

CLUE-AD: A Context-Based Method for Labeling Unobserved Entities in Autonomous Driving Data

Ruwan Wickramarachchi¹, Cory Henson², Amit Sheth¹

¹ AI Institute, University of South Carolina, Columbia, SC, USA

² Bosch Center for Artificial Intelligence, Pittsburgh, PA, USA

ruwan@email.sc.edu, cory.henson@us.bosch.com, amit@sc.edu

Abstract

Generating high-quality annotations for object detection and recognition is a challenging and important task, especially in relation to safety-critical applications such as autonomous driving (AD). Due to the difficulty of perception in challenging situations such as occlusion, degraded weather, and sensor failure, objects can go unobserved and unlabeled. In this paper, we present **CLUE-AD**, a general-purpose method for detecting and labeling unobserved entities by leveraging the object continuity assumption within the context of a scene. This method is dataset-agnostic, supporting any existing and future AD datasets. Using a real-world dataset representing complex urban driving scenes, we demonstrate the applicability of CLUE-AD for detecting unobserved entities and augmenting the scene data with new labels.

Introduction

Over the last several years, much progress has been made in the field of autonomous driving. This progress is largely due to the development and use of high-quality training datasets such as PandaSet (Scale AI 2020), NuScenes (Caesar et al. 2019), KITTI (Geiger et al. 2013), and Waymo (Sun et al. 2020). These datasets typically contain multi-modal data generated by an array of sensors such as cameras, LIDAR, and RADAR, along with high-quality annotations. Such datasets represent driving scenes as sequences and frames. A sequence represents a situation over a duration of time and is often encoded as a video. A frame represents a situation at a fixed point in time and is often encoded as an image sampled from a video sequence. Object detection and recognition typically occur for each frame.

Considering such progress, it is unfortunately still the case that objects go unobserved and unlabeled within these datasets. This may be due to several reasons, such as occlusion, low resolution, limited field-of-view, etc. (see Figure 2(a)). For example, object detection and annotation in degrading weather conditions remain a challenging and expensive task for state-of-the-art perception systems (Mirza et al. 2021). For this reason, there has been recent work on augmenting datasets to improve robustness for *specific conditions* such as rain (Halder, Lalonde, and Charette 2019; Garg and Nayar 2007), snow and fog

Copyright © 2023, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.



Figure 1: Left: A scene from Pandaset where only one cuboid label is provided for the bounding box. Right: The same scene is augmented by CLUE-AD with additional labels for the unobserved objects within the nested bounding box.

(Von Bernuth, Volk, and Bringmann 2019). Additionally, failures can occur while drawing 2D/3D bounding boxes where an object is nested within another object (see Fig.1).

CLUE is a method for detecting and labeling previously unobserved entities (UEs) in a target scene. It is designed to be dataset agnostic and robust against many failure scenarios such as full occlusion, degrading weather, sensor failures, outside field-of-view, etc. This is accomplished through the definition and analysis of the context of a scene, and by leveraging the object continuity assumption within this context. In this case, the context is defined as the sequence of scenes surrounding the target scene. Object continuity refers to the understanding, or assumption, that objects continue to exist even when they are not directly observed (Piaget and Cook 1954; Pascual-Leone and Johnson 2013). For example, in a driving scene, a pedestrian continues to exist even when they are no longer in direct view of the camera. In this case, the pedestrian could be occluded from view by standing behind another large object (e.g., a truck) or simply outside the camera’s current field of view. The object continuity assumption is a well-studied theory in Psychology that has been adopted and used in other fields, such as Semantics, to formally define and describe this basic principle of the physical world (Arp and Smith 2008; Arp, Smith, and Spear 2015).

The situation in which an object is present in a target scene, but not observed, can be inferred through analysis of its context, i.e., the sequence of scenes surrounding the target scene, including prior and/or subsequent scenes.

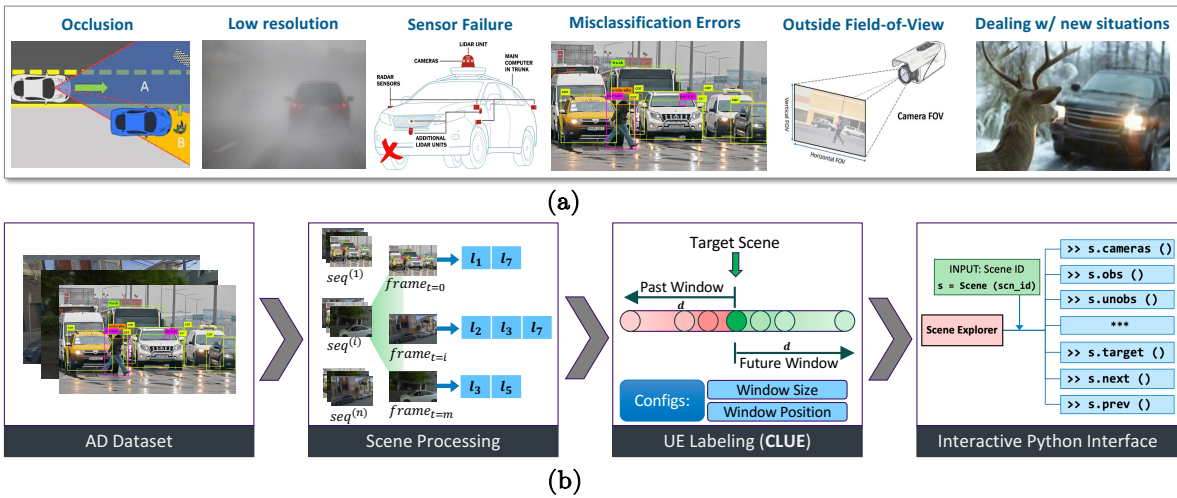


Figure 2: (a) Several situations in autonomous driving where unobserved entities may occur, (b) The overall architecture of CLUE-AD framework

Note that the concept of *scene* here refers to the spatial-temporal region surrounding the ego-vehicle. The use of this technique to detect and label previously unobserved objects provides several benefits. First, the ability to label additional objects in a scene may be used to validate the performance of existing CV labeling methods. Second, it allows for the creation of ground-truth data to be used by prediction approaches that are non-visual/semantic in nature (e.g., Knowledge-based Entity Prediction (KEP) (Wickramarachchi, Henson, and Sheth 2021, 2022)).

CLUE-AD Framework

The overall architecture of the CLUE-AD framework is shown in Figure 2(b). It illustrates the workflow starting from an existing AD dataset and ending with a Python interface that allows interaction with the dataset augmented with unobserved objects. First, the dataset is processed to generate lists of `<scene:labels>` pairs that ensure the sequential order of frames within a sequence and remove annotation redundancies due to duplicates from different detection methods (i.e., semantic segmentation, and object recognition with cuboids). Next, the main component — the CLUE method for labeling unobserved entities — is built on the premise of the object continuity assumption. It assumes that objects continue to exist even when they are either not visually observed or are (partially) observed but not annotated. This method is implemented in two primary steps: (1) Employing a scene windowing mechanism to define and analyze the context (i.e., surrounding scenes) of a target scene. The windowing mechanism can be configured to inspect k scenes (size of the window) across multiple positions — forward (i.e., future scenes), backward (i.e., past scenes), or bidirectional (i.e., both past + future scenes). (2) Executing an inference rule that yields a set of entities that are not observed in the target scene, but are observed within the specified window representing the target scene context. The dataset may now be augmented by adding additional scene labels for this

set of unobserved entities. Finally, the last component — a Python interface — provides an intuitive way to interact with the augmented dataset. It enables easy exploration of various data related to the target scene — including camera images from different angles, annotations, metadata, labels for unobserved entities, etc. — while also enabling easy navigation between adjacent scenes within the target scene context.

Demonstration

The demonstration of CLUE-AD focuses on showcasing key functionalities along with the benefits of labeling unobserved entities within an AD dataset. For this, we used Pandaset (Scale AI 2020), a real-world, open-domain, high-quality dataset with 103 sequences and 8240 frames recorded on the roads of San Francisco, CA. When CLUE is configured with a window size of 19, and window position: bidirectional (i.e., past + future scenes), the resultant dataset augmentation yields **12.49%** of scenes with at least one unobserved object. Further, we observed a gain of **1.34** unobserved objects per scene on average in the scenes augmented by CLUE-AD. Figure 1 (Right) shows an example from Pandaset, explored using the interactive Python interface, where CLUE was able to enrich the scene with two relevant, additional annotations of unobserved entities: `PersonSitting` and `PersonWithObject`.

Conclusions and Future Work

In this paper, we introduce CLUE-AD, a framework for detecting and labeling unobserved entities within AD scenes. CLUE-AD leverages the object continuity assumption across subsequent frames, and is designed to be dataset agnostic and robust to many object recognition and detection failures. Using a real-world AD dataset, we demonstrate the benefits of CLUE-AD by showing that 12.49% of scenes contain at least one unobserved object. In the near future, we plan to use the CLUE for evaluating prediction approaches that are non-visual, and more semantic in nature.

References

- Arp, R.; and Smith, B. 2008. Function, role, and disposition in basic formal ontology. *Nature Precedings*, 1–1.
- Arp, R.; Smith, B.; and Spear, A. D. 2015. *Building ontologies with basic formal ontology*. Mit Press.
- Caesar, H.; Bankiti, V.; Lang, A. H.; et al. 2019. nuScenes: A multimodal dataset for autonomous driving. *ArXiv*, abs/1903.11027.
- Garg, K.; and Nayar, S. K. 2007. Vision and rain. *International Journal of Computer Vision*, 75(1): 3–27.
- Geiger, A.; Lenz, P.; Stiller, C.; and Urtasun, R. 2013. Vision meets robotics: The kitti dataset. *The International Journal of Robotics Research*, 32(11): 1231–1237.
- Halder, S. S.; Lalonde, J.-F.; and Charette, R. d. 2019. Physics-based rendering for improving robustness to rain. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 10203–10212.
- Mirza, M. J.; Buerkle, C.; Jarquin, J.; Opitz, M.; Oboril, F.; Scholl, K.-U.; and Bischof, H. 2021. Robustness of Object Detectors in Degrading Weather Conditions. In *2021 IEEE International Intelligent Transportation Systems Conference (ITSC)*, 2719–2724.
- Pascual-Leone, J.; and Johnson, J. 2013. The psychological unit and its role in task analysis: A reinterpretation of object permanence. In *Criteria for competence*, 165–200. Psychology Press.
- Piaget, J.; and Cook, M. T. 1954. *The construction of reality in the child*. Basic Books.
- Scale AI. 2020. PandaSet Open Datasets - Scale. <https://scale.com/open-datasets/pandaset>. (Accessed on 09/20/2022).
- Sun, P.; Kretschmar, H.; Dotiwalla, X.; Chouard, A.; Patnaik, V.; Tsui, P.; Guo, J.; Zhou, Y.; Chai, Y.; Caine, B.; Vasudevan, V.; Han, W.; Ngiam, J.; Zhao, H.; Timofeev, A.; Ettinger, S.; Krivokon, M.; Gao, A.; Joshi, A.; Zhang, Y.; Shlens, J.; Chen, Z.; and Anguelov, D. 2020. Scalability in Perception for Autonomous Driving: Waymo Open Dataset. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2443–2451.
- Von Bernuth, A.; Volk, G.; and Bringmann, O. 2019. Simulating photo-realistic snow and fog on existing images for enhanced CNN training and evaluation. In *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, 41–46. IEEE.
- Wickramarachchi, R.; Henson, C.; and Sheth, A. 2021. Knowledge-infused Learning for Entity Prediction in Driving Scenes. *Frontiers in Big Data*, 4: 759110.
- Wickramarachchi, R.; Henson, C.; and Sheth, A. 2022. Knowledge-Based Entity Prediction for Improved Machine Perception in Autonomous Systems. *IEEE Intelligent Systems*, (01).