

# The role of MRI physics in brain segmentation CNNs: achieving acquisition invariance and instructive uncertainties

Pedro Borges<sup>1,2</sup>, Richard Shaw<sup>1,2</sup>, Thomas Varsavsky<sup>1,2</sup>, Kerstin Klaser<sup>2</sup>  
David Thomas<sup>3</sup>, Ivana Drobnjak<sup>1</sup>, Sebastien Ourselin<sup>2</sup>, and M Jorge Cardoso<sup>2</sup>

<sup>1</sup>Department of Medical Physics and Biomedical Engineering, UCL, UK

<sup>2</sup>School of Biomedical Engineering and Imaging Sciences, KCL, UK

<sup>3</sup>Dementia Research Centre, UCL, UK

**Abstract.** Being able to adequately process and combine data arising from different sites is crucial in neuroimaging, but is difficult, owing to site, sequence and acquisition-parameter dependent biases. It is important therefore to design algorithms that are not only robust to images of differing contrasts, but also be able to generalise well to unseen ones, with a quantifiable measure of uncertainty. In this paper we demonstrate the efficacy of a physics-informed, uncertainty-aware, segmentation network that employs augmentation-time MR simulations and homogeneous batch feature stratification to achieve acquisition invariance. We show that the proposed approach also accurately extrapolates to out-of-distribution sequence samples, providing well calibrated volumetric bounds on these. We demonstrate a significant improvement in terms of coefficients of variation, backed by uncertainty based volumetric validation.

## 1 Introduction

Magnetic Resonance Imaging (MRI) is one of the most widespread neuroimaging techniques owing to its excellent soft tissue contrast, boasting great versatility in highlighting different regions and pathologies by means of sequence selection. As a consequence, a significant body of work has emerged developing accurate processing algorithms for MR images that may arise from different sites and acquisition sequence parameters. There are those works that focus on achieving algorithms that can generalise well to all contrasts. Traditional and largely widespread techniques include probabilistic generative models [2] and multi-atlas fusion methods [17]. However, the former has strong assumptions on label intensity distributions, and the latter is predicated on lengthy processing times due to its dependence on image registration. Recent works using convolutional neural networks (CNNs), such as Billot *et al.* [3], tackle contrast agnosticism by employing a Bayesian generative segmentation model that synthesises images containing multiple different contrasts. Jog *et al.* [12] devise an approach by which networks can be made to generalise to unseen contrasts by predicting

pulse sequence parameters from such images, and simulating images of that contrast by using labelled multiparametric map datasets. Pham *et al.* [16] employ an iterative approach involving a dual segmentation-synthesis model, whereby images of unseen contrasts are segmented, used to train a synthesis network that in turn is used to generate new images of the unseen contrast from the labels in the original training set. It is important to note that, while these methods are able to segment data from unseen sites with some degree of accuracy, they do not model the interaction between acquisition parameters and the underlying anatomy explicitly - they segment what they see and not the true anatomy.

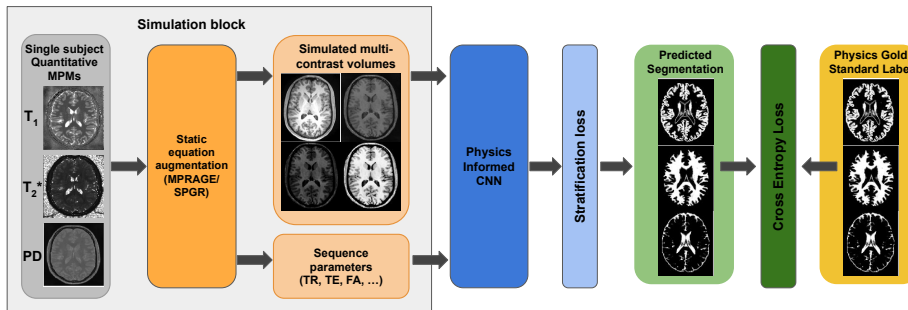
This leads to those methods that seek to harmonise measurements across sites by directly accounting for such covariates as scanner and site bias, and sequence contrast variabilities, e.g. ComBat [13] is a Bayesian framework designed to account for experimental variabilities that has been applied to cortical thickness harmonisation [7]. These classes of techniques, however, operate directly on extracted volumetric measurements and not on the images. Harmonisation has also been tackled with CycleGANs [20] [19] and domain adaptation approaches [5].

Recent work [4] proposed a means to directly introduce the physics of the MR acquisition process directly into deep learning networks in combination with pre-generated synthetic MR images based on multi-parametric MR maps (MPMs). This work achieves some agnosticism to the underlying physics by demonstrating that generated segmentations are more consistent volumetrically. This method, however, does not enforce volumetric consistency across contrasts, and has not been shown to extrapolate to out of distribution sequence parameters.

Changes in MRI acquisition parameters alter the tissue contrast, thus impacting the algorithmic ability to accurately segment images; this can be modeled via uncertainty estimation. Here, we propose to model both epistemic (ability of the model to know) and aleatoric (unknowns of the data) uncertainties. Building on existing work [4], we also introduce a new training approach and consistency loss across realisations of MRI contrasts, allowing the model to appropriately disentangle the anatomical phenotype and the MRI physics, and extrapolate to unseen contrasts without sacrificing segmentation quality.

## 2 Methods

Borges *et al.* [4] proposed that a network could be made resilient to changes in the physics parameters, and therefore be able to appropriately segment data produced by different sequences. This was achieved by generating simulated data, and passing this imaging data and associated MRI parameters to a CNN. In order to train against a "Physics Gold Standard" (PGS), i.e. a true model of the anatomy that is not influenced by the choice of acquisition parameters, the authors used a Gaussian Mixture Model of literature sourced tissue parameters for grey matter (GM), white matter (WM), and cerebrospinal fluid (CSF) on their quantitative MPMs. We build on this work and improve the algorithmic robustness, ameliorating image quality, segmentation volume consistency, and



**Fig. 1.** The training pipeline with proposed new additions of single subject batch stratification and accompanying  $L_2$  feature maps loss, and training time image simulation.

validating within and out of distribution samples paired with uncertainty derived errors.

## 2.1 Network architecture

In Borges *et al.* [4], the injection of the physics parameters into the network is done via the inclusion of two fully connected layers whose output is tiled and concatenated to the ante-penultimate convolutional layer output. We adopt a similar strategy, but take the added step of also tiling this output to an earlier region of the network, immediately preceding the first down-sampling layer. We argue that knowledge of the physics is potentially valuable information in the encoding portion of the network, and that this allows it to better disentangle the physics parameters and the subject’s phenotype.

We moved to adopt the nn-UNet architecture [10]. All networks were trained with batch size 4, on 3D patches of size  $128^3$  sampled from the simulated volumes. Networks were trained with a learning rate of  $10^{-4}$  until convergence, where convergence is defined as 7 epochs elapsing without an improvement in validation metrics, Dice score combined with coefficient of variation (CoV). We made use of two main frameworks for this work, TorchIO [15], and MONAI [1].

As the proposed method requires multi-parametric data to train the model, a more scarce resource in large numbers, a dataset comprised of 18 subjects were used for training, four for validation, and five for inference/ testing.

## 2.2 Stratification and batch homogeneity

We seek to further enforce volumetric consistencies vis-à-vis same-subject realisations generated using different sequence parameters. We therefore propose a batch stratification approach where each batch contains multiple realisations of images from a single subject. This allows for the addition of a stratification loss over the batch features of the penultimate layer of our network, which acts in addition to the standard cross-entropy segmentation loss. As the segmentation

ground truths remain consistent across same subject simulations (because the underlying anatomy is unchanging), if a given batch contains multiple simulations from a single subject (and same patch location for patch-based training networks), then the features maps at the end of the network should also be consistent across simulations. This is enforced by introducing an  $L_2$  loss over all the final feature maps for each batch. The inclusion of the physics parameters should make this tenable, as it allows for the network to learn to disentangle the anatomical phenotype and the MRI-physics related appearance.

### 2.3 Casting simulation as an augmentation layer

We adopt the same static equation multi-parametric map based simulation approach as Jog *et al.* [11], focusing on MPRAGE and SPGR sequences. The SPGR equation describing the signal  $b_S$  per voxel,  $x$  is:

$$b_S(x) = G_S PD(x) \sin\theta \frac{1 - e^{-\frac{TR}{T_1(x)}}}{1 - \cos\theta e^{-\frac{TR}{T_1(x)}}} e^{-\frac{TE}{T_2^*(x)}}, \quad (1)$$

where  $G_S$  is the scanner gain,  $TR$  the repetition time,  $T_1$  the longitudinal relaxation time,  $TE$  the echo time, and  $T_2^*$  the transverse relaxation time.

Similarly, for MPRAGE:

$$b_M(x) = G_S PD(x) \left( 1 - \frac{2e^{-\frac{TI}{T_1(x)}}}{1 + e^{-\frac{-(TI+TD+\tau)}{T_1(x)}}} \right), \quad (2)$$

where  $TD$  the delay time, and  $\tau$  the slice imaging time.

Unlike in [4], where simulated volumes are all generated prior to network training, we implemented the static equation simulation layer as an augmentation layer. Such a layer takes as input a 4D multi-parametric map, a protocol type, and a range of relevant parameters to randomly sample from, producing  $N$  (batch size) simulated volumes. This layer-based batch approach is compatible with our posed stratification model, as all generated volumes per batch belong to the same subject, permitting the utilisation of the within-batch feature consistency loss. The full training pipeline is depicted in Fig. 1.

### 2.4 Uncertainty modelling

We opt to incorporate uncertainty modelling in our framework to obtain volumetric bounds on our segmentations. We model the aleatoric uncertainty via explicit loss attenuation [14]. We modify our network architecture to include an additional convolutional block that branches off the final upsampling layer. This branch models the aleatoric uncertainty,  $\sigma_{Het}^W(x)$ . This modifies the cross-entropy loss function accordingly:

$$\hat{x}_{i,t} = f_i^W + \epsilon_t, \quad \epsilon_t \sim \mathcal{N}(0, (\sigma_i^W)^2) \quad (3)$$

$$\mathcal{L} = \sum_i \log \frac{1}{T} \sum_t c_i(-\hat{x}_{i,t,c} + \log \sum_{c'} e^{\hat{x}_{i,t,c'}}) \quad (4)$$

Where  $\hat{x}_{i,t}$  are the task logits ( $f_i^W$ ) summed with a noise sample of standard deviation equal to the predicted  $\sigma_i^W$  per voxel;  $T$  denotes the number of stochastic passes per input, and  $\sigma^W$  is defined for every voxel, per class,  $c$ . This allows for the easy extraction of volumetric bounds by repeatedly sampling from additive logit noise distributions to produce new segmentations.

The epistemic uncertainty is modelled using test-time Monte Carlo sampling via dropout. Dropout is commonly used as a regularisation technique [18], but also allows for the approximate Bayesian posterior sampling of segmentations by maintaining the random neuron switching at test-time [8]. We set a dropout level of 0.5 in all layers except for the input layer, where it is set to 0.05.

### 3 Experiments

#### 3.1 Data

We make use of a 27 subject multi-parametric early onset Alzheimer dataset, the same as in [4], for the purpose of simulating images which are used for training, validating, and testing of our models, all of which are registered to MNI space rigidly. The images contain maps of the longitudinal and effective transverse magnetisation relaxation,  $R_1$  and  $R_2^*$ , proton density,  $PD$ , and magnetisation transfer,  $MT$ . The details concerning quantitative map creation can be found in [9]. The static equation models we employ feature  $T_1$  (inverse of  $R_1$ ),  $T_2^*$  (inverse of  $R_2^*$ ), and  $PD$ .

#### 3.2 Simulation sequence details

To allow for direct comparability, we limited the ranges of the relevant parameters for simulated images at training time to those stipulated in the original work, i.e. inversion time (TI) = [600-1200] ms for MPRAGE, repetition time (TR) = [15-100] ms, echo time (TE) = [4-10] ms, and flip angle (FA) = [15-75] degrees for SPGR. For each subject, a single "Physics Gold Standard" (PGS) segmentation was used across the associated synthesized images, generated using the same process and literature values as in the original work [4].

### 4 Annealing study: Robustness and quality analysis

To ascertain the contributions of the two main additions to the underlying method, we carry out an annealing study, whereby we analyse the incremental performance increases in terms of volume consistency and Dice score, with the addition of each change. We begin with a complete physics-agnostic baseline, i.e. a standard 3D nn-UNet trained with pre-generated data (Baseline), followed by the original physics method (Phys-Base), followed by Phys-Base with the addition of batch stratification (Phys-Strat), followed lastly by Phys-Strat with the addition of the simulation augmentation scheme (Phys-Strat-Aug).

**Table 1.** Mean dice scores for Baseline, Phys-Base, Phys-Strat, and Phys-Strat-Aug on segmentation task, across inference subjects. All dice scores are estimated against a Physics Gold Standard. Standard deviations quoted in brackets. Bold values represent statistically best performances.

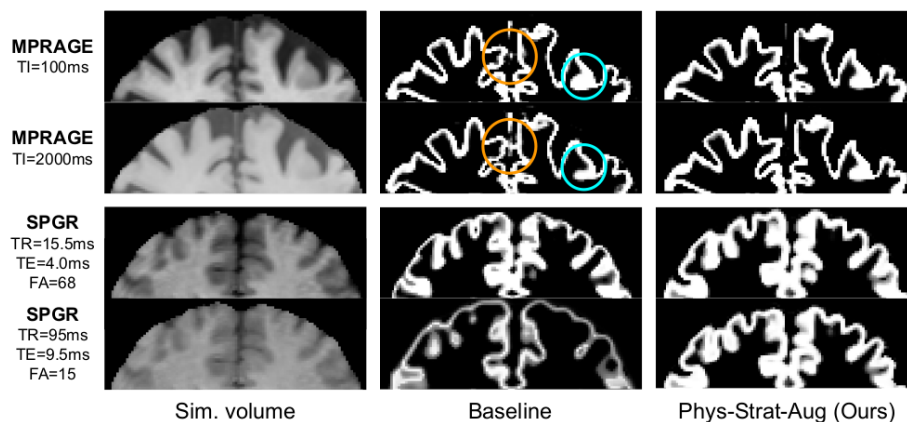
Experiments	Sequence Dice Scores							
	MPRAGE				SPGR			
	GM		WM		GM		WM	
	IoD	OoD	IoD	OoD	IoD	OoD	IoD	OoD
Baseline	0.966 (0.005)	0.956 (0.006)	0.953 (0.002)	0.934 (0.002)	0.878 (0.021)	0.872 (0.008)	0.893 (0.023)	0.873 (0.011)
Phys-Base	<b>0.971</b> <b>(0.007)</b>	0.964 (0.009)	<b>0.964</b> <b>(0.008)</b>	<b>0.959</b> <b>(0.011)</b>	0.911 (0.020)	0.872 (0.050)	0.912 (0.021)	0.880 (0.092)
Phys-Strat	<b>0.970</b> <b>(0.005)</b>	<b>0.969</b> <b>(0.005)</b>	0.958 (0.004)	<b>0.957</b> <b>(0.005)</b>	<b>0.929</b> <b>(0.015)</b>	<b>0.911</b> <b>(0.011)</b>	<b>0.922</b> <b>(0.021)</b>	0.894 (0.040)
Phys-Strat-Aug	<b>0.971</b> <b>(0.004)</b>	<b>0.971</b> <b>(0.005)</b>	<b>0.962</b> <b>(0.003)</b>	<b>0.960</b> <b>(0.004)</b>	<b>0.930</b> <b>(0.016)</b>	<b>0.913</b> <b>(0.019)</b>	<b>0.921</b> <b>(0.015)</b>	<b>0.899</b> <b>(0.019)</b>

**Table 2.** Coefficients of variation (CoV) for Baseline, Phys-Base, Phys-Strat, and Phys-Strat-Aug on segmentation task, averaged across test subjects. Standard deviations quoted in brackets. Bold values represent statistically best performances.

Experiments	Sequence CoVs ( $\times 10^3$ )							
	MPRAGE				SPGR			
	GM		WM		GM		WM	
	IoD	OoD	IoD	OoD	IoD	OoD	IoD	OoD
Baseline	6.39 (0.87)	22.50 (4.08)	14.94 (1.71)	51.12 (7.11)	61.91 (7.61)	170.10 (31.32)	32.57 (11.98)	158.93 (16.83)
Phys-Base	2.72 (2.12)	14.67 (7.30)	3.28 (2.01)	28.10 (3.98)	77.22 (34.44)	127.22 (18.61)	20.77 (9.35)	264.80 (8.52)
Phys-Strat	0.71 (0.23)	6.15 (1.51)	0.53 (0.25)	<b>3.67</b> <b>(1.34)</b>	21.83 (0.83)	59.78 (13.31)	8.60 (0.64)	59.19 (11.25)
Phys-Strat-Aug	<b>0.42</b> <b>(0.22)</b>	<b>4.74</b> <b>(1.30)</b>	<b>0.51</b> <b>(0.23)</b>	<b>3.65</b> <b>(0.62)</b>	<b>15.76</b> <b>(1.18)</b>	<b>28.88</b> <b>(9.74)</b>	<b>7.12</b> <b>(0.45)</b>	<b>44.78</b> <b>(4.22)</b>

We extend our volumetric consistency analysis by analysing out of distribution (OoD) samples. In this instance they are defined as simulated images whose sequence parameters lie outside of the training range. This not only results in images of unfamiliar contrasts, but also unseen parameters that are fed into the physics branch of the network. If our method has truly attained a measure of sequence invariance then it should be expected that both segmentation quality and volume consistency are maintained as the network should be able to extrapolate from the provided values. For MPRAGE, the OoD range is extended to [100-2000] ms, while for SPGR, the TR is extended to [10-200] ms, TE is extended to [2-20] ms, and FA is extended to [5-90] degrees.

Table 1 and Table 2 show Dice and CoV performances, respectively. We carry out signed-rank Wilcoxon tests to test for statistically significant improvements,



**Fig. 2.** Baseline and Phys-Strat-Aug comparisons. Comparing out-of-distribution MPRAGE (Top two rows) and SPGR (Bottom two rows) GM segmentations from the proposed and baseline methods. Blue circles highlight examples of significant gyrus variability. Orange circles denote regions of segmentation differences between protocols.

and bold the best model ( $p$ -value  $< 0.01$ ). Tests are carried out on CoV and Dice scores independently of each other. In instances where models may outperform baselines but are not statistically significantly different from each other, we bold both. We verify an incremental gain in CoV and Dice with each added feature, the most pronounced of which results from the addition of the stratification loss, in terms of both in and out of distribution CoVs. This is expected, as directly optimising for consistency across realisations of the same subject should more strongly enforce volume consistency, enhancing the physics invariance.

Phys-Strat-Aug boasts the best performance overall, significantly outperforming both Baseline and Phys-Base with regards to CoV. Compared to Phys-Strat, the differences are not always statistically better for MPRAGE, but are so for SPGR. With more parameters at play, an augmentation scheme should become more relevant, as sampling from the parameter space should lead to a greater extrapolating ability, as the network is no longer constrained to learn from a more discrete training set, and will experience more varied realisations.

Fig. 2 shows some qualitative results, in and out of distribution segmentation comparisons between Baseline and Phys-Strat-Aug, to convey the consistency the latter is able to achieve without compromising segmentation quality.

#### 4.1 Uncertainty measures and volumetric bounds

Given Phys-Strat-Aug’s superior performance, we train only two epistemic uncertainty models, and two aleatoric uncertainty models with this pipeline, one of each for this pipeline and a complete baseline.

At test-time we extract 50 aleatoric volume samples, and 50 epistemic volume samples for each of the networks, for both in and out of distribution simulated

images. We verify that the aleatoric samples do not contribute significantly to the volume variance in comparison to its epistemic counterpart, (an observation that was also been verified in [6]) and therefore omit it in our volumetric analysis.

Fig. 3 showcases white matter volume variations for MPRAGE and SPGR sequences, for the extended out of distribution parameter ranges, for Baseline and Phys-Strat-Aug experiments, for a single subject. For the SPGR plot, we order the points based on volumetric consistency for each experiment, thus highlighting outliers. In both instances we observe a much greater consistency in volume for Phys-Strat-Aug, itself a reflection of the aforementioned CoV results. Using the calibrated volumetric method described in [6] allows us to calculate volume percentiles for each set of dropout samples, and the errors represent the volumetric interquartile range (IQR).

The errors for the baseline do not vary in any statistically significant manner, for either sequence or tissue, independent of any volume deviation. It is a different matter for Phys-Strat-Aug, however. Specifically, for MPRAGE, we note that uncertainties are consistently larger for Phys-Strat-Aug compared to baseline, and that furthermore, Phys-Aug-Strat segmentations boast larger uncertainties for out of distribution samples. This can perhaps be explained by the additional level of uncertainty introduced by the physics, and how the presence of a physics parameter outside of the “known” further exacerbates this effect.

For SPGR, all the apparent outliers for Phys-Strat-Aug have significantly larger associated errors, while this is not the case for the Baseline. We observe that most outliers correspond to out of distribution samples boasting very low flip angles ( $< 10^\circ$ , highlighted in black in the figure). Such images will be significantly less  $T_1$ -weighted, and therefore be less familiar to the models, resulting in poorer segmentation quality, so it is reassuring that the physics-informed network’s uncertainty around these samples is larger.

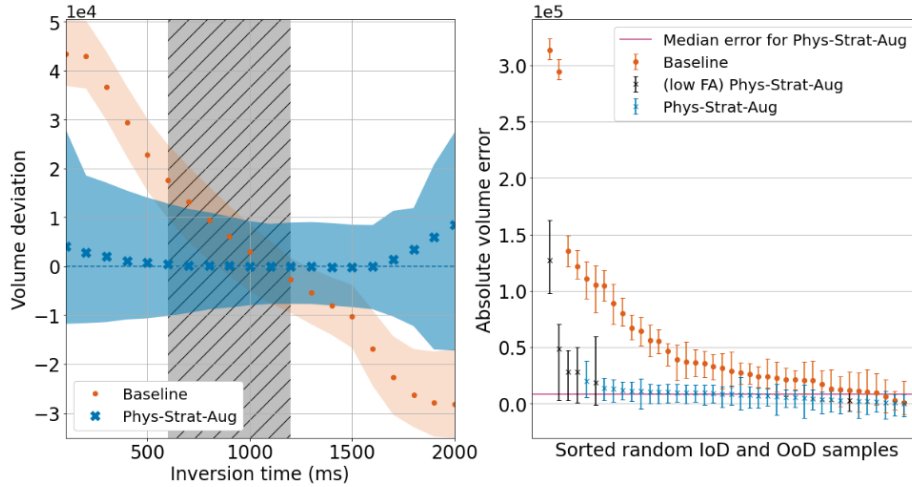
## 5 Discussion and Conclusions

In this work we demonstrated that with some well justified modifications to the training pipeline, a physics-informed network can achieve extremely constrained tissue segmentations across a wide range of contrasts, across all tissue types and investigated sequences; thus strengthening its harmonisation capabilities.

Furthermore, we also showed that it can suitably generalise to unseen domains, while maintaining volume consistency without compromising segmentation quality, and is validated by accurately quantifying the volumetric uncertainty. The uncertainty estimates further suggest that the physics knowledge grants the model an additional level of safety, as volumetric uncertainties proved to be larger for out of distribution parameter generated images.

The method is admittedly limited by those sequences that can be aptly represented as a static equation, but we argue that at the very least, for the purposes of contrast agnosticism, a wide enough range of realistic contrasts can be generated with currently implemented sequences, which should allow for our method to generalise further. Future work will therefore involve testing of our method





**Fig. 3.** Volume consistency for WM for complete baseline and Phys-Strat-Aug, for example subject. Filled plots/ Error bars correspond to IQR volumes. Left: MPRAGE. The dashed grey region denotes the TI training time parameter range (600 - 1200 ms). Right: SPGR. Black points denote samples with FA lower than  $10^\circ$  for Phys-Strat-Aug.

on multiple external datasets to ascertain generalisability and the exploration of techniques that may allow for the modelling of MR artifacts such as movement and  $B_0$  inhomogeneities, to enhance our model's utility.

**Acknowledgements** This project was funded by the Wellcome Flagship Programme (WT213038/Z/18/Z) and Wellcome EPSRC CME (WT203148/Z/16/Z).

## References

1. Project MONAI. <https://doi.org/10.5281/zenodo.4323059>
2. Ashburner, J., Friston, K.J.: Unified segmentation. *NeuroImage* **26**(3), 839–851 (2005). <https://doi.org/10.1016/j.neuroimage.2005.02.018>
3. Billot, B., Greve, D., Van Leemput, K., Fischl, B., Iglesias, J.E., Dalca, A.V.: A Learning Strategy for Contrast-agnostic MRI Segmentation. *arXiv* (2020)
4. Borges, P., Sudre, C., Varsavsky, T., Thomas, D., Drobnjak, I., Ourselin, S., Cardoso, M.J.: Physics-informed brain MRI segmentation. *Lecture Notes in Computer Science* **11827 LNCS**, 100–109 (2020). [https://doi.org/10.1007/978-3-030-32778-1\\_11](https://doi.org/10.1007/978-3-030-32778-1_11)
5. Dinsdale, N.K., Jenkinson, M., Namburete, A.I.: Unlearning Scanner Bias for MRI Harmonisation in Medical Image Segmentation. In: *Communications in Computer and Information Science*. vol. 1248 CCIS, pp. 15–25. Springer (2020). <https://doi.org/10.1007/978-3-030-52791-4>
6. Eaton-Rosen, Z., Bragman, F., Bisdas, S., Ourselin, S., Cardoso, M.J.: Towards safe deep learning: accurately quantifying biomarker uncertainty in neural network

- predictions. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **11070 LNCS**, 691–699 (2018)
7. Fortin, J.P., Cullen, N., Sheline, Y.I., Taylor, W.D., Aselcioglu, I., Cook, P.A., Adams, P., Cooper, C., Fava, M., McGrath, P.J., McInnis, M., Phillips, M.L., Trivedi, M.H., Weissman, M.M., Shinohara, R.T.: Harmonization of cortical thickness measurements across scanners and sites. *NeuroImage* **167**, 104–120 (2018). <https://doi.org/10.1016/j.neuroimage.2017.11.024>
  8. Gal, Y., Ghahramani, Z.: Dropout as a Bayesian Approximation: Representing Model Uncertainty in Deep Learning. 33rd International Conference on Machine Learning, ICML 2016 **3**, 1651–1660 (2015)
  9. Helms, G., et al.: Increased snr and reduced distortions by averaging multiple gradient echo signals in 3d flash imaging of the human brain at 3t. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine* **29**(1), 198–204 (2009)
  10. Isensee, F., Kickingereder, P., Wick, W., Bendszus, M., Maier-Hein, K.H.: No New-Net. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* **11384 LNCS**, 234–244 (2018)
  11. Jog, A., Carass, A., Roy, S., Pham, D.L., Prince B Amod, J.L.: MR Image Synthesis by Contrast Learning On Neighborhood Ensembles . <https://doi.org/10.1016/j.media.2015.05.002>
  12. Jog, A., Hoopes, A., Greve, D.N., Van Leemput, K., Fischl, B.: PSACNN: Pulse Sequence Adaptive Fast Whole Brain Segmentation (2019)
  13. Johnson, W.E., Li, C., Rabinovic, A.: Adjusting batch effects in microarray expression data using empirical Bayes methods. *Biostatistics* **8**(1), 118–127 (2007). <https://doi.org/10.1093/biostatistics/kxj037>
  14. Kendall, A., Gal, Y.: What Uncertainties Do We Need in Bayesian Deep Learning for Computer Vision? Tech. rep. (2017)
  15. Pérez-García, F., Sparks, R., Ourselin, S.: TorchIO: a Python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning (2020)
  16. Pham, D.L., Chou, Y.Y., Dewey, B.E., Reich, D.S., Butman, J.A., Roy, S.: Contrast Adaptive Tissue Classification by Alternating Segmentation and Synthesis. In: Burgos, N., Svoboda, D., Wolterink, J.M., Zhao, C. (eds.) *Simulation and Synthesis in Medical Imaging*. pp. 1–10. Springer International Publishing, Cham (2020)
  17. Sabuncu, M.R., Yeo, B.T., Leemput, K.V., Fischl, B., Golland, P.: A generative model for image segmentation based on label fusion. *IEEE Transactions on Medical Imaging* **29**, 1714–1729 (10 2010). <https://doi.org/10.1109/TMI.2010.2050897>, <https://pubmed.ncbi.nlm.nih.gov/20562040/>
  18. Srivastava, N., Hinton, G., Krizhevsky, A., Salakhutdinov, R.: Dropout: A Simple Way to Prevent Neural Networks from Overfitting. Tech. Rep. 56 (2014)
  19. Zhao, F., Wu, Z., Wang, L., Lin, W., Xia, S., Shen, D., Li, G.: Harmonization of infant cortical thickness using surface-to-surface cycle-consistent adversarial networks. In: *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. vol. 11767 LNCS, pp. 475–483. Springer (2019)
  20. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks. *Proceedings of the IEEE International Conference on Computer Vision* **2017-October**, 2242–2251 (2017)