



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** (eg. *shark*) and **action** (*bite*).
- DeepRemix is an automated way to generate remix videos from any source video.

Data

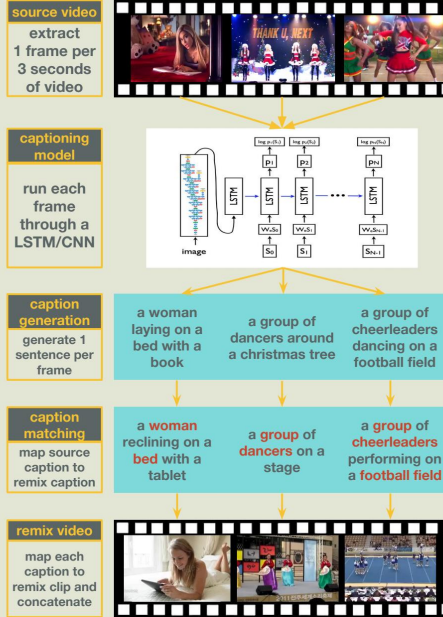
- The **COCO** dataset consists of images labeled with objects and their bounding boxes. It was used to train the image captioning model.
- The **Moments in Time** dataset consists of 3-second video clips labeled with actions.



Model

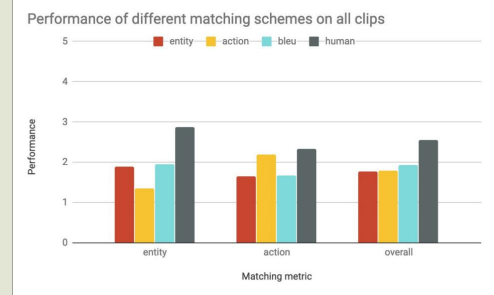
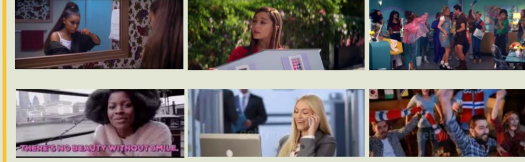
- Im2txt is a combination CNN/LSTM
- It encodes an input sequence into vector embedding using the CNN
- Then it decodes the vector embedding into a caption
- I used a pretrained model trained for 2 million iterations

Pipeline



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

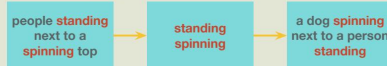
entity

- Pairs as many of the caption's **entities** as possible, sorted by rarity



action

- Pairs as many of the caption's **actions** as possible, sorted by rarity



bleu

- Algorithm used in machine translation to assess the similarity of pairs of sentences



Rey Barceló

rbarcelo@stanford.edu