Project 3: Image Recognition using Convolutional Neural Networks

Due: Friday, December 1 (see below for list of deliverables)

Project Overview and Goals

This is a team project. The teams will be announced during Monday, November 13 class.

The Digital Democracy project (www.digitaldemocracy.org) is providing a dataset of screen captures from the videos of legislative hearings that took place in the state of California (and possibly in some other states) since 2015. The images contain faces of different state legislators and also include some other labelling information.

Your goal is to

1. Build a machine learning pipeline that analyzes and properly classifies the images from the Digital Democracy dataset.
2. Use Convolutional Neural Networks and Google’s TensorFlow library as part of your machine learning pipeline.
3. Measure the accuracy of your classification methodology. A Kaggle competition will be set up for this data set. Attempt to beat the baseline accuracy of about 89%.
4. Write a comprehensive report detailing your work. For this assignment your report shall read as a paper you are intending to submit and publish at a conference or a workshop.

Digital Democracy Dataset

Digital Democracy (http://www.digitaldemocracy.org) is a project run by Cal Poly’s Institute for Advanced Technology and Public Policy. The technical co-PIs on the project are Professors Khosmood and Dekhtyar from the CSSE department. The project brings searchable transcripts of State Legislative Committee hearings to the screens of many people and organizations interested in understanding how the legislative process works at the state level, and in monitoring and improving it, as well as interested in seeing certain pieces of legislation pass or fail. At present, the Digital Democracy website tracks legislative hearings in the states of California and New York. Staring Winter 2018, the site will also allow access to Florida and Texas state legislative hearings.

The Digital Democracy process curates the process of building and up-levelling the transcripts of legislative hearings. As part of the up-leveling process, all speech in the hearings is ascribed
to specific speakers, and is tagged with a variety of features allowing for improved search and analysis of the legislative activities. As part of the process, the actual assignment of speech to specific speakers is one of the most time-consuming and error-prone parts of the up-levelling process. To speed up and improve this procedure, the Digital Democracy project is working on various automated speaker recognition techniques using text, audio of the speech and video (images) as the source material.

To aid the process of speaker identification from video, the project has collected a dataset of about 494,000 images of individual lawmakers. Each image is a 64x64 crop and downscale showing a face. Only images of lawmakers (i.e., members of State legislative bodies) are collected in the dataset (in general committee testimony may include a lot of non-legislators testifying, possibly for the first and only time in their lives.) With each lawmaker image, the following information is made available (the ground truth):

1. The unique Identifier of the person in the Digital Democracy database (Pid). This value is considered to be the main class label for each image. The label can be tied to a specific name of the person, and this information can be provided to you to provide you some context in which your learning will take place. There are a total of 398 unique lawmaker Pids in the dataset. The distribution of images to Pids is not uniform, some lawmakers have tens of thousands of images available, some only a few hundred. This is a function of the lawmakers’ activities in the legislative committees, and their tenure in the state legislature (the dataset goes back to 2015, and contains more images of the lawmakers who served in both 2015-2016 and 2017-2018 legislatures).

2. Unique Id of the video in the Digital Democracy database from which the image was taken (Vid). The Digital Democracy database (DDDB) contains a variety of metadata that can be associated with a Vid, including the date of the hearing, the legislative committee whose hearing the video represents, the bill that is being discussed in the video, and, in fact, the transcript of all speech that took place in the video. Note that a single video is typically a 20-30 min. fragment of a hearing, and a specific legislative hearing can last for hours thus being represented by multiple videos.

3. An integer number representing the position (the order) in which the image appears among other images captured from the same video. For example, Position =1 means that this is a first chronologically image captured from the video, Position = 14 means that this is the fourteenth such image. Unfortunately, the exact timestamp of the captured image is not available.

The dataset is about 3 GB in size and is available under /data/face_dataset/ on JupyterHub. You are allowed to make local copies of the dataset for your use, but we ask that you do not redistribute the dataset to any entities outside of DATA 401 and do not post the dataset in any place where it can be accessed by general public.