2024, 21(12):3914-3926
https://doi.org/10.21123/bsj.2024.9199
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

An Object Detection Model based on Augmented Reality for Iraqi Archaeology

Suha Dh. Athab**, Abdulamir Abdullah Karim

Department of Computer Science, University of Technology, Bagdad, Iraq.
*Corresponding Author.

Received 07/06/2023, Revised 24/09/2023, Accepted 26/09/2023, Published Online First 20/05/2024, Published 01/12/2024

Abstract

The culture of Iraq, which boasts a rich history, serves as evidence of the magnificence of human civilization. Nonetheless, safeguarding and highlighting this valuable cultural legacy has become a significant worry in a time characterized by technological progress. Augmented Reality (AR) offers a powerful tool for preserving and presenting historical sites. The aim of this research is to leverage AR technology as a means to ensure the continued preservation and dynamic presentation of Iraq's cultural heritage. This study explores the capabilities of CNNs as the basis of AR's development. CNN is used as an essential initial step in constructing AR systems. The proposed model utilizes a pre-trained backbone network to extract complicated spatial features from input images; additional convolutional and fully connected layers are introduced to refine these features. A new custom class called "AnchorBoxes", dynamically generates predefined anchor boxes for each feature map. Since there is not an appropriate Iraqi archeology dataset available for training deep learning models, a dataset of 2188 color images was collected. Spanning ancient Iraqi ruins, celebrated monuments, and real-time scenes combined with various objects. This dataset is subjected to manual annotation, wherein bounding boxes and labels are assigned to objects in each image. Results from the regression analysis emphasize the model's proficiency in estimating object bounding box coordinates with good precision and regression loss equal 0.008, facilitating locate-accurate object localization. The classification outcomes illuminate the model's ability to assign class labels to detected objects with high confidence. The mAP for the trained model was 0.84 and the classification loss was 0.02

Keywords: Anchor boxes, Classification, Computer Vision, Localization, Object detection.

Introduction

Iraqi culture, with its rich history, stands as a testament to the great of human civilization. Nevertheless, the preservation and highlighting of this abundant cultural heritage have emerged as a paramount challenge in a time characterized by remarkable technological progress. Augmented Reality (AR) systems are increasingly being used in the field of archaeology [1, 2] preserve historical sites [3]. AR able to blend the physical and virtual worlds offers a powerful tool for safeguarding and promoting Iraq's culture. It enables us to bridge the gap between the ancient past and the present [4]. This paper exploited Convolutional Neural Networks (CNNs) to elevate the accuracy and functionality of AR in Iraqi archaeology. [5]

This study delves into the realm of object detection, a pivotal initial step in constructing AR systems that resonate with historical significance. The proposed model employs a pre-trained backbone network,

2024, 21(12):3914-3926
https://doi.org/10.21123/bsj.2024.9199
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

adept at extracting intricate spatial features from input images.

To refine these features, additional convolutional and fully connected layers were introduced. A new custom layer, called Anchor Boxes, dynamically generates predefined anchor boxes for each feature map.

Due to the lack of Iraqi archaeology datasets, it is difficult to train deep learning models effectively. This study starts by collected dataset for Iraqi archeology and annotate it. A dataset of 2188 color images spanning ancient Iraqi ruins, celebrated monuments, and real-time scenes with various objects. This dataset, managed by the custom dataset class, experienced manual annotation, where bounding boxes and labels assigned to objects in each image.

The results of the regression analysis highlight the model's ability in estimating object bounding box coordinates good precision, facilitating identify-accurate object localization. Simultaneously, the classification outcomes show the model's ability to assign class labels to detected objects. The utilization of AR in Iraqi archaeology has the potential to revolutionize how visitors experience and understand the past. By combining AR technology with advanced deep learning, archaeologists and visitors alike can interact with historical sites and artifacts in a virtual environment, providing an immersive and educational experience.

It is essential to acknowledge the limitations of our study. Despite the promising results; the custom dataset collected for this research, may still be relatively small in comparison to more extensive datasets used in computer vision. This limitation can affect the model's ability to generalize to a broader range of archaeological contexts. Despite efforts to address data variability with the anchor box class, there may still be challenges in handling all possible variations in archaeological imagery, such as lighting conditions, weather, and terrain. The model's architecture and training parameters choice based on the available resources and computational power. This could potentially limit the model's performance compared to more complex architectures or longer training cycles. While the model has shown promise in Iraqi archaeological imagery, its ability to generalize to other regions with different archaeological characteristics remains to be tested. The transition from a successful object detection model to a practical augmented reality application involves additional challenges, such as real-time camera input processing, accurate pose estimation, and robust object tracking. These steps may introduce new limitations and complexities. Dependency on Hardware: The effectiveness of the augmented reality (AR) component will depend on the hardware used for implementation. The performance may vary on different devices, which could affect user experience. In the pages that follow; the details of the proposed approach was explored and how CNNs and AR technology together pave the way for a deeper connection with Iraq's rich cultural heritage.

Literature Review

Prior to delving into the research that underpins our work, it is essential to set the stage by exploring the overarching context of Augmented Reality (AR) and its growing significance in archaeology. Augmented reality systems can overlay digital information onto the real world, allowing users to see historical sites as they appeared in their prime and learn about their cultural significance, architectural features, and history. [6,7], driven by advancements in technology and increasing consumer demand for more immersive and interactive experiences[8,9].

Deep learning techniques improve AR capabilities and user experience [10, 11]. Convolutional Neural Networks (CNNs) have been used to improve image and object recognition in AR applications[12]. These algorithms can detect and identify objects in real time and overlay digital information on them[13].

Pose estimation is a crucial component of AR, as it determines the location and orientation of objects in the real world. “DeepPose” and “OpenPose” algorithms have been used to estimate the pose of objects in AR applications[14].

For AR applications to work properly, they must be able to identify the details of the scene in order to place digital content onto real-life objects with precision. [15]. Semantic segmentation algorithms are used for this purpose[16].

The power of deep learning was explored for automating the detection of archaeological

2024, 21(12):3914-3926
https://doi.org/10.21123/bsj.2024.9199
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

features[17], using Cold War-era CORONA Satellite Imagery and quanat shafts in Iraq's Kurdistan Region as a test case. This marks a crucial shift in archaeology towards leveraging advanced computational techniques to tackle the challenges posed by data overload and the rapid degradation of archaeological sites.

The network architecture consists of 22 convolutional, five max pooling, and 5 up-sampling layers. Convolutional layers applied without padding. The model performs well on high-density qanat images, defined as containing more than 100-labeled shafts. Two deep-learning architectures were used to evaluate and classify ancient artifacts sourced from the British Museum's collection [18].

The method compares two CNN models, InceptionResNetV2 and VGG-19, on a dataset of 55,000 ancient artifact images from 343 cultures worldwide, resulting in high accuracy for InceptionResNetV2 (65%) and poor performance for VGG-19 (12.95%). The InceptionResNetV2 models tested on artifacts of unknown origin, revealing intriguing insights for archaeology's image recognition potential.

A method that integrates a deep learning architecture into a workflow called WODAN was introduced [19]. It involves LiDAR data preprocessing, multi-class object detection with Faster R-CNN, and the conversion of results into geographical data. The model's performance, evaluated using metrics like MaF1-score, suggests its potential for detecting and categorizing archaeological objects, especially barrows and Celtic fields. Addressing dataset complexities, and incorporating domain information for broader archaeological mapping applications.

The paper provides a detailed evaluation of the model's performance, showing its strengths in detecting barrows and highlighting areas for improvement, such as detecting charcoal kilns. The model obtains an average performance of 0.49 (average MaF1 score of all experiments).

A technique for automated archaeological object detection in LiDAR data using the WODAN workflow[20] was introduced. The method uses transfer learning and an adapted Faster R-CNN model that detects and categorizes two archaeological object types (barrows, and Celtic fields) with suitable performance. Faster R-CNN demonstrates promise among top-performing methods for LiDAR-based object detection. The deep CNN approach exhibited strong performance in detecting and segmenting archaeological objects. The approach adapted well to detecting diverse archaeological structures within a specific region. The study also revealed potential improvements through negative training and emphasized the need for careful training sample selection. The deep CNN approach's overall performance assessed across 110 experimental datasets (comprising 10 datasets, each with 11 different training sizes).

In this research, the architecture of object detection, which is an initial step in the AR model suggested. The model takes an input image and a set of candidate regions (reference anchors) and outputs a set of bounding boxes and class probabilities for each object in the image.

Materials and Methods

Accurate localization and robust classification of objects within images is an essential initial requirement for the AR model. The model leverages a pre-trained backbone network to extract rich spatial features from the input images. Additional convolutional and fully connected layers employed to further refine the extracted features and enable regression and classification tasks. The suggested detection model construction required multiple steps as shown in Fig. 1.

2024, 21(12):3914-3926
https://doi.org/10.21123/bsj.2024.9199
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

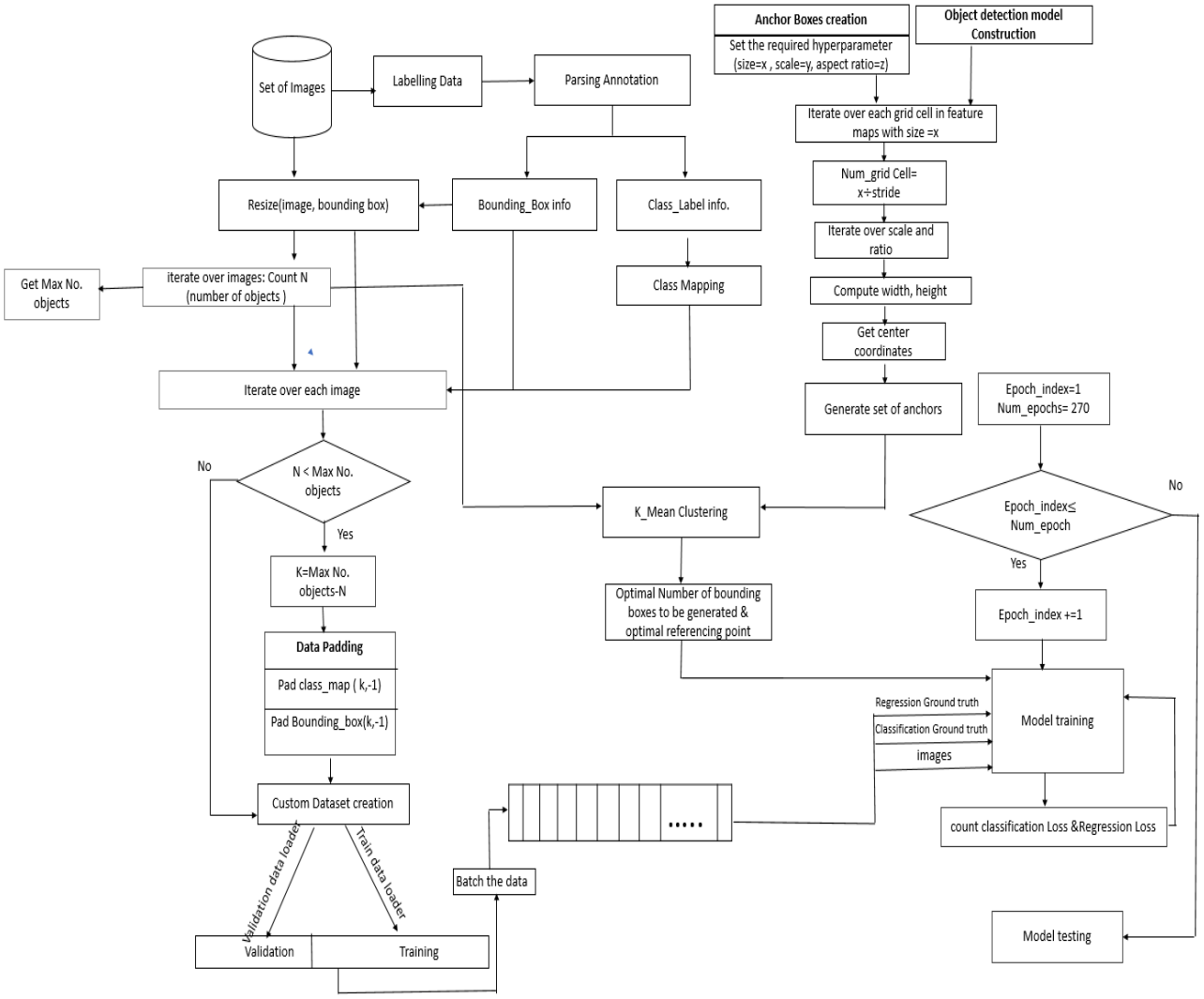

Figure 1. The steps followed during the creation of object detection for the AR model.

The model passed through multiple steps, initially; a dataset collected and manually labeled, by assigning bounding box coordinates and class labels for each object in the dataset, after that a custom class created for passing the collected data into the model. A set of reference points created for each image passed to the model using the Anchor Generator class. Finally, the model trained using a batch from the labeled data and a set of reference anchors. The following sections show a details discussion for each step in this study.

Data Collection and Preprocessing

Due to the absence of an Iraqi archeology dataset proper for training deep learning models. The first was step to collect a dataset. This data consists of 2188 color images for part of different ancient Iraqi ruins and some famous monuments. These are: Lion of Babylon, Malwiya Mosque, Mustansiriyah Madrasah, UR, Ishtar Gate, lamassu, Al_fanoos Alsahri, Ashaheed statue, Baghdad Statue, Hamorabi_statue, Liberation Square, saving culture statue, Monument Abu Jaffar Al Mansour, Monument of the Unknown Soldier, alrasafee_statue, kahramana_statue, Statue of Scheherazade and Shahryar, since its real-time images, there are some added classes such as building, car, dog, flag, motorcycle, person, and tree. **Fig. 2** shows a subset of the collected data.

2024, 21(12):3914-3926
https://doi.org/10.21123/bsj.2024.9199
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

Figure 2. A Subset of the collected images that highlights the combination of ancient ruins, famous monuments

The images pass through multiple preprocessing steps to prepare them for model training.

- Unify image names: Initially after collecting the images, the names of the dataset images unified to have the numbers starting from 0000 to 2187. Manual annotation performed on each image to assign a bounding box and give a label for each object inside each image followed by storing the annotation information in an XML file format; Fig. 3 shows an example of the annotation stage.
- Parse the annotation files: The XML file to each image is parsed to get the image ID, label, and bounding box coordinates for each object, the parsed annotations are stored later in CSV file format.
- Label Mapping Function: A mapping function established to assign each label a predefined index, commencing from index 1.
- Custom Dataset Class Creation: To enhance data access, manipulation, flexibility, and control during data loading, pre-processing, and transformations, a custom dataset class constructed. This class is designed to generate data 'on the fly during training, optimizing memory usage and enhancing performance as follows:
- Class Initialization Method: The custom dataset initialization method accepts several essential input data include: (the path to annotations and image files, mapping of class labels to their corresponding indices, the maximum number of objects per image, an optional data transformation function)
- Efficient Data Loading: The class efficiently loads the dataset by reading the CSV annotations file and precisely counting the number of objects in each image.
- Data Retrieval and Transformation: During data retrieval, the class retrieves the image based on the provided index and applies any necessary transformations. It also processes the object bounding box annotations and computes regression and classification targets for each object
- Handling varying size input: If the number of objects in an image falls below the maximum number of objects in the dataset, the class adeptly handles padding to ensure consistency.
- Structured Data Output: The resulting data presented in a structured dictionary format. This dictionary includes: (the image, regression, and classification targets, bounding boxes, and the count of objects)
- The class finally split data into training with 80% and 20% for validation. The split data is ready for batch to the models training function
- The custom dataset class offers a sophisticated and memory-efficient solution for data handling, enabling seamless integration of annotations, transformations, and data retrieval
- within the training process.

2024, 21(12):3914-3926
https://doi.org/10.21123/bsj.2024.9199
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

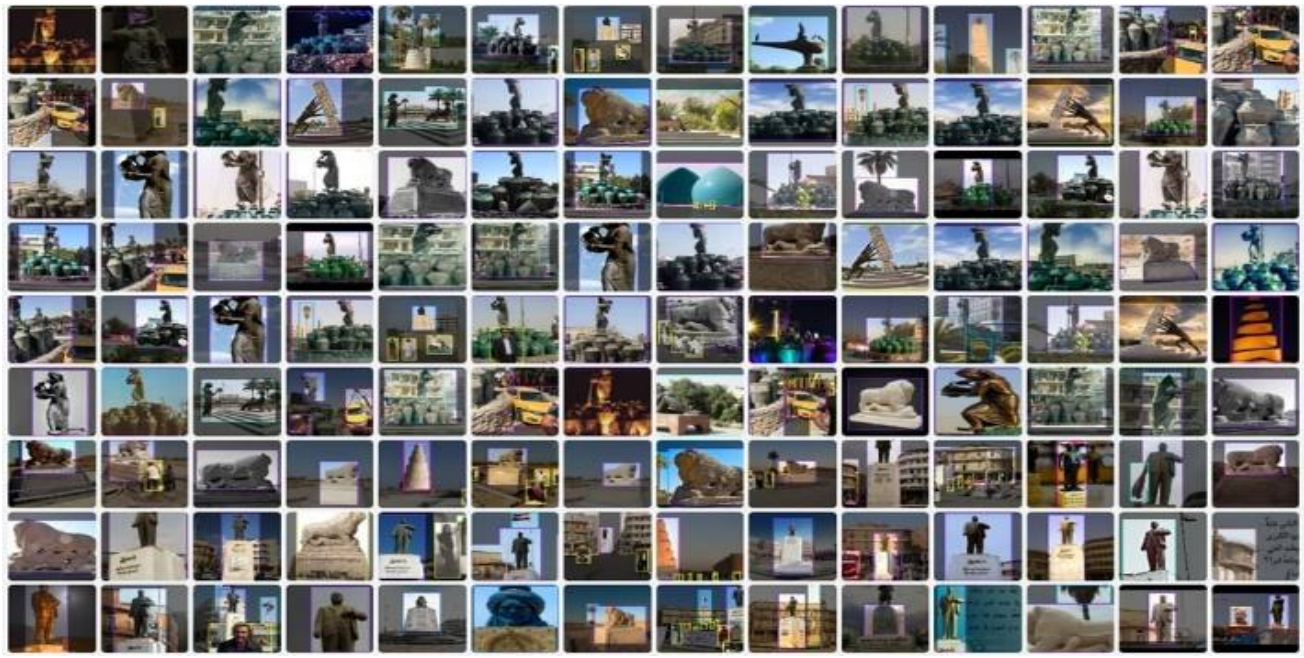

Figure 3. Samples of the dataset after annotations

Anchor Boxes Generation

A fundamental component in the object detection model is creating a dynamic set of reference bounding boxes at various scales and aspect ratios that the model will use to detect objects in an image. Once the anchor boxes generated, the next step is to predict the probability of an object being present in each of these anchor boxes and to predict the offsets required to transform these anchor boxes to fit around the object. 'Generate Anchor Boxes for Feature Map Algorithm' describes Anchor generation steps:

Generate Anchor Boxes for Feature Map Algorithm

Input: Set of scales, ratios, and feature map sizes.
Output: Set of anchor boxes

Begin
Step 1: anchors ← [] //initialize empty list.
Step 2: anchor_boxes_centers ← []
Step 3: For feat_size in feature_map_sizes
stride ← input_image_size / feat_map_size
For all y coordinates in feat_map_size
For x in range feat_map_size
center_x ←(0.5 + x) × stride
center_y←(0.5 + y) × stride
anchor_boxes_centers ← (center_x, center_y)
End For
End For
End For
Step 4: box_coords←( ) // initialize empty array
Step 5: Iterate over each scale, ratio
anchor_width, anchor_height←[anchor boxes]
width ← scale× $\sqrt{\text{ratio\_width}}$ × anchor_width
height ← scale × $\sqrt{\text{ratio\_height}}$ × anchor_height
box_coords← (width, height, (anchor_boxes_centers))
anchor ←box_coords
Step 6: Return anchors
End

Generating dynamic anchors would help improve the accuracy of the model. Based on the object sizes and aspect ratios number of anchors will be generated. The anchor generation passed through multiple steps:

- Initially, three sets were defined (anchor box sizes, aspect ratios, and scale sets) each set plays a specified role. The aspect ratio sets the shape of the anchor boxes. The scale set determines the size of the anchor boxes relative to the feature map size while the size used for choosing the feature map used to generate anchors
- Given the set definitions, anchor boxes for an image generated by sliding a set of fixed-size boxes, with different aspect ratios, across the image at various locations.
- At each location, the boxes scaled to feature map sizes, resulting in a set of anchor boxes.
- During the forward pass, the layer calculates the stride of the feature map, which determines the spacing between neighboring anchor boxes.

2024, 21(12):3914-3926
https://doi.org/10.21123/bsj.2024.9199
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

- Then compute the center coordinates of each anchor box by considering the feature map's grid structure.
- Furthermore, the class determines the width and height of each anchor box based on the specified aspect ratios and scales. These dimensions ensure flexibility in capturing objects of various shapes and sizes.
- The class takes care to clip the anchor boxes, ensuring they remain within the boundaries of the input image

Deep Learning Model Architecture Design

The architecture of the detection model used as an initial stage in the AR model shown in **Fig. 4.** The model is constructed using a carefully designed architecture comprising a pre-trained backbone network and additional layers for feature extraction and classification. The backbone network, specifically a VGG-16 network pre-trained on ImageNet, employed to extract feature maps from the input data batch. These feature maps have a size of 2048. Subsequently, a series of convolutional layers applied to these feature maps. The first convolutional layer convolves the feature map with 1024 filters, each of size (3, 3). This is followed by another convolutional layer, which further processes the feature map using 512 filters, each with a size of (3,3). In addition to these, six additional convolutional layers introduced, throughout this convolutional layer sequence, ReLU activation functions applied to the outputs, except for the eight^{h} convolutional layer, which followed by a flattened layer. This operation flattens the output feature maps from the convolutional layers into a suitable format for further processing. The flattened output then passed through fully connected layers, which serve two primary functions: regression and classification. The regression head is responsible for producing bounding box coordinates, whereas the classification head generates class probabilities. The model uses linear activations in the regression head; this function facilitates the output of precise numerical values for bounding box coordinates. Whereas softmax activation is used in the classification head to generate class probabilities, allowing for the categorical classification of objects.

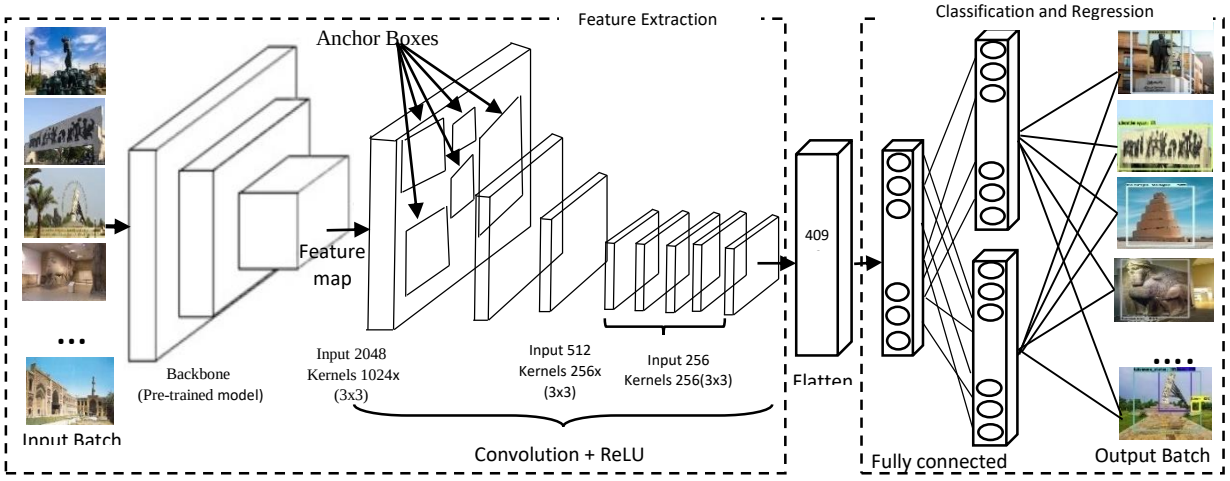

Figure 4. Layers of Object Detection with Dynamic Anchor Generation Model

Training Object Detection Model

The entire training process spans multiple epochs, where each epoch involves iterating through the entire training dataset. Each iteration updates the model's parameters to improve its predictions and reduce the loss. Over epochs, the model progressively refines its understanding of object characteristics and their positions. The train function iterates over batches of data provided from a custom dataset class. Each batch contains input images, regression targets (bounding box coordinates), classification targets (labels), boxes, and the number of objects**.** The feature map obtained from the backbone feds into eight additional convolutional layers**.** These additional layers further process the feature maps and capture many abstract representations created for the input image. The feature maps then flattened and passed through a fully connected layer (fc1) followed by a ReLU activation. The output of fc1 passed through fc2 and fc3 at the same time. Fc2 forms the regression head that used to predict the offsets for each anchor box.

2024, 21(12):3914-3926
https://doi.org/10.21123/bsj.2024.9199
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

Fc3 creates a classification head, which predicts the probabilities of each anchor box belonging to different object classes. The model returns a dictionary containing the regression and classification outputs**.** During the forward pass, the anchor box class generated the Number of anchors for a feature map as Eq. 1:

Number of anchors = number of aspect ratios × scales **1**

To get anchors equal to the number of objects in the image, the created anchors passed through k -mean clustering. The generated anchors and ground truth boxes (from the batch) concatenated to form the basis for calculating losses. Invalid classification and regression ground truth (padded) filtered out. The model defines a loss function that combines both the localization loss (measuring how accurate the model's bounding box predictions are) and the classification loss (measuring how accurate the model's class predictions are). The Intersection over Union (IoU) loss measures the dissimilarity between two bounding boxes by comparing their overlap[21]. Which is the ratio of the intersection area of the predicted bounding box and the ground truth-bounding box to the union area of the boxes. The IoU shows the mathematical representation IoU

$$\mathrm{IoU}_{\mathrm{loss}} = 1 - \frac{\min(x_{1_p},x_{1_g})\times\min(y_{1_p},y_{1_g})\times\min(x_{2_p},x_{2_g})\times\min(y_{2_p},y_{2_g})}{\max(x_{1_p},x_{1_g})\times\max(y_{1_p},y_{1_g})\times\max(x_{2_p},x_{2_g})\times\max(y_{2_p},y_{2_g})} \quad 2$$

$(x_{1_p}, y_{1_p}, x_{2_p}, y_{2_p})$ are the coordinates of the predicted bounding box, and $(x_{1_g}, y_{1_g}, x_{2_g}, y_{2_g})$ are the coordinates of the ground truth bounding box. The `min` and `max` functions are used to determine the minimum and maximum values, respectively, for each coordinate. To adapt the focal loss for multi-class classification[22], the predicted probability for each class was calculated, apply the focal loss equation independently for each class. The final loss computed by summing up the individual focal losses across all classes. The focal loss computed as Eq. 3:

$$\text{Focal loss} = -\alpha \times (1-p)^{\gamma} \times \log(p) \quad 3$$

Where α is a balancing factor that assigns different weights to different classes; which is used to address class imbalance issues by assigning higher weights to minority classes. (α) factor in the suggested model inversely set proportional to the class frequency. Compute the inverse of the class frequency and normalize it, to sum up to 1, (p) is predicted probability of the correct class for a given sample. γ is the focusing parameter that controls the rate at which the loss decreases as the predicted probability increases. A higher value of (γ) puts more emphasis on correcting misclassified samples. In this study, γ was set to 0.5. The model trained with 80% of the collected dataset and 20% set to validate the model. The network has 52,405,024 trainable parameters. The model uses a stochastic gradient descent (SGD) algorithm with the Adam update. The initial learning rate was set to 0.001.

IoU Filtering

Once anchor boxes generated for an image, The IoU (Intersection over Union) of each object in the image computed. The IoU is a measure of the overlap between the objects ground truth and the anchor box. It defines as the area of intersection between the two boxes divided by the area of union. Filtering out redundant and overlapping bounding box predictions based on their classification scores and IoU overlap. Each anchor box could assign to one of three classes: positive, negative, or ignored.

- A positive anchor box is assigned to an object if its IoU with the object is greater than or equal to a threshold value (0.5).
- A negative anchor box is assigned to the background class if its IoU with all objects in the image is less than a threshold value (usually 0.4).

The filtering process starts by creating an ‘anchor mask’ which is a binary mask that indicates which anchor boxes are positive, negative, or ignored. Specifically, the mask has the same shape as the anchor boxes and has a value of 1.0 for positive anchor boxes, -1.0 for negative anchor boxes, and 0.0 for ignored anchor boxes. The remaining non-overlapping and highly confident bounding box predictions, after filtering, considered the final detected objects in the image.

2024, 21(12):3914-3926
https://doi.org/10.21123/bsj.2024.9199
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

Results

The performance of the proposed model evaluated by monitoring the loss of boxes and labels throughout the training process. As shown in **Fig. 5**

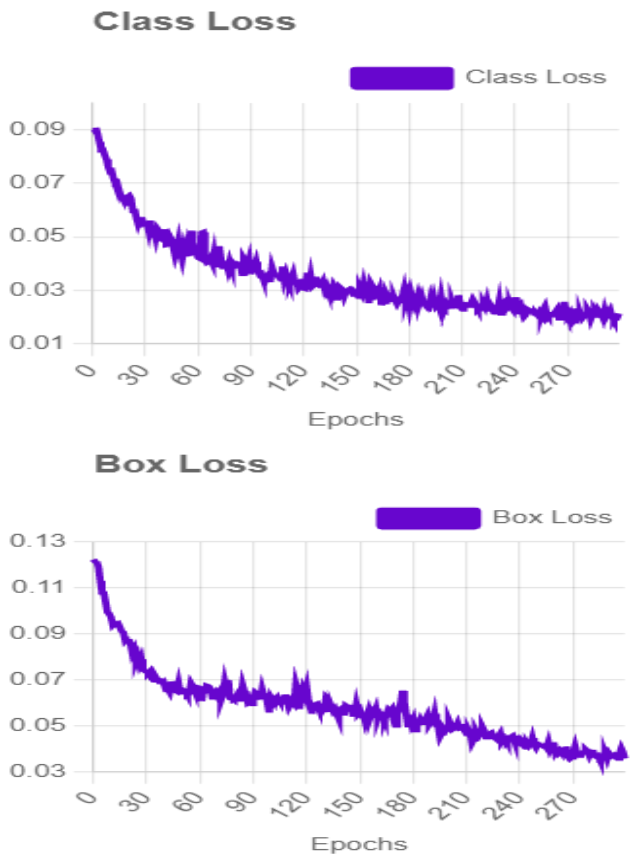

Figure 5. The left figure analyzing the progressive reduction of class labels loss; the right figure shows a decreasing box loss over 278 training epoch

The loss of box decreased steadily over the 278 training epochs to get the minimum value of 0.008 at epoch 275, indicating that the model was effectively learning to localize objects within the bounding boxes. Similarly, the loss of class labels demonstrated a consistent downward trend, indicating the model's ability to accurately classify objects. The minimum label loss value attained was 0.02 at epoch number 278. To assess the overall object detection performance, mean Average Precision mAP calculated as Average Precision AP_i values for each object class. The values of mAP [22]were computed at regular intervals during the training process Eq. 4

$$mAP = \frac{1}{n}\sum_{i=1}^{n} AP_i \qquad 4$$

Where (n) denotes the total number of object classes, the values are at regular intervals during training. The results achieved by the model demonstrate its effectiveness in addressing the object detection task to get an accurate augmented reality result. **Fig. 6** reveal the progressive improvement in object detection performance. The model's highest achieved mAP value of 84% at epoch 255 demonstrates its peak performance.

Fig. 7 presents testing results on a random subset of the dataset. Finally, Table 1 provides a concise summary of the key findings and results from various research studies related to object detection in archaeological image analysis. Each entry in the table includes information about the authors, publication year, dataset used, network architecture, and performance metrics.

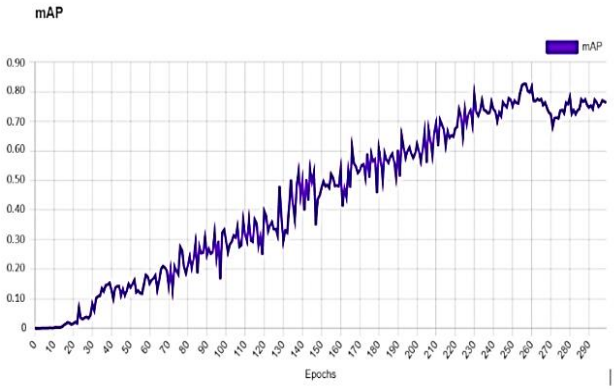

Figure 6. Tracking Object Detection Performance: Analyzing mAP Progression and Peak Performance over 278 Training Epochs

2024, 21(12):3914-3926
https://doi.org/10.21123/bsj.2024.9199
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

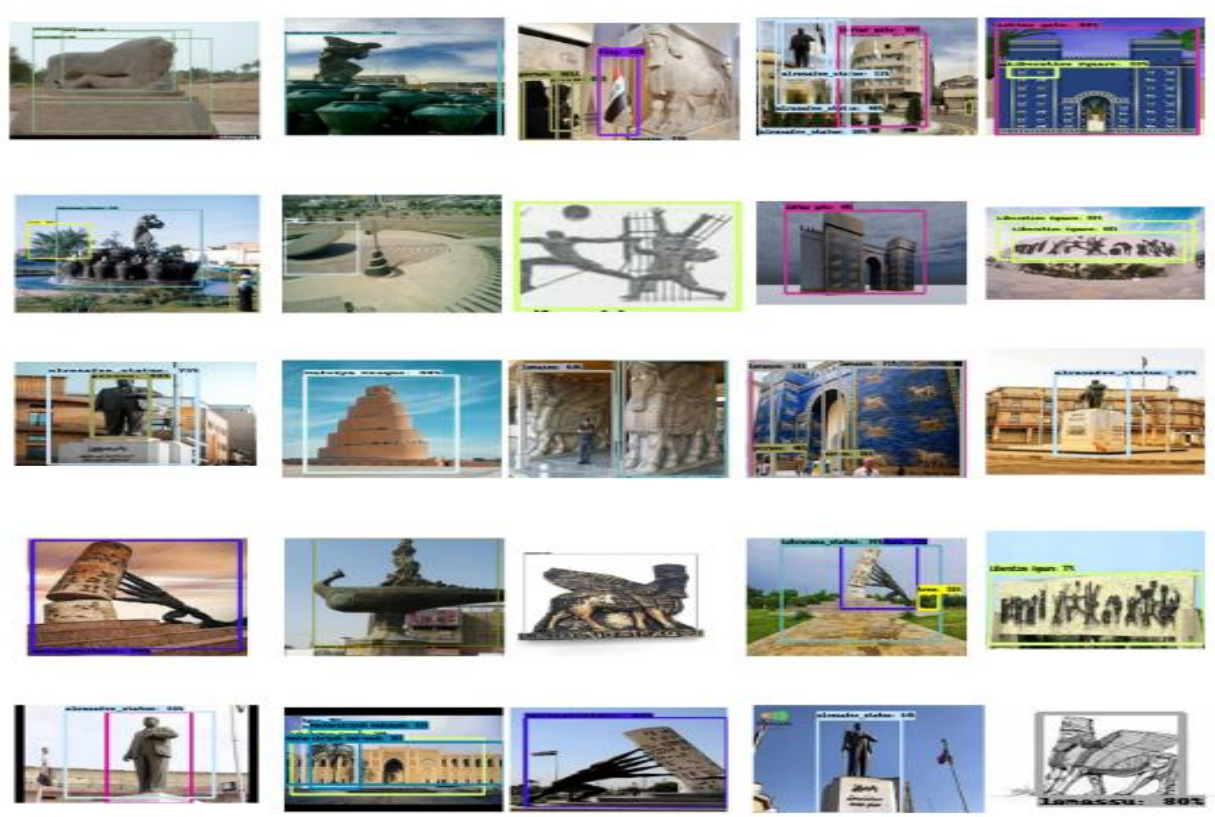

Figure7. Evaluating Performance on Random Subset of the Dataset: Insights into Testing Results

Table 1. Comparison of Object Detection Results in Archaeological Image Analysis

Author/Year	Dataset	Architecture	Results
Soroush et al. 17 2020	Cold War-era CORONA Satellite Imagery (100 labeled shaft)	The network architecture consists of 22 convolutional, 5 max pooling, and 5 up-sampling layers.	precision = 0.654, recall = 0.764, F1 score = 0.705
Lyer &Franklin 18 2022	ancient artifacts 343 of cultures worldwide with 55000 images	Two CNN models, InceptionResNetV2 and VGG-19	Accuracy for Inception ResNetV2 (65%) and performance for VGG-19 (12.95%).
Verschoof-Van der Vaart &Lambers 19 2019	LiDAR data	An adapted Faster R-CNN model with transfer learning used for multi-class object detection. The data results from the detection were converted into geographical data.	values for barrows and Celtic fields are respectively Recall (0.62–0.81) Precision (0.36–0.90) F1-scores lie between 0.49 and 0.79
Guyot et al. 20 2021	LiDAR	The method uses transfer learning, an adapted Faster R-CNN model that detects and categorizes two archaeological object types	The mAP@IoU.5 scores varied, ranging from 0.29 (experiment Ftrain10) to 0.77 (experiment Atrain80)
Suggested Method	Custom-created dataset for Iraqi archeology 2188 images	The model leverage transfer learning (VGG-16); additionally, 8 convolution layers followed by flatten and 3 fully connected	mAP 0.84 regression loss 0.008 classification loss 0.02

2024, 21(12):3914-3926
https://doi.org/10.21123/bsj.2024.9199
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

Table 1 provides a comparative view of different studies in the field of archaeological image analysis, showcasing their datasets, model architectures, and performance metrics. The proposed work, achieved a high mAP, indicating strong performance in object detection within Iraqi archaeological imagery. It demonstrates the effectiveness of the suggested model architecture.

Discussions

In this section, the interpretation of the findings and their broader implications were explored. This study in the realm of computer vision applied to Iraqi archaeology has yielded valuable insights. The model's ability to detect archaeological objects in diverse and complex imagery represents a significant advancement in the field. Our findings indicate that the custom dataset, collected to Iraqi archaeology, not only enhances the accuracy of object detection but also enables the preservation and digitization of invaluable cultural heritage. Comparing the suggested work with previous research in computer vision and object detection, it is evident that adapting the methodology to domain-specific datasets significantly improves results. This study makes several notable contributions to the field. Firstly, the development of the anchor box class. Without reference anchor boxes, the model would have to predict the size and location of every possible object in the image, which would be computationally expensive and would require a large amount of training data. The anchor boxes allow the model to handle objects of different sizes and aspect ratios, which is important in many real-world scenarios where objects can vary greatly in size and shape. Additionally Due to the absence of an Iraqi archeology dataset proper for training deep learning models, a dataset of 2188 color images was collected, spanning ancient Iraqi ruins, celebrated monuments, and real-time scenes combined with various objects. This dataset subjected to manual annotation, wherein bounding boxes and labels assigned to objects in each image. To enhance data access, manipulation, flexibility, and control during data loading, pre-processing, and transformations, a custom dataset class accurately constructed. This class designed to generate data 'on-the-fly' during training, optimizing memory usage and enhancing performance. After that in this research, a deep learning model designed and trained. The designed model testing results for both classification and localization were very promising. It is essential to acknowledge the limitations of our study. Despite the promising results, challenges persist in accurately detecting objects in highly fragmented or deteriorated archaeological remains. Future research efforts could focus on improving the model's robustness in these scenarios.Moreover, while our custom dataset is a substantial step forward, it is not exhaustive. Expanding the dataset to encompass a wider array of archaeological contexts and objects would further enhance the model's applicability. Our research carries practical implications for archaeologists and cultural heritage preservationists. The object detection model can streamline the process of identifying and cataloging artifacts in archaeological sites, potentially accelerating archaeological research and conservation efforts.

Conclusion

This scientific attempt represents a significant milestone in the realm of computer vision and archaeological research, with a particular focus on the unique context of Iraqi archaeology. This study encompassed several crucial phases, including the collection and manual annotation of a custom dataset, the development of an anchor box class to address data variability, and the training of an object detection model. The creation of a custom dataset tailored to Iraqi archaeology allowed us to lay the foundation for advanced computer vision applications in this specific domain. Manual annotation, including the assignment of class labels and precise bounding box coordinates, ensured the high quality of the training data, which is pivotal for the model's success. To tackle the challenge of varying input shapes, a specialized class was introduced, facilitating the conversion of labeled data into a structured CSV format. This approach not only improved data management but also enabled dynamic object counting and padding with -1 to standardize data shape across all images. The model training process, involving an 80% training and 20% testing data split, and an extensive 278-epoch training, resulted in a good performance. The obtained results were highly promising, demonstrating the model's effectiveness in detecting archaeological objects within Iraqi imagery. This

2024, 21(12):3914-3926
https://doi.org/10.21123/bsj.2024.9199
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

research exemplifies the integration of domain-specific data collection and advanced computer vision techniques. It signifies a substantial contribution to the preservation and study of Iraq's archaeological heritage. The suggested object detection model was the first step to create an augmented reality model the rest of the step will implemented later. The AR heavily relies on camera input to overlay virtual objects onto the real-world environment. Integrate a camera module into the model to capture live video or images as input. In addition to object detection, the model requires pose estimation algorithms to estimate the 6-DoF (degrees of freedom) camera pose. This enables accurate alignment of virtual objects with the real-world scene. Finally, implement object-tracking algorithms to track the detected objects across frames. This ensures consistent placement and interaction with virtual objects, even as the camera moves.

Acknowledgment

We would like to express our appreciation to Dr. Ahmed Athab, Senior Technical Officer at Terra Motion Ltd, for his meticulous proofreading contributions to this research.

Authors' Declaration

- Conflicts of Interest: None.
- We hereby confirm that all the Figures and Tables in the manuscript are ours. Furthermore, any Figures and images, that are not ours, have been included with the necessary permission for re-publication, which is attached to the manuscript.
- Ethical Clearance: The project was approved by the local ethical committee at University of Technology.
- No animal studies are present in the manuscript.
- No human studies are present in the manuscript.
- No potentially identified images or data are present in the manuscript.

Authors' Contribution Statement

S.D.A contributed to the acquisition of data, the design and drafting of the MS. A.A. K contributed analysis of the results, revision and proofreading

References

1. Ghasemi Y, Jeong H, Choi SH, Park K-B, Lee JYJCiI. Deep learning-based object detection in augmented reality: A systematic review. Comput Ind. 2022; 139: 103661. http://dx.doi.org/10.1016/j.compind.2022.103661
2. Khan MA, Israr SS, Almogren A, Din IU, Almogren A, Rodrigues JJ. Using augmented reality and deep learning to enhance Taxila Museum experience. J Real-Time Image Process 18 (2): 321–32. https://doi.org/10.1007/s11554-020-01038-y
3. Sweeney SK, Newbill P, Ogle T, Terry KJT. Using augmented reality and virtual environments in historic places to scaffold historical empathy.TechTrends 2018; 62: 114-8. https://doi.org/10.1007/s11528-017-0234-9
4. Blanco-Fernández Y, López-Nores M, Pazos-Arias JJ, Gil-Solla A, Ramos-Cabrer M, García-Duque J. REENACT: A step forward in immersive learning about Human History by augmented reality, role playing and social networking. Expert Syst Appl. 2014; 41(10): 4811-28. https://doi.org/10.1016/j.eswa.2014.02.018
5. Oleksy T, Wnuk A. Augmented places: An impact of embodied historical experience on attitudes towards places. Comput. Hum. Behav. 2016 Apr 1;57:11-6. https://doi.org/10.1016/j.chb.2015.12.014
6. Çakiroğlu Ü, Aydin M, Köroğlu Y, Ayvaz Kina MJILE. Looking past seeing present: teaching historical empathy skills via augmented realit.nteract. Interact Learn Environ. 2023:1-13. https://doi.org/10.1080/10494820.2023.2174142
7. Carmigniani J, Furht B, Anisetti M, Ceravolo P, Damiani E, Ivkovic MJMt, et al. Augmented reality technologies, systems and applications. Multimed. Tools Appl. 2011; 51: 341-77. https://doi.org/10.1007/s11042-010-0660-6
8. Fenais AS, Ariaratnam ST, Ayer SK, Smilovsky NJJoITiC. A review of augmented reality applied to underground construction. J Inf echnol Constr. 2020; 25: 308-24. https://doi.org/10.36680/j.itcon.2020.018
9. Ponnusamy V, Natarajan, Solutions, Applications. Precision agriculture using advanced technology of IoT, unmanned aerial vehicle, augmented reality,

2024, 21(12):3914-3926
https://doi.org/10.21123/bsj.2024.9199
P-ISSN: 2078-8665 - E-ISSN: 2411-7986
Baghdad Science Journal

and machine learning. IIOT. 2021: 207-29. https://doi.org/10.1007/978-3-030-52624-5_14
10. Lalonde J-F, editor Deep learning for augmented reality. 2018 17th Workshop on Information Optics (WIO); 2018: IEEE. https://doi.org/10.1109/WIO.2018.8643463
11. Park K-B, Kim M, Choi SH, Lee JYJR, Manufacturing C-I. Deep learning-based smart task assistance in wearable augmented reality. Robot Comput Integr Manuf. 2020; 63: 101887. https://doi.org/10.1016/j.rcim.2019.101887
12. Alsaedi EM, Farhan Ak. Retrieving Encrypted Images Using Convolution Neural Network and Fully Homomorphic Encryption. Baghdad Sci J. 2023; 20(1): 0206. https://dx.doi.org/10.21123/bsj.2022.6550
13. He Y, Ren J, Yu G, Cai YJIToWC. Optimizing the learning performance in mobile augmented reality systems with CNN. IEEE Trans Wirel Commun. 2020; 19(8): 5333-44. https://doi.org/10.1109/TWC.2020.2992329
14. Ababsa F-e, Mallem M, editors. Robust camera pose estimation using 2d fiducials tracking for real-time augmented reality systems. Proceedings of the 2004 ACM SIGGRAPH international conference on Virtual Reality continuum and its applications in industry; 2004. https://doi.org/10.1145/1044588.1044682
15. Abdullah TH, Alizadeh F, Abdullah BH. COVID-19 Diagnosis System using SimpNet Deep Model. Baghdad Sci J. 2022; 19(5): 1078. https://doi.org/10.21123/bsj.2022.6074
16. Sprute D, Viertel P, Tönnies K, König M, editors. Learning virtual borders through semantic scene understanding and augmented reality. IEEE Int Conf Intell Robots Sys.; 2019: https://doi.org/10.1109/IROS40897.2019.8967576
17. Soroush M, Mehrtash A, Khazraee E, Ur JAJRS. Deep learning in archaeological remote sensing: Automated qanat detection in the Kurdistan region of Iraq. Remote Sens. 2020; 12(3): 500. https://doi.org/10.3390/rs12030500
18. Yer A, Franklin M. AI-Powered Archaeology: Determining the Origin Culture of Various Ancient Artifacts Using Machine Learning. JSR. 2022; 11(1). https://doi.org/10.47611/jsrhs.v11i1.2465
19. Verschoof-Van der Vaart WB, Lambers K. Learning to look at LiDAR: The use of R-CNN in the automated detection of archaeological objects in LiDAR data from the Netherlands. J Comput Appl Archaeol. 2019; 2(1). https://doi.org/10.5334/jcaa.32
20. . Combined detection and segmentation of archeological structures from LiDAR data using a deep learning approach. Comput Appl Archaeol. 2021; 4(1): 1. https://dx.doi.org/10.5334/jcaa.64
21. Rahman MA, Wang Y, editors. Optimizing intersection-over-union in deep neural networks for image segmentation. ISVC; 2016: Springer. https://doi.org/10.1007/978-3-319-50835-1_22
22. Lin T-Y, Goyal P, Girshick R, He K, Dollár P, editors. Focal loss for dense object detection. Proc IEEE Int Conf Comput Vis. ; 2017. https://doi.org/10.48550/arXiv.1708.02002

نموذج الكشف عن الأشياء بالاعتماد على الواقع المعزز للآثار العراقي

سها ظاهرعذاب ،عبد الأمير عبد الله كريم

قسم علوم الحاسوب، الجامعة التكنولوجية، بغداد، العراق.

الخلاصة

تقف الثقافة العراقية، بتاريخها الغني، شاهدا على عظمة الحضارة الإنسانية. ومع ذلك، أصبح الحفاظ على هذا التراث الثقافي الغني وتقديمه مصدر قلق ملح في عصر يتميز بالتقدم التكنولوجي. الواقع المعزز هو المتصدر في عالم التكنولوجيا، يوفر أداة قوية في هذا المجال. الهدف من هذا البحث هو الاستفادة من تقنية الواقع المعزز كوسيلة لضمان استمرار الحفاظ على التراث الثقافي العراقي وعرضه بشكل ديناميكي. تستكشف هذه الدراسة قدرات الشبكات العصبية الالتفافية كأساس لتطوير الواقع المعزز. نحن نحقق في إمكاناتها في اكتشاف الأشياء، وهي خطوة أولية أساسية في صياغة أنظمة الواقع المعزز. يستخدم النموذج المقترح شبكة أساسية مدربة مسبقا لاستخراج الميزات المكانية المعقدة من الصور المدخلة، ويتم إدخال طبقات تلافيفيه إضافية ومتصلة بالكامل لزيادة تحسين هذه الميزات. تم اقتراح فئة مخصصة جديدة تسمى "مربعات الارتساء" ، تنشئ ديناميكيا مربعات ربط محددة مسبقا لكل خلية خريطة معالم. نظرا لعدم وجود مجموعة بيانات للآثار العراقية مناسبة لتدريب نماذج التعلم العميق، قمنا بجمع مجموعة بيانات من 2188 صورة ملونة، تغطي الآثار العراقية القديمة، والآثار الشهيرة ، وبعض المشاهد في الوقت الفعلي جنبا إلى جنب مع المعالم الاثرية. تخضع مجموعة البيانات هذه للوسم التوضيحي اليدوي، حيث يتم تعيين المربعات المحيطة والتسميات للكائنات في كل صورة. تؤكد نتائج تحليل الانحدار على كفاءة النموذج في تقدير إحداثيات المربع المحيط بالكائن بدقة جيدة، مما يسهل توطين الكائن بدقة وتحديد الموقع. توضح ايضا نتائج التصنيف قدرة النموذج على تعيين تسميات الفئة بثقة للكائنات المكتشفة

الكلمات المفتاحية: صناديق الارساء، التصنيف، الرؤية الحاسوبية، تحديد المكان، كشف الأجسام.