Issue III, November 2020
ISSN 2707-9481
ISBN 978-601-323-207-2 https://doi.org/10.31643/2020.010

Aziza Srazhdinova
Al-Farabi Kazakh National University, Kazakhstan
E-mail: aziza0167@gmail.com
ORCID ID 0000-0003-1963-0005

Asel' Ahmetova
Al-Farabi Kazakh National University, Kazakhstan
E-mail: baltabekova_1994@ mail.ru
ORCID ID 0000-0002-7718-6478

Sunvar Anvarov
Al-Farabi Kazakh National University, Kazakhstan
E-mail: anvarov.sunvar@gmail.com
ORCID ID 0000-0002-8330-0716

Detection and tracking people in real-time with YOLO object detector

Abstract: In this article, we wrote not a large program to solve tasks for detection and tracking objects in real-time. The program was written in Python programming language. For object detection, a convolutional neural network was used with YOLOV3 architecture. A preliminary analysis was carried out of several variations of YOLO with CNN models. In the article, we justify why we want to use YOLO, and what it is and how to use and process the model output. We will also present the code in the form of a flowchart and as a result of the program's performance, we will show a picture of the program's operation in real-time, which was launched at one of the live lectures at the University.

Keywords: Neural network, YOLO, CNN, detection.

Cite this article as: Srazhdinova A., Ahmetova A., Anvarov S., (2020). Detection and tracking people in real-time with YOLO object detector. Challenges of Science. Issue III, p.: 69-75. https://doi.org/10.31643/2020.010

Introduction

Computer vision technologies are very common. They are used for recognition of faces, pedestrians, objects, for medical analysis, navigation of autonomous cars and in many other areas. In connection with the growth of computing power and the emergence of large image databases, it became possible to train deep neural networks - neural networks with a large number of hidden layers. Convolutional Neural Networks, which each year since 2012 won the ImageNet Large Scale Visual Classification Challenge (ILSVRC) [1], was particularly successful in the task of pattern recognition. We decided to investigate various object detectors to determine the best that we will use in our program. In our work, tracking will also be used to track and count people on the premises, which is rapidly developing along with applications in retail stores, cars with automatic control, security and surveillance systems, motion capture systems, and so on.

Methods

A new approach to detecting objects is called You Only Look Once (YOLO). As the first method, completely throwing away the conveyor, it defines object detection as a regression problem in a spatially separated bounding box and probabilities of a related class, which are predicted using one neural network from complete images in one estimate [2]. In addition, YOLO selects GoogLeNet, but not VGG-16 as the network base.

The base of YOLO is also called YOLO Version 1 (YOLOv1) [3]. YOLO models detection as a regression of a problem. One convolutional network simultaneously predicts many bounding boxes and class probabilities for these boxes.

YOLO Version 2 (YOLOv2) is an improved model compared to YOLO, which retains the advantage at speed [4]. Using the new, multi-level training method of the same YOLOv2 model can work with different sizes, offering an easy compromise between speed and accuracy.

YOLO offers a completely new way of image processing, which is very different from not only Faster R-CNN, but also R-CNN and all its variants. There are key differences between YOLO and Faster R-CNN such as:

1. Framework

Although both Faster R-CNN and YOLO use CNN as the core, and their main goals are to find the best CNN-based separation method, their scope is very different from each other.

2. Speed

YOLOv3 is an advanced version of the YOLO architecture. It consists of 106 convolutional layers and better detects small objects compared to its predecessor YOLOv2. The main feature of YOLOv3 is that there are three layers at the output, each of which is designed to detect objects of different sizes.

YOLO or You Only Look Once is CNN's very popular architecture, which is used to recognize multiple objects in an image. The main feature of this architecture compared to others is that most systems apply CNN several times to different regions of the image; in YOLO, CNN is applied once to the entire image at once. The network divides the image into a kind of grid and predicts bounding boxes and the likelihood that there is a desired object for each section. The advantages of this approach is that the network looks at the entire image at once and takes into account the context when detecting and recognizing an object. Also, YOLO is 1000 times faster than R-CNN and about 100x faster than Fast R-CNN. YOLOv3 is an advanced version of the YOLO architecture. It consists of 106 convolutional layers and better detects small objects compared to its predecessor YOLOv2. The main feature of YOLOv3 is that there are three layers at the output, each of which is designed to detect objects of different sizes. An analysis of object detectors by speed, performance and accuracy was carried out (Table 1).

Table 1. Analysis of object detectors for speed, performance and accuracy

Method	mAP(%)	FPS
Faster R-CNN	74,2	12,29
YOLO	64,71	44,41
SSD512	74,74	22,86
YOLOv3	89,69	58,31

Object tracking

Simple Online and Real-Time Tracking (SORT) [5] solve the tracking problem in two stages: first, the problem of detecting objects of the required classes in the frame is solved, and then their comparison with the detections obtained in the previous frames is performed. Each detection is described by a bounding rectangular area of interest. If the detection of the same object class on consecutive frames matches, these detections belong to the same track [5].Generic Object Tracking Using Regression Network (GOTURN) [6] is a type of tracker based on convolutional neural networks (CNN). Using all the advantages of CNN trackers, GOTURN is significantly faster thanks to offline learning without online fine tuning. The GOTURN tracking system solves the problem of tracking a single target: given the object's border frame label in the first frame of the video, we track this object through the rest of the videos. Track before detect (TBD) uses a "multi-frame detection" strategy [7] to achieve the goal, and it requires both spatial and temporal information. The algorithm tracks the paths of more than one candidate during the detection process, and also estimates the a posteriori probability for each of them, which will be compared with a certain baseline value at the end of the process.

Results

For this particular task, we don't need a data set with marked data, meaning we use a pre-trained model that already specializes in similar areas. Our ability is to correctly adapt them to the situation we are interested in. Human detection is a classic application in Computer Vision, and many models are trained to recognize this standard class, achieving high performance. We chose YOLOV3 for object detection because it provides a good compromise between performance and speed [8]. We have researched, experimented, failed, experimented again, and finally achieved very good accuracy thanks to real-time tracking on a peripheral device with a small amount of computation. Detection is the first step before we can perform tracking. After detecting people using YOLOv3, we need a tracking algorithm to track these "objects" by frames. To do this, we used a very popular algorithm called SORT (Simple Online Real-Time Tracking). It determines the state of each track based on detecting the block center, block scale, block aspect ratio, and their time derivatives (i.e. speed) [9]. We presented the code used as a flowchart with comments (figure 2-4) and displayed the result via a webcam in real time (figure 5):

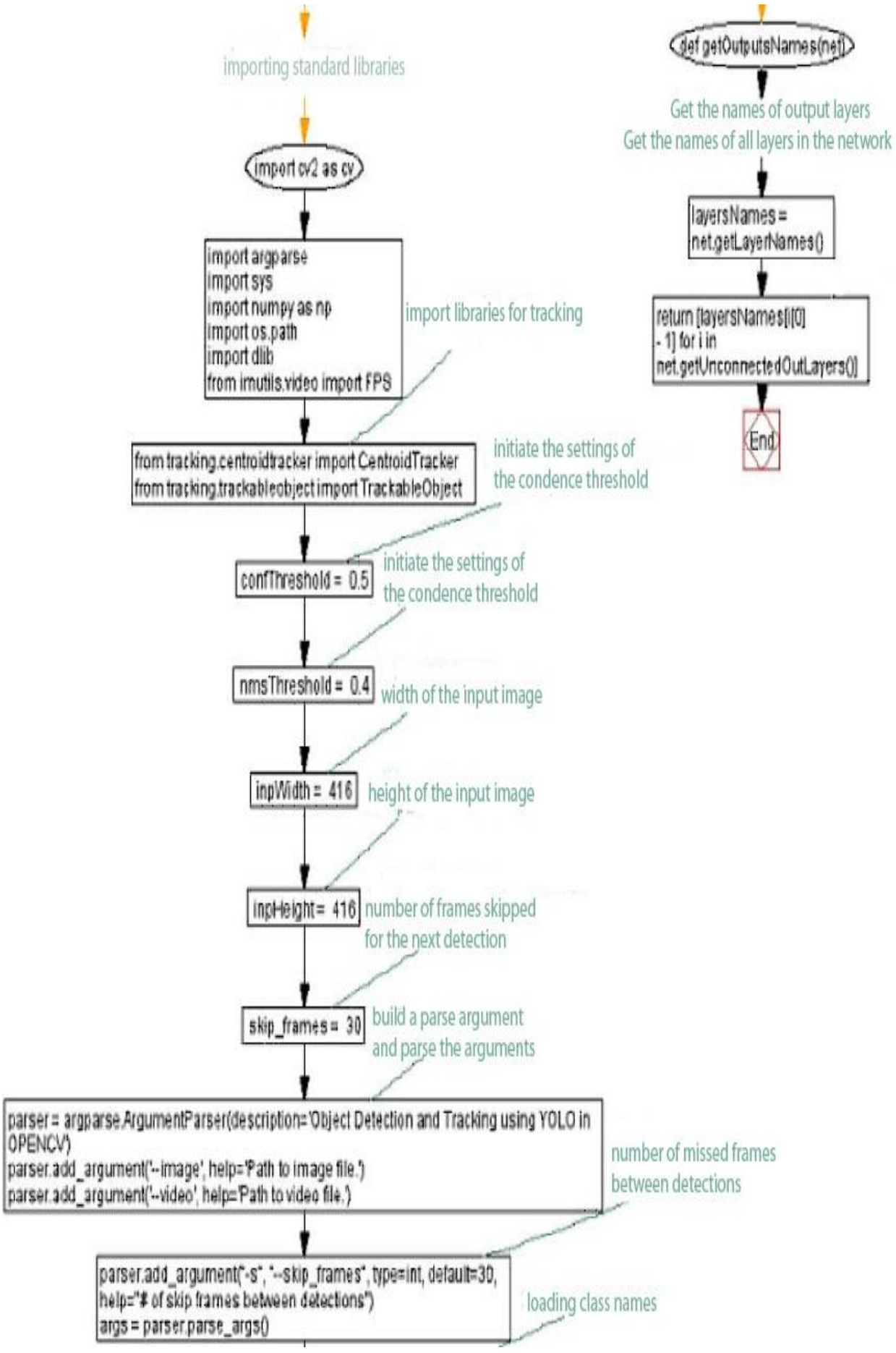

Fig. 2. Flowchart of a part of the tracking algorithm for tracking these "objects" by code frames

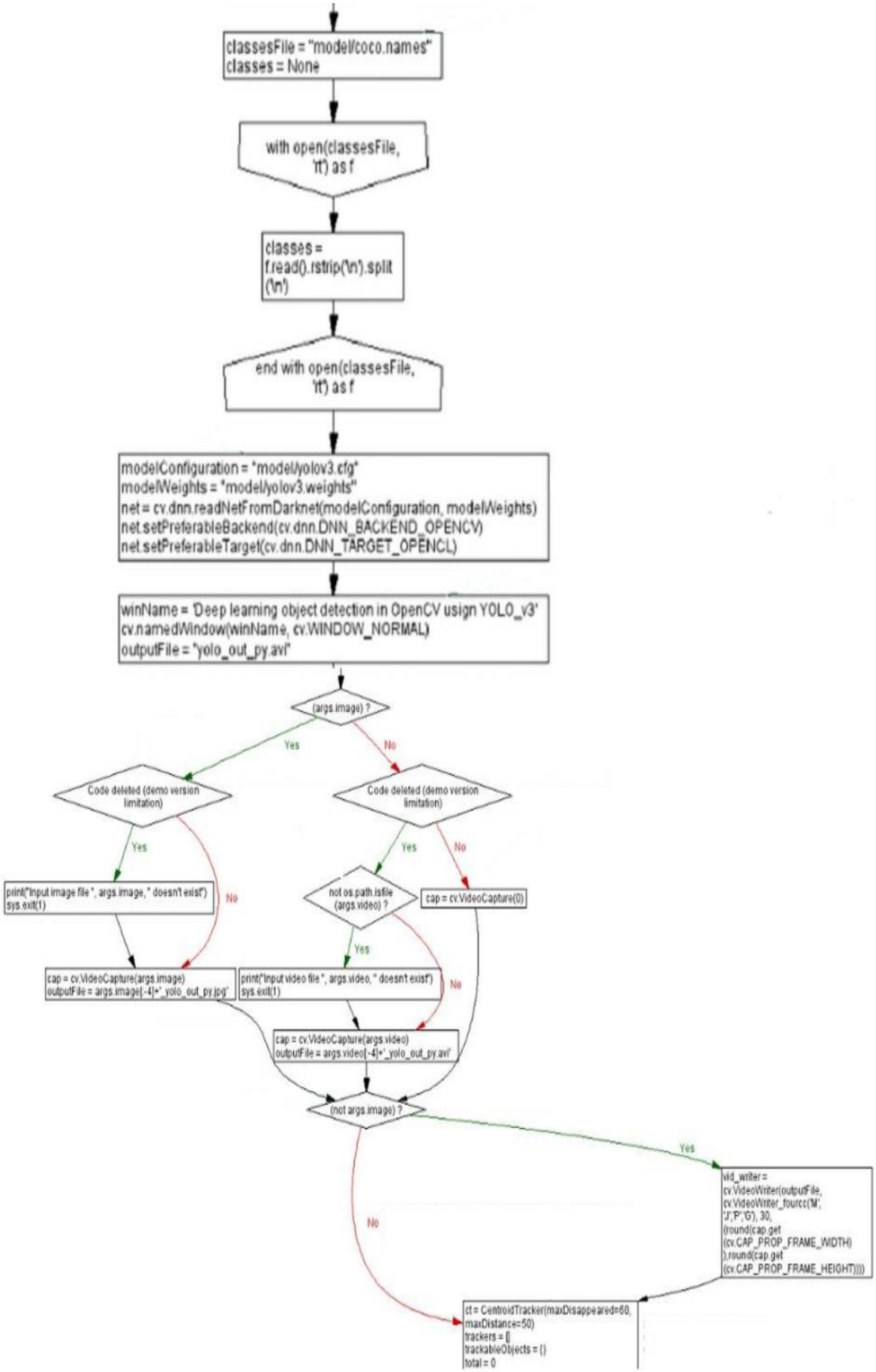

Fig. 3. Flowchart 2nd part of the tracking algorithm for tracking these "objects" by code frames

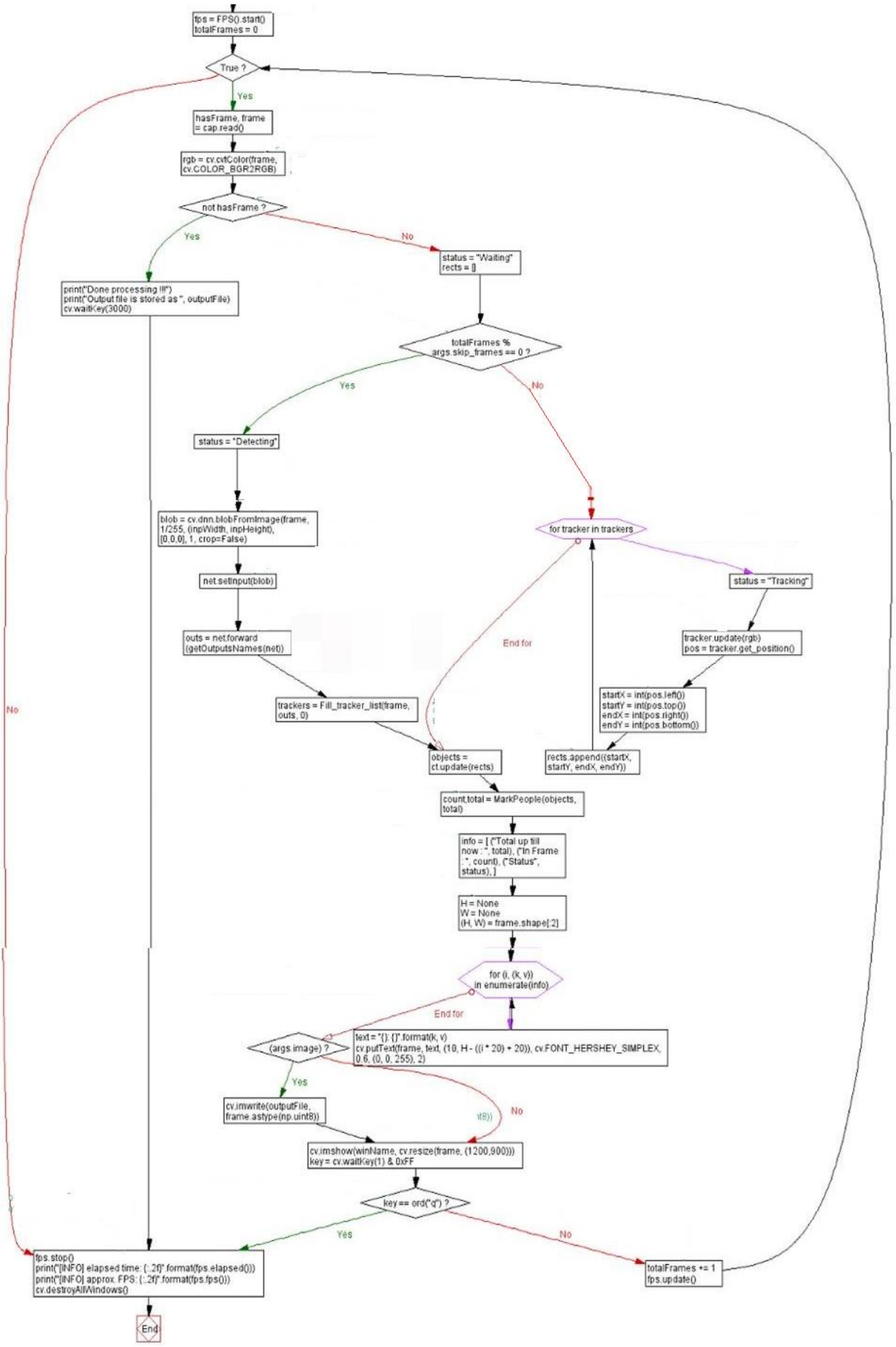

Fig. 4. Flowchart 3rd of the tracking algorithm for tracking these "objects" by code frames

Fig. 5. Practical use

Conclusion

In our article, we can use a part of the security system that in case of emergencies will track how many people were in the room, and how many people left the room to prevent adverse consequences. Currently, more and more, especially in our city, there are uncontrolled Gorenje structures, which are accompanied by deplorable results not only in material terms, but also concerning the lives of citizens. A thorough analysis of object detectors and tracking algorithms was performed to obtain the most accurate and fast-performing software product. Another advantage of digital technologies is that they can be also used in education to improve the teaching and learning processes; they also lead to human development if used properly in the right way [10-18].

Нақты уақыт режимінде YOLO нысандар детекторын қолдану арқылы адамдарды анықтау және бақылау

Азиза Сраждинова
Әл-Фараби атындағы Қазақ Ұлттық Университеті, Қазақстан
E-mail: aziza0167@gmail.com
ORCID ID 0000-0003-1963-0005

Асель Ахметова
Әл-Фараби атындағы Қазақ Ұлттық Университеті, Қазақстан
E-mail: baltabekova_1994@ mail.ru
ORCID ID 0000-0002-7718-6478

Сунвар Анваров
Әл-Фараби атындағы Қазақ Ұлттық Университеті, Қазақстан
E-mail: anvarov.sunvar@gmail.com
ORCID ID 0000-0002-8330-0716

Аннотация. Осы мақалада біз нақты уақыт режимінде нысандарды табу және қадағалау мәселесін шешуге арналған шағын бағдарлама жаздық. Бағдарлама Python бағдарламалау тілінде жазылған. Нысандарды анықтау үшін YOLOv3 архитектурасымен конвульсиялық нейрондық желі қолданылды. CNN модельдерімен YOLO бірнеше түрлеріне алдын-ала талдау жүргізілді. Мақалада біз YOLO-ны не үшін пайдаланғымыз келетіндігімізді және YOLO деген не екенін, модель шығысын қалай пайдалану және өңдеу керектігін негіздейміз. Сонымен қатар, блок-схема түріндегі кодты көрсетеміз және бағдарламаның жұмыс істеу қабілеттілігінің нәтижесі ретінде нақты уақыт режимінде бағдарлама жұмысының суреті келтіріледі, ол университетте ағынды дәрістердің бірінде іске қосылды.

Кілттік сөздер: Нейрондық желілер, YOLO, CNN, детектор қолдану.

Осы мақалаға сілтеме: Srazhdinova A., Ahmetova A., Anvarov S., (2020). Detection and tracking people in real-time with YOLO object detector. Challenges of Science. Issue III, p.: 69-75. https://doi.org/10.31643/2020.010

References

1. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C., & Fei-Fei, L. (2015). ImageNet Large Scale Visual Recognition Challenge. International Journal of Computer Vision, 115(3), pp. 211-252. https://doi.org/10.1007/s11263-015-0816-y
2. Juan Du (2018). Understanding of Object Detection Based on CNN Family and YOLO. Journal of Physics Conference Series 1004(1):012029, pp. 4-5. https://doi.org/10.1088/1742-6596/1004/1/012029
3. Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 779-788.
4. Redmon, J., & Farhadi, A. (2016). YOLO9000: better, faster, stronger. arXiv preprint arXiv:1612.08242.
5. Bewley, G. Zongyuan, F. Ramos, B. Upcroft (2016). Simple online and real-time tracking in ICIP. pp. 3464-3468.
6. David Held, Sebastian Thrun and Silvio Savarese (2016). Learning to track at 100 fps with deep regression networks. In European Conference Computer Vision (ECCV). pp. 749–765.
7. Lisha He, Lijun Xie, Tian Xie, Haibin Pan, and Yao Zheng (2012). An Effective TBD Algorithm for the Detection of Infrared Dim-Small Moving Target in the Sky Scene pp.249-251. https://doi.org/10.1007/978-3-642-35286-7_32
8. Priya Dwivedi (2019). Real-Time Person Tracking on the Edge with a Raspberry Pi.
9. Marco Cerliani (2019). People Tracking with Machine Learning.
10. Arpentieva, M. R., Kassymova, G., Kenzhaliyev, O., Retnawati, H., Kosherbayeva, A. (2019). Intersubjective Management in Educational Economy. Challenges of Science. https://doi.org/10.31643/2019.004
11. Kassymova G. K., Duisenbayeva Sh. S., Adilbayeva U. B., Khalenova A.R., Kosherbayeva A. N., Triyono M. B., Sangilbayev O. S. Cognitive Competence Based on the E-Learning. International Journal of Advanced Science and Technology Vol. 28, No.18, (2019), pp.167-177. http://sersc.org/journals/index.php/IJAST/article/view/2298
12. Kenzhaliev B.K., Kul'deev E.I., Luganov V.A., Bondarenko I.V., Motovilov I.Y., Temirova S.S. (2019). Production of Very Fine, Spherical, Particles of Ferriferous Pigments from the Diatomaceous Raw Material of Kazakhstan. Glass and Ceramics, 76(5-6), 194–198. https://doi.org/10.1007/s10717-019-00163-w
13. Triyono, B.M., Mohib, N., Kassymova, G.K., Pratama, G.N.I.P., Adinda D., Arpentieva, M.R. (2020). The Profile Improvement of Vocational School Teachers' Competencies. Vysshee obrazovanie v Rossii = Higher Education in Russia. Vol. 29, no. 2, pp. 151-158. https://doi.org/10.31992/0869-3617-2020-29-2-151-158
14. Gasanova R.R. Kassymova G.K., Arpentieva M.R., Pertiwi F. D., Duisenbayeva Sh. S., (2020). Individual educational trajectories in additional education of teachers. Challenges of Science. Issue III, p.: 59-68. https://doi.org/10.31643/2020.009
15. Kenzhaliyev, B. K., Surkova, T. Y., & Yessimova, D. M. (2019). Concentration of rare-earth elements by sorption from sulphate solutions. Kompleksnoe Ispol′zovanie Mineral′nogo syr′â/Complex Use of Mineral Resources/Mineraldik Shikisattardy Keshendi Paidalanu, 3(310), 5–9. https://doi.org/10.31643/2019/6445.22
16. Apendiyev T.A., & Abdukadyrov N.M. (2020). During the first world war germany and austria – hungary prisoners of the aulieata county. The Bulletin, 1(383), 218–225. https://doi.org/10.32014/2020.2518-1467.27
17. Apendiyev, T.A.; Zhandybaeva, S.S.; Tulebaev, T.A.; Abykenova, K.E. (2017). The Migration of Germans to Kazakhstan in the end of XIX –beginning XX century. Bylye Gody, Volume 44, Issue 2, 2 June 2017, Pages 568-575. https://doi.org/10.13187/bg.2017.2.568
18. Apendiyev, T. A., Smagulov, B. K., Kozybayeva, M. M. (2019). Study of some subethnic and genealogical groups of Kazakhs in pre-revolutionary Russian historiography (XVIII – early XX century). The Bulletin, 6(382), 346–354. https://doi.org/10.32014/2019.2518-1467.180