Phrase Localization Without Paired Training Examples
Josiah Wang, Lucia Specia

- Phrase Localization models are often trained with supervised paired training data.
- What if we don't have such training data?
- Can we still solve the problem?
- Maybe with off-the-shelf tools/model/resources?

**Experimental Results**

- **Flickr30k Entities**
  - Non-paired setting: CC, OI, CC + OI, KAC Net, SPC + PPC
  - Weak supervision: CC, OI, CC + OI
  - Strong supervision: CC, OI, CC + OI

- **ReferItGame**
  - Non-paired setting: CC, OI, CC + OI, KAC Net, SPC + PPC
  - Weak supervision: CC, OI, CC + OI
  - Strong supervision: CC, OI, CC + OI

**Example Output**

- COCO vs. OI
- A very excited drummer
- A five member band
- A long green shirt
- A red toy
- A yellow tennis suit
- Pink blanket
- Trees
- Bag below women in orange
- Hotel door
- Lamp
- Glass
- Pick his nose
- A red toy
- Older male
- Guy in yellow shirt
- A very excited drummer

**Discussion**

- Non-paired setting can be used as a strong baseline for phrase localization (or other V&L tasks)
- Paired data should be used more effectively, on top of what can be achieved with simpler methods without paired data
- Need to understand datasets better and not just blindly running complex models
- General, human-like AI: Better generalisation to different tasks