Deep Learning First Project Increment Report

Project Team 1
Billy Capps, Class ID 6
Adam Carter, Class ID 7
CS 5542 – Spring 2017

Introduction

This is the summary report of the first iteration of work on our Virtual Field Trip project. The goal for this iteration was to begin the implementation process and implement a framework that will be built upon over the remaining iterations.

Project Goals and Objectives

Our project is to create a virtual environment where students can observe an area and attempt to locate items of educational value. An example would be to tour a cave and locate a stalagmite. Teachers would be able to add new images to the collection. Through deep learning, objects in the images could be identified reducing the time teachers need to spend preparing the images. It will also allow the system to give immediate feedback. The software can either confirm the student selected the correct object or told what object they selected instead.

Our motivation is to provide students with the possibility of having first-hand experiences in places that they cannot physically visit. Caves, pyramids, the moon, and the ocean floor could be visited by students regardless of geography or income level. Scavenger hunts ensure that students are actively participating and focused on areas of educational importance.

System Features and Use Case

As stated in our Project Proposal, the Google Cardboard interface will allow students to select which image they want to explore from a library. Once the image opens, they will be given a brief introduction and then told which items to locate. It is tempting to include additional educational information while the student is exploring; however, the purpose of this project is supplement teacher instruction and not replace it. The interface needs to remain minimalist so that the student can focus on exploring the new environment. When the student believes that he or she has located the requested item and focused the cursor on it, the button on the top of the Google Cardboard device will be pressed. The system will then either congratulate the student and offer the next item to seek or tell the student what item the name of the object they selected and ask them to keep trying.

Approach

Our application is currently leveraging the following technologies:

- Android using Google Cardboard
- Spark MLib (though this is not currently linked to the Android Application
- Clarifai REST API
- Google Conversations API

Since this first iteration is a learning iteration, we chose a data set that was easy to experiment with and capture. While any 360° Panoramic Image would be useful, having easy access to the location would be required to build our first training data set. To that end, we used Adam's Game Room as a template. Using the Google Cardboard Camera, that room was captured and converted into a Cardboard-compatible image. We then developed an Android Application that could render the view using the Cardboard API, and accept trigger-based input.

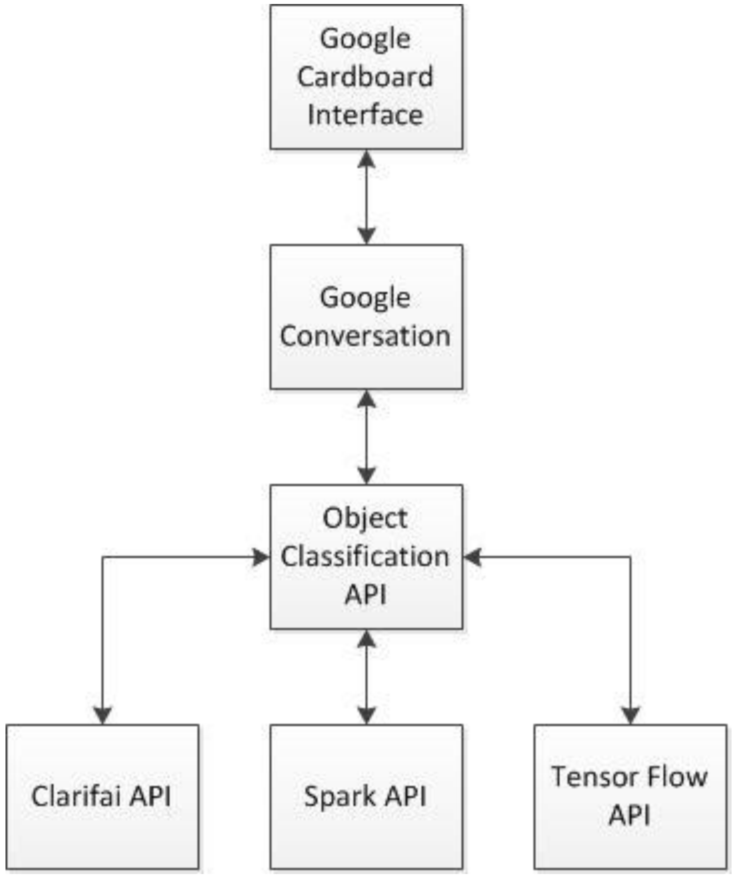
Using the Game Room again, we took videos of individual objects, then broke those video down into frames to generate training data. Currently, the Game Room dataset features 9 classifiers or distinct elements in the room. For test data, the Android Application is capable of capturing a screen shot of the center point in the VR view, which we can then use for our test data. The video used to generate the training data was sparsely sampled, and the resolution on the panorama view is not very clear, so the accuracy of the model is currently just above 53%. Technical problems with the OpenCV libraries prevented the use of a larger training data set for the Spark MLib Api.

Related Work

No new research was done into related work during this iteration, though a local news story seemed relevant. The V Form Alliance is a new startup in Kansas City whose mission is to "allow elementary and middle school students to take a virtual reality 'field trip' exploring landmarks in Kansas and Missouri that are relevant to black history." [1] Clay Middle School has a page of virtual field trips. [2] This is a collection of web pages that offer virtual tours of places with educational value. Google Cardboard Camera allows users to create their own VR photos. [3] NYT VR takes users on a journey to places of significance and includes information about the associated New York Times stories. [4] Titans of Space lets users explore our solar system using virtual reality. [5]

Application Specification

At this point in the project, most of the pieces are in the early stages of development and are largely unconnected. As future iterations proceed, we expect this connectivity to improve. We will touch on the various systems and their function.



Implementation

At this point in the project, most of the pieces are in the early stages of development and are largely unconnected. As future iterations proceed, we expect this connectivity to improve. We will touch on the various systems and their function.

Android Application

This application leverages the VrPanoramaView implemented by Google VR. This is separate from the new GvrView full VR implementation using OpenGL, which means it has limited functionality, but is sufficient for this projects' needs. This means that the native onCardboardTrigger() method cannot be used, so we instead trap any screen touch event (which is essentially what the trigger does) to take a screen shot using the Pitch and Yaw recorded by the headset to determine the current center of the Panoramic Image.

One difficulty of this task was handling images that would wrap around the 'seam' of the image, where the flat image boundaries would need to be crossed to complete the screenshot. Because the VrPanoramaView is more limited in its feature set, the native Android capability of capturing a screenshot is not available. A custom algorithm had to be written to find the current view center and grab the screen shot manually. For the case where we would wrap the scene, a separate image file is referenced that has been rotated 180° so we can cleanly capture the seam.

Spark Image Recognition

The code from Lab4 and Lab5 had to be uplifted because OS-specific errors were occurring that stem from an older version of OpenCV. To overcome this, the JavaCP and JavaCPP libraries had to be uplifted from version 0.11 to 1.3.1. This also meant that features like SIFT Feature Extraction were no longer available (as they are not freely available for patent issues). The code was converted to use the AKAZE feature extraction model, and the results were faster performance wise, but lead to smaller feature vectors and less accurate prediction. Results are descripted below in the documentation setting.

There are several current limits to this approach. First are the OS issues that prevent using a larger dataset (OpenCV is not entirely stable, and frequently causes fatal OS errors that interrupt parsing). A larger data set would surely guarantee better results, but could not be realized for this iteration. Second, I believe converting to Greyscale significantly reduces accuracy. Both SIFT and AKAZE models tend to strongly favor one image (The Star Wars posters, which are chosen significantly more frequently than any other classifier) which would appear different if processing RGB vectors separately. That change was determined to be outside of a realistic scope for this iteration. Finally, the low quality of our VR screenshots in particular makes accurate analysis more challenging. We hope to find ways to address this shortcoming in the next iteration.

Documentation

Presented are some of the screen shots taken from our application. We'll give a quick overview of some of the visual artifacts used to build our dataset, then some of the artifacts created by our application.

Basic VR Image of Game Room used as the basic for early analysis

Screenshot of Google Cardboard Viewer while running

Six images captured from our Android Application of different areas in the VR Viewer.

The Spark Code used to do our Image Training pattern used these images as part of its test data set. Because of the previously mentioned technical and issues running a larger data set and a general lack of high quality imaging, the training data is still fairly poor. Presented here is the confusion matrix generated by our model against the test data.

First is the Confusion Matrix generated by SIFT for both randomly sampled training images plus our screen shots.

```
|=================== Confusion matrix ==========================
3.0  0.0  0.0  0.0  0.0  0.0  1.0  0.0  0.0
0.0  4.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0
0.0  0.0  2.0  2.0  0.0  0.0  0.0  0.0  0.0
0.0  1.0  0.0  2.0  0.0  0.0  1.0  1.0  1.0
0.0  0.0  1.0  1.0  2.0  1.0  1.0  0.0  0.0
0.0  1.0  0.0  0.0  0.0  5.0  0.0  0.0  0.0
0.0  0.0  0.0  4.0  0.0  0.0  3.0  0.0  0.0
0.0  0.0  0.0  2.0  0.0  0.0  0.0  2.0  0.0
2.0  0.0  0.0  0.0  0.0  0.0  0.0  0.0  2.0
0.5555555555555556
```

This is the Confusion Matrix generated by AKAZE against a slightly larger training data set against the same test data set.

```
|=================== Confusion matrix ==========================
2.0  0.0  0.0  0.0  0.0  1.0  1.0  0.0  0.0
0.0  2.0  0.0  1.0  0.0  0.0  0.0  0.0  1.0
0.0  0.0  2.0  0.0  0.0  0.0  2.0  0.0  0.0
0.0  1.0  3.0  1.0  0.0  1.0  0.0  0.0  0.0
0.0  0.0  0.0  1.0  2.0  1.0  2.0  0.0  0.0
0.0  0.0  0.0  0.0  0.0  6.0  0.0  0.0  0.0
0.0  0.0  0.0  1.0  0.0  2.0  4.0  0.0  0.0
0.0  0.0  0.0  0.0  0.0  1.0  1.0  2.0  0.0
0.0  0.0  0.0  0.0  0.0  2.0  0.0  0.0  2.0
0.5111111111111111
```

Finally is the AKAZE Confusion Matrix against only our Screenshot Images.

```
|=================== Confusion matrix ==========================
0.0  0.0  0.0  0.0  0.0  1.0  1.0  0.0  0.0
0.0  0.0  0.0  1.0  0.0  0.0  0.0  0.0  1.0
0.0  0.0  0.0  0.0  0.0  0.0  2.0  0.0  0.0
0.0  0.0  2.0  0.0  0.0  0.0  0.0  0.0  0.0
0.0  0.0  0.0  1.0  0.0  1.0  2.0  0.0  0.0
0.0  0.0  0.0  0.0  0.0  2.0  0.0  0.0  0.0
0.0  0.0  0.0  0.0  0.0  1.0  2.0  0.0  0.0
0.0  0.0  0.0  0.0  0.0  1.0  1.0  0.0  0.0
0.0  0.0  0.0  0.0  0.0  2.0  0.0  0.0  0.0
0.19047619047619047
```

These results are hardly confidence inspiring, but we believe they can be improved upon for our next iteration.

Project Management

Since beginning this project, we lost a team member. That made distributing the work a little more challenging. We have also chosen not to use ZenHub for tracking the work being done on this project. Both of us have working experience, and consciously chose not to work through the overhead of a tracking tool for a two-man project. Because of this, we have also not tracked hours spent on this project.

Adam was responsible for the initial data set generation, writing the initial Android application using Cardboard viewer, and adapting the Spark code to use our new data set.

One of the largest issues encountered was the generation of a screen shot from the Cardboard Viewer that could be trigger using the cardboard trigger. The documentation around both the VrPanoramaView and the GvrView and associated classes (Google's newest VR approach that replaced the previous CardboardView API) is very poor, and new enough that there have been few questions asked or tutorials generated. It took a reasonable amount of experimentation to determine that the OpenGL libraries were both overkill and not wholly suited to our needs, but the VrPanoramaView exposes virtually none of the underlying functions to developers.

There has also been mention made about the difficulty in training a dataset. The OpenCV libraries used for image processing are not fully stable, and routinely crash above a certain volume of image data in the Feature Extraction process. Part of the next iteration, which will focus on integrating the Spark API with our application, will need to be resolved to improve the accuracy of our models.

We both contributed to the documentation effort for our system.

The next major iteration requirements are as yet undocumented, but we expect to further develop our Spark MLib API. We also plan to select a new data set that is tailored to our project goals for adaption in the next iteration.

References

[1] http://www.startlandnews.com/2017/02/virtual-reality-field-trips-offer-black-history-experiences-kc-students/
[2] http://www1.ccs.k12.in.us/clm/media-center/fieldtrips
[3] https://play.google.com/store/apps/details?id=com.google.vr.cyclops&hl=en
[4] https://play.google.com/store/apps/details?id=com.im360nytvr&hl=en
[5] https://play.google.com/store/apps/details?id=com.im360nytvr&hl=en