UDC 004.391

IMPLEMENTATION OF MEDICAL MASK RECOGNITION TECHNOLOGY IN REAL TIME USING A VIDEO CAMERA

Dmytro V. Prochukhan[1], Mykola M. Korablev[2]

[1]Separate structural subdivision Kharkiv computer and technology professional college National Technical University «Kharkiv Polytechnic Institute», Kharkiv, Ukraine
[2]Kharkiv National University of Radio Electronics, Kharkiv, Ukraine

Background. The coronavirus pandemic (COVID-19) has led to a major global health crisis due to the rapid spread of the virus. The World Health Organization has provided guidelines for protection against this disease. One of the effective recommendations is to wear a medical mask in public or crowded places. Involving human resources to monitor how these requirements are met is ineffective. It is necessary to automate the process of determining a properly dressed medical mask in real time using a video camera. The authors have developed an application that effectively copes with this task.

Objective. The purpose of thew paper is the application development of a medical mask recognition in real time using a video camera, adapted for the use in modern Ukrainian reality with high accuracy and low system requirements.

Methods. Analysis of existing analogues in the world, building a model of convolutional neural network, the architecture of which will detect and classify the image obtained from the camcorder in real time, create application architecture, develop a model in Python programming language, application testing.

Results. A convolutional neural network of its own architecture has been created. The use of the Adam algorithm to optimize learning and use binary cross entropy as a cost function is substantiated. The method of face recognition using Haar features has been improved. High rate of convolutional neural network training was obtained: the training set accuracy - 97.46%, the test set accuracy - 97.23%, the cost function value at the training set - 2.37%, the cost function value at the test set - 2.57%. An application consisting of three modules has been developed: a machine learning module, an image processing module, a video camera activation module and a mask recognition device.

Conclusions. The application effectively copes with the task of recognizing the presence of a medical mask in real time. The developed model in comparison with the analogues has a smaller size and simpler architecture without compromising accuracy or speed. The software has been successfully tested on various operating systems and hardware configurations. The application can be used in areas where there is a need to automate the process of determining the presence and accuracy of wearing a medical mask in real time.

Keywords: *convolutional neural network; detection, recognition; classification; optimization algorithm.*

Introduction

The coronavirus pandemic (COVID-19) has led to a major global health crisis due to the rapid spread of the virus. The World Health Organization has provided guidelines for protection against this disease. One of the effective recommendations is to wear a medical mask in public or crowded places. Involving human resources to monitor the compliance is ineffective.

Therefore, it is necessary to automate the process of determining a properly dressed medical mask in real time using a video camera. Solving this scientific-applied problem will increase people's safety. It will help track safety violations, promote the use of masks and create safe working conditions.

The study [1] substantiates the effectiveness of the architecture of the YOLOv3 and R-CNN models in the problem of detecting a medical mask presence on the face.

In [2], a two-stage architecture of a convolutional neural network is presented, which can detect the presence of a medical mask and is compatible with CCTV cameras.

The study [3] proposed a model of convolutional neural network, which is ahead of such models as DenseNet-121, MobileNet-V2, VGG-19, Inception-V3 in terms of recognition accuracy and efficiency. Ukrainian researchers have poorly covered the scientific-applied problem. Therefore, it is necessary to develop an application adapted for the use in modern Ukrainian reality with high accuracy and low system requirements. To solve this problem, it is necessary to build a model of a convolutional neural network, the architecture of which will allow the detection and classification of the image obtained from the camcorder in real time. Based on the developed model, an application should be developed in the Python programming language. The architecture of this application will consist of three modules: a machine learning module, an image processing module, and a module to turn on a computer video camera and mask recognition.

Development of a medical mask recognition application in real time using a video camera.

A set of 1,376 images was used to study the convolutional neural network. 690 images contain images of people with masks on their faces, and 686 images - without masks. For successful algorithms application, the input images have different resolution, a certain angle of rotation, additional noise, and blur. Features of the input images are presented in Fig.1

Fig.1 Features of the input images

The training set of masked images has 552 images, and the test set contains 138 images. The training set of images without a mask has 548 images, and the test set contains 137 images. The size of the input data is 150 × 150 × 3.

The next step in solving our problem is the step-by-step construction of a convolutional neural network, which will provide parallelization of calculations and resistance to position changes of the recognized image. To form the layers of the convolutional neural network, we use the Keras library. From the specified library we use a special Sequential model to build a stack of layers. To implement efficient spatial convolution, we use the Convolution2D layer with 100 3 × 3 filters and a ReLu activation function that speeds up learning. The MaxPooling2D layer is used for sampling by dividing the input data into rectangular areas and calculating the maximum of each area. To perform the next convolution step, apply the Convolution2D layer with 100 3 × 3 filters and the ReLu activation function again, then use another MaxPooling2D layer.

The next Flatten layer converts data from two-dimensional to one-dimensional format. To eliminate the problem of retraining of the convolutional neural network after the Flatten layer, we use the Dropout method with a value of 0.3. In the next step, we use a fully connected Dense layer, which multiplies the input by the weight matrix and then adds an offset vector. The next layer in the network is the fully connected Dense layer, which consists of 512 neurons. In this layer, ReLU is used as an activation function. The Dense source layer contains 2 neurons and a softmax activation function that determines the probability of belonging an image to one of two classes.

The constructed model of the convolutional neural network is presented in Fig.2.

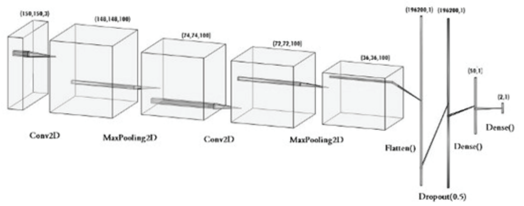

Fig.2 Convolutional network model

After building a model of a convolutional neural network, you should compile the developed network. Since two classes are used in the implementation, when compiling the model as a function of costs it is necessary to choose binary cross entropy, the value of which is an assessment of the network accuracy and the classification quality.

The lowest value of the cost function must be achieved to obtain high classification accuracy. For optimization, the Adam algorithm is used, which combines the advantages of the RMSProp and AdaGrad algorithms. The Adam optimizer is useful due to its simple implementation, low memory requirements, computational efficiency, the ability to bypass local minima and find the optimal value. During the study, the efficiency of the Adam algorithm was experimentally established in comparison with the RMSProp, AdaGrad, RMSProp and Momentum algorithms.

The ImageDataGenerator class of the Keras library processes images. In the first step of processing it is necessary to re-scale the image from the range [0; 255] to the range [0; 1]. Due to the fact that a small amount of data is not enough for quality training of the model, you need to use the method of supplementing the data, which consists of converting each image in different ways and adding new images to the data set.

In our study, the image was rotated by an angle of rotation ranging from 0 to 40 degrees (the value of the rotation_range parameter is 40). Empty areas created by rotating images are replaced by the nearest pixel values (argument fill_mode is equal to "nearest").

The study also performs vertical and horizontal image shift by 20 percent (parameters height_shift_range and width_shift_range are 0.2), image rotation in the horizontal direction (horizontal_flip = true). Images from the training and test sets are loaded using the flow_from_directory method.

When training the model using the fit_generator function, the data set was submitted 30 times. At the 18th iteration, the training set accuracy reached a maximum value of 97.46%, and the test set accuracy also reached a maximum value of 97.23%. At 18 iterations, the cost function minimum value of 2.37% was reached at the training set and 2.57% at the test set. At the 19th iteration, the process of retraining the model and the deterioration of these parameters began. At the 30th iteration, the model had an accuracy of 96.97% at the training set and 96.29% at the test set. The 30th iteration also achieved a cost function value of 3.47% at the training set and 3.89% at the test set.

Therefore, the trained model was saved after 18 iterations. After learning the convolutional neural network, we use it to further work with frames of images obtained from the webcam in an endless cycle in real time.

To recognize the face, we use the Haar sign mechanism, which is currently most often used by researchers for face recognition [4], [5].

From the upper left corner to the lower right corner in a certain image there is a search for a feature and its manifestation, which corresponds to a specific feature of Haar. Fig. 3 shows the identification of different parts of the face and face recognition.

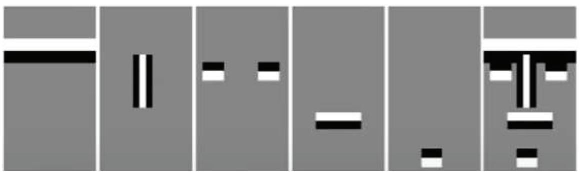

Fig.3 Convolutional network model

When specific matching features are found at the same time, then a human face can be recognized. The study used the cascade classifier Face Detection, which is a pre-trained model based on Haar characteristics.

For the selected area of the face there is a probability indicator for each of the two classes (image without a mask, image with a mask). The study derived the name of the class with a higher probability.

Using the functions of the OpenCV library, which supports GPU acceleration, the software implements real-time drawing of the square, which limits the found area of the face with the top message about the presence or absence of a mask.

Fig. 4 shows the result of determining the presence of a mask in real time.

Fig.4 Determining the presence of a mask in real time

Conclusion

A convolutional neural network of its own architecture has been created. The use of the Adam algorithm to optimize learning and use binary cross entropy as a cost function is substantiated. The method of face recognition using Haar features has been improved. High rate of convolutional neural network learning was obtained: the training set accuracy - 97.46%, the test set accuracy - 97.23%, the cost function value at the training set - 2.37%, the cost function value at the test set - 2.57%. Software consisting of three modules has been developed: a machine learning module responsible for neural network learning cycles; image processing module that performs normalization and image processing; video camera and mask detection module. The application effectively copes with the task of recognizing the presence of a medical mask in real time using a video camera. The software, compared to the analogues proposed in studies [1] - [3], has a smaller size, a simpler architecture without compromising accuracy or speed. The application has been successfully tested on various operating systems and hardware configurations. The software can be used in industries where there is a need to automate the process of determining the presence and accuracy of wearing a medical mask in real time using a video camera.

References

1. S. Singh, U. Ahuja, M. Kumar, K. Kumar, M. Sachdeva, "Face mask detection using YOLOv3 and faster R-CNN models: COVID-19 environment", *Multimedia Tools Applications*, vol.80, pp. 19753-19768, 2021 https://doi.org/10.1007/s11042-021-10711-8.

2. A. Chavda, J. Dsouza, S. Badgujar, & A. Damani, "Multi-stage CNN architecture for face mask detection". In *2021 6th International Conference for Convergence in Technology (i2ct)*, IEEE, 2021, pp. 1-8, doi: 10.1109/I2CT51068.2021.9418207.

3. H. Goyal, K. Sidana, C. Singh, A. Jain, S. Jindal, "A real time face mask detection system using convolutional neural network.", Multimedia Tool and Applications., vol. 81(11), pp.14999-15015, 2022.

4. L. Cuimei, Q. Zhiliang, J. Nan and W. Jianhua, "Human face detection algorithm via Haar cascade classifier combined with three additional classifiers", *2017 13th IEEE International Conference on Electronic Measurement & Instruments (ICEMI)*, 2017, pp. 483-487.

5. R. J. Fontana, "Recent System Applications of Short-Pulse Ultra-Wideband (UWB) Technology" *IEEE Transactions on Microwave theory and Techniques*, vol.52, no. 9, pp. 2087-2103, 2004.

Прочухан Д.В., Корабльов М.М.

Реалізація технології розпізнавання медичної маски в режимі реального часу з використанням відеокамери

Проблематика. Пандемія коронавірусу (COVID-19) призвела до значної кризи охорони здоров’я у світі через швидке розповсюдження цього вірусу. Всесвітня організація охорони здоров’я надала рекомендації щодо захисту від вказаної хвороби. Однією з ефективних рекомендацій є носіння медичної маски в громадських місцях і місцях скупчення людей. Залучення людських ресурсів для контролю за виконанням цього побажання є неефективним. Необхідно провести автоматизацію процесу визначення правильно одягнутої медичної маски в режимі реального часу з використанням відеокамери. Авторами розроблено додаток, який ефективно справляється з цією задачею.

Мета дослідження. Розробка додатку розпізнавання медичної маски в режимі реального часу з використанням відеокамери, адаптованого для використання в сучасних українських реаліях з високою точністю і низькими системними вимогами.

Методика реалізації. Аналіз існуючих аналогів у світі, побудування моделі згорткової нейронної мережі, архітектура якої дозволить провести детектування і класифікацію отриманого зображення з відеокамери в режимі реального часу, створення архітектури додатку, розробка створеної моделі на мові програмування Python, тестування додатку.

Результати дослідження. Створено згорткову нейронну мережу власної архітектури. Обгрунтовано використання алгоритму Adam для оптимізації навчання і застосування бінарної перехресної ентропії у якості функції витрат. Удосконалено метод розпізнавання обличчя з використанням ознак Хаара. Отримано високі показники навчання згорткової нейронної мережі: точність на навчальній множині — 97.46%, точність на тестовій множини — 97.23%, значення функції витрат на навчальній множині — 2.37 %, значення функції витрат на тестовій множині — 2.57 %. Розроблено додаток, що складається з трьох модулів: модуль машинного навчання, модуль обробки зображення, модуль включення відеокамери та розпізнавання наявності маски.

Висновки. Додаток ефективно справляється з задачею розпізнавання наявності медичної маски в режимі реального часу. Розроблена модель у порівнянні з аналогами має менший розмір, більш просту архітектуру, не поступаючись у точності і швидкодії. Програмне забезпечення пройшло успішне тестування на різних операційних системам та апаратних конфігураціях. Додаток може бути використано в галузях, в яких існує потреба в автоматизації процесу визначення наявності і правильності одягнутої медичної маски в режимі реального часу.

Ключові слова: *згорткова нейронна мережа; детектування; розпізнавання; класифікація; алгоритм оптимізації.*