Meta-Learning for Object Detection in High Resolution Airborne Imagery

Benjamin Rise supervised by Murat Uney, Xiaowei Huang and Industrial Supervisor Scott Batchelor

EPSRC Centre for Doctoral Training in Distributed Algorithms, University of Liverpool, Liverpool, UK

Summary

- **Context and Motivations:** This PhD project will develop novel learning strategies that allow for automatic object detection in large airborne sensor imagery with sparse labels and varying characteristics.
- Particular focus will is on high resolution airborne imagery collected by unmanned air vehicles and/or satellites.
- Missions that collect such imagery generate vast amounts of data that are labour-intensive to review and analyse.
- Therefore, it is highly desirable to leverage machine learning techniques so that objects of interest can be detected automatically, removing the requirement for human analysts to sift through mountains of data.
- **Contribution:** Most modern object detection solutions use convolutional neural networks (e.g. YOLO, Faster R-CNN), that are pretrained on data from similar tasks and then fine-tuned. These methods rely on large databases of varied, labelled data. However, in our scenario, we often have very few labels, due to the time-consuming nature of annotation. We may also lack sufficient representative data. A promising area to help address these problems is meta-learning, where models gain experience from a distribution of related tasks, improving learning performance on each individual task. This project will use meta-learning techniques to invent new algorithms, that are tailored to the problem space outlined above. This will allow for the making of an object detector, that improves on previous state of the art methods in tackling these problems.

Meta-learning

In traditional supervised machine learning, we are given a training dataset: $\mathcal{D} = \{(x_1, y_1), ..., (x_n, y_n)\}$. In this paradigm a task τ seeks to learn a function f_θ that maps x to the labels y, i.e. $\hat{y} = f_\theta(x)$. This is achieved by minimizing a loss function, so that the predictions of $\hat{y}$ are as close as possible to the true labels y:

$$\theta^* = \arg\min_\theta \mathcal{L}(D; \theta, \omega)$$

here the conditioning on ω indicates dependence on assumptions made about how to learn, for example, the optimizer, initial parameters, or the function used to make predictions.

Meta-learning changes this by also learning ω, the learning algorithm itself (learning to learn). The parameters ω have a large effect on the performance of f_θ, hence learning a superior ω can significantly improve performance on novel tasks [1]. A common way to formalise this is the so-called task-distribution view. This states that a meta-learning algorithm is an algorithm that minimizes ω over a distribution of tasks $\tau \sim p(\tau)$, where we define a task as a dataset, and a loss function $\tau = \{\mathcal{D}, \mathcal{L}\}$:

$$\omega^* = \min_\omega \underset{\tau \in p(\tau)}{\mathbb{E}} \mathcal{L}(\mathcal{D}, \omega). \tag{1}$$

Object Detection

- **Modern Object Detectors:**

In a region based object detector, a CNN generates so called region proposals or regions of interest (RoIs), which are suggestions for potential objects in an image. These suggestions are then passed to a separate "head" which classifies these proposals and adjusts the proposal so that it fits the object better. The most popular of these type of detectors is Faster R-CNN [2].

The other main category of object detectors are single shot approaches. In single stage approaches, instead of relying on some mechanism to generate region proposals, the object detector classifies and regresses bounding boxes in a single forward pass through a CNN. A common approach to this is You Only Look Once (YOLO) [3]. See figure 1, for a schematic diagram of the basic architecture of most modern object detectors.

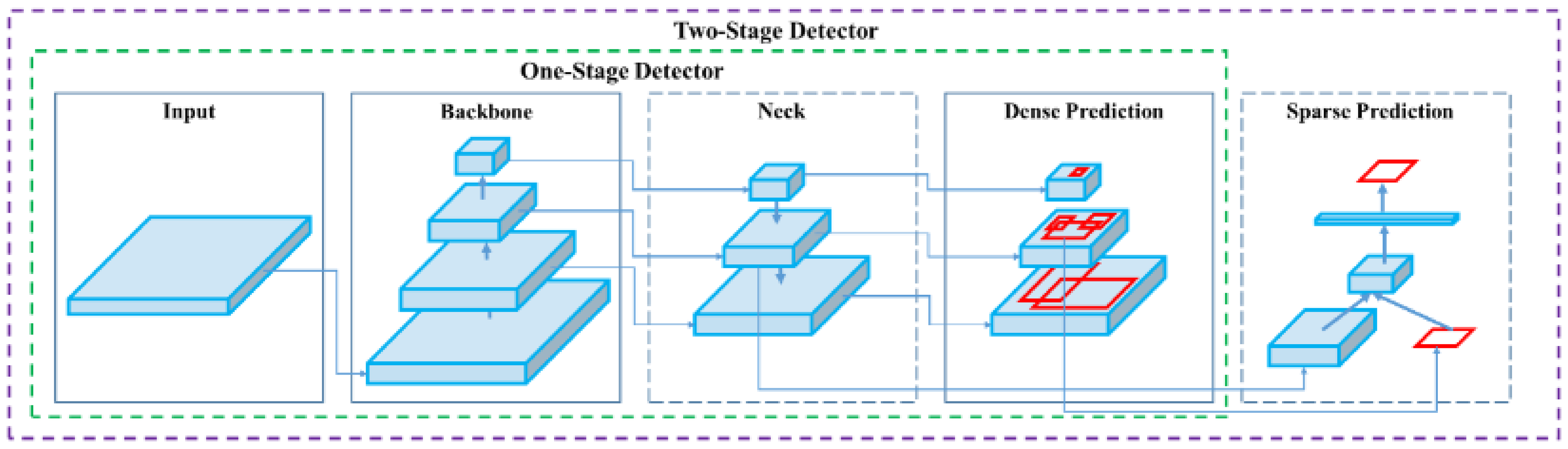

Figure 1: Schematic diagram of one-stage and two-stage object detectors [3].

- **Object detection in airborne imagery:**

While object detection in ground-based imagery has benefited from research into new deep learning approaches, transitioning such technology to overhead imagery is nontrivial [4].

Among the many challenges in training object detection algorithms for airborne imagery are: variations in image resolution, time period, look angle, typical objects they contain and their background. Another challenge is variation in look angle with respect to nadir. It is difficult to learn features from different angles and they are not directly transferable.

Figure 2: YOLT Training data [4].

Research Aims

- **Research hypothesis:**

Current approaches that fall under the umbrella of meta-learning can be used to develop effective learning strategies in airborne imaging sensor technologies.

- **Research aims:**

Current state of the art techniques use pre-trained models and transfer learning methods which are not particularly developed to the large area airborne imagery domain. To help facilitate object detection in these challenging conditions, we aim to:

- Develop meta-training techniques and sophisticated object detection architectures to help bridge the semantic gap between the same object viewed at different nadirs.
- Develop meta-learning strategies for addressing changing backgrounds.
- Develop data augmentation strategies, that will work effectively for the look angles with respect to nadir and resolutions that are typical in airborne sensor imagery.

Conclusion

The solutions developed in this project will contribute to meta-learning and provide effective learning strategies for the types of datasets that are of interest to the industry partner, Collins Aerospace. The powerful airborne imagery object detectors this project will develop will have a large range of applications, including surveillance, reconnaissance, surveying endangered animals and monitoring the effects of climate change.

References

[1] Timothy M. Hospedales et al. "Meta-Learning in Neural Networks: A Survey". In: *CoRR* abs/2004.05439 (2020). arXiv: 2004.05439. URL: https://arxiv.org/abs/2004.05439.

[2] Shaoqing Ren et al. *Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks*. 2016. arXiv: 1506.01497 [cs.CV].

[3] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao. *YOLOv4: Optimal Speed and Accuracy of Object Detection*. 2020. arXiv: 2004.10934 [cs.CV].

[4] Adam Van Etten. "You Only Look Twice: Rapid Multi-Scale Object Detection In Satellite Imagery". In: *CoRR* abs/1805.09512 (2018). arXiv: 1805.09512. URL: http://arxiv.org/abs/1805.09512.

Engineering and Physical Sciences Research Council

UNIVERSITY OF LIVERPOOL

Collins Aerospace