

CONTRAST-ENHANCED BRAIN MRI SYNTHESIS WITH DEEP LEARNING: KEY INPUT MODALITIES AND ASYMPTOTIC PERFORMANCE

Alexandre Bône^{1,*} Samy Ammari^{2,3,*} Jean-Philippe Lamarque^{2,3} Mickael Elhaik^{2,3} Émilie Chouzenoux⁴
François Nicolas¹ Philippe Robert¹ Corinne Balleyguier^{2,3} Nathalie Lassau^{2,3,†} Marc-Michel Rohé^{1,†}

¹ Guerbet Research, France

² Department of Radiology, Gustave Roussy Cancer Campus, Université Paris-Saclay, Villejuif, France

³ BioMaps (UMR1281), Université Paris-Saclay, CNRS, INSERM, CEA, Orsay, France

⁴ Center for Visual Computing, CentraleSupélec, Inria, Université Paris-Saclay, Gif-sur-Yvette, France

(*† Equal contributions)

ABSTRACT

Contrast-enhanced medical images offer vital insights for the accurate diagnosis, characterization and treatment of tumors, and are routinely used worldwide. Acquiring such images requires to inject the patient intravenously with a gadolinium-based contrast agent (GBCA). Although GBCAs are considered safe, recent concerns about their accumulation in the body tilted the medical consensus towards a more parsimonious usage. Focusing on the case of brain magnetic resonance imaging, this paper proposes a deep learning method that synthesizes virtual contrast-enhanced T1 images as if they had been acquired after the injection of a standard 0.100 mmol/kg dose of GBCA, taking as inputs complementary imaging modalities obtained either after a reduced injection at 0.025 mmol/kg or without any GBCA involved. The method achieves a competitive structural similarity index of 94.2%. Its asymptotic performance is estimated, and the most important input modalities are identified.

Index Terms— Brain MRI, gadolinium-based contrast agents (GBCA), low-dose imaging, virtual enhancement.

1. INTRODUCTION

Gadolinium and biomedical imaging. Gadolinium-based contrast agents (GBCA) are routinely used in biomedical imaging. An approximate 40% of all magnetic resonance imaging (MRI) sessions in Europe and the United States rely on GBCAs to diagnose, characterize or monitor lesions that would remain otherwise poorly visible [1, 2]. In neuroimaging, contrast-enhanced T1-weighted imaging is the cornerstone modality for the detection and precise delineation of brain glioma and metastases, which cause each year hundreds of thousands of deaths worldwide [3, 4]. GBCAs are considered safe, with less than one patient in 100,000 that experiences an adverse reaction [1]. However, the linear sub-category of GBCAs has been withdrawn from European

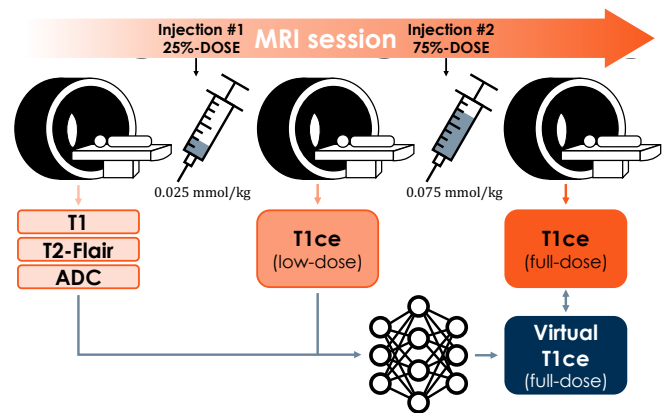


Fig. 1. Proposed deep learning approach for the synthesis of contrast-enhanced brain images.

markets in 2017 on the suspicion of abnormal accumulation of gadolinium [5]. Gadolinium has also been identified as a possible trigger of the rare nephrogenic systemic fibrosis disease in patients with renal insufficiency [6]. Macrocyclic GBCAs remain recommended if not systematic in a large number of situations in order to achieve the best diagnosis performance, but recent guidelines suggest to aim for a more parsimonious usage, especially in the case of chronic disease monitoring [7].

Deep learning approaches for low-dose MRI. In [8], the authors propose a deep learning method that predicts contrast-enhanced brain MRI slices from corresponding pairs of zero-contrast and low-dose (0.010 mmol/kg injection of GBCA) T1-weighted MRI slices, thus suggesting to decimate the usage of GBCAs in neuroimaging while maintaining a satisfying diagnostic performance. A year later, the study [9] goes even further and details a deep neural network architecture that predicts 3D patches of contrast-enhanced brain

MRIs from corresponding patches of multiple zero-contrast MRI sequences (including T1, T2, T2-Flair, diffusion and susceptibility-based contrasts), thus suggesting to completely eliminate the need for GBCAs.

Contributions. In this paper, we propose a deep learning method to synthesize brain contrast-enhanced MRIs from corresponding zero-contrast and low-dose modalities, the latter being obtained after an injection of GBCA at 0.025 mmol/kg. We share engineering tactics that allow the fast training of our architecture which directly processes entire brain images instead of slices or patches. We evaluate the method on a data set of 105 patients with mixed conditions. Finally, the relative contributions of the input modalities are systematically investigated with ablation studies, and the asymptotic performance of the method is estimated by training it on nested subsets of the total data set.

2. MATERIALS AND METHODS

Data and preprocessing. Imaging data from 150 MRI exams were collected at Gustave Roussy Hospital (Cancer Campus), France. The acquisition methodology is depicted by Figure 1: T1, T2-Flair and diffusion sequences are acquired before injecting 0.025 mmol/kg of GBCA and acquiring a contrast-enhanced T1 (T1ce), that we denote as “low-dose”. After a second injection of the remaining 75% of the standard 0.100 mmol/kg dose of GBCA, another T1-weighted sequence is acquired, and is called “full-dose” T1ce in the rest of the article. All sequences are corrected using the N4 bias field correction algorithm [10], co-registered on the MNI 152 brain atlas [11] using affine transformations, resampled to an isotropic 1mm resolution and cropped to volumes of

size $160 \times 192 \times 160$. Apparent diffusion coefficient (ADC) maps are computed from the diffusion sequences. Images are then normalized using the signal statistics of their respective brain content, identified using the HD-BET method [12]. After brain masking and extreme values clipping, all intensities are linearly mapped to the $[0, 1]$ interval.

Included patients have mixed conditions (including brain metastases, glioblastoma) and follow a mix of therapies (including surgical resection, biopsy, radiotherapy). MRI sequences were acquired on both 1.5T and 3T imaging systems. Out of 150, 25 sessions were incomplete and are therefore excluded. We choose in this work to use only 1 session per patient, further discarding 9 more MRI sessions. Finally, 11 more sessions with very large artifacts or whose preprocessing catastrophically failed were manually rejected, leading to a final heterogeneous data set of 105 sessions or patients.

Deep network architecture. Figure 2 details the proposed deep network architecture, adapted from [13], that synthesizes virtual contrast-enhanced brain MRIs from corresponding $1 \leq M \leq 4$ zero-dose and/or low-dose sequences stacked as M -channel volumes of size $160 \times 192 \times 160$. In addition to the reference configuration indicated by Figure 1 where $M = 4$ input modalities (T1, T2-Flair, ADC and low-dose T1ce) are exploited to train the deep network, additional combinations where $M < 4$ will be evaluated in Section 3.

In any case, input modalities are encoded into a hierarchy of feature maps by the successive application of 2-strided convolution filters with isotropic kernels of size 2, and residual convolution blocks of kernel size 3 without striding. The number of channels of the first feature map $C \in \{8, 16, 32\}$ is left as an hyper-parameter whose influence will be investigated in Section 3. A symmetrically-built decoder iteratively

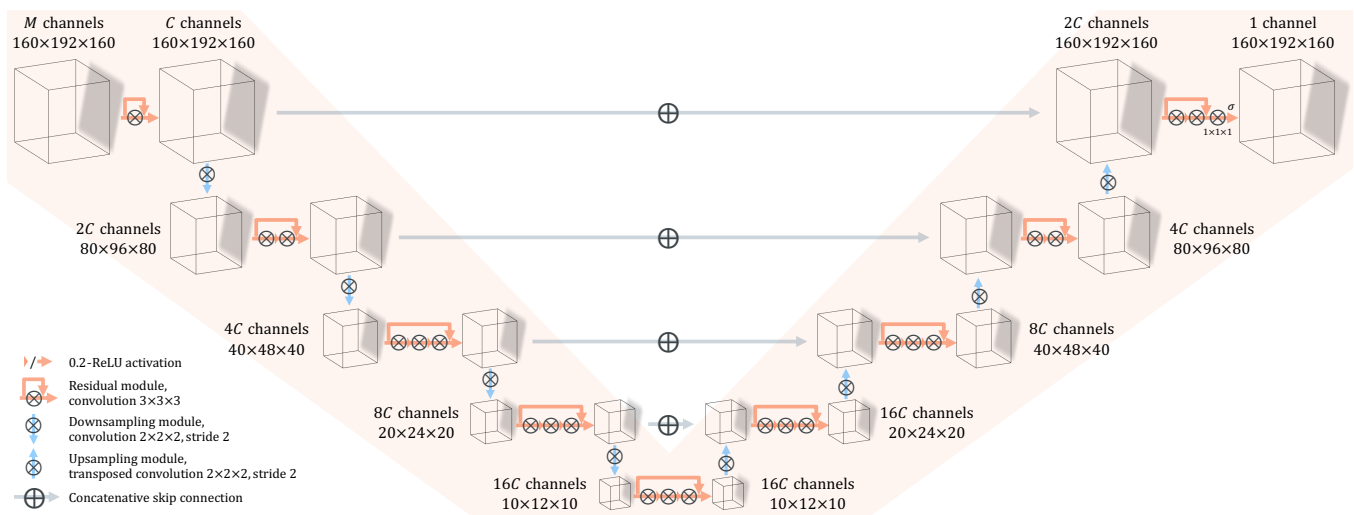


Fig. 2. Architecture of the proposed deep neural network for the synthesis of full-dose contrast-enhanced brain MRIs from zero-dose and/or low-dose sequences. Both the number of input modalities M and the network width C are left as hyperparameters.

evaluates 2-strided transposed convolution filters and residual convolution blocks of respective isotropic kernel sizes 2 and 3, up until a final convolution with kernel size 1 that finally synthesizes the virtual contrast-enhanced T1. Skip connections are taken into account by concatenation. All activation functions are 0.2-ReLU at the exception of the final sigmoid.

Training and testing the deep network. The deep network is trained to minimize the L2 loss between its predictions and the reference full-dose T1ce MRIs using the Adam optimizer with default hyper-parameters, for 300 epochs. Flip augmentation is used both during training with 50% probability for all three axes, and during validation and testing where all eight possible configurations are computed and averaged. In order to accelerate both training and inference, we leverage half-precision computing using the Nvidia-Apex library, and flip augmentations are performed directly on the GPU.

All performances are computed using the same 5-fold cross-validation approach, and are therefore directly comparable to each other. In some experiments, only a reduced fraction $F \in \{25\%, 50\%, 100\%\}$ of the 84 available training subjects in each fold are exploited, in order to evaluate the influence of the data set size and in turn estimate the asymptotic performance of the proposed method. Note that even in such cases, the test sets are kept identical. Four perfor-

mance metrics are considered: in addition to the “global” L2 loss which is directly optimized, we also compute the tumor-averaged L2 (t-L2) based on segmentation maps of enhancing lesions, as well as the peak signal-to-noise ratio (PSNR) and the structural similarity index (SSIM, see [14]) as in [8, 9].

3. RESULTS

Qualitative results for the reference configuration. We first evaluate our method using all available input modalities ($M = 4$ with T1, T2-Flair, ADC and low-dose T1ce), maximal training sets ($F = 100\%$), and intermediate network width ($C = 16$ channels). Average performance metrics are $31.6 (\times 10^{-5})$ for L2, $74.4 (\times 10^{-4})$ for t-L2, 35.3 dB for PSNR and 94.2% for SSIM, which compares favorably to the related works [8, 9], although such direct comparison of metrics on different data sets must be interpreted with care. Figure 3 presents representative examples of virtually-enhanced MRIs for three patients with different conditions. In the case of a patient with brain metastases (top row), we can see that the contrast is well-reconstituted for one of the two visible lesions (overlaid in green in the rightmost figure), when the shape of the second enhancing lesion (overlaid in red) is not perfectly captured. In the case of a patient with resected glioblastoma (middle row), we see that the enhancing border is well-located, al-

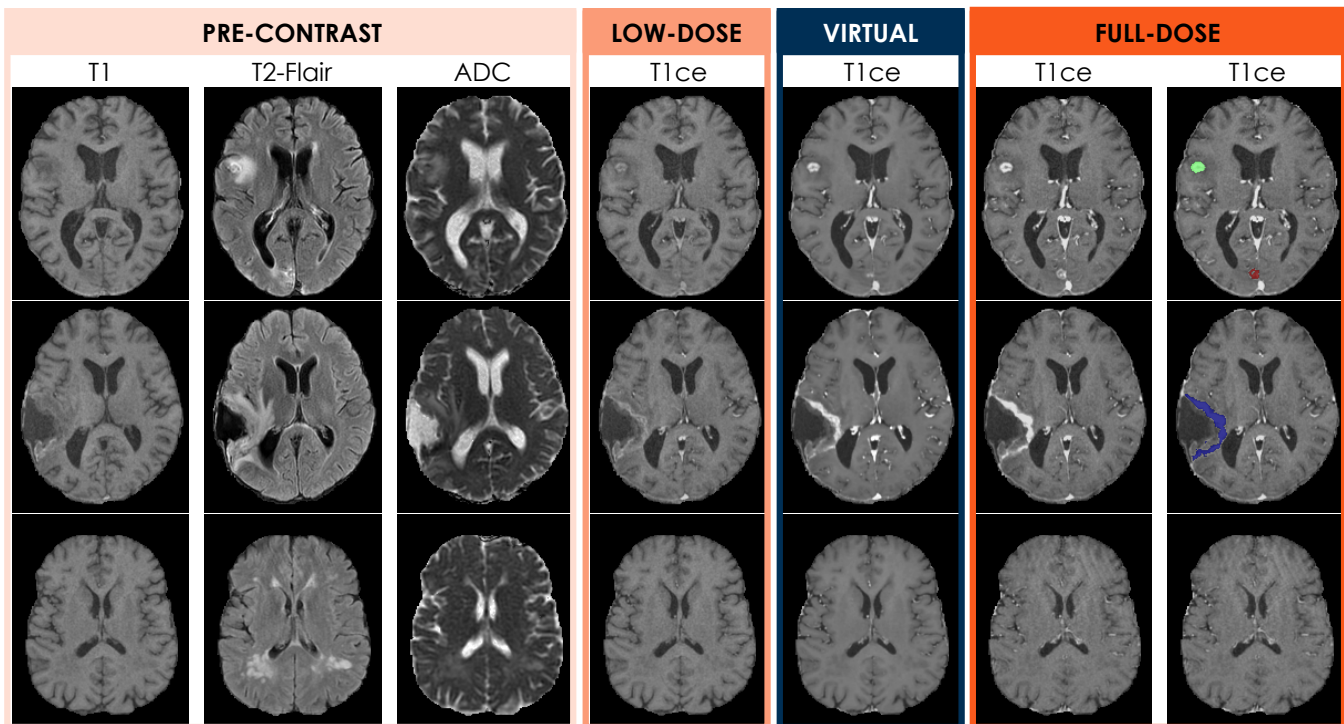


Fig. 3. Axial slices of T1, T2-Flair, ADC, low-dose T1ce, virtual T1ce and full-dose T1ce images (in columns, from left to right) for three patients with brain metastases (top row), resected glioblastoma (middle row), and no enhancing lesion (bottom row). The last column overlays manually-defined segmentation maps (in colors) of enhancing lesions on the full-dose T1ce.

T1ce _{low-dose}	T1	ADC	T2-Flair	L2 (10^{-5})	t-L2 (10^{-4})	PNSR (dB)	SSIM (%)
✓	✓	✓	✓	31.6 (11.4)	74.4 (80.2)	35.3 (1.52)	94.2 (1.31)
✗	✓	✓	✓	65.3 (30.9)	263. (185.)	32.2 (1.69)	91.6 (2.00)
✓	✗	✓	✓	34.0 (13.2)	129. (99.5)	35.0 (1.54)	94.0 (1.29)
✓	✓	✗	✓	32.2 (12.0)	84.1 (86.8)	35.2 (1.56)	94.1 (1.32)
✓	✓	✓	✗	32.0 (12.2)	79.9 (83.8)	35.2 (1.55)	94.2 (1.33)
✗	✓	✓	✗	66.9 (30.9)	264. (190.)	32.1 (1.68)	91.5 (2.01)
✓	✗	✓	✗	33.9 (13.0)	133. (102.)	35.0 (1.56)	94.1 (1.31)
✓	✓	✗	✗	31.3 (11.5)	87.1 (87.8)	35.3 (1.54)	94.2 (1.31)
✗	✓	✗	✗	65.7 (30.8)	260. (186.)	32.2 (1.71)	91.6 (1.99)
✓	✗	✗	✗	36.9 (14.3)	137. (106.)	34.6 (1.59)	93.8 (1.32)

Table 1. Means and standard deviations of the achieved L2, t-L2, PSNR and SSIM performance metrics in a sequence of scenarii where the number of input modalities M is iteratively decreased. For each configuration M , best metrics are in bold.

though slightly under-enhanced. Finally, in the case of a patient without any enhancing lesion, we can notably observe that the synthesized MRI presents a smoother and less artifacted aspect than the reference full-dose MRI, in line with [8].

Identification of key modalities. We systematically investigate the influence of input modalities by training and testing our deep learning method in varied scenarii: M is iteratively decreased according to a backward elimination procedure where at each step, the least influential modality is removed. Table 1 presents the obtained results: starting from the reference configuration previously discussed, the first modality that can be removed with the lesser impact on the performance is the T2-Flair; ADC then follows as the second less essential modality before T1. The low-dose T1ce is clearly identified as the key modality for our virtual enhancement method, whose performance systematically and significantly drops when deprived of contrast-enhanced inputs. We also interestingly observe that adding the T2-Flair and ADC modalities to the zero-contrast and low-dose T1 MRIs does not improve the global L2, PSNR and SSIM metrics, but improves the t-L2 performance on the tumor regions.

Estimation of the asymptotic performance. We finally evaluate how the performance of our method evolves when the size of the training set is artificially decreased to only a quarter ($F=25\%$), a half ($F=50\%$), or three-quarter ($F=75\%$) of its reference size ($F=100\%$). The influence of the deep network width, as defined by the number of channel dimensions of the first feature map $C \in \{8, 16, 32\}$, is jointly investigated. Figure 4 plots the evolution of the average performance metrics when these two hyperparameters are varied. The performance clearly drops when the network width of

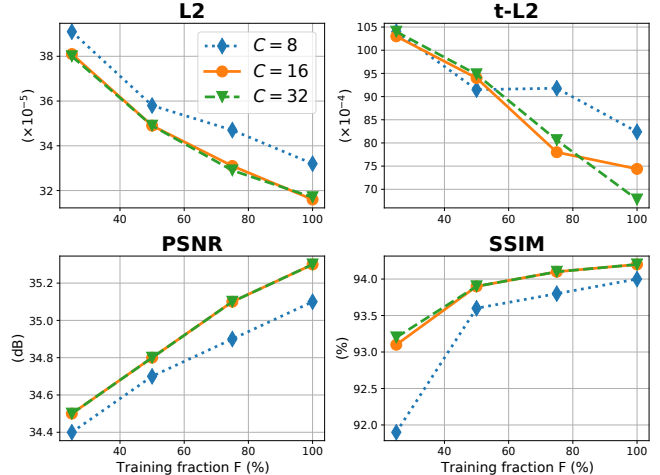


Fig. 4. Mean L2, t-L2, PSNR and SSIM performance metrics in an array of scenarii where the size of the training data set F and the deep network width C are jointly varied.

the reference configuration is decreased from $C=16$ to 8. On the other hand, doubling this number of channels to $C=32$ seems only beneficial for the t-L2 metric when all available data is exploited ($F=100\%$). We also read that the global L2, PSNR and SSIM performance metrics curves with respect to the training fraction F are approaching saturation: we may reasonably hypothesize that expanding further our data set would only yield sublinear improvements. In particular, we submit that the PSNR and SSIM metrics would respectively remain below $36.1dB$ and 94.6% if the data set size was doubled ($F=200\%$). By contrast, it is more difficult to anticipate the asymptotic performance of our method as measured by the t-L2 metric, which does not exhibit clear signs of saturation.

4. DISCUSSION AND CONCLUSION

In this paper, we detailed a deep learning method that synthesizes brain contrast-enhanced MRIs from complementary low-dose and zero-contrast modalities. The method was evaluated on a large data set of patients with mixed conditions, and showed a state-of-the-art performance with respect to the literature, both qualitatively and quantitatively. We systematically studied the influence of input modalities and clearly identified that the low-dose MRI is key, suggesting that the information provided by injecting a GBCA cannot be simply recovered by combining zero-contrast modalities. We finally investigated the influence of the data set size and the deep network width, which allowed the estimation of the asymptotic performance of our method. In order to evaluate the potential impact of our method in clinical practice, we will evaluate in future work if the synthesized images allow the correct identification of lesions by professional radiologists.

5. COMPLIANCE WITH ETHICAL STANDARDS

This research study was conducted retrospectively[‡] using human data acquired at the Gustave Roussy Hospital. Patient consent for research use was collected. Approval was granted by the Ethics Committee of the Gustave Roussy Hospital.

6. ACKNOWLEDGEMENTS

Study funded and conducted by Guerbet Research, in partnership with the Gustave Roussy Hospital.

7. REFERENCES

- [1] Val M Runge, “Safety of the gadolinium-based contrast agents for magnetic resonance imaging, focusing in part on their accumulation in the brain and especially the dentate nucleus,” *Investigative radiology*, vol. 51, no. 5, pp. 273–279, 2016.
- [2] N Tomura, K Narita, S Takahashi, T Otani, I Sakuma, K Yasuda, T Nishii, and J Watarai, “Contrast-enhanced multi-shot echo-planar flair in the depiction of metastatic tumors of the brain: comparison with contrast-enhanced spin-echo t1-weighted imaging,” *Acta Radiologica*, vol. 48, no. 9, pp. 1032–1037, 2007.
- [3] Edouard Alphandéry, “Glioblastoma treatments: an account of recent industrial developments,” *Frontiers in pharmacology*, vol. 9, pp. 879, 2018.
- [4] Achal Singh Achrol, Robert C Rennert, Carey Anders, Riccardo Soffietti, Manmeet S Ahluwalia, Lakshmi Nayak, Solange Peters, Nils D Arvold, Griffith R Harsh, Patricia S Steeg, et al., “Brain metastases,” *Nature Reviews Disease Primers*, vol. 5, no. 1, pp. 1–26, 2019.
- [5] European Medicines Agency, “Ema’s final opinion confirms restrictions on use of linear gadolinium agents in body scans,” www.ema.europa.eu, [Accessed: 2020-09].
- [6] Mark A Perazella, “Gadolinium-contrast toxicity in patients with kidney disease: nephrotoxicity and nephrogenic systemic fibrosis,” *Current drug safety*, vol. 3, no. 1, pp. 67–75, 2008.
- [7] François Lersy, Gregoire Boulouis, Olivier Clément, Hubert Desal, René Anxionnat, Jérôme Berge, Claire Boutet, Apolline Kazémi, Nadya Pyatigorskaya, Augustin Lecler, et al., “Consensus guidelines of the french society of neuroradiology (sfnr) on the use of gadolinium-based contrast agents (gbcas) and related mri protocols in neuroradiology,” *Journal of Neuroradiology*, 2020.
- [8] Enhao Gong, John M Pauly, Max Wintermark, and Greg Zaharchuk, “Deep learning enables reduced gadolinium dose for contrast-enhanced brain mri,” *Journal of magnetic resonance imaging*, vol. 48, no. 2, pp. 330–340, 2018.
- [9] Jens Kleesiek, Jan Nikolas Morshuis, Fabian Isensee, Katerina Deike-Hofmann, Daniel Paech, Philipp Kickingeder, Ullrich Köthe, Carsten Rother, Michael Forsting, Wolfgang Wick, et al., “Can virtual contrast enhancement in brain mri replace gadolinium?: a feasibility study,” *Investigative radiology*, vol. 54, no. 10, pp. 653–660, 2019.
- [10] Nicholas J Tustison, Brian B Avants, Philip A Cook, Yuanjie Zheng, Alexander Egan, Paul A Yushkevich, and James C Gee, “N4itk: improved n3 bias correction,” *IEEE transactions on medical imaging*, vol. 29, no. 6, pp. 1310–1320, 2010.
- [11] Vladimir S Fonov, Alan C Evans, Robert C McKinstry, CR Almlri, and DL Collins, “Unbiased nonlinear average age-appropriate brain templates from birth to adulthood,” *NeuroImage*, , no. 47, pp. S102, 2009.
- [12] Fabian Isensee, Marianne Schell, Irada Pflueger, Gianluca Brugnara, David Bonekamp, Ulf Neuberger, Antje Wick, Heinz-Peter Schlemmer, Sabine Heiland, Wolfgang Wick, et al., “Automated brain extraction of multisequence mri using artificial neural networks,” *Human brain mapping*, vol. 40, no. 17, pp. 4952–4964, 2019.
- [13] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi, “V-net: Fully convolutional neural networks for volumetric medical image segmentation,” in *2016 fourth international conference on 3D vision (3DV)*. IEEE, 2016, pp. 565–571.
- [14] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli, “Image quality assessment: from error visibility to structural similarity,” *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.
- [15] SC Thust, S Heiland, A Falini, HR Jäger, AD Waldman, PC Sundgren, Claudia Godi, VK Katsaros, A Ramos, N Bargallo, et al., “Glioma imaging in europe: a survey of 220 centres and recommendations for best clinical practice,” *European radiology*, vol. 28, no. 8, pp. 3306–3317, 2018.

[‡]The two-injection protocol was already in place, the first injection playing the role of preload bolus for DSC-perfusion MRI. See [15] for instance.