Deep Learning Automatic Fetal Structures Segmentation in MRI Scans with Few Annotated Datasets

Gal Dudovitch[1,5], Daphna Link-Sourani[2,5], Liat Ben Sira[3,4,5], Elka Miller[3,5], Dafna Ben Bashat[2,4,5], and Leo Joskowicz[1,5](✉)

[1] School of Computer Science and Engineering, The Hebrew University of Jerusalem, Jerusalem, Israel
josko@cs.huji.ac.il
[2] Sagol Brain Institute, Tel Aviv Sourasky Medical Center, Tel Aviv-Yafo, Israel
[3] Division of Pediatric Radiology, Tel Aviv Sourasky Medical Center, Tel Aviv-Yafo, Israel
[4] Sackler Faculty of Medicine and Sagol School of Neuroscience, Tel Aviv University, Tel Aviv-Yafo, Israel
[5] Medical Imaging, Children's Hospital of Eastern Ontario, University of Ottawa, Ottawa, Canada

Abstract. We present a new method for end-to-end automatic volumetric segmentation of fetal structures in MRI scans with deep learning networks trained with very few annotated scans. It consists of three main stages: 1) two-step automatic structure segmentation with custom 3D U-Nets; 2) segmentation error estimation, and; 3) segmentation error correction. The automatic structure segmentation stage first computes a region of interest (ROI) on a downscaled scan and then computes a final segmentation on the cropped ROI. The segmentation error estimation stage uses prediction-time augmentations of the input scan to compute multiple segmentations and estimate the segmentation uncertainty for individual slices and for the entire scan. The segmentation error correction stage then uses these estimations to locate the most error-prone slices and to correct the segmentations in those slices based on validated adjacent slices. Experimental results of our methods on fetal body (63 cases, 9 for training, 55 for testing) and fetal brain MRI scans (35 cases, 6 for training, 29 for testing) yield a mean Dice coefficient of 0.96 for both, and a mean Average Symmetric Surface Distance of 0.74 mm and 0.19 mm, respectively, below the observer delineation variability.

Keywords: Deep learning · Fetal MRI · Segmentation · Uncertainty estimation

1 Introduction

Accurate segmentation of complex structures and pathologies in volumetric images presents a great challenge in medical image processing. Recent deep learning image

Electronic supplementary material The online version of this chapter (https://doi.org/10.1007/978-3-030-59725-2_35) contains supplementary material, which is available to authorized users.

A. L. Martel et al. (Eds.): MICCAI 2020, LNCS 12266, pp. 365–374, 2020.
https://doi.org/10.1007/978-3-030-59725-2_35

classification methods have been shown to be effective for the segmentation of a variety of structures and pathologies in CT and MRI scans [1]. State-of-the-art segmentation methods for medical images are mostly based on Convolutional Neural Networks (CNN) and their variants. These include the 2D U-Net autoencoder convolution/ deconvolution architecture with skip connections [2] and its extensions to 3D [3–5]. These networks have been demonstrated in the segmentation of complex structures, e.g. adult brain and prostate in MRI scans [5, 6] and fetal structures in MRI scans [7]. The NiftyNet platform has recently been developed for deep learning segmentation [8]. Other works rely on a developing fetal brain atlas for segmentation [9]. Its drawback is that it requires a large set of fetal brain scans to create the atlas. Fetit et al. [10] uses an algorithm for the initialization of the fetal brain segmentation, which limits its generality.

The networks performance critically depend on large, high-quality annotated data, which is seldom available, if at all. Research groups must generate their own datasets for each anatomical structure, pathology and scanning protocol. This is an expensive and time-consuming task that requires significant effort and radiological expertise. This has motivated the development of interactive machine learning [11, 12] and deep learning segmentation methods [13, 14], whose aim is to reduce the amount and complexity of the user interactions required for the necessary manual error corrections. While these methods may help to reduce user interactions, they do not yet significantly reduce the user effort and the required number of annotated training datasets.

Three key and closely related issues to the automatic segmentation of structures and their subsequent validation and correction are segmentation variability, robustness, and uncertainty. Segmentation variability and robustness estimation has been researched for deep learning classification [15, 16]. For example, Monte Carlo Dropout is a regularization technique in which random selections of active neurons is used for approximate Bayesian inference. Segmentation uncertainty estimation has been performed with an ensemble of multiple models [15]. However, this method requires a large annotated datasets to train the models. Very recently, segmentation uncertainty estimation methods based on test-time augmentation have been proposed [17, 18]. However, segmentation uncertainty estimates have not been used to prioritize manual segmentation correction and to optimize the selection of scans for manual annotation to increase the model accuracy and robustness with a small training set.

In this paper, we present an end-to-end method for volumetric segmentation of fetal structures in MRI scans with deep learning networks trained with very few annotated scans. The method relies on segmentation error estimation and correction using segmentation uncertainty measures. It increases the segmentation accuracy and robustness and optimizes the total radiologists' annotation time required for creating a dataset with validated annotations, thereby bootstrapping the automatic segmentation task with very few annotated datasets.

2 Method

Our end-to-end method consists of 3 stages: 1) automatic structure segmentation; 2) segmentation error estimation; 3) segmentation error correction. The automatic structure segmentation stage computes first a region of interest (ROI) and then computes the

structure segmentation inside the ROI. The segmentation error estimation stage uses prediction-time augmentations of the input scan to compute multiple segmentations and to estimate segmentation uncertainty and error margins for individual slices and for the entire scan. The identified estimated segmentation errors are then used to prioritize the slices that require inspection and manual correction of the faulty segmentations.

The segmentation error correction stage uses individual slices corrections to automatically correct the segmentations in adjacent slices.

Custom 3D U-Net Architecture. We have developed a custom 3D U-Net architecture based on [3] and [5] for ROI localization and structure segmentation. The U-Net is an encoder/decoder architecture with residual connections whose encoding/decoding pathways classify voxels based on image patches features at different levels of abstraction. Our modifications to the standard 3D U-Net are (Fig. 1):

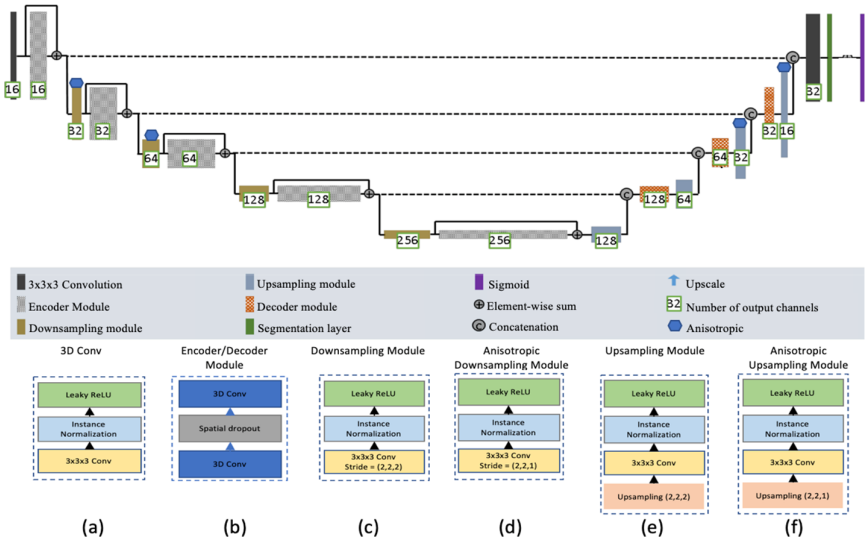

Fig. 1. Top: Architecture of our custom 3D segmentation network based on [3, 5]. The number of output channels of each unit is indicated next to it. Bottom: (a–f) network modules details.

1. Each 3D convolution layer is preceded by a leaky ReLU non-linearity activation followed by an instance normalization layer [19] (Fig. 1a). It replaces standard batch normalization and ensures classification stability for small batch sizes.
2. Each encoder layer is replaced by a residual module [20] with two 3D convolutional layers and a spatial dropout layer between them (Fig. 1b). Spatial dropout layers have been reported to yield superior results in fully convolutional networks [21].
3. Encoder modules are connected by downsampling modules (Fig. 1c) built from $3 \times 3 \times 3$ convolutions with an input stride of 2 to reduce the feature map resolution and to incorporate additional features in the encoding pathway.

4. Up-sample modules in the decoder pathway (Fig. 1e) up-sample the low resolution feature maps with a direct upscale that repeats the feature voxels twice in each spatial dimension, followed by a $3 \times 3 \times 3$ convolution that halves the number of feature maps.
5. The up-sampled features are then recombined by concatenation with the features from the corresponding level of the encoding pathway.
6. Decoder modules (Fig. 1b) recombine the features after concatenation and reduce by half the number of features maps.

A key modification of the network is anisotropic downsampling and upsampling (Fig. 1d, 1f) [22]. Scans are usually anisotropic, e.g., the slices spacing is greater than the slice pixel resolution, resulting in a mismatch between the receptive field and the scan dimensions. Anisotropic sampling enforces this match and thus increases accuracy. The anisotropic downsampling layer performs downsampling on the slices xy plane without downsampling along the z axis. This is implemented by setting the convolution stride to 2 in the xy plane and to 1 along the z axis. The downsampling reduction proceeds anisotopically until the spatial layer dimensions are equal; it then proceeds to the next layers isotropically. The decoder pathway has matching upsampling layers.

Automatic Structure Segmentation. The two-stage segmentation method consists of ROI localization followed by structure segmentation inside the ROI. Both are performed with the custom 3D U-Net described above. Each network is trained as a supervised deep learning model with ground-truth segmentations of target structure.

The ROI localization network inputs a downscaled scan and outputs a coarse segmentation with which the ROI bounding box is computed. The structure segmentation network inputs the full resolution cropped ROI scan and outputs the structure segmentations. The networks are trained with the Dice loss function from [4]. The resulting segmentation is post-processed with Gaussian filter smoothing, connected component analysis, and binarization with a preset intensity threshold.

Spatial and intensity augmentations are used to increase the data size and variety for segmentation network training, for prediction-time augmentation, and for segmentation uncertainty estimation (Fig. 2, left). Intensity augmentations include contrast, blur by Gaussian filtering, addition of Poisson noise and additive and multiplicative Gaussian noise, and coarse dropout. Spatial augmentations include slice-wise affine and elastic deformations with a smoothed displacement field. Prediction-time augmentations yield multiple segmentations that are aligned and combined by averaging or majority voting.

Segmentation Uncertainty and Error Estimation. Neural networks trained with few annotated datasets inevitably produce segmentation errors and perform poorly on out-of-distribution inputs. In these cases, the segmentation errors should be identified and manually corrected by an expert. Currently, the manual segmentation corrections are performed by examining scans and scan slices in sequential order, which is not necessarily optimal. By estimating segmentation uncertainty and detecting possible segmentation errors, the manual segmentation correction process can be optimized to reduce radiologist time and effort. We propose to use prediction-time augmentations to estimate slice-wise and scan-wise segmentation uncertainty and errors.

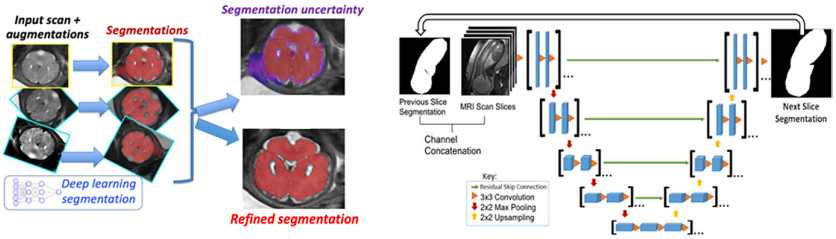

Fig. 2. Left: Prediction-time augmentations yield multiple segmentations (red). The segmentation uncertainty (blue) is then computed per voxel with the entropy of the multiple predictions (top). The final segmentation (bottom) is computed by averaging the aligned predictions. Right: slice segmentation error correction 2D U-Net architecture. The inputs are a slice, two adjacent slices above/below it, and a validated previous slice segmentation; it outputs the corrected slice segmentation. Subsequent slices can be iteratively corrected (top arrow). (Color figure online)

Slice-wise segmentation uncertainty is estimated by computing the sum of the segmentation uncertainty of each voxel v in the slice defined by the predictions binary entropy: $uncertainty(v) = P(v)\log(P(v)) + (1 - P(v))\log(1 - P(v))$ where $P(v)$ is the predicted probability of voxel v to belong to the target structure as computed by the prediction-time augmentations. The larger the voxel entropy, the higher the uncertainty value for the voxel segmentation prediction (Fig. 2 left). We use the segmentation uncertainty to distinguish between segmentation variability and segmentation error [25]. Segmentation variability is the acceptable deviation from the ground-truth: it should be removed from the segmentation uncertainty to obtain the segmentation error. We use the morphological opening operator to remove the segmentation variability around the mean segmentation contour. The estimated error value of a slice is computed by the sum of the voxels' uncertainty after filtering out the segmentation variability.

Scan-wise segmentation uncertainty measures the deviation of the individual prediction-time augmentation segmentations S_i from the mean segmentation $\bar{S}$ with an uncertainty function $u(S_i, \bar{S})$ that measures the distance of the predicted segmentation to the mean segmentation in the entire scan. Given N segmentations from the augmented scans, the mean segmentation $\bar{S} = \frac{1}{N}\sum_i^N S_i$ is computed first. Then, the uncertainty measure $u(S_i, \bar{S})$ of every segmentation S_i from the mean segmentation $\bar{S}$ is computed. Finally, the overall median uncertainty value $\bar{u} = median_i\left(\left\{u(S_i, \bar{S})\right\}\right)$ is computed. The uncertainty function is computed with standard measures, e.g., Intersection-over-Union (IoU), Dice coefficient and Average Symmetric Surface Distance (ASSD).

We use the resulting value as an estimation of the network uncertainty about its segmentation and to detect possible segmentation errors. The intuition is that the confidence of a network prediction is high when it yields small variations on the segmentations resulting from the perturbations induced by the augmentation, and small otherwise. Note that the goal is compute a measure that is well correlated with the actual segmentation error and not to accurately estimate the actual segmentation error value.

Segmentation Error Correction. Slice-wise segmentation errors are corrected by using the previous slice radiologist' validated/corrected segmentation to automatically correct segmentation errors in a slice. Our method uses a 2D U-Net (Fig. 2, right) [2];

it inputs a slice, a validated previous slice segmentation, and four adjacent scan slices (two below and two above the slice); it outputs the slice's corrected structure segmentation. The corrected segmentation can then be used to correct the subsequent slices in an iterative automatic segmentation error correction process. The order in which slices are corrected can be prioritized by the estimated slice-wise segmentation error value with the largest values shown first.

3 Experimental Results

For the experimental studies, we collected two datasets of fetal brain and fetal body MRI scans from the Sourasky Medical Center acquired as part of the routine clinical fetal assessment. The fetal body dataset consists of 64 fetal body MRI coronal scans acquired on a 1.5T GE Signa Horizon Echo speed LX MRI scanner using a torso coil with the volumetric FIESTA protocol. Each scan has 50–100 slices, 256 × 256 pixels per slice, with resolution of 1.56 × 1.56 × 3.0 mm^3. The fetal brain dataset consists of 42 fetal brain MRI coronal scans acquired on a 3T Siemens Skyra MRI scanner using a torso coil with the 3D fast imaging TrueFISP sequence. Each scan consists of 20–40 slices, 512 × 512 pixels per slice, with resolution of 0.74 × 0.74 × 3.0–5.0 mm^3.

Expert-validated ground-truth fetal body and fetal brain segmentations were created for all scans as follows. Manual segmentations of the fetal body and the fetal brain were created by two expert radiologists for 13 and 8 scans (each scan requires on average 74 and 55 min to annotate). Validated segmentations for the remaining 46 and 34 scans were created by an expert radiologist by correcting the segmentations produced by the automatic segmentation method. The original and the corrected segmentations are used in Study 2 below. To quantify the manual segmentation variability, two annotators with expertise in fetal MRI performed manual segmentations: for the fetal body, 21 scans (1,741 slices) were segmented by one annotator and 10 scans were each segmented twice by both annotators; for the fetal brain, 3 scans (97 slices) were delineated twice by both annotators. Table 1 (rows 1, 2) lists the delineation observer variability results.

We conducted two studies to evaluate our methods. The results are reported for the fetal body and fetal brain on training sets of 9 and 6 scans and test sets of 55 and 29 scans, respectively. The segmentations quality is evaluated with the Dice and ASSD.

For the fetal structures segmentation networks, the 3D patch size is set to 128 × 128 × 32 to ensure that the receptive field contains most of the scan. The first two encoder and the last two decoder layers perform anisotropic sampling to reach patches of size 32 × 32 × 32. The remaining layers perform isotropic sampling.

Study 1: Fetal Structures Segmentation. This study compares the accuracy of various segmentation architectures and quantifies the effectiveness of the automatic fetal body and fetal brain segmentation. We performed four experiments by comparing our segmentation method to: 1) an L-Net classifier [14]; 2) a standard 2D U-Net [2]; 3) our method without and 4) with prediction-time augmentations. In all cases, the segmentation results are refined with standard post-processing techniques. Table 1 (rows 3–6) shows the results. Note that prediction-time augmentation method improves the fetal body (brain) Dice score from 0.95 to 0.96 (0.95 to 0.96) and the ASSD from 1.52

Table 1. Segmentation accuracy results for the fetal body and fetal brain. The first two rows list the inter- and intra-observer manual segmentation variability; they serve as the reference for comparing the results of the segmentation methods. The next six rows list the networks architectures and methods; the columns indicate the segmentation metric scores (mean and std).

	Method	Fetal body		Fetal brain	
		Dice	ASSD (mm)	Dice	ASSD (mm)
1	Intra-observer variability	0.94 ± 0.01	0.90 ± 0.85	0.96 ± 0.01	0.26 ± 0.15
2	Inter-observer variability	0.93 ± 0.02	0.84 ± 0.78	0.96 ± 0.01	0.22 ± 0.12
3	L-Net	0.79 ± 0.08	6.31 ± 7.51	0.84 ± 0.07	1.23 ± 0.62
4	2D U-Net	0.93 ± 0.07	1.35 ± 1.70	0.94 ± 0.06	0.75 ± 1.22
5	3D U-Net	0.93 ± 0.06	1.62 ± 1.41	0.94 ± 0.05	0.64 ± 0.48
6	Two-step segmentation	0.95 ± 0.03	1.42 ± 1.32	0.95 ± 0.03	0.21 ± 0.13
7	Two-step segmentation + prediction-time augmentation	$\mathbf{0.96 \pm 0.02}$	$\mathbf{0.74 \pm 0.51}$	$\mathbf{0.96 \pm 0.02}$	$\mathbf{0.19 \pm 0.09}$
8	Previous-slice correction	0.97 ± 0.02	0.38 ± 0.31	0.97 ± 0.02	0.13 ± 0.05

to 0.74 mm (0.21 to 0.19), both below the observer variability measures. anisotropic sampling reduces the errors of Dice and ASSD by ~5% over isotropic sampling.

Study 2: Segmentation Error Estimation and Correction Prioritization. This study evaluates the segmentation error estimation and correction methods. First, note that the relatively high std with respect to the mean (Table 1, row 6) indicates that the method fails to produce accurate segmentations for a number of slices and scans, which should be identified for correction. The correlation coefficients between the estimated and the actual segmentation Dice errors computed by linear regression are 0.94 and 0.95 for the fetal brain and body, respectively. This indicates that the segmentation error estimations are reliable and can be used to identify segmentations requiring corrections.

We investigate the use of the estimated segmentation error measure to prioritize the manual segmentation errors correction process. The goal is to optimize the radiologist time and effort by correcting first the most significant segmentation errors instead of in random or sequential order, as in the current practice. We measure the test set mean segmentation error as a function of the % of the segmentations in the individual slices and scans that were corrected by the radiologist. We evaluate scan-wise and slice-wise prioritization as follows. In scan-wise prioritization, scans are ordered and corrected in descending order of their estimated scan segmentation error value. In slice-wise prioritization, the slices are partitioned into groups of five successive slices; the slice-wise segmentation error estimations of each group is averaged, and the slices are sorted by this value. This prioritization policies are compared to random and optimal scan prioritization, in which the scan segmentations are corrected in descending order of their actual segmentation error values.

Figure 3 (left) shows the results of the prioritization on the fetal body dataset. Note that observer variability accuracy is achieved by slice-wise ordered correction of 12% of the segmentations, vs. 20% and 33% in scan-wise and random order prioritization.

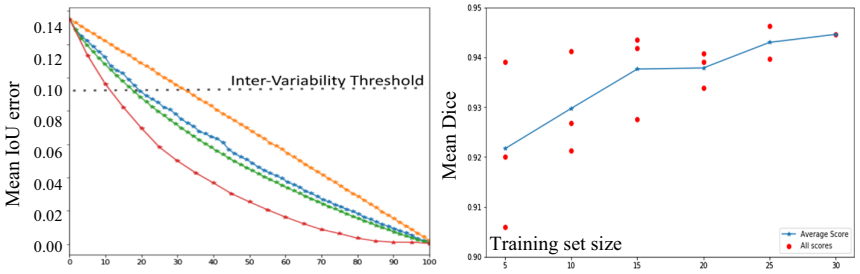

Fig. 3. Fetal body segmentation results. Left: mean IoU error (vertical axis) as a function of the # of slices corrected by the radiologist (horizontal axis). The plots show four prioritization policies: scan-wise random order (orange), scan-wise (blue) and slice-wise (red) descending order of estimated segmentation error value, and scan-wise optimal order by descending order of actual segmentation error value (green). The horizontal dotted line shows the manual segmentation observer variability. Right: mean Dice (vertical axis) as a function of the training set size (horizontal axis). Each red dot is a test result of a random training set of predefined size; blue lines show the mean test set score per training set size.

Table 1 (row 7) shows the effectiveness of using our automatic slice segmentation error correction method (Fig. 2, right), that achieves an ASSD lower by 49% and 32% from our best segmentation method for total body and brain datasets, respectively.

Study 3: Active Learning with Segmentation Error Estimation. We explore the use of segmentation error estimation for active learning [11]. The goal is to use the segmentation error estimation to select scans to augment the training set, thereby enhancing the performance of the automatic structures segmentation network.

We quantify the effect of random training sets sampling to establish a comparison baseline for training scans selection policies and to quantify the segmentation results variability of a fixed training set size. We train the fetal body segmentation networks on training sets of various sizes – 5, 10, 15, 20, 25, 30 – randomly chosen from a pool of 30 scans and tested on 30 scans. Figure 3 (right) shows the results. Note that small training sets can achieve results comparable to larger training sets when the training samples are selected differently: the best Dice (0.94) of a training set of size five is better than the worst Dice (0.93) of a training set of size 15. This suggests a potential savings in expert annotation time by judicious selection of scans added to the training set.

Finally, we quantify the addition of corrected, ground-truth annotated scans to the training dataset based on the segmentation error estimation. We test three policies for augmenting a training dataset of size 5 with 5 additional training scans based on the estimated segmentation error using the mean Dice measure of a test set of size 30. The Dice (ASSD) of the networks trained with 5 training scans is 0.92 ± 0.05 (2.1 ± 2.2). Adding 5 more scans based on their estimated segmentation errors and re-training the networks with the augmented training dataset of 10 annotated scans yields the following

Dice (ASSD) results: 0.92 ± 0.03 (2.1 ± 2.1) for the lowest estimated segmentation error, 0.93 ± 0.03 (1.7 ± 1.8) for randomly selected scans, and 0.95 ± 0.02 (1.2 ± 1.5) for the highest estimated segmentation error. These results suggest that scans with the highest estimated segmentation error should be prioritized for augmenting the training set.

4 Conclusion

We have presented a method for the end-to-end volumetric segmentation of fetal structures in MRI scan that optimizes radiologist validation and annotation time. Our method uses custom anisotropic 3D U-Net networks in a two-step process that extracts the structure ROI and computes its segmentation; the networks are trained with very few annotated scans. The segmentation error estimation stage leverages prediction-time augmentations of the input scan to compute multiple segmentations and to estimate the segmentation error for individual slices and for the entire scan based on the segmentation uncertainty estimations. These estimations are used to locate the most error prone slices and to iteratively correct the segmentations in those slices based on validated adjacent slices with a 2D U-Net slice correction network. Our method achieves state-of-the-art fetal structures segmentation results and provides effective segmentation error estimation and correction methods that enable the prioritization of the radiologist time and the effective creation of large validated datasets.

Our experimental results indicate that segmentation uncertainty and error estimation are useful for active learning and for training dataset selection and annotation optimization, thereby saving costly annotation time by utilizing the expert annotators' efforts efficiently on fewer scans. Our methods can be used to create a dataset of radiologist-validated segmentations for the accurate and robust automatic segmentation of complex structures in volumetric scans with very few annotated scans.

References

1. Litjens, G., et al.: A survey of deep learning in medical image analysis. Med. Image Anal. **42**, 60–88 (2017)
2. Ronneberger, O., Fischer, P., Brox, T.: U-Net: convolutional networks for biomedical image segmentation. In: Navab, N., Hornegger, J., Wells, W.M., Frangi, A.F. (eds.) MICCAI 2015. LNCS, vol. 9351, pp. 234–241. Springer, Cham (2015). https://doi.org/10.1007/978-3-319-24574-4_28
3. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D U-Net: learning dense volumetric segmentation from sparse annotation. In: Ourselin, S., Joskowicz, L., Sabuncu, M.R., Unal, G., Wells, W. (eds.) MICCAI 2016. LNCS, vol. 9901, pp. 424–432. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46723-8_49
4. Milletari, F., Navab, N., Ahmadi, S.A.: V-Net: fully convolutional neural networks for volumetric medical image segmentation. In: Proceedings of the 4th IEEE International Conference on 3D Vision (2016)
5. Isensee, F., Kickingereder, P., Wick, W., Bendszus, M., Maier-Hein, K.H.: Brain tumor segmentation and radiomics survival prediction: contribution to the BRATS 2017 challenge. In: Crimi, A., Bakas, S., Kuijf, H., Menze, B., Reyes, M. (eds.) BrainLes 2017. LNCS, vol. 10670, pp. 287–297. Springer, Cham (2018). https://doi.org/10.1007/978-3-319-75238-9_25

6. Karimi, D., Samei, G., Shao, Y., Salcudean, S.: A novel deep learning-based method for prostate segmentation in T2-weighted magnetic resonance imaging. arXiv:1901.09462 (2019)
7. Salehi, S.S.M., et al.: Real-time automatic fetal brain extraction in fetal MRI by deep learning. In: Proceedings of the IEEE 15th International Symposium on Biomedical Imaging – ISBI 2018, pp. 720–724 (2018)
8. Gibson, E., et al.: NiftyNet: a deep-learning platform for medical imaging. Comput. Methods Progr. Biomed. **158**, 113–122 (2018)
9. Gholipour, A., et al.: A normative spatiotemporal MRI atlas of the fetal brain for automatic segmentation and analysis of early brain growth. Sci. Rep. **7**(1), 1–13 (2017)
10. Fetit, A.E., et al.: A deep learning approach to segmentation of the developing cortex in fetal brain MRI with minimal manual labeling. In: Medical Imaging with Deep Learning (2020)
11. Veeraraghavan, H., Miller, J.V.: Active learning guided interactions for consistent image segmentation with reduced user interactions. In: Proceedings of the IEEE International Symposium on Biomed Imaging (2011)
12. Lee, N., Caban, J., Ebadollahi, S., Laine, A.: Interactive segmentation in multimodal medical imagery using a Bayesian transductive learning approach. In: Proceedings of the IEEE International Symposium on Biomedical Imaging (2009)
13. Wang, G., Li, W.: Interactive medical image segmentation using deep learning with image-specific fine tuning. IEEE Trans. Med. Imaging **37**, 1562–1573 (2018)
14. Braginsky, M., Joskowicz, L.: Interactive segmentation of structures with real-time fine-tuning of a fully convolutional neural network. M.Sc. thesis, The Hebrew University of Jerusalem (2019)
15. Lakshminarayanan, B., Pritzel, A., Blundell, C.: Simple and scalable predictive uncertainty estimation using deep ensembles. In: Advances in Neural Information Processing Systems (2017)
16. Guo, C., Pleiss, G., Sun, Y., Weinberger, K.Q.: On calibration of modern neural networks. In: Proceedings of the 34th International Conference on Machine Learning, vol. 70 (2017)
17. Wang, G., Li, W., Aertsen, M., Deprest, J., Ourselin, S., Vercauteren, T.: Aleatoric uncertainty estimation with test-time augmentation for medical image segmentation with convolutional neural networks. Neurocomputing **338**, 34–45 (2019)
18. Loquercio, A., Segu, M., Scaramuzza, D.: A general framework for uncertainty estimation in Deep Learning. arXiv preprint arXiv:1907.06890 (2019)
19. Ulyanov, D., Vedaldi, A., Lempitsky, V.: Instance normalization: the missing ingredient for fast stylization. arXiv preprint arXiv:1607.08022 (2016)
20. He, K., Zhang, X., Ren, S., Sun, J.: Identity mappings in deep residual networks. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9908, pp. 630–645. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46493-0_38
21. Tompson, J., Goroshin, R., Jain, A., Lecun, Y., Bregler, C.: Efficient object localization using convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2015)
22. Lin, H., et al.: Deep learning for low-field to high-field MR: image quality transfer with probabilistic decimation simulator. In: Knoll, F., Maier, A., Rueckert, D., Ye, J.C. (eds.) MLMIR 2019. LNCS, vol. 11905, pp. 58–70. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-33843-5_6