Critique Report on the Paper: "GAR: Graph Assisted Reasoning for Object Detection" [1]

Sumaiya Tabassum Nimi

Background Information and Purpose of the Research

Improved recognition of objects can be accomplished by taking the background scene and neighboring objects into account. Towards that end, in the paper [1], a novel scheme, called Graph Assisted Reasoning (GAR) was proposed for object detection that employed Graph Convolutional Network (GCN) for modelling the object-object and object-background relations.

Methods Proposed

The proposed object detection approach called GAR consisted of the following four steps.

- Step-1: Object proposals were generated by a backbone object detector (Faster R-CNN was employed in the paper).
- Step-2: Scene detector was trained for categorizing the background scene of the image. The background scene of an image was assumed to consist of three related information about the scene: whether its an indoor/outdoor scene, associated "Place" (like park, field, bedroom, etc.) label and the corresponding "Attribute" (like lighting, artificiality, etc.) label.
- Step-3: Rudimentary classification of the objects, using the object proposals obtained in Step-1, was done by a module called cRCN.
- Step-4: By using the raw classification scores produced in Step-3 and the scene classification information produced in Step-2 as nodes and by utilizing the prior knowledge about object-object and object-background relationship as edges, graph structure was generated, on which final object classification was done using GCN.

Strength of the Work

- Novel GCN-based scheme was proposed for object detection, probably the first work to employ graph modelling for object classification.
- The reported quantitative, as well as qualitative results demonstrated the superiority of the proposed approach for object classification, compared to the previous works.
- The proposed approach was also more efficient, compared to the other object detection approaches, in terms of reduced number of FLOPS and parameters.

Limitations of the Work

- The proposed approach was explicitly concerned about correct classification of objects based on their regular context. Hence the approach would become biased towards those contexts as found in the dataset. However,

objects appearing out of context is common in real life scenarios. The standard datasets on which the results were reported mostly consisted of images with regular contexts. Hence the proposed approach performed well. However, that did not validate that the observed superiority in terms of performance would be obtained when deployed in real-life applications also. In fact, taking into consideration the over-cautiousness of the approach regarding the context, it seemed like it would not be the case.

- The attribute information of the image background was not clearly understood. As regards to the place information, it was explicitly mentioned that the publicly available Place365 was used for training. But no such information regarding the training data used for attribute classification was provided the paper. Moreover, the meaning of the term "Attribute" also should have been discussed in more detail with more specific examples. Lack of such crucial information impeded the reproducibility of the work.

- As was observed from the quantitative results reported in Tables 1, 2 and 3, the proposed approach could not outperform the previous approaches on detection of most of the object classes.

Questions Unanswered

- The qualitative results should have included some images where objects appeared out of context. Like an object in an unusual background, or two objects co-existing in an image that did not usually co-existed together. Since such results were absent, it was not clear whether the approach would work better in those cases also.

- The paper was not coherent in terms of notations used. In fact, apparently it seemed that some notations were used in a superfluous manner. Like, a new notation, "Graph Convolutional Reasoning" was invented for the last component of the scheme, which was nothing more than a GCN. Also, notations like "cursory-cls" was used without any explanation at all. It would have been better to at least define the large number of notations that were used in the paper.

Suggested Future Studies

- As a future work, the learning rule involved could be modified by adding regularization terms so that the learning does not become too biased towards context provided in the training images. Maybe by assigning more weights to the training samples where objects appear slightly out of the regular context.

References

[1] Zheng Li, Xiaocong Du, and Yu Cao. Gar: Graph assisted reasoning for object detection. In *The IEEE Winter Conference on Applications of Computer Vision*, pages 1295–1304, 2020.