

An artificial intelligence approach to smart exam supervision using YOLO v5 and siamese network

Nabeel I. Zanoon¹, Abdullah A. Alhaj², Khalid Alkharabsheh³

¹Department of Applied Sciences, Faculty of Salt College, Al-Balqa Applied University, Aqaba, Jordan

²Department of Information Technology, School of Information Technology and Systems, The University of Jordan, Aqaba, Jordan

³Department of Software Engineering, Prince Abdullah bin Ghazi Faculty of Information and Communication Technology, Al-Balqa Applied University, As-Salt, Jordan

Article Info

Article history:

Received Sep 16, 2023

Revised Feb 24, 2024

Accepted Apr 18, 2024

Keywords:

Artificial intelligence

Exam supervision

Object detection

Siamese network

YOLO v5

ABSTRACT

Artificial intelligence has introduced revolutionary and innovative solutions to many complex problems by automating processes or tasks that used to require human power. The limited capabilities of human efforts in real-time monitoring have led to artificial intelligence becoming increasingly popular. Artificial intelligence helps develop the monitoring process by analyzing data and extracting accurate results. Artificial intelligence is also capable of providing surveillance cameras with a digital brain that analyzes images and live video clips without human intervention. Deep learning models can be applied to digital images to identify and classify objects accurately. Object detection algorithms are based on deep learning algorithms in artificial intelligence. Using the deep learning algorithm, object detection is achieved with high accuracy. In this paper, a combined model of the YOLO v5 model and network Siamese technology is proposed, in which the YOLO v5 algorithm detects cheating tools in classrooms, such as a cell phone or a book, in such a way that the algorithm detects the student as an object and cannot recognize his face. Using the Siamese network, we compare the student's face against the database of students in order to identify the student with cheating tools.

This is an open access article under the [CC BY-SA](#) license.



Corresponding Author:

Nabeel I. Zanoon

Department of Applied Sciences, Faculty of Salt College, Al-Balqa Applied University

Aqaba, Jordan

Email: dr.nabeel@bau.edu.jo

1. INTRODUCTION

Using artificial intelligence, images, and videos are analyzed and extracted by computer vision, which can help systems analyze and extract information. A machine's artificial intelligence works the same way as the human brain, and a computer's vision functions the same way as the human eye. Since human monitoring cannot reach a high level of safety and is insufficient, artificial intelligence is introduced in the analysis and monitoring of video and cameras. Artificial intelligence analyzes video content by applying computer vision and deep learning to various types of data. Through object detection, security solutions can be implemented by identifying and detecting a variety of objects, as this technology is integral to detecting people and things in videos and photos.

Surveillance cameras generate large amounts of video data, which over time becomes difficult to store or analyze, so it is necessary to provide a large amount of space for storing video clips. In order to solve this problem, the analysis and detection of objects to extract results will be the ideal solution since it reduces

storage requirements [1]. Moreover, image and video analyses provide security in several areas, since videos and images are analyzed through the process of discovering different objects. Using this method can also help track and detect important things in video clips or images so that we can predict occurrences of accidents, and report some things, and issue alerts about suspicious behavior [2]. Besides the location, object detection also determines the category of objects, such as humans, cars, mobiles, and pets, in an image or video. Object detection technology is used in several areas, including robot vision, surveillance and security, autonomous driving, and human-computer interaction [3].

Object detection algorithms combine image classification and object localization. A method of bounding boxes and assigning category names to each box has been used here, allowing the identification of multiple categories and the handling of objects that recur within the image [4]. The detection of objects began in the year 2000 with the use of the method of identifying components. A major challenge in the progress of object detection is that objects have multiple measurements and different sizes and aspect ratios. The detection of objects using deep learning began in 2014, when convolutional neural networks (CNN) technology was proposed to detect objects. With deep learning, you can predict the location of an object with high accuracy using fast and accurate algorithms. During the deep learning era, object detection algorithms were divided into two types based on their stages: two-stage detection and one-stage detection [5].

The field of deep learning focuses on object detection as a branch of object recognition. There is a difference between face detection and face recognition, as face detection considers the face an object that belongs to a category, while face recognition identifies people in photographs and videos. Detecting a face is the first step in the face recognition process. This consists of several stages, including discovery, alignment, feature extraction, and finally identifying the person [6]. To put it another way, face recognition is completely different from face detection. Face recognition involves entering an image and comparing it to all the images in the database. If the image looks similar to an existing image, the face of the person will be recognized, and his identity will be displayed as it appears in the database [7].

Several studies have focused on using deep learning techniques for object detection [8]-[11]. This field has attracted the attention of the research community, and there is an opportunity to be exploited effectively in this context. In this research, object identification technology has been combined with face recognition technology. The you only look once (YOLO) v5 model is used to detect the object, while the Siamese network recognizes the face. A monitoring system is proposed to detect cheating during exams. This system consists of two stages: the first stage is identifying the student who has a cheating tool through the YOLO v5 model, and the second stage is identifying the student, i.e., face recognition by comparing the student's face image with an image of the same student from a database that contains a set of pictures of students. Exams are monitored by human invigilators in universities. Several studies have shown that cheating using electronic tools such as cell phones poses a threat to education systems. Similarly, cheating in exams increased after the e-learning revolution during the COVID-19, as electronic educational materials became more accessible to students. The main objectives of this study are summarized by: developing electronic monitoring/invigilation methods through artificial intelligence and deep learning, enhancing examination invigilation, assisting and improving human monitoring, reducing cheating risks in universities, and tracking and identifying students who cheat.

2. BACKGROUND

2.1. Artificial intelligence and object detection

The process of object detection is one of the computer vision object detection algorithms and a field of artificial intelligence that allows surveillance systems to detect the content of images and video clips [12]. Deep learning is a model of machine learning, and it is a sub-field of artificial intelligence. Because modern science analyzes and processes large amounts of data, deep learning algorithms based on artificial neural networks (ANNs) are preferable for analyzing and processing large amounts of data [13].

Deep learning models can both classify a large amount of data and deeply-learn the features contained within the data using multilayer CNN. These data and features are extracted and used in image-based automated diagnosis in various fields. Data sets containing training images can be entered into a deep learning system, which results in automatic relearning without manual intervention [14]. Deep learning systems have contributed to advancements in the field of object detection since deep neural networks are capable of learning different features automatically. According to several studies, object detection technology based on deep learning is

superior to traditional hand-crafted features in terms of accuracy [15].

2.2. Object detection and deep learning algorithms

Traditional object detection algorithms consist of two parts: the first part extracts features and the second part classes them, and there are four main models (Haar feature, Adaboost algorithm, Hog feature, support vector machine (SVM) algorithm, deformable part model (DPM) algorithm) [16]. Object detection algorithms that utilize deep learning are categorized into three types: CNNs, recurrent neural networks (RNNs), and deep belief networks (DBNs) [17]. Compared to traditional methods of object detection, deep convolutional neural networks (DCNN) have developed rapidly in the field of object detection as they can learn image features using deep learning. There are two types of object detection algorithms in deep learning: algorithms based on handcrafted features and algorithms based on deep learning.

Object detection algorithms are important in artificial intelligence and computer vision, allowing computers to see their environments by detecting objects in images and videos [18], [19]. Object detection in deep learning requires training and testing of the input data, since these models require powerful computational resources and large datasets. Through the application and training of global data (PASCAL, ILSVRC, VOC, and MS-COCO), deep learning can achieve high results in object detection. Deep learning algorithms for detecting objects can be divided into two parts: the two-stage section (Fast region-based convolutional neural network (Fast R-CNN), region-based convolutional neural network (R-CNN)) and the single-stage section (YOLO family series, single shot multibox detector (SSD)).

The two-stage algorithms consist of two stages, where the first stage is extracting the region of interest (ROI), and the second stage is classifying and identifying the ROI. Examples of which are Faster region-based convolutional neural network (Faster R-CNN), feature pyramid networks Faster R-CNN (FPN-FRCN), and region-based fully convolutional networks (R-FCN) [20]. The single-stage object detection algorithm performs its role in only one stage, as it skips the first stage, proposing the region; instead, it proposes regions and classification at the same time in one stage because it deals with CNN through bounding boxes and classifying objects [21].

2.3. YOLO algorithm

YOLO is based on the detection of objects in real time using neural networks. It's one of the most popular algorithms and is known for its speed and accuracy [22]. YOLO is similar to a fully convolutional neural network (FCNN), which extracts results and predictions from an image (N N) once. The idea of its work is based on dividing the image into a grid ($M * M$), and each grid produces two bounding boxes and the possibility of a category in the bounding boxes [23]. As opposed to previous algorithms, YOLO uses regression to identify the interesting parts of the image rather than identifying them. It can extract all the categories and bounding boxes from an image at once, which is why it is called YOLO [24]. It was released in 2015 by Redmon *et al.*, and several versions of YOLO have been developed since then, including YOLO v1 and v2, 3 in YOLO v5 [25].

2.3.1. Model YOLO v5

YOLO's fifth version was developed by ultralytics, and it is considered a family of one-stage algorithms because it is superior to previous versions in detecting objects [26]. With the YOLO v5 model, object detection is achieved using an algorithm optimization strategy using CNN, such as auto-learning, bounding box anchors, mosaic data augmentation, and the cross-stage partial network [27]. There are several types of YOLO v5 based on the size of memory storage, from YOLO v5s to YOLO v5m and YOLO v5l to YOLO v5x, all of which follow the same principles and architecture [28]. Like other models of the YOLO family, the YOLO5 works by identifying the object in images and videos, as its components are similar to those of the YOLO4. There are three main parts of the YOLO5 architecture, namely the backbone, the neck, and the main part head.

- Backbone: A function of this part is to extract features and rich information from the images that are entered into the processing. In this case, the cross stage partial network (CSP) strategy is used to divide the feature map in the base layer into two parts and then combine them through the processing stages using a division merging strategy [29]. By using this strategy, the staging can flow through the network. There are three layers in the backbone: convolution-batchNormalization-SiLU (CBS), three convolutional modules (C3), and spatial pyramid pooling fast (SPPF). The CBS module includes three layers: convolution, batch normalization, and silu activation. As for the C3 unit, it is composed of CBS

units. SPPF is used to change the size of feature maps by combining the output of different aggregation layers [30].

- The neck: It uses PANet to generate and create distinctive pyramids. By combining feature pyramid networks (FPN) and pyramid attention network (PAN), these pyramids can detect objects of different sizes and scales. It consists of four communication layers, four wrapping layers, and five layers [31].
- The head: It is considered the final stage. YOLO v5 has the same header structure as YOLOv3 and YOLOv4. It consists of three convolutional layers that predict the location of bounding boxes. This part generates predictions based on the connecting boxes in order to discover the object. Connecting boxes are used to extract the final output from the feature maps, classification of categories, object degrees, and bounding boxes [32].

Since several previous deep learning models for object detection have been studied, the YOLO v5 model was chosen to be applied in this study. Based on previous studies, the YOLO v5 model achieved excellent results in object detection in images and videos, as well as excellent results in detecting small objects. The programming was written and implemented by PyTorch framework instead of the Darknet framework, which made it easier and faster to deal with data. For these reasons, the YOLO v5 model was chosen, in addition to some other concepts that enhanced its performance, as we will see in the next section.

2.4. Objects detection and face recognition

A computer vision system analyzes images and video clips to extract information. Objects are detected, and their location is identified by placing bounding boxes around these detected objects and then classifying them into different categories [33]. Deep learning uses the detection of objects through a training model on the input images, where the outputs are boxes surrounding the object that has been trained on. The model can be trained using training tools or applications, such as Tensor Flow, PyTorch, or Keras [33]. These algorithms use deep learning models to extract results in order to detect objects in images. Deep learning object detection is a fast and effective way to detect objects and their location in an image [34].

The human face is regarded as an object in the process of object detecting, in other words, detecting the human face in a photograph or video clip is regarded as identifying the human face as an object and not as a person's identity [35]. A facial recognition system can recognize a face in an image and identify it by comparing it to a database of stored images [36]. Different algorithms and methods can be used to identify faces in images, but in order to recognize faces in images, Siamese networks are used, which compare two inputs of images in parallel to carry out the comparison process. As the Siamese network is comprised of two neural networks that are identical to each other, it works in parallel to carry out the comparison process [37].

2.5. Siamese network

Siamese networks were proposed by Bromley at the start of the nineties as a way of verifying the matching between two images. A Siamese network works by combining two similar networks with two inputs, and it learns the extent of similarity in the data, as shown in Figure 1, and it can learn simultaneously [38]. Several operations rely on it, such as handwritten verification and face recognition by comparing the existing image with previously stored images [39]. The mechanism of the Siamese network depends of calculating similarity through binary entropy, variation function, or triple loss. It works on a binary classification of objects entered in two similar or dissimilar states [40].

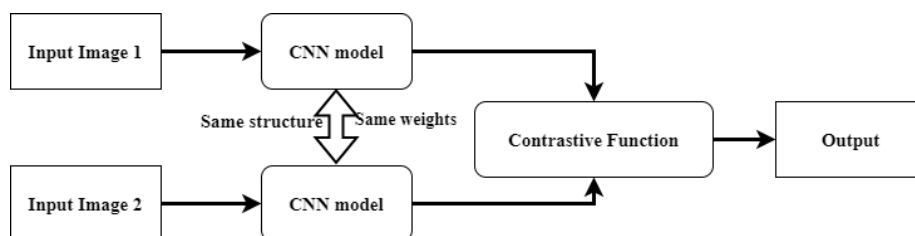


Figure 1. Siamese network architecture

As Siamese networks categorize input images according to their similarities, they are one-shot classification models that are capable of prediction based on a single training example. It works by learning similarity information between a pair of objects [41]. Learning occurs as a result of contrast loss. To get the advantage of

embedding the input image, the learning-based loss function is used during the training phase [42]. Loss functions work on the basis of pairs of objects, not individual objects, in order to complete the comparison process. The loss function causes the model to generate less similar feature embeddings if the classes are different, and more comparable feature embeddings if the aim classes are the same. In mathematical terms, we express the contrastive loss as (1):

$$Loss = (1 - y) * \frac{1}{2}(d_w)^2 + (y) * \frac{1}{2}\{max(0, m - (d_w))\}^2 \quad (1)$$

3. PROPOSED APPROACH

This paper presents a model based on the YOLO v5 model as well as the Siamese network model, as shown in Figure 2. The proposed model aims to detect cheating tools like cellular devices and books during exams in the classroom so that invigilators can identify students who are cheating on exams. At the beginning, a set consisting of 128 images (640*640 pixels) is entered into the YOLO v5 model, where the objects on which the data was trained are detected, namely GT containing (a book, a cellular device, a person). The YOLO v5 model detects images where a student has a cellular device or a book. In the event that the image does not contain any of these, it will go out to the final stage. In the event that a cellular device object or a book is detected in the image [$Obj(GT) \in (IMG(I))$] the image moves to the second stage to identify the student's identity by recognizing the student's face using the Siamese network. The first image is represented in the Siamese network inputs [$Input(1) = Img + Obj(GT)$], the second image is the student's image from the student databases [$Input(2) = Img(St(i)) \in DB$] where the two images are compared to find out the identity of the student carrying a cellular device in the classroom during the exam.

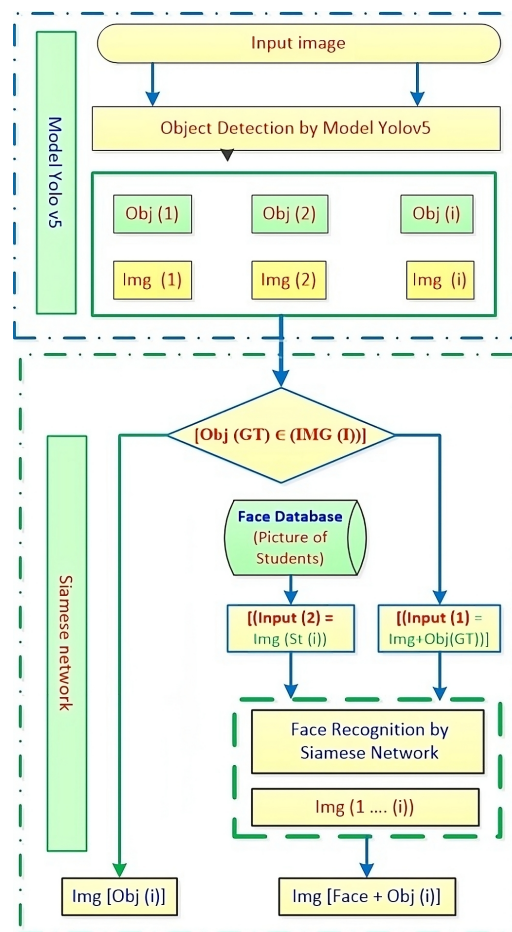


Figure 2. The proposed methodology for the smart monitoring system

4. RESULTS AND DISCUSSION

The study aims to identify students with cheating tools (a mobile device or a book) based on photos and videos collected from surveillance cameras in the exam halls. A proposed model has been proposed based on the integration of the YOLO v5 model to detect the object in the image with the Siamese neural network in order to recognize the student's face through training on the allocated data. Several steps are involved in the experiment, as follows: preparing the environment, preparing a custom dataset, training the YOLO v5 object detection model (cell phone and book), and obtaining the results by identifying the identity of the student who cheats using the Siamese network.

4.1. Environment and data preparation for training the proposed model

Two data sets were used in this study. The first dataset group contains data (images) collected via surveillance cameras located in the computer labs where the exams are held, as shown in Figure 3(a). The target object was then detected using deep learning using the YOLO v5 model. The second dataset group is a set of data groups that contains 2000 photos of students stored on the Microsoft cloud of the computer center at Al-Balqa Applied University, which are used for comparison and identification of students, as shown in Figure 3(b).

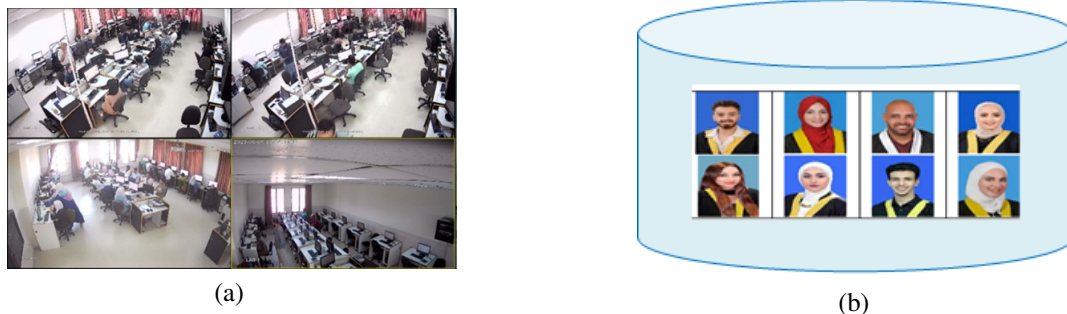


Figure 3. Two datasets were used in this study: (a) the first group contains control images and (b) the second group is a database of students' photos

4.2. Environment preparation

To train the data in YOLO v5, the data must be labeled (labeled data), but before we begin training the data, we must prepare the special data set in order to create the data in YOLO v5 format. Using Roboflow (Figure 4), we find generic computer vision datasets that can be manually annotated. We set up an account on Roboflow using the email address (dr.nabeel@bau.edu.jo) under the project name (Surveillance Systems in Classroom during Exam YOLO v5). Based on our proposed model, in which we use artificial intelligence and deep learning models. Regarding the required hardware and software environments for conducting this study, on the one hand, the configuration of the software was: Model YOLO v5 is implemented using Python 3.9 and PyTorch 1.8.0 with some additional requirements, including TensorFlow, PyTorch, OpenCV, CUDA, and NumPy, # YOLO v5 requirements, and # pip install r requirements. txt. On the other hand, the configuration of the hardware was: CPU: 12th Gen Intel Core i7-1255U Up 4.7 GHz, RAM: 12 GB, GPU: NVIDIA® GeForce RTX™ 2050 (4 GB), GPU RAM: 12 GB, HD: 512 GB SSD, and the OS: Windows 10 Pro 64-bit.

4.3. Training the proposed model

This study proposes a model based on the integration of the YOLO v5 model and the Siamese neural network, as shown in Figure 5. Training on the YOLO v5 model occurs on the first data set in order to identify objects in the images (mobile or book). Following that, a Siamese neural network is trained to identify the student. To train the Siamese neural network, two images must be entered, as the first image is the result of training the YOLO v5 model after discovering the cell phone or book and the person. The second image may represent the student database, which is the second group.

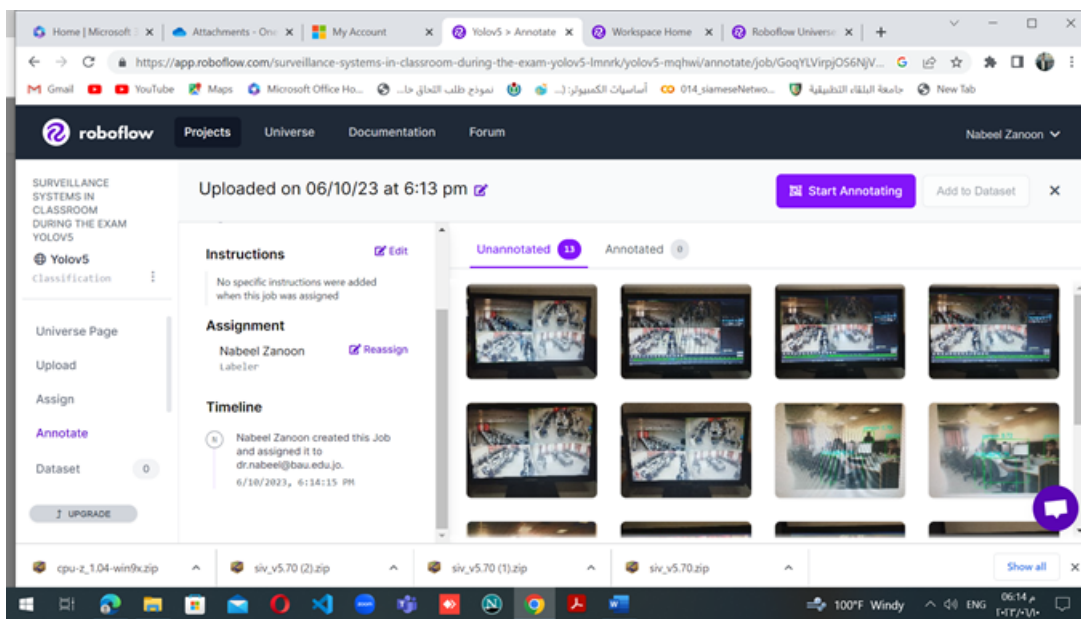


Figure 4. Uploading images to Roboflow

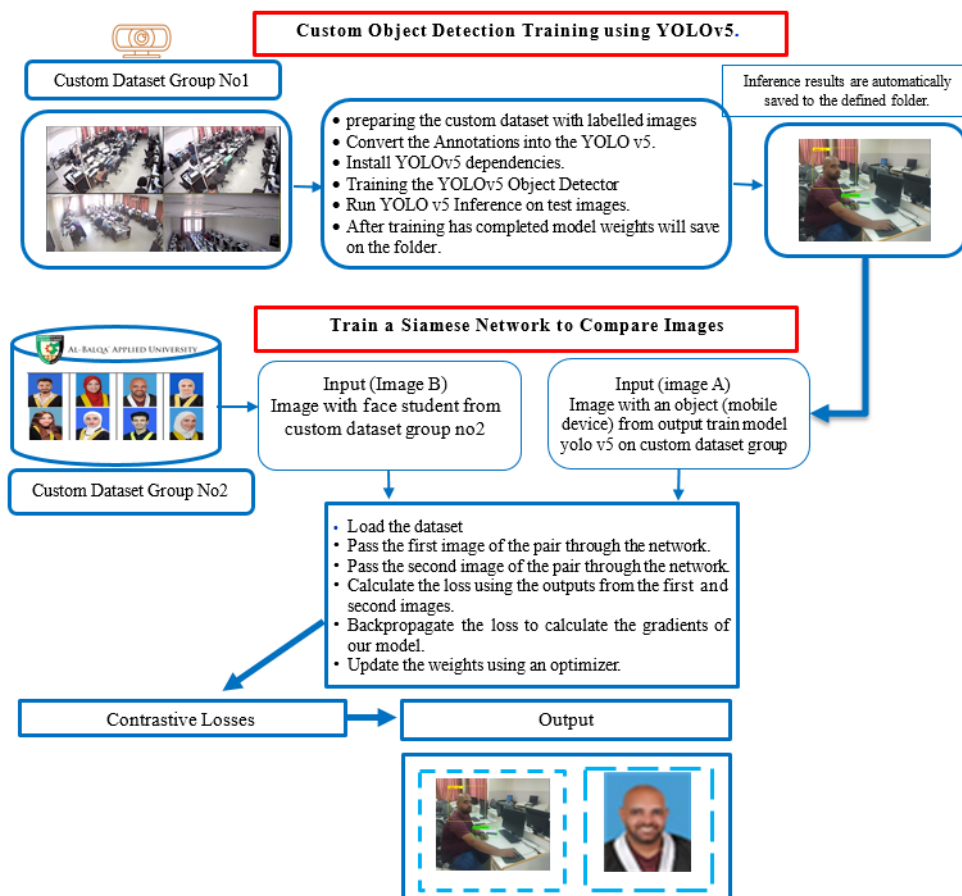


Figure 5. Flowchart of the proposed model, YOLO v5 with Siamese network

5. CONCLUSION

By combining artificial intelligence with computer vision, image processing is able to improve the ability to recognize faces and objects based on an intelligent analysis of existing images, which leads to the development of surveillance systems. The experience at university has convinced me that human invigilation capabilities must be improved urgently. Additionally, people are unable to watch multiple screens at the same time, and they cannot concentrate on the good shots in the video, so it is impossible to analyze the scenes in the video as well. We propose a system that helps universities control exams and combat cheating. We have therefore proposed a smart monitoring system based on deep learning algorithms (YOLO v5), that works better with effective accuracy and faster detection speed than other algorithms for identifying cheating tools. The YOLO v5 algorithm was trained on a set of photos and videos in order to discover cheating devices and identify students who tried to cheat. Using the Siamese network, the student was identified by comparing their photos with those in the university's database. Using the proposed model experiment in this paper, we were able to improve human observation and combat fraud cases. According to the results of the experiment, the proposed integrated model detected cases of cheating and is considered a modern and new smart monitoring technique that aids universities in combating cheating, a problem that exhausts university administrations when it comes to monitoring examinations. A future goal of ours is to transform the monitoring/invigilation system from a software application into a specialized device used in universities to monitor exams. We seek to develop the proposed monitoring system to follow up and analyze the student's movements and behaviors during the exam. As a result, we will be able to extract reports showing the discipline of students and identify students who attempted cheating.

6. CONTRIBUTIONS

The purpose of this research is to contribute to the inclusion of artificial intelligence and deep learning in intelligent monitoring systems by developing a new model that integrates artificial intelligence and machine learning with monitoring to provide security. As a result of the research, cheating during exams can be minimized by identifying cellular devices and identifying students. This strengthens systems for monitoring and invigilating exams at universities and schools. We achieved our research goals by applying the model.

ACKNOWLEDGEMENTS




The authors appreciate the support of Al-Balqa Applied University in supporting this research, providing assistance in applying the proposed model to exams conducted on the university campus, and allowing us to use the student database.

REFERENCES




- [1] G. Sreenu and S. Durai, "Intelligent video surveillance: a review through deep learning techniques for crowd analysis," *Journal of Big Data*, vol. 6, no. 1, pp. 1–27, 2019, doi: 10.1186/s40537-019-0212-5.
- [2] M. J. Iqbal, M. M. Iqbal, I. Ahmad, M. O. Alassafi, A. S. Alfakeeh, and A. Alhomoud, "Real-time surveillance using deep learning," *Security and Communication Networks*, vol. 2021, pp. 1–17, 2021, doi: 10.1155/2021/6184756.
- [3] L. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu, and M. Pietikäinen, "Deep learning for generic object detection: A survey," *International Journal of Computer Vision*, vol. 128, pp. 261–318, 2020, doi: 10.1007/s11263-019-01247-4.
- [4] J. Brownlee, "A gentle introduction to object recognition with deep learning," *Machine Learning Mastery*, vol. 5, 2019.
- [5] Z. Zou, K. Chen, Z. Shi, Y. Guo, and J. Ye, "Object detection in 20 years: A survey," *Proceedings of the IEEE*, vol. 111, no. 3, pp. 257–276, 2023, doi: 10.1109/JPROC.2023.3238524.
- [6] R. A. A. Helmi, A. T. L. Lee, M. G. M. Johar, A. Jamal, and L. F. Sim, "Quantum checkout: An improved smart cashier-less store checkout counter system with object recognition," in *2021 IEEE 11th IEEE Symposium on Computer Applications & Industrial Electronics (ISCAIE)*. IEEE, 2021, pp. 151–156, doi: 10.1109/ISCAIE51753.2021.9431839.
- [7] X. Feng, Y. Jiang, X. Yang, M. Du, and X. Li, "Computer vision algorithms and hardware implementations: A survey," *Integration*, vol. 69, pp. 309–320, 2019, doi: 10.1016/j.vlsi.2019.07.005.
- [8] A. Hanafi, L. Elaachak, and M. Bouhorma, "Machine learning based augmented reality for improved learning application through object detection algorithms," *International Journal of Electrical & Computer Engineering*, vol. 13, no. 2, pp. 1724–1733, 2023, doi: 10.11591/ijece.v13i2.pp1724-1733.
- [9] V. S. Sadanand, K. Anand, P. Suresh, P. K. A. Kumar, and P. Mahabaleshwar, "Social distance and face mask detector system exploiting transfer learning," *International Journal of Electrical & Computer Engineering*, vol. 12, no. 6, pp. 6149–6158, 2022, doi: 10.11591/ijece.v12i6.pp6149-6158.
- [10] M. Inamdar and N. Mehendale, "Real-time face mask identification using facemasknet deep learning network," *SSRN*, 2020, doi: 10.2139/ssrn.3663305.

- [11] Y. J. Wai, Z. M. Yussof, and S. I. M. Salim, "A scalable fpga based accelerator for tiny-yolo-v2 using opencl," *International Journal of Reconfigurable and Embedded Systems (IJRES)*, vol. 8, pp. 206-214, 2019, doi: 10.11591/ijres.v8.i3.pp206-214.
- [12] S. Skansi, *Introduction to Deep Learning: from logical calculus to artificial intelligence*. Cham: Springer, 2018, doi: 10.1007/978-3-319-73004-2.
- [13] Z. Sun and Z. Wu, *Handbook of research on foundations and applications of intelligent business analytics advances in business information systems and analytics*, Pennsylvania, USA: IGI Global, doi: 10.4018/978-1-7998-9016-4.
- [14] T. Hiraiwa et al., "A deep-learning artificial intelligence system for assessment of root morphology of the mandibular first molar on panoramic radiography," *Dentomaxillofacial Radiology*, vol. 48, no. 3, 2019, doi: 10.1259/dmfr.20180218.
- [15] X. Li, Z. Yang, and H. Wu, "Face detection based on receptive field enhanced multi-task cascaded convolutional neural networks," *IEEE Access*, vol. 8, pp. 174 922–174 930, 2020, doi: 10.1109/ACCESS.2020.3023782.
- [16] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. Cambridge, Massachusetts: MIT press, 2016.
- [17] C. Ning, L. Menglu, Y. Hao, S. Xueping, and L. Yunhong, "Survey of pedestrian detection with occlusion," *Complex & Intelligent Systems*, vol. 7, pp. 577–587, 2021, doi: 10.1007/s40747-020-00206-8.
- [18] P. Ostwal, "Introduction to object detection for computer vision and AI," *TagX Data*. [Online]. Available: <https://www.tagxdata.com/introduction-to-object-detection-for-computer-vision-and-ai>
- [19] C. B. Murthy, M. F. Hashmi, N. D. Bokde, and Z. W. Geem, "Investigations of object detection in images/videos using various deep learning techniques and embedded platforms—a comprehensive review," *Applied sciences*, vol. 10, no. 9, 2020, doi: 10.3390/app10093280.
- [20] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Region-based convolutional networks for accurate object detection and segmentation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 38, no. 1, pp. 142–158, Jan. 2016, doi: 10.1109/TPAMI.2015.2437384.
- [21] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "Ssd: Single shot multibox detector," in *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*. Springer, 2016, pp. 21–37.
- [22] M. A. Ahad, S. Paiva, G. Tripathi, and N. Feroz, "Enabling technologies and sustainable smart cities," *Sustainable Cities and Society*, vol. 61, 2020, doi: 10.1016/j.scs.2020.102301.
- [23] A. Ji, W. L. Woo, E. W. L. Wong, and Y. T. Quek, "Rail track condition monitoring: A review on deep learning approaches," *Intelligence & Robotics*, vol. 1, pp. 151–175, 2021, doi: 10.20517/ir.2021.14.
- [24] K. Wang, X. Li, J. Yang, J. Wu, and R. Li, "Temporal action detection based on two-stream you only look once network for elderly care service robot," *International Journal of Advanced Robotic Systems*, vol. 18, no. 4, 2021, doi: 10.1177/17298814211038342.
- [25] P. Jiang, D. Ergu, F. Liu, Y. Cai, and B. Ma, "A review of yolo algorithm developments," *Procedia Computer Science*, vol. 199, pp. 1066–1073, 2022, doi: 10.1016/j.procs.2022.01.135.
- [26] F. Jubayer et al., "Detection of mold on the food surface using yolov5," *Current Research in Food Science*, vol. 4, pp. 724–728, 2021, doi: 10.1016/j.crf.2021.10.003.
- [27] Z. Li, X. Tian, X. Liu, Y. Liu, and X. Shi, "A two-stage industrial defect detection framework based on improved-yolov5 and optimized-inception-resnetv2 models," *Applied Sciences*, vol. 12, no. 2, 2022, doi: 10.3390/app12020834.
- [28] H.-K. Jung and G.-S. Choi, "Improved yolov5: Efficient object detection using drone images under various conditions," *Applied Sciences*, vol. 12, no. 14, 2022, doi: 10.3390/app12147255.
- [29] R. Xu, H. Lin, K. Lu, L. Cao, and Y. Liu, "A forest fire detection system based on ensemble learning," *Forests*, vol. 12, no. 2, 2021, doi: 10.3390/f12020217.
- [30] L. Sumi and S. Dey, "Yolov5-based weapon detection systems with data augmentation," *International Journal of Computers and Applications*, vol. 45, no. 4, pp. 288–296, 2023, doi: 10.1080/1206212X.2023.2182966.
- [31] C. W. Reynolds, "Flocks, herds and schools: A distributed behavioral model," in *Proceedings of the 14th annual conference on Computer graphics and interactive techniques*, 1987, pp. 25–34, doi: 10.1145/37402.37406.
- [32] W. Wang, S. Li, J. Shao, and H. Jumahong, "LKC-net: large kernel convolution object detection network," *Scientific Reports*, vol. 13, no. 1, 2023, doi: 10.1038/s41598-023-36724-x.
- [33] J. Kaur and W. Singh, "Tools, techniques, datasets and application areas for object detection in an image: a review," *Multimedia Tools and Applications*, vol. 81, no. 27, pp. 38 297–38 351, 2022, doi: 10.1007/s11042-022-13153-y.
- [34] Y. Chen, L. Li, W. Li, Q. Guo, Z. Du, and Z. Xu, *AI computing systems: an application driven perspective*. Amsterdam, Netherlands: Elsevier, 2022.
- [35] Y. Kortli, M. Jridi, A. Al Falou, and M. Atri, "Face recognition systems: A survey," *Sensors*, vol. 20, no. 2, 2020, doi: 10.3390/s20020342.
- [36] NEC, "Face detection vs facial recognition – what's the difference?" *NEC Publication and Media*, 2022. Accessed: August 24, 2023. [Online]. Available: <https://www.nec.co.nz/market-leadership/publications-media/face-detection-vs-facial-recognition-whats-the-difference/>
- [37] M. S. Ryoo, K. Kim, and H. Yang, "Extreme low resolution activity recognition with multi-siamese embedding learning," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 32, no. 1, 2018.
- [38] X. Liu, X. Tang, and S. Chen, "Learning a similarity metric discriminatively with application to ancient character recognition," in *Knowledge Science, Engineering and Management: 14th International Conference, KSEM 2021, Tokyo, Japan, August 14–16, 2021, Proceedings, Part I 14*. Springer, 2021, pp. 614–626, doi: 10.1007/978-3-030-82136-4_50.
- [39] Z. Chi and B. Zhang, "A sentence similarity estimation method based on improved siamese network," *Journal of Intelligent Learning Systems and Applications*, vol. 10, no. 4, pp. 121–134, 2018, doi: 10.4236/jilsa.2018.104008.
- [40] X. Liu, "Research on the relevant matching method of internet media information-stock assets based on the combination of literal and semantic," *Journal of Education, Humanities and Social Sciences*, vol. 4, pp. 228–233, 2022, doi: 10.54097/ehss.v4i.2771.
- [41] P. Bakshi, "Siamese networks - poulami bakshi," *Medium*, 2021. Accessed: August 24, 2023. [Online]. Available: <https://poulami98bakshi.medium.com/siamese-networks-d28ac0b0836d>
- [42] D. J. Rao, S. Mittal, and S. Ritika, "Siamese neural networks for one-shot detection of railway track switches," *arXiv-Computer Science*, pp. 1-7, 2017.




BIOGRAPHIES OF AUTHORS

Dr. Nabeel I. Zanoon    received his Ph.D. in Computer Systems Engineering, from South-West State University, Kursk, Russia, in 2011. He is faculty member with Al-Balqa' Applied University since 2011; where he is currently Associate Professor and Vice Dean of Aqaba University College as well as Director of the ICDL Computer Centre and Cisco Academy Branch of Aqaba University College. He received the rank of associate professor in 2021. He has published several research in several areas, artificial intelligence, computer vision, big data, security in networks, algorithm scheduling in grid and cloud, meta-grammar, fiber optical, mobile Ad Hoc networks. He has published over 25 papers in international journals and conferences from 2011 to July 2023. He can be contacted at email: dr.nabeel@bau.edu.jo or nabeelzanoon@gmail.com.



Dr. Abdullah A. Alhaj    received B.Sc. and M.Sc. degree in computer engineering from Lviv polytechnic institute - USSR, in 1988, Ph.D. in Computer Science from Bradford University UK, in 2008. Currently, he is an associate professor in the Department of Information Technology at The University of Jordan, Aqaba branch. His research interests include computer architecture, networks, IT security, machine learning, and AI. He can be contacted at email: aa.alhaj@ju.edu.jo.



Khalid Alkharabsheh    received a B.Sc. 2002 and M.Sc. 2005 degree in computer science from Yarmouk University and Al-Balqa Applied University, Jordan, respectively. He was appointed as a lecturer at Al-Balqa Applied University (BAU) in September 2006, won an Erasmus Mundus grant to pursue his Ph.D. in 2014 from the Research Center of Intelligent Technologies (CiTIUS) at Santiago de Compostela University in Spain, and was awarded his Ph.D. in 2019. He was appointed as an assistant professor in the Department of Software Engineering in 2019 and the head of the department from 2021 to date. His research interests include machine learning, big data, software quality, empirical software engineering, software validation and verification, and design smell detection. He is currently an assistant professor and works with different research teams and committees. He can be contacted at email: khalidkh@bau.edu.jo.