Medical Image Analysis 18 (2014) 9–21

Contents lists available at ScienceDirect

Medical Image Analysis

journal homepage: www.elsevier.com/locate/media

A spatio-temporal latent atlas for semi-supervised learning of fetal brain segmentations and morphological age estimation

Eva Dittrich [a,b], Tammy Riklin Raviv [c], Gregor Kasprian [d], René Donner [a], Peter C. Brugger [e], Daniela Prayer [d], Georg Langs [a,f,*]

[a] Computational Imaging Research Lab (CIR), Department of Biomedical Imaging and Image-guided Therapy, Medical University of Vienna, Austria
[b] Machine Vision Group, Profactor GmbH, Steyr, Austria
[c] Department of Electrical and Computer Engineering, Ben-Gurion University, Israel
[d] Division of Neuroradiology, Department of Biomedical Imaging and Image-guided Therapy, Medical University of Vienna, Austria
[e] Center for Anatomy and Cell Biology, Department of Systematic Anatomy, Medical University of Vienna, Austria
[f] Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology, Cambridge, MA, USA

ARTICLE INFO

Article history:
Received 21 December 2012
Received in revised form 11 August 2013
Accepted 21 August 2013
Available online 30 August 2013

Keywords:
Fetal brain development
Spatio-temporal latent atlas
Segmentation
Magnetic resonance imaging

ABSTRACT

Prenatal neuroimaging requires reference models that reflect the normal spectrum of fetal brain development, and summarize observations from a representative sample of individuals. Collecting a sufficiently large data set of manually annotated data to construct a comprehensive *in vivo* atlas of rapidly developing structures is challenging but necessary for large population studies and clinical application. We propose a method for the semi-supervised learning of a spatio-temporal latent atlas of fetal brain development, and corresponding segmentations of emerging cerebral structures, such as the ventricles or cortex. The atlas is based on the annotation of a few examples, and a large number of imaging data without annotation. It models the morphological and developmental variability across the population. Furthermore, it serves as basis for the estimation of a structures' morphological age, and its deviation from the nominal gestational age during the assessment of pathologies. Experimental results covering the gestational period of 20–30 gestational weeks demonstrate segmentation accuracy achievable with minimal annotation, and precision of morphological age estimation. Age estimation results on fetuses suffering from lissencephaly demonstrate that they detect significant differences in the age offset compared to a control group.

1. Introduction

The analysis of the early development of a fetus in utero offers rich insights into the genesis of human anatomy. In particular, the emerging cerebral morphology is of both clinical and academic interest. The advance of novel imaging methods, such as ultra-fast Magnetic Resonance Imaging (MRI), allows for high-resolution image acquisition in utero (Garel, 2004) and the observation of the rapid fetal cerebral development. Fig. 1 shows an example of the developing brain from 20th to 30th gestational weeks (GW).

Even though fetal MRI provides a wealth of information, clinical assessment is typically performed qualitatively (Ghai et al., 2006). Both, clinicians and researchers need models capturing the developmental characteristics and variability in a large population. They can serve as basis for the study of developmental paths in healthy and patient groups, and as reference during quantitative assessment of individual cases in a clinical setting. One possible solution are spatio-temporal models and corresponding segmentations of brain structures learned from large numbers of in vivo data. A main limitation of this approach is the difficulty of acquiring complete annotation for a sufficiently large number of cases. We propose to learn a model or *atlas* of the development of a fetal anatomical structure as a latent spatio-temporal prior that connects the segmentations across subjects at different gestational ages. Starting from a small set of annotated cases, we learn segmentations for a large population together with spatio-temporal priors that link segmentation and gestational age.

The need to perform quantitative group-wise studies in neuroimaging has motivated intense research aiming at establishing accurate correspondences across individuals and labeling anatomical regions in the brain. In studies investigating the adult brain, this is typically achieved by an annotated atlas that serves as reference template (Fischl et al., 2002, 2004; Smith et al., 2004; Woolrich et al., 2009; Ashburner, 2007). Each individual is

* Corresponding author at: Computational Imaging Research Lab (CIR), Department of Biomedical Imaging and Image-guided Therapy, Medical University of Vienna, Austria. Tel.:+43 14040073725.
E-mail addresses: eva.dittrich@meduniwien.ac.at (E. Dittrich), georg.langs@meduniwien.ac.at (G. Langs).

1361-8415/$ - see front matter © 2013 Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.media.2013.08.004

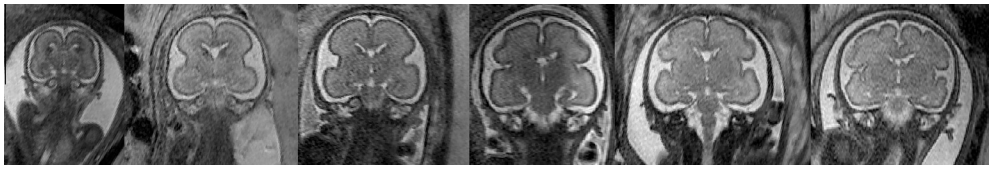

Fig. 1. A consistently positioned coronal slice of different individuals illustrating the cerebral development at GW 20, 22, 24, 26, 28, 30 (from left to right).

registered to the template, and coordinates and region labels such as Brodmann areas (Brodmann, 1909) are transferred accordingly. Templates and reference spaces range from the single subject Talairach template (Talairach and Tournoux, 1988), to the Montreal Neurological Institute (MNI) space with a template based on larger control cohorts (Evans et al., 1993; Mazziotta et al., 1995). Aside from transferring labels to new subjects, an atlas can also reveal characteristics of the population it is build from. Examples are population studies based on voxel-based morphometry or shape analysis to detect differences between adult subject groups (Ashburner and Friston, 2000; Karas et al., 2003), or fetal cohorts (Gholipour et al., 2012). Davis et al. (2007) performed kernel regression to capture spatio-temporal characteristics of the aging human adult brain from MR scans. A similar method using an adaptive kernel was published by Serag et al. (2012) for the developing brain. Spatio-temporal atlases detected developmental speed differences in two chimpanzee species (Durrleman et al., 2009, 2010). Aljabar et al. (2011) reported results on manifold learning as representation of a neonatal dataset, and Kuklisova-Murgasova et al. (2010) introduced an atlas for the developing brain from pre-term infants between the 29th and 44th GW. Encouraging results for fetal data covering a period from GW 20 to 24 were reported in Habas et al. (2010), where the authors presented a spatio-temporal atlas for a dataset of 20 fetal brains imaged in utero by T2-weighted MRI. Spatio-temporal tensor-based volume morphometry was proposed to study the sulcal formation of fetal brains (Rajagopalan et al., 2011). In a recent paper by Gholipour et al. (2012), a spatio-temporal atlas facilitated fetal brain MRI segmentation of patients and normal controls for the detection of ventricle atrophies due to pathologies. Fishbaugh et al. (2012) introduce an interesting approach that learns an atlas of the population via shape regression. In addition, subject-specific growth trajectories are estimated. To measure shape variability, the generic model is warped to individual subjects using diffeomorphic mapping. In their experiments, they analyzed the growth scenario of two different population groups. Results were reported on both a synthetic and a clinical dataset. Latent atlases have been proposed to connect segmentations in multi-modal imaging data of pathologies, such as brain-tumors (Riklin-Raviv et al., 2009). The latent atlas is learned from partially annotated imaging data to capture the varying representation of tissue properties across modalities. It learns potentially different segmentations in all modalities and a latent prior that represents the tumor presence across the registered imaging data. While Riklin-Raviv et al. (2009) accounts for variability across the data, it does not incorporate parameters such as time, or age, that might have a systematic effect on the shape, or distribution of anatomical structures. Time is crucial when observing developmental processes or disease progression. Its effect can be substantial, and the integration of parameters that have a characteristic effect is conceptually different from random variability in a population.

In this paper, we propose a method to build a spatio-temporal latent atlas capturing the development characteristics of cerebral structures during early human brain development. Instead of exhaustive annotation of anatomical structures, it simultaneously learns a spatio-temporal latent atlas and segmentations of individual structures based on a small number of annotated examples, and a large number of examples without annotation. We refer to this as a semi-supervised approach as opposed to fully-supervised approaches that use annotations on all examples, such as the leave-one-out experiments in Habas et al. (2010) or Gholipour et al. (2012). To connect the atlas with individual imaging data, we use a statistically-driven level-set segmentation framework. It translates the prior or uncertainty shared across the data as the logistic function of the corresponding level-set values, similar to Pohl et al. (2007). The spatio-temporal latent atlas is a probabilistic prior or map of the presence of an anatomical structure as a function of location and gestational age (GA). We use kernel regression for a continuous representation of the temporal domain, allowing the interpolation for GA that are not represented in the training dataset. Initial results were reported in Dittrich et al. (2011). The resulting four-dimensional atlas links the segmentations, and represents the cross-sectional component of variability in the population for a specific age, and the developmental gradient of the structure along the age axis. This is visualized in Fig. 2 for the period between GW 20 and 30. The gradient of the spatio-temporal atlas can be regarded as measure of uncertainty in two ways. In Fig. 2(a), the variability among cases of the same age is depicted as the local spatial gradient of the atlas at the surface boundary. Red areas are regions of high variability among the subjects, the gradient is low in these areas since segmentations are dissimilar across subjects. Yellow corresponds to relatively stable areas with a high local gradient of the prior. Fig. 2(b) shows the gradient of the atlas along its longitudinal axis at the structure surface. Red areas are expanding, blue areas are shrinking.

The atlas has several uses: it can serve as reference to represent characteristic development and its variability in a population, it can be used to identify deviations from a healthy population quantitatively, and it is a prerequisite for group-studies of the fetal development such as (Schöpf et al., 2012a,b), since it provides the means to establish correspondences across subjects, and age. Finally, it allows for estimating a morphological age, by matching individuals to the atlas along the age axis.

The remainder of this paper is structured as follows. In Section 2, we introduce the dataset and preprocessing methods followed by an in-depth explanation of the methodology. We define the problem, and detail methods for the learning of the latent atlas, and its use for morphological age estimation. In Sections 3 and 4 we present and discuss our experimental results in detail, and Section 5 closes with a conclusion and outlook.

2. Material and methods

2.1. Study data collection

This work is part of an ongoing collaboration with neuroradiologists and anatomists specialized on fetal MRI assessment. We took advantage of two distinct datasets. The first includes 32 fetal MR images of singleton pregnancies depicting the brain between GW 20 and 30 were retrospectively investigated. Cases suspicious for cardiac abnormalities, complex syndromes or chromosomal abnormalities were excluded from this study. The second dataset

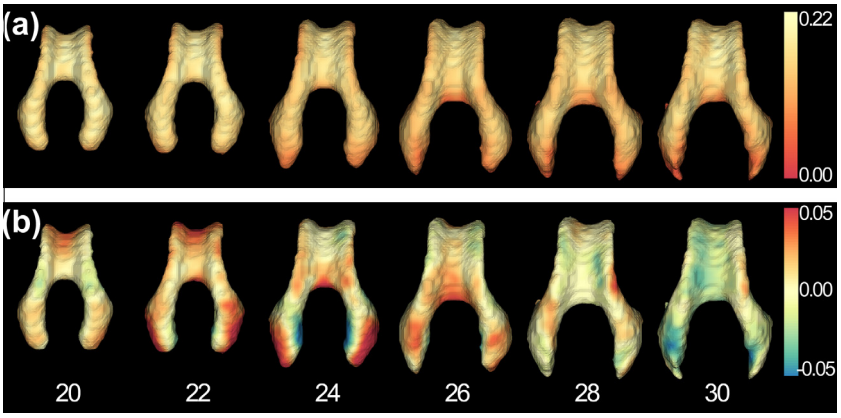

Fig. 2. Visualization of the (a) spatial and (b) temporal component of the spatio-temporal atlas for the lateral ventricles of the fetal brain (frontal horns pointing upwards, occipital horns downwards) covering GW 20–30 (ages are indicated for each column). (a) The spatial gradient of the atlas $\delta\eta(\mu)/\delta\mathbf{x}$ for each point $\mathbf{x}$ of the template surface. Red indicates areas of high variability among the subjects at this time point, whereas yellow corresponds to areas relatively similar across the population. (b) The longitudinal gradient of the atlas $\delta\eta(\mu)/\delta\mu$. Red areas correspond to a high positive gradient, i.e., local expansion of ventircular size, while blue areas indicate negative gradient, i.e., reduction in ventricular size. Yellow areas are static. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

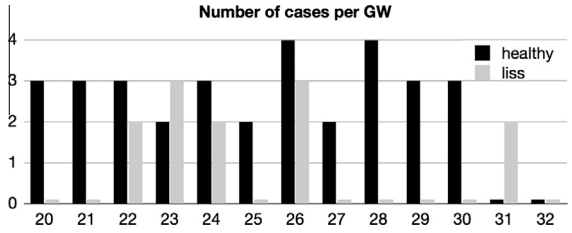

Fig. 3. Number of cases per GW. The dataset containing the healthy cases is shown in black, whereas the lissencephaly cases are depicted in gray.

contains MR images of 12 fetuses diagnosed with lissencephaly which is a certain type of malformations of the cortical development. The age distribution of both patient groups is shown in Fig. 3.

MRI acquisition. Data was acquired with a single-shot, fast spin-echo T2-weighted MRI sequence. The images are captured by a 1.5 Tesla Philips Gyroscan superconducting unit during clinical routine. In-plane resolution was 0.78–0.9 pixels per mm, slice thickness 3–4.4 mm, acquisition matrix 256×256, field of view 200–230 mm, SAR < 100%/4.0 W/kg, image acquisition time $\leqslant$ 20 s, TE 100–140 ms, TR 9000–19,000 ms. During image acquisition, neither the fetus nor the mother were sedated. The preparation of the sequences, as well as planning of exact orthogonal orientations (axial, coronal, sagittal views of the fetus), was guided either by an anatomist or an experienced fetal radiologist, both with substantial expertise in the field of fetal neuroimaging, being present at each examination.

Preprocessing. We constructed 32 isotropic 3D volumes by merging three images per patient that were acquired in axial, coronal and sagittal planes. The high resolution (HR) volumes were reconstructed using the Baby Brain Toolkit (BTK) by Rousseau et al. (2011), resulting in an isotropic voxel size between 0.78 and 0.9 mm, and an image resolution of $225 \times 288 \times 302$ voxels. An initial coarse alignment among the HR images was achieved based on the positions of the eyes and the position where the skull ends (*occipital foramen magnum*). In a subsequent step, the HR data was further registered using a registration tool (Klein et al., 2010). First, we rigidly registered the images to achieve a better alignment. Additionally, we performed a subsequent affine and spatio-temporal non-rigid registration using B-splines with a control point spacing of 32 pixels. To this end, we non-rigidly registered each case to its corresponding age-specific template. The age-specific template is composed as a weighted sum of gray-scale volumes, where the weights are given by the age, as described in Eq. (9). We evaluated the segmentation of the ventricles and cortex on both the rigid data set D_R, as well as the spatio-temporal non-rigid data set D_{NR}.

Manual segmentations. Standard of reference segmentations were generated by a physician specialized on fetal neuroradiology, who manually annotated the ventricles in all images using ITKsnap (Yushkevich et al., 2006). In this paper, the word *annotation* stands for a manually generated label, whereas *segmentation* refers to an automatic labeling of the region of interest, by the proposed algorithm.

2.2. Problem definition

Our primary goal is the construction of a spatio-temporal atlas η as probabilistic prior for the presence of an anatomical structure in space as a function of age μ. It models the development of a normal fetal brain within a continuous time period during gestation. Fig. 4 gives an overview of the proposed method. The input to our model is a set of N three dimensional MR scans $\mathbf{I}_1, \ldots, \mathbf{I}_N$ of healthy fetal brains of different individuals. We assume that the images are co-aligned, and that each image $\mathbf{I}_n : \Omega \rightarrow \mathbb{R}^+$ ($\Omega \subset \mathbb{R}^3$) is assigned a corresponding nominal age $\mu_n \in [\mu_{\min}; \mu_{\max}]$.

Let $\{\mathbf{\Gamma}_1, \ldots, \mathbf{\Gamma}_N\}$ be the binary segmentations corresponding to the images $\{\mathbf{I}_1, \ldots, \mathbf{I}_N\}$, where $\mathbf{\Gamma}_n$: $\Omega \rightarrow \{0, 1\}^V$, and V is the number of voxels in the image. In the proposed formulation each $\mathbf{\Gamma}_n$ partitions the image domain Ω into the region of interest (ROI) ω_n and the background $\Omega \backslash \omega_n$ such that the set $\{\omega_n\}$ defines the spatial locations of a particular anatomical structure across the ensemble. Let us also assume that a small fraction $k \ll N$ of the segmentations is given and can be used for training. Extraction of the remaining (unknown) segmentations $\{\mathbf{\Gamma}_{k+1}, \ldots, \mathbf{\Gamma}_N\}$ is essential for the inference of the spatio-temporal atlas η: $\Omega \rightarrow [0, 1]$ and a significant goal of itself. We emphasise that η is derived from the entire population and covers a continuous age range $\mu \in [\mu_{\min}, \mu_{\max}]$.

Let $f(\mathbf{I}_n) = F_n$ define a vector function of the image intensities $\mathbf{I}_n$ such that $F_n(x, y, z)$ is a feature vector of size d of the voxel located at (x, y, z). We assume that $\mathbf{\Gamma}_n$ generates F_n with probability $p(F_n | \mathbf{\Gamma}_n; \theta)$, where θ is a set of parameters that model the image intensity features of the entire set. We also assume that given η, the segmentations $\{\mathbf{\Gamma}_n\}$ are conditionally independent and are drawn

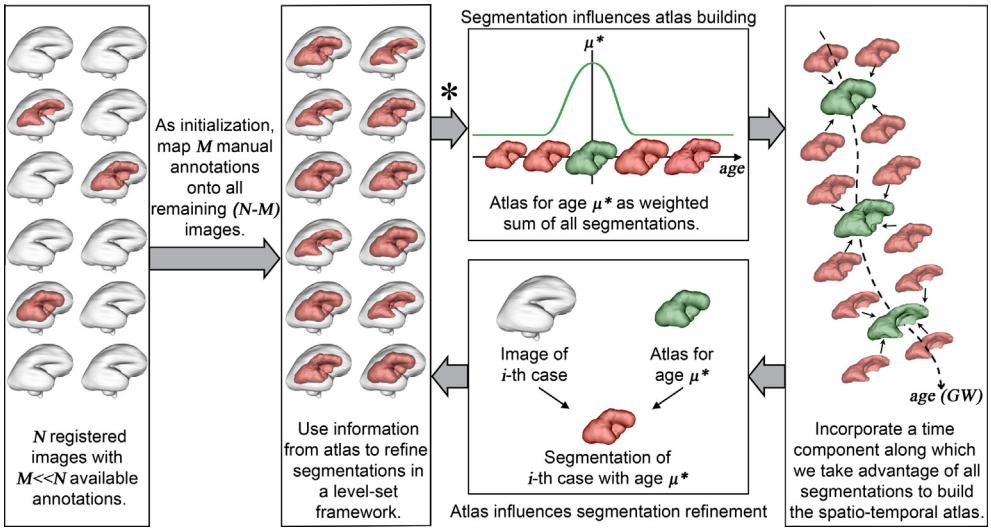

Fig. 4. Schematic overview of the proposed method. Starting with a subset of annotations, the spatio-temporal atlas and the segmentation for all images are initialized. Then, the atlas is re-estimated based on the current segmentations, and in an alternating manner, the segmentations are refined based on the current spatio-temporal atlas. After the first iteration of segmentation refinement and atlas building, the estimation of each subject's morphological age (Fig. 5) can be integrated at the position denoted by $*$.

from a probability distribution based on the age matched atlas $p(\mathbf{\Gamma}_n;\ \eta,\ \mu_n)$.

The proposed framework consists of two parts: training and testing. During training, the model parameters related to the image intensities, θ are inferred using the known segmentations $\{\mathbf{\Gamma}_1, \ldots, \mathbf{\Gamma}_k\}$ and their corresponding images $\{\mathbf{I}_1, \ldots, \mathbf{I}_k\}$. Specifically θ are the parameters of a Random Forest (RF) classifier (Breiman, 2001). Here, we make an explicit assumption that the RF parameters are consistant across the image dataset. At the test phase, both the latent atlas η, and the unknown segmentations $\{\mathbf{\Gamma}_n\}$ are jointly optimized as follows:

$$\{\hat{\eta}, \hat{\mathbf{\Gamma}}\} = \arg\max_{\{\eta,\mathbf{\Gamma}\}} \log p(F_1 \ldots F_N, \mathbf{\Gamma}_1 \ldots \mathbf{\Gamma}_N; \eta, \mu, \theta). \tag{1}$$

Using the conditional independence of F_n we get:

$$\{\hat{\eta}, \hat{\mathbf{\Gamma}}\} = \arg\max_{\{\eta,\mathbf{\Gamma}\}} \sum_{n=1}^{N} \log p(F_n, \mathbf{\Gamma}_n; \eta, \mu_n, \theta). \tag{2}$$

For a fixed value of $\hat{\eta}$, Eq. (2) implies that the segmentations can be estimated by solving $N - k$ separate MAP problems:

$$\hat{\mathbf{\Gamma}}_n = \arg\max_{\mathbf{\Gamma}_n} \left[\log p(F_n|\mathbf{\Gamma}_n; \theta) + \log p(\mathbf{\Gamma}_n; \eta, \mu_n)\right], \tag{3}$$

where

$$\log p(F_n, \mathbf{\Gamma}_n; \eta, \mu_n, \theta) = \log p(F_n|\mathbf{\Gamma}_n; \theta) + \log p(\mathbf{\Gamma}_n; \eta, \mu_n). \tag{4}$$

We then fix $\hat{\mathbf{\Gamma}}$ and estimate $\eta(\mu)$ for each μ assigned to the given image set.

$$\hat{\eta} = \arg\max_{\eta} \sum_{n=1}^{N} \log p(\mathbf{\Gamma}_n; \eta, \mu_n). \tag{5}$$

As we will show in the following, the proposed spatio-temporal model η generalizes the latent atlas approach introduced in Riklin-Raviv et al. (2010).

2.3. Generalized image likelihood model

In Riklin-Raviv et al. (2010) the image intensities were modeled by Gaussian mixtures were each voxel is represented by a single scalar – its intensity. Here, a feature vector is assigned to each of the image voxels and an RF classifier is constructed, which allows to discriminate between ROI and background feature vectors. Let v be a voxel located at (x, y, z) and let $F_n(x, y, z)$ be its associated feature vector, we will denote by p_{in} the probability that $(x, y, z) \in \omega$ and by p_{out} the probability that $(x, y, z) \in \Omega\backslash\omega$ based on the RF classifier. The image likelihood term gets the form:

$$\log p(F_n|\mathbf{\Gamma}_n, \theta) = \sum_{\{v|\mathbf{\Gamma}_n^v=1\}} \log p_{\text{in}}(F_n^v; \theta) + \sum_{\{v|\mathbf{\Gamma}_n^v=0\}} \log p_{\text{out}}(F_n^v; \theta), \tag{6}$$

2.4. The spatio-temporal latent atlas

The construction of the spatio-temporal atlas that links the segmentations of the training images and the corresponding nominal ages is the main component of the proposed framework. Consider the set of pairs $\{(\mathbf{\Gamma}_1, \mu_1) \ldots (\mathbf{\Gamma}_N, \mu_N)\}$ where $\mathbf{\Gamma}_n$ is the current estimate of the segmentation of $\mathbf{I}_n$ and μ_n is the nominal age of subject n. Note that $\mu_n \in [\mu_{\min};\ \mu_{\max}]$ and is not necessarily uniformly sampled. Hence, in order to construct a comprehensive and time-continuous atlas that represents reliably the entire age-range $[\mu_{\min};\ \mu_{\max}]$, including ages that are associated with very few (or none at all) examples in the training set, we use a non-parametric technique known as kernel regression. Specifically, we express $\eta(\mu_n)$ by the conditional expectation of $\mathbf{\Gamma}_n$ and μ_n that stand for the dependent and the independent variables, respectively:

$$E(\mathbf{\Gamma}_n|\mu_n) = \eta(\mu_n).$$

We then use the Nadaraya–Watson estimator (Nadaraya, 1964; Watson, 1964) to rewrite $\eta(\mu^*)$ for a specific age μ^* as the weighted sum of the segmentations:

$$\hat{\eta}(\mu^*)^h = \frac{\sum_{n=1}^{N} \mathcal{K}_h(\mu^* - \mu_n)\mathbf{\Gamma}_n}{\sum_{n=1}^{N} \mathcal{K}_h(\mu^* - \mu_n)}, \tag{7}$$

with $\mathcal{K}$ being a kernel with a bandwidth h. Note that the spatio-temporal atlas η is four dimensional as it is also defined over a continuous time domain. However, in practice we estimate the atlas for a set of discrete time points that correspond to the GAs of the associated examples (or the training set), or for the test sample age, when using the atlas to segment a new case. We therefore use the notation $\eta(\mu)$ to define the 3D template of the GA μ. $\eta(\mu_n)$ denotes the atlas at the age of subject n in the training set, $\eta(\mu^*)$ denotes the atlas at the specific age μ^*. Let $\mathcal{K}$ be a Gaussian kernel,

then the latent atlas estimate can then be rewritten (similar to Davis et al. (2007); Kuklisova-Murgasova et al. (2010)) as follows:

$$\hat{\eta}(\mu^*) = \frac{1}{W_{\mu^*}} \sum_{n=1}^{N} w_{*,n} \mathbf{\Gamma}_n, \tag{8}$$

where

$$w_{*,n} = e^{-\frac{(\mu^* - \mu_n)^2}{\sigma^2}}, \quad \text{and } W_{\mu^*} = \sum_{n=1}^{N} w_{*,n}. \tag{9}$$

Here σ is the standard deviation which reflects the expected variability in the ROIs due to age. For $\sigma \ll 1$, only segmentations of subjects of similar ages are considered in the atlas computation. When $\sigma \to \infty$, the segmentations are equally weighted as in the original latent atlas framework (Riklin-Raviv et al., 2010). The kernel regression technique used to create a age specific prior is similar in spirit to approaches such as Davis et al. (2007). While in Davis et al. (2007) the observed images are paired with the subjects' ages, here the dependent variables are the unknown segmentations that are estimated alternately with the construction of the latent spatio-temporal atlas. The latent atlas should facilitate the segmentations also for ages that are poorly represented by the training examples.

2.5. Linking age and atlas prior

The posterior probability $p(\mathbf{\Gamma}_n; \eta, \mu_n)$ (the right hand side term of the product in Eq. (3)) expresses the spatio-temporal dependency of $\mathbf{\Gamma}_n$. Similar to Riklin-Raviv et al. (2010) we use the log of the Bernoulli distribution to construct the associated spatio-temporal cost term that links age and atlas prior:

$$\mathcal{E}_S(\mathbf{\Gamma}_n|\eta, \mu_n) = -(\mathbf{\Gamma}_n \log(\eta(\mu_n)) + (1 - \mathbf{\Gamma}_n) \log(1 - \eta(\mu_n))), \tag{10}$$

where $\mathcal{E}_S(\mathbf{\Gamma}_n|\eta, \mu_n)$ is $-\log p(\mathbf{\Gamma}_n; \eta, \mu_n)$. It penalizes the difference between each individual voxel and the corresponding age matched atlas voxel. Voxels that are wrongly assigned to be part of the segmentation $v \in \Gamma$ are restrained by the atlas η and therefore the overall sum results in higher costs (and vice versa). Note that the spatio-temporal atlas η can be used via this term to either facilitate the evolution of the segmentation of the training examples or after the completeness of the training phase to support the segmentation of a new example as long as its associated GA is within the age-range of the training examples.

2.6. Estimation of the morphological age

The proposed spatio-temporal atlas is based on the assumption that brain development rate is similar across healthy fetuses and that age differences induce a more significant change in brain structures than the expected differences between subjects of the same age. Once the normal subjects' atlas is constructed, we can use it to determine the developmental age of a given subject based on the similarity of its brain structures to training examples with similar GAs. This process is illustrated in Fig. 5. The same mechanism can also be used to refine the age estimation of the training subjects concurrently with the segmentation and the atlas estimation process. In this context, the estimation of the morphological age of each subject can be integrated into the whole algorithm at position $*$ in Fig. 4. In other words, the developmental age (rather than the true age) of a fetus with corresponding segmentation $\mathbf{\Gamma}_*$ can be obtained by optimizing the following expression with respect to μ^*:

$$\hat{\mu}^* = \arg\max_{\mu^*} (-\log p(\mathbf{\Gamma}_*|\eta, \mu^*)). \tag{11}$$

Note that the expression above, as a function of the segmentations and the corresponding ages, is obtained by substituting Eq. (9) in Eq. (8) and using the relation in Eq. (10) while assuming that the prior probabilities $p(\mu^*)$ and $p(\mathbf{\Gamma}_*)$ are independent and constant. In practice, this means we want to minimize the costs in Eq. (10) based on the choice of the age μ^*. This optimization results in an age estimate μ^* for a segmentation $\mathbf{\Gamma}_*$ based on the atlas η. The age correction can be seen as option to extract information about how the age deviates from the nominal age, i.e. it is an indicator of the state of morphological development.

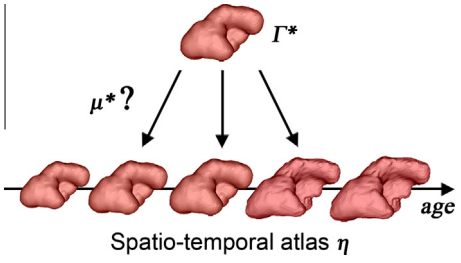

Fig. 5. Schematic overview of the morphological age estimation. The age μ^* of a segmentation $\mathbf{\Gamma}^*$ is estimated with help of the spatio-temporal atlas η.

2.7. Level-set framework

We take advantage of a probabilistic level-set formulation (Osher and Sethian, 1988) that is well established in literature for the segmentation of a single or set of aligned images. Simultaneously, we build a probabilistic spatio-temporal atlas (as illustrated in Fig. 4) from our segmentations. Let ϕ_n: $\Omega \to \mathbf{R}$ denote a level-set function relating to the region of interest (ROI) in image $\mathbf{I}_n$, and its zero-level $C_n = \{\mathbf{x}|\phi(\mathbf{x}) = 0\}$ represents the ROI's boundary. We apply a smooth approximation of the Heaviside function $\tilde{H}(\phi_n)$ (Chan and Vese, 2001) to obtain a fuzzy division of the image domain Ω into foreground and background regions. For $\tilde{H}(\phi)$ and its derivative with respect to ϕ, i.e. $\tilde{\delta}(\phi)$, we use the same formulations as in Riklin-Raviv et al. (2010). We next construct a unified cost functional of $\{\phi_n\}$. An alternating minimization process of the cost functional allows the joint estimation of the latent spatio-temporal atlas η and the segmentations $\mathbf{\Gamma}_n$ (or ϕ_n), using calculus of variations. The proposed cost functional $\mathcal{E}(\phi_1, \ldots, \phi_n, \Theta)$ is the weighted sum of the following energy terms:

$$\mathcal{E}(\phi_1, \ldots, \phi_n, \Theta) = \gamma \mathcal{E}_L + \beta \mathcal{E}_I + \alpha \mathcal{E}_S + \kappa \mathcal{E}_A, \tag{12}$$

where $\mathcal{E}_L$ is the classic smoothness term, $\mathcal{E}_I$ is an image likelihood term, $\mathcal{E}_S$ is a spatio-temporal term, $\mathcal{E}_A$ is an area term and α, β, γ, κ are the corresponding weight parameters. The level-set functions ϕ_n corresponding to the ensemble images are evolved iteratively using a gradient descent paradigm. The unknown parameters Θ are re-estimated alternately with the segmentations. In Riklin-Raviv et al. (2010) a generative model for segmentation was suggested and each of the functional terms has been evolved accordingly using the level-set framework. These energy terms are revisited hereby for self completeness. Our main contribution which is the spatio-temporal term is discussed in details in the following section.

2.8. Cost functional terms

The spatio-temporal term. The spatio-temporal term extends the spatial term suggested in Riklin-Raviv et al. (2010) to the temporal domain. Let ϕ_n correspond to the segmentation of a brain scan of a fetus at age μ_n, and let $\eta(\mu_n)$ define the latent spatial probabilistic prior associated with μ_n. The spatio-temporal term in Eq. (10) can be re-written using the continuous level-set framework as follows:

$$\mathcal{E}_S(\phi_n,\eta(\mu_n)) = -\int_\Omega \left[\widetilde{H}(\phi_n(\mathbf{x}))\log\eta(\mu_n)(\mathbf{x}) + \widetilde{H}(-\phi_n(\mathbf{x}))\log(1-\eta(\mu_n)(\mathbf{x}))\right]d\mathbf{x}, \tag{13}$$

where $\widetilde{H}(\phi_n)$ is the fuzzy approximation of the binary segmentation $\mathbf{\Gamma}_n$. The latent spatial parameters, called the latent atlas, are re-estimated at every iteration using the current segmentation estimates of the entire ensemble as follows:

$$\hat{\eta}(\mu_n) \approx \frac{1}{W_{\mu_n}}\sum_{m=1}^{N} w_{m,n}\widetilde{H}(\phi_m(\mathbf{x})). \tag{14}$$

Image likelihood term. Let p_{in} and p_{out} define probability maps of the foreground and the background regions, correspondingly. We obtain these probability maps by training a random forest classifier (Breiman, 2001; Geurts et al., 2006) on a set of training images and labels during the initialization. The training set consists of the same images that are used for the initialization of the atlas and segmentations. The random forest classifier can then be applied to predict the trained labels on any test case. As visual features F_n^v at each voxel v, we use the gray-level differences between v and its neighboring voxels being sampled according to a Gaussian distribution $\mathcal{N}(0, d_{spatial})$. The features are combined into F_n for each image $\mathbf{I}_n$. These spatial offsets $d_{spatial}$ are fixed for each decision tree.

Based on these probability maps, we compute an image likelihood term as derived in Riklin-Raviv et al. (2010):

$$\mathcal{E}_I(\phi_n,\Theta) = -\int_\Omega [\widetilde{H}(\phi_n(\mathbf{x}))\log(p_{in}(F_n;\theta)) + \widetilde{H}(-\phi_n(\mathbf{x})) \times \log(p_{out}(F_n;\theta))]d\mathbf{x}, \tag{15}$$

with p_{in} representing the result of the classifier as probability map in [0; 1], and $p_{out} = 1 - p_{in}$.

Length term. $\mathcal{E}_L$ denotes the contour length regularizer that is commonly used in the level-set literature, like Chan and Vese (2001). It controls the length of the segmentation, and restrains boundary smoothness:

$$\mathcal{E}_L(\phi_n) = \int_\Omega |\nabla\widetilde{H}(\phi_n(\mathbf{x}))|d\mathbf{x}. \tag{16}$$

Area term. The area term is denoted by $\mathcal{E}_A$ and has similar behavior as a balloon force (Cohen and Cohen, 1993). It describes the area of the shape or structure of interest:

$$\mathcal{E}_A(\phi_n) = \int_\Omega \widetilde{H}(\phi_n(\mathbf{x}))d\mathbf{x}. \tag{17}$$

2.9. Optimization of the cost functional

The cost function (Eq. (12)) is optimized by applying two steps in an alternating manner: for fixed model parameters Θ, we evolve the level set function ϕ_n using a gradient descent process:

$$\phi_n(\mathbf{x}, t+\Delta t) = \phi_n(\mathbf{x}, t) + \Delta t\frac{\partial\phi_n}{\partial t}, \tag{18}$$

where t denotes the iteration number, and Δt is the size of the update step in every iteration. The term $\phi_n = -\delta\phi_n/\delta t$ is obtained from the first variation of the unified cost functional, i.e. Eqs. (14)–(17) substituted in Eq. (12):

$$\phi_n = \delta(\phi_n)(\{\mathbf{x}\}\left\{\gamma\mathrm{div}\left(\frac{\nabla\phi_n}{|\nabla\phi_n|}\right) + \beta[\log p_{in}(F_n(\mathbf{x});\theta) - \tag{19}$$

$$- \log p_{out}(F_n(\mathbf{x});\theta)] + \alpha[\log\eta(\mu_n) - \log(1-\eta(\mu_n))] + \kappa\}. \tag{20}$$

Then, we fix the segmentations ϕ_n and update the model parameters, i.e. image information θ_n and spatio-temporal atlas $\eta(\mu_n)$. When there is no age correction the initial set of weights, w remains and the spatio-temporal atlas is re-estimated based on the updated segmentations as in Eq. (8). Otherwise, the weights can be easily recomputed based on the new age estimates.

3. Experiments

3.1. Setup

We evaluated the segmentation accuracy of the method on two cerebral structures: the ventricles, and the cortex. In overall $N = 32$ cases the ventricles were manually segmented, and in 7 cases the cortex was segmented. Segmentation and atlas building was performed with 40 iterations, and the same parameters were used for both structures: $\alpha = 0.3$, $\beta = 0.5$, $\gamma = 0.1$, $\kappa = 0.1$ for the weights of the cost terms, and $\sigma = 2$ for standard deviation in the age weighting term (Eq. (9)). A stop criterion was used that compared the costs c_i in iteration i to the previous iterations and fixed the segmentation of a case if $(c_i - c_{i-1}) - (c_{i-1} - c_{i-2}) < 100$. To measure segmentation accuracy, we computed the Dice coefficient (DC) (Dice, 1945) between each segmentation and the corresponding ground-truth (i.e. annotation). In addition, we measured the mean and median distance from each point on the segmentation surface to the closest point on the annotation surface as error E_{surf} (similar to Hausdorff (1914)).

Comparison with other segmentation approaches. In the first experiment we compare the semi-supervised spatio-temporal atlas (SSV-ST) learning with 3 alternative approaches. Table 1 summarizes the methods. We evaluate the effect of two aspects specific to the proposed method: (1) the use of only a small fraction of annotated examples for segmentation and atlas learning, instead of all but one annotated examples as for example in (Habas et al., 2010) (semi-supervised SSV-* vs. fully-supervised FSV-*). (2) The use of a spatio-temporal latent atlas (*-ST), instead of a single latent atlas (*-SNG) as for example in (Riklin-Raviv et al., 2009). In all experiments, the *training set* consists of cases for which annotations are known and used to build (FSV-ST, FSV-SNG) or to initialize (SSV-ST, SSV-SNG) the atlas, and the *test set* represents the images for which no annotation is known, and for which the segmentations have to be learned by the algorithm. For the fully supervised methods (FSV-ST and FSV-SNG), we used $M = N - 1$ annotated cases for training, and one case for testing. For the semi-supervised methods (SSV-SNG and our proposed method SSV-ST) $M \ll N$ annotated cases were used (with $M = 3$). We performed either spatio-temporal atlas building (FSV-ST and SSV-ST), or single atlas building (FSV-SNG and SSV-SNG). For FSV-ST and FSV-SNG, we re-registered all training cases based on their annotation onto each other.

To evaluate the effect of registration we performed the atlas building on rigidly registered data D_R, and on non-rigidly registered data D_{NR} (for details on the data see Section 2.1).

- **SSV-ST:** This is the proposed algorithm. In this setting, the number of known annotations used for segmenting the dataset is significantly smaller than the size of the entire dataset. We randomly choose $M = 3$ cases with $GA < 23$, $23 < GA < 27$, and $GA > 27$ respectively, and run the segmentation process including atlas update for all other cases. The experiment is repeated 20 times, with different training triplets each.
- **SSV-SNG:** This approach corresponds to the latent atlas proposed in Riklin-Raviv et al. (2010). The atlas is learned together with the segmentations, and is constructed by averaging the evolving segmentations regardless of GA. We chose the same experimental setup as in SSV-ST, resulting in the same training case triplets and 20 runs.

Table 1
We compare the proposed method (SSV-ST) with 3 alternative approaches, that differ from SSV-ST in two aspects: approaches FSV-* learn the segmentation from a data set using all but one annotated case (instead of only a small fraction of annotated cases) comparable with (Habas et al., 2010), and approaches *-SNG use a single latent atlas instead of a spatio-temporal latent atlas comparable to (Riklin-Raviv et al., 2009).

Description	Abbrev.	No. of train data	Re-estimate atlas	No. of test data
Semi-supervised spatio-temporal	SSV-ST	M	yes	(N–M)
Semi-supervised single	SSV-SNG	M	yes	(N–M)
Fully supervised spatio-temporal	FSV-ST	$(N-1)$	no	1
Fully supervised single	FSV-SNG	$(N-1)$	no	1

- **FSV-ST:** We evaluate full annotation in a leave one out cross-validation experiment. $N-1$ out of N segmentations are known and used to build the spatio-temporal atlas that remains fixed throughout the segmentation process of the remaining volume. The result demonstrates to which extent the spatio-temporal atlas generalizes to examples not included in the atlas training, and how it improves segmentation. This is similar to the algorithm used in Habas et al. (2010).
- **FSV-SNG:** Similar to the experiment above, we take advantage of $N-1$ known segmentations and perform LOOCV. However, the atlas is constructed by averaging the known segmentations regardless of the GA.

To evaluate if the algorithm generalizes well to other structures, we again applied our algorithm to the non-rigidly registered data D_{NR}, this time in order to segment the cortex. For this dataset, only 7 manual annotations of the cortex were available. As in the ventricle experiments,the atlas was initialized with 3 randomly selected training cases with cortex annotation. The experiment was performed as LOOCV, each time segmenting the cortex in one of the cases where manual annotations of the cortex are available.

Evaluating the influence of parameters. The parameters α, β, γ, κ were determined during initial cross-validation. For this purpose, FSV-ST was performed for 30 iterations on five different test images with GAs equally distributed from 20 to 30 GWs. Table 2 (column A) illustrates the parameter combinations that were evaluated (144 combinations in total). The final constellation with the highest DC on average over all test images was $\alpha = 0.3$, $\beta = 0.5$, $\gamma = 0.1$, $\kappa = 0.1$. This parameter setup was used in all experiments.

To illustrate the mutual dependencies of the parameters, we evaluated the influence of the parameters on the segmentation result in a pair-wise manner. For this, we varied two out of four parameters at a time, and subsequently fixed the others according to the previous test run. For this we performed experiment FSV-ST on a test case at GW 25 for 30 iterations. The parameters α, β, γ, κ were chosen to vary between the values given in Table 2 (column B). The resulting DC for each parameter pair is plotted as matrices in Fig. 6, showing the strong correlation between the parameters.

Table 2
Overview of the value range that was tested for each parameter. In an initial trial, experiment FSV-ST was running for 30 iterations on five different test images, and the overall mean DC was afterwards used to select the optimum parameter constellation. Between 3 and 4 values were tested for each parameter (see column A) in this experiment. Additionally, we evaluated the dependence of the parameters in a pair-wise manner by varying two parameters at a time, while the others were fixed to the previously computed optimum parameter setting. This was computed for a single test case at GW 25. Column B shows the values that were experimented with in this trial.

Parameter	(A) Initial value range	(B) Pair-wise parameter evaluation
α	{0, 0.1, 0.3}	{0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45, 0.5, 0.55}
β	{0.1, 0.3, 0.5}	{0.3, 0.35, 0.4, 0.45, 0.5, 0.55, 0.6, 0.65, 0.7, 0.75}
γ	{−0.3, −0.1, 0, 0.1}	{0, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45}
κ	{−0.3, −0.1, 0, 0.1}	{0, 0.05, 0.1, 0.15, 0.2, 0.25, 0.3, 0.35, 0.4, 0.45}

For the width of the introduced Gaussian age weight, we chose $\sigma = 2$ GW. For the random forest appearance model 32 trees proved sufficient. The random forest classifier was only trained once on the training data during the initialization, and afterwards used to predict the learned labels onto every test case.

Evaluating the estimation of morphological age estimation. To investigate the relationship between nominal age, and the morphological age predicted by the atlas, we performed an experiment using $N = 32$ annotations. The atlas was built on all cases except the test case, and its morphological age was estimated by optimizing Eq. (11) based on this atlas without knowledge of the nominal age of the test case. Optimization was performed in a multi-scale fashion, and we first computed the costs at a two-week time step between the ages 20 and 30. Afterwards, we consider the area around the local optimum for further computation, where we again compute the costs at a smaller age span, etc.

We assessed if the morphological age estimate is stable, even if the atlas training set changes. For this purpose, we selected one test case (in a leave-one-out manner) and divided the remaining cases into two distinct subsets with similar age distribution. Then, we composed 2 independent atlases, one from each subset, and estimated the morphological age of the test case with help of each atlas, again without knowledge of the nominal age of the test case. This results in two independent estimates of the test case's age.

To investigate the role of scale in the age estimation algorithm, we repeated this experiment for data after rigid registration, and after rigid + scale registration. In the latter case, we simply performed rigid registration with scale normalization in order to demonstrate that even if cases are normalized in terms of their size, the spatio-temporal atlas still encodes sufficient information to estimate the gestational age.

Evaluating morphological age characteristics lissencephaly cases. We evaluated if the relationship between morphological age and nominal age exhibits characteristic differences when comparing healthy control subjects and subjects suffering from pathology. We compared the offset of the nominal- and the morphological age in a set of 12 cases suffering from lissencephaly with the offset in a control population.

3.2. Experimental results

Segmentation accuracy. Table 3 depicts average DC and surface distances for all experiments on the rigid D_R as well as non-rigid D_{NR} data set. Starting with the results on segmenting the ventricles in D_R, the spatio-temporal latent atlas (SSV-ST) resulted in a mean/median DC of 0.56/0.58, and the mean E_{surf} of 1.58 mm. FSV-ST achieved on average the highest DC of 0.70 (median 0.73), and the smallest E_{surf} of 0.98 mm. FSV-SNG which is similar to FSV-ST but with an average atlas that does not account for temporal development reached a mean/median DC of 0.62/0.66, and a mean E_{surf} of 1.02 mm. As for the latent atlases, the single (average) atlas of SSV-SNG caused the lowest mean/median DC of 0.46/0.49, and the largest deviation of the segmentation surface from the corresponding ground truth (E_{surf} 2.76 mm on average).

As expected, non-rigid registration D_{NR} improves the results, while keeping the same relative trend among SSV-ST and the

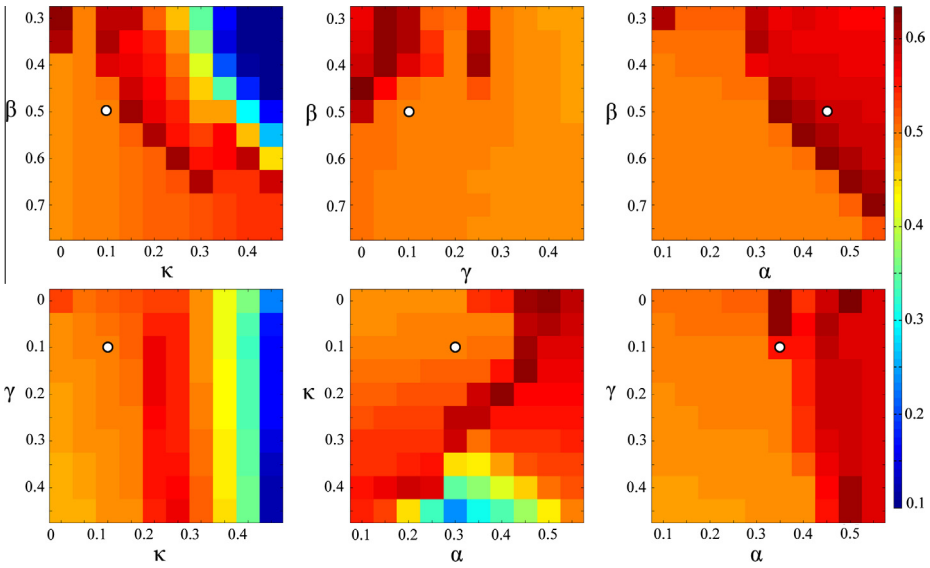

Fig. 6. This example shows the influence of the parameters α, β, γ, κ on the segmentation result of a testcase at GW 25. For each parameter pair, the other two parameters were fixed, and experiment FSV-ST was run for 30 iterations. The resulting DC was stored in each of these matrices. The white dot marks the best overall parameter setting.

Table 3
Results on the accuracy of segmenting the ventricles in both the rigid data set D_R (top), and the non-rigid set D_{NR} (bottom). Leave one out cross-validation was performed for the experimental scenarios fully supervised spatio-temporal and single, and different randomly chosen triplet combinations of annotated training data were used for our proposed method (semi-supervised spatio-temporal) and the latent atlas method (semi-supervised single). For each experiment, the overall mean (median) Dice coefficient (DC), as well as the mean (median) euclidean surface distance (E_{surf} in mm) was computed between the segmentation surface and the ground truth surface, both after initialization and after the final iteration. The results of our proposed method are highlighted in bold font.

	Single atlas	Spatio-temporal atlas
D_R		
DC		
Fully supervised	init: 0.55 ± 0.10 (0.55) final: 0.62 ± 0.17 (0.66)	init: 0.58 ± 0.10 (0.57) final: 0.70 ± 0.08 (0.73)
Semi-supervised	init: 0.38 ± 0.12 (0.40) final: 0.46 ± 0.15 (0.49)	**init: 0.48 ± 0.13 (0.48)** **final: 0.56 ± 0.13 (0.58)**
E_{surf}		
Fully supervised	init: 1.74 ± 0.50 (1.27) final: 1.02 ± 0.55 (0.90)	init: 1.61 ± 0.49 (1.27) final: 0.98 ± 0.46 (0.90)
Semi-supervised	init: 3.77 ± 1.39 (2.45) final: 2.76 ± 1.47 (1.56)	**init: 2.37 ± 1.19 (1.68)** **final: 1.58 ± 0.94 (0.90)**
D_{NR}		
DC		
Fully supervised	init: 0.64 ± 0.10 (0.64) final: 0.56 ± 0.20 (0.61)	init: 0.66 ± 0.10 (0.68) final: 0.77 ± 0.05 (0.77)
Semi-supervised	init: 0.61 ± 0.10 (0.62) final: 0.69 ± 0.08 (0.71)	**init: 0.61 ± 0.10 (0.64)** **final: 0.70 ± 0.07 (0.71)**
E_{surf}		
Fully supervised	init: 1.40 ± 0.37 (1.27) final: 1.01 ± 0.47 (0.90)	init: 1.29 ± 0.34 (1.27) final: 0.69 ± 0.11 (0.90)
Semi-supervised	init: 1.57 ± 0.37 (1.36) final: 1.02 ± 0.27 (0.90)	**init: 1.60 ± 0.40 (1.26)** **final: 0.92 ± 0.25 (0.90)**

alternative approaches. SSV-ST yielded the second highest DC with mean/median values of 0.70/0.71, and a mean E_{surf} of 0.92 mm. Again, FSV-ST resulted in the highest average/median DC of 0.77/0.77, and a mean E_{surf} of 0.69 mm. The other experiments with the single atlases, FSV-SNG and SSV-SNG, achieved mean/median DCs of 0.56/0.61, and 0.69/0.71, respectively. Their surface distances E_{surf} were 1.01 mm and 1.02 mm on average.

Fig. 7 shows the age-specific template drawn from the proposed spatio-temporal atlas for the GWs 20, 22, 24, 26, 28, 30. The first two rows represent the initialization (top row) and the resulting atlas after 40 iterations (middle row) of the proposed spatio-temporal latent atlas (SSV-ST). The color encodes the distance of the surface to the template GW. This indicates how well the template represents the variability in the learning population. In addition, the third row in Fig. 7 illustrates the spatio-temporal atlas from FSV-ST for the same ages. The rapid development of the ventricle over time is clearly visible.

For the segmentation of the cortex, we were able to achieve a mean/median DC between the automatic segmentation and ground-truth annotation of 0.81/0.89, as shown in Table 4.

Morphological age estimation. We evaluated the method to estimate the morphological age on a healthy dataset with annotations

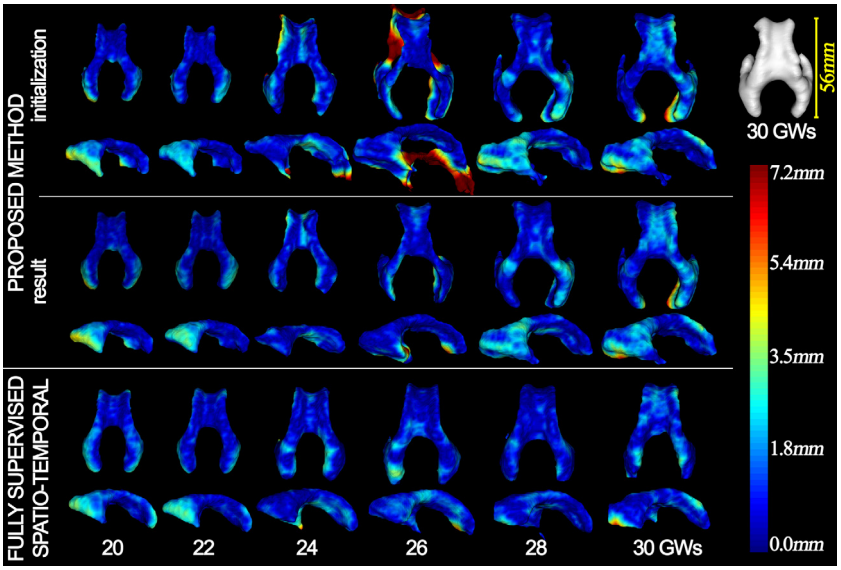

Fig. 7. 3D view of the lateral ventricles from top and right of the spatio-temporal latent atlas for the GWs 20, 22, 24, 26, 28, 30 (per column). The first row depicts the initialization, and the second row the resulting atlas after 40 iterations for our proposed method (SSV-ST). Bottom row: corresponding spatio-temporal atlas being achieved in experiment FSV-ST.

of the ventricles, and additionally, we compared the age estimation of a pathological dataset with annotations of the cortex to a healthy dataset with the same annotated structure.

We report morphological age estimate stability after rigid registration in Fig. 8(a) and after rigid registration + scale normalization in Fig. 8(b). For each case, two age estimates based on independent atlases are plotted as green and blue symbols connected by a line to illustrate the consistency of the morphological age estimate for each case, even if the atlas is based on two independent cohorts. The horizontal axis represents nominal age, the vertical axis indicates estimated morphological age. The average age estimate deviation is ±2.9 days after rigid registration, and ±6.2 days for rigid + scale normalization. The latter means that scale is not encoded in the atlas, and thus does not influence the estimate. In the majority of the cases the direction of the offset of the two morphological age estimates and the nominal age is consistent. This holds for both the atlas the encodes scale, and for the atlas that does not. Additionally, we evaluated the absolute mean (median) deviation from the nominal age of both age estimates. The results of both the rigid and scaled dataset were rather similar: 4.0 (3.5) days for the rigid dataset, and 3.9 (2.7) days for the scaled one.

In a second experiment, we estimated the morphological age of cortex segmentations in a set of 12 pathological cases and compared the deviation from their nominal ages to those of a healthy reference group with 32 cortex annotations. The annotations from the healthy set consisted of 7 manual annotations, as well as 25 segmentations that were retrieved from the proposed latent atlas-based method. The results are visualized in Fig. 9, where (a) shows the nominal age (x-axis) in contrast to the morphological age estimate (y-axis) for both the pathological group (red symbols) and the healthy control (blue symbols), and (b) illustrates the offset of the age estimate from the nominal age. The mean (median) morphological age offset of the pathological group was $-3.0 \pm 3.5(-1.9)$ days, and those of the healthy group $-0.1 \pm 0.8(0.0)$ days. The offset between the morphological age and the nominal age is significantly stronger for lissencephaly cases compared to control cases ($p = 0.014$, student's t-test, normality tested with a Lilliefors test).

Table 4
Evaluation of the cortex segmentation in the non-rigid dataset D_{NR} regarding the Dice coefficient (DC) as well as the mean (median) euclidean surface distance (E_{surf} in mm). We used the a random constellation of training case triplets to initialize the atlas. The cortex was segmented in one out of seven cases where manual annotations were available for evaluation, and the atlas remained untouched after the initialization.

Spatio-temporal atlas-based segmentation
D_{NR}
DC
init: 0.81 ± 0.07 (0.84)
final: 0.89 ± 0.06 (0.89)
E_{surf}
init: 2.33 ± 1.03 (2.06)
final: 1.21 ± 0.84 (0.90)

4. Discussion

This paper describes a method to segment cerebral structures in multiple subjects and to simultaneously learn a spatio-temporal latent atlas of their age dependent development. The latent atlas serves two purposes. Firstly, it is a prior that links segmentations across subjects of different ages, secondly, it provides a means to estimate the *morphological age* based on new observed imaging data.

The atlas as a function of age is necessary to study the substantial, systematic morphological changes of the brain during gestation. A static atlas would not represent the variability in the population with sufficient accuracy, and would neglect the age dependency of shape and appearance. We evaluated four main aspects of the proposed method: (1) Does the spatio-temporal atlas represent the variability, and the temporal development in the study population accurately. (2) Can we use the proposed method to achieve accurate segmentations, even if expert annotations are available only for a minimal number of cases. (3) Is the age-specificity of the atlas sufficient to estimate the gestational age of cases based on their mapping to the atlas. (4) What is the relative effect of the components of the cost function that segments images, and at the same time links the segmentation via the spatio-temporal latent atlas.

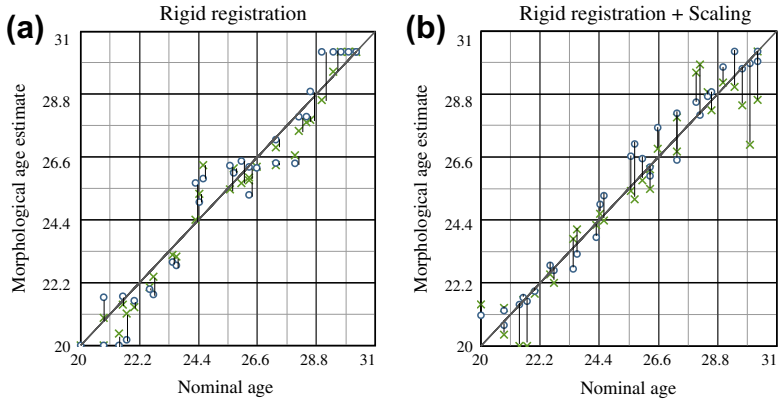

Fig. 8. Consistency of atlas based age re-estimation on segmentations of the ventricles. The nominal age is plotted on the horizontal axis, the morphological age estimate is represented on the vertical axis (both given in GWs). The gray line demonstrates the situation where the morphological age is the same as the nominal age. Green and blue symbols depict the morphological age that is computed based on two atlases learned from disjoint example sets. The corresponding morphological age estimates for a single subject retrieved from two individual atlases are connected by a line. In (a), the images were aligned via rigid registration, and in (b), we added scaling to the registration of (a) in order to achieve the same dimension for each ventricle segmentation. This eliminates the sole dependency of shape size in our age estimation. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

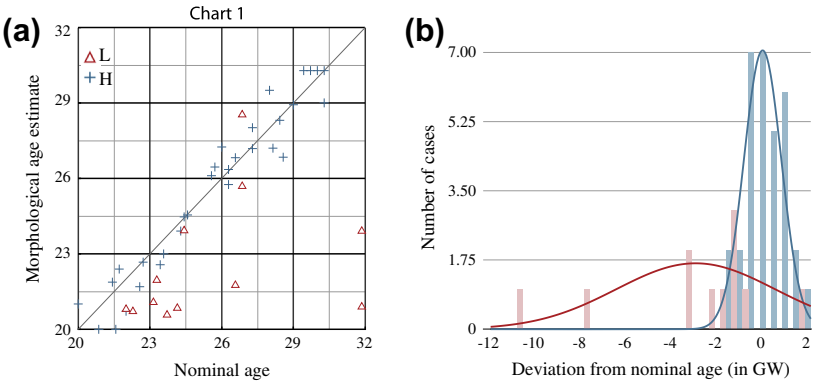

Fig. 9. Age estimation of the patient group diagnosed with lissencephaly (L) in comparison to a healthy control group (H). In (a), the nominal age is plotted on the horizontal axis, the morphological age estimate is represented on the vertical axis (both given in GWs). Blue reflects the morphological age estimation of H, whereas the estimated age of L is plotted in red. The gray line represents the situation where the morphological age is the same as the nominal age. (b) shows the deviation of both patient groups from the gray line. Obviously, cases diagnosed with lissencephaly have a larger offset than healthy cases, and are thus estimated younger than their nominal age. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

4.1. Modeling temporal development across individuals

The spatio-temporal latent atlas captures the development of a structure during gestational aging across individuals. We evaluated the atlas for both rigidly registered data D_R, and non-rigidly registered data D_{NR}. A comparison of the results on D_R and D_{NR} shows that the approach on D_{NR} achieves higher segmentation accuracy, if the latent atlas is used for segmentation. However, then, part of the variability is encoded in the deformation fields, and part in the latent atlas. In the rigid case, D_R, the entire variability of the cohort is encoded in the latent atlas. For the segmentation achieved with the spatio-temporal atlas (FSV-ST), the mean surface error E_{surf} is 1.58 mm in D_R, and 0.92 mm in D_{NR}. As expected, this is better compared to the analogous value for a single age independent average template representing all cases (FSV-SNG), bearing a mean E_{surf} of 2.76 mm in D_{NR}, and 1.02 mm in D_{NR}. While a single latent atlas as in Riklin-Raviv et al. (2010) is well suited to capture variability if no temporal component is present, it does not reflect changes that correlate with specific parameters, such as age. As expected, the accuracy is higher for non-rigidly registered data similar to Habas et al. (2010), compared to rigidly registered data. Non-rigid registration is preferable for segmentation, while the atlas in rigidly registered data encodes the entire developmental information. This suggests that for morphological age estimation or if the atlas serves as basis for statistical observations, it should be either calculated on rigidly aligned data, or should be mapped to this space, after segmentations have been learned in non-rigidly registered imaging data. It is worth noticing that in D_{NR} the distances E_{surf} of the segmentations achieved by the single atlas are closer to the results with the spatio-temporal atlas than in D_R, since the non-rigid registration that leads to D_{NR} partly accounts for the variation among the subjects over time.

Segmentation accuracy depends on data and preprocessing, and thus the comparison in this paper were performed on identically preprocessed data to evaluate the main claims regarding semi-supervised learning. Numbers should be compared only within cohorts and for identical preprocessing. Nevertheless results for the fully annotated case are comparable to published work. In Kuklisova-Murgasova et al. (2011), segmentation accuracy of the cortex in fetal brain MRIs reached a Dice coefficient (DC) of 0.85 ± 0.02. In comparison to this result, we were able to reach a DC of 0.89 ± 0.06 for the leave-one-out experiment on cortex segmentation. Habas et al. (2010) presented results for the segmentation of the ventricles with an accuracy of 0.88 ± 0.05 (DC), which is higher than our result of 0.70 ± 0.08, or 0.77 ± 0.05 respectively.

By evaluating the difference between each case used for atlas building and the corresponding age dependent atlas template in Fig. 7 we learn how well the atlas represents the variability among cases of the same age. Fig. 7 illustrates how this deviation is distributed across the surface of the learned structure in D_R. The strong initial deviation of the template at GW 26 is due to a bad initialization, but is reduced substantially during optimization. In general, there is a consistent location-specific distribution of the deviation between population and template, mirroring the variability across the population for specific parts of the structure.

There are two reasons for this observation. First, it is possible that during a specific gestational period variability across the population is particularly high in certain cerebral areas. Second, changes in developmental velocity might vary across the population, i.e., even when comparing individuals at the same nominal age we might observe them at slightly different stages of their development. This is particularly relevant when aiming to understand disease cohorts, since some diseases correlate with malformations of cortical development (Barkovich et al., 2001a,b; Raybaud and Widjaja, 2011) that might partly be explained as developmental delay.

Fig. 2 sheds some light on these two aspects. The uncertainty of the shape encoded in the spatio-temporal atlas, expressed as the gradient of the function Eq. (8), has two components. The gradient in the three spatial dimensions Fig. 2(a) represents variability in shape independent from age, or the specificity of the atlas. Red areas indicate high variability. The temporal gradient shown in Fig. 2(b) captures the longitudinal component of the atlas. For each point on the surface, red indicates strong growth, while blue indicates shrinkage. We can see how the occipital horns (*cornua occipitalia*) grow from an initially rounded shape to a sharp structure, and move outwards, in correspondence with the overall growth. The *cavum septi pellucidi* – the central part connecting the lateral ventricles – exhibits occipital growth, even after GW 20, that might be connected with a dorsal expansion of the *corpus callosum*. The *cavum septi pellucidi* broadens, and becomes thinner at the same time indicated by the red exterior parts of the lateral ventricles, and the blue interior part, indicating thinning. The ventricle grows from a length of 43 mm at GW 20 to 56 mm at GW 30, the width increases from 32 mm to 40 mm during the same period. While we cannot infer decoupled processes from the atlas, we can make two interesting observations. In some regions such as the central posterior part in GW 26–28, there is a correlation between rapid development and high spatial uncertainty. Rapid development might increase variability introduced by small speed differences across subjects. However, at other locations, e.g., the central anterior surface in GW 20–22, there is a rapid development, while at the same time the spatial variability is relatively low indicating a high degree of synchronization across the subjects. To summarize, spatial, and temporal variability are at least partially independent, and the amount of synchronization of growth among subjects varies across the ventricle surface. Related work regarding developmental delay detection was described in the context of archeological findings in Durrleman et al. (2009).

4.2. Semi-supervised spatio-temporal atlas learning and segmentation

If only part of the image data is annotated we have to perform semi-supervised learning of the atlas and segmentations. The result of SSV-ST is the corresponding central result of the proposed method, with about 10% of the data were annotated. During the optimization process, both atlas, and segmentations of all cases are learned. Table 3 shows the resulting segmentation error. In comparison to FSV-ST, accuracy decreases in SSV-ST, but overall error is still acceptable (DC 0.70 in D_{NR}, and a mean deviation of segmentation from ground truth surface of 0.92 mm). Again, the spatio-temporal latent atlas achieves higher accuracy compared with a single latent atlas (SSV-SNG). The result indicates that even if annotation is available only for an extremely small fraction of the data, spatio-temporal atlases can be learned, and outperform static atlases. This is relevant if including large amounts of data is necessary to accurately capture the variability in a population. Results of FSV-ST are comparable to work in Habas et al. (2010), that describes the development from GW 20 to GW 24.

4.3. Age specificity of the atlas, and its implications

The atlas can be used to estimate the gestational age based on morphology. Morphological age estimate and the nominal age typically agree, with a mean deviation of 4 days. While scale contributes substantially to the estimate, even atlases representing only non-rigid components of the development can predict the gestational age with comparable deviation from the nominal age. Are these deviation due to algorithmic limitations, or to actual developmental delay or advance? The relatively high repeatability when estimating the morphological age with two independent atlases (mean difference ±2.9 days for rigidly registered data) suggests that the this deviation is an valid observation.

4.4. The morphological age and disease

Measuring this offset between morphological age and nominal age is particularly relevant when studying disease since diseases can cause characteristic developmental delay (Ghai et al., 2006; von Rhein et al., 2011). This makes it clinically relevant to understand the developmental relationship of individual cases compared to the overall population, and indicates that age estimation based on morphology might be a reliable marker, when studying effects of disease. The offset between morphological age and nominal age was significantly stronger (significant developmental delay) for a pilot disease cohort, when compared to control subjects. This indicates its potential as a disease marker.

4.5. Contribution of the cost function components

In order to evaluate the influence of each parameter α (spatial weight), β (image weight), γ (length term weight), κ (balloon force weight) on the segmentation result, we performed pair-wise parameter variation, i.e. for each experiment, we fixed two parameters while the remaining parameter pair was varied within a certain interval (Table 2 (column B)). The resulting segmentation accuracy shown in Fig. 6 indicates that for all parameters, there is typically a relatively well behaved decrease of accuracy from the maximum. Furthermore they show certain relationships among the parameters. For example, it is clearly visible that the proposed method works best if α and β are assigned a relatively high value as compared to γ and δ. This means that during optimization, most information emerges from the image term, followed by the spatio-temporal atlas. Additionally, a linear correlation between parameter pairs is visible corresponding to the strong ridges of values that describes the best combination along which the parameters should be chosen. In the final experiments, the weighting of the image term, β, is assigned the highest value of 0.5, and the spatio-temporal atlas weight, α, the second highest of 0.3. γ and κ are equally set to 0.1.

4.6. Limitations

This study has limitations. The atlas building was performed on volume data reconstructed from three stacks of orthogonal slices (coronal, sagittal, axial). Therefore, artifacts introduced by the reconstruction algorithm are included in the results. It is worth

mentioning that there are alternative reconstruction, and super-resolution algorithms such as Gholipour et al. (2010); Kim et al. (2010); Kuklisova-Murgasova et al. (2011). The current formulation does not decouple the variability in the data due to developmental shifts, and actual variability. Instead it represents variability across the population and the estimated local developmental velocity in the cohort.

5. Conclusion

We propose a spatio-temporal latent atlas for the semi-supervised segmentation of cerebral structures during early brain development. Our dataset consists of MRIs from different subjects at different time points, from which we select a small subset of annotated examples for initialization. During optimization, we learn segmentations for the ventricles and cortex from a large number of non-annotated images, as well as an atlas that captures the variation among subjects and their joint development over time. We use kernel regression to build the spatio-temporal latent atlas continuously over the gestational period of 20–30 weeks. Thus, we are able to interpolate the atlas for any age within the captured time period. The spatio-temporal latent atlas requires only minimal user input, and is aimed at learning models from large numbers of examples. In addition, the atlas allows for the estimation of a subject's age based on morphology, suggesting its use when quantifying disease effects. Initial results demonstrate that the offset between morphological age and nominal age, can be a disease characteristic. Future work will focus on exploiting the learned neuro-developmental atlases for growth modeling and the assessment of pathologies.

Acknowledgements

E. Dittrich is a recipient of a DOC-fFORTE-fellowship of the Austrian Academy of Sciences. This work was partly funded by the European Union (257528, KHRESMOI, 330003 FABRIC), the Austrian Science Fund (P 22578-B19, PULMARCH), and the OeNB (13497, AORTAMOTION; 14812, FETALMORPHO).

References

Aljabar, P., Wolz, R., Srinivasan, L., Counsell, S., Rutherford, M., Edwards, A., Hajnal, J., Rueckert, D., 2011. A combined manifold learning analysis of shape and appearance to characterize neonatal brain development. IEEE Transactions on Medical Imaging 30, 2072–2086.

Ashburner, J., 2007. A fast diffeomorphic image registration algorithm. NeuroImage 38, 95–113.

Ashburner, J., Friston, K.J., 2000. Voxel-based morphometry–the methods. NeuroImage 11, 805–821.

Barkovich, A., Kuzniecky, R., Dobyns, W., 2001a. Radiologic classification of malformations of cortical development. Current Opinion in Neurology 14, 145–149.

Barkovich, A., Kuzniecky, R., Jackson, G., Guerrini, R., Dobyns, W., 2001b. Classification system for malformations of cortical development. Neurology 57, 2168–2178.

Breiman, L., 2001. Random forests. Machine Learning 45, 5–32.

Brodmann, K., 1909. Vergleichende Lokalisationslehre der Groszhirnrinde 1.

Chan, T., Vese, L., 2001. Active contours without edges. IEEE Transactions on Image Processing 10, 266–277.

Cohen, L., Cohen, I., 1993. Finite-element methods for active contour models and balloons for 2-D and 3-D images. Pattern Analysis and Machine Intelligence 15, 1131–1147.

Davis, B., Fletcher, P., Bullitt, E., Joshi, S., 2007. Population shape regression from random design data. ICCV 10, 1–7.

Dice, L., 1945. Measures of the amount of ecological association between species. Ecology 26, 297–302.

Dittrich, E., Riklin-Raviv, T., Kasprian, G., Brugger, P., Prayer, D., Langs, G., 2011. Learning a spatio-temporal latent atlas for fetal brain segmentation. Proceedings of the MICCAI Workshop on Image Analysis of Human Brain Development, 9–16.

Durrleman, S., Pennec, X., Trouve, A., Ayache, N., Braga, J., 2010. Comparison of the endocast growth of chimpanzees and bonobos via temporal regression and spatiotemporal registration. MICCAI 2010 Workshop STIA'10, pp. 1–12.

Durrleman, S., Pennec, X., Trouvé, A., Gerig, G., Ayache, N., 2009. Spatiotemporal atlas estimation for developmental delay detection in longitudinal datasets. Medical Image Computing and Computer-Assisted Intervention 12, 297–304.

Evans, A., Collins, D., Mills, S., Brown, E., Kelly, R., Peters, T., 1993. 3D statistical neuroanatomical models from 305 MRI volumes. Nuclear Science Symposium and Medical Imaging Conference 3, 1813–1817.

Fischl, B., Salat, D., Busa, E., Albert, M., Dieterich, M., Haselgrove, C., van der Kouwe, A., Killiany, R., Kennedy, D., Klaveness, S., 2002. Whole brain segmentation: automated labeling of neuroanatomical structures in the human brain. Neuron 33, 341–355.

Fischl, B., Salat, D., van der Kouwe, A., Makris, N., Ségonne, F., Quinn, B., Dale, A., 2004. Sequence-independent segmentation of magnetic resonance images. NeuroImage 23, S69–S84.

Fishbaugh, J., Prastawa, M., Durrleman, S., Piven, J., Gerig, G., 2012. Analysis of longitudinal shape variability via subject specific growth modeling. Medical Image Computing and Computer-Assisted Intervention 731, 738.

Garel, C., 2004. MRI of the Fetal Brain: Normal Development and Cerebral Pathologies 1.

Geurts, P., Ernst, D., Wehenkel, L., 2006. Extremely randomized trees. Machine Learning 63, 3–42.

Ghai, S., Fong, K., Toi, A., Chitayat, D., Pantazi, S., Blaser, S., 2006. Prenatal US and MR imaging findings of lissencephaly: review of fetal cerebral sulcal development. RadioGraphics 26, 389–405.

Gholipour, A., Akhondi-Asl, A., Estroff, J.A., Warfield, S.K., 2012. Multi-atlas multi-shape segmentation of fetal brain MRI for volumetric and morphometric analysis of ventriculomegaly. NeuroImage 60, 1819–1831.

Gholipour, A., Estroff, J., Warfield, S., 2010. Robust super-resolution volume reconstruction from slice acquisitions: application to fetal brain MRI. IEEE Transactions on Medical Imaging 29, 1739–1757.

Habas, P.A., Kim, K., Corbett-Detig, J.M., Rousseau, F., Glenn, O.A., Barkovich, A.J., Studholme, C., 2010. A spatiotemporal atlas of MR intensity, tissue probability and shape of the fetal brain with application to segmentation. NeuroImage 53, 460–470.

Hausdorff, F., 1914. Grundzuege der Mengenlehre 18.

Karas, G., Burton, E., Rombouts, S., van Schijndel, R., O'Brien, J., Scheltens, P., McKeith, I., Williams, D., Ballard, C., Barkhof, F., 2003. A comprehensive study of gray matter loss in patients with Alzheimer's disease using optimized voxel-based morphometry. NeuroImage 18, 895–907.

Kim, K., Habas, P., Rousseau, F., Glenn, O., Barkovich, A., Studholme, C., 2010. Intersection based motion correction of multislice MRI for 3-D in utero fetal brain image formation. IEEE Transactions on Medical Imaging 29, 146–158.

Klein, S., Staring, M., Murphy, K., Viergever, M., Pluim, J., 2010. Elastix: a toolbox for intensity-based medical image registration. IEEE Transactions on Medical Imaging 29, 196–205.

Kuklisova-Murgasova, M., Aljabar, P., Srinivasan, L., Counsell, S., Doria, V., Serag, A., Gousias, I., Boardman, J., Rutherford, M., Edwards, A., Hajnal, J., Rueckert, D., 2010. A dynamic 4D probabilistic atlas of the developing brain. NeuroImage 54, 2750–2763.

Kuklisova-Murgasova, M., Quagheberu, G., Hajnal, J., Schnabel, J., 2011. Unified framework for super-resolution reconstruction of 3D fetal brain MR images from 2D slices with intensity correction and outlier rejection. Proceedings of the MICCAI Workshop on Image Analysis of Human Brain Development, 48–55.

Mazziotta, J., Toga, A., Evans, A., Fox, P., Lancaster, J., 1995. A Probabilistic atlas of the human brain: theory and rationale for its development. NeuroImage 2, 89–101.

Nadaraya, E., 1964. On estimating regression. Theory of Probability and Its Applications 9, 141–142.

Osher, S., Sethian, J., 1988. Fronts propagating with curvature dependent speed: algorithms based on Hamilton–Jacobi formulations. Journal of Computational Physics 79, 12–49.

Pohl, K., Fisher, J., Bouix, S., Shenton, M., McCarley, R., Grimson, W., Kikinis, R., Wells, W., 2007. Using the logarithm of odds to define a vector space on probabilistic atlases. Medical Image Analysis 11, 465–477.

Rajagopalan, V., Scott, J., Habas, P., Kim, K., Rousseau, F., Glenn, O., Barkovich, A., Studholme, C., 2011. Spatiotemporal morphometry of adjacent tissue layers with application to the study of sulcal formation. Medical Image Computing and Computer-Assisted Intervention–MICCAI 2011, 476–483.

Raybaud, C., Widjaja, E., 2011. Development and dysgenesis of the cerebral cortex: malformations of cortical development. Neuroimaging Clinics of North America 21, 483–543.

von Rhein, M., Scheer, I., Loenneker, T., Huber, R., Knirsch, W., Latal, B., 2011. Structural brain lesions in adolescents with congenital heart disease. The Journal of Pediatrics 158, 984–989.

Riklin-Raviv, T., Leemput, K.V., Menze, B., III, W.W., Golland, P., 2010. Segmentation of image ensembles via latent atlases. Medical Image Analysis 14, 654–665.

Riklin-Raviv, T., Menze, B., Van-Leemput, K., Stieltjes, B., Weber, M., Ayache, N., III, W., Golland, P., 2009. Joint segmentation via patient-specific latent anatomy model. Proceedings of the MICCAI Workshop on Probabilistic Models for Medical Image Analysis, 244–255.

Rousseau, F., Oubel, E., Pontabry, J., Studholme, C., Koob, M., Dietemann, J., 2011. An open-source toolkit for fetal brain MR image processing. Proceedings of the MICCAI Workshop on Image Analysis of Human Brain Development 1, 96–103.

Schöpf, V., Kasprian, G., Brugger, P.C., Prayer, D., 2012a. Watching the fetal brain at 'rest'. International Journal of Developmental Neuroscience 30, 11–17.
Schöpf, V., Kasprian, G., Schwindt, J., Kollndorfer, K., Prayer, D., 2012b. Visualization of resting-state networks in utero. Ultrasound in Obstetrics & Gynecology 39, 487–488.
Serag, A., Aljabar, P., Ball, G., Counsell, S.J., Boardman, J.P., Rutherford, M.A., Edwards, A.D., Hajnal, J.V., Rueckert, D., 2012. Construction of a consistent high-definition spatio-temporal atlas of the developing brain using adaptive kernel regression. NeuroImage 59, 2255–2265.
Smith, S., Jenkinson, M., Woolrich, M., Beckmann, C., Behrens, T., Johansen-Berg, H., Bannister, P., Luca, M.D., Drobnjak, I., Flitney, D., Niazy, R., Saunders, J., Vickers, J., Zhang, Y., Stefano, N.D., Brady, M., Matthews, P., 2004. Advances in functional and structural MR image analysis and implementation as FSL. NeuroImage 23, S208–S219.
Talairach, J., Tournoux, P., 1988. Co-planar Stereotaxic Atlas of the Human Brain: 3-D Proportional System: An Approach to Cerebral Imaging.
Watson, G., 1964. Smooth regression analysis. Sankhya Series A 26, 359–372.
Woolrich, M.W., Jbabdi, S., Patenaude, B., Chappell, M., Makni, S., Behrens, T., Beckmann, C., Jenkinson, M., Smith, S.M., 2009. Bayesian analysis of neuroimaging data in FSL. NeuroImage 45, S173–S186.
Yushkevich, P., Piven, J., Hazlett, H., Smith, R., Ho, S., Gee, J., Gerig, G., 2006. User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. NeuroImage 31, 1116–1128.