Building Context Cases: Adding Contextual Information to Objects in Images

Lawrence Gates[1,*]

[1]Indiana University – Luddy School, 700 N Woodlawn Ave, Bloomington, IN, 47408, U.S.A.

Abstract

Case-Based Reasoning's (CBR) ability to handle complex data is ideal for adding contextual information to images. This work explores the idea of using hierarchical case structures to build contextual information about a given image. Utilizing image tools that identify objects, this information can be stored concretely and abstractly in the case base. The CBR system would store the relationship of past contexts to solve future contextual problems.

Keywords

Case-Based Reasoning, Object Detection, Image Segmentation, Case Hierarchy, Knowledge Engineering

1. Introduction

The primary objective of this research is to establish that captured cases from prior detection episodes can be used to provide contextual information to increase accuracy of object detection in images. Case-Based Reasoning (CBR) (e.g., [1]) is well known for its strength in enabling reasoning in domains for which rule-based analysis is difficult. This is achieved by identifying key problem features and using them to retrieve relevant prior cases to apply to the new situation. As object detection may be done at various levels of granularity, it is inherently hierarchical, motivating the use of Hierarchical CBR, which uses a case structure that divides the cases up into more manageable subparts and in an abstraction hierarchy [2].

Limited work has been done on CBR and image processing [3], but integrating CBR with deep learning for object detection has been shown to be beneficial [4]. The primary motivation for capturing the contextual information in an image through hierarchical case structures is to exploit combined information from multiple information sources, including object recognition systems using deep learning. This work will contribute to CBR by investigating the application in the machine vision area through structural case representation.

2. Research Plan

This work investigates the application of CBR to the area of machine vision. The primary objective of this research is to establish that captured cases from prior detection episodes can be used to provide contextual information to increase accuracy of object detection in images. In addition to its contributions to object detection in machine vision, the research will contribute in (1) developing new methods for hierarchical structural case representation, (2) developing knowledge-based similarity assessment criteria for images, and (3) developing new methods for similarity assessment in hierarchical contexts with image data.

For this work, a *detection episode* can involve single or multiple image operations, such as providing detection labels and pixel coordinates in the image. There are a variety of image operations that can be done to determine the contents of an image. At this stage of the research, the current image operations are object detection and image segmentation, with additional operations that are applicable in the future. Object detection and image segmentation provide different data. For example, image segmentation can

ICCBR DC'24: Doctoral Consortium at ICCBR2024, July 1, 2024, Mérida, Mexico

*Corresponding author.

gatesla@iu.edu (L. Gates)

calculate the area of a segment versus the area of a rectangular bounding box. Eventually, meta-data from images (i.e. location, angle) will be incorporated as well to provide context.

2.1. Research Objectives

Given the main objective described previously, several sub-objectives exist:

1. **Developing new methods for hierarchical structural case representation.** Traditionally, there is a single case structure in CBR holding all the features (c.f., [2, 5]). Image data is rich with features but this information cannot always be effectively captured in a flat structure or single case. The context cases representing detection episodes will require hierarchical representation. The reason we are using hierarchical representation for context in an image is to enable reasoning about scene components at different level of granularity.
   A Hierarchical CBR system is built of *abstract* and *concrete* cases in a tree-like structure [5]. The detection episodes will be stored in the concrete cases. However, the detection episodes will have to be massaged as a single type of concrete case cannot effectively contain the features. There might be a need to have multiple concrete cases as sub-cases, due to the available features. Cases involving only image segmentation might have more information (i.e. area, extent) than object detection, since object detection defines a bounding box, but not the area of the object in the bounding box. Within the CBR approach, ontologies will be used to determine relationships between objects.
2. **Developing knowledge-based similarity assessment criteria for images.** In a naturally occurring situation, certain objects may or may not commonly appear next to other objects. Object sizes can be used to disambiguate different objects in an image. This is the knowledge we want to store and retrieve. The similarity assessment will not only rely on the features, but real-life connections and encoded inferences. Knowledge about the scene will be intensive, initially starting off with hand-crafted knowledge and possibly eventually moving up to mining a large language model for relations or information about the applicability of the relations.
3. **Developing new methods for similarity assessment in hierarchical contexts with image data.** As mentioned previously, hierarchical case representations contain both *abstract* and *concrete* cases. Similarity for hierarchical CBR requires traversing the tree structure and finding the similarity between the problem case and the stored cases. The levels of abstraction influence the similarity measurement. With the idea of multiple types of concrete cases, the similarity metrics will have to contain appropriate knowledge to work with a variety of concrete cases.

Besides extracting information from the image, external knowledge can be applied in general. Objects themselves have specific properties that can be encoded. By encoding these into the system, they are easily applied to every relevant case and used in similarity assessment. One property would be an object's function: a vehicle (i.e. truck, car, van) can transport another object. A truck is an easy example to see transporting (or object is in truck), but for the case of a car, the detector might see the person (object) inside of the bounding box of the car, if the point of view is getting a side view into the vehicle window. These three objects are unique vehicles, but at a higher level of abstraction are all vehicles.

2.2. Approach

Previous research on contextual detection utilizes more probability heavy ideas; Barnea and Ben-Shahar [6] utilize the probability of an object occurring based on the location of objects present or not present. Their goal is to calculate a new confidence of each detection at a given location based on the probability of a given object appearing next to another (i.e., a keyboard appear near a monitor). Perko and Leonardis [7] probability methods add contextual information to an image by utilizing a pre-defined radius around a source object and the distance to surrounding objects from the source object (spatial object co-occurrence). The authors aimed to have the system be able represent visual context, and

combine this with object detection, utilizing context priors, feature maps, and local-appearance based object detection.

The focus of the research is the construction of the hierarchical case structure and incorporating it with the knowledge of detection episodes and encoded knowledge. Since the focus of this research is on the development of CBR methods to assist with combining information, the image processing tools it uses are "off-the-shelf" pre-trained models[1](i.e., [8, 9]). By using the off-the-shelf models, one can have a system in place to easily substitute in state-of-the-art models or models trained on specific domains.

In order to build the hierarchical cases, we have to understand what information is available from imaging tools. Understanding the difference between object detection and image segmentation was the first step. Without this information, fuller cases could not be built. These cases initially were built off of models that perform object detection and image segmentation. The definition being used for image segmentation refers to assigning a pixel to a class in the image leaving no pixel unlabelled [10]. The difference between the object detection and image segmentation is that the bounding box does not provide the exact shape of the contained object where as the pixel map contours can be used to calculate various properties (i.e., area, extent). The models being used were trained on the COCO-MS dataset by the model creators [10].

In the COCO-MS dataset, object detection is used to identify "things" in an image whereas image segmentation identifies "stuff" in the image. "Things" are identified as objects that have "a specific size and shape", such as a cars or faces [11]. "Stuff" is identified as material "defined by a homogeneous or repetitive pattern ... but has no specific or distinctive spatial extent or shade" [11]. Examples of "stuff" are grass, buildings, or trees, which is better defined by image segmentation compared to object detection. "Things" and "stuff" are not mutually exclusive as they can work in tandem with each other. They can be combined to provide essential context in the image. One such example would be vehicles ('things') can be found on a road ('stuff').

After collecting the information from the images, the hierarchical case structure needs to be constructed. The hierarchical case structure will be developed from the bottom up, starting with the concrete cases. Construction of the abstractions can be accomplished through an object oriented design approach [5]. This approach does not apply to the current state of research, as the research into planning involved a robot completing a task, whereas the research here will focus on identification. As part of the abstraction, external knowledge can be encoded. The encoding would capture relationships between objects or reasons why objects would not be near each other in an image. One such example would be a cell phone would not be near a car when the cell phone is the same size of a car in the image from a drone's view.

With the case hierarchy and external knowledge, one will have the context needed that leverages the relationship between object and knowledge-based vision systems to improve the detection. Based on relevant cases in the case hierarchy, as well as on external knowledge, the approach adjusts the certainty values for predicted objects. To evaluate this approach, we will perform an ablation study by removing some or all in various aspects: detectors used, detected object relationships, features, and knowledge. Our plan for this stage is to use a single object detection model and image segmentation model, whereas later iterations of the research will add additional models to incorporate in the case information.

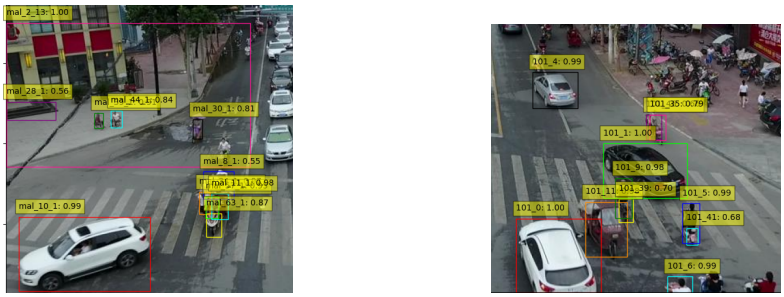
3. Progress Summary

The task domain for the project is to object identification in images collected from autonomous drones. The images are taken from a drone looking down in the dataset *VisDrone* [12].

As stated earlier, in order to build the cases, we needed to know what information we could get from off-the-shelf models. In Figure 1 are two images that have bounding boxes for object detection and image segmentation, respectively. These images show a given source detection and the 9 nearest

[1]Models used were from HuggingFace.co

detections. Image segmentation was converted from selected pixels to an overall bounding box in order to maximize space and easily compare bounding boxes in object detection.



(a) Image Segmentation (b) Object Detection

Figure 1: Example image operations for Image Segmentation (Fig. 1a) and Object Detection (Fig. 1b). Noted in the images are a source detection (red box) and nearest detections descending in colors.

The case structure can now be assembled, as the image segmentation data is available. The object detection model being used provides a location in the image, a label, and a confidence score. Image segmentation provides a mask representing the location in the image as contours. This is converted to bounding box(es), a label, and a confidence score. Before the contours are converted to bounding box(es), the following information can be calculated: area, extent, aspect ratio. At the moment, extent and aspect ratio do not have a specific use but they are available for future use.

After successfully implementing the retrieval of information from object detection and image segmentation models, implementation of the hierarchical case structure is the next step. With the information from these models, concrete cases can be constructed. Once the idea of a concrete case is developed, abstraction levels can be determined. The abstraction can be transferred into an object-oriented language in the form of inheritance [5].

In order to evaluate the effectiveness of the hierarchical case structure, multiple evaluation methods will be applied, as well as conducting an ablation study [13]. For the ablation study, the components that can be systematically removed would be features from the cases, level(s) of abstraction from the hierarchy, relational connections, and detection episode information (i.e., only using image segment or object detection). This should help show the strength of each piece in the overall structure. Additionally, since the models are "off-the-shelf", the system can be compared with different models. By comparing to different models, one would be able to see the system is not dependent on the model but uses it as a tool. Ultimately, that could show that a specific domain can be substituted in outside of the drone view, such as dash cam or animal identification in a forest.

4. Conclusion and Future Work

From the work completed in Section 3, the goal of integrating contextual information and assessing proposed object detections and justify the assessments. The current project is at the stage of starting the hierarchical case structure to be implemented. Once that has successfully been completed, the next steps would be incorporating external knowledge. The external knowledge will be useful in finding relationships between the cases and improving the similarity methods.

Following the use of external knowledge, making the hierarchical case representation approach explainable will be the next logical avenue. The structure of cases is ripe for explanation of detection decisions. Cases have been found to be a useful form in presenting explanations, as explored in a human-subjects study [14]. With the context of what is happening in an image, the system's explanation would highlight why a detection is correct or incorrect. Visual Question Answering can benefit from this approach, as it can provide relationships between items in an image[15]. The ability to explain why a detection was identified as a given label will demonstrate the power of the contextual knowledge in this approach.

5. Acknowledgements

This work was funded by the US Department of Defense (Contract W52P1J2093009).

References

[1] R. López de Mántaras, D. McSherry, D. Bridge, D. Leake, B. Smyth, S. Craw, B. Faltings, M. Maher, M. Cox, K. Forbus, M. Keane, A. Aamodt, I. Watson, Retrieval, reuse, revision, and retention in CBR, Knowledge Engineering Review 20 (2005) 215–240.

[2] B. Smyth, M. Keane, P. Cunningham, Hierarchical case-based reasoning integrating case-based and decompositional problem-solving techniques for plant-control software design, IEEE Transactions on Knowledge and Data Engineering 13 (2001) 793–812. doi:10.1109/69.956101.

[3] P. Perner, Why case-based reasoning is attractive for image interpretation, in: D. W. Aha, I. Watson (Eds.), Case-Based Reasoning Research and Development, Springer Berlin Heidelberg, Berlin, Heidelberg, 2001, pp. 27–43.

[4] J. T. Turner, M. W. Floyd, K. M. Gupta, D. W. Aha, Novel object discovery using case-based reasoning and convolutional neural networks, in: M. T. Cox, P. Funk, S. Begum (Eds.), Case-Based Reasoning Research and Development, Springer International Publishing, Cham, 2018, pp. 399–414.

[5] R. Bergmann, W. Wilke, On the role of abstraction in case-based reasoning, in: I. Smith, B. Faltings (Eds.), Advances in Case-Based Reasoning, Springer Berlin Heidelberg, Berlin, Heidelberg, 1996, pp. 28–43.

[6] E. Barnea, O. Ben-Shahar, Contextual object detection with a few relevant neighbors, in: C. V. Jawahar, H. Li, G. Mori, K. Schindler (Eds.), Computer Vision – ACCV 2018, Springer International Publishing, Cham, 2019, pp. 480–495.

[7] R. Perko, A. Leonardis, A framework for visual-context-aware object detection in still images, Computer Vision and Image Understanding 114 (2010) 700–711. doi:https://doi.org/10.1016/j.cviu.2010.03.005, special Issue on Multi-Camera and Multi-Modal Sensor Fusion.

[8] N. Carion, F. Massa, G. Synnaeve, N. Usunier, A. Kirillov, S. Zagoruyko, End-to-end object detection with transformers, CoRR abs/2005.12872 (2020). URL: https://arxiv.org/abs/2005.12872. arXiv:2005.12872.

[9] B. Cheng, A. Schwing, A. Kirillov, Per-pixel classification is not all you need for semantic segmentation, in: M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, J. W. Vaughan (Eds.), Advances in Neural Information Processing Systems, volume 34, Curran Associates, Inc., 2021, pp. 17864–17875.

[10] T.-Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, C. L. Zitnick, Microsoft COCO: Common objects in context, in: D. Fleet, T. Pajdla, B. Schiele, T. Tuytelaars (Eds.), Computer Vision – ECCV 2014, Springer International Publishing, Cham, 2014, pp. 740–755.

[11] G. Heitz, D. Koller, Learning spatial context: Using stuff to find things, in: D. Forsyth, P. Torr, A. Zisserman (Eds.), Computer Vision – ECCV 2008, Springer Berlin Heidelberg, Berlin, Heidelberg, 2008, pp. 30–43.

[12] P. Zhu, L. Wen, D. Du, X. Bian, H. Fan, Q. Hu, H. Ling, Detection and tracking meet drones challenge, IEEE Transactions on Pattern Analysis and Machine Intelligence 44 (2021) 7380–7399.

[13] P. R. Cohen, A. E. Howe, How evaluation guides AI research: The message still counts more than the medium, AI Magazine 9 (1988) 35. URL: https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/view/952. doi:10.1609/aimag.v9i4.952.

[14] L. Gates, D. Leake, K. Wilkerson, Cases are king: A user study of case presentation to explain CBR decisions, in: Case-Based Reasoning Research and Development: 31st International Conference Proceedings, Springer-Verlag, Berlin, Heidelberg, 2023, p. 153–168.

[15] M. Caro-Martinez, A. Wijekoon, B. Diaz-Agudo, J. A. Recio-Garcia, The current and future role of visual question answering in eXplainable artificial intelligence., CEUR Workshop Proceedings, 2023, pp. 172–183. URL: https://rgu-repository.worktribe.com/output/2048580.