Pattern Recognition 112 (2021) 107696

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/patcog

Client-specific anomaly detection for face presentation attack detection☆

Soroush Fatemifar[a,*], Shervin Rahimzadeh Arashloo[b], Muhammad Awais[a], Josef Kittler[a]

[a] Centre for Vision, Speech and Signal Processing, University of Surrey, Guildford, UK
[b] Department of Computer Engineering, Bilkent University, Ankara, Turkey

ARTICLE INFO

Article history:
Received 13 December 2019
Revised 23 September 2020
Accepted 7 October 2020
Available online 26 October 2020

Keywords:
Anomaly detection
Biometrics
Client-specific information
Deep convolutional neural networks
Face spoofing detection

ABSTRACT

One-class anomaly detection approaches are particularly appealing for use in face presentation attack detection (PAD), especially in an *unseen* attack scenario, where the system is exposed to novel types of attacks. This work builds upon an anomaly-based formulation of the problem and analyses the merits of deploying *client-specific* information for face spoofing detection. We propose training one-class client-specific classifiers (both generative and discriminative) using representations obtained from pre-trained deep Convolutional Neural Networks (CNN). In order to incorporate *client-specific* information, a distinct threshold is set for each client based on subject-specific score distributions, which is then used for decision making at the test time. Through extensive experiments using different one-class systems, it is shown that the use of client-specific information in a one-class anomaly detection formulation (both in model construction as well as decision boundary selection) improves the performance significantly. We also show that anomaly-based solutions have the capacity to perform as well or better than two-class approaches in the unseen attack scenarios. Moreover, it is shown that CNN features obtained from models trained for face recognition appear to discard discriminative traits for spoofing detection and are less capable for PAD compared to the CNNs trained for a generic object recognition task.

1. Introduction

Biometrics is concerned with the recognition or matching of individuals based on one or more biometric traits, such as face image, fingerprint or voice [1]. Although biometrics systems have witnessed an increase in their popularity in the past decades, their reliability is seriously challenged by spoofing attacks where an unauthorised subject tries to access the system by presenting fake biometric data. To protect a biometric system, any spoofing attacks need to be detected reliably. This requires the ability to distinguish genuine biometric system accesses from spoofing accesses, which is again a pattern recognition problem.

In the case of face recognition systems, to which we confine our discussion in this paper, spoofing attacks generally appear as print attacks, replay attacks and 3D masks. During the past couple of years, a variety of different face Presentation Attack Detection (PAD) approaches have been proposed, achieving impressive performance on benchmarking datasets. The progress made mainly owes to two factors: i) the design and deployment of more effective representations which can better capture the differences between real and fake biometrics traits, and ii) using more powerful two-class classifiers.

The majority of the approaches to face spoofing detection proposed in the literature formulate the problem as a two-class pattern recognition problem. These approaches try to learn a suitable classifier to discriminate between the real-accesses and spoofing attempts. Despite the huge advances made in this direction, the existing face PAD methods do not reliably generalise to more realistic application scenarios of unseen presentation attacks [2]. They tend to deliver degraded performance, when the type of presentation attack cannot be anticipated and may take a completely new form. The challenges of the common two-class formulation include [3]: (a) Learning an effective decision boundary due to the multi-modal nature of spoofing attack data [2] (b) Difficulties in increasing the size of the training set, as the generation of attack data is quite demanding and the data is complex to collect [4]. At the same time, it cannot cover all possible unforeseen attacks. Moreover, an increase solely in the real-access data would result in a progressively deteriorating training set imbalance [2] (c) A limited ability of the two-class systems to generalise to novel attack types [5]. Different approaches have been proposed in the literature to counteract one

☆ This work was supported in part by the EPSRC Programme Grant (FACER2VM) EP/N007743/1 and the EPSRC/dstl/MURI project EP/R018456/1.

* Corresponding author.
E-mail address: s.fatemifar@surrey.ac.uk (S. Fatemifar).

https://doi.org/10.1016/j.patcog.2020.107696
0031-3203/

or more of these shortcomings with various degrees of success. To partly compensate for some of the aforementioned inadequacies, among others, the work in Arashloo et al. [3] formulated the face PAD as an anomaly detection problem, where the real-access data was considered as *normal* and the attack-accesses were presumed to be *anomalous* observations deviating from normality. The desirable properties of the one-class anomaly detection formulation include: (a) Insulation from the undesirable effects of the spoofing data diversity on the performance, as only normal data is used to build the model [2] (b) Since only real-access data is required for training, the training set can be extended more easily.

The capacity of one-class systems to detect previously unseen novel attacks is measured in Arashloo et al. [3] using an extensive evaluation on different datasets in a *novel attack evaluation* scenario. The work in Nikisins et al. [6] followed a similar one-class face PAD approach using a Gaussian mixture model based anomaly detector, exhibiting good generalisation properties for novel types of attacks. Motivated by these observations and the desirable properties of the one-class formulation in the face spoofing detection [7], the current study also follows the one-class anomaly-based approach with distinctive contributions outlined in the next subsection.

1.1. Contributions

An aspect of classifier design which has been unexplored in Arashloo et al. [3], Nikisins et al. [6] is the use of client-specific information. Although in any spoofing detection system the representations used are selected in a way that they capture the intrinsic differences between real-accesses and spoofing attempts, such an approach does not rule out the possibility that the features used can be affected by the specific characteristics of each client [8]. From this perspective, the majority of the work on face spoofing detection, including [3,6,9,10], can be considered as *client-independent* approaches. A client-independent face PAD approach assumes that the relevant information comes from either a real-access or the attack class, whereas a client-specific method assumes that the constructed representations are additionally influenced by the identities of the subjects. In [8], it is shown that the client identity information can be deployed to devise two-class classifiers which can achieve better discrimination between the real-accesses and spoofing attacks. The use of identity information in a PAD system is justifiable from the point of view that anti-spoofing mechanisms are designed to guard biometrics systems against spoofing attacks and hence work in conjunction with them. The more critical issue is that client specific PAD systems require storing client specific models, which does not pose any problem for many applications, such as access control, smart phone unlocking, e-commerce and internet banking. However, it may not be practicable for other scenarios e.g. border control, unless additional client data is stored with the subject's photo in the passport.

Although the use of client-specific information in face recognition has been studied extensively [11–13], the deployment of such information for face spoofing detection has been limited to just a few studies [8,14,15], which all proposed by using the two-class formulation framework. This work advocates the use of such information in a one-class anomaly-detection paradigm and builds effective one-class face PAD mechanisms. The identity information for face PAD is deployed in two stages: (a) instead of a single classifier applicable to all subjects, a separate client-specific one-class classifier is designed for each individual enrolled in the dataset; (b) using the score distributions of each user, subject-specific decision thresholds are determined to make the final decision. It will be shown that the use of client-specific information, both in the model construction and in setting a decision threshold, improves the detection performance of one-class spoofing detection systems by a large margin. Motivated by the aforementioned observations, it will also be shown that both types of generative and discriminative anomaly-based approaches to face spoofing detection can benefit from client identity information to improve performance.

Second, inspired by the recent success of deep neural networks and in particular deep convolutional neural networks (CNNs) [16,17], the proposed one-class approaches are fed with representations obtained from deep pre-trained CNN models. The models employed are either pre-trained for the general object classification purposes or specifically tuned for face recognition. In this respect, a further contribution of the current study is a comparative evaluation of the applicability of different deep CNN models designed for recognition purposes to the problem of face spoofing detection based on the one-class face PAD formulation. One of these CNN models is trained for face recognition and our aim is to establish whether the extracted features perform as well as features produced by deep networks trained for the task of general visual object recognition. In the affirmative, one could use the same features for recognition and spoofing detection, which would be very interesting from the point of view of practical significance as it would simplify the design of face biometrics systems. Besides, to provide a better analysis and a greater insight regarding traditional feature extraction mechanism, we perform experiments using traditional features to compare with CNNs.

Third, we evaluate the performance of our proposed model on benchmarking anti-spoofing datasets including Replay-Attack [18] and Replay-Mobile [19] to allow fair comparisons with the state of the art. The client-specific variant of the anomaly detection solution requires a new protocol for evaluation that we introduce to set up the experiments on the new and more challenging face spoofing dataset, ROSE-Youtu [20], covering a diverse variety of illumination conditions, image acquisition devices, and spoofing attacks. We made the code publicly available for the benefit of research community.[1]

2. Related work

Print attack detection methods assume that, in contrast with genuine accesses, spoofing data exhibits abnormalities that render the observations distinguishable from the authentic counterparts. In this respect, spoofing samples constitute a class of data that is different from the normal access data. PAD countermeasures can broadly be classified into hardware-based and software-based methods [21]. While software-based methods process the data collected from a typical authentication sensor, hardware-based methods use additional hardware for anti-spoofing. The hardware-based solutions typically rely on liveness measurements (e.g. employing a specific sensor to detect attributes of living bodies), attack specific detection methods (such as depth measurements against photo and video attacks) or challenge-response mechanisms (requiring the user to respond to random requests). Software-based methods, on the other hand, use different attributes of an image sequence along with different classifiers to detect presentation attacks. Different descriptors used for this purpose include texture, motion, frequency, colour, shape or reflectance whereas the two-class classifiers employed include discriminant, regression, distance metric or other heuristics.

Among the cues conveyed by an image/image sequence, the texture is probably the one most frequently used for spoofing attack detection [22]. The use of texture is based on the assumption that face presentation attacks produce certain texture patterns which do not exist in real-access attempt data. Motion-based

[1] Code available at: https://github.com/12sf12/client-specific-anomaly-detection-PR

methods constitute another group where typically two different ways of exploiting motion are considered. The first approach focuses on intra-face variations, such as facial expressions, eye blinking, and head rotation [23,24], whereas the other alternative is to assess the consistency of the user with the environment [25]. A different category of methods is constituted by frequency-based countermeasures proposed to detect certain image artefacts in 1D or 2D Fourier transform from either a single image [26] or an image sequence [25]. Although colour does not remain constant due to inconsistencies in imaging conditions, certain colour attributes have also been used to discern attacks from real-accesses in a different group of methods [27]. Another category uses the shape as a source of information to deal with some presentation attacks [28].

Along with the use of different cues for attack detection, a variety of two-class classifiers have also been examined for face PAD. Discriminant classifiers constitute one such group of methods where Support Vector Machines are the most commonly employed technique [16,29]. Other works have also examined the linear discriminant analysis for attack detection [27,30]. Other types of classifiers using discriminant procedures include neural networks [24] and Bayesian networks [31]. Another group is the regression-based methods which try to map input descriptors directly onto their class labels [14]. A more detailed review of the recent face anti-spoofing approaches can be found in Bhattacharjee et al. [2], Ramachandra and Busch [32].

The majority of the work on spoofing detection assumes that the relevant information for the detection of an attack is independent of the class identity of the data [33,34]. Accordingly, the systems are typically designed in a client-independent fashion. However, it has been observed that the representations used for the detection of spoofing attempts are invariably affected by client-specific attributes [7,8,35]. Drawing on this observation, the work in Chingovska and dos Anjos [8] studies how much client-specific information is contained within features and its effect on the performance of different systems. Using such information, two client-specific anti-spoofing solutions, one generative and the other discriminative are built. The advocated methods outperform the client-independent methods by a large margin while demonstrating better generalisation capabilities to unseen types of attacks. The work in Yang et al. [14] proposed a person-specific anti-spoofing approach using a classifier specifically trained for each subject in an attempt to dismiss the interferences among subjects. A subject domain adaptation method was then applied to synthesise virtual features making it possible to train individual face anti-spoofing classifiers. In a different study [15], the face PAD problem is addressed by modelling radiometric distortions involved in the recapturing process. Having access to the enrolment data of each client, the exposure transformation between a test sample and its enrolment counterpart is estimated. A compact parametric representation is then proposed to model the radiometric transform and is employed as features for classification.

Although the use of client-specific information in two-class models has led to some improvements, a common drawback of these two-class approaches is their unknown capacity to generalise to spoofing attempts of different nature. As a result, the authors in Fatemifar et al. [7] proposed a new way to utilise the client-specific information to train one-class classifiers using only genuine-access data. The results of their experiments demonstrate that one-class classifiers trained with client-specific information were more robust to unseen attacks compared to client-independent and multiclass frameworks. The detection of novel attack types is particularly challenging, making it impossible to predict the performance of an anti-spoofing technique in real-world scenarios. On the other hand, as it is impossible to foresee all possible attack types and cover them in the database, one-class approaches, modelling only the real-access data, present a promising direction towards the detection of unseen attack types [36].

3. Anomaly detection

3.1. Background

Anomalies are typically known as a set of patterns/conditions which are different in some way from the majority of observations considered as normal. In this respect, anomaly detection is a problem of identifying items, events or observations which do not conform to the expected behaviour or condition [37]. A general categorisation of anomaly detection methods contains generative and non-generative types. While for the generative methods there exists a model for generating all observations, non-generative approaches lack a transparent link to the data. The non-generative methods are best represented by discriminative approaches which try to identify the class identity of an observation by partitioning the feature space. The construction of an anomaly detection mechanism can be based on normal data or both normal and anomalous observations. In this work, both generative and discriminative groups are examined for the face spoofing detection problem.

3.2. One-class classifiers

3.2.1. Mahalanobis distance

As a baseline method, in this work, it is assumed that the model representation obtained from a real-access sequence follows a single-mode Gaussian distribution with mean $\mu$and covariance matrix Σ. Once the parameters characterising the normal distribution are estimated using the real-access (normal) samples of the training set, testing for normality entails computing the Mahalanobis Distance (MD) of a test pattern $x$to the mean of the normal class as follows:

$$MD(x) = \sqrt{(x-\mu)\Sigma^{-1}(x-\mu)^T} \tag{1}$$

This distance is considered as the spoofing detection score.

3.2.2. Gaussian mixture model

A Gaussian Mixture Model (GMM) is a parametric probability density function defined as a weighted sum of MGaussian component densities given as:

$$p(x|\lambda) = \sum_{i=1}^{M} w_i g(x|\mu_i, \Sigma_i) \tag{2}$$

where $w_i, i = 1, \ldots, M$,are the mixture weights, and $g(x|\mu_i, \Sigma_i)$, $i = 1, \ldots, M$,are the component Gaussian densities. The mixture weights satisfy the constraint that $\sum_{i=1}^{M} w_i = 1$. The complete set of parameters is collectively represented by the notation,

$$\lambda = \left\{w_i, \mu_i, \Sigma_i\right\}, \quad i = 1, 2, \ldots, M \tag{3}$$

The model parameters of GMM are estimated using the Expectation Maximisation (EM) algorithm [38]. EM computes the posterior probabilities of component memberships for each sample of the training set. Using the posterior probabilities as weights, EM estimates the component means, covariance matrices, and the mixing proportions by applying the maximum likelihood principle. The EM algorithm iterates several times until convergence. Once the training data is fit to GMMs using the iterative EM, outliers are detected by measuring the minimum MD to the respective mixture components. The MD obtained for each test instance can serve as a spoofing detection score.

3.2.3. Support vector data description (SVDD)

The SVDD [39] is a one-class extension of the SVM classifier in which a hypersphere is used to enclose the genuine-access data points as tightly as possible. The objective of SVDD is to find the smallest hypersphere with centre $z$and radius $R > 0$ encompassing all training objects. It is expected that by minimising the volume of the hypersphere, the chance of accepting outliers will be decreased. Given a training set $X = \{x_1, x_2, x_3, \ldots, x_N\}$consisting of Ndata points and ϕas a function mapping data to a higher dimensional space, the SVDD primal problem is given by:

$$\begin{aligned} &\min_{R,c,\xi} R^2 + \tfrac{1}{\nu N}\textstyle\sum_i \xi_i \\ s.t. \quad &\|\phi(x_i) - z\|^2 \le R^2 + \xi_i, \quad \xi_i \ge 0, \quad \forall i. \end{aligned} \tag{4}$$

The slack variables $\xi_i \ge 0$allow a soft boundary and hyperparameter $\nu \in (0, 1]$controls the trade-off between penalties $\xi_i$and the volume of sphere. In our case, the output of One-class SVM (SVDD) is the distance of a test samples to the centre of the separating hypersphere learned by the hyperspherical model. Once the SVDD model is built, test instances lying outside the sphere, i.e. $\|\phi(x_i) - z\|^2 \ge R^2$,are the detected anomalies.

According to the underlying architecture of one-class classifiers described above, MD and GMM classifiers are categorised as generative learners whereas SVDD is a discriminant classifier. The complementary information regarding the threshold selection of each one-class classifier will be provided in the next subsection.

3.3. Client-specific anomaly detection

Previous studies formulating face spoofing detection as an anomaly detection problem [3,6] considered real-access data as the *normal* observations and the spoofing attacks as *anomalies*. The examination of different detectors revealed the merits of such an approach, particularly in the case of *unseen* attacks. However, the one-class approaches in Arashloo et al. [3], Nikisins et al. [6] implicitly make the assumption that the test statistics categorisation of a pattern was independent of the client identity. In the case of the two-class formulation of the face PAD problem, this assumption has been re-evaluated in different studies [8,14,15], with the conclusion that, using client-specific information by virtue of training different client-specific face spoofing detection classifiers, led to significant improvements in the system performance.

This work advocates a similar anomaly based client-specific spoofing detection mechanism to that proposed in Fatemifar et al. [7]. In order to train a client-specific anomaly detection model, the subject identities are required both during the training as well as the operation phase of the system. As discussed in Chingovska and dos Anjos [8], such information is readily available to a face spoofing detection engine as it would work in conjunction with a face recognition system. More specifically, the enrolment data of each subject in the recognition system can be employed to build a subject-specific spoofing detection model. In the operation phase of the spoofing detection system, the class identity information is accessible from the face verification or identification engine. While in the face verification case a test subject claims an identity, in an identification scenario the test image is compared against several models stored in the gallery, whose identities are known. In both cases, the identity of the target class is known and can be utilised by the spoofing detection system.

In summary, in the current work, during the training phase of the face spoofing detection system, a separate one-class anomaly detection classifier is trained for each subject using the enrolment data of the corresponding client while in the operation phase, the test sample is matched against the model of the claimed subject. The construction of a subject-specific classifier in this work benefits from client identity information at two levels. First as noted earlier, only the enrolment data of the subject under consideration is used to build a one-class classifier. Second, by analysing the score distributions of each subject, a subject-specific threshold is determined for the final decision-making process.

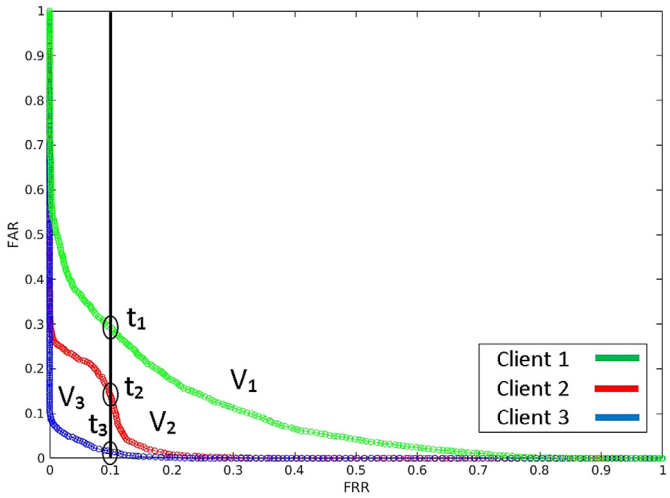

Fig. 1. An example of setting client-specific threshold using distributional curve of subjects.

3.4. Client-specific thresholds

As mentioned earlier, an anomaly detector produces a client dependent score for each biometrics trait. The principal assumption is that an attack-access would produce scores that differ from normal scores, and could be considered as outliers of the distribution of normal scores. To detect an outlier, a threshold should be defined. A common practice is to set the threshold at a predefined level of confidence. This implies a threshold rejecting a given proportion of real scores, usually 1–15%. If a validation set was available, one could set the threshold so that the False Acceptance Rate (FAR) and False Rejection Rate (FRR) were equal, producing an equal error rate (EER). In the anomaly formulation adopted here, the Half Total Error Rate (HTER) would be measured at the operating point corresponding to the selected confidence level. If no attack data is available for validation, the operating point can be selected based on an acceptable false rejection rate. This can be quite conservative, as the user can be expected to repeat the access attempt to reduce false rejection. This is also accomplished by processing multiple frames of an access attempt video footage. In this work we report the results at an operating point closest to the EER setting, assuming that some attack data is available for validation. However, it cannot be overemphasised that we do not require attack samples either for training or for enrolling new clients to the biometrics system.

As shown in Fig. 1, we compute False Acceptance Rate (FAR) and FRR according to different cut-off points for each client resulting in curves V_1,V_1,and V_3. If a confidence cut-off point is set to reject 10% of normal-accesses, as shown in Fig. 1, it cuts each $V_i$curve at a different point that corresponds to a different threshold for each subject. Once a subject-specific threshold for each client is computed, we can simply calculate the HTER by averaging FAR and FRR according to the chosen confidence cut-off point.

To demonstrate the importance of client-specific thresholds, the score distributions of three clients are presented in Fig. 2. The reported HTERs in the Table inside Fig. 2 are obtained when both client-specific and global thresholds are applied to the client-specific MD classifier. As seen in Fig. 2, the client-specific threshold of each subject can discriminate the genuine-access and attack

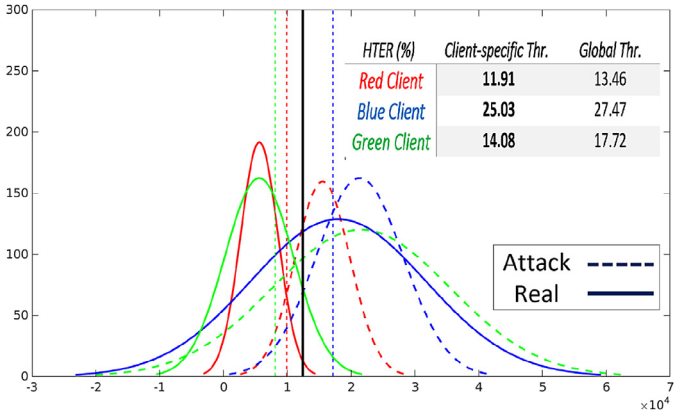

Fig. 2. The score distributions and client-specific thresholds of three subjects using *MD+GoogLeNet* system. The score distributions of real and attack samples for different clients are depicted by solid and dashed lines, respectively. The vertical colourful dashed lines and solid black line demonstrate the client-specific and global thresholds, respectively.

score distributions quite well compared to a single global threshold. The evidence of the performance improvement using client-specific thresholds can be gleaned from the fact that when data of different clients are combined, it becomes more difficult to find a common threshold satisfying all subjects jointly. Accordingly, the worst-case scenario for the face spoofing problem is the blue client having a wide real distribution and a tight attack distribution, in which a client-specific threshold obtained a better HTER, around 3%, compared to a global threshold.

3.5. Feature extraction

Motivated by the recent success of deep networks and in particular CNNs, deep pre-trained CNN models are used to derive representations for an image or image sequence. We assess the applicability of the different CNN models in the proposed client-specific one-class framework with the following two objectives in mind: (a) To confirm that CNN representations extracted by pre-trained networks outperform the conventional hand crafted features (b) To test the hypothesis that the representation extracted by a CNN model trained for the face recognition task is as good as feature extracted by CNN model trained for the generic object recognition. For practical reasons we focus on pre-trained networks, but intend to investigate the merit of end-to-end training of anomaly detectors, such as [40], in the future. The gamut of traditional image data representations used for face spoofing detection is very extensive. They include Local Binary Patterns (LBP), Local Phase Quantisation (LPQ), Image Quality Measures (IQM) features [41] and Binarised Statistical Image Features (BSIF). Their relative effectiveness has been investigated in Arashloo et al. [3]. In this paper, we adopt LBP and IQM as representatives of previous works along with HoG and FHoG feature extractors which have not been used frequently before. A description of the deep CNN models and the handcrafted feature extraction mechanisms adopted in this work is provided next.

3.5.1. GoogLeNet

GoogLeNet [42] is a 22-layer deep convolutional neural network based on the inception model. GoogLeNet achieved the state-of-the-art result for object recognition and detection in the ImageNet Large-Scale Visual Recognition Challenge 2014 (ILSVRC14) [43]. In this model, following a carefully crafted design, the depth and width of the network was increased compared to the previous networks while keeping the computational budget constant.

3.5.2. Residual neural network

Residual Neural Network (ResNet) [44] has attracted the attention of the research community as a winner of ILSVRC in 2015. The novel architecture of ResNet, which can have a very deep network of up to 152 layers (ResNet50 is used in this work), introduced a concept called skip connections to improve optimisation by enabling the flow of information across layers without attenuation. Another advantage of skip connections is to effectively simplify the network using fewer layers in the initial training stages. This leads to a speed-up in learning since there are fewer layers to propagate through.

3.5.3. VGG16

VGG16 model [45] is a pretrained CNN model which has been proven to be the state-of-the-art feature extractor. The VGG16 model which is trained on the ILSVRC benchmark for large-scale image recognition secured the first and the second places of the competition in 2014. The VGG16 has been shown to generalise well to other datasets. In terms of architecture, VGG16 consists of 13 convolutional layers and three fully connected layers, giving 16 weight layers in total.

3.5.4. VGGFace

VGGFace [46] is a deep CNN model based on the VGG model, comprised of 11 blocks, each containing a linear operator followed by one or more non-linearities such as ReLU and max pooling. The first eight blocks are convolutional, while the last three blocks are fully connected. In this network, the convolution layers are followed by a rectification layer. The model is trained on a very large scale dataset of 2.6 M images from over 2.6 K subjects. It achieved competitive results on the LFW [47] and YTF [48] face benchmarks.

3.5.5. Image quality measures

Image Quality Measures (IQMs) have been widely used in face spoofing detection [3,6]. The main assumption behind IQMs in face spoofing detection is that attacks such as 2D print attacks may be considered as an image manipulation type that can be distinguished by using IQMs. The idea of using IQMs for biometrics spoofing detection was first suggested in Galbally et al. [41]. In total 25 full-reference and blind measures have been proposed. The objective of full-reference IQMs is to measure various types of distortions of a given test image in reference with the original distortion-free image. In the area of face spoofing detection, such a reference image does not exist. However, the authors in Galbally et al. [41] proposed an approach to address this limitation and we follow the same procedure to generate the reference images. As opposed to the full-reference IQM methods, the human visual system generally does not need a reference picture to measure the image quality level. Following this principle, automatic no-reference image quality assessment methods measure the image visual quality in the absence of a reference using very complicated and challenging procedures. In this work, we use 19 images quality measures including: Mean Square Error, Peak Signal to Noise Ratio, Signal to Noise Ratio, Structural Content, Maximum Difference, Average Difference, Normalised Absolute Error, R-Averaged MD, Normalised Cross-Correlation, Total Edge Difference, Total Corner Difference, Spectral Magnitude Error, Spectral Phase Error, Gradient Magnitude Error, Gradient Phase Error, Structural Similarity Index, JPEG Quality Index, Blind Image Quality Index, and Naturalness Image Quality Estimator. The first 16 IQMs are from the full-reference category while the last three are no-reference based.

3.5.6. Local binary patterns

The original LBP operator is a powerful method of texture description whose aim is to effectively summarise the local structures of images. The LBP operator labels image pixels by thresholding the

Table 1
Dataset descriptions.

	Clients	Samples	PA instruments	Illumination conditions
Replay-Attack	50	1300	2D attacks: print, replay	2
Replay-Mobile	40	1200	2D attacks: print, replay	5
Rose-Youtu	20	3350	2D attacks: print, replay 3D attacks: Mask	5

3×3neighbourhood block of each pixel using the centre value and returning the output as a binary number. The histogram of the outputted binary numbers can be used as a texture descriptor. Later, the operator was generalised for any radius and number of points in the neighbourhood grid [49].

3.5.7. HoG

The histogram of oriented gradient (HoG) feature descriptor was first proposed in Dalal and Triggs [50] for human detection. The image is divided into a grid of overlapping rectangular blocks. The HoG descriptor for each block is based on edge information, summarised in terms of a distribution of edge orientations and the corresponding gradient magnitudes.

3.5.8. FHoG

Felzenszwalb's HoG (FHoG) [51] is an extension of HOG containing contrast sensitive and insensitive orientation channels as well as texture channels. Existing studies have shown that FHoG features achieve superior performance to the original HOG representations. As a baseline, the feature space of the original HoG is reduced in FHoG using principal component analysis. In essence, FHoG can lead to models with fewer parameters to speed up the detection and learning processes.

4. Experimental evaluation

The main aim of the experiments described in this section is to evaluate the performance of different client-specific one-class face PAD methods and compare them to the client-independent one-class approaches in an unseen attack scenario. To compare the performance of CNNs versus traditional feature extractor methods, both categories are utilised to report the results. As noted earlier, an essential pre-requisite to building a client-specific face anti-spoofing system is the availability of enrolment data for each client in the database. In other words, the enrolment and test sets should contain the same IDs. The enrolment data is originally provided for each client in Replay-Attack and Replay-Mobile datasets. However, the majority of face-spoofing datasets currently in use, including MSU-MFSD [52], Rose-Youtu, and OULU-NPU [53] lack enrolment data. Nonetheless, it is possible to modify the protocols of a number of datasets and use a subset of real-access data as an enrolment data. We opt to perform experiments on Rose-Youtu dataset, that contains more real and spoofing videos per client (25–50 real-access and 130 attack-accesses) compared to the three aforementioned datasets. It also has more illumination conditions and camera devices compared to other datasets. For instance, in case of MSU-MFSD, only three real videos are provided for each client in the test set. Regarding OULU-NPU dataset, it is again less convenient to build client-specific models using its third and fourth protocols due to the lack of real-access data. For instance, in the fourth protocol of OULU-NPU dataset, only one video is provided in the test set, which we cannot divide one video into two sets, one for testing and another one for the enrolment set.

4.1. Datasets

We provide a brief introduction of the experimental datasets used in the current work in Table 1.

To perform the experiments on the Rose-Youtu, we define a new experimental protocol which considers all different lighting conditions and cameras. In Rose-Youtu, there are either 25 or 50 real videos and 130 spoofing videos per subject. We divide the dataset into 3 subsets i.e., enrolment, validation and test. The enrolment set contains only real-accesses while validation set and test set have both real-accesses and attack-accesses. The data is split as follows: Real-accesses: enrolment (40%), validation (20%) and test (40%); Attack-accesses: validation (50%) and test (50%) The indices of videos assigned to each subset are provided along with the source code.

4.2. Implementation details

We photometrically normalise each frame of video clips based on the retinex method [54] to solve the illumination variation problem at the pre-processing level. To this end, inside the retina model function, we set the values of dogsigma1, dogsigma2, sigma1 and sigma2 to 1, 40, 10, and 30 respectively. Besides, to minimise the effect of background on the detection performance, only the face regions are used in each sequence. For this purpose, the coordinates of the faces provided along with the Replay-Attack and Replay-Mobile datasets are used to crop out a face from the entire image in each frame. For Rose-Youtu dataset, the Viola-Jones algorithm [55] is utilised to detect and crop out face regions. In case of a missing bounding box for a frame, the coordinates of the last detected face in the same sequence are used instead. To extract features, the CNNs are adapted by removing the final fully-connected layers of each model, trained to perform the classification. The adapted CNNs are then applied to the facial regions of each frame. This results in feature vectors of 1024 elements for the GoogLeNet, 2048-element feature vectors for ResNet50, 4096 elements for VGG16, and 4096-element arrays of feature for VGGFace per each frame. Likewise, traditional feature extraction methods are applied to frames and the features obtained are given to one-class classifiers. In several cases, the feature vectors obtained from different CNNs or traditional feature extraction methods are L2-normalised before feeding them to the one-class classifiers. We determine this aspect based on the development set. We also apply PCA to reduce the dimensionality by retaining 99% of the variance in the feature vectors. Each feature vector, in this case, is mapped to a lower-dimensional space using the leading eigenvectors and its components divided by the square root of the corresponding eigenvalues.

Regarding the traditional features, we use the $LBP_{16,2}$operator in all experiments. To extract HoG features, the Matlab implementation of HoG is utilised with a cell size of $8 \times 8$and block size of 2×2. For FHoG, we set bin size and the number of orientation bins to 8 and 9, respectively. For the construction of SVDD models, LIBSVM library [56]is used to build the SVDD classifiers using a linear kernel. The normalisation parameter in the kernel is determined automatically. The MD classifier is implemented as a Euclidean distance between a given test data and mean of the normal class. For the GMM classifier, the mean MD of the test sample from all mixture components is regarded as a spoofing detection score.

The value of hyperparameters are computed using 10-fold cross-validation (on the enrolment set for client-specific and

Table 2
AUC's(%) and HTER's(%) on the test set of Replay-Attack according to the frame-based scenario. The best result is marked in bold.

Replay-attack dataset												
	MD				GMM				SVDD			
	Spec		Indp		Spec		Indp		Spec		Indp	
	AUC	HTER	AUC	HTER	AUC	HTER	AUC	HTER	AUC	HTER	AUC	HTER
GoogLeNet	99.07	5.14	90.76	17.19	99.01	3.76	91.21	16.56	91.95	14.53	68.75	36.35
ResNet50	99.18	4.90	92.03	15.59	99.69	1.97	92.48	15.12	94.57	11.56	60.54	41.66
VGG16	99.20	4.80	93.78	13.26	99.53	**1.46**	90.38	17.01	95.45	9.87	69.60	35.65
VGGFace	97.85	7.23	94.09	12.75	97.04	7.27	94.35	12.79	89.98	16.83	54.77	46.50
IQM	88.12	29.25	85.24	21.37	78.21	29.82	61.13	43.08	73.80	30.28	59.63	42.14
LBP	95.01	11.52	63.68	40.61	95.89	8.88	91.50	16.09	86.48	19.59	66.96	36.25
HoG	89.07	18.45	81.65	26.79	91.30	16.11	81.03	27.39	78.10	28.00	61.94	40.77
FHoG	90.99	17.66	80.01	28.33	88.95	18.56	78.38	29.80	78.30	27.94	61.07	41.62

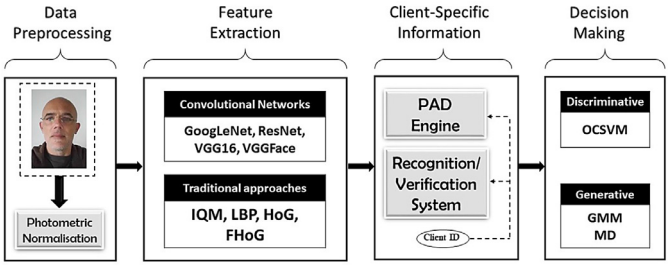

Fig. 3. The proposed approach to client-specific anomaly-detection based PAD.

client-independent anomaly classifiers and validation set for binary classifiers). We adopt the average value computed over the 10-Folds as the final hyperparameter value. Note that to train the anomaly and binary spoofing detectors, we used the computed hyperparametrs along with all samples of the enrolment set and training set, respectively. In the evaluation phase, a test query is matched only against the claimed client model. It should be reiterated that in both, client-specific and client-independent approaches, only the real-access data is used to build a one-class anomaly detection model. Hence our evaluation adheres to the *unseen* attack scenario in the sense that none of the attack types is seen during the training phase of the system. The proposed approach is depicted in a block diagram form in Fig. 3.

4.2.1. Performance measures

The performance of a spoofing detection system is commonly reported in terms of HTER, which was mentioned in Section 3.4. We also used HTER to compare the performance our proposed models with state-of-the-art approaches. For the anomaly detection approach, a more fitting way to compare the performance is using the AUC, which is an indicator of the average performance of a system across all possible decision thresholds. In order to report the performance of our proposed model with the recently standardised ISO/IEC 30107-3 terminology [32], we also consider Attack Presentation Classification Error Rate (APCER), Bonafide Presentation Classification Error Rate (BPCER) metrics, and Average Classification Error Rate (ACER). APCER is determined as the proportion of PAs incorrectly labelled as bonafide (real-accesses) presentations in a specific scenario (the PA with the highest error is reported), BPCER is defined as the proportion of bonafide presentations incorrectly listed as PAs, and ACER is computed as the mean of APCER and BPCER.

4.3. Results

The main purpose of the experiments described in this section is to demonstrate the merits of using client-specific information in conjunction with the anomaly detection paradigm. The aim is to find out whether in the guise of the client-specific formulation, anomaly-based face spoofing detection approaches become a viable alternative to two-class methods, deserving further attention in the future. As a by-product of the experiments, it will be shown that as expected, the features extracted from the deep CNN models, provide much more powerful representation for face presentation attack detection than the traditional features. As the deep CNN models and handcrafted feature extraction methods are applied frame by frame, it is possible to report the performance of different systems according to the frame-based scenario. We also report the performance of anomaly detectors using a per-video basis. For this purpose, a score level fusion approach is applied to obtain the final score for a given video. We opt for the mean fusion rule, averaging the scores of different frames in a given video to produce the final score for the video. Tables 2–4 report the performance according to the frame-based scenario for the Replay-Attack, Replay-Mobile and Rose-Youto datasets respectively. The results of the video-based scenario are given in Table 5.

As can be seen in Table 2, the client-specific *GMM+VGG16* has the lowest HTER of 1.46% and the highest AUC of 99.53% among all the other approaches. These rates are better than those achieved by the best client-independent approach, *MD+VGGFace* with HTER of 12.75% and AUC of 94.09%. The contrast between the client-specific approaches and the client-independent detectors has been observed for most of the evaluated methods. In the case of traditional feature extraction methods, client-specific *LBP+GMM* is superior to other traditional techniques with HTER of 8.88% and AUC of 95.89%. However, this rate falls far short of the best results achieved by the majority of the CNN based methods. Similar to CNNs, in traditional representation based models, client-specific methods consistently have a better performance compared to the client-independent approaches.

According to Table 3, the best performing client-specific method is the *GMM+GoogLeNet* model with the AUC measure of 94.90% and the HTER of 13.56%, whereas the AUC and HTER achieved by the best client-independent method operating on the same set of deep representations and classifier is 90.55% and 17.43%, respectively. Again here, the gap between traditional and CNN representations is considerable as the best performing traditional approach, *MD+HoG* with HTER of 34% and AUC 62.99% is inferior compared to the majority of CNN architectures. The only exception is VGGFace which is close to the HOG results. Similar to the results on the Replay-Attack dataset, all client-specific methods operating on traditional representations outperform their client-independent counterparts.

Since Rose-Youtu dataset contains more videos for the training and testing client-specific and client-independent approaches, it can provide supplementary information about the performance of the proposed client-specific approaches. As shown in Table 4, all variations of client-specific approaches outperform their client-

Table 3
AUC's(%) and HTER's(%) on the test set of Replay-Mobile according to the frame-based scenario. The best result is marked in bold.

Replay-mobile dataset												
	MD				GMM				SVDD			
	Spec		Indp		Spec		Indp		Spec		Indp	
	AUC	HTER	AUC	HTER	AUC	HTER	AUC	HTER	AUC	HTER	AUC	HTER
GoogLeNet	92.53	13.83	90.55	17.43	92.89	**13.56**	90.26	17.45	88.85	18.10	81.79	24.21
ResNet50	85.65	24.00	80.39	28.71	88.76	18.97	79.86	29.30	79.57	27.51	70.22	35.69
VGG16	88.60	20.19	82.07	25.20	89.87	17.21	82.93	23.98	86.14	19.35	77.50	30.02
VGGFace	64.47	39.03	74.20	33.01	67.77	36.61	70.35	35.55	69.57	36.83	70.46	35.49
IQM	60.59	39.93	60.01	39.51	60.53	40.91	62.78	41.71	49.59	50.13	54.48	44.73
LBP	40.07	45.81	36.52	57.85	38.80	50.12	43.15	53.80	39.86	48.99	32.64	62.70
HoG	64.79	34.00	62.99	38.49	63.33	36.42	63.94	41.10	63.01	36.06	56.48	46.44
FHoG	59.81	34.22	61.13	41.85	61.43	35.64	61.47	42.25	59.75	37.20	55.24	47.98

Table 4
AUC's(%) and HTER's(%) on the test set of Rose-Youtu according to the frame-based scenario. The best result is marked in bold.

Rose-youtu dataset												
	MD				GMM				SVDD			
	Spec		Indp		Spec		Indp		Spec		Indp	
	AUC	HTER	AUC	HTER	AUC	HTER	AUC	HTER	AUC	HTER	AUC	HTER
GoogLeNet	85.33	17.02	89.13	19.34	85.33	16.97	89.20	18.79	85.32	16.98	89.13	19.34
ResNet50	93.16	15.14	90.08	17.98	93.57	**14.69**	90.11	17.95	90.87	15.31	89.21	19.17
VGG16	91.61	16.44	87.66	19.70	91.61	16.44	87.02	20.44	91.57	16.39	87.67	19.66
VGGFace	89.49	16.08	87.15	20.17	89.49	16.08	80.24	27.59	89.50	16.03	87.31	19.94
IQM	54.84	47.07	53.55	46.11	52.67	48.51	48.67	50.02	49.68	50.89	51.18	49.63
LBP	66.95	36.25	60.12	43.24	66.95	36.26	59.49	43.22	62.58	43.28	60.28	43.78
HoG	84.32	18.91	87.05	20.19	84.32	18.92	83.71	23.68	80.49	20.99	68.53	36.40
FHoG	85.50	18.60	87.27	19.68	85.50	18.61	83.55	23.77	80.44	20.90	68.29	36.86

Table 5
Comparison between the best performing client-specific and client independent approaches according to the video-based scenario.

	Best client-specific			Best client-independent		
	Video-based scenario					
	APCER	BPCER	HTER	APCER	BPCER	HTER
Replay-Attack	0	0	0	9.54	9.14	8.45
Replay-Mobile	14.32	3.96	8.58	20.98	25.78	17.63
Rose-Youtu	17.33	10.00	8.13	20.00	0	11.48
	Frame-based scenario					
Replay-Attack	1.85	0	1.46	13.23	14.19	12.75
Replay-Mobile	23.78	5.69	13.56	32.11	12.43	17.43
Rose-Youtu	31.25	15.06	14.69	20.60	15.29	17.95

independent counterparts by a large margin. The best client-specific approach in this case, is *GMM+ResNet50* with HTER of 14.69% and AUC of 93.57% offering almost 3% improvement in HTER and AUC compared to its client-independent alternative, respectively. Similar to the findings on the other two datasets, all face spoofing detectors based on CNN representations outperform traditional ones by a considerable margin in the Rose-Youtu dataset. For instance, HTER of the best traditional method, *MD+FHoG*, is around 4% lower than the best performing client-specific CNN solution.

We also provide the results obtained from the video-based evaluation scenario to compare them with the frame-based evaluation in Table 5. We choose the best performing system in each category for the comparison. As seen in Table 5, the video-based approaches outperform frame-based counterparts in the majority of cases. The biggest differences are observed in the Replay-Attack and Rose-Youtu datasets where the client-specific approaches achieve the rates of 0% and 8.13% in terms of HTER, respectively. These numbers are by far better than the frame-based results. Note that the video-based evaluation does not always result in a better performance compared the per-frame basis [57]. The performance gain of video versus frame-based evaluation for the client-specific approaches of nearly 4% HTER reduction extend to the Replay-Mobile dataset.

4.4. Discussion

In the experiments conducted so far, it has been shown that the use of client identity information can result in large improvements in the system performance. The improvements stem not only from the deployment of client-specific classifiers, but also from the use of client-specific thresholds. The merit of the latter, compared to a global threshold, is evident from Fig. 4 showing the real score distributions of different clients from the test set of Replay-Mobile dataset, computed using *MD+GoogLeNet* system, and depicted as boxplots.

As can be seen from Fig. 4, different subjects exhibit different score distributions. Clearly, a single global threshold has no chance of being optimal for all clients. Although the best overall performing approach is found to be the GMM classifier, large im-

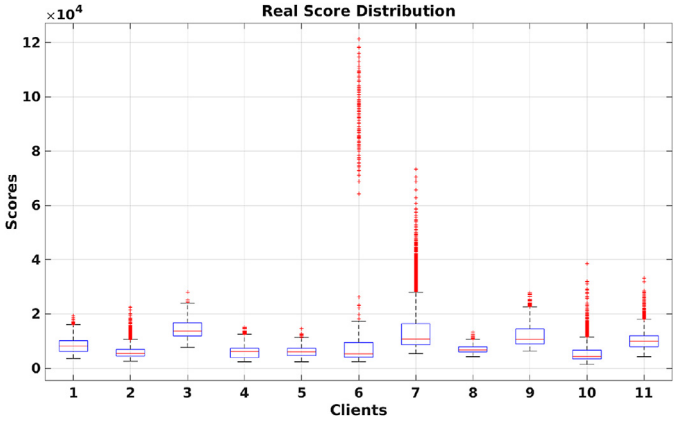

Fig. 4. The real score distributions of clients from test set of the Replay-Mobile dataset.

Fig. 5. The examples of successful and misclassified test samples using One-class GMM+ResNet50.

provements are reported by the MD classifier as well. Thus, the generative one-class classifiers used in our experiments (GMM and MD) outperform the SVDD discriminative classifier. Regarding the deep CNN models examined, the best performing CNN network is GoogLeNet that is trained for the general object recognition purposes. In contrast, VGGFace which is fine-tuned for face recognition performs worse compared to CNN networks tuning for the object recognition.

4.4.1. Operating point

Anti-spoofing systems should be designed to ensure the maximum robustness against unseen spoofing attacks. In practical anomaly detection applications, it will be challenging to set the decision thresholds since no a priory information about spoofing samples is available. As there is no guidance as to the expected rate of false positives for any selected confidence threshold, it would be advisable to err on the conservative side and set the confidence level relatively low. The expected rate of false rejections could then be mitigated by advising the user to make several access attempts. If spoofing attack samples are available for a few clients, we could use them to estimate client-specific ROC curves and merge them to get a rough indication of a suitable confidence level to set the decision thresholds that would produce the required trade-off between false rejections and false acceptances. In either cases, setting an operating point to ensure the appropriate performance is an open question which will be the subject of future studies.

4.4.2. Face tuned CNNs vs. general object recognition CNNs

As mentioned earlier, we select different CNN architectures to test the hypothesis whether the performance of VGGFace which is specific to the face recognition is as good as that enabled by the networks trained for the generic object recognition (GoogLeNet, ResNet50 and VGG16). Our experiments revealed that the generic object recognition CNNs achieved better performance compared to VGGFace. We elaborate this finding with two objectives in mind: (a) The key differences between GoogLeNet as the best performing generic object recognition CNN and VGGFace. (b) Identifying the possible reasons why VGG16 outperformed VGGFace as both VGGFace and VGG16 have the same underlying architecture [46].

In [58], an investigation of the impact of quality-related factors including compression, artefacts, blur, noise, contrast, brightness, and missing data on the performance of different CNNs is studied. Through a number of different experiments, it was found that the VGGFace is more robust to many nuisance factors compared to GoogLeNet. Therefore, it can be concluded that the VGGFace model might discard some meaningful traits conveying relevant clues for the detection of spoofing attacks to maximise the feature robustness to image distortions. In the same vein, GoogLeNet has been shown to have a marginal advantage over the VGGFace in the presence of reduced contrast [58], and this is one of the key differences between real and spoofed images. Since both VGGFace and VGG16 have the same network architectures, it is fair to perform a one-to-one comparison between them. The results obtained demonstrate that VGG16 is better than VGGFace in the majority of cases. For instance, considering the client-specific one-class approaches, VGG16 outperforms VGGFace in 8 out of 9 cases. This suggests that if a network is trained to be robust against image distortions, then it is less capable of producing representations possessing discriminative cues to differentiate between real and spoofed images.

4.4.3. Successful and failure cases

We show several success and failure cases of the proposed GMM+ResNet50 from the Replay-Mobile dataset in Fig. 5. We suspect that failure cases occur when a sudden change in the illumination condition happens. The sudden illumination change tend to increase the score value for real-accesses, but may lower the score for attack-accesses. Consequently, these test images obtain score values which are close to the threshold and make it difficult for the spoofing detectors to classify them correctly. According to Table 5, this performance degradation has affected the frame-based evaluation scenarios more than a video-based scenario. This reveals that although a number of frames are misclassified due to sudden illumination changes, the average score of all the frames in a video remains points to the correct category. It should be noted that getting a better performance using a video-based scenario is not always the case as the work of [57] obtained a superior performance using frame-based evaluation.

4.5. Cross-database spoofing detection

It is pertinent to ask, how, in anomaly based face spoofing detection, a model created using one dataset, would fare on test samples from another dataset. A pre-requisite for this is to have databases which overlap in terms of subjects. Unfortunately, in general such datasets do not exist. An exception are the Replay-Mobile and the Replay-Attack datasets which share between them three subjects. We have addressed this question by testing one of the anomaly classifiers, namely the Gaussian mixture model created for one subject (Client with ID = 1 in the training set of Replay-Mobile and Replay-Attack datasets) using the Replay-Mobile enrolment data, and tested its performance on the test set of the Replay-Attack dataset. The results in the form of score distributions are presented in Fig. 6.

Surprisingly, the scores of the spoofing attack samples are closer to the origin than the genuine-access scores. Considering

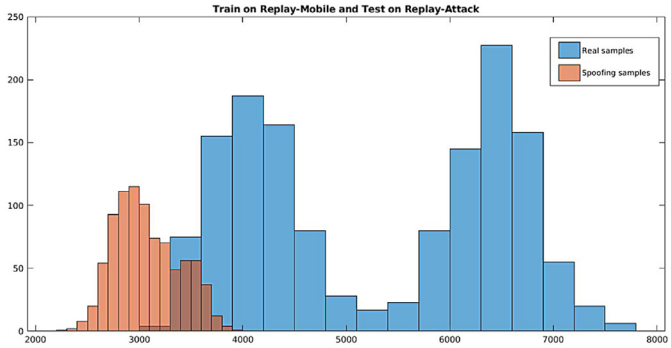

Fig. 6. The score distribution of Real samples vs. Spoofing samples in the cross-database scenario (Train on the Replay-Mobile and Test on the Replay-Attack).

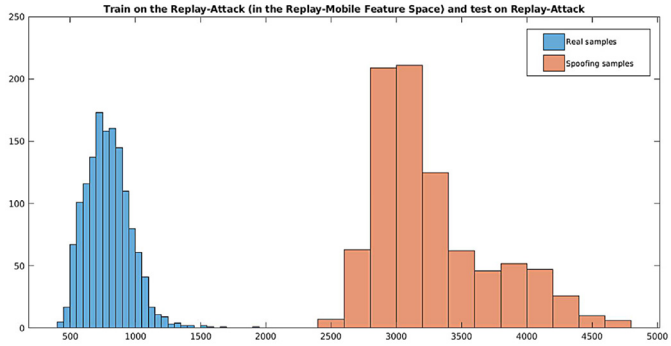

Fig. 7. The score distribution of Real samples vs. Spoofing samples in the cross-database scenario (Train on the Replay-Attack and Test on the Replay-Mobile).

that the scoring function is the Mahalanobis distance from the mean generated by the Replay-Mobile model, we observe a complete inversion of score values. Interestingly, when we created an enrolment set model using Replay-Attack training samples in the Replay-Mobile feature space, the behaviour of the test set score distributions was perfectly normal, as shown in Fig. 7.

In order to resolve this paradox, we investigated this behaviour in more detail by analysing the respective feature spaces. Recalling that the system involves a projection into a PCA space spanned by the Replay-Mobile enrolment data used for training before building one-class classifier using Gaussian mixture model. We measured the out-of-PCA space projection of the Replay-Attack enrolment samples represented by their mean vector. The result of this analysis showed that 75% of the mean vector was projected out of the PCA space. This explains the unexpected ordering of the distributions of Replay-Attack data scores with respect to the Replay-Mobile model. Interestingly, in spite of this severe curtailment of the information contained in the Replay-Attack dataset, it was still possible to create a reasonably good Gaussian mixture model for the subject in the PCA space defined by the Replay-Mobile dataset. This finding prompted the exploration of the possibility to reduce the dimensionality of the feature spaces in which models are created. More specifically, we opted for a PCA space defined by retaining 90% of the training data variance, instead of 99%. The cross dataset (Replay-Mobile vs. Replay-Attack) spoofing detection performance improved from 50% to 15% (for client with ID $= 1$). We show the result of intra vs. cross-database scenarios with the reduced PCA (%90) in Table 6 for three overlapping clients. As seen in Table 6, the HTER performance (using frame-based evaluation) degrades in the cross-database scenario. However, the gap between intra and cross-database scenario is similar to the work of [59] where the authors used Replay-Mobile and OULU-NPU datasets. This suggests that more research is needed to

Table 6
Cross-database scenario for Replay-Mobile vs. Replay-Attack dataset using frame-based HTER (%).

		Train	
		Replay-mobile	Replay-attack
Test	**Replay-Mobile**	13.98	24.52
	Replay-Attack	15.48	1.45

improve the performance of face spoofing detectors in the cross-database scenarios for anomaly based as well as binary PAD. The conclusions drawn from this exercise can be summarised as follows:

1. As part of a routine operation, it is essential to check the competence of the spoofing attack detection system to process any test data by measuring its degree of out-of-PCA space projection.
2. The lack of competence should trigger access rejection, or a process of creating a revised spoofing detection model, either from scratch, or in the existing PCA space.

4.6. Unseen attack scenario evaluation

Although there exist different methods to evaluate the generalisation capacity of PAD systems, only a few have followed the *unseen* attack type evaluation protocol [7]. In this section, we first perform supplementary experiments to investigate the potential merits of anomaly detectors compared to two-class approaches under the unseen attack scenario. Later, we compare the performance of our proposed model with the state-of-the-art techniques later in this section. A comparison of anomaly detectors and two-class approaches under the unseen attack scenario is given in Table 7. To this end, two protocols are adopted to conduct the experiments using Replay-Mobile dataset. For anomaly detectors, the unseen attack scenario is used in both protocols as anomaly detectors are trained using only real-access data. Note that, the client-specific thresholds are also predetermined for anomaly classifiers without using spoofing attacks. In the case of two-class formulation, only one type of PAs is included with real-accesses to train the binary classifiers. Note that the print-attack and video-attack are used as two PAs in the Replay-Mobile dataset. To test anomaly detectors and binary classifiers, the real-access data and only one type of PAs are considered in each protocol. As a result, the unseen PAs are used to test both anomaly detectors and binary classifiers in Protocol 1 and Protocol 2. Note that, in the case of binary classifiers, the threshold is set using the same PA type (from the validation set) used to train the classifiers.

As seen in Table 7, anomaly-based spoofing detectors outperform two-class approaches for both protocols. Besides, since only one PA is used in each protocol to evaluate the performance, it is reasonable to argue that our proposed anomaly-based solution could obtain better APCER and hence, ACER rates compared to two-class approaches. As mentioned earlier, the unseen attack protocol is a more realistic formulation of the face spoofing detection problem in real-world scenarios since impostors usually try to fool biometrics devices by using novel types of attacks rather than the previously seen spoofing attacks.

In the rest of this section, we compare our proposed model with the state-of-the-art methods. The majority of the methods which have tried a cross-database evaluation are directed more towards evaluating the generalisation capabilities of different systems subject to different imaging conditions (lighting, background, sensor-interoperability, etc.) rather than attack types. Among the other two-class systems, the works of [6,15,60] have followed the *unseen* evaluation scheme on the Replay-Attack dataset. While the

Table 7
The comparison between the performance of anomaly detectors and two-class approaches using two different protocols in terms of frame-based HTER (%).

	Replay-Mobile	
	Protocol 1 Train: Anomaly: Real data Two-class: Real data + Print PAs Test (Both Anomaly and Two-class): Real data + Video PAs	Protocol 2 Train: Anomaly: Real data Two-class: Real data + Video PAs Test (Both Anomaly and Two-class): Real data + Print PAs
One-class (GMM + GoogLeNet)	**15.35**	**5.58**
One-class (GMM + ResNet50)	23.09	10.56
Two-class (SVM + GoogLeNet)	20.02	7.68
Two-class (SVM+ ResNet50)	18.47	7.12

Table 8
A comparison of the performance of the proposed countermeasure and state-of-the-art multiclass methods. (*) indicates approaches based on an unseen attack scenario. The best result is marked in bold.

	HTER (%)
Replay-Attack	
(*) Chingovska et al. [60]	6.29
MRCNN [61]	1.6
Patch-based CNN [9]	1.25
Depth-based CNN [9]	0.75
Fusion of the two Patch and Depth CNNs [9]	0.72
Deep discriminative feature maps [62]	0.3
Attention-based two-stream CNN [63]	0.25
Image Quality Assessment [10]	0.03
Deep Learning [64]	0
(*) ***Proposed Method***	***0***
Replay-Mobile	
two-class SVM + Motion [6]	10.4
two-class SVM + Gabor [6]	9.13
Deep Pixel-wise [59]	**0**
(*) ***Proposed method***	*8.58*
Rose-Youtu	
Wavelet [20]	26.6
CoALBP [20]	16.4
Deep Learning [20]	8.0
(*) ***Proposed method***	***8.13***

evaluation scheme considered in Chingovska [60] is unseen in the sense that it excludes one of the three attack types (Print, Digital Photo and Video) during training in each of the considered scenarios, the evaluation process in Edmunds and Caplier [15] cannot be considered completely unseen as the authors use similar attack types (video replays) both for training and evaluation in some of their evaluations. The work in Nikisins et al. [6] cannot be also compared to our proposed model since their unseen attack scenario was applied to the aggregation of Replay-Attack and Replay-Mobile datasets. As a result, our comparison is limited to the work in Chingovska [60].

The best HTER on the test set obtained using a discriminative approach in Chingovska [60] is 6.29%, whereas the best HTER on the test set obtained in this work is 0%. A comparison between the performance of the proposed solutions and state of the art multiclass methods based on all possible scenarios is reported in Table 8. Note that the work of [60] is also worse than the rest of approaches reported in Table 8 since the unseen attack scenario is much harder than the seen scenario. The results reported in Table 8 are mainly based on the per-video basis. As it can be seen in Table 8, although our proposed framework is not facilitated by attack data for training anomaly detectors, it outperforms the majority of multiclass approaches. Note that in the Replay-Mobile dataset, our best approach is worse than the work of [59], the design of which does not adhere to the pure anomaly detection concept. The authors used binary supervision to train their system. Apart from the approaches reported in Table 8, our proposed model has a similar performance to the work of [34], which proposed a novel deep tree architecture for Zero-shot learning in case of the Replay-Attack dataset. It should be reiterated that our approach only uses genuine-accesses whereas zero-shot learning uses attack-accesses data along with genuine-accesses.

5. Conclusion

A novel approach to face presentation attack detection in the *unseen* attack scenario is developed. Motivated by the promising results of the one-class anomaly detection approaches, a client-specific version of the one-class methodology is proposed for the detection of face presentation attacks. Both generative and discriminative one-class classifiers utilising only positive samples (real-access data) for training are examined. It is shown that the use of client identity information in the model construction can boost the system performance of both discriminative and generative approaches significantly. Based on the score distributions of different clients, subject-specific thresholds are determined and used which further improve the performance. Different deep CNN networks and traditional features extraction approaches are compared for PAD with the conclusion that the CNN features outperform handcrafted counterparts. In addition, it is observed that the representations extracted by CNNs tuned for generic object recognition contain more information for spoofing attack detection, compared to face recognition CNNs. From the cross-database experiments, it was noted that the performance was far from perfect and future research will be needed to redress this deficiency. A comparison of the proposed one-class client-specific approaches to two-class methods in the unseen attack scenario confirmed the merits of the proposed approach. In future, we will explore the end-to-end deep learning approach using anomaly based formulation inspired by the work of [40].

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] J. Unar, W.C. Seng, A. Abbasi, A review of biometric technology along with trends and prospects, Pattern Recognit. 47 (8) (2014) 2673–2688.
[2] S. Bhattacharjee, A. Mohammadi, A. Anjos, S. Marcel, Recent Advances in Face Presentation Attack Detection, Springer International Publishing, Cham, pp. 207–228.
[3] S.R. Arashloo, J. Kittler, W. Christmas, An anomaly detection approach to face spoofing detection: a new formulation and evaluation protocol, IEEE Access 5 (2017) 13868–13882.

[4] S. Fatemifar, M. Awais, A. Akbari, J. Kittler, A stacking ensemble for anomaly based client-specific face spoofing detection, in: ICIP 2020 - 27th IEEE International Conference on Image Processing, 2020.
[5] J.J. Engelsma, A.K. Jain, Generalizing fingerprint spoof detector: learning a one-class classifier, CoRR arXiv:1901.03918 (2019).
[6] O. Nikisins, A. Mohammadi, A. Anjos, S. Marcel, On effectiveness of anomaly detection approaches against unseen presentation attacks in face anti-spoofing, in: 2018 (ICB), 2018, pp. 75–81.
[7] S. Fatemifar, S.R. Arashloo, M. Awais, J. Kittler, Spoofing attack detection by anomaly detection, in: ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2019, pp. 8464–8468.
[8] I. Chingovska, A.R. dos Anjos, On the use of client identity information for face antispoofing, IEEE Trans. Inf. Forensics Secur. 10 (4) (2015) 787–796.
[9] Y. Atoum, Y. Liu, A. Jourabloo, X. Liu, Face anti-spoofing using patch and depth-based CNNs, in: 2017 IEEE International Joint Conference on Biometrics (IJCB), 2017, pp. 319–328.
[10] E. Fourati, W. Elloumi, A. Chetouani, Anti-spoofing in face recognition-based biometric authentication using image quality assessment, Multimed. Tools Appl. 79 (1) (2020) 865–889, doi:10.1007/s11042-019-08115-w.
[11] S.R. Arashloo, J. Kittler, Class-specific kernel fusion of multiple descriptors for face verification using multiscale binarised statistical image features, IEEE Trans. Inf. Forensics Secur. 9 (12) (2014) 2100–2109.
[12] S.R. Arashloo, Multiscale binarised statistical image features for symmetric face matching using multiple descriptor fusion based on class-specific LDA, Pattern Anal. Appl. 20 (1) (2017) 113–126.
[13] A. Iosifidis, M. Gabbouj, Neural class-specific regression for face verification, IET Biom. 7 (7) (2018) 63–70.
[14] J. Yang, Z. Lei, D. Yi, S.Z. Li, Person-specific face antispoofing with subject domain adaptation, IEEE Trans. Inf. Forensics Secur. 10 (4) (2015) 797–809.
[15] T. Edmunds, A. Caplier, Fake face detection based on radiometric distortions, in: 2016 Sixth International Conference on Image Processing Theory, Tools and Applications (IPTA), 2016, pp. 1–6.
[16] X. Song, X. Zhao, L. Fang, T. Lin, Discriminative representation combinations for accurate face spoofing detection, Pattern Recognit. 85 (2019) 220–231.
[17] S. Jia, G. Guo, Z. Xu, A survey on 3D mask presentation attack detection and countermeasures, Pattern Recognit. 98 (2020) 107032.
[18] I. Chingovska, A. Anjos, S. Marcel, On the effectiveness of local binary patterns in face anti-spoofing, in: 2012 BIOSIG - Proceedings of the International Conference of Biometrics Special Interest Group (BIOSIG), 2012, pp. 1–7.
[19] A. Costa-Pazo, S. Bhattacharjee, E. Vazquez-Fernandez, S. Marcel, The replay–mobile face presentation-attack database, in: Proceedings of the International Conference on Biometrics Special Interests Group (BioSIG), 2016.
[20] H. Li, W. Li, H. Cao, S. Wang, F. Huang, A.C. Kot, Unsupervised domain adaptation for face anti-spoofing, IEEE Trans. Inf. Forensics Secur. 13 (7) (2018) 1794–1809.
[21] T. Edmunds, Protection of 2D face identification systems against spoofing attacks, Univ. Grenoble Alpes, 2017 Theses.
[22] P. Wild, P. Radu, L. Chen, J. Ferryman, Robust multimodal face and fingerprint fusion in the presence of spoofing attacks, Pattern Recognit. 50 (2016) 17–25.
[23] K. Kollreider, H. Fronthaler, J. Bigun, Non-intrusive liveness detection by face images, Image and Vision Computing 27 (3) (2009) 233–244. Special Issue on Multimodal Biometrics
[24] L. Feng, L.-M. Po, Y. Li, X. Xu, F. Yuan, T.C.-H. Cheung, K.-W. Cheung, Integration of image quality and motion cues for face anti-spoofing: a neural network approach, J. Vis. Commun. Image Represent. 38 (2016) 451–460.
[25] A. Pinto, H. Pedrini, W.R. Schwartz, A. Rocha, Face spoofing detection through visual codebooks of spectral temporal cubes, IEEE Trans. Image Process. 24 (12) (2015) 4726–4740.
[26] G. Kim, S. Eum, J.K. Suhr, D.I. Kim, K.R. Park, J. Kim, Face liveness detection based on texture and frequency analyses, in: 2012 5th (ICB), 2012, pp. 67–72.
[27] J. Galbally, S. Marcel, Face anti-spoofing based on general image quality assessment, in: 2014 22nd International Conference on Pattern Recognition, 2014, pp. 1173–1178.
[28] T. Wang, J. Yang, Z. Lei, S. Liao, S.Z. Li, Face liveness detection using 3D structure recovered from a single camera, in: 2013 International Conference on Biometrics (ICB), 2013, pp. 1–6.
[29] G. Santos, P.H. Pisani, R. Leyva, C.-T. Li, T. Tavares, A. Rocha, Manifold learning for user profiling and identity verification using motion sensors, Pattern Recognit. 106 (2020) 107408.
[30] B. Hamdan, K. Mokhtar, The detection of spoofing by 3D mask in a 2D identity recognition system, Egypt. Inform. J. 19 (2) (2018) 75–82.
[31] T. Edmunds, A. Caplier, Motion-based countermeasure against photo and video spoofing attacks in face recognition, J. Vis. Commun. Image Represent. 50 (2018) 314–332.
[32] R. Ramachandra, C. Busch, Presentation attack detection methods for face recognition systems: a comprehensive survey, ACM Comput. Surv. 50 (1) (2017) 8:1–8:37.
[33] A. Abhyankar, S. Schuckers, Integrating a wavelet based perspiration liveness check with fingerprint recognition, Pattern Recognit. 42 (3) (2009) 452–464.
[34] Y. Liu, J. Stehouwer, A. Jourabloo, X. Liu, Deep tree learning for zero-shot face anti-spoofing, CoRR arXiv:1904.02860 (2019).
[35] S. Fatemifar, M. Awais, S.R. Arashloo, J. Kittler, Combining multiple one-class classifiers for anomaly based face spoofing attack detection, in: 2019 International Conference on Biometrics (ICB), 2019, pp. 1–7.
[36] A. Hadid, Face biometrics under spoofing attacks: vulnerabilities, countermeasures, open issues, and research directions, in: 2014 CVPR Workshops, 2014, pp. 113–118.
[37] R. Domingues, M. Filippone, P. Michiardi, J. Zouaoui, A comparative evaluation of outlier detection algorithms: experiments and analyses, Pattern Recognit. 74 (2018) 406–421.
[38] A.P. Dempster, N.M. Laird, D.B. Rubin, Maximum likelihood from incomplete data via the em algorithm, J. R. Stat. Soc. Ser. B (Methodological) 39 (1) (1977) 1–38.
[39] D.M. Tax, R.P. Duin, Support vector data description, Mach. Learn. 54 (1) (2004) 45–66.
[40] L. Ruff, R. Vandermeulen, N. Goernitz, L. Deecke, S.A. Siddiqui, A. Binder, E. Müller, M. Kloft, Deep one-class classification, in: J. Dy, A. Krause (Eds.), Proceedings of the 35th International Conference on Machine Learning, Proceedings of Machine Learning Research, 80, PMLR, Stockholmsmässan, Stockholm Sweden, 2018, pp. 4393–4402.
[41] J. Galbally, S. Marcel, J. Fierrez, Image quality assessment for fake biometric detection: application to iris, fingerprint, and face recognition, IEEE Trans. Image Process. 23 (2) (2014) 710–724.
[42] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, Going deeper with convolutions, in: 2015 (CVPR), 2015, pp. 1–9.
[43] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, A.C. Berg, L. Fei-Fei, ImageNet large scale visual recognition challenge, Int. J. Comput. Vis. (IJCV) 115 (3) (2015) 211–252.
[44] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, in: 2016 (CVPR), 2016, pp. 770–778.
[45] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, CoRR arXiv:1409.1556 (2014).
[46] O.M. Parkhi, A. Vedaldi, A. Zisserman, Deep face recognition, in: British Machine Vision Conference, 2015.
[47] G.B. Huang, M. Ramesh, T. Berg, E. Learned-Miller, Labeled Faces in the Wild: A Database for Studying Face Recognition in Unconstrained Environments, Technical Report 07–49, University of Massachusetts, Amherst, 2007.
[48] L. Wolf, T. Hassner, I. Maoz, Face recognition in unconstrained videos with matched background similarity, in: CVPR 2011, 2011, pp. 529–534.
[49] T. Ojala, M. Pietikainen, T. Maenpaa, Multiresolution gray-scale and rotation invariant texture classification with local binary patterns, IEEE Trans. Pattern Anal. Mach. Intell. 24 (7) (2002) 971–987.
[50] N. Dalal, B. Triggs, Histograms of oriented gradients for human detection, in: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05), 1, 2005, pp. 886–893 vol. 1.
[51] P.F. Felzenszwalb, R.B. Girshick, D. McAllester, D. Ramanan, Object detection with discriminatively trained part-based models, IEEE Trans. Pattern Anal. Mach. Intell. 32 (9) (2010) 1627–1645.
[52] D. Wen, H. Han, A.K. Jain, Face spoof detection with image distortion analysis, IEEE Trans. Inf. Forensics Secur. 10 (4) (2015) 746–761.
[53] Z. Boulkenafet, J. Komulainen, L. Li, X. Feng, A. Hadid, OULU-NPU: a mobile face presentation attack database with real-world variations, 2017.
[54] V. Štruc, N. Pavešić, Photometric Normalization Techniques for Illumination Invariance, IGI-Global, pp. 279–300.
[55] P. Viola, M. Jones, Rapid object detection using a boosted cascade of simple features, in: Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. CVPR 2001, 1, 2001. I–I
[56] C. chung Chang, C.-J. Lin, Libsvm: a library for support vector machines, 2001,
[57] Z. Boulkenafet, J. Komulainen, A. Hadid, Face spoofing detection using colour texture analysis, IEEE Trans. Inf. Forensics Secur. 11 (8) (2016) 1818–1830.
[58] K. Grm, V. Štruc, A. Artiges, M. Caron, H.K. Ekenel, Strengths and weaknesses of deep learning models for face recognition against image degradations, IET Biom. 7 (1) (2018) 81–89.
[59] A. George, S. Marcel, Deep pixel-wise binary supervision for face presentation attack detection, in: 2019 (ICB), 2019, pp. 1–8.
[60] I. Chingovska, Trustworthy Biometric Verification under Spoofing Attacks: Application to the Face Mode, École Polytechnique Fédérale de Lausanne, 2015 Ph.D. thesis. Thèse EPFL, n 6879 (2016)
[61] Y. Ma, L. Wu, Z. Li, F. liu, A novel face presentation attack detection scheme based on multi-regional convolutional neural networks, Pattern Recognit. Lett. 131 (2020) 261–267.
[62] Y.A.U. Rehman, L.-M. Po, J. Komulainen, Enhancing deep discriminative feature maps via perturbation for face presentation attack detection, Image Vis. Comput. 94 (2020) 103858.
[63] H. Chen, G. Hu, Z. Lei, Y. Chen, N.M. Robertson, S.Z. Li, Attention-based two-stream convolutional networks for face spoofing detection, IEEE Trans. Inf. Forensics Secur. 15 (2020) 578–593.
[64] I. Manjani, S. Tariyal, M. Vatsa, R. Singh, A. Majumdar, Detecting silicone mask-based presentation attack via deep dictionary learning, IEEE Trans. Inf. Forensics Secur. 12 (7) (2017) 1713–1723.

Soroush Fatemifar, received the B.Eng. degree in Computer Hardware Engineering from the Shiraz University, Shiraz, Iran., in 2015, the M.Sc. degree in Information Technology from the University of Tehran, Tehran, Iran, in 2017.He is a current Ph.D. student in Electrical Engineering at the University of Surrey, Guildford, U.K. His research interests cover the biometrics, machine learning, anomaly detection, and also ensemble learning.

Shervin Rahimzadeh Arashloo, received the Ph.D. degree from the centre for vision, speech and signal processing, university of Surrey, U.K. He is assistant professor with the Department of Computer Engineering, Bilkent University, Ankara, Turkey and also holds a visiting research fellow position at the Centre for Vision, Speech and Signal Processing, University of Surrey, U.K. His research interests include secured biometrics, image texture analysis, anomaly detection and graphical models with applications to image and video analysis.

Muhammad Awais, received the B.Sc. computer engineering from UET Taxila in 2005, M.Sc signal processing and machine intelligence and Ph.D. in machine learning from University of Surrey in 2008 and 2011. His research interests include biometric, image processing, large scale image retrieval, medical image analysis and retrieval, computer vision, pattern recognition, machine learning and deep learning.

Josef Kittler, FREng is Distinguished Professor at University of Surrey, specialising in machine intelligence. He received his B.A. in Electrical Engineering (1971), Ph.D. in Pattern Recognition (1974), and ScD (1992), all from University of Cambridge. He conducts research in computer vision, biometrics, and machine learning. He published the Prentice Hall textbook Pattern Recognition: A Statistical Approach and more than 600 scientific papers. He is Fellow of IAPR, IEE/IET and EURASIP, and received the KS Fu Prize in 2006, and the IET Faraday Medal in 2008.