CS696E Final Presentation

CK Wijesundara

What Did I Do?

- I worked on three projects in two broad domains: facial recognition and scene segmentation
- For facial recognition:
  - I worked on a face clustering project that extracts faces from images, converts them into 128-dimensional encodings, then clusters them based on the similarity of those encodings
  - I also worked on an age classification project that attempted to classify faces into several age groups (<=12, 13 - 17, 18>=)
- For scene segmentation:
  - I worked on a background segmentation project that segments people vs. background, thus allowing you to extract backgrounds for later processing

Who Cares?

- Ideally, Dan
    - A successful age classification project would allow investigators to automate, at least a little bit, the process of sorting questionable images to see which contain underage children
    - A successful face clustering project would allow investigators to better figure out patterns of people appearing across different images without having to do as much manual investigation
    - A successful background extraction project would allow investigators to easily extract backgrounds from different datasets, which could then be used for further projects involving those background images (object detection, clustering, etc.)

What Approaches Did I Take? [1/3]

- Age classification:
  1. Input images containing people
  2. Face extractor extracts faces from those images
     - Pre-trained Dual-Shot face detector model
  3. Feed extracted faces into pre-trained age classification model
     - The three-class classification model was trained on a combined dataset of UTKFace and APPA-REAL
     - The two-class classification model was trained on a combined dataset of FairFace and APPA-REAL
  4. Output CSV file containing age classification labels for all extracted faces

What Approaches Did I Take? [2/3]

- Face clustering:
  1. Input images containing people
  2. Face extractor extracts faces from those images
     - Pre-trained Dual-Shot face detector model
  3. Feed extracted faces into pre-trained face encoding model to generate 128-dimensional encodings for each face
  4. Cluster encodings using Chinese Whispers algorithms
  5. Output face images with clustering labels

What Approaches Did I Take? [3/3]

- Background extraction:
  1. Input images containing people
  2. Pre-trained DeeplabV3+ model generates binary segmentation mask
  3. Output segmentation masks as Numpy data files

A bit of a demo...

Performance and Accuracy Results: Age Classification [1/3]

- Three-class age classification results on Kriti’s dataset:
    - Accuracy: 0.6081
    - Precision <= 12: 0.4103
    - Precision 13 to 17: 0.4192
    - Precision >= 18: 0.6800
    - Recall <= 12: 0.2871
    - Recall 13 to 17: 0.2732
    - Recall >= 18: 0.8369
- Two-class age classification results on Kriti’s dataset:
    - Accuracy: 0.6737
    - Precision Adult: 0.8446
    - Precision Child: 0.5672
    - Recall Adult: 0.5487
    - Recall Child: 0.8542
- Took approximately 12 hours to train 35 epochs of two-class model using a 1080Ti (111, 406 training images, batch sizes of 64, 8 subprocesses for data loading)
- Classification script took 74.8106 seconds to classify 13,233 LFW face images (i.e., approximately 176.8867 faces images/second) using batch size of 32, 0 subprocesses for data loading

Performance and Accuracy Results: Face Clustering [2/3]

- For having some measure of accuracy, I created a small test set of 55 face images, manually clustered them myself, then compared the clustering algorithm’s results
  - My accuracy measurement = (# correctly clustered images) / (# images) = 37/55 = 0.6727
  - Not super useful however, because the clustering could figure out more unique, better clusters than I used for making my “ground truth” clusters
- Performance metrics:
  - 63 “faces” extracted in 21.8403 seconds → ~0.3467 seconds/face (will vary depending on image resolutions and number of faces in images)
  - 2.3993 seconds to generate embeddings of 55 faces
  - 0.0243 seconds to cluster 55 faces

Performance and Accuracy Results: Background Segmentation [3/3]

- Created ground-truth portrait segmentation dataset of 20 images (scraped 20 images from Google Images and then created ground-truth masks myself)
- Average Intersection over Union was 0.8984
- Time taken to generate masks was 2.3583 seconds, which means approximately 0.1179 seconds/image (including time taken to load and initialize segmentation model before processing images)

Additional Commentary

- Age classification ended up being way harder than I thought it would be, very dataset dependent
  - Lots of facial recognition datasets don’t have any age labels
  - Facial recognition datasets that DO have age labels have different age brackets for their labels
  - They tend to use different crop/padding settings
  - Worth testing additional combinations of datasets, as well as model architectures