Comparison of Relaion-DETR and YOLO 11 Object Detection Methods

Yanxi Wang[a]
School of Computer Science and Engineering, Tianjin University of Technology, Tianjin, China

Keywords: Object Detection, Performance Comparison, Deep Learning, Relation-DETR, YOLO 11.

Abstract: This study compares the performance of two object detection methods based on deep learning, Relation-DETR and YOLO 11, using COCO datasets, in image detection tasks of animals, landscapes, people, and other categories. The experimental results show that YOLO 11 is significantly better than Relation-DETR in successful detection rate, especially in animals, people, and other categories to achieve a successful detection rate of more than 90%. However, in landscape detection tasks, the performance of both methods is not ideal, indicating the limitation of object recognition tasks in specific scenes. While YOLO 11 has the advantage in terms of detection accuracy, Relation-DETR stands out in terms of operability thanks to its user-friendly visual interface design. In order to improve the recognition accuracy, this paper proposes some improvement measures, such as thinning the data set and introducing domain-specific prior knowledge, to provide solutions. Through experiments and comparative analysis, this study provides a valuable reference for selecting appropriate object detection methods and their optimization.

1 INTRODUCTION

In today's life, object detection technology has been widely used in all aspects of life, and this technology has become a research hotspot in recent years (Wang et al., 2021), from the camera recognition of intelligent driving systems to the recognition and detection of cell level based on deep learning, all of which are inseparable from the support of object detection (Wang, 2023; Xia, 2020). However, many large-scale object detection tasks are characterized by large-scale and high engineering costs. For example, in order to ensure the normal operation of traffic, it is necessary to use object detection technology to provide computer vision assistance to the police (Wang, Chen, & Sun, 2025). Among many current methods, methods based on deep learning are widely used, and it is necessary to reduce the quality of vehicle images and improve the generalization ability of recognition to meet the needs of different scenes (Han, 2024). This requires the SkyEye system to acquire images from pictures or videos of vehicles in a short time, and to locate and recognize license plate characters according to specific algorithms for text detection (Zhang et al., 2023). However, such large-scale collection and analysis efforts are clearly not suitable for the analysis of small objects. For example: The analysis of medical images is not only the recognition of images in the 2D plane, but also the recognition of medical images such as 3D and 4D magnetic resonance images (Chen et al., 2021), which requires computer-aided detection (CAD) system to accurately identify lesions, and to have certain consideration for different organ appearances of different patients, so as to assist medical staff to measure the relevant structure and function of the current case in a short time and make a judgment on whether the disease is present (Tao et al., 2018). The work of analysing whether a single organ has a lesion is different from automobile detection. In automobile detection, there is usually a set of excellent recognition algorithms, which are undergoing continuous improvement and improvement (Zhuang, 2022), while the detection of medical images is refined to the point that each organ has its own unique algorithm, and the recognition rate of specific lesions must reach more than 90% (He, 2019).

Therefore, in this paper, we will use the relation-DETR object detection method and the classical YOLO11 object detection method to analyse the data

[a] https://orcid.org/0009-0006-8230-6012

Wang, Y.
Comparison of Relaion-DETR and YOLO 11 Object Detection Methods.
DOI: 10.5220/0013698600004670
Paper published under CC license (CC BY-NC-ND 4.0)
In *Proceedings of the 2nd International Conference on Data Science and Engineering (ICDSE 2025)*, pages 420-429
ISBN: 978-989-758-765-8
Proceedings Copyright © 2025 by SCITEPRESS – Science and Technology Publications, Lda.

respectively, and compare them. In the process of analysis, the experimental pictures are divided into four categories: animals, people, landscapes and others, and the detection results are divided into three categories: successful and accurate, successful but inaccurate, and unsuccessful. Based on these detection results and the classification of the detected objects, this paper compares the two methods and proposes some suggestions for improvement.

2 DATASET MODEL

2.1 Data Set

The COCO dataset was used as the basic data source in this experiment. Eighty images were randomly selected from the dataset and classified according to established classification criteria. The samples were divided equally into four categories: 20 animals, 20 people, 20 landscapes and 20 other objects. This balanced sample distribution provides a reliable database for subsequent experiments.

2.2 Model and Method

In this study, the Relation-DETR object detection method is used first, and 120 images are detected and analysed by combining them with the pre-trained model. The specific process is as follows. The first step is to load the pre-trained model and weight code for loading the relation-DETR model and its weights. This function defines various parameters of the model, such as embedding dimensions, number of categories, number of queries, etc. The second step is to create a graphical user interface (GUI) that facilitates interaction with the user. The code defines a “create_gui” function that creates a simple GUI that allows the user to select images and run object detection. The GUI contains the following components: Picture Display tab, Select Picture button, Picture Path input box, Run Reasoning button, and Results Display tab. The third step is to run reasoning, this step is the core function of the code, first the user clicks the "select picture" button, which will pop up a file dialog box, let the user select a picture, select picture, the picture will be displayed in the GUI picture display label, and update the picture path input box content. After the user clicks the "Run Reasoning" button, the program reads the image path in the image path input box and loads the image. After the image is pre-processed (scaled, and converted to Tensor), it is used as the input of the model for reasoning. Model outputs include bounding boxes, category labels, and confidence scores. A threshold according to the confidence score to screen out the detection results with high confidence. Finally, draw the detection result, drawing the filtered bounding box and category label on the original image. Zoom the drawn picture and display it in the picture display tab of GUI, and update the result display tab to prompt the user of the detected object.

The second is the YOLO 11 object detection method. This study still uses the method of pre-trained model to detect and reason 120 images. The first step in the code for this model is to import the model, which is used to load the model and make predictions. The second step is to import the image to be detected by setting the absolute path defined by the user and setting the size of the input image to 640×640 pixels. After processing by GPU, the detection result is saved to the image file. The third step is to use OpenCV's “show” function to display an image of the detection result and wait for the user to press Start Detection and generate a prediction result. The final step is to close all OpenCV creation windows and save the results

3 EXPERIMENTAL RESULT

3.1 The Results of the Relation-DETR based Object Detection Method

3.1.1 The Animal

The number of successful and accurate detections is 12, as shown in Figure 1 below:

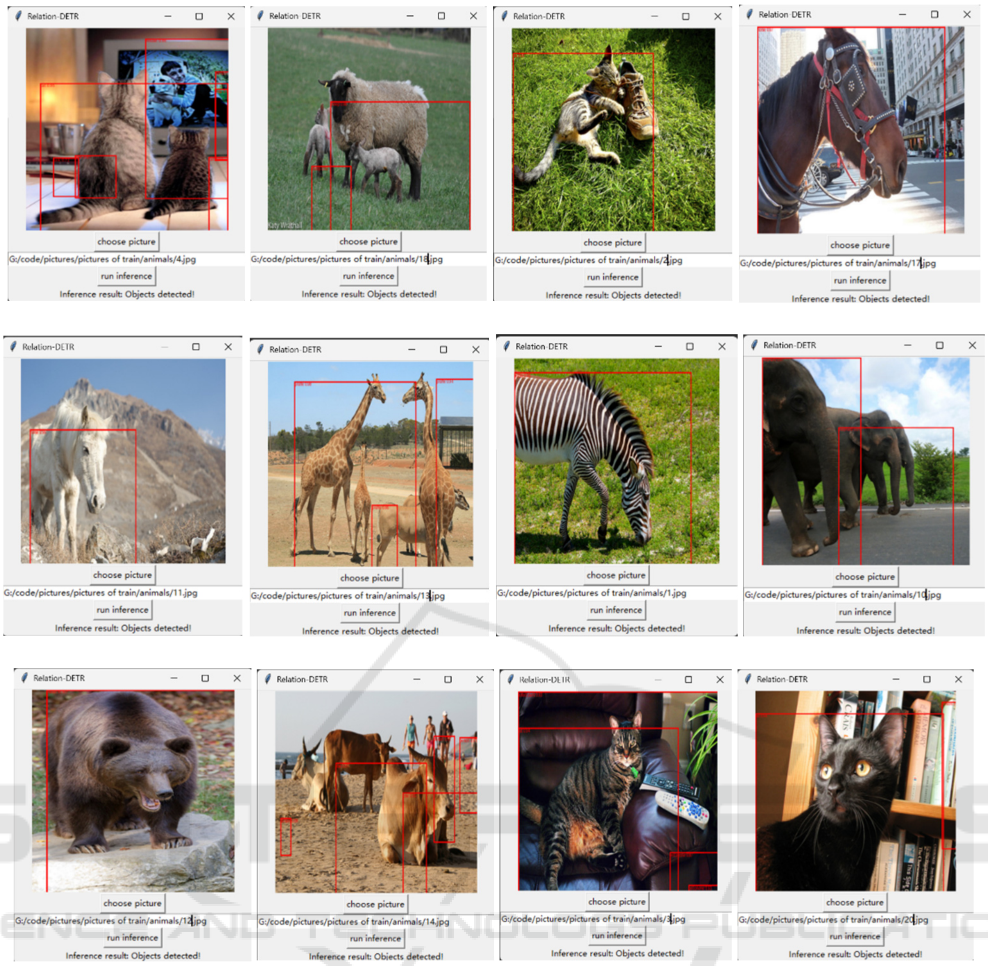
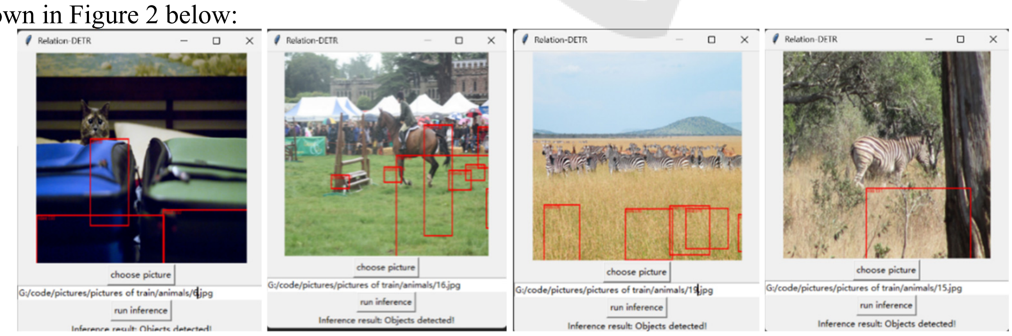


Figure 1: Successful and accurate detection of animals. (Picture credit: Original)

The number of successful inaccurate detections is 8, as shown in Figure 2 below:



SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

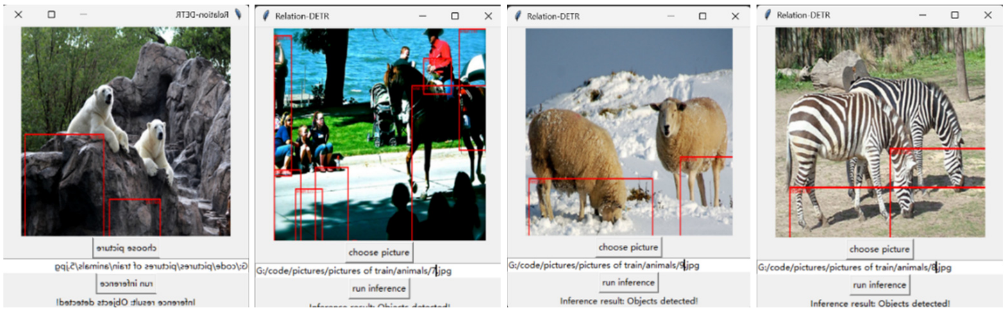


Figure 2: Successful but inaccurate detection of animals. (Picture credit: Original)

3.1.2 The Landscape

The number of successful and accurate detections is 3, as shown in Figure 3 below:

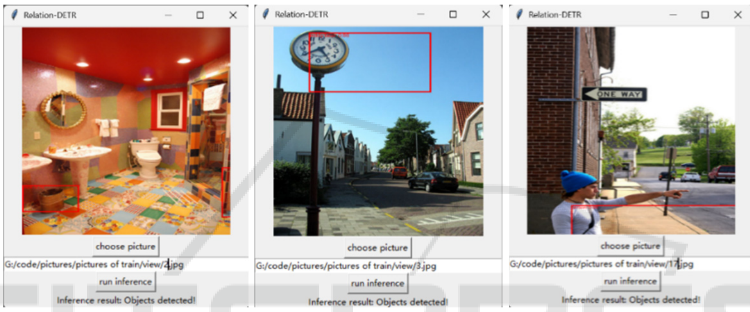
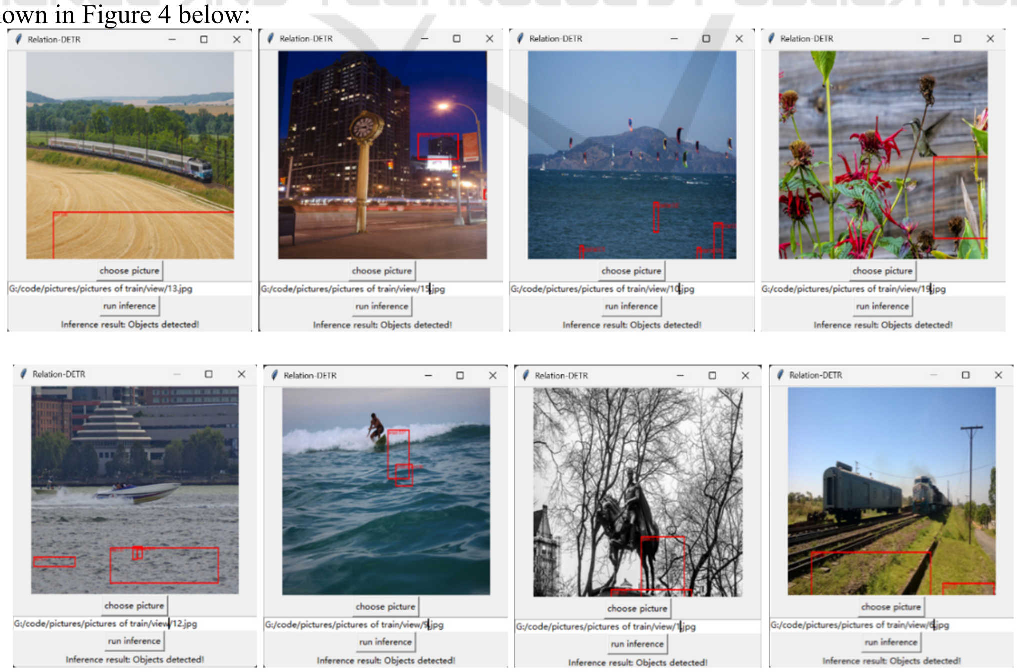


Figure 3: Successful and accurate detection of landscape images. (Picture credit: Original)

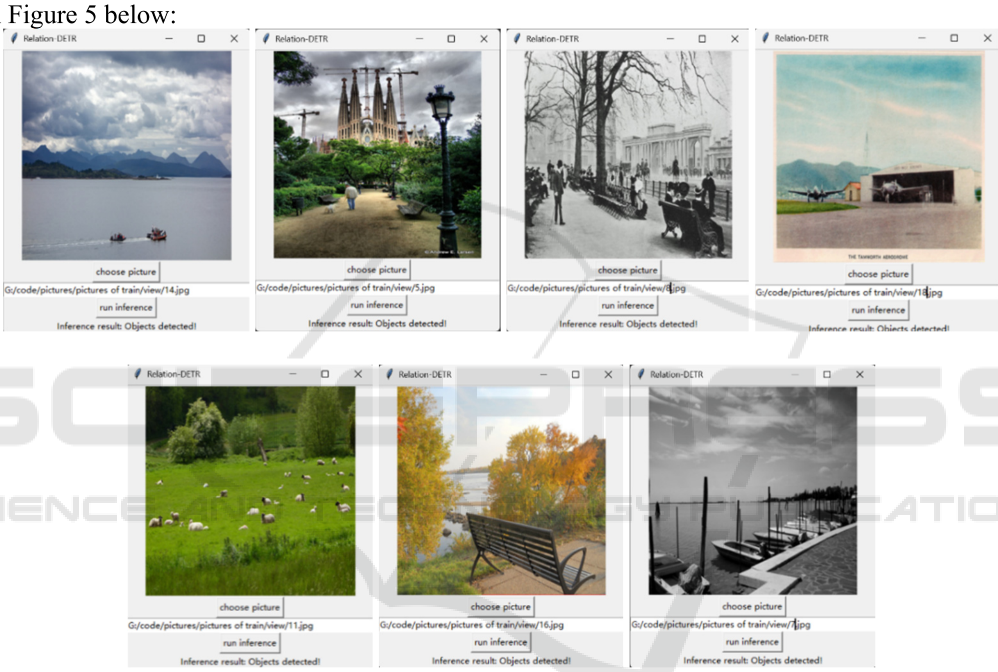
The number of successful inaccurate detection is 10, as shown in Figure 4 below:



SCITEPRESS
SCIENCE AND TECHNOLOGY PUBLICATIONS

Figure 4: Scenery successful but inaccurate detection picture. (Picture credit: Original)

The number of unsuccessful detections is 7, as shown in Figure 5 below:

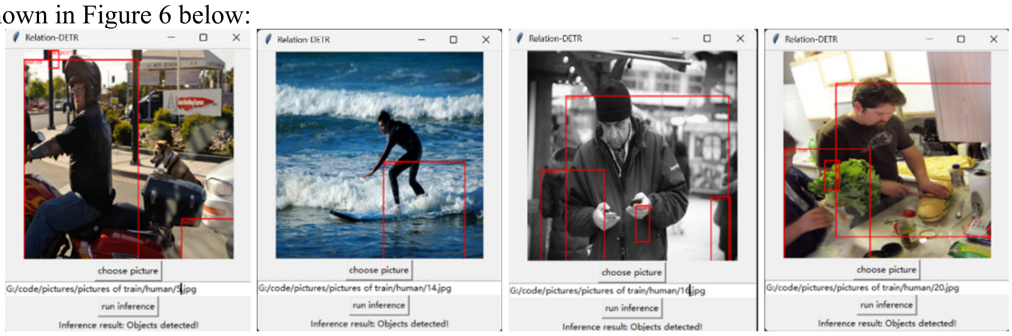


SCITEPRESS
SCIENCE AND TECHNOLOGY PUBLICATIONS

Figure 5: Unsuccessful detection of landscape images. (Picture credit: Original)

3.1.3 Character Category

The number of successful and accurate detections is 17, as shown in Figure 6 below:



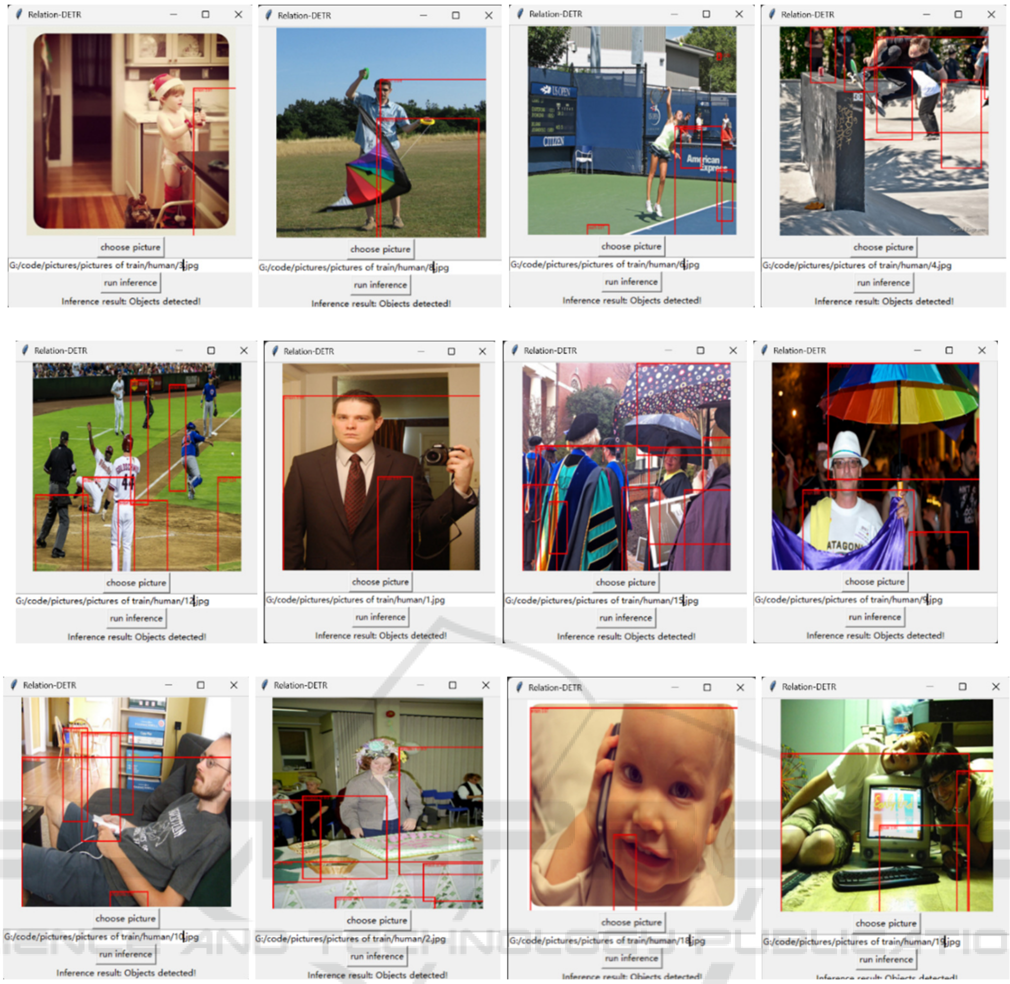


Figure 6: Successful and accurate detection of people. (Picture credit: Original)

The number of successful inaccurate detections is 3, as shown in Figure 7:

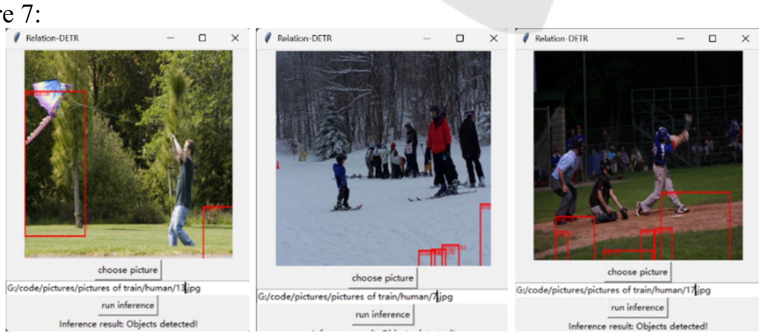


Figure 7: Successful but inaccurate detection of people. (Picture credit: Original)

The number of unsuccessful tests is 0.

3.1.4 The Other Classes

The number of successful and accurate detection is 14, as shown in Figure 8 below:

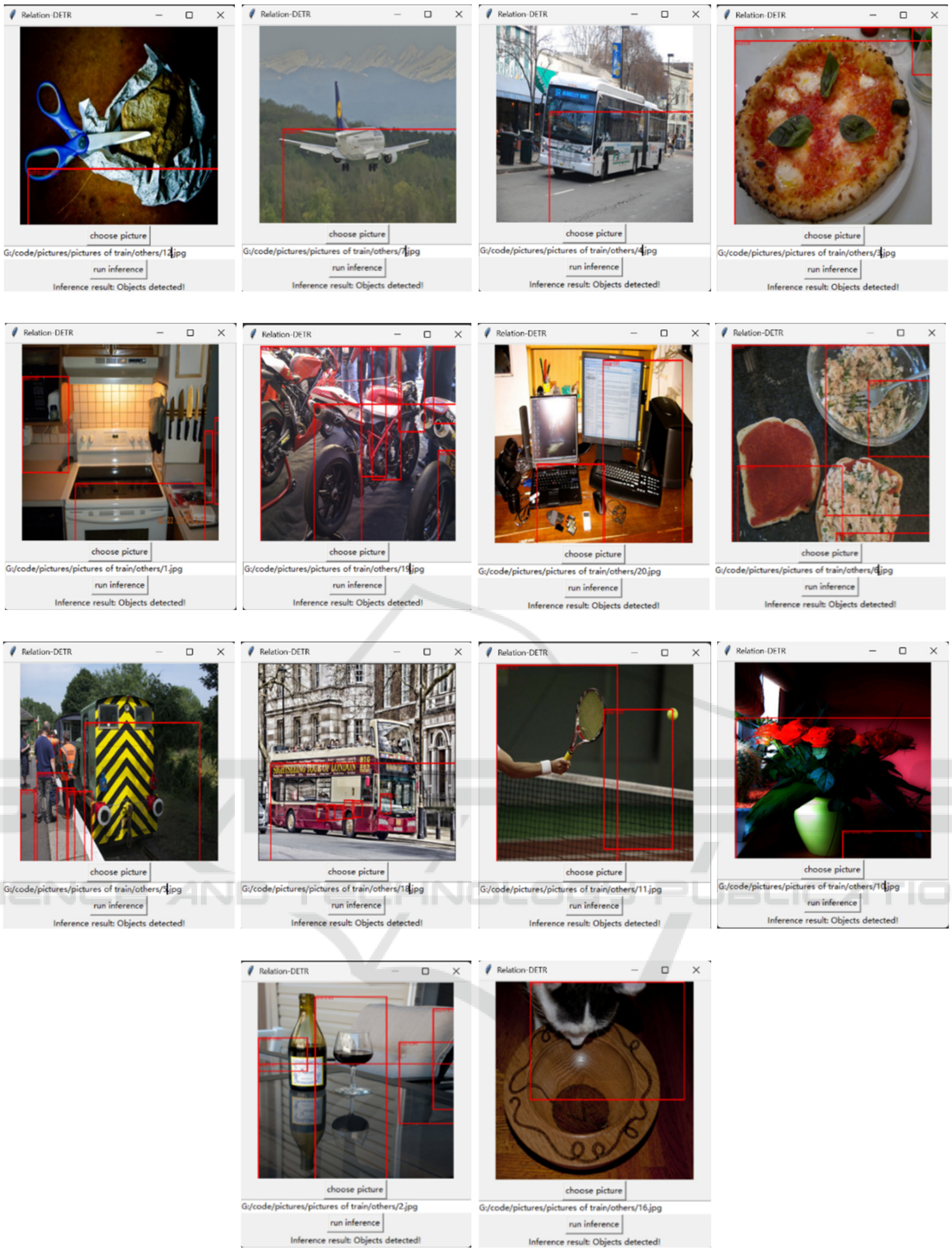


Figure 8: Other categories successfully and accurately detect pictures. (Picture credit: Original)

The number of successful inaccurate detection is 2, as shown in Figure 9 below:

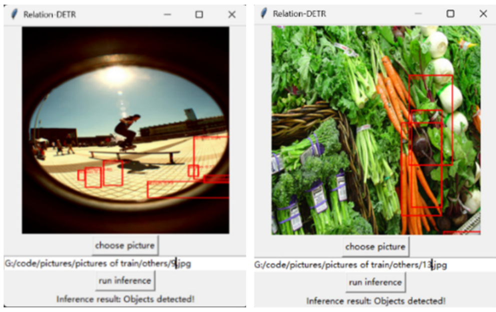


Figure 9: Other categories successfully but inaccurately detected pictures. (Picture credit: Original)

The number of unsuccessful detections is 4, as shown in Figure 10 below:

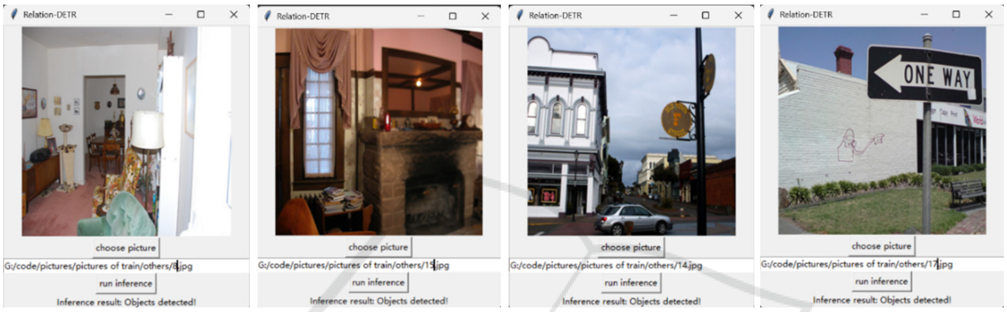


Figure 10: Other categories successfully but inaccurately detected pictures. (Picture credit: Original)

SCITEPRESS
SCIENCE AND TECHNOLOGY PUBLICATIONS

3.2 Experimental Result

In order to test the superiority of the Relation-DETR object detection method, the same data were simultaneously tested by the YOLO 11 object detection method, and the test results were as follows:

3.2.1 In the Animal Category

The number of successful and accurate detections is 20, and the remaining two types are 0, among which one picture successfully detected is shown in Figure 11:

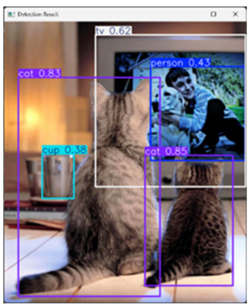


Figure 11: Successful and accurate detection of animals. (Picture credit: Original)

3.2.2 In the Landscape Category

The number of successful and accurate detections is 18, one of which is shown in Figure 12, and the other two types are 1, as shown in Figure 13 and Figure 14 respectively.

Figure 12: Successful and accurate detection of landscape images. (Picture credit: Original)

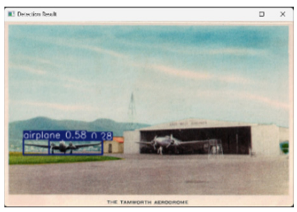


Figure 13: Scenery successfully but inaccurately detected pictures. (Picture credit: Original)

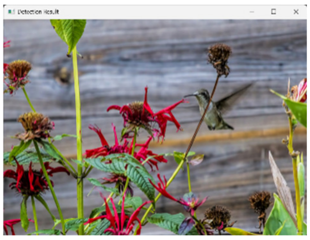


Figure 14: Unsuccessful detection of landscape images. (Picture credit: Original)

3.2.3 In the Human Category

The number of successful and accurate detections is 20, and the remaining two types are 0, of which one image is successfully detected as shown in Figure 15.

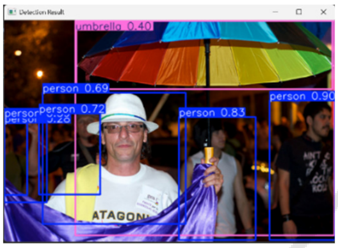


Figure 15: Successful and accurate detection of people. (Picture credit: Original)

3.2.4 In other categories

The number of successful and accurate detections is 19, one of which is shown in Figure 16, the number of successful inaccurate detections is 0, and the number of unsuccessful detections is 1, one of which is shown in Figure 17.

Figure 16: Successful and accurate detection of images in other categories. (Picture credit: Original)

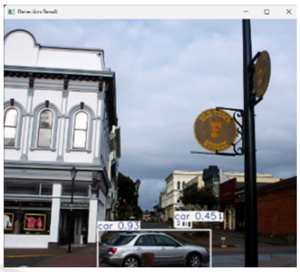


Figure 17: Other unsuccessful detection images. (Picture credit: Original)

In order to reflect the results more intuitively, the probability of the detection results of the four categories of animals, landscapes, people and general objects is statistically calculated through the statistical chart, so as to compare the Relation-DETR object detection method with the YOLO 11 object detection method. The statistical table is shown in Table 1 below.

Table 1：Statistical table of four types of comparative detection.

	Successful and accurate detection		Successful inaccurate detection		unsuccessful detection	
	Relation -DETR	YOLO11	Relation -DETR	YOLO11	Relation -DETR	YOLO11
animal	60%	100%	40%	0%	0%	0%
Landscape	15%	90%	50%	5%	35%	5%
People	85%	100%	15%	0%	0%	0%
other classes	70%	95%	10%	5%	20%	0%

From the above table statistics, it can be seen that among the four types of detection work, YOLO 11's successful detection rate is much higher than that of Relation-DETR, and the successful detection rate reaches more than 90%. Compared with the other three types, the detection accuracy of both methods is lower in landscape detection, which indicates that the object recognition model is not suitable for landscape detection.

4 CONCLUSIONS

Combined with the above research, it is found that from the experimental results, the YOLO 11 object detection method has a higher success rate and accuracy than the Relation-DETR object detection method. However, unlike YOLO 11, which loads image paths from code for identification, Relation-

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS

DETR has a user UI interface that makes it easy for non-technical people to train models and analyze results.

This interface design makes the Relation-DETR object detection method better than YOLO 11 in terms of visualization and operability, especially in projects that require presentation. After comparing the data with YOLO 11, we conclude that the advantage of Relation-DETR object detection method for object detection in large data sets is that the model can learn a wider range of features, thus having strong generalization ability, but this may also lead to insufficient recognition accuracy of the model in specific categories; in contrast, single object detection can achieve higher accuracy in specific fields, but generalization ability may be limited. For the performance of Relation-DETR in single object detection, the following improvement measures are suggested: firstly, refining the dataset to ensure that there are enough representative samples for each class; secondly, combining transfer learning technology, using the model weights pre-trained on large datasets to initialize the training of small datasets; thirdly, introducing domain-specific prior knowledge to enhance the recognition ability of the model for specific objects through feature engineering. Through the implementation of these improved schemes, this paper is expected to further improve the accuracy and generalization ability of object detection and provide a more reliable guarantee for practical applications.

REFERENCES

Chen, H. Y., Gao, J. Y., & Zhao, D. (2021). Deep learning and biomedical image analysis 2020 review. Journal of China Image Graphics, 26(3), 475-486.

Han, S. (2024). License plate recognition and speed estimation of highway speed measuring equipment based on deep learning. Electronic Components and Information Technology, 8(4), 103-106. https://doi.org/10.19772/j.cnki.2096-4455.2024.4.031

He, J. (2019). Medical image analysis and application of pneumoconiosis based on deep learning [Doctoral dissertation, Nanjing University].

Tao, P., Fu, Z., & Zhu, K. (2018). Research on medical computer-aided detection method based on deep learning. Journal of Biomedical Engineering, 35(3), 368-375.

Wang, J., Chen, Z., & Sun, J. (2025). Application of semi-supervised object detection based on fusion attention mechanism in rail transit. Locomotive Electric Drive, 1-7. https://doi.org/10.13890/j.issn.1000-128X.2025.01.104

Wang, W., Jiang, G., & Chu, Y. (2021). An overview of object detection systems from RCNN to YOLO. Journal of Qilu University of Technology, 35(5), 9-16. https://doi.org/10.16442/j.cnki.qlgydxxb.2021.05.002

Wang, Y. (2023). Research on improvement of intelligent driving target detection algorithm based on point cloud and image fusion [Doctoral dissertation, Jilin University]. https://doi.org/10.27162/d.cnki.gjlin.2023.001357

Xia, M. (2020). Cervical cancer cell medical image detection based on convolutional neural network [Doctoral dissertation, Tianjin University]. https://doi.org/10.27356/d.cnki.gtjdu.2020.003619

Zhang, T. B., Yang, Y., & Qu, Q. Q. (2023). Research on license plate detection and recognition algorithm in freeway scene. Western Communications Technology, (9), 205-207. https://doi.org/10.13282/j.cnki.wccst.2023.09.062

Zhuang, Y. (2022). Efficient and robust machine learning methods for challenging traffic video sensing applications [Doctoral dissertation, University of Washington].

SCITEPRESS SCIENCE AND TECHNOLOGY PUBLICATIONS