

Notes of “Visual Relationship Detection with Language Priors”

Jihang Li

January 17, 2020

Contents

Contents	1
5 Experiments	1

Abstract

A model is proposed to train visual models for objects and predicated individually and later combines them together to predict multiple relationships per images. Improvement is done by leveraging language priors from semantic word embeddings to finetune the likelihood of a predicted relationship.

5 Experiments

The evaluation metrics reported in this paper is **recall@100** and **recall@50** (refer to [Measuring the Objectness of Image Windows](#)). **Recall@x** computes the fraction of times the correct relationship is predicted in the top **x** confident relationship predictions.

Original Texts

Even if the prediction is correct, mAP would penalize the prediction if we do not have that particular ground truth annotation.

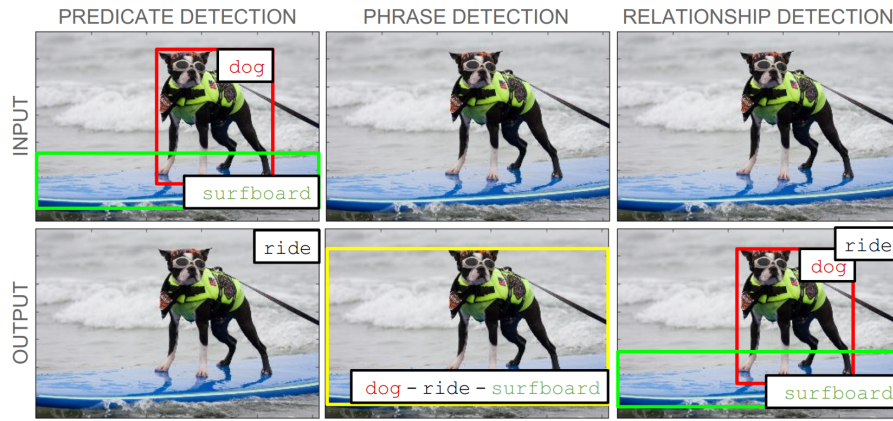


Figure 1: We evaluate visual relationship detection using three conditions: predicate detection (where we only predict the predicate given the object classes and boxes), phrase detection (where we label a region of an image with a relationship) and relationship detection (where we detect the objects and label the predicate between them).