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:

The methodology focuses on constructing a robust loss function for a YOLO model tailored for object detection tasks. The loss function is designed to address several key aspects of object detection, including bounding box prediction, object presence confidence, and class prediction.

- **Intersection Over Union (IoU) Computation**: A fundamental component for evaluating object detection performance. We calculate the IoU for pairs of predicted and ground truth boxes, aiding in bounding box regression accuracy.

- **Bounding Box Regression**: The model optimizes the location and size of bounding boxes through regression, utilizing Mean Squared Error (MSE) to minimize the differences between predicted and target values.

- **Object Presence and Confidence**: The loss accounts for object presence in two parts:
  A. Through a mask that identifies cells containing objects
  B. By penalizing incorrect confidence predictions for object presence or absence, enhancing the model's ability to discern relevant features.

- **Class Prediction**: The model also aims to correctly predict the class of each detected object. This is achieved by applying a mask to only consider predictions in cells with objects and using MSE to enforce accurate class probability distribution.

- **Loss Coefficients**: Customizable coefficients for different loss components (e.g., `l_coord` for bounding box regression, `l_noobj` for confidence predictions) allow for fine-tuning the model's sensitivity to various aspects of the detection task. In this report the `l_coord` is set to 5, and `l_noobj` is set to 0.5.

The loss function, `YoloLoss`, encapsulates these elements, balancing the contributions of each aspect to train a model that excels in localizing and classifying objects within an image.

3. Results:

The application of the described methodology yielded promising results. The model achieved a mean Average Precision (mAP) of 0.5556 on the validation set and 0.4079 on the test set of the VOCdevkit_2007 dataset.

These metrics indicate a robust ability to detect and classify objects within the dataset. The higher mAP on the validation set suggests that the model could be capturing the nuances of the data it was trained on effectively.

The slight decrease in mAP on the test set could be indicative of the challenges faced when generalizing to unseen data. Nonetheless, the results are encouraging and demonstrate the model's potential in object detection tasks.

4. Analysis & Interpretation

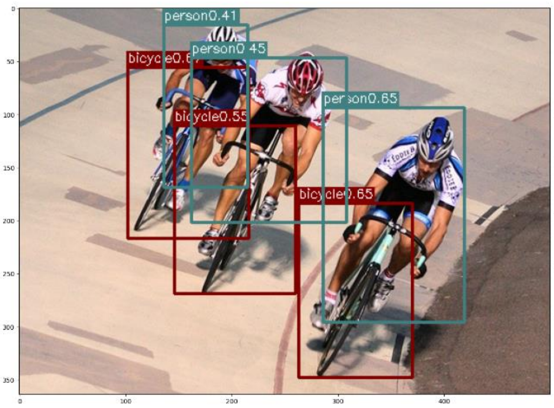
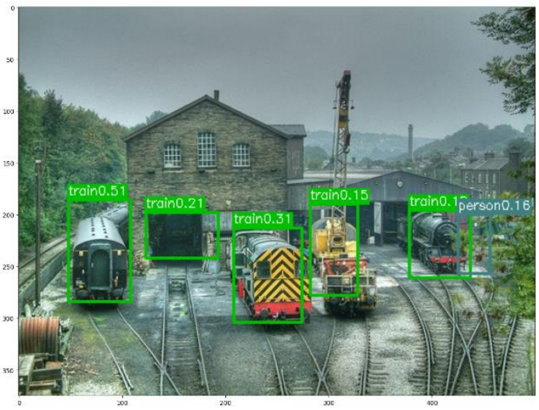
The model demonstrated a mean Average Precision (mAP) of 0.5556 on the validation and 0.4079 on the test set within the VOCdevkit_2007 dataset, indicating strong object detection capabilities, especially in the trained dataset.

5. Conclusion

The experiment showcasing the loss function's role in refining detection precision. It reflected on the blend of residual learning with real-time detection, setting a foundation for future enhancements.

6. Appendices

Include some examples of detected images showcasing the model's capabilities.


person0.41
bicycle0.6
person0.45
person0.65
bicycle0.55
bicycle0.65

train0.51
train0.21
train0.31
train0.15
train0.1
person0.16