“Web-based Gender Detection using SVM and Flask applied to DeepFake Detection”

USAMA ASHFAQ
Faculty of informtics
Otto von Guericke Magdeburg
Magdeburg, Germany
Praktikum IT Sicherheit
usama.ashfaq@st.ovgu.de

ABSTRACT

Automatic gender detection using facial images has become an interesting research area due to its vast application areas like visual surveillance, intelligent user interface, security etc. In this project, a web application was developed using Eigen images as features and Support Vector Machine as Model to classify the gender of a person as either Male or Female. The (IMDB-WIKI - 500k+ face images with age and gender labels (ethz.ch) dataset was used for training and validating the model which consist of almost 2468 Male and 3590 female images. 80 20 training validation split was used for training and validating the model. The DeepFake videos of the male and female for three different ethnic groups namely African, Asian (East) and Causian (American) were evaluated and average confidence and Standard deviation were recorded. The front end of the application was built using Flask.

INTRODUCTION

Using Support Vector Machine as a classification model has been proven in the past to produce reliable classification boundaries.

The Support Vector Machine (SVM) is basically a mathematical method that is used in the field of machine learning. It allows the classification of objects and can be used in many ways. Linear and non-linear object classification are supported. Typical areas of application include image, text or handwriting recognition”

Eigen images were used a feature for training the model. An eigen image is the name given to a set of eigenvectors when used in the computer vision problem of human face recognition. The eigen Images themselves form a basis set of all images used to construct the covariance matrix. This produces dimension reduction by allowing the smaller set of basis images to represent the original training images. Classification can be achieved by comparing how faces are represented by the basics set.

Specifically, the eigen images are the principal components of a distribution of faces, or equivalently, the eigenvectors of the covariance matrix of the set of face images, where an image with N pixels is considered a point (or vector) in N-dimensional space.

Principle Component Analyses was performed to determine the Explained variance ratio with respect to the number of the components. 50 components were considered for creating eigen images since it explains almost 80 percent of the variation as shown in the below diagram.

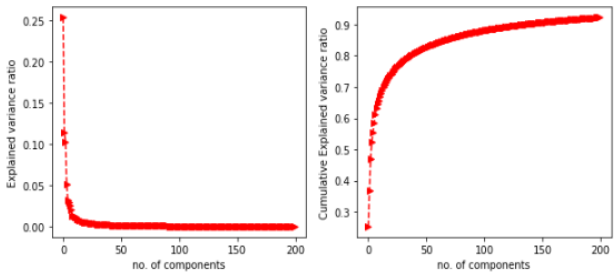

Figure 1: Principal Component Analysis

SVM model was trained using Grid Search and Hyper parameter tuning to achieve best results. 80 percent of the (IMDB-WIKI - 500k+ face images with age and gender labels (ethz.ch) dataset was used for training the model and 20 percent for validating the build SVM model.

For testing of the model, DeepFake dataset (FakeAVCeleb) was used. The three different ethnic groups i.e., Asian (East), African, Causian (American) were considered separately for both “Real Video Real Audio” and Fake Audio Fake Video” folder. 25 male and 25 female for each of the above-mentioned ethnic groups were evaluated and their confidence interval were reported.

DeepFakes are realistic - looking media content (photo, audio and video) that has been modified and falsified using artificial intelligence techniques. Although media manipulation is not a new

DeepFake dataset src (https://github.com/DASH-Lab/FakeAVCeleb)

phenomenon, deepfakes use machine learning methods, more precisely artificial neural networks, to generate fakes largely autonomously. Fake images and videos are not a new thing. For as long as photographs and film have existed, people have been fabricating forgeries designed to deceive or entertain, and this has only accelerated since the mass adoption of the internet. But now, rather than images simply being altered by editing software such as Photoshop or videos being deceptively edited, there's a new breed of machine-made fakes – and they could eventually make it impossible for us to tell fact from fiction. DeepFakes are the most prominent form of what's being called "synthetic media": images, sound and video that appear to have been created through traditional means but that have, in fact, been constructed by complex software. Deep fakes have been around for years and, even though their most common use to date has been transplanting the heads of celebrities onto the bodies of actors in pornographic videos, they have the potential to create convincing footage of any person doing anything, anywhere.

Frontend of the application was built on Flask. It is basically a web framework, it's a Python module that lets you develop web applications easily. It's having a small and easy-to-extend core: it's a microframework that doesn't include an ORM (Object Relational Manager) or such features.
It does have many cool features like url routing, template engine. It is a WSGI web app framework.

STATE OF ART

Many researchers have developed techniques for facial gender classification. Most of these techniques focus on the extraction and fusion of different types of facial features, such as:

In [1] Pritee et al. conducted a robust gender classification to occlusion by using Gabor features based. Later (2D) and PCA techniques is used to calculate features for every sub-image, one each illumination steady real Gabor space generated using both Support Vector Machine (SVM) and Gabor fitter for classification.

In [3] came up with a method which utilizes Local Block Difference Pattern (LBDP) with the assistant of Support Vector Machine (SVM) to identify the gender from the face images using FERET database. The experimental result provided to illustrate and clarify the suggested approach is an effectual method, then other similar methods.

[2] proposed a method using boosting pixel comparison of face images. They used FERET database and Support Vector Machine (SVM) for classification on images of 20_20 pixels.

Lian HC [4] obtained an accuracy of 94.81% applying local binary pattern (LBP) and SVM with polynomial kernel on the CAS-PEAL face database. According to this method, a good accuracy can be achieved if the block size for the LBP operator is correctly selected, which is really a difficult task.

Using principal component analysis (PCA), researchers in [5] processed the face image to reduce the dimensionality. After that, a good subset of eigenfeatures has been selected using genetic algorithm (GA). Here, they reported an average error rate of 11.30%. The main drawback of this method is that, the GA exhibits high computational complexity.

Deepfakes have become popular due to the quality of tampered videos and also the easy-to-use ability of their applications to a wide range of users with various computer skills from professional to novice. These applications are mostly developed based on deep learning techniques. Deep learning is well known for its capability of representing complex and high-dimensional data. One variant of the deep networks with that capability is deep autoencoders, which have been widely applied for dimensionality reduction and image compression [6]–[8]. The first attempt of deepfake creation was FakeApp, developed by a Reddit user using autoencoder-decoder pairing structure [9], [10]. In that method, the autoencoder extracts latent features of face images and the decoder is used to reconstruct the face images.

CONCEPTS

BACKEND OF APPLICATION

The training dataset looks like below with male and female images.

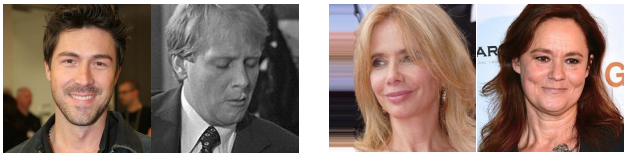

Fig3: Male and Female Training Dataset Image

The dataset initially consisted of 6058 male and female pics together but after performing data preprocessing, we have discarded those images whose size was less than 60 as shown in Figure 2, which could lead to mis-training the model and a total of 5465 images were used for training and validating the model. These images were resized to 100*100

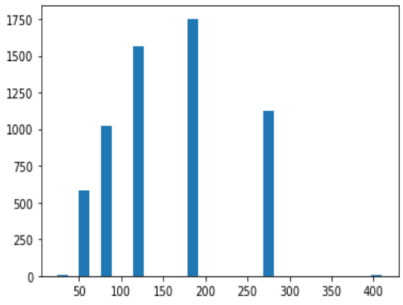

Fig2. Image size histogram

Fig 3 and Fig 4 source: IMDB-WIKI - 500k+ face images with age and gender labels (ethz.ch)
DeepFake dataset src (https://github.com/DASH-Lab/FakeAVCeleb)
Haar-cascade source: https://github.com/opencv/opencv/tree/master/data/haarcascades.
Table 1 and 2 source: https://github.com/DASH-Lab/FakeAVCeleb

After performing the Data Preprocessing the faces were detected using haar cascade. The detected faces were cropped and stored for detecting featured in the next step

Fig 4: Detected face and cropped face

The cropped images were saved with label in a data frame.
Min Max scaling was performed for normalization and missing values were removed from the training data set.

Three different features were extracted from the images i.e., PCA, SIFT, PCA-SIFT. Eigen images (PCA) is shown below.

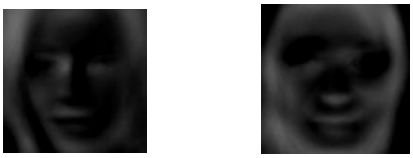

Fig 5: Eigen Images

These eigen images were feed to the Support Vector Machine model with a training validation split of 0.8 and 0.2 and the accuracy was noted.

FRONTEND OF APPLICATION

The front end of the application (GEN_REC) was build using Flask along with HTML, CSS, JAVASCRIPT and XML
The application consists of three pages or user interface.
At the start of the application, the user is directed to the Welcome page as shown below

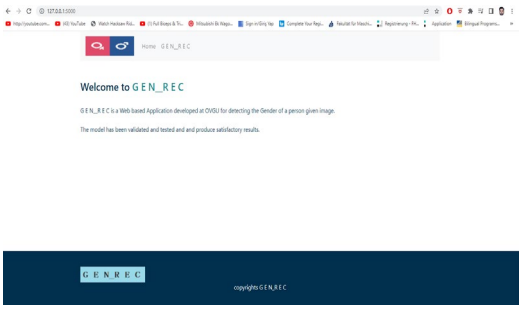

Fig 6: Welcome Page of the GEN_REC app

When the user clicks on the GEN_REC button, the user is directed to the next page where he get the concept behind the application functionality as shown in the figure below

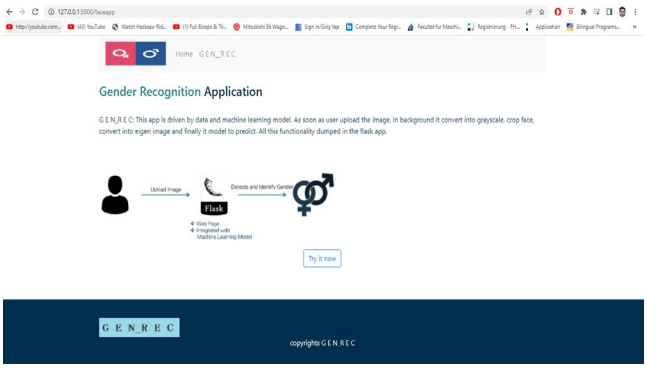

Fig 7: GEN_REC application interface

On clicking the try it now button the user is directly to the next interface where he can upload images or videos and get prediction along with the confidence score and frames as shown in the below pictures

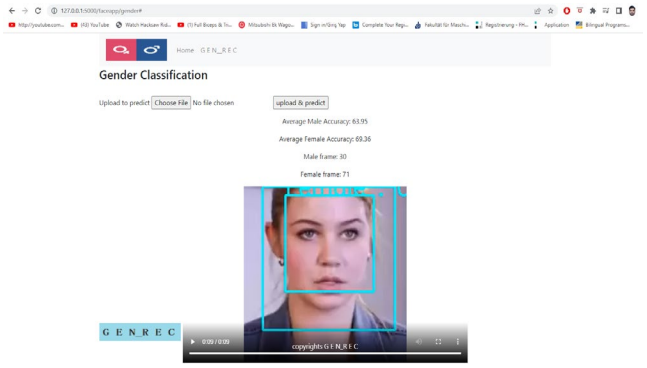

Fig3. GEN_REC app prediction page

To ensure the security and privacy of the user. The uploaded image or video get deleted automatically after the usage.

EVALUATION

Optimum values for the hyper parameters of SVM model were calculated using grid search. The following parameters were used in Grid Seacrch

{'C': [1,10,20,30,40,50,100], 'kernel' :['rbf','poly'],

'gamma': [0.1,0.05,0.01,0.001,0.002,0.005],

'coef 0' :[0,1] }

The best parameters {'C': 100, 'coef0': 0, 'gamma': 0.001, 'kernel': 'rbf'} were used for training the model. For Evaluation of the model ROC and AUC graphs have been drawn along with the confusion metrics as shown in Fig 6 and 7.

Fig 5 and Fig 6 source: IMDB-WIKI - 500k+ face images with age and gender labels (ethz.ch)

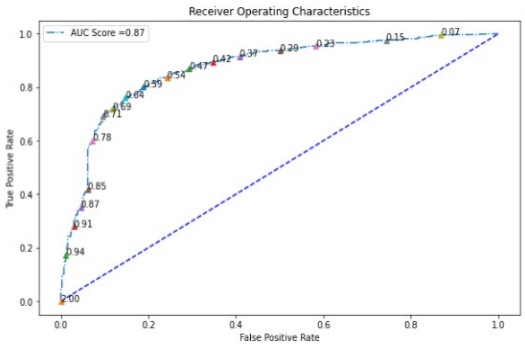

Fig 6: ROC AUC graph

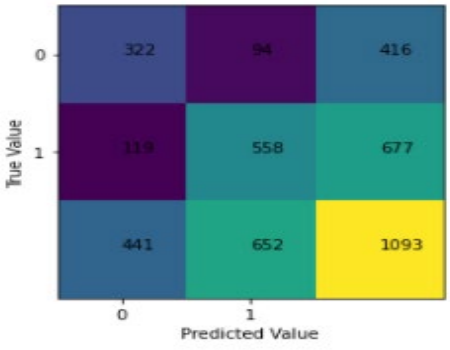

Fig 7: Confusion matrix

The testing of the model was performed using Deep Fake (FakeAVCeleb) data set as mentioned above which results in the following confidence interval table for different ethnic groups. Table 1 contains the evaluation of the Real Audio Real Video folder and Table 2 contains the evaluation of the Fake Audio Fake Video folder.

		Real Audio Real Video	
		Average	Std. Deveiation
African	Male	70.8416	7.781
	Female	76.214	10.298
Asian(East)	Male	72.56	9.985
	Female	82.42	12.947
Causian (American)	Male	75.35	7.681
	Female	78.74	9.964

Table 1: Average Confidence and Standard deviation for different ethnic groups Real Video Real Audio folder

		Fake Audio Fake Video	
		Average	Std. Deveiation
African	Male	67.51	8.75
	Female	72.93	12.89
Asian(East)	Male	68.66	8.274
	Female	81.27	12.8
Causian (American)	Male	78.5	11.04
	Female	76.58	11.165

Table 2: Average Confidence and Standard deviation for different ethnic groups Fake Video Fake Audio folder

CONCLUSION

Gender Detection using Eigen image as feature has been proven to produce satisfactory result. 85 percent accuracy on training and 80.5 percent accuracy on validation data set was recorded.

The confidence of the prediction was a little low for “Fake Video Fake Audio” than “Real Video Real Audio”. The model was able to detect the faces of the images along with their gender effectively for unseen test scenarios.

SVM model with Eigen image as features has been proven to produce satisfactory result for all considered ethnic groups

FUTURE WORK

1. Equal percentage of Male and Female Images during training the model could produce more better results.
2. During Training of the model, considering equal proportion of all the ethnic groups could lead to better results.
3. Usage of other feature like PCA-SIFT for training the model

REFERENCES

[1] Rai Preeti and Pritee Khanna, "A gender classification system robust to occlusion using Gabor features based (2D) 2 PCA", Journal of Visual Communication and Image Representation, vol. 25, no. 5, pp. 1118-1129, 2014.

[2] Shumeet Baluja and Henry A Rowley. Boosting sex identification performance. International Journal of computer vision, 71(1):111–119, 2007.

[3] Lai Chih-Chin et al., "Gender Recognition Using Local Block Difference Pattern", Advances in Intelligent Information Hiding and Multimedia Signal Processing: Pro-ceeding of the Twelfth International Conference on Intelligent Information Hiding and Multimedia Signal Processing, vol. 2, Nov., 21–23, 2016.

[4] Lian, H.-C., Lu, B.-L.: Multi-view gender classification using local binary patterns and support vector machines. In: Wang, J., Yi, Z., Zurada, J.M., Lu, B.-L., Yin, H. (eds.) ISNN 2006. LNCS, vol. 3972, pp. 202–209. Springer, Heidelberg (2006). https://doi.org/10.1007/11760023 30

[5] Sun, Z., Yuan, X., Bebis, G., Louis, S.J.: Neural-network-based gender classification using genetic search for eigen-feature selection. In: Proceedings o the 2002 International Joint Conference on Neural Networks. IJCNN’02 (Cat. No. 02CH37290), vol. 3, pp. 2433–2438. IEEE (2002)

[6] Punnappurath, A., and Brown, M. S. (2019). Learning raw image reconstruction-aware deep image compressors. IEEE Transactions on Pattern Analysis and Machine Intelligence. DOI: 10.1109/TPAMI.2019.2903062.

[7] Cheng, Z., Sun, H., Takeuchi, M., and Katto, J. (2019). Energy compaction-based image compression using convolutional autoencoder. IEEE Transactions on Multimedia. DOI: 10.1109/TMM.2019.2938345.

[8] Chorowski, J., Weiss, R. J., Bengio, S., and Oord, A. V. D. (2019). Unsupervised speech representation learning using wavenet autoencoders. IEEE/ACM Transactions on Audio, Speech, and Language Processing. 27(12), pp. 2041 2053.

[9] Faceswap: Deepfakes software for all. Available at https://github.com/deepfakes/faceswap

[10] FakeApp 2.2.0. Available at https://www.malavida.com/en/soft/fakeapp/

Fig 3 and Fig 4 source: IMDB-WIKI - 500k+ face images with age and gender labels (ethz.ch)
DeepFake dataset src (https://github.com/DASH-Lab/FakeAVCeleb)
Haar-cascade source: https://github.com/opencv/opencv/tree/master/data/haarcascades.
Table 1 and 2 source: https://github.com/DASH-Lab/FakeAVCeleb