Automated brain masking of fetal functional MRI with open data

Rutherford, Saige; Sturmfels, Pascal; Angstadt, Mike; Hect, Jasmine; Wiens, Jenna; van den Heuvel, Marion I.; Scheinost, Dustin; Sripada, Chandra; Thomason, Moriah

Automated Brain Masking of Fetal Functional MRI with Open Data

Saige Rutherford[1,2] · Pascal Sturmfels[3] · Mike Angstadt[2] · Jasmine Hect[4] · Jenna Wiens[3] · Marion I. van den Heuvel[5] · Dustin Scheinost[6,7,8] · Chandra Sripada[2] · Moriah Thomason[9,10]

Abstract

Fetal resting-state functional magnetic resonance imaging (rs-fMRI) has emerged as a critical new approach for characterizing brain development before birth. Despite the rapid and widespread growth of this approach, at present, we lack neuroimaging processing pipelines suited to address the unique challenges inherent in this data type. Here, we solve the most challenging processing step, rapid and accurate isolation of the fetal brain from surrounding tissue across thousands of non-stationary 3D brain volumes. Leveraging our library of 1,241 manually traced fetal fMRI images from 207 fetuses, we trained a Convolutional Neural Network (CNN) that achieved excellent performance across two held-out test sets from separate scanners and populations. Furthermore, we unite the auto-masking model with additional fMRI preprocessing steps from existing software and provide insight into our adaptation of each step. This work represents an initial advancement towards a fully comprehensive, open-source workflow, with openly shared code and data, for fetal functional MRI data preprocessing.

Keywords Fetal · fMRI · Functional imaging · Brain segmentation · Deep learning · Convolutional neural network · Open-source software

Introduction

Resting-state functional magnetic resonance imaging (rs-fMRI) has emerged as a powerful tool for studying the brain's network architecture development. In recent years, this methodology has been applied to the human brain *in utero*, opening a window into a functional growth period that was otherwise inaccessible. Studying fetal fMRI has the potential to illuminate the nature and manner in which the brain's network architecture is initially assembled, affording powerful new insights into neurodevelopmental origins (Jakab et al., 2014, 2015; Schöpf et al., 2012; Thomason et al., 2017; van den Heuvel et al., 2018). Despite this potential, progress has been slow due, in part, to the lack of image analysis tools tailored for fetal imaging data. Though many tools and software packages exist for fMRI analyses, these tools were designed with adult and child data in mind and encounter specific problems when applied to fetal functional data.

Progress made towards improving fetal MRI methodology can broadly be divided into image acquisition and image post-

---

Chandra Sripada and Moriah Thomason contributed equally to this work.

✉ Saige Rutherford
saige.rutherford@donders.ru.nl

✉ Chandra Sripada

✉ Moriah Thomason

1 Donders Institute, Radboud University Medical Center, Nijmegen, The Netherlands

2 Department of Psychiatry, University of Michigan, MI Ann Arbor, USA

3 Department of Electrical Engineering and Computer Science, University of Michigan, Ann Arbor, MI, USA

4 Department of Psychology, Wayne State University, Detroit, MI, USA

5 Department of Cognitive Neuropsychology, University of Tilburg, Tilburg, The Netherlands

6 Department of Radiology and Biomedical Imaging, Yale School of Medicine, New Haven, CT, USA

7 Department of Statistics and Data Science, Yale University, New Haven, CT, USA

8 Child Study Center, Yale School of Medicine, New Haven, CT, USA

9 Department of Child and Adolescent Psychiatry, New York University School of Medicine, New York, NY, USA

10 Department of Population Health, New York University School of Medicine, New York, NY, USA

processing. Image acquisition improvements have occurred mainly concerning fetal *structural* MRI, particularly in anatomical (primarily T2 HASTE; Half-Fourier Acquisition Single-shot Turbo spin Echo imaging) and diffusion tensor imaging (DTI) (Benkarim et al., n.d.; Biegon & Hoffmann, 2014). Advances have been made in inter-slice motion correction and volume reconstruction (Fogtmann et al., 2014; Gholipour et al., 2017; Studholme, 2011), mapping structural connectivity (Huang et al., 2018; A. Ouyang et al., 2015; M. Ouyang et al., 2018; Qiu et al., 2015; Song et al., 2017; Takahashi et al., 2012), and comparing different MRI signals (Seshamani et al., 2015). These strategies have made it possible to use sparse acquisition sequences, which alleviate movement concerns (Serag et al., 2017; Seshamani et al., 2014), and enable more sophisticated analytic approaches, such as morphometric (Gholipour et al., 2017; Kuklisova-Murgasova et al., 2011; Serag et al., 2012; Shi et al., 2010; Studholme, 2011) cortical folding (Wright et al., 2014), and cytoarchitectural examinations (Miller et al., 2014). In contrast, papers focusing on image post-processing improvements are markedly few and primarily focused on structural imaging. Articles suggesting possible solutions for fetal *functional* MRI data analysis are rare (Scheinost et al., 2018; Seshamani et al., 2014). In this work, we focus on methods for improving image post-processing for fetal functional MRI data.

Challenges associated with the analysis of fetal fMRI have been discussed in a growing number of studies (Schuler et al., 2018; Thomason, 2018; Thomason et al., 2014, 2015, 2017; van den Heuvel et al., 2018) and reviews (Anderson & Thomason, 2013; A. J. Robinson & Ederies, 2018; van den Heuvel & Thomason, 2016; Vasung et al., 2018). These works have focused on image characteristics: motion, size of the fetal brain, susceptibility artifacts introduced by surrounding maternal tissues, and physiological noise of both mother and fetus. Previous work has highlighted essential areas for development, but to our knowledge, no one has proposed a preprocessing pipeline for fetal functional MRI and released it in the open science framework.

The most time-consuming step in preprocessing fetal fMRI is the differentiation of the fetal brain from the surrounding maternal compartment at each acquisition time point. Differentiation is achieved by the generation of an exemplar mask that marks all in-brain voxels. This mask is critical for the entire preprocessing pipeline and subsequent activation or connectivity analyses. Tools developed to segment the adult brain, such as the Brain Extraction Tool (BET) from FSL (Jenkinson et al., 2002) and 3dSkullstrip (3dSS) from AFNI (Cox, 1996), are not effective in generating exemplar masks in fetal imaging for numerous reasons. There is surrounding maternal tissue in fetal images instead of the black background in adult brain images. The fetal brain is not in a standard orientation and shape assumptions do not hold making the atlas-based extraction priors of adult brain segmentation tools inapplicable. As a result, previous studies have relied on the manual generation of brain masks (Thomason et al., 2013, 2014, 2015, 2017; van den Heuvel et al., 2018). While manual methods are tedious and time-consuming, to date, they have been the predominant approach to achieve acceptable standards.

Here, we present an automated approach to the problem of fetal brain segmentation from surrounding tissue. Leveraging a large corpus of manually traced human fetal fMRI masks, we trained a convolutional neural network (CNN) to replace this labor-intensive preprocessing step. CNN's are a powerful tool for effectively identifying complex, non-linear patterns in spatially structured high-dimensional datasets (Lecun et al., 1998). They are increasingly utilized in image processing applications in both medical and non-medical settings (Egmont-Petersen et al., 2002; Falk et al., 2018; Zeiler & Fergus, 2014). In the context of fetal brain segmentation, prior work has investigated the application of CNN's to segment fetal structural T2-weighted volumes (Ebner et al., 2020; Ison et al., 2012; Khalili et al., 2019; Klinder et al., 2015; Link et al., 2017; Makropoulos et al., n.d.; Payette et al., 2021; Rajchl et al., 2016; Salehi et al., 2017, 2018; Serag et al., 2017; Tourbier et al., 2017). These models, however, were developed to segment the fetal brain from anatomical images and do not translate to functional time series data. Compared to structural data, functional data are typically lower resolution and due to movement require a larger quantity of individual segmentations. The majority of existing fetal MRI datasets that have been utilized for deep learning brain segmentation are collected in clinical settings where the focus is assessing fetal brain anatomy. This study uses data from a scientific research program to characterize the development of functional neural systems beginning *in utero*. The scanning protocol employed in this study prioritized data collection of functional MRI and most functional scanning sessions do not have an anatomical scan. Therefore, the existing structural brain segmentation models are not sufficient for the fetal functional MRI data used in this work. Here, we extend prior work by developing and validating a tool for automatically segmenting the fetal brain from functional MRI data. Ultimately, we connect our auto-masking model with an automated version of a previously manual preprocessing workflow. An overview of the entire suggested preprocessing stream is provided in Fig. 1.

All code discussed in this paper, along with a protocol and a tutorial python notebook, is available on GitHub (https://github.com/saigerutherford/fetal-code). All raw volumes and manually drawn brain masks used for training the convolutional neural network are hosted on OpenNeuro.org (https://openneuro.org/datasets/ds003090). This pipeline addresses the challenges mentioned above and achieves preprocessed fetal time-series data with significantly reduced manual effort. Our work embodies an initial set of guidelines
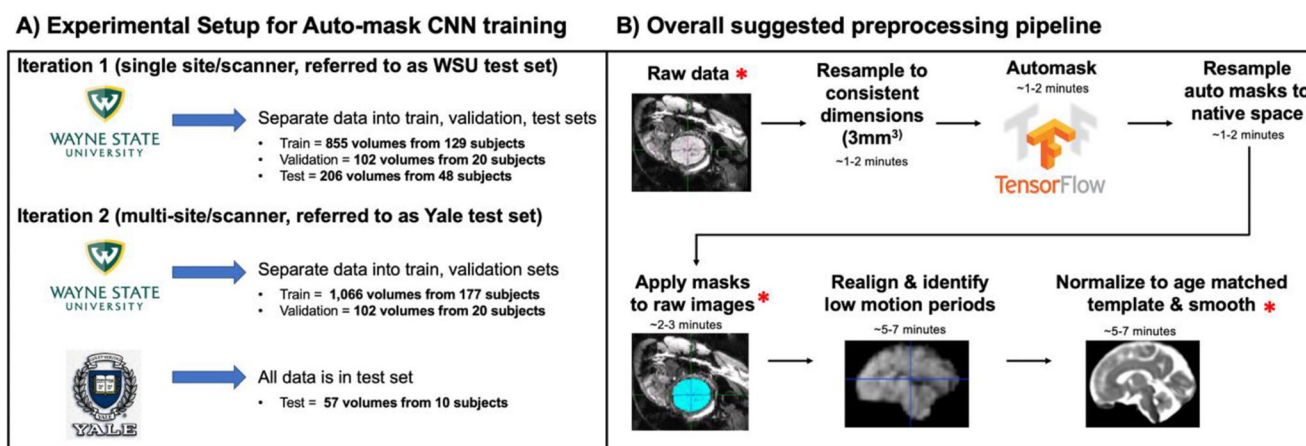
Fig. 1 Overview of the experimental pipeline for training, validation, and testing of the convolutional neural network (CNN) auto-mask model and the proposed preprocessing pipeline. (A) Details of how data were separated into training, validation, and test sets. Two iterations of the auto-mask CNN model were run to compare single-site results (iteration 1) with multi-site results (iteration 2). (B) All steps in the proposed preprocessing stream are shown, with a red asterisk representing where visually quality checking data is recommended. This workflow can be run as shell scripts from the command line and allows for user flexibility

for fetal fMRI preprocessing. The associated protocols are expected to evolve through improvements by the user community in response to new knowledge and innovations in the field.

Methods

Participants and Data

Resting-state functional MRI was obtained from two cohorts, Wayne State University (WSU) and Yale University. Eligible participants were at least 18 years of age, assessed as having uncomplicated, singleton pregnancies, and had no contraindications for MRI. WSU cohort consists of 197 fetuses (gestational age 24–39 weeks, M=30.9, SD=4.2). Twenty-one of these fetuses were scanned at two-time points *in utero*. Both time points are included in this study; however, they are counted as a single subject. WSU fetal MR examinations were performed on a Siemens Verio 3T scanner using an abdominal 4-Channel Flex Coil. Scanning protocols were acquired using echo-planar sequence (TR/TE: 2000/30; 4mm slice thickness, axial, interleaved ascending slice order, 360 volumes). Multi-echo resting-state sequences were also collected in a portion of these subjects (TR/TEs: 2000/18,34,50). The Yale University cohort contains ten fetuses scanned twice longitudinally (gestational ages 30-36 weeks, M=32.7, SD=1.9). The Yale scanner was a Siemens Skyra 3T using a 32-channel abdominal coil (TR/TE:2000/30; 3mm slices, 32 slices parallel to the bicommissural plane, 150 volumes).

Due to the lack of tools for automated segmentation of the fetal brain, research personnel were trained to manually draw fetal brain masks using BrainSuite software (Shattuck &

Leahy, 2002). In line with prior work, manually generated brain masks are used to judge the accuracy of automated segmentation methods in the present analysis.

Auto Masking

Experimental Pipeline

Due to multiple manually drawn masks per subject, WSU data were randomly separated at the subject level into training, validation, and test sets with 129, 20, and 48 subjects (855, 102, and 206 volumes), respectively. The training set was used to optimize the model. The validation set was used to gauge the network's generalization performance during training and determine when to stop training. The test set is held out and used after training was completed to evaluate the model's performance.

For those interested in using this model on unlabeled data from potentially unseen MRI scanners, we wanted to further demonstrate our model's generalizability. We re-trained the CNN a second time using the WSU data as training and validation sets (training = 177 subjects/1,066; validation = 20 subjects/102 volumes) and an additional fetal functional dataset (referred to as the Yale test set; 10 subjects/57 volumes) collected in a separate population (Yale University, New Haven, CT) on a different MRI scanner. An overview of the data splitting into single scanner and multi-scanner test sets is shown in Fig. 1A. Minimal preprocessing of the data included resampling and zero-padding the images to consistent voxel sizes ($3mm^3$) and dimensions ($96 \times 96 \times 37$). Functional time-series data are input to the auto-masking neural network as individual timepoints (i.e., a 3D NIFTI rather than a 4D NIFTI).

Network Architecture

The U-Net style CNN network architecture (Falk et al., 2018) implemented in this pipeline was adapted from Salehi et al. (2017, 2018). The architecture features repeated blocks of 3 × 3 convolutions followed by the ReLu activation function assembled into a contracting path, followed by an expanding path. In the contracting path, every second convolution is followed by a 2 × 2 max pooling operation. In the expanding path, every second convolution is followed by a 2 × 2 upsampling process using nearest-neighbor interpolation. Every other feature map in the contracting path is concatenated along the depth dimension to the corresponding map in the expanding path, helping the network learn the appropriate location of the output mask. The final layer is convolved with two 1 × 1 filters to produce an output mask with channels equal to the number of output classes.

The network separates 3D image volumes into 2D axial slices and operates on each slice independently. We chose to implement a 2D rather than 3D network to reduce computational costs. This model includes steps for converting raw NIFTI images into a format readable by the network and steps for converting the network's output into a NIFTI-formatted, 3D brain mask.

The model was implemented using Tensorflow (version 1.4.1). Training and testing of the network were performed using a GPU, but CPU testing times were also evaluated to provide an additional reference point.

Training Procedures

During training, the weights in a CNN are minimized with respect to a loss function that determines how well the network is learning from the training data. We optimized our network via per-pixel cross-entropy, with weights determined using the Adam Optimizer (Kingma & Ba, 2014). Adam is a first-order gradient method that updates the weights adaptively based on previous and current gradients. Even using an adaptive optimizer, we found that using a learning rate decay improved performance. The initial learning rate was set to 0.0001 with an exponential decay rate of 0.9, applied every 10,000 batches. The model was trained until performance no longer improved on the validation set. The 2D axial slices in the training data were augmented through 90-degree rotations and horizontal and vertical flips. While augmentations were done on 2D slices, the same rotations and flips were applied to all slices to preserve the 3D shape. These augmentations capture the non-standard orientation of the brain in fetal volumes.

Evaluation

The evaluation process was performed over multiple steps. First, we evaluated our network's ability to mask the fetal brain using the Dice coefficient. The Dice coefficient measures the percent overlap between two regions: the predicted brain region and the actual brain region. It is defined between zero and one, where zero means there is no overlap between the two areas, and one means the two regions are identical. We also report the Jaccard index, Hausdorff surface distance (Karimi & Salcudean, 2019), sensitivity (true positive rate: brain voxels are correctly identified as brain), and specificity (true negative rate: nonbrain voxels are correctly identified as nonbrain) of our network on the WSU and Yale held-out test sets. Detailed mathematical definitions of these metrics can be found in Taha and Hanbury (2015).

After training, we calculated Dice coefficients for all auto-masks that have a manually drawn counterpart, though we report values only for volumes in the test data, as performance within the train and validation datasets does not reflect model performance on new data and CNNs are known to achieve near perfect performance within the training set.

Comparison to Other Methods

Also, to aid in interpretation and benchmarking obtained evaluation metrics, we performed a secondary analysis to demonstrate that current methods for brain extraction perform poorly when applied to fetal data. We used the Brain Extraction Tool (BET) implemented in FSL, 3dSkullstrip from AFNI, and the fetal anatomical U-Net from Salehi et al. (2018), to benchmark our model. Of note, evaluation metrics can be improved by separating testing data into challenging versus non-challenging images (Salehi et al., 2017). This splitting approach is not used here as this diminishes the representativeness of estimates when applied across complex and varied data sets.

Failure Analysis

After the model performance was evaluated, we conducted a failure analysis to discover patterns of intrinsic image characteristics that may influence auto-masking performance. First, we examined the relationship between the evaluation metrics and gestational age. Next, we evaluated whether image artifacts, brain size ratio (brain volume relative to the entire image volume), or brain position in the center of imaging space influence model performance by qualitatively examining images with dice coefficients falling below 0.9.

Application of Auto-masks

Auto-masks are output as spatial probability estimates, wherein voxel values equal to one correspond to the highest probability of being brain. Probability map brain masks were then

clustered, thresholded, and binarized. These steps are taken to discard small, non-brain clusters that may have been included in the probability map brain mask. Binarized masks were then resampled back into the subject's native space and applied to the native image using an image multiplier, resulting in segmented brain volumes corresponding to each fetal fMRI data timepoint.

Other Preprocessing Steps

In addition to segmentation of the fetal brain from surrounding maternal tissue, fetal imaging preprocessing requires several additional steps: motion denoising, realignment of volumes within a time series, and group-level normalization to a fetal template. These steps are challenging because frame-to-frame displacement is elevated in fetal studies, and the fetal brain is typically not in a single standard orientation.

In prior studies by our group, a reference frame from each quiescent period was chosen to be masked. The mask would then be applied to every volume within the low movement period, not only the volume it was drawn on. Due to the time-consuming nature of manual masking, it was not feasible to mask every volume. Auto-masking's central goal is to mask an entire time-series in a fraction of the time it takes to manually draw a mask for a single volume. Realigning all volumes within a time series creates parameters (mean framewise displacement) that are used to identify low movement periods that are usable for further activation or connectivity analyses.

Typical realignment of fMRI data is done on full time series using the middle volume (in time) as a reference volume. Due to notably high movement across a fetal time series, using a single reference volume for realignment may not be optimal. In this pipeline, instead of using a 4D NIFTI file in the realignment step, we realign 3D volumes such that a new reference volume is always selected at each iteration. We selected the FLIRT FSL realignment tool (Jenkinson et al., 2002). FLIRT first estimates a linear transformation between volume n and the reference volume ($n + 1$). This step is repeated across all 3D volumes in the time-series and each of these n to ($n + 1$) transformation matrices are applied to each volume of the time series to produce a new data set comprised of realigned volumes. This step also creates a text file and plot that summarizes the six rigid-body realignment parameters across time, which can be subsequently used to identify motion outliers and perform motion censoring in later processing stages. Here, we applied the fsl_motion_outliers routine as a data-driven means of defining high and low fetal movement periods.

After masking and realignment, the individual time point 3D volumes are merged back into a 4D NIFTI, moved into group template space using linear warping, and spatially smoothed. Flexibility is built into the pipeline such that the user can define whether data are normalized to a standard reference template or age-specific fetal templates, see Serag et al. (2012). A linear normalization is implemented via FLIRT (Jenkinson et al., 2002). After normalization, all volumes are spatially smoothed with a user-specified Gaussian kernel.

Importantly, all software tools used within this preprocessing pipeline (TensorFlow, Python, AFNI, FSL) are free and open source. All commands can be implemented in a shell script, which can be run from the command line.

Quality Control

While this methodology employs fully automated techniques for preprocessing fetal resting-state fMRI data, manual quality assurance processes are necessary at crucial transition points throughout the pipeline. Specifically, our standard process includes an initial review of raw time-series data, screened as a movie. Initial inclusion criteria are that the brain is in the field of view and unobstructed by artifacts and that within the time series, there are periods of minimal fetal movement. We exclude data not meeting these criteria. However, most Wayne State University data passes this stage because long scan durations are used, and fetuses rapidly cycle through quiescent states.

Additional steps in the quality control protocol are implemented after auto-masking, realignment, and normalization. At these stages, time-series data are again visually inspected to assure that no errors were introduced during these stages of preprocessing. Several parameters that are automatically generated during the pipeline should be used in complement to manual quality checking. These parameters include realignment parameters, motion plots, and metrics from the fsl_motion_outliers command.

Results

Auto-masking Performance

Our CNN auto-mask model achieved high accuracy when delineating fetal brain from surrounding structures in fetal brain fMRI images. We evaluated the model on two held-out test sets. Applied to the Wayne State University (WSU – 206 volumes from 48 unique subjects) and Yale (57 volumes from 10 subjects) test cohorts, the models achieved a per-volume average dice coefficient of 0.94 and 0.89, respectively. The CNN's performance in terms of Dice coefficient, Jaccard coefficient, Hausdorff surface distance, sensitivity, and specificity across both test sets is summarized in Table 1. Figure 2 provides examples of high fidelity between manual and auto-masks in both test sets.
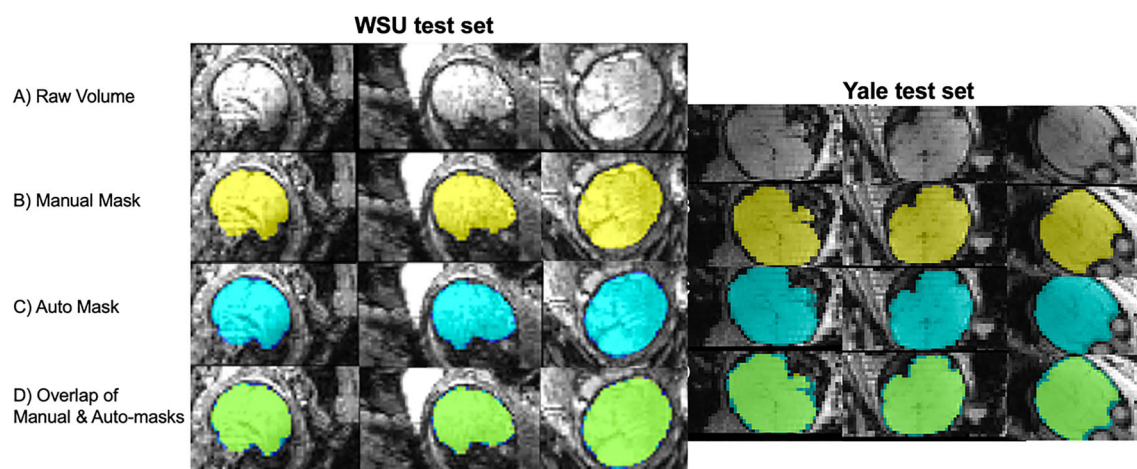
Fig. 2 Comparison of manual and automated masks. (A) Raw volume; (B) Hand-drawn mask; (C) Auto mask; (D) Conjunction of hand drawn (yellow) and auto (blue) masks, overlap between hand and auto masks shown in green. *WSU data collected in Detroit, MI, at Wayne State University. Yale data collected in New Haven, CT at Yale University*

Comparison to Adult Auto-masking Tools

As a point of reference for the Dice coefficient achieved by our method, we performed an additional analysis using several existing auto-masking tools: Brain Extraction Tool (BET), 3dSkullstrip (3dSS), and the fetal anatomical U-Net from Salehi et al. (2017, 2018). As expected, when applied to the same test set data, these tools performed significantly worse. Evaluation metrics are reported in Table 1, and examples of the masks generated using these tools are shown in Fig. 3. The fetal anatomical U-Net masks were empty in most cases, and therefore the evaluation metrics failed and are not reported in Table 1. In Fig. 2D, functional and structural fetal MRI data are shown side by side to demonstrate the substantial differences in image characteristics between modalities. These results highlight that in fetal functional data, areas of the maternal compartment have high contrast boundaries and varied image intensity creating challenges for standard adult masking routines and severely compromising performance. A network that has

been adapted for and trained on fetal functional data is necessary to achieve high-quality functional brain masks.

Age & Data Quality Failure Analysis

Examination of the effect of fetal age on the algorithm's performance revealed a significant positive association between Dice coefficient and gestational age (Fig. 4). This relationship is likely due to older fetuses having larger brain volumes (Crum et al., 2006). The correlation between gestational age and brain mask volume is highly significant and the correlation between Dice coefficient and brain mask volume is also highly significant. Auto-masks that had "low" Dice coefficients (< 0.9) were visually inspected to further understand reasons for suboptimal performance. Two out of 48 subjects in the WSU test set (4 out of 206 volumes) and two out of ten subjects (6 out of 57 volumes) in the Yale test set had dice coefficients less than 0.9. Examples of the raw and brain mask data from these low Dice subjects are shown in Fig. 3. In this study's GitHub (https://github.com/saigerutherford/fetal-code), there are video

Table 1 Performance of auto-mask model and existing masking software evaluated in two independent test sets from Wayne State University (WSU) and Yale University. Values reported are the mean (s.d.) within the test sets

	WSU Auto-mask	WSU BET	WSU 3dSS	Yale Auto-mask	Yale BET	Yale 3dSS
Dice	0.94 (+/- 0.067)	0.22 (+/- 0.13)	0.24 (+/- 0.10)	0.89 (+/- 0.13)	0.22 (+/- 0.06)	0.25 (+/- 0.08)
Jaccard	0.89 (+/- 0.069)	0.13 (+/- 0.086)	0.14 (+/- 0.07)	0.82 (+/- 0.13)	0.13 (+/- 0.03)	0.15 (+/- 0.05)
Hausdorff Distance (mm)	12.11 (+/- 22.4)	112.6 (+/- 36.7)	103.3 (+/- 26.0)	19.25 (+/- 14.5)	95.2 (+/- 30.1)	92.8 (+/- 24.2)
Sensitivity	0.90 (+/- 0.04)	0.13 (+/- 0.08)	0.14 (+/- 0.07)	0.84 (+/- 0.12)	0.12 (+/- 0.03)	0.15 (+/- 0.05)
Specificity	0.99 (+/- 0.0007)	0.99 (+/- 0.003)	0.99 (+/- 0.002)	0.99 (+/- 0.002)	0.99 (+/- 0.004)	0.99 (+/- 0.002)

Auto-mask is our proposed model. BET is Brain Extraction tool from FSL, 3dSS is 3dSkullStrip tool from AFNI
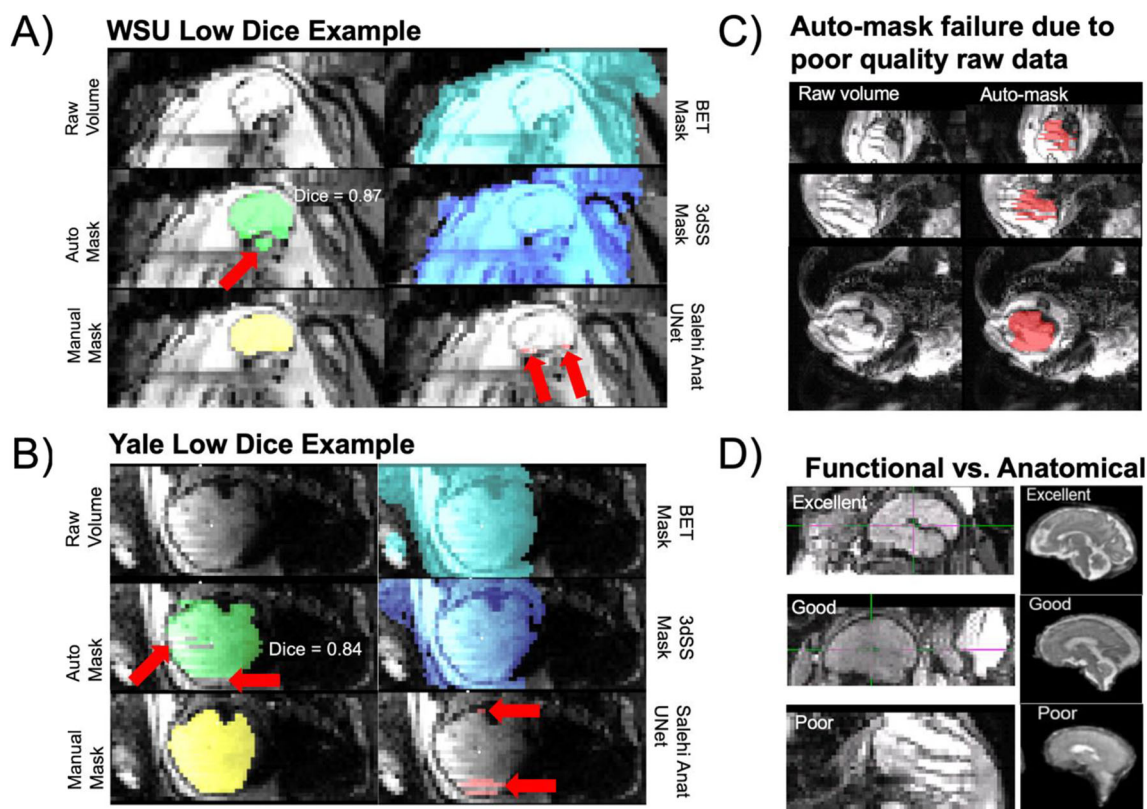
Fig. 3 Failure analysis and comparison of functional with structural fetal MRI data. (A) All data of the WSU test set subject with the lowest Dice coefficient (0.87). (B) All data of the Yale test set subject with the lowest Dice coefficient (0.84). The BET, 3dSkullStrip, and Anatomical U-Net masks do not adequately capture the fetal brain's boundary in both the WSU and Yale case. (C) Extreme failure of the auto-mask model, due to very poor quality of the raw data. (D) Comparison of data quality between fetal functional and structural MRI data to understand why models designed for brain segmentation of anatomical data do not necessarily translate to functional data. Structural fetal MRI image used with permission from Payette et al. (2021)

examples of successful and failed auto-masks which better explore the 3-dimensional space. We found that aliasing negatively impacted auto-mask performance (example shown in Fig. 3C) and that the algorithm also performed more poorly for images in which the brain had a large displacement from the image origin.

Computational Time and Hardware

An often-noted property of deep learning models is their ability to surpass human speed in completing complex tasks substantially. Our auto-masking model illustrates this acceleration. The training time refers to the wall-clock time it took our CNN model to converge to a set of weights that minimize the Dice coefficient on the validation set. The training was stopped after signs of overfitting were observed, that is, performance on the validation set was no longer increasing. The model's total training time was 3 h and 46 min on a GeForce GTX 1080 Ti GPU. Testing time refers to the time it takes to run the CreateMask.py script to load the raw input volume (NIFTI to NumPy array conversion) and output a predicted auto-mask (as a NIFTI file). The testing time to generate a new

auto brain mask is approximately 0.2 s if using a GPU, and 2.5 s if using CPU.

Other Preprocessing

An additional benefit of our approach for auto-masking all individual volumes is that multiple realignment strategies are now possible. After the fetal brain has been extracted, the time series data can enter more typical preprocessing steps for which child/adult tools have been developed. The main difference when applying these tools is that the fetal brain can have any orientation (meaning image origins may be very far apart), and fetal data exhibits substantially increased head motion.

Concerning the problem of elevated head motion, errors introduced by movement cannot, at present, be fully corrected in the fMRI time series. This fact necessitates the application of stringent criteria for retaining only low-motion volumes to ensure data integrity, which is the approach taken by most studies to date (Jakab et al., 2014, 2015; Schöpf et al., 2012; Thomason et al., 2017; van den Heuvel et al., 2018). The correct framewise displacement threshold to censor high
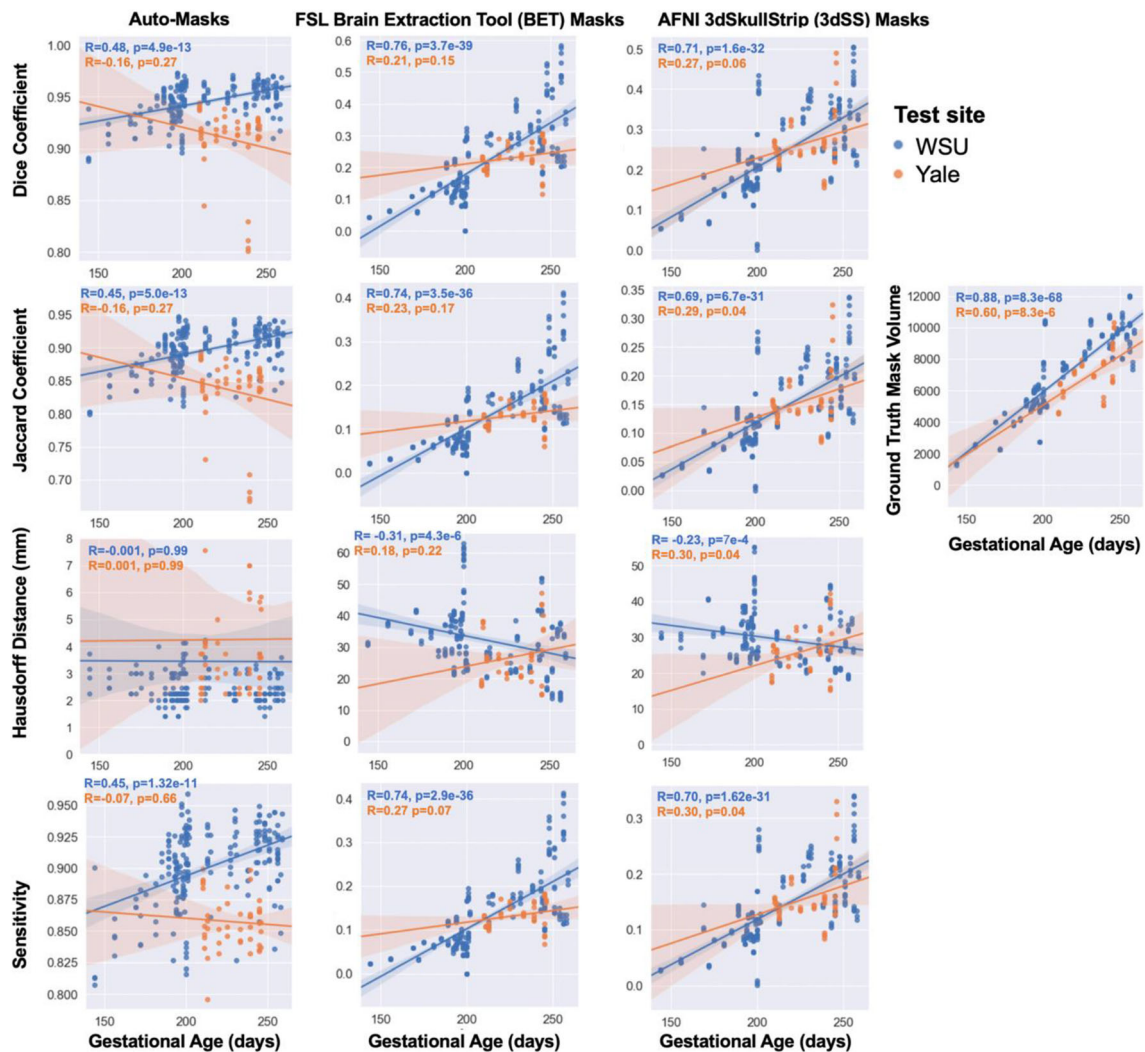
Fig. 4 Evaluation of auto-masking model. The relationships between fetal gestational age (in days) at scan are shown on the x-axes and auto-masking performance in the WSU test sets (blue) and Yale test set (orange) on the y-axes. We calculated the evaluation metrics on a per-volume basis, however, the values shown here are on a per-subject basis to examine the relationships with age

motion volumes from a BOLD time-series is an active discussion topic. The answer to the framewise displacement threshold debate is complex, and a concrete solution is outside the scope of this work. However, we tested many different censoring thresholds, and the resulting amounts of data preserved at each censoring threshold are shown in Fig. 5.

Discussion

Fetal functional MRI is an emerging field with great potential to improve understanding of human brain development. The number of papers published in this area has seen a 5-fold increase since 2010. However, methodologies for processing these complex data sets have not kept pace, and the absence of a standard publicly available processing pipeline for these data has been especially notable. Here we address this gap and

present a solution to the most cumbersome processing step, localization, and extraction of the fetal brain from surrounding tissue for each volume of a functional time series. This is a necessary step in processing human fetal fMRI data. Until now, it has been a rate-limiting factor in accurate, automated processing of fetal fMRI BOLD time-series data. This model was built by pairing a set of 1,241 hand-drawn fetal 3D brain masks with a deep learning algorithm, a U-Net convolutional neural network. Pipeline code and documentation are made available through GitHub (https://github.com/saigerutherford/fetal-code). Training, validation, and test data sets (raw volumes, hand-drawn masks, and auto-masks) are available in BIDS format on OpenNeuro.org (https://openneuro.org/datasets/ds003090). It is hoped that the release of an easy-to-use, efficient, validated auto-masking pipeline will reduce barriers for new labs to enter this area while also providing experienced labs opportunities for further optimization of their

Fig. 5 Motion summary. Framewise displacement (FD) censoring thresholds from 0.5mm – 3.8mm were tested, and the amount of data remaining for each subject is shown at each threshold. Each subplot represents a different FD threshold (bolded above the subplot). Each subject represents a point on the x-axis, and the y-axis shows the time, in minutes, remaining after removing high movement volumes
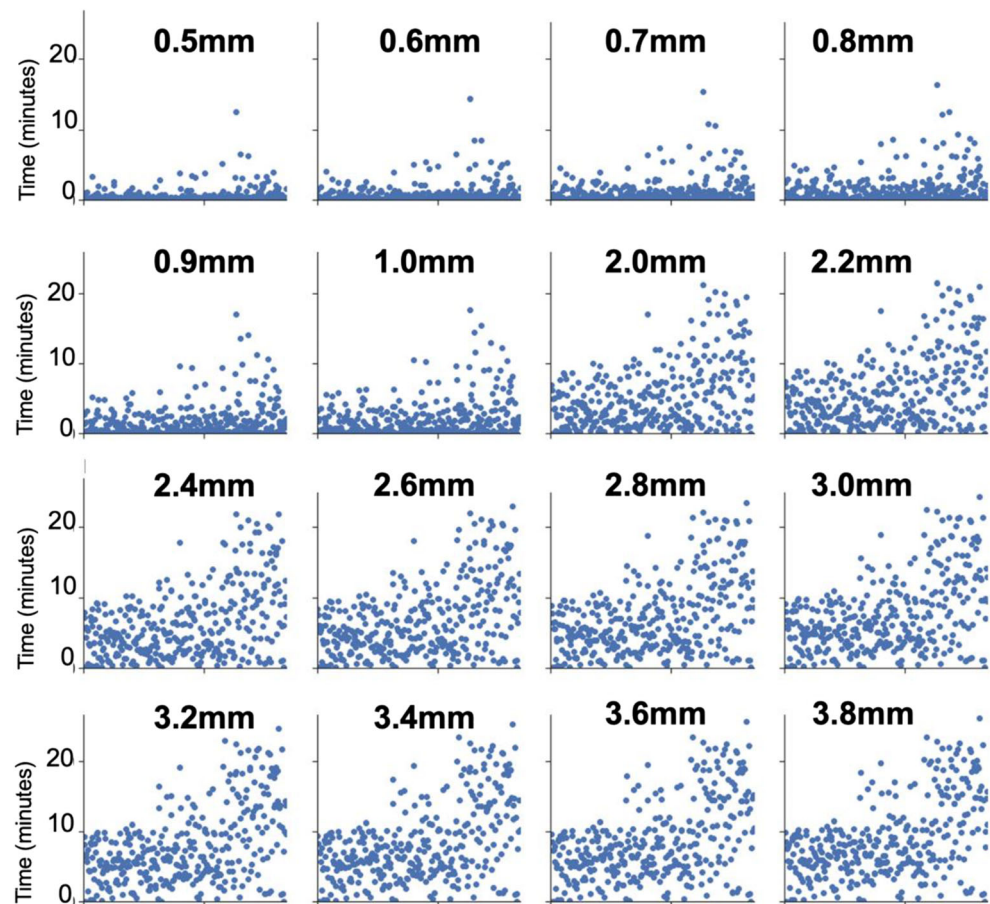
independently developed approaches. Future integration of this auto-masking model will be incorporated into BioImage Suite (https://bioimagesuiteweb.github.io/webapp/), a web-based image analysis tool that will make use of this model even more accessible to the field.

Recent pioneering work has established that deep learning approaches are effective in fetal *structural* brain segmentation (Ebner et al., 2020; Ison et al., 2012; Khalili et al., 2019; Klinder et al., 2015; Link et al., 2017; Makropoulos et al., n.d.; Payette et al., 2021; Rajchl et al., 2016; Salehi et al., 2017, 2018; Serag et al., 2017; Tourbier et al., 2017). However, fetal *functional* imaging presents a distinct set of constraints and therefore requires a different solution. In particular, the inherently lower resolution and contrast of functional time-series data, and the 4-dimensional nature of the data (~ 360 3D volumes per subject), make this a more challenging problem. Here we applied CNN methods to the most extensive fetal fMRI data set reported to date, 207 fetuses, and derived a novel method for accurate and reliable segmentation of the fetal brain from surrounding maternal tissues within a fraction of a second, 94 % accuracy in 0.2 s. These encouraging results are partially attributable to the large set of manually traced human fetal fMRI masks used to train the CNN, consistent with studies showing the correlation between CNN

performance and the training data size (Cho et al., 2015; Verghese et al., 2018).

The auto-masking model exhibited strong generalizability, which is particularly important given known biases in CNN models trained on imaging datasets from a single site/population (Tommasi et al., 2015; Zech et al., 2018). First, the trained CNN correctly classified data at two held out sets with 94 and 89 %, very similar to the training data. This finding suggests that the CNN is robust to variations in experimental procedures, scanner settings, and populations studied. Also, the training images were drawn from a wide fetal age-range, which should also enhance generalizability across fetal samples encompassing different ages.

Our aim in this project goes beyond brain segmentation; we sought to construct an automated version of a previously manual preprocessing pipeline that is standardized but flexible and readily deployable across multiple data sources. Thus, our pipeline begins with an auto-masking step, leverages existing algorithms (FSL, AFNI) that assist with applying the auto-masks, performs frame-to-frame alignment, normalization to a user-defined template, and finally smoothing. The user is referred to publicly available multi-age fetal brain templates (Serag et al., 2012) and can easily configure the tool to modify or eliminate steps. Our code's flexibility allows for potential

users to mix and match the portions of this pipeline they wish to use. For example, a user could choose an alternative re-alignment algorithm as the first step then apply the auto-masking step. The tool's open construction will allow the incorporation of future processing advances, such as surface-based registration and additive motion correction strategies.

Our work has several limitations. First, deep learning methods perform classification in high dimensional space, and consequently, results can be a "black box" with little opportunity for interpretation of axes (Cabitza et al., 2017). However, this limitation should be viewed in the context of our goal: to automatically perform a task that can take trained individuals many hours to perform manually. We are less interested in understanding computer-based brain masking mechanisms and instead focus on algorithm performance in out of sample data sets.

Another limitation is that we use direct warping, or normal-ization, of functional data to a group-averaged anatomical template. It is not clear that alternatives would improve regis-tration significantly, but one might expect registration to subject-specific anatomy, then to template space, to be a pre-ferred approach. This approach's challenge is that obtaining high-quality subject-specific high-resolution anatomical im-ages presents a different set of challenges that have been ad-dressed elsewhere (Studholme, 2015). Prior studies of fetal anatomical development (Nunes et al., 2018) demonstrate that even when trained experts apply the most advanced tech-niques to these data, there is still significant data loss and image blurring where motion effects, image artifacts, or lack of tissue contrast compromise data quality. This example ex-tends to other parts of the pipeline, where alternative opti-mized preprocessing strategies could be used. However, this work's objective is not to serve as a final fetal functional MRI preprocessing endpoint but as a backbone upon which further development can follow.

A final limitation is that ours is not a fully automated pipe-line. It requires human supervision and quality checking at several stages, which requires a certain quantity of time and level of expertise from the human supervisor. Fortunately, however, the level of involvement and associated expertise required is limited and includes looking for overt errors when viewing processed images as a continuous movie, which takes approximately 1–5 min per functional run. Assuming the en-tire time for running this pipeline is 15–30 min, including human effort, this is a 60-fold time reduction over prior methods, with manual tracing in particular requiring extensive time and substantial expertise (Thomason et al., 2017; van den Heuvel et al., 2018).

There are many future directions to continue improving fetal functional MRI preprocessing and data analysis. Regarding the auto-masking portion of the pipeline, future contributors to this pipeline should consider using 3-D instead of 2-D convolutions and directly using the probability masks

(instead of binarizing them) to utilize the model's uncertainty estimates. Regarding the realignment and quality checking portions, it may be useful to calculate the Dice coefficient between consecutive volumes (after realignment) as a metric of data quality – a dice coefficient equal to one represents two perfectly images. Other advanced alignment algorithms to register the data to template space should also be investigated and adapted for fetal data (Bozek et al., 2018; Robinson et al., 2014, 2018). It is important to benchmark new brain segmen-tation models against existing solutions. In order for benchmarking to become common practice, pre-trained models must be shared along with the code. There are expen-sive computational costs to re-training a model every time a new user wants to use it, and these costs can be eliminated by sharing pre-trained model weights. Furthermore, the code to test existing brain segmentation models must be accompanied by clear directions for setup and testing, and ideally a tutorial showing step-by-step how to successfully run the code. We encourage users who test the auto-mask pipeline introduced in this paper in their fetal data sets to report performance metrics back to the study's GitHub page. Also, if new users re-train the auto-mask model by adding additional labeled data from their site, it would be beneficial to the fetal imaging commu-nity to share these pre-trained models on our study's GitHub (via a pull request to our repository). A contributing guide on the GitHub page provides instructions for sharing additional pre-trained models. This will ensure that new users are able to access the most current, best performing auto-masking model and will lower the barriers to entry for fetal functional imag-ing. Finally, it would be beneficial to follow in the footsteps of an extremely successful adult MRI preprocessing pipeline, fMRIPrep (Esteban et al., 2018), and package the code into a container, either Docker (Merkel, 2014) or Singularity (Kurtzer, 2016). A container is a standard unit of software that packages up code and all its dependencies, so the application runs quickly and reliably from one computing environment to another (*What Is a Container?*, n.d.).

Large-scale, often multi-center, projects are becoming the new norm, and these require validated, standardized process-ing pipelines of the kind that we have developed in this work. The Developing Human Connectome Project provides just one example of a large-scale study that includes a fetal func-tional MRI component (Bozek et al., 2018; Fitzgibbon et al., 2020; Harms et al., 2018; Makropoulos et al., n.d.) and many more large-scale fetal fMRI initiatives will likely emerge in the coming years.

In sum, in this work, we leverage deep learning methods in the most significant sample of fetal fMRI data published to date to address the challenging brain segmentation problem in fetal fMRI. We unite our novel auto-masking tool with other prepro-cessing steps to initialize the first complete open-source solution to preprocessing raw fetal functional MRI time-series data.

References

Anderson, A. L., & Thomason, M. E. (2013). Functional plasticity before the cradle: A review of neural functional imaging in the human fetus. *Neuroscience & Biobehavioral Reviews*, 37(9, Part B), 2220–2232. https://doi.org/10.1016/j.neubiorev.2013.03.013.

Benkarim, O. M., Sanroma, G., Zimmer, V. A., Muñoz-Moreno, E., Hahner, N., Eixarch, E., Camara, O., Ballester, M. A. G., & Piella, G. (n.d.). Toward the automatic quantification of in utero brain development in 3D structural MRI: A review. *Human Brain Mapping*, 38(5), 2772–2787. https://doi.org/10.1002/hbm.23536.

Biegon, A., & Hoffmann, C. (2014). Quantitative magnetic resonance imaging of the fetal brain in utero: Methods and applications. *World Journal of Radiology, 6*(8), 523–529. https://doi.org/10.4329/wjr.v6.i8.523.

Bozek, J., Makropoulos, A., Schuh, A., Fitzgibbon, S., Wright, R., Glasser, M. F., Coalson, T. S., O'Muircheartaigh, J., Hutter, J., Price, A. N., Cordero-Grande, L., Teixeira, R. P. A. G., Hughes, E., Tusor, N., Baruteau, K. P., Rutherford, M. A., Edwards, A. D., Hajnal, J. V., Smith, S. M., … Robinson, E. C. (2018). Construction of a neonatal cortical surface atlas using Multimodal Surface Matching in the Developing Human Connectome Project. *NeuroImage, 179*, 11–29. https://doi.org/10.1016/j.neuroimage.2018.06.018.

Cabitza, F., Rasoini, R., & Gensini, G. F. (2017). Unintended consequences of machine learning in medicine. *JAMA, 318*(6), 517–518. https://doi.org/10.1001/jama.2017.7797.

Cho, J., Lee, K., Shin, E., Choy, G., & Do, S. (2015). How much data is needed to train a medical image deep learning system to achieve necessary high accuracy? *ArXiv:1511.06348 [Cs]*. http://arxiv.org/abs/1511.06348.

Cox, R. W. (1996). AFNI: Software for analysis and visualization of functional magnetic resonance neuroimages. *Computers and Biomedical Research, 29*(3), 162–173.

Crum, W. R., Camara, O., & Hill, D. L. G. (2006). Generalized overlap measures for evaluation and validation in medical image analysis. *IEEE Transactions on Medical Imaging, 25*(11), 1451–1461. https://doi.org/10.1109/TMI.2006.880587.

Ebner, M., Wang, G., Li, W., Aertsen, M., Patel, P. A., Aughwane, R., Melbourne, A., Doel, T., Dymarkowski, S., De Coppi, P., David, A. L., Deprest, J., Ourselin, S., & Vercauteren, T. (2020). An automated framework for localization, segmentation and super-resolution reconstruction of fetal brain MRI. *NeuroImage, 206*, 116324. https://doi.org/10.1016/j.neuroimage.2019.116324.

Egmont-Petersen, M., de Ridder, D., & Handels, H. (2002). Image processing with neural networks—A review. *Pattern Recognition, 35*(10), 2279–2301.

Esteban, O., Markiewicz, C. J., Blair, R. W., Moodie, C. A., Isik, A. I., Erramuzpe, A., Kent, J. D., Goncalves, M., DuPre, E., Snyder, M., Oya, H., Ghosh, S. S., Wright, J., Durnez, J., Poldrack, R. A., & Gorgolewski, K. J. (2018). fMRIPrep: A robust preprocessing pipeline for functional MRI. *Nature Methods, 1*. https://doi.org/10.1038/s41592-018-0235-4.

Falk, T., Mai, D., Bensch, R., Çiçek, Ö, Abdulkadir, A., Marrakchi, Y., Böhm, A., Deubner, J., Jäckel, Z., Seiwald, K., Dovzhenko, A., Tietz, O., Bosco, C. D., Walsh, S., Saltukoglu, D., Tay, T. L., Prinz, M., Palme, K., Simons, M., … Ronneberger, O. (2018). U-Net: Deep learning for cell counting, detection, and morphometry. *Nature Methods, 1*. https://doi.org/10.1038/s41592-018-0261-2.

Fitzgibbon, S. P., Harrison, S. J., Jenkinson, M., Baxter, L., Robinson, E. C., Bastiani, M., Bozek, J., Karolis, V., Grande, C., Price, L., Hughes, A. N., Makropoulos, E., Passerat-Palmbach, A., Schuh, J., Gao, A., Farahibozorg, J., O'Muircheartaigh, S.-R., Ciarrusta, J., O'Keeffe, J. C., … Andersson, J (2020). The developing Human Connectome Project (dHCP) automated resting-state functional processing framework for newborn infants. *NeuroImage, 223*, 117303. https://doi.org/10.1016/j.neuroimage.2020.117303.

Fogtmann, M., Seshamani, S., Kroenke, C., Cheng, X., Chapman, T., Wilm, J., Rousseau, F., & Studholme, C. (2014). A unified approach to diffusion direction sensitive slice registration and 3-D DTI reconstruction from moving fetal brain anatomy. *IEEE Transactions on Medical Imaging, 33*(2), 272–289. https://doi.org/10.1109/TMI.2013.2284014.

Gholipour, A., Rollins, C. K., Velasco-Annis, C., Ouaalam, A., Akhondi-Asl, A., Afacan, O., Ortinau, C. M., Clancy, S., Limperopoulos, C., Yang, E., Estroff, J. A., & Warfield, S. K. (2017). A normative spatiotemporal MRI atlas of the fetal brain for automatic segmentation and analysis of early brain growth. *Scientific Reports, 7*(1), 476. https://doi.org/10.1038/s41598-017-00525-w.

Harms, M. P., Somerville, L. H., Ances, B. M., Andersson, J., Barch, D. M., Bastiani, M., Bookheimer, S. Y., Brown, T. B., Buckner, R. L., Burgess, G. C., Coalson, T. S., Chappell, M. A., Dapretto, M., Douaud, G., Fischl, B., Glasser, M. F., Greve, D. N., Hodge, C., Jamison, K. W., … Yacoub, E. (2018). Extending the Human Connectome Project across ages: Imaging protocols for the Lifespan Development and Aging projects. *NeuroImage, 183*, 972–984. https://doi.org/10.1016/j.neuroimage.2018.09.060.

Huang, W., Bolton, T. A. W., Medaglia, J. D., Bassett, D. S., Ribeiro, A., & Ville, D. V. D. (2018). A graph signal processing perspective on

functional brain imaging. *Proceedings of the IEEE*, *PP*(99), 1–18. https://doi.org/10.1109/JPROC.2018.2798928.

Ison, M., Donner, R., Dittrich, E., Kasprian, G., Prayer, D., & Langs, G. (2012). Fully automated brain extraction and orientation in raw fetal MRI. *Workshop on Paediatric and Perinatal Imaging, MICCAI*, 17–24.

Jakab, A., Kasprian, G., Schwartz, E., Gruber, G. M., Mitter, C., Prayer, D., Schöpf, V., & Langs, G. (2015). Disrupted developmental organization of the structural connectome in fetuses with corpus callosum agenesis. *NeuroImage, 111*, 277–288. https://doi.org/10.1016/j.neuroimage.2015.02.038.

Jakab, A., Schwartz, E., Kasprian, G., Gruber, G. M., Prayer, D., Schöpf, V., & Langs, G. (2014). Fetal functional imaging portrays heterogeneous development of emerging human brain networks. *Frontiers in Human Neuroscience, 8*, 852. https://doi.org/10.3389/fnhum.2014.00852.

Jenkinson, M., Bannister, P., Brady, M., & Smith, S. (2002). Improved optimization for the robust and accurate linear registration and motion correction of brain images. *NeuroImage, 17*(2), 825–841.

Karimi, D., & Salcudean, S. E. (2019). Reducing the hausdorff distance in medical image segmentation with convolutional neural networks. *ArXiv:1904.10030 [Cs, Eess, Stat]*. http://arxiv.org/abs/1904.10030.

Khalili, N., Lessmann, N., Turk, E., Claessens, N., de Heus, R., Kolk, T., Viergever, M. A., Benders, M. J. N. L., & Išgum, I. (2019). Automatic brain tissue segmentation in fetal MRI using convolutional neural networks. *Magnetic Resonance Imaging, 64*, 77–89. https://doi.org/10.1016/j.mri.2019.05.020.

Kingma, D. P., & Ba, J. (2014). Adam: a method for stochastic optimization. *ArXiv:1412.6980 [Cs]*. http://arxiv.org/abs/1412.6980.

Klinder, T., Wendland, H., Wachter-Stehle, I., Roundhill, D., & Lorenz, C. (2015). Adaptation of an articulated fetal skeleton model to three-dimensional fetal image data. *Medical Imaging 2015: Image Processing, 9413*, 94130Q. https://doi.org/10.1117/12.2081139.

Kuklisova-Murgasova, M., Aljabar, P., Srinivasan, L., Counsell, S. J., Doria, V., Serag, A., Gousias, I. S., Boardman, J. P., Rutherford, M. A., Edwards, A. D., Hajnal, J. V., & Rueckert, D. (2011). A dynamic 4D probabilistic atlas of the developing brain. *NeuroImage, 54*(4), 2750–2763. https://doi.org/10.1016/j.neuroimage.2010.10.019.

Kurtzer, G. M. (2016). *Singularity 2.1.2—Linux application and environment containers for science*. Zenodo. https://doi.org/10.5281/zenodo.60736.

Lecun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278–2324. https://doi.org/10.1109/5.726791.

Link, D., Braginsky, M. B., Joskowicz, L., Ben Sira, L., Harel, S., Many, A., Tarrasch, R., Malinger, G., Artzi, M., Kapoor, C., Miller, E., & Ben Bashat, D. (2017). Automatic measurement of fetal brain development from magnetic resonance imaging: new reference data. *Fetal Diagnosis and Therapy*. https://doi.org/10.1159/000475548.

Makropoulos, A., Robinson, E. C., Schuh, A., Wright, R., Fitzgibbon, S., Bozek, J., Counsell, S. J., Steinweg, J., Vecchiato, K., Passerat-Palmbach, J., Lenz, G., Mortari, F., Tenev, T., Duff, E. P., Bastiani, M., Cordero-Grande, L., Hughes, E., Tusor, N., Tournier, J.-D., … Rueckert, D. (n.d.). The developing human connectome project: A minimal processing pipeline for neonatal cortical surface reconstruction. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2018.01.054.

Merkel, D. (2014). Docker: Lightweight Linux containers for consistent development and deployment. *Linux Journal, 2014*(239), 2:2.

Miller, J. A., Ding, S.-L., Sunkin, S. M., Smith, K. A., Ng, L., Szafer, A., Ebbert, A., Riley, Z. L., Royall, J. J., Aiona, K., Arnold, J. M., Bennet, C., Bertagnolli, D., Brouner, K., Butler, S., Caldejon, S., Carey, A., Cuhaciyan, C., Dalley, R. A., … Lein, E. S. (2014).

Transcriptional landscape of the prenatal human brain. *Nature, 508*(7495), 199–206. https://doi.org/10.1038/nature13185.

Nunes, R. G., Ferrazzi, G., Price, A. N., Hutter, J., Gaspar, A. S., Rutherford, M. A., & Hajnal, J. V. (2018). Inner-volume echo volumar imaging (IVEVI) for robust fetal brain imaging. *Magnetic Resonance in Medicine, 80*(1), 279–285. https://doi.org/10.1002/mrm.26998.

Ouyang, A., Jeon, T., Sunkin, S. M., Pletikos, M., Sedmak, G., Sestan, N., Lein, E. S., & Huang, H. (2015). Spatial mapping of structural and connectional imaging data for the developing human brain with diffusion tensor imaging. *Methods, 0*, 27–37. https://doi.org/10.1016/j.ymeth.2014.10.025.

Ouyang, M., Dubois, J., Yu, Q., Mukherjee, P., & Huang, H. (2018). Delineation of early brain development from fetuses to infants with diffusion MRI and beyond. *NeuroImage*. https://doi.org/10.1016/j.neuroimage.2018.04.017.

Payette, K., de Dumast, P., Kebiri, H., Ezhov, I., Paetzold, J. C., Shit, S., Iqbal, A., Khan, R., Kottke, R., Grehten, P., Ji, H., Lanczi, L., Nagy, M., Beresova, M., Nguyen, T. D., Natalucci, G., Karayannis, T., Menze, B., Cuadra, M. B., & Jakab, A. (2021). An automatic multi-tissue human fetal brain segmentation benchmark using the Fetal Tissue Annotation Dataset. *ArXiv:2010.15526 [Cs, Eess]*. http://arxiv.org/abs/2010.15526.

Qiu, A., Mori, S., & Miller, M. I. (2015). Diffusion tensor imaging for understanding brain development in early life. *Annual Review of Psychology, 66*(1), 853–876. https://doi.org/10.1146/annurev-psych-010814-015340.

Rajchl, M., Lee, M. C. H., Oktay, O., Kamnitsas, K., Passerat-Palmbach, J., Bai, W., Damodaram, M., Rutherford, M. A., Hajnal, J. V., Kainz, B., & Rueckert, D. (2016). DeepCut: object segmentation from bounding box annotations using convolutional neural networks. *ArXiv:1605.07866 [Cs]*. http://arxiv.org/abs/1605.07866.

Robinson, A. J., & Ederies, M. A. (2018). Fetal neuroimaging: An update on technical advances and clinical findings. *Pediatric Radiology, 48*(4), 471–485. https://doi.org/10.1007/s00247-017-3965-z.

Robinson, E. C., Garcia, K., Glasser, M. F., Chen, Z., Coalson, T. S., Makropoulos, A., Bozek, J., Wright, R., Schuh, A., Webster, M., Hutter, J., Price, A., Cordero Grande, L., Hughes, E., Tusor, N., Bayly, P. V., Van Essen, D. C., Smith, S. M., Edwards, A. D., … Rueckert, D. (2018). Multimodal surface matching with higher-order smoothness constraints. *NeuroImage, 167*, 453–465. https://doi.org/10.1016/j.neuroimage.2017.10.037.

Robinson, E. C., Jbabdi, S., Glasser, M. F., Andersson, J., Burgess, G. C., Harms, M. P., Smith, S. M., Van Essen, D. C., & Jenkinson, M. (2014). MSM: A new flexible framework for Multimodal Surface Matching. *NeuroImage, 100*, 414–426. https://doi.org/10.1016/j.neuroimage.2014.05.069.

Salehi, S. S. M., Erdogmus, D., & Gholipour, A. (2017). Auto-context Convolutional Neural Network (Auto-Net) for Brain Extraction in Magnetic Resonance Imaging. I*EEE Transactions on Medical Imaging, PP*(99), 1–1. https://doi.org/10.1109/TMI.2017.2721362.

Salehi, S. S. M., Hashemi, S. R., Velasco-Annis, C., Ouaalam, A., Estroff, J. A., Erdogmus, D., Warfield, S. K., & Gholipour, A. (2018). Real-time automatic fetal brain extraction in fetal MRI by deep learning. *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, 720–724. https://doi.org/10.1109/ISBI.2018.8363675.

Scheinost, D., Onofrey, J. A., Kwon, S. H., Cross, S. N., Sze, G., Ment, L. R., & Papademetris, X. (2018). A fetal fMRI specific motion correction algorithm using 2nd order edge features. *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, 1288–1292. https://doi.org/10.1109/ISBI.2018.8363807.

Schöpf, V., Kasprian, G., Brugger, P. C., & Prayer, D. (2012). Watching the fetal brain at 'rest.'. *International Journal of Developmental Neuroscience, 30*(1), 11–17. https://doi.org/10.1016/j.ijdevneu.2011.10.006.

Schuler, A.-L., Bartha-Doering, L., Jakab, A., Schwartz, E., Seidl, R., Kienast, P., Lackner, S., Langs, G., Prayer, D., & Kasprian, G. (2018). Tracing the structural origins of atypical language representation: Consequences of prenatal mirror-imaged brain asymmetries in a dizygotic twin couple. *Brain Structure and Function, 223*(8), 3757–3767. https://doi.org/10.1007/s00429-018-1717-y.

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*(3), 2255–2265. https://doi.org/10.1016/j.neuroimage.2011.09.062.

Serag, A., Macnaught, G., Denison, F. C., Reynolds, R. M., Semple, S. I., & Boardman, J. P. (2017). Histograms of oriented 3D gradients for fully automated fetal brain localization and robust motion correction in 3 T magnetic resonance images. *BioMed Research International, 2017.* https://doi.org/10.1155/2017/3956363.

Seshamani, S., Blazejewska, A. I., Gatenby, C., Mckown, S., Caucutt, J., Dighe, M., & Studholme, C. (2015). Comparing consistency of R2* and T2*-weighted BOLD analysis of resting state fetal fMRI. *Medical Imaging 2015: Biomedical Applications in Molecular, Structural, and Functional Imaging, 9417*, 94170N. https://doi.org/10.1117/12.2082236.

Seshamani, S., Cheng, X., Fogtmann, M., Thomason, M. E., & Studholme, C. (2014). A Method for handling intensity inhomogenieties in fMRI sequences of moving anatomy of the early developing brain. *Medical Image Analysis, 18*(2), 285–300. https://doi.org/10.1016/j.media.2013.10.011.

Shattuck, D. W., & Leahy, R. M. (2002). BrainSuite: An automated cortical surface identification tool. *Medical Image Analysis, 14.*

Shi, F., Yap, P.-T., Fan, Y., Gilmore, J. H., Lin, W., & Shen, D. (2010). Construction of multi-region-multi-reference atlases for neonatal brain MRI segmentation. *NeuroImage, 51*(2), 684–693. https://doi.org/10.1016/j.neuroimage.2010.02.025.

Song, L., Mishra, V., Ouyang, M., Peng, Q., Slinger, M., Liu, S., & Huang, H. (2017). Human fetal brain connectome: structural network development from middle fetal stage to birth. *Frontiers in Neuroscience, 11.* https://doi.org/10.3389/fnins.2017.00561.

Studholme, C. (2011). Mapping fetal brain development in utero using magnetic resonance imaging: the big bang of brain mapping. *Annual Review of Biomedical Engineering, 13*(1), 345–368. https://doi.org/10.1146/annurev-bioeng-071910-124654.

Studholme, C. (2015). Mapping the developing human brain in utero using quantitative MR imaging techniques. *Seminars in Perinatology, 39*(2), 105–112. https://doi.org/10.1053/j.semperi.2015.01.003.

Taha, A. A., & Hanbury, A. (2015). Metrics for evaluating 3D medical image segmentation: Analysis, selection, and tool. *BMC Medical Imaging, 15.* https://doi.org/10.1186/s12880-015-0068-x.

Takahashi, E., Folkerth, R. D., Galaburda, A. M., & Grant, P. E. (2012). Emerging cerebral connectivity in the human fetal brain: An MR tractography study. *Cerebral Cortex (New York, N.Y.: 1991), 22*(2), 455–464. https://doi.org/10.1093/cercor/bhr126.

Thomason, M. E. (2018). Structured spontaneity: building circuits in the human prenatal brain. *Trends in Neurosciences, 41*(1), 1–3. https://doi.org/10.1016/j.tins.2017.11.004.

Thomason, M. E., Brown, J. A., Dassanayake, M. T., Shastri, R., Marusak, H. A., Hernandez-Andrade, E., Yeo, L., Mody, S., Berman, S., Hassan, S. S., & Romero, R. (2014). Intrinsic functional brain architecture derived from graph theoretical analysis in the human fetus. *PLoS One, 9*(5), e94423. https://doi.org/10.1371/journal.pone.0094423.

Thomason, M. E., Dassanayake, M. T., Shen, S., Katkuri, Y., Alexis, M., Anderson, A. L., Yeo, L., Mody, S., Hernandez-Andrade, E., Hassan, S. S., Studholme, C., Jeong, J.-W., & Romero, R. (2013). Cross-hemispheric functional connectivity in the human fetal brain. *Science Translational Medicine, 5*(173). https://doi.org/10.1126/scitranslmed.3004978.

Thomason, M. E., Grove, L. E., Lozon, T. A., Vila, A. M., Ye, Y., Nye, M. J., Manning, J. H., Pappas, A., Hernandez-Andrade, E., Yeo, L., Mody, S., Berman, S., Hassan, S. S., & Romero, R. (2015). Age-related increases in long-range connectivity in fetal functional neural connectivity networks in utero. *Developmental Cognitive Neuroscience, 11*, 96–104. https://doi.org/10.1016/j.dcn.2014.09.001.

Thomason, M. E., Scheinost, D., Manning, J. H., Grove, L. E., Hect, J., Marshall, N., Hernandez-Andrade, E., Berman, S., Pappas, A., Yeo, L., Hassan, S. S., Constable, R. T., Ment, L. R., & Romero, R. (2017). Weak functional connectivity in the human fetal brain prior to preterm birth. *Scientific Reports, 7*, 39286. https://doi.org/10.1038/srep39286.

Tommasi, T., Patricia, N., Caputo, B., & Tuytelaars, T. (2015). A deeper look at dataset bias. *ArXiv:1505.01257 [Cs].* http://arxiv.org/abs/1505.01257.

Tourbier, S., Velasco-Annis, C., Taimouri, V., Hagmann, P., Meuli, R., Warfield, S. K., Bach Cuadra, M., & Gholipour, A. (2017). Automated template-based brain localization and extraction for fetal brain MRI reconstruction. *NeuroImage, 155*, 460–472. https://doi.org/10.1016/j.neuroimage.2017.04.004.

van den Heuvel, M. I., & Thomason, M. E. (2016). Functional connectivity of the human brain in utero. *Trends in Cognitive Sciences, 20*(12), 931–939. https://doi.org/10.1016/j.tics.2016.10.001.

van den Heuvel, M. I., Turk, E., Manning, J. H., Hect, J., Hernandez-Andrade, E., Hassan, S. S., Romero, R., van den Heuvel, M. P., & Thomason, M. E. (2018). Hubs in the human fetal brain network. *Developmental Cognitive Neuroscience, 30*, 108–115. https://doi.org/10.1016/j.dcn.2018.02.001.

Vasung, L., Abaci Turk, E., Ferradal, S. L., Sutin, J., Stout, J. N., Ahtam, B., Lin, P.-Y., & Grant, P. E. (2018). Exploring early human brain development with structural and physiological neuroimaging. *NeuroImage.* https://doi.org/10.1016/j.neuroimage.2018.07.041.

Verghese, A., Shah, N. H., & Harrington, R. A. (2018). What this computer needs is a physician: humanism and artificial intelligence. *JAMA, 319*(1), 19–20. https://doi.org/10.1001/jama.2017.19198.

What is a Container? | App Containerization | Docker. (n.d.). Retrieved March 19, 2021, from https://www.docker.com/resources/what-container.

Wright, R., Kyriakopoulou, V., Ledig, C., Rutherford, M. A., Hajnal, J. V., Rueckert, D., & Aljabar, P. (2014). Automatic quantification of normal cortical folding patterns from fetal brain MRI. *NeuroImage, 91*, 21–32. https://doi.org/10.1016/j.neuroimage.2014.01.034.

Zech, J. R., Badgeley, M. A., Liu, M., Costa, A. B., Titano, J. J., & Oermann, E. K. (2018). Confounding variables can degrade generalization performance of radiological deep learning models. *ArXiv: 1807.00431 [Cs, Stat].* http://arxiv.org/abs/1807.00431.

Zeiler, M. D., & Fergus, R. (2014). Visualizing and understanding convolutional networks. *Computer Vision – ECCV, 2014*, 818–833. https://doi.org/10.1007/978-3-319-10590-1_53.

Springer