Group of Objects

Experiment Presentation: Yousi Lin
Visual relationship detection with language priors

C Lu, R Krishna, M Bernstein, L Fei-Fei
ECCV2016

- This paper proposed a model that can train visual models for objects and predicates individually and later combines them together to predict multiple relationships per image.
- Their model can scale to predict thousands of types of relationships from a few examples.
Result images

- <car, behind, tree> score: 10.5
- <person, wear, glasses> score: 15.0
- <tower, has, clock> score: 5.9
Where are they looking?
Adrià Recasens*, Aditya Khosla*, Carl Vondrick, Antonio Torralba
NIPS 2015

• This paper proposed a deep neural network-based approach for gaze-following and a new benchmark dataset, GazeFollow, for thorough evaluation.

• Given an image and the location of a head, this approach follows the gaze of the person and identifies the object being looked at.

http://gazefollow.csail.mit.edu/index.html
Example of Results
Example of Results
Examples of Failures
Recognition using Visual Phrases

A. Farhadi, M. A. Sadeghi
CVPR 2011

- This paper introduced a dataset suitable for phrasal recognition that uses familiar PASCAL object categories and demonstrate significant experimental gains resulting from exploiting visual phrases.

- 32 classes (24 are objects, 8 are objects with relationships, e.g., “person riding a horse”, “dog lying on sofa”, “person riding a bicycle”, etc.)
Training YOLO with Visual Phrase Dataset

Hardware

- Processor: Intel® Xeon(R) CPU E5-2687W v3 @ 3.10GHz × 16
- Memory: 31.3 GiB
- Graphics: Quadro M4000/PCle/SSE2
- System: Linux
Problems encountered during training on YOLO…

• IOU = nan

• Loss swings a lot

• Loss is not decreasing
Good Results so far… (Loss~=$1.05$)
Thank you!