Artificial Synthesis of Sound Tracks for Silent Videos

Abstract

Swaraj Priyadarshan Dash

Roll - 12, CSE C-1

SIC - 170310417, Regd. No. - 1701209365

Objects make distinctive sounds when they are hit or scratched. These reveal aspects of an object's material properties, as well as the actions that produced them. The task is of predicting what sound an object makes when struck within a visual scene. The algorithm will produce a sound for the hit that is realistic enough to fool human viewers. This "Turing Test for sound" represents much more than just a clever computer trick. Researchers envision future versions of similar algorithms being used to automatically produce sound effects for movies and TV shows, as well as to help robots better understand objects' properties. The important performance criterion of the synthesized sound track is to be time synchronized with the input video. This provides for a realistic and believable portrayal of the synthesized sound, enough to fool a person. The basic algorithm uses a recurrent neural network to predict sound features from videos and then produces a waveform from these features with an example-based synthesis procedure. Deep Learning techniques are generally used in these kinds of problems which involve teaching computers to sift through huge amounts of data to find patterns on their own. Deep learning approaches are useful because they provide freedom to hand-design algorithms and supervise their progress. This research area is an

extension of Automatic Image Captioning projects, that combine the two sensory perceptions, in the form of images and audio. The idea, however, is the same, that is to find *context*, and build a model to learn and reproduce. The scope for this research idea is huge: it can be used to generate audio for video clips without audio, or unclear audios. It can also be used as the Foley Artist, that is responsible for creating an overlay soundtrack that helps movie scenes come alive for the audience. The topic idea is inspired by independent research done by researchers from MIT's Computer Science and Artificial Intelligence Laboratory (CSAIL) and Microsoft Research Fellowship.