Phrase Localization Without Paired Training Examples
Phrase Localization

the handles of the slide
How is this tackled?
Phrase Localization

with strong supervision
Phrase Localization
with strong supervision

- the handles of the slide
- a stuffed animal
- a few wet items
- silver shorts
Phrase Localization

with weak supervision
Phrase Localization
with weak supervision

- the handles of the slide
- a stuffed animal
- a few wet items
- silver shorts
Phrase Localization

with weak supervision

- the handles of the slide
- a stuffed animal
- a few wet items
- silver shorts
How is the supervised data used?

Really intelligent?
Memorize all the phrases with their images!!!!
Phrase Localization

with NO paired data

the handles of the slide

a stuffed animal

a few wet items

silver shorts
Phrase Localization with NO paired data

Off-the-shelf object detectors

Knowledge bases

Text corpora/Word embeddings

Generic images
Detect instances

Filter concepts

Localize
Flickr30k Entities

- Strong supervision: 65.21
- Weak supervision: 38.71
- Non-paired: 50.49

ReferItGame

- Strong supervision: 44.07
- Weak supervision: 20.91
- Non-paired: 26.48
Non-paired setting as a strong baseline
Simpler general methods vs. complex specialized methods
More effective use of paired data

...on top of what can already be achieved without. Better generalization?
Phrase Localization Without Paired Training Examples

Imperial College
London