Final Project Proposal

Team: Jennifer Crosby and Jacob King

Topic: Identification of something of interest in a video

Dataset: Video from NOAA Okeanos underwater autonomous ROV expeditions to the Marianas Trench

1. **What problem did you select and why did you select it?**

   One of the main oceanographic projects today is to discover and protect marine life and vegetation, done while mapping deep water locations through video through unmanned expeditions to areas that are otherwise impossible to see. During and shortly after each expedition, hundreds of expert scientists are available to view this video and analyze and classify what is being seen. However, they allot little time outside of the expedition window to watch the videos. At the current time, these experts must watch the entire videos in order to determine if anything captured is of use to their research. These videos are often more than four hours long, and require massive time expenditures from these experts, for sometimes 10 minutes of value. Our problem consists of helping in this project by building and training a neural network to shorten the amount of time the experts need to spend watching the video by reducing the video watched to only potentially interesting images. Ultimately, the goal is to classify what is seen and parse that video to smaller videos that would be sent to the respective scientists who have interest.

   NOAA's (National Oceanic and Atmospheric Administration) expeditions cost millions of dollars each, which makes it important that the information is utilized. Currently, however, they have too much data for people to effectively analyze and much has not been even viewed. This results in much wasted money and time for both NOAA and the experts. Our contribution will be the first step in enabling faster and more useful data analysis in the future

2. **What database/dataset will you use? Is it large enough to train a deep network?**

   We are going to use video, in the form of 1 second image increments, obtained from NOAA Federal, taken on Okeanos expeditions to the Marianas Trench. The resolution of the images are 640 pixel width x 360 height, for 20 minutes (276.5 million pixels in 1200 (+) images). Although, the final video we obtain may be different. We consider this large enough to train a deep network. We may also have more than one video sample.

3. **What deep network will you use? Will it be a standard form of the network, or will you have to customize it?**

   We will be using CNN based on AlexNet, as described in

   Ullah, J. Ahmad, K. Muhammad, M. Sajjad and S. W. Baik, "Action Recognition in Video Sequences using Deep Bi-Directional LSTM With CNN Features," in IEEE Access, vol. 6, pp. 1155-1166, 2018.

   Later portions of this project will utilize DB-LSTM network to make further improvements.

4. **What framework will you use to implement the network? Why?**

   We will use PyTorch because we will use CNN to identify which images are “interesting.” We propose to use parameters of the pre-trained CNN model, called AlexNet for feature extraction. AlexNet has five convolution layers, three pooling layers, and three fully connected layers. Each layer is followed by a norm and ReLU nonlinear activation function.

5. **What reference materials will you use to obtain sufficient background on applying the chosen network to the specific problem that you selected?**

   Ullah, J. Ahmad, K. Muhammad, M. Sajjad and S. W. Baik, "Action Recognition in Video Sequences using Deep Bi-Directional LSTM With CNN Features," in IEEE Access, vol. 6, pp. 1155-1166, 2018.

   https://towardsdatascience.com/detecting-animals-in-the-backyard-practical-application-of-deep-learning-c030d3263ba8

6. **How will you judge the performance of the network? What metrics will you use?**

   Our main performance metric will be the accuracy obtained by using a Confusion Matrix.

   Since we would prefer a false positive to a false negative because we would rather have the experts spend a couple of extra seconds viewing something irrelevant to them than have them miss an important find. To evaluate this, we will use a Miss Rate, or False negative Rate (FNR).

   We will use Cohen’s kappa because we are doing essentially binary classification. We then train our algorithm independently on both targets (Jacob’s and Jennifer’s), and use Cohen’s kappa on the results.

7. **Provide a rough schedule for completing the project.**

Monday	Tuesday	Wednesday	Thursday	Friday	Saturday	Sunday
April					11	12 1
13 Exam 2 2	14 Exam 2 3	15 Exam 2 3	16 Exam 2 4	17 Exam 2 4	18 Exam 2 4	19 Exam 2 4
20 Exam 2 due 5	21 6	22 6	23 6	24 7	25 8	26 8
27 Due Date and Presentation						

1. **Determine and request data needed (RGB, specify resolution 1280 pixels x 720 pixels , 1 second interval images, and 20 minutes time amount)**
2. **Receive data, upload data**
3. **Make and encode our manual targets; create target file**
4. **Code Neural Network Framework**
5. **Initial Train and Test**
6. **Tweak Model in terms of hyperparameters**
7. **Train and Test**
8. **Get good looking outputs and get makes PowerPoint presentation and Report**