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!

Leveraging Image-Agnostic Features

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

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

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

Feature Importance of Image-Agnostic Features

Feature Importance of Image-Agnostic Features

Feature: Fly  Eat  ... Spatial Subtypes

Effects of increasing labeled data

Effects of increasing unlabeled data

Challenges: Semantically similar phrasing / synonymous relationships

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 labeling outputs.

Experimental Results

Our approach outperforms naive 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.

Studying Visual Relationship Complexity

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

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