🔊

Automatic Audio Clip Selection

High-level idea

The audio track of a particular scene is often not the one directly recorded by the microphone during shots (can be noisy), but carefully picked from a high-quality audio database (created by sound engineers) in post-production. (An example video that does this: https://www.youtube.com/watch?v=IbiN9xIdVIM)

Can we use CLIP to identify what is occurring in the scene and automatically retrieve relevant high-quality audio clips, e.g., from Splice? For example: Given a scene of the NYC subway, we may retrieve audio clips of the subway entering the station, people chattering, or even nearby street artists performing.

Rough Timeline (Beginning of July - mid August?)

- Weeks 1-2: Literature review and description of approach (David + Anastasis will collaborate on this)
- Weeks 3-4: Implementation of interactive prototype (e.g. using Jupyter or Streamlit) of the proposed approach (David will work on this)
- Weeks 5-6: Based on the initial prototype, implement a demo branch with the feature in Runway's upcoming video editor (Anastasis + Runway will work on this)
- Weeks 7-8: Perform interviews with Runway users using the demo branch implementation (Runway to co-ordinate and schedule, David to perform the interviews)
- Weeks 9-10: Work on draft for submission for CHI22 (David).

Format

- Bi-weekly check-in meetings to discuss progress with David, Anastasis and other members of the Runway team depending on the stage of the project
- Asynchronous communication on the Runway team Slack for any feedback/discussion/debugging
- Runway will provide stipend for the duration of the 10 week project
- Runway will provide access to GPU compute for training models, running experiments, etc.

References

- CLIP - Connecting Text with Images
- Deep Audio-visual Learning: A Survey