Active learning and object detection in multimodal aerial images

Master 2/Ecole d'Ingénieur internship

Expected starting: February/March 2024 (6 months)

Scientific context The context of this internship is motivated by issues raised in studies with data collected by airborne imagery. The automation of the processing of this data, by object detection methods and supervised learning, requires annotated databases. The annotation step is therefore a task of great interest, both in machine learning (ML) and computer vision (CV). Carrying it out manually is tedious and costly in terms of time and human resources. Furthermore, in the case of multimodal images (i.e. acquired by several sensors), annotation must be performed for each modality.

Active Learning (AL) is related to semi-supervised Machine Learning in which a learning algorithm can interact at each iteration with the user to get some information about labels of new data during the training step. It is motivated by situations in which it is easy to collect unlabeled data but costly (time, money, tedious task) to (manually) obtain their labels. It stems from the idea that we should only acquire labels that actually improve the ability of the model to make accurate predictions. Instances that are more useful than others according to some performance measures have to be identified to create an optimal training dataset: well chosen, fewer representative instances are needed to achieve similar performance as if we label and use all available data. This selection process has been investigated as *selective sampling* [9]. The importance of an instance is related to a high level of both the information and uncertainty relative to the trained model, considering therefore a trade-off between *informativeness* (ability to reduce the uncertainty of a statistical model) and *representativeness* (ability to represent the whole input data space) of the selection process [6].

In remote sensing, AL has therefore become an important approach to collect informative data for object detection and supervised classification tasks, and to assist the annotation process. The effectiveness of object detection models is intricately tied to the quantity of annotated data at their disposal. To overcome this challenge, AL attempts to formulate a strategy for cherry-picking pertinent data that an annotator should annotate, as elucidated by Choi et al. [5]. This typically involves employing a scoring mechanism that is related to the model's uncertainties about the data. Computationally, ascertaining these uncertainties usually necessitates a multi-model approach. However, it's noteworthy that these ensemble techniques are resource-intensive. Hence, the overarching objective of AL lies in the formulation of a classification function that faithfully mirrors the data's contribution to the learning process.

Scientific goals In the paper by Brust et al. [3], a novel approach to object detection using deep learning is introduced. Their approach incorporates AL strategies to explore unlabeled data. The authors proposed and compared various learning metrics that are suitable for most object detectors, taking into account class imbalance.

To start this project, the first step involves evaluating the performance of a multimodal object detector (like YOLOrs [10], SuperYOLO [13], YOLOFusion [7] ...) with respect to these metrics by applying them to a single modality (RGB for example). This evaluation will be carried out under different settings, including various sizes of the initial dataset and different adjustments of algorithm parameters. Then, the aim is to extend the AL strategy to the case of multimodal images. Indeed, for each object all modalities do not contribute equally to the classification/localization tasks, one can be more informative than the other.

Finally, metrics proposed by Brust et al. [3], focus on classification uncertainty, however, the aspect of localization is overlooked. To get the uncertainty of localization, we can use a strategy like the one of the Gaussian YOLO approach [4, 5] that provides both classification and localization uncertainties which we can then use with Brust et al. metrics.

Expected work In order to address the aforementioned objectives, the work program would be:

- Bibliographical review of multimodal object detection for remote sensing [2].
- Bibliographical review of active learning algorithms, mostly considering remote sensing and multimodal object detection.
- Implementation and evaluation of state-of-the-art methods through numerical experiment, on public remote sensing datasets (eg. VEDAI [8], SeaDrones [12], DroneVehicle [11], ISPRS 2D Semantic Labeling Contest – Potsdam dataset [1] ...). The experiments will be carried out on a set of aerial images as follows:
  - Choice of a multimodal object detector that will be trained later using active learning.
  - Application of a Gaussian YOLO-like strategy to get both localization and classification uncertainties.
  - Implementation and evaluation of AL algorithms.
- Improvement of existing solutions and/or development of a new approach to extend those AL methods to multimodal cases, by exploiting also the uncertainties related to modalities.
- Dissemination: Master thesis, and if possible scientific publication (with source codes).

Supervision

- Badie Belmouhcine (non-permanent assistant professor / IRISA-OBELIX)
- Chloé Friguet (assistant professor / IRISA-OBELIX)

Candidate profile Student in computer science and/or machine learning and/or signal & image processing and/or applied statistics, with good programming skills in Python (Pytorch knowledge appreciated), knowledge of deep-learning for image analysis, and high interest to investigate machine learning methods.

Application Send your CV + Motivation letter + Internship dates / duration + Master transcripts to `chloe.friguet@irisa.fr` and `abdelbadie.belmouhcine@univ-ubs.fr` **(before 7th, December 2023)**. Potential candidates will be contacted for an interview. Feel free to contact us for any questions.

References

[1] Isprs potsdam 2d semantic labeling dataset - potsdam. https://www.isprs.org/education/benchmarks/UrbanSemLab/2d-sem-label-potsdam.aspx. Accessed: 2022-03-18.

[2] Multimodal object detection in remote sensing. In *IGARSS 2023-2023 IEEE International Geoscience and Remote Sensing Symposium*, pages 1245–1248. IEEE, 2023.

[3] Clemens-Alexander Brust, Christoph Käding, and Joachim Denzler. Active Learning for Deep Object Detection:. In *Proceedings of the 14th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications*, pages 181–190, Prague, Czech Republic, 2019. SCITEPRESS - Science and Technology Publications.

[4] Jiwoong Choi, Dayoung Chun, Hyun Kim, and Hyuk-Jae Lee. Gaussian yolov3: An accurate and fast object detector using localization uncertainty for autonomous driving. In *Proceedings of the IEEE/CVF International conference on computer vision*, pages 502–511, 2019.

[5] Jiwoong Choi, Ismail Elezi, Hyuk-Jae Lee, Clement Farabet, and Jose M Alvarez. Active learning for deep object detection via probabilistic modeling. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10264–10273, 2021.

[6] Y. Fu, X. Zhu, and B. Li. A survey on instance selection for active learning. *Knowledge and Information Systems*, 35(2):249–283, 2013.

[7] Fang Qingyun and Wang Zhaokui. Cross-modality attentive feature fusion for object detection in multispectral remote sensing imagery. *Pattern Recognition*, 130:108786, 2022.

[8] Sebastien Razakarivony and Frederic Jurie. Vehicle detection in aerial imagery: A small target detection benchmark. *Journal of Visual Communication and Image Representation*, 34:187–203, 2016.

[9] B. Settles. Active learning literature survey. Computer sciences technical report, University of Wisconsin–Madison, 2010.

[10] Manish Sharma et al. YOLOrs: Object detection in multimodal remote sensing imagery. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 14:1497–1508, 2021.

[11] Yiming Sun, Bing Cao, Pengfei Zhu, and Qinghua Hu. Drone-based rgb-infrared cross-modality vehicle detection via uncertainty-aware learning. *IEEE Transactions on Circuits and Systems for Video Technology*, 32(10):6700–6713, 2022.

[12] Leon Amadeus Varga et al. SeaDronesSee: A maritime benchmark for detecting humans in open water. In *IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 2260–2270, 2022.

[13] Jiaqing Zhang et al. SuperYOLO: Super resolution assisted object detection in multimodal remote sensing imagery. *IEEE Transactions on Geoscience and Remote Sensing*, 61:1–15, 2023.