
Bad Guy Identifier - Project Plan

11-632 MCDS Fall 2018 Capstone

Yue Cao
yuec1@andrew.cmu.edu

Advisor: Prof. Alex Hauptmann
alex+@cs.cmu.edu

1 Introduction

This project is about building a "bad guy identifier", which looks at a movie trailer or clip and spots the villains in the film, if exists.

This project will be exploratory. We plan to train on self-labeled videos fetched from YouTube or other sources, and exploit audio, video and image features. To the best of our knowledge, there is no research on similar topics for us to reference.

2 Motivation

The success and potential exhibited by deep learning in the last decade has encouraged researches on more complicated domains like autonomous driving and chatbots. These daily life scenarios incorporate many aspects of human senses, such as visual, auditory and even emotional, and requires some understanding of common sense. How well can a computer program master what is considered usual to us is still left open.

Synchronously, we are enjoying an unprecedented level of access to multimedia. Our mobile phone and numerous carefully-crafted applications makes it easy to indulge in a waterfall of images, audios and videos on the web. Our physical world is by nature multimedia, and the digital world is increasingly so. The ability to understand and analyze explicit and inexplicit multimedia data would be highly valuable for developing applications that serve human better.

Thus, motivated by what we can (possibly) do and what we want, this exploratory project takes on a question so naive that a 6-year-old human child could answer: who is the bad guy in this movie? Hopefully by the end of this project, we can gain better understanding on multimedia machine learning and acquiring commonsense.

3 Objectives

Tentative. Build an offline classifier that takes in a short video (less than 5 minutes), identify main characters and classify them into three categories: good, bad, and uncertain. Ideally, the classifier should have a performance better than 50%.

4 Timeline

- Milestone 1. [Sep. 1 - Sep. 17] Dataset preparation.
 - Setup developing environment.
 - Fetch, clean and label short videos.
 - Deliverable: a ready-to-use, clean dataset of videos, labeled with necessary and available feature at this stage, i.e. movie id and title.
- Milestone 2. [Sep. 17 - Oct. 15] Face recognition and character identification. Ground truth label.

- Identify main characters in the clip. Manually label each character indicating whether it is a villain of the movie.
 - Deliverable: for each video, use face recognition and clustering to identify its main characters appeared. For each character, manually label it with ground truth. If possible, associated each character with screen shots containing its figure or face, and/or time intervals of its appearance in the video.
- Milestone 3. [Oct. 15 - Nov. 12] First stage training using only image features.
- Milestone 4. Incorporate audio and/or text (script).
 - If necessary, add features to each video including separate audio track, timestamped scripts.
- Milestone 5. [Nov. 26 - Dec. 10] Final report and project presentation.