Box Prediction Rebalancing for Training Single-Stage Object Detectors with Partially Labeled Data

Shafin Haque
Saratoga High School
shafin1025@gmail.com

R. Austin McEver
University of California, Santa Barbara
mcever@ucsb.edu

Abstract

Partial labeling schemes, in which annotators may label some instances of classes of interest and not label other instances, can significantly reduce annotation budgets and enable machine learning algorithms that might otherwise be impossible. However, these schemes introduce noise that makes training machine learning models difficult. The Dataset for Underwater Substrate and Invertebrate Analysis (DUSIA) uses a partial labeling scheme for its training set, which consists of thousands of partially labeled video frames. To combat the challenge of training on partially labeled data, we propose Box Prediction Rebalancing for single-stage object detectors and test our method on YOLOv5, a state-of-the-art single-stage detector. We rebalance the percentage of positive and negative detections included in the loss computation of the end-to-end model, improving our model's performance and generalizability.

1 Introduction

In large-scale, real-world object detection datasets, partially labeled annotations are important as they allow annotators to be more efficient. This is prevalent in DUSIA, a public dataset containing videos of marine invertebrate species and habitats, with the goal of detecting, classifying, and counting these species. By applying machine learning methods and increasing performances on DUSIA, we can further the effort for advancement in computationally analyzing videos for efficiency from a marine scientist setting. Among the frames from DUSIA's training set that are labeled with bounding boxes, it is not guaranteed all individuals of each species of interest have bounding box labels, but all existing bounding box labels are accurately labeled. However, the validation and testing frames are fully annotated, labeling all species of interest, enabling effective evaluation for object detectors.

McEver *et al.* [2] propose their novel Context Driven Detector, a derivative of Faster R-CNN, to detect the marine invertebrates. However, we propose the use of YOLOv5 [1] due to its stronger ability to detect small objects while also scanning images at a much higher frame rate.

To enhance training on a partially annotated set, McEver *et al.* [2] introduce a novel negative region dropping method, in which proposals from Faster R-CNN's [4] Region Proposal Network (RPN) that do not cover an object with ground truth are dropped from the loss calculation of the RPN part of the two-stage detector. Because the model may be detecting objects in training frames that are actually species, but have no ground truth, by dropping negative proposals, the model learns to pay more attention to true positive labels.

Building off the ideas of negative region dropping, we propose Box Prediction Rebalancing, a novel method for single-stage object detectors that takes advantage of the multi-step loss function of YOLO originally introduced by [3] but adapted by Ultralytics [1]. Our method is applied to both YOLOv5's box regression and objectness loss and showed to boost the model's performance and generalizability on DUSIA.

Preprint. Under review.

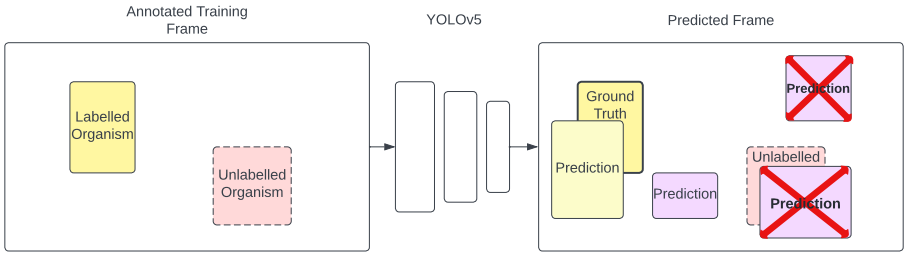

Figure 1: Illustration of the object detector and box prediction rebalancing method. The example training frame contains one labeled organism of interest and one unlabeled. The predicted frame contains one positive prediction (light yellow), but three negative ones (light purple). The red X over two of the predicted negative boxes shows how those boxes were randomly chosen to be removed and not included in the loss computation.

2 Methods

YOLO differs from Faster R-CNN in that the model does not contain an RPN. YOLO models are therefore single-stage, end-to-end networks. YOLOv5's loss function consists of three parts: box regression, objectness, and classification loss. The box regression and objectness loss update based on a tensor of the Intersection over Union (IoU) from the final predictions with the ground truth labels for a given training batch fed into the model.

In Faster R-CNN, Ren *et al.* [4] define a negative proposal as a proposal box with less than 0.3 IoU with some ground truth labels. We adapt this definition from region proposals directly to YOLO's detection during training, and randomly drop percentages of negative boxes from the tensor of IoUs, which is then used to update the loss.

YOLOv5 calculates the box regression loss based on the IoUs of the final predictions. By randomly removing a percentage of boxes that have low IoU with ground truth boxes (i.e. negative detections) and rebalancing the box predictions, we incline the model to update the box regression loss less from negative predictions, as they may, in fact, be organisms that simply do not have a ground truth label.

YOLOv5 uses a Binary Cross Entropy Objectness loss, which also uses the IoU values from the final predictions. Objectness is extremely important to our problem, as it is the probability the model thinks an object is there. So if there are species without labels, the model will penalize itself unnecessarily. We use the same process of randomly removing a percentage of negative predictions from the tensor of IoUs for predictions in a batch, and by modifying the inputs for objectness loss, we see the greatest improvement in Mean Average Precision (mAP).

3 Experiments

We run two separate experiments, where we test rebalancing the box predictions as outlined before. One experiment demonstrates the impact of removing negative detections from the box regression loss computation, and the other shows the impact of removing negative detections from the objectness loss. As an indicator of performance for our experiments, we use mAP with an IoU threshold of 0.5.

We start by tuning the hyperparameters of the YOLOv5 model on DUSIA. We find the optimal hyperparameters without box rebalancing and use these same hyperparameters for all experiments.

The default YOLOv5 model outperformed the original out-of-the-box Faster R-CNN tested by McEver *et al.* [2] as shown in Table 1. We then experiment with various rebalancing percentages for both the box regression and objectness loss, as described by Table 2 and Table 3. By randomly removing percentages of predictions from only the box regression loss, the validation mAP did not improve, however, the test mAP did increase by 0.7% at 25 percent removed. In contrast, removing percentages of negative predictions from the objectness loss increased the test mAP at all percentages tested except for 25 percent, which neither increased nor decreased the test mAP. By dropping 75% of negative predictions, the val mAP increased by 0.6% and the test mAP increased by 1.5%.

Model	val mAP	test mAP
Faster R-CNN	49.0	39.1
YOLOv5	59.0	48.3

Table 1: Performance of standard object detection models discussed on DUSIA.

removal pct	val mAP	test mAP
0%	**59.0**	48.3
10%	57.1	48.2
25%	58.1	**49.0**
75%	57.5	47.5
90%	58.1	47.8

Table 2: Affect on YOLOv5 performance by removing various percentages of box predictions from regression loss.

removal pct	val mAP	test mAP
0%	59.0	48.3
10%	57.7	48.6
25%	58.8	48.3
75%	**59.6**	**49.8**

Table 3: Affect on YOLOv5 performance by removing various percentages of box predictions from objectness loss.

Overall, box prediction rebalancing for our single-stage detector gave the model stronger generalizability, as it showed to increase test mAP from both the box and objectness loss experiments.

4 Conclusion

We propose Box Prediction Rebalancing for single-stage object detectors to combat the challenge of learning from partially annotated datasets. By randomly removing percentages of negative predictions from YOLOv5's box regression and objectness loss, our model performs better as it will learn less from negative boxes which may be true species of interest without ground truth labels. In the future, we hope to experiment with the effects of rebalancing predictions from both the box and objectness loss at the same time and also experiment with selectively targeting negative predictions to remove, instead of randomly removing percentages.

Acknowledgements

We would like to thank Professor B.S. Manjunath and the Vision Research Lab at the University of California, Santa Barbara for their support.

References

[1] Glenn Jocher. Ultralytics. `https://ultralytics.com`, 2022. Accessed: 2022-09-14.

[2] R. Austin McEver, Bowen Zhang, Connor Levenson, A. S. M. Iftekhar, and B. S. Manjunath. Context-driven detection of invertebrate species in deep-sea video, 2022.

[3] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once: Unified, real-time object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.

[4] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn: Towards real-time object detection with region proposal networks. In C. Cortes, N. Lawrence, D. Lee, M. Sugiyama, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015.