Proceedings of the Joint Austrian Computer Vision and Robotics Workshop 2020 DOI: 10.3217/978-3-85125-752-6-22

Few-shot Object Detection Using Online Random Forests

Werner Bailer, Hannes Fassold
JOANNEUM RESEARCH, DIGITAL – Institute for Information and Communication Technologies
{werner.bailer,hannes.fassold}@joanneum.at

Abstract. *We propose an approach for few-shot object detection, consisting of a CNN-based generic object detector and feature extractor, and an online random forest as a classifier. This enables incremental training of the classifier, which reaches similar performance with around 20 samples as when using 50+ training samples in batch learning.*

1. Introduction

In many practical applications for object detection, it is relevant to detect new classes or subclasses of common objects, for which only very limited training data are available. While a large amount of literature on few-shot classification has been published in recent years, the problem of few-shot detection is more challenging, as it also involves identifying candidate regions for the yet unknown object classes. The problem of few-shot detection can be discriminated into the two following cases.

Refinement of existing classes. The new class to be trained is a specific subclass of a class already supported by an object detection algorithm, e.g., classifying "truck", when the classifier already has a class "vehicle". For this approach, an existing detector and classifier for the broader class (e.g. Yolo [6], Faster R-CNN [7]) can be used, and an additional classifier to be trained/adapted for the new classes is needed.

New classes. Candidate regions for such classes will not be found by the pretrained classifier, thus another detection approach is needed. One approach to find candidate regions is to use a detector trained on "objectness", i.e. the likelihood that a regions contains a coherent object. On the identified candidate regions feature extraction and classification can be performed, similar to the first case.

We aim to enable training new object classes with only few (i.e., 5-10) labeled examples, which may also not be available all at once, but being added gradually, improve the detector over time. The contribution of this paper is thus using a CNN-based object detection framework for generic object detection and feature extraction, and train an online classifier on these features. After discussing related work in Section 2, Section 3 presents the proposed approach and results, and Section 4 concludes the paper.

2. Related work

[1] does not actually perform detection, but uses bounding box regression as proposed in SSD to improve the localisation of the region of interest. Then binary object-or-not classification as proposed in Faster R-CNN is used, and uses a modified Faster R-CNN classifier to facilitate transfer learning. The work proposes regularisation based on the probability distribution of the known classes for the new target class. [2] propose a method for few-shot classification and detection, bootstrapped from few labeled instances. The method is based on components from Faster RCNN, using Selective Search or Edge Boxes for region proposals, and iteratively adds bounding box proposals and updates classifiers. [5] propose a pipeline using faster R-CNN up to ROI pooling, and two FC layers as feature extractors. Classification is then performed using a kernel method. [4] uses FPN to create an object detection pipeline using metric learning. Classification is done different for pretrained classes (using Inception v3 [10] up to FC2), while few-shot learning is done with FPN (in the DCN variant) instead. [9] propose to train a generic object detector on ImageNet, sampling positive and negative candidate regions. This approach is suitable for generic object detection, beyond the originally trained classes. An approach based on meta-features and learning reweighting of those fea-

tures is proposed in [3]. A recent work applies fine-tuning only region proposal and classification layers on a data set consisting of many base class and few new class samples while fixing the feature extraction part of the network can outperform meta-learning approaches [11].

3. Proposed Approach

We based our approach on [9], which we use as generic object detector and feature extractor. For the classification we follow the pipeline proposed in [12], which uses online random forests proposed in [8] as a classifier. The random forest can be incrementally trained, and is able to provide good results with few training samples. We use the model pre-trained on ImageNet from [9], and evaluate it on the 12 classes dataset provided with the authors' implementation[1]. Each of the classes has between 55 and 108 training samples. We compare to a linear classifier trained on the entire set of samples, and train our online random forest based classifier with all or a fixed subset of samples per class.

With the full set of examples, the online random forest based classifier performs similarly but slightly worse than the linear classifier, with an F1 score of about 0.80. Down to about 20 samples per class, the performance stays nearly constant. With 10 samples the performance drops to around 0.70, with 5 samples to about 0.67. Only then the performance starts to degrade more quickly, arriving at only about 1.5 times better than random when using a single sample. The results are visualised in Figure 1. It is apparent, that the reduction of the F1-score is mainly due to reduced recall. In nearly all cases the loss in terms of recall is caused by misclassifying the object, while only in few cases the target object is missed in the detection stage.

4. Conclusion

Based on a recently proposed framework, which we use for generic object detection and feature extraction, we have developed an approach for few-shot object detection using an online random forest as a classifier, which makes it incrementally trainable. With about 20 samples there performance in terms of F1 score is similar to a linear classifier on the full set, and drops by about 0.13 when using only

[1]https://github.com/mahyarnajibi/SNIPER/tree/cvpr3k

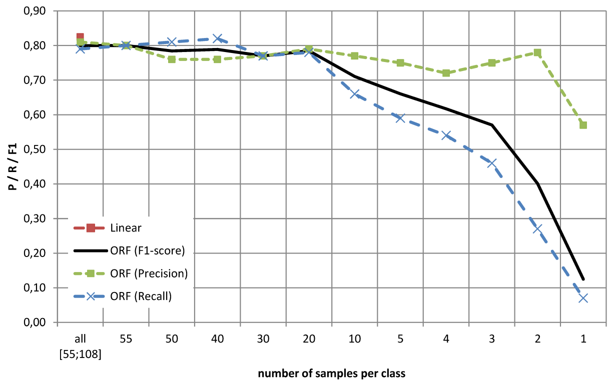

Figure 1. Detection results (F1 score, precision, recall) of the proposed approach on the 12 classes data set from [9], when trained on different numbers of samples per class. The confidence threshold is 0.15 for the online random forest classifier.

5 samples, which makes this a practically usable approach in use cases with few training samples.

Acknowledgments

This work has received funding from the European Union's Horizon 2020 research and innovation programme, under grant agreements n° 761802 MARCONI ("Multimedia and Augmented Radio Creation: Online, iNteractive, Individual") and n° 761934, Hyper360 ("Enriching 360 media with 3D storytelling and personalisation elements").

References

[1] H. Chen, Y. Wang, G. Wang, and Y. Qiao. Lstd: A low-shot transfer detector for object detection. In *32nd AAAI Conf. on AI*, 2018.

[2] X. Dong, L. Zheng, F. Ma, Y. Yang, and D. Meng. Few-example object detection with model communication. *IEEE T.PAMI*, 41(7), 2018.

[3] B. Kang, Z. Liu, X. Wang, F. Yu, J. Feng, and T. Darrell. Few-shot object detection via feature reweighting. In *Proc. ICCV*, 2019.

[4] L. Karlinsky, J. Shtok, S. Harary, E. Schwartz, A. Aides, R. Feris, R. Giryes, and A. M. Bronstein. Repmet: Representative-based metric learning for classification and few-shot object detection. In *Proc. CVPR*, 2019.

[5] E. Maiettini, G. Pasquale, L. Rosasco, and L. Natale. Speeding-up object detection training for robotics with falkon. In *RSJ IROS*, 2018.

[6] J. Redmon and A. Farhadi. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018.

[7] S. Ren, K. He, R. Girshick, and J. Sun. Faster R-CNN: Towards real-time object detection with region proposal networks. In *NIPS*, 2015.

[8] A. Saffari, C. Leistner, J. Santner, M. Godec, and H. Bischof. On-line random forests. In *Proc. ICCV Workshops*, 2009.

[9] B. Singh, H. Li, A. Sharma, and L. S. Davis. R-FCN-3000 at 30fps: Decoupling detection and classification. In *Proc. CVPR*, 2018.

[10] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna. Rethinking the inception architecture for computer vision. In *Proc. CVPR*, 2016.

[11] X. Wang, T. E. Huang, T. Darrell, J. E. Gonzalez, and F. Yu. Frustratingly simple few-shot object detection. *arXiv preprint arXiv:2003.06957*, 2020.

[12] M. Winter and W. Bailer. Incremental training for face recognition. In *Intl. Conf. on Multimedia Modeling*. Springer, 2019.