When Wikipedia meets ImageNet: Visual Entity Linking for Wikimedia Commons

Master Project Winter 2023 – AI and Intelligent Systems Group

Abstract

Multimodal settings for neural networks are becoming crucial to represent the world at large appropriately. One of the most widely-used projects relying heavily on multimodal data (in the form of images, text, and linked data) is Wikipedia. This connection between images and text is subtle, yet crucial for representing and conveying information. To strengthen this connection between images, text, and linked data, the Wikimedia Foundation introduced Structured Data on Commons, enabling Wikimedia Commons, the media- and picture-storage of Wikimedia, for structured data. Specifically, it allows for annotation of pictures on Wikimedia commons with entities from Wikidata, the structured data backbone of Wikipedia. All projects- Wikipedia, Wikimedia Commons, and Wikidata, are edited and maintained by a community of users, ensuring a higher quality of data due to a large manual effort of the communities.
Images annotated with structured data can have a large effect on downstream tasks: question answering over images could be facilitated, as well as caption generation across languages, as entities in Wikidata are language-independent.

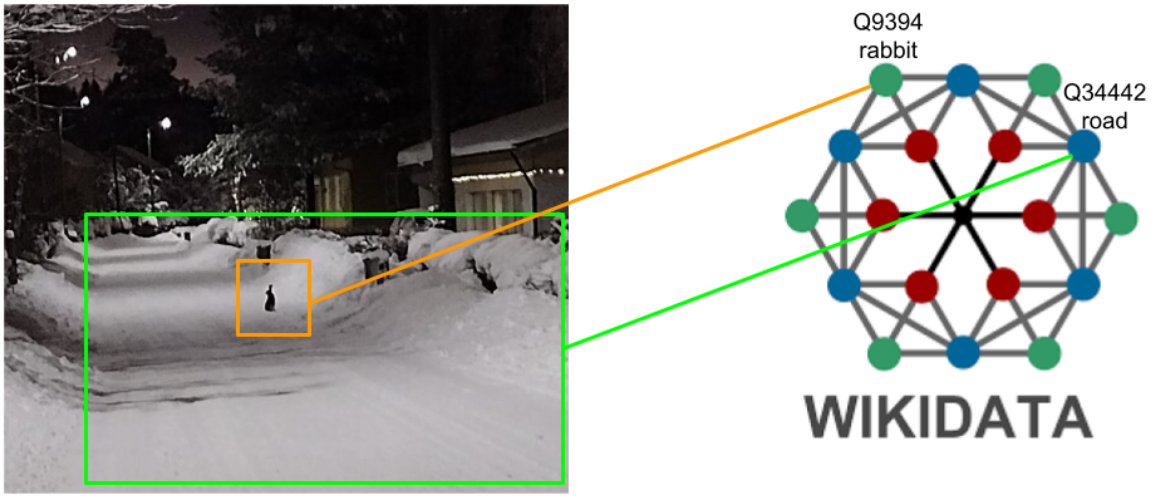

Figure 1: Visual Entity Linking, linking an object detected in an image to its corresponding entity in the knowledge graph (Wikidata).

We therefore propose a project to build a **novel dataset** based on the existing, community-annotated data on Wikimedia commons **for visual entity linking**, i.e., identifying entities in images and linking them to a knowledge graph. This data should be leveraged to identify and annotate unseen data, i.e., images that the community has not yet annotated. In this, we aim to contribute to the field of **deep learning based multimodal visual entity linking**. This automatically labelled data can be, after careful evaluation with input of the community, be **contributed back to Wikimedia Commons** to enable a central storage for freely-licensed images annotated with structured data. This data can be further used for research problems such as caption generation, question answering, and entity typing in KGs.

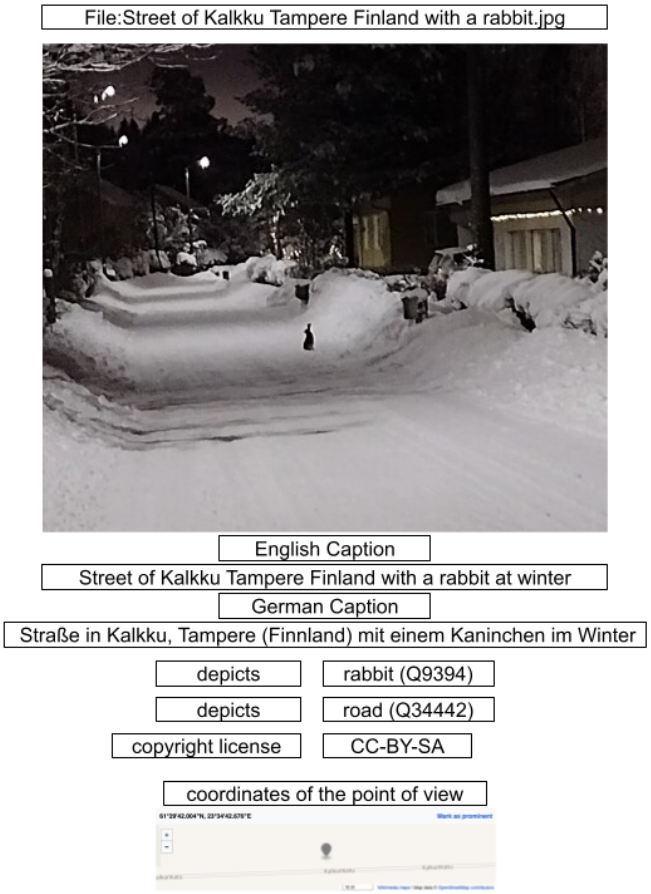

Figure 2: Simplified example of the data in Wikimedia Commons. The files have a name, multilingual captions, as well as structured data in the form of depicts*, copyright information, and location of the image contents.*

Project Approach

This project involves the following:

- Creating a novel dataset in the domain of visual entity linking
- Creating an approach for the challenges particular to the dataset created
- Deep learning based visual entity linking using the dataset created
- Evaluation of the approach with the Wikimedia Commons community
- Contributing the generated, evaluated data back to Wikimedia Commons

Project Outline

The project will follow a multi-step approach to address the research questions and achieve the project objectives. The project approach is somewhat divided into the following stages with iterations expected:

- Introductory Meeting: First discussions on the approaches, available data, literature
- Literature review: Throughout the entire project we emphasise the importance of scientific work, including basing our work in relevant literature
- Data collection: Collecting multi-modal data from Wikimedia Commons, including images, captions, structured data
- Preprocessing and analysis: The data will be analysed, answering questions around the quantity and quality of data
- Visual entity linking: In a multi-step approach we will explore different methods for entity linking, including object detection and linking as a pipeline, as well as different multi-modal model setups
- Analysis and results evaluation: The performance of the models will be evaluated and compared, leading to the best performing model
- Annotation of unseen data: We collect additional data in form of images and annotate those, leaving us with new annotations for the new images, as well as the new annotations for existing images where previous annotation might have not been exhaustive. These new annotations we aim to annotate with the community
- Community Annotation: Build an easy website for annotating our results, publicise it to the Wikimedia Commons community
- Contribute back new annotations: Build a small tool that imports the new annotations back to commons if enough annotators agreed on their accuracy
- Discussion and conclusion: The project will conclude with a discussion of the key findings and results, and a summary of the contributions and achievements of the project and how they can benefit the community of Wikimedia at large. (We aim at publishing the results in a top-tier conference venue.)

This is a general outline, and the actual project approach may be adjusted depending on the specific research question and methods addressed by the students.

To summarize, some of the research questions for the project include:

- How can we efficiently and accurately link images to entities in a knowledge graph depicted in the image?
- What is the quality of human-annotated images in terms of the coverage of its corresponding entity links to Wikidata?
- Can we contribute that data back to the community?

Long Term Questions

- Can this data be used for multilingual caption-generation?
- Can we leverage the structured data for visual question answering?
- What other data formats can support modelling information with knowledge graphs?
- Can this Image data improve the entity embeddings generated from the multimodal knowledge graph embedding models?
- Can we predict the missing semantic types of the entities from the images in the dataset?

Project Road Map

1. Dataset creation
    a. Extraction of image, caption, structured data, license information, categories from Wikimedia commons
    b. Linking to DBpedia if possible
    c. Analysis of data quantity and quality
    d. Extraction of ImageNet to Wikidata linking dataset
2. Objection detection
    a. Detect objects in test
    b. Evaluate performance
    c. Experiment with existing visual entity linking datasets
3. Visual entity linking
    a. Link objects detected to Wikidata
    b. Experiments with multi-modal models
4. Automatic evaluation
5. Community evaluation
    a. Build interface for annotation
    b. Invite community to annotate structured data
6. Contribute annotated data back to Wikimedia Commons

Related Work

Zheng, Qiushuo, et al. "Visual entity linking via multi-modal learning." Data Intelligence 4.1 (2022): 1-19. https://direct.mit.edu/dint/article/4/1/1/108470/Visual-Entity-Linking-via-Multi-modal-Learning

Gan, Jingru, et al. "Multimodal entity linking: a new dataset and a baseline." Proceedings of the 29th ACM International Conference on Multimedia. 2021. https://dl.acm.org/doi/abs/10.1145/3474085.3475400

Contact

lucie-aimee.kaffee@hpi.de, russa.biswas@hpi.de, gerard.demelo@hpi.de
This project will be supervised by the HPI Chair for Artificial Intelligence and Intelligent Systems (Dr. Lucie-Aimée Kaffee, Dr. Russa Biswas, Prof. Dr. Gerard de Melo).