

Application of Modern Object Tracking Technologies to the Task of Aortography Key Point Detection in Transcatheter Aortic Valve Implantation

V.V. Laptev¹, N.A. Kochergin²

Scientific Research Institute of Complex Problems of Cardiovascular Diseases, Kemerovo, Russia

¹ ORCID: 0000-0001-8639-8889, lptwlad1@gmail.com

² ORCID: 0000-0002-1534-264X, nikotwin@mail.ru

Abstract

Object detection, as one of the most fundamental and challenging problems in computer vision, has attracted much attention in recent years. Over the past two decades, we have witnessed the rapid technological evolution of object detection and its profound impact on the whole field of computer vision. In this paper, aortography key point detection approaches for transcatheter aortic valve implantation based on machine learning tools are discussed. The paper provides a description and analytical comparison of such popular methods as "object detection", "pose estimation". As a result of this study, a visual assessment system is proposed to facilitate the performance of the intervention procedure. The final accuracy of the proposed system reaches 79.3% with an analysis speed of 12 ms per image.

Keywords: machine learning, object detection, tracking, key points.

1. Introduction

Today in medical practice and the field of cardiology the application of methods of automatic processing of graphic data does not stop growing. The list of the most frequently used ones includes algorithms for processing anatomical structures based on Magnetic Resonance Imaging (MRI) and computed tomography (CT) data. Unfortunately, in a number of cases the application of the presented modalities is impossible. The main limitations of computed tomography include the lack of real-time mode. In its turn, such a method of contrast examination of blood vessels as fluoroscopy can be used for noninvasive diagnostics directly during interventions, possessing at the same time the possibility to reduce the radiation load up to 90%, and for visual assistance to the surgeon during interventions it is necessary to develop and implement an intelligent algorithm of data tracking and visualization.

Transcatheter aortic valve implantation (TAVI) is a relatively new and highly effective method of treating patients with moderate and high-risk aortic stenosis. Short-term and long-term patient survival after TAVI is similar to that after surgical aortic valve replacement [1, 2]. The number of TAVI procedures has steadily increased since the first procedure performed in 2002, and the indications for TAVI continue to expand [3]. During the operation, the time available for the doctor to analyze images is limited, so the development of visual assistance systems for intraoperative guidance is of paramount importance.

Some interventional angiography systems integrate commercially available software to facilitate navigation during TAVI and reduce the risk of complications. Currently, such products have been developed by Philips (HeartNavigator) [4], Siemens Healthcare (syngo Aortic Valve Guide) [5], GE Healthcare (Innova HeartVision) [6], and Paieon Inc (C-THV) [7] and have been successfully implemented in clinical practice. Existing guidance systems align a

three-dimensional anatomical model of the aortic root based on preoperative CT with live fluoroscopic images during valve positioning, ensuring optimal orientation of the angiography system and vascular access. However, these systems do not allow real-time tracking of key points and detailing the geometry of the aortic root during the operation, as they assume preoperative model reconstruction [8]. The time required to perform automated CT analysis using the HeartNavigator by constructing a three-dimensional volume is on average 2.1 minutes. Consequently, the operator still bears the responsibility for managing the device's position and deployment using aortography data and catheter tip tracking. A logical step forward is the development of visual assistance systems that enable real-time (analysis rate of at least 1 frame per second) tracking of catheter key points and aortic root contour using automated aortography image processing, regardless of the imaging equipment.

In 2021, researchers from the Department of Experimental Medicine at the Research Institute of Complex Cardiovascular Diseases proposed a method for detecting key points in aortography during transcatheter aortic valve implantation based on multitask learning [9]. It is worth noting that this method is sensitive to the presence of all points of interest in the image and is computationally complex, casting doubt on its feasibility in real-time conditions. Despite its high effectiveness on the test set, reproducing the experiment on new patients did not meet expectations.

As one of the modern methods of key point tracking, the "object detection" technology has been considered. This technology has proven itself well in solving computer vision tasks. Confining objects of analysis to a bounding box in real-time is demonstrated in works [10-12] with high accuracy metrics. The second modern tracking approach discussed in the article is based on "pose estimation." This technology has a distinctive foundation in object tracking, specifically in searching for and tracking key points of the analyzed object, making it well-suited for the task at hand. This approach has also proven effective in solving computer vision tasks [13-15].

2. Source Data

To solve the given task, a custom dataset was collected and annotated, consisting of 35 videos with a size of 1000×1000 pixels. The final sample consisted of 3730 grayscale images. Each image underwent a meticulous annotation procedure by experts. The resulting dataset was divided into two parts, with 2932 (80%) images used as the training set and the remaining 798 (20%) images used as the test set. The division was done based on patients, allowing for a more accurate evaluation of the proposed method. TAVI provided a series of anonymous images illustrating three main stages:

- Catheter positioning and delivery system Fig. 1 A.
- Initiation of capsule retraction and exposure of the prosthesis, deployment of the transcatheter aortic valve, and rotation of the drive Fig. 1 B.
- Prosthesis deployment, 1/3 valve opening Fig. 1C.

Each image was marked with a varying number of points, depending on the presence of an object of a specific class:

- Marking of key points on the catheter Fig. 1D.
- Marking of key points indicating the aortic root Fig. 1E.
- Marking of key points on the valve stent during its deployment at 1/3 Fig. 1F.
- Visualization of key points in the distal part of the delivery system through segmented aortograms Fig. 1G.
- Three-dimensional model of the target structure of the aortic valve Fig. 1H.

A maximum of 11 key points of interest (from 1 to 11 on each image) were labeled and annotated Fig. 2 D-H», namely:

1. Anatomical landmarks:

- Aortic ring, target reference for TAVI: aortic root 1 (AA1) and aortic root 2 (AA2).

- Aortosinotubular junction, an additional reference for correctly determining the plane of the aortic ring: sinotubular junction 1 (STJ1) and sinotubular junction 2 (STJ2).
2. Delivery system landmarks:
- Anchors of the delivery system, a reference determining the degree of prosthesis extraction - proximal catheter (CP).
 - Catheter bending point, a reference to the sinotubular part of the stent - middle of the catheter (CM).
 - Radiopaque marker strip of the capsule on the top of the shaft up to the distal ring, a reference for the degree of bending of the outer shaft, used to determine the degree of prosthesis extraction - distal catheter (CD).
 - Catheter tip, a reference determining the position of the catheter and the plane of the aortic ring - catheter tip (CT).
3. Additional references:
- Distal part, a reference for valve implantation, indicating the plane of the aortic ring - pigtail (PT).
 - Distal part of the self-expanding prosthesis determines the position of the stent during implantation and its deviation from the plane of the aortic root - distal part of the stent: frame edge 1 (FE1) and frame edge 2 (FE2).

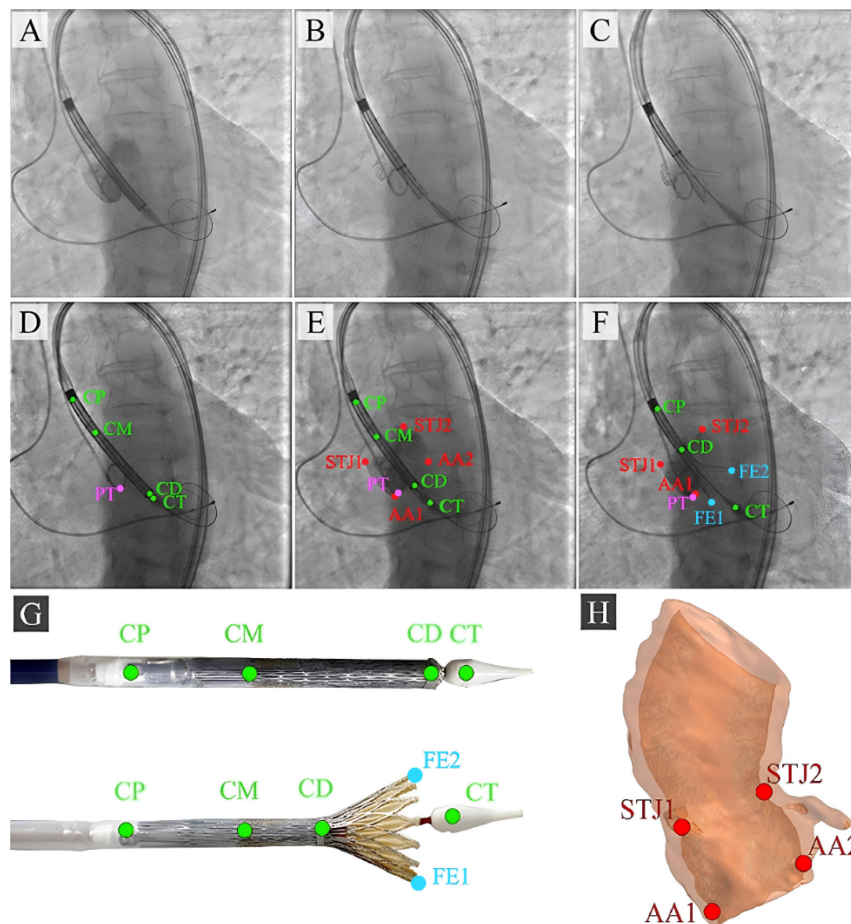


Fig. 1. An algorithm for labelling intraoperative aortography images and identifying key points for the TAVI tracking system

During the training phase of neural network models, the training set underwent a data augmentation procedure, allowing for a significant expansion of the diversity of input instances. We selected 7 data modification functions that were applied randomly:

- Random rotation of the image.
- Random zooming.

- Perspective transformation of the image.
- Vertical flipping of the image.
- Adjustment of image brightness.
- Adjustment of image contrast.
- Application of Gaussian noise.

3. Object detection

Object detection is a computer vision task aimed at identifying and locating specific objects in an image or video. The task involves determining the position and boundaries of objects in the image, as well as classifying objects into different categories. Object detection is an essential part of image recognition, along with classification and image searching.

Tracking or object detection in an image is one of the fundamental tasks in computer vision. Hence, this field has seen extensive research and development tools. One popular solution for real-time object detection is the YOLO models [16].

The YOLO family of networks has long demonstrated high-quality results and the ability to operate in real-time, which aligns with the main goal of this research. The convenience of using the provided framework is also noteworthy. As an experiment, we trained several models of different sizes with various parameters, and the experiment results are presented in Table 1. To prepare the data for training, we preprocessed the expert annotation by enclosing a key point inside a bounding box with a size of 36 by 36 pixels, which serves as the minimum detection unit for models of this type. Thus, the model learns to perform object detection with a familiar bounding box, and our focus is on its center.

Table 1. Results of detection model training experiments, YOLOV8

Model name	Optimizer	Use Batch Normalization	Speed of analysis, ms	MAP
YOLOv8n	Adam	True	4,1	0,442
	Adam	False	4,8	0,438
	SGD	True	4,1	0,46
YOLOv8s	Adam	True	7,2	0,459
	Adam	False	7,95	0,463
	SGD	True	7,2	0,468
YOLOv8m	Adam	True	9	0,482
	Adam	False	9,46	0,476
	SGD	True	9	0,519
YOLOv8l	Adam	True	18,21	0,489
	Adam	False	18,47	0,491
	SGD	True	18,2	0,521

Since the resulting quality in terms of the MAP (Mean Average Precision, a metric for measuring the accuracy of object detection) metric for the v8m and v8l models is practically indistinguishable, we chose the model with a shorter analysis time. The selected quality metric reflects a prediction accuracy between 0 and 1 based on the overlap between the two areas. For a more detailed overview of the model's performance, refer to Fig. 2. The analysis results of the image by the model are interpreted in two formats: the standard output of detected instances enclosed in bounding boxes and the visualization of detection using points. The second visualization option is more convenient when working with the image data. The presented results include 2 experiments. Also, the images in the same row are labeled with letters and the following characteristics:

- A – input image;
- B – visualization of expert annotation;
- C – visualization of the model's results.

The low performance of the models in terms of the MAP metric is related to the difficulty of determining specific classes, as confirmed by the performance indicators reflected in Table 2. It is also worth noting that these classes have a lesser impact on the original dataset.

Table 2. Results of the YOLOV8m model for each class

Class	Precision	Recall
CP	0.9645	0.99
CM	0.992	0.9846
CD	0.976	0.954
CT	0.954	0.986
PT	0.857	0.74
FE1	0.783	0.722
FE2	0.776	0.635
STJ1	0.378	0.386
STJ2	0.412	0.365
AA1	0.689	0.7476
AA2	0.678	0.7634

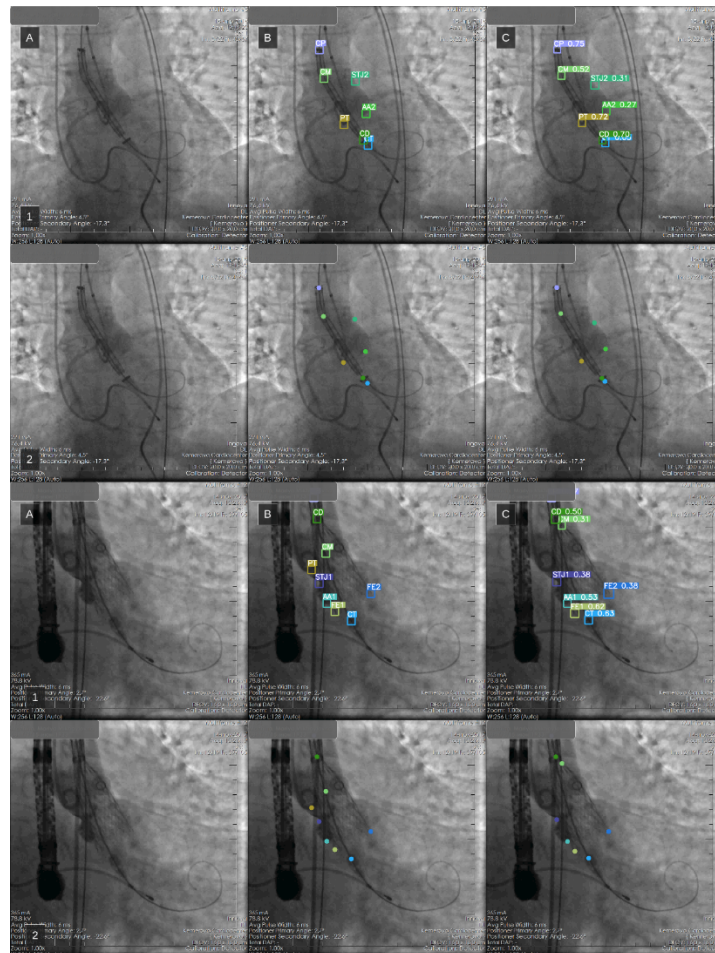


Fig. 2. Result of the «object detection» model YOLOV8m

4. Pose estimation

Pose estimation is a task that involves determining the location of specific points on an image, commonly referred to as keypoints. Keypoints can represent various parts of an object, such as joints, landmarks, or other distinctive features. The location of keypoints is usually represented as a set of coordinates. The output of a pose estimation model is a set of points representing the keypoints of the object in the image, often accompanied by confidence estimates for each point. Additionally, points related to the same object are connected by con-

necting lines (edges) and can be enclosed in a bounding box similar to the «object detection», allowing for a comprehensive evaluation of the result [17-19].

Currently, in the task of detecting and estimating poses in images, models from the YOLO family, specifically YOLOv8, are also recognized as leading, achieving accuracy scores of over 80% on the COCO dataset while maintaining image analysis speed [20].

During data preparation, we grouped points into sets (objects) to enable the detection of the entire target group. This allowed us to analyze three objects:

1. Catheter positioning in the closed state.
2. Catheter positioning in the open state.
3. Anatomical landmarks.

One drawback of this approach is the inability to obtain an unified tracking model, as the current model can only detect one object, specifying the exact number of control points. Therefore, we had to conduct training for each analyzed object individually, and the experiment results are presented in Table 3.

Table 3. Results of YOLOV8 pose estimation model training experiments

Model name	Analysis group	Speed of analysis, ms	MAP
YOLOv8n-pose	Positioning the catheter in the closed state	3,82	0,783
	Positioning the catheter in the open state		0,762
	Anatomical landmarks		0,73
YOLOv8s-pose	Positioning the catheter in the closed state	5,32	0,79
	Positioning the catheter in the open state		0,774
	Anatomical landmarks		0,75
YOLOv8m-pose	Positioning the catheter in the closed state	6,48	0,796
	Positioning the catheter in the open state		0,77
	Anatomical landmarks		0,753
YOLOv8l-pose	Positioning the catheter in the closed state	11,2	0,801
	Positioning the catheter in the open state		0,794
	Anatomical landmarks		0,785

The experiment results showed that the highest detection quality for each of the investigated objects is achieved by the YOLOv8l-pose model with an average MAP metric score of 0.793. The analysis time for one image will be 33.6 ms, meeting the conditions for real-time analysis. The details of the experimental results can be seen in Fig. 3.

Similar to the series of experiments presented in the previous section, images in the same row are labeled with letters and the following characteristics:

- A – input image.
- B – visualization of expert annotation.
- C – visualization of the model's results.

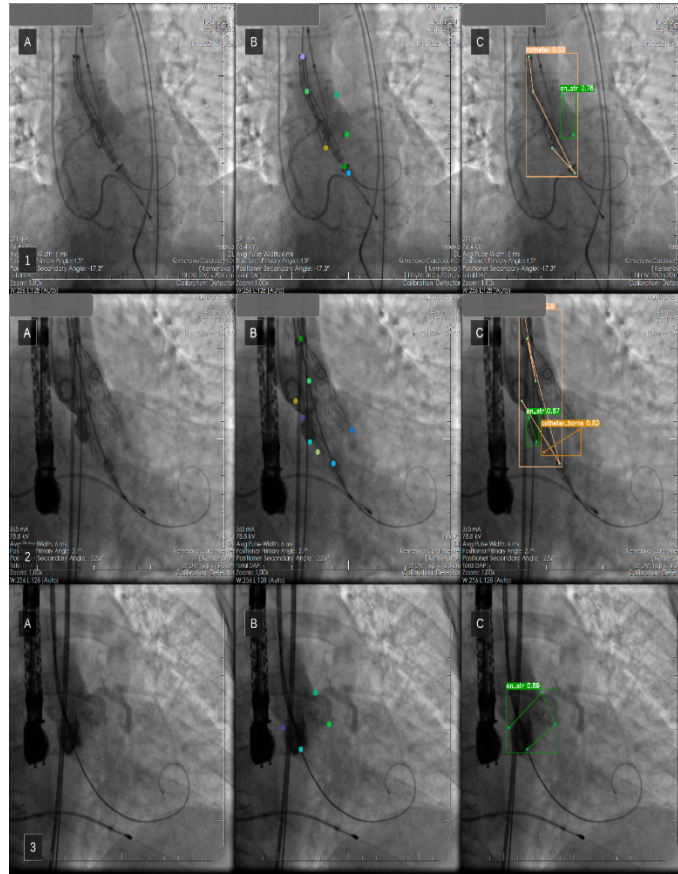


Fig. 3. Result of the «pose estimation» model YOLOV8l-pose

The network training, and testing were performed on a desktop computer

featuring an 8-core AMD Ryzen 7 5800x CPU 3.20GHz, 32 GB of RAM, and an Nvidia GeForce RTX 3060 GPU with 12GB of video memory. PyTorch v2.1 and Python v3.11 were utilized as the primary machine learning framework and language for network development, respectively.

As a result of the study, the software was developed using the API. The input data are: a DICOM data file or an image, the system automatically identifies the data type and brings it to the required format, namely a 640x640 px image in grayscale. In the role of output data, the image with anchor points applied to it is converted to the format 1000x1000 px. More details on the scheme of work can be found in Fig. 4. For ease of installation and use, the system is placed in Docker content with the definition of all necessary dependencies, which allows it to function on the following operating systems: Windows, GNU/Linux. The memory capacity of the developed software is 2.375 GB.

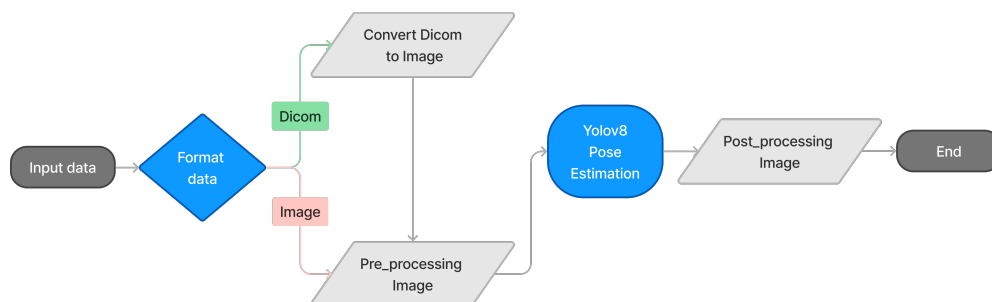


Fig. 4. Assist system operating diagram

5. Conclusion

An analysis of various approaches to the task of detecting key points in an image has been conducted. The experiments revealed that the most effective approach is based on «pose estimation» technology using the YOLOv8l-pose model. The final solution is a system based on three independent detection models with a final precision measured by the MAP quality metric of 0.793 and an analysis time not exceeding 12 ms. The presented system can recognize and track key points indicating the location of the aortic root, delivery system, and heart valve prosthesis during surgery, consolidating them into target research objects. It is anticipated that this system can be used as an auxiliary tool to optimize valve positioning and as a component of a robotic system for performing TAVI.

Acknowledgment

«The study was funded by a grant from the Russian Science Foundation № 23-75-10009, <https://rscf.ru/project/23-75-10009/>.

References

1. Abdelgawad AME, Hussein MA, Naeim H, Abuelatta R, Alghamdy S. A comparative study of TAVR versus SAVR in moderate and high-risk surgical patients: hospital outcome and midterm results. *Heart Surg Forum* 2019 (doi:10.1532/hsf.2243)
2. Baumgartner H, Falk V, Bax JJ, De Bonis M, Hamm C, Holm PJ, et al. 2017 ESC/EACTS Guidelines for the management of valvular heart disease. *Eur Heart J*, 2017 (doi:10.1016/j.rec.2017.12.013)
3. Winkel MG, Stortecky S, Wenaweser P. Transcatheter aortic valve implantation current indications and future directions. *Front Cardiovasc Med*, 2019 (doi:10.3389/fcvm.2019.00179)
4. Kocka V, Bartova L, Valoskova N, Labos M, Weichet J, Neuberg M, Tousek AP. Fully automated measurement of aortic root anatomy using Philips HeartNavigator computed tomography software: fast, accurate, or both// *Eur Heart J Suppl.* 2022 (doi: 10.1093/eurheartjsupp/suac005)
5. Jianping Gu, Wensheng Lou, Department of Interventional Radiology, Nanjing No. 1 Hospital, China, Siemens Healthcare GmbH · AT 4672-16 0720 digital, Siemens Healthcare GmbH, 2020
6. Kilic T, Yilmaz I. Transcatheter aortic valve implantation: a revolution in the therapy of elderly and highrisk patients with severe aortic stenosis. *J Geriatr Cardiol*, 2017 (doi:10.11909/j.issn.1671- 5411.2017.03.002)
7. Codner P, Lavi I, Malki G, Vaknin-Assa H, Assali A, Kornowski R. C-THV measures of self-expandable valve positioning and correlation with implant outcomes. *Catheter Cardiovasc Interv* 2014 (doi:10.1002/ccd.25594)
8. Horehledova B, Muhl C, Schwemmer C, Hendriks BMF, Eijsvoogel NG, Kietselaer BLJH, et al. Aortic root evaluation prior to transcatheter aortic valve implantation Correlation of manual and semi-automatic measurements. *PLoS One*, 2018 (doi: 10.1371/journal.pone.0199732)
9. Danilov VV, Klyshnikov KY, Gerget OM, Skirnevsky IP, Kutikhin AG, Shilov AA, Ganyukov VI and Ovcharenko EA (2021) Aortography Keypoint Tracking for Transcatheter Aortic Valve Implantation Based on Multi-Task Learning. *Front. Cardiovasc. Med* (doi: 10.3389/fcvm.2021.697737)
10. Laptev N.V., Laptev V.V., Gerget O.M. Detection of fire hazardous objects in a forest area based on dynamic features.
11. Manakov R.A., Kolpashchikov D.Y., Laptev N.V., Danilov V.V., Skirnevskiy I.P., Gerget O.M. Visual shape and position sensing algorithm for a continuum robot // 14th International Forum on Strategic Technology (IFOST-2019). Tomsk, Russia: TPU Publishing House, 2019. P. 399–402.

12. Tan M., Pang R., Le Q. V. Efficientdet: Scalable and efficient object detection //Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. – 2020. – C. 10781-10790.
13. Haralick R. M. et al. Pose estimation from corresponding point data //IEEE Transactions on Systems, Man, and Cybernetics. – 1989. – T. 19. – №. 6. – C. 1426-1446.
14. Andriluka M. et al. 2d human pose estimation: New benchmark and state of the art analysis //Proceedings of the IEEE Conference on computer Vision and Pattern Recognition. – 2014. – C. 3686-3693.
15. Ansar A., Daniilidis K. Linear pose estimation from points or lines //IEEE Transactions on Pattern Analysis and Machine Intelligence. – 2003. – T. 25. – №. 5. – C. 578-589.
16. Jiang P. et al. A Review of Yolo algorithm developments //Procedia Computer Science. – 2022. – T. 199. – C. 1066-1073.
17. Duan K. et al. Centernet: Keypoint triplets for object detection //Proceedings of the IEEE/CVF international conference on computer vision. – 2019. – C. 6569-6578.
18. Murphy-Chutorian E., Trivedi M. M. Head pose estimation in computer vision: A survey //IEEE transactions on pattern analysis and machine intelligence. – 2008. – T. 31. – №. 4. – C. 607-626.
19. Padilla R., Netto S. L., Da Silva E. A. B. A survey on performance metrics for object-detection algorithms //2020 international conference on systems, signals and image processing (IWSSIP). – IEEE, 2020. – C. 237-242.
20. Maji, Debapriya Nagori, Soyeb Mathew, Manu Poddar, Deepak. (2022). YOLO-Pose: Enhancing YOLO for Multi Person Pose Estimation Using Object Keypoint Similarity Loss. 2636-2645, 2022 (doi: 10.1109/CVPRW56347.2022.00297)