Robust Regression of Brain Maturation from 3D Fetal Neurosonography Using CRNs

Ana I.L. Namburete(✉), Weidi Xie, and J. Alison Noble

Department of Engineering Science, Institute of Biomedical Engineering, University of Oxford, Oxford, UK
ana.namburete@eng.ox.ac.uk

Abstract. We propose a fully three-dimensional Convolutional Regression Network (CRN) for the task of predicting fetal brain maturation from 3D ultrasound (US) data. Anatomical development is modelled as the sonographic patterns visible in the brain at a given gestational age, which are aggregated by the model into a single value: the brain maturation (BM) score. These patterns are learned from 589 3D fetal volumes, and the model is applied to 3D US images of 146 fetal subjects acquired at multiple, ethnically diverse sites, spanning an age range of 18 to 36 gestational weeks. Achieving a mean error of 7.7 days between ground-truth and estimated maturational scores, our method outperforms the current state-of-art for automated BM estimation from 3D US images.

1 Introduction

Estimation of fetal growth and developmental progression is paramount in obstetric care. The fetal brain undergoes a predictable sequence of structural changes across gestation: from a smooth surface, to progressively bearing more folds [1]. This process follows such a precise schedule, that any delays are indicative of impaired brain maturation. Thus, the presence of a cerebral abnormality may cause the level of brain maturation (BM) to differ from the chronological gestational age (GA). This work present a tool to automatically estimate BM from 3D ultrasound (US) images of the fetal brain from as early as 18 weeks.

Routinely used clinical methods for assessment of brain maturation are largely qualitative [2] or based on the size of a single brain structure [3]. In these examinations, obstetricians exploit the changes in texture and the emergence and evolution of structures at a given gestational timepoint to inform on BM [4]. They have to mentally fuse information from different brain structures, each of which follows a non-linear developmental trajectory, to then determine a maturational score [4]. The goal of this work is to capitalise on all available brain biomarkers, both spatially and temporally, in 3D US images to estimate BM.

Automated models have been successful in exploiting neurodevelopmental biomakers to predict age and maturation from brain images of neonates [5,6] and fetuses [7]. In [7], random regression forests (RRFs) were used to predict GA and BM from fetal ultrasound images and clinical biometric data. That approach demonstrated the ability of RRFs to map sonographic patterns visible

M.J. Cardoso et al. (Eds.): FIFI/OMIA 2017, LNCS 10554, pp. 73–80, 2017.
DOI: 10.1007/978-3-319-67561-9_8

in standard clinical US data to gestational age within ±11 days[1]. By design, RRFs are non-linear predictors that disregard any image regions which are not encapsulated within the set of hand-crafted features. Deep convolutional neural networks (DCNNs), on the other hand, have demonstrated that by not imposing priors on the data (i.e. no feature hypothesis), they are able to automatically identify relevant features for the prediction task. During training, the cost function is optimised to generate a model of high-level abstractions described by a multi-layered graph which encapsulates both linear and non-linear encodings of the data. This property makes them well-suited to making predictions about an organ as complex as the developing brain, with data as challenging as US.

Regression tasks are performed by convolutional regression networks (CRNs), and success has been achieved in estimating biological age from medical image data [8]. In this work, we follow this novel direction to apply CRNs to estimate BM from complex 3D neurosonographic data. However, while most regression models are designed to minimize the difference between ground-truth and predicted values by optimising a least-squares function, this tends to be sensitive to outliers. Within the context of ultrasound-based estimations, outliers are typically represented by images with strong acoustic shadows, partial anatomical occlusions, or variations in intensity patterns (due to acquisition protocol). They may also be represented by developmentally advanced or delayed fetal subjects, or the rare cases of fetuses affected by cerebral malformations. The L_2-norm is unlikely to perform well in such cases. In this work, we also explore an objective function that reduces emphasis on such outliers.

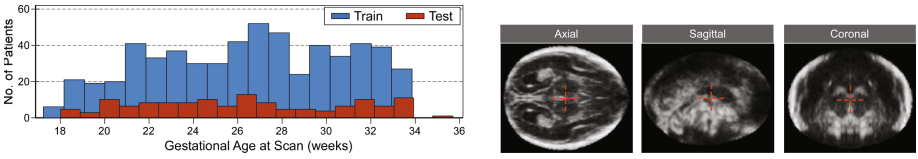
2 Methods

In the fetal period, the bright echoes visible in US images change, marking the emergence of cerebral structures [1]. This work explores two different CRN architectures for estimating BM from sonographic image patterns. The objective is to capture the brain features informative of structural brain changes and, potentially, the process of cortical folding. This section summarises the data processing pipeline and describes the architecture design.

2.1 Image Pre-processing

Fetal neurosonography is challenged by the fact that scanning through bone attenuates the ultrasound energy, and the concave shape of the skull surface refracts the signal. As the skull calcifies over the course of gestation, the US signal is increasingly affected, thereby lowering the image contrast and the visibility of anatomical boundaries. These interactions result in only the cerebral hemisphere farthest from the US probe producing an image with discernible structures. To circumvent this, a 'complete' representation of the brain was generated in our dataset by mirroring the visible hemisphere across the midsagittal plane.

[1] This result refers to the RRF model which exclusively used brain features, and did not incorporate information about fetal size (i.e. head circumference).



(a) (b)

Fig. 1. (a) Gestational age distribution of the training ($N = 609$) and validation ($N = 120$) fetal datasets. (b) Example of preprocessed US brain image, following skull masking, alignment, and mirroring of visible hemisphere.

To exclude prior knowledge about the expected size of the brain, all ultrasound images were spatially normalized and aligned to a pre-defined coordinate frame prior to processing. Specifically, the skull was manually segmented, and the brain region was cropped and rigidly aligned to a predefined coordinate frame (Fig. 1b). To reduce the effect of intensity variations, the images were then individually normalised to have zero mean and unary standard deviation.

2.2 CRN Architectures

CRNs are sequential models consisting of layers which perform operations on the input data. The input to the network is an image $\mathbf{I} : \Omega \rightarrow \mathbb{R}^3$, and the output is a real-valued scalar $y \in \mathbb{R}$ corresponding to BM score. Given a training set $\{(\mathbf{I}_i, y_i)\}_{i=1}^N$ of N data samples, the goal of training a CRN is to learn a function $\varphi(\cdot;\theta)$ and its corresponding parameters θ to map an image to a prediction of BM: $\hat{y} = \varphi(\mathbf{I};\theta)$. The function, $\varphi(\cdot;\theta)$ is minimized by a pre-defined loss function $\mathcal{L}(\mathbf{y}, \hat{\mathbf{y}} \mid \theta)$.

The input image is mapped to the output by a series of convolutional blocks (ConvBlocks). In each ConvBlock, feature extraction is performed by convolutional layers which convolve the images with a pre-defined number of filters (s) of kernel size k. In the first layer, we use a convolutional layer with filter of size $k \times k \times s$. All convolutional layers use kernels of fixed size ($k = 7$), with a sliding step of size $\delta_c = 1$. In order to improve feature generalizability, each convolution is followed by batch normalization. The batch-normalized filter outputs are then processed by a rectified linear unit (ReLU) as the non-linear activation function. Weight regularization of 10^{-4} is also used in each convolution to reduce the generalization error. In order to reduce computational burden, the image resolution is decreased by a Max-Pooling operation with a stride of $\delta_p = 2$. The last is the only layer of the network whose activation is not a ReLU; instead, it uses a linear activation to regress to GA as a proxy for BM.

We test two networks, each consisting of four pooling layers. The architecture of network CRN-2D consists of a ConvBlockA (Fig. 2b) before each pooling layer, for a total of eight two-dimensional (2D) convolutional layers between the input and output layers (Fig. 2a). Due to memory limitations associated with performing 3D convolutions, network CRN-3D was designed as a shallower network. It

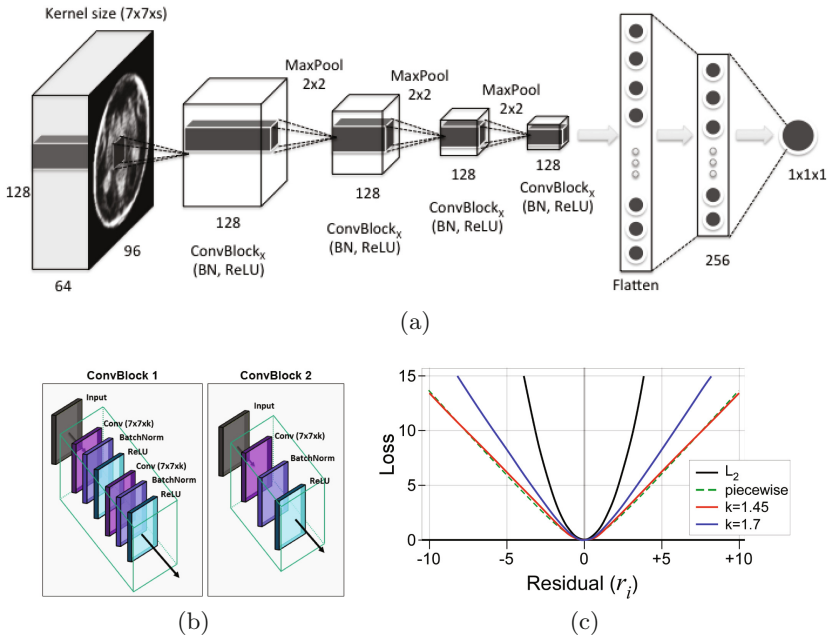


Fig. 2. (a) Basic architecture for the CRN-2D network. (b) Schematic of the components of ConvBlocks A and B. Each convolution (Conv) is followed by batch normalization (BN) and ReLU activation. (c) L_2 increases rapidly with increasing residual magnitude. We use a smooth approximation (red) of the piecewise (dashed green) Huber function. Higher values of k bias the curve towards L_2.

comprises of a ConvBlockB preceding each pooling operation, amounting to a total of four three-dimensional (3D) convolutions.

2.3 Loss Functions for Robust Maturation Regression

During the training process, the output of the CRN ($\hat{\mathbf{y}}$) is compared with the ground-truth labels $\mathbf{y}$ through a loss function $\mathcal{L}$, and the error is back-propagated to update the filter weights of all the layers. This process is repeated until convergence is reached. The goal of the loss function is to minimize the difference (or residuals, $\mathbf{r}$) between the known ($\mathbf{y}$) and estimated ($\hat{\mathbf{y}}$) values. Thus, the loss function ultimately dictates the speed of convergence, and the quality of the trained parameters.

In the back-propagation step, the magnitude of the gradient is proportional to the residual. As a result, the estimations that are close to the ground-truth (inliers) have a lower influence on the updated network parameters (θ), and the outliers yielding higher residuals have a greater influence and may thus bias the model to adapt itself to such examples. There are several options for objective functions. To achieve a robust BM estimator, we explored the effect of

different loss functions on the CRN models. Namely, the familiar least-squares estimator ($\mathcal{L}_2$),

$$\mathcal{L}_2(\mathbf{r}) = \frac{1}{n}\sum_{i=1}^{n} r_i^2 \tag{1}$$

and the Huber estimator ($\mathcal{L}_{\text{hub}}$) which is less vulnerable to outliers [9],

$$\mathcal{L}_{\text{hub}}(\mathbf{r}) = \frac{1}{n}\sum_{i=1}^{n} k^2(\sqrt{1+(r_i/k)^2}-1). \tag{2}$$

where $r_i = \hat{y}_i - y_i$, and k is the tuning constant for the Huber estimator. As shown in Fig. 2c, both the L_2 and Huber functions increase without bound as the residual r_i departs from zero. The Huber estimator approximates the L_2-norm for small residuals, but the key difference is observed at high residuals when it approximates a linear model. While the L_2-norm assigns equal weight to all residuals, the weights of the Huber estimator decline when $|r_i| > k$, thus reducing the emphasis on outliers during training. Furthermore, a small value of k increases the resistance to outliers, at the cost of performance when errors follow a normal distribution. For our experiments, we set $k = 1.45$ for CRN-2D, and $k = 1.7$ for CRN-3D.

3 Experiments

In our experiments, we explore two network architectures for the task of predicting brain maturation from 3D ultrasound images. We explore the effect of either treating the 3D input image ($\mathbf{I} \in \mathbb{R}^{n_x \times n_y \times n_z}$) as a multi-channel image, or preserving its 3D nature during feature extraction. Specifically, Fig. 2a shows that in network CRN-2D, the depth of the feature extraction kernels in the first layer extends to encompass the third dimension of the image, such that the kernel dimensions are $k \times k \times n_z$. In network CRN-3D, the CRN performs 3D convolutions throughout the image space, and the kernel dimensions are $k \times k \times k$, where $k \ll n_z$. Furthermore, we investigate the effect of the choice of loss function on output predictions. In particular, we compare the L_2-norm and Huber loss functions.

Dataset: The dataset of volumetric US images used in this work comprised of the same 447 volumes ($247 \times 190 \times 179$ voxels at a resolution of $0.6 \times 0.6 \times 0.6\,\text{mm}^3$) used in [7], the results of which constitute the baseline for automated image-based BM estimation. The sonographic images of the fetal head were obtained from the INTERGROWTH-21$^{\text{st}}$ study database [10], which were collected using a Philips HD9 curvilinear probe at a 2–5 MHz wave frequency. Images were selected from fetuses with known gestational age ranging from 18 to 36 gestational weeks, spanning the second and third trimesters of pregnancy: an active period of spatiotemporal changes visible on the cortical surface. 'True age' was defined by the last menstrual period (LMP) and confirmed by first-trimester (≤ 14 weeks) crown-rump length measurement agreeing within 7 days.

Experimental setup: The performance of each of the five proposed models was evaluated in five-fold cross-validation rounds. In each round, the volumes of 121 subjects were reserved for testing, while the remaining 489 images were used in training. As shown in Fig. 1a, the age distribution in each dataset was uniform across the gestational age range. To reduce the number of parameters optimized by the CRNs, the input volumes were resized to $128 \times 96 \times 64$ or $96 \times 64 \times 48$ voxels for networks A and B, respectively. Each CRN was re-trained on a dataset of 584 volumes, and tested on an independent set of 146 US volumes from subjects whose scans were collected as part of a multi-site, ethnically diverse study.
Implementation details: The proposed framework was implemented using the Keras framework with a Tensorflow backend, using a parallel computing architecture (CUDA, NVIDIA Corp). All testing was performed on an Intel Xeon E5-2630 CPU (2.4 GHz, 16 cores) with a NVIDIA Titan X GPU. Optimization was achieved using the RMSprop algorithm over a maximum of 100 epochs to ensure convergence [11]. The initial learning rate was set to 10^{-2}, which was decreased by a factor of ten every 20 epochs. Due to memory limitations, the CRN-2D and CRN-3D networks were trained with batches of size 32 and 10, respectively, in each epoch.

4 Results and Discussion

We validated and tested five different models for automated BM prediction from US images for the fetal brain. The results for BM prediction are summarised in Table 1, where the first row shows the result of using the RRF baseline model. The RRF model yielded an accuracy of 10.65 ± 12.65 days on the test set, which was outperformed by all CRN networks with the exception of CRN-2D trained with L_2 loss. While we observed that both 2D and 3D CRN models were able to automatically generate filters that extract and characterise patterns of brain development, notable performance gains were achieved with the 3D CRNs (< 8 days). This may be attributed to the fact that the 3D models preserve the appearance statistics and spatial distribution of structures in the dimension that is 'compressed' by the 2D models during feature extraction.

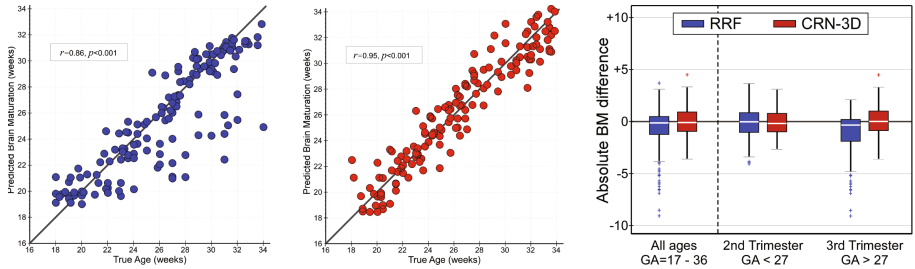
Comparing the performance of the loss function, we observed faster converge with Huber loss (i.e. 40 epochs versus 45 epochs with L_2 loss). In general, the models trained with Huber loss performed better and with reduced variance. However, the best-performing model was CRN-3D trained with L_2 loss. This may be due to the fact that our data contained few samples of subjects with brain abnormalities (i.e. outliers) and so the L_2-norm was able to model these data. Future work may explore whether the same is true when fetuses with cerebral abnormalities are included in the dataset. Figure 3 shows that when applied to the test dataset, the CRN-3D model had lower variance and fewer outliers than the RRF predictions.

It is important to note, however, that the BM predicted in this study is representative of the average level of structural development and brain appearance in fetal subjects of the same GA. Thus, it is expected that the biological variation

Table 1. Errors in predicting BM as a regression target using networks CRN-2D and CRN-3D. Mean absolute differences (± standard deviation) between ground-truth GA and predicted BM score. Results shown for the five-fold cross-validation sets (each $N = 121$), and for the independent test set ($N = 120$). CRN-3D trained with an L_2 loss function (in bold) outperformed the others.

Network	Loss function	Training time (s/epoch)	Validation error (days)	Test error (days)	No. params
RRF (baseline)	L_2-norm	60 s/tree	11.62 ± 10.43	10.65 ± 12.65	0.36 M
CRN-2D	L_2-norm	13	12.33 ± 10.09	11.43 ± 9.35	7.60 M
	Huber		10.32 ± 8.48	10.15 ± 9.88	
CRN-3D	L_2-norm	105	$\mathbf{6.93 \pm 5.46}$	$\mathbf{7.72 \pm 6.01}$	6.13 M
	Huber		7.61 ± 6.15	7.81 ± 7.01	

within a given GA is captured within our dataset. GA was used as a proxy for the ground truth BM score in the knowledge that the INTERGROWTH-21st dataset comprises of optimally healthy fetuses, with confirmed absence of neurocognitive delays on post-birth follow-up. Furthermore, our 'true age' values were annotated with a confidence level of ±7 days, so our potential accuracy is limited by this value. Therefore, our result of 7.72±6.01 days is considered a successful prediction of brain maturation. Thus, our algorithm trained on healthy fetuses, is designed to identify fetuses at risk of maturational delays.



(a) RRF (b) CRN-3D (c) $|y - \hat{y}|$

Fig. 3. Chronological gestational age plotted against BM of the (a) RRF model and (b) the best-performing CRN-3D applied to the test dataset. (c) BM estimation results for the second and third trimesters of pregnancy, comparing the best performing CRN (red) and to the baseline RRF (blue) model. (Color figure online)

5 Conclusion

This paper presents and validates an automated method to predict brain maturation from 3D US scans of the fetal brain. Two proposed models were applied to an ethnically diverse fetal cohort, ranging from 18 to 34 weeks' gestation. We

have demonstrated that without specifying a feature hypothesis or providing any information about fetal size, the CRN model is capable of identifying and characterizing neurodevelopment both spatially and temporally. The proposed CRN model is generalizable to different fetal cohorts, and capable of accurately estimating BM even in subjects where the state-of-art method fails.

References

1. Toi, A., Lister, W.S., Fong, K.W.: How early are fetal cerebral sulci visible at prenatal ultrasound and what is the normal pattern of early fetal sulcal development? Ultrasound Obstet. Gynecol. **24**(7), 706–715 (2004)
2. Monteagudo, A., Timor-Tritsch, I.E.: Normal sonographic development of the central nervous system from the second trimester onwards using 2D, 3D and transvaginal sonography. Prenat. Diagn. **29**(4), 326–339 (2009)
3. Vinkesteijn, A., Mulder, P., Wladimiroff, J.: Fetal transverse cerebellar diameter measurements in normal and reduced fetal growth. Ultrasound Obstet. Gynecol. **15**(1), 47–51 (2000)
4. Pistorius, L.R., Stoutenbeek, P., Groenendaal, F., de Vries, L., Manten, G., Mulder, E., Visser, G.: Grade and symmetry of normal fetal cortical development: a longitudinal two- and three-dimensional ultrasound study. Ultrasound Obstet. Gynecol. **36**(6), 700–708 (2010)
5. Franke, K., Luders, E., May, A., Wilke, M., Gaser, C.: Brain maturation: predicting individual BrainAGE in children and adolescents using structural MRI. NeuroImage **63**(3), 1305–1312 (2012)
6. Toews, M., Wells, W.M., Zöllei, L.: A feature-based developmental model of the infant brain in structural MRI. In: Ayache, N., Delingette, H., Golland, P., Mori, K. (eds.) MICCAI 2012. LNCS, vol. 7511, pp. 204–211. Springer, Heidelberg (2012). doi:10.1007/978-3-642-33418-4_26
7. Namburete, A.I.L., Stebbing, R.V., Kemp, B., Yaqub, M., Papageorghiou, A.T., Alison Noble, J.: Learning-based prediction of gestational age from ultrasound images of the fetal brain. Med. Image Anal. **21**(1), 72–86 (2015)
8. Štern, D., Payer, C., Lepetit, V., Urschler, M.: Automated age estimation from hand MRI volumes using deep learning. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9901, pp. 194–202. Springer, Cham (2016). doi:10.1007/978-3-319-46723-8_23
9. Huber, P.J.: Robust estimation of a location parameter. Ann. Math. Stat. **35**(1), 73–101 (1964)
10. Papageorghiou, A.T., Ohuma, E.O., Altman, D.G., Todros, T., Cheikh Ismail, L., Lambert, A., Jaffer, Y.A., Bertino, E., Gravett, M.G., Purwar, M., Noble, J.A., Pang, R., Victora, C.G., Barros, F.C., Carvalho, M., Salomon, L.J., Bhutta, Z.A., Kennedy, S.H., Villar, J.: International fetal and newborn growth consortium for the 21st century (INTERGROWTH-21st): international standards for fetal growth based on serial ultrasound measurements: the Fetal growth longitudinal study of the INTERGROWTH-21st project. Lancet **384**(9946), 869–79 (2014)
11. Tieleman, T., Hinton, G.: Lecture 6.5-RMSprop: divide the gradient by a running average of its recent magnitude. COURSERA Neural Networks Mach. Learn. **4**, 26–31 (2012)