A. Brömme, C. Busch, A. Dantcheva, C. Rathgeb and A. Uhl (Eds.): BIOSIG 2018,
Lecture Notes in Informatics (LNI), Gesellschaft für Informatik, Bonn 2018

Deep Sparse Feature Selection and Fusion for Textured Contact Lens Detection

Domenick Poster, [1] Nasser Nasrabadi, [2] Benjamin Riggan[3]

Abstract: Distinguishing between images of irises wearing textured lenses versus those wearing transparent lenses or no lenses is a challenging problem due to the subtle and fine-grained visual differences. Our approach builds upon existing hand-crafted image features and neural network architectures by optimally selecting and combining the most useful set of features into a single model. We build multiple, parallel sub-networks corresponding to the various feature descriptors and learn the best subset of features through *group sparsity*. We avoid overfitting such a wide and deep model through a selective transfer learning technique and a novel *group Dropout* regularization strategy. This model achieves roughly a four times increase in performance over the state-of-the-art on three benchmark textured lens datasets and equals the near-perfect state-of-the-art accuracy on two others. Furthermore, the generic nature of the architecture allows it to be extended to other image features, forms of spoofing attacks, or problem domains.

Keywords: feature selection, feature fusion, group sparsity, iris liveness detection, textured contact lens detection

1 Introduction

Textured contact lenses, also known as cosmetic lenses, obscure the physiological iris texture during image acquisition for biometric recognition. Wearing textured lenses has been shown to degrade the performance of iris recognition algorithms from 99% verification accuracy to nearly 20% [Ko13] due to the overwhelming occlusion caused by the lenses. This is analogous to performing facial recognition on subjects wearing masks. While robustness to spoofs like textured contact lenses remains a significant challenge, a more tractable problem is the detection of textured lenses. Even if an iris cannot be correctly identified, in the event a textured lens is detected, then a more appropriate biometric application decision can be made, such as sending individuals from checkpoints to secondary screening.

Textured lens detection is considered a liveness detection problem. Liveness detection is the binary classification of a given biometric sample as either live (genuine) or spoof (non-genuine / concealed / non-conformant). Figure 1 is a side-by-side comparison of the same subject with and without a textured lens. Irises concealed by textured lenses are spoofing

[1] Lane Department of Computer Science and Electrical Engineering, 395 Evansdale Dr, Morgantown, WV 26506, dposter@mix.wvu.edu
[2] Lane Department of Computer Science and Electrical Engineering, 395 Evansdale Dr, Morgantown, WV 26506, nasser.nasrabadi@mail.wvu.edu
[3] US Army Research Laboratory, 2800 Powder Mill Rd, Adelphi, MD 20783, benjamin.s.riggan.civ@mail.mil

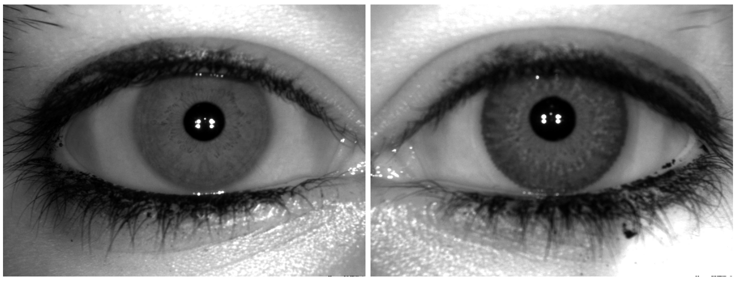

Fig. 1: An example of an iris without (left) and with (right) a textured (cosmetic) lens taken from the IIITD dataset.

attacks designed to defeat iris recognition systems. The ability to reliably detect these attacks will help mitigate reductions in security.

While many viable approaches have been developed, textured lens detection remains a challenge [CB18]. No single approach has been shown to be definitively superior. In a comparative liveness detection study, [Gr15] concludes "although designing better descriptors is certainly of interest, it is very likely that more significant improvements come from a sensible decision level fusion of existing ones."

The primary motivation of this paper is an optimal feature selection and fusion strategy for robust detection of textured contact lenses. Our hypothesis is that a combination of good features performs better than any single feature. Similarly, it stands to reason some features ought to be excluded as they may be redundant or misinformative. The decision of which features to select and how to fuse them is nontrivial. Our proposed method solves both the problems of feature selection and fusion simultaneously.

The main contribution of this paper is the development of an integrated feature selection and fusion model. The architecture is comprised of an eight-layer convolutional neural network (CNN) and several parallel multi-layer perceptrons (MLPs). We propose a $\ell_{1,2}$ *group sparsity* loss term to sparsify feature-specific sub-networks. In addition, we develop a set of techniques to train such a large network under the constraints of limited iris datasets while encouraging unbiased selection of heterogeneous image features.

The contents of the paper are structured as follows. Section 2 frames our approach in the context of related works. Section 3 discusses the model architecture and feature selection. Section 4 presents experiment details and results. Closing remarks are made in Section 5.

2 Related Works

The majority of existing research in this field has been focused on the development and utilization of feature descriptors. [Gr15] benchmarks a plethora of local feature descriptors against iris, fingerprint, and face liveness detection datasets. Our experiments draw upon BSIF, LBP, CoA-LBP, HoG, DAISY, and SID features [KR12, OPM02, NOF11, DT05, TLF10, Gr14]

The use of neural networks for iris liveness detection is becoming increasingly popular. [Me15] lead early experimentation in this area with 2-3 layer 'SpoofNet' CNNs. [Si15] applies the SpoofNet-style architecture to classify clear soft contacts, textured contacts, and no contacts. In the LivDet 2017 challenge [Ya17b], one entry is a 7-layer CNN based on the GoogLeNet/Inception architecture. [PB17] train a 5-layer CNN with 3-tuples of live and spoof images based on the triplet loss. Unlike [Ya17a] and [PB17], we consider transfer learning from pre-trained state-of-the-art (SotA) deep CNNs (DCNNs). Transfer learning has also been explored in [CR18] for iris spoof detection, however our approach differs in that we use only a portion of a larger pre-trained network. In this paper, we perform transfer learning with a subset of parameters from the VGG-16 network pre-trained on the ImageNet dataset to a smaller 8-layer CNN, allowing for a deeper yet still robust architecture despite the limited training data.

There is limited research on feature selection and fusion for liveness detection tasks. [Ga12] selects from and fuses 25 image quality features from iris images via Sequential Floating Feature Selection (SFFS) to classify print-based iris spoofing attacks. The winner of the LivDet-Iris 2013 [Ya14] competition used a combination of 14 features derived from the Gray-Level Co-occurence Matrix selected with SFFS [Se16]. [Ko16] fuse Zernike moments and LBPV features via an MLP. Our approach employs group sparsity - a regularization technique shown to have a variety of applications to neural networks including feature selection [Sc17, We16].

3 Methodology

Our model is composed of two main components: 1) the set of *feature networks* and 2) the *fusion network*. Each feature considered in our model is assigned as input to its corresponding feature network. The feature networks in our model are six MLPs (corresponding to the six features: BSIF, LBP, CoA-LBP, HoG, DAISY, and SID) and one CNN (the 8-layer VGG-based network). The output of each feature network is a *feature embedding* vector. The feature embeddings are concatenated to form the input into the fusion network. See Figure 2 for a high-level view of the architecture. The network is trained in two main phases: 1) pre-training and co-adaptation and 2) feature selection.

3.1 Pre-training and Co-adaptation

In order to ameliorate the burden of learning a large amount of parameters simultaneously, the feature networks are first independently pre-trained on the liveness detection task. Fully-connected layers with softmax classifiers are used to pre-train the feature networks. The softmax layers are removed once the feature networks have been pre-trained.

The architecture of the CNN used in our model mirrors the first seven layers of the VGG-16 network plus a 500-neuron fully-connected layer. This allows us to transplant the corresponding VGG-16 weights trained on ImageNet into our model. Our 8-layer CNN is henceforth referred to as VGG-8. The powerful, generic low-level feature extractors

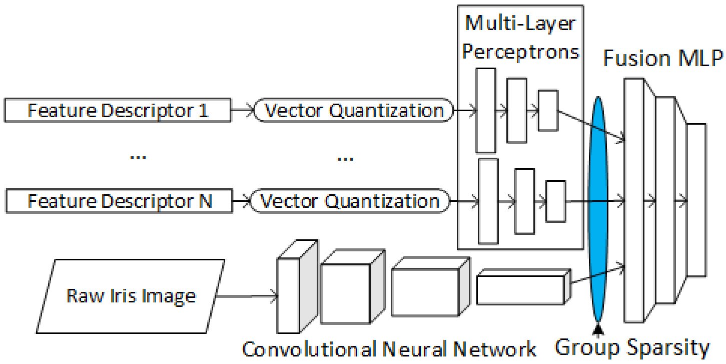

Fig. 2: High-level Architecture

adapted from VGG-16 aid in faster, more stable training as compared with a randomly initialized version of the same network, allowing us to achieve a deeper, more discriminative CNN.

After the feature networks have been pre-trained, the entire model is trained end-to-end solely on classification loss. The objective of this co-adaptation phase is to train the network to utilize all of the features as best as possible prior to inducing feature competition so that feature selection is not overly biased. In practice, however, we found the proposed network quickly achieves zero loss on the training data through attunement to only one feature - often the best performing feature. When this occurs, the network ceases to learn, despite the fact that incorporating additional features may increase performance on the test data. The parameters related to the other features are left prematurely developed and therefore at an undue disadvantage when feature selection begins. Potentially viable features, such as the VGG-8 representation, may require more time to train and co-adapt than is given, ultimately resulting in the potential deselection of a good feature.

To address this issue, we develop a novel variation of Dropout [Sr14]. Whereas Dropout operates individually on neurons within a layer, *group Dropout* affects the output of a feature network as whole. Each output feature embedding has a 50% chance to become the zero vector. As a result, the model not only learns to utilize each feature, but also learns to handle the inevitable absence of feature groups arising from the upcoming feature selection process.

3.2 Feature Selection through Group Sparsity

Let the first layer of the fusion network be denoted as the function $h(W, \mathbf{e}) \in \mathbb{R}^n$ where $\mathbf{e} \in \mathbb{R}^m$ is the concatenated feature embeddings and $W \in \mathbb{R}^{m \times n}$ is the layer's weight matrix. Given G feature networks, $W_g \in \mathbb{R}^{|\mathbf{e}_g| \times n}, \{(g, |\mathbf{e}_g|) : 1 \leq g \leq G, 1 \leq |\mathbf{e}_g| \leq m\}$, are the *feature group weights* pertaining to feature embedding $\mathbf{e}_g \in \mathbb{R}^{|\mathbf{e}_g|}$. The notation $|\mathbf{e}_g|$ refers to the size of the g-th embedding as each may vary in length. $||W_g||_2$ is the group ℓ_2 norm.

The group sparsity regularization is defined as

$$R(W) = \sum_{g=1}^{G} \frac{1}{|\mathbf{e}_g|} \left\|W_g\right\|_2 . \tag{1}$$

Given the classification loss E_D and with λ as a hyperparameter for controlling the amount of group sparsity regularization, the total loss is

$$L = E_D + \lambda R(W). \tag{2}$$

The minimization of the group sparsity regularization forces feature group weights to zero (in essence deselecting the feature), while the cross-entropy classification loss E_D mandates the network retain some capability to correctly classify images. The interplay between the two loss functions is the mechanism by which a set of best-performing features is selected.

4 Experiments

4.1 Datasets and Preprocessing

Our experiments employ the Clarkson Livdet 2013 [Ya14], the Notre Dame 1 and 2 [DBF13], and the Cogent and Vista IIITD Contact Lens [Ko16] datasets. In all cases, we follow the same steps for generating training and testing splits as established by the datasets' authors. See Table 1 for a breakdown of the datasets' statistics.

	Train		**Test**	
Dataset	Live	Spoof	Live	Spoof
Clarkson	270	400	246	440
Cogent*	1148	592	1158	610
Vista*	990	505	1020	560
ND1	2000	1000	800	400
ND2	400	200	200	100

Tab. 1: Number of images per dataset. *Based on a representative randomly generated fold.

We follow [Gr15]'s vector-quantization protocol for BSIF, LBP, CoA-LBP, SID, and DAISY descriptors. The HoG descriptor is reduced to a 200 dimensional vector via a PCA projection learned from a sample of the training data. The inputs into the six feature network MLPs are mean and variance centered from the training data. Images are resized to 224x224 and randomly horizontally flipped during training. Scikit-image is used to extract the HoG, LBP, and DAISY feature descriptors [Wa14].

4.2 Implementation Details

The feature and fusion networks are two-layer MLPs with layer sizes of [100, 50] and [500, 250] respectively. The VGG-8 network is the first 7 layers of the VGG-16 network followed by a 500-neuron fully-connected layer. Batch norm with decay = 0.9 is used for the VGG-8 layers. The activation function is ReLU. We use Adam optimizer with a learning rate = 0.001, beta1 = 0.9, beta2 = 0.999, epsilon = 1e-8. The group sparsity $\lambda =$ 0.001. Input mini-batches are generated with a balanced ratio of live and spoof samples for training. The network is implemented with the Tensorflow framework [Ab15]. Scikit-learn is utilized to perform vector quantization [Pe11]. The authors' code is made available at www.github.com/vonclites/livdet_fusion.

4.3 Results

As per ISO/IEC SC37 metrics, the Average Classification Error (ACE) is the average of the Attack Presentation Classification Error Rate and Bonafide Presentation Classification Error Rate. Table 2 compares the ACE of our model to the SotA as well as to the individual MLP feature networks after pretraining. Our model demonstrates superior performance versus the best known results obtained on each of the tested datasets. The Zernike+LBPV model and SotA for the Clarkson dataset represent the prior fusion-based approaches to textured lens detection.

Model	Clarkson	Cogent	Vista	ND1	ND2
BSIF MLP	21.19	26.87	19.27	-	-
CoA-LBP MLP	19.72	15.91	9.72	-	-
HoG MLP	26.26	25.42	7.04	2.25	10.75
LBP MLP	16.86	19.51	6.61	3.94	4.5
DAISY MLP	9.79	3.77	1.25	2.25	1.0
SID MLP	7.53	2.95	3.03	1.0	2.5
VGG_8 CNN	24.71	6.155	1.65	1.19	**0.0**
Zernike+LBPV	20.41 [Ko16]	-	-	-	-
SotA	18.17 [Ya14]	5.5 [PB17]	0.7 [PB17]	**0.1** [Gr15]	**0.0** [Gr15]
Ours	**3.25**	**1.57**	**0.22**	**0.1**	**0.0**

Tab. 2: Average Classification Error for our networks and SotA.

Figure 3 depicts the performance of our model alongside the gradual decline of the ℓ_2 norms (scaled by λ) of the feature group weights during training. As the network sparsifies and eliminates features, the average accuracy $(1 - \frac{ACE}{100})$ increases up until the point when key features are deselected. The influence of any given feature at various steps can be analyzed by re-running the model with the target feature embedding set to zero. For example, the change in accuracy at peak performance (step 10,000) from removing the LBP feature is zero. For BSIF or HOG, it is less than 0.001. Therefore, it can be concluded these features have been effectively deselected.

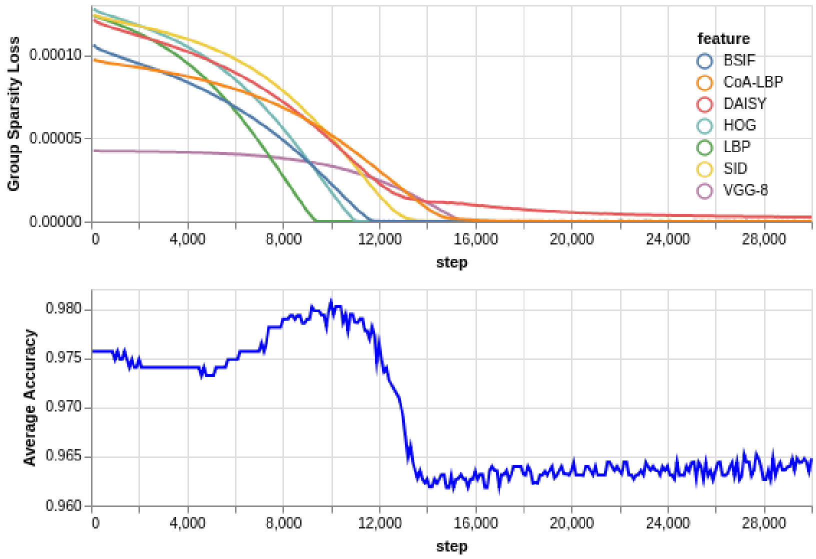

Fig. 3: Feature group weight ℓ_2 norms (top) and average accuracy (bottom) during training on a representative fold of the Cogent dataset. Co-adaptation training period not shown.

5 Conclusion

The abundance of strong features available for iris liveness detection allow for the assemblage of very powerful feature combinations. We have shown our integrated model achieves the goal of convenient, simultaneous feature fusion and selection while producing state-of-the-art results. Our generic framework is capable of incorporating new features and network architectures as they become available.

Yet, generalization to unseen datasets and types of spoofing attacks remains an unsolved problem. However, as additional data becomes available, neural network-based architectures become increasingly attractive. Nevertheless, further experimentation on the generalization capability of deep networks is a prudent avenue of future research in this field.

References

[Ab15] Abadi, Martín; Agarwal, Ashish; Barham, Paul; Brevdo, Eugene; Chen, Zhifeng; Citro, Craig; Corrado, Greg S.; Davis, Andy; Dean, Jeffrey; Devin, Matthieu; Ghemawat, Sanjay; Goodfellow, Ian; Harp, Andrew; Irving, Geoffrey; Isard, Michael; Jia, Yangqing; Jozefowicz, Rafal; Kaiser, Lukasz; Kudlur, Manjunath; Levenberg, Josh; Mané, Dandelion; Monga, Rajat; Moore, Sherry; Murray, Derek; Olah, Chris; Schuster, Mike; Shlens, Jonathon; Steiner, Benoit; Sutskever, Ilya; Talwar, Kunal; Tucker, Paul; Vanhoucke, Vincent; Vasudevan, Vijay; Viégas, Fernanda; Vinyals, Oriol; Warden, Pete; Wattenberg, Martin; Wicke, Martin; Yu, Yuan; Zheng, Xiaoqiang: , TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, 2015. Software available from tensorflow.org.

[CB18] Czajka, Adam; Bowyer, Kevin W: Presentation Attack Detection for Iris Recognition: An Assessment of the State of the Art. arXiv preprint arXiv:1804.00194, 2018.

[CR18] Chen, Cunjian; Ross, Arun: A Multi-Task Convolutional Neural Network for Joint Iris Detection and Presentation Attack Detection. In: 2018 IEEE Winter Applications of Computer Vision Workshops (WACVW). IEEE, pp. 44–51, 2018.

[DBF13] Doyle, J. S.; Bowyer, K. W.; Flynn, P. J.: Variation in accuracy of textured contact lens detection based on sensor and lens pattern. In: 2013 IEEE Sixth International Conference on Biometrics: Theory, Applications and Systems (BTAS). pp. 1–7, Sept 2013.

[DT05] Dalal, N.; Triggs, B.: Histograms of oriented gradients for human detection. In: 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05). volume 1, pp. 886–893 vol. 1, June 2005.

[Ga12] Galbally, Javier; Ortiz-Lopez, Jaime; Fierrez, Julian; Ortega-Garcia, Javier: Iris liveness detection based on quality related features. Proceedings - 2012 5th IAPR International Conference on Biometrics, ICB 2012, (LivDet 2009):271–276, 2012.

[Gr14] Gragnaniello, Diego; Poggi, Giovanni; Sansone, Carlo; Verdoliva, Luisa: Contact Lens Detection and Classification in Iris Images through Scale Invariant Descriptor. In: 2014 Tenth International Conference on Signal-Image Technology and Internet-Based Systems. IEEE, pp. 560–565, nov 2014.

[Gr15] Gragnaniello, Diego; Poggi, Giovanni; Sansone, Carlo; Verdoliva, Luisa: An Investigation of Local Descriptors for Biometric Spoofing Detection. IEEE Transactions on Information Forensics and Security, 10(4):849–863, 2015.

[Ko13] Kohli, Naman; Yadav, Daksha; Vatsa, Mayank; Singh, Richa: Revisiting iris recognition with color cosmetic contact lenses. Proceedings - 2013 International Conference on Biometrics, ICB 2013, 1, 2013.

[Ko16] Kohli, Naman; Yadav, Daksha; Vatsa, Mayank; Singh, Richa; Noore, Afzel: Detecting medley of iris spoofing attacks using DESIST. 2016 IEEE 8th International Conference on Biometrics Theory, Applications and Systems (BTAS), pp. 1–6, 2016.

[KR12] Kannala, Juho; Rahtu, Esa: Bsif: Binarized statistical image features. In: Pattern Recognition (ICPR), 2012 21st International Conference on. IEEE, pp. 1363–1366, 2012.

[Me15] Menotti, David; Chiachia, Giovani; Pinto, Allan; Robson Schwartz, William; Pedrini, Helio; Xavier Falcao, Alexandre; Rocha, Anderson: Deep Representations for Iris, Face, and Fingerprint Spoofing Detection. IEEE Transactions on Information Forensics and Security, 10(4):864–879, apr 2015.

[NOF11] Nosaka, Ryusuke; Ohkawa, Yasuhiro; Fukui, Kazuhiro: Feature extraction based on co-occurrence of adjacent local binary patterns. In: Pacific-Rim Symposium on Image and Video Technology. Springer, pp. 82–91, 2011.

[OPM02] Ojala, T.; Pietikainen, M.; Maenpaa, T.: Multiresolution gray-scale and rotation invariant texture classification with local binary patterns. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(7):971–987, July 2002.

[PB17] Pala, Federico; Bhanu, Bir: Iris Liveness Detection by Relative Distance Comparisons. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 162–169, 2017.

[Pe11] Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Prettenhofer, P.; Weiss, R.; Dubourg, V.; Vanderplas, J.; Passos, A.; Cournapeau, D.; Brucher, M.; Perrot, M.; Duchesnay, E.: Scikit-learn: Machine Learning in Python. Journal of Machine Learning Research, 12:2825–2830, 2011.

[Sc17] Scardapane, Simone; Comminiello, Danilo; Hussain, Amir; Uncini, Aurelio: Group sparse regularization for deep neural networks. Neurocomputing, 241:81–89, jun 2017.

[Se16] Sequeira, Ana Filipa Pinheiro: Liveness detection and robust recognition in iris and fingerprint biometric systems. PhD thesis, Universidade do Porto (Portugal), 2016.

[Si15] Silva, Pedro; Luz, Eduardo; Baeta, Rafael; Pedrini, Helio; Falcao, Alexandre Xavier; Menotti, David: An Approach to Iris Contact Lens Detection Based on Deep Image Representations. Brazilian Symposium of Computer Graphic and Image Processing, 2015-Octob:157–164, 2015.

[Sr14] Srivastava, Nitish; Hinton, Geoffrey; Krizhevsky, Alex; Sutskever, Ilya; Salakhutdinov, Ruslan: Dropout: A simple way to prevent neural networks from overfitting. The Journal of Machine Learning Research, 15(1):1929–1958, 2014.

[TLF10] Tola, Engin; Lepetit, Vincent; Fua, Pascal: Daisy: An efficient dense descriptor applied to wide-baseline stereo. IEEE transactions on pattern analysis and machine intelligence, 32(5):815–830, 2010.

[Wa14] van der Walt, Stéfan; Schönberger, Johannes L.; Nunez-Iglesias, Juan; Boulogne, François; Warner, Joshua D.; Yager, Neil; Gouillart, Emmanuelle; Yu, Tony; the scikit-image contributors: scikit-image: image processing in Python. PeerJ, 2:e453, 6 2014.

[We16] Wen, Wei; Wu, Chunpeng; Wang, Yandan; Chen, Yiran; Li, Hai: , Learning Structured Sparsity in Deep Neural Networks, 2016.

[Ya14] Yambay, David; Doyle, James S.; Bowyer, Kevin W.; Czajka, Adam; Schuckers, Stephanie: LivDet-iris 2013 - Iris Liveness Detection Competition 2013. In: IEEE International Joint Conference on Biometrics. IEEE, pp. 1–8, sep 2014.

[Ya17a] Yambay, D.; Walczak, B.; Schuckers, S.; Czajka, A.: LivDet-Iris 2015 - Iris Liveness Detection Competition 2015. In: 2017 IEEE International Conference on Identity, Security and Behavior Analysis (ISBA). pp. 1–6, Feb 2017.

[Ya17b] Yambay, David; Becker, Benedict; Kohli, Naman; Yadav, Daksha; Czajka, Adam; Bowyer, Kevin W.; Schuckers, Stephanie; Singh, Richa; Vatsa, Mayank; Noore, Afzel; Gragnaniello, Diego; et al.: LivDet iris 2017 Iris liveness detection competition 2017. In: 2017 IEEE International Joint Conference on Biometrics (IJCB). IEEE, pp. 733–741, oct 2017.