

0839

A Deep Learning Approach for Image Quality Assessment of Fetal Brain MRI

Sayeri Lala¹, Nalini Singh^{2,3}, Borjan Gagoski^{4,5}, Esra Turk⁴, P. Ellen Grant^{4,5}, Polina Golland^{1,3}, and Elfar Adalsteinsson^{1,2,6}

¹Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, Cambridge, MA, United States, ²Harvard-MIT Health Sciences and Technology, Massachusetts Institute of Technology, Cambridge, MA, United States, ³Computer Science and Artificial Intelligence Laboratory (CSAIL), Massachusetts Institute of Technology, Cambridge, MA, United States, ⁴Fetal-Neonatal Neuroimaging and Developmental Science Center, Boston Children's Hospital, Boston, MA, United States, ⁵Harvard Medical School, Boston, MA, United States, ⁶Institute for Medical Engineering and Science, Massachusetts Institute of Technology, Cambridge, MA, United States

Synopsis

Fetal MRI plays a critical role in diagnosing brain abnormalities but routinely suffers from artifacts resulting in nondiagnostic images. We aim to automatically identify nondiagnostic images during acquisition so they can be immediately flagged for reacquisition. As a first step, we trained a neural network to classify T2-weighted single-shot fast spin-echo (HASTE) images as diagnostic or nondiagnostic. With novel data, the average Area Under Receiver Operator Characteristic Curve was 0.84 ($\sigma = 0.04$). The neural network learned relevant criteria, identifying high contrast boundaries between areas like cerebral spinal fluid and cortical plate as relevant to determining image quality.

Problem and Proposed Approach

Because of unpredictable and severe fetal motion, 2D single-shot T2-weighted (SST2W) imaging is the dominant acquisition method to mitigate motion artifacts while achieving diagnostic image contrast¹. However, fetal motion often occurs during the k-space readouts, resulting in severe artifacts and consequently nondiagnostic images (Figure 1). Motion from slice to slice also causes artifacts in the form of incomplete volume coverage or incomplete signal recovery due to spin history effects. Current fetal MR protocols attempt to address this by repeating each orthogonal stack of 2D slices multiple times, but with no immediate guarantee that the slices are of diagnostic quality.

To address these problems, we aim to train an image quality classifier that automatically identifies nondiagnostic images. This tool can then be used during the scan to flag nondiagnostic images for reacquisition in order to deliver a diagnostic quality stack of 2D slices in a single SST2W scan.

Methods

Recent literature^{2,3} demonstrated the potential of neural networks to perform image quality assessment on MR images. Encouraged by the benefits of transfer learning in medical imaging^{3,4}, we fine-tuned a 50-layer Imagenet pretrained Residual Network on a fetal MR image quality dataset. Inspired by techniques improving classification with object segmentation⁵, we experimented with training and evaluating on two versions of the dataset: 1) the original images containing the fetal brain, other fetal organs, and maternal organs (Figure 3); and 2) brain-masked images where only the fetal brain was present in the image (Figure 5). Integrated Gradients⁶ was applied to representative results from the classifier to understand which image features contributed to the classifier's decision.

Dataset preparation

A total of 1874 images were obtained from ten previously acquired clinical scans of mothers with singleton pregnancies, ranging in gestational age between 19-37 weeks. Scans were conducted at Boston Children's Hospital (BCH) with Institutional Review Board approval. Scans were acquired using the SST2W (HASTE) sequence with TE/TR = 100 ms/1.4 s, FOV 36cm, slice thickness 3 mm, and voxel size of 1.4x1.4x3 mm³.

Under the guidance of radiologists at BCH, a research assistant identified criteria distinguishing nondiagnostic from diagnostic images (Figure 1), and labeled a novel dataset, yielding 142 nondiagnostic images and 1732 diagnostic images.

The brain-masked version of the dataset was produced using a fetal brain segmentation neural network⁷, with additional post processing to improve the segmentations. Each image was standardized prior to neural network training and evaluation.

Evaluation

A five-fold cross validation was performed, where each fold consisted of 8 subjects for training and 2 subjects for validation. Area Under the Receiver Operator Characteristic Curve (AUROC) was used to assess the performance of the neural networks.

Results and Discussion

The neural network trained on the original dataset scored an average AUROC of 0.79 and the neural network trained on the brain-masked dataset scored an average AUROC of 0.84 on the validation set, each with corresponding standard deviations of 0.04 (Figure 2). The results suggest no performance benefit from training on brain-masked images.

Integrated Gradients on representative results from the original dataset trained neural network (Figures 3, 4) demonstrated that the neural network generally learned relevant criteria for image quality assessment, as would a radiologist. The neural network identified sharp high contrast boundaries, like those separating skull and cerebrospinal fluid (CSF), CSF and cortical plate, cortical plate and white matter, as characteristic of diagnostic quality images. Noise in the white matter and blurring of boundaries between cortical plate, CSF, and white matter, seem to characterize nondiagnostic quality images. Non-relevant features outside the fetal brain, like subcutaneous fat, fetal and maternal body parts, affected the classifier's decision.

Integrated Gradients on results from the brain-masked trained neural network demonstrated that the neural network also identified high contrast boundaries as characteristic of diagnostic images (Figure 5).

The classifier took 20 ms to run on each image on a single NVIDIA Tesla K80 GPU with 4 vCPUs and 65 GB of RAM.

Conclusions

Our findings suggest that a deep learning approach could be used for image quality assessment of fetal brain MRI. Future work will involve labeling a larger dataset, experimenting with the number of quality categories, and investigating techniques for helping the classifier focus on the brain region of interest.

After achieving desirable performance, the classifier will be integrated into the SST2W acquisition. This integration may need to be semi-supervised with an MR technician reviewing the images flagged for reacquisition. A semi-supervised pipeline could then be used for online learning to further improve the neural network's performance, with the ultimate goal of having a fully automated classifier.

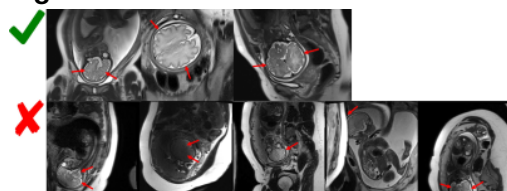
Acknowledgements

Research reported in this abstract was supported by the National Institutes of Health under award numbers R01 EB017337 and U01 HD087211, and the National Institute of Biomedical Imaging and Bioengineering (NIBIB) of the National Institutes of Health under award number 5T32EB1680.

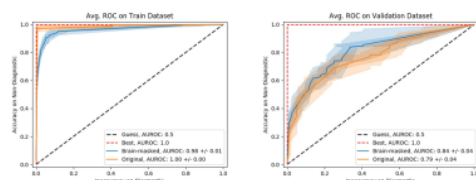
References

1. Gholipour, Ali, et al. "Fetal MRI: A Technical Update with Educational Aspirations." Concepts in Magnetic Resonance. Part A, Bridging Education and Research, vol. 43, no. 6, Nov. 2014, pp. 237–66.
2. Esses, Steven J., et al. "Automated Image Quality Evaluation of T-Weighted Liver MRI Utilizing Deep Learning Architecture." Journal of Magnetic Resonance Imaging: JMIR, vol. 47, no. 3, Mar. 2018, pp. 723–28.
3. Li, Jifan. et al. Automatic Assessment of MR Image Quality with Deep Learning. in ISMRM 0431 (2018).
4. Tajbakhsh, Nima, et al. "Convolutional Neural Networks for Medical Image Analysis: Full Training or Fine Tuning?" IEEE Transactions on Medical Imaging, vol. 35, no. 5, 2016, pp. 1299–312.
5. Girshick, Ross, Jeff Donahue, Trevor Darrell, and Jitendra Malik. 2014. "Rich Feature Hierarchies for Accurate Object Detection and Semantic Segmentation." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 580–87.
6. Sundararajan, Mukund, Ankur Taly, and Qiqi Yan. "Axiomatic attribution for deep networks." arXiv preprint arXiv:1703.01365(2017).
7. Salehi, Seyed Sadegh Mohseni, et al. "Real-Time Automatic Fetal Brain Extraction in Fetal MRI by Deep Learning." 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), 2018, doi:10.1109/isbi.2018.8363675.

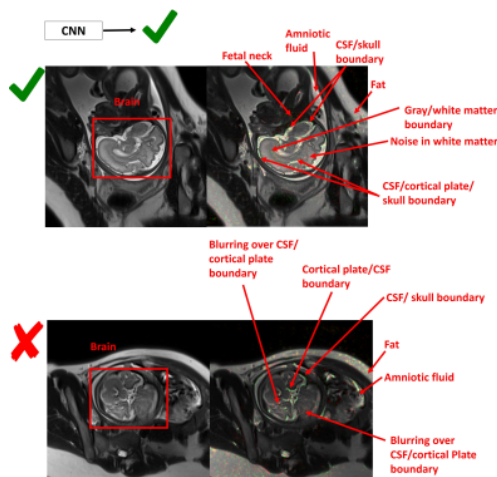
Figures



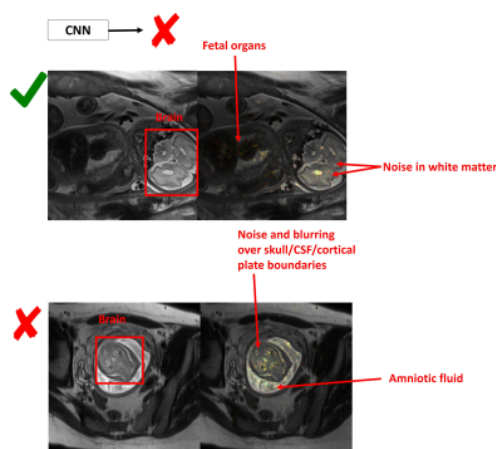
Representative diagnostic (top row) and nondiagnostic (bottom row) SST2W (HASTE) fetal brain images. Arrows point to the brain. Nondiagnostic images are characterized by artifacts obscuring brain structure, including motion (first, second images), low signal-to-noise ratio (third image), aliasing (fourth image), and fetal brain not fully present in the field of view (last image).



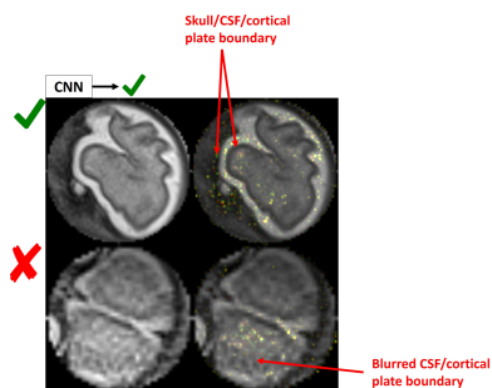
Average Receiver Operator Characteristic (ROC) curves of the neural networks on the training and validation datasets. The results suggest no performance improvement from training on brain-masked images.



Representative results of Integrated Gradients on a true diagnostic image (top row) and a true nondiagnostic image (bottom row), both labeled diagnostic by the original dataset trained classifier. The classifier learned that high contrast boundaries between CSF, skull, cortical plate, and white matter characterize diagnostic quality images, as highlighted by the green pixels. The classifier learned that motion artifacts, like noise in white matter and blurring over high contrast boundaries, characterize nondiagnostic quality images, as highlighted by the red pixels. Features outside the fetal brain, like fat, amniotic fluid, and other parts of the fetus, affected the classifier.



Representative results of Integrated Gradients on a true diagnostic image (top row) and a true nondiagnostic image (bottom row), both labeled nondiagnostic by the original dataset trained classifier. The classifier learned that motion artifacts, like noise in white matter and blurring over boundaries, characterize nondiagnostic quality images, as highlighted by the green pixels. Features outside the fetal brain, like amniotic fluid and other organs of the fetus, affected the classifier.



Representative results of Integrated Gradients on a true diagnostic image (top row) and a true nondiagnostic image (bottom row), both labeled diagnostic by the brain-masked dataset trained classifier. The classifier learned that high contrast boundaries between skull, CSF, and cortical plate characterize diagnostic quality images, as highlighted by the green pixels. The classifier does not seem to have learned that blurred boundaries characterize nondiagnostic quality images. Yellow regions correspond to conflicting feature characterizations, where the classifier deemed the region characteristic of both diagnostic and nondiagnostic quality images.