



**HAL**  
open science

## Bayesian Logistic Shape Model Inference: application to cochlear image segmentation

Zihao Wang, Thomas Demarcy, Clair Vandersteen, Dan Gnansia, Charles Raffaelli, Nicolas Guevara, Hervé Delingette

► **To cite this version:**

Zihao Wang, Thomas Demarcy, Clair Vandersteen, Dan Gnansia, Charles Raffaelli, et al.. Bayesian Logistic Shape Model Inference: application to cochlear image segmentation. *Medical Image Analysis*, In press, 75, pp.102268. 10.1016/j.media.2021.102268 . hal-03372777

**HAL Id: hal-03372777**

**<https://inria.hal.science/hal-03372777v1>**

Submitted on 11 Oct 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Bayesian Logistic Shape Model Inference : application to cochlea image segmentation

Wang Zihao<sup>a,b</sup>, Demarcy Thomas<sup>e</sup>, Vandersteen Clair<sup>b,c</sup>, Gnansia Dan<sup>e</sup>, Raffaelli Charles<sup>b,d</sup>, Guevara Nicolas<sup>b,c</sup>, Delingette Hervé<sup>a,b</sup>

<sup>a</sup>*Inria Sophia Antipolis Méditerranée, 2004 Route des Lucioles, 06902 Valbonne, FRANCE*

<sup>b</sup>*Université Côte d'Azur, 28 Avenue de Valrose, 06108 Nice, FRANCE*

<sup>c</sup>*Head and Neck University Institute, Nice University Hospital, 31 Avenue de Valombrose, 06100 Nice, FRANCE*

<sup>d</sup>*Department of Radiology, Nice University Hospital, 31 Avenue de Valombrose, 06100 Nice, FRANCE*

<sup>e</sup>*Oticon Medical, 14 Chemin de Saint-Bernard Porte, 06220 Vallauris, FRANCE*

---

## Abstract

Incorporating shape information is essential for the delineation of many organs and anatomical structures in medical images. While previous work has mainly focused on parametric spatial transformations applied on reference template shapes, in this paper, we address the Bayesian inference of parametric shape models for segmenting medical images with the objective to provide interpretable results. The proposed framework defines a likelihood appearance probability and a prior label probability based on a generic shape function through a logistic function. A reference length parameter defined in the sigmoid controls the trade-off between shape and appearance information. The inference of shape parameters is performed within an Expectation-Maximisation approach where a Gauss-Newton optimization stage allows to provide an approximation of the posterior probability of shape parameters.

This framework is applied to the segmentation of cochlea structures from clinical CT images constrained by a 10 parameter shape model. It is evaluated on three different datasets, one of which includes more than 200 patient images. The results show performances comparable to supervised methods and better than previously proposed unsupervised ones. It also enables an analysis of parameter distributions and the quantification of segmentation uncertainty including the effect of the shape model.

**Keywords:** Bayesian Inference, Image Segmentation, Shape Modeling

---

---

\*Corresponding author:

Email address: [zihao.wang@inria.fr](mailto:zihao.wang@inria.fr) (Wang Zihao)

## 1. Introduction

Several anatomical structures have a typical shape, such that a medical expert can easily recognize them from their three-dimensional representation. This is for instance the case of basal ganglia within the brain (Ashburner and Friston, 2005a), but also of abdominal structures, such as the liver or kidneys. Another emblematic example is the cochlea which is a small organ within the inner ear having a remarkable spiraling configuration where mechanical waves are transformed into electrical stimulation of the auditory nerve. The cochlea shape is complex as it completes around two and a half turns with its centerline closely resembling a logarithmic spiral helix (Cohen et al., 1996; Baker, 2008). Its segmentation from CT images of the temporal bone is challenging since those images have low resolution with respect to the anatomy of the cochlea: the cochlea dimension is about  $8.5 \times 7 \times 4.5 \text{ mm}^3$  while the typical CT voxel size is larger than  $0.2 \text{ mm}$  which is weakly visible for the fine structures of the chambers. In addition, the cochlea is filled with fluids that can be found in the vestibular system and other neighbouring structures, with similar appearance in CT images.

Supervised learning (e.g. Deep Learning) is an effective way to perform image segmentation or processing in many cases. Specifically, in inner ear CT imaging analysis, many works achieved impressive results (Lv et al., 2021; Raabid et al., 2021; Heutink et al., 2020; Wang et al., 2019; Li et al., 2021; Alshazly et al., 2019; Zhang et al., 2019; Wang et al., 2020b). However, supervised learning methods have also many limitations. First, creating dataset annotations is time consuming, possibly preventing the creation of massive training datasets. In the cochlea case, a well trained ENT surgeon would need at least ten minutes to segment each 3D cochlea volume. Second, due to the potential overfitting related to the limited training set, the output of such supervised algorithm is likely to fall outside the shape space of the structure of interest.

Shape-based image segmentation can overcome the above limitations since the optimization of the model can be done in an unsupervised or weakly supervised way. Besides, the recovered shape parameters make a natural compact representation that is useful for shape analysis and even clinical applications. In this paper, we consider shapes that are either defined as an explicit  $\mathcal{S}(\theta_S) \in \mathbb{R}^d$  or implicit  $\mathcal{S}(\theta_S, \mathbf{x}) = 0$  parametric shape models where  $\theta_S$  is a set of shape parameters and  $\mathbf{x} \in \mathbb{R}^d$ , is any point in space ( $d = 2, 3$ ).

Those parametric shape models serve to guide the delineation of such anatomical structures by constraining the shape space of the segmented object. We can roughly split the shape-based image segmentation methods into two sets of methods. A first set optimizes the shape parameters  $\theta_S$  by minimizing the sum of a regularizing term  $E_R(\theta_S)$  and an image term  $E_I(\mathcal{S}(\theta_S), I, \theta_I) : \hat{\theta}_S = \arg \min_{\theta_S} E_I(\mathcal{S}(\theta_S), I, \theta_I) + E_R(\theta_S)$  where  $\theta_I$  is a set of image parameters that may also be optimized. This iconic shape fitting principle is typically used in the classical active shape model (Cootes et al., 1995; Heimann et al., 2007) and their extensions (Cremers et al., 2003). Various generic image terms may be considered for instance as those explored in (Tsai et al., 2003). A second set of methods uses the shape model  $\mathcal{S}(\theta_S)$  as a shape prior instead of a shape space. Several shape constraints have been introduced within several image segmentation frameworks including level-sets (Chan and Zhu, 2005; Cremers, 2003), free-form deformation space (Rueckert et al., 2003b) or implicit template deformation (Prevost

et al., 2013). While those methods have greater shape flexibility for delineating structures, it is often difficult to set the coefficients weighting the shape constraint with other image terms. Those two sets of shape based segmentation methods are expressed as energy minimization problems, thus only allowing to have point estimates of shape parameters and not their posterior probabilities.

Another common shape representation consists in specifying a parametric spatial transformation  $\mathcal{T}(\theta_D) : \mathbb{R}^d \rightarrow \mathbb{R}^d$  acting on a template shape  $\mathcal{S}(\theta_0) \in \mathbb{R}^d$  leading to an indirect shape parameterization :  $\mathcal{S}(\theta_D) = \mathcal{T}(\theta_D) \circ \mathcal{S}(\theta_0)$ . This formulation of shape modeling based on a deformable template leads to solving a joint segmentation and registration problem. More precisely, several authors (Ashburner and Friston, 2005b; Pohl et al., 2006a) defined generative image and shape models and performed statistical variational inference to optimize their parameters and hyperparameters. Priors on the deformation space based for instance on minimal elastic energy (Van Leemput, 2009), were applied on triangular or tetrahedral mesh templates. Other shape priors were defined as restricted Boltzmann machines (Agn et al., 2019) or as shape-odds (Elhabian and Whitaker, 2017). In most cases, optimal shape parameters (e.g. mesh vertex positions) are obtained as maximum a posteriori but not their posterior probability. Uncertainty quantification of image registration algorithms has been tackled in some research papers (Simpson et al., 2012; Wang et al., 2018; Le Folgoc et al., 2017) based on a low dimensional representation of deformation space and Laplace approximation.

In this paper, we propose a novel Bayesian framework for shape constrained image segmentation based on parametric shape models (instead of parametric spatial transformations) where the output segmentation is driven by a shape model but without restricting it to a low dimensional space. The proposed approach is generic as it is suitable for any explicit  $\mathcal{S}(\theta_S)$  and implicit  $\mathcal{S}(\theta_S, \mathbf{x}) = 0$  parametric shape models associated with any appearance models representing the intensity distributions inside background or foreground regions. It is based on a logistic shape prior defined as the sigmoid of a shape function (e.g. signed distance map) defined over the image domain. Inferences of shape and intensity parameters are performed by maximizing the joint image and shape parameters probability  $p(\theta_S, \theta_I, I)$  with an Expectation-Maximization algorithm. We show that this optimization boils down to having the posterior label distribution as close as possible (in terms of Kullback Leibler divergence) from both the likelihood and shape prior distributions. A Gauss-Newton optimization method is introduced to optimize the shape parameters leading to closed form updates similarly to iterative reweighted least squares schemes. It outputs the most probable shape and imaging parameters but also an approximation of the posterior shape parameter probability which is essential for estimating the segmentation uncertainty.

This framework is applied to the problem of cochlea segmentation on CT images based on a parametric shape model with 10 parameters, and an imaging model defined as a mixture of Student's  $t$ -distributions. It results in the reconstruction of cochlea structures in 2 small datasets consisting of paired CT and  $\mu CT$  post-mortem images and one large dataset of nearly 200 patient CT images. We showed that the proposed framework leads to state of the art reconstruction performances as well as the recovery of consistent shape parameter distributions and the estimation of segmentation uncertainty.

The main contributions of this article are:

- A novel framework for image segmentation that combines probabilistic appearance and shape models. It is generically defined for parametric shape functions rather than parametric space transformations. The trade-off between the appearance and shape models is governed by an interpretable parameter : the reference length.
- A Gauss-Newton optimization method of the shape parameters which also produces a posterior approximation of those shape parameters.
- A method for uncertainty quantification of image segmentation which takes into account the shape uncertainty.
- A segmentation method of the cochlea in clinical CT images which provides state-of-the-art results and interpretable shape parameters.

We present below the framework of the logistic shape model (section 2), the shape and intensity models used specifically for cochlea segmentation (section 3), and the segmentation results on 3 clinical and pre-clinical datasets (section 4).

## 2. Method

### 2.1. Shape-based Generative Probabilistic Model

We consider an observed image  $I$  consisting of  $N$  voxels  $I_n \in \mathbb{R}$ ,  $n = 1, \dots, N$ , for which we seek to solve a binary segmentation problem guided by a shape model. That model is defined either as in a parametric form as  $\mathcal{S}(\theta_S) \in \mathbb{R}^d$ ,  $d = 2, 3$  or in an implicit form as  $\mathcal{S}(\theta_S, \mathbf{x}) = 0$ . In the case of parametric shape models, one can define an associated implicit function  $\text{SDM}(\mathcal{S}(\theta_S), \mathbf{x}) = 0$  as the signed distance map defined at point  $\mathbf{x}$ . Therefore, we propose to unify notations for both parametric and implicit cases by stating the existence of a *shape function*  $\hat{\mathcal{S}}(\theta_S, \mathbf{x}) \in \mathbb{R}$  whose zero level defines a shape and whose sign indicates if a point is inside (positive) or outside (negative). Note that with this hypothesis, a shape corresponds to a (smooth) manifold of co-dimension 1 without borders, thus defining a partition of the image into inside and outside regions.

A binary label variable  $Z_n \in \{0, 1\}$  is defined at each voxel specifying if voxel  $n$  belongs to the background or foreground regions. A probabilistic intensity distribution model is defined for each region  $p(I_n | Z_n = k, \theta_I^k)$ ,  $k = 0, 1$  controlled by the intensity parameter array  $\theta_I^k$ . The arrays for background ( $k = 0$ ) and foreground ( $k = 1$ ) are concatenated into the intensity parameter array  $\theta_I$ . This appearance model can be either supervised, e.g. a trained convolutional neural network, or unsupervised, e.g. a Gaussian mixture model. In the remainder, we assume the latter case and therefore we define mechanisms to optimize the appearance parameters  $\theta_I$ . In the supervised case, the steps involving the update of  $\theta_I$  should be ignored.

We enforce a spatial correlation between the label of each voxel by specifying their *a priori* dependence on the shape model  $\hat{\mathcal{S}}(\theta_S, \mathbf{x})$ . More precisely, we define a prior

probability for voxel  $n$  to belong to the foreground region as follows:

$$\begin{aligned}
 p(Z_n = 1|\theta_S) &= \sigma\left(\frac{\tilde{S}(\theta_S, \mathbf{x}_n)}{l_{\text{ref}}}\right) \\
 p(Z_n = 0|\theta_S) &= 1 - p(Z_n = 1|\theta_S) = \sigma\left(-\frac{\tilde{S}(\theta_S, \mathbf{x}_n)}{l_{\text{ref}}}\right)
 \end{aligned} \tag{1}$$

where  $\sigma(x)$  is the sigmoid (or logistic) function,  $\mathbf{x}_n$  is the position of voxel  $n$  and  $l_{\text{ref}}$  is a reference length. With that definition, the prior probability will be close to 1 inside the object, close to 0 outside and equal to 0.5 on the shape boundary. We call this formulation of the label prior, the *logistic shape model* as it combines shape information into a probability distribution through a logistic function. This definition of the shape prior is related to several prior work in the literature such as probabilistic atlases and LogOdds maps (Pohl et al., 2006b), continuous STAPLE (Commowick and Warfield, 2009), a nd label fusion (Sabuncu et al., 2010).

The quantity  $l_{\text{ref}}$  is a characteristic length which controls the slope of the prior probability next to the object boundary. This parameter also influences the trade-off between intensity and shape information in the segmentation process as discussed in section 2.5. The shape parameters  $\theta_S$  are themselves regarded as random variables with a multivariate Gaussian prior controlled by hyper-parameters  $\alpha$ :  $p(\theta_S|\alpha)$ . The intensity parameters may also optionally be considered as random variables with hyper-parameter  $\beta$  as  $p(\theta_I|\beta)$ . The shape based generative model is summarized in Fig. 3:(a).

## 2.2. Logistic Shape Model Framework

With the proposed generative model, given an image, the objective is to infer the most probable values of the intensity  $\hat{\theta}_I$  and shape parameters  $\hat{\theta}_S$  which will lead to the estimation of the posterior label probabilities given by :

$$p(Z_n = 1|I_n, \theta_I, \theta_S) = \frac{p(I_n|Z_n = 1, \theta_I^1)p(Z_n = 1|\theta_S)}{\sum_{k=0}^1 p(I_n|Z_n = k, \theta_I^1)p(Z_n = k|\theta_S)} \tag{2}$$

That posterior probability is clearly a compromise between shape information stored in the prior  $p(Z_n = 1|\theta_S)$  and appearance information stored into the likelihood  $p(I_n|Z_n = 1, \theta_I^1)$ . The *segmented region of interest* (SROI) then corresponds to voxels for which  $p(Z_n = 1|I_n, \theta_I, \theta_S) \geq \frac{1}{2}$ . In addition, the logistic shape model framework recovers the most likely shape parameters  $\hat{\theta}_S$  that corresponds to the *segmented shape instance* (SSI) which is the best fit of the shape model in that image. Finally, we will show that we can approximate the posterior shape parameter  $p(\theta_S|I)$  in order to capture the uncertainty in the shape parameter estimation.

The optimization of the intensity and shape parameters is done by maximizing the

the log-joint intensity and parameters probability :

$$\begin{aligned}
(\hat{\theta}_S, \hat{\theta}_I) &= \arg \max_{\theta_S, \theta_I} \log p(I, \theta_S, \theta_I) = \arg \max_{\theta_S, \theta_I} \mathcal{L}(\theta_S, \theta_I) \\
\mathcal{L}(\theta_S, \theta_I) &= \log p(I|\theta_S, \theta_I) + \log p(\theta_S) + \log p(\theta_I) \\
&= \sum_{n=1}^N \log \left( \sum_{k=0}^1 p(I_n|Z_n = k, \theta_I) p(Z_n = k|\theta_S) \right) + \\
&\quad \log p(\theta_S) + \log p(\theta_I)
\end{aligned} \tag{3}$$

In the log-joint probability  $\mathcal{L}(\theta_S, \theta_I)$  we have marginalized out the hidden label variables  $Z_n$  and used the conditional independence of variables  $I_n$  given  $\theta_S$ .

### 2.3. Expectation-Maximization Inference

The direct optimization of  $\mathcal{L}(\theta_S, \theta_I)$  can be done by any optimization toolbox but it is difficult due to the possible encountered overflows/underflows caused by the log-sum-exp expressions.

This is why we propose to follow the Expectation-Maximization (EM) algorithm which relaxes that optimization problem into several optimizations over simpler problems. We proceed by introducing  $N$  variables  $u_n$  that are surrogates for the posterior label probability  $p(Z_n = 1|I_n, \theta_S, \theta_I)$  such that  $u_n \in [0, 1]$ . Writing  $U = \{u_n\}$ , we introduce a new augmented criterion  $\mathcal{L}^*(\theta_S, \theta_I, U) = \log p(I, \theta_S, \theta_I) - D_{\text{KL}}(U||p(Z|I, \theta_S, \theta_I))$  by adding the negative Kullback-Leibler divergence between  $u_n$  and the posterior label  $p(Z_n|I_n, \theta_S, \theta_I)$ .

Maximizing  $(\theta_S, \theta_I, U)$  over the augmented criterion  $\mathcal{L}^*(\theta_S, \theta_I, U)$  leads to the same optima in  $(\theta_S, \theta_I)$  than the maximization of  $\mathcal{L}(\theta_S, \theta_I)$  but with simpler expressions:

$$\begin{aligned}
\mathcal{L}^*(\theta_S, \theta_I, U) &= \sum_{n=1}^N \sum_{k=0}^1 u_n^k \log(p(I_n|\theta_S, \theta_I)p(Z_n^k = k|I_n, \theta_S, \theta_I)) \\
&\quad - \sum_{n=1}^N \sum_{k=0}^1 u_n^k \log u_n^k + \log p(\theta_S) + \log p(\theta_I) \\
&= Q(U, \theta_S, \theta_I) + \sum_{n=1}^N H(u_n) + \log p(\theta_S) + \log p(\theta_I)
\end{aligned}$$

where  $Q(U, \theta_S, \theta_I) = \mathbb{E}_U(\log p(I, Z|\theta_S, \theta_I))$  is the conditional expectation of the complete marginal log-likelihood (a.k.a. evidence) and  $H(u_n)$  is the entropy of variable  $u_n$ . The quantity  $Q(U, \theta_S, \theta_I)$  is a lower bound of the log-likelihood since  $H(u_n) > 0$ .

The maximization of the augmented criterion  $\mathcal{L}^*(\theta_S, \theta_I, U)$  is performed by the successive maximization over the  $U$ ,  $\theta_I$  and  $\theta_S$  variables. The **E-step** corresponds to the maximization of  $\mathcal{L}^*(\theta_S, \theta_I, U)$  with respect to  $U$  which sets the surrogate variable  $U$  to the posterior label probability  $u_n = p(Z_n = 1|I_n, \theta_S, \theta_I)$ .

The **MI-step** optimizes the log-joint probability with respect to the appearance variables  $\theta_I$ , which is equivalent to the maximization of  $\mathcal{L}_I = -D_{\text{KL}}(U||p(I|Z, \theta_I)) + \log p(\theta_I|\beta)$ . When the appearance parameters are independent between classes, then  $\log p(\theta_I|\beta) = \sum_{k=0}^K \log p(\theta_I^k|\beta_k)$  and the MI-step splits into 2 independent maximization

over  $\theta_I^k$ ,  $k = 0, 1$  of  $\mathcal{L}_I^k = -\sum_{n=1}^N D_{\text{KL}}(u_n^k \| p(I_n | Z_n = e_k, \theta_I^k)) + \log p(\theta_I^k | \beta_k)$ . For certain well chosen intensity models such as Gaussian mixture models, this optimization leads to closed-form updates of  $\theta_I$ .

Finally, we perform the **MS-step** corresponding to the maximization over shape variables  $\theta_S$  which is equivalent to the maximization of  $\mathcal{L}_S$ :

$$\mathcal{L}_S = -D_{\text{KL}}(U \| p(Z | \theta_S)) + \log p(\theta_S | \alpha)$$

We can see that the EM algorithm preserves an interesting symmetry between shape and appearance information. Indeed, the iterative application of the E, MS and MI steps makes the posterior labels distribution  $U$  as close (in terms of KL divergence) as possible from the likelihood  $p(I | Z, \theta_I)$  and shape prior  $p(Z | \theta_S)$  that the minimization of  $D_{\text{KL}}(U \| p(Z | \theta_S)) + D_{\text{KL}}(U \| p(I | Z, \theta_I))$ . At convergence, the posterior distribution is therefore clearly a compromise between shape and appearance information.

#### 2.4. Optimization of shape parameters $p(\theta_S | I)$

The functional  $\mathcal{L}_S$  is a non trivial function of the parameters  $\theta_S$  as it combines 2 non-linear functions : the sigmoid  $\sigma()$  and the shape function  $\tilde{S}(\theta_S, \mathbf{x}_n)$ :

$$\begin{aligned} \mathcal{L}_S = & -\sum_{n=1}^N \left( u_n \log \sigma \left( \frac{\tilde{S}(\theta_S, \mathbf{x}_n)}{l_{\text{ref}}} \right) + (1 - u_n) \log \sigma \left( -\frac{\tilde{S}(\theta_S, \mathbf{x}_n)}{l_{\text{ref}}} \right) \right) \\ & + \log p(\theta_S | \alpha) + \text{cst} \end{aligned} \quad (4)$$

The functional gradient  $\nabla_{\theta_S} \mathcal{L}_S$  cannot be written in closed form since it requires the computation of the gradient of the scaled shape function at each voxel :  $\mathbf{d}_n = \frac{\nabla_{\theta_S} \tilde{S}(\theta_S, \mathbf{x}_n)}{l_{\text{ref}}} \in \mathbb{R}^{|\theta_S|}$ . Those gradient vectors may be computationally costly to compute, for instance when the shape function is based on a signed distance map of parametric shape models  $\tilde{S}(\theta_S, \mathbf{x}) = \text{SDM}(\mathcal{S}(\theta_S), \mathbf{x})$ . In that case, the  $\mathbf{d}_n$  values are computed by a costly finite difference approximation except for translation and rotation parameters for which they can be computed efficiently (see Appendix A). After combining all  $\mathbf{d}_n$  terms in a gradient matrix  $\mathbf{d} \in \mathbb{R}^{|\theta_S| \times N}$ , the functional gradient can be simplified as  $\nabla_{\theta_S} \mathcal{L}_S = -\mathbf{d}(\mathbf{u} - \boldsymbol{\mu}) + \nabla_{\theta_S} \log p(\theta_S | \alpha)$  where  $\mathbf{u} = (u_1 \dots u_N)^T \in \mathbb{R}^N$  and  $\boldsymbol{\mu} = \left( \sigma \left( \frac{\tilde{S}(\theta_S, \mathbf{x}_1)}{l_{\text{ref}}} \right) \dots \sigma \left( \frac{\tilde{S}(\theta_S, \mathbf{x}_N)}{l_{\text{ref}}} \right) \right)^T \in \mathbb{R}^N$ .

Thus, a first approach for optimizing the shape parameters is to use any quasi-Newton optimization method such as the BFGS algorithm (similarly to (Demarcy, 2017)), since it only requires the computation of the functional gradient and iteratively estimates the Hessian matrix. Yet, this generic optimization was found to be fairly time consuming and sometimes unstable.

Instead, we propose to adopt a Gauss-Newton optimization approach where we approximate the Hessian matrix by ignoring the term involving second order derivatives. More precisely, the Hessian of the functional is computed as  $\mathbf{H} = \nabla_{\theta_S}^2 \mathcal{L}_S = -\nabla_{\theta_S} \mathbf{d} \otimes (\mathbf{u} - \boldsymbol{\mu}) - \mathbf{d} \otimes \nabla_{\theta_S} \boldsymbol{\mu} + \nabla_{\theta_S}^2 \log p(\theta_S | \alpha)$ . After dropping the first term, we get the following approximate Hessian  $\mathbf{H} \approx \tilde{\mathbf{H}} = -\mathbf{d} \otimes \nabla_{\theta_S} \boldsymbol{\mu} + \nabla_{\theta_S}^2 \log p(\theta_S | \alpha)$ . When inserting the expression of the gradient of the prior, we get :  $\tilde{\mathbf{H}} = \mathbf{d} \text{Diag}(\boldsymbol{\mu} \circ (1 - \boldsymbol{\mu})) \mathbf{d}^T + \nabla_{\theta_S}^2 \log p(\theta_S | \alpha)$



where  $\circ$  is the element-wise product between two vectors. This approximate Hessian matrix is positive definite by construction and is then used to perform several Newtons steps.

The sketch of the MS step is shown as algorithm 1 where the shape parameter prior  $p(\delta\theta_S^i|\alpha)$  is arbitrarily chosen as a zero mean Gaussian distribution with covariance  $\Sigma_{\theta_S}^0$ . It consists of two intertwined loops, the innermost performing iteratively the Newton updates and updating the mean, gradient and Hessian values. The outer loop updates the shape function gradient which is potentially a costly step. In line 15 of the algorithm, the  $U$  variable is updated in an E-step in order to speed-up the convergence of the overall EM algorithm. Since the parameter range is bounded, we perform in practice a truncated Newton step as proposed in (Nash, 1984).

This Gauss-Newton approach was inspired by the iterative re-weighted least squares algorithm (Bishop, 2006) developed for solving logistic regression (LR) problems. Indeed the first term of  $\mathcal{L}_S$  is similar to the log likelihood of LR after replacing  $u_n$  with a binary variable and linearizing the shape function. The proposed approach is also related to the Fisher scoring algorithm (see (Sourati et al., 2019) as an example in medical image analysis) when the point-wise Hessian matrix of the log likelihood is replaced by its expectation thus leading to more stable evaluation. In this particular case, the approximate Hessian is not the expectation of the Hessian since the first term of  $\mathcal{L}_S$  is the expectation of the log-prior with respect to binary variable  $U$  instead of  $Z$ .

Finally, the proposed algorithm also outputs a Laplace approximation of the shape parameter posterior  $p(\theta_S|I)$  as a Gaussian distribution where the mean is the optimized shape parameter  $\theta_S^*$  and the covariance is the inverse approximate Hessian matrix  $\Sigma_{\theta_S}^* = (\hat{H})^{-1}$ .

The overall optimization finally consists in iterating a series of outer loop, each loop consisting in optimizing the shape parameters as in Alg. 1 then followed by a series of MI-steps until the relative change of intensity parameters is less than a threshold. The stopping criterion for the outer loop is the relative change of foreground intensity parameters as it is the most impactful parameter.

### 2.5. Influence of the characteristic length $l_{\text{ref}}^k$

Based on Eq.2.5 and Eq.1, it is easy to see that for infinitely small value of the characteristic length  $l_{\text{ref}} \rightarrow 0$ , then the label prior becomes more and more sharp  $p(Z_n = 1|\theta_S) \rightarrow \delta_{\text{SDM}(\mathcal{S}(\theta_S), \mathbf{x}) > 0}$  and the label posterior becomes equal to the label posterior :  $p(Z_n = 1|\theta_S, \theta_I, I_n) \rightarrow p(Z_n = 1|\theta_S)$ . Conversely, for infinitely large value of the characteristic length  $l_{\text{ref}} \rightarrow \infty$ , the label prior becomes uninformative  $p(Z_n = 1|\theta_S) \rightarrow \frac{1}{2}$  and the label posterior converges towards the appearance driven label posterior :  $p(Z_n = 1|\theta_S, \theta_I, I_n) \rightarrow p(I_n|Z_n = 1, \theta_I^1)/(p(I_n|Z_n = 0, \theta_I^0) + p(I_n|Z_n = 1, \theta_I^1))$ . Therefore the characteristic length controls the relative influence of the shape and appearance information in the probability of assigning a label.

Since it is scaling the signed distance function,  $l_{\text{ref}}$  can be interpreted as controlling how far the resulting shape given by  $p(Z_n = e_1|\theta_S, \theta_I, I_n) = 0.5$  is allowed to deviate from the reference shape given by  $\mathcal{S}(\theta_S)$ . More precisely, assuming a uniform distribution of the appearance label probability between 0 and 1, one can compute the expectation of the posterior probability for a voxel located as a distance  $d_n$  from the

---

**Algorithm 1:** MS step to compute  $p(\theta_S|I)$ 


---

```

1  $i \leftarrow 0$ ;
2  $u_n = p(Z_n = 1|I, \theta_S)$ ; // E-Step, Update U
3 repeat
4    $V \leftarrow \frac{\tilde{S}(\theta_S^i, x_n)}{l_{\text{ref}}} \in \mathbb{R}^N$ ; // Shape function
5    $\mathbf{d} \leftarrow \frac{\nabla_{\theta_S} \tilde{S}(\theta_S^i, x_n)}{l_{\text{ref}}} \in \mathbb{R}^{|\theta_S| \times N}$ ; // Shape function gradient
6    $\delta\theta_S^0 \leftarrow \mathbf{0}, t \leftarrow 0$ ;
7   repeat
8      $\mu \leftarrow \sigma(V + \mathbf{d}^T \delta\theta_S^t)$ ; // Current Prior probability
9      $\mathbf{g} \leftarrow -\mathbf{d}(\mathbf{u} - \mu) - (\Sigma_{\theta_S}^0)^{-1} \delta\theta_S^t$ ; // Functional Gradient
10     $\tilde{H} \leftarrow \mathbf{d} \text{Diag}(\mu \circ (1 - \mu)) \mathbf{d}^T - (\Sigma_{\theta_S}^0)^{-1}$ ; // Approximate Functional Hessian
11     $\Sigma^* \leftarrow (\tilde{H})^{-1}$ ; // Covariance
12     $\delta\theta \leftarrow -\Sigma^* \mathbf{g}$ ; // Truncated Gauss Newton Update
13     $\delta\theta_S^{t+1} \leftarrow \delta\theta_S^t + \delta\theta, t \leftarrow t + 1$ ; // Update shape parameters
14  until  $\|\delta\theta\|/\|\theta\| < \epsilon$ ;
15   $u_n = p(Z_n = 1|I, \theta_S)$ ; // E-Step, Update U
16   $\theta_S^{i+1} \leftarrow \theta_S^i + \delta\theta_S^{t+1}, i \leftarrow i + 1$ ; // end inner loop
17 until  $\|\delta\theta_S^{t+1}\|/\|\theta_S^{t+1}\| < \epsilon$ ;
18  $\theta_S^* \leftarrow \theta_S^i, \Sigma_{\theta_S}^* = \Sigma^*$ ; // Gaussian posterior

```

---

reference shape :

$$\begin{aligned}
\mathbb{E}(p(Z_n = 1|\theta_S, \theta_l, I_n)) &= \int_0^1 \frac{tS(\Delta_n)}{tS(\Delta_n) + (1-t)(1-S(\Delta_n))} dt \\
&= \frac{1 - \Delta_n e^{-\Delta_n} - e^{-\Delta_n}}{(e^{-\Delta_n} - 1)^2}, \quad \Delta_n = \frac{d_n}{l_{\text{ref}}}
\end{aligned}$$

Based on the graph of Fig.1, a voxel located at least at  $4l_{\text{ref}}$  inside the boundary of the reference shape  $S(\theta_S)$  ( $p < -4$ ) will have in average at least 95% probability to be classified as belonging in the object.

### 3. Application to Cochlea Shape Recovery

#### 3.1. Cochlea shape model

We use a parametric cochlea shape model which is controlled by a set of 4 deformable shape parameters  $\theta_{SD} : \{a, b, \alpha, \varphi\}$  as shown in Fig:(2). Those 4 parameters control the deformation of the centerline of the cochlea represented as a generalized cylinder and is detailed in Appendix B. In addition to those 4 *deformable* parameters, we consider the 6 pose parameters  $\theta_{SR}$  consisting of rotation  $\{rx, ry, rz\}$  (parameterizing a rotation vector) and translation  $\{tx, ty, tz\}$  values. Therefore, the total number of shape parameters is 10, controlling the rigid and non rigid (deformable) motion:  $\theta_S = \theta_{SD} \cup \theta_{SR}$ .

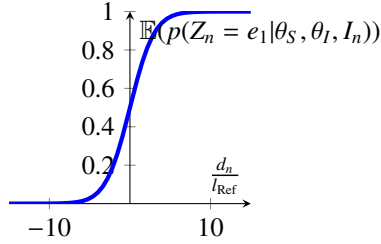


Figure 1: Expected label posterior probability as function of the normalized signed distance from the reference shape.

a b  $\alpha$   $\varphi$

Figure 2: Parametric shape model of the cochlea. (Left) Effect of the radial parameters  $a$  (red), and  $b$  (yellow) are shown with the reference position in purple; (Right) Effect of the longitudinal parameters  $\alpha$  (pink) and  $\varphi$  (blue) parameters.

To fit in our framework, signed distance map  $\text{SDM}(\mathcal{S}(\theta_S), \mathbf{x})$  from the cochlea triangular mesh surface must be created. This can be performed for instance by using VTK functions (Maurer et al., 2003) but that distance map generation may take several seconds on large volumetric images. This is why we have developed a convolution neural network, noted as DLSDM, which outputs an approximation of the signed distance map from the set of deformable shape parameters in few milliseconds on CPU (Wang et al., 2020a).

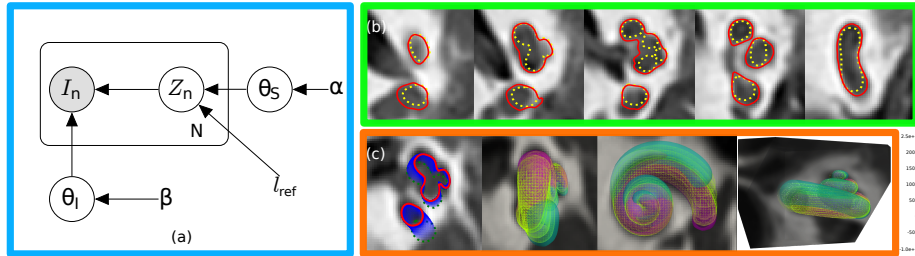


Figure 3: (a) graphical model for the shape-based generative model; (b) Cochlea segmentation on CT images is shown in solid red with the associated shape model in dashed yellow lines; (c) Evolution of the cochlea shape model during several MS steps shown as 2D contours (from dotted green to solid red) and 3D models.

### 3.2. Cochlea Appearance model

Appearance models describe the intensity patterns inside the foreground and the background classes and can be built in a supervised, semi-supervised or unsupervised

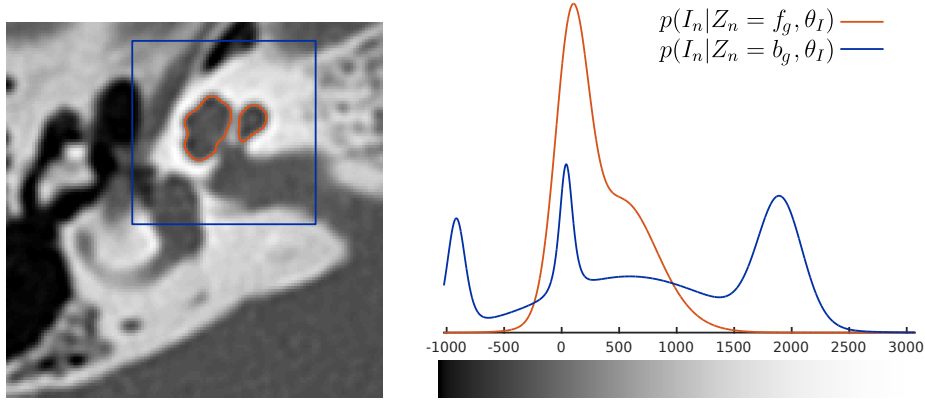


Figure 4: Example of intensity probability distributions of the foreground ( $f_g$ , in red) and the background ( $b_g$ , in blue) as functions of the Hounsfield unit.

manner. Many simple generative models such as Gaussian mixture models (GMM) with spatial corrections (Pohl et al., 2006a; Ashburner and Friston, 2005b) have been proposed in the literature to describe tissue intensity distributions. For the cochlea segmentation in CT images, we propose an unsupervised approach based on mixture of mixtures of Student's  $t$ -distributions, i.e. each background and foreground regions are described as mixtures of Student's  $t$ -distributions. Those  $t$ -distributions are generalized Gaussian distributions with heavy tails and lead to more robust estimations than GMM since they are less sensitive to extreme intensity values (Peel and McLachlan, 2000). In this context, the probability of observing intensity  $I_n$  knowing the label  $Z_n$  is parameterized as :

$$p(I_n|Z_n = k, \theta_I) = \sum_{m=1}^{M_k} \pi_m^k t(I_n|\mu_m^k, \sigma_m^k, \nu_m^k), \quad (5)$$

where  $M_k$  corresponds to the number of mixture components for the class  $k$  and mixture coefficients  $\pi_m^k$  are positive and sum to one  $\sum_{m=1}^{M_k} \pi_m^k = 1$ . The mean parameter  $\mu_i^k$ , standard deviation coefficient  $\sigma_i^k$  and degrees of freedom  $\nu_i^k$  are parameters of the Student's  $t$ -distribution defined as :

$$t(I_n|\mu, \sigma, \nu) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)} \frac{1}{\sqrt{\pi\nu}\sigma} \left(1 + \frac{(I_n - \mu)^2}{\sigma^2\nu}\right)^{-\left(\frac{\nu+1}{2}\right)}, \quad (6)$$

where  $\Gamma(\cdot)$  is the gamma function. To write the likelihood of this Student's  $t$ -distribution mixture of mixtures, we introduce a new categorical variable  $\tau_{nkm}$  which is a binary 1-of- $M_k$  encoding such that  $\tau_{nkm} = 1$  if voxel  $n$  belongs to the  $m$ -th component of region  $k$ , and  $\sum_{m=1}^{M_k} \tau_{nkm} = 1$ . The likelihood then writes as:

$$p(I_n|Z_n, \tau_n) = \prod_{k=0}^1 \prod_{m=1}^{M_k} \left[ \left( t(I_n|\mu_m^k, \sigma_m^k, \nu_m^k) \right)^{\tau_{nkm}} \right]^{Z_{nk}}$$

The inference is performed with closed-form updates of all parameters (Peel and McLachlan, 2000; Bishop, 2006) after writing the Student’s  $t$ -distribution as a Gaussian scale mixture. The total number of parameters to estimate is then  $|\theta_I| = 4(M_0 + M_1)$ . For the cochlea segmentation problem, we assume that the cochlea region mainly consists of two components ( $M_1 = 2$ ): the fluid (perilymph and endolymph) component centered around 0 HU and the bony walls centered around 500 HU. For the cochlea background, we consider 4 components ( $M_0 = 4$ ) centered around 0 HU (fluid), 2000 HU (bony labyrinth), -1000 HU (air due to pneumatization) and 600 HU (temporal bone). The corresponding initial distribution of intensity in the background and foreground regions are shown in Fig.4 and the exact initialization values are provided in Appendix C.

## 4. Results

### 4.1. Synthetic Images

We provide a 2D synthetic example to illustrate the influence of the reference length  $l_{\text{ref}}$  in the proposed segmentation algorithm. We consider the segmentation of an ellipse with Gaussian intensity distribution on both background and foreground (see Fig. 5 (Top Left)) by using a circle prior shape  $\tilde{S}(\theta_S, \mathbf{x}) = \|\mathbf{x} - \mathbf{C}\|^2 - R^2$ . It illustrates the frequent case where the parametric model used as a prior is far simpler than the shape visible in the image. The intensity model consists of two Gaussian distributions initialized with mean and variance offsets and the circle is parameterized by its center coordinates and radius. The trade-off between imaging information (leading to an ellipse) and prior shape (leading to a circle) is controlled by the  $l_{\text{ref}}$  parameter. The log-likelihood as a function of  $l_{\text{ref}}$  exhibits a single maximum for  $l_{\text{ref}} = l_{\text{opt}} = 0.021$  (Fig. 5 (Middle)) corresponding to the white circle in Fig. 5 (Left) and to the posterior label distribution in Fig. 5 (Right). The resulting segmentation is the isocontour  $p(Z_n = 1) = 0.5$ , displayed as a yellow curve in Fig. 5 (Left), which closely matches the elliptical shape except at its flat part (see arrow). This optimal value of  $l_{\text{ref}}$  corresponds to a configuration where the area of the circle is roughly equal to the area of the ellipse. A value of  $l_{\text{ref}} < l_{\text{opt}}$  leads to isocontours  $p(Z_n = e_1) = 0.5$  that fit more closely the ellipse whereas  $l_{\text{ref}} > l_{\text{opt}}$  leads to isocontours that look more like a circle.

### 4.2. Inner Ear Datasets

The evaluation of the proposed approach is studied on 3 different datasets.

*Dataset #1.* includes spiral CT temporal bone images of 210 patients from the radiology department of Nice University Hospital of size  $512 \times 512 \times 178$  corresponding to a voxel size of  $0.185\text{mm}, 0.185\text{mm}, 0.25\text{mm}$ . They have then been registered to a reference image via an automatic pyramidal blocking-matching (APBM) algorithm (Ourselin et al., 2000) from the software MedInria (Toussaint et al., 2007) followed by an image reformatting around the cochlea to the dimension  $(60, 50, 50)$  with isotropic voxel size of  $0.2\text{mm}$ . The relatively robust registration provides a rough alignment of the cochlea visible in the input image with a cochlea reference frame. From that dataset, 5 CT images were manually segmented by an ENT surgeon (see section 4.4).

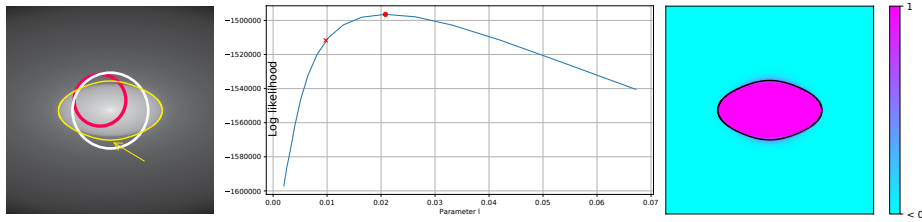


Figure 5: (Left) Input Ellipse image fitted with a circle shape : initial circle (red), final circle (white) and 0.5 isocontour of posterior label probability for optimal value of  $l_{\text{ref}}$  (yellow);(Middle) Log likelihood as a function of  $l_{\text{ref}}$  ; (Right) posterior label probability  $p(Z_n = 1 | \theta_S, \theta_I)$  for optimal value of  $l_{\text{ref}}$ ;

*Dataset #2.* includes 9 cadaveric cochlea spiral CT images acquired at the face and neck institute at Nice University Hospital with the same size and voxel spacing as dataset #1. In addition to CT images, high resolution X-ray microtomography (a.k.a.  $\mu\text{CT}$ ) images with dimension of (1035, 800, 1095) and isotropic voxel spacing of 0.02479mm were acquired each subject. The 9  $\mu\text{CT}$  and spiral CT images have been registered together as shown in Fig 6 and reformatted around the cochlea to the same physical size as for dataset #1 (i.e. 12mm, 10mm, 10mm). The cochlea and its two scala have been segmented on both CT and  $\mu\text{CT}$  images by an ENT surgeon with a semi-interactive tool (Criminisi et al., 2008). The high resolution  $\mu\text{CT}$  masks serve as ground truth information for the location of the cochlea.

*Dataset #3.* is a human bony labyrinth dataset (Wimmer et al., 2019) which includes 22 bony labyrinth CT images and their corresponding  $\mu\text{CT}$  images having isometric voxel size respectively of 0.1562mm and 0.0607mm. Those images were preprocessed and reformatted as for dataset #2 and also contains manually segmented cochlea masks.

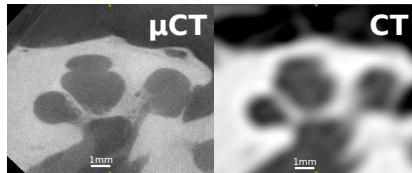


Figure 6: A visual comparison of imaging resolution between the  $\mu\text{CT}$  and conventional CT for cochlea imaging.

#### 4.3. Quantitative evaluation of segmentation on post-mortem $\mu\text{CT}/\text{CT}$ datasets #2 and #3

**Baseline Approach:** We have implemented a 3D atlas based segmentation approach and applied it on dataset #2 and #3 to get a baseline accuracy in terms of Dice score. To this end, we randomly select one image from each dataset as template image and for each input image we perform a multiscale demons deformable registration

Table 1: Computational efficiency proposed methods

	BFGS	VTK SDM	DLSDM
Mean Comput. Time	12h15min	43min	16min

Table 2: Performance metrics obtained on dataset #2 and #3.

Compared Labels		Dice Score		Symmetric Hausdorff Distance (voxel size 0.2 mm)			
				Dataset #2		Dataset #3	
		Dataset #2	Dataset #3	95%	100%	95%	100%
CT Manual	SSI	$0.74 \pm 0.02$	$0.77 \pm 0.023$	0.53	1.04	0.70	1.91
	SROI	$0.85 \pm 0.011$	$0.91 \pm 0.015$	0.34	0.82	0.36	1.68
$\mu$ CT Manual	SSI	$0.67 \pm 0.024$	$0.76 \pm 0.068$	0.68	1.48	0.67	1.96
	SROI	$0.81 \pm 0.04$	$0.91 \pm 0.019$	0.50	1.31	0.36	1.68
	CT Manual	$0.70 \pm 0.084$	$0.93 \pm 0.021$	0.50	1.34	0.19	0.74

(Vercauteren et al., 2007) (as implemented in SimpleITK 1.0.1) to estimate the deformation field. The segmented mask of the template is deformed to match the target image. The average Dice scores are 0.63 for dataset #2 and 0.68 for dataset #3.

**Logistic Shape Model Inference:** In all cases, the deformable shape parameters  $\theta_{SD}$  were initialized as ( $a = 4, b = 0.15, \alpha = 0.6, \varphi = 0.2$ ) and the pose parameters were set to zero. The  $l_{ref}$  value was set between 0.1 and 0.3 (see section 4.3) and the stopping condition is  $\frac{\Delta\theta}{\theta} < 0.1$ , thus stopping when parameter updates are less than 10% of the parameter values.

**Computational efficiency:** We analyze the computational cost of several alternative formulations of our algorithm. More precisely, in Table 1 we compare the computational time of three different implementations of our approach that differ by the choice of the quasi-Newton optimization method in the MS-step (BFGS vs Gauss-Newton) and by the algorithm used for generating signed distance maps (VTK based vs deep learning based). The various algorithms was applied on the 9 images of dataset #2 and ran on a Dell Precision 7520 computer. It is clear that the Gauss-Newton method described in Algorithm 1 is far more efficient since it uses a much better approximation of the Hessian matrix than in the generic quasi Newton approach. Furthermore, as expected, the trained deep learning method leads to a speedup factor greater than 3.

**Influence of the reference length** To assess the influence of the hyper parameter reference length:  $l_{ref}$ , we analyse the variation of the final Dice score for various reference lengths based on one image of dataset #2. The results are shown in Table 3. We see that the reference length within the range  $[0.05mm, 0.25mm]$  has a relatively small influence on the Dice score. To minimize the time of computation, we do not optimize the reference length through a greedy search but simply set its value to 0.1 for dataset #1 and #2 and 0.3 for dataset #3 for shape fitting. To compute the final hard segmentation we use a fixed reference length of 0.25.

**Robustness analysis** To study the robustness of the method, we randomly initialize the cochlea shape parameters by performing a random uniform sampling within their defined value range. Based on 10 initial random samples, we computed the average

Table 3: Influence of the hyper parameter:  $l_{ref}$  for the segmentation accuracy.

<b>Ref. Length</b>	0.05	0.1	0.15	0.2	0.25
<b>Dice Score</b>	0.84	0.85	0.85	0.82	0.84

Dice score for one image of dataset #2 and obtained a mean Dice score of  $0.81 \pm 0.1$  ( respectively of  $0.68 \pm 0.23$  ) for the Gauss-Newton method (resp. the BFGS method). This clearly shows the increased robustness with respect to initial shape values obtained by the Gauss-Newton optimization of the MS-step.

**Evaluation on CT and  $\mu$ CT images:** Datasets #2 and #3 include both CT and  $\mu$ CT images of the same subject that have been registered to each other. Furthermore the cochlea was manually or semi-automatically segmented by an expert on both modalities such that we can use those two binary maps to evaluate the accuracy of the algorithm applied on the CT image. The cochlea binary map from high resolution  $\mu$ CT images have been downsampled and represent a more reliable ground truth than the manual segmentation performed on the CT images.

The proposed algorithm using Gauss-Newton optimization and deep-learning generation of signed distance maps was applied on the 9+22 CT images of the two datasets. Fig. 3 (Right) shows the segmented cochlea in red, the associated shape model, and its evolution during the MS step. Clearly, we see that the resulting segmentation is strongly constrained by the shape model.

In Table 2, we provide two metrics between pairs of binary masks : the Dice score and the 95% and 100% symmetric Hausdorff distance (HD) (computed as the average of two distances). Furthermore, we compare the segmentations produced by the posterior label probability (SROI for  $p(Z_n|I_n, \theta_S, \theta_I) = 0.5$ ) and the ones produced by the shape model only (SSI for  $p(Z_n|\theta_S) = 0.5$ ) with both manual segmentations obtained on CT and  $\mu$ CT images. To measure the uncertainty in the manual CT segmentation, we also evaluate the metrics between both CT and  $\mu$ CT manual mask images.

The logistic shape model framework produces good segmentation results on both datasets (Dice scores of 0.81 and 0.91) and even slightly outperforms the manual CT segmentation on dataset#2 (0.81 vs 0.7) which is far more challenging dataset #3. The segmented shape instances produced by the shape model are not as accurate as the SROI for the cochlea segmentation (lower Dice score and larger HD). This confirms that the parametric geometric cochlea model is a simplified representation of the cochlea anatomy. Finally, the metrics between the 2 manual segmentations on dataset #2 (DSC of 0.7 with a 95% HD of 0.5mm) shows the difficulty of performing a manual segmentation of the cochlea due to its limited size and contrast.

#### 4.4. Semi-quantitative analysis of segmentation on clinical dataset #1

We ran the segmentation framework (with DLSDM) on the 210 CT of dataset #1 on a Dell 6145 and 6420 CPU clusters.

**Unsupervised quality control and semi-quantitative evaluation** As manual segmentations of the 210 images are not available, we propose instead an original approach to estimate our algorithm’s performance while minimizing the manual annotation effort. First, we apply the unsupervised quality control algorithm of Audelan *et al.* (Au-



delan and Delingette, 2019) on the whole dataset in order to sort the 210 segmentations according to their hypothesized performance. More precisely, this quality control algorithm computes for each image segmentation, an average distance between the segmentation provided by our algorithm and a segmentation produced by a simple generic probabilistic method. We can then generate an histogram of such average surface error (ASE) in Fig. 7. Segmentations having a low ASE correspond to those having good intensity contrast across their boundaries while those on the right tail of the distributions are considered as more challenging and suspect of including segmentation errors.

The histogram exhibits a bell shape with few outliers on its right and left tails. Furthermore, we have manually checked that this unsupervised quality control algorithms worked well on this dataset with visually better segmentations localized on the left side of the histogram. To estimate the relation between the ASE and the Dice score, we picked 5 images in order to sample the histogram at different levels of ASE corresponding to images #213, #210, #53, #264 and #143 (see Fig.7) in ascending order of ASE. Those 5 CT images were manually segmented by an ENT surgeon and the Dice scores of the segmentation produced by our algorithm was reported in Table 4. We see that the Dice score decreases as the ASE increases which indicates that the ASE may be a proper surrogate for the segmentation performance. The cochlea in image #143 was indeed found to be an outlier in terms of shape probably due to a patient malformation. Inspired by (Audelan and Delingette, 2019), we can make the hypothesis that ASE a good proxy for the Dice score as there is a monotonic relation between ASE and Dice. On this basis, we can extrapolate that the median Dice score over the whole dataset is probably above 0.82. Yet, a more thorough study with far more manual segmentations is necessary to be less speculative about the actual performance on clinical CT data.

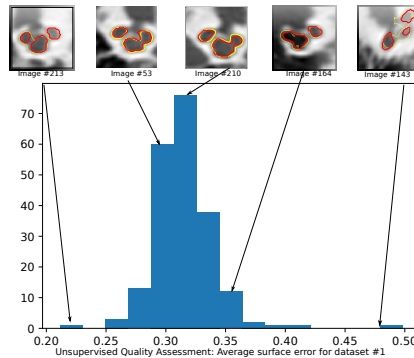


Figure 7: Average surface error of segmentations generated from dataset #1 resulting from the unsupervised quality control. Red contours correspond to the manual ground truth while yellow ones are segmentation outputs.

**Parameter analysis** The application of the algorithm on dataset #1 resulted in the estimation of  $10 \times 210$  shape parameters with 210 covariance matrices  $\Sigma_{\theta_s}^*$ . In Fig. 8(a) the histograms of the 4 deformable shape parameters are displayed in green. Interestingly, the  $a$  and  $\alpha$  parameters exhibit a bimodal distribution for which a simple explanation may be provided. Indeed, the left highest mode is probably correspond-

Table 4: Dice score for selected segmentation samples from dataset #1 based on the histogram of Fig.7. The ASE are got from automatic quality control algorithm and the DICE score are computed based on manual segmentation.

Patient ID	213	210	53	164	143
DICE Score	0.84	0.84	0.84	0.76	0.45
ASE	0.21	0.30	0.32	0.36	0.50

ing to straight centerline profiles whereas the rightmost mode may be associated with the "rollercoaster" longitudinal profiles (Avci et al., 2014). In Fig. 8(b) the average  $10 \times 10$  covariance computed as the log-Euclidean mean (Arsigny et al., 2007) is displayed showing potential correlations between pose and deformable parameters. Shape parameter  $b$  corresponding the exponent of the logarithmic center line curve has a particularly low variance such that it can be well estimated from the data. Conversely the phase parameter  $\varphi$  has much greater variance and is harder to estimate in average. The extraction of the eigenvectors of that covariance matrix confirms that most parameters are independent with other except for parameter  $a$  which is a bit correlated with the  $\varphi$  parameter and the translation term  $ty$ . The relative independence of the parameters for shape fitting indicates that we do not have an overparametrization of the cochlea shape.

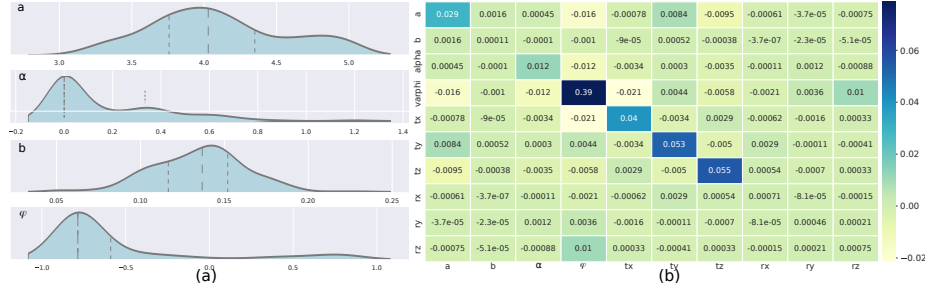


Figure 8: Distribution plots for shape parameters variance. (b) Average covariance matrix of the 10 shape parameters.

**Uncertainty Segmentation analysis** The estimated covariance matrix  $\Sigma_{\theta_S}^*$  can be used for studying the uncertainty of the output segmentation. We sampled 100 times the multivariate Gaussian approximate posterior distribution of the parameters  $p(\theta_S|I) \approx \mathcal{N}(\theta_S^*; \Sigma^*)$  and generated accordingly 100 random posterior labels  $p(Z|I, \theta_S, \theta_I)$  that are then averaged to estimate with Monte-Carlo sampling the marginal posterior  $p(Z|I, \theta_I) = \int_{\mathbb{R}^{\theta_S}} p(Z|I, \theta_S, \theta_I) p(\theta_S|I) d\theta_S$ . In Fig 9 we show several slices of the resulting probability maps with the 0.5 level curve together with the posterior probability  $p(Z|I, \theta_S^*, \theta_I)$  obtained with the most likely shape parameter  $\theta_S^*$ . We see that we have a much larger uncertainty in the resulting segmentation when accounting for the uncertainty in the shape parameters than without them. This is a far better approximation of the true uncertainty  $p(Z|I)$  than the posterior label probability  $p(Z|I, \theta_S, \theta_I)$ .

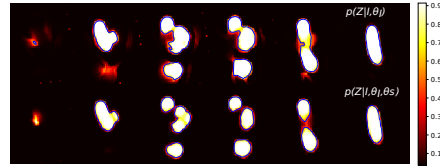


Figure 9: Marginal posterior probability  $p(Z|I, \theta_I)$  (Top) versus posterior probability  $p(Z|I, \theta_I, \theta_S)$  (Bottom) computed on patient #1 of dataset #1.

#### 4.5. Comparison with the state-of-the-art

We consider below the prior work on cochlea segmentation evaluated on clinical CT images while discarding the literature on the segmentation of  $\mu$ CT images (Kjer et al., 2014; Ruiz Pujadas et al., 2016; Pujadas et al., 2016) or of the scala tympani and vestibuli located inside the cochlea (Noble et al., 2012, 2013). Table 5 summarises the relevant publications on cochlea segmentation that are split into unsupervised and supervised methods. The former approaches are mostly based on cochlear shape fitting based on template image registration (Baker and Barnes, 2005), parametric shape model (Baker, 2008). The supervised methods are based on statistical deformation models (Ruiz Pujadas et al., 2018) and deep learning (Lv et al., 2021; Raabid et al., 2021; Heutink et al., 2020).

Quantitative comparison of performances is not straightforward due to differences in image modality (CT,  $\mu$ CT or ultra high resolution CT), in metrics (Dice, precision, mean surface error), in subject population (cadaveric vs patient) but also in the target anatomical structures (cochlea vs cochlea labyrinth). In most cases, cochlea segmentation from  $\mu$ CT images are used as ground truth information and a direct comparison between our work with (Raabid et al., 2021) is possible since they used a subset of dataset #3 which is a public database (Wimmer et al., 2019). We see that our unsupervised approach performs as well as the supervised methods with Dice scores in the range [0.85, 0.91] and outperforms previous unsupervised methods.

Table 5: Performances of prior work on cochlea segmentation. *NL* (resp. *NT*) indicates the number of training (resp. testing) images. *Unsup* (resp. *Sup*) refers to unsupervised (resp. supervised) learning methods.

Method Group	Study	Comparison	Metrics	Proposed method (Dataset#2 N=9)	Proposed method (Dataset#3 N=22)
UnSup	Baker (2008) (NT= 4)	CT	Precision	0.72 ± 0.09	<b>0.75</b> ± <b>0.03</b>
UnSup	Kjer and Paulsen (2015) (NT = 2)	post-mortem $\mu$ CT	Mean (±1 std) surface error	0.22 ±0.17	<b>0.11</b> ± <b>0.06</b>
Sup	Kjer et al. (2017) (NL = 18/NT = 14)	post-mortem CT	Dice	0.88	
Sup	Heutink et al. (2020) (NL=48/NT = 75)	Ultra-high Resolution CT	Dice	0.90 ± 0.03	
Sup	Lv et al. (2021) (NL=24/NT=6)	Cochlea Labyrinth	Dice	0.90	
Sup	Raabid et al. (2021) (NL + NT = 17)	post-mortem CT	Dice	0.85 ±0.011	<b>0.91</b> ± <b>0.03</b>

## 5. Discussion

The proposed approach relies on the definition of a generic shape function  $\tilde{S}(\theta_S, \mathbf{x})$  which can be for instance a statistical shape model, a deformation image template, or an implicit shape equation. In the case of the cochlea, it was defined as a signed distance function of a parametric shape model  $\text{SDM}(S(\theta_S), \mathbf{x})$ . This specific choice makes the computation of the shape function and its gradient fairly costly, despite the use of a fully supervised dedicated neural network (DLSDM). There are several ways to optimize its computation time. One could for instance use a supervised appearance model such as a trained neural network which would remove all MI steps in the EM algorithm and would decrease by at least a factor 2 the time of computation. Another way is to use an implicit shape model  $S(\theta_S, \mathbf{x}) = 0$  for instance based on statistical level sets (Tsai et al., 2003). The cochlea segmentation example provided in this paper relies on a fully interpretable intensity and shape parameters at the expense of its computational efficiency. Yet, one could train a deep neural regressor for predicting cochlea shape parameters and segmentation by using the segmentations generated by the proposed framework as training set.

For the cochlea segmentation, excellent results were obtained on cadaveric CT images similarly to the supervised methods. Furthermore, to the best of our knowledge, we introduced a first semi-quantitative assessment of cochlea segmentations on clinical CT images acquired on more than 200 patients. However, for a complete study, one would need to assess thoroughly the inter-rater variability of those manual segmentations and ideally combine them with other high resolution image modalities. Finally, an interesting extension of this work would be to segment the scala vestibuli and tympani in addition to the cochlea.

## 6. Conclusion

In this paper, we have presented a new probabilistic generative approach for combining shape and intensity models for image segmentation. The resulting segmentation is an interpretable compromise between a fidelity to a parametric shape space (captured in the prior distribution) and an appearance model (captured in the likelihood distribution). The proposed method goes well beyond the concept of shape fitting since it also provides an approximation to the posterior distribution of shape parameters. The use of a logistic shape model allows to control the trade-off between appearance and shape with a single parameter: the reference length. When applied to the recovery of cochlea structures from CT images, we were able to provide accurate segmentations with meaningful shape parameter distributions. Furthermore, we have shown how the approximate shape parameter posterior distribution can be exploited to provide realistic uncertainty maps.

An interesting application of the proposed approach is to perform model selection with Bayes factors, in order to estimate the optimal complexity of a parametric shape model for a given image segmentation task. Future work will also explore the application of this framework to other shape representations than explicit parametric shape models in order to find a reasonable trade-off between computational efficiency

and interpretability of shape parameters. For instance, in statistical deformation models (Rueckert et al., 2003a), the computation of shape function gradient  $\nabla_{\theta_S} \tilde{S}(\theta_S, \mathbf{x})$  is straightforward, but its shape parameters may not be meaningful besides the first modes.

## Appendix A. Gradient of shape function

In this section, we detail the computation of the shape function gradient  $\nabla_{\theta_S} \tilde{S}(\theta_S, \mathbf{x})$  when rigid and deformable shape parameters are considered. More precisely, writing the parameters controlling the non-rigid deformation as  $\theta_{SD}$ , the shape function writes as  $\tilde{S}(\theta_{SD}, \mathbf{R}\mathbf{x}_n + \mathbf{t})$ . The rotation matrix  $\mathbf{R}$  is parameterized with rotation vector  $\mathbf{r}$ , whose norm is the rotation angle and whose direction is the rotation axis. The gradients with respect to the translation and rotation vectors are then given in closed form as :

$$\begin{aligned} \nabla_{\mathbf{t}} \tilde{S}(\theta_{SD}, \mathbf{R}\mathbf{x}_n + \mathbf{t}) &= \nabla_{\mathbf{x}} \tilde{S}(\theta_{SD}, \mathbf{R}\mathbf{x}_n + \mathbf{t}) \\ \nabla_{\mathbf{r}} \tilde{S}(\theta_{SD}, \mathbf{R}\mathbf{x}_n + \mathbf{t}) &= \left( -\mathbf{R}S_{\mathbf{x}_n} \frac{\mathbf{r}\mathbf{r}^T + ((\mathbf{R})^T - \mathbf{I}_3)S_{\mathbf{r}}}{\|\mathbf{r}\|^2} \right)^T \\ &\quad \nabla_{\mathbf{x}} \tilde{S}(\theta_{SD}, \mathbf{R}\mathbf{x}_n + \mathbf{t}) \end{aligned}$$

where  $\nabla_{\mathbf{x}}$  is the spatial gradient,  $S_{\mathbf{x}}$  is the 3x3 anti-symmetric matrix associated with vector  $\mathbf{x}$ . For a deformable parameter  $\theta_{SD}$ , if the shape function is not given in an analytical form as it is the case for parametric shapes, the shape function gradient can be computed with finite differences based on a parameter increment  $\delta\theta_{SD}^i$  :

$$\begin{aligned} \nabla_{\theta_{SD}^i} \tilde{S}(\theta_{SD}, \mathbf{R}\mathbf{x}_n + \mathbf{t}) &= \frac{1}{2\delta\theta_{SD}^i} \left( \tilde{S}(\theta_{SD} + \delta\theta_{SD}^i, \mathbf{R}\mathbf{x}_n + \mathbf{t}) \right. \\ &\quad \left. - \tilde{S}(\theta_{SD} - \delta\theta_{SD}^i, \mathbf{R}\mathbf{x}_n + \mathbf{t}) \right) \end{aligned}$$

## Appendix B. Cochlea Shape Model

We are interested in the cochlea structure in CT images which is defined as a generalized cylinder, i.e. as cross-sections swept along a centerline.

**Centerline** The centerline is parameterized in a cylindrical coordinate system by its *radial*  $r(\theta_c)$  and *longitudinal*  $z(\theta_c)$  components. The range of polar angle  $\theta_c$  is  $[0, \theta_{\max}]$  where  $\theta_{\max}$  is the maximum polar angle controlling the total number of cochlear turns.

The radial component is defined piecewise with a polynomial function and a logarithmic function of the angular coordinate  $\theta_c$  in the cylindrical coordinate system as:

$$r(\theta_c) = \begin{cases} p_2\theta_c^2 + p_1\theta_c + p_0 & \text{if } \theta_c < \theta_0 \\ ae^{-b\theta_c} & \text{if } \theta_c \geq \theta_0 \end{cases} \quad (\text{B.1})$$

where  $\theta_0 = 5\pi/6$  and  $p_0 = 5$  mm. Furthermore to obtain a continuously differentiable curve, we set :

$$\begin{aligned} p_2 &= \frac{C_1\theta_0 - C_2 + p_0}{\theta_0^2} & p_1 &= \frac{-C_1\theta_0 + 2C_2 + 2p_0}{\theta_0} \\ C_2 &= ae^{-b\theta_0} & C_1 &= -C_2b. \end{aligned} \quad (\text{B.2})$$

The longitudinal component of the centerline is the sum of an exponentially damped sinusoidal and a linear function:

$$z(\theta_c) = \begin{cases} \alpha e^{-\beta\theta_c} \cos(\theta_c + \phi) + q_1\theta_c & \text{if } \theta_c < \theta_1 \\ a_2\theta_c^2 + a_1\theta_c + a_0 & \text{if } \theta_c \geq \theta_1 \end{cases}, \quad (\text{B.3})$$

where  $\beta = 0.2 \text{ rad}^{-1}$ ,  $q_1 = 0.225 \text{ mm}\cdot\text{rad}^{-1}$  and  $\theta_1 = \theta_{\max} - \pi$ . The polynomial function is used to flatten out the last half turn so that  $dz(\theta)/d\theta|_{\theta=\theta_{\max}} = 0$  and similarly  $a_2, a_1, a_0$  are set to obtain a continuously differentiable curve.

**Cross-Sections** The cross-sections are modeled by a closed *planar shape* on which a varying *affine transformation* is applied along the centerline. The scala tympani and the scala vestibuli are modeled with two half pseudo-cardioids while the cochlear cross-section corresponds to the minimal circumscribed ellipse of the union of the tympanic and vestibular cross-sections. The affine transform of cross-sections is parameterized by a *rotation*, and a *width* and *height scalings*. All cross-sectional parameters are fixed because their variability was found to be small compared to the variability of the centerline.

**Shape parameter vector** We have chosen a compact description of the cochlea shape to limit as much as possible the correlation between the shape parameters and therefore make them uniquely identifiable. Finally, only 10 free parameters are considered in  $\theta_S$  :

- 6 translation and rotation parameters :  $\mathbf{t} = (tx, ty, tz)$ ,  $\mathbf{r} = (rx, ry, rz)$
- 2 radial component parameters of the centerline,  $a$  and  $b$
- 2 longitudinal component parameters,  $\alpha$  and  $\phi$

Note that there are no free cross-section parameters which implies that  $\theta_S$  can be used to define uniquely the cochlea.

The prior probabilities on the 10 shape parameters were modeled as being an uniform distribution (uninformative prior) such that all regularization terms  $\log p(\theta_S|\alpha)$  can be ignored.

### Appendix C. Initialization of intensity parameters

The  $4*6 = 24$  initial intensity parameters for the mixture of Student's  $t$  distributions in datasets #1 and #3 are presented in Table ??.

### References

- Agn, M., af Rosenschöld, P.M., Puonti, O., Lundemann, M.J., Mancini, L., Papadaki, A., Thust, S., Ashburner, J., Law, I., Leemput, K.V., 2019. A modality-adaptive method for segmenting brain tumors and organs-at-risk in radiation therapy planning. *Medical Image Analysis* 54, 220 – 237. URL: <http://www.sciencedirect.com/science/article/pii/S1361841518305103>, doi:<https://doi.org/10.1016/j.media.2019.03.005>.

- Alshazly, H., Linse, C., Barth, E., Martinetz, T., 2019. Ensembles of deep learning models and transfer learning for ear recognition. *Sensors* 19. URL: <https://www.mdpi.com/1424-8220/19/19/4139>, doi:10.3390/s19194139.
- Arsigny, V., Fillard, P., Pennec, X., Ayache, N., 2007. Geometric means in a novel vector space structure on symmetric positive-definite matrices. *SIAM Journal on Matrix Analysis and Applications* 29, 328–347. URL: [http://www-sop.inria.fr/asclepios/Publications/Vincent.Arsigny/Arsigny\\_SIAM\\_tensors\\_07.pdf](http://www-sop.inria.fr/asclepios/Publications/Vincent.Arsigny/Arsigny_SIAM_tensors_07.pdf), doi:10.1137/050637996.
- Ashburner, J., Friston, K.J., 2005a. Unified segmentation. *NeuroImage* 26, 839 – 851. URL: <http://www.sciencedirect.com/science/article/pii/S1053811905001102>, doi:<https://doi.org/10.1016/j.neuroimage.2005.02.018>.
- Ashburner, J., Friston, K.J., 2005b. Unified segmentation. *NeuroImage* 26, 839–851. doi:10.1016/j.neuroimage.2005.02.018.
- Audelan, B., Delingette, H., 2019. Unsupervised Quality Control of Image Segmentation based on Bayesian Learning, in: *MICCAI 2019 - 22nd International Conference on Medical Image Computing and Computer Assisted Intervention*, Shenzhen, China.
- Avcı, E., Nauwelaers, T., Lenarz, T., Hamacher, V., Kral, A., 2014. Variations in microanatomy of the human cochlea. *The Journal of Comparative Neurology* 00, 1–17. URL: <http://www.ncbi.nlm.nih.gov/pubmed/24668424>, doi:10.1002/cne.23594.
- Baker, G., 2008. Tracking, modelling and registration of anatomical objects: the human cochlea. Ph.D. thesis. The University of Melbourne. URL: <http://minerva-access.unimelb.edu.au/handle/11343/37464>.
- Baker, G., Barnes, N., 2005. Model-image registration of parametric shape models: fitting a shell to the cochlea. *Insight Journal* URL: [http://users.cecs.anu.edu.au/~nmb/papers/ij2005\\_{\\_}mireg.pdf](http://users.cecs.anu.edu.au/~nmb/papers/ij2005_{_}mireg.pdf).
- Bishop, C.M., 2006. *Pattern Recognition and Machine Learning (Information Science and Statistics)*. Springer-Verlag, Berlin, Heidelberg.
- Chan, T., Zhu, W., 2005. Level set based shape prior segmentation, in: *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, pp. 1164–1170 vol. 2. doi:10.1109/CVPR.2005.212.
- Cohen, L.T., Xu, J., Xu, S.A., Clark, G.M., 1996. Improved and simplified methods for specifying positions of the electrode bands of a cochlear implant array. *The American Journal of Otology* 17, 859–865. URL: <http://cat.inist.fr/?aModele=afficheN&cpsidt=2488723>.



- Commowick, O., Warfield, S., 2009. A continuous staple for scalar, vector, and tensor images: An application to dti analysis. *IEEE transactions on medical imaging* 28, 838–46. doi:10.1109/TMI.2008.2010438.
- Cootes, T.F., Taylor, C.J., Cooper, D.H., Graham, J., 1995. Active Shape Models-Their Training and Application. *Computer Vision and Image Understanding* 61, 38–59. doi:10.1006/cviu.1995.1004.
- Cremer, D., 2003. A variational framework for image segmentation combining motion estimation and shape regularization, in: *CVPR (1)*, IEEE Computer Society. pp. 53–58.
- Cremer, D., Kohlberger, T., Schnörr, C., 2003. Shape Statistics in Kernel Space for Variational Image Segmentation. *Pattern Recognition* 36, 1929–1943.
- Criminisi, A., Sharp, T., Blake, A., 2008. GeoS: Geodesic Image Segmentation. *ECCV*, 99–112 URL: [http://link.springer.com/content/pdf/10.1007/978-3-540-88682-2\\_{\\_}9.pdf](http://link.springer.com/content/pdf/10.1007/978-3-540-88682-2_{_}9.pdf).
- Demarcy, T., 2017. Segmentation and study of anatomical variability of the cochlea from medical images. Theses. Université Côte d’Azur.
- Elhabian, S., Whitaker, R., 2017. Shapeodds: Variational bayesian learning of generative shape models, in: *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2185–2196. URL: [doi.ieeecomputersociety.org/10.1109/CVPR.2017.235](https://doi.ieeecomputersociety.org/10.1109/CVPR.2017.235), doi:10.1109/CVPR.2017.235.
- Heimann, T., Münzing, S., Meinzer, H.P., Wolf, I., 2007. A shape-guided deformable model with evolutionary algorithm initialization for 3D soft tissue segmentation. *Inf Process Med Imaging* 20, 1–12.
- Heutink, F., Koch, V., Verbist, B., van der Woude, W.J., Mylanus, E., Huinck, W., Sechopoulos, I., Caballo, M., 2020. Multi-scale deep learning framework for cochlea localization, segmentation and analysis on clinical ultra-high-resolution ct images. *Computer Methods and Programs in Biomedicine* 191, 105387. URL: <https://www.sciencedirect.com/science/article/pii/S0169260719320231>, doi:<https://doi.org/10.1016/j.cmpb.2020.105387>.
- Kjer, H., Fagertun, J., Wimmer, W., Gerber, N., Vera, S., Barazzetti, L., Lopez, N., Ceresa, M., Piella, G., Stark, T., Stauber, M., Reyes, M., Weber, S., Caversaccio, M., Ballester, M., Paulsen, R., 2017. Patient-specific estimation of detailed cochlear shape from clinical ct images. *International Journal of Computer Assisted Radiology and Surgery* 13, 389–396.
- Kjer, H.M., Paulsen, R.R., 2015. Modelling of the Human Inner Ear Anatomy and Variability for Cochlear Implant Applications. Ph.D. thesis. Technical University of Denmark (DTU).

- Kjer, H.M., Vera, S., Pérez, F., González Ballester, M.A., Paulsen, R.R., 2014. Semi-automatic anatomical measurements on microCT 3D surface models, in: International Conference on Cochlear Implants and Other Implantable Auditory Technologies, Munich, Germany, p. 711.
- Le Folgoc, L., Delingette, H., Criminisi, A., Ayache, N., 2017. Quantifying registration uncertainty with sparse bayesian modelling. *IEEE Transactions on Medical Imaging* 36, 607–617. doi:10.1109/TMI.2016.2623608.
- Li, H., Prasad, R.G.N., Sekuboyina, A., Niu, C., Bai, S., Hemmert, W., Menze, B., 2021. Micro-ct synthesis and inner ear super resolution via generative adversarial networks and bayesian inference. *arXiv:2010.14105*.
- Lv, Y., Ke, J., Xu, Y., Shen, Y., Wang, J., Wang, J., 2021. Automatic segmentation of temporal bone structures from clinical conventional ct using a cnn approach. *The International Journal of Medical Robotics and Computer Assisted Surgery* 17, e2229. URL: <https://onlinelibrary.wiley.com/doi/abs/10.1002/rcs.2229>, doi:<https://doi.org/10.1002/rcs.2229>, arXiv:<https://onlinelibrary.wiley.com/doi/pdf/10.1002/rcs.2229>.
- Maurer, C.R., Qi, R., Raghavan, V., Member, S., 2003. A linear time algorithm for computing exact euclidean distance transforms of binary images in arbitrary dimensions. *IEEE Transactions on Pattern Analysis and Machine Intelligence* , 265–270.
- Nash, S.G., 1984. Newton-type minimization via the lanczos method. *SIAM Journal on Numerical Analysis* 21, 770–788. URL: <http://www.jstor.org/stable/2157008>.
- Noble, J.H., Gifford, R.H., Labadie, R.F., Dawant, B.M., 2012. Statistical shape model segmentation and frequency mapping of cochlear implant stimulation targets in CT. *Medical Image Computing and Computer-Assisted Intervention* 15, 421–8. URL: <http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3559125&tool=pmcentrez&rendertype=abstract>.
- Noble, J.H., Labadie, R.F., Gifford, R.H., Dawant, B.M., 2013. Image-Guidance enables new methods for customizing cochlear implant stimulation strategies. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 21, 820–829.
- Ourselin, S., Roche, A., Prima, S., Ayache, N., 2000. Block Matching : A General Framework to Improve Robustness of Rigid Registration of Medical Images. *Medical Image Computing and Computer-Assisted Intervention* , 557–566.
- Peel, D., McLachlan, G.J., 2000. Robust mixture modelling using the t distribution. *Statistics and Computing* 10, 339–348. URL: <https://doi.org/10.1023/A:1008981510081>, doi:10.1023/A:1008981510081.
- Pohl, K.M., Fisher, J., Grimson, W.E.L., Kikinis, R., Wells, W.M., 2006a. A Bayesian model for joint segmentation and registration. *NeuroImage* 31, 228–239. doi:10.1016/j.neuroimage.2005.11.044.

- Pohl, K.M., Fisher, J., Shenton, M.E., Mccarley, R.W., Grimson, W.E.L., Kikinis, R., Wells, W.M., Eric, W.L., 2006b. Logarithm Odds Maps for Shape Representation. *Medical image computing and computer-assisted intervention : MICCAI ... International Conference on Medical Image Computing and Computer-Assisted Intervention* 9, 955–963. URL: <http://www.ncbi.nlm.nih.gov/pmc/articles/PMC2994060/>.
- Prevost, R., Cuingnet, R., Mory, B., Cohen, L.D., Ardon, R., 2013. Incorporating shape variability in image segmentation via implicit template deformation, in: Mori, K., Sakuma, I., Sato, Y., Barillot, C., Navab, N. (Eds.), *Medical Image Computing and Computer-Assisted Intervention – MICCAI 2013*, Springer Berlin Heidelberg, Berlin, Heidelberg. pp. 82–89.
- Pujadas, E.R., Kjer, H.M., Vera, S., Ceresa, M., Ángel González Ballester, M., 2016. Cochlea segmentation using iterated random walks with shape prior, in: Styner, M.A., Angelini, E.D. (Eds.), *Medical Imaging 2016: Image Processing*, International Society for Optics and Photonics. SPIE. pp. 778 – 786. URL: <https://doi.org/10.1117/12.2208675>, doi:10.1117/12.2208675.
- Raabid, H., Alain, L., Berihu, G.K., Guigou, C., 2021. Automatic segmentation of inner ear on ct-scan using auto-context convolutional neural network. *Scientific Reports* 1, 839 – 851. doi:<https://doi.org/10.1038/s41598-021-83955-x>.
- Rueckert, D., Frangi, A., Schnabel, J., 2003a. Automatic construction of 3-d statistical deformation models of the brain using nonrigid registration. *IEEE Transactions on Medical Imaging* 22, 1014–1025. doi:10.1109/TMI.2003.815865.
- Rueckert, D., Frangi, A.F., Schnabel, J.A., 2003b. Automatic construction of 3-d statistical deformation models of the brain using nonrigid registration. *IEEE Transactions on Medical Imaging* 22, 1014–1025. doi:10.1109/TMI.2003.815865.
- Ruiz Pujadas, E., Kjer, H.M., Piella, G., Ceresa, M., González Ballester, M.A., 2016. Random walks with shape prior for cochlea segmentation in ex vivo  $\mu$ CT. *International Journal of Computer Assisted Radiology and Surgery* 11, 1647–1659. doi:10.1007/s11548-016-1365-8.
- Ruiz Pujadas, E., Piella, G., Kjer, H.M., González Ballester, M.A., 2018. Random walks with statistical shape prior for cochlea and inner ear segmentation in micro-ct images. *Machine Vision and Applications* 29, 405–414. URL: [https://app.dimensions.ai/details/publication/pub.1093061695andhttps://backend.orbit.dtu.dk/ws/files/140598803/10.1007\\_2Fs00138\\_017\\_0891\\_x.pdf](https://app.dimensions.ai/details/publication/pub.1093061695andhttps://backend.orbit.dtu.dk/ws/files/140598803/10.1007_2Fs00138_017_0891_x.pdf), doi:10.1007/s00138-017-0891-x.
- Sabuncu, M., Yeo, B.T., Van Leemput, K., Fischl, B., Golland, P., 2010. A generative model for image segmentation based on label fusion. *IEEE transactions on medical imaging* 29, 1714–29. doi:10.1109/TMI.2010.2050897.
- Simpson, I.J., Schnabel, J.A., Groves, A.R., Andersson, J.L., Woolrich, M.W., 2012. Probabilistic inference of regularisation in non-rigid registration. *NeuroImage* 59,

- 2438–2451. URL: <https://www.sciencedirect.com/science/article/pii/S105381191101041X>, doi:<https://doi.org/10.1016/j.neuroimage.2011.09.002>.
- Sourati, J., Gholipour, A., Dy, J.G., Tomas-Fernandez, X., Kurugol, S., Warfield, S.K., 2019. Intelligent labeling based on fisher information for medical image segmentation using deep learning. *IEEE Transactions on Medical Imaging* 38, 2642–2653. doi:10.1109/TMI.2019.2907805.
- Toussaint, N., Souplet, J.C., Fillard, P., 2007. MedINRIA: Medical Image Navigation and Research Tool by INRIA, in: Proc. of MICCAI'07 Workshop on Interaction in medical image analysis and visualization, Brisbane, Australia, Australia. URL: <https://hal.inria.fr/inria-00616047>.
- Tsai, A., Yezzi, A., III, W.W., Tempany, C., Tucker, D., Fan, A., Grimson, W., Willsky, A., 2003. A shape-based approach to the segmentation of medical imagery using level sets. *IEEE Transaction in Medical Imaging* 22, 137–54.
- Van Leemput, K., 2009. Encoding probabilistic brain atlases using bayesian inference. *IEEE Transactions on Medical Imaging* 28, 822–837. doi:10.1109/TMI.2008.2010434.
- Vercauteren, T., Pennec, X., Perchant, A., Ayache, N., 2007. Diffeomorphic Demons Using ITK's Finite Difference Solver Hierarchy, in: Insight Journal – ISC/NA-MIC Workshop on Open Science at MICCAI 2007, no address, Australia. URL: <https://hal.inria.fr/inria-00616035>. source code available online.
- Wang, J., Wells, W., Golland, P., Zhang, M., 2018. Efficient laplace approximation for bayesian registration uncertainty quantification, in: 21st International Conference on Med Image Comput Comput Assist Interv.(MICCAI 2018), Granada, Spain. pp. 880–888. doi:10.1007/978-3-030-00928-1\_99.
- Wang, Z., Vandersteen, C., Demarcy, T., Gnansia, D., Raffaelli, C., Guevara, N., Delingette, H., 2019. Deep learning based metal artifacts reduction in post-operative cochlear implant ct imaging, in: Shen, D., Liu, T., Peters, T.M., Staib, L.H., Essert, C., Zhou, S., Yap, P.T., Khan, A. (Eds.), *Medical Image Computing and Computer Assisted Intervention – MICCAI 2019*, Springer International Publishing, Cham. pp. 121–129.
- Wang, Z., Vandersteen, C., Demarcy, T., Gnansia, D., Raffaelli, C., Guevara, N., Delingette, H., 2020a. A Deep Learning based Fast Signed Distance Map Generation, in: *Medical Imaging with Deep Learning*, Montréal, Canada.
- Wang, Z., Vandersteen, C., Raffaelli, C., Guevara, N., Delingette, H., 2020b. One-shot Learning Landmarks Detection. URL: <https://hal.inria.fr/hal-03024759>. working paper or preprint.
- Wimmer, W., Anschuetz, L., Weder, S., Wagner, F., Delingette, H., Caversaccio, M., 2019. Human bony labyrinth dataset: Co-registered ct and micro-ct images, surface models and anatomical landmarks. *Data in Brief* 27,

104782. URL: <https://www.sciencedirect.com/science/article/pii/S2352340919311370>, doi:<https://doi.org/10.1016/j.dib.2019.104782>.

Zhang, J., Yu, W., Yang, X., Deng, F., 2019. Few-shot learning for ear recognition, in: Proceedings of the 2019 International Conference on Image, Video and Signal Processing, Association for Computing Machinery, New York, NY, USA. p. 50–54. URL: <https://doi.org/10.1145/3317640.3317646>, doi:10.1145/3317640.3317646.