Fine-Grained Object Recognition and/or Localization in Remote Sensing Imagery

Atakan Serbes, 21200694

Department of Computer Science, Bilkent University

Email: atakan.serbes@bilkent.edu.tr

I. Problem Description

Nowadays, the amount of data collected by the aerial means, especially by the satellites, is increasing day by day with the advance of technology. Remote sensing image classification is one of the fundamental tasks of image processing in this field. Deep Convolutional Neural Networks (DCNN) [1] brought the state-of-the-art learning framework for the image recognition.

Earlier works in this field were like detecting roads [2] [3], buildings [4] [5] using traditional methods with classifiers such as random forests. Later, the DCNNs brought great results into this field. Using DCNNs for classification reached a point where increasing the accuracy became so much harder and their performance on most known datasets like UC Merced [6] already show great results. It is because the classes in older databases and applications on them worked on very distinctive features like a road, agricultural area, and coastal area.

Fine-grained object recognition on the other hand is about identifying a type of object among a large number of closely related sub-categories as mentioned in [7]. For example a recent work in the field may identify the types of the cars as small, medium, large and their colors rather than just identifying a car class. With the emerging resolution in imaging technology, it is now possible to collect more detailed images and the task of fine-grained object recognition will be more important as these data gets collected.

This is why I plan to study fine-grained object recognition in my project as it is open to further development.

II. Datasets

Many datasets were examined for the project. And to study the problem of fine-grained object recognition and/or localization, three of the dataset collections were selected. The first of the three datasets are DOTA-v1.5 Large-Scale Dataset [8] with 400,000 annotated instances with 16 categories which includes large areas like baseball diamonds, harbors as well as small parts such as small vehicles and helicopters. The second dataset considered is named PatternNet [9], a Large-Scale high resolution remote sensing dataset containing 38 different classes that vary from details like oil wells to areas like basketball courts. The third dataset may be the DLRSD [10] dense labeling dataset with 21 categories.

So, I plan to use different datasets and combine them to create a curated dataset for studying the problems of fine-grained object recognition and/or localization using deep learning. Other datasets that can be considered are [11] [12] [13] [14] and two datasets in [15].
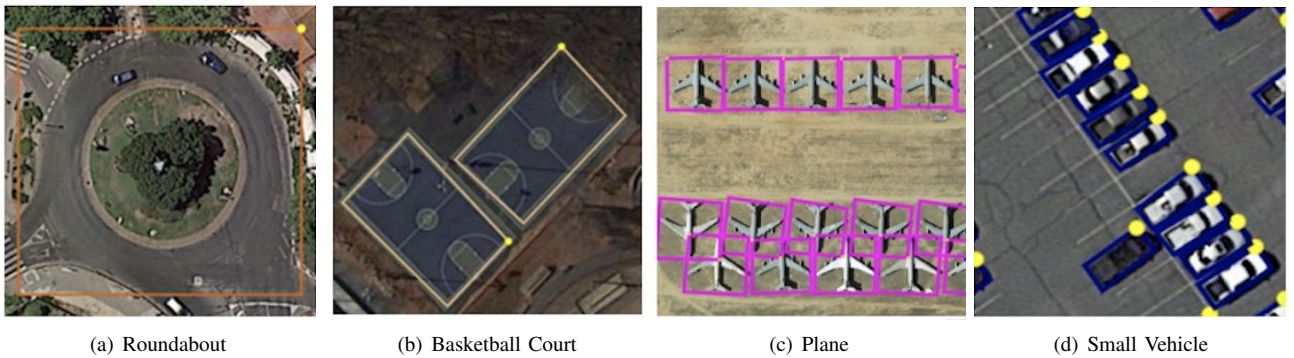

(a) Roundabout   (b) Basketball Court   (c) Plane   (d) Small Vehicle

Fig. 1: Some examples from the DOTA dataset [8]

III. Packages/Libraries to be Used

I plan to use Keras API inside TensorFlow during my project. However, there are works in the field [7] [16] which use PyTorch while working on fine-grained object recognition, so if it seems hard developing with TensorFlow, I can switch to PyTorch. Also since using these frameworks may also include using most-known libraries like pandas and numpy, I plan to use these libraries also.

References

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[2] V. Mnih and G. E. Hinton, "Learning to detect roads in high-resolution aerial images," in *European Conference on Computer Vision*. Springer, 2010, pp. 210–223.

[3] R. Alshehhi, P. R. Marpu, W. L. Woon, and M. Dalla Mura, "Simultaneous extraction of roads and buildings in remote sensing imagery with convolutional neural networks," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 130, pp. 139–149, 2017.

[4] H. G. Akçay and S. Aksoy, "Building detection using directional spatial constraints," in *2010 IEEE International Geoscience and Remote Sensing Symposium*. IEEE, 2010, pp. 1932–1935.

[5] M. Vakalopoulou, K. Karantzalos, N. Komodakis, and N. Paragios, "Building detection in very high resolution multispectral data with deep learning features," in *2015 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. IEEE, 2015, pp. 1873–1876.

[6] Y. Yang and S. Newsam, "Bag-of-visual-words and spatial extensions for land-use classification," in *Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems*. ACM, 2010, pp. 270–279.

[7] G. Sumbul, R. G. Cinbis, and S. Aksoy, "Multisource region attention network for fine-grained object recognition in remote sensing imagery," *arXiv preprint arXiv:1901.06403*, 2019.

[8] G.-S. Xia, X. Bai, J. Ding, Z. Zhu, S. Belongie, J. Luo, M. Datcu, M. Pelillo, and L. Zhang, "Dota: A large-scale dataset for object detection in aerial images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 3974–3983.

[9] W. Zhou, S. Newsam, C. Li, and Z. Shao, "Patternnet: A benchmark dataset for performance evaluation of remote sensing image retrieval," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 145, pp. 197–209, 2018.

[10] B. Chaudhuri, B. Demir, S. Chaudhuri, and L. Bruzzone, "Multilabel remote sensing image retrieval using a semisupervised graph-theoretic method," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 2, pp. 1144–1158, 2018.

[11] G. Cheng, J. Han, and X. Lu, "Remote sensing image scene classification: Benchmark and state of the art," *Proceedings of the IEEE*, vol. 105, no. 10, pp. 1865–1883, 2017.

[12] G. Cheng, J. Han, P. Zhou, and L. Guo, "Multi-class geospatial object detection and geographic image classification based on collection of part detectors," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 98, pp. 119–132, 2014.

[13] G. Cheng and J. Han, "A survey on object detection in optical remote sensing images," *ISPRS Journal of Photogrammetry and Remote Sensing*, vol. 117, pp. 11–28, 2016.

[14] G. Cheng, P. Zhou, and J. Han, "Learning rotation-invariant convolutional neural networks for object detection in vhr optical remote sensing images," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 54, no. 12, pp. 7405–7415, 2016.

[15] H. Li, C. Tao, Z. Wu, J. Chen, J. Gong, and M. Deng, "Rsi-cb: a large scale remote sensing image classification benchmark via crowdsource data," *arXiv preprint arXiv:1705.10450*, 2017.

[16] G. Sumbul, R. G. Cinbis, and S. Aksoy, "Fine-grained object recognition and zero-shot learning in remote sensing imagery," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 56, no. 2, pp. 770–779, 2018.