Segmentation from Natural Language Expressions

Ronghang Hu\textsuperscript{1}, Marcus Rohrbach\textsuperscript{1,2}, Trevor Darrell\textsuperscript{1}

\textsuperscript{1}UC Berkeley \quad \textsuperscript{2}ICSI, Berkeley
Traditional Category-based Segmentation

Semantic segmentation for class horse

Instance segmentation for class horse
Image Segmentation from Referential Expressions

Localize the following entity with bounding box: "a dark horse with a woman in a striped shirt"

**Related Work:**
- *Pixel-wise segmentation* for expression: "a dark horse with a woman in a striped shirt"

**Our Work:**
- *Segmentation* from natural language expressions
Our Model – Intuition

- **Embed the image** – spatial feature map through CNN
- **Embed the expression** – final hidden state in LSTM
- **Fully convolutional classification** – match input expression to every location on the spatial grid and up sample
Our Model – Details

input natural language expression
“right woman”

word embedding
LSTM network

$D_{\text{txt}}$ dimensional final hidden state $h_f$
L2-normalized

input image

CNN

$w \times h \times D_{im}$ spatial feature map
L2-normalized

encoded expression

fully convolutional classification network

upsampling with deconvolution

$w \times h$ low resolution response map

$W \times H$ high resolution response map

segmentation output
Experiments

Dataset:
- ReferItGame (Kazemzadeh et al. 2014)
  - Pixel-wise annotation for referential expressions

Baseline approaches as comparison:
1. Combine per-word segmentation results (bag-of-words)
2. Foreground segmentation over bounding box localization methods (e.g. GroundeR)
3. Classification over segmentation proposals (e.g. MCG)
Results

**precision@0.5 IoU**: percentage of test samples with at least 0.5 IoU

**overall IoU**: intersection-area-sum / union-area-sum on dataset
Results

query='3 people on right'

input image  score map  prediction

query='water'

input image  score map  prediction

query='bike'

input image  score map  prediction

query='guy in front'

input image  score map  prediction

Project page and code release:
http://ronghanghu.com/text_objseg
Results

query="man far right"

query="sky above the bridge"

query="big black suitcase bottom left"

query="wall above the people"

Project page and code release: http://ronghanghu.com/text_objseg