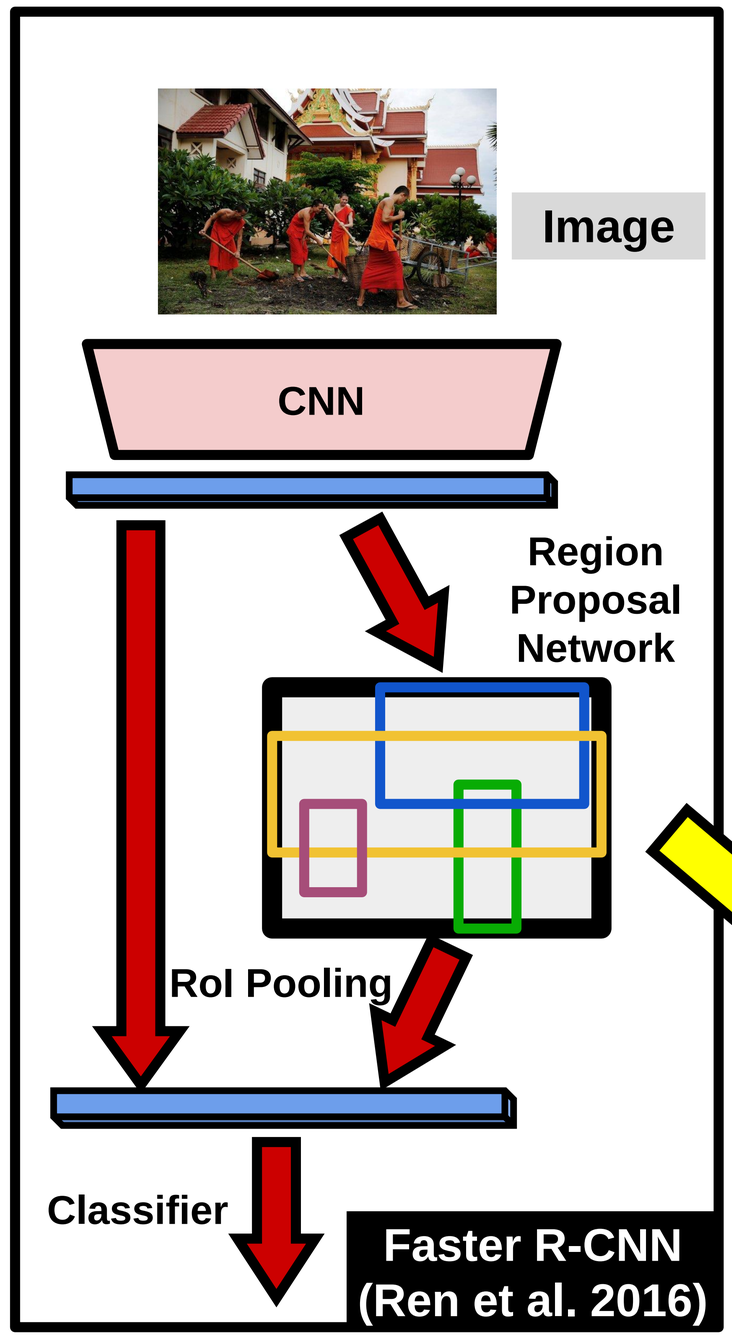
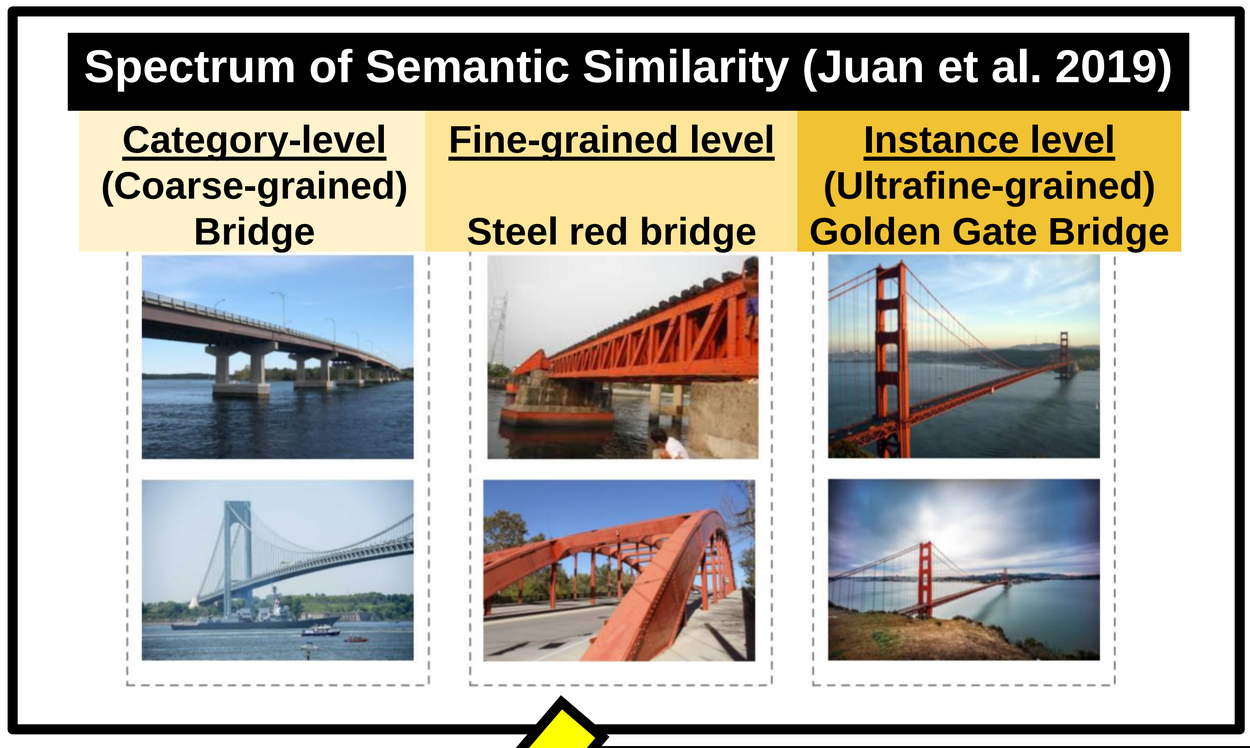
Decoupled Box Proposal and Featurization with Ultrafine-Grained Semantic Labels Improve Image Captioning and Visual Question Answering

Soravit Changpinyo, Bo Pang, Piyush Sharma, Radu Soricut

Google AI

Decoupled Box Proposal + **Featurization with Ultrafine-Grained Semantic Labels**





Vision & Language Models

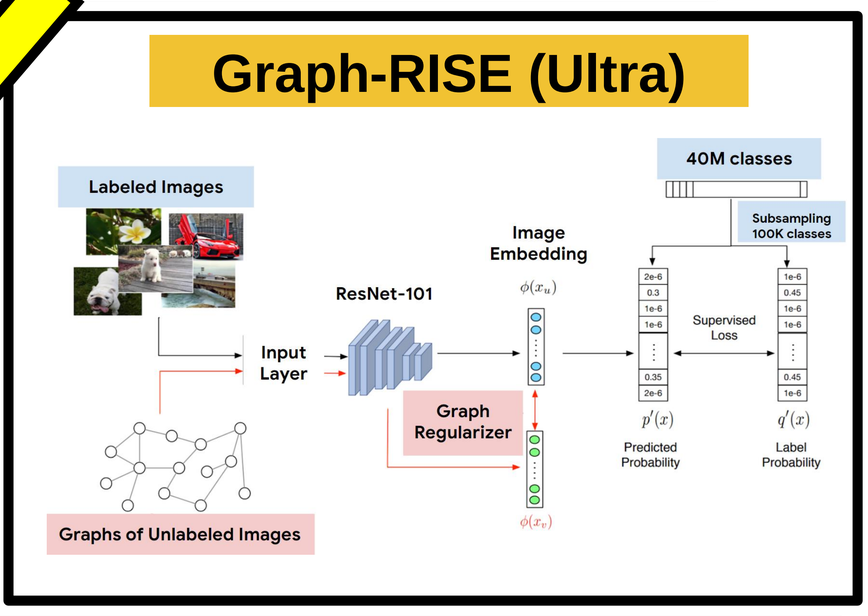
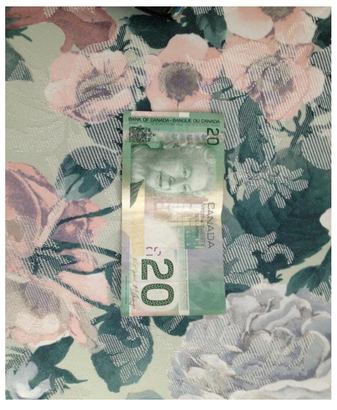


Image Captioning on Conceptual Captions (Sharma et al. 2018)

Ground-truth	monks clean a garden at a temple .	black sesame seeds on a white background
FRCNN (Visual Genome)	a woman walks through the streets .	a pile of dried flowers
Ultra	monks walking in front of a temple	black chia seeds on a white background

VQA on VizWiz (Gurari et al. 2018)



Q: How much money is this?

A: 1 dollar

A: 20

Q: What is this?

A: beer

A: bbq sauce