VisionKG: Unified Access for Integrated Visual Datasets

Jicheng Yuan
Open Distributed Systems
Technical University of Berlin
Berlin, Germany
jicheng.yuan@tu-berlin.de

Manfred Hauswirth
Open Distributed Systems
Technical University of Berlin
Berlin, Germany
manfred.hauswirth@tu-berlin.de

Danh Le-Phuoc
Open Distributed Systems
Technical University of Berlin
Berlin, Germany
danh.lephuoc@tu-berlin.de

***Abstract*—Large-scale datasets play a key role in the success of modern computer vision. However, most are narrowly tailored for specific tasks, with restricted image data distributions. There is no unified approach to organizing and accessing them across diverse sources, tasks, and taxonomies. Hence, models can only perform well in specific domains. This not only creates unnecessary overheads when building robust visual recognition systems but also introduces biases into learning systems and limits the capabilities of data-centric AI. Addressing these problems, Vision Knowledge Graph (VisionKG) is a novel framework that interlinks and organizes visual datasets via knowledge graphs. As real-world scenarios encompass diverse weather conditions and imbalanced class distributions, we will introduce a semi-supervised domain adaptive object detection architecture empowered by VisionKG to navigate these challenges. Additionally, drawing on data enriched with semantically rich descriptions in VisionKG, we will evaluate VisionKG's efficacy in boosting transparency, ensuring traceability of data sources, and fostering trust in AI systems.**

***Index Terms*—Knowledge-Graph, Data-Centric AI Systems**

I. Introduction

Computer vision has made significant advances and visual datasets have become a crucial component in building robust visual recognition systems. The performance of the underlying deep neural networks (DNNs) in the systems is influenced not only by advanced architectures but also significantly by the quality of training data. There are many available visual datasets, e.g., ImageNet, OpenImage, and MS-COCO, which offer a range of visual characteristics in different contexts to improve the generalization capabilities of advanced machine learning models [1].

However, these datasets are often published in disparate data formats, with taxonomy and annotation quality fluctuating considerably. Furthermore, labels defining objects span across diverse lexical definitions like WordNet, Freebase, or plain text. This results in semantic discrepancies across datasets. Fragmented and non-unified datasets not only compound the complexities in visual recognition system development but also infuse biases, curtailing data-centric AI's potential.

To address the data inconsistency problems mentioned above, a unified data-centric framework, VisionKG (as shown in Fig. 1), was introduced. It is built based on the Linked Data principles [2], adhering to the FAIR and open science guidelines, and encompasses various data sources. These sources have been defined and maintained by the research community, as they are widely used and have a significant impact on the development of computer vision algorithms and systems. Their popularity ensures that they will be regularly and frequently updated and extended. This makes VisionKG [3] a valuable resource for researchers and developers who require access to the newest, high-quality image data and training the models in a data-centric way [4].

II. Motivation

The field of Artificial Intelligence (AI) advanced significantly in recent years and most of the breakthroughs are data-intensive AI or its combinations with other techniques. Therefore, it is reasonable to argue that to make AI better, we need data-centric ways to create/organize (training) data and make it available to be consumed by AI models. In different AI areas, e.g., Computer Vision (CV) and Natural Language Processing (NLP) etc., and even within the same area, datasets are organized differently, e.g. stored in different formats and using different labels. This hinders not only the advancements in individual AI area but also the path towards Artificial General Intelligence (AGI). Let us make an analogy with the developments of human intelligence (HI) to demonstrate the importance of sharing data and knowledge. To advance in sciences and technologies, over thousands of years, we have accessed all kinds of knowledge and data, and this is because we have (written) languages. Based on previous discoveries, new ones were discovered and recorded in shareable forms. Similarly, the current AI era was ignited by abundant data availability. While ImageNet showcased the prowess of Deep Learning in CV, datasets from sources like Wikipedia fueled NLP models. Despite the capabilities of models like GPT or BERT that might not rely on labeled data, many downstream applications still do. The fragmentation of these datasets implies that even if AGI-ready algorithms exist, adequate training data might not. This underscores the need for a harmonized framework for dataset sharing and access.

III. Research Goals and Contributions

In pursuit of the concepts articulated earlier, we initiated the construction of a unified knowledge graph, VisionKG, tailored for CV datasets such as COCO, KITTI, and Visual Genome. VisionKG stands as an RDF-based knowledge graph encapsulating RDF statements which detail the metadata of images and the semantics of interconnected annotations. Leveraging the