PES UNIVERSITY

Project Review-4

Project Title	:	Searching a video database using Natural Language Queries
Project Guide	:	Dr Mamatha
Project Team	:	Aditeya Baral, PES1201800366 Vishesh P, PES1201800314 Anirudh HM, PES1201800131 Vinay Kirpalani, PES1201800218

Expected Outcome at the end of the project/Contest

At the end of the specified timeline, we will be able to extract all videos which are contextually or semantically related to a particular natural language query.

- The user will be able to use keywords and phrases used in natural conversational speech (including support for different languages and accents), such as a movie title or any specific scene in any video like either a protagonist or any object, and movies containing the specified described frames will be pulled out

- These frames will be analyzed and the video with the most matching frames will be picked and displayed.

Work Plan For the coming weeks

- We are in the process of creating a polished UI for the application
- We are also trying to improve the accuracies of the tags generated for every video by trying various deep learning models.
- We are simultaneously trying to improve the similarity metric and filtering of keywords to provide more relevant matches

Progress as on Date

- We have extracted the natural language query (and translated other languages into English) and then proceeded to use various API's and algorithms to extract keywords.

- We then tagged the entire database of videos by splitting each video into frames and grouped them by scene. We then performed object detection and description generation to describe each scene to give us the context of the video and even used a semi-supervised approach to perform a reverse image lookup on Google to find similar matching videos for more context and relevant searches.

- We then obtained the keywords from the context by choosing the most frequently occurring keywords describing the frames.

- The query was matched with the video by converting each into a vector space and found the similarities between the two feature spaces. This entire process was added into a pipeline to run as an executable.

Any obstacles/challenges and any Assistance Required

- Time is a constraint – a large portion of the video tagging unit takes a lot of time to generate tags since it uses various approaches and then picks the best one.

- Speech recognition – It is not that accurate when it comes to other languages apart from English, leading to mismatched outputs.

- Filtering of keywords – The semi-supervised approach tends to result in a few irrelevant keywords even after taking frequency into consideration.

Key Deliverables for the next milestone

- We hope to be able to perfect the speech recognition and tag generation.
- We will also try to reduce the time taken to run the application.
- We are trying to develop an easy to use and intuitive UI for the application.

Summary

- We have used a **novel approach** by not only using computer vision to perform object detection and tagging but have also combined it with NLP techniques such as sequence models to generate descriptions of scenes with objects such that they can be tagged and described without manually watching them. We also used word embeddings to obtain relevant searches on not just a contextual level but also at a semantic level.
- Our work is going as planned by our team and we have finished about 90% of the expected project.
- We hope to continue to work on this and finish our planned objectives and start work on the final integration, which will be our next goal
- We are also working on the efficiency and are trying to reduce latency and bring in faster results.
- Our code can be found at https://bit.ly/intelnlq

Thank You