Face Recognition System Development

Thomas Wang Wei Hong (A0111393N)

Abstract

Traditional username and password – based web portal login systems have long been plagued by many problems, such as hijacking and stealing of sensitive personal information by hackers etc. These systems needed to be augmented by additional security measures to guard against unauthorized access. A face recognition system would serve as a good two – factor authentication measure to protect users' sensitive personal information, given the technology's maturity and reliability. However, a subscription to commercial providers of face recognition systems is prohibitively expensive, especially for start – ups which are constantly facing cash flow issues. As such, this project is born, with the objective of providing an in – house face recognition solutions at much lower costs than commercial providers. However, even face recognition systems have also been plagued by problems, mainly the use of a physical medium by malicious personnel (such as photo cut – outs etc.) to trick the system into granting access to the victim's personal information, which is widely known as face spoofing. In view of this, an in – house face spoof detection system will also be developed alongside the face recognition system to secure the face recognition system and maintain its reliability amidst the increased frequency and threat of cyber – attacks. This report would provide an overview of the entire development process, from conceptualization to creation of proof – of – concept and performance analysis of the systems to assess their readiness for deployment.

Introduction

Age – related Macular Degeneration (AMD) is an eye disease that affects the macula, which is the central part of the retina responsible for detailed central vision, and that AMD usually affects people over the age of 50 [1]. Based on statistics published by the World Health Organization (WHO), AMD is the leading cause of vision loss and blindness in people over the age of 60, and it is currently estimated that over 200 million people worldwide already have AMD or are at risk of developing it [2, 3]. Although there is no treatment regime that can reverse the effects of AMD, its early detection and intervention can help prevent the disease from getting worse, thus helping to prolong a patient's vision. At the time of writing of this report, there is currently no home – based vision monitoring method for the early detection of AMD, which then translates to many issues for patients, such as having to deal with frequent and burdensome clinic visits, unpredictable disease recurrence etc.

The goal of Occutrack Medical Solutions Pte Ltd (hereinafter referred to as "Occutrack") is to tap on this market and address the above issues, by introducing a web – based gaze tracking system that displays specialized moving patterns on a computer screen and with the help of a specialized webcam, analyze the user's eye's reactions to these patterns to assess the severity of AMD impairment for the user. The most important aspect of this system is that it is web – based and can be used on any computers, thus allowing users to take the tests from the comfort of their home (especially for elderlies who will make up the bulk of Occutrack's users), which would greatly facilitate the early detection of AMD and allow timely intervention measures to be taken, thus preserving the user's sights, thus aligning with Occutrack's goals.

In order to access Occutrack's services, the user would have to first sign up for an account with Occutrack. It is known that the traditional process of user authentication via the use of passwords is riddled with many issues, such as the need for complex passwords which are difficult to remember, vulnerability of elderly to malicious phishing attacks and resistance from users caused by the need to constantly reset passwords etc. This leads to a need for a two – factor authentication method (2FA) as an additional safeguard for user authentication.

Face recognition has long been used as a user authentication method in many industries due to its high level of maturity [4]. As such, Occutrack has decided to adopt face recognition as a form of 2FA on top of password protection for its user authentication system, which is the primary objective of this project.

However, it is also known that malicious attackers can use fake faces to trick the face recognition system into granting unauthorized access to users' personal information, which is known in layman terms as face spoofing. In fact, the authors of [4] have talked about how the advent of big data has made it easier for malicious attackers to perform face spoofing attacks on face recognition system. In light of this, it is also necessary to develop a Presentation Attack Detection (PAD) system to augment the face recognition system to ensure its reliability against face spoofing attacks, and this would be the secondary objective of this project.

Data

The development of both face recognition and PAD systems cannot be possible without the utilization of data for model training and validation. In fact, two separate datasets will be required to train the models, one for the face recognition system and the other for the PAD system. Since Occutrack is a Med – Tech start – up which has no data of its own, it must depend on external data sources for the time being in order to build proof – of – concept face recognition and PAD models.

Face Recognition Dataset

A quick search of the literature has revealed that there exists several large – scale face recognition datasets that have been used to train and assess face recognition models. A few examples of them would be the CASIA – WebFace dataset [5], the MegaFace dataset [6] etc. However, most of these datasets are currently shelved by the authors and no longer released for download and use. After some extensive search, a new dataset called the Asian Face Dataset (hereinafter referred to as AFD) [7] has been found and deemed suitable for use. The AFD consists of 360,000 images for 2019 unique Asian identities under various poses and illumination, of which a small sample provided by the authors is shown in Figure 1 below.

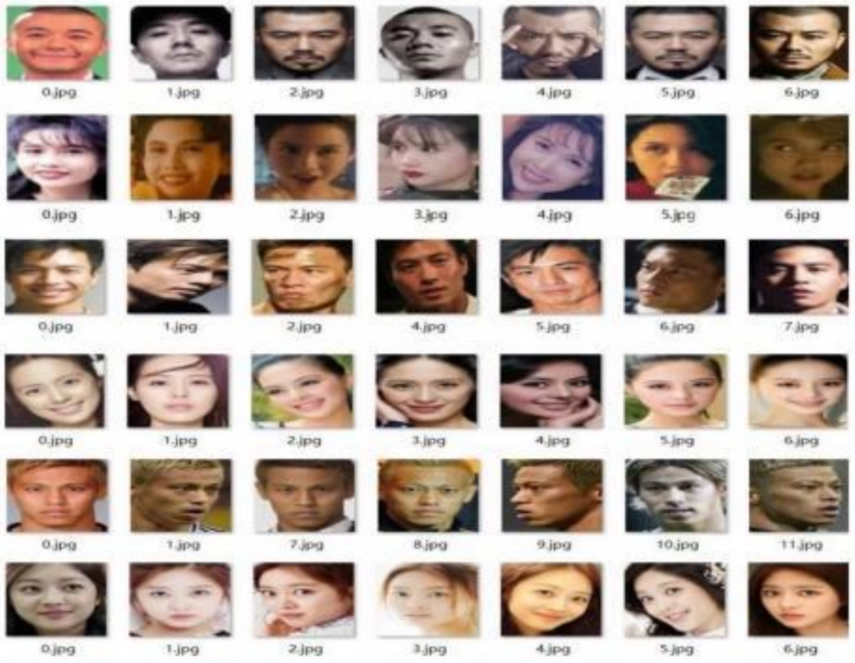


Figure 1: Sample of Faces Present in the AFD

The authors have released this dataset for public use in a bid to aid the general community (academic and non – academic) in creating and improving face recognition models' performance, and it is available on GitHub for download[1].

Although the AFD is ethnically biased towards Asians (Chinese in particular), it is still appropriate for Occutrack's use case, since the main target users would be Singaporeans and Chinese nationals, who are predominantly Asians. The authors of the AFD did not provide a train – test split, thus Occutrack performed the split on its own in a random manner using a 4 : 1 ratio for each unique identity in order to create a training and testing set for model performance evaluation.

Face Spoof Dataset

After some extensive search of the literature, a large face spoof detection dataset, the CelebA – Spoof dataset [8], has been deemed suitable, and is currently available for download[2]. This dataset consists of 625,537 images from 10,177 subjects, while consisting of many spoof types, such as print, paper cut, replay and 3D attacks. Figure 2 contains an example of the spoofing types contained in the dataset.

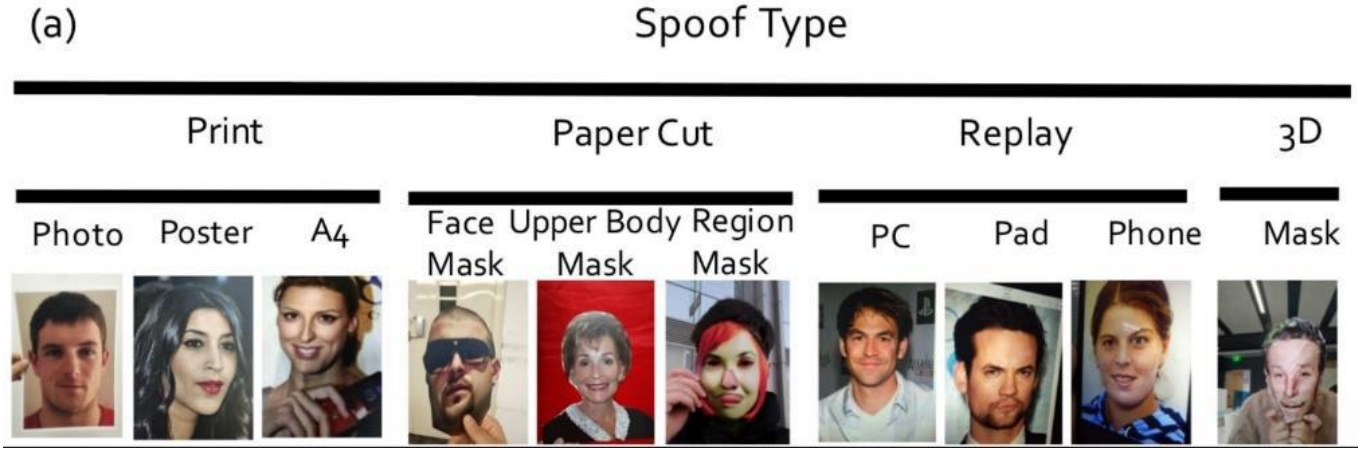


Figure 2: Spoof Types in the CelebA – Spoof Dataset

The original CelebA – Spoof dataset is trimmed for practical purposes, such that it excludes paper cut and 3D mask attacks, which is deemed highly unlikely to be encountered for the web – based gaze – tracking application in the real – world settings. The authors of the dataset have provided a training and a test set, of which Occutrack has further split the training set into a training – validation set (in a 4 : 1 manner randomly) in order to assess any trained face spoof detection model performance before proceeding to obtain the final test performance on the test set.

Methods and Algorithms

Face Recognition Component

The entire face recognition pipeline can be summarized into four stages, as outlined in [9]. These four stages are face detection, face alignment, face representation extraction and representation similarity comparison to obtain the identity of any given face, in the respective order. Figure 3 below shows a simple outline of the face recognition pipeline in graphical form.

[1] https://github.com/X-zhangyang/Asian-Face-Image-Dataset-AFD-dataset
[2] https://github.com/ZhangYuanhan-AI/CelebA-Spoof

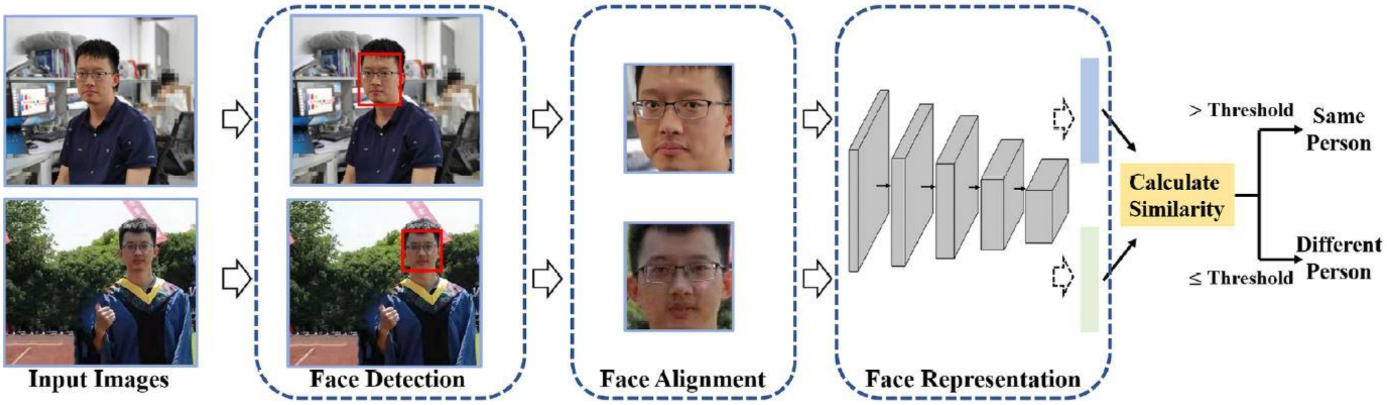


Figure 3: Simple Illustration of Face Recognition Pipeline (Figure Extracted From [9])

In this project, the main objective is to train a face recognition model that can perform the face representation extraction component well, since the face detection and face alignment can be performed by open – source packages reliably, such as the Python MediaPipe package[3].

Since a proof – of – concept face recognition pipeline needs to be developed in a very short amount of time, it is decided that well – known Convolutional Neural Networks (CNNs) that have been pre – trained on the ImageNet dataset [10] should be fine – tuned with the AFD to obtain a CNN that can then be used for the facial representation extraction task, with this approach validated in [11]. The pre-trained CNNs are extracted up to the Global Average Pooling layer (or a similar layer), after which the final 1000 – node Fully Connected (FC) layer is replaced with a 2019 – node FC layer to train the CNN models using the classification paradigm, after which the 2019 – node FC layer would then be removed to obtain the CNN as facial representation extractor.

The CNN architectures which have been tested for the face recognition component are DenseNet201 [12], EfficientNetV2 [13], InceptionV3 [14], InceptionResNet-V2 [15], MobileNetV2 [16], ResNet152 [17], ResNet152V2 [18], ResNetRS152 [19], RegNetX064 [20], RegNetY064 [20] and Xception [21].

The models are fine – tuned using the original size of the images in the AFD, which is 160 x 160 pixels, of which data normalization is performed to scale the pixel values to the range of 0 to 1. Data augmentation is also performed for the training data, whereby the following transformations would be performed at random: horizontal flip, rotation of up to 20 degrees in either clockwise or anti-clockwise direction, translation in the horizontal and vertical directions of up to 20% of image width / height, shearing of up to 20 degrees, and zooming in of the image of up to 20% of image width / height. The only operation performed on the testing data is data normalization of the pixel values. The batch size used varies according to the amount of computational resources available after a particular model is used. The optimizer used is the Adam optimizer [22] with an initial learning rate of 0.001 which is then exponentially decayed after the first 10 epochs of training. Early stopping is implemented with a patience of 20 epochs, with the option to restore the best weight for the validation dataset enabled. Reduction of learning rate on loss value plateau is also implemented, with a reduction of 25% and patience of 10 epochs together with cooldown of 5 epochs and lowest learning rate being 1e-07. There are many loss functions introduced in [9] which could be used for model training under the classification paradigm, such as CosFace [23] and ArcFace [24]. However, these loss functions required some form of hyper-parameter tuning, which given the time constraint and multiple objectives of this project, is not feasible. In the end, the ubiquitous categorical cross-entropy loss function is used as it is parameter-free and still yields decent performance for the classification task.

[3] https://developers.google.com/mediapipe/solutions/examples

PAD Component

Similar to the face recognition component, well – known Convolutional Neural Networks (CNNs) that have been pre – trained on the ImageNet dataset [10] should be fine – tuned with the CelebA – Spoof dataset to obtain a CNN that can be used for PAD, except that a 2 – node FC layer is used instead of a 2019 – node FC layer, and that this 2 – node FC layer would be retained after model training instead of being discarded.

The CNN architectures which have been tested for the PAD component are DenseNet201 [12], InceptionResNet-V2 [15], MobileNetV2 [16] and RegNetY064 [20], which are fewer than that for the face recognition component due to time constraints.

The training methodology is exactly the same as that for the face recognition component, so the exact details will not be repeated here.

Performance Evaluation

Face Recognition Component

The metric used to assess the model performance would be the categorical cross – entropy loss and the accuracy of the model. The accuracy metric is used here because the number of images for each identity in the AFD is roughly similar, thus class imbalance should not be as issue in this case and hence the accuracy metrics would be reliable. The performance of each tested model for each metric is detailed in Table 1 below, with the best performing metric highlighted in bold.

Model	Categorical Cross – Entropy Loss	Accuracy (%)
DenseNet201	0.8297	85.37
EfficientNetV2	2.0847	59.16
InceptionV3	1.2026	82.49
InceptioResNetV2	0.8731	83.95
MobileNetV2	1.5070	78.37
ResNet152	1.2074	79.06
ResNet152V2	1.0381	80.65
ResNetRS152	**0.7901**	**87.39**
RegNetX064	0.8053	86.37
RegNetY064	0.7961	85.97
Xception	1.0481	81.89

Table 1: Face Recognition Model Performance for Each Metric (Best Performing Metric Highlighted in **Bold**)

Although the ResNetRS152 model yields the best performance, its learning process is quite unstable (which applies to most of the other models as well), as can be seen from the plot in Figure 4 below. On the other hand, even though the RegNetY064 model is not as accurate as the ResNetRS152 model, it has the second lowest categorical cross – entropy loss value, and its learning process is very stable for the test set, as can be seen from the plot in Figure 4 below. The features learnt by the RegNetY064 model would be stable and more suitable for face recognition purposes than that of the ResNetRS152 model, and as such, the RegNetY064 model is adapted for use in the proof – of – concept face recognition pipeline.

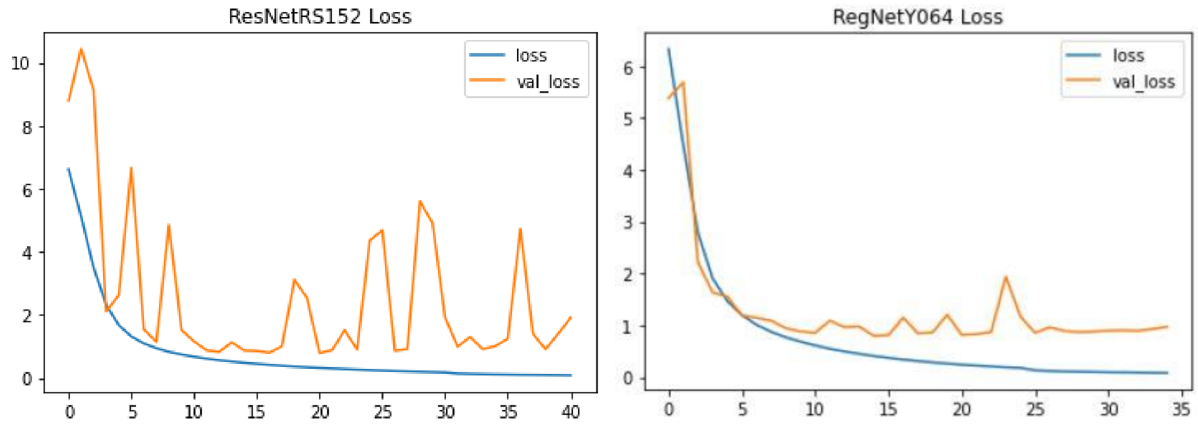


Figure 4: Model Learning Process for the ResNetRS152 Model (Left) and the RegNetY064 Model (Right)

Occutrack has recently developed a proof – of – concept web portal which supports the traditional password login with face recognition component as a 2FA method, which is then used to test the performance of the trained face recognition model. The web portal is illustrated in Figure 5 below.

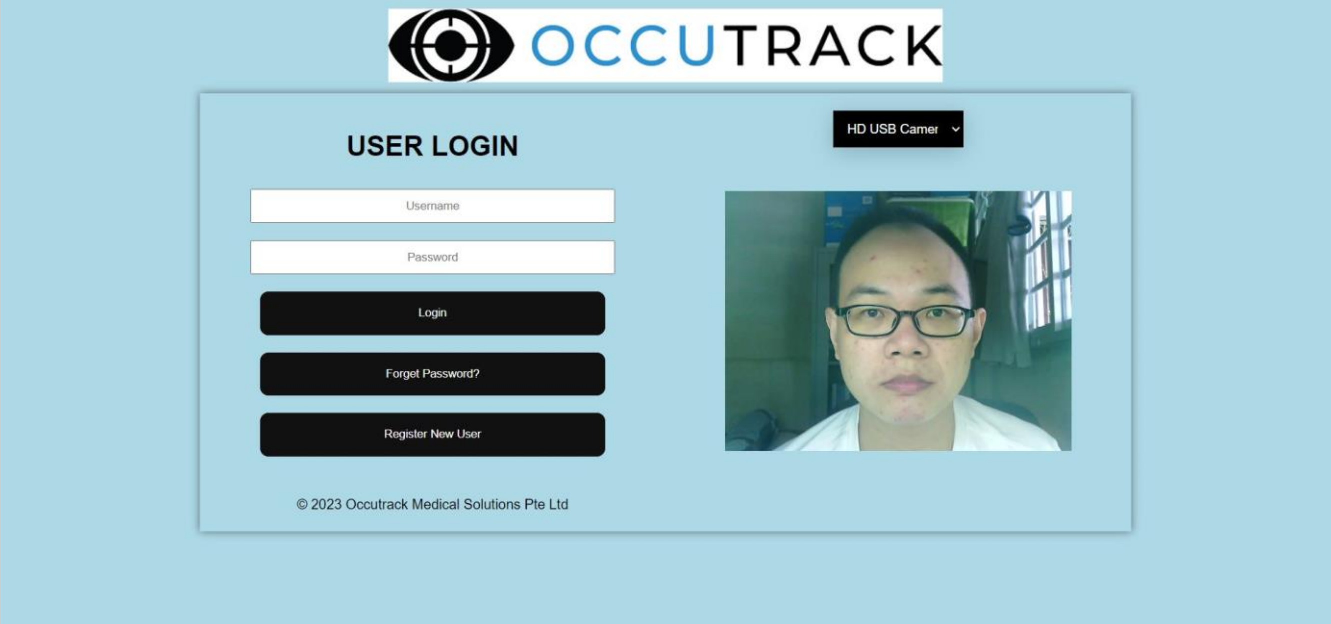
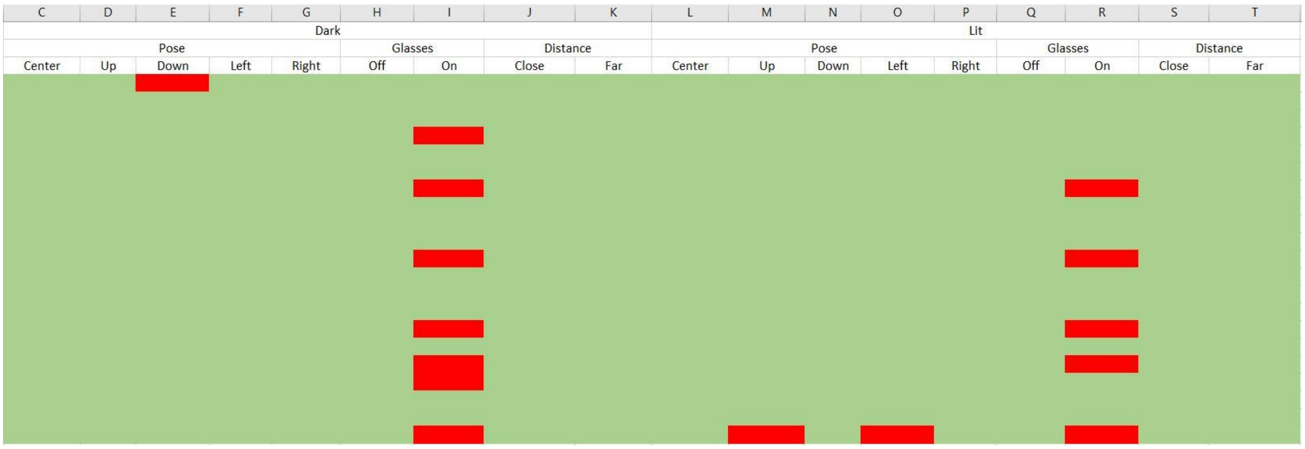


Figure 5: Web Portal to Test Face Recognition Model Performance

The face recognition scheme under this arrangement is known as the verification scheme, since the identity of a face is now provided via the username, of which the function of the face recognition model is to verify that the person shown on the webcam feed is the same one as that linked to the username. More details on the verification scheme can be found in [9].

The face recognition model is tested using Occutrack's specialized gaze – tracking webcam, together with the support from the organic staff of Trendlines Medical Singapore (hereinafter referred to as "TMS"), the incubator responsible for incubating Occutrack since the early stage, and TMS' other portfolio companies, totalling approximately 20+ people. The verification scheme is tested under different conditions, such as pose, lighting etc., with the data collected in the form of an excel sheet table as shown in Figure 6 below. The red cells indicate cases where the face recognition model fails to recognize the user when he / she tried to login to his / her account, while the green cells indicate cases where the face recognition model works as intended.



C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
Dark									Lit								
Pose					Glasses		Distance		Pose					Glasses		Distance	
Center	Up	Down	Left	Right	Off	On	Close	Far	Center	Up	Down	Left	Right	Off	On	Close	Far
🟩	🟩	🟥	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
🟩	🟩	🟩	🟩	🟩	🟩	🟥	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
🟩	🟩	🟩	🟩	🟩	🟩	🟥	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟥	🟩	🟩
🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
🟩	🟩	🟩	🟩	🟩	🟩	🟥	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟥	🟩	🟩
🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
🟩	🟩	🟩	🟩	🟩	🟩	🟥	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟥	🟩	🟩
🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
🟩	🟩	🟩	🟩	🟩	🟩	🟥	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟥	🟩	🟩
🟩	🟩	🟩	🟩	🟩	🟩	🟥	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
🟩	🟩	🟩	🟩	🟩	🟩	🟥	🟩	🟩	🟩	🟥	🟩	🟥	🟩	🟩	🟥	🟩	🟩

Figure 6: Excel Table for Test Data Collected from Face Recognition Model Assessment

It is found that each user can be verified with ease under most conditions encountered in the real – world settings. However, it is noted that the face recognition model struggles to identify the individuals when they registered with the system without the use of glasses, but then subsequently tried the system while wearing glasses.

This can be attributed to the fact that most, if not all, of the individuals contained in the AFD do not wear glasses, thus the face recognition model does not get the chance to learn the features that allowed it to be invariant to glasses.

PAD Component

It is known that there are specialized metrics which are used to assess the different aspects of the performance of PAD models. After a quick literature search, the metrics used in the paper "LBP and CNN Feature Fusion for Face Anti-Spoofing" has been adopted [25]. The four metrics used are Attack Presentation Classification Error Rate (APCER), Bonafide Presentation Classification Error Rate (BPCER), Average Classification Error Rate (ACER), and Accuracy, with equations (1) – (4) showing how these metrics are calculated.

$$APCER = \frac{FP}{FP + TN} \tag{1}$$

$$BPCER = \frac{FN}{FN + TP} \tag{2}$$

$$ACER = \frac{APCER + BPCER}{2} \tag{3}$$

$$Accuracy = 100 \times (1 - \frac{APCER \times \#Spoof + BPCER \times \#Real}{\#Spoof + \#Real}) \tag{4}$$

where FP represents False Positives, TP represents True Positives, FN represents False Negatives and TN represents True Negatives, # Spoof represents number of spoof images in the test set and # Real represents number of real images in the test set, with spoof images as the negative class and real images as the positive class.

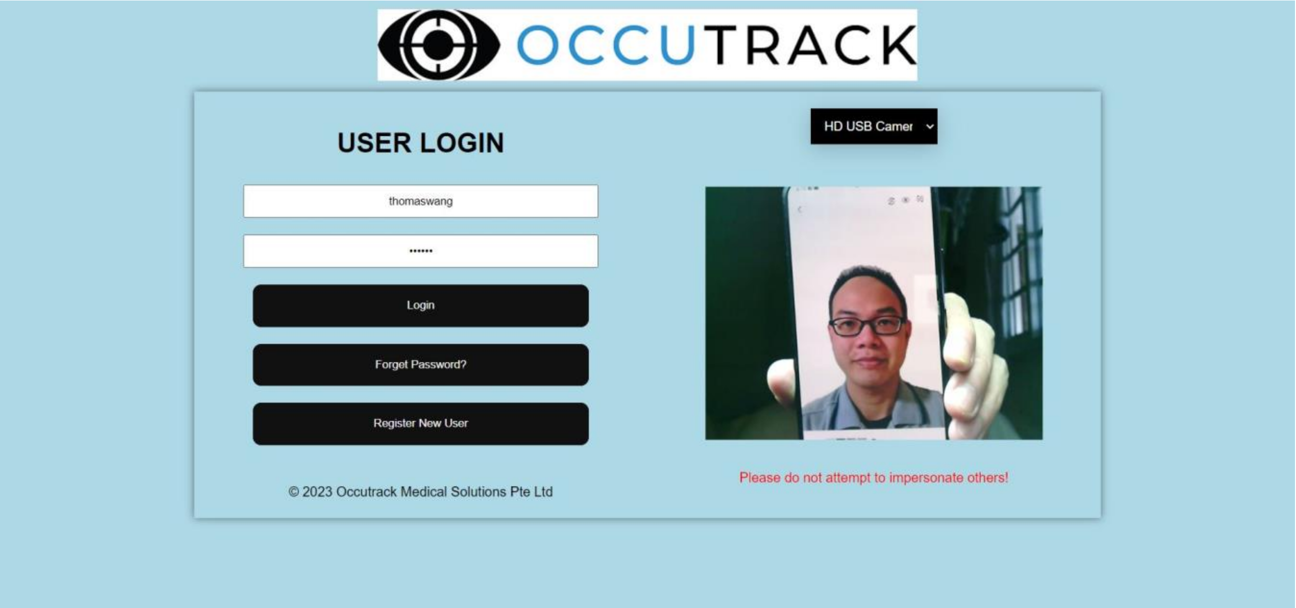
The performance of each tested model for each metric is detailed in Table 2 below, with the best performing metric highlighted in bold.

Metrics \ Model	DenseNet201	InceptionResNetV2	MobileNetV2	RegNetY064
APCER	**0.00437**	0.00794	0.00880	0.00554
BPCER	0.37880	0.35563	**0.23322**	0.37149
ACER	0.19159	0.18179	**0.12101**	0.18852
Accuracy	0.85972	0.86608	**0.90998**	0.86201

Table 2: PAD Model Performance for Each Metric (Best Performing Metric Highlighted in **Bold**)

Although the performance of the MobileNetV2 is the best in terms of most of the metrics, it still lacks significantly behind the DenseNet201 in terms of detecting spoofed faces, which is considered more important than the other metrics. As such, DenseNet201 is still chosen as the PAD model.

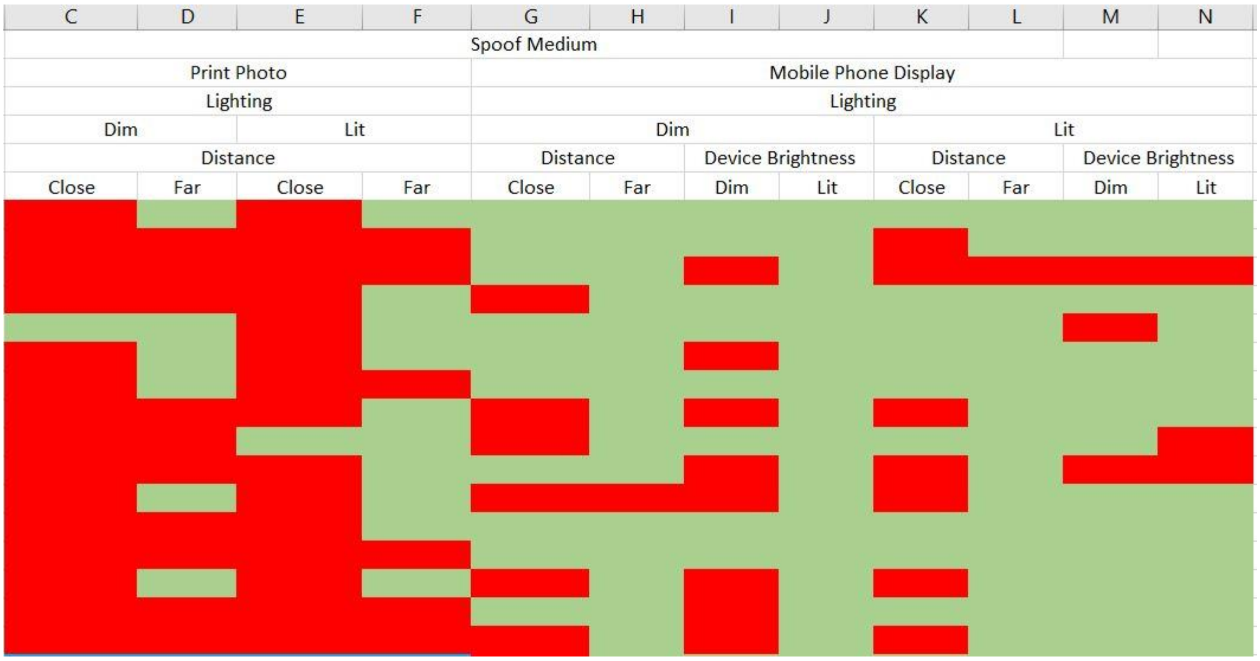
The same proof – of – concept web portal is also used to test the performance of the PAD, of which an alert would be triggered (indicated as red text) when it detected a spoofed face, as shown in Figure 7 below.



Please do not attempt to impersonate others!

Figure 7: Demonstration of PAD model in Action for the Web Portal

The PAD model is tested using the same webcam and personnel from the face recognition model component, with the PAD model tested under various conditions such as lighting and medium etc., and the data collected in the form of an excel sheet table as shown in Figure 8 below. The red cells indicate cases where the PAD model fails to recognize that the presented medium is a spoof medium, and the green cells indicate cases where the PAD model works as intended.



C	D	E	F	G	H	I	J	K	L	M	N
Spoof Medium											
Print Photo				Mobile Phone Display							
Lighting				Lighting							
Dim		Lit		Dim				Lit			
Distance				Distance		Device Brightness		Distance		Device Brightness	
Close	Far	Close	Far	Close	Far	Dim	Lit	Close	Far	Dim	Lit
🟥	🟩	🟥	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
🟥	🟥	🟥	🟥	🟩	🟩	🟩	🟩	🟥	🟩	🟩	🟩
🟥	🟥	🟥	🟥	🟩	🟩	🟥	🟩	🟥	🟥	🟥	🟥
🟥	🟥	🟥	🟩	🟥	🟩	🟩	🟩	🟩	🟩	🟩	🟩
🟩	🟩	🟥	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟥	🟩
🟥	🟩	🟥	🟩	🟩	🟩	🟥	🟩	🟩	🟩	🟩	🟩
🟥	🟩	🟥	🟥	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
🟥	🟥	🟥	🟩	🟥	🟩	🟥	🟩	🟥	🟩	🟩	🟩
🟥	🟥	🟩	🟩	🟥	🟩	🟩	🟩	🟩	🟩	🟩	🟥
🟥	🟥	🟥	🟩	🟩	🟩	🟥	🟩	🟥	🟩	🟥	🟥
🟥	🟩	🟥	🟩	🟥	🟥	🟥	🟩	🟥	🟩	🟩	🟩
🟥	🟥	🟥	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
🟥	🟥	🟥	🟥	🟩	🟩	🟩	🟩	🟩	🟩	🟩	🟩
🟥	🟩	🟥	🟩	🟥	🟩	🟥	🟩	🟥	🟩	🟩	🟩
🟥	🟥	🟥	🟥	🟩	🟩	🟥	🟩	🟩	🟩	🟩	🟩
🟥	🟥	🟥	🟥	🟥	🟩	🟥	🟩	🟥	🟩	🟩	🟩

Figure 8: Excel Table for Test Data Collected from PAD Model Assessment

It is apparent that the PAD model is not performing as well as the face recognition model, especially for the detection of spoof print medium. This phenomenon is understandable, as it is visually more difficult to perceive whether an image is spoofed or real just by looking only at the face region in an image than to attempt to tell different faces apart.

This can be attributed to the fact that the PAD model is trained using the CelebA – Spoof dataset which is captured from imaging sensors which are very different from Occutrack's specialized webcam, thus resulting in a very huge performance gap when applied in the real world setting for Occutrack's use case.

Business Insights and Discussion

It is a widely known fact that commercial face recognition and PAD systems are very costly owing to the amount of technology, labour and resources used to develop them. In addition, these systems can only function through the use of Application Programming Interfaces (APIs), which might not be compatible with the web application that Occutrack would be using.

In developing its own face recognition and PAD systems, Occutrack can potentially save significant costs on its web application by not having to acquire these systems from a third party, which in turns translates to greater take – up rate for the general society since Occutrack can now lower the price and make its services much more affordable for its customers, hence allowing more people to have the chance to monitor their eye conditions from the comfort of their home and achieve Occutrack's goal of preserving the sight of its customers.

Although the performance of the in – house developed face recognition and PAD systems is still significantly inferior by commercial standards, it is believed that future Research and Development (R&D) efforts should be able to bring these systems up to standards and achieve its goal of independence from third – party software.

Conclusion

This project is initiated with the goal of developing Occutrack's in – house face recognition and PAD system to avoid incurring costs associated with third – party software. Although the performance of the developed systems still far short of expectations, it is still considered a reasonably successful project considering the amount of resources that Occutrack has as a Med – Tech start – up to develop these systems on top of its other objectives. It is believed that future R&D efforts would be able to bring the performance of these systems up to standards, thus realizing Occutrack’s goal of providing affordable and reliable eye monitoring services to preserve the sight of its customers.

References

[1] Grassmann F, Mengelkamp J, Brandl C, Harsch S, Zimmermann ME, Linkohr B, Peters A, Heid IM, Palm C, Weber BH. A deep learning algorithm for prediction of age-related eye disease study severity scale for age-related macular degeneration from color fundus photography. Ophthalmology. 2018 Sep 1;125(9):1410-20.

[2] Glatz M, Riedl R, Glatz W, Schneider M, Wedrich A, Bolz M, Strauss RW. Blindness and visual impairment in Central Europe. PLoS One. 2022 Jan 13;17(1):e0261897.

[3] Bourne R, Steinmetz JD, Flaxman S, Briant PS, Taylor HR, Resnikoff S, Casson RJ, Abdoli A, Abu-Gharbieh E, Afshin A, Ahmadieh H. Trends in prevalence of blindness and distance and near vision impairment over 30 years: an analysis for the Global Burden of Disease Study. The Lancet global health. 2021 Feb 1;9(2):e130-43.

[4] Chang HH, Yeh CH. Face anti-spoofing detection based on multi-scale image quality assessment. Image and Vision Computing. 2022 May 1;121:104428.

[5] Yi D, Lei Z, Liao S, Li SZ. Learning face representation from scratch. arXiv preprint arXiv:1411.7923. 2014 Nov 28.

[6] Kemelmacher-Shlizerman I, Seitz SM, Miller D, Brossard E. The megaface benchmark: 1 million faces for recognition at scale. In Proceedings of the IEEE conference on computer vision and pattern recognition 2016 (pp. 4873-4882).

[7] Xiong Z, Wang Z, Du C, Zhu R, Xiao J, Lu T. An asian face dataset and how race influences face recognition. InAdvances in Multimedia Information Processing–PCM 2018: 19th Pacific-Rim Conference on Multimedia, Hefei, China, September 21-22, 2018, Proceedings, Part II 19 2018 (pp. 372-383). Springer International Publishing.

[8] Zhang Y, Yin Z, Li Y, Yin G, Yan J, Shao J, Liu Z. Celeba-spoof: Large-scale face anti-spoofing dataset with rich annotations. In Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part XII 16 2020 (pp. 70-85). Springer International Publishing.

[9] Du H, Shi H, Zeng D, Zhang XP, Mei T. The elements of end-to-end deep face recognition: A survey of recent advances. ACM Computing Surveys (CSUR). 2022 Sep 14;54(10s):1-42.

[10] Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L. Imagenet: A large-scale hierarchical image database. In 2009 IEEE conference on computer vision and pattern recognition 2009 Jun 20 (pp. 248-255). IEEE.

[11] Gwyn T, Roy K, Atay M. Face recognition using popular deep net architectures: A brief comparative study. Future Internet. 2021 Jun 25;13(7):164.

[12] Huang G, Liu Z, Van Der Maaten L, Weinberger KQ. Densely connected convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition 2017 (pp. 4700-4708).

[13] Tan M, Le Q. Efficientnetv2: Smaller models and faster training. In International conference on machine learning 2021 Jul 1 (pp. 10096-10106). PMLR.

[14] Szegedy C, Vanhoucke V, Ioffe S, Shlens J, Wojna Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE conference on computer vision and pattern recognition 2016 (pp. 2818-2826).

[15] Szegedy C, Ioffe S, Vanhoucke V, Alemi A. Inception-v4, inception-resnet and the impact of residual connections on learning. In Proceedings of the AAAI conference on artificial intelligence 2017 Feb 12 (Vol. 31, No. 1).

[16] Sandler M, Howard A, Zhu M, Zhmoginov A, Chen LC. Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE conference on computer vision and pattern recognition 2018 (pp. 4510-4520).

[17] He K, Zhang X, Ren S, Sun J. Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition 2016 (pp. 770-778).

[18] He K, Zhang X, Ren S, Sun J. Identity mappings in deep residual networks. In Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part IV 14 2016 (pp. 630-645). Springer International Publishing.

[19] Bello I, Fedus W, Du X, Cubuk ED, Srinivas A, Lin TY, Shlens J, Zoph B. Revisiting resnets: Improved training and scaling strategies. Advances in Neural Information Processing Systems. 2021 Dec 6;34:22614-27.

[20] Radosavovic I, Kosaraju RP, Girshick R, He K, Dollár P. Designing network design spaces. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition 2020 (pp. 10428-10436).

[21] Chollet F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE conference on computer vision and pattern recognition 2017 (pp. 1251-1258).

[22] Kingma DP, Ba J. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980. 2014 Dec 22.

[23] Wang H, Wang Y, Zhou Z, Ji X, Gong D, Zhou J, Li Z, Liu W. Cosface: Large margin cosine loss for deep face recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition 2018 (pp. 5265-5274).

[24] Deng J, Guo J, Xue N, Zafeiriou S. Arcface: Additive angular margin loss for deep face recognition. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition 2019 (pp. 4690-4699).

[25] Singh RP, Dash R, Mohapatra RK. LBP and CNN feature fusion for face anti-spoofing. Pattern Analysis and Applications. 2023 May;26(2):773-82.