

面向结构感知的 多模态知识图谱补全

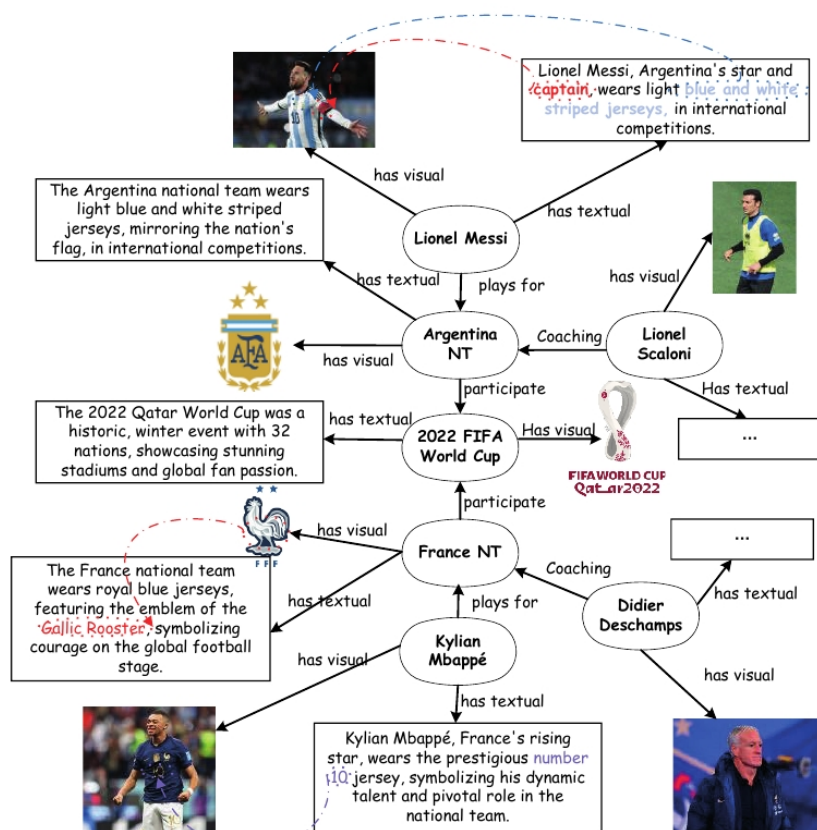
Towards Structure-aware Model for Multi-modal Knowledge Graph Completion

Linyu Li, Zhi Jin, Yichi Zhang, et al.

28 May 2025

结构 / 视觉 / 文本

现有 MMKGC 忽略了两个关键事实



细粒度跨模态关系广泛存在

细粒度交互不足

简单拼接、平均或 token 化后直接融合，难以显式捕获图像—文本之间的局部语义对应。

结构模态是推理主干

移除图结构后，MyGO 与 OTKGE 的 MRR / Hits@1 均显著下降；视觉与文本更适合作为辅助。

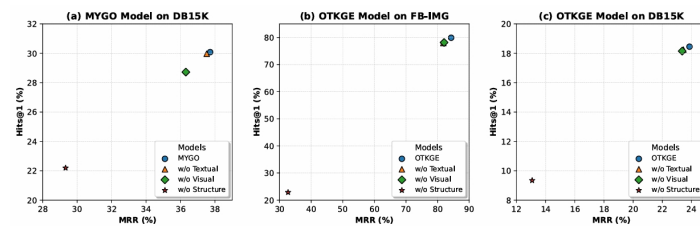
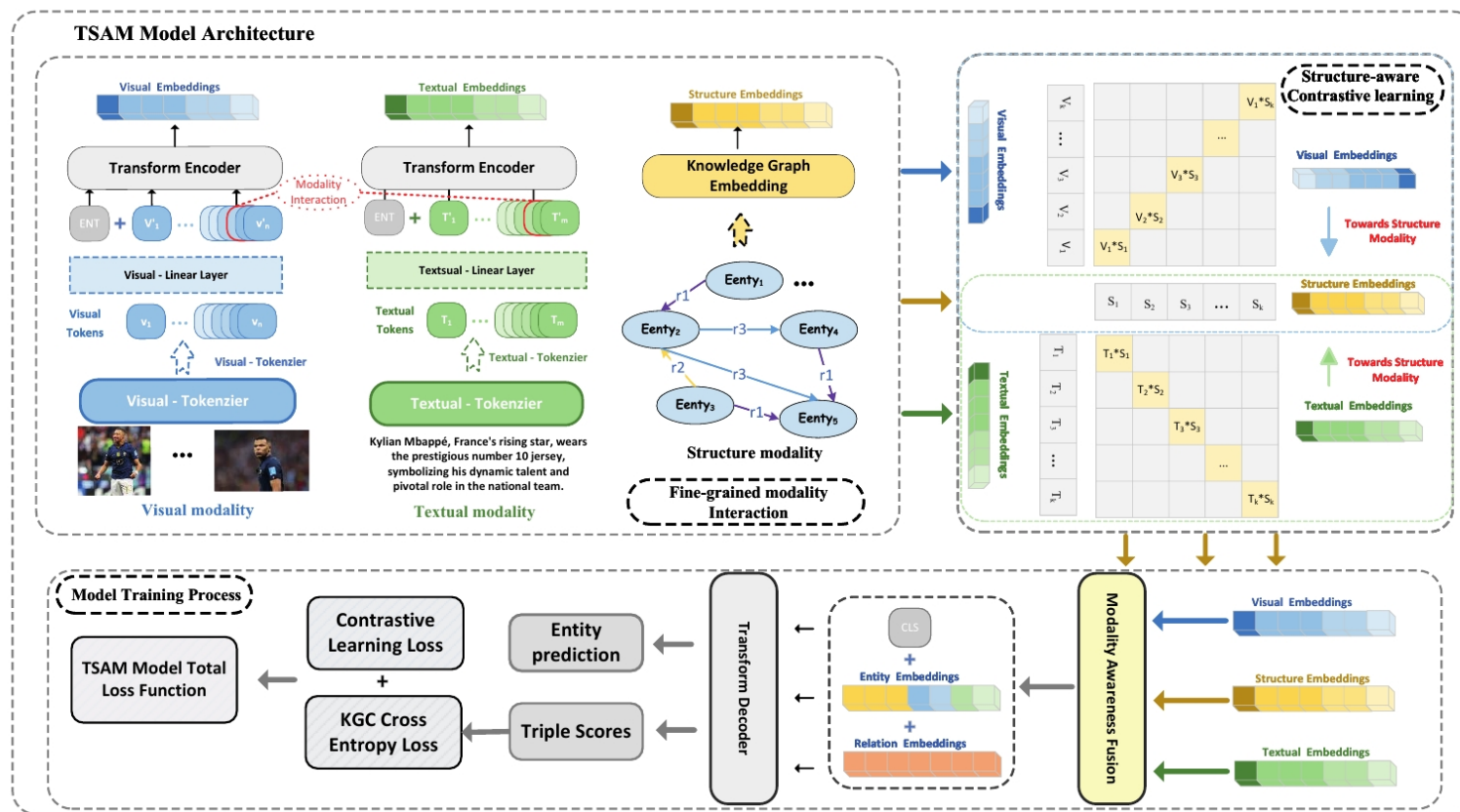


Fig. 2. Performance of the MyGO [25] and OTKGE [42] models in terms of MRR and Hits@1 after completely removing all modality knowledge.

核心问题：如何在融合中保留结构主导性，同时利用细粒度多模态信息？

03 TSAM: 以结构为锚点的细粒度多模态融合



FgMAF

视觉 / 文本先被离散为 token，经线性投影和 Transformer 编码，再由模态注意力融合。

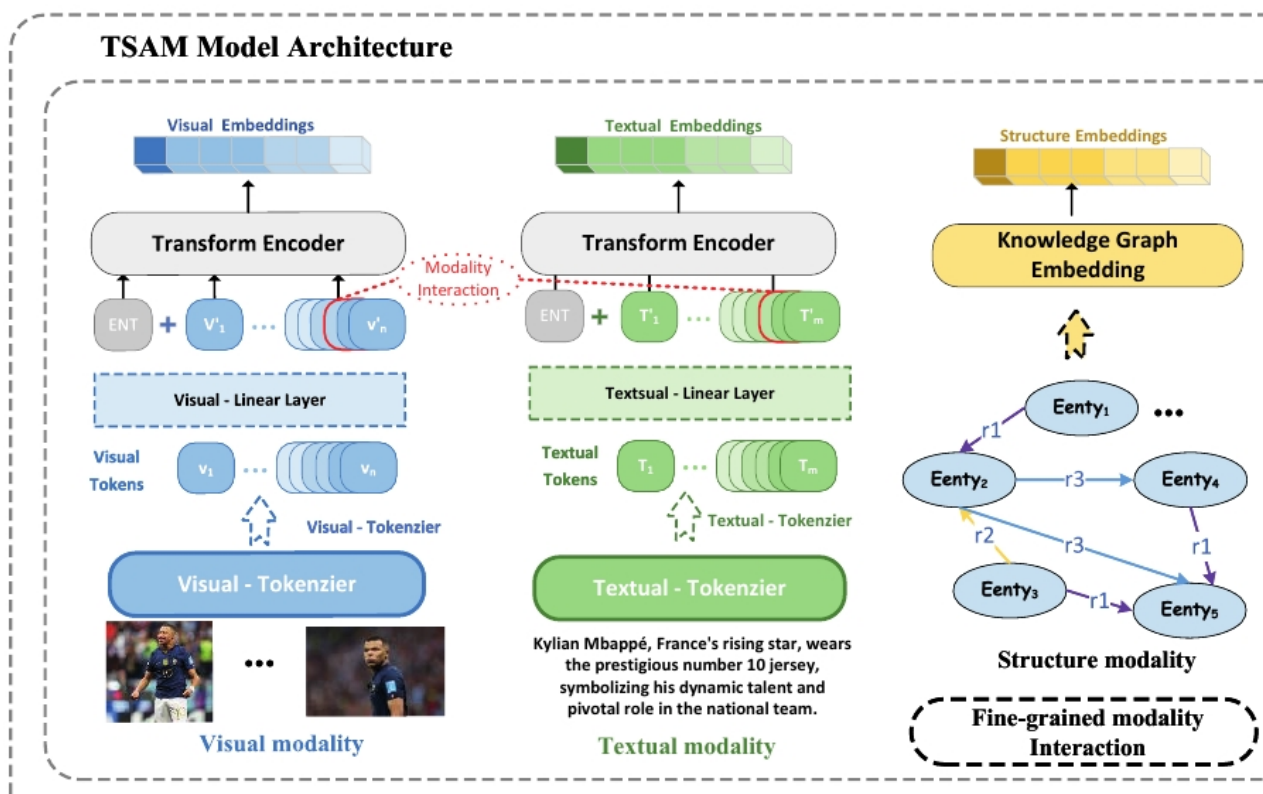
SaCL

分别执行结构-视觉、结构-文本对比学习，让其他模态向结构空间对齐并抑制噪声。

训练目标

$$L = \text{预测交叉熵} + \text{LST} + \text{LSV}$$

FgMAF: 先统一 token 空间，再做模态感知融合



离散化

BEIT-V2 将图像转为视觉 token；BERT 将实体描述转为文本 token。

统一表示空间

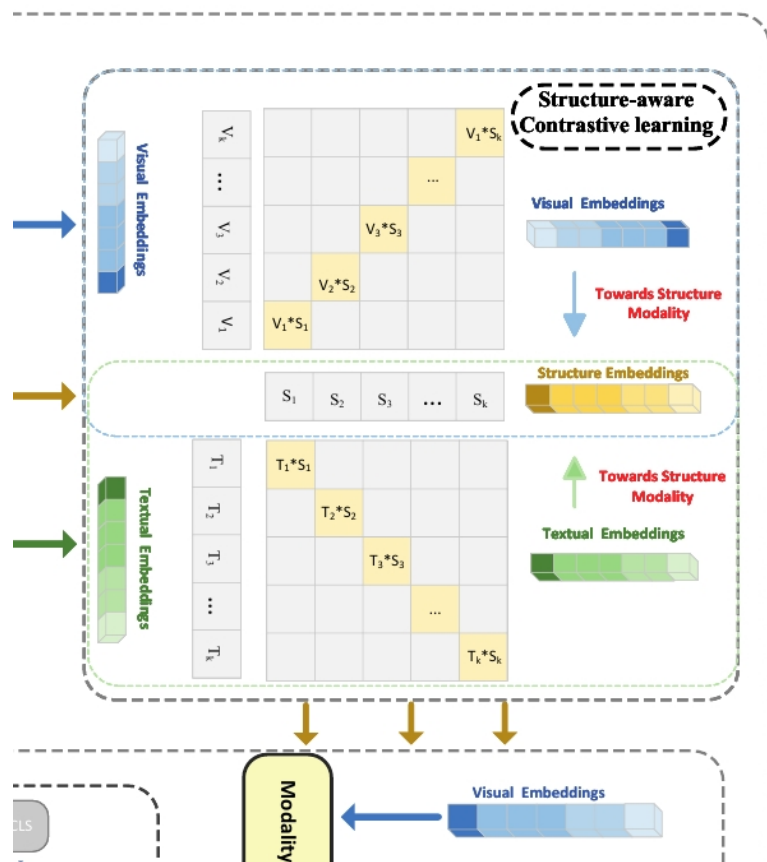
通过 gv / gt 线性投影统一维度，再由 Transformer 分别得到 ev_{vis} 与 et_{txt} ；KGE 提供 $estr$ 。

模态感知

注意力权重 α_s / α_v / α_t 随实体动态变化，避免固定比例融合。

$$ef = \text{stack}(\alpha_s \cdot estr, \alpha_v \cdot ev_{vis}, \alpha_t \cdot et_{txt})$$

05 SaCL: 以结构表示为锚点执行双向对比学习



结构是锚点

视觉与文本不是彼此对齐，而是分别向结构空间收敛。

正样本

同一实体的 S_i-V_i 与 S_i-T_i ；目标是提高跨模态同实体相似度。

负样本

从同一 mini-batch 随机采样 K 个其他实体，拉开错误配对的距离。

$$LSV = LS \rightarrow V + LV \rightarrow S$$

$$LST = LS \rightarrow T + LT \rightarrow S$$

实验设置覆盖三套公开 MMKGC 基准

TABLE I

STATISTICAL OF THE DB15K [57] AND MKG-W/Y [51] DATASETS.

Dataset	$ \mathcal{E} $	$ \mathcal{R} $	#Vis	#Text	#Train	#Valid	#Test
DB15K	12,842	279	12,818	12,842	79,222	9,902	9,904
MKG-W	15,000	169	14,463	14,123	34,196	4,276	4,274
MKG-Y	15,000	28	14,244	14,305	21,310	2,665	2,663

评价指标

MRR**Hits@1****Hits@3****Hits@10**

均为排名类指标；数值越高越好

实现与对比

多模态编码

BEIT-V2 生成视觉 token，BERT 生成文本 token。

主干与变体

BERT-base 为主编码器 / 解码器；另测 BERT-large、RoBERTa、DeBERTa、LLaMA-7 B。

对比规模

3 个单模态 KGC + 10 个 MMKGC 基线；Adam 优化，8×V100 32GB。

08 主结果：TSAM 在三套数据上几乎全面领先

TABLE II

THE EXPERIMENTAL RESULTS OF TSAM AND THE BASELINE MODEL ON THREE MMKG DATASETS. ♣ REPRESENTS THE EXPERIMENTAL RESULTS THAT WE REPRODUCED THROUGH ITS SOURCE CODE. THE REST OF THE BASELINE RESULTS ARE FROM THE SOURCE PAPERS OF THEIR RESPECTIVE MODELS AND THE REPORT IN [25] AND [64].

		DB15K				MKG-W				MKG-Y			
Model		MRR	Hit@1	Hit@3	Hit@10	MRR	Hit@1	Hit@3	Hit@10	MRR	Hit@1	Hit@3	Hit@10
KGC	TransE	24.86	12.78	31.48	47.07	29.19	21.06	33.20	44.23	30.73	23.45	35.18	43.37
	Tucker	33.86	25.33	37.91	50.38	30.39	24.44	32.91	41.25	37.05	34.59	38.43	41.45
	RotatE	29.28	17.87	36.12	49.66	33.67	26.80	36.68	46.76	34.95	29.10	38.35	45.30
MMKGC	IKRL	26.82	14.09	34.93	49.09	32.36	26.11	34.75	44.07	33.22	30.37	34.28	38.60
	RSME	29.80	24.20	32.10	49.40	29.20	23.40	32.00	40.40	34.40	33.80	36.10	38.60
	AdaMF	32.51	21.31	39.67	51.68	34.27	27.21	37.86	47.21	38.06	33.49	40.44	45.48
	OTKGE	23.86	18.45	25.89	34.23	34.36	34.36	36.25	44.88	35.51	31.97	37.18	41.38
	VISTA	30.42	22.49	33.56	45.94	32.91	26.12	35.38	45.61	30.45	24.87	32.39	41.53
	QEB	28.18	14.82	36.67	51.55	32.38	25.47	35.06	45.32	34.37	29.49	36.95	42.32
	IMF	32.25	24.20	36.00	48.19	34.50	28.77	36.62	45.44	35.79	32.95	37.14	40.63
	MMRNS	32.68	23.01	37.86	51.01	35.03	28.59	37.49	47.47	35.93	30.53	39.07	45.47
	SNAG♣	36.30	27.40	41.10	<u>53.00</u>	<u>37.30</u>	<u>30.20</u>	<u>40.50</u>	<u>50.30</u>	39.10	34.7	41.08	46.70
	Native♣	34.30	25.08	39.48	51.35	36.84	29.94	40.06	49.39	<u>39.21</u>	<u>35.03</u>	<u>41.21</u>	46.25
	MyGO	<u>37.72</u>	<u>30.08</u>	<u>41.26</u>	52.21	36.10	29.78	38.54	47.75	38.44	35.01	39.84	44.19
TSAM(Ours)		40.50	32.60	44.36	55.44	40.07	33.29	42.53	52.72	39.80	35.28	41.29	<u>46.44</u>
improvement		7.37%	8.38%	7.51%	4.60%	7.43%	10.23%	5.01%	4.81%	1.5%	0.7%	0.2%	-

40.50

DB15K MRR (+7.37%)

+10.23%

MKG-W Hits@1 相对提升

11 / 12

指标取得最优；MKG-Y Hits@10 例外

提升在 MRR / Hits@1 更明显，说明模型不仅整体排序更好，也提高了精确命中。

消融：结构对齐比融合本身更关键

TABLE III
ABLATION EXPERIMENTS OF TSAM ON DB15K AND MKG-W DATASETS

	DB15K			
	MRR	Hit@1	Hit@3	Hit@10
TSAM	40.50	32.60	44.36	55.44
w/o FgMAF	39.83	31.90	43.73	54.55
w/o SaCL	38.83	31.11	42.87	53.95
w/o $\mathcal{L}_{ST}$	39.33	31.62	42.99	54.31
w/o $\mathcal{L}_{SV}$	39.37	31.65	42.93	54.36

	MKG-W			
	MRR	Hit@1	Hit@3	Hit@10
TSAM	40.07	33.29	42.53	52.72
w/o FgMAF	38.88	32.60	41.54	51.02
w/o SaCL	37.30	31.23	39.45	48.82
w/o $\mathcal{L}_{ST}$	38.63	32.26	41.12	50.95
w/o $\mathcal{L}_{SV}$	38.51	32.16	40.83	50.6

TABLE IV
EXPERIMENTAL RESULTS OF THE TSAM MODEL USING DIFFERENT MESSAGE FUNCTIONS ON THE MKG-W DATASET.

Model	MRR	Hit@1	Hit@3	Hit@10
TSAM-Tucker	40.07	33.29	42.53	52.72
TSAM-RotateE	37.49	31.09	39.67	49.46
TSAM-TransE	36.77	30.59	38.74	48.29

-6.92%

移除 SaCL: MKG-W MRR

-2.97%

移除 FgMAF: MKG-W MRR

两条对比损失都必要

去掉 LST 或 LSV 均导致下降，说明结构—文本与结构—视觉对齐不可互相替代。

Tucker 最适合作为结构编码器

MKG-W MRR: Tucker 40.07, RotateE 37.49, TransE 36.77。

负样本：K=16 已接近最优，继续增大主要增加显存

TABLE V
PARAMETER SENSITIVITY EXPERIMENT OF THE NUMBER OF NEGATIVE
SAMPLES K

Neg_Num	DB15K				
	MRR	Hit@1	Hit@3	Hit@10	Mem. Usage
K=8	39.50	31.78	43.03	54.65	16.77G
K=16	40.50	32.60	44.36	55.44	17.47G
K=32	40.07	32.36	43.44	55.00	20.24G
K=64	40.16	32.24	43.93	55.59	20.99G

Neg_Num	MKG-W				
	MRR	Hit@1	Hit@3	Hit@10	Mem. Usage
K=8	39.80	33.27	42.23	52.24	19.26G
K=16	40.07	33.29	42.53	52.72	19.96G
K=32	40.10	33.29	42.70	52.72	22.55G
K=64	40.20	33.37	43.05	52.71	24.23G

17.47G

DB15K · K=16 显存

20.99G

DB15K · K=64 显存

性能—效率折中

DB15K: K=16 的 MRR 为 40.50; K=64 反而为 40.16。

MKG-W: K=64 仅比 K=16 高 0.13 MRR, 却多占约 4.27G 显存。

模态 token：数量增加有效，但高区间收益趋缓

F. Experiments with different numbers of modality tokens

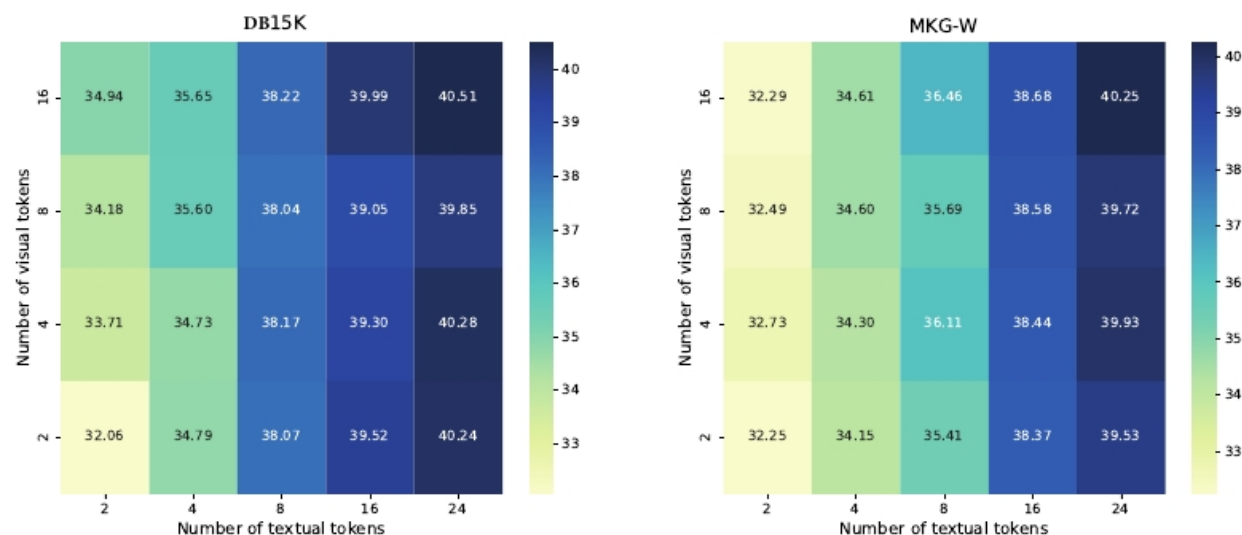


Fig. 6. The MRR performance of the TSAM model on the DB15K and MKG-W datasets with varying numbers of modality tokens.

总体趋势

增加视觉与文本 token 能补充实体 / 关系特征, DB15K 与 MKG-W 的 MRR 均明显提升。

边际收益递减

16-24 个文本 token 区间提升开始放缓, 冗余特征与计算开销逐渐显现。

实践建议

优先保证多模态信息的平衡覆盖, 再根据资源预算增加 token。

案例：结构强约束问题中，TSAM 的排名优势最明显

TABLE VI
CASE STUDY EXPERIMENTS ON THE MKG-W DATASET

Case 1: (<i>J.R.R.Tolkien</i> , <i>ethnic group</i> , ?)	
TSAM	Rank of the correct tail entity " <i>English people</i> " : <u>1</u>
MyGo	Rank of the correct tail entity " <i>English people</i> " : <u>2</u>
Case 2: (<i>Oasis</i> , <i>influenced by</i> , ?)	
TSAM	Rank of the correct tail entity " <i>The Beatles</i> " : <u>1</u>
MyGo	Rank of the correct tail entity " <i>The Beatles</i> " : <u>9</u>
Case 3: (<i>Son of Paleface</i> , <i>cast member</i> , ?)	
TSAM	Rank of the correct tail entity " <i>Bing Crosby</i> " : <u>1</u>
MyGo	Rank of the correct tail entity " <i>Bing Crosby</i> " : <u>36</u>
Case 4: (<i>Paris Underground</i> , <i>director</i> , ?)	
TSAM	Rank of the correct tail entity " <i>Gregory Ratoff</i> " : <u>1</u>
MyGo	Rank of the correct tail entity " <i>Gregory Ratoff</i> " : <u>5119</u>

1 vs 5119

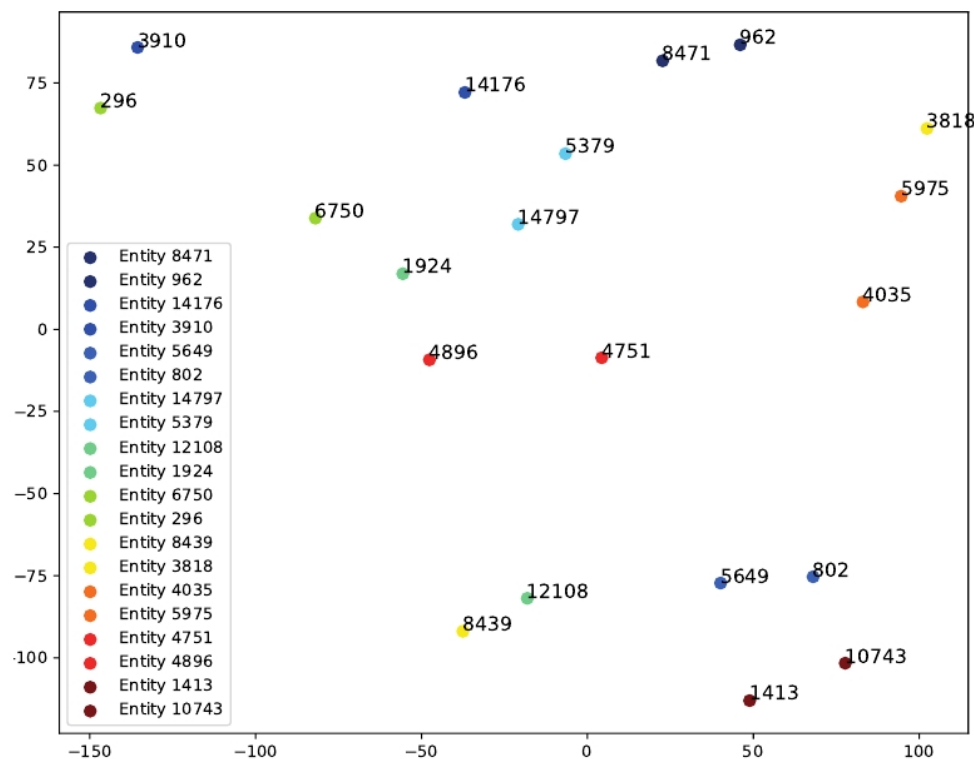
Paris Underground → director

四个案例中，TSAM 均将正确实体排在第 1；
MyGO 的排名分别为 2 / 9 / 36 / 5119。

为什么差距这么大？

SaCL 保住了结构主导性。

可视化：相关实体被拉到更近的表达空间



结论

同一三元组的实体对在 t-SNE 空间中更接近；
若干印度电影相关实体也形成邻近簇。

FgMAF

补充细粒度视觉 / 文本语义。

SaCL

以结构为锚点对齐多模态表示。

局限：大规模动态知识图谱上的可扩展性仍待验证。