Paper ID # 718

Recovering the Unbiased Scene Graphs from the Biased Ones

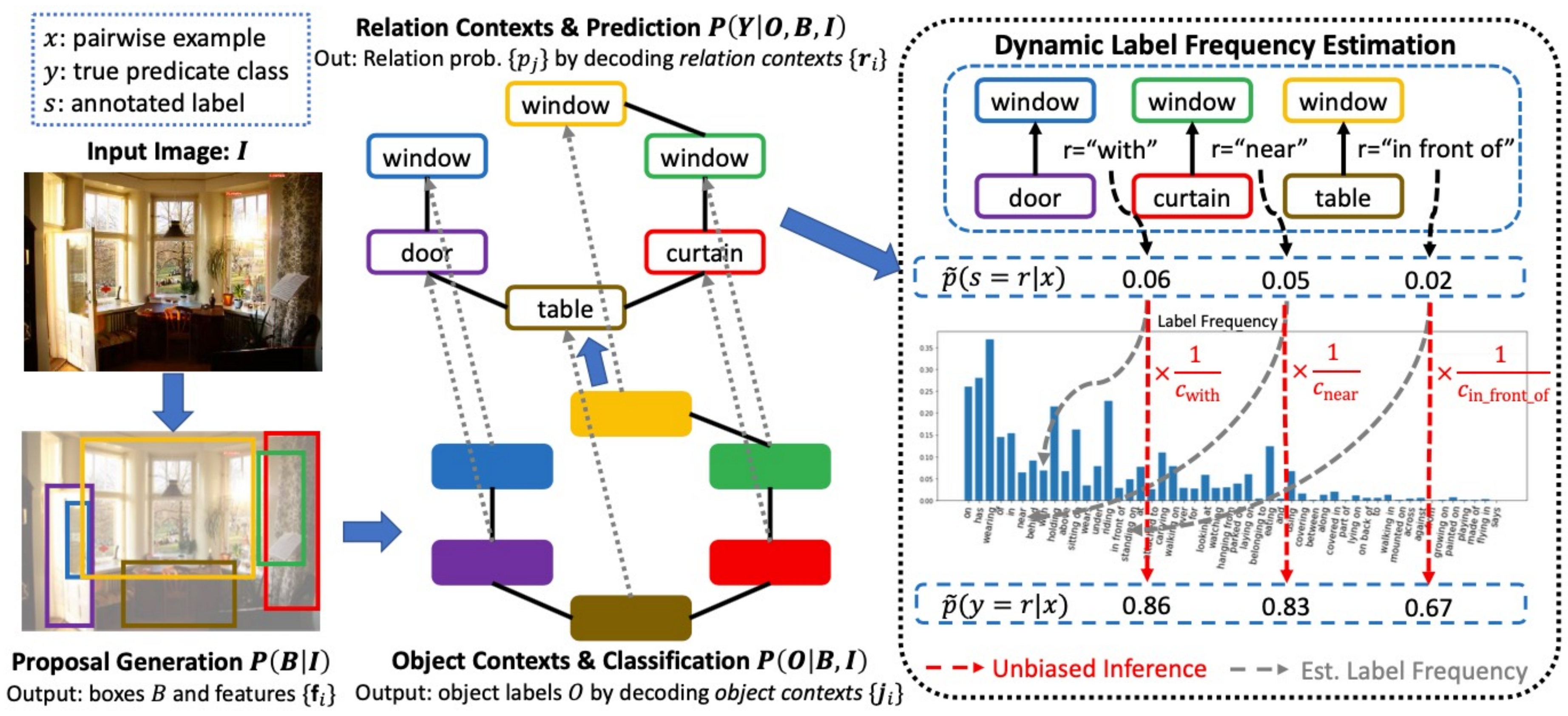
Meng-Jiun Chiou[1], Henghui Ding[2], Hanshu Yan[1],
Changhu Wang[2], Roger Zimmermann[1] and Jiashi Feng[1]
[1]National University of Singapore [2]ByteDance AI Lab

https://github.com/coldmanck/recovering-unbiased-scene-graphs/

• Introduction

- Given input images, ***scene graph generation (SGG)*** [1] aims to produce comprehensive, graphical representations describing visual relationships among salient objects.
- Recently, more efforts have been paid to the long tail problem in SGG; however, the imbalance in the fraction of missing labels of different classes, or ***reporting bias***, exacerbating the long tail is rarely considered and cannot be solved by the existing debiasing methods.

- In this paper we show that, due to the missing labels, SGG can be viewed as a ***Learning from Positive and Unlabeled data (PU learning)*** problem, where the *reporting bias can be removed by taking the difference among the chance of objects being labeled into account*.
- We propose ***Dynamic Label Frequency Estimation (DLFE)*** to obtain accurate label frequency estimates for debiasing SGG. We show that DLFE significantly **alleviates the long tail and achieves state-of-the-art debiasing performance** on the VG dataset, **producing prominently more informative and less-biased scene graphs**.
- We propose **DLFE**: augmenting data with random flipping, and averaging over multiple epochs, to introduce more samples.



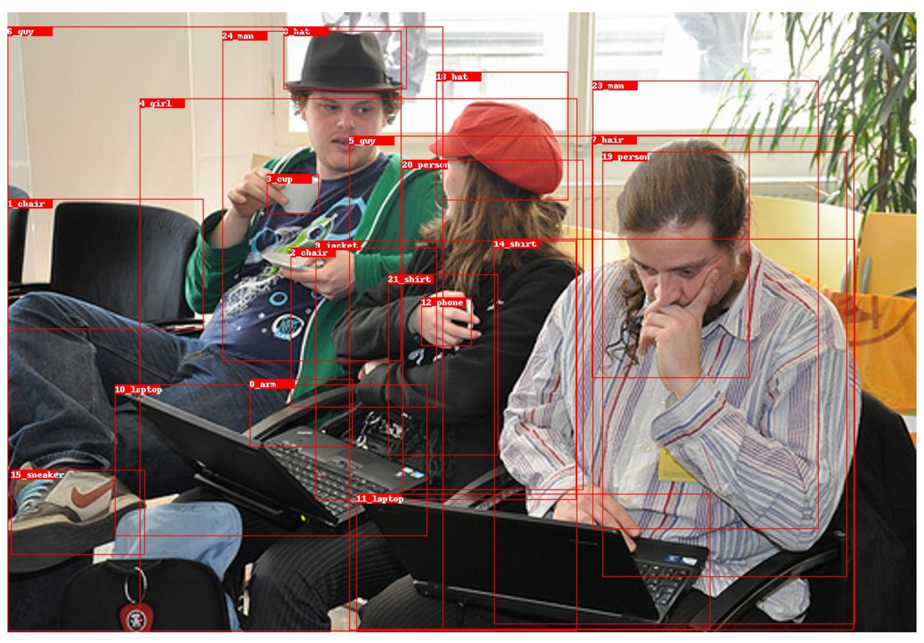
Coupled with the SGG Model from [1]

Motivation

- **Not All Long Tails are Equal**: Unlike long tails in other vision tasks like image classification, the long tail in SGG is significantly affected by the **imbalance** in **missing labels**.

1. Missing labels:

- Caused by **the cubic number of possible visual relations**
- For each image: N objects, K predicate classes → $KN(N-1)$ possibilities
- Missing label bias causes the **predicted probabilities to be under-estimated**



2. Imbalance in missing labels (reporting bias):

- Because easier predicates (*e.g.*, on) get annotated more than harder ones (*e.g.*, parked on) during collection (*e.g.*, VG dataset)
- The predicted probabilities of the hard ones are thus under-estimated more than the easy ones, causing the long tail in the VG dataset.
- **[Definition] Label frequency**: the per-class fraction of labeled, positive examples in all the examples, which can be estimated (details below).

Methodology

- We draw inspiration from ***Learning from Positive and Unlabeled Data*** [2]. Examples in VG dataset are a set of triplets $\{(x, y, s)\}$.
  - x be an example (candidate object pair)
  - $y \in \{0,..,K\}$ be its true class, where K is the number of predicate classes.
  - $s \in \{0,..,K\}$ is a label and $s = 0$ if x is unlabeled.
  - When $r \neq 0,\ s = r \Rightarrow y = r$; When $r = s = 0$, y can be 0 (background) or any natural number.
- **Biased probability** of a non-background class r $(r \neq 0)$:

$$p(s=r|x) = p(y=r, s=r|x) = p(y=r|x)p(s=r|y=r,x)$$

- **Unbiased probability** can be recovered w/ SCAR assumption [2]:

$$p(y=r|x) = \frac{p(s=r|x)}{p(s=r|y=r,x)} \approx \frac{p(s=r|x)}{\underbrace{p(s=r|y=r)}_{\text{label frequency}}}$$

Experiments

- Is DLFE more effective in estimating label frequency?

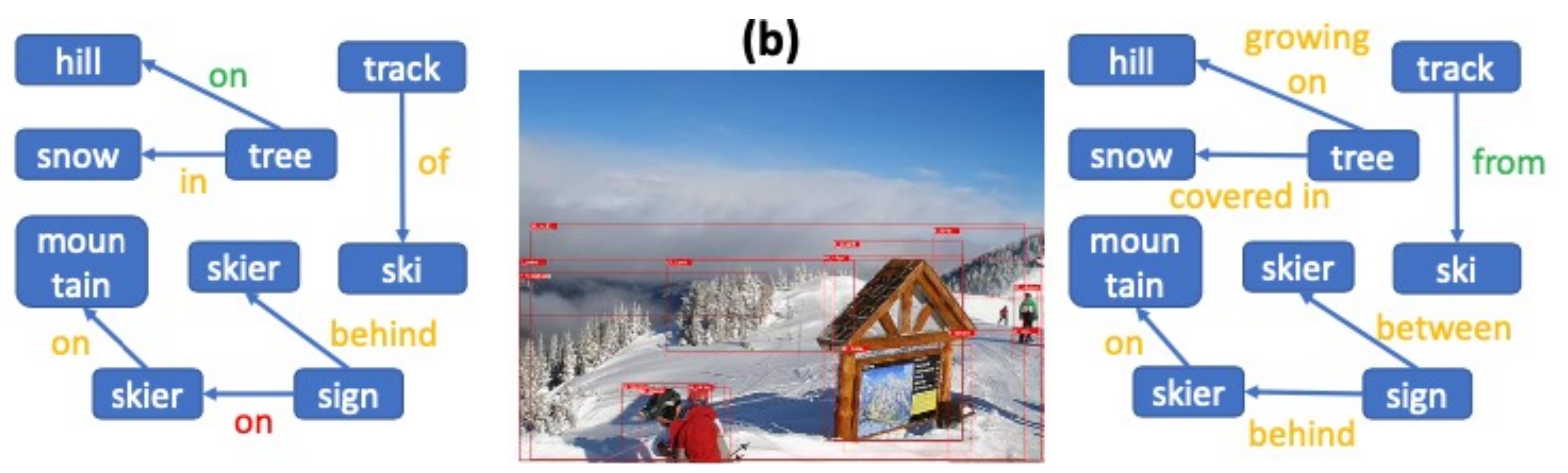
	Predicate Classification (PredCls)			Scene Graph Classification (SGCls)			Scene Graph Detection (SGDet)		
Model	ng-mR@20	ng-mR@50	ng-mR@100	ng-mR@20	ng-mR@50	ng-mR@100	ng-mR@20	ng-mR@50	ng-mR@100
MOTIFS [37, 53]	19.9	32.8	44.7	11.3	19.0	25.0	7.5	12.5	16.9
MOTIFS-Train-Est [12]	24.4	38.9	50.5	17.1	**26.1**	**32.8**	8.9	14.1	18.9
MOTIFS-DLFE	**30.0**	**45.8**	**57.7**	**17.6**	25.6	32.0	**11.7**	**18.1**	**23.0**
VCTree [37, 38]	21.4	35.6	47.8	12.4	19.1	25.5	7.5	12.5	16.7
VCTree-Train-Est [12]	25.0	39.1	52.4	21.0	**32.2**	**39.4**	8.1	13.0	17.1
VCTree-DLFE	**29.1**	**44.6**	**56.8**	**21.6**	31.4	38.8	**11.7**	**17.5**	**22.5**

- Does DLFE help in debiasing scene graph generation?

	Predicate Classification (PredCls)		
Model	ng-mR@20	ng-mR@50	ng-mR@100
KERN [5]	-	36.3	49.0
GB-Net-$\beta^{\diamond}$ [51]	-	44.5	58.7
MOTIFS† [41, 53]	19.9	32.8	44.7
MOTIFS-Reweight‡	20.5	33.5	44.4
MOTIFS-L2+uKD‡ [41]	-	36.9	50.9
MOTIFS-L2+cKD‡ [41]	-	37.2	50.8
MOTIFS-TDE† [37]	18.7	29.0	38.2
MOTIFS-PCPL† [45]	25.6	38.5	49.3
MOTIFS-STL† [4]	15.7	29.4	43.2
MOTIFS-DLFE	**30.0**	**45.8**	**57.7**
VCTree† [38, 41]	21.4	35.6	47.8
VCTree-Reweight‡	20.6	32.5	41.6
VCTree-L2+uKD‡ [41]	-	37.7	51.7
VCTree-L2+cKD‡ [41]	-	38.4	52.4
VCTree-TDE† [37]	20.9	32.4	41.5
VCTree-PCPL† [45]	25.1	38.5	49.3
VCTree-STL† [4]	16.8	31.8	45.1
VCTree-DLFE	**29.1**	**44.6**	**56.8**

	Head Recalls		Middle Recalls		Tail Recalls	
Model	R@50	R@100	R@50	R@100	R@50	R@100
MOTIFS† [37, 53]	65.9	**78.6**	30.0	45.4	3.3	9.7
MOTIFS-Reweight†	57.4	69.2	30.7	43.0	13.3	21.5
MOTIFS-TDE† [37]	48.3	60.8	34.9	46.1	1.8	5.3
MOTIFS-PCPL† [45]	**66.5**	77.6	41.8	**55.2**	6.0	13.2
MOTIFS-STL† [4]	56.4	70.0	24.1	39.8	9.6	21.2
MOTIFS-DLFE	61.9	72.4	**42.8**	54.2	**31.8**	**44.6**
VCTree† [37, 38]	**67.5**	**79.8**	34.3	50.0	5.5	12.7
VCTree-Reweight†	61.6	73.4	28.3	38.3	9.0	14.3
VCTree-TDE† [37]	54.8	67.5	37.9	49.1	2.5	5.4
VCTree-PCPL† [45]	64.5	75.9	**42.6**	**54.2**	6.9	16.1
VCTree-STL† [4]	57.6	71.1	26.1	41.8	13.8	23.5
VCTree-DLFE	57.5	68.3	36.0	48.2	**26.5**	**38.1**

- Visualizing debiased scene graphs



Conclusion

- we are the first to tackle long-tailed SGG with the **cause (unbalanced missing labels)** instead of its superficial effect (long tail distribution).
- We view SGG as a PU problem and we remove the reporting bias by recovering the per-class unbiased probabilities from the biased ones.
- We propose DLFE which provides more reliable label frequency estimates using augmented data and averages over multiple epochs
- We show that DLFE is more effective in estimating label frequencies, and SGG models with DLFE achieves SOTA debiasing performance in VG dataset and produce significantly more balanced scene graphs.
- **Scan the top-most QR code to check out the source code!**

Reference

[1] Zellers et al. "Neural Motifs: Scene Graph Parsing with Global Context." In CVPR 2018.

[2] Elkan et al. "Learning classifiers from only positive and unlabeled data". In SIGKDD 2008.