

Face Anti-spoofing using Hybrid Residual Learning Framework

Usman Muhammad and Abdenour Hadid
Center for Machine Vision and Signal Analysis (CMVS)
University of Oulu, Finland

Abstract

Face spoofing attacks have received significant attention because of criminals who are developing different techniques such as warped photos, cut photos, 3D masks, etc. to easily fool the face recognition systems. In order to improve the security measures of biometric systems, deep learning models offer powerful solutions; but to attain the benefits of multilayer features remains a significant challenge. To alleviate this limitation, this paper presents a hybrid framework to build the feature representation by fusing ResNet with more discriminative power. First, two variants of the residual learning framework are selected as deep feature extractors to extract informative features. Second, the fully-connected layers are used as separated feature descriptors. Third, PCA based Canonical correlation analysis (CCA) is proposed as a feature fusion strategy to combine relevant information and to improve the features' discrimination capacity. Finally, the support vector machine (SVM) is used to construct the final representation of facial features. Experimental results show that our proposed framework achieves a state-of-the-art performance without fine-tuning, data augmentation or coding strategy on benchmark databases, namely the MSU mobile face spoof database and the CASIA face anti-spoofing database.

1. Introduction

With applications in speech recognition, signature verification, and fingerprint identification, biometric systems are extensively used in our daily lives. Among different biometrics, face is commonly used in security systems. A simple face recognition algorithm can recognize or verify the identity of a subject from a digital photo by comparing and analyzing facial patterns. Due to increase of electronic devices, people are expecting for secure and convenient ways to access their personal information. On the other hand, criminals are active for spoofing by masquerade or concealing one's identity to gain illegitimate access and advantages.

In this regard, a high security requirement for face authentication is needed. The face spoofing attacks can be as simple as a printed photo paper, facial cosmetic makeup, replayed video and 3D masks [34], which can easily spoof a face recognition system to access secure information illegally. Therefore, anti-spoofing problem has attracted great attentions to the researchers to improve the security measures of public places and companies.

Color-based [31, 38], or texture-based [8] methods are used widely in the face anti-spoofing domain. However, these methods require high resolution images to extract detailed information, which is needed for discriminating genuine faces from attack faces. Alternatively, the Bag of Visual Words (BoVW) and Fisher Vector encoding methods have been frequently used [9], in which the input is a set of handcrafted features and the output is a set of learned features. Several popular approaches for mapping visual and textual features to the same latent space [13, 16, 28] rely on canonical correlation analysis (CCA) [32]. An extended work of CCA is introduced in [2], where a deep CCA model (DCCA) can be end-to-end trainable. It is regarded as a nonlinear extension of traditional linear CCA. However, if dataset has limited samples, DCCA is easy to overfitting. To improve general nonlinear setting via a kernelization procedure, Kernel CCA (KCCA) is introduced [1]. The conventional KCCA does not allow feature selection or capturing of multiple canonical components. Thus, Multi-set canonical correlation analysis (MCCA) [17] and discriminative of CCA (DCCA) [33] were introduced to deal with multi-feature information fusion. Despite the great success of these methods, a large semantic gap between the low-level features and the high-level semantic meanings still exist, which limits the ability to deal with the complex attack images. Under this situation, the focus is shifting to high-dimensional Convolutional Neural Networks (CNNs) features rather than acquiring handcrafted (low-level) or mid-level features.

Recently, a significant progress has been made for learning high-level semantic features due to development of the convolutional neural network (CNN), since it has a natural ability to efficiently encode spectral and spatial information

based on stack of convolutional filters. Deep learning architectures such as GoogLeNet [36], AlexNet [21], ResNet-18, ResNet-50 [18] and VGG-19 [29] have been utilized not only in classic problems such as face recognition, image segmentation, object detection, but also in many other practical applications [10]. However, the backpropagation process, the stochastic gradient descent (SGD) strategy or training a deep CNN from the start is challenging especially with limited training data. Recent works have demonstrated that the pretrained CNNs which are trained on large datasets such as ImageNet can be efficiently use and transfer to many other classification tasks. Inspired by the work proposed in [4, 26], we focus on extracting activation vectors and present a hybrid residual learning framework for face anti-spoofing. The main contribution of this paper is presenting a solution to exploit more benefits of a pretrained CNN for face anti-spoofing. To this end, contribution is summarized as follows.

1) We present a hybrid residual learning framework, which has stronger discrimination ability than single model features. Using this framework, we make the ResNet model more suitable for spoofing problem for face recognition system.

2) We propose a low-rank approximation scheme based on principal components and linear Canonical Correlation Analysis to integrate final outputs of recurrent residual learning, which outperforms the performance compared to state-of-the-art fusion methods.

The rest of the paper is organized as follow: The recent works are reviewed in Section 2. We discuss the proposed framework in Section 3, and present experimental results in Section 4. Finally, we draw the conclusion in Section 5.

2. Related work

A well-known approach to enhance the CNN performance is the feature fusion, which is an efficient step in face anti-spoofing. Nguyen *et al.* [26] fused deep learned features with handcrafted features by simply concatenating them. The work reported in [10], adopted discriminant correlation analysis to fuse the features extracted from two fully-connected layers of a pretrained CNN. The Multi Kernel Learning (MKL) and Metric Learning (ML) have been proved as effective fusion strategies to fuse different layers of CNN model [4, 25]. Two streams: patch-based CNN, and depth-based CNN is proposed by extracting the CNN features and holistic depth maps from the face images in [3]. The CNN was served as a feature extractor in [23], where the convolutional layer features were chosen to distinguish the real and fake faces. In order to reduce the dimensions, the block principle component analysis (PCA) method was utilized. Two kind of features such as static and dynamics were separately extracted by using the CNN network and finally fusion is performed to fuse facial features [27]. A

combined approach based on Short-Term Memory (LSTM) units with Convolutional Neural Networks (CNN) is proposed to conduct a joint prediction for multiple frames of a video [39]. Feng *et al.* [11] adopted Shearlet-based approach to develop an image quality-based liveness features and dense optical flow maps to estimate the facial motions. Then, a bottleneck feature fusion strategy is proposed that can integrate different liveness features effectively.

Recently, a multilayer feature fusion method based on fisher kernel coding is introduced [22]. The authors showed that the extracted features from fully connected layers can improve the discrimination of features representation. However, they argue that integration of deeper CNN model such as AlexNet, VGG-M, VGG-S, and CaffeNet, can achieve a better performance but simple combination strategy achieves less than 1% in terms of accuracy. Therefore, we pay particular attention to their arguments, and can easily arrive at the following question: can we integrate different model features effectively and efficiently to address face spoofing detection? To our knowledge, this question still remains unclear, except for the concurrent works [4, 10, 23] with ours.

3. The Proposed Approach

The activation vectors extracted from pretrained CNN are typically much faster and easier than training from the start. However, selecting a suitable CNN model as a feature extractor is very crucial because inappropriate choice of CNN model may yield poor results. If the model is too small, it can limit detection performance. In order to find the most suitable feature descriptors for our task, we select some of the most popular CNN models as a feature extractor and compare their performance in Table 1. Among of them, we choose two variants of ResNet for the final detection. In this section, we explain the propose framework based on pretrained deep residual learning with PCA and the canonical correlation analysis (CCA).

3.1. The Deep Residual Learning Model

We use ResNet-50 and Resnet-101 because these networks achieve better performance on the proposed datasets. The deep residual learning framework is designed to address the degradation problem, and achieves 3.57% error on the ImageNet test set. Each unit can be expressed in the following formulas [18]:

$$y_1 = a(x_l) + F(x_l, W_l) \quad (1)$$

$$x_{l+1} = F(y_1) \quad (2)$$

where x_l and x_{l+1} are the input and output of the i -th unit, and F is a residual function. Each variant contains different size of convolutional layers (blocks), where each unit of layer receives inputs from a set of units located in small

neighborhood in the previous layer. The fully-connected (fc) layer is the final layers, where each neuron in this layer is linked to all the numbers in the previous layer. The fc layer has an output size of 1000 and the output of the very last layer is fed to a 1000-way softmax which produces probabilities over 1000 object categories.

3.2. Canonical Correlation Analysis with PCA

Feature fusion is the process of concatenating two or more than two feature vectors into a single one. Sun *et al.* [32] introduced a robust feature fusion method based on Canonical Correlation Analysis (CCA). We aim to maximize the correlations of two features sets, which are extracted from ResNet-50 and ResNet-101, and combine them with more discriminative power. Conventional feature fusion methods such as serial fusion [24] and parallel fusion [41] strategy simply concatenate or integrate several kinds of features together. Recently, Discriminant Correlation Analysis (DCA) has been introduced that includes the class information during fusion [14], but integrating different feature vectors is always a challenge and still an open problem. Despite the simplicity of such methods, the fusion techniques that adopt such strategies may still not perform well. Features should be equally represented, normalized or measured before applying fusion.

To alleviate the aforementioned challenge, three regularization methods are used commonly such as ridge-type regularization, low-rank approximation by principal components, and the ridge-type regularization [20]. In our work, we use low-rank approximation by principal components. Let D_1 and D_2 be two feature sets of ResNet framework. The PCA extracts certain eigenvectors from D_1 and D_2 , represented by S_1 and S_2 , respectively. After calculating eigenvectors, it projects columns of D_1 and D_2 onto the column spaces of S_1 and S_2 and achieves the reduced-rank approximation of kernel data.

$$\tilde{D}_1 = D_1' S_1 \quad \text{and} \quad \tilde{D}_2 = D_2' S_2 \quad (3)$$

The Canonical Correlation Analysis is performed on reduced kernel $\tilde{D} = [\tilde{D}_1 \tilde{D}_2]$ instead of the full data kernel matrix D . The PCA approach corresponds to using the following linear independent systems for discretization:

$$\phi_\nu(a^{(1)}) = \sum_{i=1}^n d_1(a^{(1)}, a_i^{(1)}) u_{i\nu}^{(1)}, \nu = 1, \dots, m_1, \quad (4)$$

and

$$\psi_\mu(a^{(2)}) = \sum_{i=1}^n d_2(a^{(2)}, a_i^{(2)}) u_{i\mu}^{(2)}, \mu = 1, \dots, m_2, \quad (5)$$

where $[u_{i\nu}^{(1)}] = S_1$ and $[u_{i\mu}^{(2)}] = S_2$. The principal components method seeks to provide optimal linear independent

basis for discretization. After applying PCA on each feature set, we aim to maximize the correlation among them. In other words, projections of the variables onto these two vectors are mutually maximized. For further detail, mathematical explanation is given below.

Suppose that $A \in Q^{n \times m}$ and $B \in Q^{p \times m}$ represent two matrices, each consists of m training feature vectors from two different sets. n and p are defined as dimensions of each vector.

Let consider that $W_{xx} \in Q^{n \times p}$ and $W_{yy} \in Q^{n \times p}$ define the within-sets covariance matrices of A and B and $W_{xy} \in Q^{n \times p}$ contains the between-set covariance matrix (note that $W_{yx} = W_{xy}^T$). The overall covariance matrix $(n+p) \times (n+p)$ is then calculated:

$$W = \begin{pmatrix} cov(x) & cov(x, y) \\ cov(y, x) & cov(y) \end{pmatrix} = \begin{pmatrix} W_{xx} & W_{xy} \\ W_{yx} & W_{yy} \end{pmatrix} \quad (6)$$

The correlation of the matrix W between these two sets of vector does not ensure a consistent pattern [15]. The objective of CCA is to find the linear combinations, $\hat{A} = R_x^T A$ and $\hat{B} = R_y^T B$, which maximizes the pair-wise correlations across the two feature sets:

$$corr(\hat{A}, \hat{B}) = \frac{cov(\hat{A}, \hat{B})}{var(\hat{A}) \cdot var(\hat{B})} \quad (7)$$

Where $cov(\hat{A}, \hat{B}) = R_x^T Q_{xy} R_y$, $var(\hat{A}) = R_x^T Q_{xx} R_x$ and $var(\hat{B}) = R_y^T Q_{yy} R_y$. In order to find the local maxima and minima, and to maximize the covariance between \hat{A} and \hat{B} , the Lagrange multipliers method is used to satisfy the following constraints $var(\hat{A}) = var(\hat{B}) = 1$. Both transformation matrices, R_x and R_y , are then computed by using the eigenvalue equations:

$$\begin{cases} W_{xx}^{-1} W_{xy} W_{yy}^{-1} W_{yx} \hat{R}_x = Q^2 \hat{R}_x \\ W_{yy}^{-1} W_{yx} W_{xx}^{-1} W_{xy} \hat{R}_y = Q^2 \hat{R}_y \end{cases} \quad (8)$$

where \hat{R}_x and \hat{R}_y represent eigenvectors and Q^2 is the diagonal matrix of eigenvalues and it is denoted as squares of the canonical correlations. The number of non-zero eigenvalues can be find in each equation, that is $x = rank(W_{xy} \leq \min(m, n, p))$, which will be fixed in descending order, $z_1 \geq z_2 \geq \dots \geq z_d$. As mentioned earlier, both the transformation matrices, R_x and R_y , compose of the sorted eigenvectors corresponding to the non-zero eigenvalues. $\hat{A}, \hat{B} \in Q^{x \times m}$ are denoted as canonical variants. The sam-

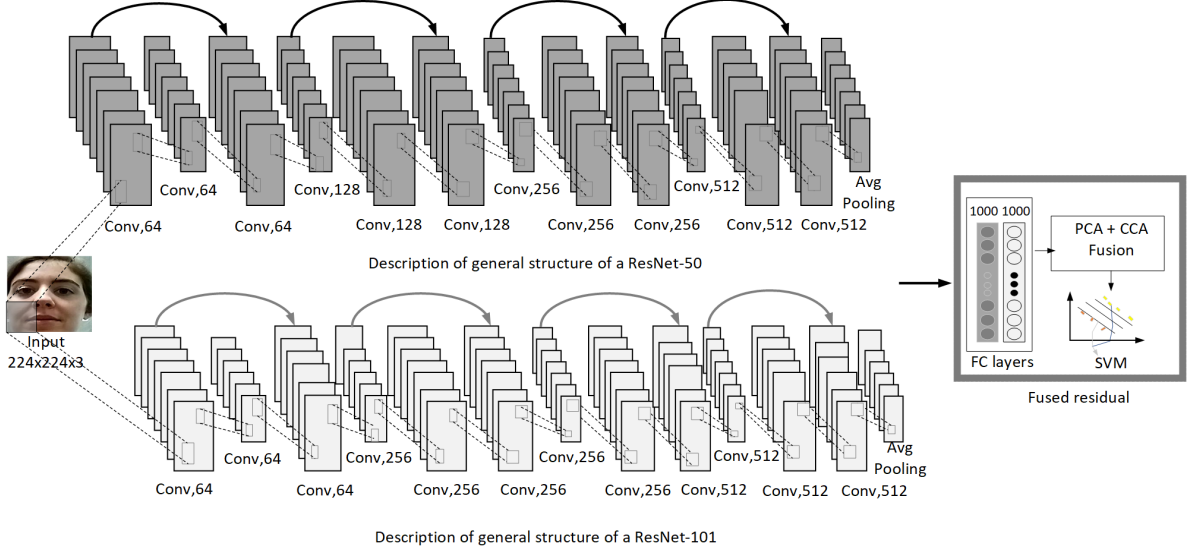


Figure 1. Block diagram of visual concept extraction method with PCA based canonical correlation analysis

ple covariance matrix denoted in Eq.(6) will be of the form:

$$\begin{aligned}
 {}^*W = \begin{pmatrix} 1 & 0 & \dots & 0 & z_1 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & 0 & z_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 & 0 & 0 & \dots & z_d \\ \hline z_1 & 0 & \dots & 0 & 1 & 0 & \dots & 0 \\ 0 & z_2 & \dots & 0 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & z_d & 0 & 0 & \dots & 1 \end{pmatrix} \quad (9)
 \end{aligned}$$

In Eq.(9), the matrix contains the linear indices of each nonzero element only on their corresponding indices. In addition, it explains that the canonical variants are uncorrelated within each other because of identity matrices in the upper left and lower right corners. Hence, two choices are available to perform feature-level fusion either by concatenation or summation of the transformed feature vectors:

$$F_1 = \begin{pmatrix} {}^*A \\ {}^*B \end{pmatrix} = \begin{pmatrix} R_x^T A \\ R_y^T B \end{pmatrix} = \begin{pmatrix} R_x & 0 \\ 0 & R_y \end{pmatrix}^T \begin{pmatrix} A \\ B \end{pmatrix} \quad (10)$$

or

$$F_2 = {}^*A + {}^*B = R_x^T A + R_y^T B = \begin{pmatrix} R_x \\ R_y \end{pmatrix}^T \begin{pmatrix} A \\ B \end{pmatrix} \quad (11)$$

where F_1 and F_2 are called the Canonical Correlation Discriminant Features (CCDFs). Hence, by using the above Eq.(10) and Eq.(11), fusion is possible to perform by concatenation or summation. Both final outputs layers of each ResNet model were used for PCA and CCA. In our work, we report the results by using Eq.(10). We also used Eq.(11) to fuse the features but achieve low accuracy. Hence, we do

not report those results. The support vector machine (SVM) is utilized as one-against-all strategy to perform detection. In an overall view, the block diagram in Fig.1, represents the procedure of fusing ResNet with CCA subspace.

4. Experiments

The first dataset, MSU Mobile Face SpooF Database [38], was collected from 280 videos of fake and real faces. The videos were recorded from 35 subjects using two kinds of cameras. The duration of each video is about nine seconds. Different types of attacks such as high definition replay attacks, printed attacks and mobile replay attacks were used to generate attack faces. For the performance evaluation, 15 subjects are used in the training set and the rest for the test.

The second dataset used in the experiments is the CASIA Face Anti-Spoofing database [42]. The videos were collected from 50 subjects, where the images from fake faces were taken in the high resolution recordings of the original faces. Three fake face attacks were made: cut photo attacks, warped photo attacks and video attacks. Three kinds of image qualities have been used for recording such as low, normal and high. In order to perform experiments, the dataset is splitted into two subsets for training and testing (20 and 30, respectively).

4.1. Experimental Setup

The strategy of adopting a pretrained CNN model as feature extractor is simple, since no data augmentation or fine-tuning is necessary. Moreover, one just needs to choose a CNN model and the features can be extracted from any layer, such as convolutional, pooling or fully-connected.

Table 1. Results of different feature extractors and fusion strategies for MSU MFSD dataset using frame based evaluation metric

Pre-trained CNN model	Accuracy(%)	Feature Size
VGG-19	85.10±0.20	4096
AlexNet	86.45±0.30	4096
GoogLeNet	84.78±0.40	1000
VGG-16	85.80±0.40	4096
ResNet-18	87.69±0.20	1000
ResNet-50	91.63±0.40	1000
ResNet-101	90.20±0.40	1000
DenseNet-201	89.80±0.30	1000
Inception-ResNet-v2	84.80±0.50	1000
ResNet(50+101) with Serial Fusion	90.10±0.20	2000
ResNet(50+101) with Parallel Fusion	89.66±0.20	1000
ResNet(50+101) with DCA	85.66±0.20	23
The Proposed ResNet(50+101) with PCA + CCA	96.05±0.70	643

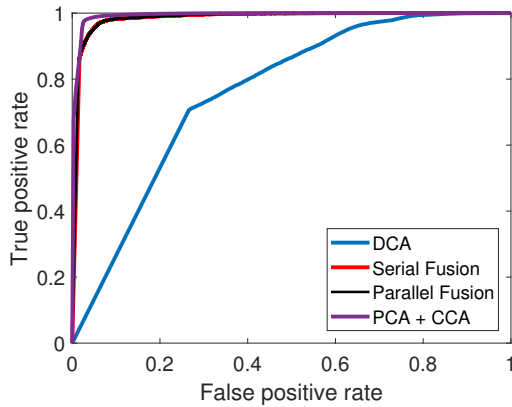


Figure 2. ROC curve of the CASIA database with the proposed framework. The higher the curve, the better.

In this paper, several pretrained CNN models, GoogLeNet [36], AlexNet [21], ResNet-18, ResNet-50 [18], VGG-16, VGG-19 [29], much deeper network ResNet-101 [18], Inception-ResNet-v2 [35] and DenseNet-201 [19], are applied as off-the-shelf CNN features extractors. Table 1 shows the experimental results for the MSU MFSD dataset. After that, we choose ResNet-50 and ResNet-101 because these models achieve better performance than others. All the face images are scaled to 224×224 for the pre-defined size requirement of the network. For each dataset, the features are learned on training set and then evaluated on testing set.

4.2. Results and Analysis

We observe a remarkable improvement in Table 1 for MSU MFSD database comparing the results with the baseline fusion strategies as described in section 3.2. The proposed approach improves four to five percent accuracy than the individual models. We also compare the performance on CASIA FASD database, the proposed fusion framework achieves the best accuracy (98.20 ± 0.50).

Table 2 provides a comparison between the results of our

Table 2. Comparisons between the proposed approach and state-of-the-art methods on CASIA and MSU MFSD dataset using frame based evaluation metric.

Method	CASIA	MSU
	EER (%)	EER (%)
Texture analysis [8]	2.1	4.9
Motion mag+LBP [5]	14.4	-
DMD [37]	21.7	-
CNN [40]	7.4	-
IQA [12]	32.4	-
HSV + YCbCr fusion [9]	2.8	2.2
Multiscale Fusion [6]	4.2	6.9
IDA [38]	-	8.5
Colour LBP [7]	7.1	10.6
SPMT + SSD [30]	0.04	-
The proposed method	0.02	0.04

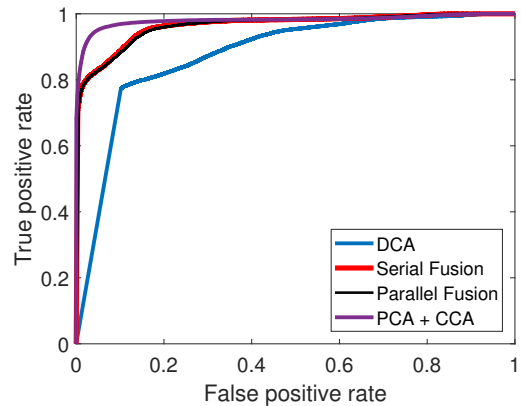


Figure 3. ROC curve of the MSU MFSD database with the proposed framework. The higher the curve, the better.

proposed approach with several state-of-the-art approaches on MSU MFSD and CASIA FASD database. In [8], authors emphasize the importance of different colour representations and fuse LPQ and CoALBP descriptors by concatenating their resulting histograms. Another fusing scheme is introduced based on SURF and Fisher vector encoding, where fusion is performed using HSV and YCbCr colour spaces [9]. A new multiscale space is introduced to expand images using different filtering schemes. Then, feature histograms are concatenated to fuse them into final feature vector [6]. LBP based descriptor is proposed to investigate the most suitable color space for face spoofing in [7]. In addition to texture based methods, different variants of Image Distortion Analysis (IDA) features (color moments, blurriness, specular reflection, and color diversity) are developed to capture the image distortion in the spoof face images. The fusion is performed by simply concatenating them [38]. Other approaches including, detection with motion magnifi-

cation [5], face spoofing using visual dynamics [37], detection based on general image quality assessment [12], CNN for face anti-spoofing [40] and discriminative representation combinations for accurate spoofing detection [30].

Making a comparison with above methods, the proposed fusion framework achieves the best equal error rate (0.02) for CASIA and (0.04) for MSU database. It proves that the performance of our proposed framework competes the state-of-the-art methods by a fair margin. For further analysis, the ROC (Receiver Operating Characteristics) curves of CASIA and MSU datasets are shown in Fig. 2, and Fig. 3, respectively. The true positive rate (TPR) determines how many correct positive results exist among all positive samples which are taken during the test. In contrast to TPR, the false positive rate (FPR) illustrates how many incorrect positive results appear among all negative samples accessible during the test. As can be seen, our proposed method yields the highest point in the upper left corner.

5. Conclusions

Deep learning models provide a powerful solution for face anti-spoofing. However, fine-tuning or constructing a CNN model from the start remains challenging for face spoofing images in consideration of limited training data and time consuming process. Therefore, this work focuses on exploiting the potential of off-the-shelf CNN features, and gives a robust proposal by combining ResNet with PCA based canonical correlation analysis. In comparison with texture based fusion methods and baseline fusion strategies, the proposed framework works much better than existing fusion and state-of-the-art methods.

Although it achieved a better performance, the proposed method still has room for improvement. For instance, the computational cost is not taken into consideration. Therefore, our future research aims to determine on computational complexity and concrete strategies for fusing multiple layer features. We will try to exploit pruning approaches to compress convolutional neural networks, to further decrease the features dimension and accelerating the detection speed.

Acknowledgment

The financial support of the Academy of Finland is acknowledged.

References

- [1] S. Akaho. A kernel method for canonical correlation analysis. *arXiv preprint cs/0609071*, 2006.
- [2] G. Andrew, R. Arora, J. Bilmes, and K. Livescu. Deep canonical correlation analysis. In *International Conference on Machine Learning*, pages 1247–1255, 2013.
- [3] Y. Atoum, Y. Liu, A. Jourabloo, and X. Liu. Face anti-spoofing using patch and depth-based cnns. In *Biometrics (IJCB), 2017 IEEE International Joint Conference on*, pages 319–328. IEEE, 2017.
- [4] A. Bansal, R. Ranjan, C. D. Castillo, and R. Chellappa. Deep features for recognizing disguised faces in the wild. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 10–106. IEEE, 2018.
- [5] S. Bharadwaj, T. I. Dhamecha, M. Vatsa, and R. Singh. Computationally efficient face spoofing detection with motion magnification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 105–110, 2013.
- [6] Z. Boulkenafet, J. Komulainen, X. Feng, and A. Hadid. Scale space texture analysis for face anti-spoofing. In *Biometrics (ICB), 2016 International Conference on*, pages 1–6. IEEE, 2016.
- [7] Z. Boulkenafet, J. Komulainen, and A. Hadid. Face anti-spoofing based on color texture analysis. In *Image Processing (ICIP), 2015 IEEE International Conference on*, pages 2636–2640. IEEE, 2015.
- [8] Z. Boulkenafet, J. Komulainen, and A. Hadid. Face spoofing detection using colour texture analysis. *IEEE Transactions on Information Forensics and Security*, 11(8):1818–1830, 2016.
- [9] Z. Boulkenafet, J. Komulainen, and A. Hadid. Face anti-spoofing using speeded-up robust features and fisher vector encoding. *IEEE Signal Processing Letters*, 24(2):141–145, 2017.
- [10] S. Chaib, H. Liu, Y. Gu, and H. Yao. Deep feature fusion for vhr remote sensing scene classification. *IEEE Trans. Geosci. Remote Sens*, 55(8):4775–4784, 2017.
- [11] L. Feng, L.-M. Po, Y. Li, X. Xu, F. Yuan, T. C.-H. Cheung, and K.-W. Cheung. Integration of image quality and motion cues for face anti-spoofing: A neural network approach. *Journal of Visual Communication and Image Representation*, 38:451–460, 2016.
- [12] J. Galbally and S. Marcel. Face anti-spoofing based on general image quality assessment. In *Pattern Recognition (ICPR), 2014 22nd International Conference on*, pages 1173–1178. IEEE, 2014.
- [13] Y. Gong and S. Lazebnik. Comparing data-dependent and data-independent embeddings for classification and ranking of internet images. 2011.
- [14] M. Haghghat, M. Abdel-Mottaleb, and W. Alhalabi. Discriminant correlation analysis: Real-time feature level fusion for multimodal biometric recognition. *IEEE Transactions on Information Forensics and Security*, 11(9):1984–1996, 2016.
- [15] M. Haghghat, M. Abdel-Mottaleb, and W. Alhalabi. Fully automatic face normalization and single sample face recognition in unconstrained environments. *Expert Systems with Applications*, 47:23–34, 2016.
- [16] D. R. Hardoon, S. Szedmak, and J. Shawe-Taylor. Canonical correlation analysis: An overview with application to learning methods. *Neural computation*, 16(12):2639–2664, 2004.

- [17] M. A. Hasan. On multi-set canonical correlation analysis. In *Neural Networks, 2009. IJCNN 2009. International Joint Conference on*, pages 1128–1133. IEEE, 2009.
- [18] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.
- [19] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger. Densely connected convolutional networks. In *CVPR*, volume 1, page 3, 2017.
- [20] S.-Y. Huang, M.-H. Lee, and C. K. Hsiao. Kernel canonical correlation analysis and its applications to nonlinear measures of association and test of independence. *Institute of Statistical Science: Academia Sinica, Taiwan*, 2006.
- [21] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In *Advances in neural information processing systems*, pages 1097–1105, 2012.
- [22] E. Li, J. Xia, P. Du, C. Lin, and A. Samat. Integrating multi-layer features of convolutional neural networks for remote sensing scene classification. *IEEE Transactions on Geoscience and Remote Sensing*, 55(10):5653–5665, 2017.
- [23] L. Li, X. Feng, Z. Boulkenafet, Z. Xia, M. Li, and A. Hadid. An original face anti-spoofing approach using partial convolutional neural network. In *Image processing theory tools and applications (IPTA), 2016 6th international conference on*, pages 1–6. IEEE, 2016.
- [24] C. Liu and H. Wechsler. A shape-and texture-based enhanced fisher classifier for face recognition. *IEEE transactions on image processing*, 10(4):598–608, 2001.
- [25] Q. Liu, R. Hang, H. Song, and Z. Li. Learning multiscale deep features for high-resolution satellite image scene classification. *IEEE Transactions on Geoscience and Remote Sensing*, 56(1):117–126, 2018.
- [26] D. T. Nguyen, T. D. Pham, N. R. Baek, and K. R. Park. Combining deep and handcrafted image features for presentation attack detection in face recognition systems using visible-light camera sensors. *Sensors*, 18(3):699, 2018.
- [27] K. Patel, H. Han, and A. K. Jain. Cross-database face anti-spoofing with robust feature representation. In *Chinese Conference on Biometric Recognition*, pages 611–619. Springer, 2016.
- [28] N. Rasiwasia, J. Costa Pereira, E. Coviello, G. Doyle, G. R. Lanckriet, R. Levy, and N. Vasconcelos. A new approach to cross-modal multimedia retrieval. In *Proceedings of the 18th ACM international conference on Multimedia*, pages 251–260. ACM, 2010.
- [29] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.
- [30] X. Song, X. Zhao, L. Fang, and T. Lin. Discriminative representation combinations for accurate face spoofing detection. *Pattern Recognition*, 85:220–231, 2019.
- [31] L. Souza, L. Oliveira, M. Pamplona, and J. Papa. How far did we get in face spoofing detection? *Engineering Applications of Artificial Intelligence*, 72:368–381, 2018.
- [32] Q.-S. Sun, S.-G. Zeng, Y. Liu, P.-A. Heng, and D.-S. Xia. A new method of feature fusion and its application in image recognition. *Pattern Recognition*, 38(12):2437–2448, 2005.
- [33] T.-K. Sun, S.-C. Chen, Z. Jin, and J.-Y. Yang. Kernelized discriminative canonical correlation analysis. In *Wavelet Analysis and Pattern Recognition, 2007. ICWAPR'07. International Conference on*, volume 3, pages 1283–1287. IEEE, 2007.
- [34] X. Sun, L. Huang, and C. Liu. Multispectral face spoofing detection using vis–nir imaging correlation. *International Journal of Wavelets, Multiresolution and Information Processing*, 16(02):1840003, 2018.
- [35] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi. Inception-v4, inception-resnet and the impact of residual connections on learning. In *AAAI*, volume 4, page 12, 2017.
- [36] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.
- [37] S. Tirunagari, N. Poh, D. Windridge, A. Iorliam, N. Suki, and A. T. Ho. Detection of face spoofing using visual dynamics. *IEEE transactions on information forensics and security*, 10(4):762–777, 2015.
- [38] D. Wen, H. Han, and A. K. Jain. Face spoof detection with image distortion analysis. *IEEE Transactions on Information Forensics and Security*, 10(4):746–761, 2015.
- [39] Z. Xu, S. Li, and W. Deng. Learning temporal features using lstm-cnn architecture for face anti-spoofing. In *Pattern recognition (acpr), 2015 3rd IAPR asian conference on*, pages 141–145. IEEE, 2015.
- [40] J. Yang, Z. Lei, and S. Z. Li. Learn convolutional neural network for face anti-spoofing. *arXiv preprint arXiv:1408.5601*, 2014.
- [41] J. Yang, J.-y. Yang, D. Zhang, and J.-f. Lu. Feature fusion: parallel strategy vs. serial strategy. *Pattern recognition*, 36(6):1369–1381, 2003.
- [42] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, and S. Z. Li. A face antispoofing database with diverse attacks. In *Biometrics (ICB), 2012 5th IAPR international conference on*, pages 26–31. IEEE, 2012.