

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

Soravit (Beer) Changpinyo



Google AI

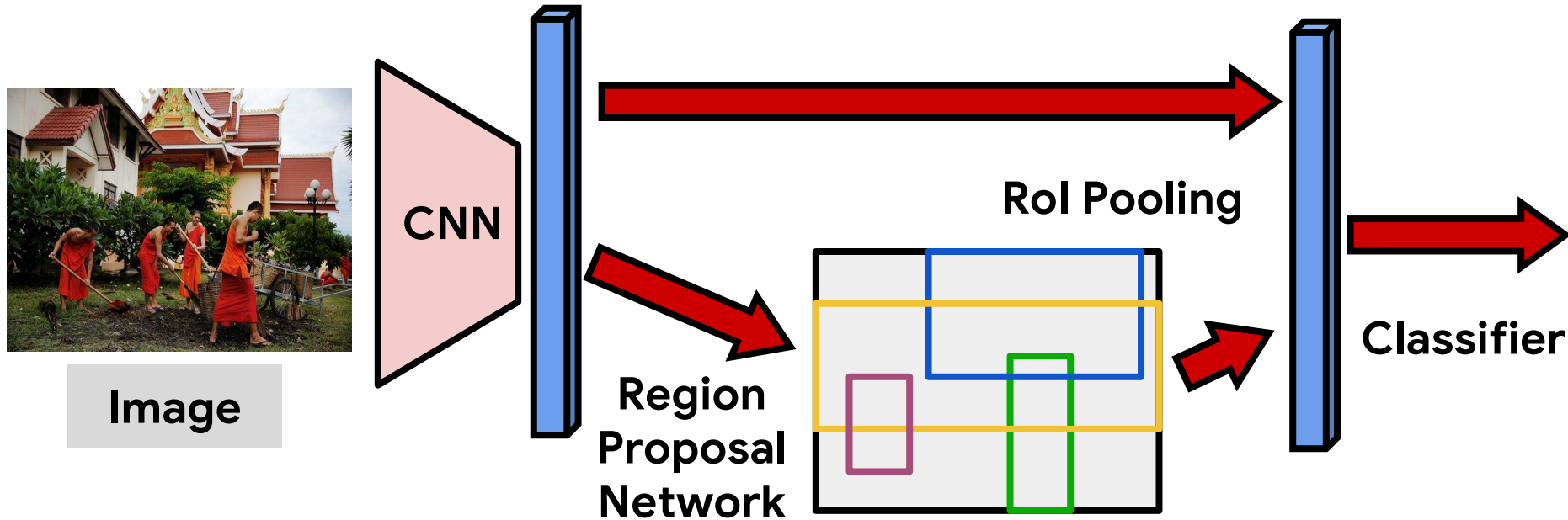


Google AI
Language

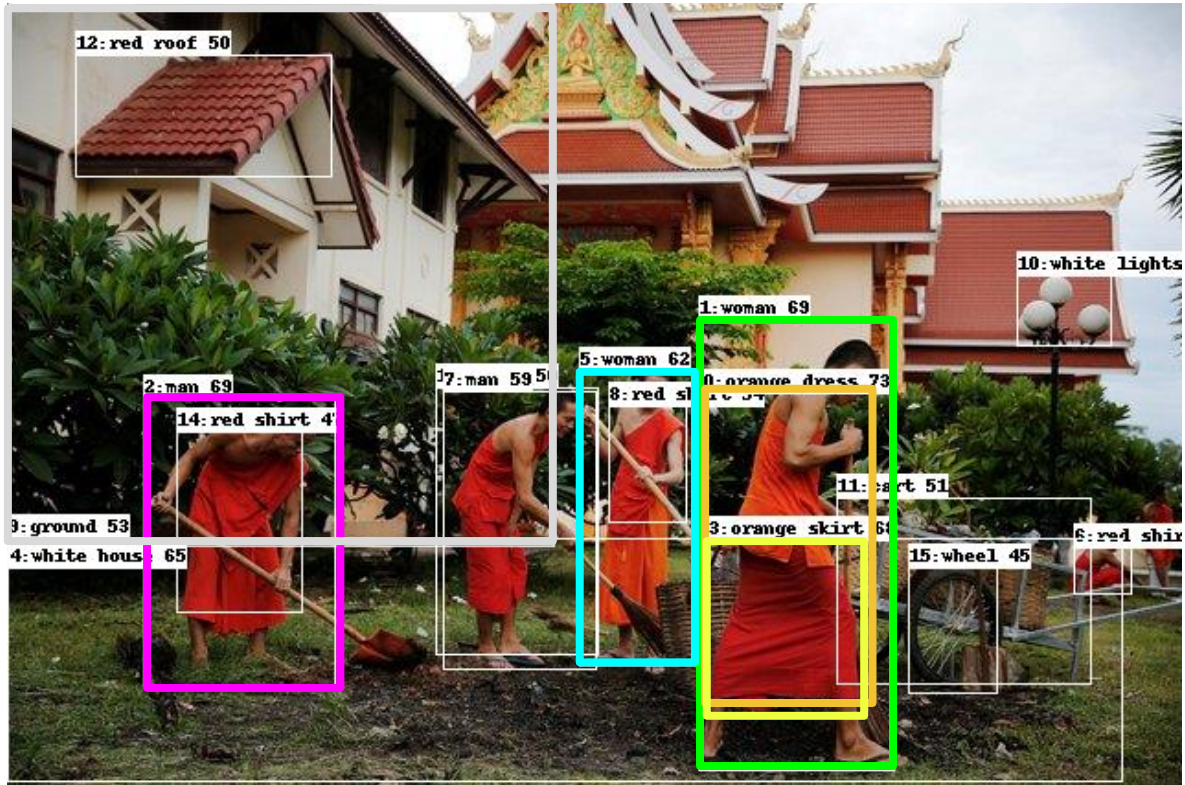
Joint work with Bo Pang, Piyush Sharma, and Radu Soricut

Faster R-CNN

Trained to predict object + attributes of Visual Genome



Detected Objects by Faster R-CNN (Visual Genome)



orange dress

woman

man

orange skirt

white house

woman

Each box comes with <rank>: <optional attribute class> <object class> <score x 100>

Ultrafine-grained Semantic Labels

Spectrum of Semantic Similarity

Category-level (Coarse-grained)

Fine-grained level

Instance level (Ultrafine-grained)

Bridge



Steel red bridge



Golden Gate Bridge



Graph-RISE on Internal Data

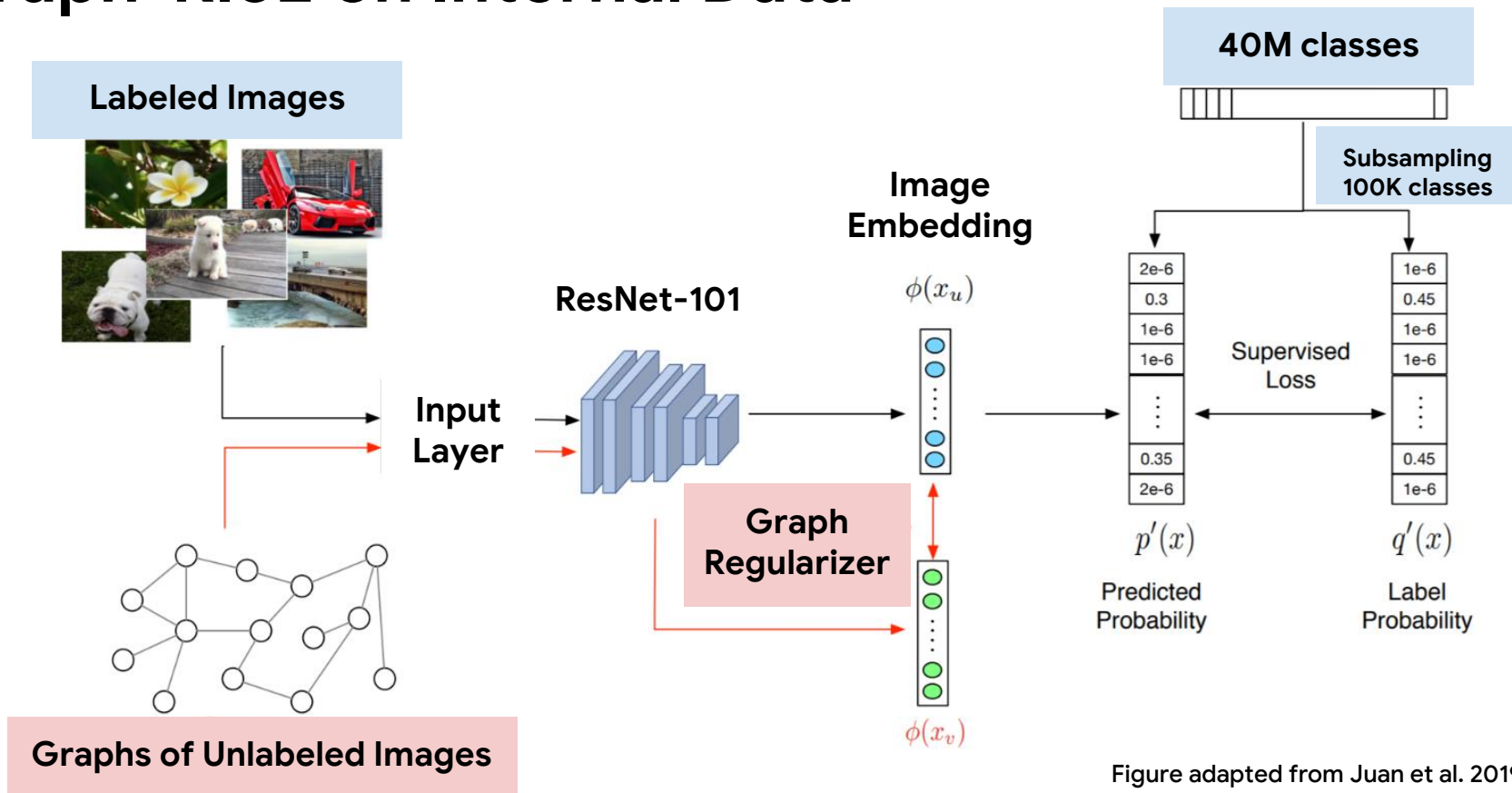
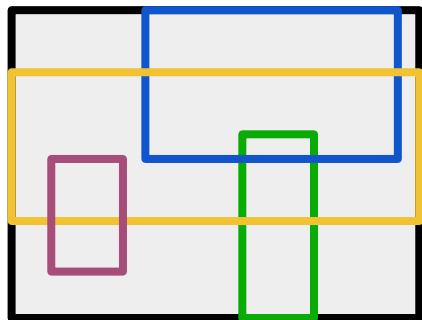


Figure adapted from Juan et al. 2019 5

Our Work (To be presented at EMNLP 2019)



Category-level
(Coarse-grained)



Fine-grained
level



Instance level
(Ultrafine-grained)



Decoupled Boxes + Ultrafine-Grained Featurization

=

Better Image Captioning & Visual Question Answering

Image Captioning on Conceptual Captions



**“monks clean a garden
at a temple .”**

Region Faster-RCNN

“a woman walks through the streets .”

Region Ultra

“monks walking in front of a temple”

Informative Words *induced* by Ultra



“mughal structure
on the way”



“lollipop on a
yellow
background”



“green algae in
the sea”



“breakfast with
coffee and
croissants on a
white wooden
background .”

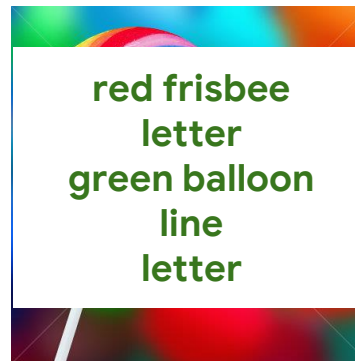
Possible to predict from Top 5 FRCNN objects?



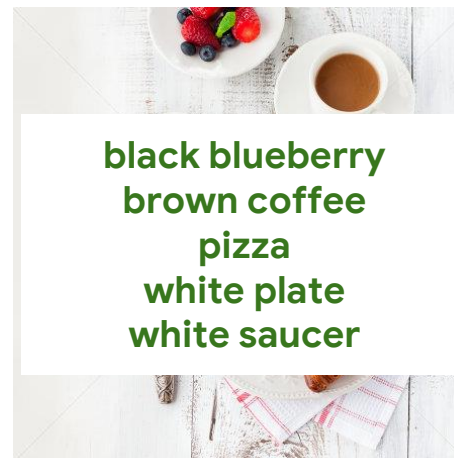
“mughal structure
on the way”



“green algae in
the sea”



“lollipop on a
yellow
background”



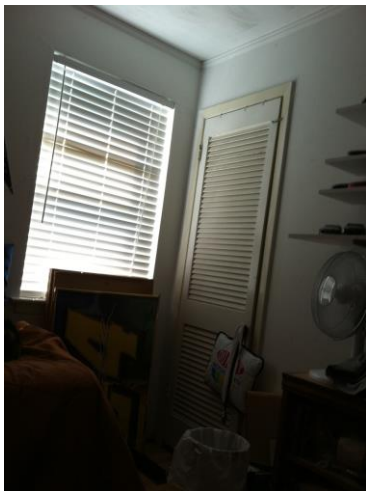
“breakfast with
coffee and
croissants on a
white wooden
background .”

VQA on VizWiz

Data: Gurari et al. 2018

Model: Anderson et al. 2018, Jiang et al. 2018

yes / no



“Is it sunny outside?”

yes
yes

number



“How much money is this?”

1 dollar
20

other



“What is this?”

beer
bbq sauce

unanswerable



“What does it say on this card?”

unanswerable
unanswerable

FRCNN

Ultra