Scientific
Research
Publishing

AI in Recognition of Surgery Tools

Ciprian Dragne[1], Teodor Buliga[2]

[1]Institute of Solid Mechanics of Romanian Academy, Bucharest, România
[2]SANADOR Clinical Hospital, Bucharest, România
Email: ciprian_dragne@yahoo.com

Abstract

AI techniques are proving unbeatable in the fight for supremacy in methods of evaluation, optimization, control, object recognition, sensor monitoring, image interpretation, machine learning, rapid data access and storage, speech recognition, process automation, new learning capabilities, etc., all aspects are related to how computers will be used from now on in the future. The medical field, especially robotic surgery, will fully benefit from all the advantages of AI techniques. Monitoring surgical operations, more precise and robotic control of medical instruments, reconstruction of the anatomy and surgical scene, advanced visualization and segmentation techniques, are already methods that are applied and that use some AI techniques, but still timidly and selectively in medical areas. The article presented here proposes innovative methods for recognizing medical instruments used in surgical operations, evaluating their positioning, including in 3D space, obtaining position information based on advanced labeling information stored in the database, and CAD models of focused objects.

Keywords

AI, Robotic Surgery, Recognition, Advanced Labelling, 3D

1. Introduction

Robotic surgery allows to the medical teams to perform many types of complex medical procedures with more operational precision at the injury site, smaller incisions, lower risk of infection, less pain for the patient, reduced blood loss, faster rehabilitation, better control of the movement of medical instruments, more flexibility in actions, more complete and adequate storage of information, monitoring of the entire process, rapid access to pre-operative information, better ergonomics and reduced physical strain for the medical team. That is why, at least to date, robotic surgery proves to be the best option among all minimally invasive tech-

niques.

Medicine today is at the intersection of two significant trends. The first trend is the increasing attractiveness of robotic medical procedures, which have beneficial effects on medical outcomes. The second trend is the increasing cost of expenses, as a result of the introduction of new technologies that include new materials, complex mechanical parts, and advanced software applications. Artificial Intelligence (AI) has come to help in both trends.

The use of artificial intelligence, in particular the deep-learning subtype, has made it possible to use labeled big data, along with significantly improved computing power and cloud storage. In medicine, in particular, this is starting to have an impact on many levels: improve quality of the medical procedures, fast and accurate image interpretation, less medical errors, better access to remote diagnostic methods (telemedicine), automation of medical procedures, dedicated machine learning, 3D reconstruction, advanced sensors monitoring, storing medical information for post-treatment or educational purposes and information distribution worldwide [1].

Storing information and interpreting it quickly has become an increasingly difficult task, even in the medical domain. That is why a medical history remains crucial for timely treatments, but also very useful for current interpretations.

Education, lifestyle, eating habits, environmental factors, and the balance of all factors related to a normal life, all play a major role in the success of any general medical action. Understanding all risk factors and their risk weight can play a decisive role in the medical act. Life-threatening risks can be diagnosed at early stages using predictive analysis. This allows doctors to make prompt decisions and act effectively on the results to save lives, improve rehabilitation and reduce unnecessary costs.

Looking deeper, there are notable deficiencies in any healthcare system that have long-lasting effects on healthcare and are responsible for the overall level of healthcare quality. These include a considerable number of diagnostic errors, mistakes in establishing proper treatments, an enormous waste of resources, inefficiency in workflow, inequities and inadequate waiting time for the patient to reach a specialist.

Interpreting current medical results, most being in form of data values and medical imaging maps, also requires considerable effort from the healthcare providers. The best monitoring methods for critical medical cases imply continuous monitoring, both sensory and imaging [2].

Since the early stages of scanning and loading medical images into a computer, researchers have sought to build automated analysis systems. Measuring distances, scaling images, improving quality and contrast, advanced area selection and color segmentation, comparing results with reliable databases for similarity, and assessing the severity of the medical case, medical reconstruction, etc., were just a few of the improvements brought to the system of visualization of medical results.

The father of AI, John McCarthy, defines intelligence as the ability of humans to achieve certain goals that you set for yourself in life or that life itself sets for you. John McCarthy also provides that AI is the science of making machines with some intelligence, especially through intelligent IT programs. AI is about using computers to understand human intelligence also. But AI doesn't have to be limited to the classical methods of human intelligence, and will actually bring many improvements to it, through the way we learn and understand everything in the future [3].

No other theory or unifying paradigm established so far in history has guided research as much as AI has [4].

A sub-field of AI is machine learning. This has an unprecedented success with statistical machine learning in the 2010s when it eclipsed all other approaches. However, this approach is largely sub-symbolic, neat but weak, and narrow in approach. Critics argue that these drawbacks could be revisited by future generations of AI researchers [5].

AI is a research tool and an opportunity for humanity!

In another form of definition, artificial intelligence is a field that combines computer science routines and data sets to enable faster problem-solving. The subfields of machine learning and deep learning contain AI algorithms that aim to create expert systems that make predictions, classifications and better selections based on input data.

A deep learning neural network consists of digitized inputs (e.g. sets of values, images, sounds, etc.), which are passed through multiple layers of "neurons" which progressively detect certain well-defined specific features, and which ultimately provide the output matrix with the best similarity from all points of view [6].

Creativity has partially passed from human abilities to those of computing machines.

There is a huge interest in AI applications, including the processing of images, especially given large data sets, substantial advances in machine learning, and the opportunities presented by quantum computing and the need to develop new algorithms.

In addition to developing new AI methods themselves, there are many opportunities and challenges for the scientific community interested in imaging to develop new storage methods that include partial storage and image generation, including the development of a standard nomenclature, better ways to share images and standards for validating the use of AI programs across different imaging platforms and subject populations. These fields attracted particular attention for the application of AI in the radiology medical domain [7] [8].

1.1. Limitations and Challenges of AI

Despite all the predictions, expectations, and estimates about AI technology, there are also formidable obstacles, pitfalls, and ethical questions that remain unanswered or unresolved. AI also needs to be critically evaluated. It is recognized that there

is an exaggerated enthusiasm and empathy for artificial intelligence programs, which often leads to a limited overview or even a lack of critical scientific vision, especially when it comes to validating and preparing for the implementation of programs aimed at the medical care of a patient.

A recent example is IBM Watson Health's cancer software (known as Watson for Oncology). Used by hundreds of hospitals worldwide to recommend treatments for cancer patients, it was shown that the algorithm is built based on a small, unrealistic number of synthetic cases with very limited (real-world) data from oncologists. Many of the recommendations turned out to be erroneous treatment recommendations, such as suggesting the use of Bevacizumab in a patient with severe bleeding, which is an explicit contraindication and marked "black" on the drug's box for precisely that purpose [9].

But the involvement of PCs and artificial intelligence in the workflow of some medical institutions cannot stop here. In fact, AI will be increasingly in demand in all areas of life. The requirements of evolution and IoT (Internet of Things) will lead to new ideas and innovations by adding new ways for study and branches for research, interesting new chapters in the encyclopedia of the evolution of Information Technology (IT).

IoT is a network of physical objects that contain specially developed local devices with sensors, simpler or more complex devices, with local software, cloud applications, cloud data, and other technologies, to connect and exchange data with other devices and systems via the Internet.

1.2. Advanced Data Labeling

In machine learning, data labeling is the process of marking raw data (images, text files, videos, etc.) and identifying by adding one or more labels with meaningful and informative information to provide context of data, so that both we and the mathematical model can "learn something" from it.

For example, labels can indicate whether a photo contains a particular object (e.g. a pedestrian, a car, a traffic sign, a medical tool, etc.), which words were spoken in an audio recording and in what language, or whether an X-ray contains a tumor or other medical clues. Data labeling is needed for a variety of cases, including computer vision, malignant disease recognition of skin defects, natural language processing, and speech recognition [10].

A general observation is that it is not quite possible to build an object detection architecture with both higher accuracy and better efficiency across a wide spectrum of object classes, reference size, zooming, resource constraints, image quality, object pose or orientation, etc. [11]-[13]. All these problems should be carefully and systematically studied for various AI model design choices of detector architectures.

In our method, labeling contains more data related to particular object classification, object 2D orientation, and particular object poses.

A custom model database was developed to contain more detailed images and

labels for surgical tools in various positions.

Figure 1 shows a CAD model of a Stewart Hexapod platform for controlling and positioning the CAD surgical instruments tested in this paper.
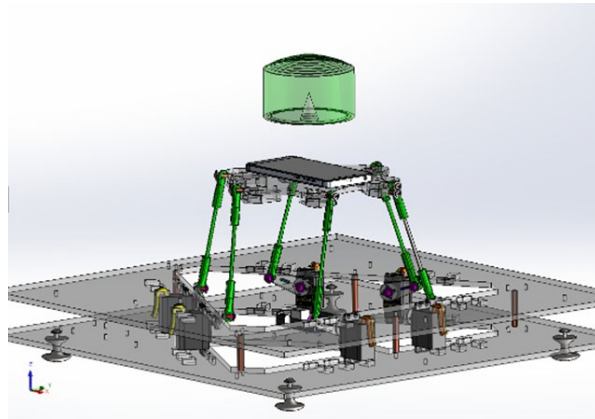

Figure 1. CAD assembly used for assessment of objects pose.

2. Methodology

The current paper presents many aspects of a complete research, with the following steps in the development of applications:

1) Object detection by common methods (YOLO, ROBOFLOW) and with images of real objects in data-sets;

2) Object detection by improved methods (A + CNN) for better confidence of results;

3) Object detection with image data-sets from CAD models;

4) Advanced labeling for additional information;

5) Special CAD application for computer vision.

Each previous step was developed to improve the overall detection efficiency in various aspects in object detection of complex assemblies and orientation of objects in 2D and 3D space.

The global efficiency of the object recognition techniques depends on:

1) Neural Network technique implemented: ANN, CNN, or other technique with additional methods to increase efficiency;

2) The number of the reference objects in the database (Ex: pretrained COCO 2017 dataset has 330k images);

3) The images size for each reference (32×32, 200×200, 320×320, 640×640, 768×768, 1024×1024, 1280×1280, 1536×1536);

4) Database format (TF, ONNX, etc.);

5) Number of classes implemented in database;

6) Software and version used to test (CNET, Local Python, etc.);

7) Additional libraries used to test (additional libraries to manage structures, to process images, etc.);

8) CPU, GPU and RAM capabilities;

9) RUN mode, latency;

10) Number of Epochs (Ex: 300);

11) Original image size and object scale;

12) Scene type: image or video;

13) Inference performances;

14) Fine tune: Detection, Segmentation, Classification, Pose or OBB;

15) Number of detections in one frame;

16) Image quality and augmentation processing: colors, illumination, contrast, HDR, adaptive contrast, black-hole-areas, color channels and swap;

17) Partial scanning of areas can improve the quality of object detection;

18) Labeling definitions and information stored for additional data;

19) Complexity of labeling implemented system: amount of information, structured data, etc.;

20) Percentages for Training Set, Validation Set and Test Sets;

21) Proper selection of references for training, validation or test groups: manually, random or specific techniques for selection;

22) Additional references in set based on images processing: rotate, skew, etc.;

23) Additional references in custom set based on augmented images processing;

24) Additional images based on 3D object CAD design.

A particular aspect of object detection in medical applications is the detection of small objects. This issue is a computer vision problem because we aim to accurately identify objects that are small in a video feed or images, even if each object is small relative to the photo size.

Detecting small objects is hard because detection models form specific features by aggregating pixels in convolutional layers. The most successful object detection techniques recommend higher-resolution images, even with a size of larger than 1280px. The management of such a dataset became difficult using custom computers.

Every new object detection technique comes with additional parameters that need to be set for better global or local performance. A major problem in computer image processing is the multiple data required to compute in cloud. This multiple datum depends on the areas to be observed and several parameters for image quality and object detection performance.

Some detection techniques are used instead of running on the whole scene, to split it into smaller parts (slices), run the model on each one, and then stitch the results together. This method even took much more time, the model gives better detection performances. Such a technique is SAHI (Slicing-Aided Hyper-Inference).

2.1. Common Object Detection Models

There are few references for surgical tools in common database sets; most of them are particular reference objects. For comparison reasons, the object detection model was also tested using YOLO11 and the ROBOFLOW online platform. Convolutional Neural Networks (CNNs) were initially developed as standardized

structures—fully connected stacked convolutional layers [14].

Various AI techniques were developed to achieve models with higher performance in object detection. From object detection technique in a single pass for speed (YOLO = You Only Look Once) with various sub-methods for improvements in recognition based on augmentation, scaling, the fusion of features between layers, the technique for optimization of memory used or semantic segmentation capabilities was only a few improvements during years in developed on YOLO methods [15].

In contrast, two-stage or multi-stage methods, such as Region-based Convolutional Neural Networks (R-CNNs), first generate region proposals and only then perform classification. These methods offer high precision but are computationally expensive in resources and time for study. Convolutional Neural Networks techniques that use studies on regions came to helping in decomposition and to improve overall detection problem performances, but also to split the global problem into a few other sub-problems easier to manage [16].

Improved methods were also developed over the years to increase the efficiency of object detection studies, like feature extraction, methods for transfer learning, or import and export of neural network models between various platforms [17].

Various on-line applications were also developed for researchers to study the object detection techniques, for labeling images, and to test their models.

The advantages of using common detection applications and models are the possibility of developing new models in the future with more connected users, more references saved in a dataset and using a specific standard.

An example with object detection test using YOLO pretrained dataset is presented in Figure 2, which shows that common applications and datasets have a limited number of classes and references for a specific object. Only one type of class object can be recognized using pre-trained datasets.

Figure 2. Object detection using pre-trained dataset and YOLO-local-application.

The use of common applications can also limit us through certain constraints in developing our research methods, limitations in the way of presenting and providing information regarding the objects subject to detection, limitations in imposing our objective functions, limitations in using a specific number of hidden layers, as well as limitations in their connectivity.

2.2. Custom Dataset for Object Detection

Reusing a pre-trained network is not the best method for our intention to detect

specific objects. The first reason is the limited number of references and label information in datasets for our objects [18] [19].

Figure 3 shows a partially custom dataset that uses 300 images were initially split into 3 classes: Fenestrated forceps, Needles driver, and Scissors.
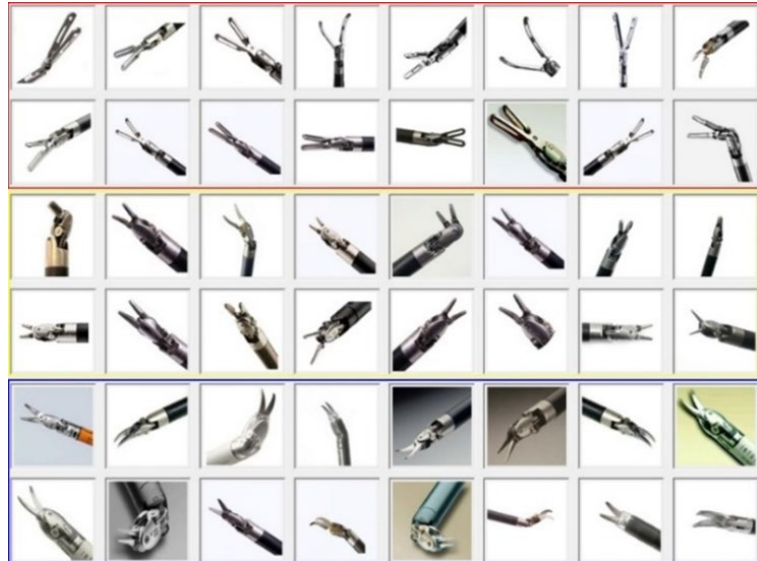

Figure 3. Custom dataset split into 3 classes.

Using custom dataset and YOLO11 training was obtained the results from Figure 4. For this image, object detection has no results using pretrained YOLO11, Google NET or ROBOFLOW applications.
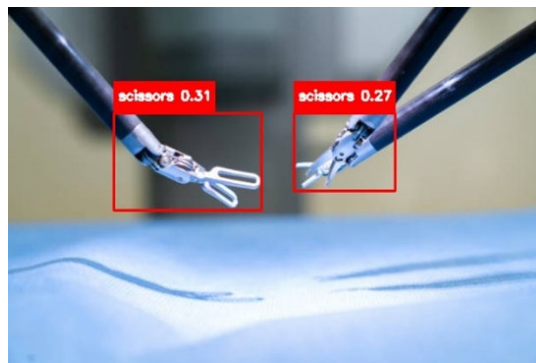
Figure 4. Object detection using custom dataset split into 3 classes.

It was also a big challenge to thoroughly verify the performance of our object detection model. Even established applications can have data processing errors for custom models. There should be checked bounding boxes, selection scaling, anchors, labeling, etc., for almost every AI training. Object detection training performances using custom model obtained with ROBOFLOW are shown in Figure 5.

RoboflowTrain Metrics
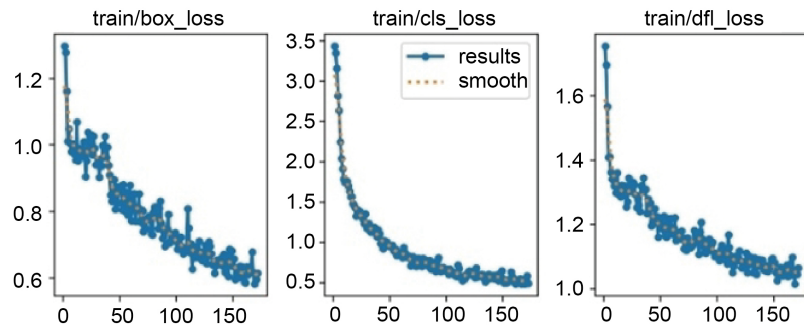
Validation Set   Test Set   Training Graphs

Figure 5. Object detection using custom dataset split into 3 classes.

During testing was observed also the possibilities to detect more than one object. This problem is due to the initial image set defined with labeling areas with too obvious similarities between object classes. A such example is shown in **Figure 6**.

Figure 6. ROBOFLOW object detection using custom dataset with 3 classes.

For problems of similarities between various areas of various objects, the initial set was redefined with 5 classes like in **Figure 7**:

1) Fenestrated forceps;
2) Needles driver;
3) Scissors;
4) Surgery stick;
5) Gears.

Performances in object detection was also extracted from custom training using YOLO11. **Figure 8** shows some specific parameters from training of the custom AI model with 5 classes. A special representation of YOLO procedure is the Auto-
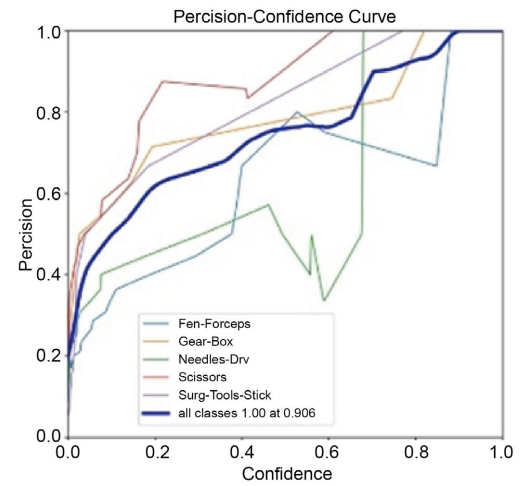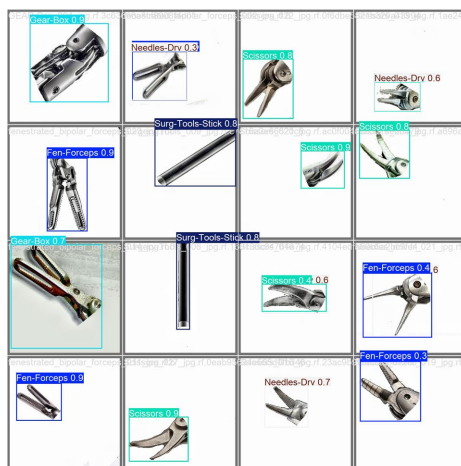
test prediction for used custom dataset (Figure 8(c)).

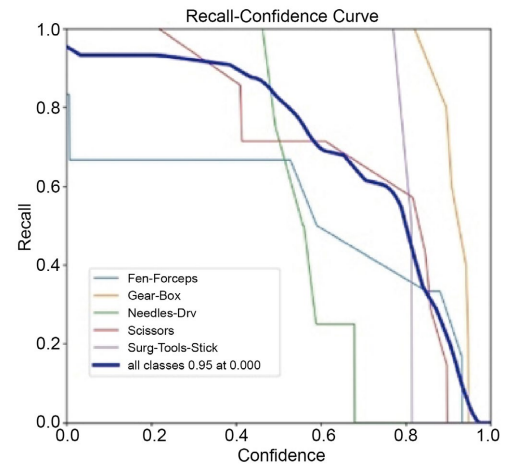**Figure 7.** Custom dataset arranged in 5 classes.
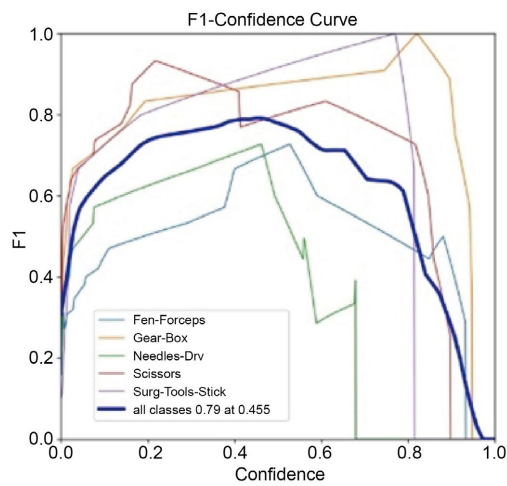
(a)

(b)

(c)

(d)

Figure 8. Performances in training of custom dataset with YOLO11: F1, Precision, Auto-test-Prediction, and R-curves.

Even though the dataset does not have too many images, and therefore the division between test, training, and validation images leads to even fewer images, it is noticeable that for some objects, the auto-test prediction is quite perfect (90%). The rest of the information of the training needs to be analyzed to add more images to the AI model and for future improvements.

The paper doesn't focus on a benchmark between YOLO and ROBOFLOW. We want to have a comparison only at confidence results obtained with ROBOFLOW-web-app and YOLO-local-app. Any other type of parameter for comparisons between YOLO and ROBOFLOW was not our concern.

2.3. Improve Images Quality to Improve Object Detection Technique

Limitations in common object detections observed during testing led to develop a method to improve the efficiency of the object recognition techniques.

One of the major problems observed in object detection results was the low performance in case of difficult capture conditions related to object illuminations and contrast.

For this reason, during this research an improved CNN technique developed was to implement preprocessing methods that use adaptive contrast and adaptive illumination for input images to improve segmentation [20]-[23]. In the next sentences for this method, we refer to A + CNN (Adaptive plus CNN).
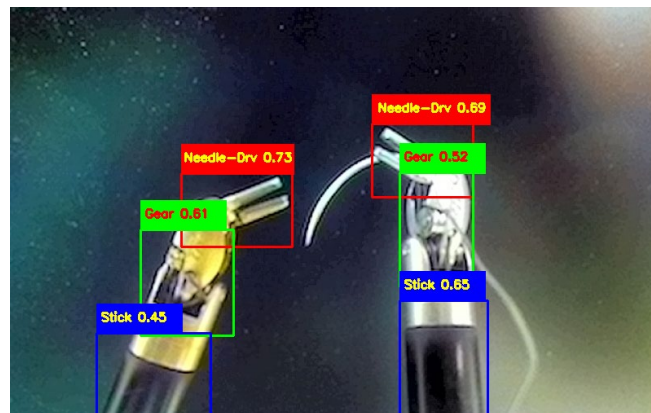

Figure 9. Object detection using A + CNN.

Figure 9 shows object detection using A + CNN based on custom CAD dataset with 5 classes. Contrast and illumination of initial image was improved before detection.

A + CNN technique can be applied also inside layers if layers use selective areas for object detection because final results in adaptive contrast and illuminations depend on the areas selected for processing. These results will be explained in future papers.

Because the best explanation is given by images, in **Figure 10**, a few examples are presented in using CNN with adaptive illumination and adaptive contrast in

technique for colorization of noir images.

Figure 10. A + CNN in colorization of noir images (on left—original images, on right—colorized image using adaptive illumination and contrast).

2.4. Object Orientation Using AI Strategies

One of the most common object orientation detection techniques is based on OBB-Detection, an oriented object detection library, which is based on the MM-detection project [24]. The Oriented Bounding Boxes (OBBs) object detection is used to detect objects oriented at various angles. The output is a set of rotated bounding boxes that exactly enclose the objects in the image, along with class labels and information data (angle, confidence, "size" of object, contour area, etc.).

Why is object orientation important for us? Because the main goal of this research is to detect specific instruments in the surgical scenes, their position, and the direction of action of these instruments. Figure 11 shows two types of orientation labelling info that can be used during detection: bounding box and local coordinate system.
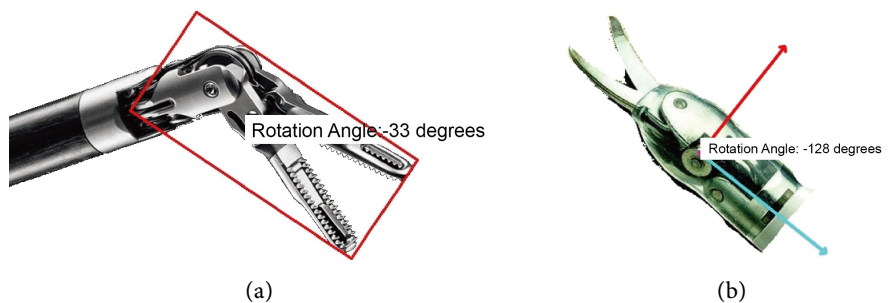

(a)   (b)

Figure 11. Results of orientation detection implemented in our method.

Object orientation detection can be evaluated based on the matrix transformation between the input and the reference. Object orientation detection focuses on iden-

tifying key features, such as edges, corners, or specific shapes, to under-stand the orientation of the object and to highlight the most important such features. Verifying the alignment of an object involves ensuring that it is exactly where it is supposed to be or should be, with the correct position and direction, including with all necessary corrective measures that depend on the viewing angle of the focused object.

The position and alignment of an object can require varying different values of precision depending on the field of activity. For example, in production, even a small misalignment can cause product defects. Similarly, in robotics, precise positioning is required for tasks such as accurately reaching a specific target location. In the medical field, precise localization and orientation are understood to be even more important.

The object orientation detection method even allows a selective assessment of particular objects with specific information (see Figure 12) based on specific requirements: confidence, angle, size, object color, etc. [25]-[31]. An example of using AI object detection in the evaluation of distances and object sizes is presented in the paper [32].

The object orientation can even be represented selectively when we focus only on specific classes. An example is shown in Figure 12.
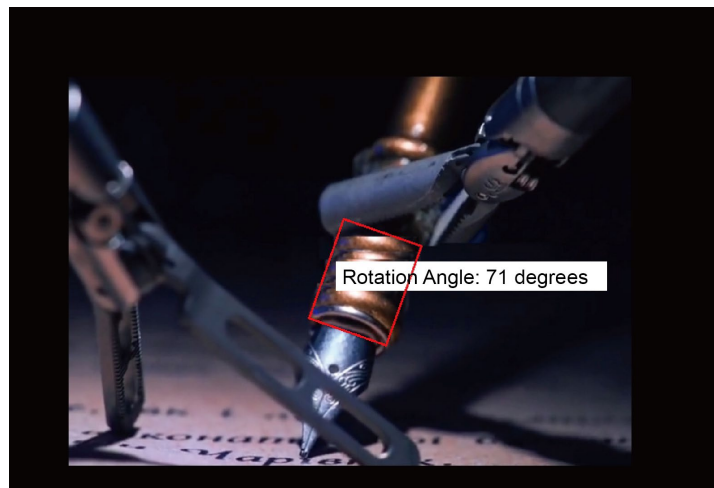

Figure 12. Selective evaluation based on specific object detected.

2.5. Advanced Labeling

At the end of any study, there is always a need to obtain the maximum amount of information with as little technical documentation as possible. One of such technique is an advanced labeling method for object detection.

Advanced labeling is a combination of information about classification, confidence, localization, orientation, spatial orientation, tracking, etc. that can be done with a custom application for object detection. Figure 13 shows a such example where some results with confidence depend on classes. The global confidence is between 61% for gear, 83% for stick, and 78% for forceps.
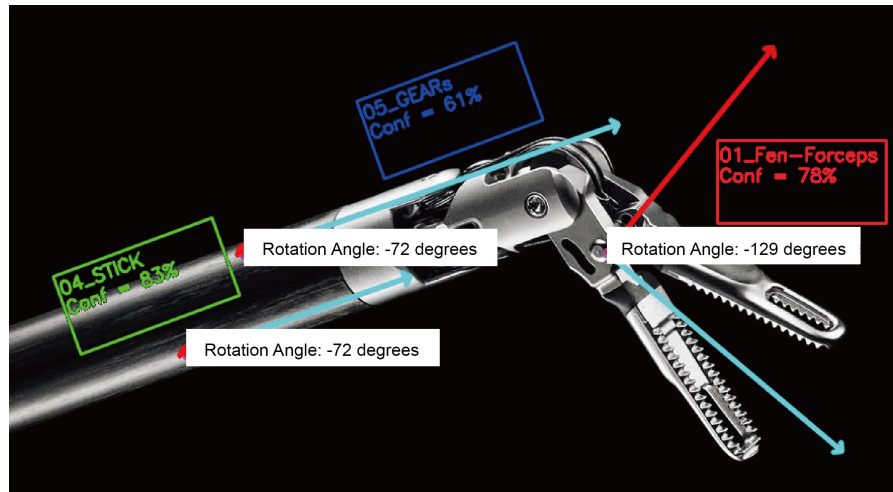
Figure 13. Advanced labelling tools.

2.6. Using CAD Models for Pretraining

One of the main issues in object detection is the limited number of references in dataset. Preparing the images and labeling also takes time for pre-processing.

This sub-chapter shows an innovative method of using CAD models to obtain images for object detection techniques.

The reference images for AI training are obtained from CAD models of focused objects for detection. Figure 14 shows a comprehensive list of 60 images from a dataset for AI training with only two classes. Using images obtained from a CAD model of focused objects for a custom dataset is a good method when there are not enough images of the real object. Results from object detection using ROBOFLOW can be viewed in Figure 15.

Using only CAD-set images was successful even for some difficult images for testing in object detection, like the image from Figure 16.

The CAD-set can be combined with the real images dataset to complement each other.

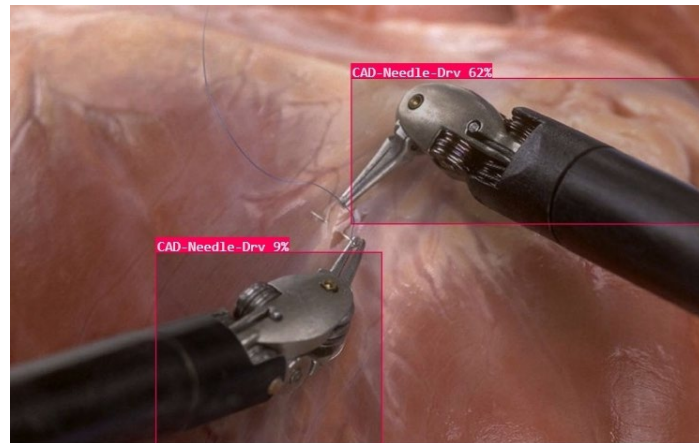**Figure 14.** Reference images in CAD-set for object detection.

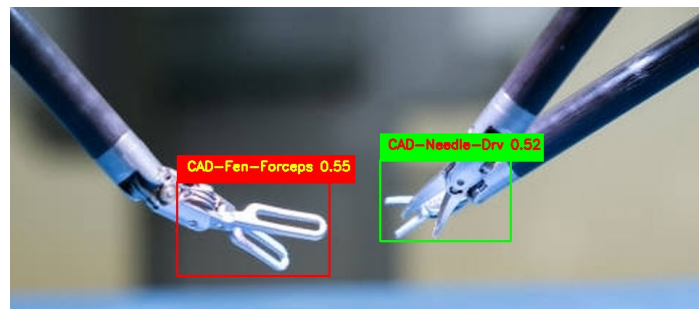Figure 15. Object detection based on CAD-set using ROBOFLOW.

Figure 16. Object detection based on CAD-set using A + CNN.

2.7. Multi-Scope Vision

In all situations, it appears a necessity to have also a different view from the scene, a different view angle, or a different view direction, because of insufficient data about the object from the scene when using limited number of object views.

This issue can be improved by using stereoscopic cameras. In Figure 17, two types of stereo-devices can be viewed used to capture stereoscopic images.
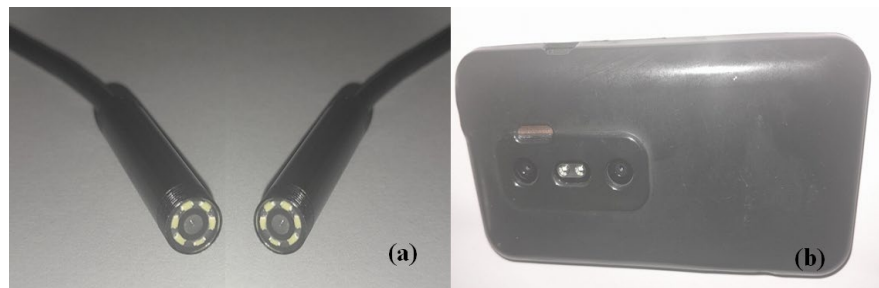

Figure 17. Stereoscopic cameras used for testing: (a) Stereo-endoscope; (b) Stereoscopic camera.

Figure 18 shows a CAD design of a surgical robot that use four video cameras (quad-scope view). This concept has special methods implemented for spatial evaluation of the scene.
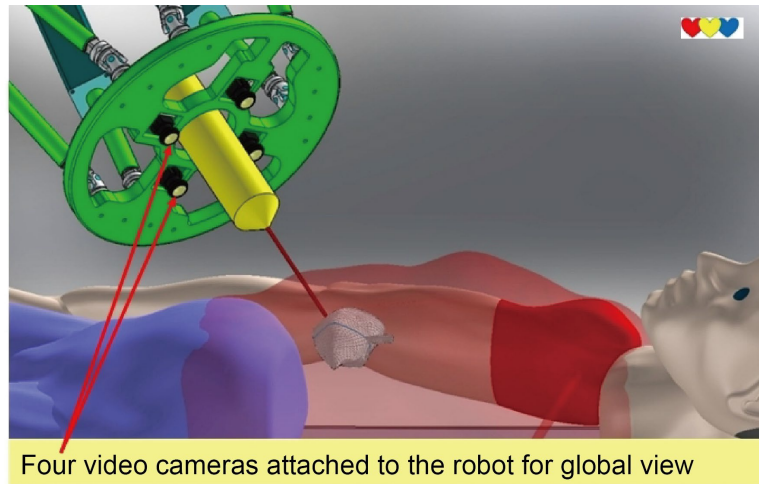
Four video cameras attached to the robot for global view

Figure 18. CAD model assembly of a robot with four video cameras.

2.8. Special CAD Application for Computer Vision

After obtaining real success with CAD-dataset in object detection, we propose in this sub-chapter also a CAD application add-on special designed for computer vision technique used in object detection. By controlling the CAD models and using special applications for object detection, we actually obtain software for CAD detection. This method is also an innovative method for object detection.

Disadvantages of CAD-dataset:

1) Is needed to define in CAD model of every object for detection;

2) Modeling in CAD is taking more time than a single picture;

3) Real images offer more realistic view of the object;

4) Hard to code for almost engineers.

Advantages of CAD application in object detection:

1) Is not necessary to store images of the objects. In this case is stored the CAD models of the objects;

2) No limitations in number of pose of objects. Any poses of objects can be obtained using CAD application;

3) CAD models take up a small portion of space on your HDD, much smaller than the size of all the images in the dataset;

4) CAD object offer much more detailed information about object (pose position, orientation, size, zooming, etc.);

5) CAD applications (Figure 19) in object detection also help us benefit from all the capabilities of a genuine CAD system: viewing objects in any direction, augmented visualization, improved color, adequate lighting, object sectioning, zooming at any scale, exploding of assemblies, transparency of specific areas, etc.;

6) CAD models can be created now very quick using ultrasonic scanning methods.

Figure 19 shows a special application developed also for computer vision. Managing the local coordinate systems orientation, we obtain the necessary similarity between a real object and its CAD twin.
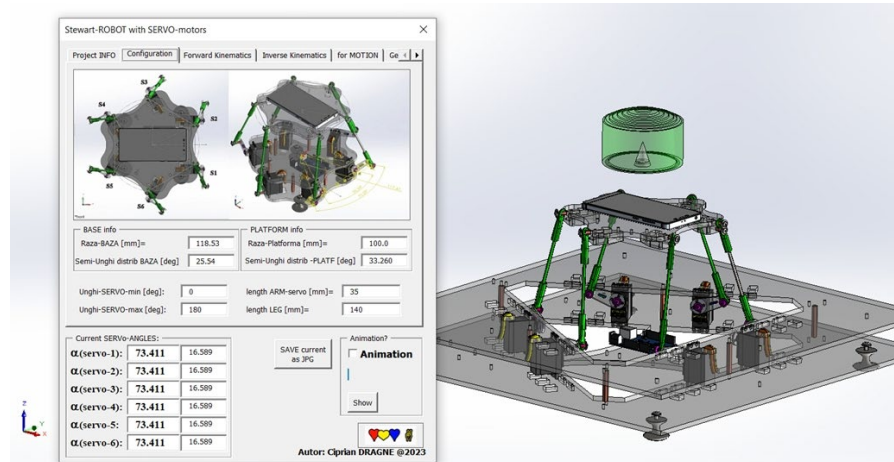
Figure 19. SOLIDWORKS Add-on for CAD control, computer vision and object detection.

Reconstruction in 3D of scenes and the development from now of digital twins will lead to increased interest in CAD models for every object around us. 3D reconstruction is applied very often in medical domain today [29].

3. Case Studies

For assessment of proposed methods are presented the next case studies:
1) Test using real images dataset;
2) Test images using CAD dataset;
3) Special CAD application for object detection;
4) Reverse engineering between real images dataset and CAD-dataset.

3.1. Test Using Real Images Dataset

Results for few object detection cases are presented in Figure 20 and Figure 21. The images were collected from various sources on the www, processed and labeled for this research. Confidence varies from 6% in Figure 20(c) to 71% in Figure 20(d). AI test was made with ROBOFLOW for (a, c, and d) and with YOLO11 in image (b).

3.2. Test Images Using CAD Dataset

Object detection results from Figure 21 are made with A + CNN for (a), and RO-BOFLOW for (b, c, and d).
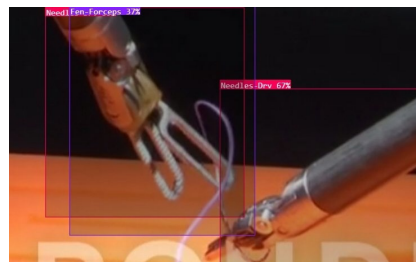
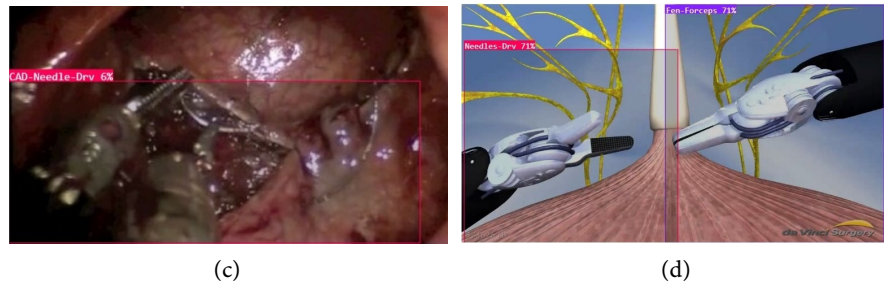(a)   (b)

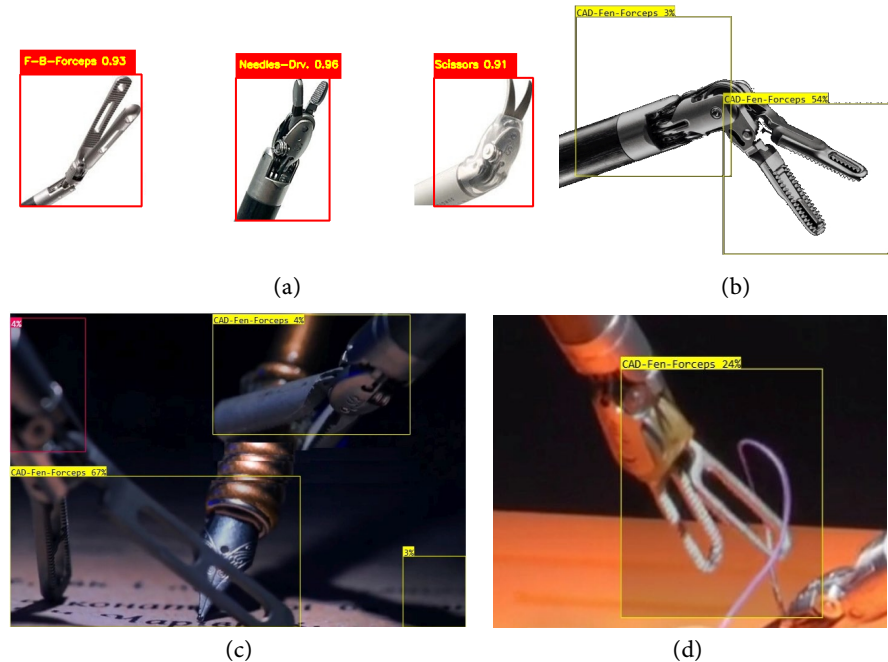Figure 20. Object detection results—various objects, YOLO and ROBOFLOW.

Figure 21. Object detection with CAD-set, using A + CNN and ROBOFLOW.

As we can observed during this paper, a ROBOFLOW was preferred. Even is an online web application ROBOFLOW has capability to create a quick AI training for custom data-set because the solving is made on multiple servers with multiple processors. In comparison, a local AI training with YOLO takes 10 hours for all images custom set, and with ROBOFLOW was 10 times much faster.

3.3. Special CAD Application for Object Detection

Based on CAD application to create CAD objects of similar real object, was developed also a new innovative method for using directly CAD in object detection. Application presented in Figure 1 to control CAD assembly orientation was extended with functions to compare virtual CAD models with real images of focused object. The application was developed under SOLIDWORKS-addon and with connectivity capabilities to run other external codes (MATLAB or PYTHON).

Connectivity between MATLAB, PYTHON, EXE, and SW add-ons is achieved directly through DDE (Dynamic Data Exchange) capabilities, STDIN/STDOUT

capabilities, or any other user-defined data transfer type.

Figure 22 shows object detection results using directly CAD application. We can observe in Figure 22(b) the errors in 2D angle orientation in comparison with Figure 13, and also advanced labelling for 3D spatial orientation (Euler Angles).
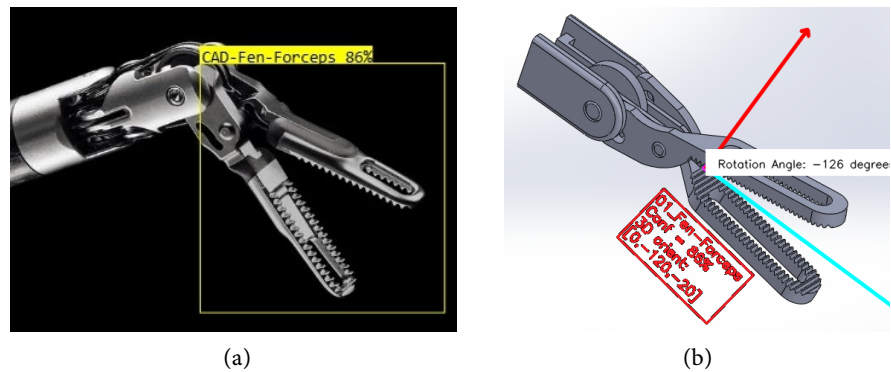

(a)   (b)

Figure 22. Object detection using CAD-application.

The CAD application for object detection has two types of assessments:

A) Search assembly CAD orientation and compare the camera image with focused object for detection;

B) Use AI in image comparison between target image and an internal data-set of CAD assembly internally and locally generated.

For the method from point A), was coded Multi-scale Structural Similarity Index Measure (MS-SSIM method) implemented in Python as an external application for the main CAD-addon. The MS-SSIM method is very accurate and sensitive in image processing [33]. Because the methodology is under-development, as acceptance criterion was used SSIM score higher than 50%. For example, between images from Figure 22, 66% SSIM score was obtained. There are more methods to use for improvements in images comparisons, which will be evaluated into the future research [34].

3.4. Reverse Engineering between Real Images Dataset and CAD-Dataset

Validation of the CAD-dataset can be done also using real images dataset. By reverse engineering from real images to CAD-set or CAD-addon can improve research evaluation of object detection, classification, segmentation or orientation.

The combination between the real images set and the SOLIDWORKS application (virtual model) is a real, fully functional, Digital-Twins.

3.5. Multi-Scope Images

Figure 23 shows object detection results when was used two images in the same detection session. These two images are obtained from a stereoscopic device from Figure 17. But this technique can be applied also using multiple cameras, with more success for detection and improvements in global efficiency.
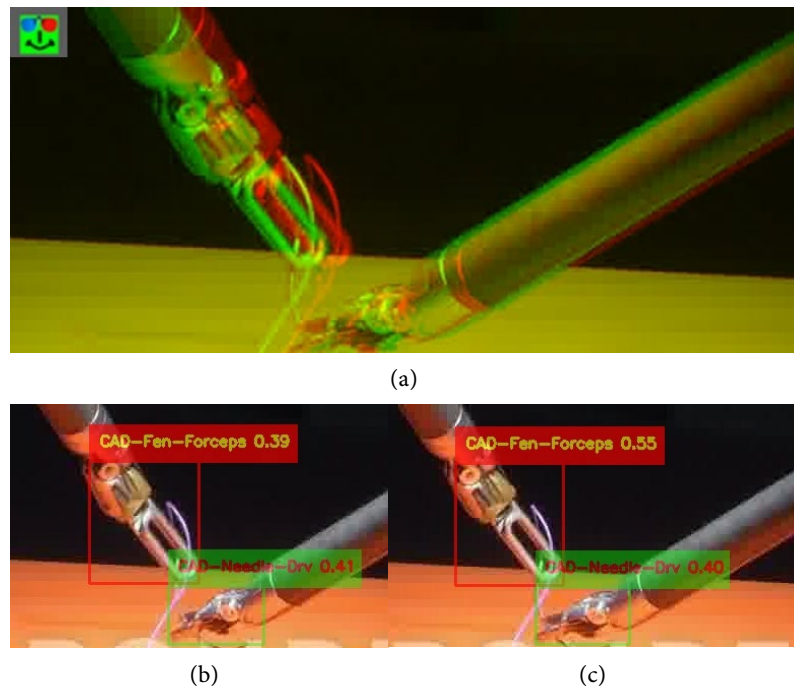
Figure 23. Object detection on stereoscopic images.

It can be observed that there are some differences between the left and right images if we are using A + CNN object detection technique.

4. Conclusions

In this paper, we systematically study neural network architecture design choices for efficient object detections of surgical tools, and propose improvements in CNN, a custom database, advanced labeling system and special CAD methods for detection of objects and object orientation, in order to improve accuracy and efficiency. The amount of research required proved to be considerable for this paperwork.

Object detection efficiency depends on the model database set. So, it has no absolute relevance in all cases.

Detecting of the small objects is also a challenging and important problems in computer vision in general and in medical domain that use AI applications. Using CAD-set can improve these issues [35].

For each phase, it was necessary to develop also custom IT tools for image preprocessing, labeling, quick extract of training information, pre- and post-visualization, etc.

Advantages of using custom model:

1) Selection of the best and adequate images;

2) Researchers learn in detail all methodologies involved, step-by-step;

3) Very easy to add new references.

Advantages of custom AI technique:

1) Custom labels with additional data;

2) Additional info about deep inside methodology related to disparity and all other variables workflow;

3) Free to test and no limitations in additional methods for improvement.

Difficulties in developing of surgical detection technique:

1) Small parts;

2) Small differences between classes;

3) Small number of images for references.

Advantages in using on-line applications for AI detection:

1) Could made faster training than local computers training because of server resources involved;

2) Additional techniques involved to improve and increase volume of dataset;

3) Advanced verification of performance criteria for dataset quality;

4) Work using augmentations methods—generate augmented versions of your source images to simulate different lighting conditions, camera angles, or contrast settings.

Disadvantages in using on-line applications for AI detection:

1) Not free for full training and object detection test;

2) Limitations in using number of references in datasets;

3) Limitations in using inside implemented methods;

4) Custom datasets are not entirely secure and new privacy standards are needed.

Good practice recommendations:

1) Add detailed labels;

2) Add specific details from images into dataset to improve object recognition (specific gear, specific bolts, specific shapes, etc.).

CAD datasets for object detection give us more additional advantages, already explained in paper. But, CAD applications in object detection also help us benefit from all the capabilities of a genuine CAD system: viewing objects in any direction, augmented visualization, improved color, adequate lighting, object sectioning, zooming at any scale, exploding of assemblies, transparency of specific areas, etc.

A twinning of object detection technique using real images with CAD models or CAD software itself is a more successful method for attempts to streamline object detection techniques, evaluate their positioning and orientation, segmentation of images to be processed or 3D spatial orientation.

The global performance results at this moment are around 80% in confidence, and the main reason is the small number of references in dataset. Future research intends to implement parallel computing and GPU (graphic processor) acceleration techniques. However, using the CAD-set method shows good perspectives.

AI computer vision will be transforming everything from simple capturing or observing the world to object detection, from estimating to increase security and detect unappropriated actions and revealing faults objects, from counting objects to change management in logistics or manufacturing, from offering new ideas until to change the entire educational processes, learning, testing scientific research,

assembly data and information's, from global medical data until data focus on patient needs and new requirements, etc.

Even the conclusion seems to be too lengthy, for someone who wants to learn from our research, it is better in this way. A such complex research has many aspects that need to be studied, observed, notified, learned from it, focused, and even reloaded, returned, made changes, etc. The learning process should be a continuous improvement, studying papers, understanding ideas, etc. So, let's communicate and educate each other, and let our ideas inspire you. A new, astonishing chapter in education has just started and will change us fundamentally.

The complete task of the simulations was developed based on Google-NET documentation [36], UE recommendations [37], PYTHON software [38], MATLAB software [39], SOLIDWORKS Educational [40], user-defined programming routines [41], and on-line applications for object detection [42] [43]. Some images for testing were selected from Pond5 website database [44].

Acknowledgements

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Topol, E.J. (2019) High-Performance Medicine: The Convergence of Human and Artificial Intelligence. *Nature Medicine*, **25**, 44-56.
https://doi.org/10.1038/s41591-018-0300-7

[2] Litjens, G., Kooi, T., Bejnordi, B.E., Setio, A.A.A., Ciompi, F., Ghafoorian, M., *et al.* (2017) A Survey on Deep Learning in Medical Image Analysis. *Medical Image Analysis*, **42**, 60-88. https://doi.org/10.1016/j.media.2017.07.005

[3] McCarthy, J. (2000) Concepts of Logical AI. In: Minker, J., Ed., *Logic-Based Artificial Intelligence*, Springer, 37-56. https://doi.org/10.1007/978-1-4615-1567-8_2

[4] McCarthy, J. (1990) Generality in Artificial Intelligence. In: Lifschitz, V., Ed., *Formalizing Common Sense*, Ablex, 226-236.

[5] Dragne, C. (2024) Advanced Concurrent Engineering Using CAD Software-Classic, Avant-Garde and AI. *International Journal of Advanced Multidisciplinary Research and Studies*, **4**, 279-285. https://doi.org/10.62225/2583049x.2024.4.4.3038

[6] Hosny, A., Parmar, C., Quackenbush, J., Schwartz, L.H. and Aerts, H.J.W.L. (2018) Artificial Intelligence in Radiology. *Nature Reviews Cancer*, **18**, 500-510.
https://doi.org/10.1038/s41568-018-0016-5

[7] Zaidi, D. (2017) AI Is Transforming Medical Diagnosis, Prosthetics, and Vision Aids. Venture Beat.
https://venturebeat.com/ai/ai-is-transforming-medical-diagnosis-prosthetics-and-vision-aids/

[8] Singh, R., Kalra, M.K., Nitiwarangkul, C., Patti, J.A., Homayounieh, F., Padole, A., *et al.* (2018) Deep Learning in Chest Radiography: Detection of Findings and Presence

of Change. *PLOS ONE*, **13**, e0204155. https://doi.org/10.1371/journal.pone.0204155

[9] Ross, C. and Swetlitz, I. (2018) IBM's Watson Supercomputer Recommended "Unsafe and Incorrect" Cancer Treatments; Internal Documents Show. Stat News. https://www.statnews.com/2018/07/25/ibm-watson-recommended-unsafe-incorrect-treatments/

[10] Bamba, Y., Ogawa, S., Itabashi, M., Kameoka, S., Okamoto, T. and Yamamoto, M. (2021) Automated Recognition of Objects and Types of Forceps in Surgical Images Using Deep Learning. *Scientific Reports*, **11**, Article No. 22571. https://doi.org/10.1038/s41598-021-01911-1

[11] Tan, M., Pang, R. and Le, Q.V. (2020) EfficientDet: Scalable and Efficient Object Detection. 2020 *IEEE/CVF Conference on Computer Vision and Pattern Recognition* (*CVPR*), Seattle, 13-19 June 2020, 10781-10790. https://doi.org/10.1109/cvpr42600.2020.01079

[12] Chollet, F. (2017) Xception: Deep Learning with Depthwise Separable Convolutions. 2017 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Honolulu, 21-26 July 2017, 1251-1258. https://doi.org/10.1109/cvpr.2017.195

[13] He, K., Girshick, R. and Dollar, P. (2019) Rethinking ImageNet Pre-Training. 2019 *IEEE/CVF International Conference on Computer Vision* (*ICCV*), Seoul, 27 October-2 November 2019, 4918-4927. https://doi.org/10.1109/iccv.2019.00502

[14] Szegedy, C., Sermanet, P., Reed, S., Anguelov, D., *et al*. (2015) Going Deeper with Convolutions. 2015 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Boston, 7-12 June 2015, 1-9. https://doi.org/10.1109/cvpr.2015.7298594

[15] Jegham, N., Koh, C.Y., Abdelatti, M. and Hendawi, A. (2024) Evaluating the Evolution of YOLO (You Only Look Once) Models: A Comprehensive Benchmark Study of YOLO11 and Its Predecessors. arXiv: 2411.00201.

[16] Girshick, R., Donahue, J., Darrell, T. and Malik, J. (2014) Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation. 2014 *IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, 23-28 June 2014, 580-587. https://doi.org/10.1109/cvpr.2014.81

[17] Tan, M. and Le, Q. (2019) EfficientNET: Rethinking Model Scaling for Convolution-al Neural Networks. *International Conference on Machine Learning*, Long Beach, 9-15 June 2019, 6105-6114.

[18] Zhou, P., Ni, B., Geng, C., Hu, J. and Xu, Y. (2018) Scale-Transferrable Object Detection. 2018 *IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 18-23 June 2018, 528-537. https://doi.org/10.1109/cvpr.2018.00062

[19] Huang, R., Pedoeem, J. and Chen, C. (2018) YOLO-LITE: A Real-Time Object Detection Algorithm Optimized for Non-GPU Computers. 2018 *IEEE International Conference on Big Data* (*Big Data*), Seattle, 10-13 December 2018, 2503-2510. https://doi.org/10.1109/bigdata.2018.8621865

[20] Agossou, B.E., Pedersen, M., Raja, K., Vats, A. and Floor, P.A. (2025) Influence of Color Correction on Pathology Detection in Capsule Endoscopy. In: Palaiahnakote, S., *et al*., Eds., *Pattern Recognition. ICPR 2024 International Workshops and Challenges*, Springer, 365-379. https://doi.org/10.1007/978-3-031-88220-3_26

[21] Lin, T., Goyal, P., Girshick, R., He, K. and Dollar, P. (2017) Focal Loss for Dense Object Detection. 2017 *IEEE International Conference on Computer Vision* (*ICCV*), Venice, 22-29 October 2017, 2980-2988. https://doi.org/10.1109/iccv.2017.324

[22] Watine, L., Floor, P.A., Pedersen, M., Nussbaum, P., Ahmad, B. and Hovde, Ø. (2023) Enhancement of Colour Reproduction for Capsule Endoscopy Images. 2023 *11th European Workshop on Visual Information Processing* (*EUVIP*), Gjovik, 11-14 September

2023, 1-6. https://doi.org/10.1109/euvip58404.2023.10323058

[23] Tian, J., Jin, Q., Wang, Y., Yang, J., Zhang, S. and Sun, D. (2024) Performance Analysis of Deep Learning-Based Object Detection Algorithms on COCO Benchmark: A Comparative Study. *Journal of Engineering and Applied Science*, **71**, Article No. 76. https://doi.org/10.1186/s44147-024-00411-z

[24] Chen, K., Wang, J., Pang, J., Cao, Y., Xiong, Y., Li, X. and Lin, D. (2019) MM-Detection: Open MMlab Detection Toolbox and Benchmark. arXiv: 1906.07155.

[25] Yao, Y., Cheng, G., Wang, G., Li, S., Zhou, P., Xie, X., *et al.* (2023) On Improving Bounding Box Representations for Oriented Object Detection. *IEEE Transactions on Geoscience and Remote Sensing*, **61**, Article ID: 5600111. https://doi.org/10.1109/tgrs.2022.3231340

[26] Wang, K., Wang, Z., Li, Z., Su, A., Teng, X., Liu, M. and Yu, Q. (2023) Oriented Object Detection in Optical Remote Sensing Images Using Deep Learning: A Survey. arXiv: 2302.10473.

[27] Yi, J., Wu, P., Liu, B., Huang, Q., Qu, H. and Metaxas, D. (2021) Oriented Object Detection in Aerial Images with Box Boundary-Aware Vectors. 2021 *IEEE Winter Conference on Applications of Computer Vision* (*WACV*), Waikoloa, 3-8 January 2021, 2150-2159. https://doi.org/10.1109/wacv48630.2021.00220

[28] Reyes-Hernández, J.C., Alomar, A., Rubio, R., Piella, G. and Sukno, F. (2024) OBBaby-Face: Oriented Bounding Box for Infant Face Detection. In: Fred, A., *et al.*, Eds., *Deep Learning Theory and Applications*, Springer, 336-350. https://doi.org/10.1007/978-3-031-66705-3_22

[29] Ding, S., Mannan, M.A. and Poo, A.N. (2004) Oriented Bounding Box and Octree Based Global Interference Detection in 5-Axis Machining of Free-Form Surfaces. *Computer-Aided Design*, **36**, 1281-1294. https://doi.org/10.1016/s0010-4485(03)00109-x

[30] Chang, J., Wang, W. and Kim, M. (2010) Efficient Collision Detection Using a Dual OBB-Sphere Bounding Volume Hierarchy. *Computer-Aided Design*, **42**, 50-57. https://doi.org/10.1016/j.cad.2009.04.010

[31] Hoang, D., Chen, L. and Nguyen, T. (2016) Sub-OBB Based Object Recognition and Localization Algorithm Using Range Images. *Measurement Science and Technology*, **28**, Article ID: 025401. https://doi.org/10.1088/1361-6501/aa513a

[32] Dragne, C., Todiriţe, I., Iliescu, M. and Pandelea, M. (2022) Distance Assessment by Object Detection—For Visually Impaired Assistive Mechatronic System. *Applied Sciences*, **12**, Article No. 6342. https://doi.org/10.3390/app12136342

[33] Wang, Z., Simoncelli, E.P. and Bovik, A.C. (2003) Multiscale Structural Similarity for Image Quality Assessment. *The 37th Asilomar Conference on Signals, Systems & Computers*, Vol. 2, 1398-1402.

[34] Sampat, M.P., Wang, Z., Gupta, S., Bovik, A.C. and Markey, M.K. (2009) Complex Wavelet Structural Similarity: A New Image Similarity Index. *IEEE Transactions on Image Processing*, **18**, 2385-2401. https://doi.org/10.1109/tip.2009.2025923

[35] Dragne, C. (2023) CAX Software—The Next Level in Computer Aided Technology. *International Journal of Advances in Engineering and Management*, **5**, 551-561.

[36] Google API Homepage. https://storage.googleapis.com/openimages/web/index.html

[37] UE Website—Industry 5.0. https://ec.europa.eu/info/research-and-innovation/research-area/industrial-research-and-innovation/industry-50_en

[38] PYTHON Homepage. https://www.python.org/

[39] MATLAB Homepage. https://www.matlab.com/

[40] SOLIDWORKS Homepage. https://www.solidworks.com/

[41] GRABCAD Author Codes. https://grabcad.com/ciprian.dragne

[42] ROBOFLOW Website. https://roboflow.com/

[43] YOLO11 Ultralytics Homepage. https://docs.ultralytics.com/models/yolo11/

[44] Pond5 Website Database. https://www.pond5.com/