VISTA: Visual-Textual Knowledge Graph Representation Learning

Jaejun Lee, Chanyoung Chung, Hochang Lee,
Sungho Jo, and Joyce Jiyoung Whang*
School of Computing, KAIST

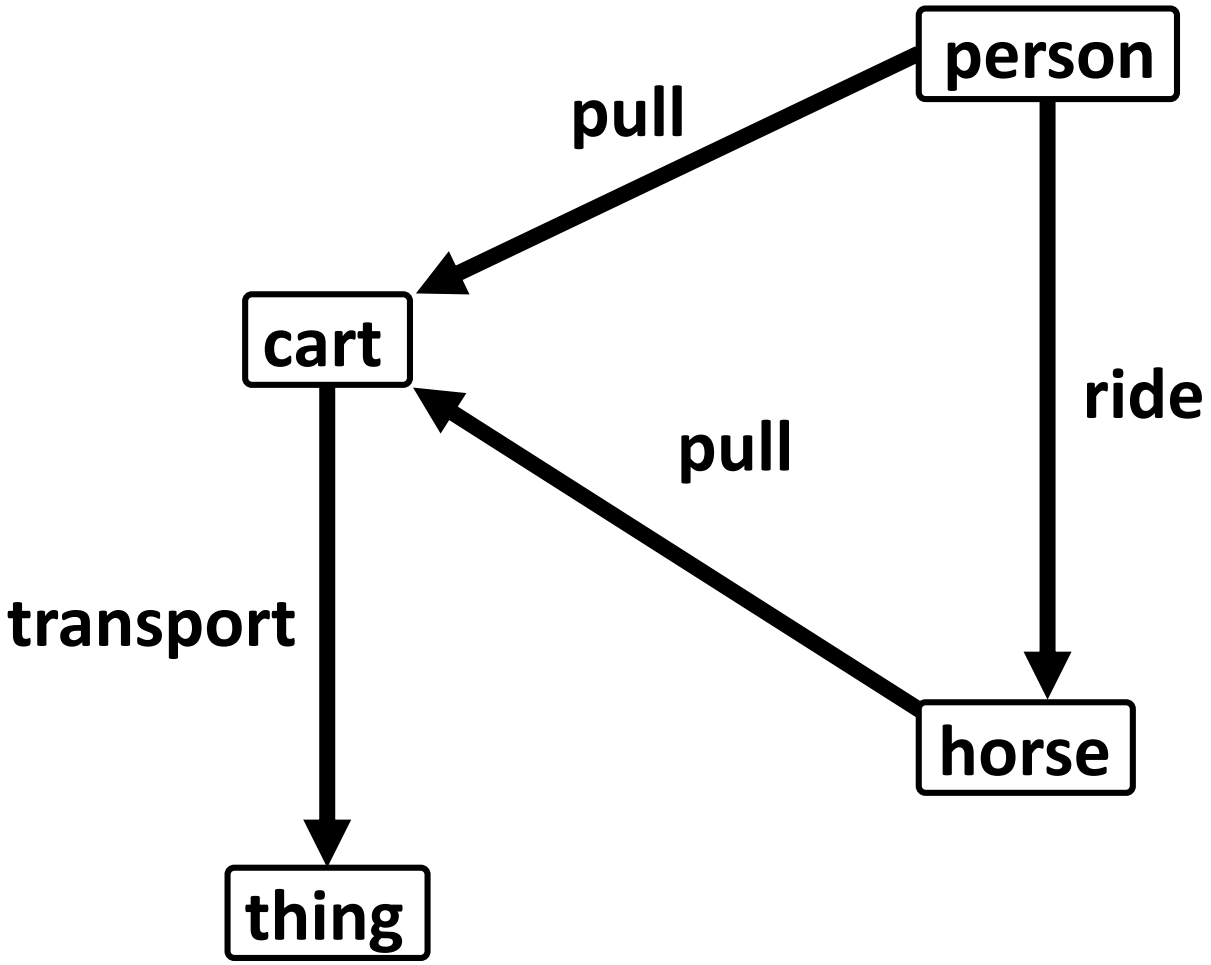
* Corresponding Author

The 2023 Conference on Empirical Methods in Natural Language Processing (EMNLP 2023)

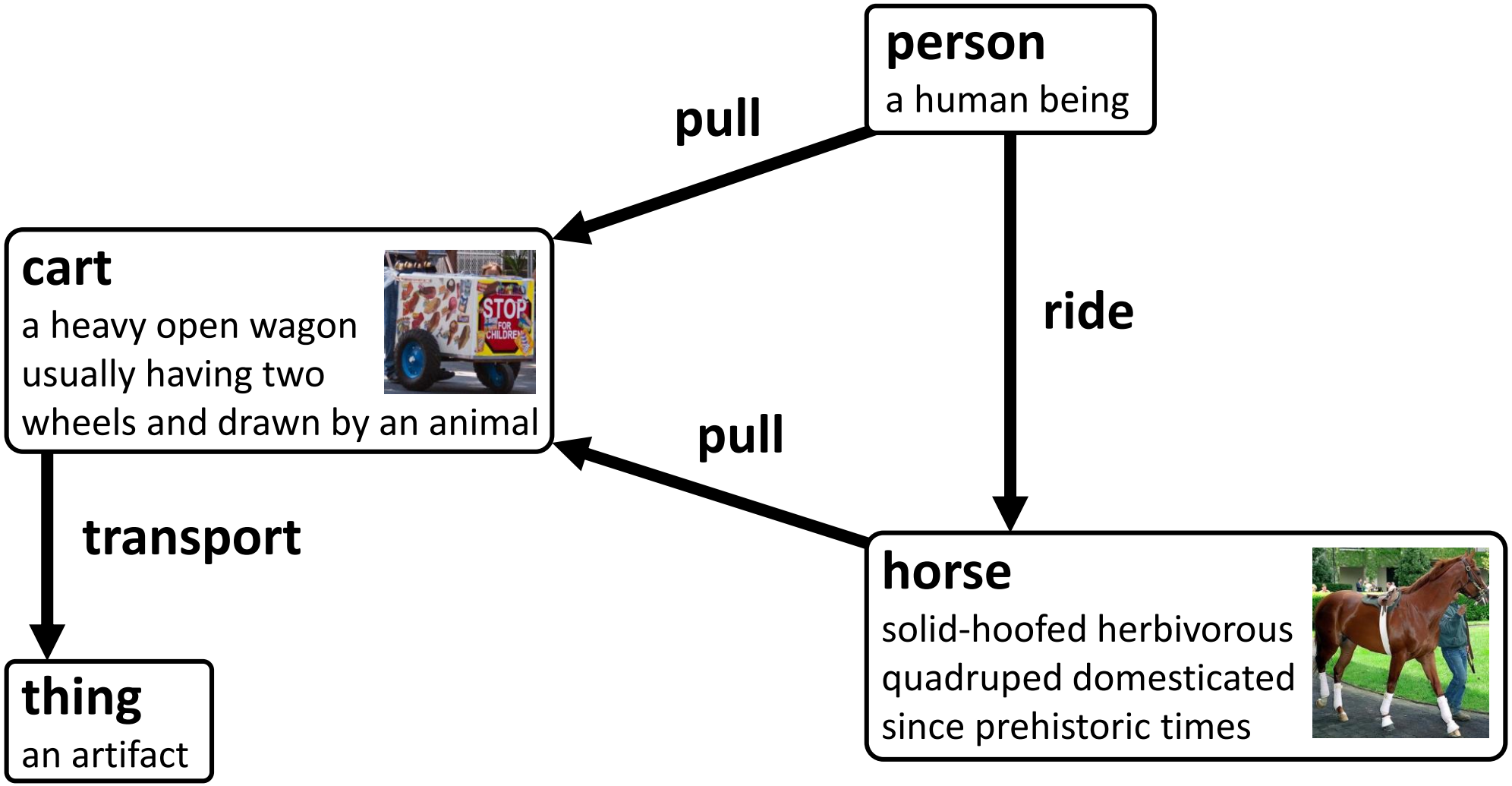
KAIST

KAIST BDILab
BIG DATA INTELLIGENCE

Knowledge Graphs



Multimodal Knowledge Graphs



Visual-Textual Knowledge Graphs (VTKGs)

pull
cause to move by pulling

person
a human being

cart
a heavy open wagon usually having two wheels and drawn by an animal

pull
cause to move by pulling

ride
sit and travel on the back of animal, usually while controlling its motions

transport
move something or somebody around; usually over long distances

thing
an artifact

horse
solid-hoofed herbivorous quadruped domesticated since prehistoric times

Link Prediction on VTKGs

pull
cause to move by pulling

person
a human being

cart
a heavy open wagon usually having two wheels and drawn by an animal

⟨horse, pull, ?⟩

pull
cause to move by pulling

ride
sit and travel on the back of animal, usually while controlling its motions

transport
move something or somebody around; usually over long distances

horse
solid-hoofed herbivorous quadruped domesticated since prehistoric times

thing
an artifact

pull

Contributions

- Define **V**isual-**T**extual **K**nowledge **G**raphs **(VTKGs)**
  - Create two real-world datasets: **VTKG-C** and **VTKG-I**

- Propose **VIS**ual-**T**extu**A**l (**VISTA**) knowledge graph representation learning method
  - VISTA utilizes the **visual and textual features of relations and entities**
  - Define an entity encoder, a relation encoder, and a triplet decoder

- VISTA outperforms **10 different** state-of-the-art knowledge graph completion methods, including multimodal knowledge graph representation learning methods

Creating Real-World VTKGs

VRD



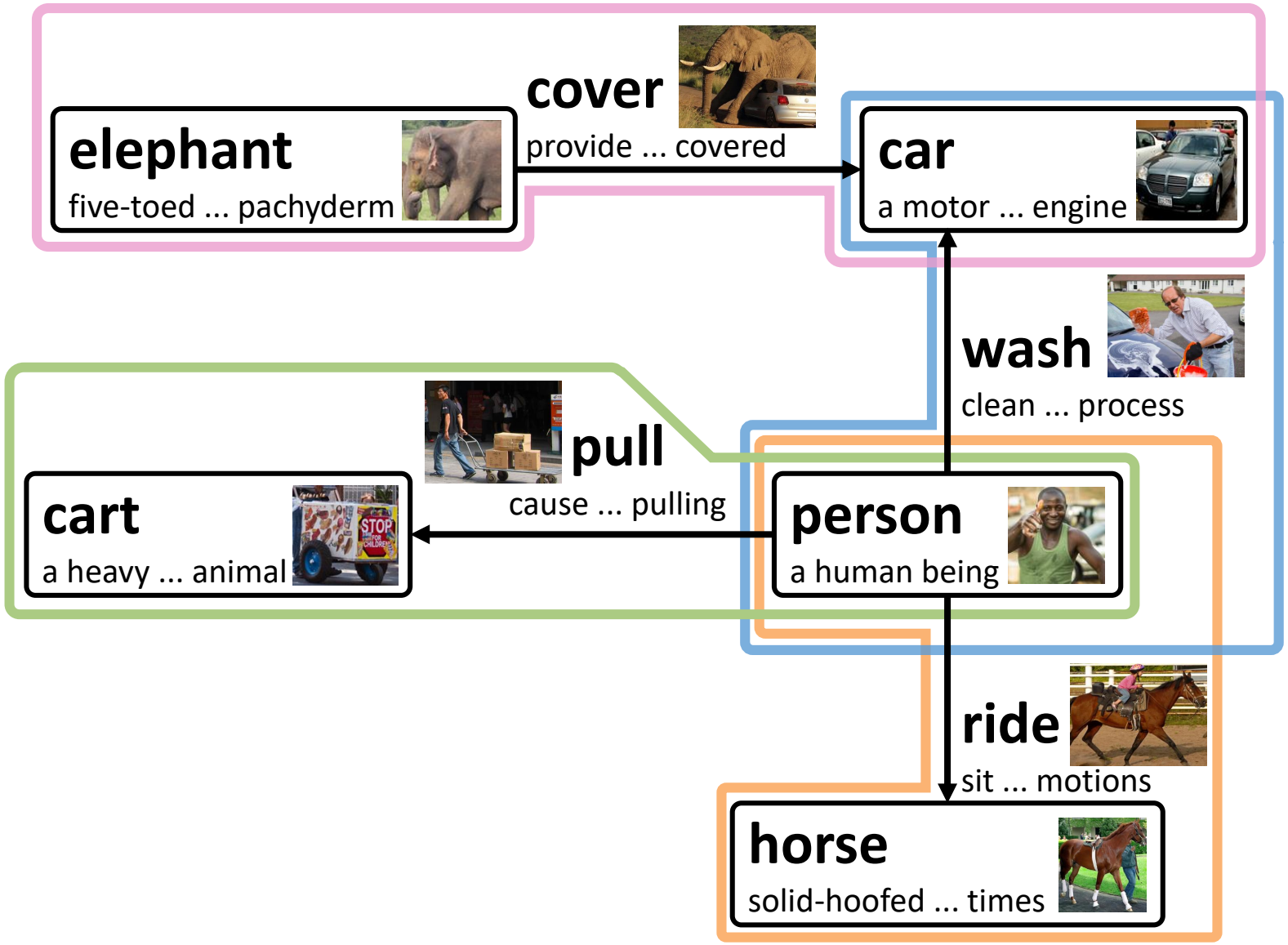
HICO-DET



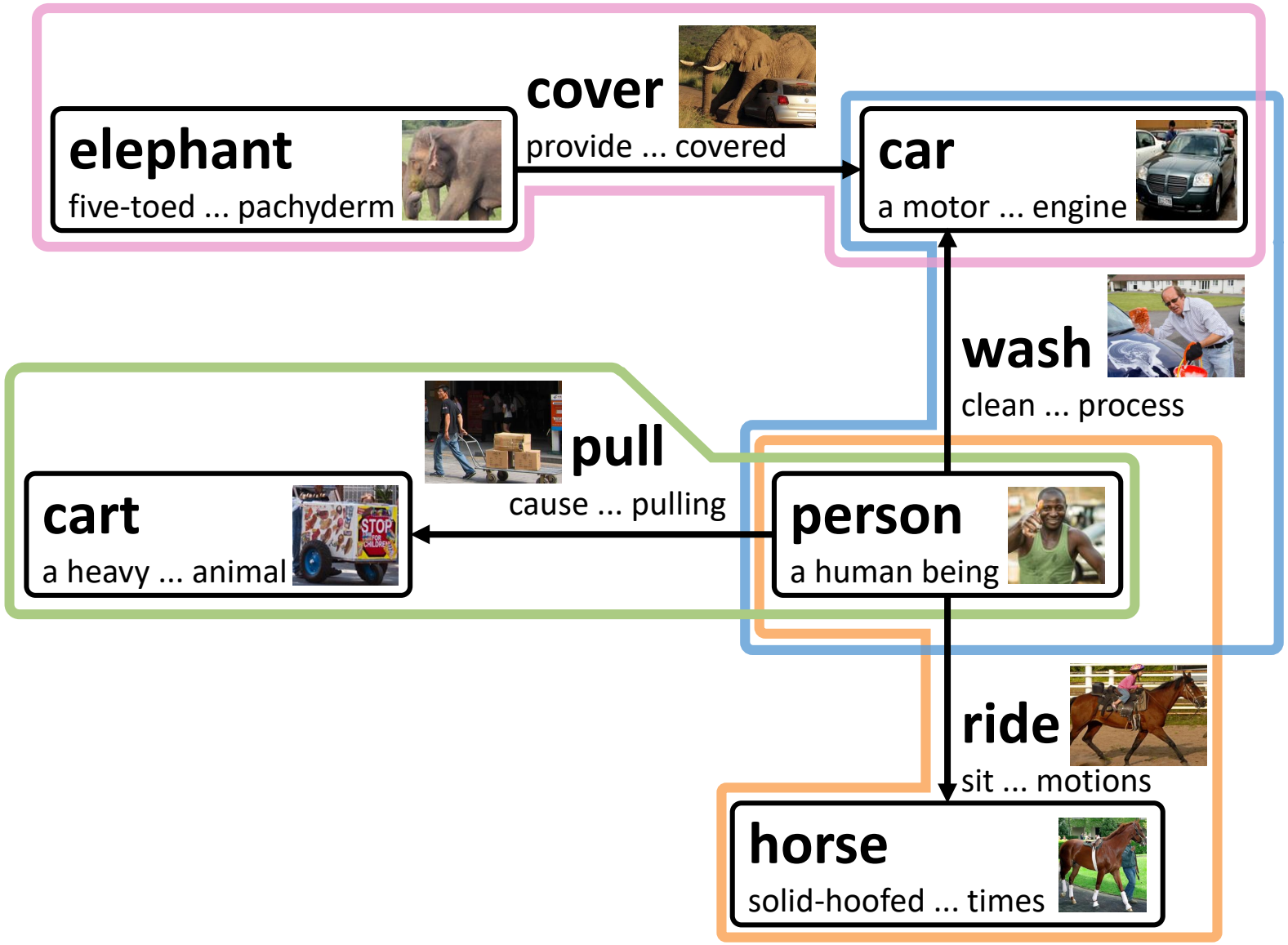
UnRel

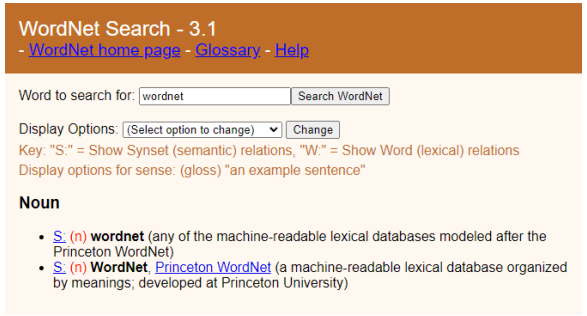


Creating Real-World VTKGs: VTKG-I



Creating Real-World VTKGs: VTKG-C





WordNet

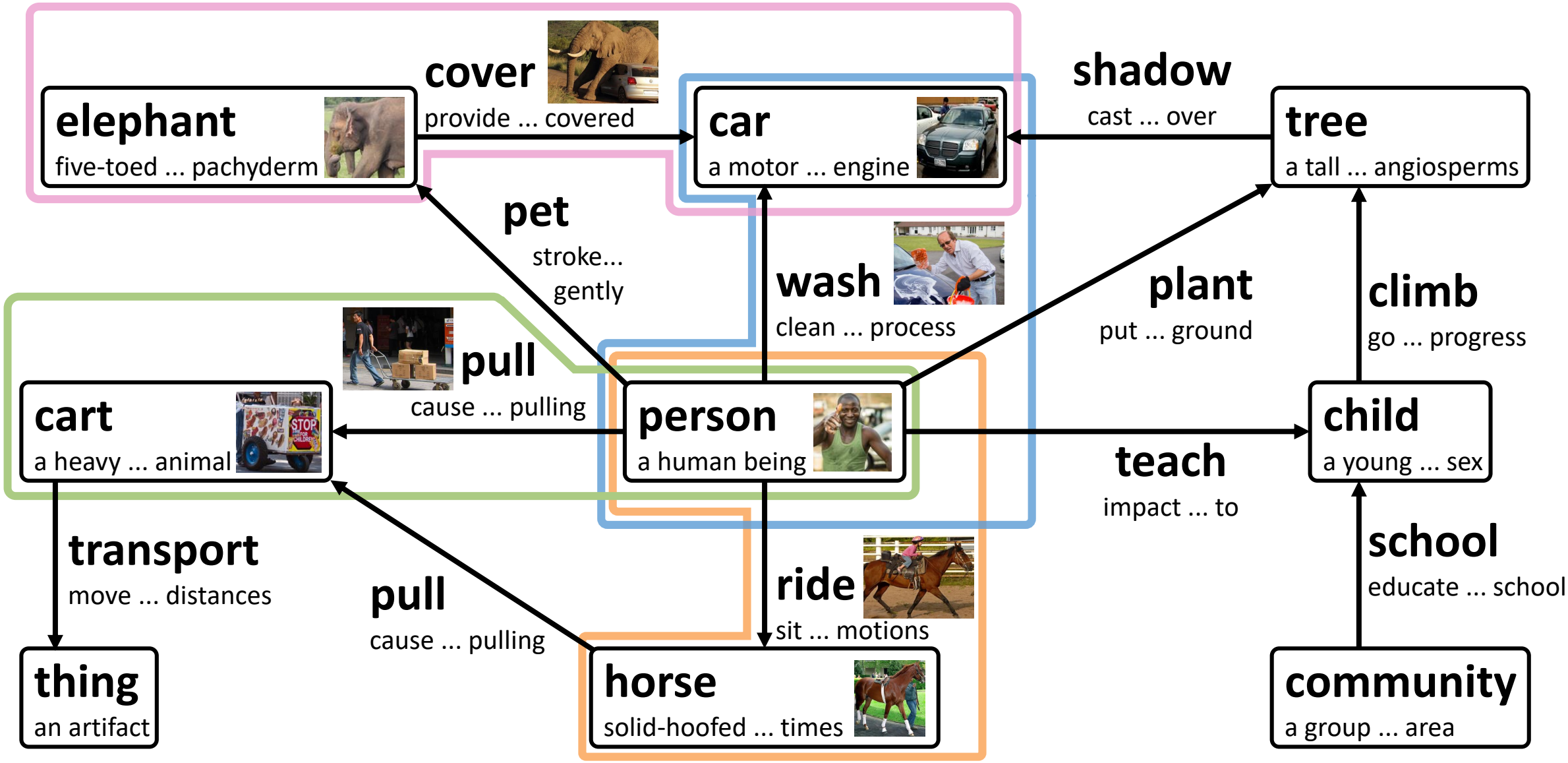
ConceptNet
An open, multilingual knowledge graph

ConceptNet

VisKE

VisKE

Creating Real-World VTKGs: VTKG-C



Extracting Visual and Textual Features of Entities

Visual Features of horse

ViT-Base → $\mathbf{h}_{\text{vis},1}$

ViT-Base → $\mathbf{h}_{\text{vis},2}$

ViT-Base → $\mathbf{h}_{\text{vis},3}$

ViT-Base → $\mathbf{h}_{\text{vis},4}$

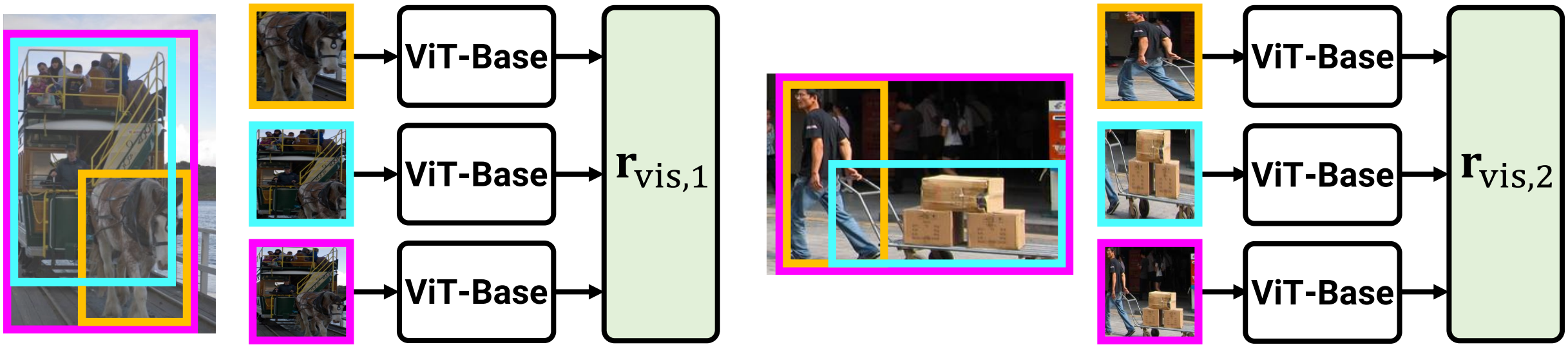
Textual Feature of horse

solid-hoofed herbivorous quadruped
domesticated since prehistoric times

→ $\text{BERT}_{\text{BASE}}$ → $\mathbf{h}_{\text{txt}}$

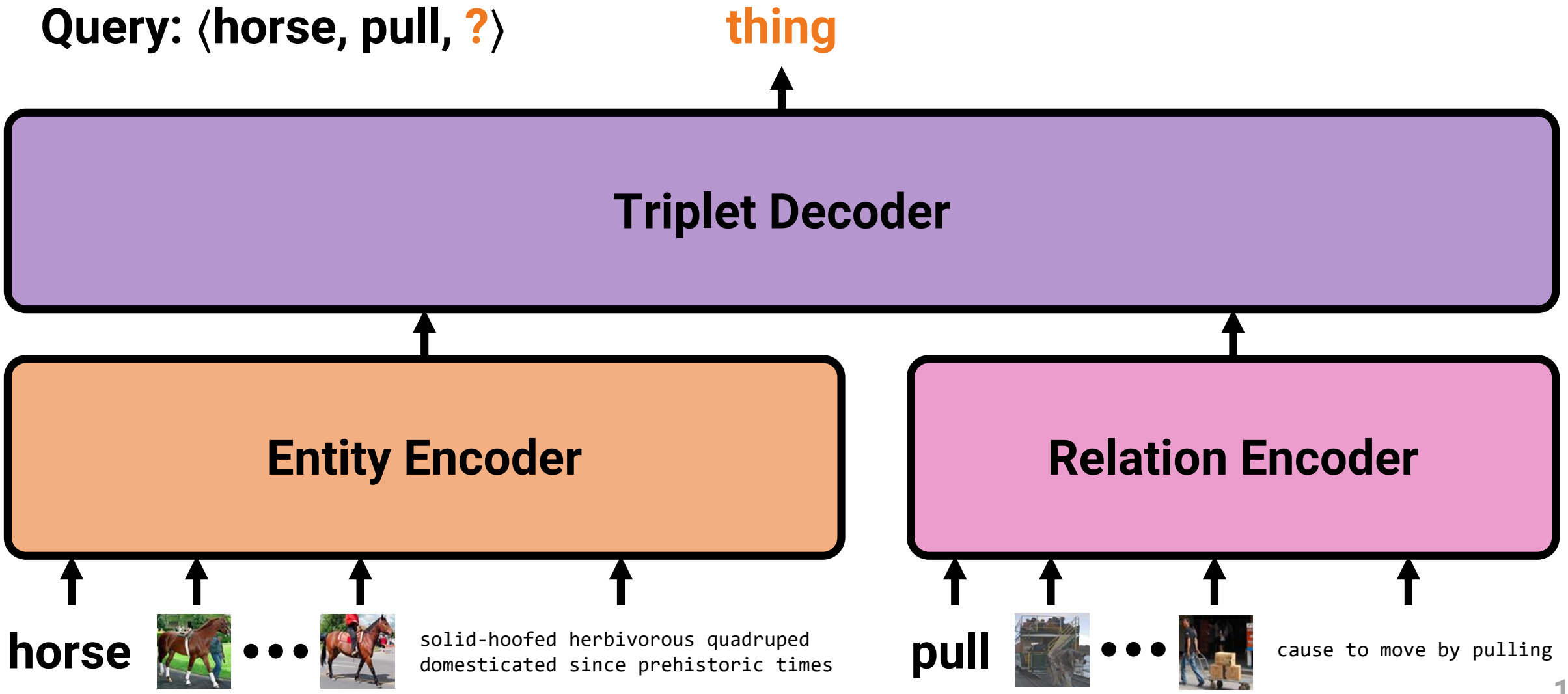
Extracting Visual and Textual Features of Relations

Visual Features of pull

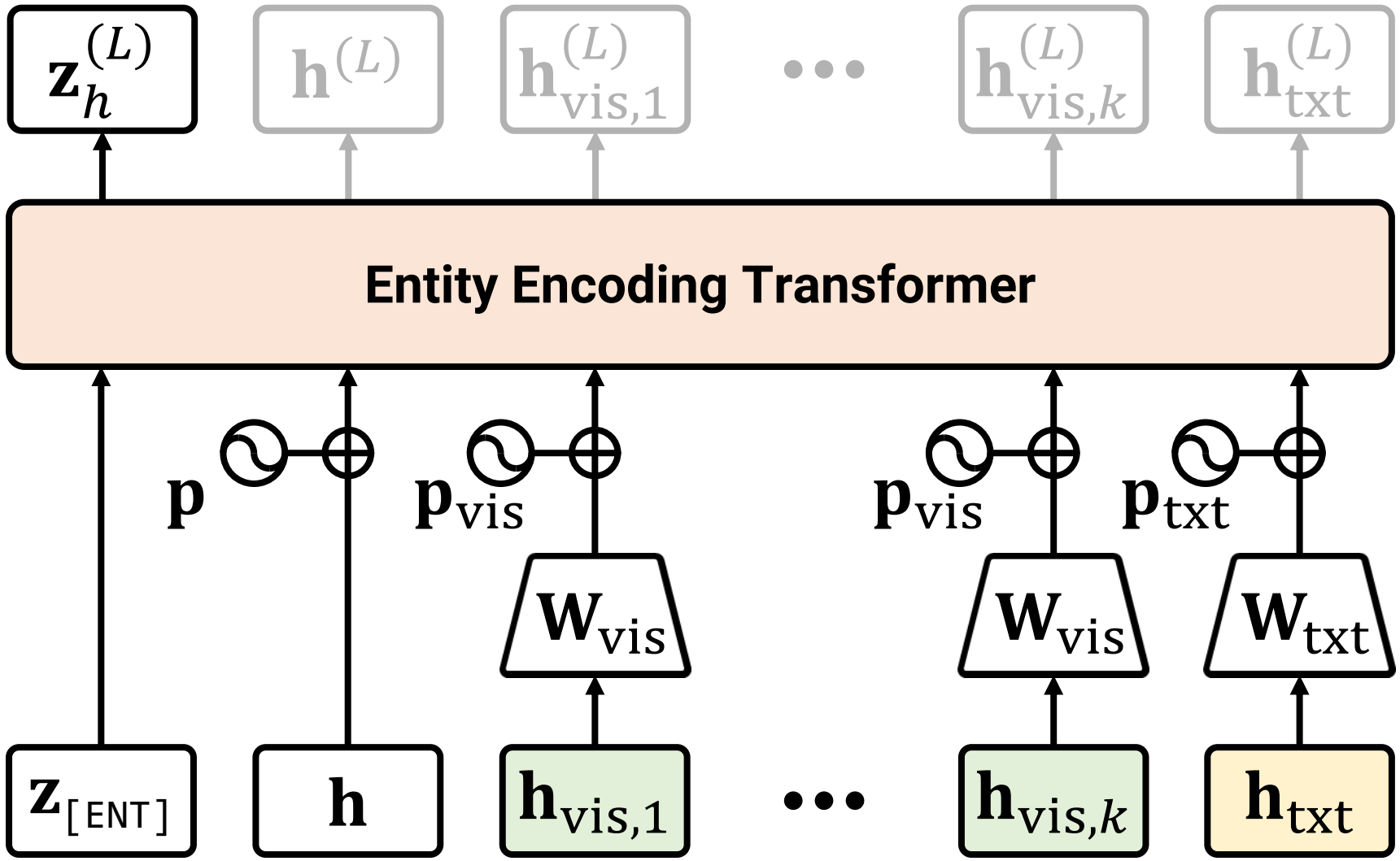


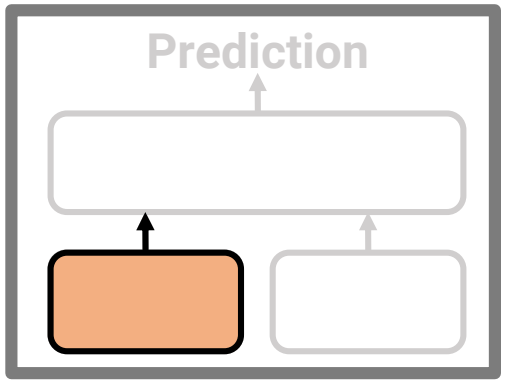
Textual Feature of pull

Overview of VISTA

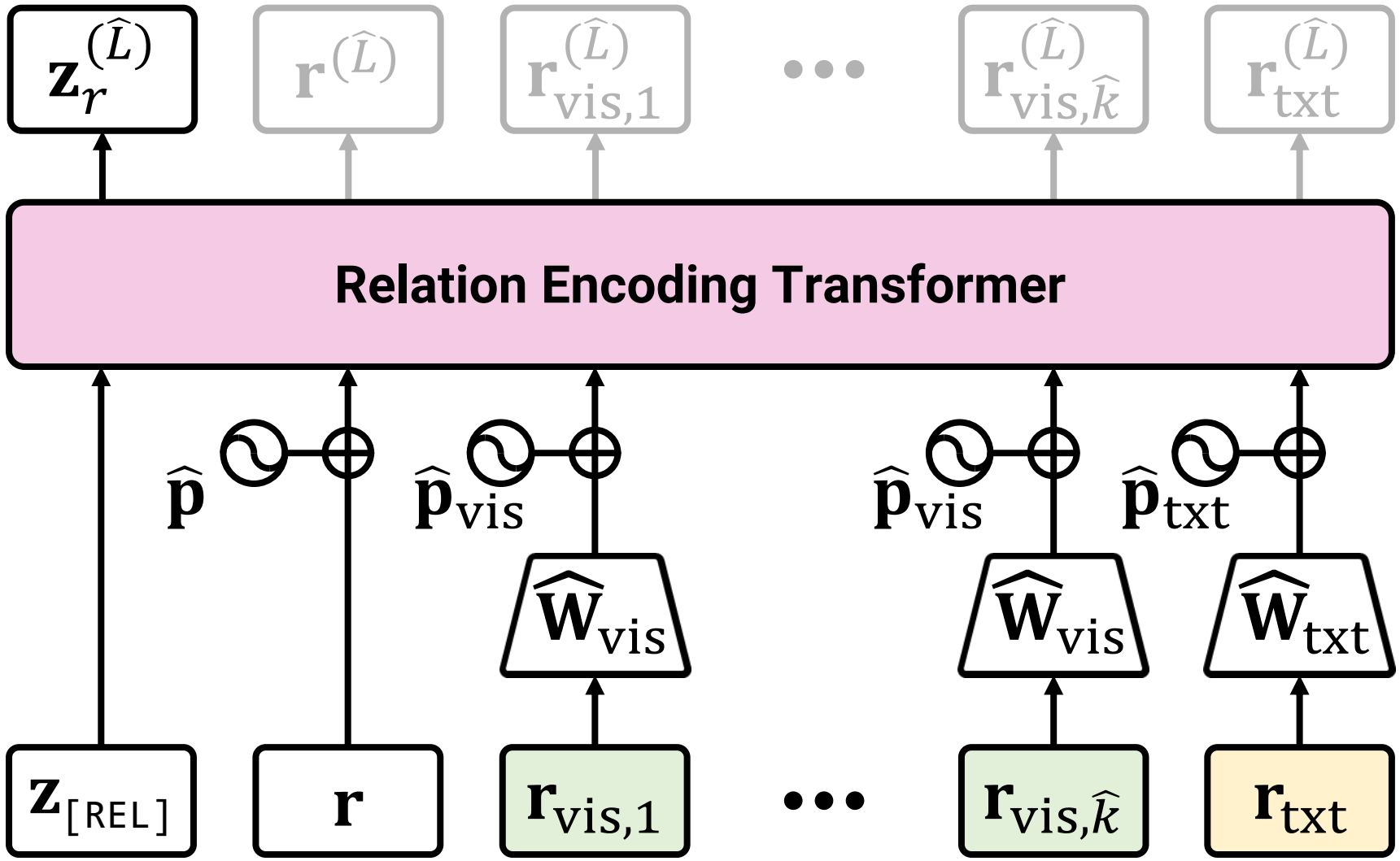


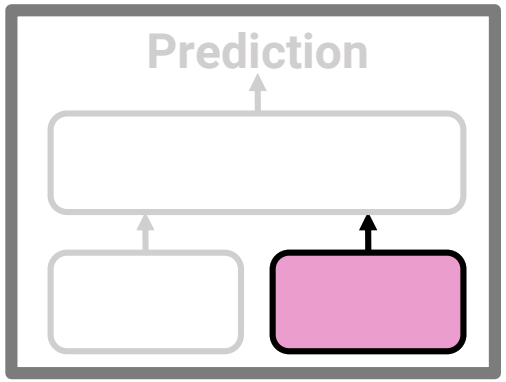
Entity Encoder



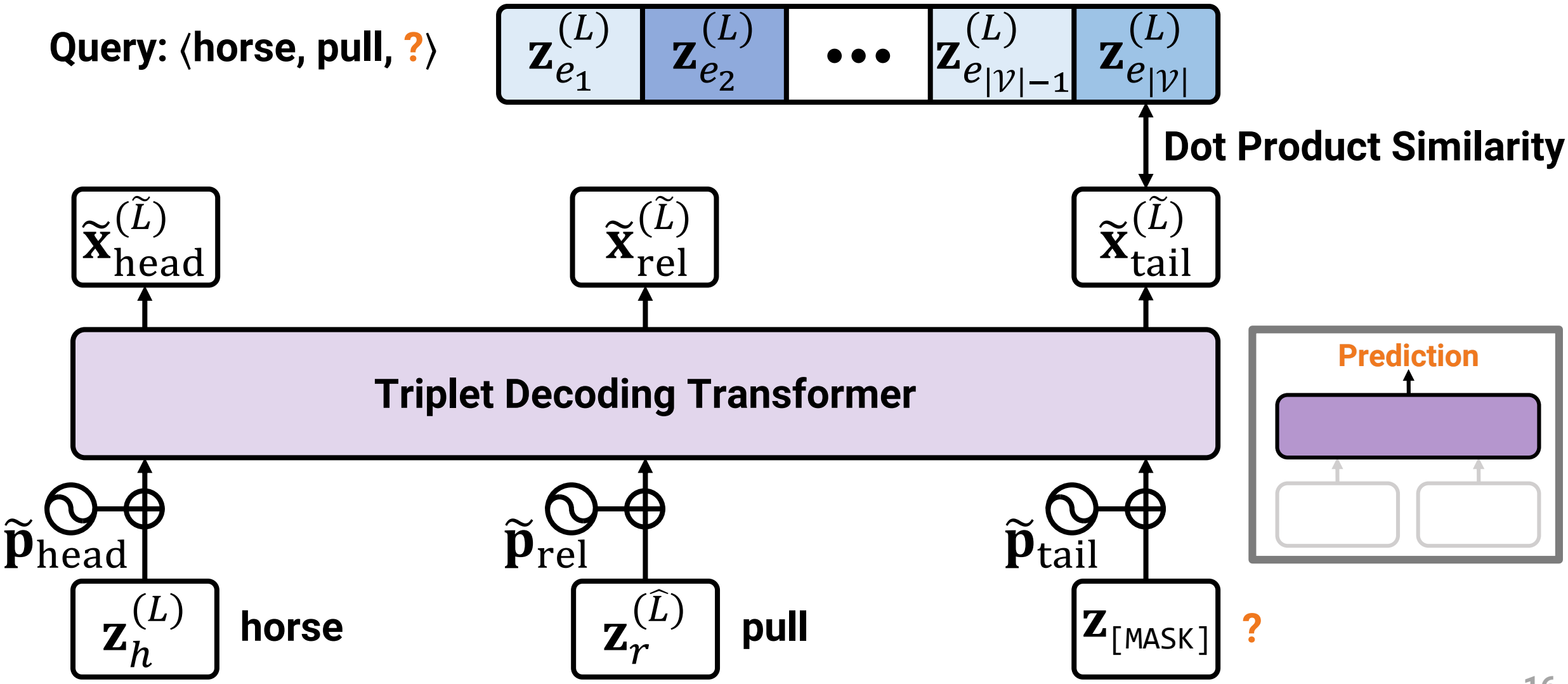


Relation Encoder





Triplet Decoder



Experiments

- Datasets
  - Create two **Visual-Textual Knowledge Graphs (VTKGs)**
    - VTKG-I, VTKG-C
  - Two Benchmark Multimodal Knowledge Graphs
    - WN18RR++ (WN18RR with corrections), FB15K237

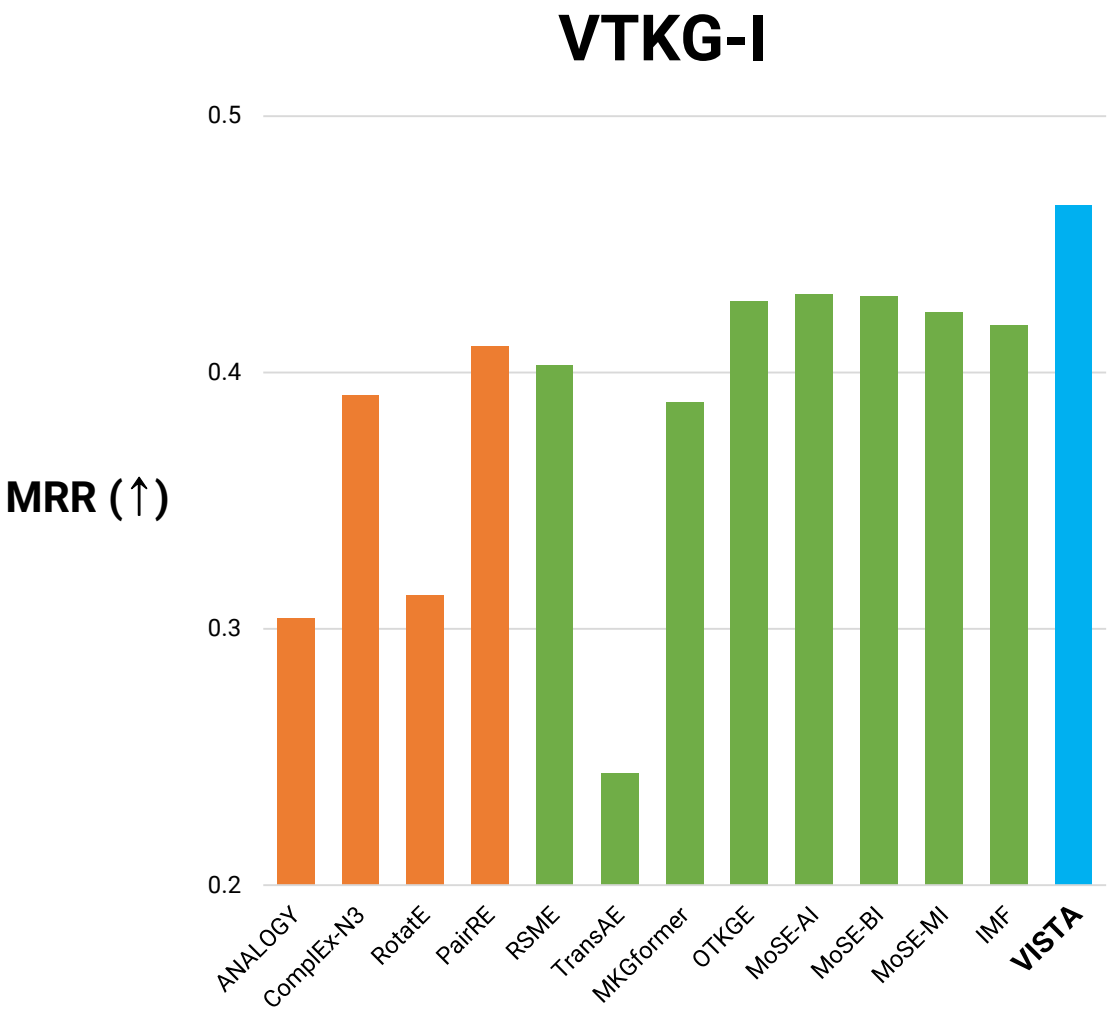
No. of Images ↓ No. of Text Descriptions ↙

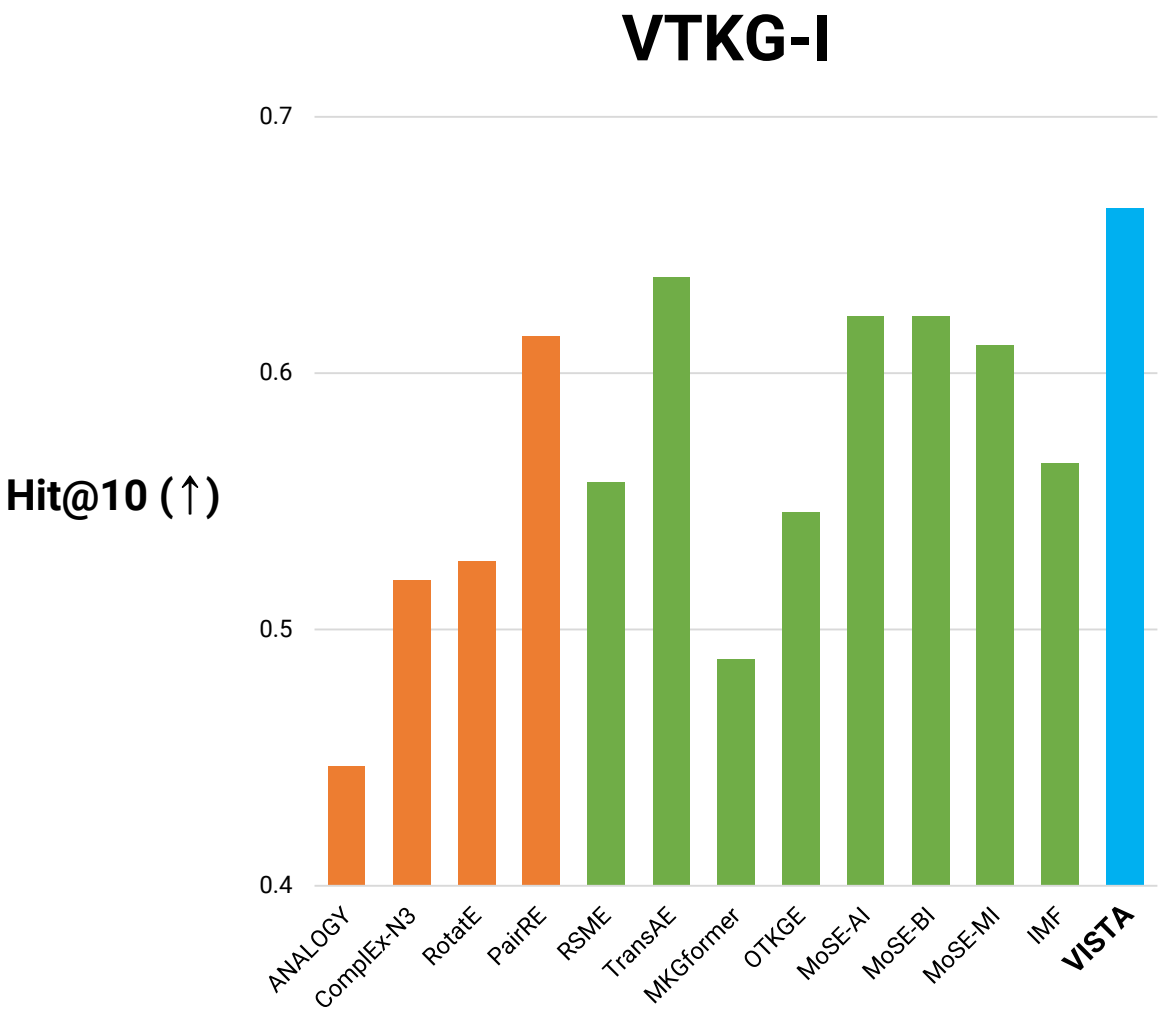
	$\lvert\mathcal{V}\rvert$	$\lvert\mathcal{R}\rvert$	$\lvert\mathcal{T}\rvert$	$\lvert\mathcal{I}\rvert$	$\lvert\mathcal{D}\rvert$
VTKG-I	181	217	1,316	390,658	383
VTKG-C	43,267	2,731	111,491	461,007	45,401
WN18RR++	41,105	11	93,003	70,349	41,105
FB15K237	14,541	237	310,116	145,944	14,515

Experiments

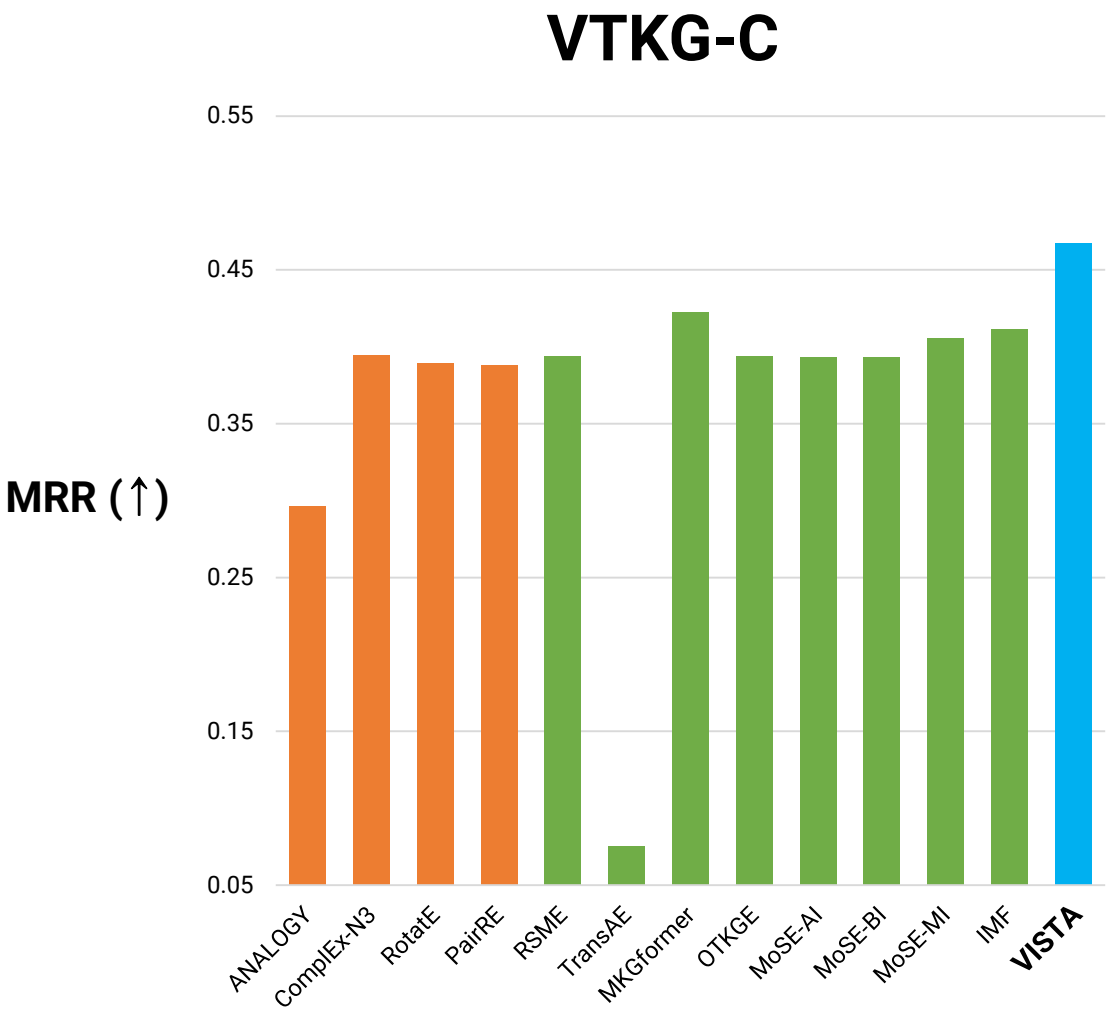
- Comparison with **10 baseline methods**
  - Knowledge Graph Embedding Methods
    - ANALOGY (ICML 2017)
    - ComplEx-N3 (ICML 2018)
    - RotatE (ICLR 2019)
    - PairRE (ACL 2021)
  - Multimodal Knowledge Graph Representation Learning Methods
    - RSME (MM 2021)
    - TransAE (IJCNN 2019)
    - MKGformer (SIGIR 2022)
    - OTKGE (NeurIPS 2022)
    - MoSE (EMNLP 2022)
    - IMF (TheWebConf 2023)

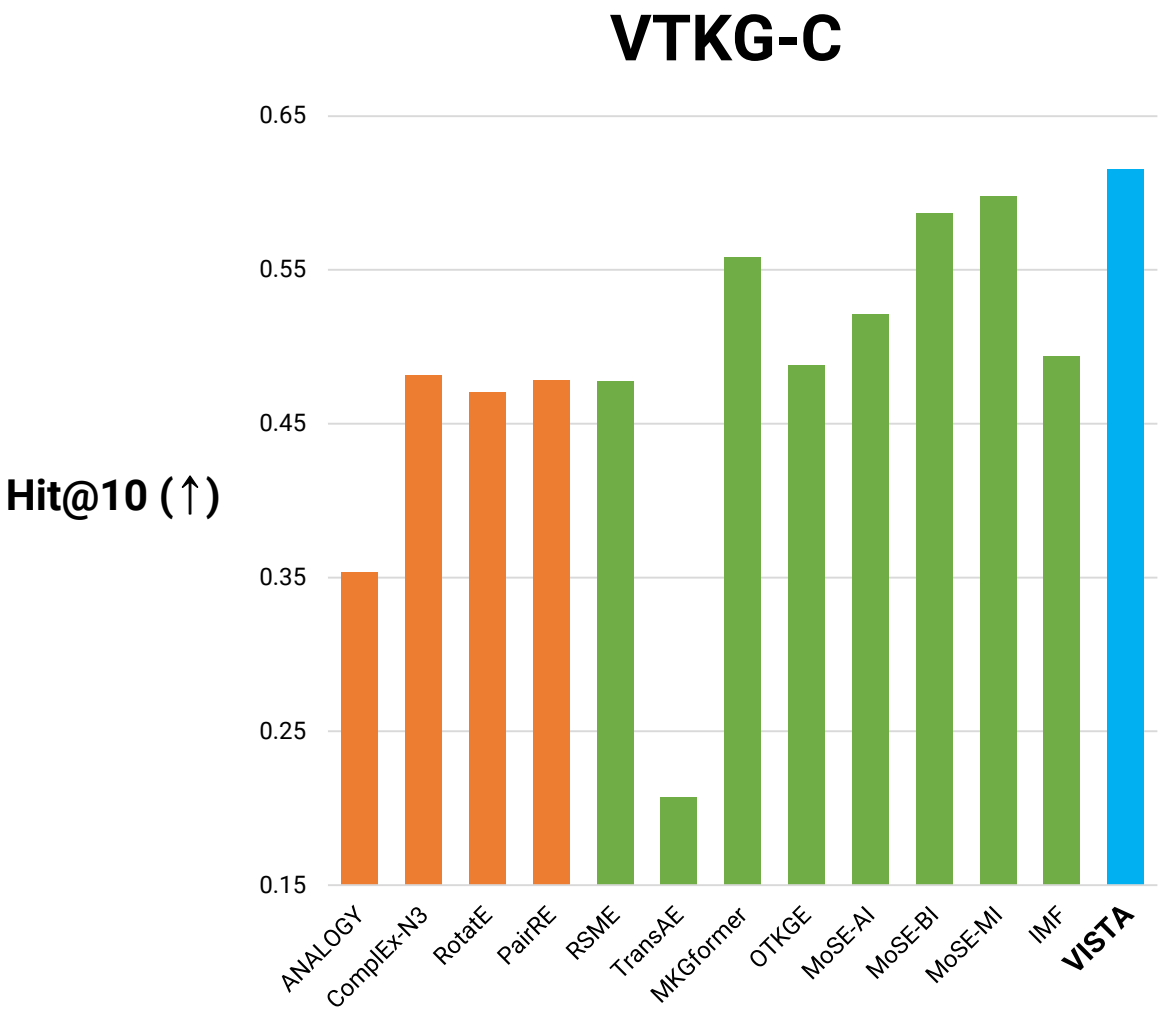
Knowledge Graph Completion Performance



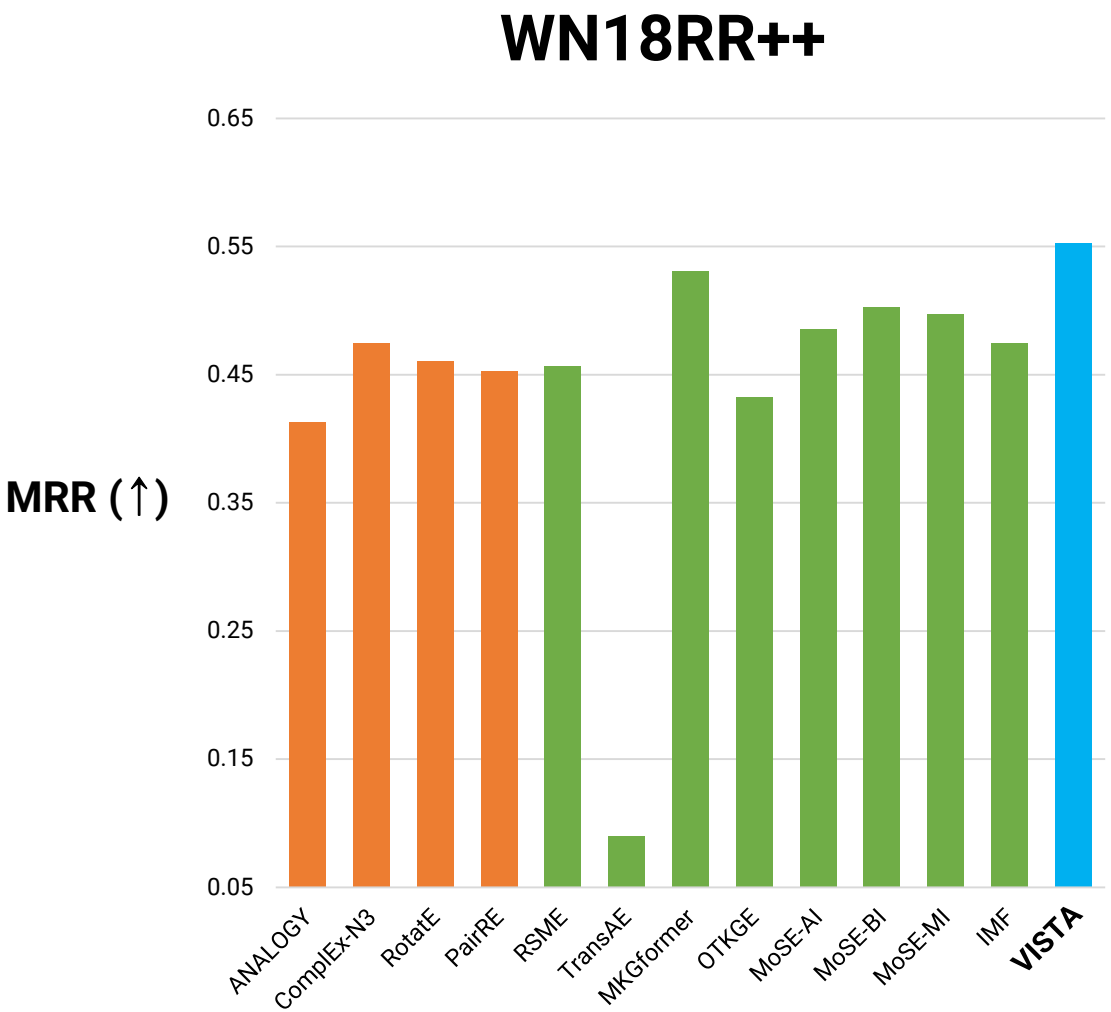


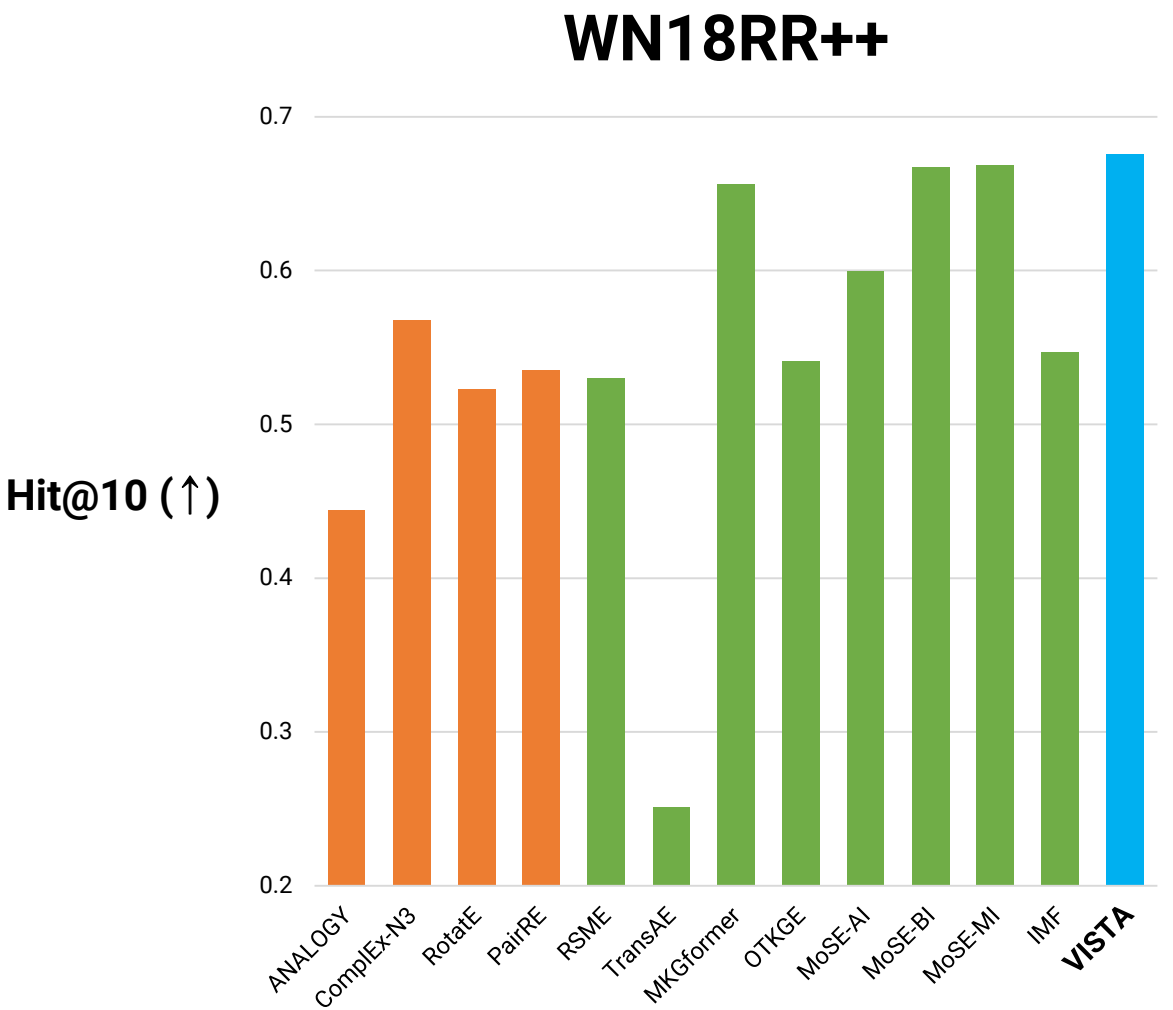
Knowledge Graph Completion Performance



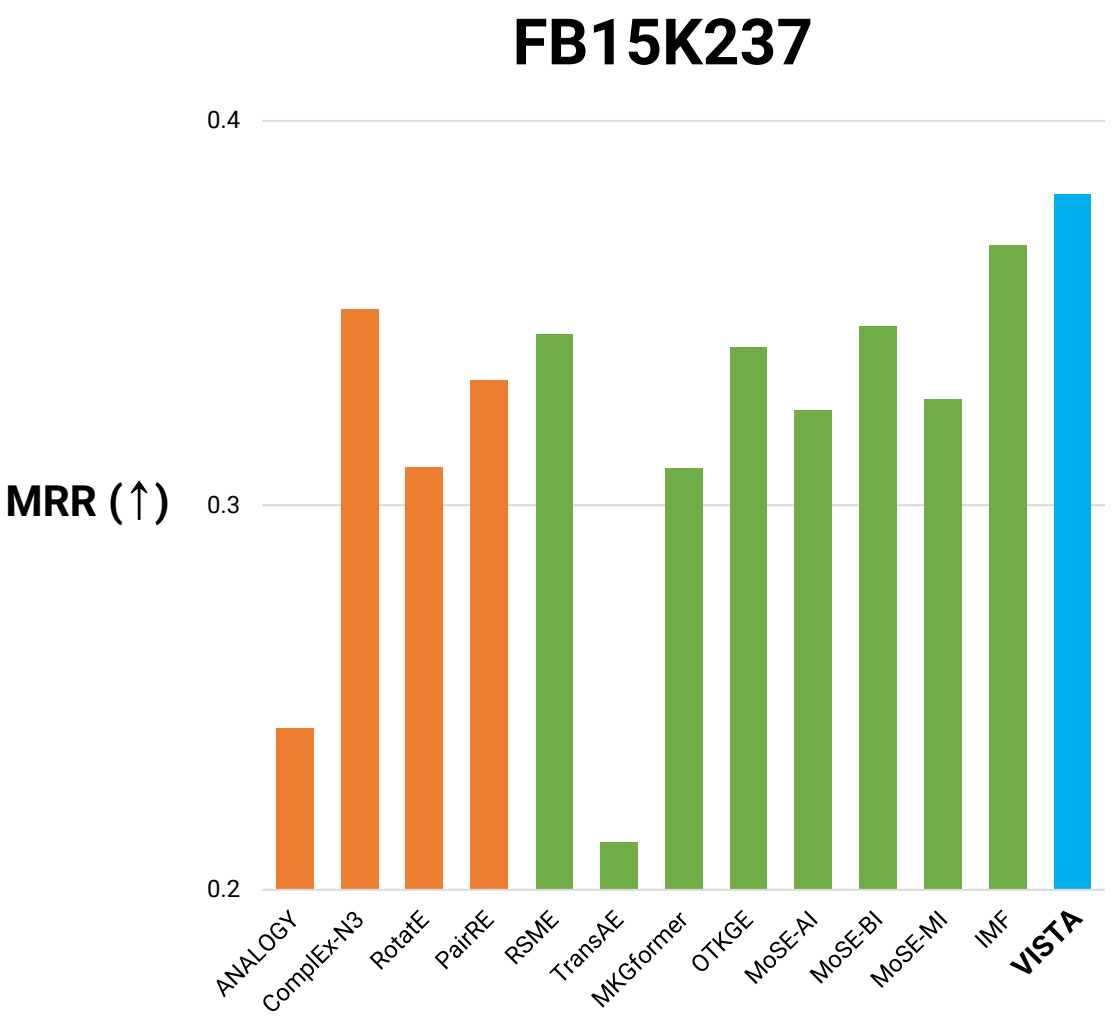


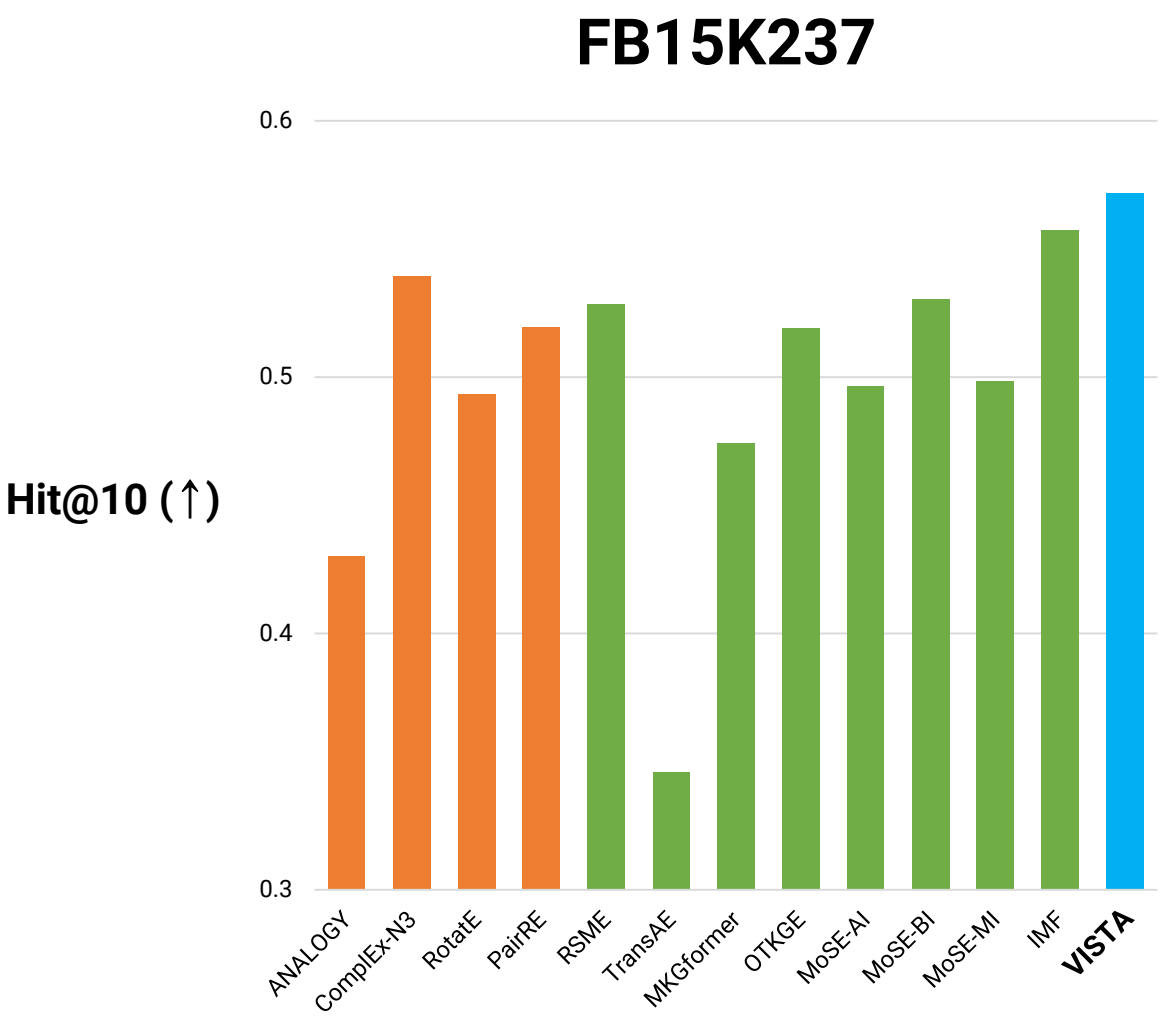
Knowledge Graph Completion Performance



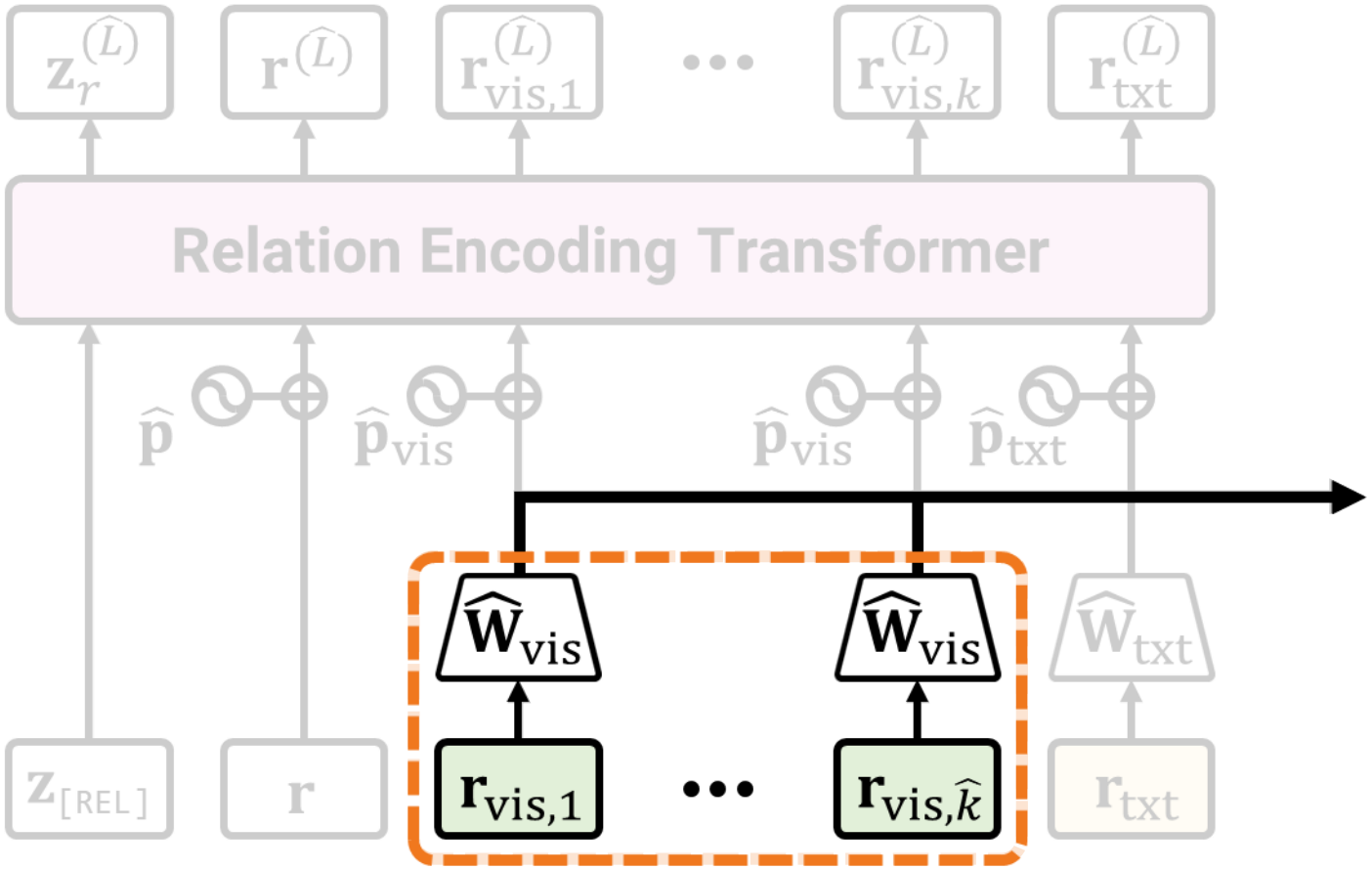


Knowledge Graph Completion Performance



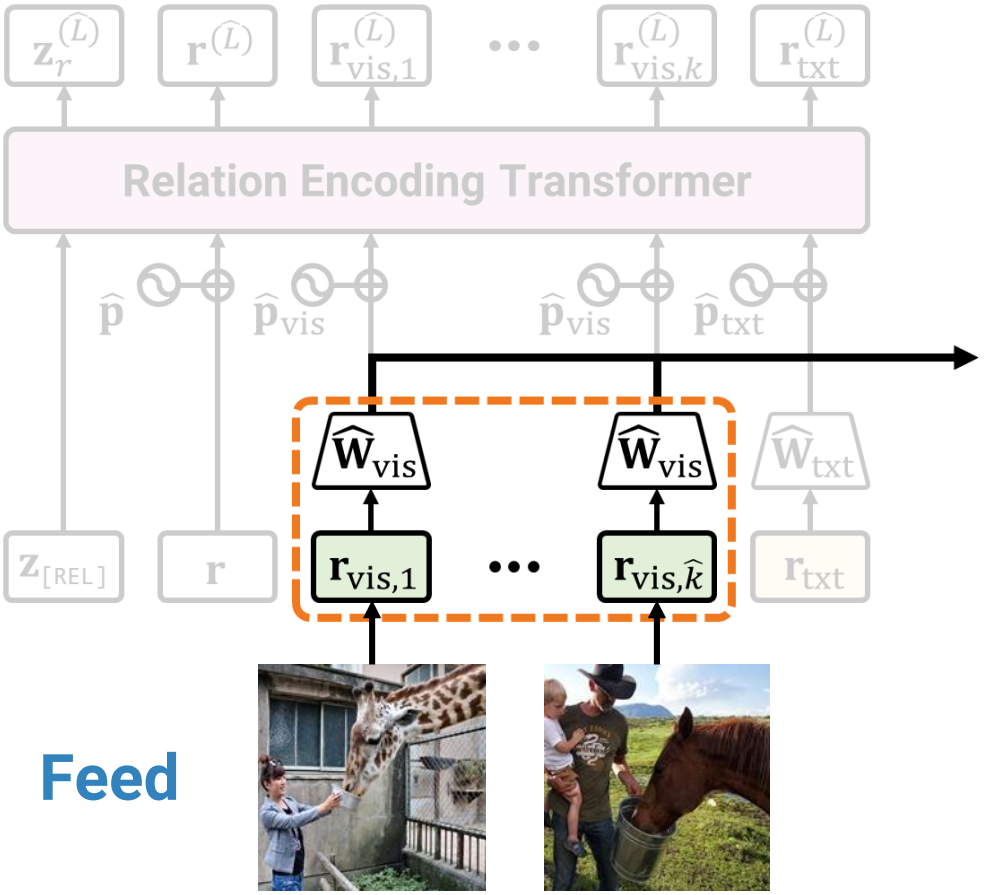


Visual Representation Vectors of Relations

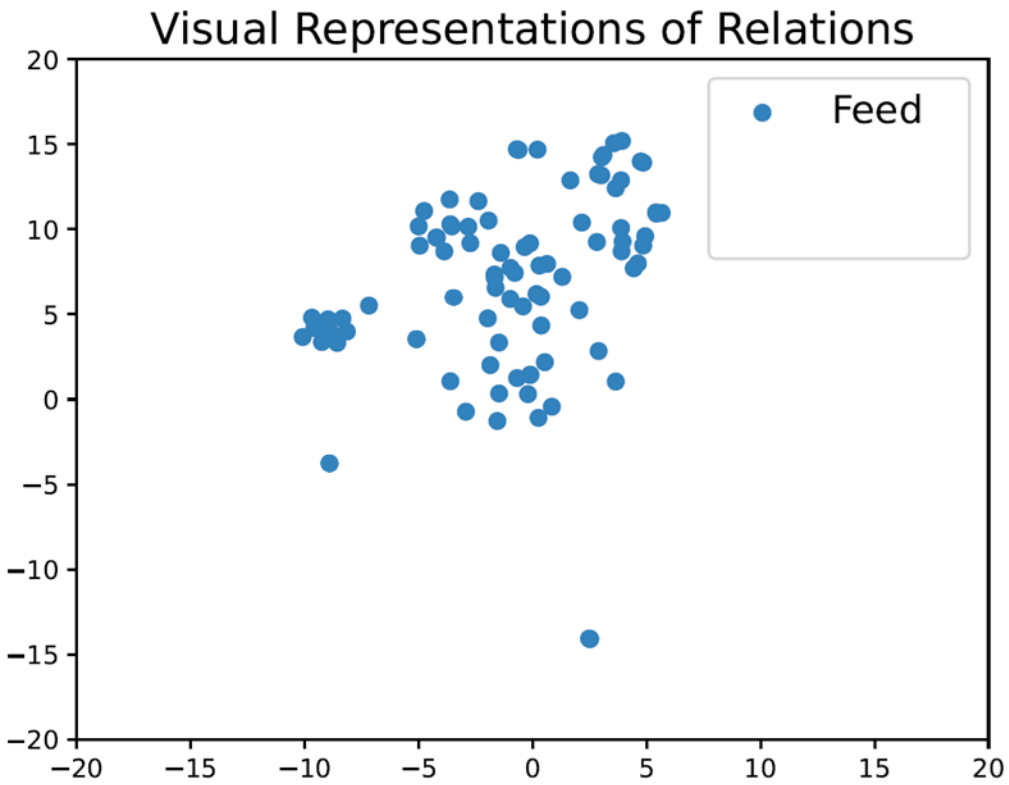


Visual representation vectors of relations

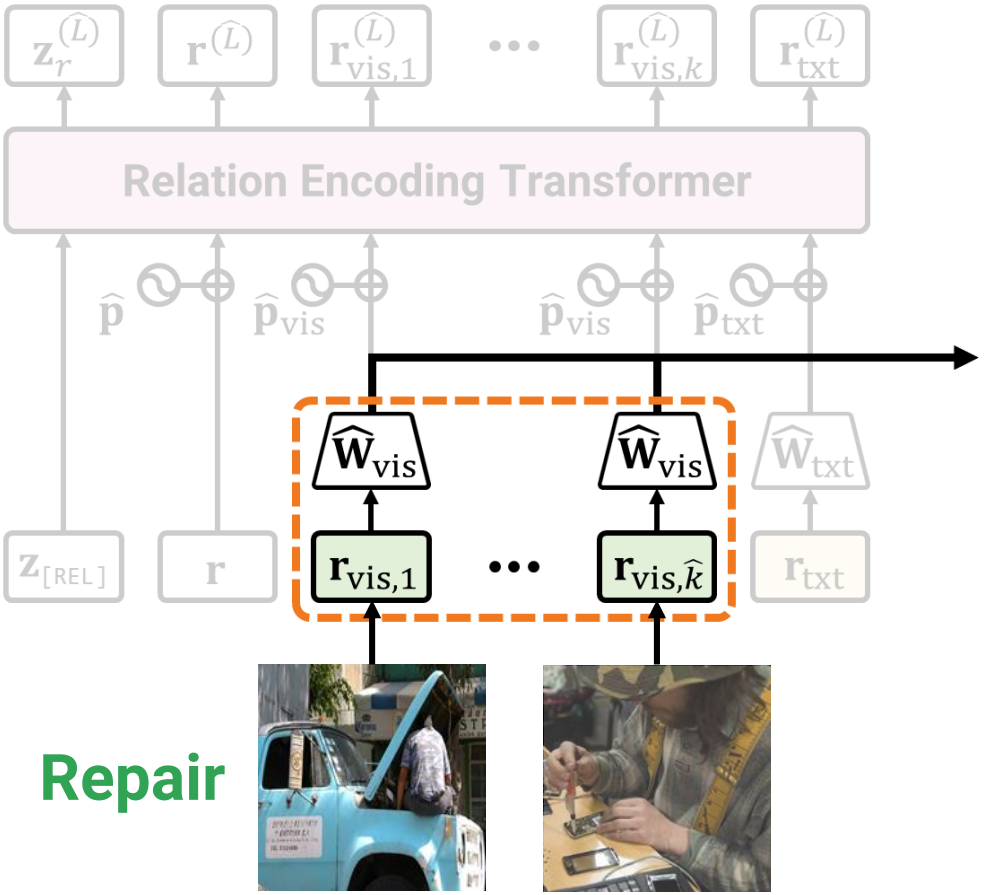
Visual Representation Vectors of Relations



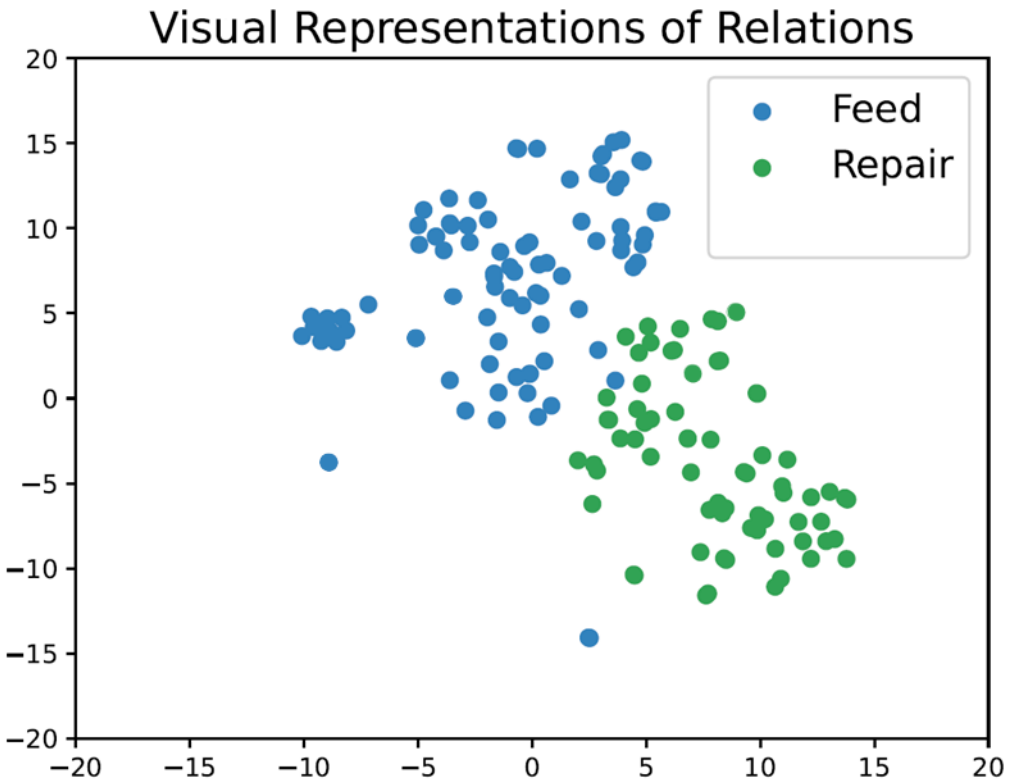
Visual representation vectors of Feed



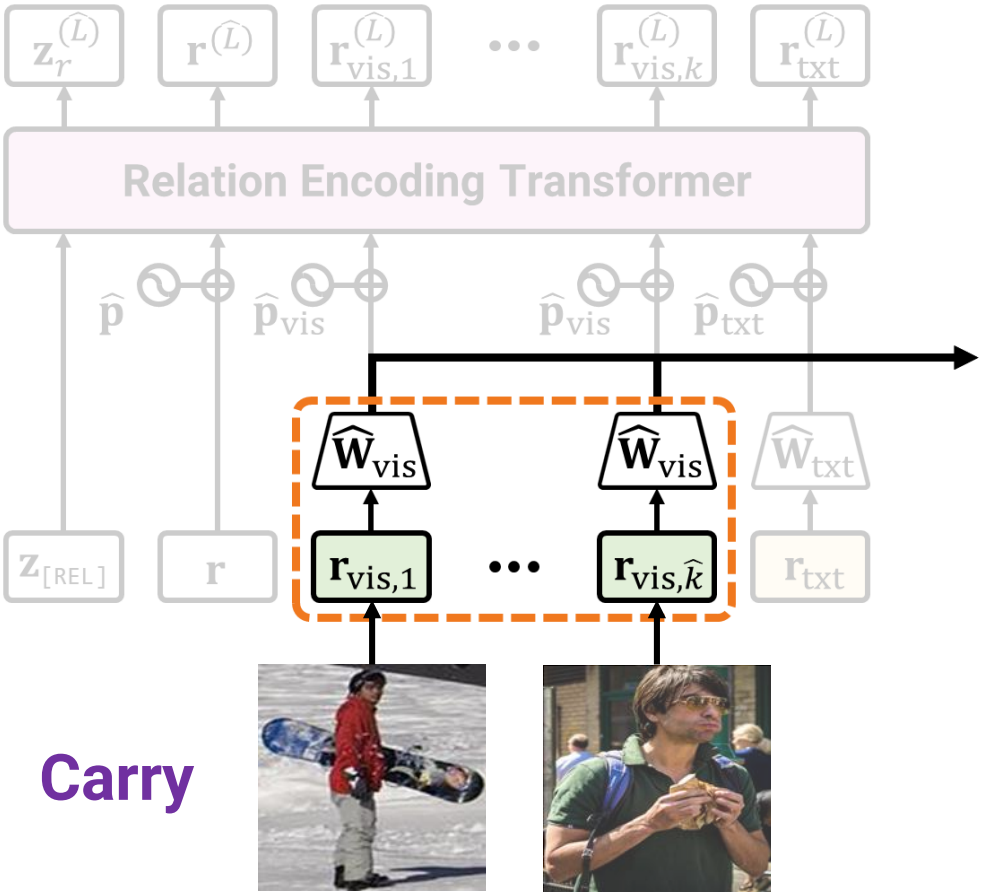
Visual Representation Vectors of Relations



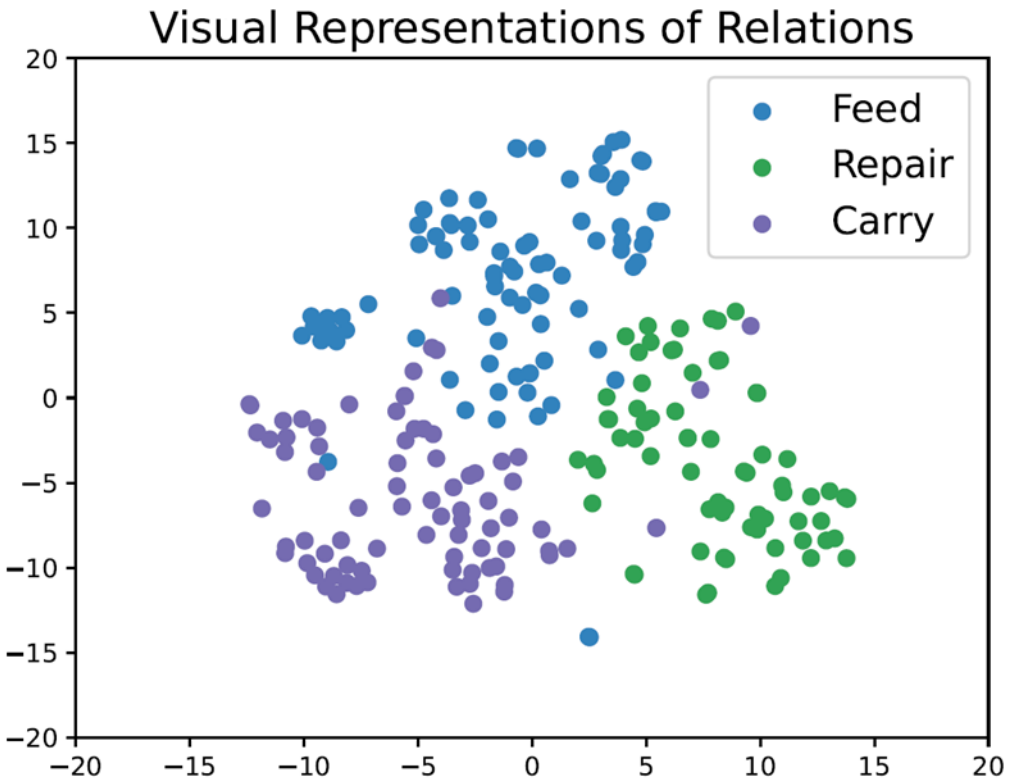
Visual representation vectors of Repair



Visual Representation Vectors of Relations



Visual representation vectors of Carry



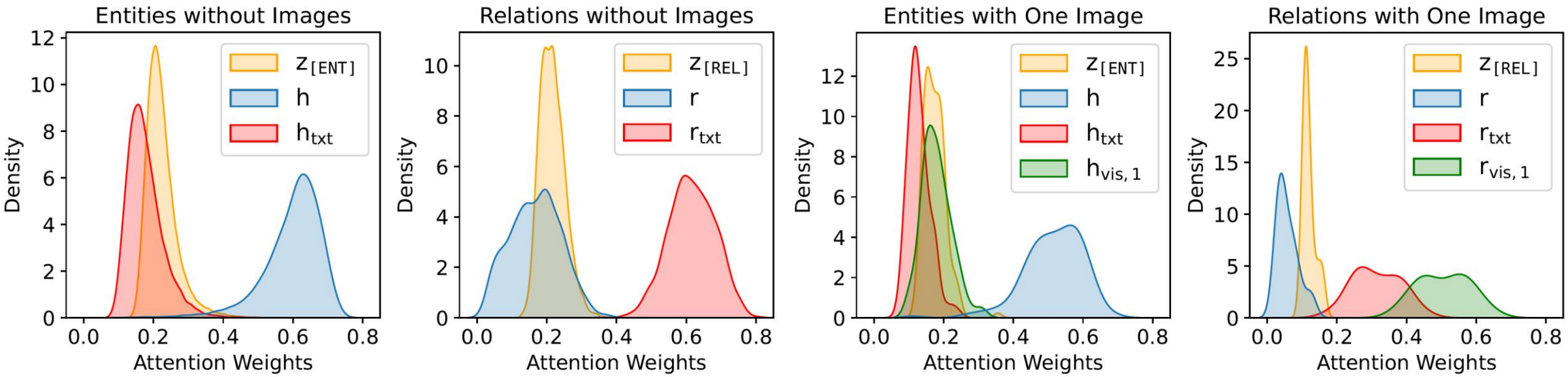
Top Similar Entities/Relations

- BERT returns **abstract concepts**; ViT returns **visually expressible concepts**.
- VISTA successfully returns the most semantically close entities and relations to the queries by utilizing **both texts and images**.

Query		BERT	ViT	VISTA
dark_red	1	incense	leisure_wear	orange
	2	coloring	sportswear	red
	3	buffer	sweatshirt	crimson
have	1	move	straddle	keep
	2	influence	hop_on	hold
	3	begin	inspect	incorporate

Attention Weights

- When images are not given, **learnable vectors** have relatively high attention weights in entities whereas **textual features** play the crucial role in relations.
- When an image is given, **learnable vectors** still have high importance in entities whereas **visual features** tend to have high contributions in relations.



Conclusion

- **V**isual-**T**extual **K**nowledge **G**raphs (**VTKGs**)
  - Visually expressible triplets are augmented by images
  - Both entities and relations have textual descriptions

- Propose **VIS**ual-**T**extu**A**l (**VISTA**) knowledge graph representation learning method to solve knowledge graph completion problems in real-world VTKG datasets

- VISTA takes into account the visual and textual features of entities and relations

- VISTA substantially outperforms 10 different state-of-the-art methods

Our datasets and codes are available at:

https://github.com/bdi-lab/VISTA

◀ GitHub

You can find us at:

{jjlee98, chanyoung.chung, jjwhang}@kaist.ac.kr

https://bdi-lab.kaist.ac.kr

◀ BDILab

KAIST BDILab
BIG DATA INTELLIGENCE