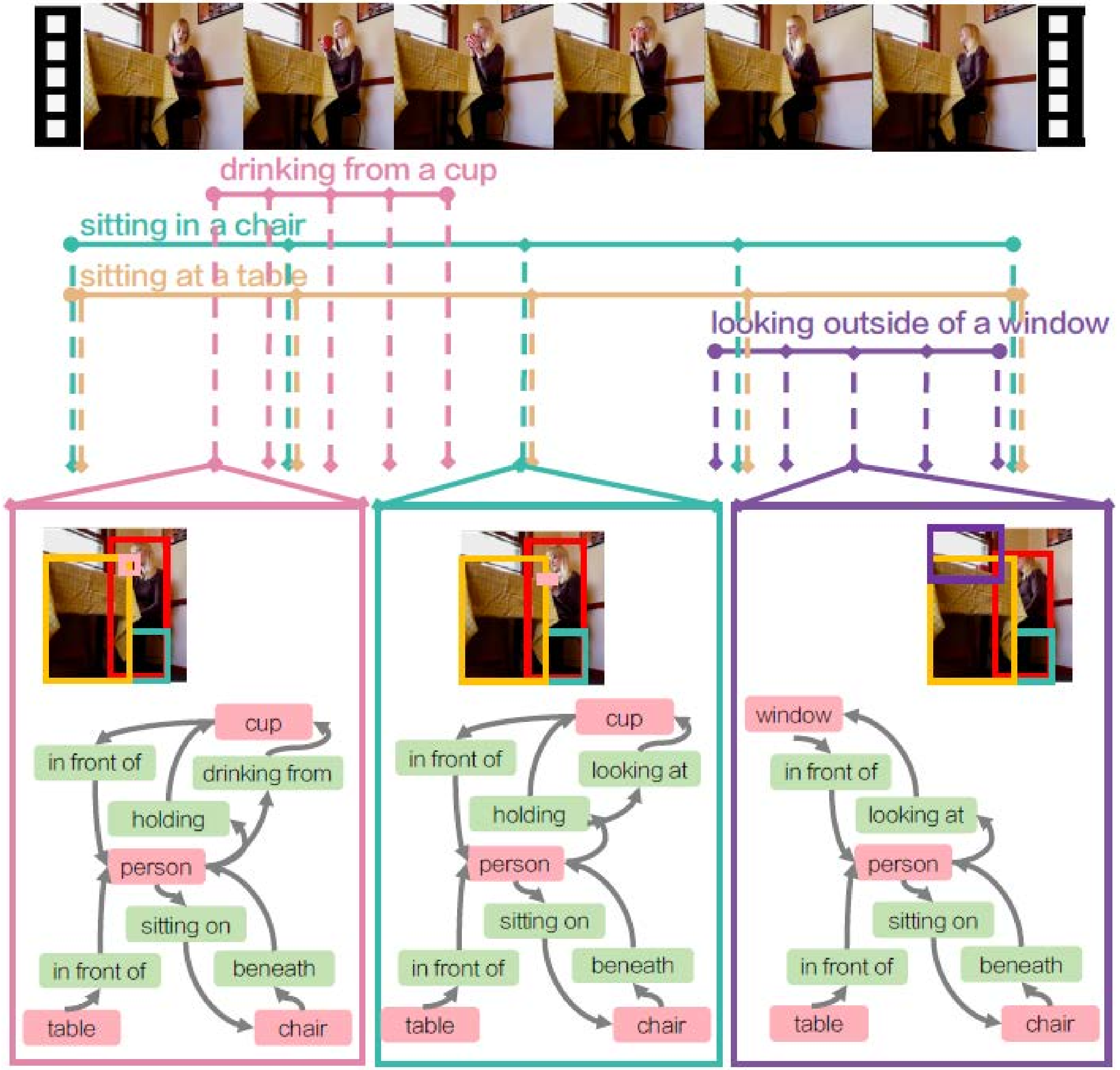
静态场景图生成中数据集的长尾分布问题

汇报人：崔奕宸

2021/4/14

工作介绍



Method	Backbone	Pre-train	mAP
I3D + NL [10, 72]	R101-I3D-NL	Kinetics-400	37.5
STRG [73]	R101-I3D-NL	Kinetics-400	39.7
Timeception [31]	R101	Kinetics-400	41.1
SlowFast [23]	R101	Kinetics-400	42.1
SlowFast+NL [23, 72]	R101-NL	Kinetics-400	42.5
LFB [75]	R101-I3D-NL	Kinetics-400	42.5
SGFB (ours)	R101-I3D-NL	Kinetics-400	**44.3**
SGFB Oracle (ours)	R101-I3D-NL	Kinetics-400	**60.3**

ground truth scene graph

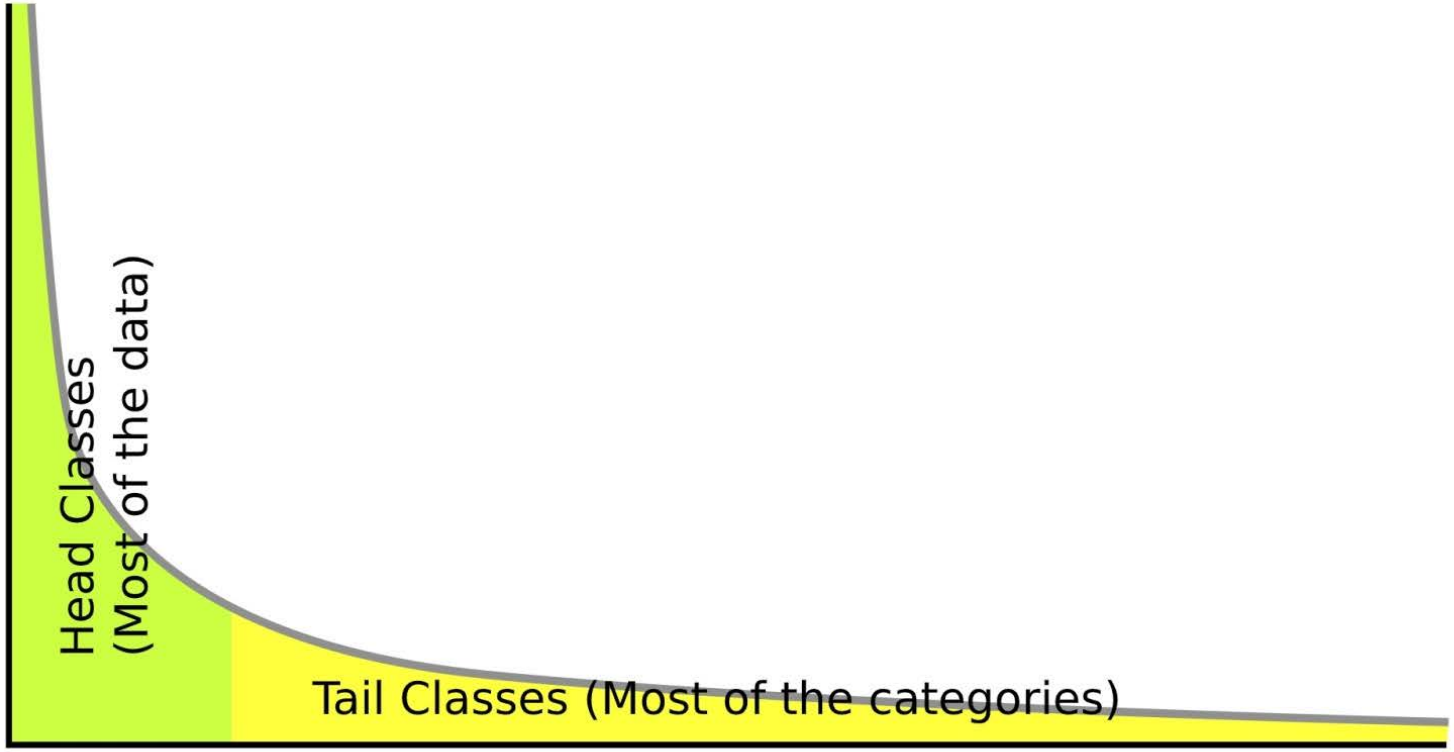
视频抽帧 → 有时序关系的多个静态场景图 → 视频中人的行为

工作介绍

四个优化方向

Methods	Efficient Graph Features Refinement	Efficient Graph Generation	Long-tailed Dataset Distribution	Efficient Loss Function Definition
Deep Relational Networks	✓			
Iterative Message Passing	✓			
MSDN	✓			
Factorizable Net		✓		
Neural Motifs	✓		✓	
Graph R-CNN	✓	✓		
Knowledge-embedded Routing Network	✓		✓	
Graph Contrastive Losses				✓
VCTree	✓		✓	
External Knowledge and Image Reconstruction	✓		✓	
Prediction with Limited Labels			✓	
Counterfactual Critic Multi-agent Training				✓
Visual Relationships as Functions			✓	
Unbiased SGG from Biased Training			✓	

什么是长尾问题？



- 自然情况下，数据呈现**长尾分布**
- 训练数据的分布受到**人工的均衡**
- 直接利用长尾数据会**对头部数据过拟合**，在**预测时忽略尾部的类别**。

什么是长尾问题？

- 2010年就有论文涉及长尾问题：

  [2011 CVPR]Unbiased look at dataset bias

- 长尾问题是较热门的研究方向

  1. Scene Graph Generation
  2. Multi-Label Classification
  3. Object Detection
  4. Instance Segmentation

  … …

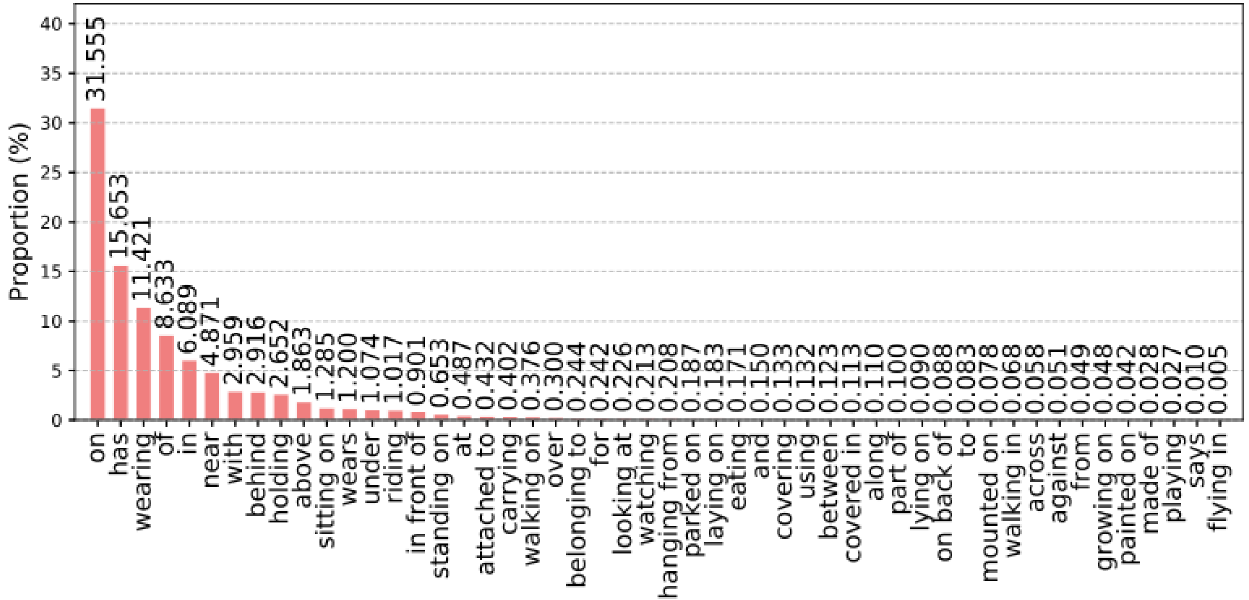
2020

Title	Venue	Type	Code	Star
Rethinking the Value of Labels for Improving Class-Imbalanced Learning	NeurIPS	Other	PyTorch(Author)	153
Balanced Meta-Softmax for Long-Tailed Visual Recognition	NeurIPS	CLW	PyTorch(Author)	
Long-Tailed Classification by Keeping the Good and Removing the Bad Momentum Causal Effect	NeurIPS	Other	PyTorch(Author)	
Forest R-CNN: Large-Vocabulary Long-Tailed Object Detection and Instance Segmentation	ACM-MM	Other	PyTorch(Author)	
Mitigating Dataset Imbalance via Joint Generation and Classification	ECCV-W	Other	PyTorch(Author)	
Seesaw Loss for Long-Tailed Instance	ECCV-W	Other	-	
Balanced Activation for Long-tailed Visual Recognition	ECCV-W	Other	-	
Imbalanced Continual Learning with Partitioning Reservoir Sampling	ECCV	Other	PyTorch(Author)	
Feature Space Augmentation for Long-Tailed Data	ECCV	Aug	-	
The Devil is in Classification A Simple Framework for Long-tail Instance Segmentation	ECCV	Aug	-	
Distribution-Balanced Loss for Multi-Label Classification in Long-Tailed Datasets	ECCV	CLW	PyTorch(Author)	91
Solving Long-tailed Recognition with Deep Realistic Taxonomic Classifier	ECCV	Other	-	
Learning From Multiple Experts_Self-paced Knowledge Distillation for Long-tailed Classification	ECCV	TL	-	
Rethinking Class-Balanced Methods for Long-Tailed Visual Recognition from a Domain Adaptation Perspective	CVPR	CLW	-	
Equalization Loss for Long-Tailed Object Recognition	CVPR	CLW SLW	PyTorch(Author)	116
Domain Balancing: Face Recognition on Long-Tailed Domains	CVPR	Other	-	
BBN: Bilateral-Branch Network with Cumulative Learning	CVPR	Other	PyTorch(Author)	360

场景图生成中的长尾问题

问题1：数据集的长尾分布

- 头部是大量简单关系、尾部是类别繁多的复杂关系

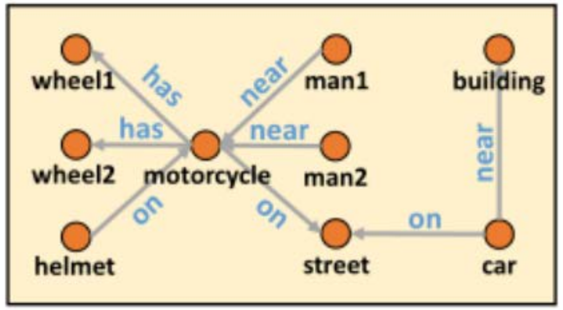


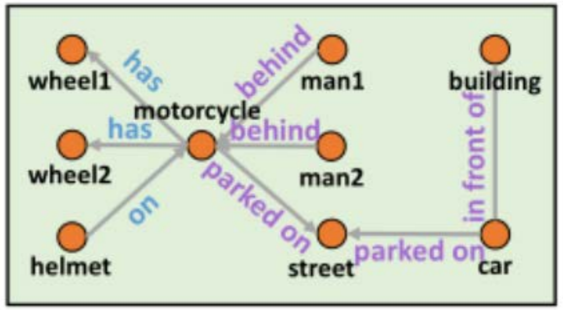
Benchmark Dataset Visual Genome中不同关系的数量分布

问题2：评价指标

- SGG metrics： Recall@k（R@k）
- 只要预测对最高频的几个关系类别performance就很高， 而完全不考虑低频类

- **结果**





Methods

A. 统计信息（Statistical Information）

- [CVPR 2018] Neural Motifs: Scene Graph Parsing with Global Context
- [CVPR 2019] Knowledge-Embedded Routing Network for Scene Graph Generation
- [CVPR 2019] External Knowledge and Image Reconstruction

B. 半监督学习和小样本学习

- [CVPR 2019] Learning to Compose Dynamic Tree Structures for Visual Contexts
- [ICCV 2019] Prediction with Limited Labels
- [ICCV 2019] Visual Relationships as Functions

C. 因果推理（Causal Inference）

- [CVPR 2020 oral] Unbiased Scene Graph Generation from Biased Training

Neural Motifs:
Scene Graph Parsing with Global Context

[CVPR 2018]

Rowan Zellers[1] Mark Yatskar[1,2] Sam Thomson[3] Yejin Choi[1,2]
[1]Paul G. Allen School of Computer Science & Engineering, University of Washington
[2]Allen Institute for Artificial Intelligence
[3]School of Computer Science, Carnegie Mellon University

Neural Motifs [CVPR 2018]

- 第一篇关注到Visual Genome数据集存在bias的论文
- 没有直接去解决长尾问题，但提出了一个简单有效的baseline，充分说明了数据集中存在bias

- 主要工作
  1. **深入分析Visual Genome数据集的统计信息**
  2. **根据统计信息提出了一个极为简单的baseline，表现优于state-of-art方法**
  3. 提出一个新的模型Stacked Motif Network（SMN）

Neural Motifs [CVPR 2018] – VG数据集分析

- VG数据集中object和关系的类别、实例数量

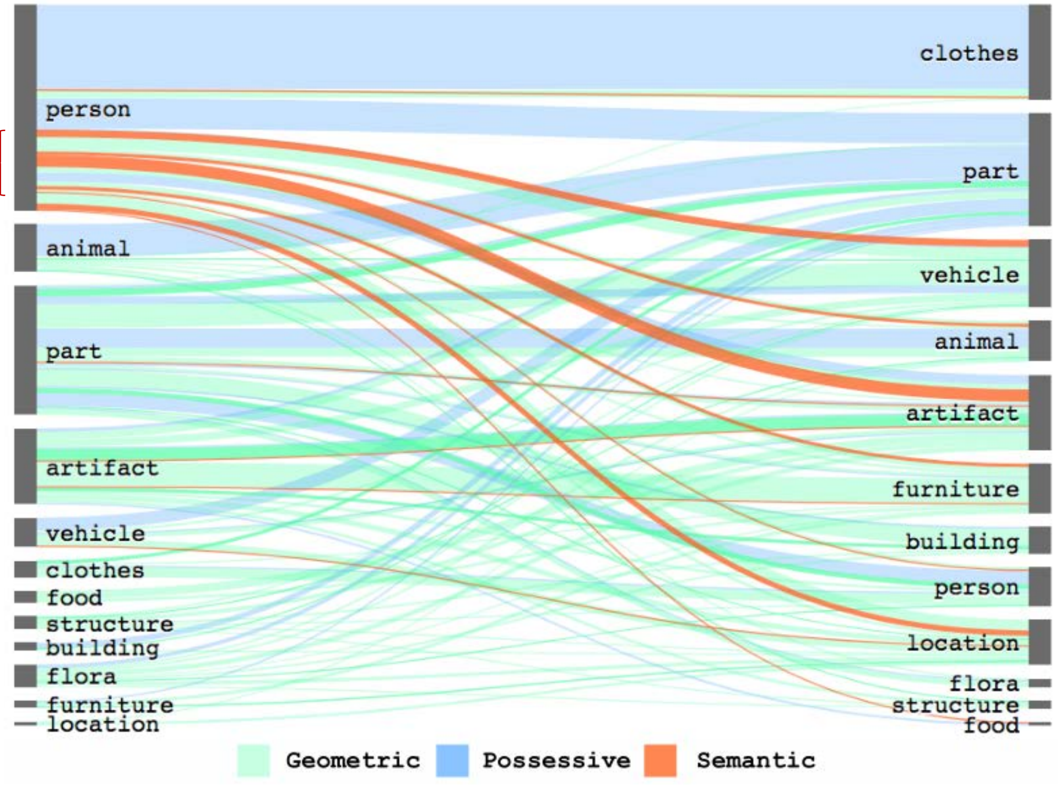
90.9%的关系是Geometric或Possessive类别

Type	Examples	Classes	Instances
	Entities		
Part	arm, tail, wheel	32	200k (25.2%)
Artifact	basket, fork, towel	34	126k (16.0%)
Person	boy, kid, woman	13	113k (14.3%)
Clothes	cap, jean, sneaker	16	91k (11.5%)
Vehicle	airplane, bike, truck,	12	44k (5.6%)
Flora	flower, plant, tree	3	44k (5.5%)
Location	beach, room, sidewalk	11	39k (4.9%)
Furniture	bed, desk, table	9	37k (4.7%)
Animal	bear, giraffe, zebra	11	30k (3.8%)
Structure	fence, post, sign	3	30k (3.8%)
Building	building, house	2	24k (3.1%)
Food	banana, orange, pizza	6	13k (1.6%)
	Relations		
Geometric	above, behind, under	15	228k (50.0%)
Possessive	has, part of, wearing	8	186k (40.9%)
Semantic	carrying, eating, using	24	39k (8.7%)
Misc	for, from, made of	3	2k (0.3%)

24/40的关系类别是Semantic，仅拥有8.7%的关系实例。

- 进一步可视化VG数据集的object类型和关系类型的分布

part和clothes几乎都是Possessive关系



几乎所有Semantic关系由人主导的

Furniture和building类型object几乎都是Geometric关系

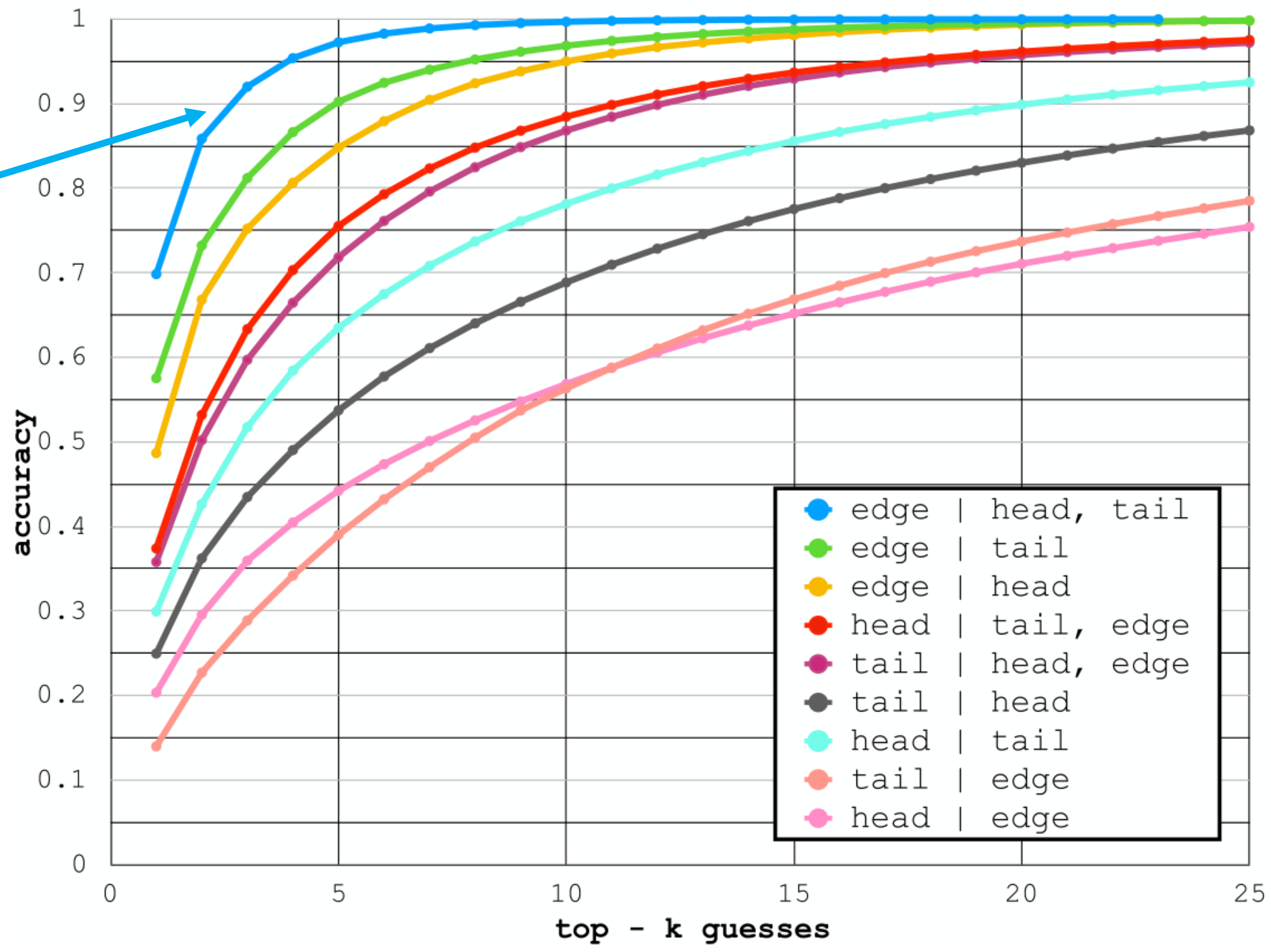
表明常识先验（common sense priors）对于预测关系有着重要作用

Neural Motifs [CVPR 2018] – VG数据集分析

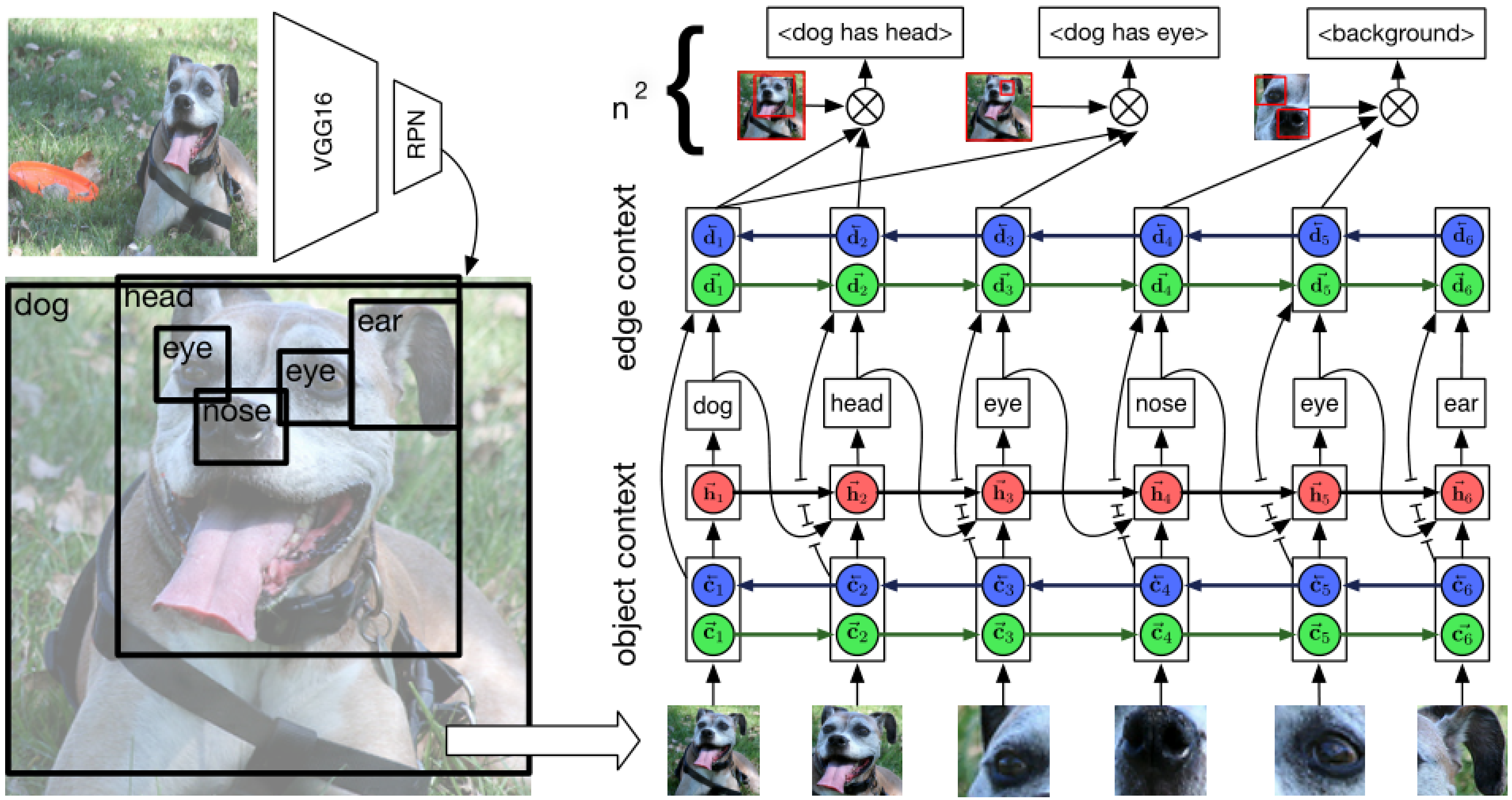
- 先验概率测试

给定(subject, object)的label，关系的label是高度确定的

证明先验知识在预测关系时至关重要。



Neural Motifs [CVPR 2018] – Stacked Motif Network（MotifNet）



Neural Motifs [CVPR 2018] – 实验

- 数据集

  Visual Genome数据集，遵循广泛采用的分割方式，即包含150个最常见的object类别和50个最常见的predicate类别。

- Metrics

  Recall@k（R@k）

- Sub task

  PredCls

  SGCls

  SGDet（SGGen）

Neural Motifs [CVPR 2018] – 实验

	Model	Scene Graph Detection			Scene Graph Classification			Predicate Classification			Mean
		R@20	R@50	R@100	R@20	R@50	R@100	R@20	R@50	R@100	
models	VRD [29]		0.3	0.5		11.8	14.1		27.9	35.0	14.9
	MESSAGE PASSING [47]		3.4	4.2		21.7	24.4		44.8	53.0	25.3
	MESSAGE PASSING+	14.6	20.7	24.5	31.7	34.6	35.4	52.7	59.3	61.3	39.3
	ASSOC EMBED [31]⋆	6.5	8.1	8.2	18.2	21.8	22.6	47.9	54.1	55.4	28.3
	FREQ	17.7	23.5	27.6	27.7	32.4	34.0	49.4	59.9	64.1	40.2
	FREQ+OVERLAP	20.1	26.2	30.1	29.3	32.3	32.9	53.6	60.6	62.2	40.7
	MOTIFNET-LEFTRIGHT	21.4	27.2	30.3	**32.9**	**35.8**	**36.5**	**58.5**	**65.2**	**67.1**	**43.6**
ablations	MOTIFNET-NOCONTEXT	21.0	26.2	29.0	31.9	34.8	35.5	57.0	63.7	65.6	42.4
	MOTIFNET-CONFIDENCE	**21.7**	**27.3**	**30.5**	32.6	35.4	36.1	58.2	65.1	67.0	43.5
	MOTIFNET-SIZE	21.6	**27.3**	30.4	32.2	35.0	35.7	58.0	64.9	66.8	43.3
	MOTIFNET-RANDOM	21.6	**27.3**	30.4	32.5	35.5	36.2	58.1	65.1	66.9	43.5

- 引入两个frequency baseline
  - **FREQ：使用P(Relation | Subject, Object)预测关系**
  - FREQ-OVERLAP：在FREQ的基础上，要求两个RoI相交，才算这两个object有关系

Scene Graph Generation with External Knowledge and Image Reconstruction

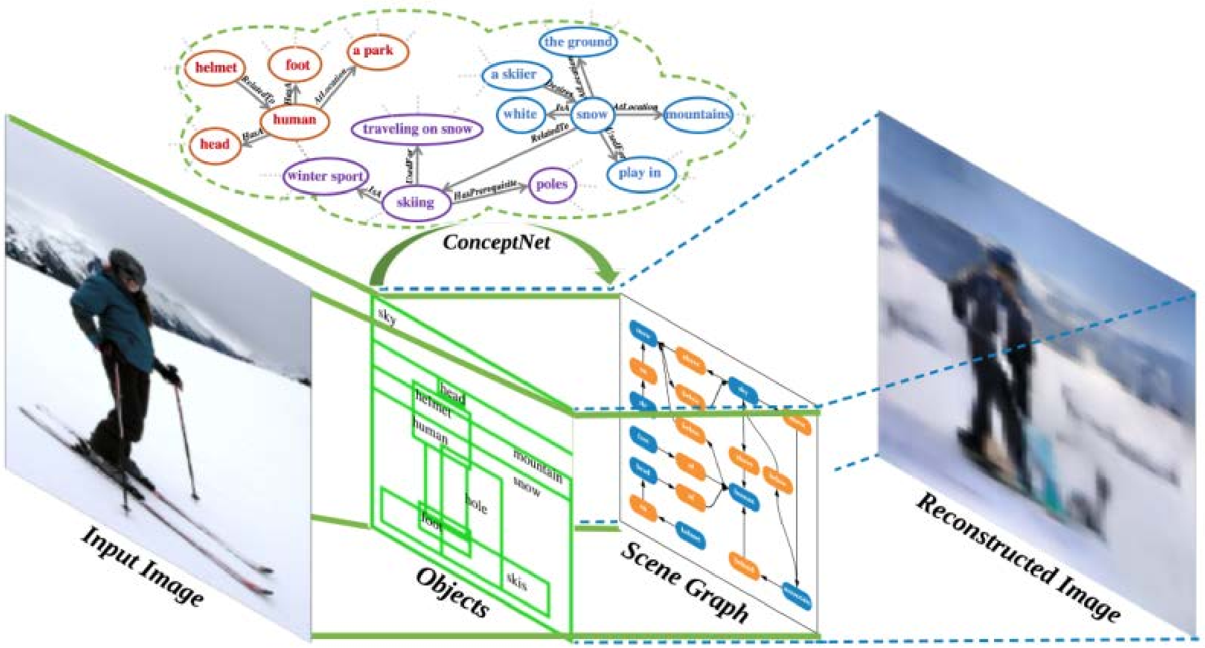
[CVPR 2019]

Jiuxiang Gu[1]*, Handong Zhao[2], Zhe Lin[2], Sheng Li[3], Jianfei Cai[1], Mingyang Ling[4]
[1] ROSE Lab, Interdisciplinary Graduate School, Nanyang Technological University, Singapore
[2] Adobe Research, USA [3] University of Georgia, USA [4] Google Cloud AI, USA
{jgu004, asjfcai}@ntu.edu.sg, {hazhao, zlin}@adobe.com
sheng.li@uga.edu, mingyangling@google.com

SGG with External Knowledge and Image Reconstruction [CVPR 2019]

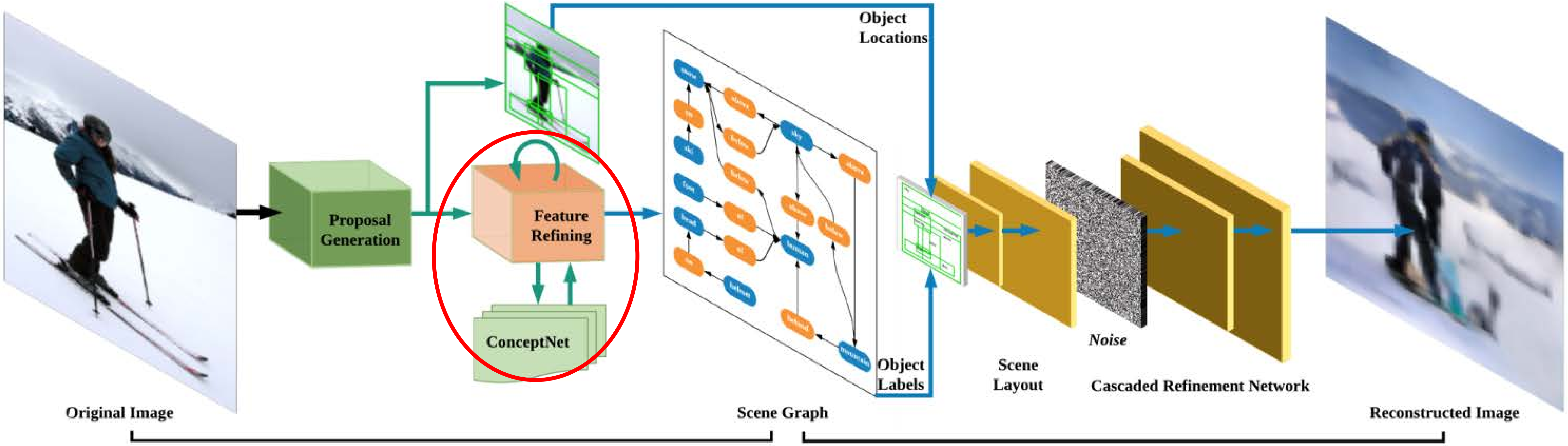
- 本文尝试将统计信息实际应用到模型中，使用的是外部知识库的方法。

- **主要工作**
    1. 提出了一个新模型，引入了外部知识库利用偏差，同时，为了解决无意义和缺失标注，引入了一个辅助图像重建路径。



SGG with External Knowledge and Image Reconstruction [CVPR 2019]

- **Pipeline**



$$Knowledge\ Bases \begin{cases} manual\ effort\ (e.g., Wikipedia, DBpedia) \\ automatic\ extraction (e.g., ConceptNet) \end{cases}$$

SGG with External Knowledge and Image Reconstruction [CVPR 2019]

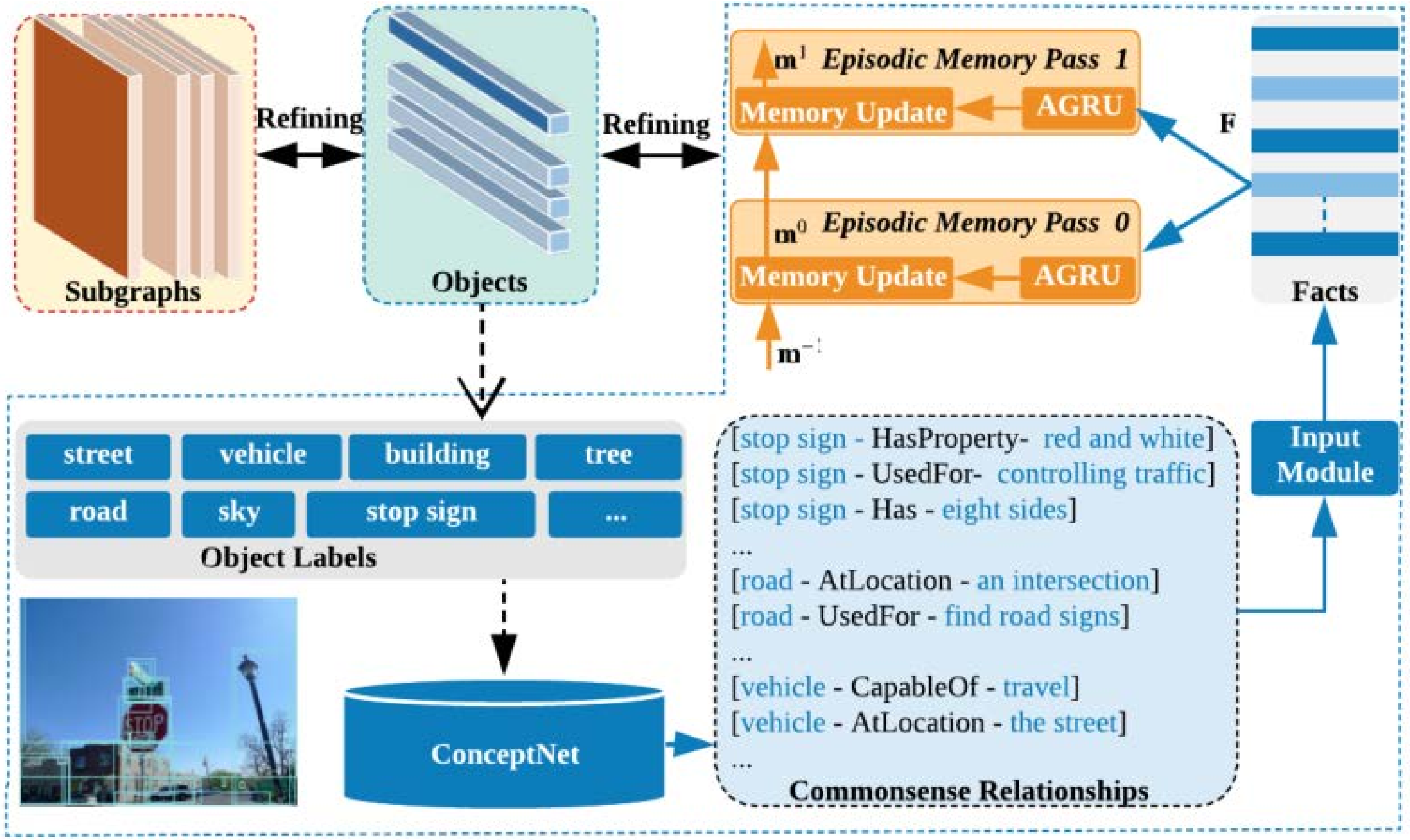
$$o_i \rightarrow a_i \xrightarrow{retrieve} \langle a_i, a_{i,j}^r, a_j^o, w_{i,j} \rangle, j \in [0, K-1]$$

$$\langle a_i, a_{i,j}^r, a_j^o, w_{i,j} \rangle \rightarrow [X^0, \dots, X^{T_a - 1}]$$

$$x^t = W_e X^t$$

$$h_k^t = RNN(x_k^t, h_k^{t-1}), t \in [0, T_a - 1]$$

$$h_k^{T_a - 1} \rightarrow f_k^i \rightarrow F$$



SGG with External Knowledge and Image Reconstruction [CVPR 2019]

Dataset	Model	PhrDet		SGGen	
		Rec@50	Rec@100	Rec@50	Rec@100
VRD [29]	ViP-CNN [27]	22.78	27.91	17.32	20.01
	DR-Net [5]	19.93	23.45	17.73	20.88
	U+W+SF+LK: T+S [45]	26.32	29.43	19.17	21.34
	Factorizable Net [25]	26.03	30.77	18.32	21.20
	KB-GAN	**27.39**	**34.38**	**20.31**	**25.01**
VG-MSDN [26]	ISGG [41]	15.87	19.45	8.23	10.88
	MSDN [26]	19.95	24.93	10.72	14.22
	Graph R-CNN [42]	–	–	11.40	13.70
	Factorizable Net [25]	22.84	28.57	13.06	16.47
	KB-GAN	**23.51**	**30.04**	**13.65**	**17.57**

Knowledge-Embedded Routing Network for Scene Graph Generation

[CVPR 2019]

Tianshui Chen
Sun Yat-Sen University
DarkMatter AI Research
tianshuichen@gmail.com

Weihao Yu
Sun Yat-Sen University
weihaoyu6@gmail.com

Riquan Chen
Sun Yat-Sen University
sysucrq@gmail.com

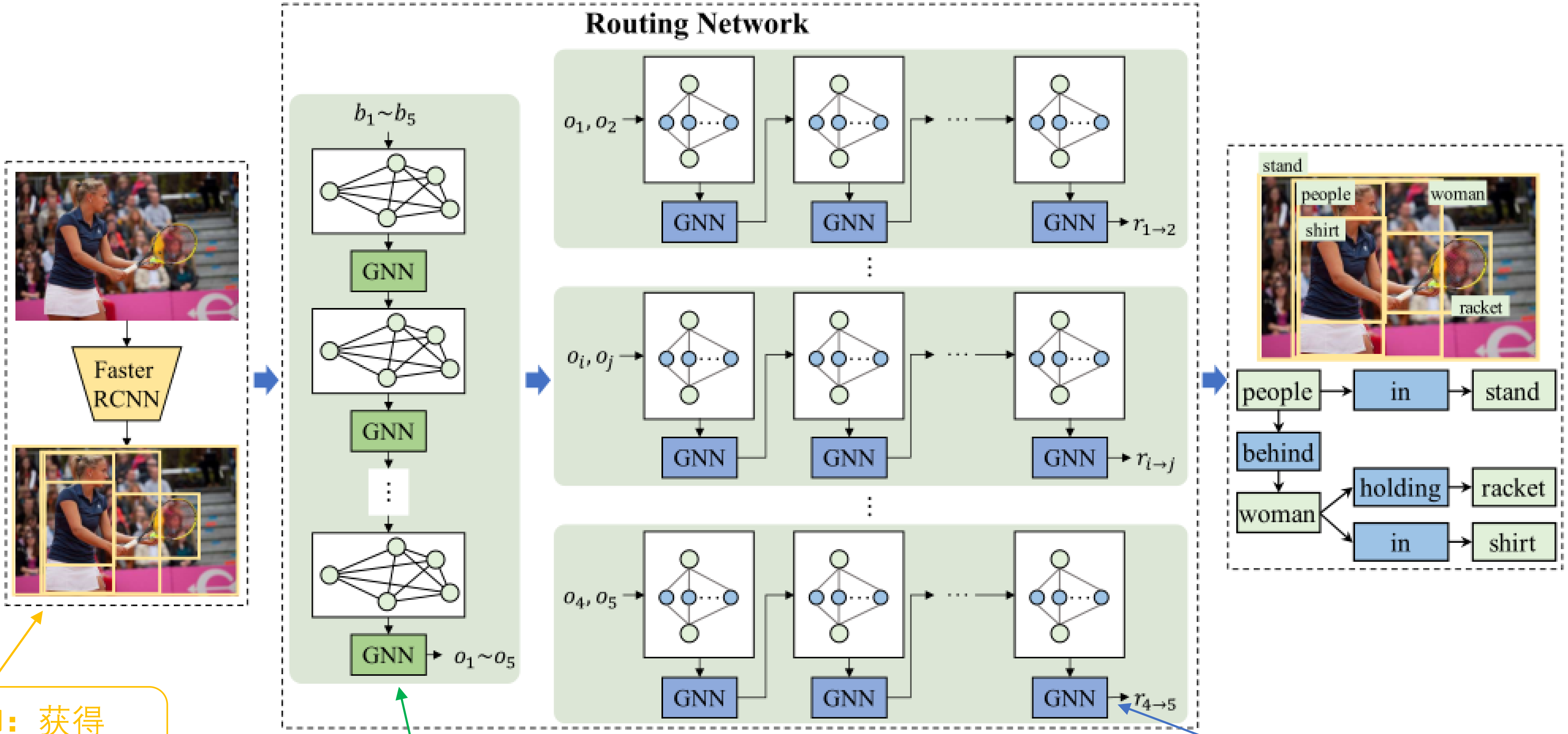
Liang Lin*
Sun Yat-Sen University
DarkMatter AI Research
linliang@ieee.org

KERN [CVPR 2019]

- **Neural Motifs**是先前作品中表现最好的方法，但**没有将统计信息实际应用到模型中**。
- **KERN则是第一个在其pipeline中明确建模统计相关性（统计信息），**为解决长尾问题迈出了第一步。

- **主要工作**
  1. 提出了一个新模型Knowledge-Embedded Routing Network， 明确建模统计相关性
  2. 提出了新的Metrics——Mean Recall@k

KERN [CVPR 2019] - Knowledge-Embedded Routing Network

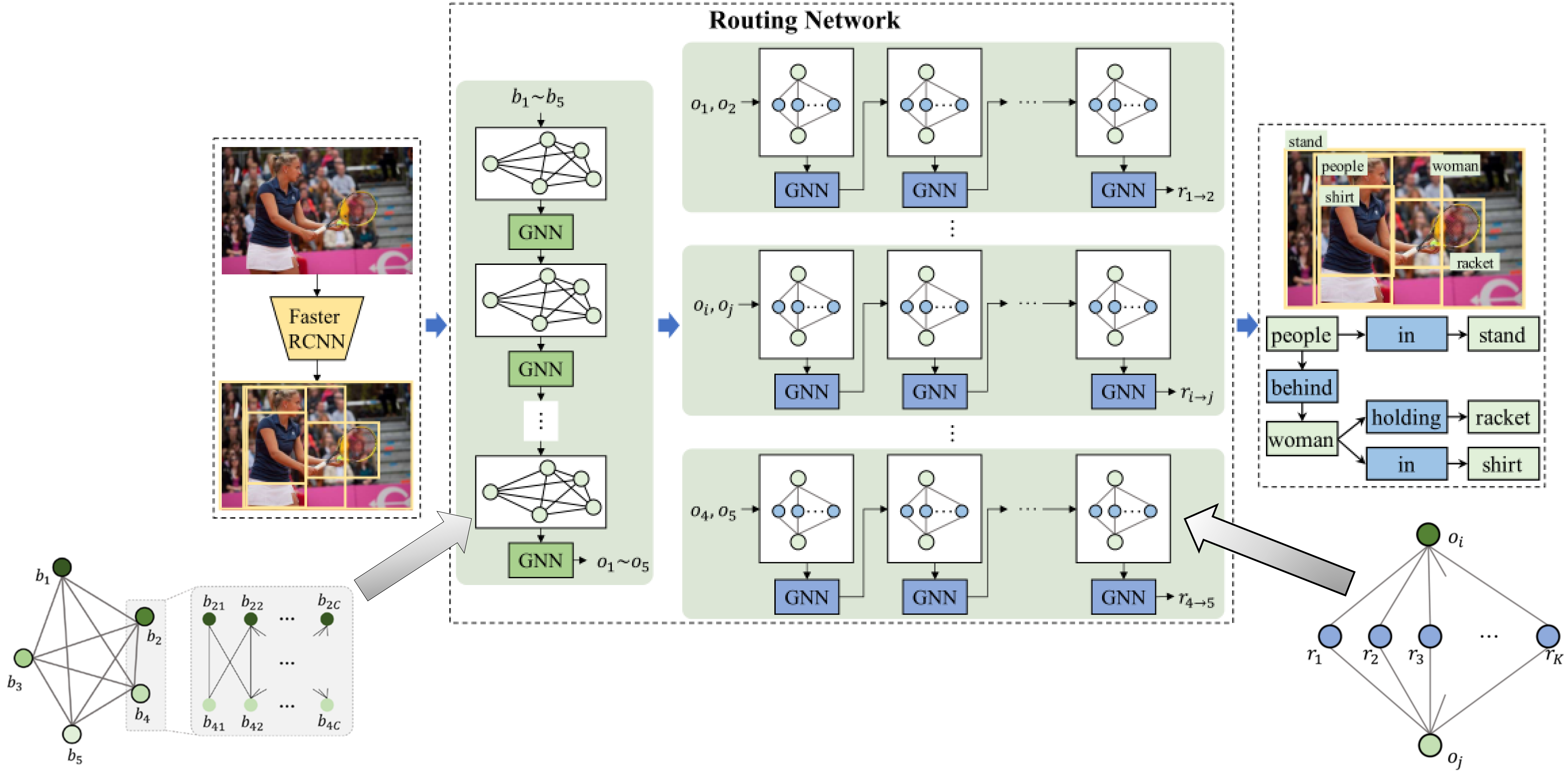


Faster RCNN： 获得 region proposal

构建一个图来关联这些候选区域，并使用GNN预测每个区域的类别标签。

构建另一个图将(subject, object)与所有可能的关系相关联，并使用GNN预测关系类别。

KERN [CVPR 2019] - Knowledge-Embedded Routing Network



模型构建这两个图时，进行了Knowledge-Embedded（建模统计信息）

KERN [CVPR 2019] - Knowledge-Embedded Routing Network

根据object的统计相关性矩阵M_c来关联节点。

- $\boldsymbol{M_C}$

  c'类和c类object共存概率：$m_{cc'}$

  统计所有共存概率，得：$M_C \in R^{C\times C}$；

  C：object类别的数量

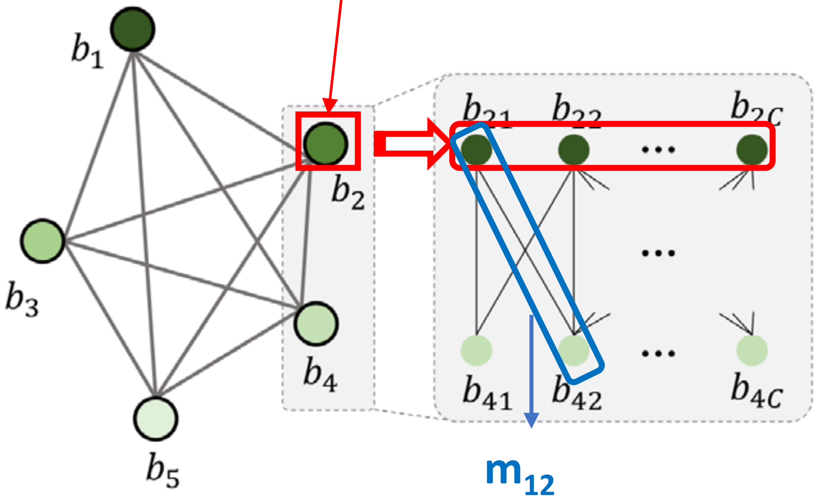
- **关联候选区域**

  区域b_i复制C次：节点C $= \{b_{i1}, b_{i2}, \ldots, b_{iC}\}$

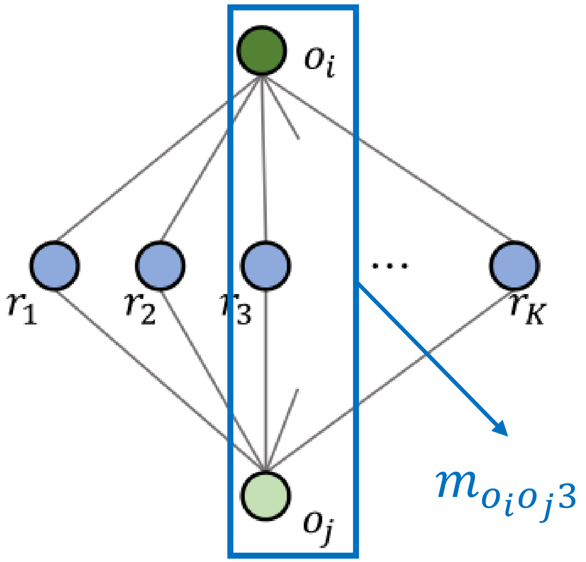
  b_{ic}：区域b_i与类别c的相关性

  $m_{cc'}$用于将节点b_{ic}和节点$b_{jc'}$关联起来

- gated GNN：$o_1, o_2, \ldots$



以**结构化图的形式**表示(subject, object)及其关系的统计相关性。



- **构建结构化图**

  类别c和c'的object之间**所有可能存在的关系的概率**

  $\{m_{cc'1}, m_{cc'2}, \ldots, m_{cc'K}\}$。

  $m_{o_i o_j k}$：o_i、o_j和关系k之间的相关性

- 经过gated GNN，获得最终预测结果

KERN [CVPR 2019] – 实验

- 数据集

  Visual Genome数据集，遵循广泛采用的分割方式，即包含150个最常见的object类别和50个最常见的predicate类别。

- Metrics

  Recall@k、**mean Recall@k**

- **约束条件**

  有约束（Constraint）

  无约束（No constraint）

- Sub task

  PredCls

  SGCls

  SGDet（SGGen）

KERN [CVPR 2019] – 实验

mR@k

分别为每个关系的样本计算R@k，然后对所有关系的R@k进行平均

- mR@k

	Method	SGGen		SGCls		PredCls		
		mR@50	mR@100	mR@50	mR@100	mR@50	mR@100	Mean
Constraint	IMP [30]	0.6	0.9	3.1	3.8	6.1	8.0	3.8
	IMP+ [30, 33]	3.8	4.8	5.8	6.0	9.8	10.5	6.8
	FREQ [33]	4.3	5.6	6.8	7.8	13.3	15.8	8.9
	SMN [33]	5.3	6.1	7.1	7.6	13.3	14.4	9.0
	Ours	**6.4**	**7.3**	**9.4**	**10.0**	**17.7**	**19.2**	**11.7**
Unconstraint	AE [23]	1.6	2.5	6.0	7.8	15.1	19.5	8.8
	IMP+ [30, 33]	5.4	8.0	12.1	16.9	20.3	28.9	15.3
	FREQ [33]	5.9	8.9	13.5	19.6	24.8	37.3	18.3
	SMN [33]	9.3	12.9	15.4	20.6	27.5	37.9	20.6
	Ours	**11.7**	**16.0**	**19.8**	**26.2**	**36.3**	**49.0**	**26.5**

- R@k

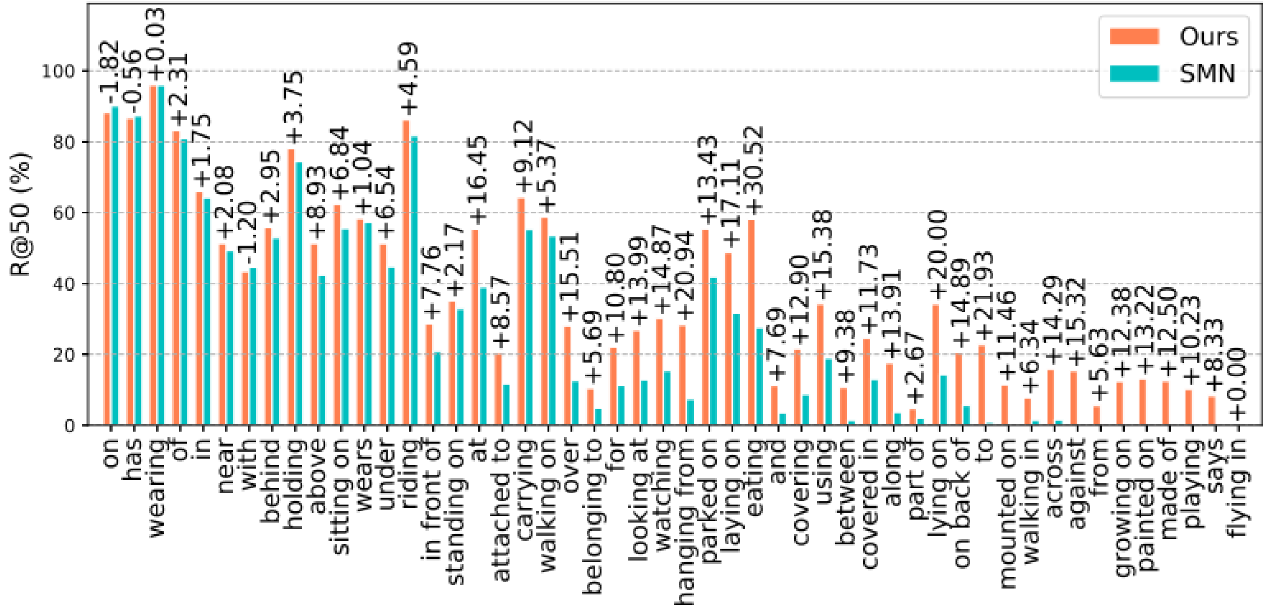
	Methods	SGGen		SGCls		PredCls		
		R@50	R@100	R@50	R@100	R@50	R@100	Mean
Constraint	VRD [19]	0.3	0.5	11.8	14.1	27.9	35.0	14.9
	IMP [30]	3.4	4.2	21.7	24.4	44.8	53.0	25.3
	IMP+ [30, 33]	20.7	24.5	34.6	35.4	59.3	61.3	39.3
	FREQ [33]	23.5	27.6	32.4	34.0	59.9	64.1	40.3
	SMN [33]	**27.2**	**30.3**	35.8	36.5	65.2	67.1	43.7
	Ours	27.1	29.8	**36.7**	**37.4**	**65.8**	**67.6**	**44.1**
No constraint	AE [23]	9.7	11.3	26.5	30.0	68.0	75.2	36.8
	IMP+ [30, 33]	22.0	27.4	43.4	47.2	75.2	83.6	49.8
	FREQ [33]	25.3	30.9	40.5	43.7	71.3	81.2	48.8
	SMN [33]	30.5	35.8	44.5	47.7	81.1	88.3	54.7
	Ours	**30.9**	**35.8**	**45.9**	**49.0**	**81.9**	**88.9**	**55.4**

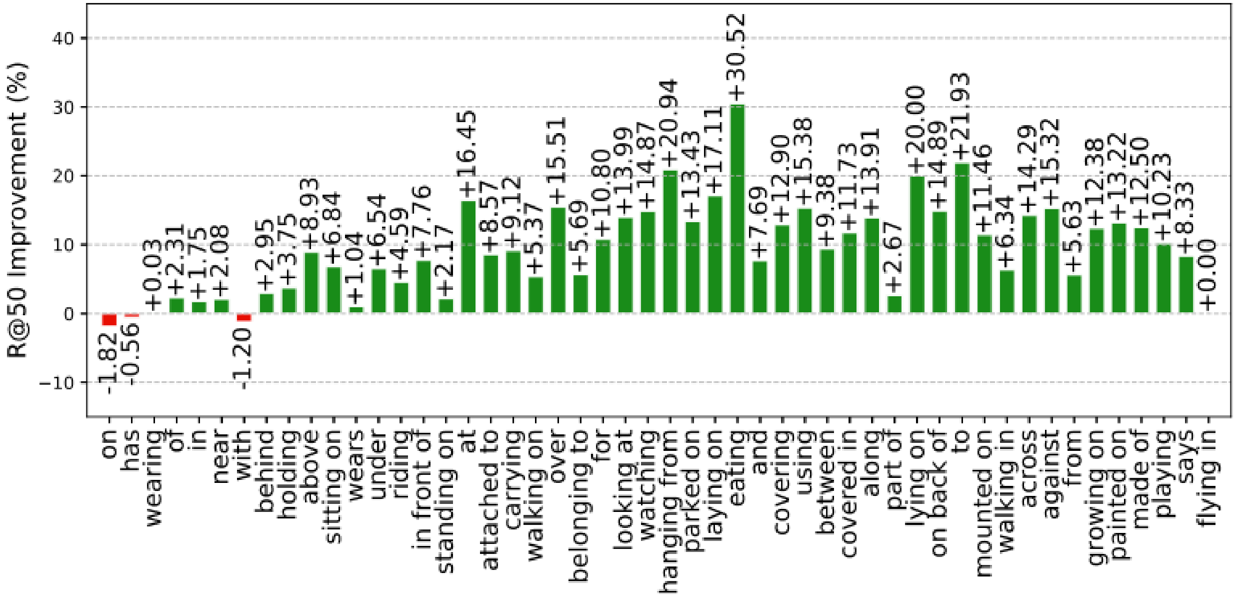
KERN [CVPR 2019] – 实验

- 消融实验

Methods	SGGen		SGCls		PredCls		
	mR@50	mR@100	mR@50	mR@100	mR@50	mR@100	Mean
Ours w/o rk & w/o ok	5.1	5.8	6.1	6.5	10.5	11.5	7.6
Ours w/o rk	5.2	5.9	6.5	6.9	11.1	12.0	7.9
Ours	**6.4**	**7.3**	**9.4**	**10.0**	**17.7**	**19.2**	**11.7**
	R@50	R@100	R@50	R@100	R@50	R@100	Mean
Ours w/o rk & w/o ok	25.2	27.9	33.9	34.8	58.7	61.0	40.3
Ours w/o rk	25.5	28.0	34.3	35.2	59.2	61.5	40.6
Ours	**27.1**	**29.8**	**36.7**	**37.4**	**65.8**	**67.6**	**44.1**

- KERN VS Motifs





Visual Relationships as Functions: Enabling Few-Shot Scene Graph Prediction

[ICCV 2019]

Tianshui Chen
Sun Yat-Sen University
DarkMatter AI Research
tianshuichen@gmail.com

Weihao Yu
Sun Yat-Sen University
weihaoyu6@gmail.com

Riquan Chen
Sun Yat-Sen University
sysucrq@gmail.com

Liang Lin*
Sun Yat-Sen University
DarkMatter AI Research
linliang@ieee.org

Visual Relationships as Functions [ICCV 2019]

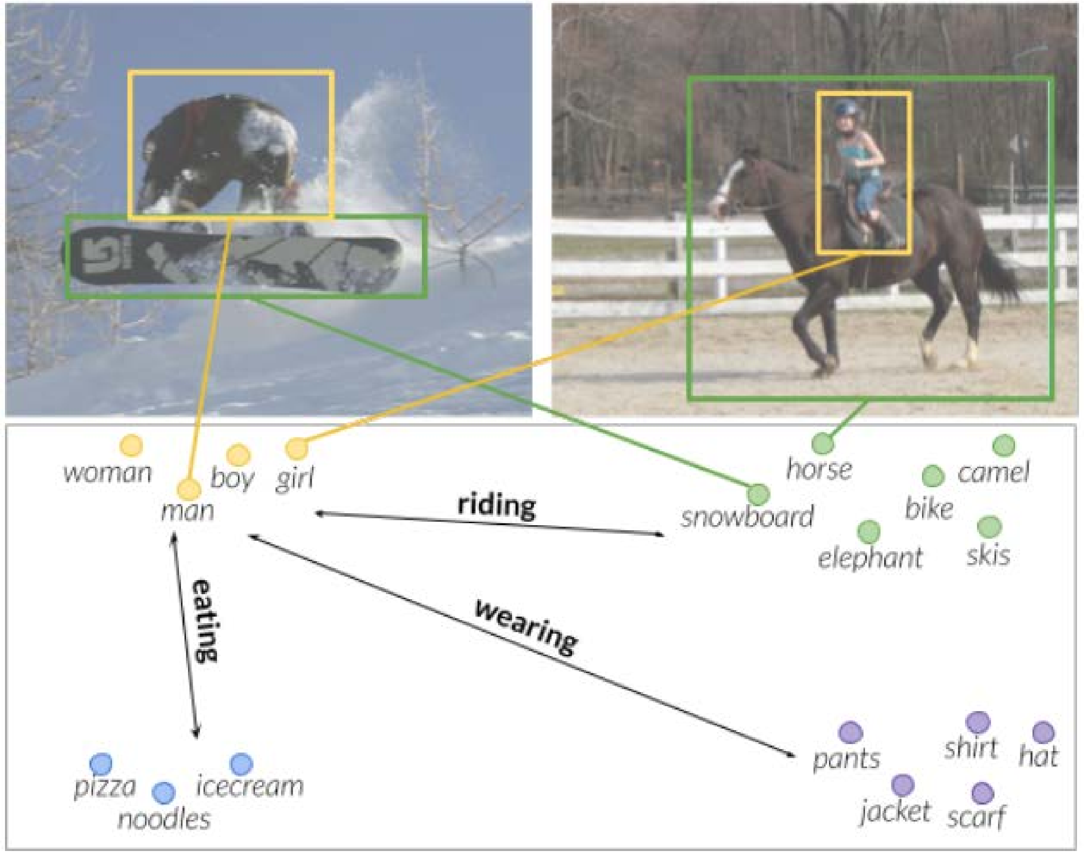
- 上述论文使用统计信息来获得期望的结果，本文和下一篇文章则使用了小样本学习和半监督技术来处理数据集的不平衡。

- **主要工作**
  1. 本文使用小样本学习方面的进展，**先从频繁的关系类别中学习对象表示，然后使用对象表示来预测稀有的关系类别**。
  2. 在一个graph convolution framework中学习对象表示，然后用对象表示来构建few-shot predicate classifiers

Visual Relationships as Functions [ICCV 2019]

<person – riding – horse>

<person – ? – snowboard>

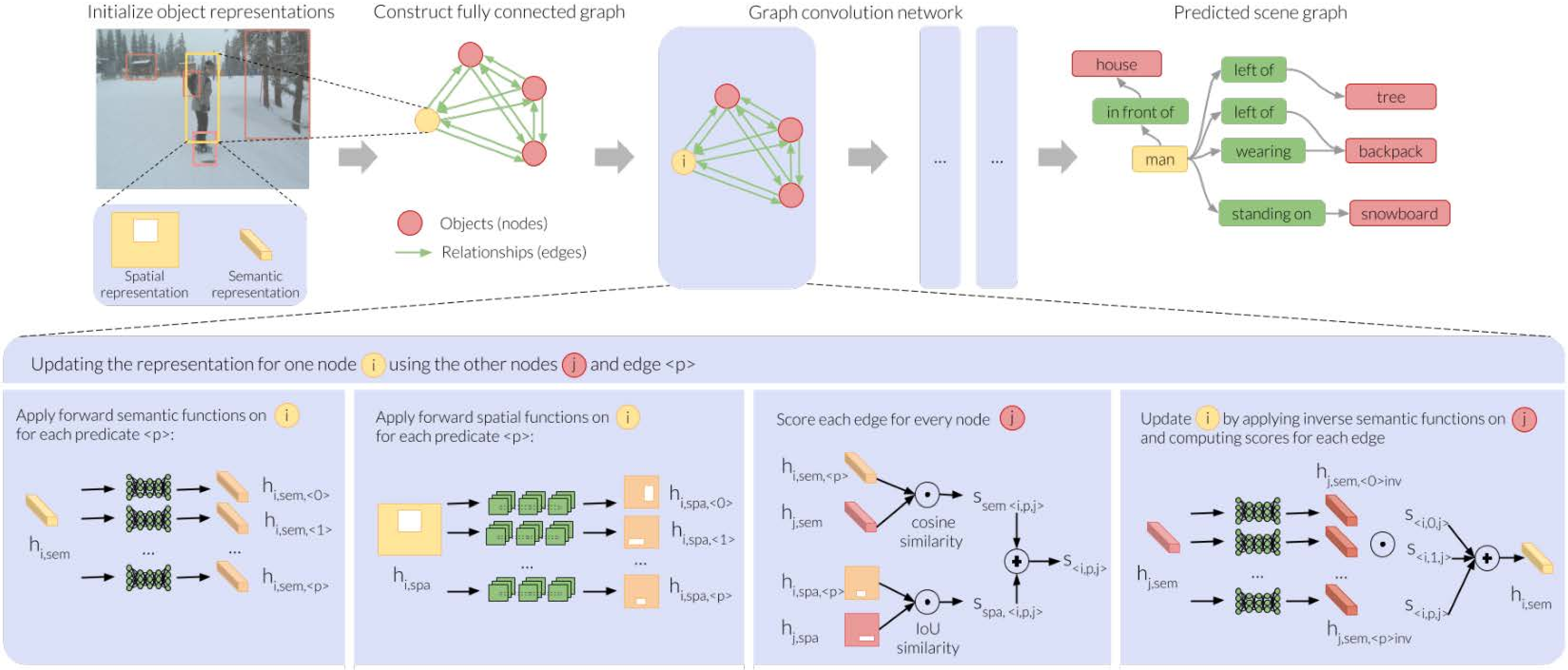


Forward predicate function learns to structure object representations
Inverse predicate function learns to structure the subjects.

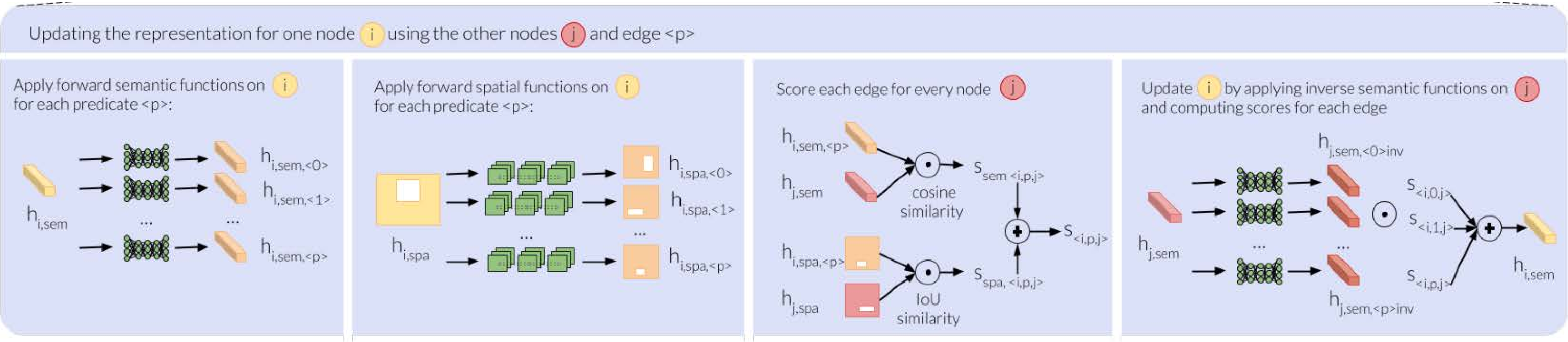
Rare predicate: laying on
<dog – laying on - bed> <dog – laying on - beach>
<dog – laying on - bench> <dog – laying on - desk>

⟹ <dog – laying on - layable>

Visual Relationships as Functions [ICCV 2019]



Visual Relationships as Functions [ICCV 2019]



$$h_i^t = (h_{i,sem}^t, h_{i,spa}^t)$$

$$m_{i,sem}^{t+1} = \sum_{p \in \mathcal{P}} \sum_{j \neq i} M_{sem}(h_{i,sem}^t, h_{j,sem}^t, e_{ijp})$$

$$\boxed{M_{sem}(h_{i,sem}^t, h_{j,sem}^t, e_{ijp}) = s_p^l(h_i^t, h_j^t) f_{sem,p^{-1}}^{-1}(h_{j,sem}^t)}$$

$$U_{sem}(h_{i,sem}^t, m_{i,sem}^{t+1}) = W_0 h_{i,sem}^t + \frac{1}{|\mathcal{P}|(|\mathcal{V}|-1)} m_{i,sem}^{t+1} \tag{10}$$

$$h_i^{t+1} = (U_{sem}(h_{i,sem}^t, m_{i,sem}^{t+1}), h_{i,spa}) \tag{11}$$

$$s_p(h_i^t, h_j^t) = \alpha s_{p,sem}(h_{i,sem}^t, h_{j,sem}^t) + (1-\alpha) s_{p,spa}(h_{i,spa}, h_{j,spa}), \tag{6}$$

$$s_{p,sem}(h_{i,sem}^t, h_{j,sem}^t) = \cos\left[f_{sem,p}(h_{i,sem}^t), h_{j,sem}^t\right], \tag{7}$$

$$s_{p,spa}(h_{i,spa}, h_{j,spa}) = \mathrm{IoU}\left[f_{spa,p}(h_{i,spa}), h_{j,spa}\right],$$

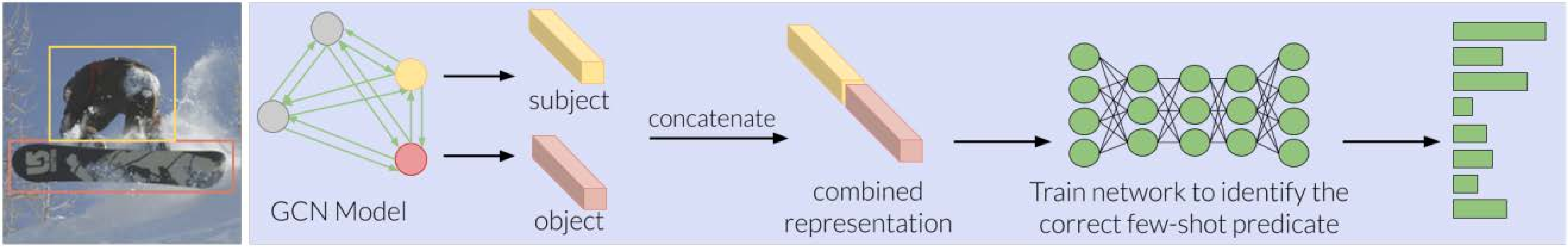
$f_{sem,p}(\cdot) = MLP$

$f_{spa,p}(\cdot) = convolution\ layers$

6 layers + ReLu

Visual Relationships as Functions [ICCV 2019]

- **Few-shot predicate framework**



Now, when training with few examples of rare predicates, such as ***driving***, we can rely on the semantic embeddings for objects that were clustered by ***riding***.

Visual Relationships as Functions [ICCV 2019] – 实验

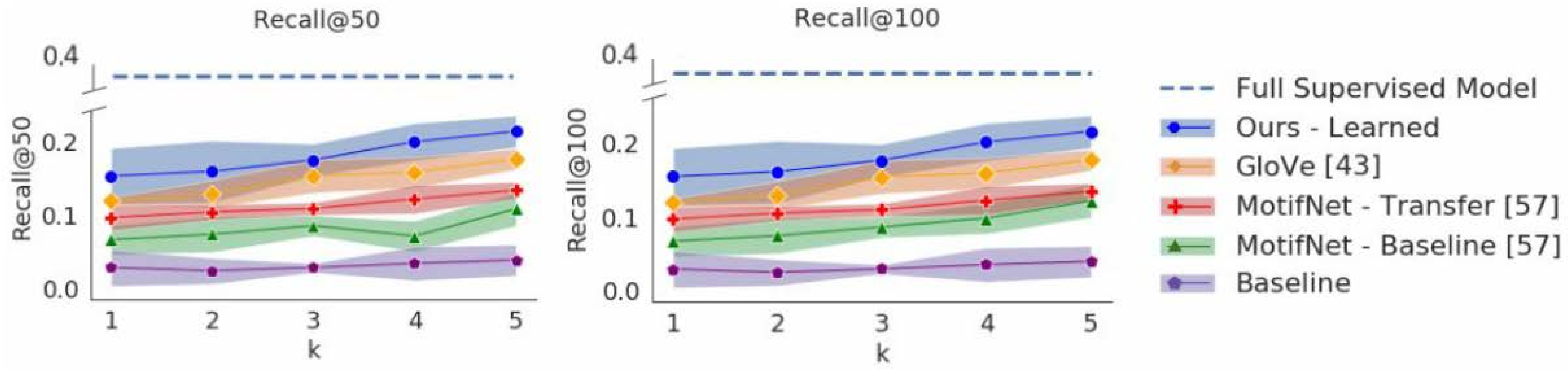


Figure 4. We show Recall@1 and Recall@50 results on k-shot predicates. We outperform strong baselines like transfer learning on MotifNet [57], which also relies on linguistic priors.

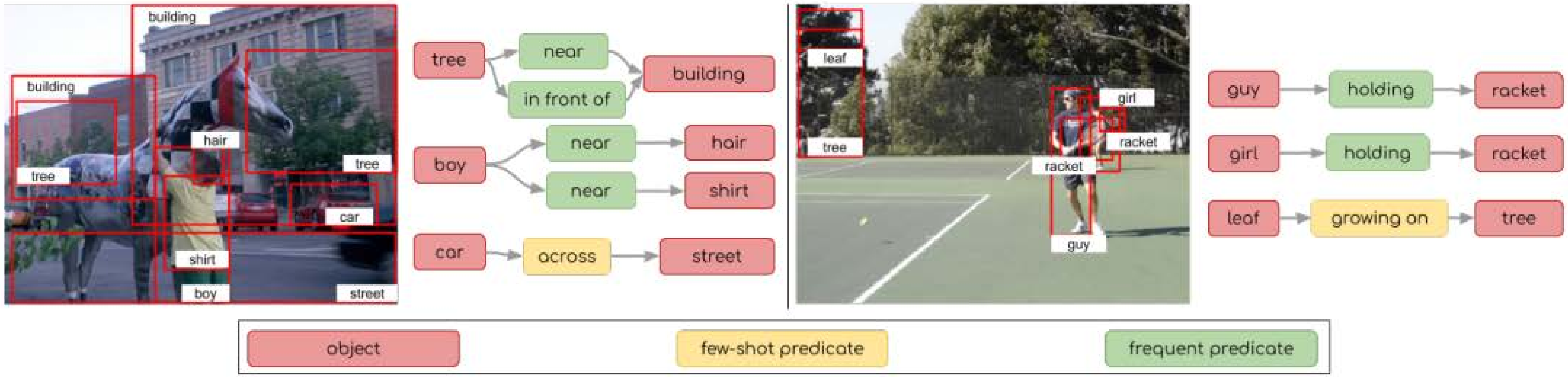


Figure 5. We show example scene graphs predicted with frequent as well as rare relationships, a feat previous models have not tackled.

Scene Graph Prediction with Limited Labels

[ICCV 2019]

Vincent S. Chen, Paroma Varma, Ranjay Krishna, Michael Bernstein, Christopher Ré, Li Fei-Fei
Stanford University

SGG with Limited Labels[ICCV 2019]

- 本文从半监督学习的角度，把潜在的未标注信息标注出来，然后将增强后的数据用于训练现有的模型。

- **主要工作**
  1. 把Data Programming技术应用到视觉关系检测上
  2. 设计了一种Labeling Function，描述了针对Visual Relation标注的经验性规则。

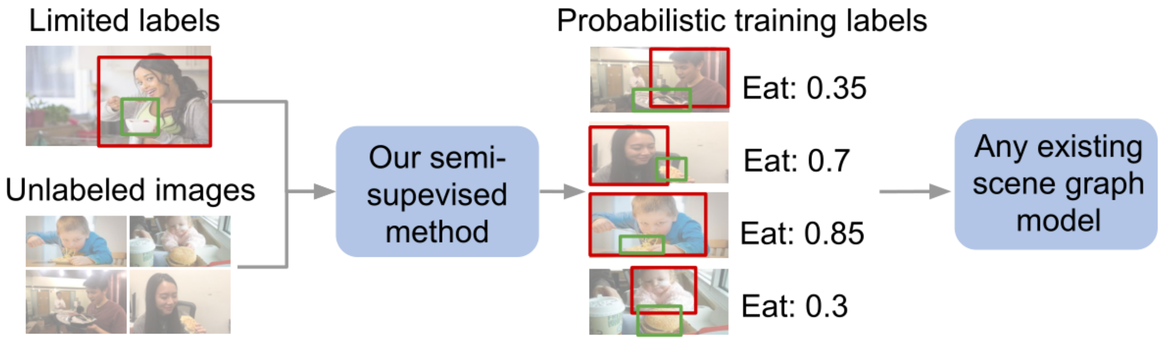
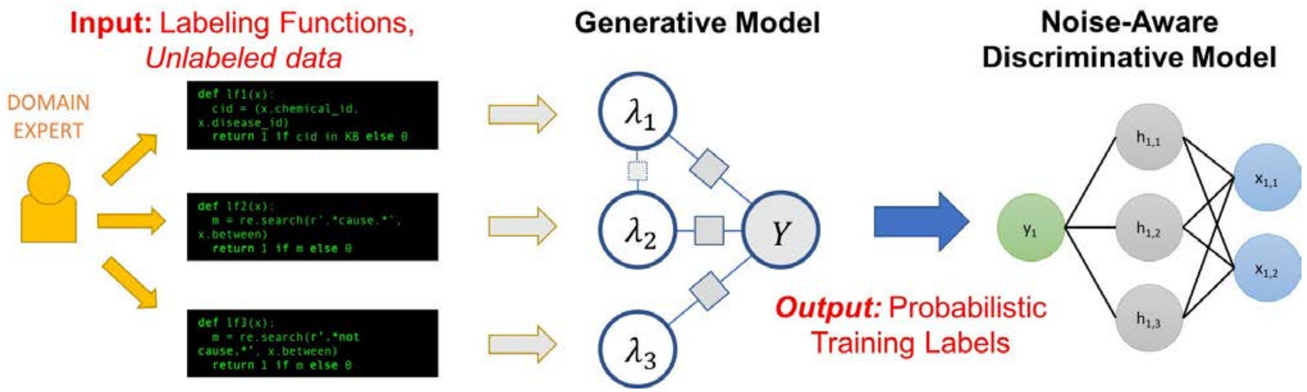


Figure 1. Our semi-supervised method automatically generates probabilistic relationship labels to train any scene graph model.

SGG with Limited Labels[ICCV 2019]

- **Data Programming**

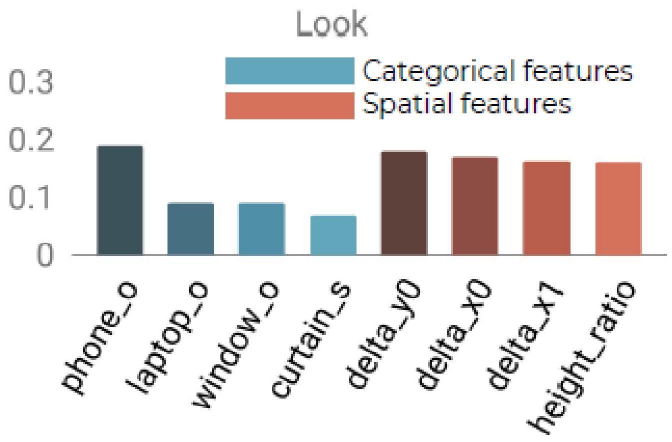


- **Labeling Function**

将人的一般经验抽象成可供参考的简单判别规则。

使用决策树构建。

predicate特征（Image-Agnostic） $= categorical\ feature + spatial\ feature$



SGG with Limited Labels[ICCV 2019]

- **pipeline**

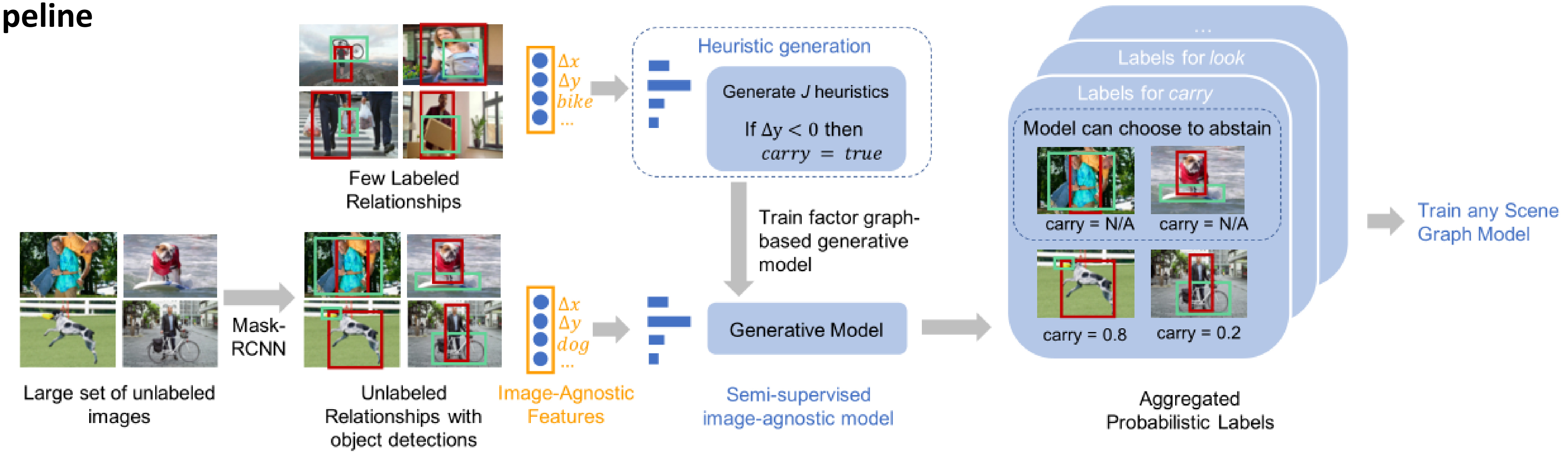
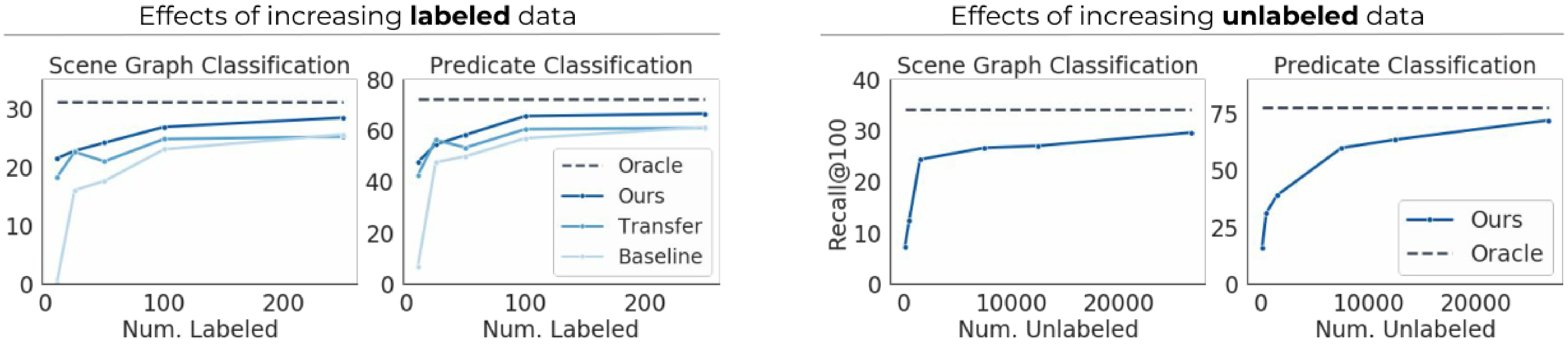


Figure 6. For a relationship (e.g., `carry`), we use image-agnostic features to automatically create heuristics and then use a generative model to assign probabilistic labels to a large unlabeled set of images. These labels can then be used to train any scene graph prediction model.



Unbiased Scene Graph Generation from Biased Training

[CVPR 2020 oral]

Kaihua Tang[1], Yulei Niu[3], Jianqiang Huang[1,2], Jiaxin Shi[4], Hanwang Zhang[1]

[1]Nanyang Technological University, [2]Damo Academy, Alibaba Group, [3]Renmin University of China, [4]Tsinghua University

kaihua001@e.ntu.edu.sg, niu@ruc.edu.cn, jianqiang.jqh@gmail.com

shijx12@163.com, hanwangzhang@ntu.edu.sg

Unbiased SGG [CVPR 2020 oral]

- 先前介绍的所有技术都需要从头开始训练才能解决问题，本质上都提出了一个完整的模型
- 本文则提出了一个独特的框架，**可以附加在任何SGG模型之上**，以处理数据集中存在的偏差。

- **主要工作**
  1. 设计了一个无偏预测的**推理算法Causal TDE Inference**
  2. 设计了一个新的通用SGG框架

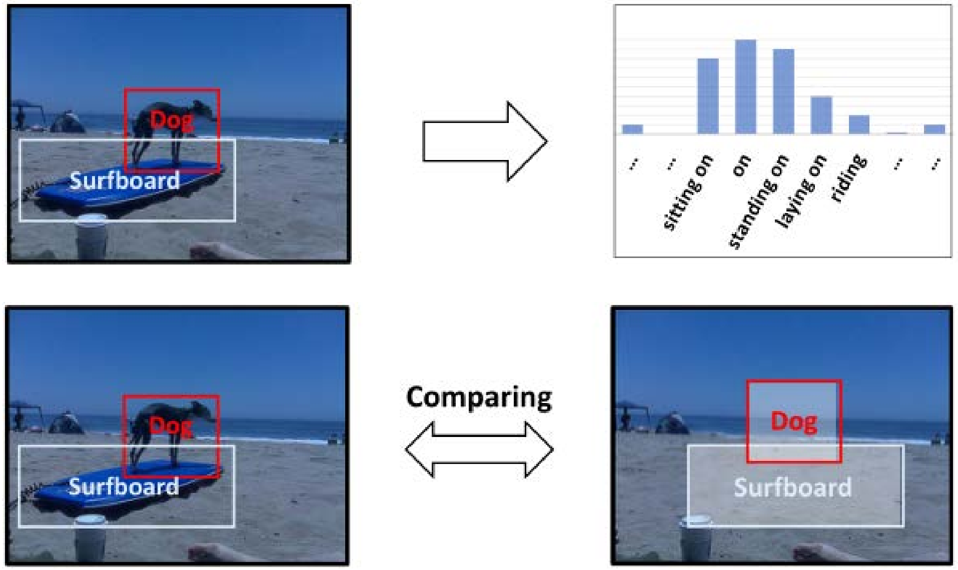
Unbiased SGG [CVPR 2020 oral]

- 论文提出赋予机器**反事实思维（counterfactual thinking）**

  If I had not seen the content, would I still make the same prediction?

贝叶斯之父Judea Pearl
《The Book of Why》

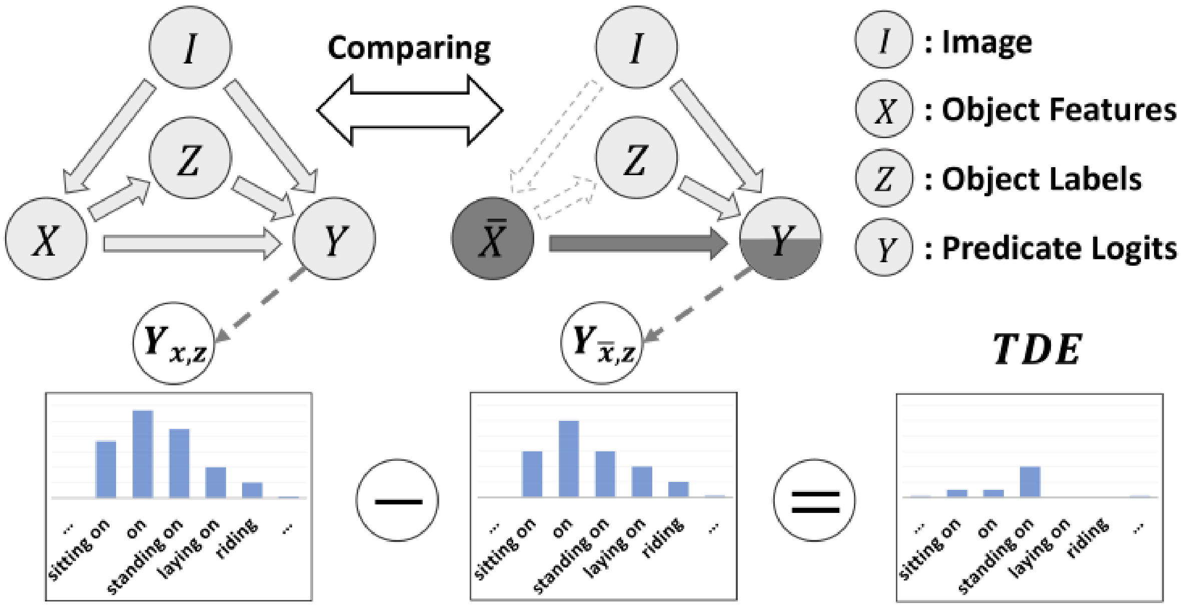
- 反事实思维运用到场景图生成



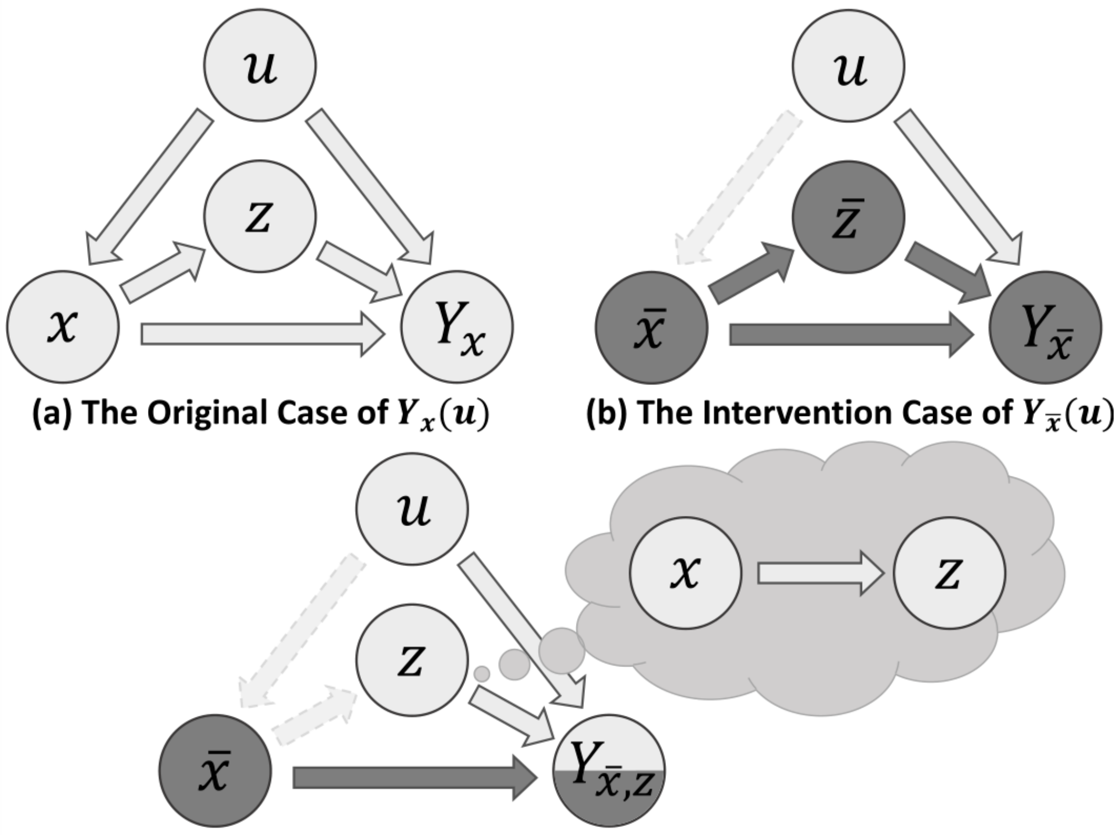
去除object和subject的
visual features

- 基于*causal inference*中的**Total Direct Effect (TDE)**

  因果图上的反事实运算



Unbiased SGG [CVPR 2020 oral]



(a) The Original Case of $Y_x(u)$ (b) The Intervention Case of $Y_{\bar{x}}(u)$

(c) The Counterfactual Case of $Y_{\bar{x},z}(u)$

原始因果图 → 被干预的原始因果图 → 反事实因果图

- **干预**

  切断因果父代影响，指定值

  $Y_{\bar{x},z}(u) = Y_x(\boldsymbol{do}(\boldsymbol{X} = \bar{\boldsymbol{x}})|u)$

- **反事实**

  被干扰后取原值

$$Y_{\bar{x},z} = SUM(\bar{x}, z, u)$$

$$\boldsymbol{TDE} = \boldsymbol{Y_x(u)} - \boldsymbol{Y_{\bar{x},z}(u)}$$

Make the machine “think” twice

$Y_x(u) = y$

$Y_{\bar{x},z}(u) = y(\bar{x}, z)$

模型不可见

$y^{\dagger} = y - y(\bar{x}, z)$

Unbiased SGG [CVPR 2020 oral] – 实验

			Predicate Classification			Scene Graph Classification			Scene Graph Detection		
Model	Fusion	Method	mR@20	mR@50	mR100	mR@20	mR50	mR100	mR@20	mR50	mR100
IMP+ [64, 6]	-	-	-	9.8	10.5	-	5.8	6.0	-	3.8	4.8
FREQ [71, 55]	-	-	8.3	13.0	16.0	5.1	7.2	8.5	4.5	6.1	7.1
MOTIFS [71, 55]	-	-	10.8	14.0	15.3	6.3	7.7	8.2	4.2	5.7	6.6
KERN [6]	-	-	-	17.7	19.2	-	9.4	10.0	-	6.4	7.3
VCTree [55]	-	-	14.0	17.9	19.4	8.2	10.1	10.8	5.2	6.9	8.0
MOTIFS†	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	**6.5**	**8.4**	**9.8**
		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	**18.5**	**25.5**	**29.1**	**9.8**	**13.1**	**14.9**	5.8	8.2	**9.8**
	GATE	Baseline	12.2	15.5	16.8	7.2	9.0	9.5	5.2	7.2	8.5
		TDE	**18.5**	**24.9**	**28.3**	**11.1**	**13.9**	**15.2**	**6.6**	**8.5**	**9.9**
VTransE†	SUM	Baseline	11.6	14.7	15.8	6.7	8.2	8.7	3.7	5.0	6.0
		TDE	**17.3**	**24.6**	**28.0**	**9.3**	**12.9**	**14.8**	**6.3**	**8.6**	**10.5**
	GATE	Baseline	13.6	17.1	18.6	6.6	8.2	8.7	5.1	6.8	8.0
		TDE	**18.9**	**25.3**	**28.4**	**9.8**	**13.1**	**14.7**	**6.0**	**8.5**	**10.2**
VCTree†	SUM	Baseline	11.7	14.9	16.1	6.2	7.5	7.9	4.2	5.7	6.9
		TDE	**18.4**	**25.4**	**28.7**	**8.9**	**12.2**	**14.0**	**6.9**	**9.3**	**11.1**
	GATE	Baseline	12.4	15.4	16.6	6.3	7.5	8.0	4.9	6.6	7.7
		TDE	**17.2**	**23.3**	**26.6**	**8.9**	**11.8**	**13.4**	**6.3**	**8.6**	**10.3**

Thanks

汇报人：崔奕宸

2021/4/14