Facial Liveness Testing: For The Web

Student Name: Ryan Collins

Supervisor Name: Prof A. Krokhin

Submitted as part of the degree of MEng Computer Science to the

Board of Examiners in the Department of Computer Sciences, Durham University

February 14, 2019

***Abstract* —**

Context/Background

TODO context

Aims

TODO

Method

TODO

Results

TODO

Conclusions

TODO

***Keywords* —** Facial liveness, convolutional neural networks, image quality metrics

I INTRODUCTION

Currently, username and password authentication is commonplace throughout the web. However, username and password based authentication systems have a number of problems. Some common passwords can be broken using dictionary attacks, especially if they consist partially or entirely of a word in a standard dictionary. Furthermore, the process of shoulder surfing is possible (watching out for someone's password, and how they type it).

An easy to use system is necessary to remove the choice from the user (in terms of password), relying on the user being automatically detected, and several confirmation methods to ensure the user is indeed who they say they are (and not just someone spoofing the system). Before such a system is developed, a facial liveness testing method must be found that operated in near real-time, and that is fairly accurate.

II RELATED WORK

III SOLUTION

A Image Quality Assessment based liveness test

For 2D spoofing attacks, spoofed images are typically lower quality than the real images, and thus by measuring the image quality one can train a classifier to detect real and spoofed images respectively.

The method used, based on the work of **(author?)** [1], implements 24 different metrics with varying differences, and produces a vector for each image. Initially, classification was done using a Support Vector Machine (SVM), but after experimentation this proved to be fairly unreliable (yielding 70% accuracy on the test set). The classifier was later changed to use Linear Discriminant Analysis (LDA) which yielded a much improved accuracy (96% accuracy on the test set).

TODO: give more accuracy figures of accuracy here, I can't remember the exact numbers

.

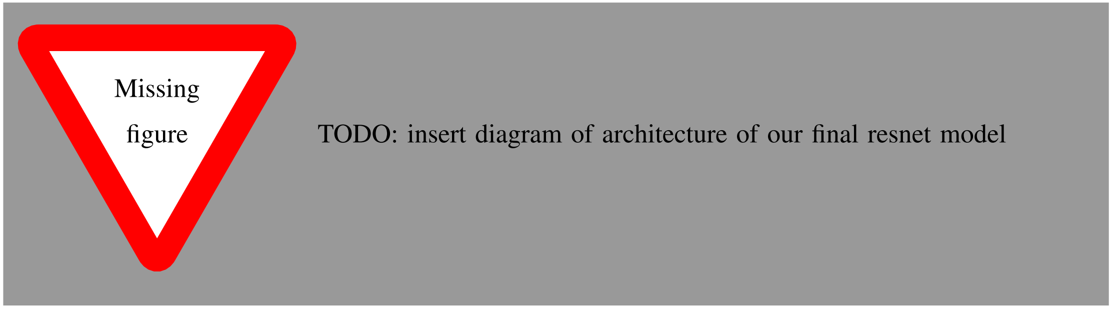
B Residual Network based 2D liveness test

Recently, 2D convolutional neural networks have had great success in image classification tasks. Therefore, it might be possible to train a residual neural network (resnet) to classify for facial liveness tasks.

In order to simplify the process of training, an existing resnet model (ResNet50) was used, with only the final convolutional layer being set to trainable. This is because the initial convolutional layers contain the standard features contained within images, while the final one learns bundles of features. Internal feed forward activations use relu, while the external output uses the softmax activation function

Training was completed using the categorical cross-entropy loss function (as this is considered multiclass). We yield a 2-tuple output from this model, which is the probability of each possible case. We take the value with the highest probability as the true outcome.

The output of this ResNet model is then fed into a 2D Max Pooling layer, which then feeds into a feed forward neural network. Initially, the model was trained using the Adam optimiser, but this yielded poor accuracy (75% accuracy). Utilising the standard gradient descent (SGD) optimiser with a low learning rate yielded far greater accuracy.



TODO: insert diagram of architecture of our final resnet model

TODO: insert citation for trying pretrained imagenet

C A system for preventing 3D spoofing attacks

While the systems before might go partially towards preventing 3D spoofing attacks, though primarily considering the 2D image, we now propose a method that is designed for classifying facial liveness based on a 3D point cloud.

C.1 Point Cloud Reconstruction

In order to classify an image/video, a 3D point cloud needs to be created, containing many 3D points (x, y, z) of a user's face. While 3d reconstruction is easier with videos (using structure from motion or other multiview based methods), there also exist image-based reconstruction methods such as vrn (**(author?)** 2) which are more specific and designed for reconstructing faces based on images.

C.2 3D point cloud classification

Once the 3D reconstruction is obtained, one can then classify this using some model to produce the fake/real metric.

For points, PointNet is a model that can be used

EXPLAIN POINTNET

However, as we are using voxels, there is a more specialised architecutre called Voxnet that is designed for classifying 3d volume-based objects. VoxNet takes in a point cloud and converts this to an occupancy grid. This is then fed through two convolutional layers, pooled, and then goes through a dense layer before reaching the classifier output (a dense layer with the k outcomes).

Instead of training on our existing datasets, we first train a VoxNet model on the SUOD dataset, in order to learn the basic features surrounding 3D classification. This largely uses the tools provided in the original implementation of VoxNet, but with a custom Keras implementation. This pretrained model is then saved, for use in our final implementation.

Using this original VoxNet implementation, the last dense layer is then ignored, with it being extended to contain further dense layers, and an eventual output classifier (for fake/real respectively).

C.3 How this overall system functions

Each image is first preprocessed: the image is fed through the 3D reconstruction network, and the output is a set of Voxels. Using these voxels, they are then resized into a (24 x 24 x 24) shape to provide a basis for the network. From here, they are fed through several convolutional layers, before then being fed through several dense layers for classification.

Missing figure

INSERT FIGURE HERE CONTAINING PIPELINE

D Visualisation and Demonstration

In order to visualise the overall outcome of facial liveness, a generic model

IV RESULTS

TODO results

V EVALUATION

TODO evaluation

VI CONCLUSIONS

References

[1] J. Galbally, S. Marcel, and J. Fierrez. Image quality assessment for fake biometric detection: Application to iris, fingerprint, and face recognition. *IEEE Transactions on Image Processing*, 23(2):710–724, Feb 2014.

[2] Aaron S Jackson, Adrian Bulat, Vasileios Argyriou, and Georgios Tzimiropoulos. Large pose 3d face reconstruction from a single image via direct volumetric cnn regression. *International Conference on Computer Vision*, 2017.