Journal

Kaitlyn Chen (kjc256), Kevin Zhou(klz23), Greg Bodik(gnb23)

Week of 9/26 - 10/02

objectives

- Do background research on necessary technologies:
  - openCV
  - HMM
  - CNNs
  - Identify data sources
- Narrow down project scope
  - specify project goals and desired outcome
  - determine course of action
  - write design doc accordingly
- Think of punny name for project

progress

- Kevin:
  - Met with Nikhil to talk about possible project ideas and gain insight into the necessary technologies such as openCV and the process of converting video input to images
  - Read and watched some openCV tutorials
  - Read about ALR
  - Design doc
  - Trying to find GRID data
- Greg:
  - Meeting with Nikhil to learn about openCV and gain ideas on project scope
  - Found source for data - includes words and sentences from TED and TEDx talks (http://www.robots.ox.ac.uk/~vgg/data/lip_reading/index.html#applications)
  - Did background research on ALR field and required pipelines
  - Added above info to design doc
- Kaitlyn:
  - Met with Nikhil to talk about possible project ideas and gain insight into the necessary technologies such as openCV
    - converting video input to images and using HMM's
  - Read articles about traditional ALR systems

- Wrote design doc

Week of 10/03 - 10/09 (Prelim season 😩)

objectives

- Think of punny name for project
- Continue to do background research on necessary technologies:
  - openCV
  - HMM
  - CNNs
- Acquire datasets (LRS3)

progress

- Kaitlyn
  - Read articles about CNNs
  - Completed openCV tutorial for Open CV basic
- Greg
  - Set up access to CDS server
  - Completed open CV tutorial for Open CV basic
  - Worked on incorporating NLP into predicting word out of list of possibilities instead of predicting the next word
- Kevin
  - Completed openCV tutorial for Open CV basic
  - Read articles about CNNs

Week of 10/10 - 10/16

objectives

- Continue to do background research on necessary technologies:
  - openCV
  - CNNs
- Write a design doc
- Figure out remote processing

progress

- Kaitlyn
  - Read paper:
    - End-to-End Audiovisual Speech Recognition (Petridis et al. 2018)
  - Looked at and tried out demos of automatic lip readers

- Greg -
  - Read Papers:
    - "Lip Reading in the Wild" (Chung and Zisserman 2017)
    - "Lip Reading sentences in the Wild" (Chung 2017)
    - "LRS3-TED: a large-scale dataset for visual speech recognition" (Afouras 2018)
    - "Listening with Your Eyes: Towards a Practical Visual Speech Recognition System Using Deep Boltzmann Machines"
  - Figure out accessing data set through the server (So much wget documentation 😥)
  -
- Kevin
  - Read papers:
    - "Lip Reading sentences in the Wild" (Chung 2017)
    - "Lip Reading Using Convolutional Neural Networks with and without Pre-Trained Models" (Ozcan and Basturk 2019)
-

Week of 10/17 - 10/23

objectives

- Continue to do background research on necessary technologies:
  - openCV
  - CNNs
    - Loss functions
    - learning rates
    - Data augmentation
- Figure out remote processing and download/merge files from LRS3 dataset

progress

- Kaitlyn
  - Looked at the API for the face_recognition Python library
  - Read about convolutional neural networks
- Greg -
  - Read Paper: "Automatic Lip-Reading System Based on DeepConvolutional Neural Network and Attention-Based Long Short-Term Memory" (Lu, Li 2019)
  - Created subset of LRS3 data (don't need all 4000 videos for running basic scripts)
  - EDA for LRS3 data
  - Discussed the issue of word boundaries with Tushar

- Kevin
  - Reached out to the Computational Linguistics Lab to ask for advice on the lip reading process in general as well as the word boundary issue
  - Received a response that discussed LSTMs and got the responder's contact information

Week of 10/24 - 10/30

objectives
- Organize dataset on CDS server
- Learn how to extract/process lips from still images containing a face
- Read literature to determine optimal size/dimension for lip boundaries based on past research

progress
- Kaitlyn
  - Wrote a script that would take an image from a local directory and crop it to ~10 pixels around the subject's mouth (if there was a face in the original image)
    - Using `face_recognition`
  - Created a super sick PowerPoint showing off our extract lips script
  - McDonalds with the boys
- Greg
  - Collected sample images for testing lip extraction script
  - All 125gb of data comprise thousands of folders, one for each ted talk.
    - Each video file broken into a sub video for a single sentence. Hence "Lip Reading Sentences"
- Kevin
  - Emailed professors Weinberger and Artzi
    - Received responses from both, who pointed us to potential helpful papers or people to contact, such as CUVL and "LipNet: End-To-End Sentence-Level Lipreading" (Assael et al 16)
  - Read through the above paper ^

Week of 10/31 - 11/06

objectives
- Continue to do background research on CNNs
- Extract lip images from training dataset
- Try out lip reading demo on CDS server

- Determine algorithm to decide which frames of a video we want to feed into our model

progress
- Kaitlyn
  - Wrote a script that would take a video from a directory, convert it to images of each individual frame, and store the images in a separate directory
  - Passed that processed data through the extract lips script and discarded a video from the pre-training set if the speaker's lips could not be "extracted" in every frame
- Greg
  - Met with Nikhil to discuss feasibility of Lip Reading CNN right now
- Kevin
  - Read the parts of the papers that talked about how to decide which frames to select for the model

Week of 11/07 - 11/20 (Study Period for Semi-finals 💀)

Week of 11/21 - 11/27 (Thanksgiving 🦃)

objectives
- Redefine project scope
- Research Keras documentation
- Research Haar cascade

progress
- Kaitlyn
  - Reconsidered project scope to include possibility of face sentiment analysis to improve/enhance eventual lip reading model
  - Trained 2-layer model on 28,273 training images through 50 epochs
- Greg
  - Reconsidered project scope to include possibility of face sentiment analysis to improve/enhance eventual lip reading model
  - Found data needed for face sentiments
  - Started EDA on said data
  - Rebalanced train and test sets, found weights, excluded non-viable classes
- Kevin
  - Reconsidered project scope to include possibility of face sentiment analysis to improve/enhance eventual lip reading model

- Found articles and models about face sentiment analysis

Week of 11/28 - 12/04

objectives

- Improve our facial sentiment analysis model (add layers)
- Adjust model weights for each emotion
- Expand dataset artificially through image augmentation

progress

- Kaitlyn
  - Pair programmed a sequential convolutional neural network using Keras API for tensorflow
  - Asked Nikhil for ways to improve our model and what layers to add to it
  - Researched other possible computer vision projects such as age prediction
- Greg
  - Started building CNN with just two layers
  - Original data included 7 classes, removed "disgusted" class because it was less than 2 % of the data.
  - Modified CNN architecture to differentially weigh inputs by how that class' representation in the data (ie spend more time training on underrepresented classes).
- Kevin
  - Started creating the GUI first using Java Swing and then switching to Python's tkinter
  - Created the buttons and the formatting layout through the Button and Canvas widgets

Week of 12/05 - 12/11

objectives

- Compare different CNN architectures like ResNet50 and VGG19 to our model
- Create a GUI
- Create a D3.js visualization of our model's accuracy + confusion matrices
- Think of a better name for the project

progress

- Kaitlyn
  - Pair programmed and added more layers to the convolutional neural network after talking to Nikhil

- Altered training dataset using image augmentation
- Debug web cam input to model
- Greg
  - Researched Haar Cascades to create face bounding box in the end GUI
  - Install all Keras dependencies locally and on server for testing the gui implementation
  - Connect model weights (.h5 file) to GUI
  - Use cv2.videoCapture to launch camera
- Kevin
  - Implemented launching the camera and loading a file through GUI button clicks
  - Researched Haar Cascades for face bounding box
  - Got scripts to run through button clicks

Possible project names:
- Face Off
- IMAGEin
- Lips and Chips
- A moment on the lips, a lifetime on the hips
- XPress

Week of 12/12 - 12/18

objectives
- Complete the final report
- Create a powerpoint for final presentation
- Connect model (backend) to GUI (frontend)
- ASSEMBLE EVERYTHING

progress
- Kaitlyn
  - Wrote part of the Overview, Background, Methods and Results, Lip Extraction, and Future Work sections of the final paper
  - Attempted to build an age prediction model using the UTKFace dataset 😞
    - Had trouble parsing the ages of the image subjects into different directories
  - Added slides to the final presentation
- Greg
  - Finally got to use what we learned about CNNS to refine emotion classifier

- Trained CNN so many times in Colab and Kaggle notebook 😳
- Compared 5 convolution model accuracy to standard processing pipelines like VGG16, VGG19, and ResNet50
- Graphed model loss and accuracy
- Graphed confusion matrix for model test predictions
- Used summary data from Keras to build d3 vizualization for model accuracy
- Generated paper figures in matplotlib
- Wrote paper
- Made presentation
- Kevin
    - Fully connected the updated models to the GUI and debugged issues with the file reading
    - Added introduction and future features buttons to the GUI, which open a text box when clicked
    - Added pictures to the GUI
    - Researched ResNets for information on how to potentially improve our models
    - Worked on the paper and the presentation slides