- ORIGINAL ARTICLE -

# Automatic Ear Detection and Segmentation over Partially Occluded Profile Face Images

## Detección y Segmentación Automática de Oídos en Imágenes de Rostro con Vista Lateral Parcialmente Ocluıda.

Celia Cintas[1,2], Claudio Delrieux[2], Pablo Navarro[1,2,3], Mirsha Quinto-Sánchez[4], Bruno Pazos[1,2,3], and Rolando Gonzalez-José[1]

[1]*Instituto Patagónico de Ciencias Sociales y Humanas. Centro Nacional Patagónico, CONICET.*
{cintas, pnavarro, bpazos, rolando}@cenpat-conicet.gob.ar
[2]*Laboratorio de Ciencias de las Imágenes. Departamento de Ingeniería Eléctrica y Computadoras. Universidad Nacional del Sur y CONICET. cad@uns.edu.ar*
[3]*Departamento de Informática, Facultad de Ingeniería, Universidad Nacional de la Patagonia San Juan Bosco. Trelew-Chubut, Argentina*
[4]*Ciencia Forense, Facultad de Medicina, Universidad Nacional Autónoma de México.*
mirsha@cienciaforense.facmed.unam.mx

## Abstract

Automated, non invasive ear detection in images and video is becoming increasingly required in several contexts, including nonivasive biometric identification, biomedical analysis, forensics, and many others. In biometric recognition systems, fast and robust ear detection is a crucial step within the recognition pipeline. Existing approaches to ear detection are susceptible to fail in the presence of typical everyday situations that prevent a crisp imaging of the ears, like partial occlusions, ear accessories, or uncontrolled camera and illumination conditions. Even more, most of the proposed solutions work efficiently only within a previously detected rectangular region of interest, which limits their applicability and lowers the accuracy of the overall detection. In this paper we evaluate the use of Convolutional Neural Networks (CNNs) together with Geometric Morphometrics (GM) for automatic ear detection in the presence of partial occlusions, and a Convex Hull algorithm for the ear area segmentation. A CNN was trained with a set of ear images landmarked by experts using GM to achieve high consistency. After training, the CNN is able to detect ears over profile faces, even in the presence of partial occlusions. We analyze the performance of the proposed ear detection and segmentation method over partially occluded ear images using the CVL Dataset.

**Keywords:** biometrics, convex hull, deep learning, ear detection, occlusion

## Resumen

La detección automática del pabellón auditivo en imágenes y video, es una funcionalidad crecientemente requerida en varios contextos. Entre ellos podemos citar: identificación biométrica no invasiva, análisis biomédicos, estudios forenses, entre otros. En los sistemas de reconocimiento biométrico, la detección rápida y confiable del pabellón auditivo es un paso fundamental dentro del procesamiento. Las aproximaciones existentes con respecto a esta detección no son robustas, siendo susceptibles a fallas en la presencia de oclusiones parciales, accesorios como aros o piercings, o condiciones desfavorables en la cámara o la iluminación. Además, gran parte de los sistemas biométricos de la actualidad asumen que el dato de entrada será la región de interés que contiene el pabellón auditivo, lo cual limita su uso y reduce la exactitud global de reconocimiento. En este trabajo se evalúa el uso de redes convolucionales (*Convolutional Neural Networks o CNNs)* junto con Morfometría Geométrica para la detección automática del pabellón auditivo y la segmentación de los píxeles correspondientes al mismo mediante el uso de un algoritmo de *Convex Hull.* Luego del entrenamiento, la red CNN puede detectar el pabellón auditivo sobre imágenes de rostro en vista lateral, inclusive en la presencia de oclusiones parciales. Se analiza la performance del método de detección y segmentación de orejas sobre imágenes con oclusiones parciales correspondientes al conjunto de datos CVL.

**Palabras claves:** aprendizaje profundo, biometría, convex hull, detección de oidos, oclusiones

## 1 Introduction

Ears have several biometric advantages in recognition tasks over other anatomical structures such as fingerprints, iris patterns, or faces [1]. Ear shots can be easily taken with non intrusive methods (even from

afar). Also the ear's anatomical features vary only slightly with aging, and its shape is not influenced by facial expressions [2, 3]. However, ear images are subject to potential acquisition problems such as partial occlusion by hair, earrings, or earphones, which are almost certain to happen when images are taken in the open. Few fully automatic ear recognition algorithms have been proposed in the literature, mostly due to the lack of robust detection techniques capable to locate the ears in the input image, which is key to identify ears efficiently. According to Emersic et al. [4], the absence of automatic ear detection approaches is one of the most important factors hindering a wider deploy of ear recognition technology.

Traditional proposals in the literature take advantage of the ear's specific shape, frequently using engineered geometric features. This aims to detect specific border configurations often present in ears' images, like the occurrence of certain characteristic edges, curvature dispositions or frequency patterns, using image processing techniques. In general, since the ears' biometric properties are not fully leveraged in these procedures, the performance is only acceptable in strictly controlled acquisition contexts (illumination and camera position), not being robust under homographies or changes in luminance conditions. Also, these methods require several special case considerations.

A much less explored strategy for ear detection and recognition is to represent the ears' shape and phenotypic attributes in the form of landmark coordinates. In particular, landmarking based on Geometric Morphometrics (GM) provides a robust methodology for shape analysis and evaluation [5]. Manual landmarking, however, is not feasible for a massive sample, since it takes considerable supervised time, increases the likelihood of operation mistakes due to operator visual fatigue, intra- and inter-observer error, and is prone to distractions or confusions during the landmarking sequence. Automatic 2D or 3D landmark acquisition appears to be a promising venue to explore since it may overcome both limitations (the lack of robustness in most ear detection and recognition proposals, and difficulties associated to manual landmarking).

This work introduces a flexible and versatile method for automatic ear detection based on selection of 2D landmarks. Even though the main intended use of this method is on population and quantitative genomic, biomedical or forensic studies based on 2D data [6], it is easily adaptable to be useful in other contexts, like biometric identification. The main contributions of this paper are the following:

- An evaluation of the CNN over an open dataset, not previously used for training or validation.

- An analysis of the performance of the CNN over different occlusion settings over a new dataset with 219 images from [7] with the corresponding ROI annotated by the authors.

- Comparison of results with other proposed ear detection techniques in the literature, in particular the effect on detection quality under progressive occlusions.

- An ear segmentation technique based on Geometric Morphometrics and Convex Hull methods.

## 2 Related Work

This section briefly summarizes the state of the art in automatic ear detection in 2D images. A more thorough description of current advances in ear detection, feature extraction and biometric recognition methods can be found in [8] and [4]. Most ear detection approaches rely on shape properties of the external ear's morphology, like the occurrence of certain characteristic edges, curvature dispositions, or frequency patterns. Among the most widespread ideas, the use of shape models appears to be extensively used. Shape models aim to recognize specific distributions of shape descriptors that are frequent in the object under study, in this case the ear's surface. For instance, Chen and Bhanu [9] propose to detect image regions with large local curvatures with a technique they call step edge magnitude. Then, template matching is performed with typical shapes of the outer helix and anti-helix. Later, in [10] the number of possible ear candidates was narrowed by detecting skin regions first before the helix template matching is applied, also reducing spurious detections. This method, however, by its very nature is not robust under homographies, making it unsuitable for most applications where a careful and calibrated acquisition may not be performed. Following a similar shape-based approach, Attarchi et al. [11] use contour lines for ear detection. Their proposal locates first the outer contour of the ear using a search method that finds the longest connected edge in the region of interest. Once located, this contour can be used to define a triangle formed by the outermost points in the top, bottom and left positions of the contour. Finally, geometric properties of this triangle, for instance the barycenter, can be used as a reference point for image alignment. Although less prone to break under homographies, this method still requires noise-free and white-balanced images to perform adequately.

Another method, related to edge detection properties, was proposed by Ansari et al. [12]. First, they apply an edge detector in which the edges are marked as either convex or concave segments, since the most likely candidates for the ear's outer contour are convex edges. After that, the algorithm connects the contour segments and selects the shape enclosing the largest area as being the outer ear contour. Like other akin tracking algorithms, several special cases must be accounted for, thus leading to very complex algorithms. In a similar vein, Prakash and Gupta in [13] combine skin segmentation and hierarchy edges. After being detected, the edges located in the skin region

are decomposed into edge segments. An edge connectivity graph is constructed, integrating all these edge segments. The connectivity graph is finally used to compute the convex hull of the set of edge segments, which encloses the ear's outer shape. Also significant is the proposal of Yan and Browyer [14], who developed an ear detection method which starts by locating the concha (an anatomic part of the ear, see Fig. 1), which is set as the initial shape for an active contour used for determining the ear's outer boundary.

Pflug et al. [15] use a combination of depth images and texture. Their method starts with a preprocessing step, where edges and shapes are extracted from the texture and the depth image, and edges and shapes are fused together in the image domain. In the next step, the components are combined with each other to find ear candidates and rank them according to a computed score. Finally, the enclosing rectangle of the best ear candidate is returned as the ear region. Like the other methods already mentioned, the main disadvantage of these shape-model approaches is the fact that they require specifically engineered features, which makes them less flexible or adaptable to other detection problems, and also renders them fragile under homographies and luminance changes.

A different approach regards the ear detection problem as an instance of a pattern recognition problem instead of focusing on the unique geometric features of the ear, In this approach, the first stage uses image processing techniques to extract features present in the image, followed by a second stage in which pattern recognition techniques are applied over the feature set to perform detection tasks. This approach is in general more robust under homographies and luminance changes, depending on the feature space used for the ear representation in the first stage. Among the proposals based on pattern recognition approaches, we can mention Abaza et al. [16] and Islam et al. [17], which use weak classifiers based on Haar-wavelets over regions of the image to find correlation with previously learned patterns. These weak classifiers are then combined with a standard AdaBoost procedure for ear localization. Yuan et al. [18] propose a dictionary-based sparse representation and classification scheme, intended to work with partially occluded ear imagery. An identity occlusion dictionary encodes occluded parts in the source image to perform ear recognition. A non-negative dictionary that includes a Gabor feature-set extracted from ear images improves the sparseness of the coding representation, thus circumventing the expense of a conventional occlusion dictionary. In [19], Kumar et al. take advantage of the sparse representation of the finite (discrete) Radon transform based local orientation information. The neighborhood relationship of gray-levels in the normalized ear images is encoded as the dominant gray-level feature orientations in a local region using a local Radon transform. In [20] the authors develop an approach that encodes

reliable phase information using 2D quadrature filtering. They extensively evaluated both quaternionic and monogenic quadrature filters and develop a new quaternionic-code-based approach for the ear identification. These proposals based on pattern recognition techniques are more recent and tend to outperform shape-model methods.

## 3 Methods and Implementation

Given the aforementioned limitations of the current proposals in ear detection algorithms, we propose the use of Geometric Morphometrics together with Deep Learning algorithms, as presented in [6] for ear detection, and Convex Hull methods for posterior segmentation. A set of 2735 manually landmarked images, each with 45 interest points (landmarks and semi-landmarks), the landmark configuration is described in Fig. 1, was obtained to train a convolutional neural network, using specific learning techniques to achieve a high generalization rate and to avoid overfitting. An overview of the network structure used in this case is detailed in Table 1. For more information regarding the network training please refer to [6].

In this paper we used a subset from CVL Face Database [7], with 219 profile face images associated with a Region Of Interest (ROI) where the ear is located, used as a ground truth in this work. After the ear is landmarked the convex hull of these landmarks is calculated and used to segment the images resulting on a set of pixels corresponding only to the ear structure.

### 3.1 Geometric Morphometrics

Geometric Morphometrics (GM) provide a set of methods for the quantitative analysis of the size and shape of objects. GM is widely used in the study of biological organisms [23], specially humans [24]. Methods in GM propose to quantify the shape of each specimen according to the location in space of a set of 2D or 3D reference points or *landmarks* that are homologous across individuals. The specific configuration and anatomical descriptions of human ears are shown in Fig. 1.

### 3.2 Convolutional Neural Networks

In recent years, the computer vision literature has witnessed many research efforts in descriptor engineering. A sought-for advantage of these descriptors, when applied to recognition purposes, is that they require to use the same operator to all locations in the image. In this way, the design of workflows for specific recognition purposes is greatly simplified. Moreover, and as more data becomes available, learning-based methods are increasingly outperforming engineered features, because they can discover and optimize features without supervision for the specific task at hand [25, 26].
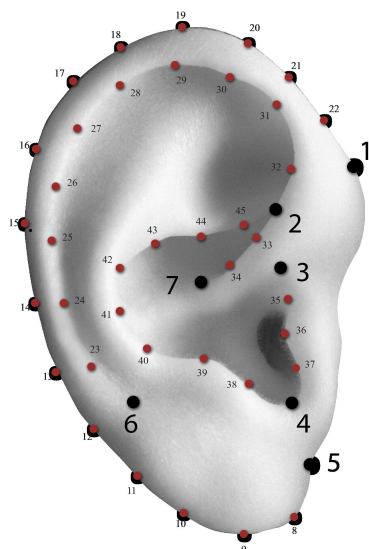
| Number | Name |
|---|---|
| 1 | Otobasion superiorious |
| 2 | Concha superiorious |
| 3 | Tragus superiorious |
| 4 | Intertragic incisure |
| 5 | Otobasion inferiorous |
| 6 | Helix basal border |
| 7 | Crus Helix |
| 8 to 45 | Semi landmarks |

Figure 1: Landmark and semi-landmark configuration and anatomical description [21, 22].

Convolutional neural networks (*CNNs*) [27, 28] constitute the state of the art in many computer vision problems, since they were shown to be very effective for large-scale image classification [26, 29, 30]. Their outstanding performance is based in four core concepts: local connections, shared weights, pooling, and the use of several layers [25]. However, since the amount of learnable parameters in these nets is huge, special care must be taken to avoid overfitting (*i.e.*, the network may just memorize the examples, without generalization). In CNNs, the connectivity patterns between some of the layers are constrained in a way such to facilitate the processing of input data that comes in the form of multiple arrays, for example 2D arrays containing pixel intensities, or 3D for video or volumetric images. Images commonly exhibit high correlation between values in a local group, forming distinctive local patterns that are easily detected. To take advantage of these properties, CNNs contain two types of layers: *convolutional* and *pooling* layers.

A convolutional layer is parametrized by a set of learnable filters. The feature maps are taken as input and then a convolution is applied to each with the set of filters to produce a stack of output feature maps. To reduce the dimensionality of the features maps, a pooling layer is located between convolutional layers. Pooling layers eliminate non-maximal values by

computing some aggregation function (typically the maximum or the mean) across small local regions of the input [31]. The main purpose of this pooling is to reduce the computational cost in the remaining layers, reducing the dimensionality of the feature maps and providing a form of translational invariance.

### 3.3 Dataset

Images used in this paper are a subset of the CVL Face Database [7]. The images consist of a lateral view of the head, with a $640 \times 480$ pixel resolution, taken with Sony Digital Mavica under uniform illumination conditions, and with background removal. The dataset contains 219 images (corresponding to 114 persons) from [7], each with an associate rectangle corresponding to the ROI (Region of Interest) where the ear is located, for ground truth purpose[1]. In some of the images, ears were partially occluded by hair, earrings or a combination of both. To further evaluate partial occlusion, black boxes were added at random locations in the images. An example of individual images of the dataset with the ROI associated can be seen in Fig. 2. It is worth mentioning that the network used in this paper was trained with a different dataset (privately owned by the CANDELA Consortium [32]), with different pixel resolution, non uniform background, and different and varying illumination conditions. In this way, the instances used in this paper were never seen by the network neither in the training nor in the validation stage.

### 3.4 Preprocessing

Prior to the CNN-based processing, images are converted to a single channel (i.e., luminance). The luminance histogram is stretched to black-out at most 2% of the pixels in the ROI, and to white-out at most 1% of the pixels. Then, the image is resampled to a final size of $96 \times 96$ pixels, using bilinear downsampling. For testing robustness with respect to partial occlusion, we generate black rectangles located randomly over the resampled image with different areas. The implementation of the occlusion method can be seen in the repository, and an example of the occluded image with a configuration of 5 boxes with a size of $15 \times 15$ pixels each one can be seen in Fig. 6.

### 3.5 CNN Architecture and Training

In Table. 1 the best performing architecture is shown. The input layer takes a single-channel profile face image of size $96 \times 96$ pixels, with brightness scaled to $[0, 1]$. This is followed by a convolution layer with square filters, and after a *max* pooling and dropout

---

[1]The images and ROI for the network can be downloaded from https://github.com/celiacintas/tests_landmarks, please note that the images belong to CVL Database and should be properly cited if used [7].
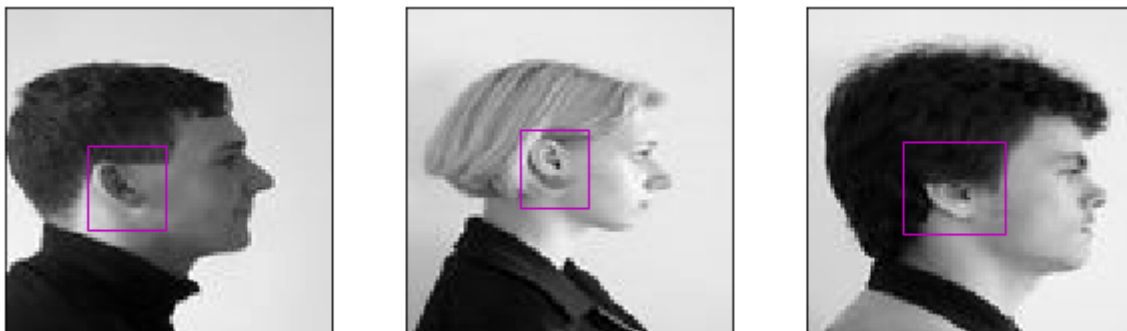
Figure 2: Sample of the images used, with the associated ROIs superimmposed.

Table 1: Structure and parameters used for building the CNN for automatic landmarking over profile faces.

| Name | Type | Size | Filters | Pool | Dropout | Units |
|------|------|------|---------|------|---------|-------|
| C1 | Convolutional | 4x4 | 32 | - | - | - |
| M2 | Max pooling | - | - | 2x2 | - | - |
| D3 | Dropout | - | - | - | 0.1 | - |
| C4 | Convolutional | 2x2 | 64 | - | - | - |
| M5 | Max pooling | - | - | 2x2 | - | - |
| D6 | Dropout | - | - | - | 0.2 | - |
| C7 | Convolutional | 2x2 | 128 | - | - | - |
| M8 | Max pooling | - | - | 2x2 | - | - |
| D9 | Dropout | - | - | - | 0.3 | - |
| F10 | Fully connected | - | - | - | - | 1024 |
| D11 | Dropout | - | - | - | 0.5 | - |
| F12 | Fully connected | - | 1024 | | | |
| O13 | Output | - | - | - | - | 90 |

layer. This structure is repeated three times to obtain features at different levels of abstraction, with different filter size, number of feature maps, and probability values. The convolutional layers C1, C4, C7 have 32, 64 and 128 filters of size $4 \times 4$, $2 \times 2$ and $2 \times 2$. All *max* pooling layers are of size $2 \times 2$, and the probability values used for D3, D6, D9 and D11 are (resp.) 0.1, 0.2, 0.3 and 0.5. After the feature extraction layers, the architecture contains two fully connected linear layers with 1024 units each (F10 and F12 in the diagram), and a dropout layer in between (D11). The output layer (O13) contains 90 output units (45 $[x, y]$ pairs) for the predicted position of the landmarks and semi-landmarks. The implementation used Python and the Lasagne library [33][2]. This allows the use of GPU acceleration without considerable programming effort. The training of the network took roughly 10 hours using NVIDIA GeForce GTX 1080 cards. Also, the learning curves showing the training set error and the validation set error can be seen at Fig. 3.

### 3.6 Pixel Segmentation based on Convex Hull

Usually ear detection methods return a rectangular ROI [34, 35], that contains the pixels corresponding to the ear among others from the background. This is rather impractical in several cases, since the amount of pixels that require ulterior processing (in a recognition step, for instance) is much larger, and also because the ROI could contain background undesired information that may add clutter. Using the landmarking configuration described in 1, instead, a more precise and practical ROI based on a convex hull can be calculated over landmarks. An example of this process can be seen in Fig. 4. A convex hull of a set of points $S$ in $n$ dimensions is the intersection of all convex sets containing $S$. For $N$ points $x_1, \ldots x_N \in S$ the convex hull $Conv(S)$ is given by:

$$\text{Conv}(S) = \left\{ \sum_{i=1}^{|S|} \alpha_i x_i \; \middle| \; (\forall i : \alpha_i \geq 0) \wedge \sum_{i=1}^{|S|} \alpha_i = 1 \right\} \tag{1}$$

Convex hulls over $n$ points (planar or otherwise) can be computed with very low complexity algorithms. In our case we use the combined two-dimensional Quickhull algorithm with the general-dimension Beneath-Beyond algorithm developed by [36]. Subsequently,

---

[2]The code is available at `https://github.com/celiacintas/tests_landmarks/blob/master/testing_output_ears_JCST.ipynb`
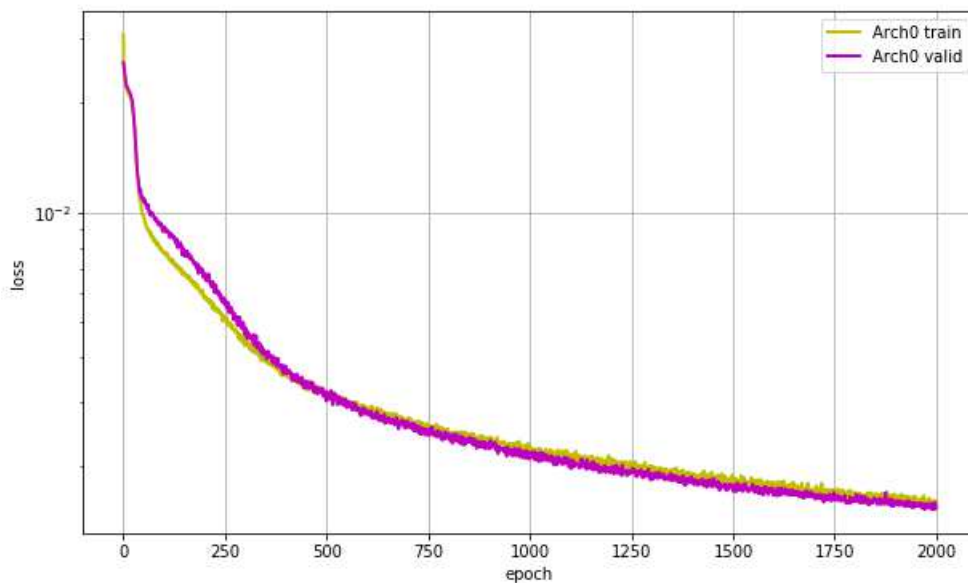
Figure 3: Loss curves over validation and training set of the CNN detailed in Table 1. The yellow line represent error on the training set, and the violet line represent the validation error, for the best performing network.
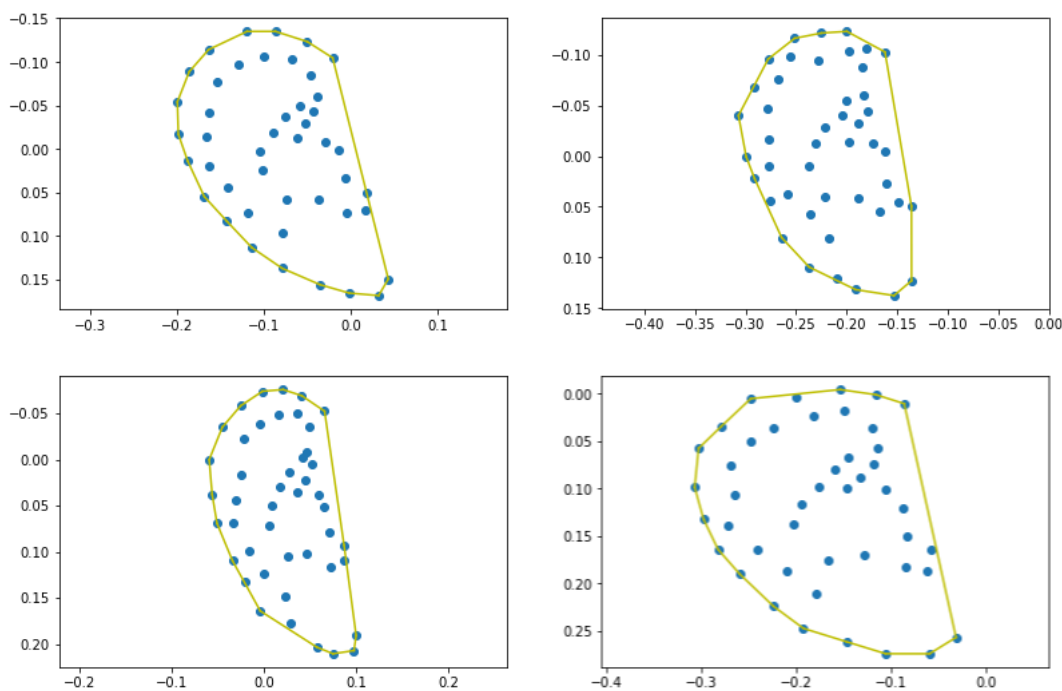


Figure 4: Convex Hulls calculated over automatic placed landmarks.

the convex hull is used as a mask for the ROI, obtaining as a final result only the pixels corresponding to the ear, see Fig. 5. The code for convex hull calculation and posterior segmentation along with the full output dataset can be seen in the repository.

## 4    Performance Assessment of Ear Detection

To evaluate the performance of our method we used as a ground truth the geometry of the ROI, and as a second geometry, the one formed by the convex hull of the landmarks coordinates described in Eq. 1. If the latter is completely inside the former, the ear was correctly detected. Since both regions are convex, this

Table 2: Performance with different types of occlusion.

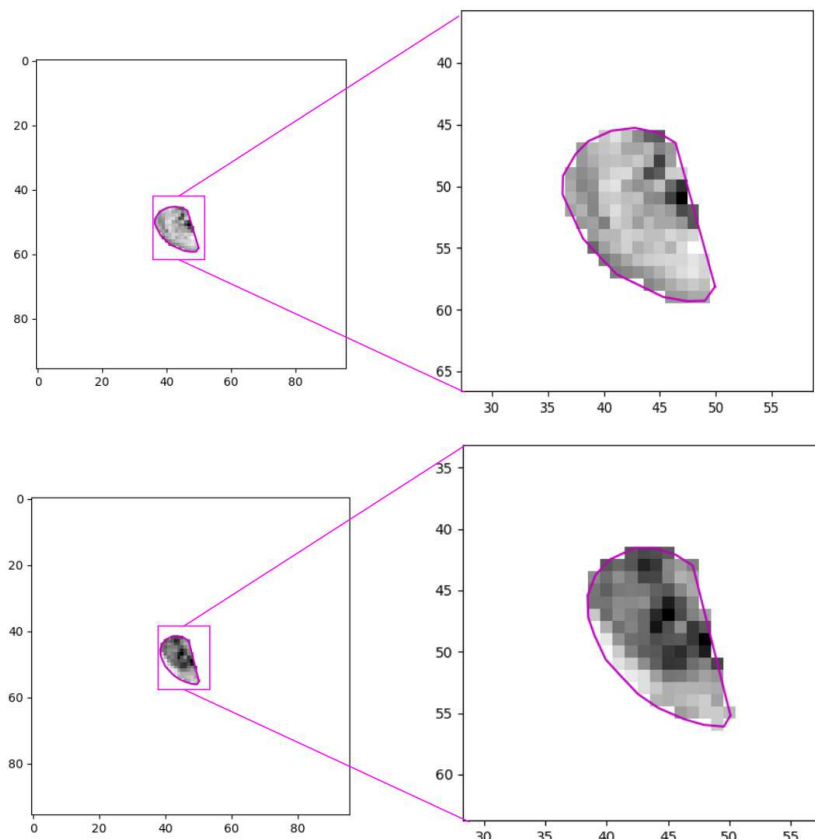| # of boxes | % of image occluded | Accuracy CNN | Accuracy Viola-Jones |
|---|---|---|---|
| 0 | 0 | **0.941** | 0.803 |
| 3 | 10.1 | **0.900** | 0.602 |
| 5 | 17.9 | **0.868** | 0.474 |
| 7 | 23.7 | **0.832** | 0.461 |



Figure 5: Pixels corresponding to Convex Hull mask based on automatic placed landmarks.

means that all the detected landmarks must lie within the ROI. Otherwise the ear was incorrectly detected.

For comparison, we also included results using the Haar-based ear detector using cascades for left and right ear available at OpenCV library. In Table 2 we put together the outcomes of the Haar-based and our method under progressive occlusions in the images. As can be seen, our method clearly outperform Viola-Jones, being also much more robust under increasingly larger occluded areas. The accuracy of our method degrades gracefully when occlusion increases, being still significant with as much as 24% occlusion, while Viola-Jones' performance lowers rather drastically.

In Fig. 6 it can be noticed that even though in some of the images the ear is partially occluded, the landmarking is still sound. The full test set landmarked by the CNN, the full structure and an analysis of the net's behavior can be accessed in the aforementioned repository.

## 5   Discussion and Conclusions

We analyze the feasibility of using Geometric Morphometrics and CNNs for ear detection. For this purpose we used a specific CNN previously trained with supervised landmarks [6]. Detection was evaluated over an open dataset, not previously used for training or validation. The CNN was further evaluated over different settings with incremental partial occlusions. Ear detection was still adequate even with 24% of the image occluded. Finally we propose an alternative ROI segmentation method, based on using the convex hull determined by the set of landmarks detected by the CNN. The resulting ROI greatly enhances the quality and reduces the computational burden of subsequent tasks like people identification and other biometric applications, since it is a robust and clean pixel set without any background clutter that may mistify further analysis.

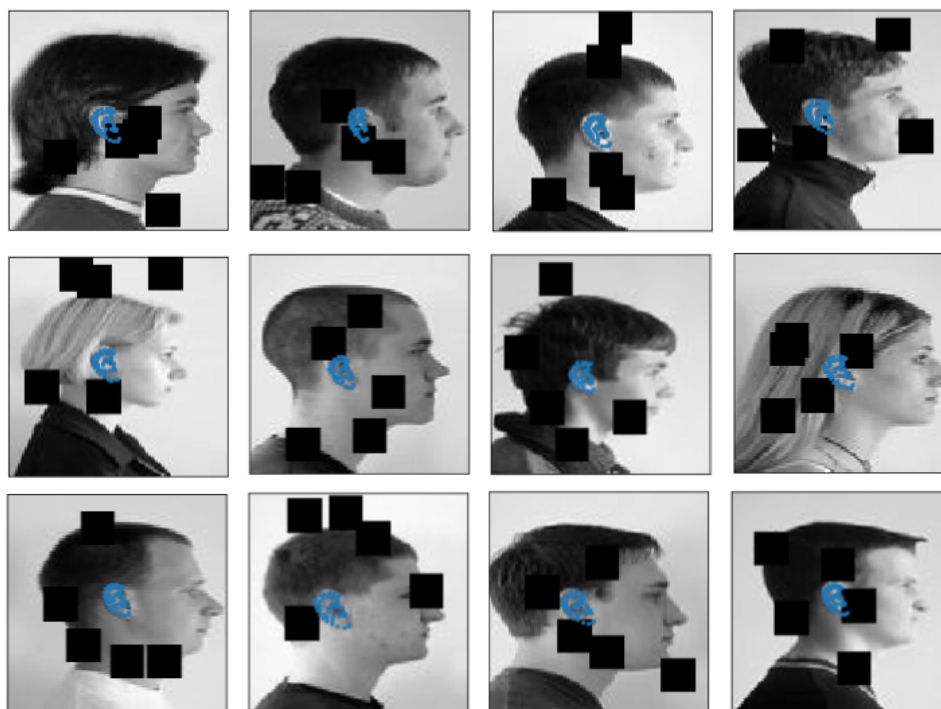We are currently working on people identification

Figure 6: Results over images randomly chosen from the CVL Face Database with partial occlusion with five randomly located black boxes of size $15 \times 15$ `http://www.lrv.fri.uni-lj.si/facedb.html` [7].

using ear biometrics. Both the convex ROI method and the plain landmarks appear to be promising venues. The latter provides significant qualitative information, such as relative distances and angles among landmarks. Also, as initial studies suggest ([6]), the relative importance of the landmarks' coordinates is uneven, therefore reducing the amount of information required for correct identification.

## Competing interests

The authors have declared that no competing interests exist.

## References

[1] I. Alberink and A. Ruifrok, "Performance of the FearID earprint identification system.," *Forensic science international*, vol. 166, pp. 145–54, mar 2007.

[2] C. Sforza, G. Grandi, M. Binelli, D. G. Tommasi, R. Rosati, and V. F. Ferrario, "Age- and sex-related changes in the normal human ear.," *Forensic science international*, vol. 187, pp. 110.e1–7, may 2009.

[3] M. I. S. Ibrahim, M. S. Nixon, and S. Mahmoodi, "The effect of time on ear biometrics," in *2011 International Joint Conference on Biometrics (IJCB)*, pp. 1–6, IEEE, oct 2011.

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

[5] M. L. Zelditch, D. L. Swiderski, H. D. Sheets, and W. L. Fink, "Geometric morphometrics for biologists," *Elsevier*, vol. 59, no. 3, p. 457, 2004.

[6] C. Cintas, M. Quinto-Sánchez, V. Acuña, C. Paschetta, S. De Azevedo, C. Silva de Cerqueira, V. Ramallo, C. Gallo, G. Poletti, M. C. Bortolini, S. Canizales-Quinteros, R. Francisco, G. Bedoya, A. Ruiz-Linares, R. González-José, and C. Delrieux, "Automatic ear detection and feature extraction using Geometric Morphometrics and Convolutional Neural Networks," *IET Biometrics*, dec 2016.

[7] F. Solina, P. Peer, B. Batagelj, S. Juvan, and J. Kovač, "Color-based face detection in the" 15 seconds of fame" art installation," *Proceedings of Mirage 2003, Conference on Computer Vision / Computer Graphics*, pp. 38–47, 2003.

[8] S. Pflug and C. Busch, "Ear biometrics: a survey of detection, feature extraction and recognition methods," *IET Biometrics*, vol. 1, no. 2, p. 114, 2012.

[9] H. Chen and B. Bhanu, "Shape Model-Based 3D Ear Detection from Side Face Range Images," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recogni-*

*tion (CVPR'05) - Workshops*, vol. 3, pp. 122–122, IEEE.

[10] H. Chen and B. Bhanu, "Contour matching for 3D ear recognition," in *Seventh IEEE Workshop on Applications of Computer Vision, WACV 2005*, pp. 123–128, 2007.

[11] S. Attarchi, K. Faez, and A. Rafiei, *Advanced Concepts for Intelligent Vision Systems*, vol. 5259 of *Lecture Notes in Computer Science*. Berlin, Heidelberg: Springer Berlin Heidelberg, oct 2008.

[12] S. Ansari and P. Gupta, "Localization of ear using outer helix curve of the ear," *International conferenceon computing:the theoryand applications*, pp. 688–692, 2007.

[13] S. Prakash and P. Gupta, "An efficient ear localization technique," *Image and Vision Computing*, vol. 30, no. 1, pp. 38–50, 2012.

[14] P. Yan and K. W. Bowyer, "Biometric recognition using 3D ear shape.," *IEEE transactions on pattern analysis and machine intelligence*, vol. 29, no. 8, pp. 1297–308, 2007.

[15] A. Pflug, A. Winterstein, and C. Busch, "Robust localization of ears by feature level fusion and context information," in *Biometrics (ICB), 2013 International Conference on*, pp. 1–8, IEEE, 2013.

[16] A. Abaza, C. Hebert, and M. A. F. Harrison, "Fast learning ear detection for real-time surveillance," *IEEE 4th International Conference on Biometrics: Theory, Applications and Systems, BTAS 2010*, 2010.

[17] S. M. S. Islam, M. Bennamoun, and R. Davies, "Fast and fully automatic ear detection using cascaded adaboost," *2008 IEEE Workshop on Applications of Computer Vision, WACV*, 2008.

[18] L. Yuan, W. Liu, and Y. Li, "Non-negative dictionary based sparse representation classification for ear recognition with occlusion," *Neurocomputing*, vol. 171, pp. 540–550, 2016.

[19] A. Kumar and T.-S. T. Chan, "Robust ear identification using sparse representation of local texture descriptors," *Pattern recognition*, vol. 46, no. 1, pp. 73–85, 2013.

[20] T.-S. Chan and A. Kumar, "Reliable ear identification using 2-d quadrature filters," *Pattern Recognition Letters*, vol. 33, no. 14, pp. 1870–1881, 2012.

[21] R. Purkait and P. Singh, "A test of individuality of human external ear pattern: its application in the field of personal identification.," *Forensic science international*, vol. 178, no. 2-3, pp. 112–8, 2008.

[22] I. Ercan, S. T. Ozdemir, A. Etoz, D. Sigirli, R. S. Tubbs, M. Loukas, and I. Guney, "Facial asymmetry in young healthy subjects evaluated by statistical shape analysis," *Journal of Anatomy*, vol. 213, no. 6, pp. 663–669, 2008.

[23] P. Mitteroecker and P. Gunz, "Advances in Geometric Morphometrics," 2009.

[24] D. Slice, "Modern morphometrics," in *Modern morphometrics in physical anthropology*, pp. 1–46, 2005.

[25] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, 2015.

[26] S. Dieleman, K. W. Willett, and J. Dambre, "Rotation-invariant convolutional neural networks for galaxy morphology prediction," *Monthly Notices of the Royal Astronomical Society*, vol. 450, no. 2, pp. 1441–1459, 2015.

[27] K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position," *Biological Cybernetics*, vol. 36, no. 4, pp. 193–202, 1980.

[28] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proceedings of the IEEE*, vol. 86, no. 11, pp. 2278–2323, 1998.

[29] A. Toshev and C. Szegedy, "DeepPose: Human Pose Estimation via Deep Neural Networks," in *Computer Vision and Pattern Recognition (CVPR)*, pp. 1653–1660, 2014.

[30] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Advances In Neural Information Processing Systems*, pp. 1–9, 2012.

[31] Y.-L. Boureau, J. Ponce, and Y. Lecun, "A Theoretical Analysis of Feature Pooling in Visual Recognition," *Proceedings of the 27th International Conference on Machine Learning (2010)*, pp. 111–118, 2010.

[32] A. Ruiz-Linares, K. Adhikari, V. Acuña-Alonzo, M. Quinto-Sanchez, C. Jaramillo, W. Arias, M. Fuentes, M. Pizarro, P. Everardo, F. de Avila, J. Gómez-Valdés, P. León-Mimila, T. Hunemeier, V. Ramallo, C. C. Silva de Cerqueira, M.-W. Burley, E. Konca, M. Z. de Oliveira, M. R. Veronez, M. Rubio-Codina, O. Attanasio, S. Gibbon, N. Ray, C. Gallo, G. Poletti, J. Rosique, L. Schuler-Faccini, F. M. Salzano, M.-C. Bortolini, S. Canizales-Quinteros, F. Rothhammer,

G. Bedoya, D. Balding, and R. Gonzalez-José, "Admixture in Latin America: geographic structure, phenotypic diversity and self-perception of ancestry based on 7,342 individuals.," *PLoS genetics*, vol. 10, p. e1004572, sep 2014.

[33] S. Dieleman, J. Schlüter, C. Raffel, E. Olson, S. K. Sønderby, D. Nouri, D. Maturana, M. Thoma, E. Battenberg, J. Kelly, J. D. Fauw, and M. H. et al., "Lasagne: First release.," Aug. 2015.

[34] L. yuan and Z. C. Mu, "Ear detection based on skin-color and contour information," in *2007 International Conference on Machine Learning and Cybernetics*, vol. 4, pp. 2213–2217, Aug 2007.

[35] P. Viola and M. Jones, "Robust real-time object detection," *International Journal of Computer Vision*, vol. 57, pp. 137–154, 2001.

[36] C. B. Barber, D. P. Dobkin, and H. Huhdanpaa, "The quickhull algorithm for convex hulls," *ACM Transactions on Mathematical Software*, vol. 22, no. 4, pp. 469–483, 1996.