

Anthropometric Landmark Detection in 3D Head Surfaces Using a Deep Learning Approach

Helena R. Torres^{ID}, Pedro Morais, Anne Fritze, Bruno Oliveira, Fernando Veloso^{ID}, Mario Rüdiger, Jaime C. Fonseca, and João L. Vilaça^{ID}

Abstract—Landmark labeling in 3D head surfaces is an important and routine task in clinical practice to evaluate head shape, namely to analyze cranial deformities or growth evolution. However, manual labeling is still applied, being a tedious and time-consuming task, highly prone to intra-/inter-observer variability, and can mislead the diagnose. Thus, automatic methods for anthropometric landmark detection in 3D models have a high interest in clinical practice. In this paper, a novel framework is proposed to accurately detect landmarks in 3D infant's head surfaces. The proposed method is divided into two stages: (i) 2D representation of the 3D head surface; and (ii) landmark detection through a deep learning strategy. Moreover, a 3D data augmentation method to create shape models based on the expected head variability is proposed. The proposed framework was evaluated in synthetic and real datasets, achieving accurate detection results. Furthermore, the data augmentation strategy proved its added value, increasing the method's performance. Overall, the obtained results

Manuscript received May 29, 2020; revised September 11, 2020 and October 27, 2020; accepted October 29, 2020. Date of publication November 4, 2020; date of current version July 20, 2021. This work was supported under the projects "NORTE-01-0145-FEDER-024300" and "NORTE-01-0145-FEDER-000045", supported by the Northern Portugal Regional Operational Programme (Norte2020), under the Portugal 2020 Partnership Agreement, through the European Regional Development Fund (FEDER). It was also funded by national funds, through the FCT – Fundação para a Ciência e Tecnologia and FCT/MCTES in the scope of the Project UIDB/05549/2020 and UIDP/05549/2020, and by FCT and the European Social Found, through Programa Operacional Capital Humano (POCH), in the scope of the Ph.D. under Grants SFRH/BD/136670/2018, SFRH/BD/136721/2018, and SFRH/BD/131545/2017. (*Corresponding author: Helena R. Torres.*)

Helena R. Torres and Bruno Oliveira are with the 2Ai – School of Technology, IPCA, Barcelos, Portugal, with the Algoritmi Center, School of Engineering, University of Minho, 4800-058 Guimarães, Portugal, with the Life and Health Sciences Research Institute (ICVS), School of Medicine, University of Minho, 4710-057 Braga, Portugal, and also with the ICVS/3B's - PT Government Associate Laboratory, 4710-057 Braga/Guimarães, Portugal (e-mail: htorres@ipca.pt; boliveira@ipca.pt).

Pedro Morais and João L. Vilaça are with the 2Ai – School of Technology, IPCA, Barcelos, Portugal (e-mail: pmorais@ipca.pt; jvilaça@ipca.pt).

Anne Fritze and Mario Rüdiger are with the Department for Neonatology and Pediatric Intensive Care, Children's Hospital, Medical Faculty of TU Dresden 01307, Germany (e-mail: anne.fritze@uniklinikum-dresden.de; mario.ruediger@uniklinikum-dresden.de).

Fernando Veloso is with the 2Ai – School of Technology, IPCA, Barcelos, Portugal, and also with the Department of Mechanical Engineering, School of Engineering, University of Minho, 4800-058 Guimarães, Portugal (e-mail: fveloso@ipca.pt).

Jaime C. Fonseca is with the Algoritmi Center, School of Engineering, University of Minho, Guimarães, Portugal (e-mail: jaime@dei.uminho.pt).

Digital Object Identifier 10.1109/JBHI.2020.3035888

demonstrated the robustness of the proposed method and its potential to be used in clinical practice for head shape analysis.

Index Terms—Cranial deformities, convolutional networks, deep learning, head growth, landmark detection.

I. INTRODUCTION

TO MEASURE and quantify features of individual heads is of great importance in clinical practice to evaluate head shape, namely for diagnosis of cranial deformities or head growth evolution analysis. Concerning cranial deformities, one of the most common is deformational plagiocephaly (DP) which refers to an asymmetrical distortion of an infant's skull [1]. Studies showed that 1 in 5 infants suffer from DP, with an even higher prevalence in prematurely born infants (38%) [2]. Moreover, it is indicated that 73% of extremely preterm infants have symmetrical head deformations [2]. Besides cosmetic concerns, infants with deformities can also experience developmental delay [3]. Thus, early diagnosis by analyzing the head shape is essential [4], where cranial indexes are estimated from anthropometric landmarks of surfaces generated by 3D digital techniques or medical images such as magnetic resonance (MR), ultrasound, or radiography [4], [5]. Moreover, head analysis using landmarks is also used to evaluate head growth from the gestational period to childhood [6]. However, manual landmark labeling is a time-consuming and tedious task, highly prone to intra-/inter-observer variability, which can lead to a misdiagnose of cranial deformities. Thus, fully automatic methods to detect anthropometric landmarks are frequently claimed. With automated landmark detection methods, manual analysis is avoided, reducing diagnosis biases resulted from user interaction and allowing to broad head analysis to non-specialized clinicians. Plus, it allows automating the current process for head analysis, where automatized measurements can be performed daily to closely and continually monitor head shape and evolution.

This work focus on the development of a novel deep learning (DL) method to automatically detect anthropometric landmarks in 3D head surfaces of infants. High robustness, accuracy, and efficiency are mandatory to this solution, making it attractive to clinical practice. Moreover, since specific clinical indicators are extracted from these landmarks, automatic evaluation of cranial deformities is possible [4]. A preliminary approach for landmark detection was proposed by the authors in [7]. However, it used a non-optimal and computational expensive unfolding strategy,

while requiring the model in a specific orientation. Plus, only tests in synthetic heads were performed.

Hereto, the main contributions of this paper are:

- A rotational invariant DL-based method to detect anatomical landmarks in 3D head models using a computationally efficient 2D representation based on explicit functions;
- Use of meaningful shape descriptors to embed information to the 2D representation of the head surface;
- A novel 3D data augmentation strategy to robustly generate a high number of head models using clinical landmarks;
- Evaluation of the proposed method in a clinical database for head shape and cranial deformities analysis;
- The authors will make publicly available a 3D synthetic database (*i.e.*, first public database of infant head surfaces).¹

The rest of the paper is organized as follows. In Section II, an overview of the literature is presented. In Section III, the proposed methodology is described. The 3D data augmentation approach is presented in Section IV. The experiments performed and the obtained results are described in Section V. In Section VI, the method's performance is discussed, and the conclusions are presented in Section VII.

II. RELATED WORK

Landmark detection in 3D models is a topic of interest in the literature. Although landmark detectors have already been explored in 3D facial surfaces, only a few were focused on clinical relevant anatomical landmarks from entire head models. For facial surfaces, one of the first approaches was proposed by Colombo *et al.* [8], where triplet candidates of salient facial features are detected using surface curvature information. The triplets are then processed using a principal component analysis (PCA)-based classifier to select the final landmark points. PCA was also used in [9] to select the best facial features in cases of high shape variation. The curvature of a 3D shape was also explored by D'Hose *et al.* in [10]. Here, instead of curvature analyses based on the normal direction, the curvature information was retrieved from Gabor wavelets [11]. Moreover, other approaches focused on other geometric descriptors than the curvature to characterize the head model [12]–[14]. However, the abovementioned approaches usually required the facial model in a specific orientation, while threshold-based or traditional classification approaches are needed to process the retrieved geometric descriptors.

Some researchers applied a comparison strategy between the target case and a mean shape model. In [15], [16], a Point Distribution Model (PDM) was used. The method fits a PDM to the facial surface using candidate vertices extracted from low-level feature maps. The landmarks are then detected by finding the transformation that minimizes the deviation of the facial surface to the PDM. A Constrained Local Model (CLM) was proposed in [17], where a statistical model that learns both global

variations of 3D shapes and local changes at each landmark was developed. In [18], a local descriptor was added to the CLM model. Other approaches are deformable matching techniques, where a template mesh with labeled landmarks are used to estimate the landmarks in a target surface using registration techniques. In [19], [20], a deformable registration is used to align the template mesh of a facial surface and the target shape. The main disadvantages of the deformable techniques are the dependence on the initialization and the high computational cost usually required.

Other methods do not use directly the 3D model. In [21] and [22], the 3D mesh of the facial surface is converted into a depth image, which is posteriorly processed to detect the landmarks. The same approach was explored in [23]. However, these projection-based representations usually only include low model information such as depth variations, missing relevant geometric information of the model.

Limitations of the abovementioned methods, related to its accuracy and robustness, must be considered. On the one hand, methods that strongly rely on a mean shape are not versatile. On the other hand, deformable techniques are susceptible to local minimums and to the initialization. Moreover, traditional shape descriptors-based methods are sensitive to a threshold value or to its post-processing methodology. Finally, projection-based methods may fail to deal with more complex models.

Although facial models and corresponding automatic methods could be used to extract landmarks, e.g., for recognition tasks, it cannot be used to extract all relevant clinical indexes for cranial deformities evaluation. For clinical tasks, processing of a more complex full head model is required. Plus, non-clinical surfaces are typically acquired from one source, in opposition to the clinical practice where multiple sources with high user dependence are used for head model generation and evaluation. Thus, non-clinical datasets usually present superior quality, easing the detection task. In this sense, landmark detectors for clinical head models are claimed.

III. METHODS

The proposed method relies on a two-stage approach (Fig. 1). The first stage creates a 2D representation of a 3D head model (Section III-A). Here, the 3D model is represented as an explicit function, decreasing the detection complexity. Then, shape descriptors are included in the 2D representation, creating anatomically enhanced 2D feature maps. The second stage relies on a DL approach to detect the anthropometric landmarks on the images created in the first stage (Section III-B). Due to its importance for cranial asymmetry evaluation [24], nine landmarks were used: (Fig. 2): glabella (GL), sellion (SL), subnasal (SN), right and left exocanthions (R-EX/L-EX), endocanthions (R-ED/L-ED), and tragions (R-TR/L-TR).

A. 2D Images Creation From 3D Surfaces

1) *2D Representation of the 3D Surface*: Due to the advances in 3D sensing technologies, 3D data can be represented in different ways. Two common data types are 3D point clouds and meshes. The first one consists of a set of unstructured 3D points

¹<https://myneurogrowth.ipca.pt/index.php/the-expected-outputs/3d-synthetic-head-model-datasets/>

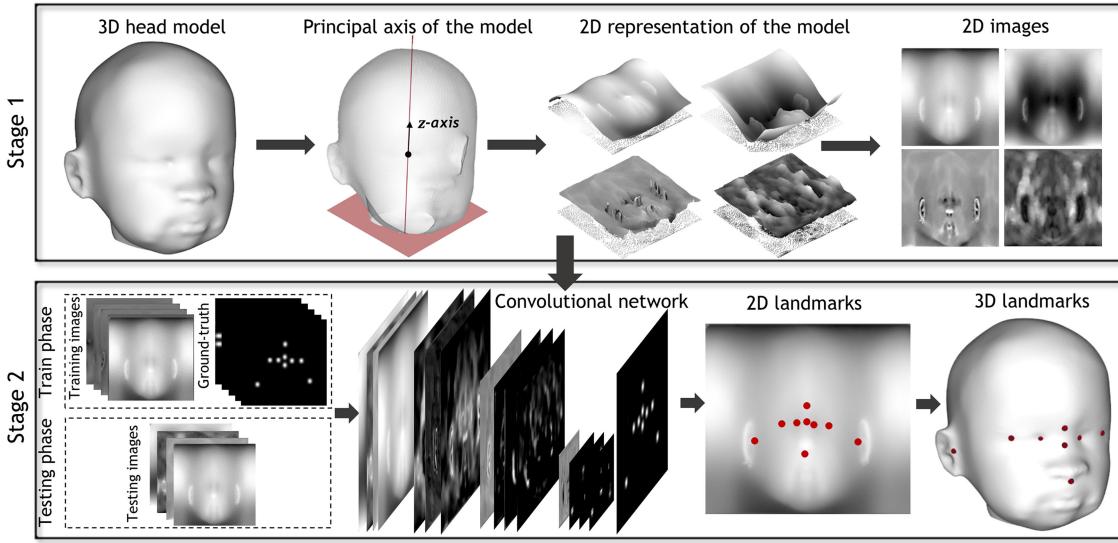


Fig. 1. Overview of the proposed 3D landmark detection method.

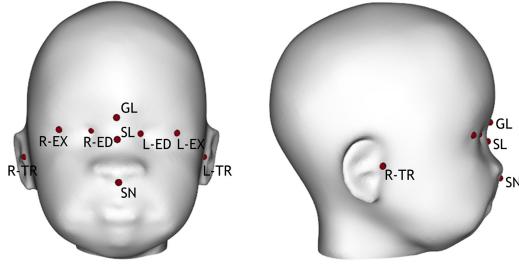


Fig. 2. Anthropometric landmarks. (A) Coronal view; (B) Sagittal view.

that approximate the geometry of 3D objects. However, due to the absence of connectivity information, geometric analysis of the model is hampered. To solve this issue, 3D meshes, composed of a set of vertices and faces describing points connections, are used. However, this type of data can result in irregular surfaces that frequently require complex and computational demanding mathematical models to process them [25]. Some works attempt to extend meshes to graph-structured data, but the mechanisms to process it are not yet fully clear in the literature [26]. Another approach is to use 2D representations to perform 3D shape analysis. 2D representations of a 3D shape include projection-based or multi-view approaches. While in the first only depth variations are usually captured [27], the latter requires independent processing of a pre-defined number of views, which may be computationally demanding. Plus, these approaches usually result in loss of geometric information. Another type of 2D representation can be obtained using geometric functions, where geometric shape descriptors can be easily included in the representation. Thus, this type of 2D representation of a 3D mesh model is exploited in this work.

To create the 2D representation, the concept of geometric functions is used by representing the 3D model as an explicit function, *i.e.*, by using a 2D function to represent the 3D surface in space [28]. Geometrically, this implies that one of the

coordinates of the points on the model is given explicitly as a function of the remaining coordinates. Such explicit relation can be mathematically defined for a N -dimensional problem as:

$$g : \mathbb{R}^{n-1} \mapsto \mathbb{R}, p_1 = g(\mathbf{p}^*), \quad (1)$$

where \mathbf{p} is a point of coordinates $\{p_1, \dots, p_N\}$ within the model and $\mathbf{p}^* = \{p_2, \dots, p_N\}$. Thus, one can obtain a representation function which has fewer dimensions than the original model, reducing the dimensionality of the problem. Although this limits the topology of the 3D model, it fits well in the case of the head model due to its convex-like and smooth shape.

To define the geometric function, a coordinate system needs to be defined. In this work, a cylindrical coordinate system was chosen. Thus, the cartesian coordinates of each point $\mathbf{p} = \{x, y, z\}$ are converted in cylindrical coordinates $\mathbf{p} = \{\rho, \varphi, z\}$, where ρ represents the radial distance from the z -axis to \mathbf{p} , φ is the azimuth angle, and z is the height of \mathbf{p} . To correctly perform this transformation, the z -axis must be well-defined for all head models. The z -axis was defined using the model principal axis, defined as the vector with origin \mathbf{o} (model's centroid) and a direction perpendicular to the neck's region plane. The neck region is detected in a pre-processing step by assuming that an open region is always found at the neck, and by searching boundary edges, *i.e.*, model holes. Then, the vertices of the largest hole are used to fit a neck's region plane. Note that head rotation variations in defined z -axis are possible and later solved by the invariant DL approach. Indeed, head rotation variation in defined z -axis will only result in modification of the 2D representation throughout the azimuth coordinate. Plus, once the representation is always defined around the axis, head orientation in other directions do not affect the representation.

Due to the explicit formulation of the 3D model, the 2D representation is obtained by defining the radial coordinate ρ of the points as a function of the azimuth angle φ and height z , as shown in Fig. 3. However, the conventional definition

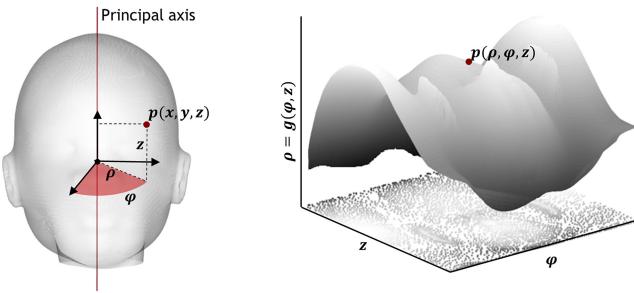


Fig. 3. Representation of a 3D head model as an explicit function in the spherical domain. **(A)** 3D head model in the cartesian domain and correspondence between cartesian coordinates (x, y, z) with cylindrical coordinates (ρ, φ, z) ; **(B)** Explicit function in the cylindrical domain.

of the coordinate ρ in the cylindrical space only provides information about the radial distance from the z -axis to a given point, offering reduced spatial information. Thus, different shape descriptors were applied to the 2D representation by redefining the coordinate ρ as the descriptors' value. Thus, coordinate ρ will be termed henceforward as ρ' . It might be emphasized that in complex concave regions, multiple 3D points can be used to represent the same $\{\rho, \varphi\}$ coordinate pair. For that, the 3D point with a higher radius (r) is used in the 2D representation.

2) Image Creation Using Shape Descriptors: Shape descriptors of a 3D model concern information about its features, usually retrieved from the model's surface or its interior. Once texture information is not usually included in 3D models, these descriptors are important for surface processing. In the present method, three shape descriptors were studied and applied to the 2D representation. The first descriptor consists of the distance between each point of the 3D head model and its center o , being defined as:

$$d(p) = \sqrt{\sum_n^N [p_n - o_n]^2}, \quad (2)$$

where n represents each coordinate of the point in the 3D world. This descriptor, henceforward termed as depth descriptor, provides useful information about the surface, once retrieves information regarding its salient parts (e.g., nose and ears).

Another shape descriptor highly used to characterize a 3D mesh is the mean curvature. To calculate the mean curvature at each point of the model, one needs to average the principal curvatures k_1 and k_2 of the surface, which measure the maximum and minimum bending of the model at each point. Thus, the mean curvature descriptors can be described as:

$$c(p) = \frac{k_1(p) + k_2(p)}{2}, \quad (3)$$

where k_1 and k_2 are calculated using concepts of differential geometry. Using this descriptor, high curvatures regions, such as the nose, the mouth, or the outer region of the ears, can be differentiated from flatter ones. The reader is kindly directed to [29] for further details on principal curvature computation.

The third descriptor quantifies the level of detail in different regions of the model. It can be clearly understandable that

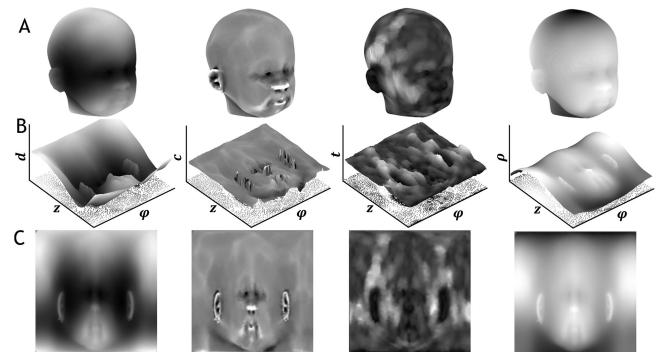


Fig. 4. 3D model representation based on shape descriptors. **(A)** Original model colorized with descriptors' values. **(B)** Explicit functions with descriptors as the explicit coordinate. **(C)** 2D image resulted from the representations. The first, second, third, and fourth columns concern the depth d , curvature c , texture t , and radial ρ descriptors, respectively.

complex regions require a higher number of polygons to fully recover the details of the region. Thus, a shape descriptor that quantifies this level of detail can be seen as a measure of a texture-like property of the mesh, and it is defined as:

$$t(p, s) = \frac{1}{Q} \sum_{q=1}^Q \sqrt{\sum_n^N [p_n - s_{q,n}]^2}. \quad (4)$$

where s is a list of the Q points in the neighborhood of p , q is the neighbor number, and s_q is the q neighboring point. Note that, for a given p only vertices that share at least one face with p are considered.

After computing the shape descriptors in the 3D space, each one is transformed into a 2D representation by assigning their values for each point as a function of the azimuth angle φ and height z . Four representations are used. Three of them consisted of shape descriptors, *i.e.*, $\rho'_1(p) = d(p)$, $\rho'_2(p) = c(p)$, and $\rho'_3(p) = t(p)$. The fourth relies on the radial distance ρ directly extracted from the cylindrical space $\rho'_4(p) = \rho(p)$. **Fig. 4(A)** presents an illustration of the 3D models colorized with shape descriptors' values. **Fig. 4(B)** shows the four 2D representations. As a final step, a 2D image is created for each representation, by defining a grid with fixed size and interpolating the scattered data of the 2D representation for each grid position (**Fig. 4(C)**). All descriptors were normalized between 0 to 1, ensuring the same scale across all 2D images.

B. Deep Learning for Landmark Detection

After image creation, a DL strategy was applied to detect the landmarks, where a convolutional neural network (CNN) with the 2D images as input and the location of each landmark as output was used. The architecture used for the CNN is presented in **Fig. 5**. The network is split into four branches, one for each 2D input image. Each branch is firstly initialized and fine-tuned by the first layers of the VGG-19 [30], which is used to perform an initial image analysis in a transfer learning approach, generating a set of feature maps V_r with $r \in \{1 \dots 4\}$. The generated feature maps are then concatenated in a global one. Next, a

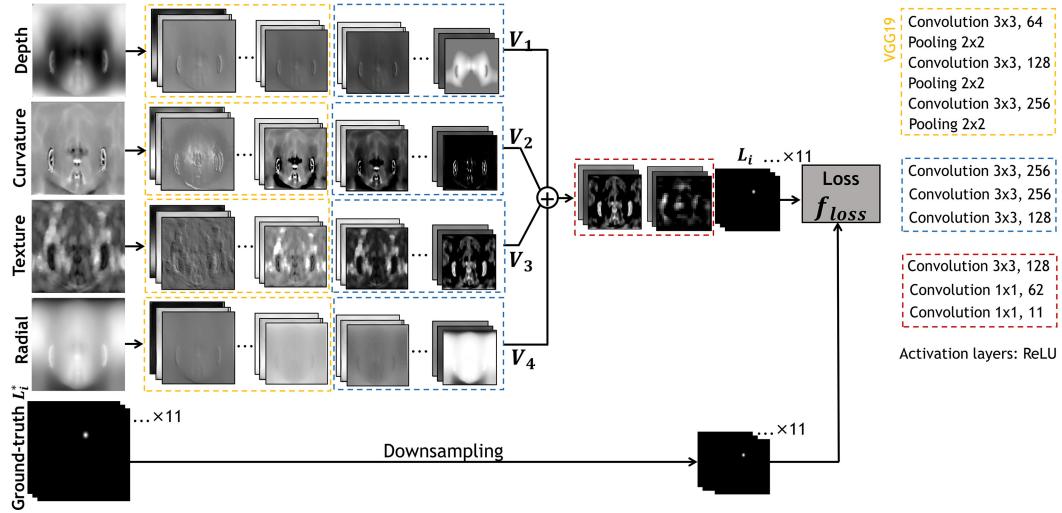


Fig. 5. Architecture of the convolutional network used in the proposed framework.

feed-forward network is used to predict confidence maps L that encode the probability of occurrence of a landmark at a given image position. These confidence maps are described as $L = (L_1, L_2, \dots, L_M)$, where $L_i, i \in \{1 \dots M\}$ represents the map for the landmark i and M is the number of landmarks.

At the training phase, confidence maps L^* are generated using manual labeling of landmark positions. For each landmark, the respective confidence map is computed by applying a Gaussian-like function where the maximum of the gaussian map represents the landmark position:

$$L_i^*(x, y) = \exp\left(-\frac{(x - x_i)^2 + (y - y_i)^2}{\sigma^2}\right), \quad (5)$$

where (x_i, y_i) corresponds to the ground-truth position of landmark i , $x \in \{1, \dots, h\}$, and $y \in \{1, \dots, w\}$, being h and w the image size. σ controls the spread of the Gaussian function. Each map represents the belief that a landmark point occurs in a given image pixel (x, y) , being assigned to each a probability $L_i^*(x, y)$ of being the respective landmark. The pixel with the higher probability value is the ideal landmark position. To train the method, a loss function f_{loss} that calculates the Euclidean distance between predicted and ground-truth confidence maps was applied at the end of the feed-forward network:

$$f_{loss} = \sum_{i=1}^M \|L_i - L_i^*\|_2^2, \quad (6)$$

where L_i and L_i^* are the prediction and ground truth maps for landmark i , respectively.

At the testing phase, each landmark (in the unlabeled case) is detected by estimating the respective confidence maps. Besides estimation of the maps in the original images, two additional shifts are performed to the images, being confidence maps also inferred for the shifted images. This allows minimizing the influence of landmarks that can be closely located in the image borders, once this positioning results in incomplete image information around the landmark, hampering the inference

process. The final confidence map is estimated by averaging all confidence maps, being the optimal position defined as the maximum of the respective average confidence map, *i.e.*, its peak, after a non-maximum suppression processing. In the final step, the obtained landmarks are transferred to the 3D world, by applying the inverse of the image to mesh transformation.

IV. 3D DATA AUGMENTATION

Fig. 6 illustrates the proposed data augmentation method. Firstly, PCA analysis is applied to a synthetic dataset to find its principal modes of variations. Next, transformations are applied to warp each head model to the principal modes of variation.

A. Principal Landmark Shape Modes Creation

The 3D data augmentation method starts by using the PCA to examine shape variations presented in a dataset. However, instead of using the entire head surface for the statistical analysis, M -point landmark models were used, corresponding to the nine landmarks plus two additional landmarks placed in the parietal side of the head. The mathematical representation of one landmark shape model is obtained by concatenating the landmarks into a representation vector given as:

$$\mathbf{v} = [p_{x,1}, p_{y,1}, p_{z,1}, \dots, p_{x,M}, p_{y,M}, p_{z,M}]. \quad (7)$$

From this representation vector, a matrix \mathbf{V} with shape $S \times 3M$ is then defined as $\mathbf{V} = [\mathbf{v}_1^T, \dots, \mathbf{v}_S^T]^T$, where S represents the number of samples of head models. The mean shape model $\bar{\mathbf{v}}$ is then obtained by averaging all the representation vectors:

$$\bar{\mathbf{v}} = \frac{1}{S} \sum_{c=1}^S \mathbf{v}_c. \quad (8)$$

After creation of \mathbf{V} and mean shape $\bar{\mathbf{v}}$, PCA analysis is implemented to produce principal shape variation modes of the head surfaces. Thus, the covariance matrix \mathbf{V}_c of \mathbf{V} is computed, followed by calculation of its eigenvectors \mathbf{E} and

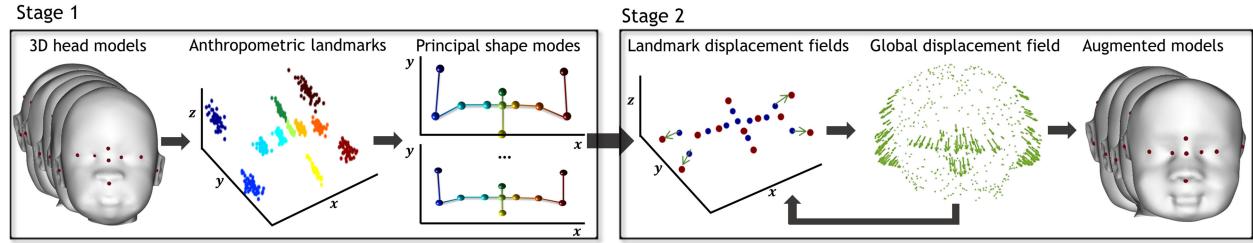


Fig. 6. Overview of the proposed data augmentation approach.

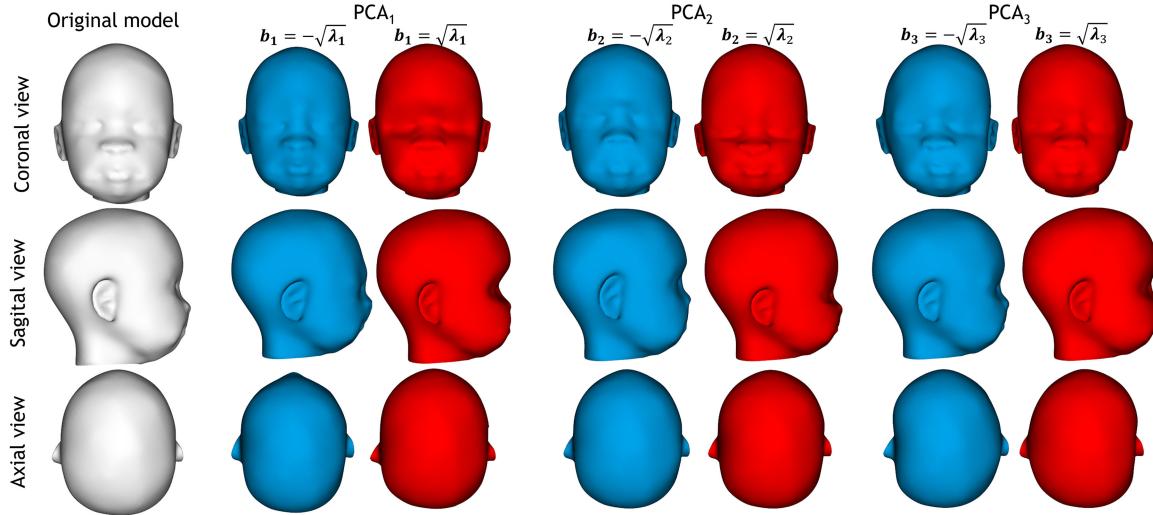


Fig. 7. Example of augmented models generated from one synthetic head model.

corresponding eigenvalues λ . One deformed landmark shape model can be then modeled as $v' = \bar{v} + Eb$, where b is a weight vector that represents principal modes of variation of the landmark models. By modifying the weights b , new landmark models can be created. In fact, by setting $b_m = \pm\sqrt{\lambda_m}$ and the remaining b weights to zero, extreme shape deformations for each mode of variation m are obtained [31]. In this work, the first three modes of variations were used, originating a total of six principal landmark shape models of the population of head surfaces.

B. Head Surface Deformation

Free model deformation strategies can be a solution to obtain augmented models. However, extreme deformations which produce unrealistic models must be prevented. Thus, a transformation method that warps an head surface based on the landmark shape models (Section V-A) is proposed. Once the principal shape models include realistic head shapes only, the risk of unrealistic 3D results is avoided (Fig. 7).

The warping method first creates a sparse displacement field computed between the landmarks of the original surface and the landmark shape model resulted from PCA analysis, estimating then a dense displacement map using a diffusion-like approach.

For each landmark i , the initial displacement field is given by:

$$\begin{aligned} & \{u_0(x_i, y_i, z_i), v_0(x_i, y_i, z_i), k_0(x_i, y_i, z_i)\} \\ &= \frac{1}{M} \sum_{i=1}^M \{x_i - x'_i, y_i - y'_i, z_i - z'_i\}, \end{aligned} \quad (9)$$

where (x_i, y_i, z_i) corresponds to the 3D coordinates of landmark i of the original model and (x'_i, y'_i, z'_i) is the position of landmark i in one principal landmark shape model. u_0, v_0, k_0 is the initial map different of zero only at $\{x_i, y_i, z_i\}$. Using the displacement field, denoted $\{u_{t,i}, v_{t,i}, k_{t,i}\}$ to simplify, a map of displacement fields for the entire surface is computed by performing multiple convolutions at each iteration t with a Gaussian function G centered in each landmark:

$$\begin{aligned} & \{u_t(x, y, z); v_t(x, y, z), k_t(x, y, z)\} \\ &= \frac{1}{M} \sum_{i=1}^M G_{\sigma,i}(x, y, z) * \{u_{t-1,i}, v_{t-1,i}, k_{t-1,i}\}, \end{aligned} \quad (10)$$

where the σ parameter controls the deformation. The map at each iteration t is obtained by applying a gaussian diffusion process to the previous map ($t - 1$). In this procedure, the displacement fields computed for the landmarks are used to estimate the remaining displacement fields. To evolve the head

model towards the principal shape model, an iterative approach is performed to minimize an energy functional given as:

$$E = \sum_{i=1}^M ((x_{t,i} - x'_i)^2 + (y_{t,i} - y'_i)^2 + (z_{t,i} - z'_i)^2). \quad (11)$$

The diffusion procedure is performed until a minimum value defined for E is reached or during a fixed number of iterations. Moreover, if a decrease of the energy criterion is obtained at one iteration, the surface's update is considered successful and the deformation parameter is kept. However, if an increase of E is verified, σ is multiplied by a factor $\alpha < 1$, resulting in a more conservative deformation in the next iteration.

V. EXPERIMENTS AND RESULTS

A. Dataset Creation

Due to the inexistence of public datasets of infant's heads, two different datasets of head surfaces were constructed, namely a synthetic and a real one. For the synthetic dataset, a 3D human modeling software (*MakeHuman*) was used [32]. Different full-body infant models were initially generated using random characteristics, including anatomical features and facial expressions. Then, the 3D meshes were post-processed to keep only the head. To evaluate the method's performance in clinical head surfaces, a real dataset was constructed. A set of MR images of infant's heads were acquired as described previously in [33] and processed. For each MR, the head surface was created by applying a binarization process followed by clipping of non-relevant areas. For both synthetic and real datasets, ground-truth was created by manually labeling the anthropometric landmarks. The datasets were divided into:

- Basic augmented synthetic training dataset: composed of 1250 models. From these 1250 models, 50 were synthetically created, being the remaining originated from scaling techniques. The scaling factors were randomly chosen according to the head parameters range described in [34]. This dataset was used to train the DL approach;
- Deformation-based augmented synthetic training dataset: composed by 1250 head models. From these 1250 models, 50 were synthetically created, being the remaining originated from the data augmentation strategy proposed in this work. This dataset was used to train the DL approach;
- 3DMM-based augmented synthetic training dataset: with 1250 models. 50 were synthetically created and 600 were originated from scaling techniques. The remaining 600 were created using the concept of morphable model (3DMM) [35]. A PCA-based morphable model was generated from the 50 synthetic models and new models were constructed by modifying the first twenty principal component weights and applying scaling techniques to them. This dataset was used to train the DL approach;
- Synthetic validation dataset: composed of 15 head models synthetically generated, used to find the best parameters for the DL, to analyze overfitting problems, and to detect the ideal epoch for early stopping of the training;

- Synthetic testing dataset: composed of 40 head models synthetically generated, being used to test the performance of the proposed detection method;
- Real testing dataset: composed of 40 head models originated from MR images, used to test the method's performance in a real scenario. Image acquisition was performed on a 1.5 T MR system (Siemens, Berlin, Germany) and different types of sequences were processed. For MR processing, all volumes were made isotropic (0.5 mm). Both normal and deformed cases are presented in the dataset.

B. Implementation Details

In the method's first stage, 2D images with a size of 256×256 pixels were created. Regarding the DL, experiments were carried out using the validation dataset to get the optimal settings. The learning rate was experimentally set to 0.0004. For the optimization, the Adam solver was used with a regularization term of 0.01. For the batch size, 10 images were used. Besides the proposed augmentation strategy, a 2D data augmentation strategy was also applied during training, allowing small rotations (until $\pm 5^\circ$) and scaling (resize factors between 0.95 and 1.05) of the image. Furthermore, head model rotations in the principal axis were simulated by shifting the 2D images by a random number of pixels. Concerning the shifted images used for inference, a shift of 30 and 60 pixels was applied, considering the total image size.

C. Detection Performance

Two types of analyses were performed to validate the proposed method, namely landmark detection and clinical validation. In the first analysis, different studies were conducted: i) the influence of each descriptor in the method's performance; and ii) the influence of the proposed data augmentation. Both synthetic and real cases were used. A comparison between the proposed method and state-of-the-art landmark detectors was also performed in the real database.

1) Landmark Detection Evaluation Metrics: The performance of the landmark detector was validated by comparing automatic results with manual analysis through the mean error, defined as the distance between estimated landmarks and the ground truth, and precision rate, which is the percentage of detections with a mean error below a threshold.

2) Accuracy Assessment: The first study evaluates the influence of each shape descriptor in the method's performance. Thus, instead of using the proposed multi-branch network configuration only, a single-branch network configuration was also evaluated for each descriptor, being conducted five experiments:

- **Eval₁**: single-branch network configuration with images resulted from the depth shape descriptor (d) as input;
- **Eval₂**: single-branch network configuration with images resulted from the curvature shape descriptor (c) as input;
- **Eval₃**: single-branch network configuration with images resulted from the texture shape descriptor (t) as input;
- **Eval₄**: single-branch network configuration with images resulted from the radial shape descriptor (ρ) as input;

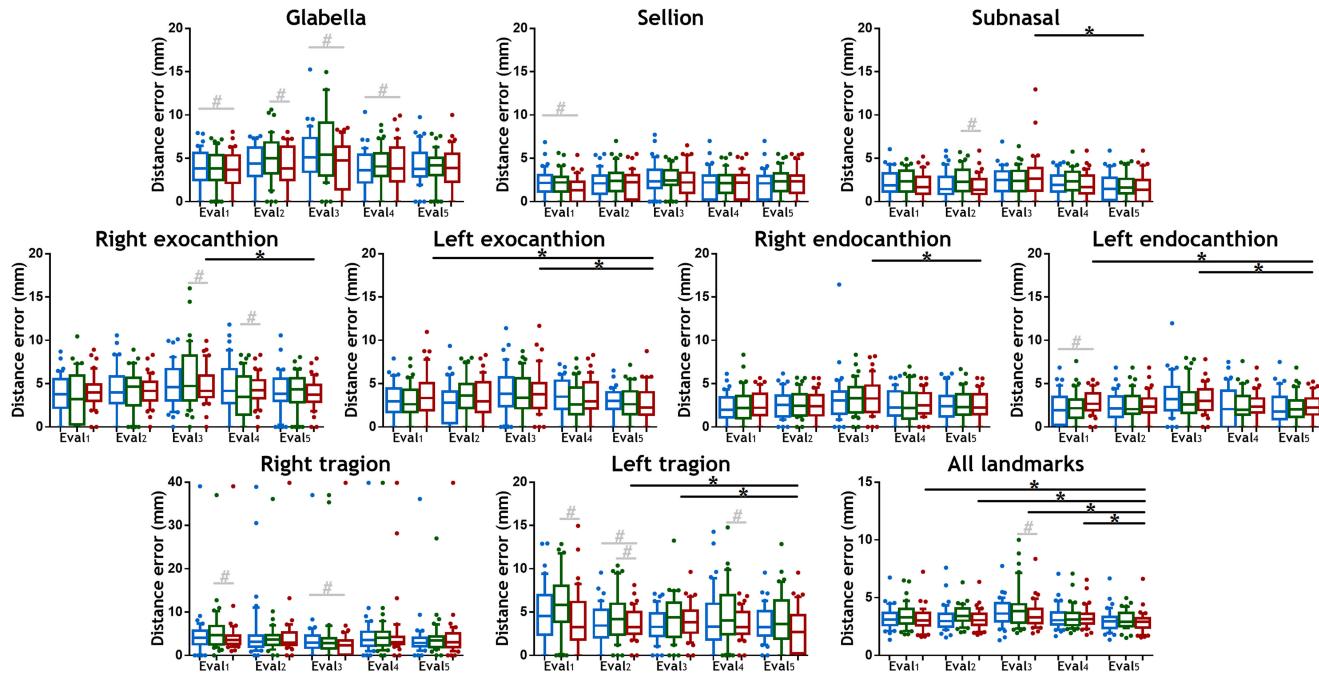


Fig. 8. Detection performance in the synthetic dataset. The ends of the whiskers represent the 10th and 90th percentiles. # $p < 0.05$ in a two-tailed paired t -test between basic/3DMM-based (blue/green boxplots) and proposed deformation-based (red boxplots) augmentation; * $p < 0.05$ in a two-tailed paired t -test assessed between each single-branch configuration and multi-branch configuration with augmentation (red boxplots).

- **Eval₅:** proposed multi-branch network configuration combining the images resulted from the shape descriptors.

To compare each single-branch configuration with the multi-branch approach, a two-tailed paired t -test ($p < 0.05$) was used to evaluate statistically significant differences. Moreover, to evaluate the influence of the 3D data augmentation strategy, the experiments were performed three times, first using the basic augmented synthetic dataset for training, second using the 3DMM-based augmented dataset, and third using the deformation-based augmented one (Section IV). Again, statistically significant differences were detected through a two-tailed paired t -test ($p < 0.05$) between pairs.

Fig. 8 and **Fig. 9** summarize the performance of the method in the synthetic and real testing dataset, respectively, assessed in terms of mean error. **Fig. 10** shows the average precision rate for different thresholds, using the proposed multi-branch method. Detection results examples are presented in **Fig. 11**. Plus, to evaluate the method's performance in the real dataset against an inter-observer variability, a second observer manually labeled the landmarks in each model, being this comparison presented in **Table I**.

3) Comparison With State-of-the-Art Detectors: To compare the proposed approach with the state-of-the-art, three different detectors were implemented and applied to the real dataset, namely a registration-based, a multi-view CNN, and a point cloud-based strategy. The registration-based approach, originally proposed in [36], relies on the Go-ICP approach [37] followed by a deformable registration. The synthetic dataset was used as labeled template models. The multi-view approach, termed Deep-MVLM, was proposed in [38] and predicts 3D landmarks using different 2D model views that are fed to a DL

network. For the point cloud-based approach, the PointNet [39] was used. Here, a modification to the original method was implemented, where probability maps are generated as the output of the network. For that, equation (6) was used as loss function. For the latter two methods, training was performed using the augmented synthetic dataset. For all methods, a pre-processing was applied by aligning the models using the principal axis found by neck positioning. **Table II** reports the performance of the different detectors.

D. Clinical Validation

The clinical validation of the proposed method targeted the diagnose of head deformities. For that, different cranial asymmetry indexes were estimated using the detected landmarks and compared with the manual ones. In the next subsections, the methodology used to estimate the indexes is described, along with a description of the selected indexes.

1) Head Coordinate System Estimation: To estimate the cranial asymmetry indexes from the detected landmarks, the definition of a coordinate system to standardize the head position/rotation is needed [6]. The coordinate system origin is defined as the midpoint of both TR. From the origin, x and y -axes are defined as the vectors from the origin through the L-TR and SL, respectively. The z -axis is the vector from the origin through the top of the head, such that it is perpendicular to the x and y -axes. A measurement plane parallel to the xy plane is generated, where the clinical measurements are estimated. Moreover, four head quadrants are defined by the coronal and sagittal planes of the coordinate system.

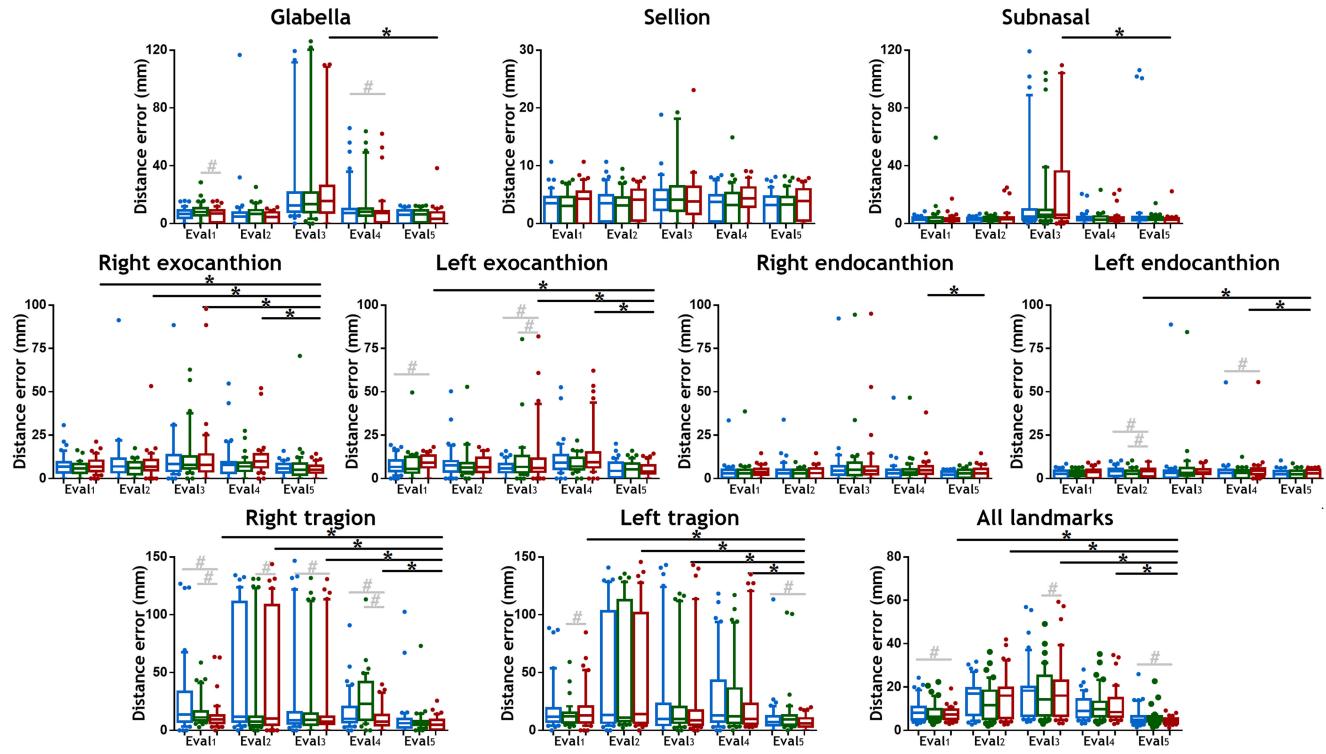


Fig. 9. Detection performance in the real dataset. The ends of the whiskers represent the 10th and 90th percentiles. # $p<0.05$ in a two-tailed paired t-test between basic/3DMM-based (blue/green boxplots) and proposed deformation-based (red boxplots) augmentation; * $p<0.05$ in a two-tailed paired t-test assessed between each single-branch configuration and multi-branch configuration with augmentation (red boxplots).

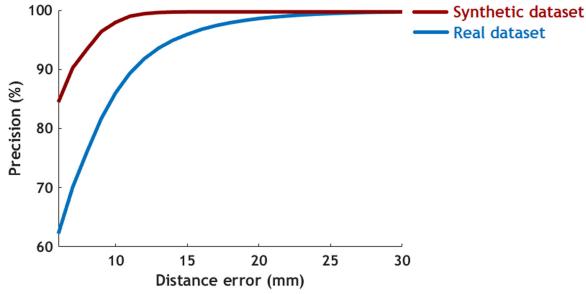


Fig. 10. Method's precision for different threshold errors.

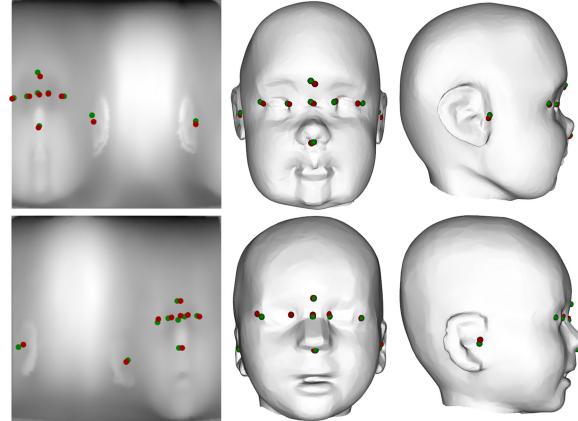


Fig. 11. Detection results examples using the multi-branch approach. First row: synthetic model; Second row: real model; The ground-truth and the detected landmarks are presented in green and red, respectively.

2) Cranial Asymmetry Indexes: Different anthropometric measurements can be estimated on the previous measurement plane, namely cranial width, cranial length, and oblique transcranial diagonals. The cranial width/length corresponds to the maximum head width/length on the measurement plane, while the oblique transcranial diagonals correspond to a diagonal from the anterior/posterior left to the posterior/anterior right on the measurement plane and are measured with an angle α . The volume of each head quadrant can also be computed. In this work, four cranial asymmetry indexes estimated for the measurements were evaluated, namely the cephalic index (ratio between the cranial width and length), the transcranial diagonal difference (difference between transcranial diagonals), anterior cranial symmetry (ratio of the smaller anterior quadrant volumes to the largest anterior quadrant volume), and posterior cranial

asymmetry (ratio of the smaller posterior quadrant volumes to the largest posterior quadrant volume) [4].

3) Clinical Assessment: To perform the clinical validation, a comparison between the cranial indexes estimated using the detected landmarks and from the manual ones was established. This analysis can be seen in Fig. 12, where it is presented the Bland-Altman plot analysis of the automatic method against manual analysis for the cranial indexes. In the analyses, the biases (average difference between estimations) and limits of

TABLE I
ACCURACY OF THE PROPOSED METHOD IN THE REAL DATASET ASSESSED IN TERMS OF ERROR (MM)

	GL	SL	SN	R-EX	L-EX	R-ED	L-ED	R-TR	L-TR
Eval₁	6.3±4.5	3.7±2.9	3.1±3.0	7.6±4.7	9.2±4.9*	3.9±3.1	3.5±2.8	12.5±13.4*	18.9±18.8*
Eval₂	4.0±3.8*	3.7±2.6	4.1±5.8	8.5±8.4	10.2±17.9	6.2±17.1	11.5±28.1*	43.5±53.0*	43.0±51.8*
Eval₃	29.3±37.5*	9.5±24.5	28.3±40.7*	13.2±19.8*	13.5±21.7	8.4±16.6	9.4±25.1	23.4±40.2*	32.5±47.3*
Eval₄	9.2±13.5	5.3±8.3	3.9±5.0*	11.5±10.1*	14.1±14.1*	5.7±6.2	5.5±8.4	16.5±27.0*	28.6±40.7*
Eval₅ (Proposed)	4.5±6.8	3.6±2.6	2.7±3.6	5.3±3.5*	4.9±3.9*	3.3±3.3	2.9±2.3*	6.0±5.7	7.1±5.6
Inter-observer	6.2±3.0	3.4±2.3	2.3±1.5	7.3±3.8	6.8±3.7	3.2±1.7	3.8±1.8	6.1±2.7	8.8±3.5

* $p < 0.05$ in a two tailed paired t -test assessed between the detection results and the inter-observer variability.

TABLE II
COMPARISON OF THE PROPOSED DETECTOR WITH STATE-OF-THE-ART DETECTORS (ERROR IN MM)

	GL	SL	SN	R-EX	L-EX	R-ED	L-ED	R-TR	L-TR	Average
Proposed	4.5±6.8	3.6±2.6	2.7±3.6	5.3±3.5	4.9±3.9	3.3±3.3	2.9±2.3	6.0±5.7	7.1±5.6	4.5±4.1
Alignment-based	4.6±3.4	3.6±2.5	4.0±2.6*	7.4±13.0	10.2±14.2	5.2±8.2	3.7±2.2	14.8±20.2*	16.2±19.2*	7.8±9.5*
Deep-MVLM	6.3±3.0	2.3±1.1*	3.3±2.1	8.8±11.1*	10.3±12.3	4.8±8.0	3.1±1.7	15.4±23.0*	10.2±19.1	6.3±9.1*
Prob-PointNet	3.8±2.7	3.7±3.3	4.3±2.4*	6.4±3.1	5.0±3.8	3.3±3.3	2.7±3.2	7.5±5.4	7.2±5.5	4.9±3.6

* $p < 0.05$ in a two tailed paired t -test assessed between the state-of-the-art method and the proposed detector.

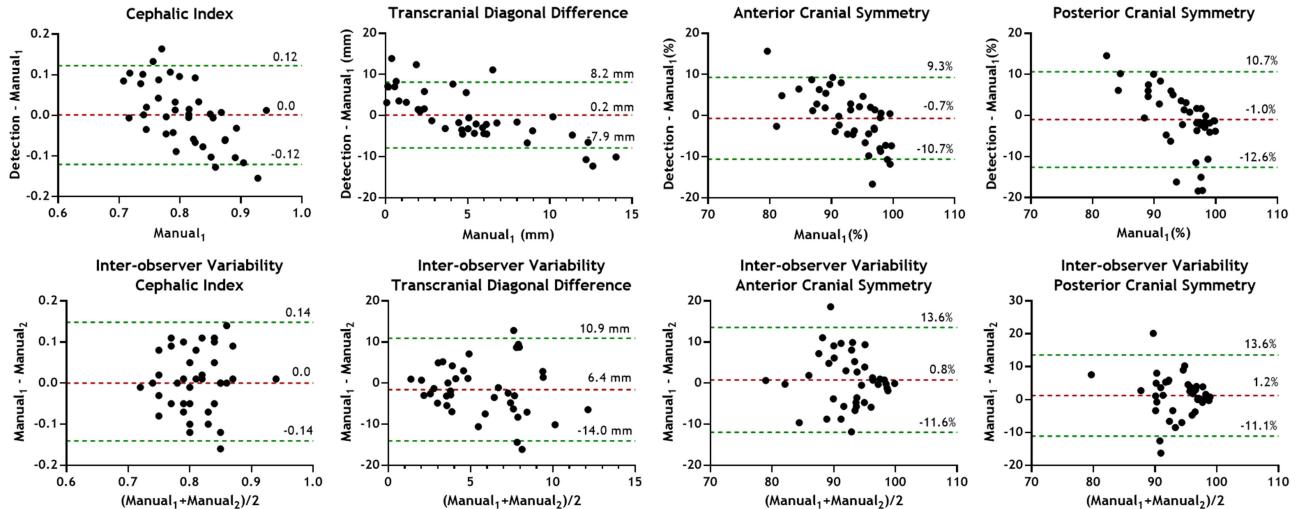


Fig. 12. Bland-Altman analysis for the estimated cranial indexes in the real dataset.

agreement (LOA, 1.96σ) were assessed. Moreover, the difference between observers is presented. For each pair, a two-tailed paired t -test and f -test were used to detect statistically significant differences ($p > 0.05$) in the biases and LOAs, respectively.

E. Runtime Analysis

Concerning the computational burden of the first stage of the method, is in mean ~ 2.2 s in a MATLAB implementation. Regarding the DL method, it took on average 0.1s to process one 2D image. Thus, the average processing time is 2.5s for each model. The runtime analysis was performed on a computer with a CPU: i7-7700HQ @ 3.56GHz, RAM: 16GBytes @ 1.2GHz, and GPU: GTX1070 @ 1.443GHz. The average time required for the manual labeling of all the landmarks by an experienced observer is ~ 30 s.

VI. DISCUSSION

An automatic method to detect anthropometric landmarks in 3D head surfaces was proposed along with a new strategy for 3D data augmentation. The results obtained in different experiments show the potential of the methods. Concerning the method's accuracy in a synthetic testing dataset, Fig. 8 shows the good accuracy of the method for the different evaluations conducted. The obtained results corroborated the feasibility of 3D landmark detection using a 2D representation. For an accurate detection, the 2D representation must preserve important head features. The proposed shape descriptors appear to accurately preserve salient geometric properties of the 3D shape, providing meaning to the 2D representation by characterizing different surface features so that specific characteristics can be correctly distinguished. Moreover, the proposed multi-branch configuration

showed superior results for most landmarks when compared with a single-branch one, achieving an average mean error lower than 3 mm for both training with and without the proposed augmentation. This is explained since the proposed multi-branch approach increases the anatomical information used in the DL. Indeed, some features of the 3D model are more representative in one given shape descriptor, while other features can be more evident in others. Thus, the multi-branch configuration can effectively extract features and details from different 2D representations. Plus, despite the higher performance in terms of mean error, statistically significant differences were obtained between *Eval₅* and the remaining experiments, corroborating the added value of using multiple descriptors simultaneously.

To evaluate the detection method in a real scenario, a validation was also done in the clinical dataset. Fig. 9 and Table I demonstrate the performance of the automatic method in this database, where results comparable to the inter-observer variability were obtained. Based on the results, the proposed detector outperformed the inter-observer variability for some landmarks, including the TRs that present the highest errors. Statistically significant differences were obtained between the multi-branch approach (*Eval₅*) and the four single-branch experiments, where *p*-values lower than 0.05 (two-tailed *t*-test) were obtained, indicating superior performance of the multi-branch network also for the real dataset. When compared with Fig. 8, it can be seen that the real dataset benefits more from the multi-branch approach. In a single-branch configuration, some shape descriptors fail to detect specific landmarks in challenging real models, once its independent information is not enough for the detection, which can be easily understood by Fig. 9 and Table I. However, when together with other descriptors, an individual descriptor provides additional information to describe landmark local surface properties. Thus, when fusing multiple maps, the method performance is improved, showing high performance for this automatic task. Moreover, when compared with the state-of-the-art, Table II shows that the proposed detector outperforms the other methods in locating most of the landmarks, achieving the lowest mean average error. The better performance of the proposed detector when compared with the 3D alignment strategy and the probabilistic point cloud-based approach corroborates the feasibility of the proposed 2D representation, which also proved to be more accurate than the projection-based multi-view approach. This can be explained once the proposed representation explicitly contains information of different shape descriptors whereas traditional multi-view approaches lose geometric information. Plus, once the proposed approach produces a structured and ordered representation, an advantage over raw point-clouds is also verified.

By comparing the synthetic and real results, it was observed an inferior performance at the real database. This difference can be in part explained by the training stage, where only synthetic cases were used. To improve the detection, training with real and synthetic data could be an interesting option that is expected to be explored in the future. However, the obtained performance is still appetitive for clinical practice.

Besides landmark detection, a 3D augmentation strategy was proposed. Analyzing Fig. 8 and Fig. 9, it can be seen that the

results using a training process only with basic augmentation produce not as good results when compared with a training with the proposed augmented models. When compared with a 3DMM-based dataset representation, a better performance was also obtained using the proposed augmented dataset, considering the average error for all the landmarks. This suggests that deforming a model based on the principal shape variations of anthropometric landmarks produces a training dataset robust and variable enough to produce good detection results. Plus, once the anthropometric landmarks are in full correspondence across all models, pre-processing of the entire models to be in full correspondence, which may be challenging in a 3DMM-based approach, is avoided. Moreover, statistical differences between basic/3DMM-based and deformation-based augmentation box-plots were achieved for some experiments. Furthermore, once the augmented models are generated by transformations into principal shape models of a training population, unrealistic models that could hinder the training are avoided. In detail, to analyze the realism of the augmented models, a quantitative evaluation of the generated models by estimating the traditional clinical indexes was performed, obtaining for all indexes values within normal references.

Clinical validation of cranial asymmetry indexes was also performed. When evaluating statistically significant biases between automatic results, a *p*-value higher than 0.05 was achieved for all clinical measurements in a two-tailed *t*-test against 0. Thus, no statistically significant biases (red dashed line in Fig. 12) were obtained by the automatic method for the measurements. Moreover, when assessing the agreement between manual observers and automatic measurements, no statistically significant biases were obtained for the anterior and posterior cranial symmetry indexes (*p* > 0.05 in a two-tailed F-Test against the intra-observer), corroborating the high performance of the automatic method for this task.

Finally, an important aspect to evaluate is the computational burden of the method, which only takes 2.5 s to detect all landmarks, showing a clear advantage of the framework over the tedious and time-consuming manual estimation of the cranial indexes and landmark labeling. Despite the proposed method was only evaluated in nine landmarks, the current solution is independent of the number of landmarks. Thus, it can be easily expanded for more landmarks without increasing the computational time, allowing to expand the current method to other head shape analysis tasks or other anatomical studies, increasing the clinical interest of this application. Moreover, the proposed approach has also potential to be applied in other non-clinical tasks, e.g., face recognition and can be applied to different types of non-clinical data, e.g., facial surfaces.

One should notice that a current limitation of the proposed approach is the detection of landmarks located at non-featured and flat regions. At these regions, different shape/spatial descriptor must be studied in future work. Other limitation is the lack of benchmarks of infant's head to evaluate the method performance in other clinical databases and against other possible solutions. As a first approach to create pediatric head benchmarks, we will freely release our synthetic benchmark, promoting its application by other researcher teams.

VII. CONCLUSION

An automatic method for detecting landmarks in infant's head models was presented. The results corroborated the accuracy and robustness of the method. The 2D representation approach correctly differentiates specific characteristics of the landmarks. Moreover, the proposed network is suitable for landmark location prediction. Overall, the results show the advantage of the proposed framework over the traditional manual estimation of cranial indexes and landmark labeling. The performed clinical validation corroborated the clinical added value of the current method for the automatic shape analysis of infants' heads. Moreover, the method can also be applied to different landmark detection tasks.

REFERENCES

- [1] W. K. Peitsch, C. H. Keefer, R. A. LaBrie, and J. B. Mulliken, "Incidence of cranial asymmetry in healthy newborns," *Pediatrics*, vol. 110, no. 6, Dec. 2002, Art. no. e72.
- [2] S. If, M. Rüdiger, D. Konstantelos, K. Wahls, and W. Burkhardt, "Early head development prevalence of head deformities in preterm infants at term equivalent age ☆," *Early Hum. Dev.*, vol. 89, pp. 1041–1047, 2013.
- [3] A. L. C. Martiniuk, C. Vujojich-Dunn, M. Park, W. Yu, and B. R. Lucas, "Plagiocephaly and developmental delay: A systematic review," *J. Dev. Behav. Pediatr.*, vol. 38, no. 1, pp. 67–78, 2017.
- [4] H. Aarnivala *et al.*, "Accuracy of measurements used to quantify cranial asymmetry in deformational plagiocephaly," *J. Cranio-Maxillofac. Surg.*, vol. 45, no. 8, pp. 1349–1356, 2017.
- [5] M. J. Cho, L. L. Borchert, and A. A. Kane, "Diagnostic yield of routine skull radiographs in infants with deformational plagiocephaly," *Cleft Palate-Craniofacial J.*, vol. 54, no. 5, pp. 497–501, 2017.
- [6] P. Meyer-Marcotty, H. Böhm, C. Linz, J. Kochel, A. Stellzig-Eisenhauer, and T. Schweitzer, "Three-dimensional analysis of cranial growth from 6 to 12 months of age," *Eur. J. Orthod.*, vol. 36, no. 5, pp. 489–496, 2014.
- [7] H. R. Torres *et al.*, "Deep learning-based detection of anthropometric landmarks in 3D infants head models," in *Proc. SPIE Med. Imag.*, p. 1095034, 2019.
- [8] A. Colombo, C. Cusano, and R. Schettini, "3D face detection using curvature analysis," *Pattern Recognit.*, vol. 39, no. 3, pp. 444–455, 2006.
- [9] A. S. Mian, M. Bennamoun, and R. Owens, "Keypoint detection and local feature matching for textured 3D face recognition," *Int. J. Comput. Vis.*, vol. 79, no. 1, pp. 1–12, 2008.
- [10] J. D'Hose, J. Colineau, C. Bichon, and B. Dorizzi, "Precise localization of landmarks on 3D faces using gabor wavelets," in *Proc. IEEE Conf. Biometrics : Theory , Appl. Syst.*, pp. 1–6, 2007.
- [11] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with gabor wavelets," in *Proc. - 3rd IEEE Int. Conf. Autom. Face Gesture Recognit.*, pp. 200–205, 1998.
- [12] E. Vezzetti, F. Marcolin, S. Tornincasa, L. Ulrich, and N. Dagnes, "3D geometry-based automatic landmark localization in presence of facial occlusions," *Multimed. Tools Appl.*, vol. 77, no. 11, pp. 14177–14205, 2018.
- [13] A. Abu, C. G. Ngo, N. I. A. Abu-Hassan, and S. A. Othman, "Automated craniofacial landmarks detection on 3D image using geometry characteristics information," *BMC Bioinf.*, vol. 19, no. 13, p. 548, 2019.
- [14] S. Gupta, M. K. Markey, and A. C. Bovik, "Anthropometric 3D face recognition," *Int. J. Comput. Vis.*, vol. 90, no. 3, pp. 331–349, 2010.
- [15] P. Nair and A. Cavallaro, "Region segmentation and feature point extraction on 3D faces using a point distribution model," in *Proc. - Int. Conf. Image Process.*, 2006, pp. 85–88.
- [16] P. Nair and A. Cavallaro, "3-D face detection, landmark localization, and registration using a point distribution model," *IEEE Trans. Multimed.*, vol. 11, no. 4, pp. 611–623, Jun. 2009.
- [17] T. Baltrušaitis, P. Robinson, and L. P. Morency, "3D Constrained local model for rigid and non-rigid facial tracking," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, 2012, pp. 2610–2617.
- [18] M. C. El Rai, C. Tortorici, H. Al-Muhairi, H. Al Safar, and N. Werghi, "Landmarks detection on 3D face scans using local histogram descriptors," in *Proc. 18th Mediterr. Electrotech. Conf. Intell. Effic. Technol. Serv. Citizen*, 2016, pp. 18–20.
- [19] S. Liang, J. Wu, S. M. Weinberg, and L. G. Shapiro, "Improved detection of landmarks on 3D human face data," in *Proc. Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. EMBS*, pp. 6482–6485, 2013.
- [20] F. M. Sukno, J. L. Waddington, and P. F. Whelan, "3-D Facial landmark localization with asymmetry patterns and shape regression from incomplete local features," *IEEE Trans. Cybern.*, vol. 45, no. 9, pp. 1717–1730, Sep. 2015.
- [21] X. Cheng and F. Da, "3D Facial landmark localization based on two-step keypoint detection," in *Proc. Int. Conf. Audio, Lang. Image Process.*, 2018, pp. 406–412.
- [22] T. Terada, Y. Chen, and R. Kimura, "3D Facial landmark detection using deep convolutional neural networks," in *Proc. 14th Int. Conf. Nat. Comput. Fuzzy Syst. Knowl. Discov.*, 2018, pp. 390–393.
- [23] S. Z. Gilani, A. Mian, and P. Eastwood, "Deep, dense and accurate 3D face correspondence for generating population specific deformable models," *Pattern Recognit.*, vol. 69, pp. 238–250, 2017.
- [24] J. F. Teichgraeber *et al.*, "Deformational posterior plagiocephaly: Diagnosis and treatment," *Cleft Palate-Craniofacial J.*, vol. 39, no. 6, pp. 582–586, 2002.
- [25] E. Ahmed *et al.*, "A survey on deep learning advances on different 3D data representations," *Comput. Vis. Pattern Recognit.*, vol. 1, no. 1, pp. 1–35, 2018.
- [26] Q. Li, Z. Han, and X. M. Wu, "Deeper insights into graph convolutional networks for semi-supervised learning," in *Proc. 32nd AAAI Conf. Artif. Intell.*, 2018, pp. 3538–3545.
- [27] A. S. Gezawa, Y. Zhang, Q. Wang, and L. Yunqi, "A review on deep learning approaches for 3d data representations in retrieval and classifications," *IEEE Access*, vol. 8, pp. 57566–57593, 2020.
- [28] Q. Duan, E. D. Angelinib, and A. F. Lainea, "Real-time segmentation by active geometric functions," *Comput. Methods Programs Biomed.*, vol. 98, no. 3, pp. 223–230, 2011.
- [29] E. Abbena, S. Salamon, and A. Gray, *Modern Differential Geometry of Curves and Surfaces with Mathematical*, Chapman & Hall/CRC, London, 1997.
- [30] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. Comput. Vis. Pattern Recognit.*, 2014, pp. 1–14.
- [31] P. Perakis, I. C. Society, and G. Passalis, "3D Facial landmark detection under large yaw and expression variations," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 7, pp. 1552–1564, Jul. 2013.
- [32] M. Bastioni, S. Re, and S. Misra, "Ideas and methods for modeling 3D human figures: The principal algorithms used by makehuman and their implementation in a new approach to parametric modeling," in *Proc. 1st Bangalore Annu. Compute Conf.*, 2008, pp. 1–6.
- [33] W. Burkhardt, D. Schneider, G. Hahn, D. Konstantelos, H. G. Maas, and M. Rüdiger, "Non-invasive estimation of brain-volume in infants," *Early Hum. Dev.*, vol. 132, pp. 52–57, May 2019.
- [34] F. Veloso *et al.*, "Synthetic infant head shapes with deformational plagiocephaly: Concept and 3D model parameterization," in *Proc. Int. Conf. Serious Games Appl. for Health*, pp. 1–7, 2019.
- [35] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3D faces," in *Proc. 26th Annu. Conf. Comput. Graph. Interact. Tech. SIGGRAPH 1999*, pp. 187–194, 1999.
- [36] B. Oliveira *et al.*, "Automatic strategy for extraction of anthropometric measurements for the diagnostic and evaluation of deformational plagiocephaly from infant's head models," in *Proc. SPIE Med. Imag. Int. Soc. Opt. Photon.*, vol. 10950, p. 10950A, Mar. 2019.
- [37] J. Yang, H. Li, D. Campbell, and Y. Jia, "Go-ICP: A globally optimal solution to 3D ICP point-set registration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 38, no. 11, pp. 2241–2254, Nov. 2016.
- [38] R. R. Paulsen, K. A. Juhl, T. M. Haspang, T. Hansen, M. Ganz, and G. Einarsson, "Multi-view consensus CNN for 3D facial landmark placement," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinf.)*, vol. 11361, no. LNCS, pp. 706–719, 2019.
- [39] C. R. Qi, H. Su, K. Mo, and L. J. Guibas, "PointNet: Deep learning on point sets for 3D classification and segmentation," in *Proc. - 30th IEEE Conf. Comput. Vis. Pattern Recognit.*, 2017, pp. 77–85.