulting matrix (*Fig. 3*). With the help of max pooling layers, insensitivity to the small input image distortions is achieved, as well as the dimension of subsequent layers is reduced [16].

The last CNNs layers are several fully connected layers. These are some of the simplest layers, in which every neuron in one layer is connected to every neuron in another layer. Thus, an image (most often a three-layer – RGB) is fed to the neural network input, and the output is a class to which the image belongs. Often, the number of the neural network outputs corresponds to the classes' number.

In the problem under consideration, a photo of a trading floor section taken in real time is fed to the neural network input. Two outputs are enough. The maximum signal at the first output should be set when there is a masked person on the image (or if there are no people in the frame), and at the second – if there is a person without a breathing mask.

The purpose of this paper is to obtain the trained model of a Convolutional Neural Network (CNN) for identifying people without breathing masks acceptable prediction accuracy.

To achieve the goal, it is necessary to solve the following tasks:

1. Choose a CNN architecture suitable for classifying images with human faces.

2. Collect and label the dataset required to train and test the model.

3. Train a CNN model and propose the intelligent system scheme for identifying people without breathing masks.

4. Perform testing on images that were not used in training process and evaluate the main quality factors of the resulting model.

1. BRIEF OVERVIEW AND CHOICE OF THE CNN ARCHITECTURE FOR IMAGE CLASSIFICATION

AlexNet is a convolutional neural network that has had a major impact on the development of machine learning, especially computer vision. This neural network won the ILSVRC in 2012 [17]. AlexNet collected the latest techniques at that time to improve network performance. So, it turned out that the use of the Rectified Linear Unit (ReLU) (*Fig. 4*) activation function instead of the more traditional sigmoid and hyperbolic tangent functions allowed reducing the learning epochs number by six times. ReLU is currently the most commonly used activation function as it overcomes the gradient fading problem inherent in other activation functions.

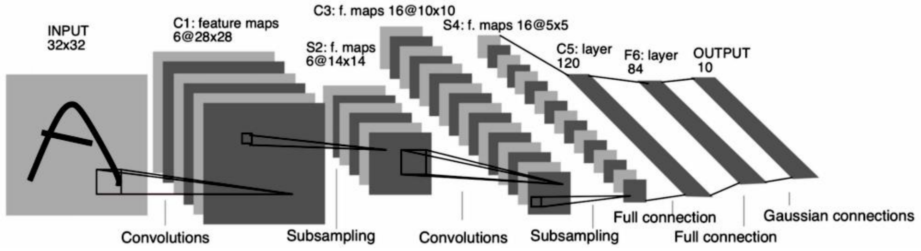


Fig. 1. **Architecture of LeNet-5**
Source: **compiled by the authors**

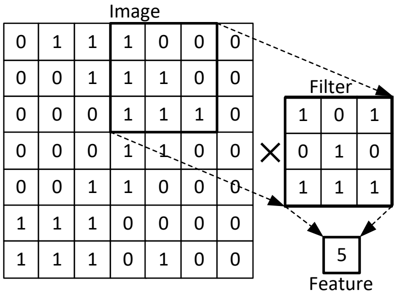


Fig. 2. **An example of convolution**
***Source:* compiled by the authors**

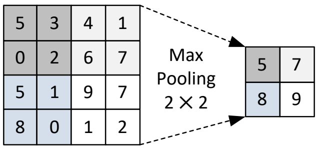


Fig. 3. **An example of max pooling**
***Source:* compiled by the authors**

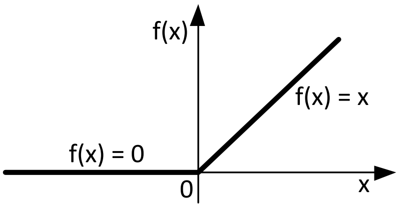


Fig. 4. **ReLU activation function**
***Source:* compiled by the authors**

Also, AlexNet uses such an important technique as dropout [18]. The essence of this technique is to randomly disable each neuron on a given layer with probability in each epoch. Dropout addresses one of the main deep neural networks problems – overfitting [19]. The essence of this technique is to randomly disable each neuron in a given layer with some probability in each epoch.

Another architecture that won the ILSVRC in 2013 is ZFNet [20]. The main innovation of this architecture is the filter visualization technique – a deconvolution network, consisting of operations inverse to convolution. An example of a deconvolution network is shown in Fig. 5 [21]. The ideas behind the ZFNet architecture have become a significant contribution to the CNNs development.

In 2014, work [22] proposed the architecture called VGG or VGGNet. The main and distinctive idea of this structure is to keep the filters as simple as possible. Therefore, all convolution operations are performed using a filter of size 3 and a step of 1, and all subsampling operations are performed using a filter of size 2 and a step of 2 (*Fig. 6*).

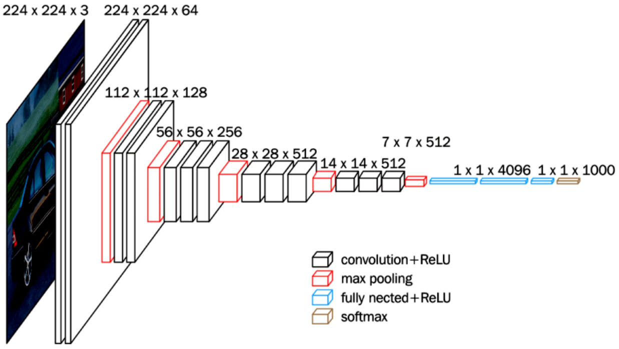


Fig. 6. **Architecture of VGG Net**
Source: compiled by the authors

The authors of work [22] showed that a layer with a 7×7 filter is equivalent to three layers with 3×3 filters, and in the latter case 55 % fewer parameters are used. Likewise, a 5×5 filter layer is equivalent to two 3×3 filter layers, which save 22 % of the network parameters.

Along with the convolutional modules simplicity, the network differs from previous solutions in a greater depth. The most important idea, first proposed in [22], is to overlay convolutional layers without subsampling layers. The overlay of convolutional layers provides a rather large receptive field, but the parameters number is much less than in networks with large filters.

VGG Net is used as part of more complex networks for object detection [23], semantic segmentation [24] and other tasks [25].

Thus, the main development of convolutional network architectures has been to simplify filters and increase network depth. In 2014, a different approach was proposed in [26], called the GoogleNet architecture. The inception module is one oGoogleNet main achievements (Fig. 7) [26].

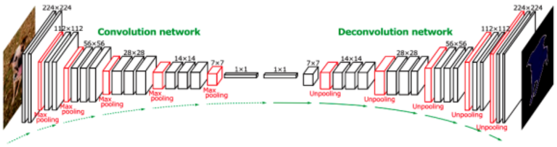


Fig. 5. **Deconvolution network**
***Source:* compiled by the authors**

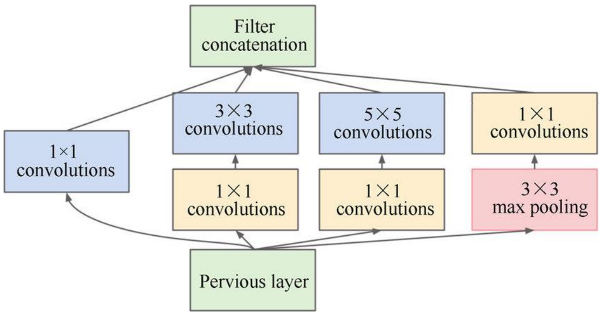


Fig. 7. **The inception module**
Source: **compiled by the authors**

The inception module uses several parallel branches that calculate different properties based on the same input data, and then merges the results. Another feature of this module is the use of 1×1 convolutional layers. As shown in [27], 1×1 convolution is an easy way to reduce the property map dimension.

The developers of the ResNet (residual network) network [28] noticed that with an increase in the number of layers, the convolutional neural network can begin to degrade. This training problem is called the vanishing gradient problem [29]. The crux of the problem is that a chain rule is used when working with back propagation method. If the gradient has a small value at the network's end, then it can take an infinitesimal value by the time it reaches the beginning of the network. This can lead to the impossibility of network training [30].

In [28], it was assumed that if a CNN has reached its accuracy limit on a certain layer, then all subsequent layers will have to degenerate into an identical transformation, but this does not happen due to the complexity of training deep networks. As a result of research [28], it was proposed to introduce the residual block. When it is used, the input data is passed over a shortcut, bypassing the transformation layers and added to the result (*Fig.* 8).

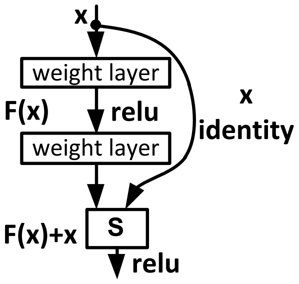


Fig. 8. **The residual block**
Source: **compiled by the authors**

After the ResNet effectiveness recognition, updated versions of the Inception network were presented: Inception-v4 and Inception-ResNet [31].

After analyzing the features of the presented CNN architectures, it was decided to use the VGG-16 architecture. Number 16 in the name VGG-16 refers to the fact that this has 16 layers that have some weights (the deeper VGG-19 architecture is also known [32]).

The VGG-16 advantage is its ease of implementation. One of the VGG-16 disadvantages was the slow learning speed. Currently, due to the computer technology development, this drawback has been eliminated. This architecture is well documented and is used to perform classification tasks of medium complexity. The VGG-16 training dataset does not have to be large.

2. COLLECT AND LABEL THE DATASET

In supervised learning, the neural network is trained on a labeled dataset and predicts responses, which are used to evaluate the accuracy of an algorithm on the training data. Data collection and labeling (in the classification case, dividing it into separate classes) is a very important step.

The dataset was obtained from the filming results from security cameras of the shopping malls in Tokyo. Since several people can be in the same frame, each photo was previously divided into small sections. Then the resulting images were saved as JPG files.

Thus, individual people must be detected before classification can be performed. OpenCV [33] Histograms of Oriented Gradients (HOG) [34] is used to detect people. This model is well known, so a pre-trained version of it was used. An example of an image obtained with a pre-trained model is shown in Fig. 9.

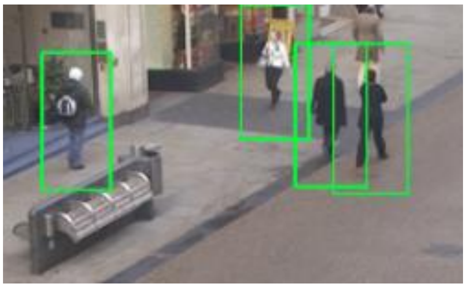


Fig. 9. **The residual block**
Source: **compiled by the authors**

Cutting of separate image parts is performed in accordance with the found bounding boxes. As shown in *Fig. 9*, individual people's bounding boxes overlap. This overlap makes it impossible to accurately separate individuals. However, for a VGG-16 based classifier, such noise will not pose a significant problem.

After automatically slicing the images along the bounding boxes, dataset labelers manually split all images into 2 folders. The first folder corresponds to the first class ("a person in a breathing mask"), the second – to the second class ("a person without a breathing mask"). Sample images assigned to each of the classes are shown in Fig. 10.

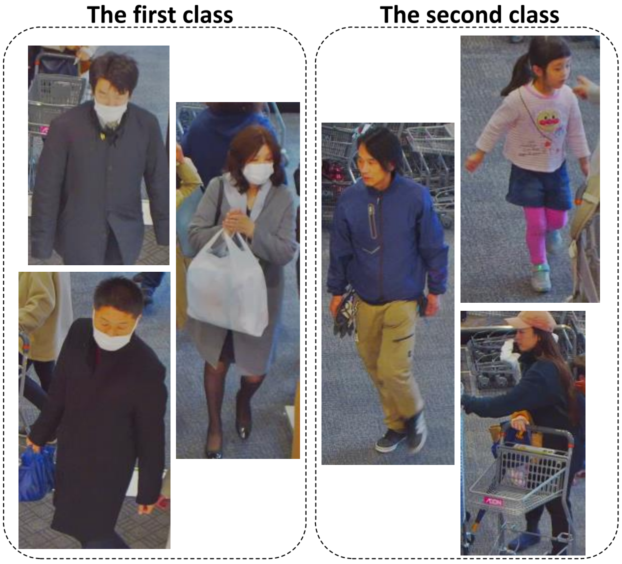


Fig. 10. **Sample images assigned to each of the classes**
***Source:* compiled by the authors**

In total, 2000 images for the training dataset were collected for each class. The validation and test datasets consist of 500 images each.

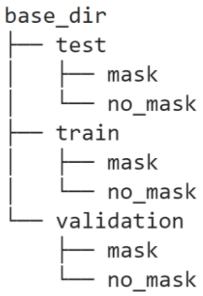
An intelligent identification system for people without breathing masks should contain a classifier based on a VGG-16. And its general functioning is provided according to a certain scheme.

3. AN INTELLIGENT SYSTEM FOR IDENTIFYING PEOPLE WITHOUT BREATHING MASKS

A key element of the intelligent system is the trained VGG-16 model. CPU is not suitable for training the VGG-16 in a reasonable amount of time. This task requires a powerful GPU or cloud services. In this regard, Google Colab is a convenient and free solution for machine learning problems. Google Colab is a cloud service based on Jupyter Notebook [35]. Google Colab provides everything for machine learning right in web browser and gives free access to fast GPUs.

The VGG-16 direct implementation is done using TensorFlow, an open source machine learning software library. Additionally, the Keras neural network library is used as an add-on over TensorFlow. This library greatly simplifies the developer's job as it is designed to be compact, modular, and extensible.

The training data was uploaded to Google Drive and divided into separate folders in accordance with the structure shown in Fig. 11.



```
base_dir
├── test
│   ├── mask
│   └── no_mask
├── train
│   ├── mask
│   └── no_mask
└── validation
    ├── mask
    └── no_mask
```

Fig. 11. **The structure of the data folders**
***Source:* compiled by the authors**

The scheme of layers connection for the VGG-16 network used to identify people without a breathing mask is shown in *Fig. 12*. In this figure, the following designations are adopted: Conv2d – 2D convolutional layers (this layer creates a convolution kernel that is convolved with the layer input to produce a tensor of outputs), Max_Pooling2d – max pooling layers for 2D spatial data, Flatten – input data flattering layer, Dense – regular densely-connected layers.

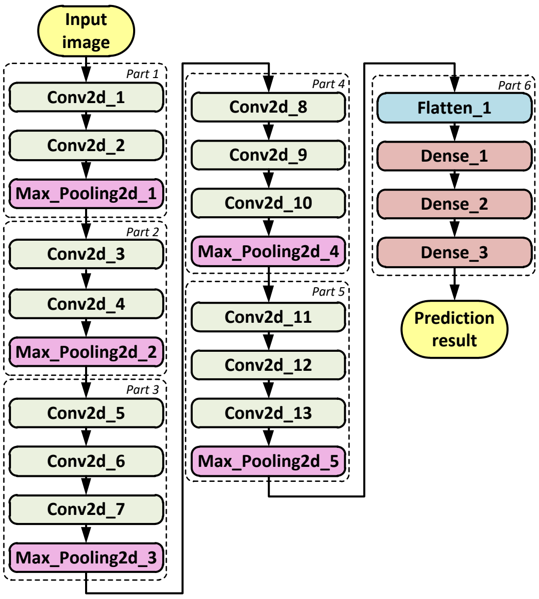


Fig. 12. **The scheme of layers connection for the VGG-16 network**
***Source:* compiled by the authors**

Conventionally, the VGG-16 neural network architecture can be divided into 6 parts:

– convolutional part 1, consisting of two convolutional layers Conv2d_1, Conv2d_2 and one subsampling layer Max_Pooling2d_1;

– convolutional part 2, consisting of two convolutional layers Conv2d_3, Conv2d_4 and one subsampling layer Max_Pooling2d_2;

– convolutional part 3, consisting of three convolutional layers Conv2d_5, Conv2d_6, Conv2d_7 and one subsampling layer Max_Pooling2d_3;

– convolutional part 4, consisting of three convolutional layers Conv2d_8, Conv2d_9, Conv2d_10 and one subsampling layer Max_Pooling2d_4;

– convolutional part 5, consisting of three convolutional layers Conv2d_11, Conv2d_12, Conv2d_13 and one subsampling layer Max_Pooling2d_5;

– flatten and dense part 6, consisting of one flatten layer Flatten_1 and three fully connected (dense in Keras) layers Dense_1, Dense_2, Dense_3.

Each input image is scaled to the dimension of the input convolutional window (224×224). In this case, scaling along the abscissa and ordinate axes can be performed with different coefficients, distorting the image. However, for the considered neural network, such distortion is not significant.

After passing the data through the five convolutional neural network parts, the result is flattened and represented as a vector of 25088 values (*Fig. 12*). The neural network has 2 outputs (according to the number of classes).

The summary representation shows the neural network features (Table 1). It uses a ReLU activation function to produce a probability distribution over the output classes. The total number of trainable parameters is 134268738.

Table 1. **Summary representation of the network**

Layer	Output shape	Parameters number
Conv2d_1	(224, 224, 64)	1792
Conv2d_2	(224, 224, 64)	36928
Max_pooling2d_1	(112, 112, 64)	0
Conv2d_3	(112, 112, 128)	73856
Conv2d_4	(112, 112, 128)	147584
Max_pooling2d_2	(56, 56, 128)	0
Conv2d_5	(56, 56, 256)	295168
Conv2d_6	(56, 56, 256)	590080
Conv2d_7	(56, 56, 256)	590080
Max_pooling2d_3	(28, 28, 256)	0
Conv2d_8	(28, 28, 512)	1180160
Conv2d_9	(28, 28, 512)	2359808
Conv2d_10	(28, 28, 512)	2359808
Max_pooling2d_4	(14, 14, 512)	0
Conv2d_11	(14, 14, 512)	2359808
Conv2d_12	(14, 14, 512)	2359808
Conv2d_13	(14, 14, 512)	2359808
Max_pooling2d_5	(7, 7, 512)	0
Flatten_1	(25088)	0
Dense_1	(4096)	102764544
Dense_2	(4096)	16781312
Dense_3	(2)	8194

Source: **compiled by the authors**

The image augmentation technique is a well-known approach for increasing the training dataset size. Even a small original dataset can be significantly expanded using augmentation.

In this work, Keras ImageDataGenerator is used to perform augmentation in automatic mode. Keras ImageDataGenerator provides a host of different augmentation techniques like standardization, rotation, shifts, flips, brightness change and other. However, the main benefit of Keras ImageDataGenerator is that it increases the dataset size in real time, right during training.

Three Python generators were created to perform augmentation: train generator, validation generator and test generator. Each of them generated a data stream for training, validation and testing, respectively.

When training VGG-16 network, the following parameters were used: optimizer='RMSprop', loss='categorical crossentropy', metrics='acc', learning rate=10^{-5}, epochs=20, steps per epoch=200. That is, the 'RMSprop' optimization algorithm is used.

RMSProp (Root Mean Square Propagation) is a method in which the learning rate is adapted for each of the parameters. The idea is to divide the learning rate for a weight by the moving average of the recent gradients for that weight.

The categorical crossentropy loss function is a very good measure of how distinguishable two discrete probability distributions are from each other.

Training in the Google Colab cloud service with GPU hardware acceleration lasted 8926 seconds. The learning curves obtained from the model training results are shown in Fig. 13 and Fig. 14.

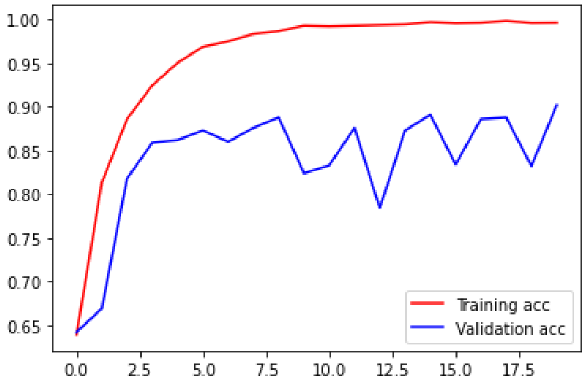


Fig. 13. **Training and validation accuracy**

Source: **compiled by the authors**

As follows from Fig. 13, at the end of the last (twentieth) epoch, the training accuracy is 0.9962 and validation accuracy is 0.9020. This means that 99.62 % of the images from the training dataset and 90.2 % – from the validation (unused for training) dataset are classified correctly. Such quality factors are suitable for practical application of the trained model.

At the end of the training, the loss function reaches 0.0105 in the training dataset and 0.0723 in the validation sample (Fig. 14). The peaks in the validation loss function are due to the use of dropout.

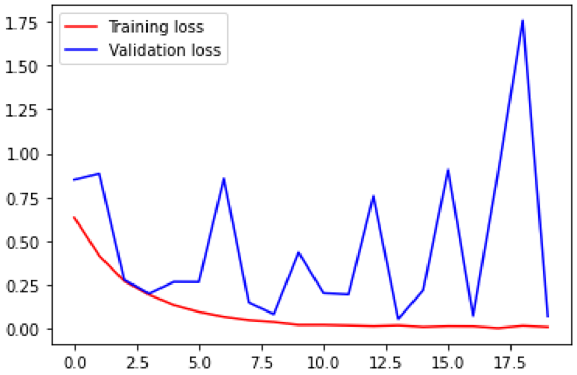


Fig. 14. **Training and validation loss**
***Source:* compiled by the authors**

The trained model is saved to a binary PB (protobuf) file. The PB file stores the actual program or model, as well as a set of named signatures, each of which identifies a function that takes tensor input and produces tensor output. In this form, the model can be easily transferred to a mobile device, for example, Intel Movidius Neural Compute Stick (NCS) [36]. The NCS module can be easily integrated into a surveillance system. For its connection, it only requires a microcomputer with a USB port.

The intelligent system scheme for identifying people without breathing masks is shown in Fig. 15. Initial data in the form of video streams comes from *N* cameras installed in different locations of the shopping mall.

The first data preprocessing is to extract separate frames from video streams. There are many software tools for extracting frames from a video stream. The standard set of OpenCV functions is most often used in machine learning. In particular, for video capturing task, the VideoCapture class from the OpenCV library is used. Thus, at the first stage of preprocessing, each video stream is divided into separate images *Frame1.1*, *Frame1.2*, …, *Frame1.M*; *Frame2.1*, *Frame2.2*, …, *Frame2.M*; …; *FrameN.1*, *FrameN.2*, …, *FrameN.M* (Fig. 15).

The second preprocessing task is to detect individual people. For this, as at the stage of training data preparation, HOG from the OpenCV library is used. Thus, images of the same size obtained at the first stage (after VideoCapture) are divided into smaller ones of various sizes. Each video stream generates a different images number to process (the more people in the frame, the more images). For example, the first video stream gives *P* images (*Image1.1*, *Image1.2*, … , *Image1.P*), the second – *Q* (*Image2.1*, *Image2.2*, … , *Image2.Q*), the *N*-th – *S* (*ImageN.1*, *Image N.2*, … , *ImageN.S*).

Trained VGG-16 model receives images *Image1.1*, *Image 1.2*, …, *Image1.P*; *Image2.1*, *Image2.2*, …, *Image2.Q*; …; *ImageN.1*, *Image N.2*, …, *ImageN.S* as input. At the output, each image is associated with one of the classes: 0 (“a person in a breathing mask”) or 1 (“a person without a breathing mask”). It should be noted, that the constructed model has two outputs. The first output shows the probability that the image is of a person wearing a breathing mask. The second is the probability of not having a breathing mask. To get a response in the form of 0 or 1, the maximum argument function (argmax) from the NumPy library was used.

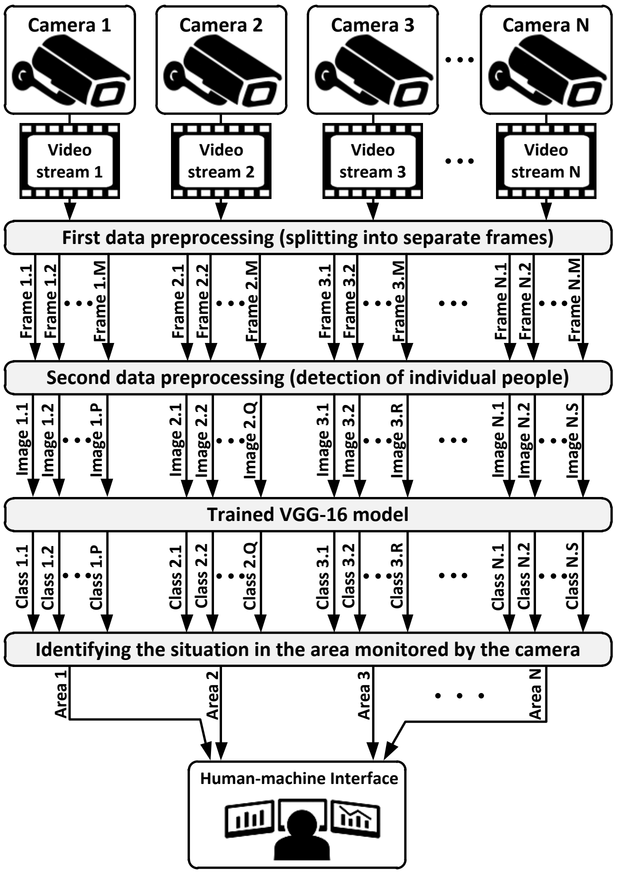


Fig. 15. **The scheme of the intelligent system for identifying people without breathing masks**
***Source:* compiled by the authors**

The identification of the situation in the areas monitored by cameras is carried out as a common logical task. Each camera gives the number of recognized classes (0 or 1), conventionally equal to the people number in the frame. For example, *P* values for the first camera (*Class1.1*, *Class1.2*, … , *Class1.P*). If in such a sample there is at least one “1”, then the area monitored by the camera is considered problematic. The logical conclusion is formed according to the following formulas:

$$Area\,1 = Class\ 1.1 \vee Class\ 1.2 \vee \ldots \vee Class\ 1.P\ ,$$
$$Area\,2 = Class\ 2.1 \vee Class\ 2.2 \vee \ldots \vee Class\ 2.Q\ ,$$
$$Area\,3 = Class\ 3.1 \vee Class\ 3.2 \vee \ldots \vee Class\ 3.R\ ,$$
$$\ldots$$
$$AreaN\ = ClassN\ .1 \vee ClassN\ .2 \vee \ldots \vee ClassN\ .S\ .$$

The security post regularly receives a problem area report via the human-machine interface. Thus, a security service consisting of a limited personnel number can effectively monitor compliance with the mask regime in the context of the COVID-19 pandemic.

4. TRAINED MODEL TESTING

A sample of 500 images was used to test the model. These images were not used in the training and validation process.

An example of prediction on test images is shown in Fig. 16 and Fig. 17. Integrated test results are illustrated by the confusion matrix presented in Table 2.

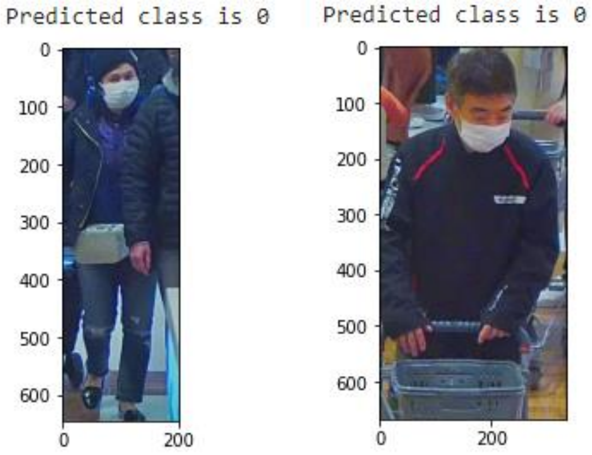


Fig. 16. **An example of a person in a breathing mask prediction**
***Source:* compiled by the authors**

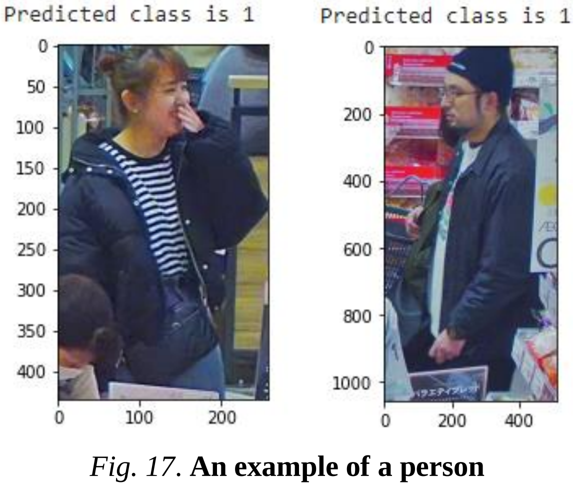


Fig. 17. **An example of a person without a breathing mask prediction**
***Source:* compiled by the authors**

Machine learning conventions presented in *Table 2:*

– *TP* – a true positive is an outcome where the model correctly predicts the positive class;

– *TN* – a negative is an outcome where the model correctly predicts the negative class;

– *FP* – an outcome where the model incorrectly predicts the positive class;

– *FN* – an outcome where the model incorrectly predicts the negative class.

Table 2. **Confusion matrix**

	Predicted	
	Positive (mask)	Negative (no mask)
Actual True	*TP* = 486	*FN* = 14
Actual False	*FP* = 47	*TN* = 453

***Source:* compiled by the authors**

Metrics calculated by confusion matrix:

$$Precision\ = \frac{TP}{TP+FP} = \frac{486}{486+47} = 0.91,$$
$$Recall\ = \frac{TP}{TP+FN} = \frac{486}{486+14} = 0.97,$$
$$F1 = 2\cdot\frac{Precision\ \cdot Re\,call}{Precision\ + Re\,call} = 2\cdot\frac{0.91\cdot 0.97}{0.91+0.97} = 0.94.$$

Precision is the fraction of relevant instances among the retrieved instances. *Recall* is the fraction of the relevant instances total amount that was actually retrieved. The *F1* score is the harmonic mean of the precision and recall. The highest *F1* value achieved with ideal precision and recall is 1.

Quality metrics (*Precision*, *Recall* and *F1* score) above 0.9 indicate good predictive quality. Models with such metrics are considered ready for implementation in production.

CONCLUSIONS

The main developments in the field of convolutional neural networks over the past decade are analyzed. The VGG-16 architecture was chosen for the research. The VGG-16 architecture was chosen for the research, since it is easy to implement, does not require large amounts of data for training, and achieves acceptable accuracy in image classification problems of medium complexity.

The dataset was collected and labeled for training and testing the neural network model. OpenCV HOG algorithm was proposed to detect an individual people in the image.

The CNN model was trained and the intelligent system scheme for identifying people without breathing masks was developed. The direct implementation of a neural network is done using TensorFlow, an open source machine learning software library. Additionally, the Keras neural network library is used as an add-on over TensorFlow. According to the results of training 99.62 % of the images from the training dataset and 90.2 % – from the validation dataset were classified correctly. Such quality indicators are suitable for practical application of the trained model.

As a result of assessing the predication quality, the following metrics were obtained: *Precision*=0.91, *Recall*=0.97, *F1*=0.94. Models with such metrics are considered ready for implementation in production. Thus, the proposed neural network architecture and the model trained on its basis are completely suitable for use in the intelligent system for identifying people without breathing masks.

REFERENCES

1. Ajit, A., Acharya, K. & Samanta, A. “A Review of Convolutional Neural Networks”. *International Conference on Emerging Trends in Information Technology and Engineering (ic-ETITE).* Vellore: India. 2020. p. 1–5. DOI: https://doi.org/10.1109/ic-ETITE47903.2020.049.

2. Zhou, Y., Chen, S., Wang, Y. & Huan, W. “Review of Research on Lightweight Convolutional Neural Networks”. *IEEE 5th Information Technology and Mechatronics Engineering Conference (ITOEC).* Chongqing: 2020. p. 1713–1720. DOI: https://doi.org/10.1109/ITOEC49072.2020.9141847.

3. Elhassouny, A. & Smarandache, F. “Trends in deep Convolutional Neural Networks Architectures: a Review”. *International Conference of Computer Science and Renewable Energies (ICCSRE).* Agadir: Morocco. 2019. p. 1–8. DOI: https://doi.org/10.1109/ICCSRE.2019.8807741.

4. Muhammed, M. A. E., Ahmed, A. A. & Khalid, T. A. “Benchmark Analysis of Popular ImageNet Classification Deep CNN Architectures”. *International Conference on Smart Technologies for Smart Nation (SmartTechCon).* Bangalore: 2017. p. 902–907. DOI: https://doi.org/10.1109/ SmartTechCon.2017.8358502.

5. Deng, J., Dong, W., Socher, R., Li, L., Kai Li & Li Fei-Fei. “ImageNet: A Large-scale Hierarchical Image Database”. *IEEE Conference on Computer Vision and Pattern Recognition*. Miami, FL.: 2009. p. 248–255. DOI: https://doi.org/10.1109/ CVPR.2009.5206848.

6. Rajalakshmi, M., Saranya, P. & Shanmugavadivu, P. “Pattern Recognition-Recognition of Handwritten Document Using Convolutional Neural Networks”. *IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS).* Tamilnadu: India. 2019. p. 1–7. DOI: https://doi.org/10.1109/ INCOS45849.2019.8951342.

7. Wan, X., Song, H., Luo, L., Li, Z., Sheng, G. & Jiang, X. “Pattern Recognition of Partial Discharge Image Based on One-dimensional Convolutional Neural Network”. *Condition Monitoring and Diagnosis (CMD).* Perth, WA. 2018. p. 1–4. DOI: https://doi.org/10.1109/CMD.2018.8535761.

8. Lee, J., Lee, S. & Yang, S. “An Ensemble Method of CNN Models for Object Detection”. *International Conference on Information and Communication Technology Convergence (ICTC)*. Jeju: 2018. p. 898–901. DOI: https://doi.org/10.1109/ ICTC.2018.8539396.

9. Mane, S. & Mangale, S. “Moving Object Detection and Tracking Using Convolutional Neural Networks”. *Second International Conference on Intelligent Computing and Control Systems (ICICCS).* Madurai: India. 2018. p. 1809–1813. DOI: https://doi.org/10.1109/ICCONS.2018.8662921.

10. Marmanis, D., Schindler, K., Wegner, J. D., Datcu, M. & Stilla, U. “Semantic Segmentation of Aerial Images with Explicit Class-boundary Modeling”. *IEEE International Geoscience and Remote Sensing Symposium (IGARSS).* Fort Worth, TX. 2017. p. 5165–5168. DOI: https://doi.org/10.1109/ IGARSS.2017.8128165.

11. Tao, H., Li, W., Qin, X. & Jia, D. “Image Semantic Segmentation Based on Convolutional Neural Network and Conditional Random Field”. *Tenth International Conference on Advanced Computational Intelligence (ICACI).* Xiamen: 2018. p. 568–572. DOI: https://doi.org/10.1109/ICACI.2018.8377522.

12. Yang, J. & Li, J. “Application of Deep Convolution Neural Network”. *14th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP)*. Chengdu: 2017. p. 229–232. DOI: https://doi.org/10.1109/ICCWAMTIP.2017.8301485.

13. Yenter, A. & Verma, A. “Deep CNN-LSTM with Combined Kernels from Multiple Branches for IMDb Review Sentiment Analysis”. *IEEE 8th Annual Ubiquitous Computing, Electronics and Mobile Communication Conference (UEMCON).* New York: NY. 2017. p. 540–546. DOI: https://doi.org/10.1109/ UEMCON.2017.8249013.

14. Li, P., Li, J. & Wang, G. “Application of Convolutional Neural Network in Natural Language Processing”. *15th International Computer Conference on Wavelet Active Media Technology and Information Processing (ICCWAMTIP).* Chengdu: China. 2018. p. 120–122. DOI: https://doi.org/10.1109/ICCWAMTIP.2018.8632576.

15. Lecun, Y., Bottou, L., Bengio, Y. & Haffner, P. “Gradient-based Learning Applied to Document Recognition”. *Proceedings of the IEEE.* 1998; Vol. 86 No.11: 2278–2324. DOI: https://doi.org/10.1109/5.726791.

16. Goodfellow, I., Bengio, Y. & Courville, A. “Deep Learning”. *MIT Press.* Massachusetts: US. 2016. 802 p.

17. Krizhevsky, A. “Learning Multiple Layers of Features from Tiny Images”, *Technical Report TR-2009.* University of Toronto. Toronto: 2009. 58 p.

18. ByungSoo Ko, Kim, H., Kyo-Joong Oh & Choi, H. “Controlled dropout: A Different Approach to Using Dropout on Deep Neural Network”. *IEEE International Conference on Big Data and Smart Computing (BigComp).* Jeju: 2017. p. 358–362. DOI: https://doi.org/10.1109/BIGCOMP.2017.7881693.

19. Srivastava, N., Hinton, G., Krizhevsky, A. & Salakhutdinov, R. “Dropout: A Simple Way to Prevent Neural Networks from Overfitting”. *Journal of Machine Learning Research.* 2014; 15(1): 211–252.

20. Zeiler, M.D. & Fergus, R. “Visualizing and Understanding Convolutional Networks”. *European conference on computer vision. Springer International Publishing.* 2014. p. 818–833. DOI: https://doi.org/10.1007/978-3-319-10590-1_53.

21. Noh, H., Hong, S. & Han, B. “Learning Deconvolution Network for Semantic Segmentation”. *IEEE International Conference on Computer Vision (ICCV).* Santiago: 2015. p. 1520–1528. DOI: https://doi.org/10.1109/ICCV.2015.178.

22. Simonyan, K. & Zisserman, A. “Very Deep Convolutional Networks for Large-Scale Image Recognition”. *arXiv preprint,* arXiv: 1409.1556. 2014. 14 p.

23. Girshick, R. “Fast R-CNN”. *2015 IEEE International Conference on Computer Vision (ICCV).* Santiago: 2015. p. 1440–1448. DOI: https://doi.org/10.1109/ICCV.2015.169.

24. Dai, J., He, K. & Sun, J. “Instance-Aware Semantic Segmentation via Multi-task Network Cascades”. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* Las Vegas: NV. 2016. p. 3150–3158. DOI: https://doi.org/10.1109/ CVPR.2016.343.

25. Zhang, K., Li, T., Liu, B. & Liu, Q. “Co-Saliency Detection via Mask-Guided Fully Convolutional Networks With Multi-Scale Label Smoothing”. *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).* Long Beach: CA, USA: 2019. p. 3090–3099. DOI: https://doi.org/10.1109/CVPR.2019.00321.

26. Szegedy, C., Liu, W., Jia, Y. et al. “Going Deeper with Convolutions”. *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* Boston: MA. 2015. p. 1–9. DOI: https://doi.org/10.1109/CVPR.2015.7298594.

27. Lin, M, Chen, Q. & Yan, S. “Network In Network”. *CoRR abs/1312.4400. arXiv: 1312.4400.* 2013. 10 p.

28. He, K., Zhang, X., Ren, S. & Sun, J. “Deep Residual Learning for Image Recognition”. *IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* Las Vegas: NV. 2016. p. 770–778. DOI: https://doi.org/10.1109/ CVPR.2016.90.

29. Squartini, S., Hussain, A. & Piazza, F. “Preprocessing Based Solution for the Vanishing Gradient Problem in Recurrent Neural Networks”. *Proceedings of the International Symposium on Circuits and Systems, ISCAS '03.* Bangkok: 2003. p. 713–716. DOI: https://doi.org/10.1109/ ISCAS.2003.1206412.

30. Hochreiter, S. “The Vanishing Gradient Problem During Learning Recurrentneural Nets and Problem Solutions”. *Int. J. Uncertain. Fuzziness Knowledge Based Syst.* 1998; 6: 107–116.

31. Szegedy, C., Ioffe, S., Vanhoucke, V. & Alemi, A. “Inception-v4, Inception-ResNet and the Impact of Residual Connections on Learning”. *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*. 2017. p. 4278–4284.

32. Wen, L., Li, X., Li, X. & Gao, L. “A New Transfer Learning Based on VGG-19 Network for Fault Diagnosis”. *IEEE 23rd International Conference on Computer Supported Cooperative Work in Design (CSCWD).* Porto: Portugal. 2019. p. 205–209. DOI: https://doi.org/10.1109/CSCWD. 2019.8791884.

33. Noble, F. K. “Comparison of OpenCV's Feature Detectors and Feature Matchers”. *23rd International Conference on Mechatronics and Machine Vision in Practice (M2VIP).* Nanjing: 2016. p. 1–6. DOI: https://doi.org/10.1109/M2VIP.2016.7827292.

34. Dalal, N. & Triggs, B “Histograms of Oriented Gradients for Human Detection”. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05).* San Diego: CA, USA: 2005; Vol. 1: 886–893. DOI: https://doi.org/10.1109/CVPR.2005.177.

35. Cardoso, A., Leitao, J. & Teixeira, C. “Using the Jupyter Notebook as a Tool to Support the Teaching and Learning Processes in Engineering Courses”. *Auer M., Tsiatsos T. (eds) The Challenges of the Digital Transformation in Education. ICL 2018. Advances in Intelligent Systems and Computing, Publ. Springer.* Cham: 2019; Vol. 917: 227–236. DOI: https://doi.org/10.1007/978-3-030-11935-5_22.

36. Pester, A. & Schrittesser, M. “Object Detection with Raspberry Pi3 and Movidius Neural Network Stick”. *5th Experiment International Conference (exp.at'19). Funchal (Madeira Island).* Portugal: 2019. p. 326–330. DOI: https://doi.org/10.1109/ EXPAT.2019.8876583.

Conflicts of Interest: the authors declare no conflict of interest

Received 02.08.2020
Received after revision 15.09.2020
Accepted 21.09.2020

DOI: https://doi.org/10.15276/aait.03.2020.2
УДК 004.931

ІНТЕЛЕКТУАЛЬНА СИСТЕМА НА БАЗІ ЗГОРТКОВОЇ НЕЙРОННОЇ МЕРЕЖІ ДЛЯ ІДЕНТИФІКАЦІЇ ЛЮДЕЙ БЕЗ ДИХАЛЬНИХ МАСОК

Олексій Іванович Шеремет[1)]
ORCID: https://orcid.org/0000-0003-1298-3617, sheremet-oleksii@ukr.net
Олександр Євгенович Коробов[2)]
ORCID: https://orcid.org/0000-0003-4530-7535, korobov.alex@gmail.com
Олександр Валентинович Садовой[3)]
ORCID: https://orcid.org/0000-0001-9739-3661, sadovoyav@ukr.net
Юлія Віталіївна Сохіна[3)]
ORCID: https://orcid.org/0000-0002-4329-5182, jvsokhina@gmail.com
[1)]Донбаська державна машинобудівна академія, бул. Машинобудівників, 39. Краматорськ, 84313, Україна
[2)]AI-labs, Майдан незалежности, 2. Киев, 01001,Украина
[3)]Дніпровський державний технічний ун-т, вул. Дніпробудівська, 2. Кам’янське, 51918, Україна

АНОТАЦІЯ

Пандемія COVID-19 має величезний вплив на окремих людей та громади. Багато організацій стикаються зі значними перебоями роботи та проблемами, що вимагають негайного реагування та вирішення. Соціальна дистанція, дихальні маски та захист очей, як профілактичні заходи проти поширення COVID-19 за відсутності ефективної противірусної вакцини, відіграють важливу роль. Заборона здійснення покупок без маски у супермаркетах та торгівельних центрах є обов’язковою в більшості країн. Однак при великій кількості покупців охорона не може перевірити наявність дихальних масок на всіх. Необхідно запровадити інтелектуальні засоби автоматизації, які допоможуть працювати охороні. У зв'язку з цим у статті пропонується сучасне рішення – інтелектуальна система ідентифікації людей без дихальних масок. Запропонована інтелектуальна система працює разом із системою відеоспостереження. Система відеоспостереження має структуру, що включає відеокамери, записуючі пристрої (жорсткі диски) та монітори. Відеокамери знімають зони продажу і передають відеозображення на записуючі пристрої, які, в свою чергу, фіксують те, що відбувається і відображають відео з камер безпосередньо на моніторі. Основною ідеєю запропонованого рішення є використання інтелектуальної системи класифікації зображень, періодично отримуваних з камер системи відеоспостереження. Розроблений класифікатор поділяє потік зображення на два класи. Перший клас – «людина в дихальній масці», а другий – «людина без дихальної маски». Коли з’являється зображення другого класу, тобто особа, яка зняла дихальну маску або зайшла в супермаркет без дихальної маски, служба безпеки негайно отримає повідомлення із зазначенням проблемної зони. Інтелектуальна система класифікації зображень заснована на згортковій нейронній мережі VGG-16. На практиці ця архітектура демонструє гарні результати класифікації зображень з великою схожістю. Для навчання моделі нейронної мережі була використана хмарна служба Google Colab – безкоштовний сервіс на базі Jupyter Notebook. Навчена модель базується на відкритій платформі машинного навчання TensorFlow. Ефективність запропонованого рішення підтверджується правильною обробкою практично отриманого набору даних. Точність класифікації становить вище 90 %.

Ключові слова: інтелектуальна система; VGG-16; згорткова нейронна мережа; TensorFlow; класифікація зображень; точність; функція втрат; машинне навчання

DOI: https://doi.org/10.15276/aait.03.2020.2
УДК 004.931

ИНТЕЛЛЕКТУАЛЬНАЯ СИСТЕМА НА БАЗЕ СВЕРТОЧНОЙ НЕЙРОННОЙ СЕТИ ДЛЯ ИДЕНТИФИКАЦИИ ЛЮДЕЙ БЕЗ ДЫХАТЕЛЬНЫХ МАСОК

Алексей Иванович Шеремет[1)]
ORCID: https://orcid.org/0000-0003-1298-3617, sheremet-oleksii@ukr.net

Александр Евгеньевич Коробов[2)]
ORCID: https://orcid.org/0000-0003-4530-7535, korobov.alex@gmail.com
Александр Валентинович Садовой[3)]
ORCID: https://orcid.org/0000-0001-9739-3661, sadovoyav@ukr.net
Юлия Витальевна Сохина[3)]
ORCID: https://orcid.org/ 0000-0002-4329-5182, jvsokhina@gmail.com
[1)]Донбасская государственная машиностроительная академия, бул. Машиностроителей, 39. Краматорск, 84313, Украина
[2)]AI-labs, площадь Независимости, 2. Киев, 01001,Украина
[3)]Днепровский государственный технический университет, ул. Днепростроевская, 2. Каменское, 51918,Украина

АННОТАЦИЯ

Пандемия COVID-19 имеет огромное влияние на отдельных людей и сообщества. Многие организации сталкиваются со значительными перебоями работе и проблемами, требующими немедленного реагирования и решения. Социальная дистанция, дыхательные маски и защита глаз, как профилактические меры против распространения COVID-19 при отсутствии эффективной противовирусной вакцины, играют важную роль. Запрет совершения покупок без маски в супермаркетах и торговых центрах является обязательным в большинстве стран. Однако при большом количестве покупателей охрана не может проверить наличие дыхательных масок на всех. Необходимо внедрить интеллектуальные средства автоматизации, которые помогут в работе охраны. В связи с этим в статье предлагается современное решение – интеллектуальная система идентификации людей без дыхательных масок. Предложенная интеллектуальная система работает вместе с системой видеонаблюдения. Система видеонаблюдения имеет структуру, включающую видеокамеры, записывающие устройства (жесткие диски) и мониторы. Видеокамеры снимают зоны продаж и передают видеоизображение на записывающие устройства, которые, в свою очередь, фиксируют происходящее и отражают видео с камер непосредственно на мониторе. Основной идеей предлагаемого решения является использование интеллектуальной системы классификации изображений, периодически получаемых с камер системы видеонаблюдения. Разработанный классификатор разделяет поток изображения на два класса. Первый класс – «человек в дыхательной маске», а второй – «человек без дыхательной маски». Когда появляется изображение второго класса, то есть человек, снявший дыхательную маску или зашедший в супермаркет без дыхательной маски, служба безопасности немедленно получит сообщение с указанием проблемной зоны. Интеллектуальная система классификации изображений основана на сверточной нейронной сети VGG-16. На практике эта архитектура демонстрирует хорошие результаты классификации изображений с большим сходством. Для обучения модели нейронной сети была использована облачная служба Google Colab – бесплатный сервис на базе Jupyter Notebook. Обученная модель базируется на открытой платформе машинного обучения TensorFlow. Эффективность предложенного решения подтверждается правильной обработкой практически полученного набора данных. Точность классификации составляет выше 90 %.

Ключевые слова: интеллектуальная система; VGG-16; сверточная нейронная сеть; TensorFlow; классификация изображений; точность; функция потерь; машинное обучение

ABOUT THE AUTHORS

Oleksii I. Sheremet – Dr. Sci. (Eng.), Prof., Head of the Department of Electromechanical Systems of Automation and Electric Drive, Donbas State Engineering Academy. 39, Mashinobudivnykiv Blvd. Kramatorsk, 84313, Ukraine
sheremet-oleksii@ukr.net. ORCID: https://orcid.org/0000-0003-1298-3617
***Research field*:** Image Processing, Data Analysis, Artificial Intelligence Methods and Systems, Automatic Control Systems, Automated Electric Drive

Олексій Іванович Шеремет д-р техніч. наук, зав. каф. Електромеханічних систем автоматизації. Донбаська державна машинобудівна академія, бул. Машинобудівників, 39. Краматорськ, 84313, Україна

Oleksandr Ye. Korobov – CEO AI-labs, AI Engineer and Software Developer, AI-labs, 2, Maidan Nezalezhnosti Sq. Kyiv, 01001,Ukraine
korobov.alex@gmail.com. ORCID: https://orcid.org/0000-0003-4530-7535
***Research field*:** Image Processing, Data Analysis, Artificial Intelligence Methods and Systems

Олександр Євгенович Коробов, Генеральний директор AI-labs, фахівець зі штучного інтелекту та розробник програмного забезпечення, AI-labs, Майдан незалежности, 2. Киев, 01001,Украина

Oleksandr V. Sadovoi – Dr. Sci. (Eng.), Prof. of the Department of Electrical Engineering and Electromechanics, Dnipro State Technical University, 2, Dniprobudivska St. Kamyanske, 51918,Ukraine
sadovoyav@ukr.net. ORCID: https://orcid.org/0000-0001-9739-3661
***Research field*:** Optimal Control of Electromechanical Systems

Олександр Валентинович Садовой – д-р техніч. наук, проф. каф. Електротехніки та електромеханіки. Дніпровський державний технічний ун-т, вул. Дніпробудівська, 2. Кам'янське, 51918, Україна

Yuliia V. Sokhina – Cand. Sci. (Eng.), Associate Prof. of the Department of Electrical Engineering and Electromechanics, Dnipro State Technical University, 2, Dniprobudivska St. Kamyanske, 51918,Ukraine
jvsokhina@gmail.com. ORCID: https://orcid.org/ 0000-0002-4329-5182
***Research field*:** Research Area: Optimal Control of Electromechanical Systems

Юлія Віталіївна Сохіна – канд. техніч. наук, доц. каф. Електротехніки та електромеханіки, Дніпровський державний технічн ий ун-т, вул. Дніпробудівська, 2. Кам'янське, 51918, Україна