Analysis of Object Detection Model Performance through Intersection over Union (IoU) Metrics

Giovanni Mantovani and Andrea Palmieri

May 19, 2024

1 Introduction

The objective of this lab report is to explore the performance of object detection models using Intersection over Union (IoU) as a key evaluation metric. By examining the IoU scores, we aim to assess the accuracy and reliability of the models in detecting and localizing objects within images. This analysis will help in understanding the strengths and limitations of the models and provide insights into potential areas for improvement.

2 Methodology

2.1 Metrics

The primary metric used to evaluate the models' performances is the Intersection over Union (IoU).

IoU (Intersection over Union): Measures the overlap between the predicted bounding box and the ground truth bounding box. It is defined as the area of overlap divided by the area of union between the two bounding boxes. IoU values range from 0 to 1, where 1 indicates a perfect overlap.

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}} \tag{1}$$

Where:

- Area of Overlap is the area where the predicted and ground truth bounding boxes intersect.
- Area of Union is the total area covered by both the predicted and ground truth bounding boxes.

In this report, we will analyze images from the test set with the highest and lowest IoU scores to understand the model's detection performance and identify common challenges in object detection.

2.2 Dataset

The Uno Cards dataset, provided by Adam Crawshaw and released by Roboflow, consists of 8,992 images of Uno cards, featuring a total of 26,976 labeled examples. These images are taken against various textured backgrounds, providing a diverse set of conditions for object detection tasks. The dataset is released under a modified MIT license, as detailed here.

2.3 Model Used for Testing

We evaluate the performance of a single object detection model using the Intersection over Union (IoU) metric on the test set. The model used is as follows:

- **Model:** Faster R-CNN with ResNet50 backbone pre-trained on COCO dataset.

  Architecture and Considerations:
  Faster R-CNN with ResNet50 backbone is a robust model for object detection. The ResNet50 backbone is pre-trained on the COCO dataset. In our configuration, only the final classifier layers were finetuned while retaining the pre-trained weights for the rest of the network. This approach balance computational efficiency with detection accuracy by leveraging the pre-trained features of ResNet50.

Figure 1: Example image from the Uno Cards dataset

2.4 Training

Training was performed on the training set using a custom function that iteratively updates the model parameters. The model performed a forward pass to compute the loss, which was then backpropagated to update the model weights. This approach enabled the model to retain the learned representations from the source domain (pre-trained COCO weights) while adapting to the target task of object detection. By progressively fine-tuning the layers, the aim was to preserve the high-level features learned by the network while facilitating the learning of task-specific features from the target dataset.

2.5 Testing

The model was evaluated using a designated test set. For each image, only bounding boxes with a confidence score above the detection threshold of 0.8 were considered. To quantify the model's object detection performance, we calculated a score for each image, which was the average Intersection over Union (IoU) of the top three bounding boxes.

Initially, the score for each image was computed by averaging the IoUs of the detected bounding boxes, dividing by 3 even if the image had fewer than three bounding boxes above the detection threshold. This method penalized images that did not meet the threshold for three bounding boxes, resulting in lower scores. These images are identified as the "Top worst images" in our results: see Figures 7, 8, 9, 10, and 11.

Subsequently, we recalculated the scores by only considering images with at least three bounding boxes above the 0.8 threshold. This approach produced higher scores as it focused on images where the model detected more objects with high confidence.

The summary and comparison of these scores are detailed in the Results section, see Table 1. This dual approach provides a comprehensive assessment of the model's performance, highlighting both its strengths and areas needing improvement.

3 Results

The 5 images with the highest Intersection over Union score all achieved a score of 0.996 to 0.994. The images are the following: 2 3 4 5 6.

Image Name	IoU
Top 1 worst	0.650
Top 2 worst	0.655
Top 3 worst	0.660
Top 4 worst	0.662
Top 5 worst	0.663
Top 1 worst with 3 bbox	0.930
Top 2 worst with 3 bbox	0.933
Top 3 worst with 3 bbox	0.950
Top 4 worst with 3 bbox	0.951
Top 5 worst with 3 bbox	0.958
Top 1 best	0.996
Top 2 best	0.996
Top 3 best	0.996
Top 4 best	0.995
Top 5 best	0.995

Table 1: Intersection over Union values for the images

Figure 2: Best 1

Figure 3: Best 2

Figure 4: Best 3

Figure 5: Best 4

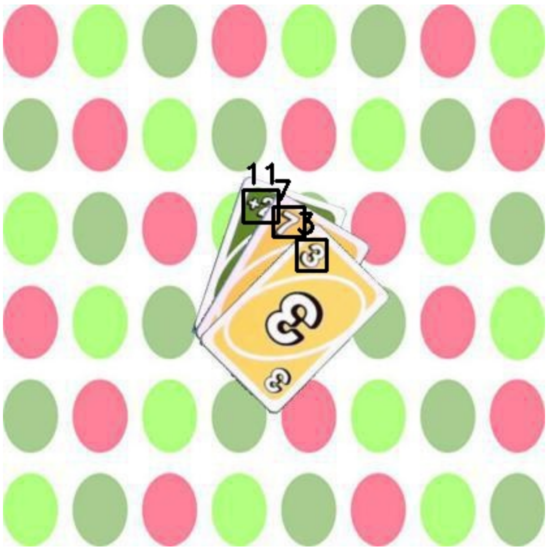

Figure 6: Best 5

Figure 7: Top 1 worst

Figure 8: Top 2 worst

Figure 9: Top 3 worst

Figure 10: Top 4 worst

Figure 11: Top 5 worst

Figure 12: Top 1 worst with 3 bbox above threshold

Figure 13: Top 2 worst with 3 bbox above threshold

Figure 14: Top 3 worst with 3 bbox above threshold

Figure 15: Top 4 worst with 3 bbox above threshold

Figure 16: Top 5 worst with 3 bbox above threshold

4 Interpretation of results

4.1 Top 5 Performances

The images with the highest average IoU scores are listed in Table 1. Upon examining these top-performing images, we notice that the Uno cards have, in some cases, a larger size relative to the whole image. This larger size likely helps the model in detecting and localizing the numbers more accurately, contributing to the higher IoU scores.

4.2 Worst 5 Performances

The images with the lowest average IoU scores are listed in Table 1. For the worst-performing images with two bounding boxes, the main problem appears to be the overlapping of cards, particularly in images ranked as Worst 2, Worst 3, and Worst 4. This overlapping creates confusion for the model, leading to inaccurate bounding boxes.

In the context of the worst cases with three bounding boxes, the threshold highlights additional issues related to the background. For instance:

- Complex Background Patterns: Worst 2 with three bounding boxes shows a complex background pattern that interferes with the model's ability to distinguish the cards.
- Color Similarity and glare: Worst 1 and Worst 5 with three bounding boxes have background colors similar to the colors of the Uno cards. This color similarity might make it difficult for the model to differentiate between the cards and the background.

5 Conclusion

The results of our investigation emphasize the impact of various image characteristics on IoU scores. High IoU scores are associated with images where Uno cards are large relative to the image size, making them easier to detect and localize accurately. On the other hand, low IoU scores highlight challenges such as overlapping objects and complex backgrounds. Overlapping cards, particularly in the worst-performing images, lead to confusion and inaccurate bounding boxes. Additionally, backgrounds with intricate patterns or colors similar to the Uno cards impacts negatively the model's ability to distinguish between the cards and the background, further reducing IoU scores.