

Domain-Agnostic Stroke Lesion Segmentation Using Physics-Constrained Synthetic Data

Liam Chalcraft¹[0000–0003–3363–6454], Jenny Crinion²[0000–0001–8080–6562],
Cathy J. Price¹[0000–0001–7448–4835], and John Ashburner¹[0000–0001–7605–2518]

¹ Department of Imaging Neuroscience, University College London

² Institute of Cognitive Neuroscience, University College London
l.chalcraft@cs.ucl.ac.uk

Abstract. Segmenting stroke lesions in MRI is challenging due to diverse acquisition protocols that limit model generalisability. In this work, we introduce two physics-constrained approaches to generate synthetic quantitative MRI (qMRI) images that improve segmentation robustness across heterogeneous domains. Our first method, **qATLAS**, trains a neural network to estimate qMRI maps from standard MPRAGE images, enabling the simulation of varied MRI sequences with realistic tissue contrasts. The second method, **qSynth**, synthesises qMRI maps directly from tissue labels using label-conditioned Gaussian mixture models, ensuring physical plausibility. Extensive experiments on multiple out-of-domain datasets show that both methods outperform a baseline UNet, with **qSynth** notably surpassing previous synthetic data approaches. These results highlight the promise of integrating MRI physics into synthetic data generation for robust, generalisable stroke lesion segmentation. Code is available at <https://github.com/liamchalcraft/qsynth>

Keywords: Segmentation, Synthetic Data, Quantitative MRI

1 Introduction

Segmenting brain pathologies in MRI is vital for clinical and research applications, yet remains challenging due to the variability of acquisition protocols across hospitals. Recent work shows that even for diffusion-weighted MRI - arguably the modality of choice in acute stroke - dedicated CNNs can equal human raters when trained on large, carefully curated cohorts [17]. While public datasets achieve high performance using standardised sequences (e.g., T1w, T2w, FLAIR) with similar parameters, clinical data rarely conforms to these ideal conditions [21, 6]. Existing domain adaptation methods either require prior target domain knowledge or many unlabelled images [10, 25], and synthetic approaches like SynthSeg [2] can produce unrealistic contrasts for heterogeneous stroke lesions [1, 8].

To address these limitations, we propose generating synthetic images using quantitative MRI (qMRI) parameters and physics-based forward models to ensure physical plausibility. qMRI provides voxel-level tissue properties - proton

density (PD), longitudinal relaxation rate (R_1), effective transverse relaxation rate (R_2^*), and magnetisation transfer (MT) - that enable the simulation of diverse MRI sequences while preserving tissue characteristics. Since acquiring full qMRI data is time-consuming, recent deep learning methods estimate qMRI maps from standard sequences [4, 24, 3], making large-scale synthetic data generation feasible.

We introduce two methods for domain-agnostic stroke lesion segmentation. The first, **qATLAS**, trains a qMRI estimation model on MPAGE images to augment the ATLAS dataset [16] with simulated sequences. The second, **qSynth**, extends previous synthetic approaches [1, 8] by sampling qMRI maps from intensity priors derived from real data, ensuring greater physical realism. By embedding MRI physics into data synthesis, our framework bridges the gap between synthetic and clinical data, thereby improving segmentation robustness across diverse imaging domains.

2 Methods

We propose two methods, **qATLAS** and **qSynth**, for domain-agnostic stroke lesion segmentation. Both leverage qMRI parameter maps to generate diverse, physics-constrained training data, enhancing robustness and generalisability.

2.1 qATLAS: Estimating qMRI from MPAGE

qATLAS fits an nnU-Net [14] that predicts four qMRI maps (PD, R_1 , R_2^* , MT) from a single MPAGE[†]. Training used 51 subjects (22 healthy, 29 stroke; 80/20 split) with 3D-EPI ground-truth maps from hMRI [22]. Thus any routine MPAGE can be converted to qMRI and fed to our physics-based simulator.

Diverse MPAGE images were simulated using NiTorch with parameters: $T_R \sim \mathcal{U}(1.9, 2.5)$ s, $T_I \sim \mathcal{U}(0.6, 1.2)$ s, $T_E \sim \mathcal{U}(2, 4)$ ms, $\alpha \sim \mathcal{U}(5^\circ, 12^\circ)$, and $B_0 \sim \mathcal{U}(0.3, 7)$ T.

Data Augmentation: We applied standard augmentations (elastic/affine deformations, bias field, Gibbs ringing, Rician noise, and random cropping to 192^3 voxels) using MONAI [7]. Figure 1 illustrates examples of the augmented training data.

Model Architecture and Training: Our U-Net comprises five encoder stages (channels: 24, 48, 96, 192, 384), each with two residual units, using GELU activations [13], instance normalisation [23], linear upsampling, and 0.1 dropout. The network outputs four channels corresponding to PD, R_1 , R_2^* , and MT. To enforce positivity, PD, R_1 , and R_2^* are computed as the exponential of their raw outputs, while MT is computed as 100 times the sigmoid of its raw output.

Training was performed over 200,000 iterations (batch size 1) using AdamW (lr= 10^{-4} , $\beta_1 = 0.9$, $\beta_2 = 0.999$, weight decay=0.01). An initial 20,000 iterations

[†]Estimating R_2^* from one T1 is ill-posed; errors mainly affect simulated T2w/FLAIR.

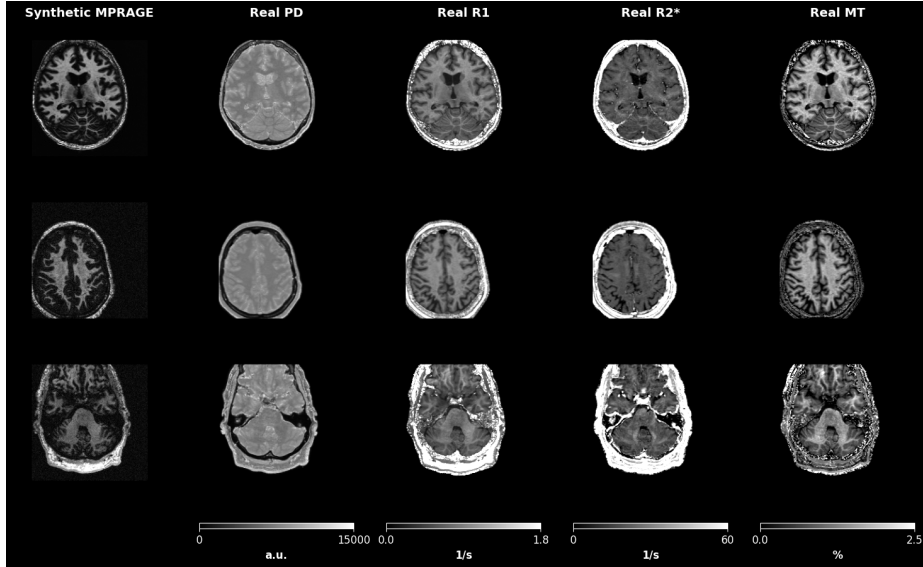


Fig. 1: Examples of training data for qMRI parameter map prediction in the qATLAS method.

used L2 loss; thereafter, we employed a combined loss $L = L1(y, \hat{y}) + L2(y, \hat{y}) + L1(\nabla y, \nabla \hat{y}) + L2(\nabla y, \nabla \hat{y})$ augmented by a perceptual LPIPS loss [27] (weight 0.1) using features from a pretrained Med3D encoder [9].

Figure 2 displays examples of predicted qMRI maps from input MPRAGE images in the ATLAS dataset [16].

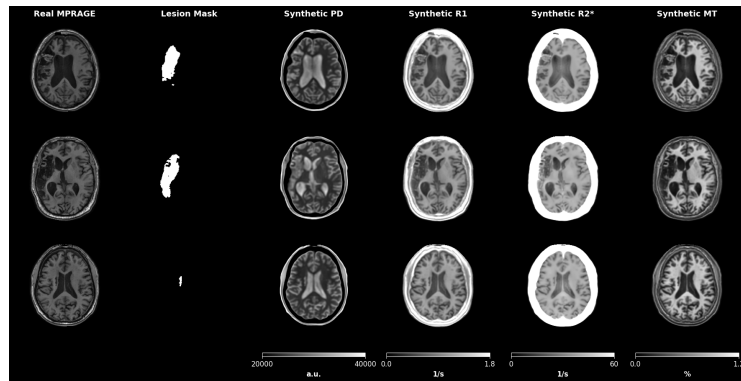


Fig. 2: Examples of predicted qMRI parameter maps from input MPRAGE images of the ATLAS dataset in the qATLAS method.

Simulation of MRI Sequences: We used the estimated qMRI maps from the ATLAS dataset to simulate diverse MRI sequences via a physics-based generative model (see Section 2.3), thereby creating the qATLAS dataset for segmentation training.

2.2 qSynth: Synthesising qMRI Maps from Tissue Labels

The **qSynth** method generates synthetic qMRI maps directly from segmentation labels using label-conditioned Gaussian Mixture Models (GMMs). We first define prior distributions for each qMRI parameter (PD, R_1 , R_2^* , MT) for different tissue types (gray matter, white matter, cerebrospinal fluid, and lesions). These priors are estimated from the population used in qATLAS (for anatomy or pathology where real data is not available, it is feasible that parameters could be estimated based on values reported in literature). For each voxel in a segmentation label, the corresponding qMRI parameters are sampled from the appropriate prior, thereby producing a synthetic qMRI map that reflects realistic tissue properties. Healthy tissue maps are obtained using the Multibrain SPM toolbox [5], and lesion masks are generated by overlaying random binary lesion maps onto these healthy tissue maps (see [8] for further details).

Simulation of MRI Sequences: The synthetic qMRI maps are then used with our physics-based generative model (Section 2.3) to simulate diverse MRI sequences, resulting in the **qSynth** dataset for training an alternative segmentation model.

2.3 Physics-Based Generative Model

Both qATLAS and qSynth use a physics-based generative model to simulate realistic MRI images from qMRI parameter maps via standard signal equations. We simulate several common MRI sequences, including Fast Spin-Echo (FSE), Gradient-Echo (GRE), Fluid-Attenuated Inversion Recovery (FLAIR), and Magnetisation-Prepared Rapid Gradient Echo (MPRAGE).

For example, the FSE signal is computed as:

$$S_{\text{FSE}} = B_1 \text{PD} \left(1 - e^{-R_1 T_R}\right) e^{-R_2 T_E},$$

where B_1 is the receive field strength, PD is the proton density, R_1 and R_2 ($\approx R_2^*$) are the relaxation rates, and T_R and T_E are the repetition and echo times, respectively. Similar equations are used for GRE, FLAIR, and MPRAGE sequences.

Acquisition parameters (e.g., T_R , T_E , T_I , T_X , T_D , and α) were sampled from physically plausible distributions; please refer to our code repository[§] for the complete sampling details.

[§]<https://github.com/liamchalcroft/qsynth>

To mimic realistic MRI data, we add Rician noise by perturbing the simulated signal S_{MRI} with independent Gaussian noise on the real and imaginary components:

$$S_{\text{noisy}} = \sqrt{(S_{\text{MRI}} + n_r)^2 + n_i^2}, \quad n_r, n_i \sim \mathcal{N}(0, \sigma^2).$$

All simulations and noise additions were implemented using the NiTorch library[¶]. Additional augmentations - such as elastic deformations, bias field variations, axis flips, Gaussian noise, low-resolution reslicing, and random cropping to $192 \times 192 \times 192$ voxels - were applied using MONAI [7].

2.4 Segmentation Model Training

We trained nnUNet-based segmentation models [14] on four settings: (i) real MPRAGE images from the ATLAS dataset (**baseline**), (ii) the **qATLAS** dataset (from MPRAGE-derived qMRI maps), (iii) a synthetic data model using the public **Synth** method [8], and (iv) the **qSynth** method. The **qSynth** method differs from **Synth** in that **Synth** directly samples the random tissue intensities, while **qSynth** uses intensity priors to first synthesise qMRI parameter maps, which may then be used to generate realistic tissue intensities via random variations of real forward models. For **Synth** and **qSynth**, we additionally evaluated models trained on a mixture of synthetic and real ATLAS images.

Model Architecture and Training Details: All models use PReLU activations [12] with one residual unit per block. The **qATLAS** model performs binary segmentation (background vs. stroke lesion), while the **qSynth** model predicts additional healthy tissue classes (gray matter, white matter, GM/WM partial volume, and cerebrospinal fluid). Baseline and **qATLAS** models were trained on the ATLAS dataset (N=419/105/131 train/validation/test), whereas **Synth** and **qSynth** models used ATLAS lesion labels combined with healthy tissue labels from OASIS-3 (N=2579/100 train/validation) as detailed in [8].

All models were optimised using a combined Dice and cross-entropy loss with the AdamW optimiser [19] ($\text{lr}=10^{-4}$, $\beta_1 = 0.9$, $\beta_2 = 0.999$, weight decay=0.01) and a learning rate scheduler $\eta_n = \eta_0 (1 - \frac{n}{N})^{0.9}$. Training was run for 700,000 iterations with a batch size of 1.

3 Experiments

We evaluated segmentation models on four datasets: ATLAS, comprising 131 subjects (isotropic MPRAGE) [16] for in-domain evaluation; ARC, with 229 subjects (T1w, T2w, FLAIR) [11, 15] for out-of-domain testing; PLORAS, consisting of 406 subjects (106 T2w, 300 FLAIR) from various UK hospitals; and ISLES 2015, including 28 subjects (T1w, T2w, FLAIR, DWI) [20] with acute lesions.

[¶]<https://github.com/balbasty/nitorch>

All scans were resliced to 1 mm, histogram-normalised and z -scored, then segmented with a 192^3 sliding window (50 % overlap, Gaussian blending) and 8-flip TTA [26]. For multi-modal sets (ARC, ISLES15) we averaged per-voxel logits across contrasts before soft-max and post-processing (reported as 'Ensemble'). Dice and HD95 (95-percentile Hausdorff, mm) were computed on predictions/GT padded to 256^3 ; only the *lesion* channel is reported for **Synth/qSynth**.

Dataset caveat. **Synth/qSynth** are trained on a much larger multi-site cohort with multi-class labels, whereas **qATLAS** and the baseline use **ATLAS** only (binary masks). Thus absolute score differences combine physics, training-set size, label granularity and pre-stripped skulls in ISLES15; isolating each factor is left for future work.

4 Results

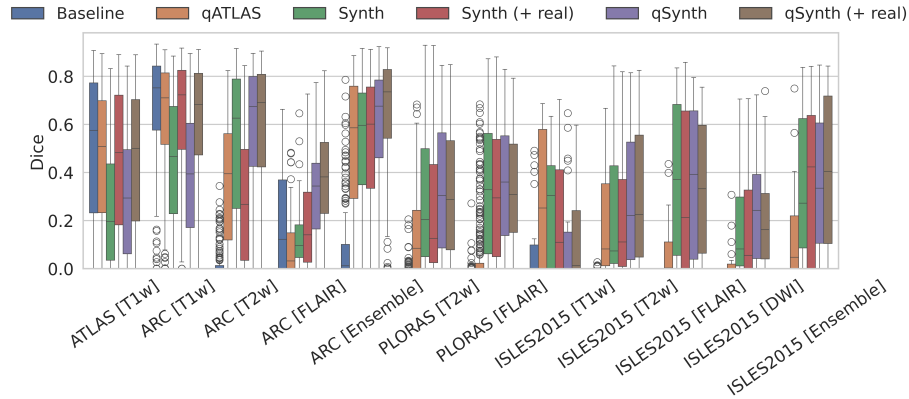


Fig. 3: Dice metric performance for all datasets.

ATLAS: The ATLAS test set is considered in-domain. The **baseline** model is expected to perform optimally, while the qMRI generator model must accurately reproduce MPRAGE details to match this performance. As shown in Tables 1 & 2 and Figure 3, the **baseline** outperformed the qMRI-based methods, with no statistically significant differences except between **Synth** and **qSynth**. Notably, **qSynth** significantly outperformed **Synth**, indicating that the physics-constrained pipeline reduces domain shift between simulated and real data.

ARC: The ARC dataset exhibits moderate domain shift in the T1w channel and larger shifts in T2 and FLAIR. As shown in Tables 1 & 2 and Figure 3, **qATLAS** underperformed relative to the **baseline** (though not significantly). For T1w, **qSynth** initially lagged behind **Synth** but reached comparable performance when real data was integrated; in T2w, **qSynth** outperformed all models, and in

Table 1: Median **Dice** (top line) and 95 % CI (tiny line). Best, second, and third best are **bold**, underlined, and *italic*.

Dataset	Modality	Baseline	qATLAS	Synth	Synth + Real	qSynth	qSynth + Real
ATLAS	T1w	0.575 (0.522–0.628)	0.508 (0.460–0.556)	0.197 (0.155–0.239)	0.482 (0.431–0.534)	0.294 (0.250–0.338)	<i>0.501</i> (<i>0.450–0.551</i>)
ARC	T1w	0.752 (0.713–0.790)	<i>0.710</i> (<i>0.673–0.747</i>)	0.467 (0.430–0.504)	0.723 (<u>0.684–0.762</u>)	0.394 (0.359–0.429)	0.683 (0.643–0.723)
	T2w	0.000 (0.000–0.007)	0.395 (0.363–0.426)	<i>0.626</i> (<i>0.588–0.665</i>)	<u>0.268</u> (0.233–0.302)	<u>0.674</u> (0.640–0.709)	0.691 (0.655–0.726)
	FLAIR	0.122 (0.076–0.169)	0.032 (0.005–0.059)	0.096 (0.070–0.123)	<i>0.141</i> (<i>0.097–0.185</i>)	<u>0.344</u> (0.300–0.388)	0.382 (0.332–0.431)
	Ensemble	0.012 (0.000–0.032)	0.586 (0.549–0.622)	0.597 (0.562–0.631)	<i>0.602</i> (<i>0.566–0.638</i>)	<u>0.676</u> (0.642–0.709)	0.735 (0.701–0.770)
PLORAS	T2w	0.000 (0.000–0.006)	0.085 (0.049–0.122)	<i>0.204</i> (<i>0.150–0.258</i>)	0.126 (0.073–0.179)	0.304 (0.253–0.356)	0.288 (0.238–0.338)
	FLAIR	0.000 (0.000–0.002)	0.000 (0.000–0.017)	<u>0.328</u> (<u>0.298–0.358</u>)	0.294 (0.264–0.325)	0.361 (0.333–0.388)	<i>0.309</i> (<i>0.283–0.335</i>)
ISLES15	T1w	0.000 (0.000–0.064)	0.252 (<u>0.142–0.363</u>)	0.304 (0.213–0.396)	<i>0.110</i> (<i>0.012–0.208</i>)	0.002 (0.000–0.078)	0.013 (0.000–0.087)
	T2w	0.000 (0.000–0.002)	0.082 (0.000–0.170)	0.074 (0.000–0.184)	<i>0.111</i> (<i>0.007–0.216</i>)	0.222 (<u>0.114–0.329</u>)	0.225 (0.117–0.333)
	FLAIR	0.000 (0.000–0.000)	0.004 (0.000–0.052)	0.372 (<u>0.255–0.489</u>)	0.212 (0.085–0.340)	0.392 (0.281–0.504)	<i>0.332</i> (<i>0.226–0.439</i>)
	DWI	0.000 (0.000–0.000)	0.001 (0.000–0.027)	<i>0.082</i> (<i>0.000–0.168</i>)	0.056 (0.000–0.144)	0.243 (0.155–0.331)	0.162 (0.076–0.249)
	Ensemble	0.000 (0.000–0.000)	0.048 (0.000–0.123)	0.272 (0.157–0.388)	0.423 (0.302–0.545)	<i>0.335</i> (<i>0.222–0.448</i>)	0.404 (0.284–0.524)

FLAIR and ensemble evaluations, it achieved statistically significant improvements, highlighting its robustness under domain shift.

Why do real scans help more in some contrasts than others? We currently model the entire lesion with a *single* Gaussian intensity prior. This approximation is adequate for T2w or FLAIR images, where stroke lesions are nearly isointense within the lesion core, so mixing in real data offers only marginal gain and can even dilute the synthetic coverage. In T1w scans stroke appearance is markedly more heterogeneous, so a single-mode prior under-represents the true lesion distribution. Injecting real ATLAS images therefore supplies the missing high-variance examples and yields the largest Dice improvement for T1w. Future work could address this imbalance by using multi-modal lesion priors.

PLORAS: The PLORAS dataset, containing real clinical data from UK hospitals, highlights the superiority of **qSynth** over **baseline** and **qATLAS** models as seen in Tables 1 & 2. **qSynth** also demonstrated moderate but consistent improvements over **Synth**, reflecting its capacity to generalise effectively to diverse real-world clinical scenarios.

ISLES 2015: The ISLES 2015 dataset, featuring co-registered T1w, T2w, FLAIR, and DWI channels with acute stroke lesions, posed unique challenges. DWI channels, absent in qMRI simulations, were unseen during **qATLAS**/**qSynth** training. Nevertheless, **qSynth** significantly outperformed all models, including **Synth**, which could have trained on DWI-like data. **qSynth** achieved the highest Dice

Table 2: 95th-percentile Hausdorff distance (**HD95, mm**) - median on top, 95 % CI beneath. Lowest, 2nd-lowest, and 3rd-lowest scores are **bold**, underlined, and *italic*.

Dataset	Modality	Baseline	qATLAS	Synth	Synth+Real	qSynth	qSynth+Real
ATLAS	T1w	19.7 (10.2–29.3)	<i>34.4</i> (26.5–42.2)	63.7 (59.6–67.9)	<u>22.6</u> (12.0–33.3)	51.3 (43.5–59.2)	38.0 (28.2–47.8)
ARC	T1w	8.8 (2.5–15.0)	<i>11.8</i> (5.6–18.1)	57.1 (52.9–61.3)	11.0 (4.0–18.0)	31.0 (26.6–35.5)	13.0 (6.8–19.2)
	T2w	<u>74.4</u> (71.4–77.4)	54.5 (50.7–58.2)	<i>46.1</i> (42.2–50.0)	<u>64.0</u> (60.5–67.4)	39.0 (34.8–43.2)	43.0 (38.7–47.2)
	FLAIR	58.5 (49.5–67.4)	59.7 (50.7–68.7)	57.6 (49.2–65.9)	<i>52.2</i> (43.3–61.1)	50.3 (41.3–59.2)	44.0 (34.5–53.5)
	Ensemble	48.2 (40.1–56.3)	16.4 (10.9–22.0)	44.0 (39.9–48.2)	20.1 (14.8–25.3)	31.6 (27.2–36.0)	<i>20.7</i> (16.2–25.2)
PLORAS	T2w	70.9 (65.9–76.0)	59.6 (54.4–64.8)	67.5 (61.3–73.7)	70.3 (64.9–75.8)	60.4 (54.6–66.3)	<i>60.7</i> (55.2–66.3)
	FLAIR	66.8 (61.2–72.4)	65.4 (58.4–72.3)	<i>69.7</i> (66.8–72.6)	<u>74.3</u> (71.4–77.2)	<u>73.3</u> (70.4–76.2)	73.0 (70.2–75.7)
ISLES15	T1w	69.8 (30.0–109.7)	51.5 (40.0–63.0)	58.6 (48.5–68.6)	52.5 (29.9–75.0)	<i>56.3</i> (19.6–93.1)	56.3 (17.8–94.8)
	T2w	63.1 (57.8–68.4)	59.4 (43.3–75.4)	72.3 (65.7–78.9)	<u>75.0</u> (68.4–81.7)	<i>66.1</i> (58.3–73.8)	67.7 (59.8–75.6)
	FLAIR	<u>71.8</u> (38.5–105.2)	<i>65.8</i> (51.2–80.5)	56.1 (49.5–62.7)	56.1 (38.7–73.4)	64.5 (55.0–73.9)	66.0 (57.5–74.5)
	DWI	83.7 (57.3–110.2)	<i>75.9</i> (71.2–80.6)	83.7 (76.3–91.0)	84.9 (77.2–92.6)	<u>75.7</u> (67.9–83.5)	74.4 (67.1–81.7)
	Ensemble	256.0 (0.0–256.0)	60.2 (25.9–94.4)	<i>60.1</i> (49.7–70.5)	47.3 (24.2–70.4)	67.5 (59.7–75.3)	56.3 (48.1–64.5)

scores for T2w and FLAIR modalities, while both **Synth** and **qSynth** showed strong ensemble performance. For T1w, **qATLAS** and **Synth** achieved the best results. The dataset’s skull-stripped images likely conferred an advantage to **Synth**/**qSynth** models compared to other methods.

5 Conclusion

We proposed two physics-aware generators for stroke-lesion segmentation training: **qATLAS**, which derives qMRI maps from a single MPRAGE, and **qSynth**, which samples qMRI from learned intensity priors. **qATLAS** boosts T1-weighted performance but degrades on contrasts dominated by non- R_1 parameters; **qSynth** delivers the most consistent cross-domain gains, yet both remain below fully in-domain models. Future work should explore domain adaptation techniques or domain-conditioning via hypernetworks to further close this gap. Additionally, our framework can be extended to other tasks (e.g., registration, super-resolution), other pathologies (e.g. glioblastoma), and other anatomies (e.g., cardiac, abdomen). We also plan to integrate recent advances in more realistic lesion simulations [18] to further enhance synthetic data fidelity. Overall, embedding MRI physics in data generation represents a promising step toward robust, generalisable segmentation in diverse clinical scenarios.

Acknowledgments. LC is supported by the EPSRC-funded UCL Centre for Doctoral Training in Intelligent, Integrated Imaging in Healthcare (i4health) (EP/S021930/1), and the Wellcome Trust (203147/Z/16/Z and 205103/Z/16/Z). CP is funded by Wellcome (203147/Z/16/Z, 205103/Z/16/Z and 224562/Z/21/Z). This research was supported by NVIDIA and used NVIDIA RTX A6000 48GB.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

Bibliography

- [1] Billot, B., Cerri, S., Leemput, K.V., et al.: Joint segmentation of multiple sclerosis lesions and brain anatomy in MRI scans of any contrast and resolution with CNNs. In: 2021 IEEE 18th International Symposium on Biomedical Imaging (ISBI). IEEE (Apr 2021). <https://doi.org/10.1109/isbi48211.2021.9434127>
- [2] Billot, B., Greve, D.N., Puonti, O., et al.: SynthSeg: Segmentation of brain MRI scans of any contrast and resolution without retraining. *Medical Image Analysis* **86**, 102789 (May 2023). <https://doi.org/10.1016/j.media.2023.102789>
- [3] Borges, P., Fernandez, V., Tudosiu, P.D., et al.: Unsupervised Heteromodal Physics-Informed Representation of MRI Data: Tackling Data Harmonisation, Imputation and Domain Shift, p. 53–63. Springer Nature Switzerland (2023). https://doi.org/10.1007/978-3-031-44689-4_6, http://dx.doi.org/10.1007/978-3-031-44689-4_6
- [4] Borges, P., Shaw, R., Varsavsky, T., et al.: Acquisition-invariant brain mri segmentation with informative uncertainties (2021)
- [5] Brudfors, M., Balbastre, Y., Flandin, G., et al.: Flexible Bayesian Modelling for Nonlinear Image Registration, p. 253–263. Springer International Publishing (2020). https://doi.org/10.1007/978-3-030-59716-0_25
- [6] Brzus, M., Boes, A.D., Bruss, J., Johnson, H.J.: Leveraging high-quality research data for ischemic stroke lesion segmentation on clinical data. In: 2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI). p. 1–5. IEEE (Apr 2023). <https://doi.org/10.1109/isbi53787.2023.10230775>, <http://dx.doi.org/10.1109/ISBI53787.2023.10230775>
- [7] Cardoso, M.J., Li, W., Brown, R., et al.: Monai: An open-source framework for deep learning in healthcare (2022)
- [8] Chalcraft, L., Pappas, I., Price, C.J., Ashburner, J.: Synthetic data for robust stroke segmentation (2024), <https://arxiv.org/abs/2404.01946>
- [9] Chen, S., Ma, K., Zheng, Y.: Med3d: Transfer learning for 3d medical image analysis (2019)
- [10] Dorent, R., Kujawa, A., Ivory, M., et al.: Crossmoda 2021 challenge: Benchmark of cross-modality domain adaptation techniques for vestibular schwannoma and cochlea segmentation. *Medical Image Analysis* **83**, 102628 (Jan 2023). <https://doi.org/10.1016/j.media.2022.102628>, <http://dx.doi.org/10.1016/j.media.2022.102628>
- [11] Gibson, M., Newman-Norlund, R., Bonilha, L., et al.: The Aphasia Recovery Cohort, an open-source chronic stroke repository. *Scientific Data* **11**(1), 1–8 (2024). <https://doi.org/10.1038/s41597-024-03819-7>, <http://dx.doi.org/10.1038/s41597-024-03819-7>
- [12] He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: Surpassing human-level performance on ImageNet classification (2015)
- [13] Hendrycks, D., Gimpel, K.: Gaussian error linear units (gelus) (2023)

- [14] Isensee, F., Jaeger, P.F., Kohl, S.A.A., et al.: nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods* **18**(2), 203–211 (Dec 2020). <https://doi.org/10.1038/s41592-020-01008-z>
- [15] Johnson, L., Newman-Norlund, R., Teghipco, A., et al.: Progressive lesion necrosis is related to increasing aphasia severity in chronic stroke. *NeuroImage: Clinical* **41** (1 2024). <https://doi.org/10.1016/j.nicl.2024.103566>
- [16] Liew, S.L., Lo, B.P., Donnelly, M.R., et al.: A large, curated, open-source stroke neuroimaging dataset to improve lesion segmentation algorithms. *Scientific Data* **9**(1) (Jun 2022). <https://doi.org/10.1038/s41597-022-01401-7>
- [17] Liu, C.F., Hsu, J., Xu, X., et al.: Deep learning-based detection and segmentation of diffusion abnormalities in acute ischemic stroke. *Communications Medicine* **1**(1) (Dec 2021). <https://doi.org/10.1038/s43856-021-00062-8>, <http://dx.doi.org/10.1038/s43856-021-00062-8>
- [18] Liu, P., Aguila, A.L., Iglesias, J.E.: Unraveling normal anatomy via fluid-driven anomaly randomization (2025), <https://arxiv.org/abs/2501.13370>
- [19] Loshchilov, I., Hutter, F.: Decoupled weight decay regularization (2019)
- [20] Maier, O., Menze, B.H., von der Gabelntz, J., et al.: ISLES 2015 - a public evaluation benchmark for ischemic stroke lesion segmentation from multispectral MRI. *Medical Image Analysis* **35**, 250–269 (Jan 2017). <https://doi.org/10.1016/j.media.2016.07.009>
- [21] de la Rosa, E., Reyes, M., Liew, S.L., et al.: A robust ensemble algorithm for ischemic stroke lesion segmentation: Generalizability and clinical utility beyond the isles challenge (2024), <https://arxiv.org/abs/2403.19425>
- [22] Tabelow, K., Balteau, E., Ashburner, J., et al.: hmri – a toolbox for quantitative mri in neuroscience and clinical research. *NeuroImage* **194**, 191–210 (Jul 2019). <https://doi.org/10.1016/j.neuroimage.2019.01.029>, <http://dx.doi.org/10.1016/j.neuroimage.2019.01.029>
- [23] Ulyanov, D., Vedaldi, A., Lempitsky, V.: Instance normalization: The missing ingredient for fast stylization (2017)
- [24] Varadarajan, D., Bouman, K.L., van der Kouwe, A., et al.: Unsupervised learning of mri tissue properties using mri physics models (2021)
- [25] Wang, D., Shelhamer, E., Liu, S., et al.: Tent: Fully test-time adaptation by entropy minimization (2021)
- [26] Wang, G., Li, W., Ourselin, S., Vercauteren, T.: Automatic Brain Tumor Segmentation Using Convolutional Neural Networks with Test-Time Augmentation, p. 61–72. Springer International Publishing (2019). https://doi.org/10.1007/978-3-030-11726-9_6, http://dx.doi.org/10.1007/978-3-030-11726-9_6
- [27] Zhang, R., Isola, P., Efros, A.A., et al.: The unreasonable effectiveness of deep features as a perceptual metric (2018)