Amir Sotoodeh
CS5990
3/7/19

Progress Report #1

Several weeks in, I was able to gain a significant amount of understanding of deep neural network models commonly utilized for a facial recognition task. Additionally, I was able to decide that the LFW (labeled faces in the wild) dataset would be best suited for this task for any additional training that may be needing during the transfer learning process for the outer layers of the network. However, we are not limited to this dataset alone; should an additional dataset be discovered (i.e Youtube Faces), these images may also be utilized.

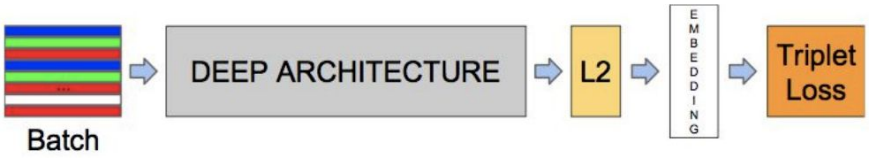
As such, the key components of this task will involve a significant amount of image processing as well as require a fundamental understanding of how metrics of facial boundaries/features are calculated. I will be utilizing a Microsoft Research paper, [1] *Deep Residual Learning for Image Recognition*, for such understanding as well as make use of Davis King's software library: dlib and Opencv for obtaining facial metrics/image processing functions. Additionally, with the specified requirement of running the machine learning model on a live feed of webcam images, it is necessary to consider the complexity of the deep network of choice. Thus, the final model for use has not yet been decided upon, but has been narrowed down to a select few [3]: ResNet-34, VGG Variant(s), FaceNet. Throughout the research phase, it seems to me that there is a common method used throughout all facial recognition projects. They seem to follow they same steps:

1. Detection of face (DNN detector, HOG and Linear SVM, Haar Cascade, etc)
2. Region of face is translated into multidimensional vector/quantification (Dlib, OpenCV)
3. Numerical representation of face is used for classification. (DNN, kNN, etc)

Given the aforementioned procedure, it may be wise to consider a DNN not for classification, but rather for the generation of a n-dimensional vector to represent a given image of an individual's face for the sake

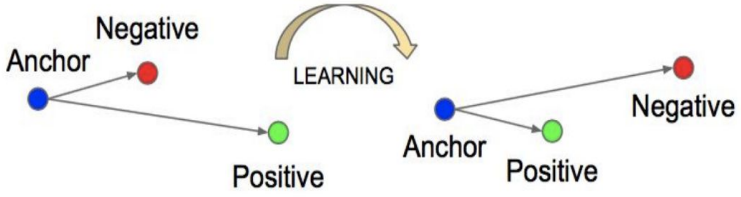
of performance and frame-rate. FaceNet uses this very concept to produce the vector based embeddings of an image (much like word2vec) and its output is then followed with computing the Euclidean distances of the images for "classification".

Learning about this process sparked a huge interest for me because I had realized that facial recognition is not explicitly a classification problem; I had come to this conclusion when I discovered that it would be an extremely difficult task to utilize images of individuals as a "final" class for a deep neural network. Rather, it would be much more simple and computationally efficient to utilize an algorithm such as k-NN, or clustering to make the final decision about an image's "classification". In reality, it is simply the computation of vectors that results in a similarity calculation between two faces.



In my research about FaceNet , I discovered how the architecture is laid out as well as the loss that is utilized. FaceNet trains on a mapping of face images to its respective Euclidean space where its distance corresponds to how similar a face is to another [2]. By using a Deep CNN to be optimized to produce accurate vectors/embeddings, an quick and efficient algorithm could be used for the classification of its output. The most essential part, training, is completed using what is known as a triplet:

1. Image of a face [Anchor]
2. Another image of the same face [Positive Exemplar]
3. Image of a different face [Negative Exemplar]



In the next few weeks, I will finalize my research about the logistics of the FaceNet architecture and begin to work on the facial segmentation/detection aspect of the project. Once this crucial

preprocessing step is complete, I will begin importing the selected model of choice and begin to test the accuracy of the model to see how it performs in real-time.

References

[1] Zhang, et al., *Deep Residual Learning for Image Recognition* (Dec 2015). Retrieved From https://arxiv.org/pdf/1512.03385.pdf

[2] Schroff, et al., *FaceNet: A Unified Embedding for Face Recognition and Clustering*. Retrieved From https://arxiv.org/pdf/1503.03832.pdf

[3] Wang, et al., *Deep Face Recognition: A Survey*. Retrieved From https://arxiv.org/pdf/1804.06655.pdf