Rahul Torlapati
DATA 606 SP'20
Deliverable II

Detecting Deep Fake Faces

Related Work and Literature Review

Previous research in detecting fake images relied heavily on handcrafted features to analyze tampered regions which were inefficient and time-consuming. Currently a large number of approaches are being designed to detect fake images. GoogLeNet (Inception v1) is a pretrained convolutional neural network that is 22 layers deep is one of such approaches. Although GoogLeNet has 22 layers it consists of 12x less parameters [1]. GoogLeNet was a significant improvement in image classification when compared to ZFNet and AlexNet which were considered previous benchmark. Another such approach is MANFA which is a hybrid framework which that uses Adaptive Boosting (Adaboost) and eXtreme Gradient Boost (XGBoost) [2]. SwapMe, an iOS application and an open source application called FaceSwap utilize this technique.

Pixel-level analysis is another such approach which tackles the problem of failing to capture high quality features, which generally lead to sub-optimal solutions. Pixel-level analysis is a pixel-level segmentation task which evaluates multiple architectures on both segmentation and classification tasks [4]. Another approach is to train GoogLeNet to detect tampering artifacts in a face classification stream and train a patch-based triplet network to leverage features capturing local noise residuals and camera characteristics as a second stream [3].

Preliminary Work

CelebA dataset has 202,599 images and Fake Faces dataset has about a Million image of which 200,000 will be used in order to maintain a balanced dataset. Due to lack of computational resources I will only be using a limited number of images for the preliminary work and will use the entire dataset for the final implementation of the project. For the preliminary phase of the project, I've used a type of CNN (Convolutional Neural Network) to detect eyes, nose, mouth and facial areas in a person's face. The dataset folder has an additional file list_landmarks_align_celeba221.csv which contains information as to the location of eyes, nose, mouth and facial areas in an image. I've compared the output generated by the CNN to the landmarks file and the results are similar.

Fig 1 – Landmarks

Fig 2 – CNN output.

Fig 3 – Landmarks

Fig 4 – CNN output

References

1. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2017) Going Deeper with Convolutions.
2. L. Minh Dang, Syed Ibrahim Hassan, Suhyeon Im and Hyeonjoon Moon (2019) Face image manipulation detection based on a convolutional neural network
3. Zhou, P., Han, X., Morariu, V.I., & Davis, L.S. (2017). Two-Stream Neural Networks for Tampered Face Detection. *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 1831-1839.
4. Jia Li, Tong Shen, Wei Zhang, Hui Ren, Dan Zeng, Tao Mei (2019) Zooming into Face Forensics: A Pixel-level Analysis.