



**NANYANG**  
TECHNOLOGICAL  
UNIVERSITY



# Unbiased Scene Graph Generation from Biased Training

**Kaihua Tang<sup>1</sup>, Yulei Niu<sup>3</sup>, Jianqiang Huang<sup>1,2</sup>, Jiaxin Shi<sup>4</sup>, Hanwang Zhang<sup>1</sup>**

<sup>1</sup>Nanyang Technological University, <sup>2</sup>Damo Academy, Alibaba Group,

<sup>3</sup>Renmin University of China, <sup>4</sup>Tsinghua University

**Github:** <https://github.com/KaihuaTang/Scene-Graph-Benchmark.pytorch>

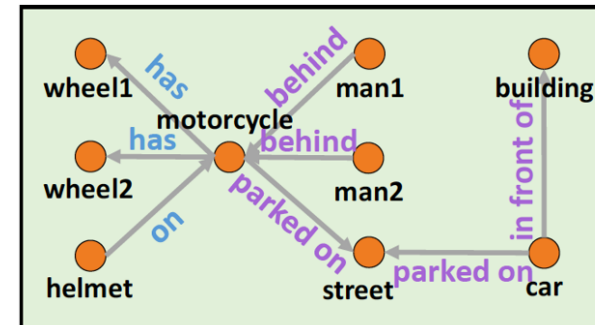
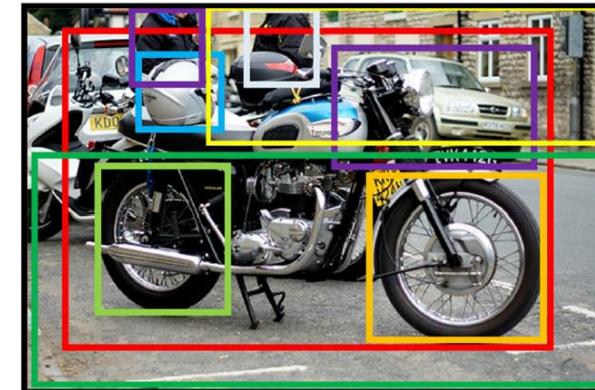
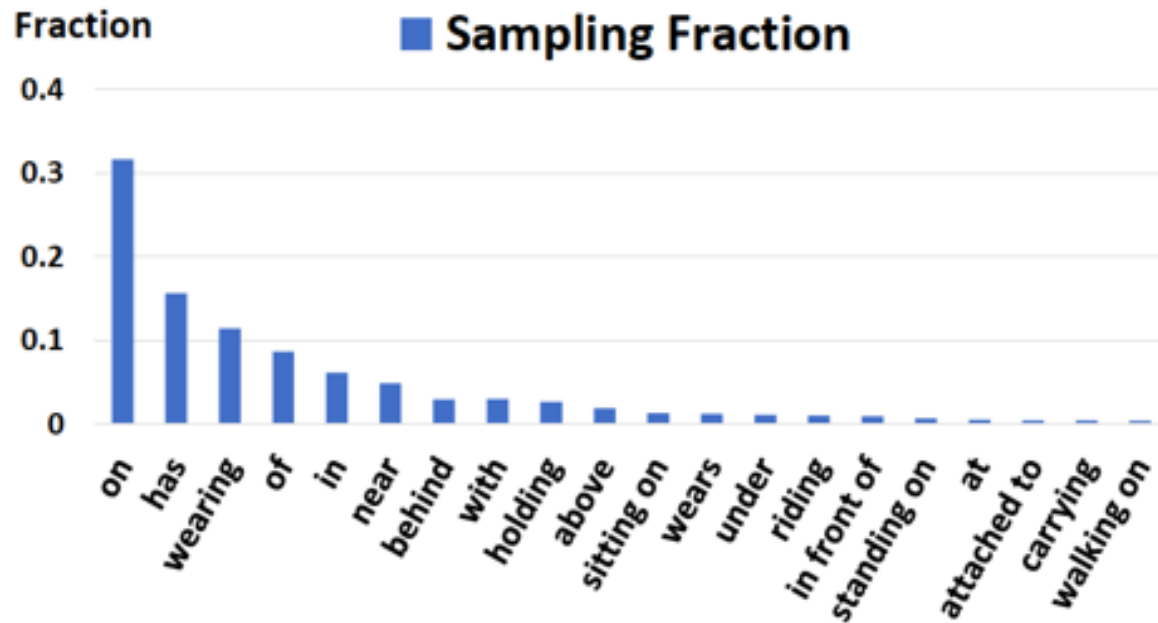
# Contents

---

- Motivation
- **Three Steps** Towards the Unbiased Inference:
  - Causal Graph
  - Counterfactual Thinking
  - Total Direct Effect
- Advantages
- Quantitative Results
- Qualitative Examples

# Motivation

- Existing Problems of SGG
  - a) Long-Tailed Distribution of Dataset
  - b) Trivial Predicates for Dominating Relationships

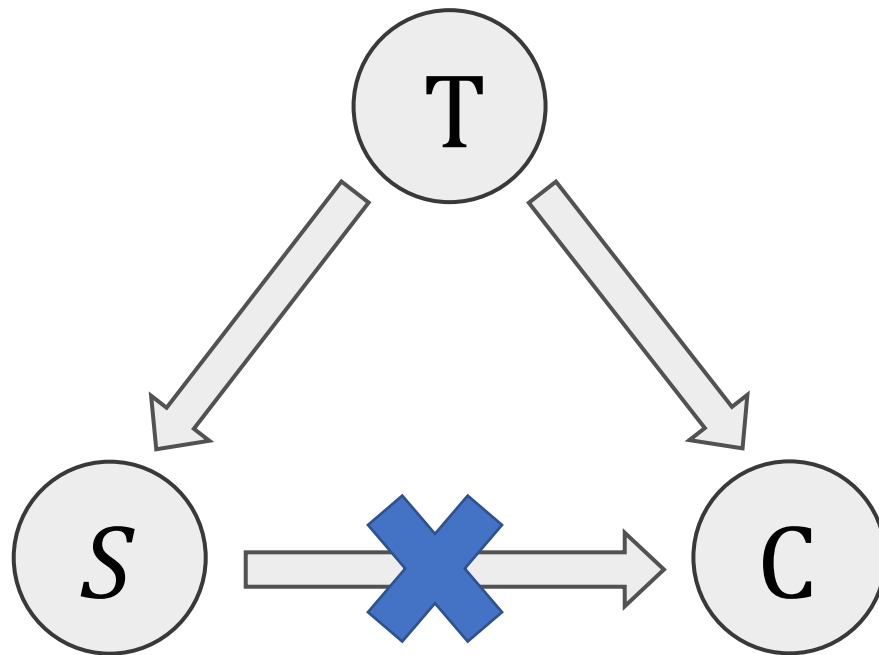


(a) Distribution of Predicates in VisualGenome (b) Example of the Proposed Unbiased SG

# Prerequisite: Causal Graph

---

- Definition: Causal Graph (a.k.a., structural causal model) is a directed acyclic graph  $G = \{N, E\}$ , indicating how a set of variables  $N$  interact with each other through the causal link  $E$ .



$T$  : Temperature

$S$  : Ice-cream Sales

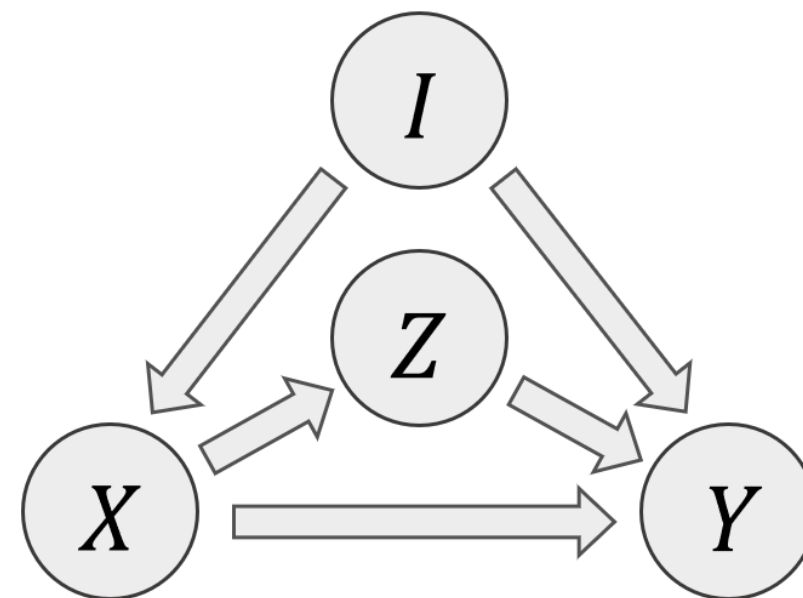
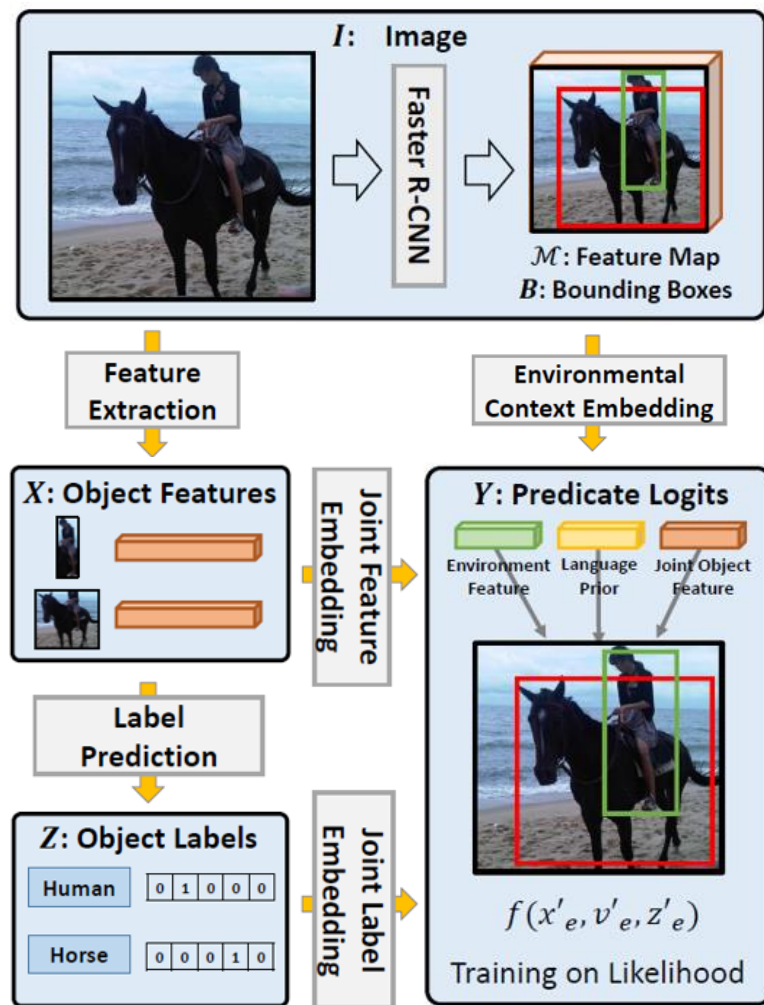
$C$  : Crime

**Common Strategy:** Take as much related inputs as possible.

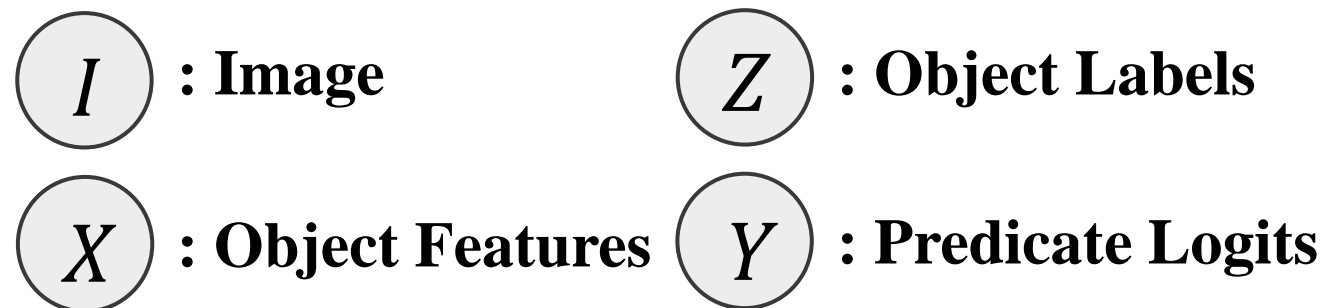
[1] Judea Pearl, Dana Mackenzie, The Book of Why

[2] Judea Pearl, Madelyn Glymour, and Nicholas Jewell, Causal Inference in Statistics: A Primer

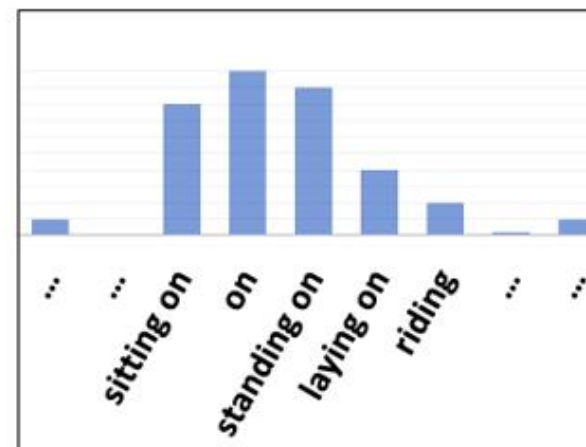
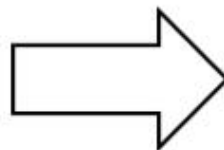
## Causal Graph of SGG: The Road Map



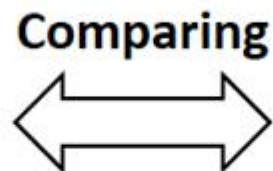
Valid edge can still be the shortcut



## Counterfactual Thinking: Identify the Source of Bias



(a) Biased Generation Based on Likelihood

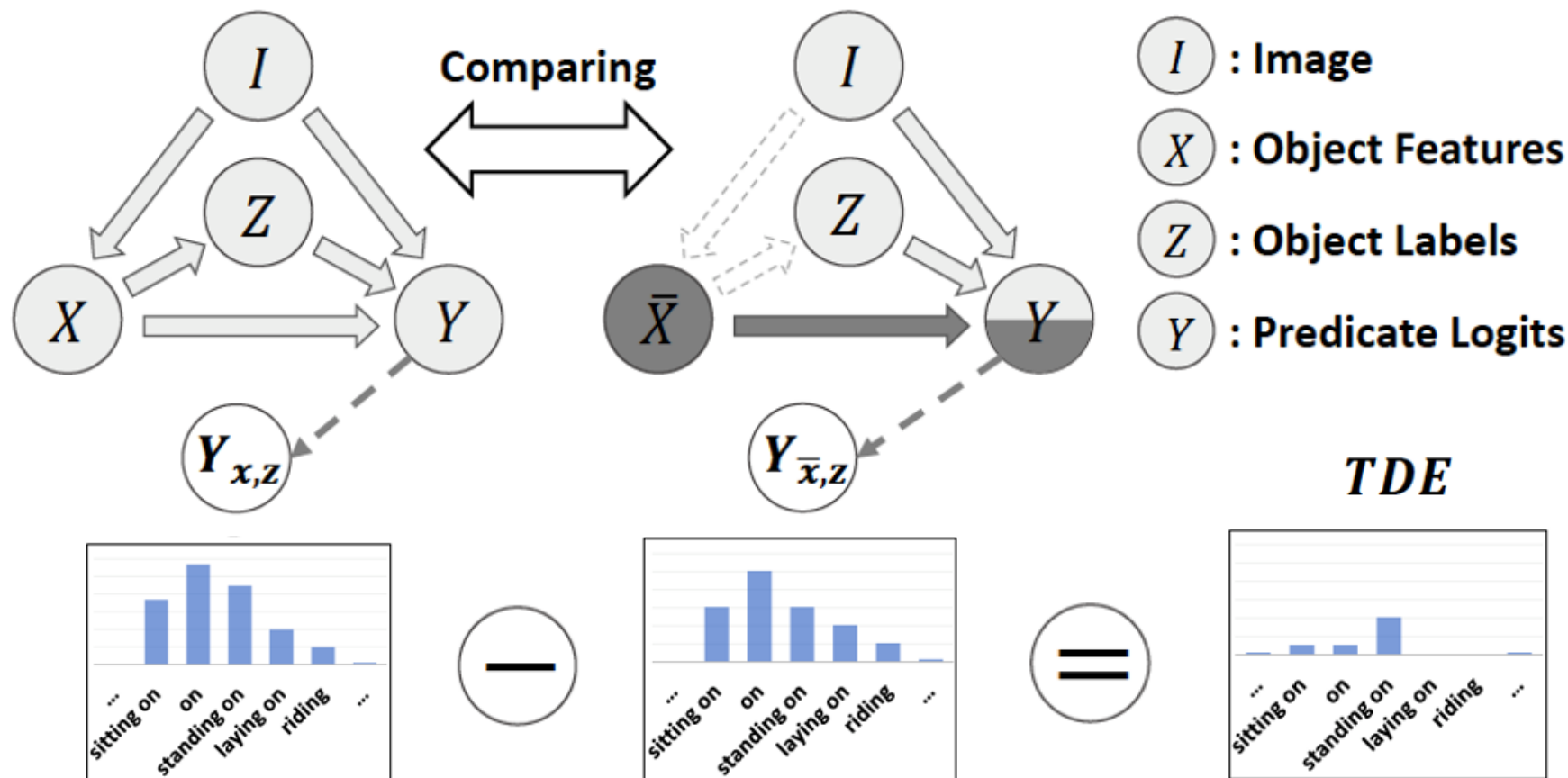


(b) An Intuitive Example of Counterfactual Thinking

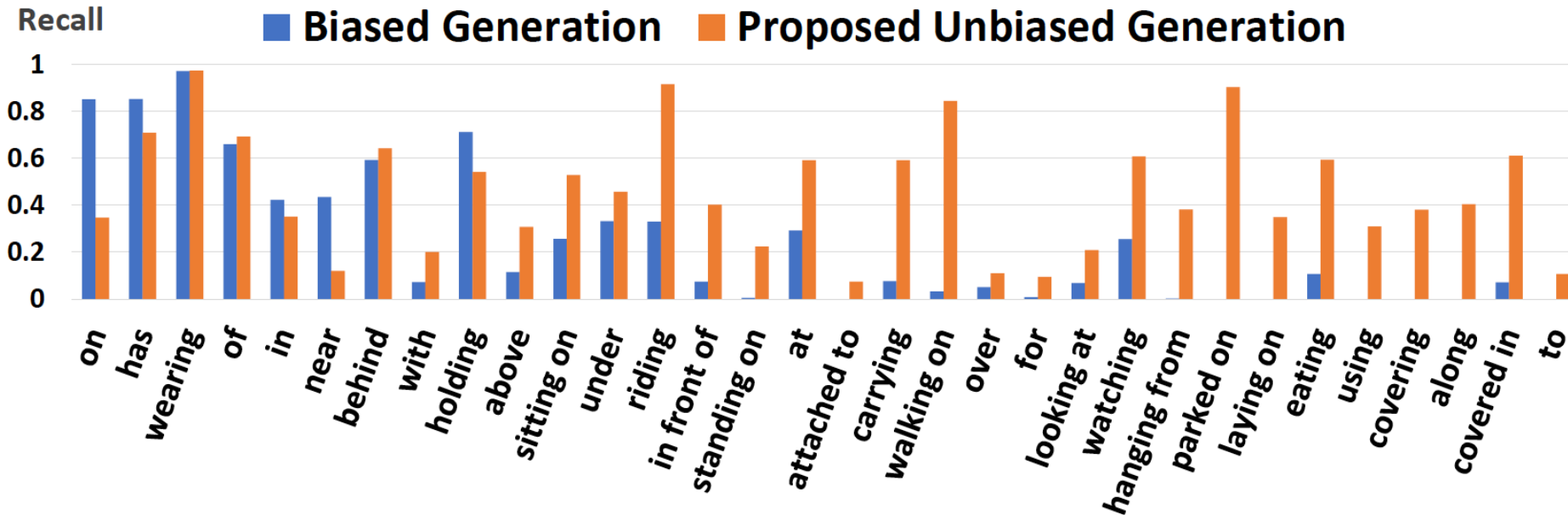


# Total Direct Effect (TDE): Unbiased Inference

- Note that the training is still conventional biased training without any re-balancing tricks (re-weight / re-sampling).



# Advantages



- **Advantages:**

- **Model-Agnostic:** Consistent improvements across different models and fusion functions.
- **Simple:** Respecting the conventional biased training.
- **Effective:** Significantly improving mean Recall@K, Zero-Shot Recall@K, Sentence-to-Graph Retrieval



# Quantitative Results: Mean Recall@K

Model	Fusion	Method	Predicate Classification			Scene Graph Classification			Scene Graph Detection		
			mR@20	mR@50	mR100	mR@20	mR50	mR100	mR@20	mR50	mR100
IMP+	-	-	-	9.8	10.5	-	5.8	6.0	-	3.8	4.8
FREQ	-	-	8.3	13.0	16.0	5.1	7.2	8.5	4.5	6.1	7.1
MOTIFS	-	-	10.8	14.0	15.3	6.3	7.7	8.2	4.2	5.7	6.6
KERN	-	-	-	17.7	19.2	-	9.4	10.0	-	6.4	7.3
VCTree	-	-	14.0	17.9	19.4	8.2	10.1	10.8	5.2	6.9	8.0
MOTIFS <sup>†</sup>	SUM	Baseline	11.5	14.6	15.8	6.5	8.0	8.5	4.1	5.5	6.8
		Focal	10.9	13.9	15.0	6.3	7.7	8.3	3.9	5.3	6.6
		Reweight	16.0	20.0	21.9	8.4	10.1	10.9	<b>6.5</b>	<b>8.4</b>	<b>9.8</b>
		Resample	14.7	18.5	20.0	9.1	11.0	11.8	5.9	8.2	9.7
		X2Y	13.0	16.4	17.6	6.9	8.6	9.2	5.1	6.9	8.1
		X2Y-Tr	11.6	14.9	16.0	6.5	8.4	9.1	5.0	6.9	8.1
		TE	18.2	25.3	29.0	8.1	12.0	14.0	5.7	8.0	9.6
		NIE	0.6	1.1	1.4	6.1	9.0	10.6	3.8	5.1	6.0
		TDE	<b>18.5</b>	<b>25.5</b>	<b>29.1</b>	<b>9.8</b>	<b>13.1</b>	<b>14.9</b>	5.8	8.2	<b>9.8</b>
	GATE	Baseline	12.2	15.5	16.8	7.2	9.0	9.5	5.2	7.2	8.5
		TDE	<b>18.5</b>	<b>24.9</b>	<b>28.3</b>	<b>11.1</b>	<b>13.9</b>	<b>15.2</b>	<b>6.6</b>	<b>8.5</b>	<b>9.9</b>
VTransE <sup>†</sup>	SUM	Baseline	11.6	14.7	15.8	6.7	8.2	8.7	3.7	5.0	6.0
		TDE	<b>17.3</b>	<b>24.6</b>	<b>28.0</b>	<b>9.3</b>	<b>12.9</b>	<b>14.8</b>	<b>6.3</b>	<b>8.6</b>	<b>10.5</b>
	GATE	Baseline	13.6	17.1	18.6	6.6	8.2	8.7	5.1	6.8	8.0
		TDE	<b>18.9</b>	<b>25.3</b>	<b>28.4</b>	<b>9.8</b>	<b>13.1</b>	<b>14.7</b>	<b>6.0</b>	<b>8.5</b>	<b>10.2</b>
VCTree <sup>†</sup>	SUM	Baseline	11.7	14.9	16.1	6.2	7.5	7.9	4.2	5.7	6.9
		TDE	<b>18.4</b>	<b>25.4</b>	<b>28.7</b>	<b>8.9</b>	<b>12.2</b>	<b>14.0</b>	<b>6.9</b>	<b>9.3</b>	<b>11.1</b>
	GATE	Baseline	12.4	15.4	16.6	6.3	7.5	8.0	4.9	6.6	7.7
		TDE	<b>17.2</b>	<b>23.3</b>	<b>26.6</b>	<b>8.9</b>	<b>11.8</b>	<b>13.4</b>	<b>6.3</b>	<b>8.6</b>	<b>10.3</b>

# Quantitative Results: Zero-Shot & Retrieval

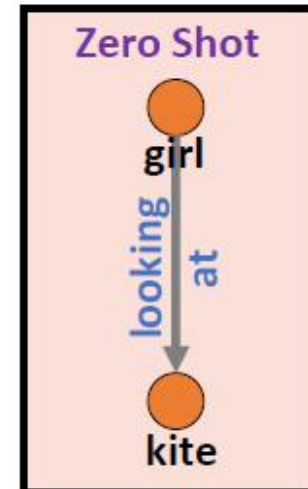
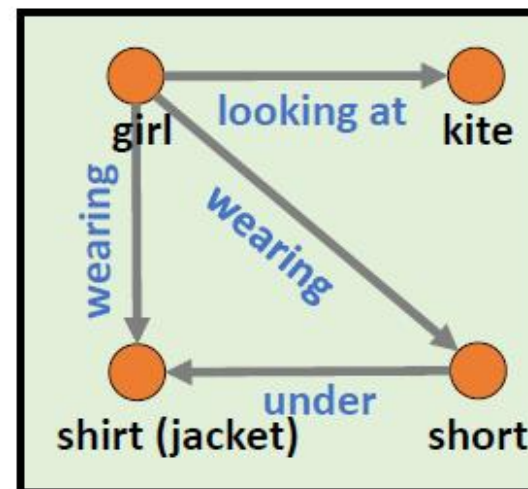
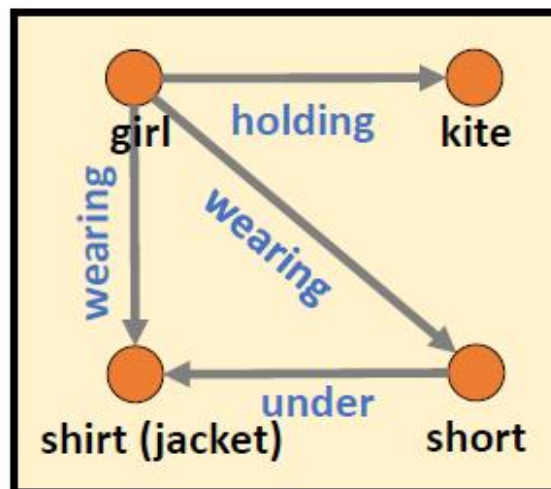
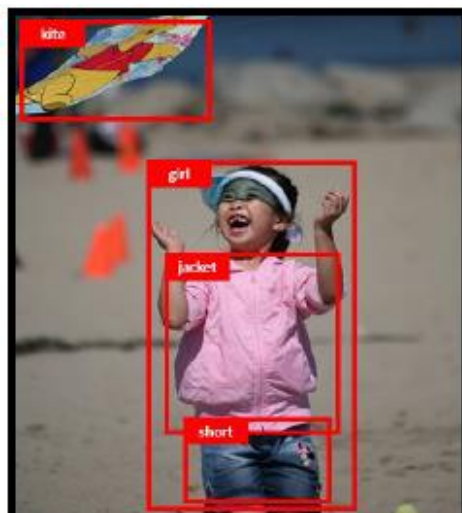
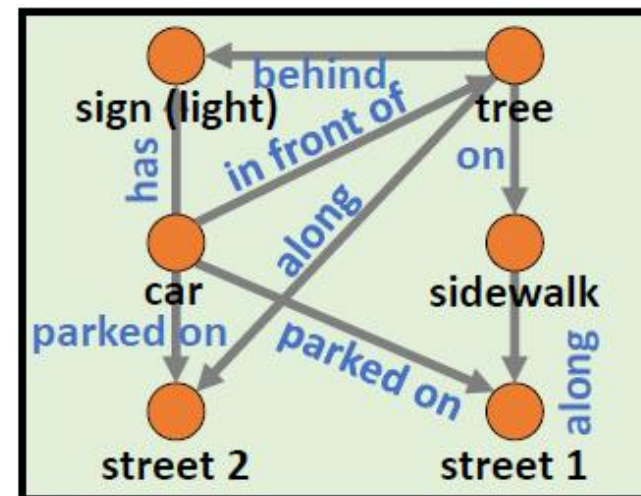
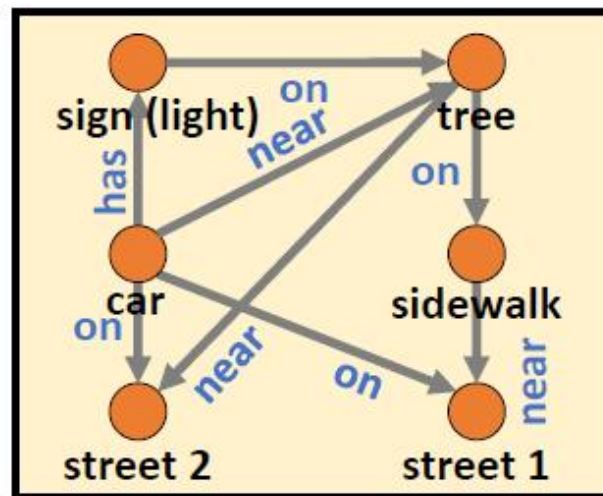
Zero-Shot Relationship Retrieval			PredCls	SGCls	SGDet
Model	Fusion	Method	R@50/100	R@50/100	R@50/100
MOTIFS <sup>†</sup>	SUM	Baseline	10.9 / 14.5	2.2 / 3.0	0.1 / 0.2
		Focal	10.9 / 14.4	2.2 / 3.1	0.1 / 0.3
		Reweight	0.7 / 0.9	0.1 / 0.1	0.0 / 0.0
		Resample	11.1 / 14.3	2.3 / 3.1	0.1 / 0.3
		X2Y	11.8 / 17.6	2.3 / 3.7	1.6 / 2.7
		X2Y-Tr	13.7 / 17.6	3.1 / 4.2	1.8 / 2.8
		TE	14.2 / 18.1	1.4 / 2.0	1.4 / 1.8
		NIE	2.4 / 3.2	0.2 / 0.4	0.3 / 0.6
		TDE	<b>14.4 / 18.2</b>	<b>3.4 / 4.5</b>	<b>2.3 / 2.9</b>
	GATE	Baseline	7.4 / 10.6	0.9 / 1.3	0.2 / 0.4
		TDE	<b>7.7 / 11.0</b>	<b>1.9 / 2.6</b>	<b>1.9 / 2.5</b>
VTransE <sup>†</sup>	SUM	Baseline	11.3 / 14.7	2.5 / 3.3	0.8 / 1.5
		TDE	<b>13.3 / 17.6</b>	<b>2.9 / 3.8</b>	<b>2.0 / 2.7</b>
	GATE	Baseline	4.2 / 5.9	1.9 / 2.6	<b>1.9 / 2.6</b>
		TDE	<b>5.3 / 7.9</b>	<b>2.1 / 3.0</b>	<b>1.9 / 2.7</b>
VCTree <sup>†</sup>	SUM	Baseline	10.8 / 14.3	1.9 / 2.6	0.2 / 0.7
		TDE	<b>14.3 / 17.6</b>	<b>3.2 / 4.0</b>	<b>2.6 / 3.2</b>
	GATE	Baseline	4.4 / 6.8	2.5 / 3.3	1.8 / 2.7
		TDE	<b>5.9 / 8.1</b>	<b>3.0 / 3.7</b>	<b>2.2 / 2.8</b>

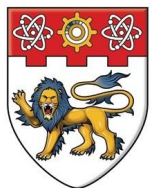
**Zero-Shot Recall@K**

Sentence-to-Graph Retrieval								
Gallery Size			1000			5000		
Model	Fusion	Method	R@20	R@100	Med	R@20	R@100	Med
MOTIFS <sup>†</sup>	SUM	Baseline	11.6	39.9	155	3.1	12.1	708
		Focal	10.9	39.0	163	2.9	11.1	737
		Reweight	9.7	36.8	159	3.0	11.4	725
		Resample	13.1	43.6	124	2.5	13.4	593
		X2Y	14.3	44.8	125	3.5	14.6	556
		X2Y-Tr	14.5	45.6	114	3.9	16.8	525
		TE	15.9	49.9	100	4.4	16.9	469
		NIE	6.7	29.2	202	1.6	8.6	1050
		TDE	<b>17.0</b>	<b>53.6</b>	<b>91</b>	<b>5.2</b>	<b>18.9</b>	<b>425</b>
	GATE	Baseline	13.7	45.6	143	4.4	16.2	618
		TDE	<b>20.8</b>	<b>59.2</b>	<b>72</b>	<b>5.2</b>	<b>21.3</b>	<b>325</b>
VTransE <sup>†</sup>	SUM	Baseline	12.3	42.3	129	<b>3.6</b>	15.0	596
		TDE	<b>14.7</b>	<b>48.4</b>	<b>106</b>	<b>3.6</b>	<b>16.3</b>	<b>483</b>
	GATE	Baseline	12.9	41.8	136	3.8	14.3	634
		TDE	<b>18.5</b>	<b>50.4</b>	<b>110</b>	<b>4.5</b>	<b>19.1</b>	<b>486</b>
VCTree <sup>†</sup>	SUM	Baseline	9.9	37.4	150	3.1	11.5	745
		TDE	<b>19.0</b>	<b>57.0</b>	<b>82</b>	<b>5.0</b>	<b>20.0</b>	<b>385</b>
	GATE	Baseline	13.4	44.1	121	3.7	13.6	583
		TDE	<b>19.1</b>	<b>55.5</b>	<b>87</b>	<b>5.1</b>	<b>20.3</b>	<b>395</b>

**SG-Based Retrieval**

# Qualitative Examples





**NANYANG**  
TECHNOLOGICAL  
UNIVERSITY

**DAMO**  
ALIBABA DAMO ACADEMY



# Thank You

**Code Link**



**Paper Link**

