Fetal MRI by Robust Deep Generative Prior Reconstruction and Diffeomorphic Registration

Lucilio Cordero-Grande[ID], Juan Enrique Ortuño-Fisac[ID], Alejandra Aguado del Hoyo[ID], Alena Uus,
Maria Deprez[ID], Andrés Santos[ID], *Senior Member, IEEE*, Joseph V. Hajnal[ID],
and María J. Ledesma-Carbayo[ID], *Senior Member, IEEE*

Abstract— **Magnetic resonance imaging of whole fetal body and placenta is limited by different sources of motion affecting the womb. Usual scanning techniques employ single-shot multi-slice sequences where anatomical information in different slices may be subject to different deformations, contrast variations or artifacts. Volumetric reconstruction formulations have been proposed to correct for these factors, but they must accommodate a non-homogeneous and non-isotropic sampling, so regularization becomes necessary. Thus, in this paper we propose a deep generative prior for robust volumetric reconstructions integrated with a diffeomorphic volume to slice registration method. Experiments are performed to validate our con- tributions and compare with methods in the literature in a cohort of 72 fetal datasets in the range of 20-36 weeks gestational age. Quantitative as well as radiological assess- ment suggest improved image quality and more accurate prediction of gestational age at scan is obtained when com- paring to state of the art reconstruction methods. In addi- tion, gestational age prediction results from our volumetric reconstructions are competitive with existing brain-based approaches, with boosted accuracy when integrating infor- mation of organs other than the brain. Namely, a mean absolute error of 0.618 weeks ($R^2 = 0.958$) is achieved when combining fetal brain and trunk information.**

Lucilio Cordero-Grande is with the Biomedical Image Technologies, ETSI Telecomunicación, Universidad Politécnica de Madrid, 28040 Madrid, Spain, also with CIBER-BBN, ISCIII, 28029 Madrid, Spain, and also with the Centre for the Developing Brain and Biomedical Engineering Department, School of Biomedical Engineering and Imaging Sciences, King's College London, King's Health Partners, St Thomas' Hospital, SE1 7EH London, U.K. (e-mail: lucilio.cordero@upm.es).

Juan Enrique Ortuño-Fisac, Andrés Santos, and María J. Ledesma-Carbayo are with the Biomedical Image Technologies, ETSI Teleco-municación, Universidad Politécnica de Madrid, 28040 Madrid, Spain, and also with CIBER-BBN, ISCIII, 28029 Madrid, Spain (e-mail: juanen@die.upm.es; andres@die.upm.es; mj.ledesma@upm.es).

Alejandra Aguado del Hoyo is with the Department of Pediatric Radiol-ogy, Mother and Child Hospital Gregorio Marañón, 28009 Madrid, Spain (e-mail: aaguadodelhoyo@hotmail.com).

Alena Uus, Maria Deprez, and Joseph V. Hajnal are with the Cen-tre for the Developing Brain and Biomedical Engineering Department, School of Biomedical Engineering and Imaging Sciences, King's Col-lege London, King's Health Partners, St Thomas' Hospital, SE1 7EH London, U.K. (e-mail: alena.uus@kcl.ac.uk; maria.deprez@kcl.ac.uk; jo.hajnal@kcl.ac.uk).

I. INTRODUCTION

MAGNETIC Resonance Imaging (MRI) is indicated when both Central Nervous System (CNS) and non CNS fetal anomalies are suspected on ultrasound [1], [2]. When compared to ultrasound, MRI is a unique instrumental technique for studying fetal development due to enlarged Field Of View (FOV), superior soft tissue contrast, and lack of shadowing. Basic examination protocols involve collecting T_2-weighted slices along the main axes of the fetal organs. In the case of brain imaging, aspirations for more quantita-tive imaging motivated the development of Slice to Volume (SV) reconstruction techniques aiming to obtain a volumetric representation of the fetal brain from the set of acquired slices [3], [4]. If SV reconstruction is available, acquisitions may be redesigned for collecting the most diverse and efficient set of orientations considering both scanning limitations and reconstruction properties. Free reformatting of the imaging plane may be even more important for non-brain applications, where, due to motion, limited resolution, Signal to Noise Ratio (SNR), and scanning time, obtaining the principal axes of various target organs while planning the scans may be difficult or infeasible. However, 3D reconstruction of the whole uterus faces an increased complexity due to different non-rigid sources of motion, consequently rendering rigid motion models [3], [4], [5], [6], [7], [8] ill suited.

Despite non-rigid motion transformation models for SV registration problems have been proposed some time ago [9],

the predominance of fetal brain applications has delayed their incorporation to whole body fetal image reconstruction algorithms. After some preliminary works applying rigid correction models outside the brain, [10] proposed a method using patch-based rigid registration to approximate non-rigid deformations. Although this approach may benefit from a flexible motion model, continuity, smoothness and invertibility of motion cannot be guaranteed as no mechanism is provided for assembling per-patch estimates. Improved performance was shown in [11] when using deformable registration based on free-form deformations on a hierarchical B-Spline grid, reconstruction based on super-resolved weighted Gaussian interpolation, bias correction, and global and local outlier rejection based on normalized cross correlation and structural similarity indexes respectively. In this case, a continuous and smooth motion model is proposed but invertibility is not guaranteed, which could compromise the physical plausibility of the estimated motion fields. In addition, the regularization scheme is based on an edge-preserving term relying on a piecewise constant image assumption, which is prone to spurious boundaries or non-natural image texture due to noise amplification effect or when this assumption is violated. As reported in [9], other explored alternatives for deformable motion models in SV registration include usage of thin plate splines and finite element meshes.

On the other hand, application of Deep Learning (DL) methodologies in the orbit of fetal SV reconstruction includes methods for rigid alignment of slices to brain templates [5], [6], rigid motion tracking [8], automatic localization of the fetal brain [7], or image quality assessment [12]. Distinctly, in this work we focus on the integration of DL architectures within the reconstruction formulation. Most efforts on DL for inverse problems have focused on learning a mapping between an approximate inverse of a fully characterized measurement operator and a Ground Truth (GT) reconstruction, which is used at test time to mitigate spurious residuals in the reconstruction, typically by unrolled schemes [13]. However, generation of GT reconstructions is problematic in our application as it would involve the acquisition of oversampled datasets in a sensible population. In addition, as the measurements are affected by motion, learning may be biased by the reconstruction method employed to build the training data.

Difficulties with the GT are also an issue for validation, differences between reconstruction methods are often subtle or difficult to summarize, and improved reconstructions may not necessarily impact a particular clinical application. For these reasons, we have combined generic measures of data quality and radiological scoring with a task-based validation strategy based on assessing the Gestational Age (GA) prediction performance. This is identified as a clinically relevant task because GA knowledge is critical for fetal development characterization from imaging and errors or uncertainties in GA predictions could be indicative of developmental abnormalities [14].

In this work we introduce the Robust Generative Diffeomorphic Slice to Volume Reconstruction (RGDSVR) method which is devised for multi-slice MRI scans with data acquired along different slice orientations and tested for a single-shot Fast Spin Echo (FSE) T_2-weighted sequence, as typically used in fetal MRI. Our contributions include an efficient version of the Large Deformation Diffeomorphic Metric Mapping (LDDMM) [15] framework adapted to the computational requirements of joint deformable registration and reconstruction problems which guarantees smooth and invertible motion fields. In addition, we propose a robust explicit inverse formulation of the reconstruction with regularization based on an untrained generative model able to preserve both the edges of the images and structured image statistics, the so-called Deep Decoder (DD) [16]. Finally, we show that free reformatting of whole body fetal SV reconstructions can be leveraged for accurate estimates of GA at scan. The source code and exemplary data required to reproduce the main results of the paper is made available at https://github.com/lcorgra/RGDSVR/releases/tag/1.0.0.

II. METHODS

Common structural fetal MRI protocols are based on the acquisition of a series of stacks of slices along different orientations (see Fig. 1). Thus, we start from a data array $\mathbf{y} = (y_{m_1^l m_2^l m_3^l})$ encompassing several stacks $1 \leq l \leq N_l$, where $m_1^l m_2^l m_3^l$ indexes a voxel in stack l, respectively along the readout, phase encode and slice directions, and N_l denotes the total number of stacks in the acquisition. Then, we formulate the reconstruction problem as the recovery of a volume $\mathbf{x}$ from $\mathbf{y}$ via

$$(\hat{\mathbf{x}}, \hat{\boldsymbol{\phi}}, \hat{\boldsymbol{\theta}}) = \underset{\mathbf{x}, \boldsymbol{\phi}, \boldsymbol{\theta}}{\operatorname{argmin}} f(\mathbf{A}\boldsymbol{\Phi}^{\boldsymbol{\phi}}\mathbf{x} - \mathbf{y}) + g(\mathbf{x} - \boldsymbol{\Theta}^{\boldsymbol{\theta}}\mathbf{z}) + h(\boldsymbol{\phi}), \quad (1)$$

with f a robust loss function, g a reconstruction regularizer, and h a motion regularizer. In this formulation, we model the image formation by a measurement operator $\mathbf{A}$, the fetal motion by a diffeomorphic image warping operator $\boldsymbol{\Phi}^{\boldsymbol{\phi}}$ with deformation described by the parameter vector $\boldsymbol{\phi}$, and use a generative DD network operator $\boldsymbol{\Theta}^{\boldsymbol{\theta}}$ with learnable parameters $\boldsymbol{\theta}$ and fixed random input $\mathbf{z}$ for regularization.

In § II-A we describe the robust loss function f, in § II-B the measurement operator $\mathbf{A}$, in § II-C the temporal structure of the acquisition, in § II-D the registration methodology to obtain $\boldsymbol{\Phi}^{\boldsymbol{\phi}}$, in § II-E the deep generative prior $\boldsymbol{\Theta}^{\boldsymbol{\theta}}\mathbf{z}$ used as a regularizer, in § II-F the fetal MRI cohort used to test our reconstructions, in § II-G the adopted reconstruction implementation, and in § II-H the GA estimation procedure.

A. Robust Reconstruction

The residuals of the reconstruction are denoted by

$$\mathbf{r} = \mathbf{A}\boldsymbol{\Phi}\mathbf{x} - \mathbf{y} = (r_{m_1^l m_2^l m_3^l l}) \quad (2)$$

The corresponding squared residuals map is denoted by $\mathbf{r}^2 = (r^2_{m_1^l m_2^l m_3^l l})$. We use a smooth version of the Welsch metric due to its strong robustness to outliers [17]:

$$f(\mathbf{r}) = \frac{k^2 \sigma^2}{2} \left\| 1 - \exp\left(-\frac{\mathcal{G}\mathbf{r}^2}{k^2 \sigma^2}\right) \right\|_1, \quad (3)$$
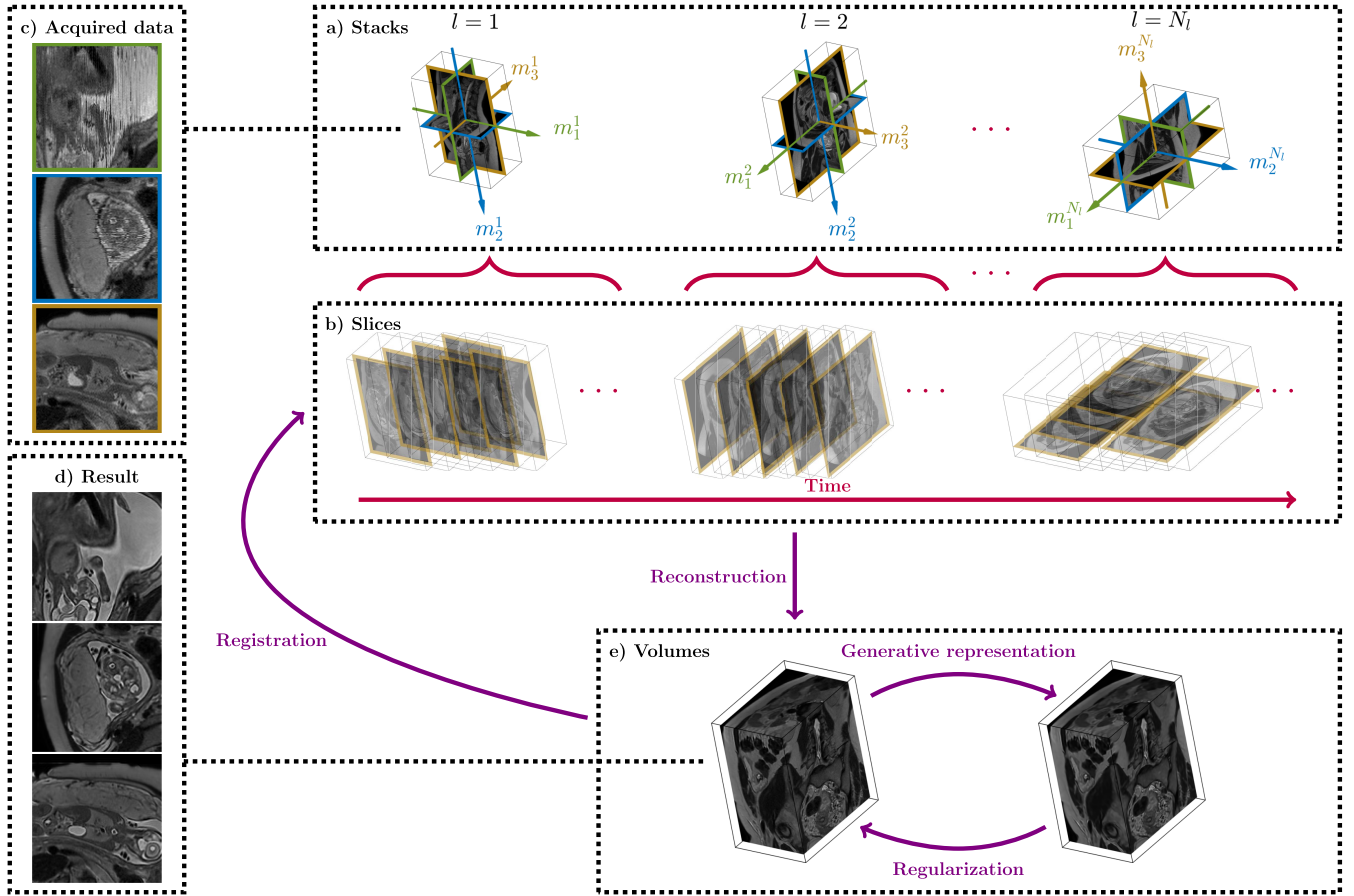
Fig. 1.   Overview of SV reconstruction and registration for whole body fetal imaging. **a)** Stacks of slices are acquired along different orientations, here with axial, sagittal and coronal examples. Corresponding readout, phase encoding and slice axes and planes are color coded in green, blue and yellow, respectively. **b)** Slices collection follows different interleaving configurations in time. **c)** Acquired data may include corrupted slices and appear discontinuous in the slice direction due to motion. **d)** Volumetric reconstructions achieving a dense representation of the fetal anatomy can be obtained by accounting for these measurement factors. **e)** For this sake, the algorithm alternates between robust motion compensated SV reconstruction, multi-scale registration of reconstructed volume to acquired slices for motion refinement, projection of the reconstruction solution to a natural image generative representation, and regularization of the reconstruction using this projection.

with $\mathcal{G}$ an in-plane Gaussian smoothing operator to weight the contribution of the observation in a voxel according to the residuals in its neighborhood, σ a robust estimation of scale, $\sigma^2 = 2.1981 \, \text{median}(\mathcal{G}\mathbf{r}^2)$ [17], and k a tuning constant.

We can solve for $\mathbf{x}$ using Iteratively ReWeighted Least Squares (IRWLS) [17], [18]. Defining the encoding operator as $\mathbf{E} = \mathbf{A}\boldsymbol{\Phi}$, the update at iteration $i+1$, which can be computed via conjugate gradient, is

$$\hat{\mathbf{x}}_{(i+1)} = (\mathbf{E}^T \mathbf{W}_{(i)} \mathbf{E})^{-1} \mathbf{E}^T \mathbf{W}_{(i)} \mathbf{y}, \qquad (4)$$

with $\mathbf{W}_{(i)}$ a diagonal matrix of weights. The diagonal entries for the weights, $\mathbf{w}_{(i)}$, are obtained from the squared residuals and the updated scale at iteration i, $\mathbf{r}_{(i)}^2$ and $\sigma_{(i)}^2$ as [17], [18]:

$$\mathbf{w}_{(i)} = \mathcal{G} \exp\left(-\frac{\mathcal{G}\mathbf{r}_{(i)}^2}{k^2 \sigma_{(i)}^2}\right). \qquad (5)$$

To improve IRWLS convergence we use homotopy continuation [19]. The computed weights are modified for reconstruction according to:

$$\mathbf{W}_{(i)} = \text{diag}(\mathbf{w}_{(i)})^{\tau_{(i)}},$$
$$\tau_{(0)} = 0, \quad \tau_{(i+1)} = \min((\tau_{(i)} + 1)/N_i, 1), \qquad (6)$$

so an Ordinary Least Squares (OLS) problem is solved in the first iteration ($\tau = 0$, $\mathbf{W} = \mathbf{I}$, the identity matrix) and the target formulation is gradually reached after N_i iterations ($\tau = 1$). In Fig. 2 we show an example including a sampled slice, residuals at first and last iterations of the reconstruction and corresponding weights.

B. Measurement Operator

The measurement operator $\mathbf{A}$, is defined for each stack l as $\mathbf{A}_l = \mathcal{D}_l \mathbf{T}_l$, where $\mathbf{T}_l$ is a rigid transformation accounting for the orientation of the stack l and $\mathcal{D}_l$ is the slice sampling operator. We model blurring and discretization in $\mathcal{D}_l$ using a Gaussian slice profile and according to the slice thickness and slice separation of stack l [20]. Modeling and estimation of image inhomogeneities had a small impact in the tested data, so it is left out of this manuscript. However, inhomogeneities may become significant at higher field strengths.

C. Temporal Structure of the Scan

Multi-slice scans are typically collected in a non-sequential manner where consecutively acquired slices are located distant
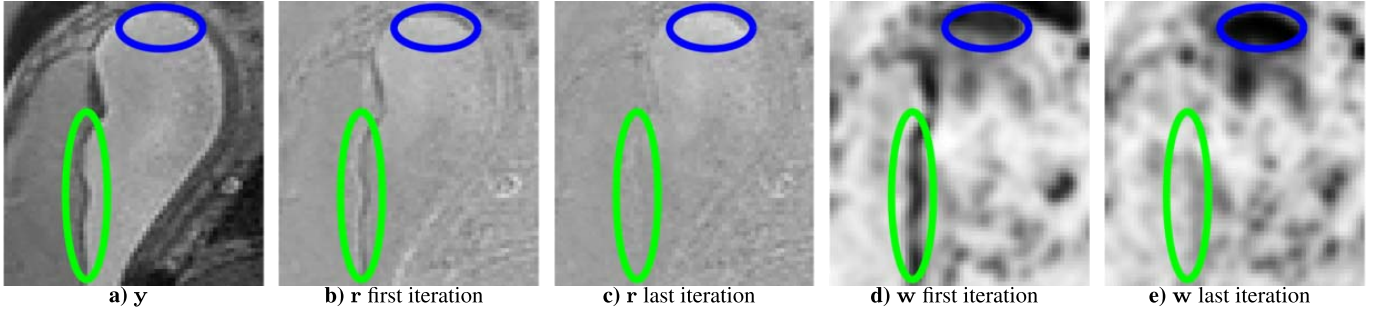
Fig. 2. **a)** Measured data **y**, residuals **r** at **b)** first and **c)** last iterations of the reconstruction, and weights **w** at **d)** first and **e)** last iterations of the reconstruction. The green ellipse encloses an area around the the placenta and amniotic sac boundary with high residuals when starting the reconstruction –see **b)**–. These become smaller at last iteration –see **c)**– due to corrected motion, which increases the reliability of the data in that area, as shown by the corresponding weights. The blue ellipse encloses an area where flow artifacts cause a relative enhancement of the amniotic fluid so, driven by anomalously high residuals, weights are kept low in the last iteration –see **e)**–.
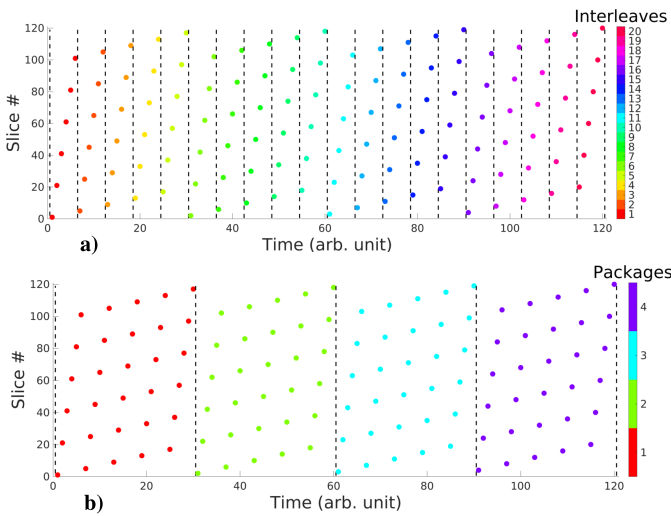

Fig. 3. Slice acquisition order for an exemplary stack highlighting **a)** interleave and **b)** package structure. Arbitrary units (arb. unit) in time refer to the moments the slices are acquired.

to each other. An example of the particular slice order used in a given stack of the acquisition is shown in Fig. 3a. We define an interleave as a series of slices acquired within a single FOV sweep, so in this example we have 20 interleaves. As illustrated in Fig. 3b, we can also group our slices into so-called packages, constructed by considering the gap between the first slices in two consecutive sweeps, with 4 packages per stack in our sequence.

For motion estimation we define an operator $\mathbf{P}_s^{(j)}$, $1 \leq s \leq N_s^j$, to extract the slices associated to the motion state s. Motion states are constructed to correspond to a given stack, package or interleave depending on the multi-scale motion estimation level $j \in \{\texttt{STACK}, \texttt{PACKAGE}, \texttt{INTERLEAVE}\}$ at which we are operating, with N_s^j the number of states at level j. We start by estimating a deformation per stack. Then, we propagate these estimates as starting transformations for estimations at the package level. Finally, we propagate the latter to estimate a deformation per-interleave, which serves to accommodate different slice deformations because subsequent slices are acquired with a substantial gap.

D. Diffeomorphic Registration

In the LDDMM framework [15] a transformation between two images φ_1 is given as the end point of the flow of a vector field $v_t(\varphi_t) = \varphi_t'$, where $\varphi_0 = \mathrm{id}$ with id the identity function. The diffeomorphic registration between the source and target volumes I_0 and I_1 is posed as the variational minimization of:

$$E(v) = \frac{1}{\sigma_\mathrm{I}^2}\|I_0(\varphi_1^{-1}) - I_1\|_{L_2}^2 + \int_0^1 \|v_t\|_V^2 dt, \qquad (7)$$

where σ_I represents the image noise variance and norms are taken in the spaces of square integrable functions L_2 and velocity fields V. Then, by enforcing a certain smoothness on V, φ_t is guaranteed to lie in the space of diffeomorphisms. In practice, smoothness is promoted by an operator $\mathcal{L} = (-\beta\Delta + \mathrm{id})^\gamma$ with Δ the Laplacian and β and γ controlling the level and properties of smoothness respectively.

Ensuring smooth and invertible mappings by the LDDMM framework may confront numerical difficulties [21]. For efficiency, we adopt single-step temporal integration and track the invertibility based on the sufficient Lipschitz condition [22]:

$$q(p) = \max_{p'} \frac{\|(\varphi(p) - \varphi(p')) - (p - p')\|_2^2}{\|p - p'\|_2^2} < a, \qquad (8)$$

with p, p' a pair of coordinates and Lipschitz constant $a < 1$. A risk of local non-invertibility (with maximum for p' taken in a discretized neighborhood of p) is used in each step to modulate the registration gradient descent step size by $\max(a - \max_p(q(p)), 0)$.

Hence, in our joint SV reconstruction and diffeomorphic registration formulation, we face a series of volume to slice [11] registration subproblems from source reconstructions to target measurements:

$$\hat{\boldsymbol{\phi}}_s^{(j)} = \operatorname*{argmin}_{\boldsymbol{\phi}_s^{(j)}} \frac{1}{\sigma_\mathrm{I}^2}\|\mathbf{P}_s^{(j)}(\mathbf{A}\ \boldsymbol{\Phi}^{\boldsymbol{\phi}_s^{(j)}}\mathbf{x} - \mathbf{y})\|_2^2 + (\mathcal{L}\boldsymbol{\phi}_s^{(j)})^T\boldsymbol{\phi}_s^{(j)}.$$
$$(9)$$

$\boldsymbol{\Phi}^{\boldsymbol{\phi}_s^{(j)}}$ is the image warping operator, implemented by linear interpolation according to the diffeomorphism induced by $\boldsymbol{\phi}_s^{(j)}$, which parametrizes a band-limited field [23]. $\mathcal{L}$ represents the discrete version of the smoothness operator $\mathcal{L}$. (9) corresponds
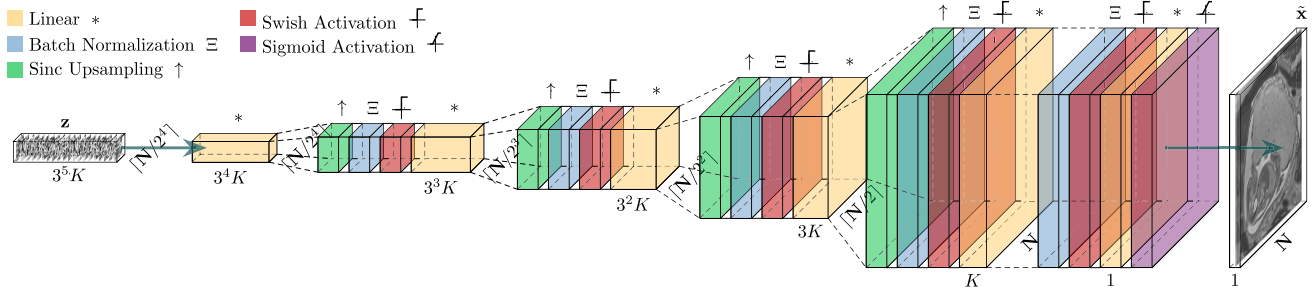
Fig. 4. DD architecture used for reconstruction regularization. Considering a network with D scales, scaling ratio U and channel compression ratio per-scale C, the input is a random array of size $\lceil N/U^{D-1} \rceil \times C^D K$, with $N = (N_1, N_2, N_3)$ the grid dimensions of the output volume, K the number of channels at the finest scale, and $\lceil \cdot \rceil$ the ceiling function. The coarsest scale comprises simply a linear layer. The remaining scales are connected by sinc upsampling layers and formed by blocks of batch normalization, activation and linear layers. At the finest scale, we use two such blocks followed by sigmoid activation. Linear layers at scale $0 \leq d < D$ map $C^{d+1} K$ input to $C^d K$ output features but for the additional layer at full resolution that performs the final $K \to 1$ mapping.

to the optimization with respect to the motion parameters of (1) with $\tau = 0$ in (6) and regularization

$$h(\boldsymbol{\phi}^{(j)}) = \sigma_I^2 \sum_s (\mathcal{L} \boldsymbol{\phi}_s^{(j)})^T \boldsymbol{\phi}_s^{(j)}. \quad (10)$$

Although the robust cost function presented in § II-A could also be used for motion estimation, we have observed that OLS are effective in escaping local optima of the motion parameters. The Hilbert gradient in the space of diffeomorphisms V [15] is obtained by:

$$(\nabla_{\boldsymbol{\phi}_s^{(j)}} E)_V = \mathcal{L}^{-1} \left(\frac{2}{\sigma_I^2} (\nabla \boldsymbol{\Phi}^{\boldsymbol{\phi}_s^{(j)}} \mathbf{x}) \cdot (\mathbf{A}^T \mathbf{P}_s^{(j)} \mathbf{r}) \right) + 2 \boldsymbol{\phi}_s^{(j)}. \quad (11)$$

E. DD Regularization

The DD [16] is a deep architecture designed for the efficient representation of natural images without using training data. The network is based on the concatenation of upsampling, convolution, activation and batch normalization layers where 1×1 convolutions (i.e., linear layers) are shown to be the most cost-effective. Its parameters are optimized by fitting to a particular cost function using fixed random inputs $\mathbf{z}$, and it has shown competitive results when compared to supervised methods in applications such as image compression, denoising, inpainting, and reconstruction [24].

The proposed DD architecture is depicted in Fig. 4. We have performed a series of modifications to the original network. First, swish units are used instead of rectified linear units, according to the results in [25]. Second, batch normalization is applied before rather than after activation, following the recommendations in [26]. Third, we use sinc rather than bilinear upsampling to better preserve fine detailed structures. Of independent interest, these changes have been tested for the 2D image compression experiment in [16] showing noticeable improvements for a wide range of compression rates. Finally, we parametrize our architecture using the number of scales D, the scaling ratio U, the channel compression ratio per-scale C (defined as the ratio of input and output features of the linear layers), and the number of output channels of the first linear layer at full resolution K. Similar to [16],

we empirically fix $D = 5$ and $U = 2$, and K is determined by prescribing a given minimum compression rate S for the network. However, we use $C = 3$ instead of $C = 1$, as this provides a thinner network at the finest scale, which reduces peak memory consumption, a limiting factor in 3D. Namely, this modification allows the 3D problem to be fitted in the memory of our GPU card with minimum loss of detail.

The DD output is used as the expected value of the reconstruction for the following generalized Tikhonov regularizer:

$$g(\mathbf{x} - \boldsymbol{\Theta}^{\boldsymbol{\theta}} \mathbf{z}) = \lambda \|\mathbf{x} - \boldsymbol{\Theta}^{\boldsymbol{\theta}} \mathbf{z}\|_2^2. \quad (12)$$

For a series of fixed parameters of the network at reconstruction iteration i, $\hat{\boldsymbol{\theta}}_{(i)}$, we can compute the DD output

$$\tilde{\mathbf{x}}_{(i)} = \boldsymbol{\Theta}^{\hat{\boldsymbol{\theta}}_{(i)}} \mathbf{z} \quad (13)$$

mapped to match the dynamic range of the reconstruction at previous iteration, and solve the reconstruction problem

$$\hat{\mathbf{x}}_{(i+1)} = \underset{\mathbf{x}}{\operatorname{argmin}} f(\mathbf{A} \boldsymbol{\Phi} \mathbf{x} - \mathbf{y}) + \lambda \|\mathbf{x} - \tilde{\mathbf{x}}_{(i)}\|_2^2 \quad (14)$$

by extending the solution in (4) with the Tikhonov term. Then, the parameters of the network can be refined according to:

$$\hat{\boldsymbol{\theta}}_{(i+1)} = \underset{\boldsymbol{\theta}}{\operatorname{argmin}} \|\boldsymbol{\Theta}^{\boldsymbol{\theta}} \mathbf{z} - \hat{\mathbf{x}}_{(i+1)}\|_2^2. \quad (15)$$

This scheme penalizes the deviation of the reconstruction from the space of natural or structured images (like medical images) because the DD fits structured images faster than noise (the so-called implicit bias of the architecture) [27], and corresponds to a relaxed version of the formulation in [16] and [24]. The Tikhonov regularizer in (12) is consistent with the fitting cost in (15) and, when compared with more complicated regularizers, does not introduce significant computational overload to the reconstruction.

F. Materials

We have tested our reconstruction algorithm in a cohort of 72 fetal cases consented as participants in the iFIND project (ISRCTN16542843) [28] with GA distribution ranging from $20 + 2$ to $36 + 0$ weeks shown in Fig. 5a. The cohort includes controls, fetuses with suspected abnormalities, and cases with
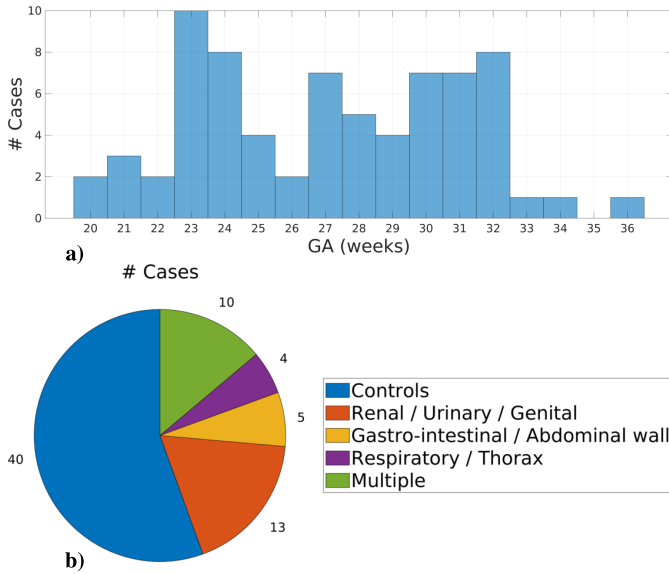
Fig. 5. Cohort characteristics. **a)** GA and **b)** abnormalities distribution.

Algorithm 1 Motion Compensated Robust Reconstruction With Deep Generative Regularization

---

1: **Inputs: y, A, P, $\mathcal{L}$, $\mathcal{G}$; Outputs: x, ϕ, θ, $\tilde{\mathbf{x}}$, w**
2: $\phi \leftarrow \mathbf{0}$, $\mathbf{x} \leftarrow \mathbf{0}$, $\tilde{\mathbf{x}} \leftarrow \mathbf{0}$, $\theta \leftarrow$ random, $\mathbf{z} \leftarrow$ random
3: **for** $j \in \{\text{STACK}, \text{PACKAGE}, \text{INTERLEAVE}\}$ **do**
4: $\quad \tau = 0$, $\mathbf{W} \leftarrow \mathbf{I}$
5: $\quad$ **while** 1 **do**
6: $\quad\quad \mathbf{x} \xleftarrow{(14)} \mathbf{W}, \tilde{\mathbf{x}}, \mathbf{A}, \phi^{(j)}, \mathbf{x}, \mathbf{y}$ (reconstruction)
7: $\quad\quad \theta \xleftarrow{(15)} \mathbf{z}, \theta, \mathbf{x}$ (DD fitting)
8: $\quad\quad \tilde{\mathbf{x}} \xleftarrow{(13)} \mathbf{z}, \theta$ (DD inference)
9: $\quad\quad$ **if** $\tau = 1$ and update in $\phi^{(j)}$ lower than δ_ϕ **then**
10: $\quad\quad\quad$ **break**
11: $\quad\quad$ **end if**
12: $\quad\quad \phi \xleftarrow{(9)} \mathbf{P}^{(j)}, \mathcal{L}, \mathbf{A}, \phi^{(j)}, \mathbf{x}, \mathbf{y}$ (motion estimation)
13: $\quad\quad \mathbf{w} \xleftarrow{(5)} \mathcal{G}, \mathbf{A}, \phi^{(j)}, \mathbf{x}, \mathbf{y}$ (weight computation)
14: $\quad\quad \mathbf{W}, \tau \xleftarrow{(6)} \mathbf{w}, \tau$ (weight relaxation)
15: $\quad$ **end while**
16: **end for**

---

incidental findings. In Fig. 5b we show the distribution of abnormalities, including controls or cases without any detected anomaly; with renal, urinary or genital abnormalities; with gastrointestinal abnormalities including abdominal wall defects; with chest abnormalities including respiratory, cardiac and thoracic; and with multiple abnormalities including skeletal and CNS anomalies.

Images were acquired on a PHILIPS INGENIA 1.5 T with a 24-channel receive coil using a single-shot FSE sequence. For each subject $N_l = 5$ stacks were acquired, including axial (one repeat), sagittal (two repeats) and coronal (two repeats) orientations. Data was collected with an in-plane resolution of 1.25 mm isotropic, 2.5 mm slice thickness and 1.25 mm slice separation. Sensitivity encoding acceleration was set to 2 with half scan 0.575 and echo time $T_E = 80$ ms. Number of slices was variable for adequate coverage of the targeted FOV in the range [100, 160] with number of packages and interleaves in the order of those in Fig. 3a. Total acquisition time was on average $T_A = 11'20''$.

G. Implementation Details

We start by performing an OLS reconstruction of the acquired stacks on a 2.5 mm grid. A Region Of Interest (ROI) containing the whole uterus is drawn on the resulting volumes by the 3D implicit model tool [29] in the SEG3D software [30]. This is used to define the FOV of the reconstruction and has no effect on the quality of the final result. Then, the pipeline proceeds as described in Algorithm 1. At each motion estimation scale j the IRWLS is reset ($\tau = 0$) and the method alternates the estimation of the reconstructed image $\mathbf{x}$, DD parameters θ, prior $\tilde{\mathbf{x}}$, motion ϕ, and weights $\mathbf{w}$ and $\mathbf{W}$. This is repeated till reaching the target formulation ($\tau = 1$) and convergence of motion estimates as dictated by maximum update below a threshold δ_ϕ.

The algorithm confronts different subproblems for which a series of parameters need to be selected. First, for robust reconstruction we use a full width half maximum of 2.5 mm for $\mathcal{G}$, appropriate to reduce the variance in data reliability estimation and keep good localization properties, with the simple choice $k = \sqrt{2}$ providing strong suppresion of outliers and acceptable SNR penalty. Second, for registration we use $\beta = 2.5$ and $\gamma = 2.5$, in agreement with values suggested in the literature [15], [23], $a_{\text{STACK}} = 0.45$, $a_{\text{PACKAGE}} = 0.65$ and $a_{\text{INTERLEAVE}} = 0.85$, which have been observed to prevent early saturation of convergence and guarantee numerical invertibility, and $\sigma_I = 0.1$, chosen by visual assessment of plausibility of deformations. Third, for regularization we use $S = 2$, achieving a strong suppression of spurious structures while maintaing good data fidelity, visually tuned $\lambda = 0.4$, and run 180 epochs using the Adam optimizer with learning rate $5 \cdot 10^{-3}$ for refining the network parameters. Finally, $N_i = 5$ iterations are enough for improved IRWLS convergence by homotopy continuation and δ_ϕ set to half the in-plane acquisition resolution provides full motion estimation convergence. We refer the reader to the source code for further details.

H. GA Prediction

We propose to estimate the GA from the reconstructions by the following steps:

1) Obtain a set of slices by uniformly reformatting a reconstructed ROI using N_r rotations evenly distributed in the 3D rotation group. Rotations are performed around the ROI center after windowing the corresponding volumes to prevent boundary artifacts, and three centered slices are extracted along the main reoriented planes.

2) Extract a set of deep features using a given pretrained model. We have considered ShuffleNet [31], ResNet-18 [32], GoogLeNet [33], ResNet-50 [32], MobileNet-v2 [34], ResNet-101 [32], and DenseNet-201 [35], all trained in the ImageNet database [36]. Slices extracted in previous step are spatially

zero-padded, replicated in the channel dimension to match the input sizes of the models, and z-score normalized.

3) Taking the deep features as predictors, use zero-correlation constrained linear GA regression [37].

4) At inference time, ensemble the estimates for the $3N_r$ slices into a final GA prediction by taking their median.

III. EXPERIMENTS AND RESULTS

In § III-A we run synthetic experiments analyzing the regularization and registration schemes introduced in this paper. In § III-B we quantify the relative impact of the main constituents of our proposal by an ablation study. In § III-C we compare with existing approaches using quantitative image quality metrics, resolution analysis, and radiological rating. In § III-D we characterize the performance for varying levels of motion. Finally, potential clinical utility is studied in § III-E by presenting comparative results on GA estimation from the reconstructions.

A. Synthetic Experiments

The reconstruction with minimum data corruption as assessed by the average Normalized Cross Correlation (NCC) of adjacent input slices [11] is used as a synthetic GT $\mathbf{x}^*$. This NCC indicator, which has been reported to agree with human observer quality ratings, is used hereafter to rank the acquired image quality of cases in our cohort. First, we study the DD regularization scheme in § II-E using L-curve analysis [38]. We simulate the acquired data using a motionless forward model and adding Gaussian noise for a SNR of 30 dB in the acquired stacks, in gross agreement with the SNR levels observed in real images. In Fig. 6a we show an L-curve traced from the regularization $\|\mathbf{x} - \tilde{\mathbf{x}}\|_2^2$ and residual term $\|\mathbf{r}\|_2^2$ norms for different values of the regularization parameter λ –see (14)– when using an OLS reconstruction. Reconstructions are repeated 5 times with different noise realizations and average results are reported. In Fig. 6b we show the curvature of the L-curve κ and in Fig. 6c we present the reconstruction errors with respect to the GT $\|\mathbf{x} - \mathbf{x}^*\|_2^2$, both as a function of λ. We observe that the maximum curvature $\lambda = 0.2$ matches that of minimum error. Thus, L-curve may be appropriate for predicting the regularization parameter for SV reconstruction, although increased computational burden is a limiting factor. In addition, our formulation involves dynamically modifying the system matrix and uses IRWLS instead of OLS, scenarios where the properties of the L-curve are less documented. Moreover, these differences may explain the discrepancies between predicted λ and that determined visually, $\lambda = 0.4$.

Second, we study motion estimation at different levels of motion by synthesizing random diffeomorphisms with independent and identically distributed Gaussian entries in the space of velocities V and windowed rigid motion (elliptic Tukey window, taper ratio 0.5) with random rotation axes and center of rotation in the center of the FOV. We simulate the acquired data by applying our forward model for different slice subsampling rates corresponding to the whole stack
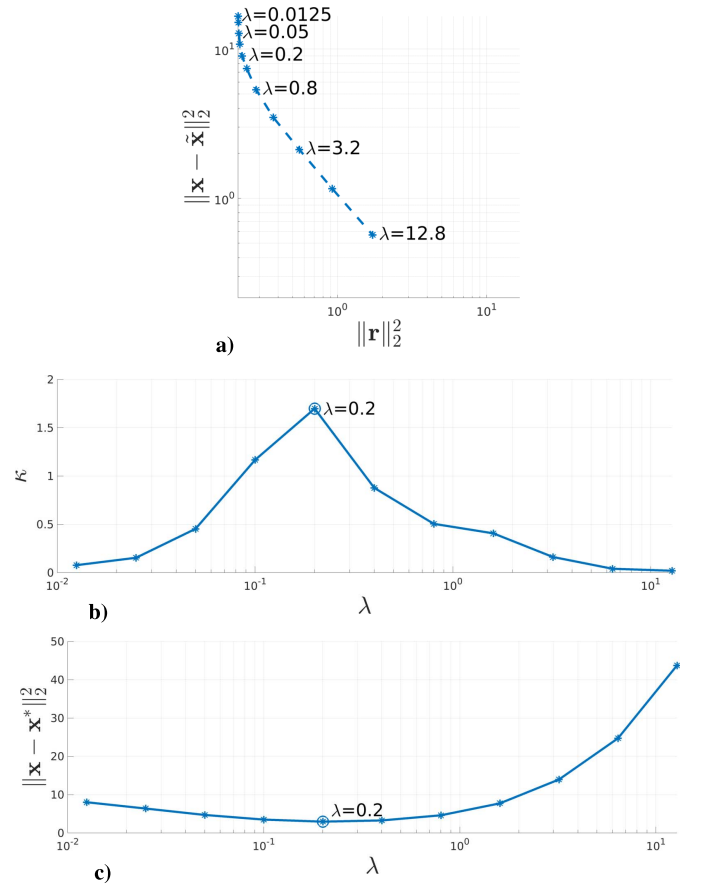

Fig. 6. L-curve analysis of DD regularization for synthetic data. **a)** L-curve; **b)** Curvature of the L-curve; **c)** Reconstruction error.
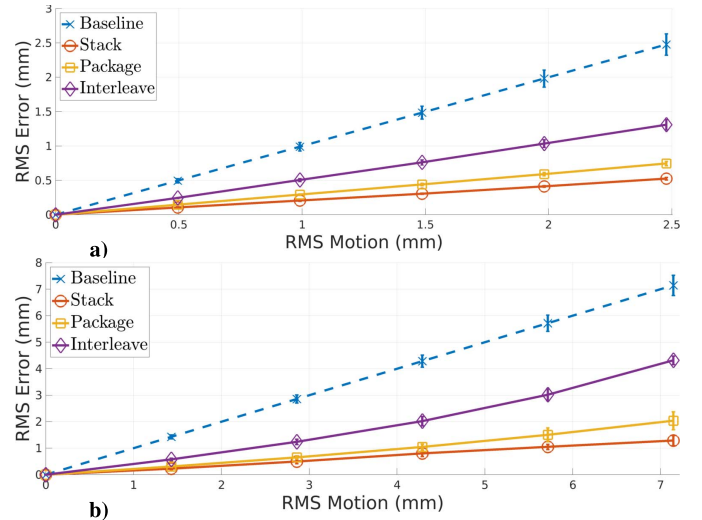

Fig. 7. RMS error of motion estimation at the stack, package and interleave levels (solid lines) as assessed from synthetic data. Baseline RMS level of simulated motion is depicted for reference (dashed line). **a)** Random diffeomorphisms; **b)** Rigid motion.

(no subsampling), package ($4\times$ subsampling) and interleave ($20\times$ subsampling) levels. In Fig. 7 we show the Root Mean Square (RMS) error for motion estimation using the GT reconstruction to match the simulated slices, including
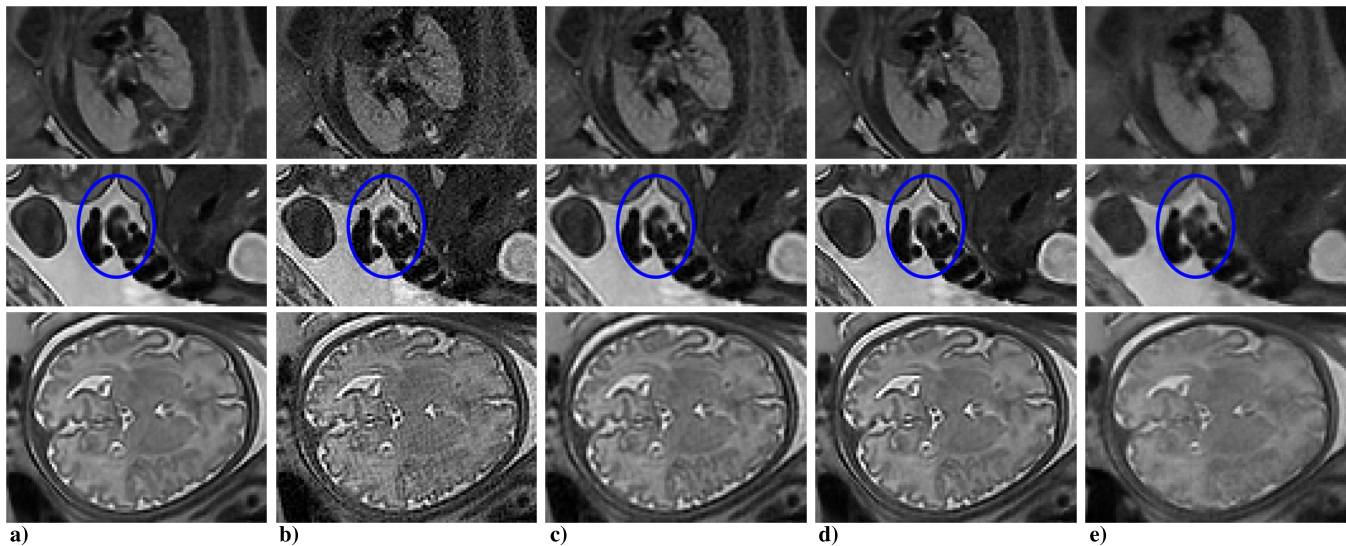
a)   b)   c)   d)   e)

Fig. 8. Comparison of different reconstruction alternatives in the case with median degradation. From top to bottom, coronal, sagittal and axial planes in the mother's geometry. Reconstructions **a)** based on the full model **(x)**; **b)** without regularization $(\mathbf{x}_{\lambda=0})$; **c)** with handcrafted regularization $(\mathbf{x}_{\mathcal{H}})$; **d)** without the robust formulation $(\mathbf{x}_{\tau=0})$; **e)** without motion correction $(\mathbf{x}_{\sigma_I \to \infty})$. The blue ellipse highlights local artifacts, noise or blurring around the umbilical chord when taking off any of the main components of our formulation.

the mean values when warping to 5 stacks simulating the orientations in our actual data and corresponding error bars. Maximum errors were also computed with similar appearance to RMS curves. Simulations at different deformation and rotation levels, with averaged maximum motion excursions and minimum Jacobian respectively around 6 mm and 0.7 for random diffeomorphisms and 16 mm and 0.5 for rigid motion, may span some typically observed motion ranges, as larger levels start to introduce unrealistic deformations as assessed from visual checks. We see that for both motion models stable motion estimates can be obtained for motion states defined at high slice subsampling rates, as those given by the interleave pattern in Fig. 3a. However, the higher the subsampling, the smaller the accuracy, which justifies the temporal multi-scale motion estimation scheme described in II-C. Comparing appearance of errors when diffeomorphism generation and estimation models are matched (Fig. 7a) and when using rigid motion for motion generation (Fig. 7b), we observe close to linear estimation error in the former and supralinear trends in the latter, especially at the interleave level.

B. Ablation Study

To assess the impact of the main components of the proposed reconstruction scheme, we compare the reconstructions using the full model (**x**) versus not using regularization –i.e., $\lambda = 0$ in (14)– $(\mathbf{x}_{\lambda=0})$; using a handcrafted regularizer $g(\mathbf{x}) = \|\mathcal{H}\mathbf{x}\|_2^2$ where $\mathcal{H}$ is the superposition of finite difference penalizers of different orders (0 to 32) whose weights have been visually tuned for trading off noise suppression and resolution loss $(\mathbf{x}_{\mathcal{H}})$; not using the robust formulation –i.e., fixing $\tau = 0$ in (6)– $(\mathbf{x}_{\tau=0})$; and not using motion correction –i.e., $\sigma_I \to \infty$ in (11)– $(\mathbf{x}_{\sigma_I \to \infty})$. The convergence criterion has been adjusted for comparable number of iterations for all alternatives. Table I shows results of Leave One Out Cross-Validation (LOOCV), where we measure the agreement

TABLE I
MEAN ± STD RESULTS OF ABLATION STUDY BY
LOOCV. BEST RESULTS ARE SHADED

	PSNR (dB)	NCC	SSIM
$\mathbf{x}$	24.925 ± 2.049	0.890 ± 0.063	0.756 ± 0.088
$\mathbf{x}_{\lambda=0}$	24.671 ± 2.044	0.884 ± 0.065	0.745 ± 0.091
$\mathbf{x}_{\mathcal{H}}$	24.817 ± 2.043	0.887 ± 0.064	0.752 ± 0.090
$\mathbf{x}_{\tau=0}$	24.501 ± 2.072	0.879 ± 0.066	0.738 ± 0.087
$\mathbf{x}_{\sigma_I \to \infty}$	23.458 ± 1.707	0.849 ± 0.066	0.662 ± 0.086

between the stack with minimum motion and the propagation to the acquisition space of the reconstructions obtained by completely excluding this stack [4]. Note this is a challenging scenario as we are left with 4 most corrupted stacks only. Results of Peak SNR (PSNR), in dB, NCC and Structural SIMilarity Index (SSIM) are shown for all 72 cases in our cohort. Results without motion correction $(\mathbf{x}_{\sigma_I \to \infty})$ are clearly worse for all compared metrics and removing the robust formulation $(\mathbf{x}_{\lambda=0})$ also shows an impact on performance as assessed by all considered metrics. Differences are smaller for different regularization choices although we observe that the DD regularizer (**x**) provides improvements over the hand-crafted regularizer $(\mathbf{x}_{\mathcal{H}})$ or not using regularization $(\mathbf{x}_{\lambda=0})$. However, improvements of the full model versus any of the alternatives are highly significant when performing two-tailed sign tests for all metrics used for comparison ($p < 1 \cdot 10^{-7}$).

To aid interpretation of Table I we report in Fig. 8 visual results for the scan with median level of acquired data corruption [11]. Additional results for representative subjects with different levels of data corruption are included in a Supplemental Video showing reconstructions in 3D. We observe that reconstructions without regularization (Fig. 8b) present a noisy appearance. Noise can be mitigated either by the regularizer penalizing high-frequency content (Fig. 8c) or by our proposed DD regularizer (Fig. 8a), but resolution is better
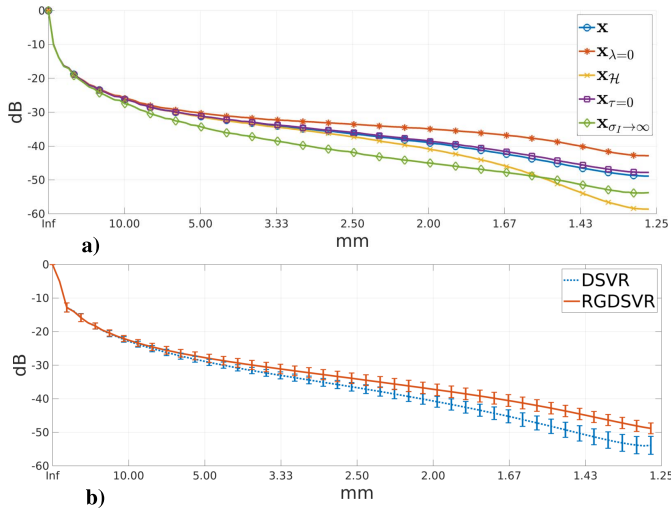
Fig. 9. PSD comparisons. **a)** Reconstruction alternatives in the case with median degradation (uterus ROI); **b)** DSVR and RGDSVR (trunk ROI).

preserved when using the DD regularizer, for instance at the boundaries between the fetal body and the amniotic fluid in the area enclosed in blue. We observe that differences are mainly affecting fine details of the reconstruction, which may not be captured in the LOOCV experiment due to imperfect mapping of the reconstructions or noise and artifacts in the excluded stack. This hypothesis is supported by the L-curve experiment in § III-A, where we have observed improvements approaching 5 dB when introducing DD regularization, much higher than differences with no regularization in Table I. Ability to resolve fine detailed structures is evident when comparing to results without motion compensation (Fig. 8e). Non-robust reconstructions (Fig. 8d) look similar to robust reconstructions (Fig. 8a). However, impact of non-suppressed artifacts is noticeable locally, as in the area within the blue ellipse, where more uniform fluid background is observed when the robust formulation is adopted.

In Fig. 9a we show the Power Spectral Density (PSD) averaged along the three axes of the reconstruction grid for the aforementioned reconstruction alternatives applied to the same subject. Reconstructions without regularization ($\mathbf{x}_{\lambda=0}$) present the largest power at high spatial frequencies with disruption of power law of attenuation above approximately 1.5 mm, probably stemming from a Gibbs ringing filter applied as part of the scanner k-space reconstructions. However, we know from image inspection in Fig. 8b that a significant amount of the energy at high frequencies is contributed by noise. Noise reduction was effective when using the handcrafted regularizer, but we observe here ($\mathbf{x}_{\mathcal{H}}$) that this comes at the price of strong suppression of high frequency signal components. The DD-based regularized reconstructions ($\mathbf{x}$) lie somewhere in between both scenarios, likely with better signal preservation at high frequencies and strong noise reduction. Small differences are observed between non-robust ($\mathbf{x}_{\tau=0}$) and robust reconstructions, but the non-robust version presents slower PSD decay rates consistent with reduced suppression of artifactual structures. Motion compensation has an impact

TABLE II
MEAN ± STD RESULTS OF COMPARISON WITH THE LITERATURE BY LOOCV. BEST RESULTS ARE SHADED

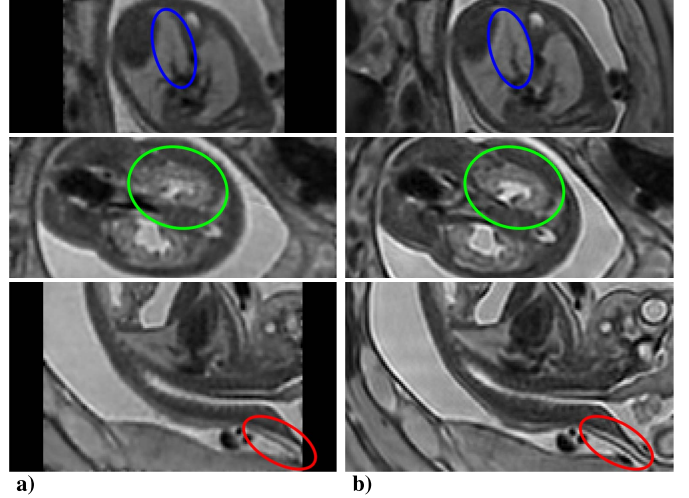|	PSNR (dB)	NCC	SSIM
SVR	22.989 ± 2.357	0.821 ± 0.118	0.673 ± 0.100
PVR	23.351 ± 1.672	0.845 ± 0.074	0.671 ± 0.077
SRR	23.252 ± 2.278	0.836 ± 0.104	0.686 ± 0.104
DSVR	23.518 ± 2.224	0.845 ± 0.105	0.706 ± 0.103
RGDSVR	24.925 ± 2.049	0.890 ± 0.063	0.756 ± 0.088

Fig. 10. Three orthogonal views with reconstructions using **a)** DSVR and **b)** RGDSVR. Areas enclosed by the blue (top panel), green (central panel) and red (bottom panel) ellipses suggest that our method provides improved conspicuity of the main vessels, sharper contrast and better delineation of the bowel loops, and increased resolution of stacked tissue layers in the neck, respectively.

in moderate to high spatial frequencies, with power enhancements versus non-compensated reconstructions ($\mathbf{x}_{\sigma_I \to \infty}$) above 5 dB at 2 mm, for instance.

C. Comparison With Other Methods

Table II reports results of LOOCV for Slice to Volume Reconstruction (SVR) [4], Patch to Volume Reconstruction (PVR) [10], Super-Resolution Reconstruction (SRR) [7], Deformable Slice To Volume Reconstruction (DSVR) [11], and our RGDSVR method, again using all cases in our cohort. We observe that RGDSVR generally provides higher metrics than other approaches. Statistical significance is assessed by two-tailed sign tests which have shown highly significant differences when comparing RGDSVR with any of the alternatives for all metrics used for comparison ($p < 5 \cdot 10^{-10}$). As for the remaining approaches, results are in line with those reported in [11], with DSVR generally overtaking previous methods. Therefore, in what follows experiments are focused on RGDSVR and DSVR.

We compare the resolution provided by RGDSVR and DSVR. The reconstruction grid orientation can be different for both methods, so we have used sinc interpolation to reformat our datasets into the grid of the reference method. In Fig. 9b we show the mean±std PSD of both methods across our whole cohort. The curves suggest better preserved moderate
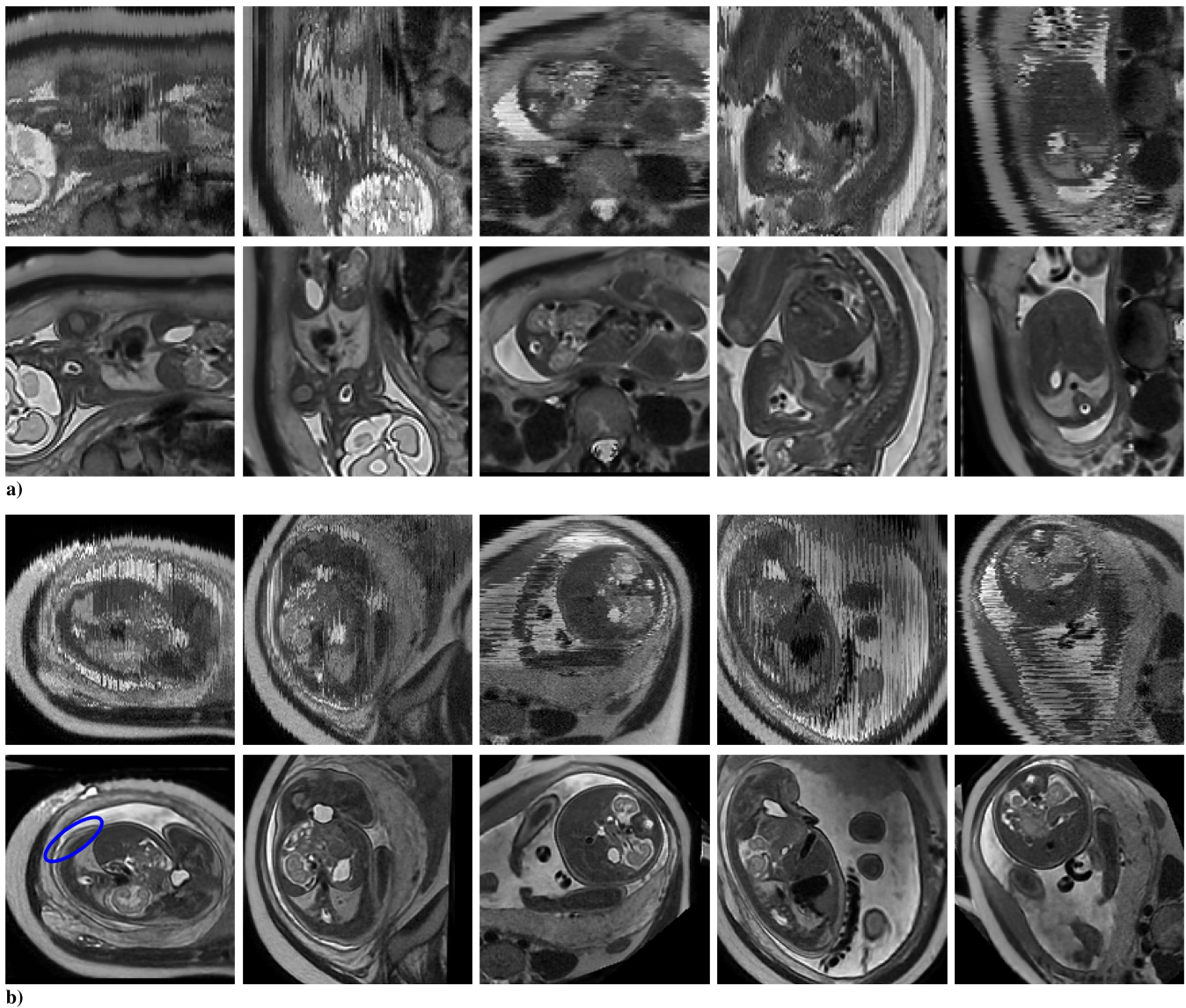
Fig. 11. Input data and reconstructions for **a)** low and **b)** high motion examples, corresponding respectively to cases with 25- and 75-percentile degradation. Top rows show the input data for different stacks in planes containing the slice direction, with larger degradation observed in **b)** when compared to **a)**. Bottom rows show the 3D reconstructions for the same planes. Retrieval of consistent and sharp information about the fetal anatomy is generally observed, with minor inconsistencies and artifacts in the high motion example (as in the area enclosed by the blue ellipse).

to high spatial frequency information when using RGDSVR, with highly significant differences as assessed by a two-tailed sign test for structures below 10 mm ($p < 10^{-8}$).

A radiologist with 11 years' experience in fetal MRI was asked to select the higher quality image in each of three orthogonal pairs of matched slices extracted around the center of the trunk FOV from DSVR and RGDSVR. Preferences were averaged per-subject and this resulted in 62 cases were RGDSVR ranked better (86.11 %) and 10 cases were DSVR ranked better (13.89 %). In Fig. 10 we visually illustrate the differences between DSVR and RGDSVR by selecting areas where these can be clearly perceived. Despite reconstructions are not spatially matched due to arbitrary definitions of reference deformations, we observe that RGDSVR tends to provide sharper results for comparable levels of noise and artifacts. This is illustrated by the ellipses on top of the main vessels,

the bowel, and the neck, respectively in the top, central and bottom panels.

D. Performance for Varying Levels of Motion

In Fig. 11 we show examples of input data and reconstructions for low (Fig. 11a) and high (Fig. 11b) motion cases, corresponding respectively to subjects with 25- and 75-percentile acquired data degradation. Visual results are presented for the 5 acquired stacks along planes containing the slice direction and centered on relevant information about the fetal anatomy. We observe that reconstructions (bottom rows) are able to suppress the inconsistencies in the input data (upper rows) and produce sharp results for most imaged structures.

We have computed the Pearson correlation ρ of the PSNR, NCC and SSIM metrics considered in the LOOCV experiments, assumed to be related to the reconstruction quality,

TABLE III

MAE (WEEKS) AND R^2 IN GA ESTIMATION USING DIFFERENT RECONSTRUCTIONS, ROIS, AND DEEP FEATURES. BEST RESULTS ARE SHADED

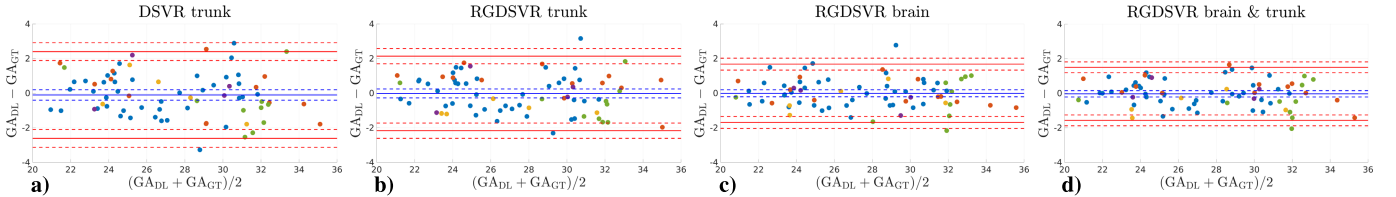|	DSVR trunk		RGDSVR trunk		RGDSVR brain		RGDSVR brain & trunk	
	MAE	R^2	MAE	R^2	MAE	R^2	MAE	R^2
ShuffleNet	1.115	0.874	1.012	0.895	0.764	0.941	0.632	0.956
ResNet-18	1.234	0.829	1.133	0.876	0.712	0.941	0.674	0.947
GoogLeNet	1.472	0.771	1.049	0.885	0.714	0.945	0.694	0.945
ResNet-50	1.164	0.841	0.955	0.910	0.734	0.937	0.686	0.948
MobileNet-v2	1.201	0.841	1.137	0.863	0.712	0.945	0.705	0.943
ResNet-101	1.218	0.839	1.015	0.900	0.731	0.943	0.677	0.949
DenseNet-201	1.045	0.888	0.931	0.918	0.683	0.950	0.618	0.958

Fig. 12.　Bland-Altman plots for DenseNet-201. **a)** DSVR trunk. **b)** RGDSVR trunk. **c)** RGDSVR brain. **d)** RGDSVR brain & trunk. Solid blue line: mean differences. Solid red lines: 95% limits of agreement. Dashed lines: corresponding approximate confidence intervals [41].

against the average NCC of adjacent input slices, which is related to acquired data quality, for reconstructions without and with motion correction. We observe stronger correlation between input and output data quality without ($\rho_{PSNR} = 0.349$, $\rho_{NCC} = 0.371$ and $\rho_{SSIM} = 0.556$) than with motion correction ($\rho_{PSNR} = 0.330$, $\rho_{NCC} = 0.303$ and $\rho_{SSIM} = 0.439$). Thus, correlation is still important but becomes weaker when introducing motion correction, which may indicate that resulting reconstructions are less sensitive to the artifacts and inconsistencies in the original data.

E. Clinically-Oriented Application: GA Prediction

Despite potential gains of proposed reconstructions are suggested by the experiments in § III-C, lack of GT makes direct comparison of methods very challenging. In addition, it is very difficult to faithfully reproduce the different sources of corruption in real data as well as complex patterns of fetal and mother motion by simulations. Therefore, we have resorted to a task-based validation where we perform comparisons on a clinically-oriented GA prediction problem. To ensure comparisons are performed using the same FOV, we assess the performance of RGDSVR and DSVR for GA estimation using a trunk ROI corresponding to the FOV used as the target reconstruction volume in the experiments in [11]. As existing methods for fetal GA estimation from MRI [14], [39], [40] are based on brain data, we also test the GA estimation performance using a brain ROI from our reconstructions. In all cases we use the GA estimation method described in § II-H with 3D space spanned by slices from $N_r = 200$ random volume reorientations. Finally, to investigate the added value of non-brain features, we combine brain and trunk ROIs by using $N_r = 100$ random reorientations each. In this case, common regression and z-score normalization weights are computed at training using slices from both ROIs, and joint median ensemble of estimates is used at inference.

We perform a 6-fold cross validation using the 72 cases in our cohort. Different GA regression alternatives are compared by computing the Mean Absolute Error (MAE) and coefficient of determination R^2 with results reported in Table III. GA predictions using RGDSVR consistently outperform predictions using DSVR for comparable trunk ROI. Best results in both cases are obtained using DenseNet-201, respectively with MAE (R^2) 0.931 (0.918) and 1.045 (0.888). We also observe consistently better results for all models when using the brain rather than the trunk ROI, with best figures 0.683 (0.950) provided again by DenseNet-201. These results compare favourably with those reported by [39], 0.751 (0.947), and [14], 0.767 (0.920), and unfavourably with those in [40], 0.508 (0.992). Further to this, with the exception of R^2 for poorly performing MobileNet-v2 and GoogLeNet, additional improvements are consistently observed for all models when combining volumetric brain and trunk information, a unique feature of the proposed technique, with MAE (R^2) 0.618 (0.958) for DenseNet-201. However, GT GA is obtained from the clinical records which is not free from errors. Therefore, in Fig. 12 we provide Bland-Altman plots of agreement [41] between DenseNet-201 based predictions (GA_{DN}) and GT GA (GA_{GT}) for the considered reconstructions and ROIs combinations. Results are color coded according to the abnormality categories legend in Fig. 5b. We observed no statistical significance at $p = 0.05$ in GA discrepancies between both methods when comparing controls and cases with anomalies with an unequal variances t-test for any of the alternatives considered in Fig. 12.

IV. DISCUSSION

We have proposed a novel methodology for robust whole-body fetal MRI reconstruction relying on diffeomorphic motion estimation to capture plausible deformations of the fetal organs and high-quality regularization using a deep

generative model. We have quantitatively and qualitatively characterized the impact of these components in the reconstructions. Quantitative comparisons with other fetal MRI reconstruction methods have shown improved performance for PSNR, NCC and SSIM metrics when using stack-wise LOOCV, which has been replicated by radiological rating. Noticeable differences in reconstruction sharpness have been demonstrated by resolution analysis and a potentially strong impact of the reconstruction method in the clinical utility of fetal MRI has been showcased by a GA prediction task. Finally, we have provided a conceptually simple GA prediction method based on free reformatting the 3D reconstructions for 2D deep feature extraction and correlation constrained linear regression, showing competitive accuracy with respect to existing approaches.

Similar to [3], we have built our cost function using the robust regression via M-estimators framework, so weights for outlier mitigation are directly derived from the cost function [18], instead of using independent metrics as in [11]. In addition, the introduction of the smoothing operator $\mathcal{G}$ simplifies previous combinations of voxelwise and slicewise weights [3], [11] by assuming that motion-induced degradation of magnetization as well as uncorrected non-rigid motion have a regional nature.

In the non-rigid registration setting, reproducible morphometry may be compromised by geometric distortions introduced by the algorithm, which can be alleviated by the diffeomorphic constraint. Of particular interest, we should highlight the brain; despite the prominent use of rigid motion models in the past, non-rigid components may become appropriate to model non-linearities of the scanner gradient fields [42]. Although our motion model should probably be refined for the brain, for instance via decoupled motion models [43], distortion levels in the reconstructions are small enough so as to lead to accurate GA estimates.

Taking into account the limitations for generating GT datasets, we have opted for an unsupervised application of deep architectures for regularizing our reconstructions. However, there are alternatives for integration of DL into reconstruction problems [13] and DL could also be applied for extending the capture range of registration or refining the characterization of outliers. On the other hand, the memory footprint of the 3D DD architecture is a strong limitation, so deep network regularizers based on implicit representations [44] could be considered.

3D reconstruction errors in whole-body fetal MRI may arise from multiple causes. We may encounter errors due to (a) small inaccuracies in motion estimation, (b) inconsistencies in magnetization of different slices, (c) multiple poses of the fetal body throughout the scan, and (d) the fetus moving continuously across the examination. We believe that in cases (a) and (b) information coming from complementary stacks can resolve the ambiguities in many instances, so the robust formulation with deep generative regularization generally provides satisfactory solutions. However, artifacts in the reconstructions may be strong in cases (c) and (d), as the method may struggle to find a good direction for high quality convergence.

Our reconstruction tool can be applied to different acquisition settings. Its flexibility has been confirmed by preliminary application to cardiac and animal brain studies. However, results could show variable quality. In this regard, reconstruction quality is ultimately determined by available scan time, as enlarged sampling redundancy gives more flexibility to implement robust reconstructions with improved SNR. For a fixed total acquisition time, there are different acquisition choices that may impact the reconstruction reliability. Importantly, [45] studies the comparative performance of overlapped single orientation scans versus multi-oriented scans in terms of resolution retrieval, with clear benefits observed in the latter (see also [46]). In our context, this result may suggest replacing the repeated sagittal and coronal stacks by new orientations, perhaps whilst changing the overlapping factors. However, implementation of optimal sampling schemes is often limited by hardware specifications of the scanner, inherent complexity of fetal imaging in vivo, computational requirements of reconstruction algorithms, or need of harmonization with protocols currently in place.

Our GA estimation method leverages free reformatting of volumetric reconstructions to obtain a dense set of slices covering the fetal structures in the variable spatial configurations they can adopt due to fetal motion. We have shown that deep feature extraction using pre-trained models combined with correlation constrained linear regression provides accurate results for this task. Our results look competitive to existing methods [14], [39], [40], particularly when complementing brain features with trunk information, but there are differences in the cohorts considered. Most notably, existing methods use larger cohorts, including single-sequence data from 220 [40] and 289 subjects [39] and multi-sequence data from 764 subjects [14]. Small sample sizes in our case precluded isolation of a subset of subjects for testing, a potentially important limitation when compared to [14] and [40].

In the future, we plan to follow a practical roadmap to facilitate the application of our algorithm in clinical scenarios. This may include automated ROI extraction, refining and further testing the algorithm and GA estimation using additional cohorts and acquisition protocols, and move towards a comprehensive analysis pipeline by integrating techniques for whole-body fetal segmentation and atlas construction [47].

V. CONCLUSION

We have proposed a method for robust whole-body fetal and placenta MRI based on diffeomorphic registration and deep generative regularization. Volumetric reconstructions are obtained from a set of motion-affected and possibly corrupted single-shot slices. Our proposal provides alternative solutions to existing methods for the different subproblems faced in this application. These are validated by an ablation study and improved quantitative metrics, radiological scoring and conspicuity have been observed when comparing with existing methods in a cohort of 72 fetal subjects. A GA estimation task is defined to assess the clinical utility of our technique, for which we propose a simple method leveraging the 3D information, which produced competitive results. For usual levels

of motion, our reconstructions provide dense and consistent representations of the fetal anatomy. Therefore, the proposed methods may find application in 3D fetal MRI morphometry, developmental assessment, or fetal surgery planning.

REFERENCES

[1] American College of Radiology. (2020). *ACR-SPR Practice Parameter for the Safe and Optimal Performance of Fetal Magnetic Resonance Imaging (MRI), Resolution 45*. [Online]. Available: https://www.acr.org/-/media/ACR/Files/Practice-Parameters/mr-fetal.pdf

[2] C. L. Herrera, J. J. Byrne, H. R. Clark, D. M. Twickler, and J. S. Dashe, "Use of fetal magnetic resonance imaging after sonographic identification of major structural anomalies," *J. Ultrasound Med.*, vol. 39, no. 10, pp. 2053–2058, Oct. 2020.

[3] A. Gholipour, J. A. Estroff, and S. K. Warfield, "Robust super-resolution volume reconstruction from slice acquisitions: Application to fetal brain MRI," *IEEE Trans. Med. Imag.*, vol. 29, no. 10, pp. 1739–1758, Oct. 2010.

[4] M. Kuklisova-Murgasova, G. Quaghebeur, M. A. Rutherford, J. V. Hajnal, and J. A. Schnabel, "Reconstruction of fetal brain MRI with intensity matching and complete outlier removal," *Med. Image Anal.*, vol. 16, no. 8, pp. 1550–1564, 2012.

[5] B. Hou et al., "3-D reconstruction in canonical co-ordinate space from arbitrarily oriented 2-D images," *IEEE Trans. Med. Imag.*, vol. 37, no. 8, pp. 1737–1750, Aug. 2018.

[6] S. S. M. Salehi, S. Khan, D. Erdogmus, and A. Gholipour, "Real-time deep pose estimation with geodesic loss for image-to-template rigid registration," *IEEE Trans. Med. Imag.*, vol. 38, no. 2, pp. 81–470, Feb. 2019.

[7] M. Ebner et al., "An automated framework for localization, segmentation and super-resolution reconstruction of fetal brain MRI," *NeuroImage*, vol. 206, Feb. 2020, Art. no. 116324.

[8] A. Singh, S. S. M. Salehi, and A. Gholipour, "Deep predictive motion tracking in magnetic resonance imaging: Application to fetal imaging," *IEEE Trans. Med. Imag.*, vol. 39, no. 11, pp. 3523–3534, Nov. 2020.

[9] E. Ferrante and N. Paragios, "Slice-to-volume medical image registration: A survey," *Med. Image Anal.*, vol. 39, pp. 101–123, Jul. 2017.

[10] A. Alansary et al., "PVR: Patch-to-volume reconstruction for large area motion correction of fetal MRI," *IEEE Trans. Med. Imag.*, vol. 36, no. 10, pp. 2031–2044, Oct. 2017.

[11] A. Uus et al., "Deformable slice-to-volume registration for motion correction of fetal body and placenta MRI," *IEEE Trans. Med. Imag.*, vol. 39, no. 9, pp. 2750–2759, Sep. 2020.

[12] A. Largent et al., "Image quality assessment of fetal brain MRI using multi-instance deep learning methods," *J. Magn. Reson. Imag.*, vol. 54, no. 3, pp. 818–829, Sep. 2021.

[13] G. Ongie, A. Jalal, C. A. Metzler, R. G. Baraniuk, A. G. Dimakis, and R. Willett, "Deep learning techniques for inverse problems in imaging," *IEEE J. Sel. Areas Inf. Theory*, vol. 1, no. 1, pp. 39–56, May 2020.

[14] W. Shi et al., "Fetal brain age estimation and anomaly detection using attention-based deep ensembles with uncertainty," *NeuroImage*, vol. 223, Dec. 2020, Art. no. 117316.

[15] M. F. Beg, M. I. Miller, A. Trouvé, and L. Younes, "Computing large deformation metric mappings via geodesic flows of diffeomorphisms," *Int. J. Comput. Vis.*, vol. 61, no. 2, pp. 139–157, 2005.

[16] R. Heckel and P. Hand, "Deep decoder: Concise image representations from untrained non-convolutional networks," 2018, *arXiv:1810.03982*.

[17] P. W. Holland and R. E. Welsch, "Robust regression using iteratively reweighted least-squares," *Commun. Statist.-Theory Methods*, vol. 6, no. 9, pp. 813–827, 1977.

[18] R. A. Maronna, R. D. Martin, V. J. Yohai, and M. Salibián-Barrera, *Robust statistics, Theory and Methods (With R)* (Wiley Series on Probability & Statistics), 2nd ed. Hoboken, NJ, USA: Wiley, 2019.

[19] C. S. Burrus, J. A. Barreto, and I. W. Selesnick, "Iterative reweighted least-squares design of FIR filters," *IEEE Trans. Signal Process.*, vol. 42, no. 11, pp. 2926–2936, Nov. 1994.

[20] D. C. Noll, F. E. Boada, and W. F. Eddy, "A spectral approach to analyzing slice selection in planar imaging: Optimization for through-plane interpolation," *Magn. Reson. Med.*, vol. 38, no. 1, pp. 151–160, Jul. 1997.

[21] A. Mang and L. Ruthotto, "A Lagrangian Gauss–Newton–Krylov solver for mass- and intensity-preserving diffeomorphic image registration," *SIAM J. Sci. Comput.*, vol. 39, no. 5, pp. B860–B885, Jan. 2017.

[22] M. Chen, W. Lu, Q. Chen, K. J. Ruchala, and G. H. Olivera, "A simple fixed-point approach to invert a deformation field," *Med. Phys.*, vol. 35, no. 1, pp. 81–88, Jan. 2008.

[23] M. Zhang and P. T. Fletcher, "Fast diffeomorphic image registration via Fourier-approximated Lie algebras," *Int. J. Comput. Vis.*, vol. 127, no. 1, pp. 61–73, Jan. 2019.

[24] M. Z. Darestani and R. Heckel, "Accelerated MRI with un-trained neural networks," *IEEE Trans. Comput. Imag.*, vol. 7, pp. 724–733, 2021.

[25] P. Ramachandran, B. Zoph, and Q. V. Le, "Searching for activation functions," 2017, *arXiv:1710.05941*.

[26] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," in *Proc. Int. Conf. Mach. Learn.*, Jul. 2015, pp. 448–456.

[27] R. Heckel and M. Soltanolkotabi, "Denoising and regularization via exploiting the structural bias of convolutional generators," in *Proc. Int. Conf. Learn. Represent.*, Apr. 2020, pp. 1–31.

[28] (2021). *iFIND: Intelligent Fetal Imaging and Diagnosis*. [Online]. Available: http://www.ifindproject.com

[29] G. Turk and J. F. O'Brien, "Modelling with implicit surfaces that interpolate," *ACM Trans. Graph.*, vol. 21, no. 4, pp. 855–873, Oct. 2002.

[30] CIBC. (2016). *Seg3D: Volumetric Image Segmentation and Visualization*. Scientific Computing and Imaging Institute (SCI). [Online]. Available: http://www.seg3d.org

[31] X. Zhang, X. Zhou, M. Lin, and J. Sun, "ShuffleNet: An extremely efficient convolutional neural network for mobile devices," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Sep. 2018, pp. 6848–6856.

[32] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2016, pp. 770–778.

[33] C. Szegedy et al., "Going deeper with convolutions," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 1–9.

[34] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L.-C. Chen, "MobileNetv2: Inverted residuals and linear bottlenecks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Apr. 2018, pp. 4510–4520.

[35] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 2261–2269.

[36] (2021). *ImageNet*. [Online]. Available: http://www.image-net.org

[37] M. S. Treder, J. P. Shock, D. J. Stein, S. du Plessis, S. Seedat, and K. A. Tsvetanov, "Correlation constraints for regression models: Controlling bias in brain age prediction," *Frontiers Psychiatry*, vol. 12, Feb. 2021, Art. no. 615754.

[38] P. C. Hansen, "The L-curve and its use in the numerical treatment of inverse problems," in *Computational Inverse Problems in Electrocardiology* (Advances in Computational Bioengineering), no. 5. Billerica, MA, USA: WIT Press, 2000, pp. 119–142.

[39] L. Liao et al., "Multi-branch deformable convolutional neural network with label distribution learning for fetal brain age prediction," in *Proc. IEEE 17th Int. Symp. Biomed. Imag.*, Apr. 2020, pp. 424–427.

[40] J. Hong et al., "Optimal method for fetal brain age prediction using multiplanar slices from structural magnetic resonance imaging," *Frontiers Neurosci.*, vol. 15, Oct. 2021, Art. no. 714252.

[41] J. M. Bland and D. G. Altman, "Measuring agreement in method comparison studies," *Stat. Methods Med. Res.*, vol. 8, no. 2, pp. 135–160, Jun. 1999.

[42] S. J. Doran, L. Charles-Edwards, S. A. Reinsberg, and M. O. Leach, "A complete distortion correction for MR images: I. Gradient warp correction," *Phys. Med. Biol.*, vol. 50, no. 7, pp. 1343–1361, Apr. 2005.

[43] A. Uus et al., "Deformable slice-to-volume registration for reconstruction of quantitative T2* placental and fetal MRI," in *Proc. Int. Workshop Adv. Simplifying Med. Ultrasound* Lecture Notes in Computer Science, vol. 12437. Lima, Peru, Sep. 2020, pp. 222–232.

[44] R. Fathony, A. K. Sahu, D. Willmott, and J. Z. Kolter, "Multiplicative filter networks," in *Proc. Int. Conf. Learn. Represent.*, Vienna, Austria, May 2021, pp. 1–11.

[45] R. Z. Shilling, S. Ramamurthy, and M. E. Brummer, "Sampling strategies for super-resolution in multi-slice MRI," in *Proc. IEEE 15th Int. Conf. Image Process.*, San Diego, CA, USA, Oct. 2008, pp. 2240–2243.

[46] E. V. Reeth, I. W. K. Tham, C. H. Tan, and C. L. Poh, "Super-resolution in magnetic resonance imaging: A review," *Concepts Magn. Reson. A*, vol. 40, no. 6, pp. 306–325, Nov. 2012.

[47] J. Torrents-Barrena et al., "Assessment of radiomics and deep learning for the segmentation of fetal and maternal anatomy in magnetic resonance imaging and ultrasound," *Academic Radiol.*, vol. 28, no. 2, pp. 173–188, Feb. 2021.