Scene Graph Prediction with Limited Labels

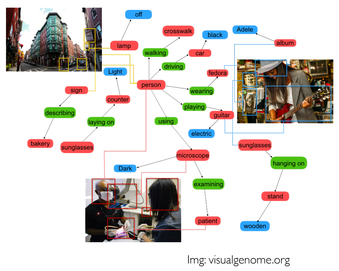
Vincent S. Chen (vincentsc@cs.stanford.edu), Paroma Varma, Ranjay Krishna, Michael Bernstein, Christopher Ré, Li Fei Fei

The Long Tail of Visual Relationships

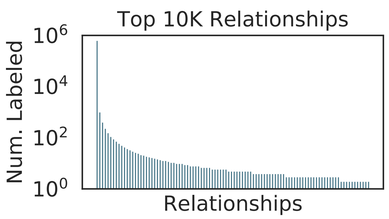
Problem

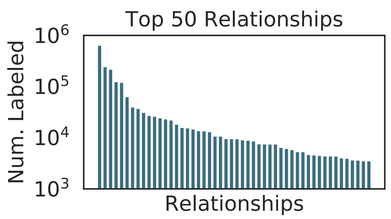
Scene graph datasets are **incomplete** due to annotator error +cost.

State-of-the-art models **only consider the top 50 relationships** and ignore the rest!



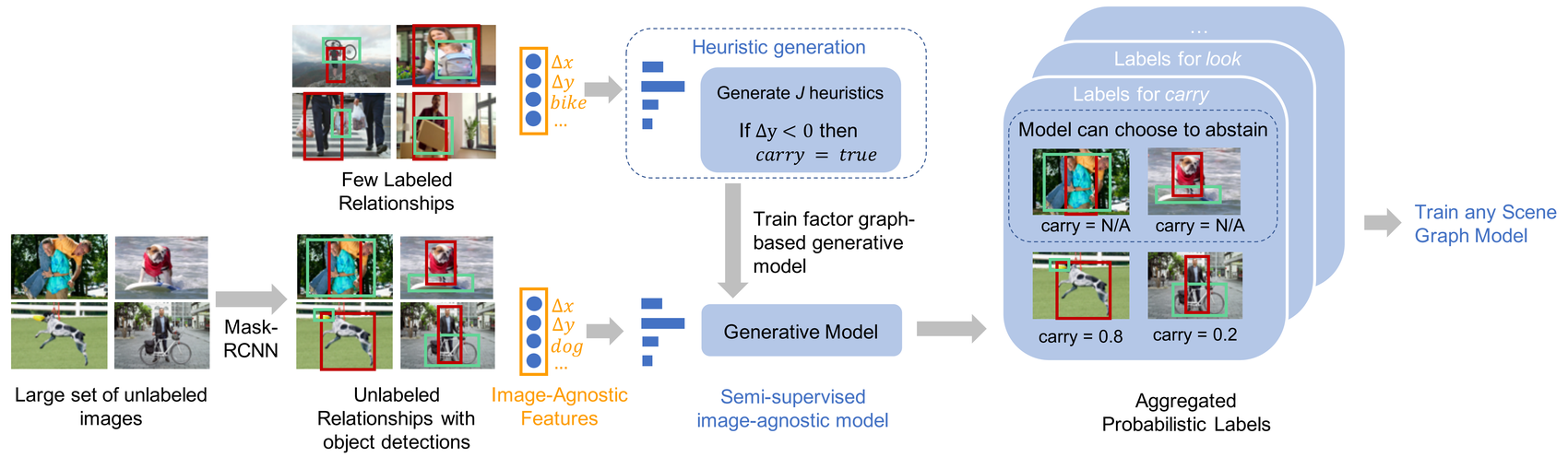
Img: visualgenome.org





Generating Labels for Structured Predictions

1. Feature extraction: Extract spatial + categorical features based on object bounding boxes.
2. Heuristic Generation: Using a limited set of labeled data, automatically generate noisy heuristics.
3. Generative Model: Combine the noisy heuristics' outputs into probabilistic training labels.



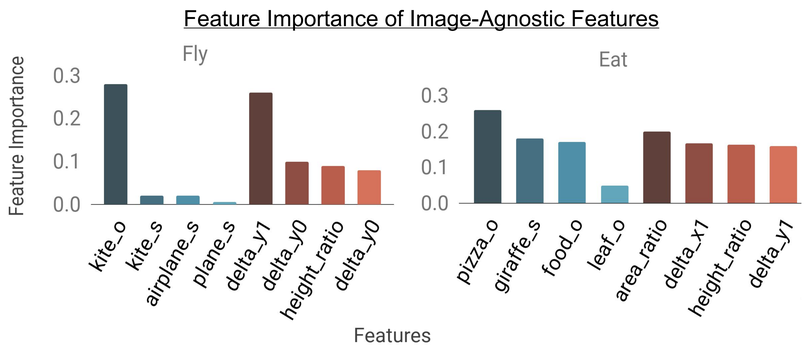
Leveraging Image-Agnostic Features

Inspiration: *Textual relation extraction* relies on *document-agnostic* heuristics.

```
<Tokyo, capitol of, Japan>
```

Challenge: *Visual relationships* vary based on context within an image!

Idea: Leverage *image-agnostic features* in heuristics for label generation.

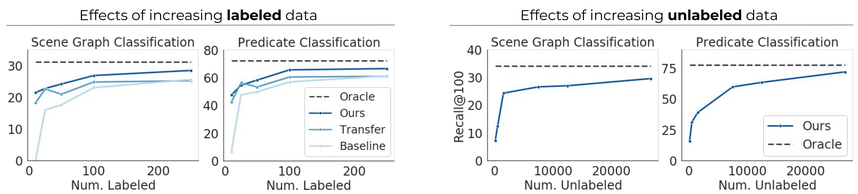


Experimental Results

Our approach outperforms **naïve baselines** (B, F+O) using only limited labels, **semi-supervised methods** (DT, LP) relying on image-agnostic features to learn patterns over labeled/unlabeled data, and **transfer learning**, which pretrains on the set of available relationships and fine-tunes on limited relationship labels.

	Scene Graph Detection			Scene Graph Classification			Predicate Classification		
Model	R@20	R@50	R@100	R@20	R@50	R@100	R@20	R@50	R@100
Baseline [$n = 10$]	0.00	0.00	0.00	0.04	0.04	0.04	3.17	5.30	6.61
Freq+Overlap	10.16	10.84	10.86	9.90	9.91	9.91	20.39	20.90	22.21
Transfer Learning	11.99	14.40	16.48	17.10	17.91	18.16	39.69	41.65	42.37
Decision Tree	11.11	12.58	13.23	14.02	14.51	14.57	31.75	33.02	33.35
Label Propagation	6.48	6.74	6.83	9.67	9.91	9.97	24.28	25.17	25.41
Ours (Categ. + Spat.)	**17.67**	**18.69**	**19.28**	**20.91**	**21.34**	**21.44**	**45.49**	**47.04**	**47.53**
Oracle [$n_{\text{ORACLE}} = 108n$]	24.42	29.67	30.15	30.15	30.89	31.09	69.23	71.40	72.15

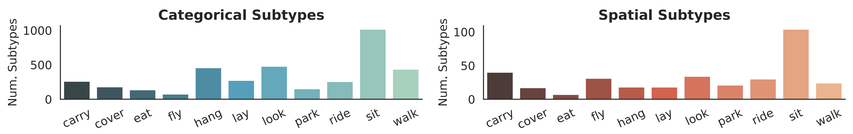
Our approach improves with increasing **labeled and unlabeled data**.



Challenges: Semantically similar phrasing / synonymous relationships

Studying Visual Relationship *Complexity*

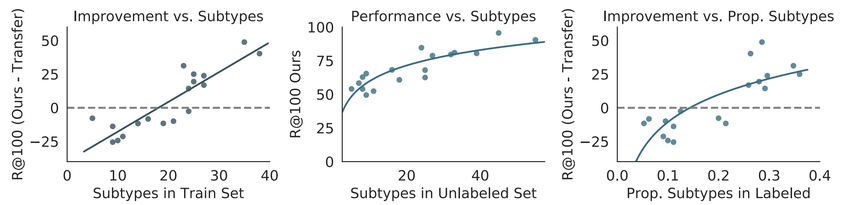
Relationship subtypes capture the different ways that a visual relationship manifests in the dataset.



 # Categorical Subtype:
Count the number of object categories for this relationship

 # Spatial Subtypes:
Mean shift clustering over spatial features, computed over bounding boxes

Trends: With ↑ relationship complexity (defined by subtypes), our weak supervision approach improves relative to transfer learning



Code: https://github.com/vincentschen/limited-label-scene-graphs

Snorkel: snorkel.org