UDC 621.391

A RESEARCH OF THE LATEST APPROACHES TO VISUAL IMAGE RECOGNITION AND CLASSIFICATION

Lysechko V. P. – PhD, Professor, Professor of Transport Communication Department, Ukrainian State University of Railway Transport, Kharkiv, Ukraine.

Sadovnykov B. I. – Postgraduate student of Transport Communication Department, Ukrainian State University of Railway Transport, Kharkiv, Ukraine.

Komar O. M. – PhD, Associate Professor, Associate Professor of National Aviation University, Kyiv, Ukraine.

Zhuchenko O. S. – PhD, Associate Professor, Associate Professor of Transport Communication Department, Ukrainian State University of Railway Transport, Kharkiv, Ukraine.

ABSTRACT

Context. The paper provides an overview of current methods for recognizing and classifying visual images in static images or video stream. The paper will discuss various approaches, including machine learning, current problems of these methods and possible improvements. The biggest challenges of the visual image retrieval and classification task are discussed. The main emphasis is placed on the review of such promising algorithms as SSD, YOLO, R-CNN, an overview of the principles of these methods, network architectures.

Objective. The aim of the work is to analyze existing studies and find the best algorithm for recognizing and classifying visual images for further activities.

Method. Primary method is to compare different factors of algorithms in order to select the most perspective one. There are different marks to compare, like image processing speed, accuracy.

There are a number of studies and publications that propose methods and algorithms for solving the problem of finding and classifying images in an image [3–6]. It should be noted that most promising approaches are based on machine learning methods.

It is worth noting that the proposed methods have drawbacks due to the imperfect implementation of the Faster R-CNN, YOLO, SSD algorithms for working with streaming video. The impact of these drawbacks can be significantly reduced by applying the following solutions: development of combined identification methods, processing of edge cases – tracking the position of identified objects, using the difference between video frames, additional preliminary preparation of input data. Another major area for improvement is the optimization of methods to work with real-time video data, as most current methods focus on images.

Results. As an outcome of the current research we have found an optimal algorithm for further researches and optimizations.

Conclusions. Analysis of existent papers and researches has demonstrated the most promising algorithm for further optimizations and experiments. Also current approaches still have some space for further. The next step is to take the chosen algorithm and investigate possibilities to enhance it.

KEYWORDS: machine learning, computer vision, image processing, convolutional neural networks, visual image recognition, visual image classification, algorithms, telecommunication systems.

INTRODUCTION

The field of computer vision is a promising area for the development of visual image processing systems. An example of the implementation of such a system is the image classification system, namely the analysis of medical images, the solution of which opened up the possibility of developing systems for the automatic detection of pathology in patient images [1]. Another example is the process of production automation based on automated quality control of products based on photographs [2].

The problem of recognition and classification of images in a fixed image or video stream is complex and important for many potential and existing practical applications in various fields of activity, primarily in the operation of video surveillance systems as an element of a telecommunications system.

The object of study is the process of recognition and classification of object on the video.

The subjects of study are algorithms to detect and classify objects on the video or image.

The purpose of this work is to review and analyze existing methods and approaches for recognizing and classifying visual patterns in images in order to identify possible ways to improve their performance.

1 PROBLEM STATEMENT

The task of recognizing and classifying visual images in an image is a complex one, and its solution consists of several separate steps, as shown in Fig. 1.

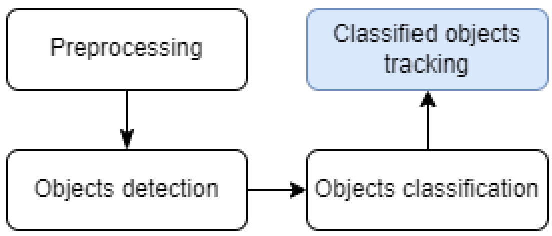

Figure 1 – Flowchart of the general method of object recognition and classification

From Figure 1, we can conclude that the method can be the same for images and video, since video is a set of images (frames) that change at a certain interval, which depends on the number of frames per second. That is, to perform the task with the input data in the form of video,

DOI 10.15588/1607-3274-2024-1-13

it is enough to search and classify objects on each frame. There is a possibility that the next frame contains few or many differences from the previous one, or no differences at all. This provides additional room for optimizations and accuracy improvement.

2 REVIEW OF THE LITERATURE

The first step is to process the input image [7]. This is a common practice in the field of computer vision, as the image usually has redundant data. First of all, these are different encodings, formats of image pixels, which make sense in everyday use, but have no impact in the case of analysis. It makes sense to convert all images to one specific format for the algorithm to work with them.

The next component to be removed is redundant color data. In many cases, converting an image to grayscale improves the accuracy of the algorithms. Simplifying the color model also provides a number of benefits in terms of resource usage – the computations become simpler, so less CPU time is consumed and memory is used more economically due to less color data.

The last possible processing is physical transformations of the image, such as resizing, cropping, flipping, and mirroring. A certain constant image size that the algorithm works with greatly simplifies implementation and adds versatility, so it is advisable to resize all input images to a certain format. It should be noted that this operation may lead to the loss of some data, so it is necessary to use the most accurate algorithms, which may lead to some deterioration in execution time.

"Cropping" the image is not appropriate in this case because the algorithm is aimed at finding objects, i.e. it is not known whether the area to be cropped contains an object or part of it.

Flipping and reflecting the input image significantly increases the amount of computation required, as each transformation is a new image to process. This step can improve accuracy, but significantly degrade speed, which can be critical for an algorithm that potentially needs to work in real time.

The next step after image preprocessing is to search for potential objects in the image. The algorithm chosen for this task must satisfy a number of criteria, such as sufficient accuracy, the ability to work with real-time data, work with images of low quality and size, and resistance to noise, changes in angles and ambient light [8].

Classical computer vision algorithms or machine learning methods are potential approaches to solving this problem.

Computer vision algorithms consist of a set of mathematical operations and transformations of image pixel data, and are usually unchanged regardless of the input data. Any modifications to suit specific input data or environments must be done manually. This significantly reduces the flexibility of the solution as it potentially needs to be adjusted to work optimally with different data sets. It is possible that there will be several parallel implementations of the algorithm for slightly different input data. The lack of self-adaptation is a disadvantage of classical algorithms and can be significant in solving the problem of object detection and classification, as it can significantly complicate the practical use of methods based on these algorithms.

An alternative approach in terms of solution architecture is to use machine learning methods instead of classical computer vision algorithms, as the focus shifts to network building rather than image data operations. The main advantage of machine learning methods is their flexibility. They are able to adapt to the input data on their own, which greatly improves their practical use. That is, we have one neural network architecture that can adjust its weights according to the input data. This reduces the number of edge cases, and the algorithm can find logical connections in the data on its own. Thus, deep learning algorithms are a good candidate for solving the problem of finding and classifying objects in an image.

Convolutional neural networks are a special class of neural networks for image processing [9]. These networks consist of interconnected layers of neurons. The main goal of this architecture is to simulate the processes that occur in the human brain when analyzing images. Convolutional neural networks are used to solve such tasks as face recognition, image classification, search for anomalies in medical images, etc. This type of network automatically extracts important image features in numerical form from pixel data and makes predictions based on them. Which image features will be extracted depends on the input data and the task set at the model training stage. Each layer of the convolutional model works with one level of image features or details, for example, the first layers work with low-level details such as borders or texture. Each subsequent layer of the model works with more abstract features.

Searching for objects in an image and classifying them can be considered as two separate tasks, which is shown in the flowchart of the general method. The result of the search is information about the coordinates of the object in the image, while the result of the classification is the type or class of the object. Thus, you first need to find the position of the objects, and then classify them.

Modern machine learning algorithms are divided into two types depending on whether they perform search and classification together or separately. Thus, there are one-pass detectors that find objects in an image and determine their class in one cycle of image analysis and two-pass detectors that first analyze the image for the presence and position of objects and then classify the found objects. Good examples of one-pass models are YOLO, SSD, and for two-pass models, the R-CNN family of models. In any case, these algorithms simultaneously find the position of the objects and their type, i.e., they perform steps 2 and 3 from Figure 1.

Currently, there are several modifications of the R-CNN algorithm, namely Fast R-CNN and Faster R-CNN [10–12]. The main purpose of the modifications was to speed up the algorithm and improve accuracy, since the original version of R-CNN processed one image for about 40 sec. [13]. This level of speed is not enough for real-

DOI 10.15588/1607-3274-2024-1-13

time data processing, so the next modification of Fast R-CNN reduced this time to about 2.5 seconds. Further improvements in the next iteration of the network, called Faster R-CNN, made it possible to reduce the image processing time even further, to about 0.2 seconds per image [14].

In the following, it makes sense to consider only the latest iteration of the algorithm, namely Faster R-CNN, since it is the latest development of this family of algorithms and has the highest speed, which is a significant indicator for the task of object recognition and classification in streaming video.

Let's look at the general architecture of the Faster R-CNN neural network in Fig. 2.

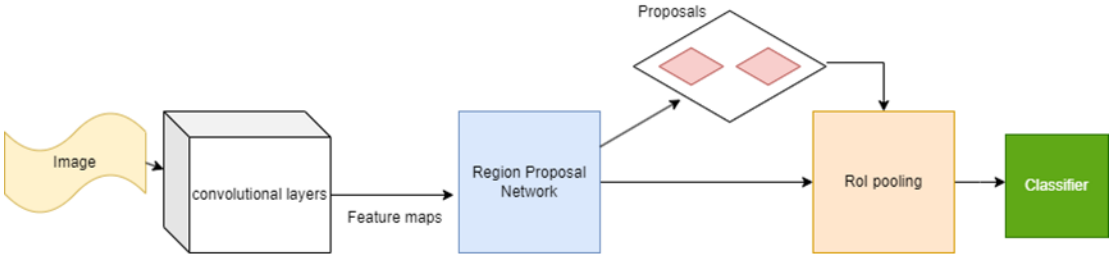

Figure 2 – structure of the Faster R-CNN network

Figure 2 shows that the network consists of two main modules. The first module is a deep convolutional network that provides suggestions for possible regions of the image that contain objects and is abbreviated as RPN (region proposal network). This optimization is the result of a study that demonstrated that based on the features extracted by the convolutional neural network, it is possible to make assumptions about regions containing objects and that working with pixel data directly is not required. RPN is an important addition to the model, as it allows to use GPU resources to search for potential regions of the image with objects. In previous iterations, the algorithm for finding potential regions was executed on the CPU. The Selective search algorithm was used for this task. Despite the fact that it is a greedy algorithm that combines groups of pixels based on low-level image features, it is slower than RPN, so these changes give a significant increase in the efficiency of the neural network as a whole.

The number of regions proposed by RPN is on average less than that of the selective search algorithm, 2000 vs. 300 on the PASCAL VOC 2007 dataset, which significantly reduces the number of detector calls and leads to a reduction in overall image processing time.

The second module is the Fast R-CNN detector, which processes the provided regions and draws conclusions about the presence and type of objects on them. An important architectural solution is the use of shared convolutional layers in the network that provides the regions and the detector. This reduces the requirements for the required amount of RAM and somewhat simplifies the overall network architecture. The disadvantage of this solution is additional complexity during training, since the same convolutional layers must be used for two different components.

Let us consider the general steps of Faster R-CNN. First, the input image, which has already passed the preprocessing stage, is transferred to the convolutional neural network to search for regions with objects on them. Regions, namely rectangles around potential objects, are determined using features calculated by convolutional layers, rather than pixel data per line as in previous iterations of the algorithm. Next, a smaller feature map is extracted from the calculated features for the regions and passed to the Faster R-CNN detector, which in turn de termines the presence of the object, class, and corrects the boundaries occupied by the object in the image.

From the network architecture and algorithm, it follows that search and classification are actually performed in two steps, so this algorithm is classified as a two-pass detector. Taking into account the speed of Faster R-CNN, this model is close to the possibility of real-time image processing, but still inferior to single-pass detectors [14]. When working with streaming video, additional processing optimizations are possible, so the algorithm may work faster with them. An important advantage of Faster R-CNN is its high accuracy and the ability to recognize small objects.

Another promising algorithm for the task of searching and classifying objects in an image is YOLO (You Only Look Once) [15]. Similar to R-CNN, there are many iterations and variations of this algorithm. One of the most recent versions of the model is YOLOv8 [16].

Unlike R-CNN, YOLO is not a complex algorithm consisting of several separate and interchangeable parts, but a monolithic model that performs the task of finding and classifying an object in an image. From the first to the eighth version of the model, a very significant number of changes took place. Due to the structure of the model, the main changes were in the framework model, for example, changing the Darknet24 framework model to the Darknet53 model in the third version, changes in the model training algorithm, input data, training parameters, and adjustments to the overall architecture of the model.

The main feature of the YOLO family of models is a constant focus on balancing speed and accuracy in order to provide sufficient efficiency for real-time operation without significant loss of accuracy. This balance changes slightly between different iterations of the model. The latest versions of the algorithm have several versions with different balances between these characteristics to opti-

DOI 10.15588/1607-3274-2024-1-13

mize performance on different types of instruments. The lightweight models are optimized for use on embedded devices and have the lowest accuracy in order to work in resource-constrained environments, while larger models require more resources but have higher accuracy.

The first step of the model is to divide the image into small cells of size S x S, where S depends on the model configuration. For each cell, the confidence that an object is located there and the boundaries of that object are calculated. The confidence also reflects how accurately the boundaries are calculated. Also, each cell represents a specific class of object inside. Only one class is calculated for a cell. This is a disadvantage of the model that if several small objects are in a cell, only one will be found and classified. Next, cells of the same type are merged to form the final boundaries and position of the found object, as shown in Fig. 3.

Figure 4 shows that according to the algorithm, objects are detected and classified in one pass through the image, so this algorithm is considered a single-pass detector. According to well-known studies, the speed of the latest versions of YOLO algorithms, including YOLOv8, exceeds 100 frames per second, so they can be used for real-time image processing [17].

SSD (Single Shot Detector) is an algorithm for recognizing and classifying visual images in an image that uses convolutional neural networks [18]. The main goal of developing the algorithm was to increase the speed of operation compared to the previous advanced YOLO algorithm and improve the recognition accuracy. The goal was to achieve an accuracy similar to two-pass detectors such as R-CNN or Faster R-CNN and to allow the algorithm to work in real time.

Consider the architecture of the SSD model in Fig. 4.

Figure 4 shows that the algorithm consists of two main parts. The first part is the main deep neural network, which is used to calculate image features. To calculate features, it is possible to use a trained classifier model. To do this, the top classifier layer of the network is removed to access the feature maps. The initial SSD implementation uses the VGG16 network without the classifier layer [19].

The next part is several convolutional SSD layers that process the image feature maps to find object boundaries and classify them.

The SSD model is also a single-pass detector like YOLO, so there are a number of similar steps in their algorithms. First, SSD divides the input image into cells, and each cell is responsible for searching for an object in that area of the image. The search is a calculation by each cell of the probability of finding an object of a particular class in that region. Then the boundaries of the found objects are formed. Unlike YOLO, image boundaries are formed using possible object boundaries, or anchor

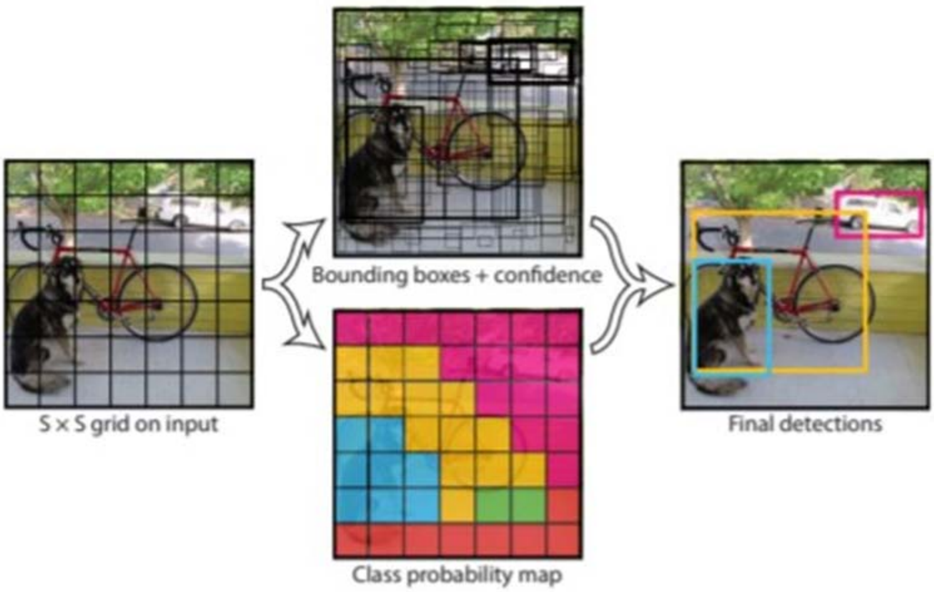

Figure 3 – YOLO algorithm flow

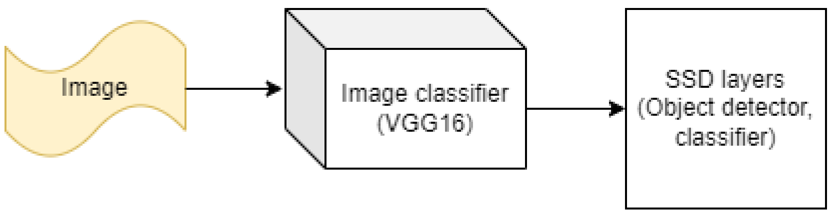

Figure 4 – SSD network structure

DOI 10.15588/1607-3274-2024-1-13

boundaries, determined at the training stage. Each anchor boundary has its own size, shape, and position inside the cell. Thus, the boundary that has the largest common area with a potential object is considered its boundary, and the object type is determined using this boundary. It is important to note that feature maps in different layers of the convolutional network are analyzed. That is, the size of feature maps changes, which leads to a change in the size of the image region described by each feature. Thus, the size of the cells into which the image is divided is different, while the size of the anchor cells remains fixed, which allows you to find objects of different sizes. Due to the anchor boundaries, SSD can find several objects in one cell, unlike YOLO. After that, for each boundary, the probability of finding an image of a particular class in it is displayed.

A critically small object is a visual object whose set can be located in a single cell into which the algorithm divides the image, i.e., one cell can contain 2 or more critically small objects. The main advantage of SSD is the ability to recognize critically small objects. According to a well-known study, the SSD model working with input images of 300 x 300 pixels has an accuracy of 3% higher than the Faster R-CNN, which worked on the basis of the VGG16 model [18]. The processing speed of the SSD300 was 59 frames per second, which is sufficient for real-time work.

The last optional step is to track the positions of the detected and classified objects. This step is only possible for video, and will allow you to avoid using the search algorithm for each frame. The main advantage of this approach is resource savings due to the use of a faster computing algorithm for tracking images on streaming video. Potential image tracking algorithms will be discussed in more detail in the following works.

3 MATERIALS AND METHODS

Let's consider a few metrics to setup a common performance measurement system for different algorithms.

The first crucial part of the neural network algorithm's efficiency is accuracy. Usually it displays how many correct answers network gave. In the current case it can be a combination of the correct class labels for detected objects and correctness of detected objects positions and borders.

Another important aspect is the time consumption for detection and classification. There can be plenty of options like time spend per frame, but the best one for the current problem is frame processed per second metric. We want to use an algorithm in real time processing, the primary characteristic for real time video is an amount of frames per second, in order to map algorithm's performance to real world, we can calculate amount of frames algorithm process per second. So it's straightforward to make a conclusion, can it process real time video data or not if we use an FPS as an efficiency metric.

4 EXPERIMENTS

Let's gather experimental data related to discussed algorithms.

We will use two primary metrics discussed previously: accuracy and FPS.

Figure 5 shows a diagram of the accuracy of the considered algorithms for searching and classifying objects.

It can be seen that the YOLO algorithm provides the ability to process images in real time with an accuracy lower by about 1.26 times relative to the slower Faster R-CNN algorithm.

Figure 6 shows a comparative speed chart of the methods.

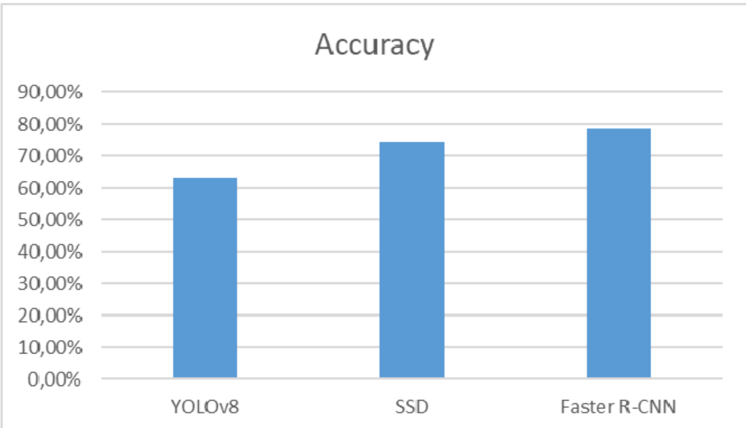

Figure 5 – Diagram of the accuracy of algorithms

DOI 10.15588/1607-3274-2024-1-13

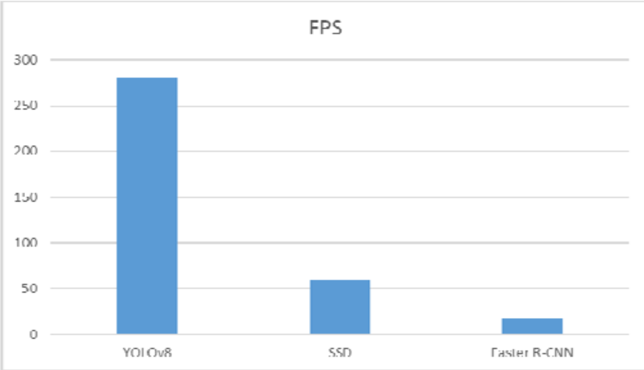

Figure 6 – Algorithms speed chart

5 RESULTS

As a result of the analysis of modern methods for searching and classifying visual images in an image, three promising neural network models were selected for further experimental research on the possibilities of optimizing them for more efficient work with streaming video data.

From the graphs shown in Figures 5 and 6, the Faster R-CNN method has a higher accuracy of approximately 80%, but insufficient speed for real-time work, 18 times lower than YOLO. SSD has a speed 5 times lower than YOLO, but sufficient for real-time operation and accuracy similar to Faster R-CNN.

Thus, for real-time work, it is advisable to choose the SSD algorithm, as it has a better balance between accuracy and speed. In addition, it makes sense to experiment with other algorithms, since accuracy and speed depend on the dataset and the task, it is possible that the algorithms will demonstrate different characteristics for the current task.

REFERENCES

1. Yue W., Liu S., Li Y. An Efficient Pure CNN Network for Medical Image Classification, *Applied Sciences,* 2023, No. 13(16), P. 9226.
2. Cui W., Zhang Y., Zhang X., Li L., Liou F. Metal Additive Manufacturing Parts Inspection Using Convolutional Neural Network, *Applied Sciences,* 2020, No. 10(2), P. 545.
3. Lysechko V. P., Syvolovskyi I. M., Shevchenko B. V. et al. Research of modern NoSQL databases to simplify the process of their design, *Academic journal: Mechanics Transport Communications*, 2023, Vol. №21, Issue 2, article №2363, pp. 234–242
4. Lysechko V. P., Zorina O. I., Sadovnykov B. I. et al. Experimental study of optimized face recognition algorithms for resource – constrained, *Academic journal: Mechanics Transport Communications*, 2023, Vol. №21, Issue 1, article №2343, pp. 89–95.
5. Mohana, Ravish Aradhya H. V Design and Implementation of Object Detection, Tracking, Counting and Classification Algorithms using Artificial Intelligence for Automated Video Surveillance Applications, Conference, *24th International Conference on Advanced Computing and Communications,* 2022, pp. 56–60.
6. Feroz A., Sultana M., Hasan R. et al. Object Detection and Classification from a Real-Time Video Using SSD and YOLO Models, *Computational Intelligence in Pattern Recognition,* 2021, 405 p.
7. Seker M., Köylüoğlu Y., Celebi A., Bayram B. Effects of Open-Source Image Preprocessing on Glaucoma and Glaucoma Suspect Fundus Image Differentiation with CNN [Electronic resource], 2021, Access mode: https://doi.org/10.21203/rs.3.rs-1695441/v1.
8. Sadovnykov B., Zhuchenko O., Perets K. Overview of state-of-the-art image object detection and classification approaches, *Collection of scientific papers of UkrDUZT International scientific and technical conference "Development of scientific and innovative activity in transport",* 2023, Issue 177. Kharkiv, UkrDUZT, pp. 46–48.
9. Sharada K., Alghamdi W., Karthika K., Alawadi A., Nozima G., Vijayan V. Deep Learning Techniques for Image Recognition and Object Detection, *E3S Web of Conferences* 2023, Vol. 399, Article Number 04032, pp. 234–243.
10. Girshick R., Donahue J., Darrell T., Malik J. Rich feature hierarchies for accurate object detection and semantic segmentation, *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2014, pp. 86–114.
11. Girshick R. Fast R-CNN, *IEEE International Conference on Computer Vision (ICCV)*, 2015, pp. 112–123.
12. Ren S., He K., Girshick R. et al. Faster R-CNN: Towards real-time object detection with region proposal networks, *Neural Information Processing Systems (NIPS)*, 2015.
13. Mijwil M., Aggarwal K., Doshi R. et al. The Distinction between R-CNN and Fast R-CNN in Image Analysis: A Performance Comparison, *Asian Journal of Applied Sciences,* 2022, No. 10(5), pp. 429–437.
14. Dong W. Faster R-CNN and YOLOv3: a general analysis between popular object detection networks, *Journal of Physics Conference Series,* 2023, No. 2580(1), № 012016.
15. Redmon J., Divvala S., Girshick R., Farhadi A. You Only Look Once: Unified, Real-Time Object Detection, *IEEE Conference on Computer Vision and Pattern Recognition*, 2016, 118 p.
16. Terven J., Cordova-Esparza D. A comprehensive review of YOLO: from YOLOv1 and beyond, *arXiv: 2304.00501,* 2023, 125 p.
17. Li Y., Fan Q., Huang H. A Modified YOLOv8 Detection Network for UAV Aerial Image Recognition, *MDPI Innovative Urban Mobility*, 2023, pp. 35–45.
18. Liu W., Anguelov D., Erhan D., Szegedy C., Reed S., Fu C., Berg A., SSD: Single Shot MultiBox Detector, *arXiv 1512.02325,* 2016.
19. Simonyan K., Zisserman A. Very Deep Convolutional Networks for Large-Scale Image Recognition, *arXiv:1409.1556*, 2014, pp 202–212.

Received 05.01.2024.
Accepted 28.02.2024.

DOI 10.15588/1607-3274-2024-1-13

УДК 621.391

ДОСЛІДЖЕННЯ НОВІТНІХ ПІДХОДІВ ДО РОЗПІЗНАВАННЯ ТА КЛАСИФІКАЦІЇ ВІЗУАЛЬНИХ ЗОБРАЖЕНЬ

Лисечко В. П. – канд. техн. наук, професор, професор кафедри транспортних комунікацій, Український державний університет залізничного транспорту, Харків, Україна.

Садовников Б. І. – аспірант кафедри транспортних комунікацій, Український державний університет залізничного транспорту, Харків, Україна.

Комар О. М. – канд. техн. наук, доцент, доцент Національного авіаційного університету, Київ, Україна.

Жученко О. С. – канд. техн. наук, доцент, доцент кафедри транспортних комунікацій, Український державний університет залізничного транспорту, Харків, Україна.

АНОТАЦІЯ

Актуальність. У статті представлено огляд сучасних методів розпізнавання та класифікації візуальних образів на статичних зображеннях або у відеопотоці. Будуть розглянуті різні підходи, включаючи машинне навчання, поточні проблеми цих методів та можливі вдосконалення. Обговорюються найбільші проблеми пошуку та класифікації візуальних зображень. Основний акцент зроблено на огляді таких перспективних алгоритмів, як SSD, YOLO, R-CNN, огляді принципів роботи цих методів, мережевих архітектур.

Мета. Метою роботи є аналіз існуючих досліджень та пошук найкращого алгоритму розпізнавання та класифікації візуальних зображень для подальшої діяльності.

Метод. Основним методом є порівняння різних факторів алгоритмів з метою вибору найбільш перспективного. Існують різні показники для порівняння, такі як швидкість обробки зображень, точність.

Існує ряд досліджень та публікацій, в яких пропонуються методи та алгоритми розв'язання задачі пошуку та класифікації образів на зображенні [3–6]. Слід зазначити, що найбільш перспективні підходи базуються на методах машинного навчання.

Варто зазначити, що запропоновані методи мають недоліки, пов'язані з недосконалою реалізацією алгоритмів Faster R-CNN, YOLO, SSD для роботи з потоковим відео. Вплив цих недоліків можна суттєво зменшити шляхом застосування наступних рішень: розробка комбінованих методів ідентифікації, обробка крайніх випадків – відстеження положення ідентифікованих об'єктів, використання різниці між відеокадрами, додаткова попередня підготовка вхідних даних. Іншим важливим напрямком вдосконалення є оптимізація методів для роботи з відеоданими в реальному часі, оскільки більшість сучасних методів орієнтовані на зображення.

Результати. В результаті проведеного дослідження було знайдено оптимальний алгоритм для подальших досліджень та оптимізацій.

Висновки. Аналіз існуючих робіт та досліджень показав найбільш перспективний алгоритм для подальших оптимізацій та експериментів. Також існуючі підходи все ще мають певний простір для розвитку. Наступним кроком є робота над обраним алгоритмом та дослідження можливостей його вдосконалення.

КЛЮЧОВІ СЛОВА: машинне навчання, комп'ютерний зір, обробка зображень, згорткові нейронні мережі, розпізнавання візуальних образів, класифікація візуальних образів, алгоритми, телекомунікаційні системи.

ЛІТЕРАТУРА

1. An Efficient Pure CNN Network for Medical Image Classification / W. Yue, S. Liu, Y. Li // Applied Sciences. – 2023. – No.13(16). – P. 9226.
2. Metal Additive Manufacturing Parts Inspection Using Convolutional Neural Network/ [W.Cui, Y. Zhang, X. Zhang et al.] // Applied Sciences. – 2020. – No. 10(2). – P. 545.
3. Research of modern NoSQL databases to simplify the process of their design/ [V. P. Lysechko, I. M. Syvolovskyi, B. V. Shevchenko et al.] // Academic journal: Mechanics Transport Communications. – 2023. – Vol. №21, Issue 2, article №2363. – P. 234–242.
4. Experimental study of optimized face recognition algorithms for resource – constrained / [V. P. Lysechko, O. I. Zorina, B. I. Sadovnykov et al.] // Academic journal: Mechanics Transport Communications. – 2023. – Vol. №21, Issue 1, article №2343. – P. 89–95
5. Design and Implementation of Object Detection, Tracking, Counting and Classification Algorithms using Artificial Intelligence for Automated Video Surveillance Applications, Conference / [Mohana, H. V Ravish Aradhya] //24th International Conference on Advanced Computing and Communications. – 2022. – P. 56–60.
6. Object Detection and Classification from a Real-Time Video Using SSD and YOLO Models / [A. Feroz, M. Sultana, R. Hasan et al.] // Computational Intelligence in Pattern Recognition. – 2021. – 405 p.
7. Effects of Open-Source Image Preprocessing on Glaucoma and Glaucoma Suspect Fundus Image Differentiation with CNN [Electronic resource]/ [M. Seker, Y. Köylüoğlu, A. Celebi, B. Bayram]. – 2021 – Access mode: https://doi.org/10.21203/rs.3.rs-1695441/v1.
8. Sadovnykov B. Overview of state-of-the-art image object detection and classification approaches / B. Sadovnykov O. Zhuchenko, K. Perets // Collection of scientific papers of UkrDUZT International scientific and technical conference “Development of scientific and innovative activity in transport”. –2023. – Issue 177. – Kharkiv, UkrDUZT. – P. 46–48.
9. Deep Learning Techniques for Image Recognition and Object Detection/ [K. Sharada, W. Alghamdi, K. Karthika et al.] // E3S Web of Conferences. – 2023. – Vol. 399, Article Number 04032. – P. 234–243.
10. Rich feature hierarchies for accurate object detection and semantic segmentation / [R. Girshick, J. Donahue, T. Darrell, J. Malik] // IEEE Conference on Computer Vision and Pattern Recognition (CVPR). – 2014. – P. 86–114.
11. Girshick R. Fast R-CNN / R. Girshick // IEEE International Conference on Computer Vision (ICCV). –2015. – P. 112–123.
12. Faster R-CNN: Towards real-time object detection with region proposal networks/ [S. Ren, K. He, R. Girshick et al] // Neural Information Processing Systems (NIPS). – 2015.
13. The Distinction between R-CNN and Fast R-CNN in Image Analysis: A Performance Comparison/ [M. Mijwil, K. Aggarwal, R. Doshi et al.] // Asian Journal of Applied Sciences 2022. – No. 10(5). – P. 429–437.
14. Dong W. Faster R-CNN and YOLOv3: a general analysis between popular object detection networks / W. Dong // Journal of Physics Conference Series. – 2023. – No. 2580(1), № 012016.
15. You Only Look Once: Unified, Real-Time Object Detection / [J. Redmon, S. Divvala, R. Girshick, A. Farhadi] // IEEE Conference on Computer Vision and Pattern Recognition. – 2016. – 118 p.
16. Terven J. A comprehensive review of YOLO: from YOLOv1 and beyond / J. Terven, D. Cordova-Esparza // arXiv: 2304.00501. – 2023. – 125 p.
17. Li Y. A Modified YOLOv8 Detection Network for UAV Aerial Image Recognition/ Y. Li, Q. Fan, H. Huang // MDPI Innovative Urban Mobility. – 2023. – P. 35–45.
18. SSD: Single Shot MultiBox Detector // [W. Liu, D. Anguelov, D. Erhan et al.] // arXiv 1512.02325. – 2016.
19. Simonyan K. Very Deep Convolutional Networks for Large-Scale Image Recognition / K. Simonyan, A. Zisserman // arXiv:1409.1556. – 2014. – P. 202–212.

DOI 10.15588/1607-3274-2024-1-13