

Matrix-Regularized One-Class Multiple Kernel Learning for Unseen Face Presentation Attack Detection

Shervin Rahimzadeh Arashloo 

Abstract—The functionality of face biometric systems is severely challenged by presentation attacks (PA's), and especially those attacks that have not been available during the training phase of a PA detection (PAD) subsystem. Among other alternatives, the one-class classification (OCC) paradigm is an applicable strategy that has been observed to provide good generalisation against unseen attacks. Following an OCC approach for the unseen face PAD from RGB images, this work advocates a matrix-regularised multiple kernel learning algorithm to make use of several sources of information each constituting a different view of the face PAD problem. In particular, drawing on the one-class null Fisher classification principle, we characterise different deep CNN representations as kernels and propose a multiple kernel learning (MKL) algorithm subject to an (r, p) -norm ($1 \leq r, p$) matrix regularisation constraint. The proposed MKL algorithm is formulated as a saddle point Lagrangian optimisation task for which we present an effective optimisation algorithm with guaranteed convergence. An evaluation of the proposed one-class MKL algorithm on both general object images in an OCC setting as well as on different face PAD datasets in an unseen zero-shot attack detection setting illustrates the merits of the proposed method compared to other one-class multiple kernel and deep end-to-end CNN-based methods.

Index Terms—Unseen face presentation attack detection, one-class Fisher null projection, multiple kernel learning, matrix regularisation, zero-shot learning.

I. INTRODUCTION

ATTACKS made at the sensor level (presentation attacks) present a challenge to the operability of face biometric systems. Typical instances of presentation attacks include print attacks, video replay attacks, etc. The common approach to the problem is to collect both genuine and presentation attack samples to train a two-class classifier. Such an approach, however, implicitly assumes the problem a close-set recognition task and lies on the premise that all presentation attacks that might be encountered by a biometric system in a real-world setting can be anticipated and covered in the training set. Typically, such closed-set binary classifiers would be inclined towards flagging those observations which are similar to the negative samples as PA's and others as genuine samples. Nevertheless,

for an unseen attack, it could not be ensured that the feature space representation of the observation would resemble those previously seen by the system during the training stage. As such, there would be a high chance that the novel unseen attack may fall onto the wrong side of the decision boundary of the learned closed-set two-class classifier, putting the functionality of the biometric system at high risk in real-world applications. Accordingly, the face PAD problem may be better characterised as an open-set recognition task in a real-world setting, necessitating a different approach to be dealt with. The importance of unseen attacks has not only been identified in the context of face PAD [1]–[6] but also in other biometric modalities [7]–[11], motivating lots of intensive research on the problem. Among others options, one potential approach to the problem is that of one-class classification (OCC) [12], [13]. A fundamental difference between the OCC and the conventional two-/multi-class classification formalism is that in an OCC setting the classifier mainly uses observations from a single, typically target (i.e. normal/positive) class for training. In this case, genuine biometric samples may be considered as normal/target observations while PA's are regarded as anomalies (a.k.a. novelties, outliers, etc.). Since in an OCC approach the training set is primarily formed from genuine observations, the trained system is less biased towards any particular attack type, and thus, may possess a higher capacity to detect unseen attacks.

An effective strategy to boost the classification performance is to combine multiple information sources which might exist for the problem at hand. A well studied fusion approach among others is that of multiple kernel fusion where observations are projected onto a reproducing kernel Hilbert space (RKHS) to construct multiple kernels followed by a combination of multiple base kernels. In this context, different kernels may represent different views of similarity captured via different kernel functions or correspond to different representations obtained from different modalities/sources. A successful application of this strategy for the unseen face PAD problem is presented in [14] where a one-class multiple kernel fusion approach is developed. Although the kernel fusion method in [14] is shown to be effective in detecting unseen attacks, a shortcoming of this approach is that all the individual base kernels are considered to be equally important, and thus, are weighted equally in the composite kernel. Such a strategy, ignores any potential differences that might exist in the discriminatory information content

Manuscript received April 23, 2021; revised July 21, 2021 and August 16, 2021; accepted August 26, 2021. Date of publication September 10, 2021; date of current version September 24, 2021. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Clinton Fookes.

The author is with the Department of Computer Engineering, Faculty of Engineering, Bilkent University, 06800 Ankara, Turkey (e-mail: s.rahimzadeh@cs.bilkent.edu.tr).

Digital Object Identifier 10.1109/TIFS.2021.3111766

1556-6021 © 2021 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See <https://www.ieee.org/publications/rights/index.html> for more information.

between different views of the problem characterised via kernels.

In the context of kernel methods when multiple base kernels are available, an optimum combination of all kernels, obtained via the so-called a multiple kernel learning (MKL) algorithm [15], [16] is desired. Although other alternatives exist [17], [18], the common approach for MKL is to assume the composite kernel a linear fusion of multiple base kernels [19]–[23] and translate the MKL task into one of finding optimal linear combination weights. In this context, different prior assumptions and regularisations on kernel weights have been considered among which a vector ℓ_p -norm ($p \geq 1$) is one widely used regularisation scheme. An immediate implication of an ℓ_p -norm regularisation is providing a controlling mechanism over the sparsity of the kernel weight vector. While there has been lots of intensive work on ℓ_p -norm MKL in a multi-class setting [19], [24]–[26], and the vector ℓ_p -norm regularisation has been a popular choice to impose sparsity on kernel weights [24], [25], yet, other alternatives exist. In particular, a matrix-norm regularisation for multi-class classification scenarios has been considered in different studies [27]–[29]. It is known that since the ℓ_p -norm regulariser operates on individual kernels separately, it does not explicitly take into account any possible interaction between kernels. A matrix-norm regulariser, on the other hand, is better suited to capture inter-kernel pairwise couplings, leading to improved performance as observed in different multi-class classification problems.

Inspired by the observations above, in this study, a one-class matrix (r, p) -norm constrained multiple kernel learning technique is proposed for the unseen face PAD problem. In particular, drawing on the one-class Fisher null projection [30]–[32], the one-class mixed-norm MKL problem is formulated as a convex min-max optimisation task for which an effective optimisation approach is presented. In the context of unseen face PAD, an important characteristic of the proposed MKL approach is that it only uses target (i.e. genuine/positive) training samples. Utilisation of only genuine samples for training is advantageous as it removes any potential bias towards any particular attack type, and consequently, improves the generalisation capability for the detection of unseen attacks. On the other hand, by virtue of a matrix-norm regularisation constraint, the proposed one-class multiple kernel learning algorithm allows to combine multiple representations in an effective fashion to derive a kernel-based OCC classifier with better unseen PAD capabilities.

The existing work on matrix-regularised MKL only considers the multi-/binary-class classification problem [27]–[29] and require training data from multiple classes to operate. This is a fundamental limitation with regards to the unseen zero-shot face PAD problem since, for this problem, training data from only a single class is utilised. In this context, and in contrast to [27]–[29], the proposed matrix-regularised multiple kernel learning algorithm operates in a pure one-class learning setting. A further algorithmic and mathematical difference between the proposed approach and those in [27]–[29] is that the formulation of the multiple kernel learning problem in the current study is based on the one-class Fisher null

classification principle as opposed to other work in [27]–[29] which operate based on an SVM formulation. As validated through the experiments, a one-class Fisher null formulation offers a superior classification performance as compared with an SVM-based one-class classifier.

The main contributions of the current study are as follows:

- We propose a one-class (r, p) -norm multiple kernel learning algorithm based on the one-class Fisher null method for the unseen face PAD problem and pose the corresponding problem as a saddle point optimisation task. A matrix-norm regularisation scheme not only provides a mechanism to tune into the inherent sparseness of a particular problem but also facilitates modelling interactions between kernels, improving the detection performance of the final one-class classifier.
- We derive an effective method to optimise the saddle point optimisation problem associated with the proposed one-class MKL algorithm.
- We carry out an evaluation of the proposed MKL algorithm on both general object image and face presentation attack datasets in an unseen attack scenario and compare its performance against the baseline and other methods from the literature including multiple kernel and deep one-class end-to-end learning approaches. By virtue of an optimal combination of multiple representations of the problem posed as a one-class MKL task, the proposed approach provides superior performance compared with the existing methods.

The article is structured as follows. An overview of related work on face PAD with an emphasis on the methods focusing on the detection of unseen attacks is presented in Section II. In Section III, a background on the one-class Fisher null projection is presented. In Section IV, once a short summary of the existing one-class MKL algorithms is provided, we present our new one-class matrix-regularised (r, p) -norm MKL algorithm. The results of an assessment of the proposed one-class MKL algorithm on different databases are provided in Section V. Finally, in Section VI, we provide conclusions.

II. PRIOR ART

The face PAD problem has been addressed using hardware-, challenge-response- or software-based mechanisms [33]–[36] among which the software-based approaches have received more popularity. Distinct from presentation attacks, there exist other attack types that might be directed towards a classification system such as those considered in [37], [38]. The face PAD methods classify an image/video using different inherent representations obtained from the signal content. The current study follows a software-based formalism for face PAD based on visible spectrum RGB images. In terms of software-based methods, texture is regarded as the most commonly used cue for PAD [39], [40] while other approaches based on motion [41]–[43] also exist. Frequency-based methods try to detect PA's based on their frequency content in the Fourier domain [43]–[46] while other work [47], [48] use colour and shape information for face PAD. There exist other studies [49] where a statistical approach is used to model noise for face PAD.

A well known group of methods focuses on deep convolutional neural networks [50]–[52] to detect presentation attacks.

From a classification perspective, the conventional approach to the face PAD problem is that of two-class classification. The widely used two-class formulations include the linear discriminant analysis [47], [53], Support Vector Machines [54], [55], neural networks [42] or convolutional neural networks [50], [51] as well as Bayesian networks [56] and Adaboost-based [57] approaches. In contrary to the two-class classification-based methods, there exist regression-based techniques that try to project input representations to the corresponding labels [58].

A strong alternative to the common two-class approach for face PAD, is that of one-class classification which has been found to be especially useful with regards to unseen attacks [1]. One instance of the approaches in this group is that of [2] which uses a GMM to learn the distribution of genuine observations based on image quality metrics. Other study [3] considers the One-Class SVM and Auto-Encoder one-class classifiers for the detection of unseen attacks. The authors in [4], [59] advocate a client-specific modelling approach to train one-class learners separately for each subject in the dataset. Other study [60] addresses the unknown PAD problem in a zero-shot classification framework using a tree network. A triplet focal loss in the context of a metric learning approach is considered in [6]. The work in [61] considers a sum fusion rule for one-class classifier combination. While a fused system illustrates some improvement compared to the single best classifier deployed, the improvements were very limited. For this purpose, in the same study, the authors also applied a weighted sum fusion rule over classifier scores to further boost the performance where the fusion weights were learned using both positive and negative samples. Such a weight tuning mechanism, however, raises concerns regarding the generalisation of such weights to unseen attacks. Moreover, as the fusion mechanism operates solely on similarity/dissimilarity scores produced by individual learners, any other potentially useful information inherent to each individual learner is ignored. Other work [14] proposed a multiple kernel fusion method over one-class kernel learners where it is demonstrated that a kernel fusion approach could significantly improve the detection performance compared to each individual kernel. One advantage of a kernel fusion approach is that the discriminatory information of each individual representation is not summarised solely as a single score as is the case when a classifier fusion is practised over classifier scores. Nevertheless, in [14], the relative importance of each individual kernel is ignored and all kernels are weighted equally in the composite kernel which may compromise the performance.

In contrast to the existing multiple classifier fusion systems such as [61] which are based on scores generated by different one-class learners, the proposed multiple kernel learning algorithm operates on features in a RKHS. As such, the combined system have access to a more informative and richer data representation compared to multiple classifier systems which may only access classifier (dis)similarity scores. As verified through experiments, this improves the capability of the proposed MKL system to detect unseen attacks. On the other

hand, some other studies employ both positive and negative samples to weigh different classifiers. The use of negative training PA samples for weight learning, however, may pose difficulties in detecting novel attacks as such weights have been optimised for certain attack types. In comparison, our proposed MKL algorithm, consistent with the presumed evaluation setting of unseen attacks, solely uses positive samples for training and operates in a pure one-class framework which enhances its generalisation towards novel attacks. Finally, as noted previously, compared to other multiple kernel fusion approaches for face PAD [14] which ignore potential discrepancies in the discriminatory information content of different representations, our multiple kernel learning algorithm learns the intrinsic sparsity of the problem and encodes relative importance of different kernel representations for improved detection performance.

III. BACKGROUND

The Fisher classification principle tries to maximise the following ratio [62]:

$$\mathbf{F}(\boldsymbol{\beta}) = \frac{\boldsymbol{\beta}^\top \boldsymbol{\Sigma}_b \boldsymbol{\beta}}{\boldsymbol{\beta}^\top \boldsymbol{\Sigma}_w \boldsymbol{\beta}}$$

where $\boldsymbol{\Sigma}_b$ and $\boldsymbol{\Sigma}_w$ respectively stand for the between- and within-class scatter matrices, and $\boldsymbol{\beta}$ denotes the discriminant. The Fisher null approach [63] corresponds to the theoretically optimal Fisher discriminant $\boldsymbol{\beta}_* = \arg \max_{\boldsymbol{\beta}} \mathbf{F}(\boldsymbol{\beta})$ that provides the best separability between classes (in a Fisher sense) and yields a between-class scatter that is positive and a within-class scatter of zero value, i.e.

$$\begin{aligned} \boldsymbol{\beta}_*^\top \boldsymbol{\Sigma}_b \boldsymbol{\beta}_* &> 0 \\ \boldsymbol{\beta}_*^\top \boldsymbol{\Sigma}_w \boldsymbol{\beta}_* &= 0 \end{aligned} \quad (1)$$

The *one-class* Fisher null-space approach [30], [31] operates by adapting the Fisher null classification principle to a one-class setting. This is realised by representing the negative class by an artificial sample at the origin and utilising only positive training observations.

The standard approach to find the optimal one-class Fisher null discriminant involves solving a generalised eigenvalue problem [30], [31]. The work in [32], however, reformulates the one-class Fisher null method as a *regression* problem in the RKHS (reproducing kernel Hilbert space). A regression-based reformulation is not only favourable for bypassing the computationally demanding eigen-decomposition of dense matrices but also makes it possible to regularise the discriminant for improved performance. The one-class regression-based Fisher null algorithm [32] operates as follows. Suppose there are n positive training samples x_i 's, $i = 1, \dots, n$ and $\mathbf{v}(\mathbf{x}_i)$ denotes the RKHS feature vector for \mathbf{x}_i . If $\boldsymbol{\theta}$ is the optimal solution to

$$\min_{\boldsymbol{\theta}} \frac{1}{n} \sum_{i=1}^n (1 - \boldsymbol{\theta}^\top \mathbf{v}(\mathbf{x}_i))^2 \quad (2)$$

then for a suitably chosen $\mathbf{v}(\cdot)$ which yields a kernel matrix which is strictly positive-definite, the projection $\boldsymbol{\theta}^\top \mathbf{v}(\cdot)$ corresponds to a one-class kernel Fisher null projection. cf. [14] for a proof. The Gaussian (RBF) kernel on a dataset with no

duplicate samples provides a strictly positive-definite kernel matrix.

Eq. 2 corresponds to an 'unregularised' one-class kernel Fisher null-space classifier. While there exists other alternatives, in [32], it has been observed that imposing a Tikhonov regularisation on the Fisher null projection leads to better generalisation performance. A Tikhonov regularisation on θ is enforced as

$$\min_{\theta} \|\theta\|_2^2 + \frac{\delta}{n} \sum_{i=1}^n (1 - \theta^\top \mathbf{v}(\mathbf{x}_i))^2 \quad (3)$$

where δ is the regularisation variable. In kernel methods, a dual space representation is commonly preferred due to its convenience. The dual form of the optimisation task in Eq. 3 may be readily shown to be

$$\max_{\omega} -\sigma \omega^\top \mathbf{1} + 2\omega^\top \mathbf{1} - \omega^\top \mathbf{K} \omega \quad (4)$$

where \mathbf{K} is the kernel matrix, $\sigma = n/\delta$, and $\mathbf{1}$ represents an n -dimensional unity vector. The optimal solution to the problem in Eq. 4 is a one-class Tikhonov-regularised Fisher null projection in the RKHS and has been observed to outperform many one-class learning approaches in different settings [32].

IV. ONE-CLASS MULTIPLE KERNEL LEARNING

In this section, first, we provide a brief summary of the existing work on one-class MKL and then present our new matrix-regularised one-class multiple kernel learning algorithm.

A. Related Work on One-Class MKL

There is a large body of study concentrating on multiple kernel learning in a multi-class setting. The work in [19], [21], [24] provide a good background and taxonomy of different techniques for multi-class multiple kernel learning. Despite its importance, the *one-class* multiple kernel learning problem has been rarely addressed in the literature, except for a few exceptions. As an example, the authors in [64] propose an ℓ_1 -norm multiple kernel learning approach based on the support vector data description method. Drawing on the fact that in an SVM classifier tighter class boundaries are derived by using a larger number of support vectors, the authors also propose *slim* counterparts of their approach through modifications of the cost function so that tighter class boundaries are preferred over loose boundaries. Other study [65], addresses the multiple kernel learning task via an ℓ_1 -norm regularised formulation to encourage sparsity. The work in [25] studied multiple kernel learning in multi-class classification settings and proposed two optimisation techniques for SVM-based classification where OCC is flagged as a especial scenario. However, the authors do not provide any experimental evaluation for OCC. Other work [16] formulates an ℓ_1 -norm SVM-based multi-class MKL as a semi-infinite linear programme. The authors then discuss extensions of their formulations to OCC without carrying out any experimental analysis in a one-class setting. Other study [66], proposes a localised MKL algorithm where as an alternative to fixed kernel weights over the whole space of samples, a parametric function is deployed to assign locally

optimal weights to the kernels. The regularisation which is implicitly assumed is that of an ℓ_1 -norm regularisation which is imposed through specific gating functions associated with kernel weights.

In summary, the existing one-class MKL methods either focus on fixed-norm regularisation schemes or use a vector ℓ_p -norm constraint for learning kernel weights. In contrast, in the current study, we propose a matrix-based mixed (r, p) -norm regularisation for learning kernel weights in an OCC setting. The proposed (r, p) -norm regularisation is shown to possess the potential to lead to substantial improvements over the existing one-class MKL methods while providing a superior performance compared to other methods including end-to-end one-class deep networks.

B. The Proposed One-Class MKL Approach

Following the majority of the existing work on MKL, in this study, the composite kernel is assumed as a linear combination of multiple base kernels. In this work, kernel weights are constrained via an (r, p) -norm in a one-class setting. The mixed (r, p) -norm ($1 \leq r, p$) is a generalisation of the ordinary vector-norm to matrices. The (r, p) -norm for a matrix Φ is defined as

$$\|\Phi\|_{r,p} = \left(\sum_i \left(\sum_j |\Phi_{ij}|^r \right)^{p/r} \right)^{1/p} \quad (5)$$

where the j and i indices run over the rows and columns of matrix Φ and Φ_{ij} represents the element in the i^{th} row and j^{th} column. That is, one first applies $\|\cdot\|_r$ to each column of Φ and then applies $\|\cdot\|_p$ to the result to obtain $\|\Phi\|_{r,p}$. Given an input vector π , in order to apply a mixed (r, p) -norm on π , one may consider Φ as $\Phi = \pi \pi^\top$. As observed in a multi-class multiple kernel learning paradigm [27]–[29], a mixed-norm regularisation enables interactions between different kernels by introducing inter-kernel cross coupling terms which are absent in the ordinary vector-norm regularisation scheme. As such, a mixed-norm regularisation offers a better modelling capability to benefit from such interactions for improved performance. From a different perspective, a mixed-norm regularisation introduces additional flexibility into the model by providing further potential solution loci in the kernel weight space. In order to visualise this effect, in Fig. 1, we have plotted 2D unit ℓ_p -norm balls in the first quadrant for $p \in \{1, 2, 4, 8\}$. The green solid curves in this figure represent these balls. In the same figure, we have also plotted the unit (r, p) -norm balls for the same possible choices of r and $p \in \{1, 2, 4, 8\}$. As the (r, p) -norm encapsulates the ℓ_p -norm as a special case when $r = p$, the unit (r, p) -norm balls include the solution loci provided by an ℓ_p -norm. However, a mixed (r, p) -norm provides additional potential solutions in the kernel weight space when $r \neq p$. The balls corresponding to the case when $r \neq p$ are depicted as blue dotted curves in Fig. 1. As it is evident from the figure, a mixed (r, p) -norm regularisation provides further potential loci in the kernel weight space compared to an ℓ_p -norm regularisation, and thus, leads to an increased modelling capability which if suitably deployed may improve the performance. In the proposed multiple kernel

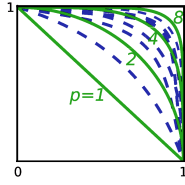


Fig. 1. The green solid curves correspond to the unit ℓ_p -norm balls (in 2D) for $p \in \{1, 2, 4, 8\}$. The unit (r, p) -norm balls for $r, p \in \{1, 2, 4, 8\}$, not only include the green solid curves, but also yield the blue dotted curves as additional and potential solution loci.

learning algorithm given J base kernels, \mathbf{K} in Eq. 4 (the kernel matrix) would be substituted by $\sum_{j=1}^J \pi_j \mathbf{K}_j$ where π_j 's represent kernel weights. Throughout the sequel, we assume that the kernels to be combined have positive contribution to the combined system. That is, the corresponding kernel weights are strictly positive: $\pi_j > \tau, \forall j$ where τ is an arbitrary but otherwise fixed small positive real number. However, for mathematical tractability, we relax the strict positivity constraint to a non-negativity constraint in the subsequent formulations. Representing \mathbf{K} as $\sum_{j=1}^J \pi_j \mathbf{K}_j$ and optimising over $\boldsymbol{\pi}$ (the kernel weights), under mixed matrix-norm regularisation and non-negativity constraints, the optimisation task for the proposed method shall be

$$\begin{aligned} \min_{\boldsymbol{\pi}} \max_{\boldsymbol{\omega}} & -\sigma \boldsymbol{\omega}^\top \boldsymbol{\omega} + 2\boldsymbol{\omega}^\top \mathbf{1} - \boldsymbol{\omega}^\top \left(\sum_{j=1}^J \pi_j \mathbf{K}_j \right) \boldsymbol{\omega} \\ \text{s.t. } & \boldsymbol{\pi} \geq 0, \quad \left\| \boldsymbol{\pi} \boldsymbol{\pi}^\top \right\|_{r,p} \leq 1 \end{aligned} \quad (6)$$

where the non-negativity constraint ensures that a valid combined kernel is obtained while the (r, p) -norm constraint introduces sparsity as well as interactions between kernels.

C. Optimisation

We assume \mathbf{u} to be a J -element vector where the j^{th} element is defined as $u_j = \boldsymbol{\omega}^\top \mathbf{K}_j \boldsymbol{\omega}$. The optimisation problem in Eq. 6 then reads

$$\begin{aligned} \min_{\boldsymbol{\pi}} \max_{\boldsymbol{\omega}} & -\sigma \boldsymbol{\omega}^\top \boldsymbol{\omega} + 2\boldsymbol{\omega}^\top \mathbf{1} - \boldsymbol{\pi}^\top \mathbf{u} \\ \text{s.t. } & \boldsymbol{\pi} \geq 0, \quad \left\| \boldsymbol{\pi} \boldsymbol{\pi}^\top \right\|_{r,p} \leq 1 \end{aligned} \quad (7)$$

For fixed $\boldsymbol{\omega}$, the objective function above is linear w.r.t. $\boldsymbol{\pi}$ and the set of constraints induced by $p \geq 1$ and $r \geq 1$ can be represented as a supremum of linear functions of $\boldsymbol{\pi} \boldsymbol{\pi}^\top$, and hence, forms a convex set [29], [67]. As a result, for fixed $\boldsymbol{\omega}$, the optimisation problem is convex in $\boldsymbol{\pi}$. On the other hand, for fixed $\boldsymbol{\pi}$, the objective function is concave in $\boldsymbol{\omega}$. Consequently, the generalised minimax theorem [68], [69] may be deployed to switch the order of optimisation without affecting the result:

$$\max_{\boldsymbol{\omega}} \left\{ -\sigma \boldsymbol{\omega}^\top \boldsymbol{\omega} + 2\boldsymbol{\omega}^\top \mathbf{1} + \min_{\boldsymbol{\pi} \in \mathcal{C}} -\boldsymbol{\pi}^\top \mathbf{u} \right\} \quad (8)$$

where $\mathcal{C} = \{\boldsymbol{\pi} \mid \boldsymbol{\pi} \geq 0, \left\| \boldsymbol{\pi} \boldsymbol{\pi}^\top \right\|_{r,p} \leq 1\}$. Eq. 8 suggests that the optimisation may be first performed in $\boldsymbol{\pi}$ and then with respect to $\boldsymbol{\omega}$.

For optimisation in $\boldsymbol{\pi}$, the Lagrangian of the minimisation subproblem may be formed as

$$\mathcal{L} = \gamma \left(\left\| \boldsymbol{\pi} \boldsymbol{\pi}^\top \right\|_{r,p} - 1 \right) - (\mathbf{u} + \boldsymbol{\mu})^\top \boldsymbol{\pi} \quad (9)$$

where the Lagrange multipliers are non-negative, i.e. $\gamma \geq 0$ and $\boldsymbol{\mu} \geq 0$. The KKT optimality criteria in this case may be written as

$$\nabla_{\boldsymbol{\pi}} \mathcal{L} = 0 \quad (10a)$$

$$\boldsymbol{\mu}^\top \boldsymbol{\pi} = 0 \quad (10b)$$

$$\boldsymbol{\pi} \geq 0 \quad (10c)$$

$$\gamma \left(\left\| \boldsymbol{\pi} \boldsymbol{\pi}^\top \right\|_{r,p} - 1 \right) = 0 \quad (10d)$$

From (10a) one obtains

$$\begin{aligned} -\mathbf{u} - \boldsymbol{\mu} + \gamma \left[\frac{\left\| \boldsymbol{\pi} \right\|_p \left\| \boldsymbol{\pi} \right\|_r}{\left\| \boldsymbol{\pi} \right\|_p^p} |\boldsymbol{\pi}|^{p-1} \odot \text{sign}(\boldsymbol{\pi}) \right. \\ \left. + \frac{\left\| \boldsymbol{\pi} \right\|_r \left\| \boldsymbol{\pi} \right\|_p}{\left\| \boldsymbol{\pi} \right\|_r^r} |\boldsymbol{\pi}|^{r-1} \odot \text{sign}(\boldsymbol{\pi}) \right] = 0 \end{aligned} \quad (11)$$

where \odot represents element-wise (Hadamard) multiplication. Using (10c) we have

$$-\mathbf{u} - \boldsymbol{\mu} + \gamma \left[\frac{\left\| \boldsymbol{\pi} \right\|_p \left\| \boldsymbol{\pi} \right\|_r}{\left\| \boldsymbol{\pi} \right\|_p^p} \boldsymbol{\pi}^{p-1} + \frac{\left\| \boldsymbol{\pi} \right\|_r \left\| \boldsymbol{\pi} \right\|_p}{\left\| \boldsymbol{\pi} \right\|_r^r} \boldsymbol{\pi}^{r-1} \right] = 0 \quad (12)$$

Due to the form of the minimisation problem it is clear that at the optimum the elements of $\boldsymbol{\pi}$ must be as large as possible. Since $\left\| \boldsymbol{\pi} \boldsymbol{\pi}^\top \right\|_{r,p}$ is convex, maximising the elements of $\boldsymbol{\pi}$ leads to the maximisation of $\left\| \boldsymbol{\pi} \boldsymbol{\pi}^\top \right\|_{r,p}$ whose optimum lies on the boundary of the feasible set \mathcal{C} specified by $\left\| \boldsymbol{\pi} \boldsymbol{\pi}^\top \right\|_{r,p} = 1$. Since $\left\| \boldsymbol{\pi} \boldsymbol{\pi}^\top \right\|_{r,p} = \left\| \boldsymbol{\pi} \right\|_p \left\| \boldsymbol{\pi} \right\|_r$ [29], at the optimum we have $\left\| \boldsymbol{\pi} \right\|_p \left\| \boldsymbol{\pi} \right\|_r = 1$. Eq. 12 may now be rewritten as

$$\mathbf{u} + \boldsymbol{\mu} = \gamma \left(\frac{\boldsymbol{\pi}^{p-1}}{\left\| \boldsymbol{\pi} \right\|_p^p} + \frac{\boldsymbol{\pi}^{r-1}}{\left\| \boldsymbol{\pi} \right\|_r^r} \right) = \gamma \boldsymbol{\pi} \odot \left(\frac{\boldsymbol{\pi}^{p-2}}{\left\| \boldsymbol{\pi} \right\|_p^p} + \frac{\boldsymbol{\pi}^{r-2}}{\left\| \boldsymbol{\pi} \right\|_r^r} \right) \quad (13)$$

Let's define the new variables $\bar{\mathbf{u}}$ and $\bar{\boldsymbol{\mu}}$ as

$$\bar{\mathbf{u}} = \mathbf{u} \odot \left(\frac{\boldsymbol{\pi}^{p-2}}{\left\| \boldsymbol{\pi} \right\|_p^p} + \frac{\boldsymbol{\pi}^{r-2}}{\left\| \boldsymbol{\pi} \right\|_r^r} \right)^{-1} \quad (14)$$

and

$$\bar{\boldsymbol{\mu}} = \boldsymbol{\mu} \odot \left(\frac{\boldsymbol{\pi}^{p-2}}{\left\| \boldsymbol{\pi} \right\|_p^p} + \frac{\boldsymbol{\pi}^{r-2}}{\left\| \boldsymbol{\pi} \right\|_r^r} \right)^{-1} \quad (15)$$

Eq. 13 may then be expressed as

$$\bar{\mathbf{u}} + \bar{\boldsymbol{\mu}} = \gamma \boldsymbol{\pi} \quad (16)$$

and thus:

$$\boldsymbol{\pi} = (\bar{\mathbf{u}} + \bar{\boldsymbol{\mu}}) / \gamma \quad (17)$$

According to (10b) it should hold that

$$\boldsymbol{\mu}^\top \boldsymbol{\pi} = \frac{1}{\gamma} \sum_{j=1}^J \mu_j (\bar{u}_j + \bar{\mu}_j) = 0 \quad (18)$$

Since $\boldsymbol{\pi} \geq 0$, $\mathbf{u} \geq 0$, $\boldsymbol{\mu} \geq 0$, we have $\bar{\mathbf{u}} \geq 0$ and $\bar{\boldsymbol{\mu}} \geq 0$. Next, we show that for Eq. 18 to hold, one must have $\boldsymbol{\mu} = 0$. For the proof, we use contradiction and assume not

Algorithm 1 Matrix-Regularised One-Class Multiple Kernel Method

```

1:  $\omega = (\sum_{j=1}^J J^{\frac{-p-r}{2rp}} \mathbf{K}_j + \sigma \mathbf{I})^{-1} \mathbf{1}$ 
2: repeat
3:    $\mathbf{u} = [\omega^\top \mathbf{K}_1 \omega, \dots, \omega^\top \mathbf{K}_J \omega]$ 
4:    $\bar{\mathbf{u}} = \mathbf{u} \odot (\frac{\pi^{p-2}}{\|\pi\|_p^p} + \frac{\pi^{r-2}}{\|\pi\|_r^r})^{-1}$ 
5:    $\pi = \bar{\mathbf{u}} / \sqrt{\|\bar{\mathbf{u}}\|_r \|\bar{\mathbf{u}}\|_p}$ 
6:    $\omega = (\sum_{j=1}^J \pi_j \mathbf{K}_j + \sigma \mathbf{I})^{-1} \mathbf{1}$ 
7: until convergence
8: Output:  $\omega$  and  $\pi$ 

```

all elements of μ are zero and $\mu_j = \epsilon > 0$ for an arbitrary index j . This assumption leads to

$$\mu^\top \pi = \frac{\epsilon}{\gamma} \left(\frac{\pi_j^{p-2}}{\|\pi\|_p^p} + \frac{\pi_j^{r-2}}{\|\pi\|_r^r} \right)^{-1} (u_j + \epsilon) > 0 \quad (19)$$

which contradicts the requirement of $\mu^\top \pi = 0$. As a result, μ cannot have any non-zero elements and hence $\mu = 0$ which leads to $\bar{\mu} = 0$. Using Eq. 17, π is then derived as

$$\pi = \bar{\mathbf{u}} / \gamma \quad (20)$$

where according to the relation $\|\pi\|_p \|\pi\|_r = 1$, we have

$$\gamma = \sqrt{\|\bar{\mathbf{u}}\|_r \|\bar{\mathbf{u}}\|_p} \quad (21)$$

Once π is determined, in order to maximise the cost function in ω , its partial derivative may be set to zero to yield:

$$\omega = \left(\sum_{j=1}^J \pi_j \mathbf{K}_j + \sigma \mathbf{I} \right)^{-1} \mathbf{1} \quad (22)$$

Note that ω in the equation above is given in terms of π which itself depends on ω . In other words, ω in Eq. 22 is expressed in terms of itself. In order to find the optimal ω , let us define $\left(\sum_{j=1}^J \pi_j \mathbf{K}_j + \sigma \mathbf{I} \right)^{-1} \mathbf{1} = f(\omega)$. The optimal ω must then satisfy $\omega = f(\omega)$. A fixed-point iteration [70] may then be applied to determine ω . The approach described above is provided in Algorithm 1 where ω is initially set to a weight vector (i.e. π) which has equal elements and a unit matrix mixed-norm.

In the proposed approach, thanks to the convexity of the problem, π is determined exactly. For optimisation w.r.t. ω , a fixed-point iteration is applied. It can be shown that for a sufficiently large regularisation parameter σ , $f(\omega) = \left(\sum_{j=1}^J \pi_j \mathbf{K}_j + \sigma \mathbf{I} \right)^{-1} \mathbf{1}$ is Lipschitz continuous with a Lipschitz constant smaller than 1, and hence, the proposed approach summarised as Algorithm 1 converges to a unique fixed point regardless of any initial guess for ω . cf. Appendix for a proof.

V. EXPERIMENTAL ASSESSMENT

The proposed one-class MKL algorithm (denoted as “ (r, p) -norm MK-FN”) is evaluated on different datasets and compared against other approaches in this section. The comparison includes both one-class kernel-based and end-to-end

deep learning-based approaches. The multiple-kernel-based approaches in the comparison are constructed using the “product” and “average” rules for kernel fusion (corresponding respectively to the geometric and arithmetic mean of kernel matrices) applied to the Gaussian process method (GP) [71], the Fisher null approach (FN) [30], [31], and to the kernel principal component analysis for OCC (KPCA) [72]. In addition, the SVDD-based multiple kernel learning approach (MK-SVDD) [64], the multiple kernel learning one-class SVM algorithm (MK-OCSVM) [65] and their ‘slim’ variants [64] denoted as Slim-MK-SVDD and Slim-MK-OCSVM are included in the comparisons. Moreover, we include state-of-the-art one-class deep learning approaches in the comparisons. The rest of this section is arranged as follows.

- In Section V-A, we provide the implementation details.
- The convergence behaviour of the proposed approach is analysed in Section V-B.
- In Section V-C, the proposed method is examined for abnormality detection on the Abnormality-1001 dataset [73] and for novelty detection on the Caltech256 dataset [74].
- The experimental results of an assessment of the proposed matrix (r, p) -norm one-class MKL method for “unseen” face presentation attack detection on the Replay-Mobile [75], Oulu-NPU [76], MSU-MFSD [77] and Replay-Attack [78] databases are presented and discussed in Section V-D.

A. Implementation Details

In the following experiments, the regularisation parameter σ in the proposed method is selected from $\{10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 10^2\} \times n$ where n denotes the number of positive training observations. We use a Gaussian kernel function to form the kernel matrices. The width of the Gaussian kernel is selected from $\{\frac{1}{4}M, \frac{1}{2}M, M\}$ where M corresponds to the average over all pairwise Euclidean distances among all positive training observations. Parameters r and p are selected from $\{32/31, 16/15, 8/7, 4/3, 2, 4, 8, 10\}$. The parameters of the proposed method are set on a separate validation set. For the SVM-based MKL approaches, the parameters are set on the validation set as suggested in [64].

B. Convergence Characteristics

The convergence characteristics of the proposed MKL approach is analysed in this section. To this end, we randomly select a single class from the Caltech256 object dataset [74] and identify it as the target class. We then use Algorithm 1 to learn kernel weights corresponding to seven deep CNN features derived using the pre-trained Resnet50 [79], Googlenet [80], Alexnet [81], Vgg16 [82], Densenet201 [83], Mobilenetv2 [84] and Nasnetlarge [85]. We repeat the experiment 100 times for a number of different combinations of (r, p) , namely for $(r, p) \in \{(32/31, 32/31), (4/3, 4/3), (8, 8), (32/31, 8)\}$, each time initialising π to a random vector with an (r, p) -norm equal to one. We define the error as the l_2 -norm of the change in ω in the course of optimisation.

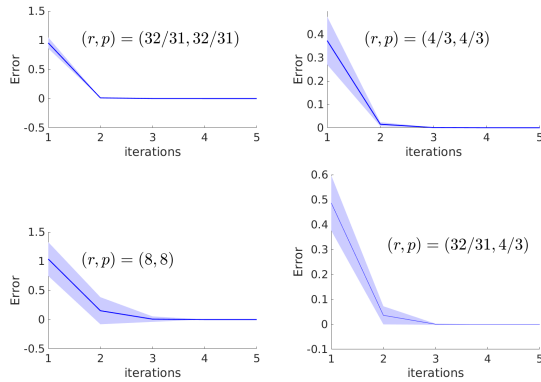


Fig. 2. Convergence curves of the proposed approach for a sample one-class MKL problem for different (r, p) -norm regularisations.

A zero change represents convergence. The results of this experiment are depicted in Fig. 2. From Fig. 2, it may be seen that the proposed approach typically converges in 5 iterations regardless of the regularisation imposed. It is worth noting that similar behaviour has been observed for other (r, p) combinations.

C. General Object Image One-Class Classification

An assessment of the proposed MKL approach for abnormality and novelty detection on different databases is presented in this section.

1) *Abnormality Detection*: One of the frequently used databases for abnormality detection is that of 1001 Abnormal Objects dataset [73] comprised of 1001 images of 6 different object categories from the PASCAL dataset [86]. On this dataset, the task is to label each image as abnormal or normal. However, since the pattern of the abnormality is not known ahead of time, training is conducted on observations from the target/normal class only. We build seven kernel matrices based on the representations obtained from the pre-trained Resnet50 [79], Googlenet [80], Alexnet [81], Vgg16 [82], Densenet201 [83], Mobilenetv2 [84] and Nasnetlarge [85]. In order to enable a fair comparison with other approaches, the protocol proposed in [73] is followed for evaluation. In this experiment, we not only include other kernel fusion methods but also consider the state-of-the-art approaches from the literature inclusive of deep one-class methods. A comparison of different approaches in terms of AUC (Area Under the ROC Curve) on this database is provided in Table I. From this table it may be verified that the proposed matrix-norm MKL method improves over the previous best reported result by a large margin. The previous best reported performance on this dataset corresponds to the work in [87] with an average AUC of 95.6% which is much inferior compared to the proposed method with an AUC of 99.2%. Compared to fixed kernel fusion rules, the proposed MKL learning algorithm also provides substantial improvements. The SVM-based MKL approaches also perform worst than the proposed method. In particular, the best SVM-based MKL approach on this dataset is the MK-SVDD-Slim with an average AUC of 94.1% whose performance is more than 5% worst than the performance of the proposed MKL algorithm. In summary, the proposed

TABLE I
COMPARISON OF DIFFERENT OCC APPROACHES FOR ABNORMALITY DETECTION ON THE ABNORMALITY-1001 DATASET

Method	AUC (MEAN±STD%)
Product-FN	90.1 ± 0.3
Average-FN	94.7 ± 0.2
Product-GP	84.4 ± 0.3
Average-GP	93.5 ± 0.2
Product-KPCA	87.3 ± 0.3
Average-KPCA	91.6 ± 0.2
MK-SVDD	92.3 ± 0.2
MK-OCSVM	91.4 ± 0.3
MK-SVDD-Slim	94.1 ± 0.1
MK-OCSVM-Slim	93.9 ± 0.1
Graphical Model [74]	87.0 ± <i>n.a.</i>
Adj. Graphical Model [74]	91.1 ± <i>n.a.</i>
OCNN [89]	88.5 ± 1.4
Autoencoder [90]	67.4 ± 1.2
OC-CNN [91]	84.3 ± <i>n.a.</i>
DOC [88]	95.6 ± 3.1
This work	99.2 ± 0.2

(r, p) -norm one-class MKL method not only performs better than fixed-rule and SVM-based multiple kernel systems but also outperforms the state-of-the-art end-to-end deep learning methods.

2) *Novelty Detection*: In novelty detection, one is interested in quantifying the novelty of a test sample based on the observations previously enrolled to the system. Since the typical characteristics of a novel observation are not available a priori, the training is very often based on only positive samples using one-class classification techniques. The Caltech 256 dataset [74] is one of the commonly employed databases for novelty detection that encapsulates images of objects from 256 categories for a total of 30607 samples. Similar to the previous experiment on abnormality detection, using the aforementioned seven pre-trained deep convolutional networks, we construct seven kernels matrices. In order to perform a fair comparison against the existing techniques, the protocol introduced in [91] is followed where each class is considered as the target/normal category and the rest as novel observations. For the first 40 classes of the database the experiment is repeated and the performance is measured in terms of the area under the ROC curve (AUC). The results corresponding to this experiment are tabulated in Table II where we have included both multiple kernel and deep learning methods. The following observations from Table II may be made. First, thanks to a multiple kernel representation, all the multiple kernel methods outperform end-to-end deep learning methods by a large margin. Second, compared to fixed-rule multiple kernel approaches, the proposed (r, p) -norm MKL method performs better. Third, the proposed MKL algorithm performs better than other MKL alternatives based on an SVM formulation. An last but not the least, the classification performance of the proposed method is better than the state-of-the-art one-class deep learning methods in the literature. In this context, the best performing deep OCC method is that of DOC-VGG16 [87] with an average AUC of 98.1% compared to the proposed approach with an AUC of 99.6%.

TABLE II
COMPARISON OF DIFFERENT OCC APPROACHES FOR NOVELTY
DETECTION ON THE CALTECH256 DATASET

Method	AUC (MEAN±STD%)
Product-FN	99.3 ± 0.6
Average-FN	99.3 ± 0.6
Product-GP	99.3 ± 0.6
Average-GP	99.3 ± 0.6
Product-KPCA	99.3 ± 0.8
Average-KPCA	99.2 ± 0.8
MK-SVDD	99.1 ± 0.7
MK-OCSVM	99.1 ± 0.8
MK-SVDD-Slim	99.2 ± 0.6
MK-OCSVM-Slim	99.1 ± 0.8
OCNN-AlexNet [89]	82.6 ± 15.3
OCNN-VGG16 [89]	88.5 ± 14.4
Autoencoder [90]	62.3 ± 7.2
OCSVM-VGG16 [88]	90.2 ± 5.0
DOC-VGG16 [88]	98.1 ± 2.2
DOC-AlexNet [88]	95.9 ± 2.1
This work	99.6 ± 0.5

D. Unseen Face Presentation Attack Detection

An assessment of the proposed method for face PAD in an unseen PA setting is conducted in this section. The databases utilised for this purpose are as follows.

1) *The OULU-NPU Database [76]*: incorporates 4950 genuine and attack video samples from 55 individuals captured with 6 different devices in three sessions under different illuminations and background settings. The data incorporates previously unseen acquisition conditions, attack types as well as input sensors. The video sequences are divided into 3 subject-disjoint sets for training, development and testing. For evaluation, four different protocols are introduced where the forth protocol is known to be the most challenging one which is used in the current work.

2) *The Replay-Mobile Dataset [75]*: contains 1190 video sequences of both attack and bona fide (genuine) data corresponding to 40 subjects which are recorded using two different devices in different illumination settings. Three disjoint subdivisions for training, development and testing in addition to an enrolment set exist in this database.

3) *The Replay-Attack Dataset [78]*: provides 1300 video sequences of attack and genuine data from 50 subjects. Attacks are generated using a high definition iPad screen, a mobile phone or a printed image. Three randomly divided subject-disjoint subsets for training, development and testing are available in this database.

4) *The MSU-MFSD Dataset [77]*: includes 440 video sequences captured from either photo or video attack attempts from 55 subjects that are recorded by two different recording devices. The publicly available subset of this dataset, provides data from 35 individuals. The database is divided into two partitions for training and testing which are subject-disjoint.

The standard ISO metrics for measuring the performance of a PAD system are [92]: 1) attack presentation classification error rate (APCER) that corresponds to the ratio of misclassified attack presentations using the same presentation

attack instrument species; and 2) bona fide presentation classification error rate (BPCER) that represents the misclassified percentage of bona fide presentations. For performance reporting, the highest APCER over all PAIS's (presentation attack instrument species) is used:

$$APCER = \max_{PAIS} APCER_{PAIS} \quad (23)$$

The all-inclusive performance of a PAD system can be expressed as ACER (the Average Classification Error Rate):

$$ACER = \left(BPCER + \max_{PAIS} APCER_{PAIS} \right) / 2 \quad (24)$$

In order to enable a comparison to the existing methods in the literature, the performance of the proposed approach is also gauged in terms of Half Total Error Rate (HTER), and the AUC (the Area Under the ROC Curve).

In this work, we use the features suggested in [14] to construct multiple kernel matrices as they have been found to be useful for face PAD. Moreover, using a similar set of features enables a fair comparison to other similar multiple kernel methods. These features correspond to deep representations obtained using the pre-trained VGG16 [82], ResNet50 [79] and GoogleNet [80] extracted from four facial regions giving rise to a total of 12 kernels. The regions correspond to the whole face, eyes and the nose region, nose and surroundings and the regions around the nose and the mouth. In all the following experiments, we follow a client-specific modelling approach as advocated in [59] and evaluate the proposed approach in an unseen attack setting using “only” genuine (bona fide) data for training. As a client-specific modelling approach is pursued, for each test subject, the data from all the other individuals serves as the validation data to tune the parameters. The results corresponding to this experiment are tabulated in Table III, IV, V and VI for the Oulu-NPU, Replay-Attack, MSU-MSFD and Replay-Attack datasets, respectively. In addition to the OCC unseen face PAD approaches from the literature, 10 different multiple kernel systems introduced earlier are evaluated as baseline methods. For a fair comparison, all multiple kernel methods are fed with similar features as that of the proposed MKL method. Deep one-class face PAD methods from the literature are also included in the comparison.

a) *Summary of detection performances*: Based on the performances reported in tables III, IV, V and VI, a number of observations may be made. First, on the Oulu-NPU and Replay-Mobile datasets which are relatively more challenging datasets, the proposed matrix-norm MKL Fisher null method clearly demonstrates its advantage. In this respect, on the Oulu-NPU dataset, while fix kernel fusion rules (i.e. geometric and arithmetic mean) do provide reasonable results, yet, the proposed matrix-norm MK-FN approach provides an outstanding ACER of 2.5 ± 2.2 compared to the best fix fusion rule (Product-KPCA and Product-FN) with an ACER of 4.5 ± 5.3 . It is worth noting that neither one of the MKL methods based on the one-class SVM, does not provide any advantage compared to the examined fixed-rule kernel fusion methods in Table III. In comparison with the best reported result in the literature which is due to the OCA-FAS

TABLE III

COMPARISON OF DIFFERENT APPROACHES FOR THE UNSEEN FACE PRESENTATION ATTACK DETECTION ON PROTOCOL IV OF THE OULU-NPU DATASET (MEAN \pm STD %)

Method	BPCER	APCER	ACER
Product-FN	2.5 \pm 4.1	6.6 \pm 6.8	4.5 \pm 5.3
Average-FN	0.8 \pm 2.0	9.2 \pm 7.3	5.0 \pm 3.9
Product-GP	2.5 \pm 4.2	9.2 \pm 9.7	5.8 \pm 6.4
Average-GP	1.7 \pm 4.1	10.8 \pm 8.0	6.2 \pm 4.4
Product-KPCA	2.5 \pm 4.2	6.6 \pm 6.8	4.5 \pm 5.3
Average-KPCA	0.0 \pm 0.0	10.8 \pm 7.3	5.4 \pm 3.6
MK-SVDD	2.5 \pm 4.2	11.7 \pm 10.8	7.1 \pm 6.2
MK-OCSVM	2.5 \pm 4.2	13.3 \pm 11.7	7.9 \pm 6.4
Slim-MK-SVDD	1.7 \pm 4.1	10.8 \pm 8.0	6.2 \pm 4.4
Slim-MK-OCSVM	1.7 \pm 4.1	10.8 \pm 8.0	6.2 \pm 4.4
SAPLC [94]	6.6 \pm 5.5	11.9 \pm 6.9	9.3 \pm 4.4
OCA-FAS [95]	5.9 \pm 4.5	2.3 \pm 2.5	4.1 \pm 2.7
The work in [96]	1.7 \pm 2.6	5.8 \pm 4.9	3.7 \pm 2.1
The work in [97]	9.7 \pm 4.8	11.3 \pm 3.9	9.8 \pm 4.2
The work in [14]	0.8 \pm 2.0	11.6 \pm 13.6	6.2 \pm 6.8
This work	0 \pm 0	5.0 \pm 4.5	2.5 \pm 2.2

TABLE IV

COMPARISON OF DIFFERENT APPROACHES FOR THE UNSEEN FACE PRESENTATION ATTACK DETECTION ON THE REPLAY-MOBILE DATASET (HALF TOTAL ERROR RATE (HTER) %)

Method	HTER
Product-FN	8.8
Average-FN	7.3
Product-GP	8.8
Average-GP	7.6
Product-KPCA	9.8
Average-KPCA	9.1
MK-SVDD	11.1
MK-OCSVM	10.8
Slim-MK-SVDD	7.6
Slim-MK-OCSVM	8.1
The work in [59]	8.5
Weighted-Fusion [61]	9.9
GoogleNet+MD [4]	13.7
The work in [14]	11.8
This work	6.7

method [94] with an ACER of 4.1 ± 2.7 , the proposed method also yields a better detection performance.

On the Replay-Mobile database, while the proposed matrix-norm MKL algorithm achieves a HTER of 6.7%, the best fixed-rule kernel fusion system of Average-FN achieves a HTER of 7.3% which underlines the effectiveness of the proposed matrix-norm MK-FN method. Similar to the Oulu-NPU dataset, on the Replay-Mobile dataset neither one of the MKL methods based on SVM does not provide any performance gain compared to the fixed-rule multiple kernel learning methods examined. In comparison with the best reported performance in the literature which is due to the method in [59] with a HTER of 8.5%, the proposed matrix-regularised multiple kernel Fisher null method also performs better.

On the MSU-MFSD and Replay-Attack datasets, almost all the multiple kernel systems including the fixed-rule kernel

TABLE V

COMPARISON OF DIFFERENT APPROACHES FOR THE UNSEEN FACE PRESENTATION ATTACK DETECTION ON THE MSU-MFSD DATASET (AREA UNDER THE ROC CURVE (AUC) %)

Method	AUC
Product-FN	100
Average-FN	100
Product-GP	100
Average-GP	100
Product-KPCA	100
Average-KPCA	100
MK-SVDD	100
MK-OCSVM	100
Slim-MK-SVDD	100
Slim-MK-OCSVM	100
IMQ+OCSVM [1]	67.7
BSIF+OCSVM [1]	75.6
LBP+NN [3]	81.6
LBP+GMM [3]	81.3
LBP+OCSVM [3]	84.5
LBP+AE [3]	87.6
DTL [5]	93.0
The work in [14]	100
This work	100

TABLE VI

COMPARISON OF DIFFERENT APPROACHES FOR THE UNSEEN FACE PRESENTATION ATTACK DETECTION ON THE REPLAY-ATTACK DATASET (AREA UNDER THE ROC CURVE (AUC) %)

Method	AUC
Product-FN	100
Average-FN	100
Product-GP	100
Average-GP	100
Product-KPCA	99.9
Average-KPCA	100
MK-SVDD	100
MK-OCSVM	100
Slim-MK-SVDD	100
Slim-MK-OCSVM	100
IMQ+OCSVM [1]	80.7
BSIF+OCSVM [1]	81.9
LBP+NN [3]	91.2
LBP+GMM [3]	90.1
LBP+OCSVM [3]	87.9
LBP+AE [3]	86.1
DTL [5]	99.8
MD [4]	99.7
The work in [14]	100
This work	100

fusion systems and the proposed (r, p) -norm MK-FN approach achieve a perfect performance which emphasises the utility if a multiple kernel system.

b) Discussion: Compared with the existing multiple kernel or deep end-to-end approaches, the proposed method obtains a better performance. The superior performance of the proposed approach for the unseen zero-shot face PAD problem may be justified as follows.

In comparison to the existing one-class multiple kernel learning algorithms, the proposed one-class MKL algorithm

operates based on the Fisher classification principle whereas the existing one-class MKL algorithms are based on an SVM formulation. The superiority of a Fisher-based one-class classification framework as compared with an SVM formulation may be verified by comparing the classification performance of the fixed fusion rules applied to the Fisher null method with those of the SVM-based MKL algorithms. More importantly, the proposed multiple kernel learning algorithm infers optimal kernel weights for a kernel fusion subject to a matrix-norm constraint which, as discussed previously, offers higher flexibility to the MKL procedure and also enables inter-kernel interactions. This is in contrast to the existing one-class MKL algorithms which are limited to vector-norm regularisation constraints.

When compared to the existing end-to-end one-class deep networks, the proposed one-class MKL algorithm performs better since a multiple kernel learning method benefits from an optimal combination of multiple representations whereas the existing OCC approaches typically train a single network to yield a ‘single representation’ for classification. An optimal combination of multiple representations possesses a higher capacity to lead to a better classification performance, as confirmed via experiments on multiple datasets.

E. Computational Complexity

The computationally dominant component of the proposed matrix-norm one-class MKL algorithm is step 3 in Algorithm 1 performing a matrix-vector multiplication that incurs a time complexity of $\mathcal{O}(Jn^2)$. In addition, a naïve calculation of the inverse matrix for step 6 of Algorithm 1 leads to a time complexity of $\mathcal{O}(n^3)$. However, a matrix inversion operation can be performed in $\mathcal{O}(n^2)$ time benefiting from the incremental Cholesky decomposition and the Sherman’s march algorithm [97], [98].

One particularly appealing attribute of the proposed MKL approach is that of parallelizability. In this respect, the matrix-vector multiplications may be computed by benefiting from parallel processing units to yield large speed-ups. In a similar fashion, a parallel implementation of matrix inversion is applicable to obtain significant improvements in the running time [99], [100]. In order to illustrate this, we have measured the CPU and GPU timings for the vector-matrix-vector multiplication (step 3 of Algorithm 1) and for the matrix inversion operation (step 6 of Algorithm 1) for different numbers of training samples on Matlab R2021a. The results are tabulated in Table VII and Table VIII for a machine with 64-bit 4GHz CPU, 32 GB memory and with a GeForce GTX 1080Ti GPU operating on Windows 10.

As may be observed from Table VII, for the vector-matrix-vector multiplication (step 3 of the proposed algorithm) more than 24 times speed-up again may be achieved by porting the operations onto a GPU. The speed-up gain may be more than 100 times when the number of training observations is 5000 or more. Regarding the matrix inversion operation (Table VIII), the relative speed-up gain achieved is more than 8 times when the number of training samples 1000. The speed-up gain would be around 3 times when increasing the number of training samples towards 5000, 10000 or to 15000.

TABLE VII

COMPARISON OF CPU AND GPU TIMINGS (IN SECONDS) FOR VECTOR-MATRIX-VECTOR MULTIPLICATION (STEP 3 OF THE PROPOSED ALGORITHM)

no. of training samples	n=1000	n=5000	n=10000	n=15000
Elapsed time on CPU	0.0037	0.0845	0.3367	0.7420
Elapsed time on GPU	1.56e-04	7.67e-04	0.0025	0.0054
Speed-up gain	24.20	110.17	134.68	137.41

TABLE VIII

COMPARISON OF CPU AND GPU TIMINGS (IN SECONDS) FOR MATRIX INVERSION OPERATION (STEP 6 OF THE PROPOSED ALGORITHM)

no. of training samples	n=1000	n=5000	n=10000	n=15000
Elapsed time on CPU	0.0289	0.2553	0.9830	2.2178
Elapsed time on GPU	0.0035	0.0851	0.3131	0.7036
Speed-up gain	8.12	2.99	3.13	3.15

F. Remarks

- In the current study, as the main objective was to demonstrate the efficacy of the proposed one-class MKL algorithm to improve the performance of a multiple kernel system for face PAD, we fed the proposed approach with similar representations as those of previous studies [14] to accurately gauge any performance benefits brought by the proposed MKL algorithm. Nevertheless, one may consider a richer pool of representations for improved detection performance as future research directions.
- While we presented a zero-shot face PAD algorithm, the proposed approach can be generalised to benefit from any seen attacks by constructing a separate one-class learner for each different attack type. In this case, a test sample may be either classified as bona fide, or as one of the previously seen types of attack or as an unseen attack.
- The matrix-regularised MKL approaches presented in [27]–[29] require training data from all the classes, and thus, cannot be applied to the zero-shot one-class face PAD problem considered in the current study. In this context, the proposed method in this work is innovatively and deliberately designed to fill this gap by being operable in a one-class classification setting, i.e. be trainable using only samples from a single class to be applicable to the zero-shot unseen face PAD setting. Please note that, although one may consider a “multi-class” extension of the proposed matrix-regularised MKL algorithm, nevertheless, such a formulation is not desired as a multi-class approach neither fits the evaluation settings of the problem addressed in the current study nor is it the preferred approach for the unseen face PAD problem as observed in other studies [1], [3], [5], [14], [61], [93]–[96].

VI. CONCLUSION

The face presentation attack detection problem in an unseen zero-shot attack setting was addressed. To this end and motivated by the success of multiple kernel methods, a matrix-regularised one-class MKL algorithm was presented.

We formulated the corresponding MKL problem as an optimisation task and presented an effective method to find the optimal solution. Through comprehensive evaluations on different datasets, including general object image and four different face PAD datasets, it was illustrated that the proposed approach not only outperforms fixed-rule and SVM-based MKL methods but also performs better than the existing one-class end-to-end deep learning approaches, supporting the efficacy of a mixed-norm multiple kernel learning method for the unseen face PAD problem.

APPENDIX

CONVERGENCE OF THE PROPOSED ONE-CLASS MATRIX-NORM MKL ALGORITHM

A fixed-point iteration is convergent if the function is Lipschitz continuous with a Lipschitz constant which is smaller than one [70]. Hence, for convergence, it is sufficient to demonstrate that for a sufficiently large regularisation parameter σ , we have

$$\|f(\omega^{m+1}) - f(\omega^m)\|_2 < \|\omega^{m+1} - \omega^m\|_2 \quad (25)$$

where $f = \left(\sum_{j=1}^J \pi_j \mathbf{K}_j + \sigma \mathbf{I}\right)^{-1} \mathbf{1}$ and ω^m and ω^{m+1} represent two consecutive discriminants generated at iterations m and $m+1$, respectively. Let us assume $\|\omega^{m+1} - \omega^m\|_2 = \kappa$. If $\kappa = 0$, then the sequence ω^m is converged. As a result, let us analyse the case $\kappa > 0$. Considering the LHS in Eq. 25, we have

$$\begin{aligned} & \|f(\omega^{m+1}) - f(\omega^m)\|_2 \\ & \leq \|f(\omega^{m+1})\|_2 + \|f(\omega^m)\|_2 \\ & \leq \sqrt{n} \left\| \left(\sum_{j=1}^J \pi_j^{m+1} \mathbf{K}_j + \sigma \mathbf{I} \right)^{-1} \right\|_2 \\ & \quad + \sqrt{n} \left\| \left(\sum_{j=1}^J \pi_j^m \mathbf{K}_j + \sigma \mathbf{I} \right)^{-1} \right\|_2 \end{aligned} \quad (26)$$

where the first inequality is based on the triangle inequality and the second is due to the fact that the norm of a product of a matrix and a vector is smaller than/equal to the product of their norms [97]. Furthermore, the l_2 -norm of an inverted matrix is $1/z$ where z is the smallest eigenvalue of the original matrix [97]. As a result

$$\|f(\omega^{m+1}) - f(\omega^m)\|_2 \leq \sqrt{n}(1/z_1 + 1/z_2) \quad (27)$$

where z_1 and z_2 denote the minimum eigenvalues of $\left(\sum_{j=1}^J \pi_j^{m+1} \mathbf{K}_j + \sigma \mathbf{I}\right)$ and $\left(\sum_{j=1}^J \pi_j^m \mathbf{K}_j + \sigma \mathbf{I}\right)$, respectively. It is known that adding a constant to the main diagonal shifts the eigenvalues of a matrix by the same value. As a result, if the minimum eigenvalues of $\sum_{j=1}^J \pi_j^{m+1} \mathbf{K}_j$ and $\sum_{j=1}^J \pi_j^m \mathbf{K}_j$ are z'_1 and z'_2 , respectively, it holds that $z_1 = z'_1 + \sigma$ and $z_2 = z'_2 + \sigma$. In the worst case, $z'_1 = 0$ and $z'_2 = 0$, and hence, $z_1 = \sigma$ and $z_2 = \sigma$ and consequently

$$\|f(\omega^{m+1}) - f(\omega^m)\|_2 \leq 2\sqrt{n}/\sigma \quad (28)$$

By using the upper bound in Eq. 28 in Eq. 25, the sufficient condition for convergence is obtained as

$$\sigma > \frac{2\sqrt{n}}{\kappa} \quad (29)$$

By choosing σ sufficiently large, the above inequality is satisfied, and hence, $f(\omega)$ would be Lipschitz continuous with a Lipschitz constant smaller than one and as a result the proposed one-class MKL algorithm would be convergent. In the discussion above, we analysed the worst case scenario where the combined kernel matrices were not strictly positive-definite. However, by choosing a Gaussian kernel function for a duplicate-free training set, the combined kernel matrix will be strictly positive definite (i.e. $z'_1 > 0$ and $z'_2 > 0$). In this case, the lower bound for σ that would lead to convergence would be smaller than that of Eq. 29. ■

REFERENCES

- [1] S. R. Arashloo, J. Kittler, and W. Christmas, "An anomaly detection approach to face spoofing detection: A new formulation and evaluation protocol," *IEEE Access*, vol. 5, pp. 13868–13882, 2017.
- [2] O. Nikisins, A. Mohammadi, A. Anjos, and S. Marcel, "On effectiveness of anomaly detection approaches against unseen presentation attacks in face anti-spoofing," in *Proc. Int. Conf. Biometrics (ICB)*, Feb. 2018, pp. 75–81.
- [3] F. Xiong and W. Abdalmegeed, "Unknown presentation attack detection with face RGB images," in *Proc. IEEE 9th Int. Conf. Biometrics Theory, Appl. Syst. (BTAS)*, Oct. 2018, pp. 1–9.
- [4] S. Fatemifar, S. R. Arashloo, M. Awais, and J. Kittler, "Spoofing attack detection by anomaly detection," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2019, pp. 8464–8468.
- [5] Y. Liu, J. Stehouwer, A. Jourabloo, and X. Liu, "Deep tree learning for zero-shot face anti-spoofing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2019, pp. 4675–4684.
- [6] D. Perez-Cabo, D. Jimenez-Cabello, A. Costa-Pazo, and R. J. Lopez-Sastre, "Deep anomaly detection for generalized face anti-spoofing," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2019, pp. 1591–1600.
- [7] A. Rattani and A. Ross, "Automatic adaptation of fingerprint liveness detector to new spoof materials," in *Proc. IEEE Int. Joint Conf. Biometrics*, Sep. 2014, pp. 1–8.
- [8] A. Rattani, W. J. Scheirer, and A. Ross, "Open set fingerprint spoof detection across novel fabrication materials," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 11, pp. 2447–2460, Nov. 2015.
- [9] Y. Ding and A. Ross, "An ensemble of one-class SVMs for fingerprint spoof detection across different fabrication materials," in *Proc. IEEE WIFS*, Dec. 2016, pp. 1–6.
- [10] P. M. Ferreira, A. F. Sequeira, D. Pernes, A. Rebelo, and J. S. Cardoso, "Adversarial learning for a robust iris presentation attack detection method against unseen attack presentations," in *Proc. Int. Conf. Biometrics Special Interest Group (BIOSIG)*, Sep. 2019, pp. 1–7.
- [11] M. Sahidullah *et al.*, *Introduction to Voice Presentation Attack Detection and Recent Advances*. Cham, Switzerland: Springer, 2019, pp. 321–361.
- [12] V. Chandola, A. Banerjee, and V. Kumar, "Anomaly detection: A survey," *ACM Comput. Surv.*, vol. 41, no. 3, pp. 15:1–15:58, Jul. 2009.
- [13] S. S. Khan and M. G. Madden, "One-class classification: Taxonomy of study and review of techniques," *Knowl. Eng. Rev.*, vol. 29, no. 3, pp. 345–374, Jan. 2014.
- [14] S. R. Arashloo, "Unseen face presentation attack detection using sparse multiple kernel Fisher null-space," *IEEE Trans. Circuits Syst. Video Technol.*, early access, Dec. 22, 2020, doi: 10.1109/TCSVT.2020.3046505.
- [15] G. R. G. Lanckriet, T. D. Bie, N. Cristianini, M. I. Jordan, and W. S. Noble, "A statistical framework for genomic data fusion," *Bioinformatics*, vol. 20, no. 16, pp. 2626–2635, 2004.
- [16] S. Sonnenburg, G. Rätsch, C. Schäfer, and B. Schölkopf, "Large scale multiple kernel learning," *J. Mach. Learn. Res.*, vol. 7, vol. 57, pp. 1531–1565, Jul. 2006. [Online]. Available: <http://jmlr.org/papers/v7/sonnenburg06a.html>

- [17] C. Cortes, M. Mohri, and A. Rostamizadeh, "Learning non-linear combinations of kernels," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 22, Y. Bengio, D. Schuurmans, J. Lafferty, C. Williams, and A. Culotta, Eds. Red Hook, NY, USA: Curran Associates, 2009, pp. 396–404.
- [18] Y. Gu, T. Liu, X. Jia, and J. A. Benediktsson, "Nonlinear multiple kernel learning with multiple-structure-element extended morphological profiles for hyperspectral image classification," *IEEE Trans. Geosci. Remote Sens.*, vol. 54, no. 6, pp. 3235–3247, Jun. 2016.
- [19] M. Gönen and E. Alpaydm, "Multiple kernel learning algorithms," *J. Mach. Learn. Res.*, vol. 12, pp. 2211–2268, Jul. 2011. [Online]. Available: <http://jmlr.org/papers/v12/gonen11a.html>
- [20] Q. Mao, I. W. Tsang, S. Gao, and L. Wang, "Generalized multiple kernel learning with data-dependent priors," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 26, no. 6, pp. 1134–1148, Jun. 2015.
- [21] S. S. Bucak, R. Jin, and A. K. Jain, "Multiple kernel learning for visual object recognition: A review," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 7, pp. 1354–1369, Jul. 2014.
- [22] H. Xia, S. C. H. Hoi, R. Jin, and P. Zhao, "Online multiple kernel similarity learning for visual search," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 3, pp. 536–549, Mar. 2014.
- [23] Y.-Y. Lin, T.-L. Liu, and C.-S. Fuh, "Multiple kernel learning for dimensionality reduction," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 6, pp. 1147–1160, Jun. 2011.
- [24] F. Yan, J. Kittler, K. Mikolajczyk, and A. Tahir, "Non-sparse multiple kernel Fisher discriminant analysis," *J. Mach. Learn. Res.*, vol. 13, no. 1, pp. 607–642, Mar. 2012. [Online]. Available: <http://jmlr.org/papers/v13/yan12a.html>
- [25] M. Kloft, U. Brefeld, S. Sonnenburg, and A. Zien, " ℓ_p -norm multiple kernel learning," *J. Mach. Learn. Res.*, vol. 12, no. 26, pp. 953–997, 2011. [Online]. Available: <http://jmlr.org/papers/v12/kloft11a.html>
- [26] G. R. G. Lanckriet, N. Cristianini, P. Bartlett, L. E. Ghaoui, and M. I. Jordan, "Learning the kernel matrix with semidefinite programming," *J. Mach. Learn. Res.*, vol. 5, pp. 27–72, Jan. 2004. [Online]. Available: <http://www.jmlr.org/papers/v5/lanckriet04a.html>
- [27] C. Hinrichs, V. Singh, J. Peng, and S. Johnson, "Q-MKL: Matrix-induced regularization in multi-kernel learning with applications to neuroimaging," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 25, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Red Hook, NY, USA: Curran Associates, 2012, pp. 1430–1438.
- [28] Y. Han, K. Yang, Y. Yang, and Y. Ma, "Localized multiple kernel learning with dynamical clustering and matrix regularization," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 2, pp. 486–499, Feb. 2018.
- [29] Y. Han, Y. Yang, X. Li, Q. Liu, and Y. Ma, "Matrix-regularized multiple kernel learning via (r, p) norms," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 29, no. 10, pp. 4997–5007, Oct. 2018.
- [30] J. Liu, Z. Lian, Y. Wang, and J. Xiao, "Incremental kernel null space discriminant analysis for novelty detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 4123–4131.
- [31] P. Bodesheim, A. Freytag, E. Rodner, M. Kemmler, and J. Denzler, "Kernel null space methods for novelty detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 3374–3381.
- [32] S. R. Arashloo and J. Kittler, "Robust one-class kernel spectral regression," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 32, no. 3, pp. 999–1013, Mar. 2021.
- [33] T. Edmunds, "Protection of 2D face identification systems against spoofing attacks," M.S. thesis, Grenoble GIPSA-LAB, School Electron., Elect. Eng., Automat. Signal Process., Univ. Grenoble Alpes, Grenoble, France, Jan. 2017.
- [34] S. Bhattacharjee, A. Mohammadi, A. Anjos, and S. Marcel, *Recent Advances in Face Presentation Attack Detection*. Cham, Switzerland: Springer, 2019, pp. 207–228.
- [35] R. Ramachandra and C. Busch, "Presentation attack detection methods for face recognition systems: A comprehensive survey," *ACM Comput. Surveys*, vol. 50, no. 1, pp. 1–37, Apr. 2017.
- [36] S. Marcel, M. S. Nixon, J. Fierrez, and N. W. D. Evans, Eds., *Handbook Biometric Anti-Spoofing—Presentation Attack Detection* (Advances in Computer Vision and Pattern Recognition), 2nd ed. Cham, Switzerland: Springer, 2019.
- [37] K. Ma, Q. Xu, J. Zeng, X. Cao, and Q. Huang, "Poisoning attack against estimating from pairwise comparisons," *IEEE Trans. Pattern Anal. Mach. Intell.*, early access, Jun. 8, 2021, doi: 10.1109/TPAMI.2021.3087514.
- [38] J. Bai, B. Wu, Y. Zhang, Y. Li, Z. Li, and S.-T. Xia, "Targeted attack against deep neural networks via flipping limited weight bits," in *Proc. Int. Conf. Learn. Represent.*, 2021.
- [39] Z. Boulkenafet, J. Komulainen, and A. Hadid, "On the generalization of color texture-based face anti-spoofing," *Image Vis. Comput.*, vol. 77, pp. 1–9, Sep. 2018.
- [40] F. Peng, L. Qin, and M. Long, "Face presentation attack detection using guided scale texture," *Multimedia Tools Appl.*, vol. 77, no. 7, pp. 8883–8909, Apr. 2018.
- [41] K. Kollreider, H. Fronthaler, and J. Bigun, "Non-intrusive liveness detection by face images," *Image Vis. Comput.*, vol. 27, no. 3, pp. 233–244, 2009.
- [42] L. Feng, L. Po, and Y. Li, "Integration of image quality and motion cues for face anti-spoofing: A neural network approach," *J. Vis. Commun. Image Represent.*, vol. 38, no. 1, pp. 451–460, Jul. 2016.
- [43] A. Pinto, H. Pedrini, W. R. Schwartz, and A. Rocha, "Face spoofing detection through visual codebooks of spectral temporal cubes," *IEEE Trans. Image Process.*, vol. 24, no. 12, pp. 4726–4740, Dec. 2015.
- [44] G. Kim, S. Eum, J. K. Suhr, D. I. Kim, K. R. Park, and J. Kim, "Face liveness detection based on texture and frequency analyses," in *Proc. 5th IAPR Int. Conf. Biometrics (ICB)*, Mar. 2012, pp. 67–72.
- [45] A. Pinto, W. R. Schwartz, H. Pedrini, and A. Rocha, "Using visual rhythms for detecting video-based facial spoof attacks," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 5, pp. 1025–1038, May 2015.
- [46] T. Wang, J. Yang, Z. Lei, S. Liao, and S. Z. Li, "Face liveness detection using 3D structure recovered from a single camera," in *Proc. ICB*, Jun. 2013, pp. 1–6.
- [47] J. Galbally and S. Marcel, "Face anti-spoofing based on general image quality assessment," in *Proc. Int. Conf. Pattern Recognit.*, Aug. 2014, pp. 1173–1178.
- [48] Z. Boulkenafet, J. Komulainen, and A. Hadid, "Face anti-spoofing based on color texture analysis," in *Proc. ICIP*, Sep. 2015, pp. 2636–2640.
- [49] H. P. Nguyen, A. Delahaies, F. Retraint, and F. Morain-Nicolier, "Face presentation attack detection based on a statistical model of image noise," *IEEE Access*, vol. 7, pp. 175429–175442, 2019.
- [50] G. B. de Souza, J. P. Papa, and A. N. Marana, "On the learning of deep local features for robust face spoofing detection," *CoRR*, vol. abs/1806.07492, pp. 1–8, Oct. 2018.
- [51] J. Yang, Z. Lei, and S. Z. Li, "Learn convolutional neural network for face anti-spoofing," *CoRR*, vol. abs/1408.5601, Aug. 2014.
- [52] A. George, Z. Mostaani, D. Geissenbuhler, O. Nikisins, A. Anjos, and S. Marcel, "Biometric face presentation attack detection with multi-channel convolutional neural network," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 42–55, 2020.
- [53] B. Hamdan and K. Mokhtar, "The detection of spoofing by 3D mask in a 2D identity recognition system," *Egyptian Informat. J.*, vol. 19, no. 2, pp. 75–82, Jul. 2018.
- [54] G. Heusch and S. Marcel, *Remote Blood Pulse Analysis for Face Presentation Attack Detection*. Cham, Switzerland: Springer, 2019, pp. 267–289.
- [55] L. Li, X. Feng, Z. Xia, X. Jiang, and A. Hadid, "Face spoofing detection with local binary pattern network," *J. Vis. Commun. Image Represent.*, vol. 54, pp. 182–192, Jul. 2018.
- [56] T. Edmunds and A. Caplier, "Motion-based countermeasure against photo and video spoofing attacks in face recognition," *J. Vis. Commun. Image Represent.*, vol. 50, pp. 314–332, Jan. 2018.
- [57] M. Killioğlu, M. Taşkıran, and N. Kahraman, "Anti-spoofing in face recognition with liveness detection using pupil tracking," in *Proc. IEEE 15th Int. Symp. Appl. Mach. Intell. Inform. (SAMII)*, Jan. 2017, pp. 000087–000092.
- [58] J. Yang, Z. Lei, D. Yi, and S. Li, "Person-specific face antispoofing with subject domain adaptation," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 4, pp. 797–809, Apr. 2015.
- [59] S. Fatemifar, S. R. Arashloo, M. Awais, and J. Kittler, "Client-specific anomaly detection for face presentation attack detection," *Pattern Recognit.*, vol. 112, Apr. 2021, Art. no. 107696. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S0031320320304994>
- [60] A. J. Yaojie Liu, J. Stehouwer, and X. Liu, "Deep tree learning for zero-shot face anti-spoofing," *CoRR*, vol. abs/1904.02860, Jun. 2019.
- [61] S. Fatemifar, M. Awais, S. R. Arashloo, and J. Kittler, "Combining multiple one-class classifiers for anomaly based face spoofing attack detection," in *Proc. Int. Conf. Biometrics (ICB)*, Jun. 2019, pp. 1–7.
- [62] E. Alpaydm, *Introduction to Machine Learning* (Adaptive Computation and Machine Learning), 3rd ed. Cambridge, MA, USA: MIT Press, 2014.

- [63] L. Chen, H. Liao, M. Ko, J. Lin, and G. Yu, "A new LDA-based face recognition system which can solve the small sample size problem," *Pattern Recognit.*, vol. 33, no. 10, pp. 1713–1726, 2000.
- [64] G. Loosli and H. Aboubacar, "Using SVDD in SimpleMKL for 3D-shapes filtering," Clermont-Universite, Universite Blaise Pascal, Marseille, France, Tech. Rep., Jul. 2016. [Online]. Available: <https://hal.archives-ouvertes.fr/hal-01593595>
- [65] A. Rakotomamonjy, F. Bach, S. Canu, and Y. Grandvalet, "SimpleMKL," *J. Mach. Learn. Res.*, vol. 9, pp. 2491–2521, Nov. 2008.
- [66] M. Gönen and E. Alpaydm, "Localized algorithms for multiple kernel learning," *Pattern Recognit.*, vol. 46, no. 3, pp. 795–807, Mar. 2013.
- [67] S. P. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2014.
- [68] J. von Neumann, "Über ein ökonomisches gleichungssystem und eine verallgemeinerung des brouwerschen fixpunktsatzes," in *Proc. Wein 8th Ergebn. Math. Kolloq. Wein*, 1937, pp. 73–83.
- [69] D. Du and P. Pardalos, *Minimax and Applications* (Nonconvex Optimization and its Applications). Norwell, MA, USA: Kluwer, 1995.
- [70] J. Hoffman and S. Frankel, *Numerical Methods for Engineers and Scientists*, 2nd ed. Boca Raton, FL, USA: CRC Press, 2001.
- [71] M. Kemmler, E. Rodner, E.-S. Wacker, and J. Denzler, "One-class classification with Gaussian processes," *Pattern Recognit.*, vol. 46, no. 12, pp. 3507–3518, 2013.
- [72] H. Hoffmann, "Kernel PCA for novelty detection," *Pattern Recognit.*, vol. 40, no. 3, pp. 863–874, 2007.
- [73] B. Saleh, A. Farhadi, and A. Elgammal, "Object-centric anomaly detection by attribute-based reasoning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 787–794.
- [74] G. Griffin, A. Holub, and P. Perona, "Caltech256 image dataset," Caltech, Pasadena, CA, USA, Tech. Rep. 7694, 2007. [Online]. Available: http://www.vision.caltech.edu/Image_Datasets/Caltech256/
- [75] A. Costa-Pazo, S. Bhattacharjee, E. Vazquez-Fernandez, and S. Marcel, "The replay-mobile face presentation-attack database," in *Proc. Int. Conf. Biometrics Special Interest Group (BIOSIG)*, Sep. 2016, pp. 1–7.
- [76] Z. Boulkenafet, J. Komulainen, L. Li, X. Feng, and A. Hadid, "OULU-NPU: A mobile face presentation attack database with real-world variations," in *Proc. IEEE Int. Conf. Autom. Face Gesture Recognit.*, Washington, DC, USA, May 2017, pp. 612–618.
- [77] D. Wen, H. Han, and A. K. Jain, "Face spoof detection with image distortion analysis," *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 4, pp. 746–761, Apr. 2015.
- [78] I. Chingovska, A. Anjos, and S. Marcel, "On the effectiveness of local binary patterns in face anti-spoofing," in *Proc. Int. Conf. Biometrics Special Interest Group*, Sep. 2012, pp. 1–7.
- [79] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [80] C. Szegedy *et al.*, "Going deeper with convolutions," in *Proc. IEEE Conf. CVPR*, Jun. 2015, pp. 1–9.
- [81] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Adv. Neural Inf. Process. Syst.*, vol. 25, F. Pereira, C. J. C. Burges, L. Bottou, and K. Q. Weinberger, Eds. Red Hook, NY, USA: Curran Associates, 2012, pp. 1097–1105.
- [82] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *CoRR*, vol. abs/1409.1556, Sep. 2014.
- [83] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269.
- [84] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4510–4520.
- [85] B. Zoph, V. Vasudevan, J. Shlens, and Q. V. Le, "Learning transferable architectures for scalable image recognition," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8697–8710.
- [86] M. Everingham, L. V. Gool, C. K. I. Williams, J. Winn, and A. Zisserman, "The Pascal visual object classes (VOC) challenge," *Int. J. Comput. Vis.*, vol. 88, pp. 303–308, Sep. 2009.
- [87] P. Perera and V. M. Patel, "Learning deep features for one-class classification," *IEEE Trans. Image Process.*, vol. 28, no. 11, pp. 5450–5463, Nov. 2019.
- [88] R. Chalapathy, A. K. Menon, and S. Chawla, "Anomaly detection using one-class neural networks," *CoRR*, vol. abs/1802.06360, Feb. 2018. [Online]. Available: <http://arxiv.org/abs/1802.06360>
- [89] R. Hadsell, S. Chopra, and Y. LeCun, "Dimensionality reduction by learning an invariant mapping," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 2, Jun. 2006, pp. 1735–1742.
- [90] P. Oza and V. M. Patel, "One-class convolutional neural network," *IEEE Signal Process. Lett.*, vol. 26, no. 2, pp. 277–281, Feb. 2019.
- [91] P. Perera and V. M. Patel, "Learning deep features for one-class classification," *CoRR*, vol. abs/1801.05365, May 2018. [Online]. Available: <http://arxiv.org/abs/1801.05365>
- [92] *Information Technology—Biometric Presentation Attack Detection—Part 3: Testing and Reporting*, Standard ISO/IEC 30107-3:2017, International Organization for Standardization, 2017.
- [93] W. Sun, Y. Song, C. Chen, J. Huang, and A. C. Kot, "Face spoofing detection based on local ternary label supervision in fully convolutional networks," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 3181–3196, 2020.
- [94] Y. Qin, W. Zhang, J. Shi, Z. Wang, and L. Yan, "One-class adaptation face anti-spoofing with loss function search," *Neurocomputing*, vol. 417, pp. 384–395, Dec. 2020.
- [95] H. Feng *et al.*, "Learning generalized spoof cues for face anti-spoofing," *CoRR*, vol. abs/2005.03922, May 2020.
- [96] H. Chen, G. Hu, Z. Lei, Y. Chen, N. M. Robertson, and S. Z. Li, "Attention-based two-stream convolutional networks for face spoofing detection," *IEEE Trans. Inf. Forensics Security*, vol. 15, pp. 578–593, 2020.
- [97] G. W. Stewart, *Matrix Algorithms: Basic Decompositions*, vol. 1. Philadelphia, PA, USA: SIAM, 2001.
- [98] S. R. Arashloo and J. Kittler, "Class-specific kernel fusion of multiple descriptors for face verification using multiscale binarised statistical image features," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 12, pp. 2100–2109, Dec. 2014.
- [99] P. Benner, P. Ezzatti, E. S. Quintana-Ortí, and A. Remón, "Matrix inversion on CPU-GPU platforms with applications in control theory," *Concurrency Computation, Pract. Exper.*, vol. 25, no. 8, pp. 1170–1182, Jun. 2013, doi: [10.1002/cpe.2933](https://doi.org/10.1002/cpe.2933).
- [100] D. Yu, S. He, Y. Huang, G. Yu, and L. Yang, "A fast parallel matrix inversion algorithm based on heterogeneous multicore architectures," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Dec. 2015, pp. 903–907.



Shervin Rahimzadeh Arashloo received the Ph.D. degree from the Centre for Vision, Speech and Signal Processing (CVSSP), University of Surrey, Guildford, U.K., in 2010. He is currently an Assistant Professor with the Department of Computer Engineering, Bilkent University, Ankara, Turkey, and a Visiting Research Fellow with CVSSP, University of Surrey. His research interests include pattern recognition, machine learning, and signal processing.