Harnessing Pre-trained ResNet for YOLO-Based Object Detection: A Loss Function Journey

Ren-Di Wu

whats2000mc@gmail.com

1. Introduction:

The field of computer vision consistently seeks efficient object detection methods. This study explores adapting a pre-trained ResNet model within the YOLO framework, focusing on developing a compatible loss function to enhance object localization and classification accuracy.

2. Methodology:

This study's methodology is centered around a custom loss function for the YOLO-based object detection model, comprising four main components:

A. **Class Prediction Loss (`get_class_prediction_loss`):**

- This function is crucial for accurate class prediction of each detected object.
- It applies a mask to focus on predictions in cells containing objects, using Mean Squared Error (MSE) to ensure precise class probability distribution.
- The loss is calculated only for those cells that actually contain objects, thereby enhancing classification accuracy.

B. **No-Object Loss (`get_no_object_loss`):**

- Essential for accurately predicting the absence of objects in certain cells.
- This function penalizes incorrect confidence predictions for cells that do not contain objects.
- It ensures that the model does not falsely detect objects where there are none, improving the overall reliability of the detection.

C. **Containment Confidence Loss (`get_contain_conf_loss`):**

- This loss component refines the model's confidence in its predictions regarding object containment.
- By applying MSE loss to the confidence scores of bounding boxes, it ensures that the model accurately gauges its certainty in the presence of objects.
- This function contributes significantly to reducing false positives and improving detection confidence.

D. **Regression Loss (`get_regression_loss`):**

- Focused on the accuracy of bounding box coordinates.

- Utilizes MSE loss for the center coordinates and dimensions of the bounding boxes.
- This function is pivotal for ensuring that the model precisely predicts the location and size of each detected object.

Customizable coefficients like `l_coord` (5) for bounding box regression and l_noobj (0.5) for no-object confidence predictions allow for nuanced calibration of the model's sensitivity to different detection aspects.

3. Results:

The adaptation of the YOLO model, using a pre-trained ResNet as the backbone, led to significant achievements in object detection. On the VOCdevkit_2007 validation set, the model reached a mean Average Precision (mAP) of 0.5605. When employing the Exponential Moving Average (EMA) technique, the model's performance was further enhanced, achieving a mAP of 0.5697. (Figure 2)

In the test set, comprising 4950 images, the model exhibited a mAP of 0.4295. This test set performance, while slightly lower than the validation results, still underscores the model's effectiveness in generalizing to new, unseen data.

4. Analysis & Interpretation

The experiment demonstrates the model's strong capability in object detection, particularly when trained on a well-structured dataset. The EMA model's superior performance in the validation set is indicative of the benefits that come with more stable and consistent training methodologies.

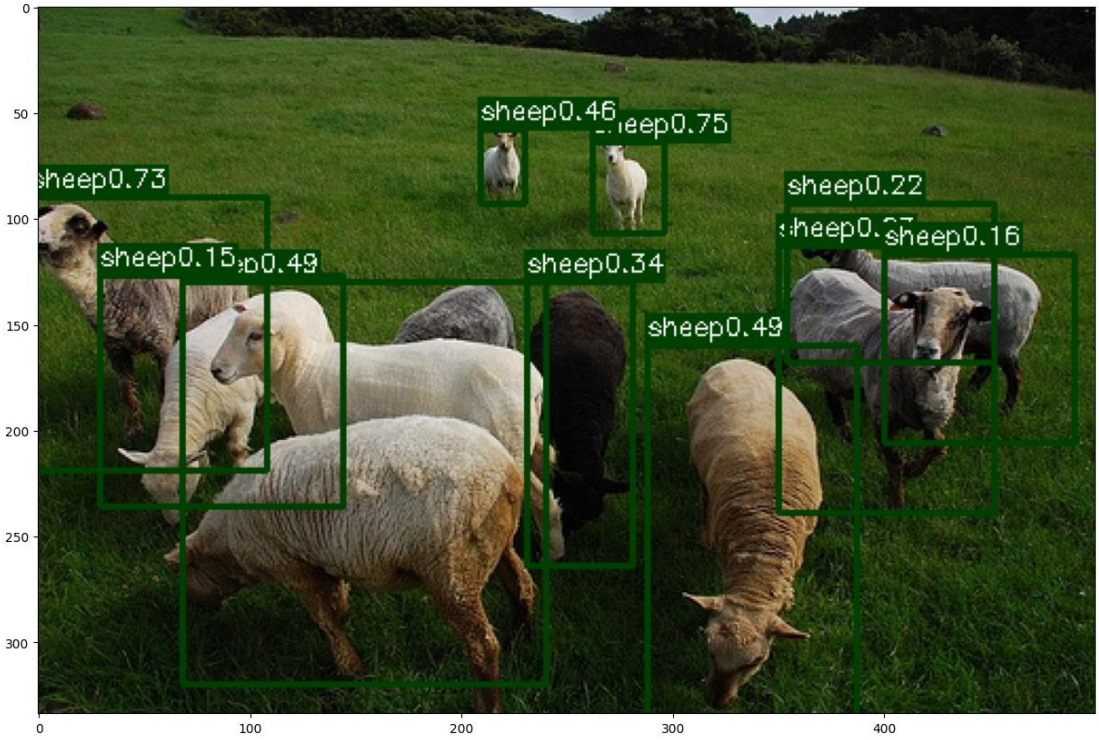
The slight drop in mAP on the test set compared to the validation set could indicate areas for improvement in model generalization. This discrepancy also highlights the importance of considering diverse datasets and scenarios in training to enhance the model's robustness and applicability to real-world scenarios.

5. Conclusion

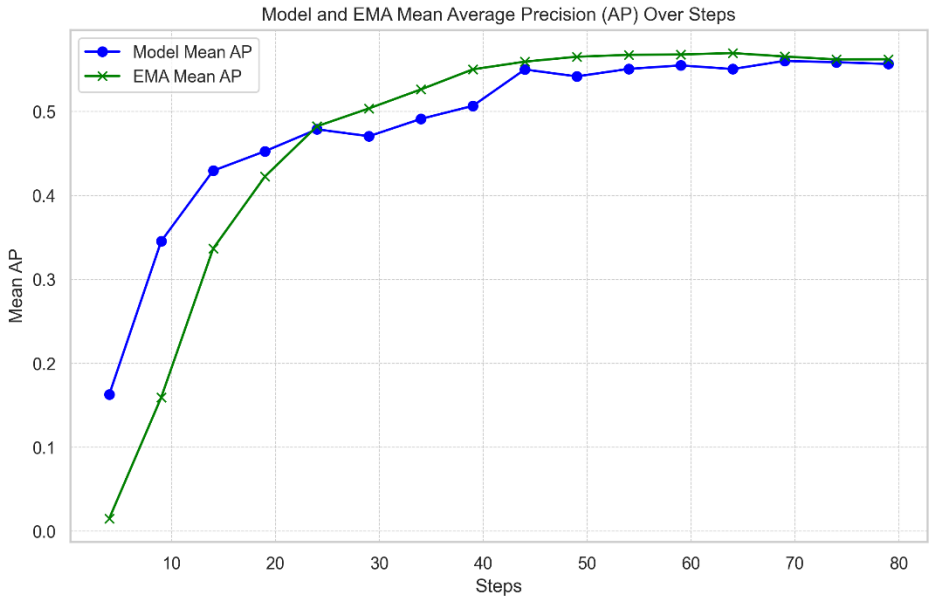
This study successfully demonstrates the effectiveness of leveraging pre-trained networks within the YOLO framework, enhanced by advanced training techniques like EMA. The achieved mAPs, both on the validation and test sets, establish the model as a robust tool for object detection tasks. Future work will focus on further improving the model's generalization capabilities and exploring additional advancements in loss functions and training strategies.

6. Appendices

This section provides additional visual evidence to complement the findings discussed in the report:



^ Figure 1: Detected Images Examples



^ Figure 2: MAP overview

Harnessing Pre-trained ResNet for YOLO-Based Object Detection: A Loss Function Journey