Ontolabeling: Re-Thinking Data Labeling For Computer Vision

Nicola Croce
Technical Product Manager
Deepen AI
Santa Clara, CA 95054
nico@deepen.ai

Marcos Nieto
Principal Researcher
Vicomtech
Parque Científico y Tecnológico de Gipuzkoa, Paseo Mikeletegi 57, 20009 Donostia / San Sebastián (Spain)
mnieto@vicomtech.org

Abstract

Over the last decade, developments in computer vision tasks have been driven by image, video, and multimodal benchmark datasets fueling the growth of machine learning methods for object detection, classification, and scene understanding.

Such advances have, however, created static, goal-specific and heterogeneous datasets, with little to none emphasis on the used taxonomies and semantics behind the class definitions, making them ill-defined, and hardly mappable to each others. This approach hinders and limits the long-term usability of datasets, their intercompatibility, extensibility, and the ability to repurpose them.

In this work we propose a new methodology for data labeling, which we call **Ontolabeling**[1], that detaches data structure from semantics, creating two data model layers. The first layer organizes spatio-temporal labels for multi-sensor data, while the second layer makes use of ontologies to structure, organize, maintain, extend and repurpose the semantics of the annotations.

Our approach is supported by an open source toolkit that enables label management (create, read, update, and delete) following the proposed Ontolabeling principles.

1 Introduction

In the last five years several voluminous multimodal Autonomous Driving (AD) datasets and toolsets have been open sourced from major OEMs and Tier-1 aiming to foster research and development in object detection, semantic scene understanding and other AD-related areas [3][6][7][8][9][11][12][14][15].

However, current research and engineering pain points are still on methodologies to manage extremely large volumes of data, and creating pipelines to process such data with semi-supervised approaches (e.g. labeling to create training sets and ground truth). Problems related to the interoperability of datasets will arise when new projects and AI functions demand different labels, richer level of detail or even different perspectives (e.g. labeling objects vs labeling actions and events).

[1] The Ontolabeling methodology and its resulting annotation JSON schema structure have been used as the basis of the soon to be released ASAM OpenLABEL standard for labeling datasets in the automated driving domain.

35th Conference on Neural Information Processing Systems (NeurIPS 2021), Sydney, Australia.

In addition, in the field of autonomous driving, data labeling has become a challenging engineering task, where multi-sensor vehicle set-ups produce complex data with images, point clouds, odometry entries, vehicle dynamics, often non-synchronized and with numerous coordinate systems and transforms involved.

Motivated by this lack of interoperability and data complexity, we propose the concept of Ontolabeling as a methodology and toolset to progress in the definition of semantic and standardised labeling principles. The Ontolabeling approach pillars are: (i) a JSON schema to structure labels as class instances, with attributes and relations to provide the necessary richness to describe complex AD-related scenes in multi-sensor set-ups; and (ii) the utilisation of ontologies as external resources of knowledge, where the semantics behind the classes and relations are universal and accessible for logical inference and reasoning mechanisms.

We complement our contribution with an open source toolkit providing support for the implementation of the Ontolabeling methodology. Next sections detail the differences between tradidional labeling approaches and our proposed Ontolabeling methodology, and reference the produced toolkit and related material.

2 Data Labeling: current approaches

In the context of computer vision for visual learning tasks, *data annotation*, or *data labeling* can be summarized as the process of enriching raw data, for example, data streams from multiple sensors, such as cameras, LiDAR, radar, with descriptive metadata. This metadata, or *annotation instance* annotates the content of the raw data by for example isolating and summarizing static or dynamic objects populating a video, actions they are performing, or environmental conditions, etc. Additional administrative meta-information about the data may also be included. Traditionally, labels are serialized and stored in text-based type of files such as XML, CSV or JSON files, that uses pointers to refer to a specific image, frame in a video, point in a point cloud, etc., as illustrated in Figure 1.

Labels produced this way contain the following information:

- **a semantic tag** (e.g. "car") usually with a clear meaning in human natural language
- **a labeling construct** such as a geometry to isolate concepts, e.g. objects to be detected in visual data (e.g. a bounding box around a car)
- **a reference** to a file or stream that is being labeled
- **additional metadata**: unique ID, pointers to frames where the labeled object is present, sensor calibration parameters, etc.

But what does a specific semantic tag actually mean? Usually, very little importance is given to the list of concepts that are labeled and their definitions. These specifications are often to be found in the documentation accompanying the labelled dataset, they are expressed in natural language, and they are almost always structured as a flat list. In sum, there is nothing in the process that rigorously, systematically, and formally organizes and manages the knowledge contained in the "semantic tag" of the labels.

This approach poses several limitations to the utility of labeled datasets:

First, it makes the labeling production error-prone and subject to non-objective interpretation of the semantics of the labelled classes.

Second, it limits dataset reutilisation: if a dataset has been labeled for a specific object detection task, for example, it cannot easily be repurposed to another, e.g. less granular task, since labeled concepts are not organized in a parent-child or other type of relationships.

Third, datasets using different labeling taxonomies cannot be easily composed with or mapped to each other, as no formalized semantics exist within and across labeled concepts. In addition, no direct form of semantic validation of the content of labels can be applied without costly post-processing steps.

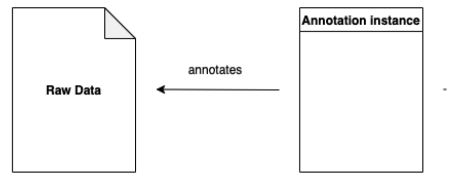

Figure 1: Conceptual overview data labeling: an annotation instance annotates specific raw data

3 Ontolabeling: managing labels' knowledge

Taking inspiration from the semantic web [4] and linked data principles [5], we propose a new methodology to label data for visual learning tasks, where the semantics of the labels are formalized and organized in external ontologies [10] that can be referenced within the annotation instance, where labels, attributes and relations of elements follow the labeling structure schema. Figure 2 depicts the Ontolabeling paradigm. Within the Ontolabeling methodology, labels can still be generally considered to be composed by the same parts described for labeling, with the only difference that the **semantic tag** part is substituted or complemented by an URI (Uniform Resource Identifier) pointing to a specific semantic concept defined in a specific ontology. This way, an "isInstanceOf" relationship between the labeled entity at the annotation file and the class at the ontology is established.

Using an ontology external to the static labeling file where labels are serialized to organize and orchestrate brings the following advantages:

First, the semantics of labels are rigorously and coherently organized, defined and orchestrated, making knowledge management much more efficient and less error-prone than traditional approaches relying on unstructured and loose documentation.

Second, the ontology can model relationships among different labeled classes, including hierarchical "SubclassOf", "PartOf", "ComposedBy" type of relationships. Action and event relationships can link objects and maneuvres, e.g. "isSubjectOfAction" can determine that a certain instance of a "pedestrian" participates in a "zebraCrossing" action. A Machine Learning practitioner can thus easily repurpose the same training dataset for computer vision tasks at different level of granularity or focused on different perspectives (object detection, action recognition, analytics).

Third, different datasets using different (or evolving) ontologies can be smoothly mapped to each other just by modeling their relationships (e.g. "sameAs" or other hierarchical concepts) in a combined ontology, without the need to re-touch and modify the labels in the first place. This can save time, costs and effort, and enables the composability of different datasets.

Fourth, while an annotation schema (e.g. JSON schema), can validate the structure of an annotation instance, it cannot validate its semantic content. The structure validation just guarantees that the machines can parse labels and proceed with their processing. The Ontolabeling approach, instead, paves the path towards an ontology-based validation of the labels semantics, reducing human errors in labels, improving data interpretation, and increasing the overall datasets quality.

3.1 Limitations

Ontologies are traditionally non-trivial to work with. Several different ontology modeling languages exist and organizing knowledge in a coherent manner is a task that usually requires a certain level of expertise, domain knowledge and familiarity with specific tooling. However, Ontolabeling enforces the reutilisation of ontologies as external resources, which may be created, maintained, extended and shared in public repositories as the result of joint efforts, so the labeling engineering tasks can deputize the responsibility of creating a solid knowledge model. As a result, the benefits outlined above can still be reaped by the non expert user, and advantages significantly surpass the overhead cost of managing ontologies.

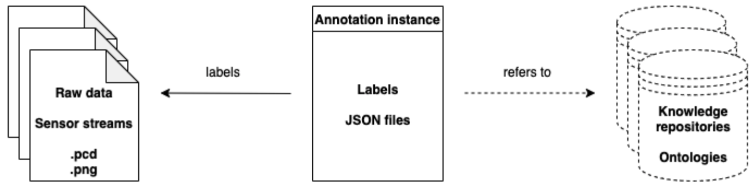

Figure 2: conceptual overview Ontolabeling: an annotation instance annotates specific raw data, referring to an external knowledge repository (ontology) to define, organize, structure and maintain the semantics of the labels.

3.2 Societal and ethical considerations

Our proposed methodology introduces a paradigm where increased importance and value is given to labels semantics, stimulating deeper reflections about categorization and knowledge modeling. For specific use cases and datasets this might lead to in-depth categorization choices that may exacerbate different form of bias and/or discrimination: a more articulated ontology may pose the risk of more articulated bias. On the other hand, the authors acknowledge that putting more emphasis and attention to the semantics of the labels used in visual datasets will trigger the ML community to think deeper about their categorization choices, raising awareness and helping uncovering hidden bias. Moreover, future work may address the developments of guidelines on how to ethically design labeling ontologies for sensitive applications.

4 Implementation and toolkit

The Ontolabeling concepts are implemented into the open source Video Content Description (VCD) toolkit[2] [13] available under the MIT License, which also includes an example ontology, the Automotive Global Ontology (AGO)[16] designed for labeling autonomous driving datasets.

The VCD toolkit is a multi-language library (Python, Typescript, C++) which provides CRUD operations for data labeling. Client applications can use the library APIs to manage labels following the structure JSON schema, and create links to semantic concepts in the AGO or any other ontology. The repository includes some converters from popular datasets (KITTI [8], Mapillary [12]), and example mapping of ontologies and label instances into a graph database (Neo4j) using the neosemantics plugin[3] for further reasoning and inference capabilities.

The concepts of Ontolabeling are being adopted by the ASAM e.v. OpenLABEL standard[1] and ASAM e.v. OpenXOntology standard[2].

5 Conclusions

In this paper we present the Ontolabeling methodology for semantic data labeling, specially devised for the automotive domain. Current approaches and industrial practices lack an appropriate distinction between label structure and semantics. As the volume and variety of datasets in the domain grow (also inside a single company/institution running multiple labeling projects), format and meaning incompatibilities will expose a number of problems. This situation will render data unusable or at least limit its usability to specific purposes rather than facilitating reutilisation, interoperability, maintenance or fusion.

In our methodology, we summarize the principles to separate data labeling into two layers. The first layer is governed by structure schemas, which guarantee compatibility at machine level. The second layer provides the semantics of labels by means of linked data definitions of classes in ontologies. In Ontolabeling, any label is an instance of a class defined (and related to other classes) in ontologies. Such subtle concept makes any labeled content to be easily extended, contextualized and re-used by means of creating the necessary ontology-level mappings.

[2]https://github.com/Vicomtech/video-content-description-VCD

[3]https://neo4j.com/labs/neosemantics/

Developments that support the Ontolabeling methodology include the open source toolkit VCD which we make available as a programming multi-language API to support labeling tools to create structured content, and an example ontology designed for the ADAS/AD development domain. In addition, the proposed concepts are being adopted by the ASAM OpenLABEL standardization project where the format of labels is being defined as a JSON schema, while the semantics are separated from structure and defined in the ASAM OpenXOntology standardization project.

References

[1] ASAM e.v. OpenLabel V1.0.0 standardization project. `https://www.asam.net/project-detail/asam-openlabel-v100/`. Accessed: 2021-09-28.

[2] ASAM e.v. OpenXOntology V1.0.0 standardization project. `https://www.asam.net/project-detail/asam-openxontology/`. Accessed: 2021-09-28.

[3] Siddharth Agarwal, Ankit Vora, Gaurav Pandey, Wayne Williams, Helen Kourous, and James McBride. Ford multi-av seasonal dataset. *The International Journal of Robotics Research*, 39(12):1367–1376, 2020.

[4] Tim Berners-Lee, James Hendler, and Ora Lassila. The semantic web. *Scientific american*, 284(5):34–43, 2001.

[5] Christian Bizer. The emerging web of linked data. *IEEE intelligent systems*, 24(5):87–92, 2009.

[6] Holger Caesar, Varun Bankiti, Alex H Lang, Sourabh Vora, Venice Erin Liong, Qiang Xu, Anush Krishnan, Yu Pan, Giancarlo Baldan, and Oscar Beijbom. nuscenes: A multimodal dataset for autonomous driving. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11621–11631, 2020.

[7] Marius Cordts, Mohamed Omran, Sebastian Ramos, Timo Rehfeld, Markus Enzweiler, Rodrigo Benenson, Uwe Franke, Stefan Roth, and Bernt Schiele. The cityscapes dataset for semantic urban scene understanding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3213–3223, 2016.

[8] Andreas Geiger, Philip Lenz, Christoph Stiller, and Raquel Urtasun. Vision meets robotics: The kitti dataset. *The International Journal of Robotics Research*, 32(11):1231–1237, 2013.

[9] Jakob Geyer, Yohannes Kassahun, Mentar Mahmudi, Xavier Ricou, Rupesh Durgesh, Andrew S Chung, Lorenz Hauswald, Viet Hoang Pham, Maximilian Mühlegg, Sebastian Dorn, et al. A2d2: Audi autonomous driving dataset. *arXiv preprint arXiv:2004.06320*, 2020.

[10] Tom Gruber. Ontology., 2018.

[11] Xinyu Huang, Peng Wang, Xinjing Cheng, Dingfu Zhou, Qichuan Geng, and Ruigang Yang. The apolloscape open dataset for autonomous driving and its application. *IEEE transactions on pattern analysis and machine intelligence*, 42(10):2702–2719, 2019.

[12] Gerhard Neuhold, Tobias Ollmann, Samuel Rota Bulo, and Peter Kontschieder. The mapillary vistas dataset for semantic understanding of street scenes. In *Proceedings of the IEEE international conference on computer vision*, pages 4990–4999, 2017.

[13] Marcos Nieto, Orti Senderos, and Oihana Otaegui. Boosting ai applications: Labeling format for complex datasets. *SoftwareX*, 13:100653, 2021.

[14] Matthew Pitropov, Danson Evan Garcia, Jason Rebello, Michael Smart, Carlos Wang, Krzysztof Czarnecki, and Steven Waslander. Canadian adverse driving conditions dataset. *The International Journal of Robotics Research*, 40(4-5):681–690, 2021.

[15] Pei Sun, Henrik Kretzschmar, Xerxes Dotiwalla, Aurelien Chouard, Vijaysai Patnaik, Paul Tsui, James Guo, Yin Zhou, Yuning Chai, Benjamin Caine, et al. Scalability in perception for autonomous driving: Waymo open dataset. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2446–2454, 2020.

[16] Itziar Urbieta, Marcos Nieto, Mikel García, and Oihana Otaegui. Design and implementation of an ontology for semantic labeling and testing: Automotive global ontology (ago). *Applied Sciences*, 11(17):7782, 2021.