Gazi AKADEMİK YAYINCILIK

AIS ARTIFICIAL INTELLIGENCE STUDIES

ARTIFICIAL INTELLIGENCE STUDIES

Digitization of Printed Multiple Choice Questions Using Object Detection Methods: A Yolov7-Based Approach

Yiğit Çetinkaya*[a], Adem Tekerek[b]

[a,*] Havelsan
06510 - Ankara, Türkiye
ORCID: 0009-0007-5570-2754

[b] Gazi University,
Technology Faculty,
Dept. of Computer Engineering
06560 - Ankara, Türkiye
ORCID: 0000-0002-5979-4197

* Corresponding author.
e-mail: ycetinkaya@havelsan.com.tr

ABSTRACT

In education, multiple-choice questions are widely used in exams. Hard copies of multiple-choice questions are usually used to prepare students for exams. However, when hard-copy questions are intended to be used in electronic environments, difficulties are experienced because editing cannot be done, and statistical and mathematical operations cannot be performed with hard-copy questions. This study uses object detection methods to transfer hard-copy multiple-choice questions to digital media. Within the scope of the study, a new approach is presented to determine the locations of multiple-choice questions and convert them to text using the YOLOv7 algorithm. There are 16 different class labels, such as question stem, question number, and question options, in the proposed YOLO model. During the data preprocessing phase, 1,140 images were manually labeled using the Roboflow image data labeling system, resulting in the creation of a unique dataset. As a result of model training, it was seen that the model classified approximately 94% of the questions correctly. The data detected with the developed model was converted to XML format using the algorithms of the YOLOv7 library. Educational institutions can use this approach to extract data and understand visual information.

Keywords: Convolutional Neural Networks, Real-Time Object Detection, YOLO

Anahtar Kelimeler: Konvolüsyonel Sinir Ağları, Gerçek Zamanlı Nesne Tespiti, YOLO

Submitted: 06.10.2024
Revised: 18.12.2024
Accepted: 18.12.2024

doi: 10.30855/AIS.2024.07.02.04

Nesne Tespit Yöntemlerini Kullanarak Basılı Çoktan Seçmeli Soruların Dijitalleştirilmesi: YOLOv7 Tabanlı Bir Yaklaşım

ÖZ

Eğitimde, çoktan seçmeli sorular sınavlarda yaygın olarak kullanılmaktadır. Çoktan seçmeli soruların basılı kopyaları genellikle öğrencileri sınavlara hazırlamak için kullanılır. Ancak, basılı kopya soruların elektronik ortamlarda kullanılmak istendiğinde, düzenleme yapılamadığı ve istatistiksel ve matematiksel işlemler gerçekleştirilemediği için zorluklar yaşanmaktadır. Bu çalışma, basılı çoktan seçmeli soruları dijital ortama aktarmak için nesne tespit yöntemlerini kullanmaktadır. Çalışma kapsamında, YOLOv7 algoritmasını kullanarak çoktan seçmeli soruların konumlarını belirlemek ve metne dönüştürmek için yeni bir yaklaşım sunulmaktadır. Önerilen YOLO modelinde sorunun gövdesi, soru numarası ve soru seçenekleri gibi 16 farklı sınıf etiketi bulunmaktadır. Veri ön işleme aşamasında, Roboflow görüntü veri etiketleme sistemi kullanılarak 1.140 görüntü manuel olarak etiketlenmiş ve böylece benzersiz bir veri seti oluşturulmuştur. Model eğitimi sonucunda, modelin soruların yaklaşık %94'ünü doğru sınıflandırdığı görülmüştür. Geliştirilen model ile tespit edilen veriler, YOLOv7 kütüphanesinin algoritmaları kullanılarak XML formatına dönüştürülmüştür. Eğitim kurumları, bu yaklaşımı veri çıkarmak ve görsel bilgiyi anlamak için kullanabilirler.

To cite this article: Y. Çetinkaya and A. Tekerek, "Digitization of Printed Multiple Choice Questions Using Object Detection Methods: A Yolov7-Based Approach," *Artificial Intelligence Studies,* vol.7, no.2, pp. 66-82, 2024. doi: 10.30855/AIS.2024.07.02.04

1. Introduction

Hard-copy multiple-choice questions are widely used in education to measure students' knowledge levels. These questions are essential for evaluating student achievements and the effectiveness of educational institutions' curricula. However, such tests, traditionally prepared with hard copies, are time-consuming and costly. Manual writing, reviewing, and categorizing questions increases the process's complexity. With the development of technology, the need to digitize hard-copy printed materials and use them in electronic environments has emerged. Digitizing and processing hard copies brings with it specific difficulties. For this reason, artificial intelligence technologies are increasingly being used to optimize and improve these processes. In this study, technologies that can be used for digitizing printed multiple-choice test questions were examined. YOLOv7, an advanced object detection algorithm, was used to digitize multiple-choice test questions. This algorithm stands out because it identifies and classifies different multiple-choice question patterns. In exams with printed multiple-choice questions, it is generally impossible to perform detailed analyses. However, object detection methods can provide an essential solution in digitizing and processing these questions. This study aims to digitize and classify hard-copy multiple-choice questions using the YOLOv7 algorithm.

The model developed within the scope of the study aims to make test preparation and evaluation processes more efficient for educational institutions and educators. In the rest of the article, the working principles, areas of use, experimental results, and obtained findings of the YOLOv7 algorithm are explained in detail. The study aims to develop a new approach to digitizing traditional printed multiple-choice questions with object detection methods from artificial intelligence techniques. In this context, the proposed YOLO model can recognize different classes, such as question stem, question number, and question options. The data set obtained for training the model was created with the manual labeling method and trained using GeForce GTX TITAN X GPU. According to the study results, it was observed that the developed model correctly classified approximately 94% of the fundamental questions. This success reveals that object detection methods offer significant potential in the digitization and processing multiple-choice questions in educational institutions. The study is an important step in providing a more effective and efficient learning environment by pushing the boundaries in data extraction and understanding in the field of visual information in education. The increasing popularity of automation and artificial intelligence technologies for predicting exam questions makes reviewing previous studies in this area important. The potential to make more effective and accurate predictions than traditional methods direct researchers to these new technologies. Conventional image processing techniques and object detection algorithms have generally been used in the studies in the literature. However, these studies usually focus on limited question types and are inadequate in dealing with different question patterns. The prominent feature of this study is the use of the YOLOv7 object detection algorithm to predict multiple-choice test questions. YOLOv7 is a robust deep-learning algorithm for recognizing and classifying different question patterns, and this study has shown that the algorithm can effectively predict test questions. The study's contribution to the literature is creating an artificial intelligence model with the YOLOv7 algorithm to digitize multiple-choice test questions.

2. Materials and Method

This study uses the YOLOv7 object detection algorithm to identify multiple-choice printed test questions. YOLOv7 is an advanced deep-learning model that offers high accuracy and speed in real-time object detection tasks. In the scope of the study, a unique dataset is created to evaluate the algorithm's effectiveness. This dataset is designed to include various features required for accurate identification and classification of test questions.

2.1. Dataset and Data Collection

The dataset must encompass a wide range of question types and patterns to improve the model's performance. Consequently, it includes questions from various printed test books, ensuring comprehensive coverage. This variety will help the model recognize and classify different question patterns effectively.

Criteria for Dataset Creation: When compiling the dataset, specific criteria were established to categorize question patterns. These include:

- **Question Title:** The name or identifier of the question.
- **Question Text:** The actual content of the question.
- **Question Answer:** The correct response to the question.
- **Question Format:** The style of the question (e.g., multiple-choice, true/false).
- **Question Complexity:** An assessment of the difficulty level.

Labeling Methods: Question patterns can be identified through manual or automatic labeling. While manual labeling tends to yield more precise results, it is time-consuming and labor-intensive. Conversely, automatic labeling is quicker and easier but may lack the accuracy of manual methods. In this study, manual labeling was prioritized to ensure the highest quality of data and minimize potential inaccuracies.

Preprocessing Steps: Before training the model, several preprocessing steps were applied to the dataset to ensure data quality and consistency. These steps included:

- **Text Normalization:** Standardizing question text by removing special characters and converting all text to lowercase.
- **Tokenization:** Breaking down the question text into individual tokens (words) to facilitate analysis.
- **Filtering:** Removing irrelevant or duplicate entries to maintain dataset integrity.
- **Data Augmentation:** Applying techniques such as synonym replacement or slight rephrasing to enhance dataset diversity and robustness.

Model Utilization: The YOLOv7 model was employed for object recognition in this study. Developed on Google Colab using deep learning frameworks such as TensorFlow or PyTorch, the model underwent training using groups of test questions. This process enabled the model to learn how to identify and classify various question patterns effectively.

Dataset Structure: The dataset is a vital component for the success of the proposed artificial intelligence model. With a total of 1140 pages, the dataset was divided as follows:

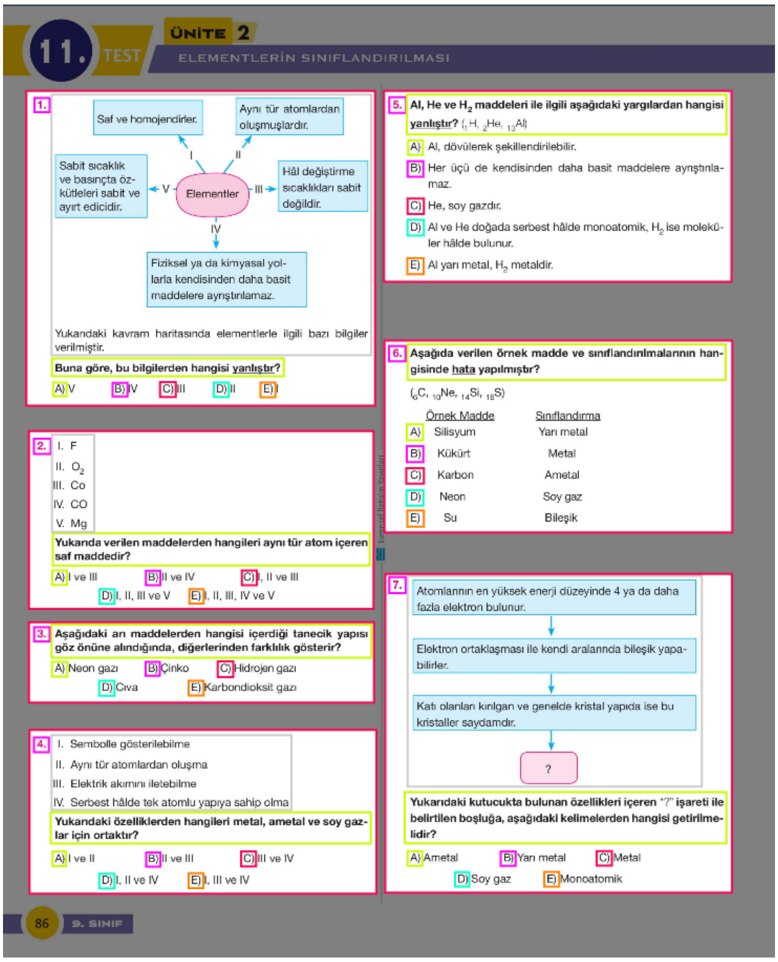
Table 1. Dataset and training, validation, and test rates

Data Split	Number of Pages	Percentage (%)
Training	880	77%
Validation	125	11%
Test	135	12%
Total	**1140**	**100%**

Although 1,140 images were manually labeled, expanding the diversity of the dataset would be beneficial for enhancing the model's generalizability. To achieve this, the dataset was sourced from a variety of educational institutions across different levels, including elementary, middle, and high schools. This approach ensured that the dataset reflects a broader range of question types and formats. Additionally, multiple institutions were involved in providing the data, further increasing the diversity and variability of the dataset. By expanding the dataset with samples from different educational contexts, we aimed to create a model that can generalize better across different environments and exam formats.

More thorough and precise labeling of the dataset will significantly enhance the performance of the YOLOv7 model. The dataset was meticulously collected and labeled from various printed books, with a strong focus on accurately identifying the question models represented by each test question. This collection and labeling process constituted the most time-consuming and challenging aspect of the study.

Visual Representation: Figure 1 illustrates a scanned page of multiple-choice test questions, showcasing the raw data used in the dataset.



11. TEST

ÜNİTE 2

ELEMENTLERİN SINIFLANDIRILMASI

1. Saf ve homojendirler. — I
Aynı tür atomlardan oluşmuşlardır. — II
Hâl değiştirme sıcaklıkları sabit değildir. — III
Fiziksel ya da kimyasal yollarla kendisinden daha basit maddelere ayrıştırılamaz. — IV
Sabit sıcaklık ve basınçta özkütleleri sabit ve ayırt edicidir. — V
Elementler

Yukarıdaki kavram haritasında elementlerle ilgili bazı bilgiler verilmiştir.

Buna göre, bu bilgilerden hangisi yanlıştır?

A) V B) IV C) III D) II E) I

2. I. F
II. O_2
III. Co
IV. CO
V. Mg

Yukarıda verilen maddelerden hangileri aynı tür atom içeren saf maddedir?

A) I ve III B) II ve IV C) I, II ve III D) I, II, III ve V E) I, II, III, IV ve V

3. **Aşağıdaki arı maddelerden hangisi içerdiği tanecik yapısı göz önüne alındığında, diğerlerinden farklılık gösterir?**

A) Neon gazı B) Çinko C) Hidrojen gazı D) Cıva E) Karbondioksit gazı

4. I. Sembolle gösterilebilme
II. Aynı tür atomlardan oluşma
III. Elektrik akımını iletebilme
IV. Serbest hâlde tek atomlu yapıya sahip olma

Yukarıdaki özelliklerden hangileri metal, ametal ve soy gazlar için ortaktır?

A) I ve II B) II ve III C) III ve IV D) I, II ve IV E) I, III ve IV

5. **Al, He ve H_2 maddeleri ile ilgili aşağıdaki yargılardan hangisi yanlıştır?** ($_1H$, $_2He$, $_{13}Al$)

A) Al, dövülerek şekillendirilebilir.
B) Her üçü de kendisinden daha basit maddelere ayrıştırılamaz.
C) He, soy gazdır.
D) Al ve He doğada serbest hâlde monoatomik, H_2 ise moleküler hâlde bulunur.
E) Al yarı metal, H_2 metaldir.

6. **Aşağıda verilen örnek madde ve sınıflandırılmalarının hangisinde hata yapılmıştır?**

($_6C$, $_{10}Ne$, $_{14}Si$, $_{16}S$)

	Örnek Madde	Sınıflandırma
A)	Silisyum	Yarı metal
B)	Kükürt	Metal
C)	Karbon	Ametal
D)	Neon	Soy gaz
E)	Su	Bileşik

7. Atomlarının en yüksek enerji düzeyinde 4 ya da daha fazla elektron bulunur.
↓
Elektron ortaklaşması ile kendi aralarında bileşik yapabilirler.
↓
Katı olanları kırılgan ve genelde kristal yapıda ise bu kristaller saydamdır.
↓
?

Yukarıdaki kutucukta bulunan özellikleri içeren "?" işareti ile belirtilen boşluğa, aşağıdaki kelimelerden hangisi getirilmelidir?

A) Ametal B) Yarı metal C) Metal D) Soy gaz E) Monoatomik

86 9. SINIF

Figure 1. Multiple choice test question paper used for education

2.2. YOLO

An object detection model such as YOLO (You Only Look Once) is a robust algorithm that uses convolutional neural networks. YOLO detects objects by analyzing an image in a single pass, making it an ideal deep-learning model for real-time object detection. YOLO is used to recognize objects in an image and determine their locations. Object detection involves locating and classifying objects in an image that people can identify. The main goal of YOLO is to quickly and efficiently identify objects in an image and determine their locations. This is a method that can be used in real-time applications. The basic principle of YOLO is based on dividing an image into small squares and analyzing each square. To identify objects in each square, the network tries to detect one or more objects. Thus, detecting all objects in the image simultaneously using a single network is possible. YOLO can be used in many areas, such as security cameras, autonomous vehicles, object recognition systems, video analysis, object detection, and tracking.

For example, an autonomous vehicle can use YOLO to detect objects in its surroundings and classify various objects for road safety. YOLOv7 is trained on a larger and more complex dataset than other versions. However, more recent versions like YOLOv8 have been used in the literature due to their improved performance and accuracy. Studies utilizing YOLOv8 for object detection and classification are available in the literature [17]. When this project was being developed, YOLOv7 had just been released. This single-stage object detection algorithm uses a new backbone architecture to more accurately estimate the size and angle of objects and produce a higher confidence score. To understand how the YOLO model works, it is necessary to consider the basic steps. First, the input image is converted to a specific size and fed into a deep neural network called the "backbone." This backbone passes the image through several convolutional and pooling layers. The convolutional layers use learned filters to extract different features of the image; these filters detect edges, corners, and patterns in the image. In each layer, the convolution process acts as a kernel on the feature map from the previous layer, emphasizing different features. Pooling layers are used to reduce and summarize the feature maps after convolution. For example, max pooling selects the most prominent feature in a particular region, preserving the essential features of that region and reducing its size. This process helps in learning the features of the image and extracting necessary information. Then, a section called the "detection head" receives the feature map produced by the backbone. This section uses features to detect objects of different sizes, where tasks such as classification and estimation of bounding box coordinates are performed. Unlike models based on region proposals, the YOLO model tries to detect all objects at once without subjecting the image to heavy processing. This approach provides efficiency in real-time applications.

Figure 2 shows the default architecture of the YOLOv7 model. YOLOv7 is an optimized Convolutional Neural Network (CNN) model for object detection. Due to the wider and longer layers of our developed model, we cannot go into the details here. However, in summary, our model is customized for more complex datasets and includes additional layers and optimizations that differ from the standard structure.

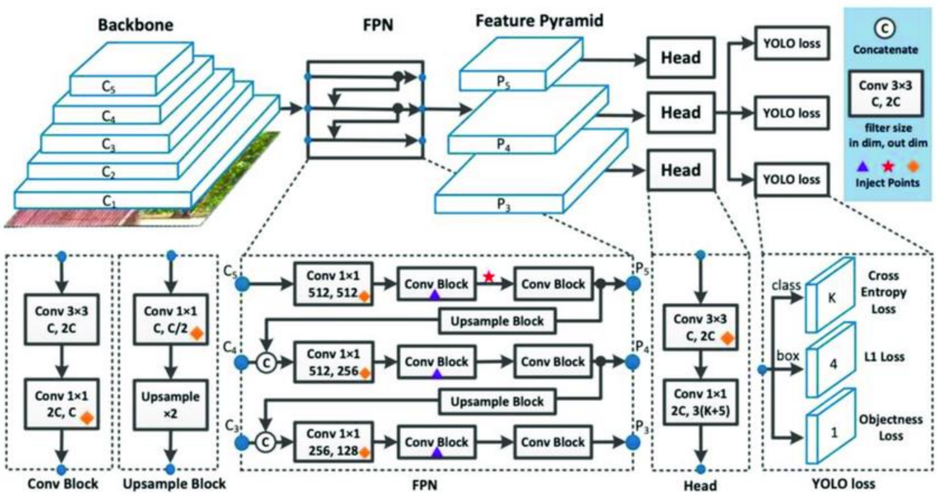


Figure 2. Default YOLOV7 Model Structure

The structure of our model is more complex and comprehensive than YOLOv7. It contains numerous layers to process the input data. The main components of the model include various convolution layers (Conv), activation layers (Sigmoid), multiplication operations (Mul), and concatenation operations (Concat). In summary, the layers and structure listed below enhance our model's deep learning capacity, enabling effective performance on larger datasets:

- Inputs: input.1
- Outputs: 688, onnx::Sigmoid_512, onnx::Sigmoid_575, onnx::Sigmoid_637
- Layers:
    - 74 Conv layers
    - 74 Sigmoid layers
    - 74 Mul layers
    - 12 MaxPool layers
    - 8 Concat layers
    - 1 Resize layers
    - 12 Constant layers

This structure increases the depth and complexity of our model, allowing us to achieve better results in object detection tasks.

3. Materials and Method

This study proposes a new approach using the YOLOv7 (You Only Look Once) algorithm for digitizing and converting hard-copy multiple-choice questions into text. YOLOv7 is an algorithm known for its fast and accurate object detection capacity, and these features have been effectively used in detecting components of multiple-choice exam questions. The main components of the proposed model are dataset and labeling, model training, object detection and location determination, and conversion of questions to XML format. The dataset consists of 1140 hard-copy multiple-choice question pages, as shown in Figure 1. Each question page was manually labeled using the Roboflow labeling system. During the labeling process, 16 class labels were defined, such as question stem, question number, and question options. This meticulous labeling process is critical for the model to work accurately.

Model training was performed using the YOLOv7 algorithm. YOLOv7's deep learning capabilities and fast processing capacity enable the model to work accurately. The YOLOv7 model is used to identify objects (question number, question stem, options, etc.) in the images of exam questions. The model detects the location of each class label, ensuring that the questions are transferred to the digital environment correctly. The detected data is converted to XML format using the algorithms of the YOLOv7 library. This conversion allows the data to be stored and used in an editable and analyzable format. The XML format is widely used for data extraction and analysis for educational institutions, facilitating the digital processing of multiple-choice exam questions. This approach reduces the digitalization of hard-copy exam questions and enables using this data in statistical and mathematical operations. Educational institutions can conduct a more efficient and effective exam preparation by digitizing exam questions with this model. The development stages of the proposed model are as follows.

Stage 1: Distinction of multiple choice question patterns and tag classes

The objects detected by the model developed using the YOLOv7 algorithm are divided into two main groups: question pattern (whole_question_box) and other classes (answer_a, answer_b, answer_c, answer_d, question_number_box, etc.). This distinction ensures that the classification process is more systematic and error-free. At this stage, the question pattern and other classes are placed in two arrays. Thus, the components of each question are grouped in an orderly manner. This step increases the success of the verification and classification processes to be carried out in later stages.

Stage 2: Control Algoritması

The control algorithm is used to detect whether there are any problems on the page by checking whether the question number is within the question pattern. If there is a problem on the page, the questions on the page are not used, and the next page is processed. This process ensures that the program works more stably and correctly. The algorithm first uses the question number to determine whether the question page is problem-free. It checks whether the question number is within the question pattern. If the question number is not within the pattern, it is assumed that there is an error on the question page, and the page is skipped without processing. This control mechanism ensures that the faulty pages are detected and not processed, thus increasing the program's overall performance and ensuring its accuracy.

Stage 3: Filtering and Sorting

After the filtering algorithms are applied, the data is searched again within the question pattern and placed in the arrays in an orderly manner. Any unauthorized intervention in the data flow may negatively affect the book where the question cut is made. In stage 3, the data is searched within the question pattern after completing the filtering processes. The obtained data is placed in the arrays in an orderly manner, and necessary precautions are taken to prevent data shifts. At this stage, to correct the errors made during labeling, the data belonging to the pure data is reduced by a few pixels, while the question pattern is enlarged by a few pixels. This process prevents errors that may occur due to incorrect labeling.

Stage 4: OCR and Sorting Algorithms

OCR (Optical Character Recognition) and sorting algorithms are used in the final stage. Data locations are sent to the OCR algorithm, and the incoming locations are cut off on the page and made smaller. Tesseract OCR and EasyOCR are used during this process. If the detection fails, different manipulation processes are applied. After the OCR process is successful, the data is sent to the sorting algorithm. The complicated incoming locations are corrected and sorted according to the question number. Finally, the data is added to the temporary data structure created to hold the data of the book pages, and the temporary page structure is reset. The process is carried out using the OCR algorithm to detect questions containing figures and formulas. The relevant numbers are detected and added to the data set by searching for specific words in the translated text. This process ensures that the questions on the question pages are classified correctly.

These stages and algorithms enable the YOLOv7 model to correctly detect and classify printed multiple-choice questions. The model's performance has reached high accuracy rates thanks to the hyperparameters and detailed control algorithms optimized during training.

The proposed method includes the following steps:

1. **Data Collection and Labeling:** Test questions are collected and labeled with relevant question patterns.

2. **Model Training**: The YOLOv7 model is trained using the labeled dataset, and a classification model is developed based on different question patterns.

3. **Test Question Prediction:** The trained YOLOv7 classification model predicts question patterns in new test questions.

For successful model training, the model should be trained with a correctly labeled dataset. The labeled dataset should be as large as possible. The performance of the model can be improved using training parameters. The training parameters determine how the model is trained. The performance of the model should be evaluated using regular tests. The training process can be repeated if the model's performance is unsatisfactory. The trained YOLOv7 model can be used to predict question patterns in new test questions. This process is done by giving the model an image of a test question and allowing it to predict question patterns. Predictions can be evaluated using a classification matrix. The classification matrix shows how the expected question patterns match the actual question patterns. In this way, we can manage the classification most accurately. Then, we can decide on the order and management of the data.

The proposed method has several advantages over traditional methods:

- Efficient: It can predict the question patterns of multiple test questions at the same time.
- Accurate: It is trained on a labeled dataset.
- Versatile: It can be used to predict the question patterns of different question types.

The proposed method can be used to improve the process of preparing and evaluating test questions. The potential applications of this method are:

- Automatic preparation of test questions
- Evaluation of the quality of test questions
- Determination of the difficulty level of test questions

Proposed training and runtime stages are presented in Figure 3 and Figure 4.

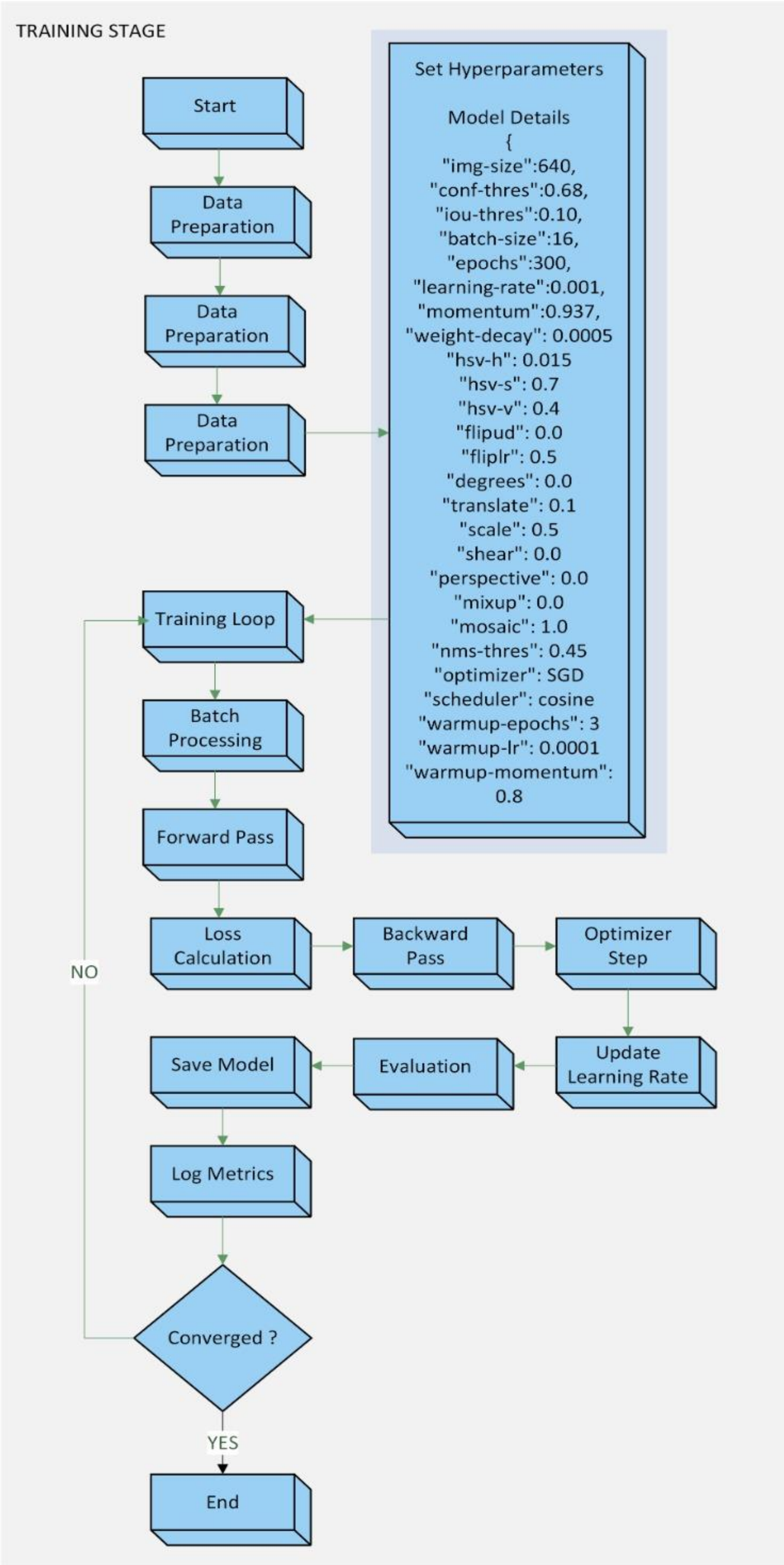


Figure 3. Training Structure

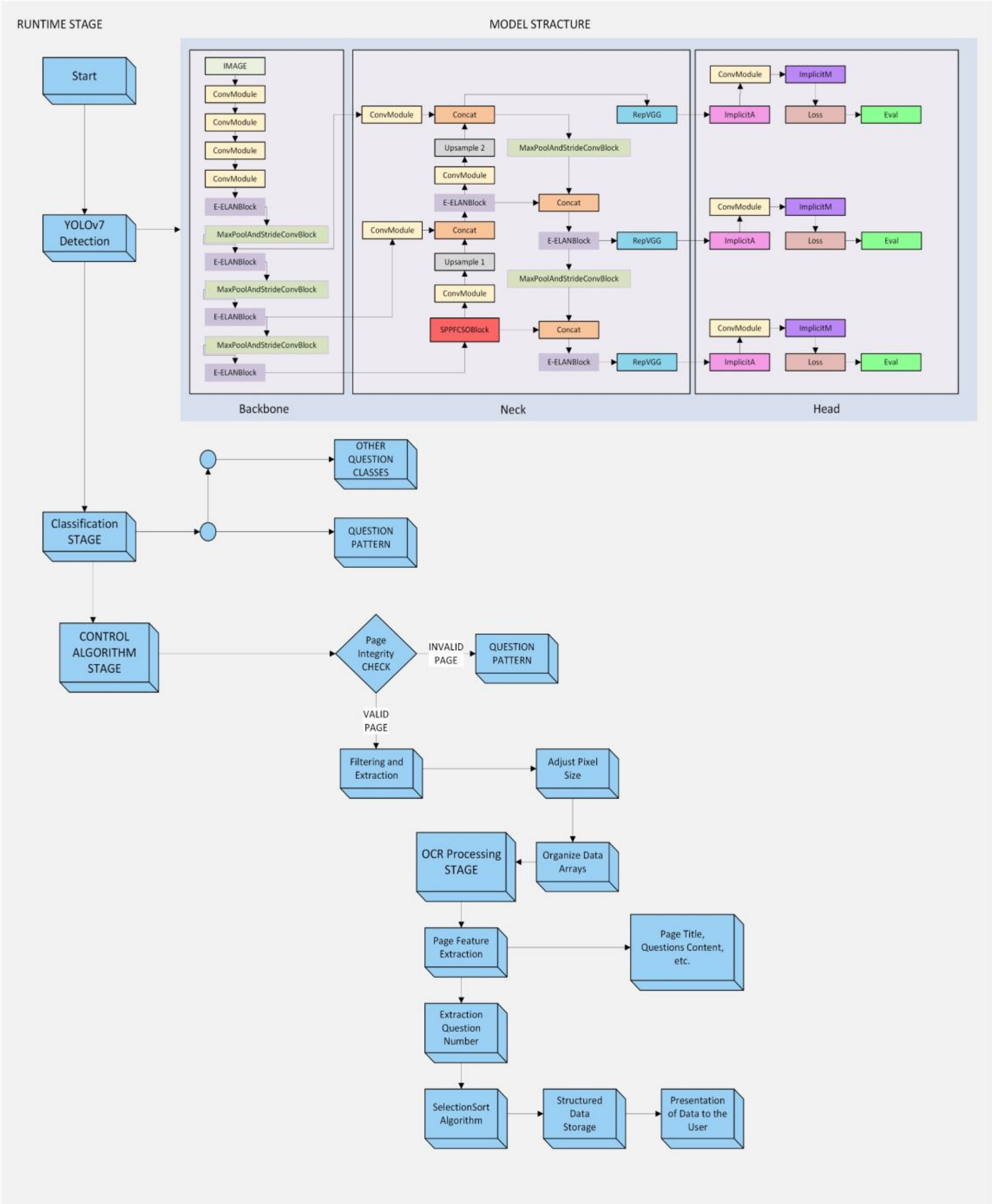


Figure 4. Runtime Structure

4. Experiments Results

This section explains the dataset used in the study, the training, validation, and testing stages, and the experimental results. Detecting and classifying printed multiple-choice questions using the YOLOv7 model will be discussed in detail. The dataset consists of 1140 pages collected from different sources and manually labeled, with each page containing an average of 5 or 6 test questions. Of these, 77% (880 pages) is reserved for training, 11% (125 pages) for validation, and 12% (135 pages) for testing.

The YOLOv7 model is trained on Google Colab using deep learning frameworks such as TensorFlow and PyTorch. The training focused on improving the model's ability to identify and classify different question patterns. In this study, transfer learning was applied by utilizing pre-trained weights from earlier YOLO models. This helped to accelerate the training process and improve the model's performance by fine-tuning these pre-trained weights on the specific dataset. Previous YOLO models served as references throughout the training process. Transfer learning allows the new model to learn faster and more efficiently by using the weights learned in previous models as a starting point and fine-tuning them on the new dataset.

In this study, the effects of transfer learning on training duration and model performance have been thoroughly discussed. For example, transfer learning led to a faster training process and an increase in overall accuracy. The training speed was significantly faster compared to training from scratch. Additionally, the pre-trained model weights were optimized to better adapt to the diversity of the dataset, helping the model generalize more effectively across different question types and formats.

The dataset is divided into three parts:

- Training Set: 880 pages (77%)
- Validation Set: 125 pages (11%)
- Testing Set: 135 pages (12%)

The breadth and diversity of the dataset are critical for enhancing the model's overall performance. This division follows established practices in machine learning, including the Pareto rule, commonly referenced in the literature for guiding data splits.

The hyperparameters used in the training of the YOLOv7 model were carefully selected to optimize its performance, summarized in Table 2.

Table 2. Hyperparameters of the proposed model

Hyperparameters	**Value**
Learning Rate	0.001
Batch Size	16
Epoch	250
Optimizer	Adam

These hyperparameters are designed to ensure the model achieves high accuracy rates across training, validation, and testing sections. Regular performance monitoring during training allowed for timely adjustments to these hyperparameters. For instance, if the learning rate was found to be too high, leading to instability in loss reduction, it was subsequently lowered. Similarly, the batch size was adjusted to balance training efficiency and convergence speed. Data augmentation techniques, such as random cropping, rotation, and scaling, were also employed to enhance the dataset and improve the model's robustness. The optimization of these hyperparameters involved a systematic approach, initially testing a range of values through grid search and random search methods to evaluate model performance across various combinations.

The OCR technologies (Tesseract OCR, EasyOCR etc.) and data augmentation techniques used in the model's development process ensure that the model gives successful results on different data sets. During the training process, the model's performance was regularly evaluated on the validation set, and the development of the model was monitored. Below are some critical performance metrics and losses recorded during the training process

Table 3. Training Performance Metrics Over Epochs

Epoch	**Accuaracy(%)**		**Loss(%)**	
	Training	**Validation**	**Training**	**Validation**
1	72	70	0.45	0.47
50	85	83	0.30	0.32
100	90	88	0.20	0.25
150	92	90	0.15	0.22
200	93	92	0.10	0.18
250	94	93	0.05	0.15

In Figure 5, the loss of training encountered by the proposed model in the training and validation stages. As the training losses decreased, the validation losses also decreased, and the model's generalization ability increased. The decrease in the Box Loss, Objectness Loss, and Classification Loss values shows that the model's performance increased and the errors decreased.

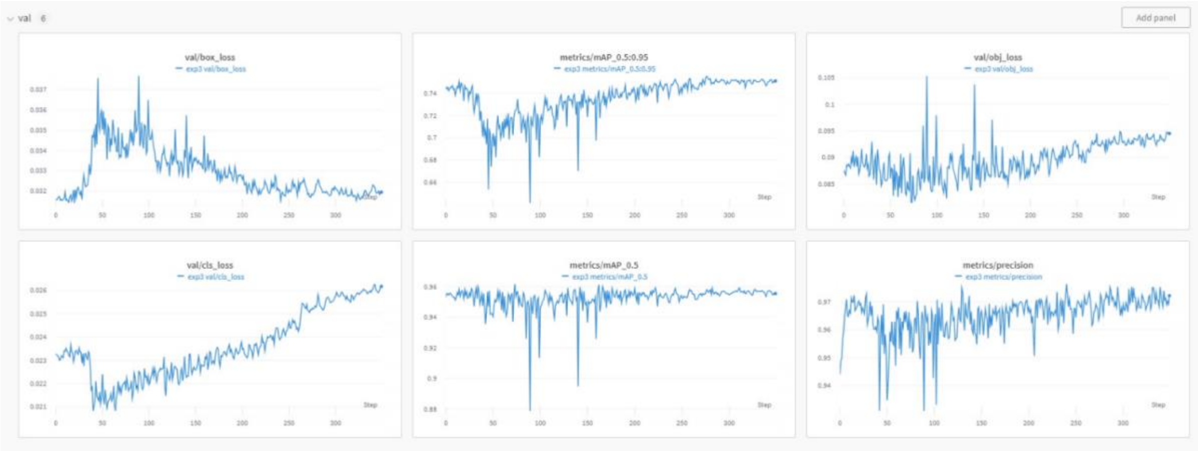


Figure 5. Training and Validation Loss Plot

The model's performance was evaluated using various metrics at the end of the training process. The model's accuracy, precision, recall rate, and F1 score on the training and validation sets are summarized in Table 4 below.

Table 4. Model Performance Metrics

Hyperparameters	**Value**	**Validation Set (%)**
Accuaracy	94	93
Precision	96	95
Recall	92	91
F1-Score	94	93

The losses and performance metrics monitored throughout the training process show that the model's overall performance is high and gives successful results on the validation set. The model's performance was evaluated using accuracy, precision, recall, and F1 score metrics. The tables below show the model's performance in the training, validation, and testing stages.

Table 5. Performance Metrics of the YOLOv7 Model

Metric	**Training (%)**	**Validation (%)**	**Test (%)**
Accuaracy	94	94	93
Precision	96	96	95
Recall	92	92	91
F1-Score	94	94	93

The model's accuracy was 94%, precision was 96%, recall was 92% and F1 score was 94%. These results show that the model exhibited high performance and could classify the test questions correctly. The confusion matrix was used to evaluate the model's classification performance in more detail. The confusion matrix shows the relationship between the predicted and actual values of the model. This matrix shows the model's true positive, true negative, false positive, and false negative classifications. The matrix is used to analyze in which cases the model was successful and in which cases it made mistakes.

The confusion matrix given in Figure 6 shows the model's prediction performance for each class in detail. The diagonal cells in the graph represent the classes that the model predicted correctly, and the high values indicate that the model's classification success is high. For example, the correct prediction rates for "answer_a," "answer_b," "answer_c," and other answer options are pretty high, which shows that the model recognized these classes correctly. The cells "background FN" and "background FP" in the graph show the false positive and false negative predictions made by the model for background objects. The low values of these values indicate that the model's errors regarding background objects are minimal. Overall, the confusion matrix plot shows that the model achieves high accuracy rates across classes and that errors are minimal. These results confirm that the model can effectively recognize and classify different components of multiple-choice questions.

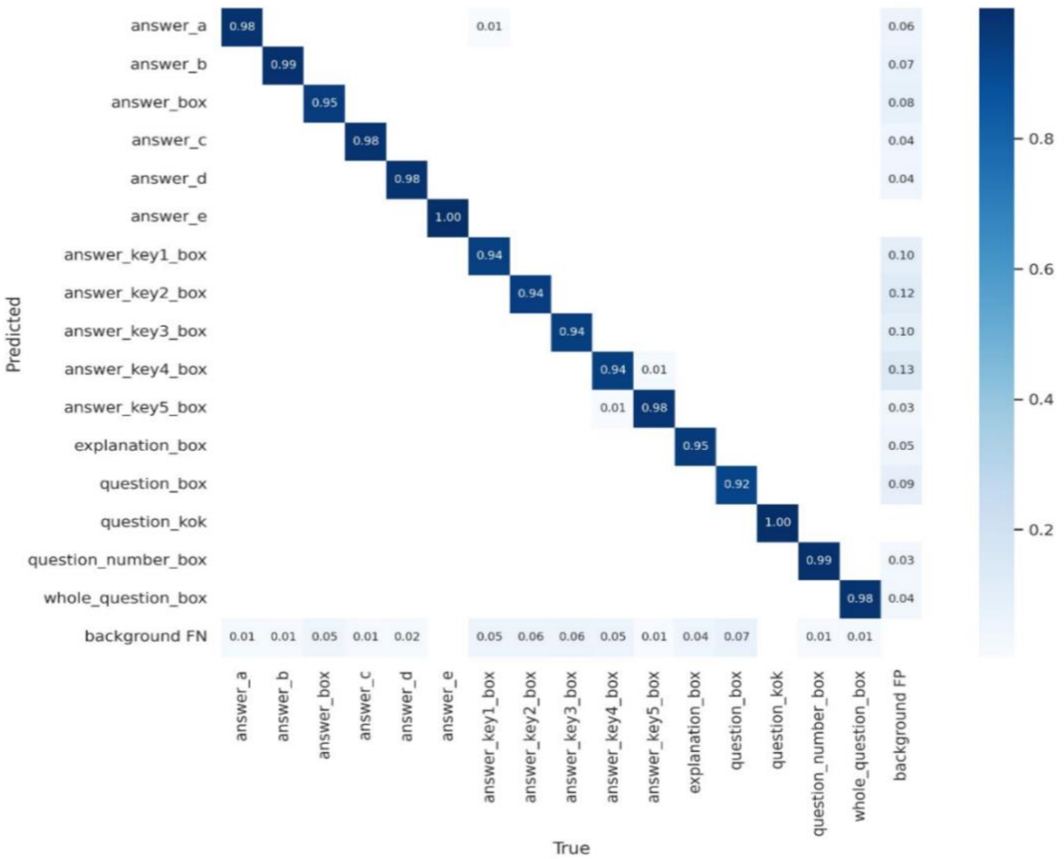


Figure 6. Confusion Matrix

The confusion matrix graph in Figure 6 shows the model's prediction performance for each class in detail. The diagonal cells in the graph represent the classes that the model predicted correctly, and high values in these cells indicate that the model's classification success is high. For example, the correct prediction rates for "answer_a", "answer_b", "answer_c" and other answer options are quite high, indicating that the model correctly recognized these classes. The "background FN" and "background FP" cells in the graph show the model's false negative and false positive predictions for background objects. The low values indicate that the model's errors regarding background objects are minimal.

The confusion matrix shows that the proposed model achieves high accuracy rates across various classes and minimal errors. These results confirm that the model can effectively recognize and classify different components of multiple-choice questions. The losses encountered by the model during the training and validation stages are summarized in Table 6 below.

Table 6. Training and Validation Loss

Loss Type	**Training Loss**	**Validation Loss**
Accuaracy	94	93
Precision	96	95
Recall	92	91

Figure 7 shows the performance of the model in correctly delimiting objects (Box Loss), correctly predicting the existence of objects (Objectness Loss), and correctly classifying objects (Classification Loss). The differences between the training and validation losses are used to evaluate the model's generalization ability and the training process's effectiveness. Performance metrics, including accuracy, precision, and recall rates support the experimental results of the model. The graphs below summarize how the model performed during the training process and the results.

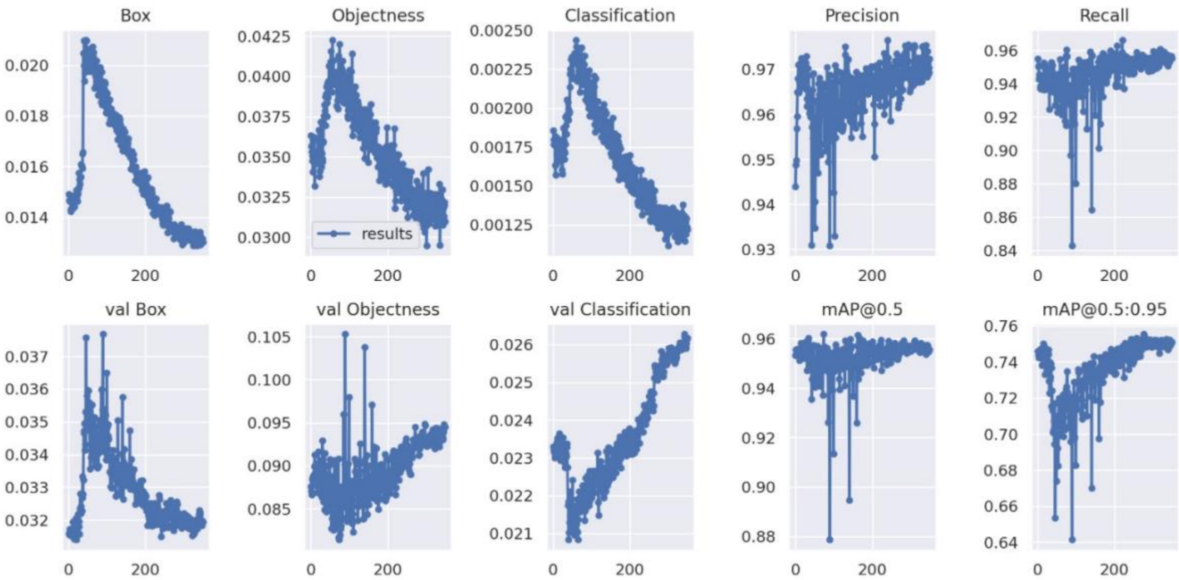


Figure 7. Model Performance

The accuracy rate of the proposed model is 94%, the precision rate is 96%, and the completeness rate is over 92%. The losses and performance metrics of the model during the training and validation processes are shown in Figure 1 and Figure 4. These graphs reveal how the accuracy and losses of the model change throughout the training process. Figure 1 and Figure 4 show how the model's training and validation losses (loss) decrease over time. Box Loss, Objectness Loss, and Classification Loss graphs reveal that the accuracy and performance of the model increase continuously throughout the training process, and the losses decrease. Val Box, Objectness, and Val Classification graphs show that the losses in the validation phase decrease similarly, and the generalization ability of the model increases. It is shown that the YOLOv7-based deep learning model can correctly classify the test data and generalize what it learns from the training data to the test data.

The precision rate of the model could be slightly higher, which may indicate that the model falsely rejects true positives in some cases. This can be improved by training the model with more data or adjusting the model parameters. Figure 5 shows how the model performed on an actual test question. In this figure, it can be seen that the model correctly identified the different components of the multiple-choice questions. Each question and answer box was correctly identified and marked. The model could accurately identify the question stem, the options, and the correct answer. Figure 8 gives an example of detecting and marking parts of questions in a test book using the proposed model.

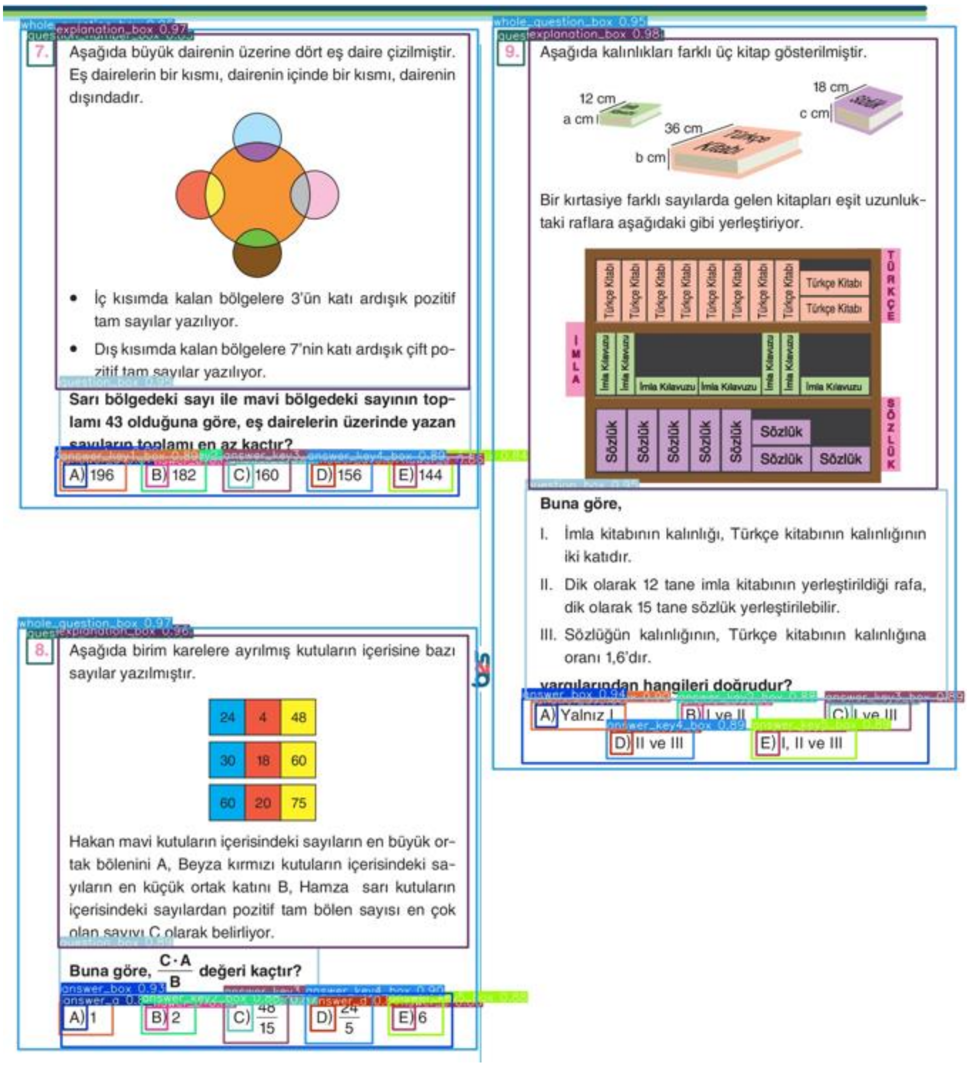


Figure 8. Test Question Output Sample

This study demonstrates the applicability and accuracy of the model in real-world scenarios. The model successfully recognizes and classifies various question types and formats, proving to be an effective tool for use in educational environments.

5. Experiments Results

This study evaluates the effectiveness of the YOLOv7 object detection algorithm in digitizing printed multiple-choice exam questions. It was observed that the model achieved high accuracy rates and successfully performed the classification thanks to the hyperparameters optimized in the training and validation processes. The model's performance was verified by testing on various datasets and was proven to work with high precision. One of the essential contributions of this study to the literature is that it significantly reduces the labor force in digitizing printed exam questions. While digitizing a book with traditional methods can take hours, the developed model reduces this time by 95%. This speed and accuracy provide an excellent advantage for educational institutions. The developed model is currently actively used in the sector, and academic institutions accelerate and make exam preparation and evaluation processes more efficient thanks to this system. The system's success is not limited to saving time and labor but also increases the accuracy and reliability of the exam processes. However, further development of this system is needed, particularly in terms of expanding its application to other types of test questions and materials. One key area for improvement is increasing the dataset by incorporating more labeled resources, which will enable the model to better generalize across different types of exam formats and question structures. As more data is labeled and integrated, the model's performance can be enhanced to handle even more complex or varied question formats. In addition, the performance and scope of the system can be expanded by integrating with other object detection and optical character recognition (OCR) technologies. Future research should focus on these enhancements to improve the model's accuracy and applicability, ensuring it can address a wider range of exam formats and materials. In conclusion, this study shows that the YOLOv7 object detection algorithm has great potential in the education sector and can significantly contribute to digitizing printed exam questions. Using such innovative technologies for educational institutions increases efficiency and improves the quality of education.

Additionally, this system is actively employed in educational platforms, significantly reducing the workload that used to require teams of 30-35 individuals. What previously demanded manual labor is now efficiently handled by the program, which automatically categorizes and labels exam questions according to their respective classes. This automated process not only increases operational efficiency but also enhances the overall accuracy and speed of exam preparation and grading.

Conflict of Interest Statement

No conflicts of interest were declared by the authors.

References

[1] J. Baek, B. Lee, D. Han, S. Yun, and H. Lee, "Character region awareness for text detection," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), Long Beach, CA, USA, June 2019, pp. 9365-9374.

[2] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, June 2016, pp. 770-778.

[3] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in Proc. Eur. Conf. Computer Vision (ECCV), Amsterdam, Netherlands, Oct. 2016, pp. 21-37.

[4] Z. Liu, Y. Lin, Y. Cao, H. Hu, Y. Wei, Z. Zhang, S. Lin, and B. Guo, "Swin Transformer: Hierarchical vision transformer using shifted windows," in Proc. IEEE Int. Conf. Computer Vision (ICCV), Montreal, QC, Canada, Oct. 2021, pp. 10021-10030.

[5] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in Advances in Neural Information Processing Systems (NIPS), Montreal, QC, Canada, Dec. 2015, pp. 91-99.

[6] A. Bochkovskiy, C. Wang, and H. M. Liao, "YOLOv4: Optimal speed and accuracy of object detection," arXiv preprint arXiv:2004.10934, 2020. [Online]. Available: https://arxiv.org/abs/2004.10934. [Accessed: Sept. 18, 2024].

[7] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," arXiv preprint arXiv:2010.11929, 2020. [Online]. Available: https://arxiv.org/abs/2010.11929. [Accessed: Sept. 18, 2024].

[8] M. Liao, Z. Wan, and C. Yao, "Real-time scene text detection with differentiable binarization," in Proc. AAAI Conf. Artificial Intelligence, vol. 34, no. 7, June 2020. [Online]. Available: https://www.aaai.org/ocs/index.php/AAAI/AAAI20/paper/view/6350. [Accessed: Sept. 18, 2024].

[9] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," arXiv preprint arXiv:1804.02767, 2018. [Online]. Available: https://arxiv.org/abs/1804.02767. [Accessed: Sept. 18, 2024].

[10] C. Wang, A. Bochkovskiy, and H. M. Liao, "YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors," arXiv preprint arXiv:2207.02696, 2022. [Online]. Available: https://arxiv.org/abs/2207.02696. [Accessed: Sept. 18, 2024].

[11] X. Zhou, D. Wang, and P. Krähenbühl, "Objects as points," arXiv preprint arXiv:1904.07850, 2019. [Online]. Available: https://arxiv.org/abs/1904.07850. [Accessed: Sept. 18, 2024].

[12] C. Shorten and T. M. Khoshgoftaar, "A survey on image data augmentation for deep learning," Journal of Big Data, vol. 6, no. 1, pp. 60, Dec. 2019. doi: 10.1186/s40537-019-0197-0.

[13] F. Chollet, Deep Learning with Python, 1st ed., New York, NY, USA: Manning Publications, 2017.

[14] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning, Cambridge, MA, USA: MIT Press, 2016.

[15] A. Radford, K. Narasimhan, T. Salimans, and I. Sutskever, "Improving language understanding by generative pre-training," OpenAI, June 2018. [Online]. Available: https://openai.com/research/gpt. [Accessed: Sept. 18, 2024].

[16] R. Smith, "An overview of the Tesseract OCR engine," arXiv preprint arXiv:2009.12355, Sept. 2020. [Online]. Available: https://arxiv.org/abs/2009.12355. [Accessed: Sept. 18, 2024].

[17] G. Arslan, F. Aydemir, and S. Arslan, "Enhanced license plate recognition using deep learning and block-based approach," Journal of Scientific Reports-A, vol. 058, pp. 57-82, 2023.