DeepRemix: an automated remix video generator

**Introduction**
- Given a source video, a *remix video* is a collection of clips that match the original video’s *entity* (ag. share) and *action* (e.g.,
- DeepRemix is an automated way to generate remix videos from any source video.

**Data**
- The COCO dataset consists of images labeled with objects and their bounding boxes. It was used to train the image captioning model.
- The Memnent dataset consists of 3 second video clips labeled with actions.

**Model**
- Im2txt is a combination CNN/LSTM
- It encodes an input sequence into vector embedding using the CNN
- Then it decodes the vector embedding into a caption
- I used a pretrained model trained for 2 million iterations

**Pipeline**
- **source video**
  - extract 1 frame per 3 seconds of video
- **captioning model**
  - run each frame through an LSTM/CNN
- **caption generation**
  - generate 1 sentence per frame
  - a woman lying on a bed with a book
  - a group of dancers around a Christmas tree
  - a group of cheerleaders dancing on a football field
- **caption matching**
  - map source caption to remix caption
  - a woman reclining on a bed with a tablet
  - a group of dancers on a stage
  - a group of cheerleaders performing on a football field
- **remix video**
  - map each caption to remix clip and concatenate

**Results**
- Respondents were shown 4 clip pairs: [source, remix]
- Each remix clip was generated by a *matching scheme*: entity, action, bleu, or human.
- They were asked how well the clips match by [entities | actions | overall]. This was the matching metric.

**Caption Matching Schemes**
- **entity**
  - Pairs as many of the caption’s *entities* as possible, sorted by rarity
  - people standing next to a spinning top
  - top people
  - people on top of a building
- **action**
  - Pairs as many of the caption’s *actions* as possible, sorted by rarity
  - people standing next to a spinning top
  - standing spinning
  - a dog spinning next to a person standing
- **bleu**
  - **Algorithm used in machine translation to assess the similarity of pairs of sentences**
  - people standing next to a spinning top
  - friends spinning on top of a building
  - people on top of a spinning ride

Rey Barceló
rbarcelo@stanford.edu