Spatial Feature-wise Linear Modulation (SpFiLM) for Contrast Agent-Aware Brain Parcellation

Pushpendra Singh[1 [0009−0004−8956−1732]], Joshua R. Astley[1 [0000−0002−6552−5436]], Roman Rodionov[2], John Duncan[2], Tom Vercauteren[1 [0000−0003−1794−0456]], and Rachel Sparks[1 [0000−0003−1553−7903]]

[1] School of Biomedical Engineering and Imaging Sciences, King's College London, London, UK
[2] Department of Epilepsy, Queen Square Institute of Neurology, University College London, London, UK

Abstract. Most automated brain parcellation tools are developed and validated on T1-weighted (T1w) MRI. Yet, some clinical workflows for which parcellation is relevant only use contrast-enhanced T1w (T1ce) MRI, on which T1w-trained models are less accurate. We present a unified network that parcellates both pre- and post-contrast agent T1w MRI reliably, trained on a combination of the two with conditioning that spatially modulates its response differently for each. Feature-wise Linear Modulation (FiLM) is a known approach for input-based modulation in networks. It applies a per-channel scale and shift uniformly across the input. However, the appearance change between pre- and post-contrast varies locally across the brain, making FiLM suboptimal for our use case. In this work, we introduce Spatial FiLM (SpFiLM), a conditioning layer whose modulation varies spatially, assembling a voxel-wise scale and shift from image-derived spatial patterns. Using a cohort of 134 patients with paired T1w and T1ce MRI parcellated into 106 classes, the addition of SpFiLM layers in a UNet increased the mean Dice on the test set of 25 patients from 80.2% to 84.1%, a 4.9% relative improvement. Adding SpFiLM layers led to the best performance on both pre- and post-contrast MRI, even when controlling for network parameter counts.

Keywords: Brain parcellation · Contrast conditioning · Feature-wise modulation · MRI segmentation · Spatial FiLM

1 Introduction

Brain parcellation assigns every voxel in the brain a label corresponding to a specific anatomic structure. It supports downstream clinical measurements such as region volumes and cortical thickness, both established biomarkers for neurodegenerative disease [16], and helps define targets for radiotherapy planning [2] and stereotactic intracerebral procedures [26]. FreeSurfer [6], FMRIB Software Library (FSL) [12], Advanced Normalization Tools (ANTs) [24], and

deep learning-based FastSurfer [9] are widely used brain parcellation tools designed and validated on T1-weighted (T1w) MRI.

Contrast-enhanced T1w (T1ce) MRI is acquired after administration of the contrast-agent gadolinium. As gadolinium does not cross the blood-brain barrier, structures where this barrier is lacking or disrupted show increased uptake (the intravascular compartment, choroid plexus, pituitary gland, and tumours), making T1ce useful for delineating blood vessels, tumour margins, and active lesions [23]. Brain parcellation tools trained on T1w MRI have systematically different morphometric estimates between paired T1w and T1ce MRI scans [15].

Several methods adapt neural networks to images of different appearance. Domain adaptation retrains a network to a new target domain [21]. Pre- to post-contrast domain adaptation have been proposed using distribution matching [1] or adversarial and paired-consistency training [19]. However, domain adaptation techniques require a dedicated post-training adaptation stage to adjust network weights using target-domain data. A single network that is reliable for both pre- and post-contrast images would remove the need to adjust network weights, or acquire and register an additional scan. This would simplify the training process, since one neural network could be applied to images regardless of contrast agent usage. A unified contrast agent-aware model could reuse common features of shared anatomy while learning contrast agent-specific differences, making more efficient use of limited training data.

Domain-randomisation methods can increase network generalisation, potentially eliminating performance differences between images of different domains, by training on synthetic images of randomised appearance so the network becomes agnostic to acquisition [3]. Such invariance suits arbitrary or unknown acquisitions, but discards usable information, since it cannot exploit systematic, anatomically localised differences between pre- and post-contrast agent MRI scans. Knowledge distillation can transfer shared anatomical knowledge across modalities of the same anatomy [27], typically assuming paired datasets. Network conditioning can work on unpaired data by modulating a network response using information about the input images. Feature-wise Linear Modulation (FiLM) [20] is one established conditioning method that has proven effective for MRI segmentation [14]. FiLM applies one scale and shift parameter to each feature channel, treating image differences as a global adjustment. However, the appearance differences between T1w and T1ce MRI are not uniform.

We introduce SpFiLM, a conditioning layer illustrated in Fig. 1, with spatially varying modulation to account for local appearance differences between T1w and T1ce MRI. In place of the channel-wise scale and shift of FiLM, it learns spatially varying scale and shift parameters but can recover channel-wise FiLM as a special case. We show that this spatial variation consistently improves brain parcellation on T1w and T1ce MRI.

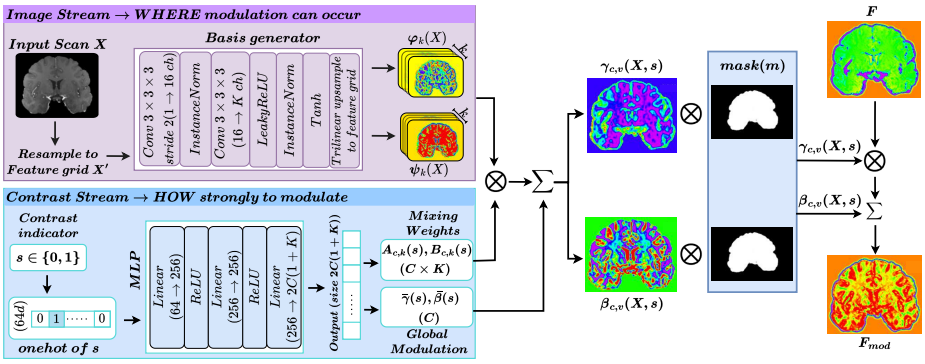

Fig. 1. Overview of Spatial FiLM feature modulation block. Spatial bases $\boldsymbol{\varphi}(X)$ and $\boldsymbol{\psi}(X)$ are learned from the image. Contrast agent indicator s generates only the recombination coefficients $(\bar{\gamma}, \mathbf{A}, \bar{\beta}, \mathbf{B})$. Anatomy thus determines where modulation can occur, while the contrast agent indicator determines how strongly it is applied.

2 Methodology

For a set of unpaired scans, each scan $X \in \mathbb{R}^{1\times D\times H\times W}$ had an associated contrast indicator $s \in \{0, 1\}$. A convolutional neural network was trained to map (X, s) to a dense label map assigning each voxel $v \in X$ to a class. At both training and inference the network receives a single scan X with its indicator s, never both contrasts jointly; s is known from the acquisition record and is an input, not a prediction.

2.1 Channel-wise FiLM

FiLM [20] adds a conditioning layer after each convolutional block in the encoding part of a neural network. For a feature map $F \in \mathbb{R}^{C\times D'\times H'\times W'}$ with channels indexed by c, FiLM applies the per-channel transformation:

$$\mathrm{FiLM}(F_{c,v}) \;=\; \big(1+\gamma_c(s)\big)\, F_{c,v} \;+\; \beta_c(s). \tag{1}$$

The scale γ_c and shift β_c are produced from s by a small network. Following Perez et al. [20], the contrast indicator s is encoded with no learned parameters: the two contrast indicators map to the first two standard basis vectors of $\mathbb{R}^{64}$. This dimension matches the width of the learned embedding it is ablated against (Section 4.2), so the two differ only in whether the mapping is learned. A three-layer MLP with 256, 256, and $2C$ outputs maps the one-hot vector to one scale and one shift value for each of the C channels. Because these values carry no spatial information, every voxel in a channel applies the same scale and shift.

2.2 Spatial FiLM

SpFiLM extends Equation (1) by allowing γ_c and β_c to vary spatially, expressed as $\gamma_{c,v}$ and $\beta_{c,v}$ where voxels are indexed by v. However, an unconstrained Sp-

FiLM would require $\mathcal{O}(CD'H'W')$ parameters (4.6×10^{10} at our 128^3 patch size), would tie the layer to the feature grid it was trained on, and would impose no structure on the feature modulation. Addressing both points, we add structure to $\gamma_{c,v}$ and $\beta_{c,v}$ using a rank-K factorisation to modulate features using patterns learned from the training dataset. Feature modulation is composed of bases inferred from the image X, to represent spatial patterns, and contrast coefficients computed from the contrast indicator s, to represent patterns between contrast indicator types.

Image-conditioned bases. Each conditioning layer contains two independent basis generators, $\boldsymbol{\varphi}(X)$ and $\boldsymbol{\psi}(X)$, which produce the spatial bases for the scale and shift modulation terms, respectively. Each generator computes K basis maps to capture spatial patterns present in the input image. The generators are shared across all contrast indicator types, allowing common anatomical spatial representations to be learned jointly rather than separately.

A conditioning layer modulates the feature map $F \in \mathbb{R}^{C\times D'\times H'\times W'}$ at the encoder stage where it is inserted, with D', H', and W' denoting the spatial dimensions at that stage. The input scan X is resampled to that resolution, giving $X' \in \mathbb{R}^{1\times D'\times H'\times W'}$. Each basis generator then processes X' using a lightweight CNN consisting of two $3 \times 3 \times 3$ convolutional blocks with instance normalisation, containing 16 and K output channels, respectively, separated by a LeakyReLU activation and followed by a final tanh activation that constrains the basis values to $[-1, 1]$.

The first convolution is applied with stride 2, so the basis maps are learned on a feature grid one resolution level coarser than F. They are then returned to the feature resolution by trilinear interpolation. Learning the bases on this coarser grid acts as a regulariser, encouraging smooth, spatially coherent modulation fields that operate over anatomical regions rather than introducing high-frequency, per-voxel variations. After interpolation, the resulting basis maps, $\boldsymbol{\varphi}(X')$ and $\boldsymbol{\psi}(X') \in \mathbb{R}^{K\times D'\times H'\times W'}$, are spatially aligned voxel-for-voxel with the feature map F.

Contrast-conditioned coefficients. A three-layer multi-layer perceptron (MLP) with 256, 256, and $2C(1+K)$ neurons maps the 64-dimensional contrast indicator embedding of s to spatially-invariant channel shifts ($\bar{\boldsymbol{\gamma}}, \bar{\boldsymbol{\beta}} \in \mathbb{R}^C$) and mixing coefficient matrices ($\mathbf{A}, \mathbf{B} \in \mathbb{R}^{C\times K}$).

Conditioning parameter map. The final conditioning parameter maps in SpFiLM are the combination of the spatially-invariant channel shifts, rank-K spatial bases and contrast coefficients computed by:

$$\gamma_{c,v}(X, s) \;=\; \bar{\gamma}_c(s) \;+\; \sum_{k=1}^{K} A_{ck}(s)\, \varphi_{k,v}(X), \tag{2}$$

and similarly $\beta_{c,v}(X, s) = \bar{\beta}_c(s) + \sum_{k=1}^{K} B_{ck}(s)\, \psi_{k,v}(X)$. With K=0, the sum vanishes and Equation (2) collapses to channel-wise FiLM in Equation (1).

Brain mask gating. For brain parcellation, only appearance changes within the brain are relevant for the task. Brain mask gating confines the feature modulation to anatomically meaningful voxels using a binary mask, obtained from skull-stripping (Section 2.3). Skull-stripping the input does not by itself achieve this: the spatially-invariant terms $\bar{\gamma}_c(s)$ and $\bar{\beta}_c(s)$ are applied at every voxel by construction, so without gating the layer is not the identity outside the brain. The brain mask is resampled as m to each feature map F with nearest-neighbour interpolation. The final feature modulation becomes:

$$\mathrm{SpFiLM}(F_{c,v}) \;=\; \big(1 + m_v\,\gamma_{c,v}(X,s)\big)\,F_{c,v} \;+\; m_v\,\beta_{c,v}(X,s). \tag{3}$$

2.3 Implementation Details

The backbone tested was a standard 3D U-Net [22,5] (five levels widths (32, 64, 128, 256, 512), Instance Normalization [25], LeakyReLU) with deep supervision via three auxiliary heads. All input scans were skull-stripped using HD-BET [10] and z-normalised within the resulting brain mask. Unless stated otherwise, models were trained on the combined T1w and T1ce (the *both* regime), treated as unpaired so the two scan types (pre- and post contrast scans) of a patient never co-occurred in a training iteration. The 134 patients (Section 3.1) were split once at the patient level into 97/12/25 (train/val/test), giving 194/24/50 scans with no patient shared across splits. Training used AdamW, 128^3 patches, a deep-supervised Dice plus cross-entropy loss, and standard spatial and intensity augmentation. The SpFiLM modulation was computed in full precision with its fields γ and β clamped to $[-5, 5]$ for numerical stability. At inference, scans were parcellated using a 128^3 sliding window with 50% overlap and Gaussian patch weighting. Full training details are available in the released code at https://github.com/p-singh-kcl/spatial_film_parcellation.

3 Experimental Design

3.1 Dataset

We used a private dataset collected at the National Hospital for Neurology and Neurosurgery (NHNN), London, UK, from a prospective imaging study (REC 20/LO/0966, IRAS 278210) of 230 adult patients with drug-resistant epilepsy. Each patient was scanned with and without gadolinium contrast agent using a 1 mm isotropic MPRAGE acquisition. Patients with anatomy distorting pathology including focal lesions, tumours, and prior surgical resection were excluded from analysis, resulting in a total of 134 patients.

Pseudo ground truth brain parcellations were generated by multi-atlas label fusion, an approach previously used to bootstrap training labels for whole-brain segmentation [7]. The OASIS-TRT-20 Mindboggle-101 atlases [13] were propagated to each T1w MRI using NiftyReg [17], and scan-specific labels were computed via geodesic information flow (GIF) label fusion [4]. Since the atlases are T1w, gadolinium enhancement precludes fusion on T1ce: each T1ce was rigidly

Table 1. Test-set Dice on 25 T1w and 25 T1ce, with params in millions. Best in **bold**, second best underlined. Short names denote Ch-FiLM (channel-wise FiLM), Attn (Attention U-Net), Swin (SwinUNETR), SpFiLM-n (narrow SpFiLM), and nnU+SpFiLM (nnU-Net+SpFiLM). †Param-matched to SpFiLM K=8. ‡Param-matched to U-Net. nnU-Net is ResEnc-L at 300 epochs and the conditioned models use brain-masked input. All ± terms are across-patient standard deviations (n=25 per contrast, n=50 for All); they exceed the between-model differences, hence the paired tests. HD rows report the 95th-percentile Hausdorff distance (HD95, mm). *Per-patient mean Dice differs from SpFiLM (proposed) on both contrasts (paired t-test, Bonferroni α=0.05/16≈0.003 over 8 models × 2 contrasts, n=25).

	Baselines				SotA		Proposed		
	U-Net*	Wide†*	Ch-FiLM*	Attn*	Swin*	nnU-Net*	SpFiLM-n‡*	SpFiLM	nnU+SpFiLM
Params	22.6	27.6	23.6	22.7	35.1	27.5	22.6	27.7	30.7
T1w	82.9 ± 2.0	83.0 ± 2.0	85.9 ± 2.0	85.9 ± 2.2	84.6 ± 2.2	85.7 ± 2.2	86.4 ± 2.0	86.6 ± 2.0	**87.4 ± 2.0**
T1ce	77.6 ± 3.0	77.7 ± 3.0	80.8 ± 3.1	81.1 ± 3.4	79.5 ± 3.4	78.9 ± 3.3	81.4 ± 3.1	**81.6 ± 3.0**	**81.6 ± 3.3**
All	80.2 ± 3.4	80.4 ± 3.4	83.4 ± 3.6	83.5 ± 3.5	82.1 ± 3.5	82.3 ± 3.3	83.9 ± 3.6	84.1 ± 3.6	**84.5 ± 3.2**
HD T1w	2.22 ± 0.50	2.17 ± 0.48	2.18 ± 0.55	2.14 ± 0.48	2.43 ± 0.65	2.15 ± 0.55	2.13 ± 0.46	2.10 ± 0.49	**1.93 ± 0.36**
HD T1ce	2.42 ± 0.49	2.40 ± 0.47	2.43 ± 0.56	2.40 ± 0.49	2.60 ± 0.61	2.41 ± 0.59	2.32 ± 0.47	2.39 ± 0.60	**2.17 ± 0.52**

registered to its T1w and labels propagated by nearest-neighbour interpolation, leaving sub-voxel registration error as the dominant source of label noise.

3.2 Model Comparisons

We compared SpFiLM against four controlled baselines sharing the same U-Net backbone to isolate individual design choices. **U-Net**, the backbone without conditioning. **Wide U-Net** adds convolution kernels to match SpFiLM's parameter count. **Channel-wise FiLM** applies the channel-wise conditioning in Equation (1). **Attention U-Net** [18] adds multi-head attention on the decoder skip connections, applying a spatial modulation without conditioning.

4 Experimental Results

As reported in Table 1, SpFiLM was the best performing model for the full dataset and for T1w and T1ce individually, reaching 84.1 mean Dice, a gain of 3.9% over the U-Net baseline and 0.7% over channel-wise FiLM. Every baseline and state-of-the-art model differed significantly from SpFiLM on both contrasts (per-patient paired t-tests, Bonferroni-corrected; Table 1). On T1w, SpFiLM lowered HD95 against the U-Net baseline, channel-wise FiLM and SwinUNETR under the paired test used for Dice; on T1ce no HD95 difference reached significance. Parameter count alone cannot explain the increase in model performance. Wide U-Net, param count matched to SpFiLM, performed similarly to the baseline, whereas narrow SpFiLM K=8, matched to the baseline, clearly exceeded it. Attention U-Net applied spatial modulation without conditioning and only matched channel-wise FiLM. Fig. 2 shows the learned fields are spatially structured and differ between T1w and T1ce.

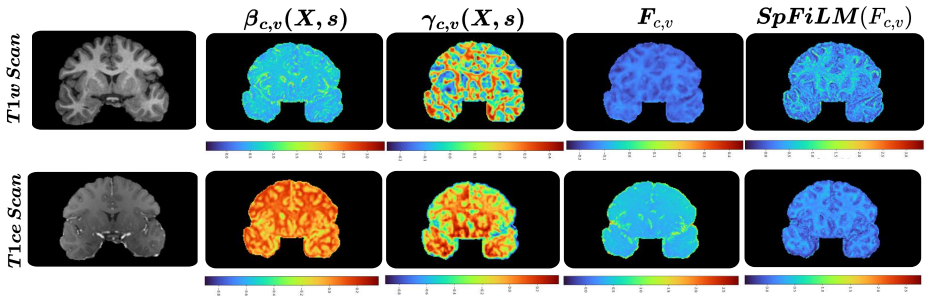

Fig. 2. SpFiLM on a representative subject and feature for T1w and T1ce. **Modulation fields**, left to right: input scan; channel-averaged shift $\beta_{c,v}$; scale $\gamma_{c,v}$; encoder features $F_{c,v}$; modulated features SpFiLM($F_{c,v}$).

4.1 Comparison with the State of the Art

The baselines in Table 1 share SpFiLM's backbone and bound its design choices but not its standing against established methods. We trained two widely used architectures on the same split and protocol, nnU-Net ResEnc-L [11] and SwinUNETR [8] (Table 1). On a plain U-Net backbone SpFiLM already exceeded nnU-Net and the larger SwinUNETR, so its gains were not an artefact of a weak backbone. Inserting the SpFiLM layer into the nnU-Net encoder raised its mean Dice and improved T1ce most, reproducing the pooling benefit of Section 4.2 on a stronger backbone and confirming the conditioning is backbone-agnostic.

4.2 Ablation Study

One-hot encoding. The proposed model uses a frozen one-hot encoding, in which the contrast indicator type contributes no trainable parameters. We compared this against a learned embedding, a two-layer MLP with ReLU that maps the contrast indicator type s into $\mathbb{R}^{64}$ where the MLP is trained jointly with the network. As shown in Table 3, the learned embedding did not improve over the frozen one-hot encoding on either contrast, model gains come from the spatial factorisation mechanism rather than the choice of contrast encoding.

Rank K. We evaluated the effect the number of bases, $K \in \{1, 2, 4, 8, 16\}$, had on model performance. As shown in Table 3, lower K reduced SpFiLM performance. A single basis ($K = 1$) gives one shared spatial pattern for all channels, too restrictive and slightly below channel-wise FiLM ($K = 0$). $K \geq 2$ gives enough independent spatial degrees of freedom to recover the full gain. Because accuracy is already flat for $K \ll C$, the factorisation removes parameters the task does not use: the conditioning layers hold 5.0M parameters at K=8 against 181M for full-rank mixing (K=C).

Table 2. Contrast-specialist training (each model trained only on the contrast it is evaluated on) versus joint training on *both* contrasts. Mean Dice (%) on the test set.

	T1w		T1ce	
Model	Specialist	Both	Specialist	Both
U-Net	82.6	**82.9**	77.5	**77.6**
SpFiLM	86.2	**86.6**	80.7	**81.6**

Table 3. Ablation and sensitivity analysis for SpFiLM. Rank K (left) and insertion level (middle) are sensitivity sweeps, showing stable performance for $K \geq 2$ and across depth: conditioning only the shallowest level already recovers most of the improvement (83.7 versus 84.1). Contrast encoding (right) is an ablation. † marks the proposed configuration. Mean Dice (%) on the held-out test set.

K	T1w	T1ce	Mean	Levels	T1w	T1ce	Mean	Encoding	T1w	T1ce	Mean
1	85.6	80.4	83.0	Top-1	86.2	81.2	83.7	One-hot†	86.6	81.6	84.1
2	86.4	81.4	83.9	Top-2	86.2	81.3	83.8	Learned	86.4	81.5	84.0
4	86.6	81.5	84.0	Top-3	86.5	81.4	83.9				
8†	86.6	81.6	84.1	All†	86.6	81.6	84.1				
16	86.6	81.5	84.1								

T1w and T1ce model training. We compared *specialist* models, each trained only on the contrast it is evaluated on, against models trained on the combined T1w and T1ce data (*both*), for the unconditioned U-Net and for SpFiLM (Table 2). Pooling improved SpFiLM on T1ce by +0.9% but the U-Net by only +0.1%. The *both*-regime models were trained on the same data as each other, so this difference is attributable to contrast conditioning exploiting cross-contrast information rather than the larger training set, equally available to both. On T1w the gains were small for both (+0.4% and +0.3%), confining the benefit to T1ce.

5 Conclusion

We presented SpFiLM, a conditioning layer that adjusts a spatially varying feature modulation to image contrast, with channel-wise FiLM recovered as the K=0 special case. Ablation studies attribute SpFiLM's improvement to its contrast-conditioned spatial modulation, not to parameter count, spatial modulation without conditioning, or the choice of contrast encoding. SpFiLM surpassed nnU-Net and SwinUNETR baselines, and improved nnU-Net further when used in the nnU-Net backbone, indicating the mechanism is backbone-agnostic. A single network trained jointly on T1w and T1ce MRI, with no per-contrast models, was the best performing model on both contrasts and, unlike the unconditioned baseline, converted the pooled cross-contrast data into a gain on the clinically-acquired T1ce.

However, our study has limitations. The cohort is from a single centre and scanner, limiting generalisability. The reference labels are atlas-generated rather

than manually annotated, so the atlas accuracy bounds the model and our findings. Because gadolinium is not given to healthy volunteers, paired pre- and post-contrast scans are available only from patients with a clinical indication, we therefore excluded patients with anatomy-distorting pathology (Section 3.1) so that contrast-handling errors are not confounded with pathology-induced label errors. Evaluating SpFiLM on pathological cohorts and validating it across institutions and scanners are natural next steps, as is a direct comparison against an unfactorised SpFiLM, which we defer to a journal extension. Finally, the method is not specific to brain parcellation, any segmentation task where contrast-agent uptake produces spatially varying intensity could benefit from SpFiLM.

Acknowledgments. This work was funded by the Engineering and Physical Sciences Research Council [Grant Number EP/Y035364/1] and a Hinduja PhD Scholarship and the National Institute for Health and Care Research (NIHR) Invention for Innovation Programme (NIHR208398). The views expressed are those of the author(s) and not necessarily those of the NIHR or the Department of Health and Social Care.

Disclosure of Interests. TV is co-founder and shareholder of Hypervision Surgical who have no interest in this work.

References

1. Bermudez, C., Remedios, S.W., Ramadass, K., McHugo, M., Heckers, S., Huo, Y., Landman, B.A.: Generalizing deep whole-brain segmentation for post-contrast MRI with transfer learning. Journal of Medical Imaging **7**(06) (Dec 2020). https://doi.org/10.1117/1.jmi.7.6.064004
2. Bibault, J.E., Giraud, P.: Deep learning for automated segmentation in radiotherapy: a narrative review. British Journal of Radiology **97**(1153), 13–20 (Dec 2023). https://doi.org/10.1093/bjr/tqad018
3. Billot, B., Greve, D.N., Puonti, O., Thielscher, A., Van Leemput, K., Fischl, B., Dalca, A.V., Iglesias, J.E.: 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
4. Cardoso, M.J., Modat, M., Wolz, R., Melbourne, A., Cash, D., Rueckert, D., Ourselin, S.: Geodesic information flows: Spatially-variant graphs and their application to segmentation and fusion. IEEE Transactions on Medical Imaging **34**(9), 1976–1988 (Sept 2015). https://doi.org/10.1109/tmi.2015.2418298
5. Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O.: 3D u-net: Learning dense volumetric segmentation from sparse annotation. In: Medical Image Computing and Computer-Assisted Intervention – MICCAI 2016, pp. 424–432. Lecture Notes in Computer Science, Springer International Publishing, Cham (2016)
6. Fischl, B.: FreeSurfer. NeuroImage **62**(2), 774–781 (Aug 2012). https://doi.org/10.1016/j.neuroimage.2012.01.021
7. Guha Roy, A., Conjeti, S., Navab, N., Wachinger, C., Alzheimer's Disease Neuroimaging Initiative: QuickNAT: A fully convolutional network for quick and accurate segmentation of neuroanatomy. Neuroimage **186**, 713–727 (Feb 2019)

8. Hatamizadeh, A., Nath, V., Tang, Y., Yang, D., Roth, H., Xu, D.: Swin unetr: Swin transformers for semantic segmentation of brain tumors in mri images. arXiv preprint arXiv:2201.01266 (2022)
9. Henschel, L., Conjeti, S., Estrada, S., Diers, K., Fischl, B., Reuter, M.: FastSurfer - a fast and accurate deep learning based neuroimaging pipeline. NeuroImage **219**, 117012 (Oct 2020). https://doi.org/10.1016/j.neuroimage.2020.117012
10. Isensee, F., Schell, M., Pflueger, I., Brugnara, G., Bonekamp, D., Neuberger, U., Wick, A., Schlemmer, H.P., Heiland, S., Wick, W., Bendszus, M., Maier-Hein, K.H., Kickingereder, P.: Automated brain extraction of multisequence MRI using artificial neural networks. Hum. Brain Mapp. **40**(17), 4952–4964 (Dec 2019)
11. Isensee, F., Wald, T., Ulrich, C., Baumgartner, M., Roy, S., Maier-Hein, K., Jäger, P.F.: NnU-net revisited: A call for rigorous validation in 3D medical image segmentation. In: Lecture Notes in Computer Science, pp. 488–498. Lecture Notes in Computer Science, Springer Nature Switzerland, Cham (2024)
12. Jenkinson, M., Beckmann, C.F., Behrens, T.E., Woolrich, M.W., Smith, S.M.: FSL. NeuroImage **62**(2), 782–790 (Aug 2012). https://doi.org/10.1016/j.neuroimage.2011.09.015
13. Klein, A., Tourville, J.: 101 labeled brain images and a consistent human cortical labeling protocol. Frontiers in Neuroscience **6** (2012). https://doi.org/10.3389/fnins.2012.00171
14. Lemay, A., Gros, C., Vincent, O., Liu, Y., Cohen, J.P., Cohen-Adad, J.: Benefits of linear conditioning for segmentation using metadata. In: Heinrich, M., Dou, Q., de Bruijne, M., Lellmann, J., Schläfer, A., Ernst, F. (eds.) Proceedings of the Fourth Conference on Medical Imaging with Deep Learning. Proceedings of Machine Learning Research, vol. 143, pp. 416–430. PMLR (07–09 Jul 2021), https://proceedings.mlr.press/v143/lemay21a.html
15. Lie, I.A., Kerklingh, E., Wesnes, K., van Nederpelt, D.R., Brouwer, I., Torkildsen, Ø., Myhr, K.M., Barkhof, F., Bø, L., Vrenken, H.: The effect of gadolinium-based contrast-agents on automated brain atrophy measurements by FreeSurfer in patients with multiple sclerosis. European Radiology **32**(5), 3576–3587 (Jan 2022). https://doi.org/10.1007/s00330-021-08405-8
16. Marek, J., Bachurska, D., Wolak, T., Borowiec, A., Sajdek, M., Maj, E.: Quantitative brain volumetry in neurological disorders: from disease mechanisms to software solutions. Polish Journal of Radiology **90**, 299–306 (June 2025). https://doi.org/10.5114/pjr/203781
17. Modat, M., Ridgway, G.R., Taylor, Z.A., Lehmann, M., Barnes, J., Hawkes, D.J., Fox, N.C., Ourselin, S.: Fast free-form deformation using graphics processing units. Computer Methods and Programs in Biomedicine **98**(3), 278–284 (June 2010). https://doi.org/10.1016/j.cmpb.2009.09.002
18. Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., Glocker, B., Rueckert, D.: Attention u-net: Learning where to look for the pancreas. In: Medical Imaging with Deep Learning (2018), https://openreview.net/forum?id=Skft7cijM
19. Orbes-Arteaga, M., Varsavsky, T., Sudre, C.H., Eaton-Rosen, Z., Haddow, L.J., Sørensen, L., Nielsen, M., Pai, A., Ourselin, S., Modat, M., Nachev, P., Cardoso, M.J.: Multi-domain Adaptation in Brain MRI Through Paired Consistency and Adversarial Learning, pp. 54–62. Springer International Publishing (2019). https://doi.org/10.1007/978-3-030-33391-1_7
20. Perez, E., Strub, F., De Vries, H., Dumoulin, V., Courville, A.: FiLM: Visual reasoning with a general conditioning layer. Proc. Conf. AAAI Artif. Intell. **32**(1) (Apr 2018)

21. Rebsamen, M., McKinley, R., Radojewski, P., Pistor, M., Friedli, C., Hoepner, R., Salmen, A., Chan, A., Reyes, M., Wagner, F., Wiest, R., Rummel, C.: Reliable brain morphometry from contrast-enhanced T1w-MRI in patients with multiple sclerosis. Human Brain Mapping **44**(3), 970–979 (Oct 2022). https://doi.org/10.1002/hbm.26117
22. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation, pp. 234–241. Springer International Publishing (2015). https://doi.org/10.1007/978-3-319-24574-4_28
23. Smirniotopoulos, J.G., Murphy, F.M., Rushing, E.J., Rees, J.H., Schroeder, J.W.: Patterns of contrast enhancement in the brain and meninges. RadioGraphics **27**(2), 525–551 (Mar 2007). https://doi.org/10.1148/rg.272065155
24. Tustison, N.J., Cook, P.A., Klein, A., Song, G., Das, S.R., Duda, J.T., Kandel, B.M., van Strien, N., Stone, J.R., Gee, J.C., Avants, B.B.: Large-scale evaluation of ANTs and FreeSurfer cortical thickness measurements. NeuroImage **99**, 166–179 (Oct 2014). https://doi.org/10.1016/j.neuroimage.2014.05.044
25. Ulyanov, D., Vedaldi, A., Lempitsky, V.: Instance normalization: The missing ingredient for fast stylization (2016). https://doi.org/10.48550/ARXIV.1607.08022
26. Vakharia, V.N., Sparks, R., Miserocchi, A., Vos, S.B., O'Keeffe, A., Rodionov, R., McEvoy, A.W., Ourselin, S., Duncan, J.S.: Computer-assisted planning for stereoelectroencephalography (SEEG). Neurotherapeutics **16**(4), 1183–1197 (Oct 2019)
27. Zhu, S., Chen, Y., Chen, W., Jiang, S., Zhou, G., Wang, Y., Qin, F., Wang, C., Tian, Q.: No modality left behind: Adapting to missing modalities via knowledge distillation for brain tumor segmentation. Medical Image Analysis **112**, 104108 (2026). https://doi.org/https://doi.org/10.1016/j.media.2026.104108