Development of an Application for Recognizing Emotions Using Convolutional Neural Networks

Viktor Pomazan, Iryna Tvoroshenko, Volodymyr Gorokhovatskyi

Department of Informatics
Kharkiv National University of Radio Electronics,
Kharkiv, Ukraine
e-mail: viktor.pomazan@nure.ua; iryna.tvoroshenko@nure.ua; volodymyr.horokhovatskyi@nure.ua

Abstract: This paper explores the potential of using convolutional neural networks (CNN) for emotion recognition in marketing research of advertising. Discusses how CNN can be trained to accurately recognize emotions conveyed in advertisements and how this information can be used to gain insights into consumer behavior. By analyzing consumer emotions in response to different types of advertising, marketers can better tailor their advertising campaigns to elicit positive emotional responses and improve overall effectiveness. Overall, the paper demonstrates the potential of CNN-based emotion recognition as a valuable tool to optimize advertising strategies and improve consumer engagement.

Keywords—convolutional neural networks (CNN); deep learning; emotion analysis; emotion recognition; marketing; neural networks

1. INTRODUCTION

Emotions play a crucial role in our daily lives, influencing our decision-making processes, social interactions, and overall well-being. As such, understanding and accurately detecting emotions is of great importance in various fields, including marketing [1].

Emotional marketing helps the author to understand the most compelling aspects of human characteristics, which can lead to the emergence of aspirational drives that motivate and encourage personal satisfaction and emotional fulfillment.

Marketers can leverage emotional information to gain a deeper understanding of their customer's preferences, improve brand perception, and ultimately increase sales. With the rise of digital marketing and the abundance of data generated by online interactions, there is a growing interest in developing automated tools for emotion recognition [2]. In this paper, we explore recent advances in emotion recognition technology and its potential applications in marketing. Specifically, we examine the challenges associated with accurately detecting emotions; discuss the different modalities used in emotion recognition, and review recent studies on emotion recognition in marketing. The insights gained from this research have the potential to revolutionize how marketers understand and engage with their customers.

So, considering the very rapid development of the field of emotional marketing and the need to automate the processes of processing and transferring data about the client, it is necessary to develop an application that will be needed by the end consumer, which will have a simple and intuitive interface and high accuracy of input data processing.

The purpose of the work is to develop a program for recognizing emotions in marketing research on the effectiveness of advertising using convolutional neural networks.

The object of research is emotion recognition with convolutional neural networks.

2. RELATED WORKS

In recent years, many researches have been conducted to understand how consumer emotions affect decision-making and behavior. As a result of these researches, it became clear that emotions play an important role in marketing strategies and that they can be used to improve product quality, increase sales, and increase consumer satisfaction.

Many publications consider various methods and approaches to emotion recognition, including the use of machine learning, analysis of electro-skin signals and other physiological indicators, as well as data collection from social networks and cloud services. The examination of the most recent investigation and literature concerning this subject revealed that the concerns addressed are not adequately expounded upon, and the primary scientific outcomes have not been put into practical practice and necessitate further investigation.

In the papers [1] and [2], which address the issue of advertising research, greater attention is paid to marketing strategies in this area. The authors of both papers extensively examine various marketing approaches and strategies used in advertising. They also analyze existing trends in this field, summarize the experience of other researchers, and draw conclusions about which marketing strategies are most effective in the context of advertising research. These papers can be useful for those interested not only in the research of advertising but also in the application of marketing strategies in their work.

The papers [3-10] are dedicated to convolutional neural networks and their application in various tasks. They describe the structure and operation of convolutional neural networks, including their main layers and architectures, as well as explain how they can be used in different situations.

The papers [11-17] are about the use of convolutional neural networks for the task of emotion recognition is described. The authors of these works focused on developing deep architectures that are capable of automatically extracting features from facial images, allowing for effective determination of a person's emotional state. These works emphasize the significance of using convolutional neural networks to achieve high performance in emotion recognition tasks.

The papers also describe using of convolutional neural networks in various fields, such as image processing, speech recognition, natural language processing, and others. They provide examples of using convolutional neural networks in tasks such as object recognition in images, text classification, and real estate price prediction.

It is also worth noting that privacy and ethical issues are important in the field of emotion recognition. This topic is described in paper [18]. It is an analysis of ethical issues related to emotion-oriented technologies. The authors of the paper discuss the importance of understanding the unique ethical problems that may arise as a result of the development and use of such technologies.

In general, the analysis of publications on emotion recognition in marketing research shows that this is an active and promising area of research that can lead to improved product quality and increased consumer satisfaction. However, ethical and confidential aspects in this area must be taken into account.

3. MATERIALS AND METHOD

The development of an application for recognizing emotions in marketing research on the effectiveness of advertising involves a combination of machine learning, software engineering, and marketing research expertise. The goal of the application is to accurately recognize and measure emotions in response to advertising stimuli, allowing marketers to optimize their campaigns and better understand the needs and desires of their target audience.

The development consists of these steps:

Step 1: Found and Collected Data from Which Convolution Neural Networks Can Learn

Input data for learning convolutional neural networks was collected with the fer2013 dataset.

The data consists of 48×48 pixel grayscale images of faces. The faces have been automatically registered so that the face is more or less centered and occupies about the same amount of space in each image. The task is to categorize each face based on the emotion shown in the facial expression into one of seven categories (0=Angry, 1=Disgust, 2=Fear, 3=Happy, 4=Sad, 5=Surprise, 6=Neutral).

The training set consists of 28709 examples and the public test set consists of 3589 examples. The dataset was downloaded and added to the development environment for further use in convolutional neural network learning.

Step 2: Created a Convolution Neural Network with the Ability to Learn in the Fer2013 Dataset

Convolutional neural networks [3] (CNN's) are a specific type of artificial neural networks (ANN's), that has demonstrated high performance on various visual tasks, including image classification, image segmentation, image retrieval, object detection, image captioning, face recognition, pose estimation, traffic sign recognition, speech processing, neural style transfer, etc. [4].

The convolution operation is one of the fundamental building blocks of a convolutional neural network. The parameters of a convolutional layer are composed of a set of learnable filters (kernels) [3-6]. Each filter is small in spatial dimensions (width and height) but extends through the full depth of the input volume. Typical filter sizes can be 3×3, 5×5 or 7×7.

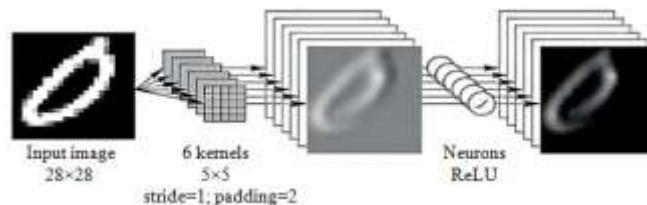The depth of a grayscale image is equal to 1, while for a color image, it is 3 (RGB) color channels (Fig. 1, [12]).

Fig. 1. *Convolutional layer*

During forward propagation, each convolution filter operates on the input volume in width and height, computing the dot product between the filter weights and input at any position, followed by a non-linear activation function (sigmoid, tanh, ReLU, etc.), and the resulting outputs are called feature maps [3-8].

An object map (also known as an activation map) gives the filter's responses at each spatial position. We stack these activation maps along the depth dimension and produce the output volume [3]. The depth of the output volume is the number of filters used in the convolution operation. Each filter learns something different from the input data, edges, dots, and colors.

The stride of convolution is the number of iterations we use on input. When the stride is 1, we move filters one pixel at a time. When the stride is 2, filters jump by 2 pixels when computing them, producing smaller output volumes in space.

Padding controls the output size. Applying convolution to the input reduces the output size, resulting in information loss. We pad the input volume with zeros around the border. There are two common options: valid convolution and same convolution. Valid convolution means no padding; the same convolution means the output size stays the same as the input size.

The output size is computed as follows:

$$x = \frac{n+2p-f}{s+1}, \qquad (1)$$

where n – the number of filters;
$\quad\quad\quad p$ – the amount of padding;
$\quad\quad\quad f$ – the filter size;
$\quad\quad\quad s$ – the stride.

CNN's often use pooling layers after convolutional layers. Their function is to reduce dimensionality, also known as subsampling. The hyperparameters of a pooling layer represent the filter size and strides [3]. The most common pooling layer is the one with a filter size of 2 and a stride of 2. There are two common types of pooling layers: max pooling and average pooling, which take the maximum and average values, respectively [3-8].

Max pooling is more frequently used than average pooling. The pooling layer does not have any trainable parameters. The intuition behind max pooling is that it detects the presence of a feature in a particular region [3-6, 13]. An example of a convolutional layer followed by a pooling layer is shown in Fig. 2 [12].
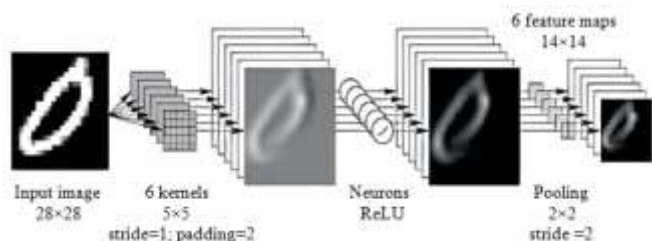

Fig. 2. *Pooling layer*

Fully connected layer. After several convolutional and pooling layers, a CNN usually ends with several fully connected layers. The tensor produced by these layers is transformed into a vector and then fed into several neural network layers [3-6, 19]. The tensor that we have at the output of these layers is transformed into a vector and then we add several neural network layers.

The fully connected layers typically are the last few layers of the architecture as shown in Fig. 3 [12], the dropout regularization technique can be applied in the fully connected layers to prevent overfitting.

The final fully connected layer in the architecture contains the same amount of output neurons as the number of classes to be recognized [13].
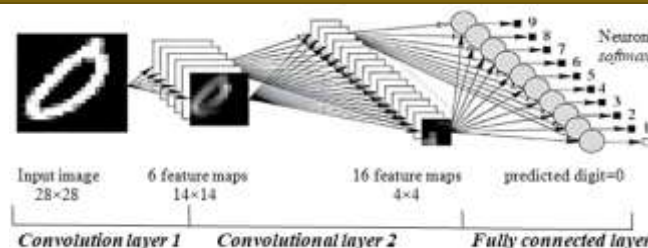

Fig. 3. *Fully connected layer*

The learning process of a CNN on the fer2013 dataset involves several steps, which are outlined below.

Data preprocessing. The first step was preprocessing of the dataset. This may involve tasks such as resizing the images to a specific size, normalizing the pixel values, and splitting the dataset into training, validation, and testing sets.

Modeling architecture. In the next step was to define the architecture of the CNN. This involves deciding on the number of layers, the number of filters in each layer, the size of the filters, and the activation functions to be used. In the case of the fer2013 dataset, the architecture might be involved several convolutional layers followed by max-pooling layers and then fully connected layers.

For this task was developed CNN which consists of 15 layers: 10 convolution layers, 4 Batch Normalization layers, 4 Dropout layers, and 2 AveragePooling2D layers, as well as GlobalAveragePooling2D and Activation layers.

All convolutional layers have ReLU activation Fig. 2, and the last layer has Softmax activation, called "predictions". The number of filters in the convolutional layers gradually increases from 16 to 256, and the size of the convolution kernels decreases from 7×7 to 3×3.

The CNN design is shown in Fig. 4-Fig. 7.

Training model. The CNN is then trained on the training set using a loss function, such as cross-entropy, and an optimization algorithm, such as stochastic gradient descent. During training, the weights of the network are adjusted to minimize the loss function.

Validation model. After each epoch of training, the performance of the network is evaluated on the validation set. This allows for early stopping or adjusting hyperparameters to prevent overfitting.

Testing model. Once training is complete, the final performance of the CNN is evaluated on the testing set. This provides an estimate of the network's ability to generalize to new, unseen data.

Hyperparameter Tuning. The performance of the CNN on the validation set can be used to tune hyperparameters such as learning rate, number of filters, and filter size.
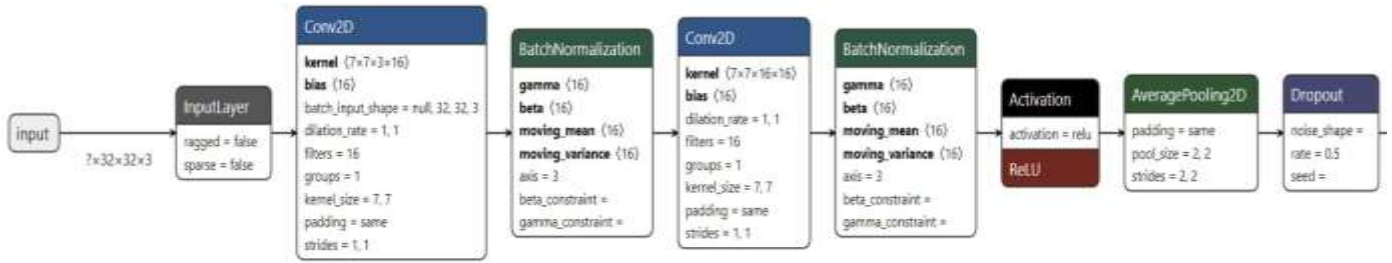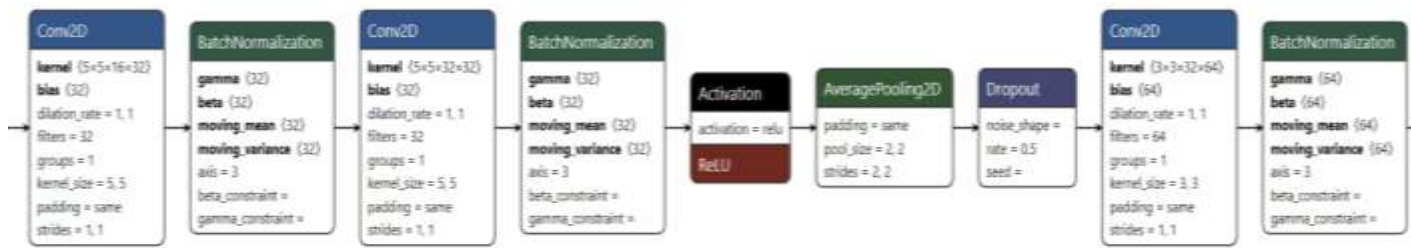
Fig. 4. *CNN's design (layer 1)*
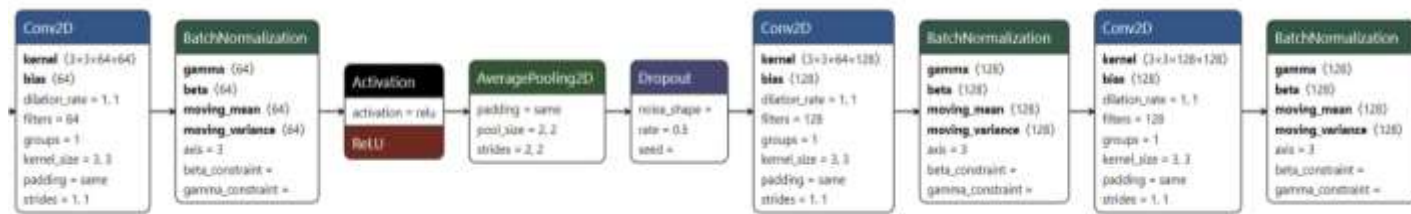

Fig. 5. *CNN's design (layer 2)*
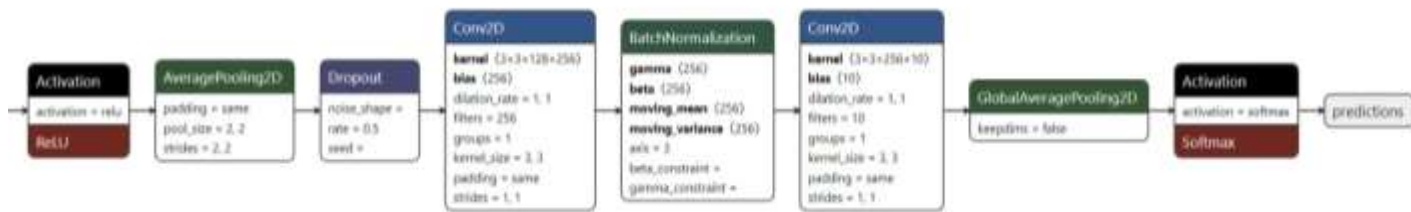

Fig. 6. *CNN's design (layer 3)*

Fig. 7. *CNN's design (layer 4)*

Overall, the learning process of a CNN on the fer2013 dataset involved a combination of data preprocessing, model architecture design, training, validation, testing, and hyperparameter tuning. With proper tuning and optimization, a well-designed CNN can achieve high accuracy in recognizing facial expressions in the fer2013 dataset [3-8, 15].

This model shares some similarities with the Visual Geometry Group (VGG) architecture but also differs in several aspects. The similarities include the use of sequential convolutional layers with ReLU activation and batch normalization [20].

Key differences from VGG:

1. Convolutional Layer Structure: The model includes convolutional layers with different filter sizes: 7×7, 5×5, and 3×3. VGG, on the other hand, exclusively used 3×3 filters.

2. Dimensionality Reduction: VGG employed max-pooling layers with a 2×2 filter and stride of 2 after each block of convolutional layers to reduce dimensionality by half [20].

In this model, average pooling is used instead with a 2×2 filter and stride of 2, employing a different method of spatial feature averaging.

3. Number of Filters: VGG typically used more filters in each convolutional layer, starting from 64 and doubling them in each block. This model utilizes smaller numbers of filters: 16, 32, 64, 128, and 256 [20].

4. Global Average Pooling: Instead of fully connected layers before the final softmax layer, this model employs global average pooling (GlobalAveragePooling2D), which averages features spatially rather than using fully connected layers.

5. Size of Final Convolutional Layer: In VGG, the last convolutional layer had a 7×7 filter size, whereas, in this model, the filter size is determined by the num_classes parameter [20].

Overall, this model can be considered a simplified version of VGG, utilizing different filter sizes, modified dimensionality reduction, fewer filters, and a different final layer approach.

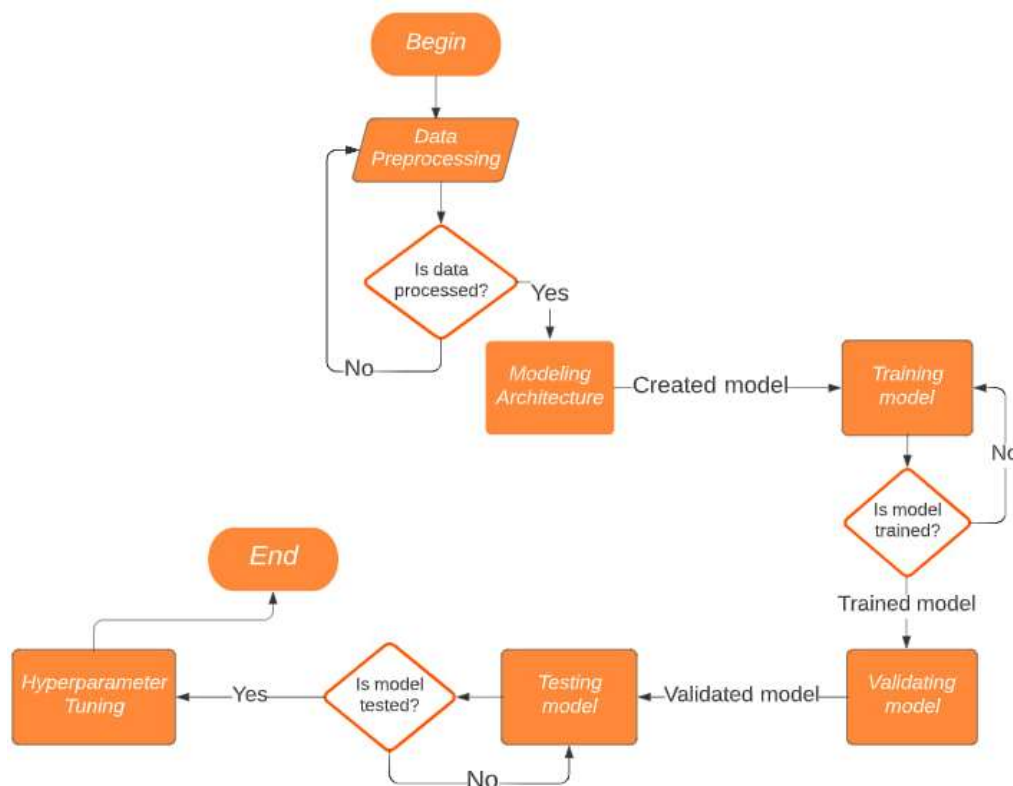The block diagram of the learning process of the convolutional neural network is shown in Fig. 8.

Step 3: Prepared a PYQT Desktop Project for Importing Convolutional Neural Networks

PyQt5 is a set of Python libraries for creating a graphical interface based on the Qt5 platform. The Qt library is one of the most powerful GUI libraries. To develop a PyQt application with two windows for displaying video from a camera and a file, two QMainWindow windows need to be created with a QLabel widget in each window. One window should have a button to open a dialog window to select a video file, and the other window should use the OpenCV library to capture video frames from the camera. A QTimer can be used to update the camera frames and a function to update the frames from the video file. Finally, controls should be added to allow the user to switch between the cameras and file modes.

Step 4: Has Been Imported Convolutional Neural Network to the PYQT Project

For importing a convolutional neural network in a PyQT project, has been installed a deep learning framework that is used to train a convolutional neural network. After that has been loaded the training convolutional neural network model is to be imported into a PyQT project. Import the loaded model into the PyQT project using the appropriate framework API functions that were used. Adjust the model if necessary, such as setting the correct input data size. Used the model to perform inference in a PyQT project, such as investigating accusations (Fig. 9).
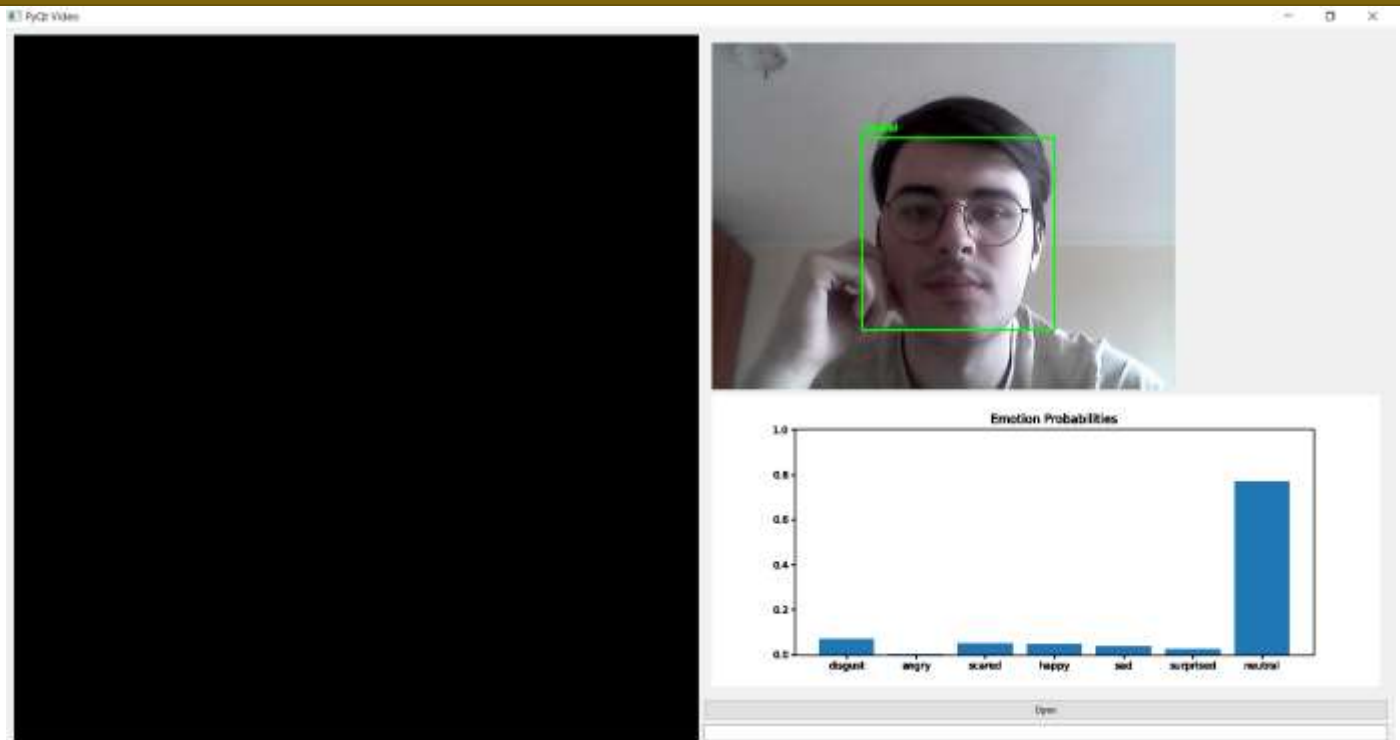

Fig. 8. *Block diagram of the learning process of the neural network*

Fig. 9. *Result of emotion recognition*

Fig. 9 is shown the base result of emotion recognition. The main program window is opened. This window contains all the main program control elements.

The real-time facial emotion detection result is displayed on the main window. For this, a video camera connected to the computer is used. The program analyzes the image, detects the person's face, and determines the emotion they are expressing. This could be happiness, sadness, fear, surprise, etc.

In the case of Fig. 9, it's neutral emotion. That means that the neural network was successfully learned and ready for testing. Also was created histogram to show emotion probabilities in percent for a better understanding of emotions in the current time. There is also a button on the main window that allows a video file to be opened. After clicking this button, the user can select a video file on their computer and open it in the program. After that, the data about recognized emotions will be transmitted to a text file for further processing by specialists.

Step 5: Testing

For testing the application for recognizing emotions in marketing research on the effectiveness of advertising we use the fer2013 dataset for emotion recognition. Application may be run by the client and followed up every little bug or mistake, especially in CNN accuracy in the recognition process. When the application is started it must see one modal window with two small windows and a button for choosing video from the device. When the button is clicked, it's opened the file system, and from that can be chosen the video file for a marketing advertisement experiment. When the video is chosen from Explorer, it's opened in a high modal window, in the lower starts showing recording from the device camera. And while the application is working it will show the emotions in a square around its face and emotions probability in the histogram [21-23]. After the video is started playing the timing of the video and the application working are compared, and with each new frame information about the current changes in the emotions on the screen is added to the text file, with the exact timing of the video on which this change occurred, for further comparison, to understand at what point in the video the customer changed dramatically in emotions and based on this, certain conclusions were made regarding the quality and advertising and possible.

4. THE RESULTS OF THE TESTING

During the research, an experiment was conducted, which showed that among 20 people who tried the program, 17 of them agreed that the emotions were recognized and displayed correctly in all cases, while 3 thought that in 3-5 cases out of 20, the emotions were identified incorrectly. Each of the subjects showed one emotion 5 times to determine the accuracy of recognition. Table 1 was shown the average recognition accuracy for each emotion and the error percentage with the rest of the presented emotions.

The result of the testing is shown in Table 1.

Table 1: Result of testing

Emotion	disgust	angry	scared	happy	sad	surprised	neutral
disgust	88.06%	1.34%	1.45%	0.43%	4.33%	1.43%	1.01%
angry	3.56%	92.86%	1.34%	1.57%	2.22%	1.87%	1.08%
scared	2.89%	2.01%	79.75%	0.98%	3.75%	2.80%	0.57%
happy	1.11%	1.45%	1.42%	94.50%	0.84%	5.04%	2.23%
sad	2.38%	1.12%	10.01%	1.14%	84.33%	2.12%	1.65%
surprised	1.11%	0.43%	4.98%	1.33%	1.98%	85.13%	0.34%
neutral	0.89%	0.79%	1.05%	1.19%	2.55%	1.61%	93.12%
Accuracy	88.25%						

Also was compared our neural network with existing convolutional neural network models. Out CNN was compared with three other CNN architectures AlexNet [3], VGGNet [3], and ResNet [11]. The AlexNet [3] architecture was the first work that popularized Convolutional Networks in Computer Vision, it was the winner of the ImageNet ILSVRC competition in 2012, and it had a 15.4% top-5 error rate of 26.2% for the next lowest network [3].

The input image resolution is 224×224, the architecture consists of 5 convolutional layers, three 2×2 max-pooling layers, and 2 fully connected layers. As shown in Fig. 5, the filter size of the first convolution layer is (11, 11, 3, 96) with a stride of 4, and the first output shape is (55, 55, 96). As a regularization technique, dropout is used in the fully connected layers to reduce overfitting [5]. The total number of parameters in AlexNet [3] is 60 million.

The VGG Network [3] was introduced in 2014 by Karen Simonyan and Andrew Zisserman. At that time, it was considered a very deep network. Its main contribution was in showing that the depth of the network is a critical component to achieving better recognition or classification accuracy in CNN's.

ResNet [11] features special skip connections and the use of batch normalization [24] after every convolution layer. Deeper models are harder to optimize, the solution is to use a skip connection, which allows taking the activation from one layer and feeding it to another layer. Using that enables us to train very deep networks and avoid vanishing and exploding gradient problems. To reduce the number of parameters, the ResNets [11] do not have fully connected layers, besides fully connected layers to output the 1000 classes [25].

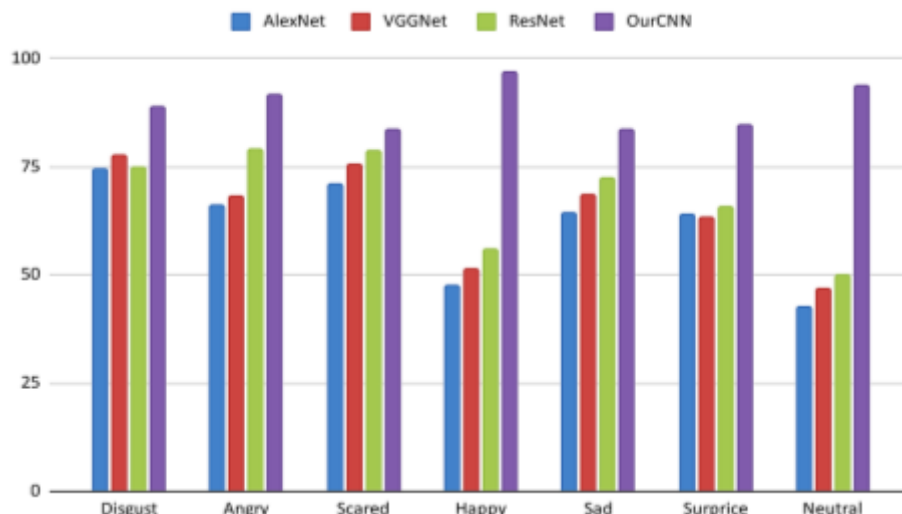The result of the comparison is shown in Fig. 10.

Fig. 10. *Result of comparison of different CNN architectures*

5. DISCUSSION

On the test dataset used in this research, the accuracy of the new model exceeded that of AlexNet, VGGNet, and ResNet by an average of 15%-20%. In addition, the new model required less memory and learned faster, which can be an important factor when using neural networks in practice.

The results of this research indicate that the new model can be effectively applied to solve problems in the field of emotional marketing.
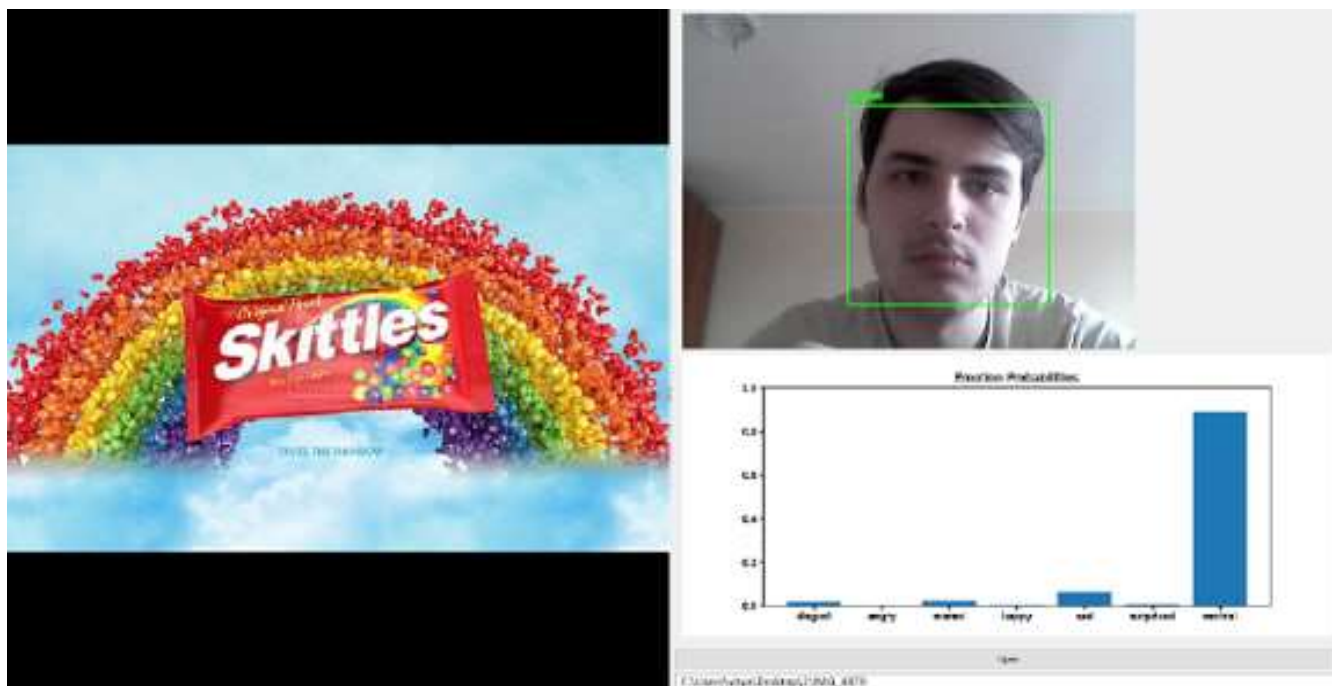
The result of the testing is shown in Fig. 11-Fig. 16.

Fig. 11. *Result of emotion recognition (example 1)*

Fig. 12. *Information about recognition process per frame (example 1)*
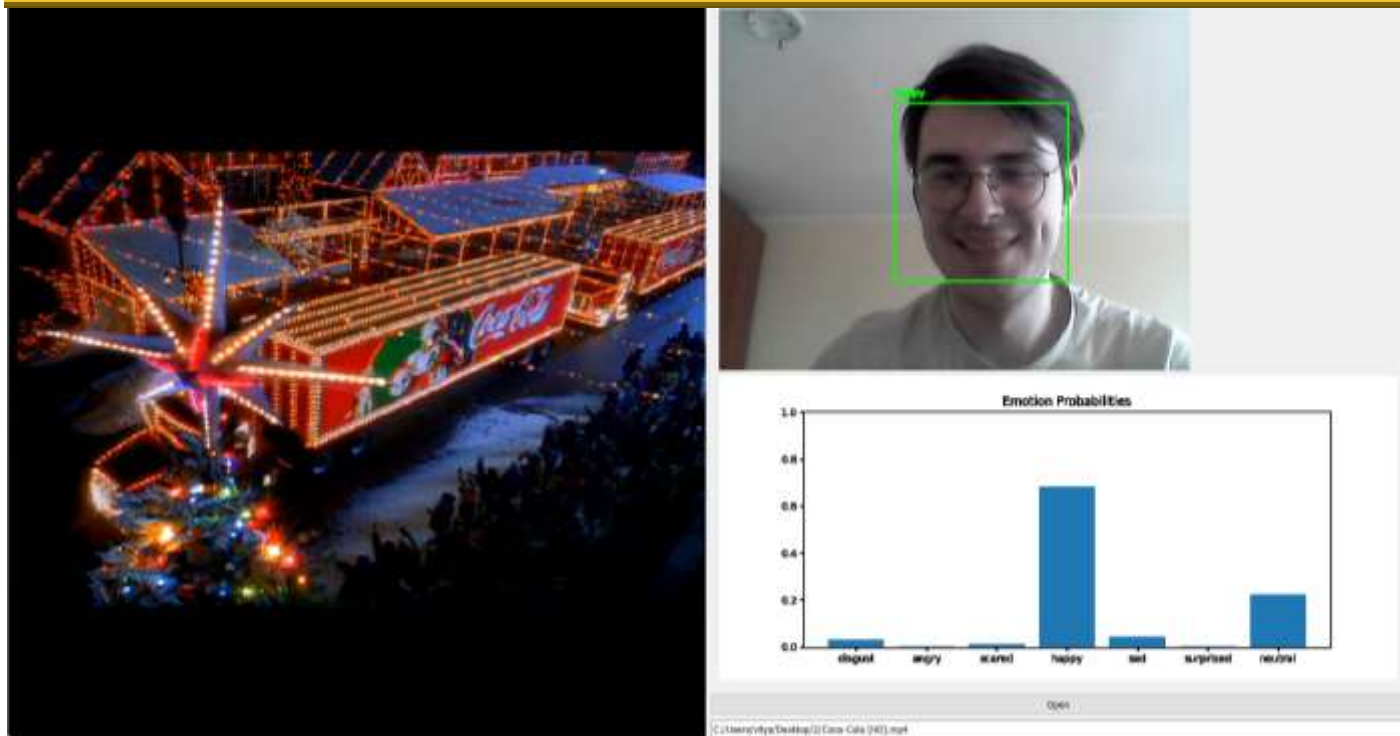
Fig. 13. *Result of emotion recognition (example 2)*

Fig. 14. *Information about recognition process per frame (example 2)*

Fig. 15. *Result of emotion recognition (example 3)*

Fig. 16. *Information about recognition process per frame (example 3)*

In Fig. 11-Fig. 16, the results of the program are shown. The user can view different advertising videos and the program will analyze their emotional reaction to each one in real-time using emotion recognition algorithms. The results of image processing will be recorded in a text file that can be analyzed later.

Fig. 11 and Fig. 12 are shown that have neutral emotion with accuracy of 85%-90%, Fig. 13 and Fig. 14 – happy with 70%-77% of accuracy, and Fig. 15 and Fig. 16 disgust with 60%-70% of accuracy. Also, to enable more precise further analysis, the histogram displayed the emotions that arise in a percentage ratio to each other during processing, so as not to miss any minor processing details.

Additionally, the program can display the recognition results on a histogram below the video camera recording, so that it can be monitored their emotional reactions to advertising videos in real-time.

If the client reacts neutrally to the advertising video, the histogram will show no significant changes in the level of emotions. If the customer reacts positively or negatively, this will be displayed on the histogram accordingly.

6. Conclusion

The obtained results show that emotions can be correctly identified by the program in 88.25% of cases, taking into account that emotions are individual things and are manifested differently for each person. As a result, it was shown that emotions can be successfully recognized by the program, the neural network can be trained correctly, and there is room for improvement results up to 90% – 95%.

This method has demonstrated its effectiveness compared to other methods of emotion recognition, where the average recognition rate reaches 65% – 75%.

In the future, it is planned to improve the dataset on which the neural network is trained for more accurate detection of emotions and to add new emotions to the dataset, so that users of the service have a greater understanding of the change of emotions during the research of people's reactions to this or that advertisement, to further improve it.

During the development of the application, was being made a comparison of performance with analogs such as Beyond Verbal, Noldus FaceReader, RealEyes, and others were conducted. Overall, the developed application has several advantages and a few drawbacks that will be improved with the advancement of technology. Among the main advantages, it is worth highlighting the accuracy of the program, which provides a more accurate assessment of emotional reactions than traditional research methods. The program allows for the collection of a large amount of data, which improves the statistical significance of research results. In addition, the use of the program allows for a reduction in the cost of research, as it does not require the hiring of a large number of specialists to conduct traditional research methods. The main drawback of the developed application for recognizing emotions in marketing research is its sensitivity to noise, as the program may be sensitive to noise and changes in lighting, which can lead to inaccurate results. Additionally, the program may be sensitive to age and race, as the developed application may not accurately recognize emotions in people of a certain age and race, but there is potential to address this issue in the future.

In addition, work is already underway on the function of recognizing emotions of several people at once; development is underway on improving the work of a convolutional neural network for more accurate recognition of emotions during marketing research on the effectiveness of advertising.

7. Acknowledgment

8. References

[1] Archer-Brown, C. & Kietzmann, J. (2018). Strategic knowledge management and enterprise social media, J. Knowl. Manag., vol. 22, no. 6, pp. 1288-1309.

[2] Dao, B. (2020). The power of emotional marketing: Case study of visit Lapland tours (bachelor's thesis, degree programmed in tourism, bachelor of hospitality management). [Online]. Available: https://www.theseus.fi/handle/10024/340599.

[3] Bezdan, T., & Bacanin, N. (2019). Convolutional neural network layers and architectures, in Sinteza 2019 – Int. Sci. Conf. Inf. Technol. Data Relat. Res., Belgrade, Singidunum University, Serbia, pp. 445-451.

[4] Yamashita, R., Nishio, M., Do, R. K. G., & Togashi, K. (2018). Convolutional neural networks: An overview and application in radiology, Insights Imaging, vol. 9, pp. 611-629.

[5] Krizhevsky, A., Sutskever, I. & Hinton, G. E. (2012). Imagenet classification with deep convolutional neural networks, Adv. Neural Inf. Process. Syst., pp. 1097-1105.

[6] Scherer, K. R. (1984). Approaches to emotion. Chapter: On the nature and function of emotion: A component process approach; Scherer, K. R., Ekman, P., Eds.; Taylor and Francis Group: New York, NY, USA.

[7] Nour, N., Elhebir, M. & Serestina V. (2020). Face expression recognition using convolution neural network (CNN) models, Int. J. Grid Comput. & Appl. (IJGCA), vol. 11, no. 1/2/3/4. [Online]. Available: https://aircconline.com/ijgca/V11N4/11420ijgca01.pdf.

[8] Velasco, M., Justo, R., & Inés Torres, M. (2022). Automatic identification of emotional information in Spanish TV debates and human-machine interactions, Appl. Sci., vol. 12, no. 4, p. 1902.

[9] Cun, Y. Le *et al*. (1990). Handwritten digit recognition with a back-propagation network, Adv. Neural Inf. Process. Syst., pp. 396-404. [Online]. Available:

https://proceedings.neurips.cc/paper/1989/file/53c3bce6 6e43be4f209556518c2fcb54-Paper.pdf.

[10] Wang, Z. J. *et al*. (2020). CNN explainer: learning convolutional neural networks with interactive visualization, IEEE Trans. Vis. Comput. Graph., vol. 27, no. 2, pp. 1396-1406.

[11] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition, in Proc. IEEE Comput. Vis. Pattern Recognit. (CVPR), Las Vegas, NV, USA, pp. 770-778.

[12] Fan, Y., Lu, X., Li, D., & Liu, Y. (2016). Video-based emotion recognition using CNN-RNN and C3D hybrid networks, in Proc. 18th ACM Int. Conf. Multimodal Interact., pp. 445-450.

[13] Couchot, J., Couturier, R., Guyeux, C., & Salomon, M. (2016). Steganalysis via a convolutional neural network using large convolution filters for Embedding Process with Same Stego Key, pp. 1-24. [Online]. Available: https://arxiv.org/pdf/1605.07946.pdf.

[14] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. & Salakhutdinov, R. (2014). Dropout: a simple way to prevent neural networks from overfitting, J. Mach. Learn. Res., vol. 15, pp. 1929-1958.

[15] Jain, N., Kumar, S., Kumar, A., Shamsolmoali, P., & Zareapoor, M. (2018). Hybrid deep neural networks for face emotion recognition, Pattern Recognit. Lett., vol. 115, pp. 101-106.

[16] Mellouk, W., & Handouzi, W. (2020). Facial emotion recognition using deep learning: review and insights, Procedia Comput. Sci., vol. 175, pp. 689-694.

[17] Cai, Y., Zheng, W., Zhang, T., Li, Q., Cui, Z., & Ye, J. (2016). Video based emotion recognition using CNN and BRNN, in: Tan, T., Li, X., Chen, X., Zhou, J., Yang, J., Cheng, H. (Eds), Pattern Recognit., CCPR 2016, Commun. Comput. Inf. Sci., vol 663, Springer, Singapore. [Online]. Available: https://doi.org/10.1007/978-981-10-3005-5_56.

[18] Khaireddin, Y., & Chen, Z. (2021). Facial emotion recognition: state of the art performance on FER2013, arXiv:2105.03588. [Online]. Available: https://doi.org/10.48550/arXiv.2105.03588.

[19] Goldie, P., Döring, S., & Cowie, R. (2011). The ethical distinctiveness of emotion-oriented technology: four long-term issues, in: Cowie, R., Pelachaud, C., Petta, P. (Eds), Emotion-Oriented Syst., Cogn. Technol., Springer, Berlin, Heidelberg. [Online]. Available: https://doi.org/10.1007/978-3-642-15184-2_39.

[20] Kabir, S., Islam, R. U., Hossain, M. S., & Andersson, K. (2020). An integrated approach of belief rule base and deep learning to predict air pollution. Sens., vol. 20, no 7, p. 1956.

[21] Gorokhovatskyi, V., Peredrii, O., Tvoroshenko, I., & Markov, T. (2023). Distance matrix for a set of structural description components as a tool for image classifier creating. Adv. Inf. Syst., vol. 7, no. 1, pp. 5-13. Гороховатський, В. О., Передрій, О. О., Творошенко, І. С., & Марков, Т. Є. (2023). Матриця відстаней для множини компонентів структурного опису як інструмент для створення класифікатора зображень, Сучасні інформаційні системи, 7(1), С. 5-13.

[22] Gorokhovatskyi, V., Tvoroshenko, I., Kobylin, O., & Vlasenko, N. (2023). Search for visual objects by request in the form of a cluster representation for the structural image description, Adv. Electr. Electron. Eng., vol. 21, no. 1, pp. 19-27.

[23] Daradkeh, Y. I., Gorokhovatskyi, V., Tvoroshenko, I., & Zeghid, M. (2022). Tools for fast metric data search in structural methods for image classification, IEEE Access, vol. 10, pp. 124738-124746.

[24] Ioffe, S. & Szegedy, C. (2015). Batch normalization: accelerating deep network training by reducing internal covariate shift, in Proc. 32nd Int. Conf. Mach. Learn., vol. 37, pp. 448-456.

[25] Gorokhovatsky, V. O., Pupchenko, D. V., & Solodchenko, K. G. (2018). Analysis of the properties, characteristics and results of using the latest detectors to determine specific points of the image, Control, Nav. Commun. Syst., vol. 1, p. 47.