SOFT LABELS FOR RAPID SATELLITE OBJECT DETECTION

Matthew Ciolino and Grant Rosario and David Noever

PeopleTec, Inc, 4901 Corporate Dr NW, Huntsville, AL 35805, USA

ABSTRACT

Soft labels in image classification are vector representations of an image's true classification. In this paper, we investigate soft labels in the context of satellite object detection. We propose using detections as the basis for a new dataset of soft labels. Much of the effort in creating a high-quality model is gathering and annotating the training data. If we could use a model to generate a dataset for us, we could not only rapidly create datasets, but also supplement existing open-source datasets. Using a subset of the xView dataset, we train a YOLOv5 model to detect cars, planes, and ships. We then use that model to generate soft labels for the second training set which we then train and compare to the original model. We show that soft labels can be used to train a model that is almost as accurate as a model trained on the original data.

KEYWORDS

Soft Labels, Object Detection, Datasets

1. INTRODUCTION

Learning representations is a powerful tool in artificial intelligence. Deep learning has always been used to learn representations of images, text, and audio but furthermore can be used for transfer learning [1] and pre-training [2]. In this way one models strength can be used to improve performance on another task. As is commonplace for many object detection backbones, a pre-trained feature extraction network [3] is used to initialize the backbone of a model. This stabilizes training, improves convergence speed, and improves performance [4]. Soft labels attempt this by abstracting the transfer of information to the dataset level.

Soft labels use a well-trained model to completely generate the training data for another model. A clear use of soft labels is in model distillation [5] where the final layer in a large neural network containing the class probabilities is used as the ground truth to train a smaller network on. This effectively teaches a smaller model what a large model learned in a teacher-student [6] relationship. In this paper, we use soft labels to train an object detection model on satellite imagery and then rapidly create a dataset of soft labels.

1.1. Background

Dataset creation can be a time-consuming and costly endeavour [7]. While a handful of high-quality satellite object detection datasets exist [8], they may not be sufficient for the task at hand. By automating the process of annotations through soft labels, we can cut down on the time it takes to make a dataset. Various problems and solutions arise with soft labels and here are some examples:

David C. Wyld et al. (Eds): ICDIPV, CBIoT, ICAIT, WIMO, NC, CRYPIS, ITCSE, NLCA, CAIML -2023
pp. 239-245, 2023. CS & IT - CSCP 2023 DOI: 10.5121/csit.2023.131318

1.1.1. Missing Objects

One intuitive problem with soft labels for object detection is that low confidence detections will be filtered not allowing a model to learn finer grain details for objects. While this would leave the objects in question to be designated as background objects, Wu et al. [9], show that with the right modification to the bounding box loss function, a model can improve instead of worsening. To prove this, they dropped 30% of the ground truth labels and found a 5% drop in performance while when they weigh high-quality bounding boxes higher an increase of 3% is found as compared to a baseline.

1.1.2. Dataset Creation

One effective use of soft labels for object detection is subtyping. Subtyping is the process of taking a class and breaking it down into smaller classes. For example, a car can be broken down into a sedan, truck, and SUV. In our past work, Rosario et al. [10] show that with a simple car detection model and classifying cars by color, we can create a dataset of soft labels for car colors.

1.1.3. Overfitting

Overfitting is a common problem in deep learning. While there are many ways to combat overfitting, Zhang et al. [11] proposed using an online label smoothing to generate soft labels more reliably. In each epoch during training, they mix hard labels and the previous epochs soft labels to iteratively improve the soft labels of each object detected. That mixture is governed by the cross-entropy classification loss with the original distribution of soft labels being uniform. This method is shown to better define and separate classes on image datasets bringing a 2.1% gain on performance for top-1 error on VOC [12] and COCO.

1.1.4. Student Teacher Models

Using the soft labels from a teacher model, a student model can outperform models trained on partially labelled COCO data. Xu et al. [13] in this semi-supervised object detection (SSOD) paper, presented 2 techniques: soft teacher, where the teacher model is actually updated by the student using an exponential moving average, and a training strategy where the student/teacher models are trained using images with different augmentations. This use of soft labels improves state-of-the-art performance by 8.43% on 1% to 10% labelled coco data.

1.2. Contributions

The above methods are all examples of how soft labels can be used to improve a model. In this paper, we approach a far simpler task of investigating the performance drop-off using 100% soft labels as compared to the complete dataset. Ground truth versus soft label dataset can be seen in [Figure 1]. In this paper we attempt to answer 3 research questions that pertained to the nuances of soft labels:

- How does the soft labelling affect performance?
- Can this be used to categorize datasets automatically?
- What confidence value trains the best soft label model?

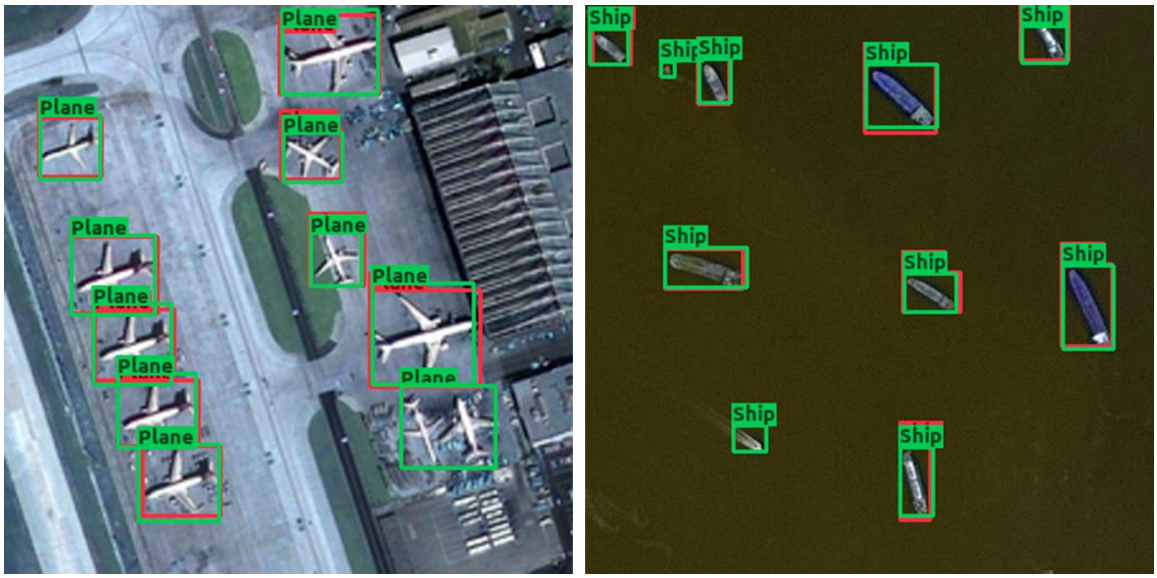

Figure 1. Ground Truth Labels (green) vs Soft Labels (red) from trained model

2. EXPERIMENT

To answer these questions, we lay out the following experiments [Figure 2] involving a subset of the xView dataset [14]. We sub-select xView for moving objects (vehicles, planes, and ships) leaving 34,252 images collected from parking lots, marine ports, and airports. Approximately one-third of the pictures were background, meaning they contained no examples but contributed to the training "null" set. Our data contains 37,712 instances of ships, 174,779 instances of cars, and 18,052 instances of planes. These counts are randomly split across 3 sets for train_1/train_2/valid sets containing 40/40/20 percent of the data respectively.

YOLOv5s [15] containing 7,027,720 parameters was trained on the train_1/train_2 set for 3 epochs with a batch size of 16 using the pretrained COCO [16] model as a checkpoint. Using those models, we soft-labelled the opposite training set to generate soft-labelled train_2/train_1. All models trained used the unseen valid set as the test set to compare metrics. We use the mAP, F1, and losses to evaluate the models.

- mAP is the mean of the average precision value for recall value over 0 to 1 for each class at a given IoU. [17]
- The IoU threshold is the intersection over union of the predicted bounding box and ground truth bounding box. [18]
- F1 is the harmonic mean of precision and recall at different confidence thresholds. [19]

We also look at train/test loss. For loss, YOLOv5 uses a combination [20] of bounding box loss, objectness loss, and classification loss to train the model.

- The bounding box loss is the mean squared error between the predicted bounding box and the ground truth bounding box.
- The objectness loss measures the probability that an object exists in a proposed region of interest through binary cross entropy.
- The classification loss is the cross-entropy loss between the predicted class and the ground truth class.

A detailed summary of YOLOv5 can be found here [21]. The test set which the results are all based on is broken down as follows: Across the 7,896 images and 58,165 instances in the test set, 8,710 instances (15%) are ships, 45,025 instances (77%) are cars, and 4,430 instances (8%) are planes.

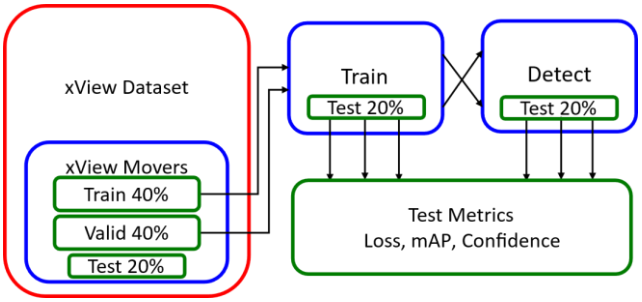

Figure 2. Experiment Flow with xView Movers Dataset

3. RESULTS

While ground truth models performed the best [Table 1], the soft label models consistently came within a 6% difference for mAP and F1 score. Interestingly, the F1 score remains very consistent (.68 to .70) with the optimal confidence value acting inversely to the confidence threshold of the soft-labelled training set. When looking at losses of the test set after the final epoch of training, we see a more interesting story that might tell us why the metrics are lower.

While the soft label trained model losses are both higher and lower than the ground truth model, we can derive insight from the relative differences. Both bounding box loss and objectness loss seem relatively consistent at an average difference of 11.35%. This is in stark contrast to the classification loss which averages a difference of 44.59%. Since classification loss has the highest change (still the lowest component of the loss) it can be said that the unlabelled objects in the soft label training sets account for the loss of performance.

We can look at the per-class metrics [Figure 3] to get a better look at what might be going on. As shown, most of the decrease in performance comes from soft-labelled planes. For planes trained on 0.5 confidence soft labels, all metrics dropped on average 8.49%. Interestingly, cars, which are the smallest but most abundant object in the dataset performed only 0.65% worse across all metrics than the ground truth models. Overall, per-class metrics dropped 4.44% across all metrics.

Table 1. Test Set Metrics on xView Movers

Model	**Conf**	**mAP50**	**mAP95**	**F1-Score**	**box loss**	**obj loss**	**class loss**
Train Set 1		0.76251	0.45069	.71 @ .419c	0.052841	0.03327	0.0058541
Train Set 1 Soft	0.3	0.73275	0.42357	.70 @ .482c	0.042647 (-21.35%)	0.029469 (-12.11%)	0.0033020 (-55.74%)
Train Set 1 Soft	0.5	0.70798	0.40687	.69 @ .318c	0.044096 (-18.04%)	0.030910 (-7.35%)	0.0041233 (-34.69%)
Train Set 2		0.7747	0.4609	.72 @ .421c	0.0415	0.027994	0.002451
Train Set 2 Soft	0.3	0.72932	0.42006	.70 @ .503c	0.042767 (+3.01%)	0.031548 (+11.93%)	0.0033173 (+30.04%)
Train Set 2 Soft	0.5	0.71681	0.41578	.70 @ .308c	0.044652 (+7.32%)	0.030846 (+9.69%)	0.0044479 (+57.89%)

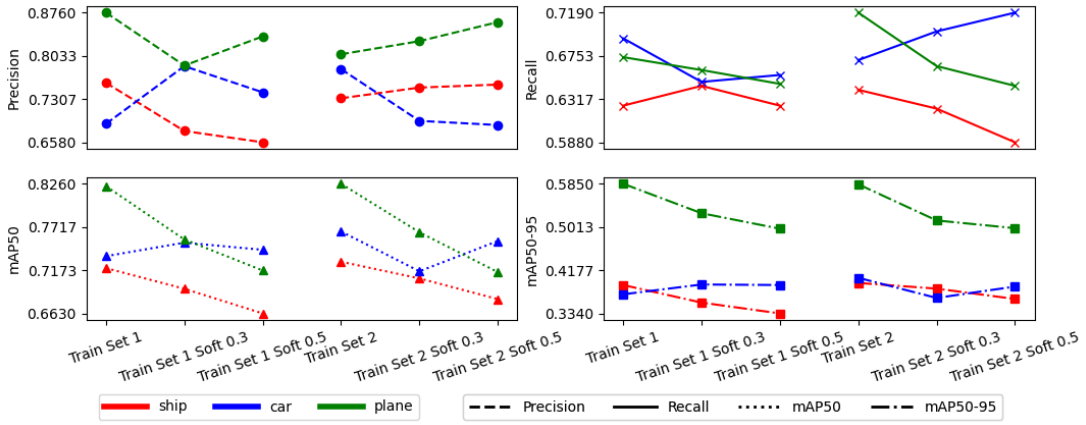

Figure 3. Per Class Metric Drop (color and line type map to class and metric)

4. DISCUSSION

While this performance drop is statistically significant in certain cases, i.e., planes, the benefits of soft labels outweigh this loss in performance by providing additional data in a low-cost and efficient manner as well as potentially increasing model knowledge in the cases of transfer learning. Moreover, this loss in performance can usually be remedied by either balancing the dataset or increasing data for lagging labels, as evidenced by the statistically insignificant drop in car labelling performance.

5. CONCLUSIONS

Models can be trained exclusively on soft labels with a less than 6% drop in mAP as compared to ground truth labels on the same dataset. Regardless of the confidence threshold used to create the soft labels, mAP/F1 scores remain within 1% of each other. This suggests that the soft labels are not overfitting to the training data. These results validate that rapid object detection datasets can be created with soft labels and that soft labels can be used to train models with a high degree of accuracy.

FUTURE WORK

In future work, we would like to use soft labels to improve the performance of the model on the test set by supplementing existing data instead of training solely on the soft labels. We would also like to use soft labels from one generalized model that could categorize any satellite image you are looking at into a dataset.

During experimentation OWL-ViT [22] was released. This Open-Vocabulary Object Detection model can be used to soft label. We compare YOLOv5 models with the COCO 2017 validation set with ground truth, 0.1 confidence OWL-ViT inference, and 0.25 confidence OWL-ViT inference. The mAP at 0.5 IoU for the ground truth is 0.2225, OWL-ViT at 0.25 confidence is .1844 (-17.12%), and OWL-ViT at 0.10 confidence is .1534 (-31.05%). More info can be found in the appendix section of the preprint.

ACKNOWLEDGEMENTS

The authors would like to thank the PeopleTec Technical Fellows program for encouragement and project assistance.

REFERENCES

[1] F. Zhuang, Z. Qi, K. Duan, D. Xi, Y. Zhu, H. Zhu, H. Xiong, andQ. He, "A comprehensive survey on transfer learning," Proceedings ofthe IEEE, vol. 109, no. 1, pp. 43–76, 2020.

[2] H. H. Mao, "A survey on self-supervised pre-training for sequentialtransfer learning in neural networks," arXiv preprint arXiv:2007.00800,2020.

[3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learningfor imagerecognition," ComputerScience, 2015.

[4] Y. Chen, T. Yang, X. Zhang, G. Meng, X. Xiao, and J. Sun, "Detnas:Backbone search for object detection," Advances in Neural InformationProcessing Systems, vol. 32, 2019.

[5] G. Chen, W. Choi, X. Yu, T. Han, and M. Chandraker, "Learning efficientobject detection models with knowledge distillation," Advances in neuralinformation processing systems, vol. 30, 2017.

[6] C. H. Nguyen, T. C. Nguyen, T. N. Tang, and N. L. Phan, "Improvingobject detection by label assignment distillation," in Proceedings ofthe IEEE/CVF Winter Conference on Applications of Computer Vision,2022, pp. 1005–1014.

[7] R. Incze, "The cost of machine learning projects,"Sep 2019. [Online]. Available: https://medium.com/cognifeed/the-cost-of-machine-learning-projects-7ca3aea03a5c

[8] Chrieke, "Chrieke/awesome-satellite-imagery-datasets: list of satelliteimage training datasets with annotations for computer visionand deep learning." [Online]. Available: https://github.com/chrieke/awesome-satellite-imagery-datasets

[9] Z. Wu, N. Bodla, B. Singh, M. Najibi, R. Chellappa, and L. S.Davis, "Soft sampling for robust object detection," arXiv preprintarXiv:1806.06986, 2018.

[10] G. Rosario, D. Noever, and M. Ciolino, "Soft-labeling strategies forrapid sub-typing," arXiv preprint arXiv:2209.12684, 2022.

[11] C.-B. Zhang, P.-T. Jiang, Q. Hou, Y. Wei, Q. Han, Z. Li, and M.-M. Cheng, "Delving deep into label smoothing," IEEE Transactions onImage Processing, vol. 30, pp. 5984–5996, 2021.

[12] M. Everingham, L. V. Gool, C. K. I. Williams, J. M. Winn, andA. Zisserman, "The pascal visual object classes (voc) challenge." Int. J.Comput. Vis., vol. 88, no. 2, pp. 303–338, 2010. [Online]. Available: http://dblp.uni-trier.de/db/journals/ijcv/ijcv88.html#EveringhamGWWZ10

[13] M. Xu, Z. Zhang, H. Hu, J. Wang, L. Wang, F. Wei, X. Bai, andZ. Liu, "End-to-end semi-supervised object detection with soft teacher,"in Proceedings of the IEEE/CVF International Conference on ComputerVision, 2021, pp. 3060–3069.

[14] D. Lam, R. Kuzma, K. McGee, S. Dooley, M. Laielli, M. Klaric,Y. Bulatov, and B. McCord, "xview: Objects in context in overheadimagery," arXiv preprint arXiv:1802.07856, 2018.

[15] G. Jocher, A. Chaurasia, A. Stoken, J. Borovec, Y. Kwon, K. Michael,and J. Fang, "ultralytics/yolov5: v6. 2-yolov5 classification models,apple m1, reproducibility, clearml and deci. ai integrations," Zenodo.org, 2022.

[16] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan,P. Doll´ar, and C. L. Zitnick, "Microsoft coco: Common objects incontext," in European conference on computer vision. Springer, 2014,pp. 740–755.

[17] J. Hui, "Map (mean average precision) for object detection,"Apr 2019. [Online]. Available: https://jonathan-hui.medium.com/map-mean-average-precision-for-object-detection-45c121a31173

[18] Baeldung, "Intersection over union for object detection,"Sep 2022. [Online]. Available: https://www.baeldung.com/cs/object-detection-intersection-vs-union

[19] "F-score," May 2019. [Online]. Available: https://deepai.org/machine-learning-glossary-and-terms/f-score

[20] P. Lihi Gur Arie, "The practical guide forobject detection with yolov5 algorithm," Apr2022. [Online]. Available: https://towardsdatascience.com/the-practical-guide-for-object-detection-with-yolov5-algorithm-74c04aac4843

[21] Ultralytics, "Yolov5 (6.0/6.1) brief summary · issue 6998 ·ultralytics/yolov5." [Online]. Available: https://github.com/ultralytics/yolov5/issues/6998

[22] M. Minderer, A. Gritsenko, A. Stone, M. Neumann, D. Weissenborn,A. Dosovitskiy, A. Mahendran, A. Arnab, M. Dehghani, Z. Shen et al.,"Simple open-vocabulary object detection with vision transformers,"arXiv preprint arXiv:2205.06230, 2022.

AUTHORS

Matthew Ciolino has experience in deep learning and computer vision. He received his bachelor's from Lehigh University in Mechanical Engineering.

Grant Rosario has research experience in embedded applications and autonomous driving applications. He received his Masters from Florida Atlantic University in Computer Science and his Bachelors from Florida Gulf Coast University in Psychology.

David Noever has research experience with NASA and the Department of Defense in machine learning and data mining. He received his Ph.D. from Oxford University, as a Rhodes Scholar, in theoretical physics.