ІНФОРМАЦІЙНО-КОМУНІКАЦІЙНІ СИСТЕМИ

UDC 004.931004.94
DOI https://doi.org/10.17721/ISTS.2021.1.61-67

D. N. Andriievskyi, orcid.org/0000-0002-7644-9328,
hoku.dev01@gmail.com
V. L. Shevchenko, orcid.org/0000-0002-7435-3533,
gii2014@ukr.net
V. Petrivskyi, orcid.org/0000-0001-9298-8244,
vovapetrivskyi@gmail.com
Taras Shevchenko National University of Kyiv, Kyiv, Ukraine

MOBILE APPLICATION DEVELOPMENT FOR BLIND PEDESTRIANS TO PREVENT ROAD DANGERS

According to the last statistic researches approximately more than one billion people worldwide live with some form of visual impairment. In turn, visual impairments limit people's ability to perform daily functions and affect their quality of life and ability to interact with the world around them. In the article mobile application development for blind pedestrians to prevent road dangers is presented. Short overviews of similar applications like Alexa, Via Opta Nav, and Object Detector are described. Each of described programs has disadvantages like limited use area, real-time object detection absence, use third-party or physical devices need. As a result, the main task of the study is to investigate modern hazard classification algorithms, improve the accuracy of the algorithm and develop software that will be able to identify hazards in real-time, does not require physical devices, and is operated using the simplest possible interface. For solving presented above problem solution based on MobiNetV2 and InceptionV3 open-source models for defining objects in a photo modification is presented. The presented solution consists of several steps like image input with further preprocessing, optimization and result processing. For the image input hosts receive data from the file system or local memory, perform any preprocessing, and then transmit the preprocessed data to the TPU cores. Preprocessing calls the parser, which in turn calls the parser function, where images are preprocessed. For the optimization stochastic gradient descent optimization and momentum optimizer are used. As a result, method of image classification for real-time hazard identification has been further developed. A model layer was developed that interprets the unbalanced results of the model and provides the necessary results to prevent accidents, which increased accuracy by 20%. A mobile application for road hazard recognition for blind pedestrians has been developed using the above. Presented results confirm the efficiency of the described approach. Also, described model and approach can be improved in further investigations.

Keywords: *image processing; road conditions; pedestrian safety; danger determining; visual impairments.*

1. INTRODUCTION

Visual impairments can limit people's ability to perform daily functions and affect their quality of life and ability to interact with the world around them. Blindness, the most severe form of visual impairment, limits people's ability to perform daily functions and move without assistance. Rehabilitation of proper quality allows people with visual impairments to varying degrees to enjoy life, achieve goals, be active and productive in today's society.

It is estimated that approximately 1.3 billion people worldwide live with some form of visual impairment. In terms of long-distance vision, mild visual impairments are observed in 188.5 million people, from moderate to severe - in 217 million, while 36 million people are affected by blindness. As for nearby vision, 826 million people live with such visual impairments.

On Tuesday, on the eve of World Vision Day (2019), which is celebrated on October 10, experts published for the first time a detailed report on vision problems. Its authors conclude that the number of visually impaired and blind people is increasing, in particular due to limited access to relevant health services in low- and middle-income countries.

2. ANALYSIS OF EXISTING STUDIES AND TASK STATEMENT

2.1. EXTENSION SHOW AND TELL FOR ALEXA

Short Description: Amazon has expanded the capabilities of the audio assistant, and now a visually impaired user can bring the item to the gadget and ask Alex what it is. Currently, this feature is only available to users in the US who use Echo Show devices [1].

The disadvantages are access only for the United States and the fact that you need to have a physical device to use.

2.2. VIA OPTA NAV

Short description: An application with step-by-step navigation. Despite the fact that the application has a slightly different direction, we could not skip such an important product. This program was developed in partnership with visually impaired users [2].

You can add waypoints to the map to increase the accuracy of the route. At any time, the user can specify the location, the remaining distance and get a description of the streets crossing the route. Even with the screen reader turned off, the program voices the necessary information using the built-in function of text-to-speech. The application was developed by the medical corporation Novartis [2].

The disadvantage is that it is still a slightly different but related direction and therefore the application does not contain the functionality of real-time object detection.

2.3. OBJECT DETECTOR

Short description: The application allows the blind user to navigate in space, determine which objects are nearby, which road signs are nearby and inform about doors and ladders. This is the most related application, which was relied on almost all the time of development [3].

The application uses intelligent algorithms to identify objects. How it works: you need to launch the application, select the mode "Signs", "Objects" or "Doors and stairs", bring the camera of the smartphone in front of you, the result is displayed in visual (high contrast letters), sound (voice assistant) and tactile (special vibrations for selected objects) forms. The application can be used by both blind and deafblind people. The application supports voice assistant and Braille displays [3].

The disadvantages are a rather complex interface, many options and the inability to identify multiple groups of objects in real time.

The most important drawback that is present in all competitors is the complexity of the interface. A blind person cannot see the buttons on the screen and is unlikely to navigate the program. The goal is to avoid confusing the interface or interrupting object recognition just because the user has moved to another part of the application. It is better if there are few buttons and the best option – when there are no buttons at all. The user launches the application and immediately begins to perform the main function – a warning of danger.

An equally important disadvantage is the need to use third-party and/or physical devices. Firstly, not all blind people have the money to buy devices, as it is difficult for people with visual impairments to find work. Secondly, although a physical device has (or may have) relatively high accuracy, it can fail and give more error than the device's camera. And if the camera of the device fails, the application will not work, and it will not create an error of malfunction.

The main task of the study is to investigate modern hazard classification algorithms, improve the accuracy of the algorithm and develop software that will be able to identify hazards in real time, does not require physical devices and is operated using the simplest possible interface. The software must be implemented using widely available mobile devices and adapted for use on special embedded processors.

3. DESCRIPTION OF THEORETICAL METHODS USED

There are currently many models for defining objects in a photo, some of which are open-source ML models that can be integrated into the iOS app. These models have similar functionality, but different data sets, so they may produce different results. But the study identified two main ones –MobiNetV2 and InceptionV3. The first is advertised by Apple in its object recognition instructions [4], and the second has a well-known developer – Google. The second model was used during the work as it has a higher level of confidence, and the error in the results is unacceptable.

InceptionV3 is an image classification model, which has many different classifiers in implementation (or "under the hood"). This provides a fairly high accuracy of object recognition, but due to their large variety on the streets, can give an error. Therefore, to ensure an accurate result, two methods were investigated – input image processing and processing of model results.

Input image processing. This method did not work very well as it did not bring any statistical improvement. His main idea is to change the image according to some criteria:

- RGB tone change (increase of red, blue and/or green dots),
- cleaning background and secondary colors, highlighting image borders.

However, due to the great complexity of the images, such processing gives unexpected results and more often gives an error due to the fact that the error of the basic model is added to the error of preprocessing. Due to this, the processing of the input image was rejected.

Processing of model results. This method does not increase the error because it does not affect the results of object definition at all, but only processes them. The method searches for and ranks keywords for each hazard by:

- hazard priority,
- detection position.

Also, a very important part of the task is the definition of different groups of hazards at the same time, because the existing implementations of hazard identification in the movement of a blind pedestrian are aimed at only one group (e.g., road signs, ladders, etc.).

Inception v3 is a widely used image recognition model that has been shown to achieve an accuracy of over 78.1% in the ImageNet dataset. (The lowest error rate was obtained, giving the model first place in the image classification in ILSVRC (ImageNet Large Scale) 2015. ImageNet is a data set of more than 15 million high-resolution images from approximately 22,000 categories. ILSVRC uses a subset of ImageNet of about 1000 images in each of 1000 categories [5].

The model is the culmination of many ideas developed by many researchers over the years. It is based on the original work: "Rethinking the original architecture for computer vision" Szeged and others [5].

The model itself consists of symmetrical and asymmetrical building blocks, including convolutions, middle joints, maximum joints, concates, screenings, and fully bonded layers. Losses are calculated using Softmax [6].

The TPU version of Inception v3 is written using the TPUEstimator, an API designed to facilitate development, so you can focus on the models themselves rather than the basic hardware details. The API performs most of the low-level grunge work required to run TPU models behind the scenes, while automating common features such as saving and restoring checkpoints [6]. Proposed approach can be seen as a part of an Internet of Things [7] technologies. Besides that, Big Data can be transported in the network and pressing issue is such network stability [8, 9].

3.1. INPUT PROCESSOR

Each TPU cloud device has 8 cores and is connected to a host (CPU). Larger slices have multiple hosts. Other larger configurations interact with multiple hosts. For example, v2-256 communicates with 16 hosts [6].

Hosts receive data from the file system or local memory, perform any preprocessing, and then transmit the preprocessed data to the TPU cores. We consider these three steps of data processing performed by the host separately, and call them phases: 1) storage, 2) preprocessing, 3) transmission [6].

To ensure good performance, the system must be balanced. No matter how long the CPU takes to receive the images, decode them, and perform the appropriate preprocessing, it should ideally be slightly smaller or about the same as that spent by the TPU on the computation. If the CPU takes longer than the TPU to perform the three phases of data processing, then the execution will be associated with the host. (Note: Because TPUs are so fast, this may be unavoidable for some very simple models.) [6].

Storage begins with the creation of a data set and includes reading TFRecords from the repository. Nested data sets are created and their elements are displayed alternately [6].

Preprocessing calls the parser, which in turn calls the parser function, where images are preprocessed [6].

The transfer includes images that return a string, labels. TPUEstimator accepts the returned values and automatically transmits them to the device [6].

3.2 PREPROCESSING

Image preprocessing [10] is an important part of the system and can greatly affect the maximum accuracy that the model achieves during training. At a minimum, the image must be decoded and resized according to the model. In the case of the beginning of the image must be 299x299x3 pixels [6].

However, simple decoding and resizing will not be enough to get good accuracy. The ImageNet training data set contains 1,281,167 images. One passage through a set of educational images is called an epoch. During training, the model will need several passes through the training data set to improve its image recognition capabilities. In the case of Inception v3, the number of required epochs will be somewhere in the range of 140 to 200 depending on the total batch size [6].

It is extremely advantageous to constantly change the images before submitting them to the model, and do it in such a way that a particular image is slightly different in each era. On the one hand, a well-designed preprocessing step can significantly improve model recognition capabilities. On the other hand, too simple a pretreatment step can create an artificial ceiling with the maximum accuracy that the same model can achieve during training [6].

3.3. OPTIMIZER

The current model demonstrates the following optimizer options: SGD and momentum [6, 11].

Stochastic gradient descent (SGD) is the simplest type of optimization: the importance is pushed in the negative direction of the gradient. Despite its simplicity, some models can still get good results. The dynamics of updates can be written as follows [6, 11]:

$$w_{k+1} = w_k - \alpha \nabla f(w_k), \tag{1}$$

where w_{k+1}, w_k – importances, α – coefficient that is equal 0.9.

Momentum is a popular optimizer that often leads to faster convergence than SGD can achieve. This optimizer updates the importance, like SGD, but also adds a component in the direction of the previous update. The dynamics of the update is given by [6]:

$$z_{k+1} = \beta z_k + \nabla f(w_k), \tag{2}$$

where z_{k+1}, z_k – components in the direction of the previous update, β – coefficient that is equal 0.9, w_{k+1}, w_k – importances, α – coefficient that is equal 0.9.

In conclusion, we can say that this model is very functional and powerful, it is used in various spheres of life, but it cannot fully satisfy the condition of the problem. We need high accuracy for complex and object-laden images. Therefore, it was decided to develop and apply a service to process the results of the model.

3.4. MODEL RESULTS PROCESSION

As mentioned earlier, the program has a service processing results. Its main task is to process the results according to some criteria: priority of the result, level of danger, keywords and position in determining.

The essence of the proposed algorithm:

We have the set $Y = \{x_1, x_2, \dots, x_n\}$ containing model results.

First, we set two variables – precision and keywords. The model produces results in the form of an array of terms. We need to choose only the first few elements and do not waste time on the last elements. Therefore, we take only a certain number (precision) elements and receive the set $Y' = \{x_1, x_2, \dots, x_{Precision}\}$.

Keywords – a dictionary with keywords – objects of danger and values – arrays of keywords.

Then the essence is very simple – we go through all the values of the arrays of keywords in keywords (variable) in descending order of priority of the danger-key and look for the entry of texts with the original raw array.

Keywords' check:

$$\forall x \in Y'. \tag{3}$$

4. ANALYSIS OF RESULTS

During the work, the ability of the program to perform its main function, namely – determining the danger on the road for blind pedestrians was tested and the following results were highlighted in Table 1:

***Table* 1**

Results

Name	**Service is turned off**	**Service is turned on**	**Total number**
Correctly identified traffic lights	5	7	10
Mistakenly identified traffic lights	1	0	-
Not identified traffic lights	5	3	10
Traffic lights identification accuracy	50 %	70 %	100 %
Correctly identified autos	8	10	10
Mistakenly identified autos	2	0	-
Not identified autos	2	0	10
Auto's identification accuracy	80 %	100 %	100 %
Total accuracy	69.5 %	89.5 %	100 %

The table clearly shows the improvement of the results after turning on the service: the accuracy of traffic lights has increased by 20%, the accuracy of car detection – also by 20%. In addition, the number of misdiagnosed hazards has been reduced to 0. However, for different data sets, the increase in accuracy may differ slightly for different objects. In this case, it is not entirely clear which is more important, the accuracy of traffic lights, or the accuracy of the car. For the general case, we enter weights that reflect the importance of each of the objects:

K1 = 0.35 – coefficient of traffic lights identification danger;

K2 = 0.65 – coefficient of autos identification danger.

These coefficients were chosen as such because the chance of stumbling on the roadway (car) of a blind pedestrian is higher than he has to stumble on a traffic light. The coefficients were determined by experts.

Next, we make a convolution of the multicriteria assessment of accuracy to scalar with the help of additive convolution [9]:

$$A = \sum_{i=1}^{2} d_i K_i, \tag{4}$$

where d_i – hazard percentage.

Calculating the overall accuracy by formula 1:

$$A_{f_1} = 0.35 * 50 + 0.65 * 80 = 69.5,$$
$$A_{f_2} = 0.35 * 70 + 0.65 * 100 = 89.5,$$
$$\Delta A_f = A_{f_1} - A_{f_2} = 20.$$

So, we have a difference in overall accuracy of 20%.

As shown on the Figure 1, by turning on service accuracy increases in all aspects of hazard detection (both in traffic lights and autos). Graph shows accuracy with turned off service percentage of accuracy with turned on.

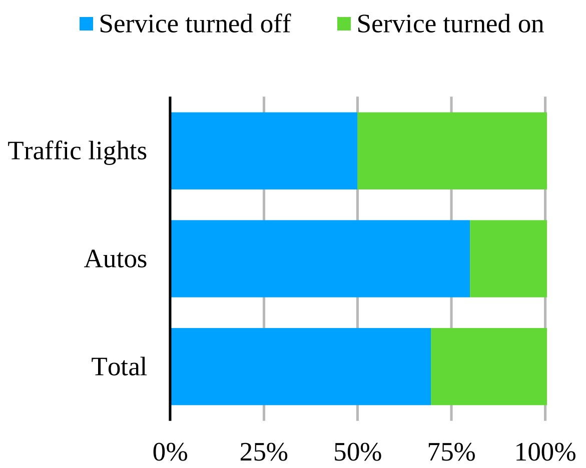


Fig. 1. Graph of improving the accuracy of hazard identification when turning on the service

During the work, a mobile iOS application was developed for the application of the above service in practice. The application uses the service and camera of the device to analyze the environment and danger messages.

The application contains 3 main components:

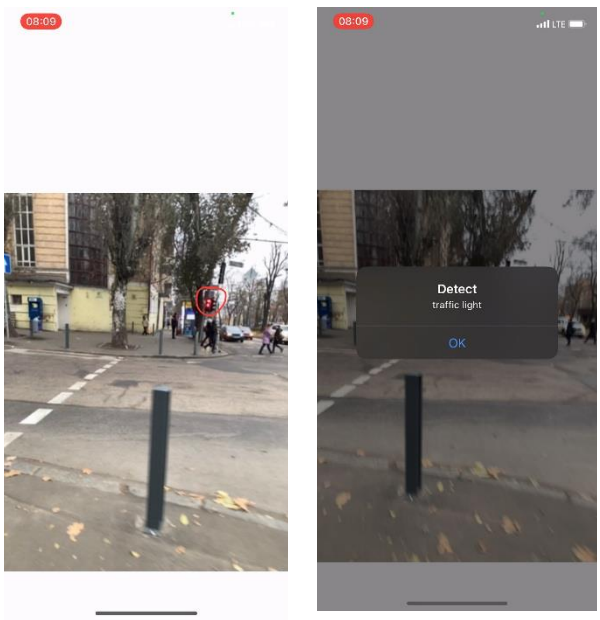
- UI-interface or user interaction part;
- Model results processing service;
- Inception V3 model.

The interface was designed to be as simple as possible so that visually impaired people would not find it difficult to use it.

Main goals of iOS application are:

- accuracy;
- scalability;
- simplicity.

As said before, iOS App has very simple interface. Also, service and its architecture were designed such as adding new road hazard type will not take much time. And great amount of accuracy was introduced by developing model result handler service.



a) b)

Fig. 2. Traffic light detection: a) a pedestrian approaches a traffic light; b) the program recognizes it and displays a warning

As shown on the Figure 2, mobile app can detect traffic light (hazard) in real time by using camera.

5. CONCLUSION

1. The method of image classification for real-time hazard identification has been further developed.

2. A model layer was developed that interprets the unbalanced results of the model and provides the necessary results to prevent accidents, which increased accuracy by 20%.

3. A mobile application for road hazard recognition for blind pedestrians has been developed using the above model and layer of work with it.

4. The application was tested under normal conditions for pedestrians, namely on city streets.

5. The application has a great extensibility, so in the future it is possible to easily add new types of hazards.

6. The software is implemented using widely available mobile devices and adapted for use on special embedded processors.

REFERENCES

[1] App Store, *Amazon Alexa App*, [Online]. Available at: https://apps.apple.com/us/app/amazon-alexa/id944011620

[2] App Store, *Nav by ViaOpta App*, [Online]. Available at: https://apps.apple.com/ua/app/nav-by-viaopta/id908435532?l=ru

[3] App Store, *Object detector App*, [Online]. Available at: https://apps.apple.com/ua/app/определитель-предметов/id1485796154?l=ru

[4] Apple, *Object detection tutorial*, [Online]. Available at: https://developer.apple.com/documentation/vision/classifying_images_with_vision_and_core_ml

[5] Sik-Ho Tsang, *Inception V3 description*, 2015, [Online]. Available at: https://sh-tsang.medium.com/review-inception-v3-1st-runner-up-image-classification-in-ilsvrc-2015-17915421f77c.

[6] Google, *Inception V3 description*, [Online]. Available at: https://cloud.google.com/tpu/docs/inception-v3-advanced.

[7] D. Evans, The Internet of Things How the Next Evolution of the Internet Is Changing Everything, Cisco Internet Business Solutions Group, 2011, [Online]. Available at: https://www.cisco.com/c/dam/en_us/about/ac79/docs/innov/IoT_IBSG_0411FINAL.pdf.

[8] V. Petrivskyi, G. Dimitrov, V. Shevchenko, O. Bychkov, M. Garvanova, G. Panayotova, P. Petrov, "Information Technology for Big Data Sensor Networks Stability Estimation," Information & Security: An International Journal, Vol. 47, Issue 1, pp. 141–154, 2020.

[9] M. Brazhenenko, V. Petrivskyi, O. Bychkov, I. Sinitcyn and V. Shevchenko, "Enabling Big Data Query with Modern CAD Systems Redundant Data Stores," in CADSM 2021, 16th International Conference on the Experience of Designing and Application of CAD Systems (CADSM), Lviv, Ukraine, February 22–26, 2021.

[10] H. Jeong, K. Park and Y. Ha, "Image reprocessing for Efficient Training of YOLO Deep Learning Networks," in 2018 IEEE International Conference on Big Data and Smart Computing (BigComp), Shanghai, China, January 15–18, 2018, pp. 635–637.

[11] G. Panayotova, G.P. Dimitrov, P. Petrov and O. Bychkov, "Modeling and dataprocessing of information systems," in Artificial Intelligence and Pattern Recognition (AIPR), Xiamen, China, September 19-21, 2016, pp. 154–158.

Стаття надійшла до редколегії 10.06.2021

Розробка мобільних додатків для незрячих пішоходів для запобігання небезпекам на дорозі

Згідно з останніми статистичними дослідженнями більше мільярда людей у всьому світі мають певні вади зору. У свою чергу, порушення зору обмежують здатність людей виконувати щоденні функції і впливають на якість їхнього життя та здатність взаємодіяти з навколишнім світом. У статті представлено розробку мобільного додатка для незрячих пішоходів для запобігання небезпекам на дорозі. Описано короткий огляд подібних програм, таких як: Alexa, Via Opta Nav та Object Detector. Кожна з описаних програм має недоліки, наприклад, обмежена область використання, відсутність виявлення об'єктів у реальному часі, застосування сторонніх або фізичних пристроїв. Як результат, основним завданням цієї роботи є дослідження сучасних алгоритмів класифікації небезпек, підвищення точності алгоритму та розроблення програмного забезпечення, яке зможе ідентифікувати небезпеки в режимі реального часу, що не потребує фізичних пристроїв і експлуатується за допомогою максимально простого інтерфейсу. Для розв'язання описаної вище проблеми використано моделі з відкритим кодом MobiNetV2 та InceptionV3 для визначення об'єктів. Подане рішення складається з декількох етапів, таких як: введення зображення з подальшим попереднім обробленням, оптимізація та оброблення результатів. Для введення зображень хости отримують дані з файлової системи або локальної пам'яті, виконують будь-яке попереднє оброблення, а потім передають попередньо оброблені дані в ядра TPU. Попереднє оброблення викликає парсер, який у свою чергу, викликає функцію синтаксичного аналізатора, де зображення попередньо обробляються. Для оптимізації використовують стохастичну оптимізацію градієнтного спуску й оптимізатор імпульсу. У результаті дослідження отримав подальший розвиток метод класифікації зображень для ідентифікації небезпеки в режимі реального часу. Розроблено модельний шар, який інтерпретує незбалансовані результати моделі та забезпечує необхідні результати для запобігання аваріям, що підвищило точність на 20 %. Розроблено мобільний додаток із використанням наведеної вище моделі для розпізнавання небезпеки дорожнього руху для незрячих пішоходів. Представлені результати підтверджують ефективність описаного підходу. Крім того, описану модель і підхід можна вдосконалити в подальших дослідженнях.

Ключові слова: *оброблення зображень; умови роду; пішохідний сейф; визначення небезпеки; порушення зору.*

Denys Andriievskyi,
Bachelor student of Department of Program System and Technologies of Taras Shevchenko National University of Kyiv.

Денис Андрієвський,
бакалавр кафедри програмних систем і технологій Київського національного університету імені Тараса Шевченка.

Viktor Shevchenko,
Dr. Sc., Professor,
Professor of Department of Program System and Technologies of Taras Shevchenko National University of Kyiv.

Віктор Шевченко,
доктор технічних наук, професор, професор кафедри програмних систем і технологій Київського національного університету імені Тараса Шевченка.

Volodymyr Petrivskyi,
PhD student of Department of Program System and Technologies of Taras Shevchenko National University of Kyiv.

Володимир Петрівський,
аспірант кафедри програмних систем і технологій Київського національного університету імені Тараса Шевченка.