Project 2(A)

AutoNarrate_v0.1: A deep learning based tool that narrates the silent parts of a movie, describing it for blind audience.

What you are given:

- A GUI, README with instructions on how to execute, and well-documented code
- Automatically detects silent parts of the movie,
- Uses 'show and tell' deep learning model to caption silent frames [1]
- Also tried using I3D (a two streamed 3D ConvNet) [2] but that produced inaccurate recognition
- Merges the result of above two to form meaningful sentences
- An audio narration module that converts these sentences to speech

What you are required to do:

1. Understand, compile and run the model on 5 videos (10 *points*)
2. Propose and implement some deep learning based changes to improve narration accuracy and realism. Some examples can be more meaningful sentences, proper/improper nouns, more accurate activity recognition, gesture recognition etc. (10 *points*)
3. Develop a deep learning model that uses movie's subtitles/screenplay and video to identify names of each character in the movie(assuming the name occurred at least once in the movie) (10 *points*)

Instructions:

- Major portions of parts 2 and 3 must use deep learning module except for small tweaks and trivial techniques
- Slides describing the project's modules and sample videos and captions are also given
- Each student can pick one and only one of the two projects to work on. Students must inform the choice of their project to TA simply by an email titled Your Roll No: Project 2(A) or Project 2(B) within two days
- All projects are to be submitted electronically to the TA
- Each submission must include a report named P2_YourRollNo.pdf and all code in a file named P2_YourRollNo.zip. Title of the email should be DLF18P2 Your RollNo.
- For Project 2(A), report must include:

- Evidence of running the code successfully in (1), along with details of any changes you had to make. One line description of the 3 major modules implementing the main parts of the algorithm
- Description and justification of improvement(s) you chose to do in (2), ALL the changes in code/external modules you had to use to implement these improvements and how close their results were to the desired output
- Description and design of your method to perform (3) (including details of inputs you will use, any training data and its collection) and its results

References:

[1] O. Vinyals, A. Toshev, S. Bengio, and D. Erhan, "Show and Tell: A Neural Image Caption Generator," Nov. 2014.

[2] J. Carreira and A. Zisserman, "Quo Vadis, Action Recognition? A New Model and the Kinetics Dataset," May 2017.