KI DATA TOOLING
The Data Kit for Automotive AI

Focus on the Challenges: Analysis of a User-friendly Data Search Approach with CLIP in the Automotive Domain

Philipp Rigoll, Patrick Petersen,
Hanno Stage, Lennart Ries, Eric Sax | FZI

Data Storage, Analysis & Discoverability

Semantic search for images

The processing of large image data sets has become a key competence in the development of driver assistance systems. In this context, it is important to have an automatable and versatile approach to search these data sets. We analyze the use of a state-of-the-art neural network to embed text and images in a vector space. This approach enables semantic searchability with an understandable text-based description. In our experiments, we show how the method can be used in the development of driver assistance systems.

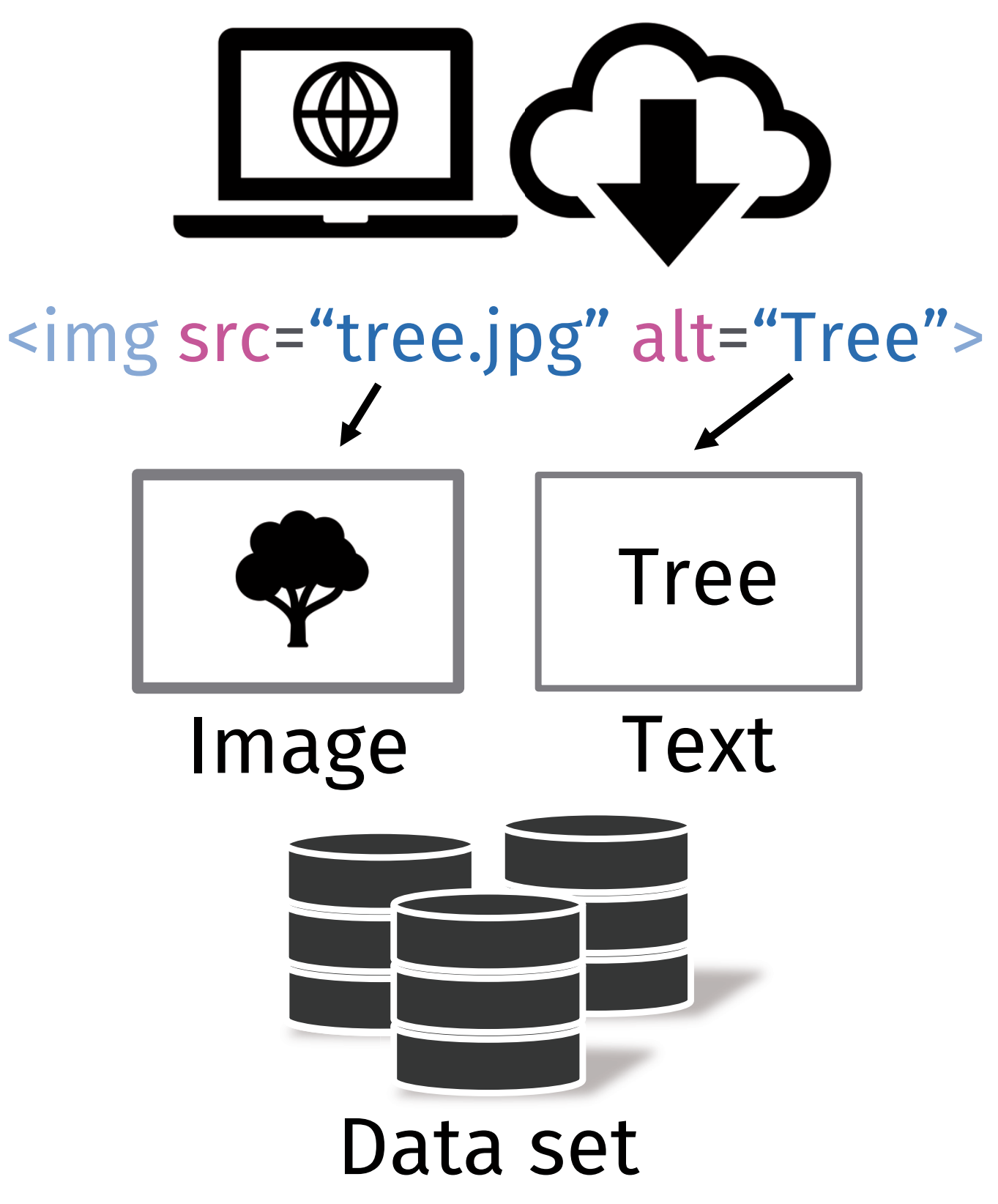

Figure 1: Creation of the CLIP [2] data set

CLIP

We investigate whether CLIP [2], a state-of-the-art network, can be used to solve this problem. CLIP is able to represent images and texts as vectors. CLIP was trained in such a way that similar vectors represent semantically similar images and texts. During the search, images are queried with text or image input. The corresponding CLIP encoder transfers the input into a vector representation. The search is then performed by finding similar image vectors in an initially built vector database.

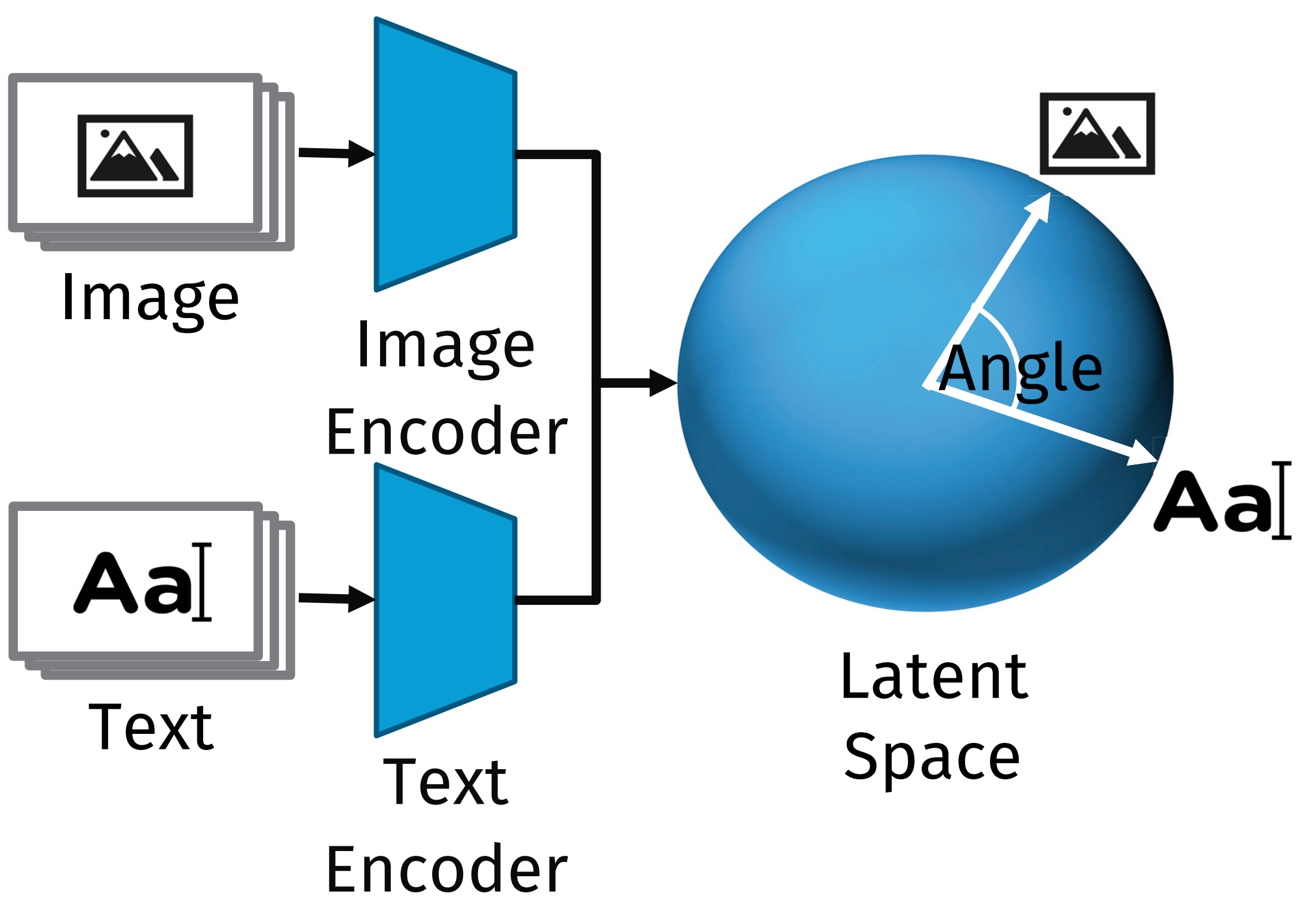

Figure 2: Training of CLIP [2] with a joint latent space for images and text descriptions

Search options

With this method it is possible to search for similar images. But it is also possible to search for situational knowledge or partial aspects of an image by creating a textural prompt. This process can be supported by WordNet [4] where interrelated words are connected in a graph.

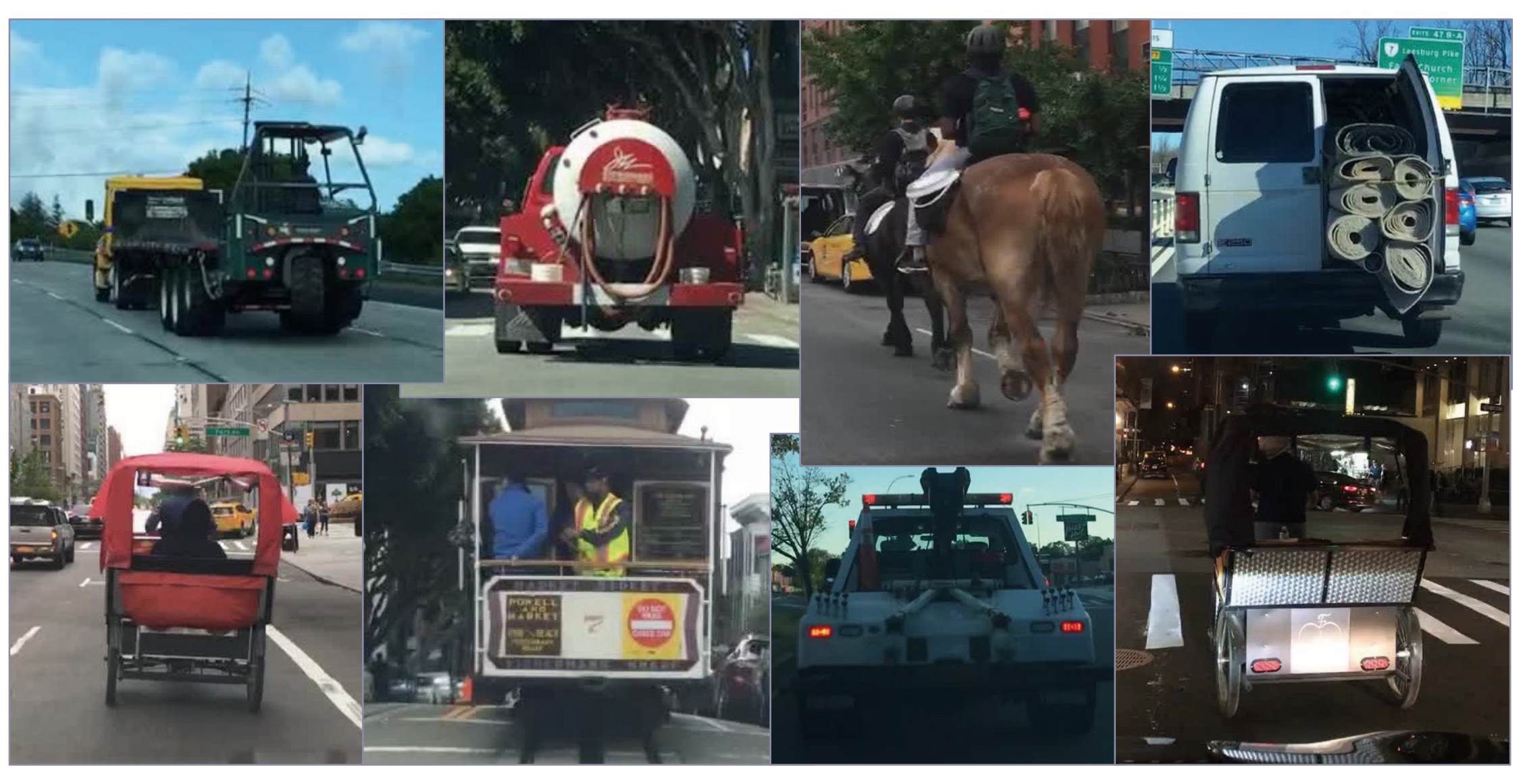

Figure 3: Selection of images found in the BDD100k data set [3] with our method when starting with the word carriage and searching for linked words in WordNet [4] [1]

References:

[1] Rigoll, Philipp, et al. "Focus on the Challenges: Analysis of a User-friendly Data Search Approach with CLIP in the Automotive Domain." arXiv preprint arXiv:2304.10247 (2023).
[2] Radford, Alec, et al. "Learning transferable visual models from natural language supervision." International conference on machine learning. PMLR, 2021.
[3] Yu, Fisher, et al. "Bdd100k: A diverse driving dataset for heterogeneous multitask learning." Proceedings of the IEEE/CVF conference on computer vision and pattern recognition. 2020.
[4] Miller, George A. "WordNet: a lexical database for English." *Communications of the ACM* 38.11 (1995): 39-41.

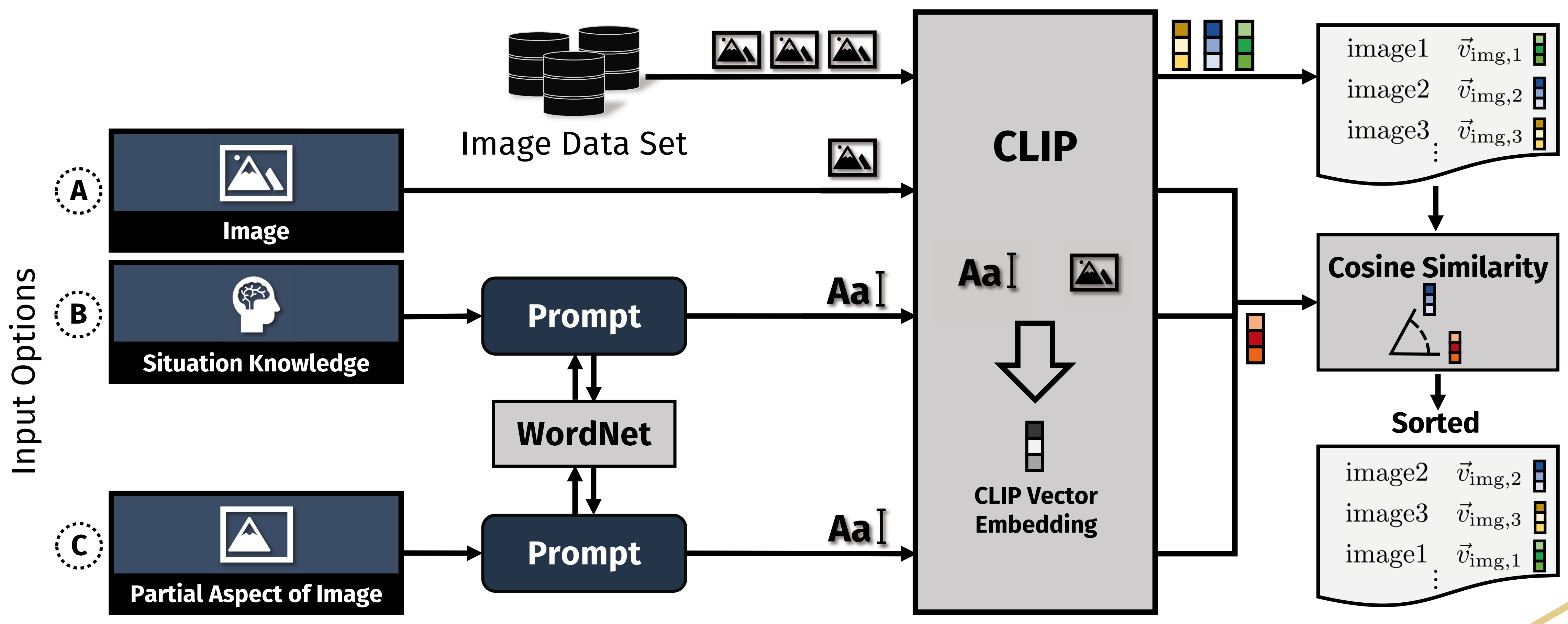
Image Data Set
CLIP
image1 $\vec{v}_{\text{img},1}$
image2 $\vec{v}_{\text{img},2}$
image3 $\vec{v}_{\text{img},3}$
Input Options
A
Image
B
Situation Knowledge
C
Partial Aspect of Image
Prompt
WordNet
Prompt
Aa
Aa
Aa
CLIP Vector Embedding
Cosine Similarity
Sorted
image2 $\vec{v}_{\text{img},2}$
image3 $\vec{v}_{\text{img},3}$
image1 $\vec{v}_{\text{img},1}$

Figure 4: Overview of the proposed method for textual and image-based search of similar images [1]

Partners

BMW GROUP
BOSCH
Continental
Valeo
ZF
Ansys
AVL
dSPACE
DLR
FKFS RESEARCH IN MOTION.
FZI
BERGISCHE UNIVERSITÄT WUPPERTAL
TH Aschaffenburg university of applied sciences
Technische Universität Braunschweig
Technische Universität München TUM
UNIKASSEL VERSITÄT
UNIVERSITÄT PASSAU

External partners

BIT TECHNOLOGY SOLUTIONS
MACKEVISION
ejct

For more information contact:

philipp.rigoll@fzi.de

KI Data Tooling is a project of the KI Familie. It was initiated and developed by the VDA Leitinitiative autonomous and connected driving and is funded by the Federal Ministry for Economic Affairs and Climate Action.

KI FAMILIE

Supported by:
Federal Ministry for Economic Affairs and Climate Action
on the basis of a decision by the German Bundestag

www.ki-datatooling.de   @KI_Familie   KI Familie