Towards Accessories-Aware Ear Recognition

Žiga Emeršič*, Nil Oleart Playà†, Vitomir Štruc‡ and Peter Peer*
*Faculty of Computer and Information Science, University of Ljubljana, Slovenia, EU
E-mail: {ziga.emersic, peter.peer}@fri.uni-lj.si
†Barcelona School of Telecommunications Engineering, Universitat Politcnica de Catalunya, Spain, EU
E-mail: nil.oleart@alu-etsetb.upc.edu
‡Faculty of Electrical Engineering, University of Ljubljana, Slovenia, EU
E-mail: vitomir.struc@fe.uni-lj.si

***Abstract*—Automatic ear recognition is gaining popularity within the research community due to numerous desirable properties, such as high recognition performance, the possibility of capturing ear images at a distance and in a covert manner, etc. Despite this popularity and the corresponding research effort that is being directed towards ear recognition technology, open problems still remain. One of the most important issues stopping ear recognition systems from being widely available are ear occlusions and accessories. Ear accessories not only mask biometric features and by this reduce the overall recognition performance, but also introduce new non-biometric features that can be exploited for spoofing purposes. Ignoring ear accessories during recognition can, therefore, present a security threat to ear recognition and also adversely affect performance. Despite the importance of this topic there has been, to the best of our knowledge, no ear recognition studies that would address these problems. In this work we try to close this gap and study the impact of ear accessories on the recognition performance of several state-of-the-art ear recognition techniques. We consider ear accessories as a tool for spoofing attacks and show that CNN-based recognition approaches are more susceptible to spoofing attacks than traditional descriptor-based approaches. Furthermore, we demonstrate that using inpainting techniques or average coloring can mitigate the problems caused by ear accessories and slightly outperforms (standard) black color to mask ear accessories.**

***Index Terms*—ear accessories, accessories removal, ear recognition, ear biometrics, biometrics**

I. Introduction

Ear recognition techniques have several advantages over recognition approaches using competing biometric modalities as emphasized in a recent survey [1]. However, some recent studies, such as [2], suggest that partial occlusions of the ear region, the presence of ear accessories and variable facial poses under which the images are captured are three of the main factors adversely affecting ear recognition performance. The absence of dedicated mechanisms for accounting for the presence of ear accessories in particular is one of the leading shortcomings of existing ear recognition techniques.

Ear accessories contain their own identifiable characteristics. This means that they not only occlude usable information, i.e. biometric ear traits, but also present a feature-rich source of non-identity related data. To make matters worse, such accessories can easily be replicated and used in so-called presentation attacks where the goal is to spoof ear recognition systems. Consequently, the common use of ear accessories poses one of the most problematic aspects of ear recognition.

Due to the fact that ear accessories themselves contain certain features that can be used to identify a subject (e.g., a person carries the same earrings in all enrollment images), an classification model can implicitly learn to use the information from the accessories to distinguish between subjects. This can lead to two types of problems, P1 and P2:

- *P1:* A person does not wear the correct (or any) type of ear accessories in the probe image and the classification model fails to recognize the person. This problem affect the overall recognition performance of the classification model and limits in usability.
- *P2:* A person wears the same type of ear accessories as used by some other person during the enrollment stage and gets recognized as the person originally wearing the ear accessory. This characteristic (P2) can be exploited within a presentation-type of attack [3] on the classification model, where an attacker tries to impersonate a target identity and, consequently, affects the security of an ear recognition system.

To alleviate the outlined issues, ear recognition systems need to be aware of the presence of ear accessories and incorporate mechanisms to ignore them when performing recognition. In related fields, such as face recognition, such mechanisms, e.g., [4]–[6], typically first detect the occluded areas and then either remove them from the image or replace them with suitable surrogates. Such an approach is expected to circumvent problem (P2), because features related to ear accessories are removed and the basis for presentation attacks is eliminated from the image. However, the first problem (P1) still persists and measures need to be taken to ensure that ear accessories are not taken into account when learning recognition models, so accessories need to be removed from the training (and enrollment) images as well.

In general, there are multiple options for ear accessories removal. A straight forward approach is to replace image regions corresponding to ear accessories with a uniformly colored patches (e.g., black or white patches). Another possibility is trying to incorporate information from the area surrounding the ear accessories into the surrogate region using inpainintg techniques. Both of these approaches have certain advantages

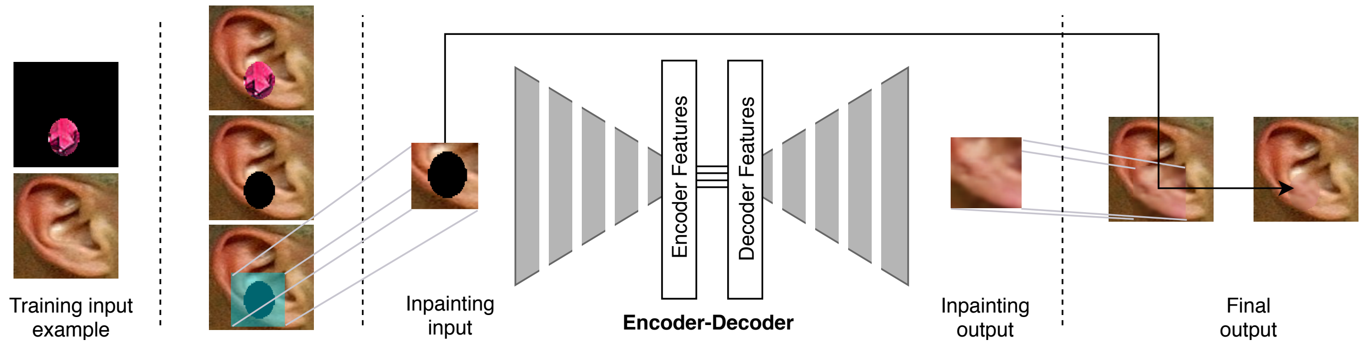

Fig. 1: Illustration of the inpainting process used in this work. We assume that the location and shape of the ear accessories is known and try to replace the ear accessory in the input image with a surrogate region produced by a context encoder. The result (on the out most right) is an inpainted images where the ear accessory has been removed.

and disadvantages

Using uniformly colored regions, for example, can introduce new edges and shapes that the recognition system can erroneously use to distinguish and classify samples. However, this can also happen when replacing accessories with realistically-looking textures. Such textures can introduce new, non-biometric features that the classification system can wrongfully assume to be important. The resulting images may look better to the human observer compared to images masked with uniformly colored areas, but can again adversely affect the recognition process. Due to this issues we evaluate both approaches in this work (i.e., inpainting and masking).

Generally speaking, there are two goals when performing ear accessories removal:

- *Noise removal:* The first goal is to remove all non-biometric features, while not introducing new class-biased information. This can be evaluated by observing the recognition performance which should not change significantly compared to images without ear accessories.
- *Realistic appearance:* In some domains, keeping realistically looking images of ears is preferable to introducing artificial blobs and other non-naturally looking areas – this is important if we plan to use the processed images with removed earrings in some other, already developed systems, e.g. we wish to perform accessories removal before inputting the image into an ear segmentation system, or would like to use it in conjuncture with a face detector. In cases like this it is key that the replaced areas appear natural and do not stand out.

Our goal in this work is to answer two basic research questions related to ear accessories removal, which, to the best of our knowledge, have not been addressed in the literature before: 1) Do accessories removal techniques help recognition performance? and 2) Which methods of ear accessories removal work best? When trying to remove accessories from ear images, the first step is typically to detect whether ear accessories are present in the input image in the first place and to localize the accessory region using segmentation techniques [7], [8]. However, evaluating the potential of accessories removal techniques in conjunction with detection and segmentation may lead to biased results, as the detection/segmentation performance also affects accessories removal. We, therefore, limit ourselves to oracle-type of experiments in this work and assume that the location and shape of the accessories is already known and binary masks exist that can be used to conceal the ear accessories. In practice, different detection and segmentation approaches can be used in the final recognition pipeline to generate the actual masks - see, e.g., [7]–[9].

The main contributions we make in this work are:

- We conduct (for the first time) and empirical investigation into the effects of ear accessories on the performance of ear recognition techniques and study the possibility of using accessories for presentation attacks. We use recent convolutional-neural-network-based (CNN-based) approaches as well as traditional ear recognition approaches for our experiments.
- We present a novel approach to ear accessories removal using CNN-based inpainting [10] of textures that yields realistically looking images. A diagram of the process is illustrated in Fig. 1.
- We introduce a novel dataset of ear images with annotated ear accessories containing artificially generated and superimposed earrings that is made freely available to the research community from http://ears.fri.uni-lj.si.

II. Related Work

Although ear recognition has gained on popularity in recent years [11]–[15], ear accessories and their effects on recognition has, to the best of our knowledge, not yet been studied comprehensively in the literature. This is in stark contrast to other biometric problems, where studies regarding occlusions and accessories are common and have contributed significantly to the performance and most of all robustness of biometric systems [4], [5], [16]–[19]. In [20], [21], for example, the authors studied the effect of the presence of face accessories, such as eyeglasses, scarfs and hats, on face recognition performance using Local Binary Patterns (LBPs). These works are related to ours in the sense that the authors address a conceptually similar problem with similar characteristics.

Many solutions to the problem of texture replacement (also tailored specifically toward accessories removal) have been proposed in the literature. In the context of biometric recognition, these have most commonly been applied for face recognition. Early approaches were based on a statistical analysis of the textures surrounding the accessories areas [22], [23], whereas more recent inpainting methods are mostly based on CNNs [5], [10], [16], more traditional approaches [4] or a combination ob both, such as [24].

However, the goal of ear accessories removal is in essence to remove discriminative features from accessories and replacing them with artificially generated surrogates [25], produced for example by generative neural networks (GNNs). Along these lines, the authors of [5], [16] use generative adversarial networks (GAN) for image inpainting. GANs, however, tend to produce unrelated data with high probability, if not constantly constrained by the unwanted image [16]. Auto-encoders, on the other hand, tend to generate overly smoothed images [16]. One of the possible solutions to this problem is to condition the auto-encoder on the corrupted images and using a so-called context encoder. In [10], the authors use such an context encoder to predict missing data and generate sharp and realistically looking surrogate images. We use a similar idea and also build our inpainting process around context encoders.

III. Methodology

The key questions regarding ear-accessories-aware recognition are i) which accessories removal techniques are most suitable, ii) to what extent do removal techniques help, and iii) are accessories in ears images as problematic as they have been proven to be in other areas of biometrics? To answer these questions we conduct a number of experiments using the methodology presented in the remainder of this section.

A. Dataset Preparation

We first built a suitable dataset for our experiments using a subset of images from the Unconstrained Ear Recognition Challenge (UERC) test dataset [7]. The UERC dataset was first presented as a test dataset for the Unconstrained Ear Recognition Challenge held in conjunction with the International Joint Conference on Biometrics (IJCB) 2017. It contains images acquired from the Internet, which exhibit variability in terms of illumination conditions, different pose angles, occlusions, and most importantly accessories. However, because the subset of images that contain accessories is small (i.e. 189 images) we generate an artificially augmented dataset where a set of different earrings is superimposed over the existing ear images. There are nine source images of earrings, but are always resized and changed in color, resulting in many different variations. A few examples of the resulting images are shown in Fig. 2.

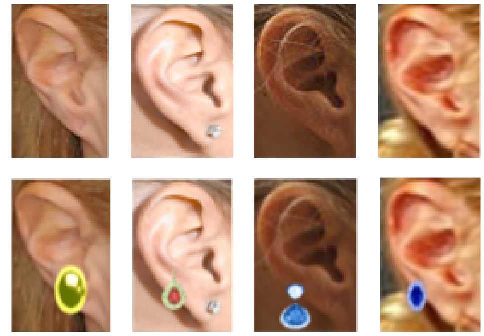

Fig. 2: Sample images from the generated dataset. The top rows shows original images from the UERC dataset and the lower row shows images with artificially superimposed earrings.

In order to follow realistic earrings positions, the earrings are generated in locations using the following rules:

- the vertical position of the earring is selected randomly in the lower half of the image,
- the horizontal position is set randomly, so that the earrings are at least 20% of the image width away from the left and right image border,
- the earring size is set randomly for the height and width with ratios ranging from 0.8 to 1.5 of the original earring dimensions.

With the outlined procedure we can generate a practically infinite amount of ear images, however, for our experimental dataset we limited to the original size of 4,104 ear images. Since the images are artificially generated, a binary mask indicating the location and shape of the ear accessories is also included in the dataset for each image. The dataset is publicly available from: http://ears.fri.uni-lj.si.

B. Removal of Ear Accessories

We remove ear accessories by replacing the unwanted image pixels corresponding to earrings with three different approaches:

- *Inpainting:* replacement of the ear accessories with naturally looking surrogate regions,
- *Fixed-color:* replacement of the ear accessories with black overlay color - this is the most straight-forward approach, and
- *Adaptive-color:* replacement of the ear accessories with the average color of the image – to avoid too sharp edges at the borders of the new area while still ensuring the absence of unwanted accessory features.

We describe all three approaches in detail in the remainder of this section.

Replacing accessories with CNN-based inpainting: For the accessories removal we use Context Encoders initially proposed in [10]. The reason we rely on context encoders is that they ensure realistic surrogate regions for the inpainting process and that an open source implementation is readily available from https://github.com/jazzsaxmafia/Inpainting.

Context encoders represent CNN-based auto-encoders that are conditioned on the input image. They can easily be trained to generate the missing contents of an arbitrary input image, including ear images. The architecture consists of encoder-decoder pipeline, where the encoder uses an input image with missing areas and extracts features. The features are then fed to the decoder through the fully-connected layer which generates a complete image, consequently filling in the missing areas [10]. The three main parts of a context encoder are as shown in Fig. 1 and are the following:

- *The encoder:* derived from the first five layers of AlexNet [10], [26] that accepts 227×227 RGB inputs.
- *The fully connected intermediate layer:* a simplified fully-connected layer, where some part are not really fully connected for performance reasons.
- *The decoder:* uses the feature representations from the fully connected layer in five upsampling and convolutional layers to generate the final inpainting output [27].

For the loss function we use an adversarial loss within a GAN framework. This means the context encoder is consecutively learning an adversarial discriminative model D to provide loss gradients to the generative model. The learning procedure is a process where the adversarial discriminator D takes in the prediction of the generator G and the ground truth, and then tries to distinguish between them. At the same time G tries to confuse D by producing samples that appear more an more realistic (i.e., similar to the actual ground truth) as the training procedure progresses [10].

The pipeline for the inpainting method is shown in Fig. 1 and is the following:

1) In the first step, the mask of the ear accessory is defined - here we assume this is known, but in practice the mask can be generated automatically using techniques, such as [8], [9].
2) Next, the bounding box for the ear accessory is defined. The area of this bounding box is stored as a new image and used with the CNN inpainting model. During training, these cropped areas (without the ear accessories) are used to train the model – both the original cropped area and the area with the overlaid mask are needed by the training procedure. At test-time the rectangular area with the masked-out ear accessory serves as an input for the inpainting model.
3) Once the masked and cropped region is processed by the inpainting model, the resulting inpainted image reinsert into the original image.
4) In the last step, the prediction is merged with the ear-accessory mask to produce smoother baounderies in the final inpainted ear image.

Our inpainting pipeline can be seen to be related to [28], where the authors used GNNs for face deidentification. In their pipeline they swap originals with generated faces. This is analogue to our pipeline, however instead of faces we are "deidentify" accessories and instead of GNNs we use context encoders.

Replacing accessories with uniformly colored areas: Because inpainting produces realistically looking areas it also means that potentially unwanted features are created. The premise for using uniformly-colored patches (or masks) is that no new features are introduced. Using areas like these, the classification system has no non-biometric information to rely on, that would adversely impact the recognition performance. This is the reason that we include these options in our tests. However, although these areas contain no textures, they still have certain shapes and edges. Due to this reason we consider two options:

- The whole masked region is set to 0, resulting in black color everywhere where ear accessories used to be (i.e., fixed-color replacement),
- The whole masked region is colored uniformly with the average pixel value of the whole image. The area corresponding to the ear accessory is omitted from the average color information and each color channel average is calculated separately (i.e., adaptive-color replacement).

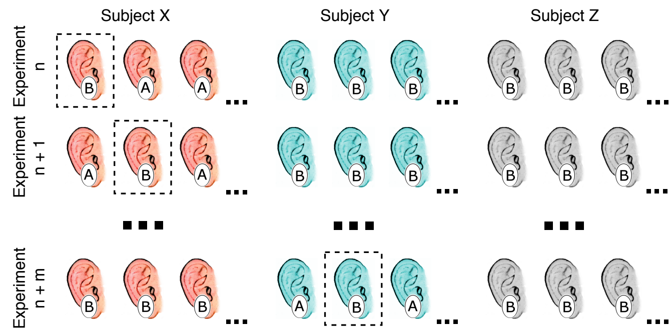

Fig. 3: Illustration of the experimental setup for the spoofing-attack experiments. For each identification attempt the earring of the probe image is copied over to the images all other gallery identities, whereas the gallery images of the same subject get a different earring. The dotted line indicates the probe image in the current identification attempt. Subjects are color coded and accessories are marked with letters (best viewed in color).

C. Recognition

We use two feature representations for our recognition experiments. The first is based on a CNN-based recognition model exploiting a VGG model architecture that was already used as a baseline for UERC [7]. The second representation relies on Local Binary Patterns (LBP) and is again taken from the UERC toolkit. The parameters of both feature extractors were left to the default values (as shipped with the UERC toolkit) and were not altered for the experiments, as our goal was not to optimize recognition performance itself, but to study the effect of ear accessories and different replacement/removal techniques on two representative ear-recognition techniques.

CNN-based recognition with VGG-16: The VGG network (or model) [29] is an example of so-called very deep CNN models and in the most common configuration consists of a total of 16 network layers (VGG-16). The VGG model has been successfully applied to numerous recognition problems, including ear recognition (see e.g., [7], [11], [30], [31]) and has been shown to ensure competitive performance on challenging ear datasets. Although newer architectures were shown to outperform the model for certain recognition tasks [32], we select VGG-16 for our tests to show how one the most widely used CNN architectures performs with accessories removal techniques and how susceptible it is to presentation attacks based or ear accessories.

In our experiments we use the output of the second fully connected layer (referred to as *FC6*) of the VGG-16 model as our image representation, instead of feeding the features forward to the next fully connected layer (textitFC7) and the final softmax layer at the end. This enables us to use VGG-16 open-set recognition problems, where identities in the test set are different from the identities used to train the model. To measure similarities we use the cosine distance in the experiments.

LBP-based recognition: Local Binary Patterns (LBP) [33] represent one of the most popular (hand-crafted) feature-extraction methods used for recognition purposes [1], [34]–[36]. We used the implementation available as a part of the AWE [1] and UERC [7] toolkits. The use of the LBP descriptor for ear recognition is mainly motivated by its computational simplicity and the fact that the texture of the ear is highly discriminative. Many successful ear recognition techniques have been presented in the literature exploiting LBPs either as stand-alone texture representations or in combination with other techniques, e.g., [37]–[39]. The technique used in this work uses uniform LBPs (with a radius of 1 and 8 neighbors) extracted from partially overlapping image blocks as an image representation and again the cosine distance for similarity measurements. A more detailed description of the approach is available from [7].

IV. Experiments & Results

In this section we describe the experiments used to evaluate the impact of ear accessories on ear recognition techniques and assess the usefulness of different accessories-removal techniques.

Experimental setup. We perform two types of experiments:

- *Standard identification experiments* (1:N matching, where N denotes the number of identities), where either the original UERC ear images (without artificial accessories) or images with superimposed ear accessories are used. The goal of these experiments is to establish the baseline performance for the tested recognition techniques and evaluate the impact of the presence of ear accessories on the recognition performance.
- *Spoofing attempt experiments* in an identification scenario, where ear accessories, or more precisely earrings, are copied from the probe to the gallery images to artificially increase the similarity of the images between subjects. Specifically, for each identification test the earring of the given probe image is superimposed onto the gallery images of all target subjects. To make the problem

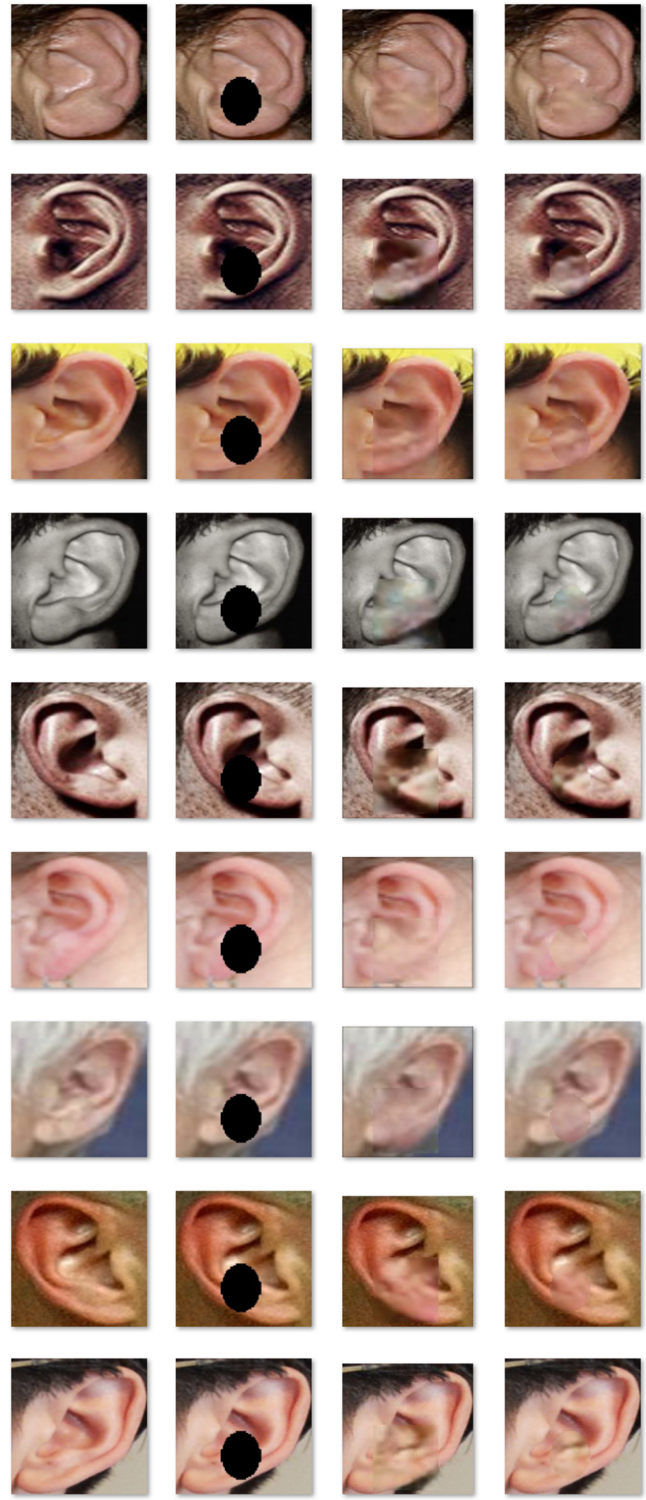

Fig. 4: Sample inpainting results. For easier comparison, inpainting in these examples was done on the same region each time. The first column contains the original images, the second column are the input images, the third are the outputs of the inpainting model, and the last column shows the outputs of the pipeline after masking.

TABLE I: Comparison of the Rank-1, Rank-5 and AUCMC score generated during the experiments. The approach most affected by presentation attacks is highlighted in gray and the best performing accessories removal approach for each method (VGG-16 vs. LBP) and performance metric (Rank-1 vs. Rank-5 vs. AUCMC) is shown in bold.

Exp. type	Data	Method	Rank-1	Rank-5	AUCMC
Standard	Original	VGG-16	12.06	28.85	82.66
		LBP	14.90	29.91	77.28
	Accessories	VGG-16	4.89	15.06	71.08
		LBP	12.84	27.29	75.90
Spoofing	Attack	VGG-16	0.00	0.00	1.05
		LBP	3.84	9.06	55.98
	Inpainting	VGG-16	10.45	**25.85**	**81.50**
		LBP	**14.29**	27.96	**76.57**
	Fixed color	VGG-16	9.01	22.79	78.69
		LBP	13.62	28.02	75.78
	Adapt. color	VGG-16	**11.01**	24.68	80.47
		LBP	14.01	**28.46**	76.13

harder and capitalize on the role of ear accessories, a different earring is placed over the gallery images of the true identity (i.e., the gallery images corresponding to the identity of the probe image). The overall idea of this experiment is illustrated in Fig. 3. The goal of this series of experiments is to test how susceptible the two feature extraction approaches are to accessory-based presentation attacks and evaluate how useful accessory removal techniques are as spoofing counter measures.

We adopt a similar *all-vs-all* protocol as for UERC, where a total of 1800 images belonging to 180 subjects is used for the experiments. Each of the 1800 images is used once as the probe and is matched against all remaining 1799 images for each identification attempt. The dataset contains an additional (subject disjoint) training set of $2,304$ images that are used to train the inpainting network. This training set is further divided into train and validation set for the learning procedure with the ratio of $7:3$ – $1,728$ images in the training and 576 images in the validation set. The inpainting model is trained from scratch on a desktop PC with a Titan Xp GPU using stohastic gradient descend (SGD) with a learning rate of 0.002, weight decay rate of 10^{-5}, and a momentum 0.9.

Qualitative evaluation of the inpainting model. Some sample outputs of the inpainting model are shown in Fig. 4. Note that the output in the last column look reasonably realistic, a closer look, however, reveals some features that are different from the ground truth on the far left. This means that the inpainting produces results in images that look realistic to a human observer, but are not necessary more useful for recognition purposes than the more simple maksing techniques (using either fixed or adaptive color masking). This observation is also validated by the experimental results presented in the next section.

Recognition experiments. For the recognition (identification) experiments, we use all three accessories removal strategy introduced in Section III-B. We report the results in terms of the rank-1 and rank-5 recognition rates as well as the normalized area under the Cumulative Match Score Curve (AUCMC) similarly to [7]. We also provide CMC curves of the experiments for a more detailed picture of the performance.

From the results in Table I and Fig. 5 we see that the LBP-based technique performs slightly better than the VGG model on the original UERC images without any ear accessories (see row of Table I labeled *Original*). Once accessories are randomly added (row of Table I labeled *Accessories*) to the images, the recognition performance drops for both techniques suggesting that accessories have in general an adverse effect on the recognition performance. The observed performance drop is significantly larger for the VGG model than for the LBP-based technique, which shows that accessories affect the learned features to a larger extent than the hand-crafted features.

The impact of the ear accessories becomes even more extreme in the case of presentation attacks, where the earrings in the gallery were intentionally copied over from the probe images to simulate spoofing attempts. In this case (see Table I row labeled as *Attack*), the features generated by the VGG-16 network are rendered virtually useless, as the rank-1 as well as the rank-5 recognition rate both drop to 0. For the LBP-based approach, the drop in performance is still in the range of 75%, but, nevertheless, not as extreme as with the CNN-based features. This is an important finding, as it shows that using traditional approaches based on hand-crafted features in real life deployment might still be viable, despite the recent advancements in deep learning. The results suggest that simply by matching ear accessories (which often cover a large are of the ear images) between probe and gallery images it is possible to spoof existing ear recognition systems.

We observe that accessories removal techniques can largely mitigate the impact of accessories on the recognition performance even within the more challenging spoofing scenario. There still exists a performance gap between the original images and the processed images with masked or inpainted accessories, which is slightly larger for the CNN-based features than for the hand-crafted ones, but overall all assessed accessories removal techniques help with the recognition performance significantly. It is interesting to observe that the inpainting technique, despite being computationally the most demanding and producing the best visual results, has only a slight advantage over the much simpler masking techniques. As we can see from the lower part of Table I and the CMC plots in Fig. 5, the difference in the recognition performance ensured by the three tested accessories removal techniques is minimal.

CONCLUSION

In this work, we showed for the first time that ear accessories present a real problem for ear recognition techniques that not only significantly affects performance, but can also be exploited as the basis for presentation attacks. We also showed

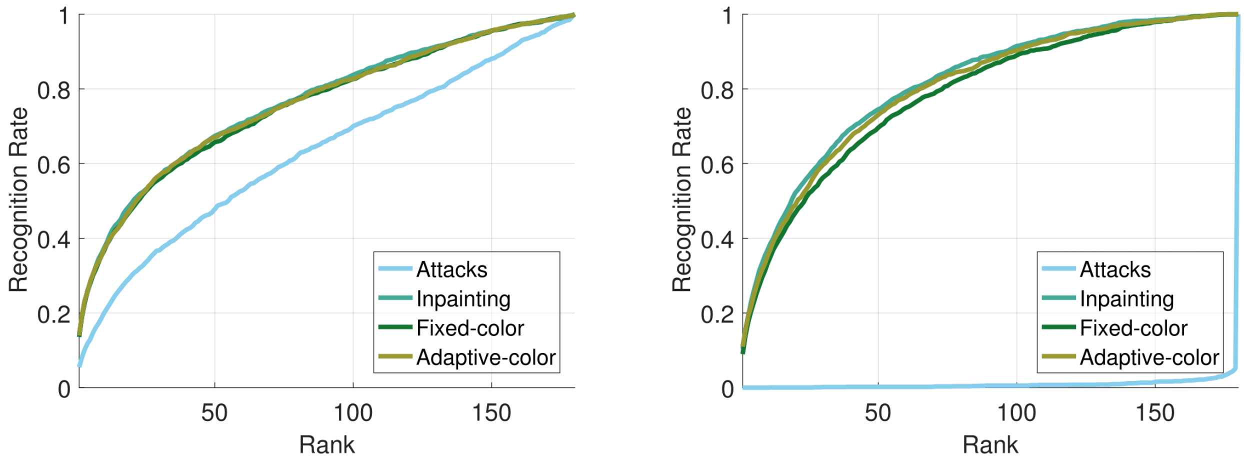

Fig. 5: Cumulative match score curves (CMC) generated during the spoofing experiments without any counter measures (denoted as *Attacks*) and with the three ear accessories removal techniques (denoted as *Inpainting, Fixed-color,* and *Adaptive-color* - see Section III-B for details) using: LBP features (left), and VGG-16 features (right). The plots show that the learned CNN-based features are much more susceptible to presentation attacks than the hand-crafted LBP features and that all three accessories removal technique represent effective counter measures, which ensure similar recognition performance.

that especially CNN-based approaches are highly susceptible to these kinds of attacks. Furthermore, our experiments suggested that accessories removal techniques can be used to efficiently mitigate the impact of accessories on ear recognition systems.

As part of our future work we plan to incorporate the best performing ear accessories removal technique into a complete ear recognition pipeline with together with an CNN-based ear accessories segmentation model that we are currently working on and that will automate all steps of the inpainting procedure.

Acknowledgements

This research was supported in parts by ARRS (Slovenian Research Agency) Research Program P2-0250 (B) Metrology and Biometric Systems, ARRS Research Program P2-0214 (A) Computer Vision, and the RS-MIZŠ and EU-ESRR funded GOSTOP. The Titan Xp GPU used for this research was donated by the NVIDIA Corporation.

References

[1] Ž. Emeršič, V. Štruc, and P. Peer, "Ear recognition: More than a survey," *Neurocomputing*, vol. 255, pp. 26–39, 2017.

[2] Ž. Emeršič, B. Meden, P. Peer, and V. Štruc, "Covariate analysis of descriptor-based ear recognition techniques," in *Bioinspired Intelligence (IWOBI), 2017 International Conference and Workshop on*. IEEE, 2017, pp. 1–9.

[3] J. Galbally, S. Marcel, and J. Fierrez, "Biometric antispoofing methods: A survey in face recognition," *IEEE Access*, vol. 2, pp. 1530–1552, 2014.

[4] M. Jampour, C. Li, L.-F. Yu, K. Zhou, S. Lin, and H. Bischof, "Face inpainting based on high-level facial attributes," *Computer Vision and Image Understanding*, vol. 161, pp. 29–41, 2017.

[5] Y.-A. Chen, W.-C. Chen, C.-P. Wei, and Y.-C. F. Wang, "Occlusion-aware face inpainting via generative adversarial networks," in *Image Processing (ICIP), 2017 IEEE International Conference on*. IEEE, 2017, pp. 1202–1206.

[6] V. Štruc, S. Dobrisek, and N. Pavesic, "Confidence weighted subspace projection techniques for robust face recognition in the presence of partial occlusions," in *International Conference on Pattern Recognition (ICPR)*, 2010, pp. 1334–1338.

[7] Ž. Emeršič, D. Štepec, V. Štruc, P. Peer, A. George, A. Ahmad, E. Omar, T. E. Boult, R. Safdari, Y. Zhou, S. Zafeiriou, D. Yaman, F. I. Eyiokur, and H. K. Ekenel, "The unconstrained ear recognition challenge," *International Joint Conference on Biometrics (IJCB)*, 2017.

[8] A. Ahmad, D. Lemmond, and T. E. Boult, "Chainlets: A new descriptor for detection and recognition," in *IEEE Winter Conference on Applications of Computer Vision (WACV)*. IEEE, 2018, pp. 1897–1906.

[9] Ž. Emeršič, L. L. Gabriel, V. Štruc, and P. Peer, "Convolutional encoder–decoder networks for pixel-wise ear detection and segmentation," *IET Biometrics*, vol. 7, no. 3, pp. 175–184, 2018.

[10] D. Pathak, P. Krahenbuhl, J. Donahue, T. Darrell, and A. A. Efros, "Context encoders: Feature learning by inpainting," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2536–2544.

[11] Y. Dogucan, E. Fevziye, and H. Ekenel, "Domain adaptation for ear recognition using deep convolutional neural networks," *IET Biometrics*, 2018.

[12] Y. Zhang, Z. Mu, L. Yuan, and C. Yu, "Ear verification under uncontrolled conditions with convolutional neural networks," *IET Biometrics*, vol. 7, no. 3, pp. 185–198, 2018.

[13] S. Dodge, J. Mounsef, and L. Karam, "Unconstrained ear recognition using deep neural networks," *IET Biometrics*, vol. 7, no. 3, pp. 207–214, 2018.

[14] E. E. Hansley, M. P. Segundo, and S. Sarkar, "Employing fusion of learned and handcrafted features for unconstrained ear recognition," *IET Biometrics*, vol. 7, no. 3, pp. 215–223, 2018.

[15] I. I. Ganapathi and S. Prakash, "3d ear recognition using global and local features," *IET Biometrics*, vol. 7, no. 3, pp. 232–241, 2018.

[16] R. A. Yeh, C. Chen, T. Y. Lim, A. G. Schwing, M. Hasegawa-Johnson, and M. N. Do, "Semantic image inpainting with deep generative models," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 5485–5493.

[17] M. Wang, Z. Hu, Z. Sun, S. Zhao, and M. Sun, "Varying face occlusion detection and iterative recovery for face recognition," *Journal of Electronic Imaging*, vol. 26, no. 3, p. 033009, 2017.

[18] Y. Wen, W. Liu, M. Yang, Y. Fu, Y. Xiang, and R. Hu, "Structured occlusion coding for robust face recognition," *Neurocomputing*, vol. 178, pp. 11–24, 2016.

[19] D. Lin and X. Tang, "Quality-driven face occlusion detection and recovery," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*, June 2007, pp. 1–7.

[20] R. Min, A. Hadid, and J.-L. Dugelay, "Improving the recognition of faces occluded by facial accessories," in *Automatic Face & Gesture Recognition and Workshops (FG 2011), 2011 IEEE International Conference on*. IEEE, 2011, pp. 442–447.

[21] ——, "Efficient detection of occlusion prior to robust face recognition," *The Scientific World Journal*, vol. 2014, 2014.

[22] H. Igehy and L. Pereira, "Image replacement through texture synthesis," in *Image Processing, 1997. Proceedings., International Conference on*, vol. 3. IEEE, 1997, pp. 186–189.

[23] Y. Tsin, Y. Liu, and V. Ramesh, "Texture replacement in real images," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, vol. 2. IEEE, 2001, pp. II–II.

[24] H. Fujishige, J. Miyao, and T. Kurita, "Image inpainting by recursive estimation using neural network and wavelet transformation," in *International Conference on Neural Information Processing*. Springer, 2017, pp. 652–661.

[25] B. Meden, Ž. Emeršič, V. Štruc, and P. Peer, "k-same-net: k-anonymity with generative deep neural networks for face deidentification," *Entropy*, vol. 20, no. 1, p. 60, 2018.

[26] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097–1105.

[27] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 3431–3440.

[28] B. Meden, R. C. Mall, S. Fabijan, H. K. Ekenel, V. truc, and P. Peer, "Face deidentification with generative deep neural networks," *IET Signal Processing*, vol. 11, no. 9, pp. 1046–1054, 2017.

[29] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[30] H. Earnest, P. Segundo, and S. Sarkar, "Employing fusion of learned and handcrafted features for unconstrained ear recognition," *IET Biometrics*, 2018.

[31] Ž. Emeršič, D. Štepec, V. Štruc, and P. Peer, "Training convolutional neural networks with limited training data for ear recognition in the wild," in *Proceedings of the 12th IEEE International Conference on Automatic Face and Gesture (FG 2017)*, 2017.

[32] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-v4, inception-resnet and the impact of residual connections on learning." in *AAAI*, vol. 4, 2017, p. 12.

[33] M. Pietikäinen, A. Hadid, G. Zhao, and T. Ahonen, *Computer Vision Using Local Binary Patterns*, ser. Computational Imaging and Vision. Springer, 2011.

[34] A. Pflug and C. Busch, "Ear biometrics: a survey of detection, feature extraction and recognition methods," *Biometrics*, vol. 1, no. 2, pp. 114–129, 2012.

[35] A. Abaza, A. Ross, C. Hebert, M. A. F. Harrison, and M. Nixon, "A Survey on Ear Biometrics," *ACM Computing Surveys*, vol. 45, no. 2, pp. 1–22, 2013.

[36] R. Purkait, "Role of External Ear in Establishing Personal Identity - A Short Review," *Austin Journal of Forensic Science and Criminology*, vol. 2, no. 2, pp. 1–5, 2015.

[37] A. Pflug, P. N. Paul, and C. Busch, "A comparative study on texture and surface descriptors for ear biometrics," in *Proceedings of the International Carnahan Conference on Security Technology*. IEEE, 2014, pp. 1–6.

[38] Y. Guo and Z. Xu, "Ear recognition using a new local matching approach," in *Proceedings of the International Conference on Image Processing*. IEEE, 2008, pp. 289–292.

[39] A. Benzaoui, A. Kheider, and A. Boukrouche, "Ear description and recognition using ELBP and wavelets," in *Proceedings of the International Conference on Applied Research in Computer Science and Engineering*, 2015, pp. 1–6.