Graph R-CNN for Scene Graph Generation

Jianwei Yang1, Jiasen Lu1, Stefan Lee1, Dhruv Batra1,2, Devi Parikh1,2

1. Georgia Tech; 2. Facebook AI Research (FAIR); © Equal Contribution

Graph R-CNN Faster R-CNN

1. BACKGROUND

Scene Graph Generation

- 1) Object Detection
- 2) Relationship Prediction
- 3) Attribute Prediction

2. MOTIVATIONS AND CONTRIBUTIONS

Motivations:
1. The scene graph is naturally sparse, i.e., few objects have relationships with each other.
2. Existence of relationships between objects highly depends on the object categories;
3. Type of relationships between objects highly depends on the context.

Contributions:
1. Proposed a relation proposal network (RePN) to learn to sparsify the densely connected scene graph;
2. Proposed attentional GCNs (aGCNs) to incorporate the context;
3. Proposed a new metric for evaluating scene graph generation;

3. GRAPH R-CNN: OUR FRAMEWORK

Scene Graph Generation Factorization:

Relation Proposal Network: Predicts relationship

Graph Labeling: Assigns relationships to objects

Attentional GCNs:
- Input Object: \(i\)
- Scene Graph: \(S\)
- Graph Vertices: \(V\)
- Relationship: \(E\)
- Scene Graph Edge: \(O\)

Relation Proposal Network (RePN)

GCN layer with residual connection:

Update object representations:

Nonlinear function: Learnable parameters: Inputs from last layer

Attention on edges (learning affinities):

Attentional GCNs (aGCNs) on scene graph:

Object and semantic sense emerges in aGCNs

4. EXPERIMENT

Settings:
Models: Baseline: Faster R-CNN with VGG-16
Training: two-stage training
Dataset: Visual Genome: train (75011), test (9244)
Annotation Relation: Relation (50), Relation (50)
Metrics: SGGen and

SGGen: recall of relationship triplets
Predl: object location given
Predl: object location and label given

Scene Graph Generation

Table 1: Quantitative comparison with previous work

Table 2: Object relationship emergence in aGCNs

Table 3: Top multiple categories improve after adding RePN

Table 4: Top multiple categories improve after adding RePN

Merits:
1. Our model beats previous models with significant margin;
2. Object detection is improved due to the benefit from RePN;
3. The common-sense emerges in semantic-level aGCNs.

Quality Results

Qualitative Results

Graph R-CNN Faster R-CNN

Graph R-CNN Faster R-CNN

Graph R-CNN Faster R-CNN

Graph R-CNN Faster R-CNN

Graph R-CNN Faster R-CNN

Graph R-CNN Faster R-CNN

Graph R-CNN Faster R-CNN

Graph R-CNN Faster R-CNN

Graph R-CNN Faster R-CNN

Graph R-CNN Faster R-CNN