



# Going beyond voxel-wise deconvolution in perfusion MRI: learning and leveraging spatio-temporal regularities with the stU-Net

Julien Veron Vialard, Marc-Michel Rohé, Philippe Robert, François Nicolas,  
Alexandre Bône

## ► To cite this version:

Julien Veron Vialard, Marc-Michel Rohé, Philippe Robert, François Nicolas, Alexandre Bône. Going beyond voxel-wise deconvolution in perfusion MRI: learning and leveraging spatio-temporal regularities with the stU-Net. 2021. hal-03128610

**HAL Id: hal-03128610**

**<https://hal.science/hal-03128610>**

Preprint submitted on 2 Feb 2021

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# GOING BEYOND VOXEL-WISE DECONVOLUTION IN PERFUSION MRI: LEARNING AND LEVERAGING SPATIO-TEMPORAL REGULARITIES WITH THE stU-NET

Julien Veron Vialard<sup>1,2,3</sup> Marc-Michel Rohé<sup>1</sup> Philippe Robert<sup>1</sup> François Nicolas<sup>1</sup> Alexandre Bône<sup>1</sup>

<sup>1</sup> Guerbet Research, France

<sup>2</sup> CentraleSupélec, France

<sup>3</sup> Stanford University, United States of America

## ABSTRACT

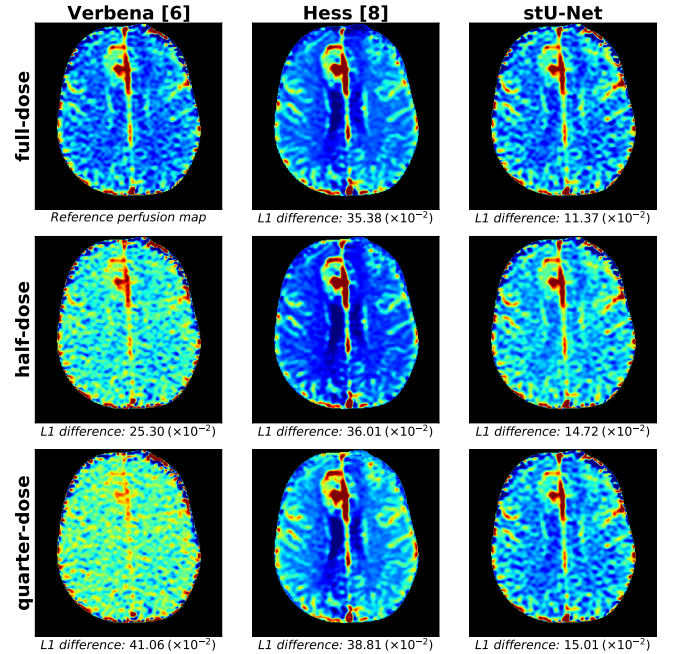
We present a deep learning method able to accurately predict quantitative perfusion maps from raw perfusion sequences, even in low-dose regimes of gadolinium-based contrast agent. The proposed stU-Net architecture is composed of a 2-dimensional decoder and a 3-dimensional encoder able to exploit spatio-temporal regularities by learning convolutional filters that jointly act on the spatial and temporal domains. The method is evaluated on a public data set containing 49 patients with brain tumors, using an original approach to simulate low-dose perfusion sequences.

**Index Terms**— Perfusion imaging, glioma, gadolinium, low-dose, deep learning, simulation.

## 1. INTRODUCTION

*From raw perfusion sequences to quantitative maps.* Perfusion MRI is an advanced biomedical imaging technique that allows to visualize and quantify the blood supply of an organ, such as the brain or the heart. The surveys [1] and [2] show that perfusion MRI is widely used in clinical practice, notably in neuroimaging for the initial diagnosis and treatment planning of stroke and glioma. Dynamic susceptibility contrast (DSC) is the most widespread perfusion technique: a gadolinium-based contrast agent (GBCA) is injected intravenously to the patient and rapid repeated T2\* imaging is performed to obtain a temporal sequence of 3D images. Quantitative maps of relevant parameters, such as the cerebral blood volume (CBV) or the cerebral blood flow (CBF), are computed from these raw perfusion sequences before analysis by the radiologist [3, 4]. Different perfusion post-processing software solutions are freely or commercially available [4, 5, 6]. They rely on microvascular models that are fitted in a voxel-wise fashion. The estimation of quantitative perfusion maps is classically framed as a deconvolution problem, requiring the preliminary estimation of the so-called arterial input function (AIF) by manually or semi-automatically delineating a large artery in the raw perfusion sequence [4].

*Limits of classical deconvolution approaches.* The robust and



**Fig. 1.** Cerebral blood volume (CBV) maps computed from full, half and quarter-dose DSC-perfusion sequences with either the classical deconvolution approach of Verbena [6], the deep learning method proposed in [8], and our stU-Net.

repeatable estimation of perfusion quantitative maps is notoriously difficult for three main reasons [2, 4]: (i) raw sequences typically exhibit low signal-to-noise ratios (SNR), (ii) deconvolution approaches are mathematically sensitive to noise, and (iii) manual inputs are required from the user. In addition, recent clinical guidelines advocate for more parsimonious GBCA injections [7]. As reducing the injected volumes below the standard 0.1 mmol/kg dose would further degrade the SNR of raw perfusion sequences [4], the practical feasibility of low-dose perfusion remains unclear today.

*Deep learning approaches to perfusion.* In [8] and [9], the authors propose deep learning methods able to reproduce

reference quantitative maps directly from the corresponding raw perfusion sequences, in a fully-automatic fashion. Such model-free, purely data-driven approaches notably circumvent the classical deconvolution approach, and are shown to be more robust to noise. One fundamental strength of these approaches is the ability for the neural network to learn by itself powerful regularizing priors. For instance, in [8] the “spatial correlation” module learns adapted 2-dimensional (2D) convolution filters that enforce some level of spatial smoothness in the produced quantitative perfusion maps, therefore enhancing the robustness of the method when exposed to noisy data. Similarly, [9] resort to 2D convolutional neural network (CNN) architectures able to learn and leverage the natural spatial regularity of imaging data. By contrast, these methods differ in their management of the temporal dimension of raw perfusion sequences: [8] resort to a “sequence encoder” module based on 1D convolution filters, when [9] treats each frame of the perfusion sequence as an independent color channel.

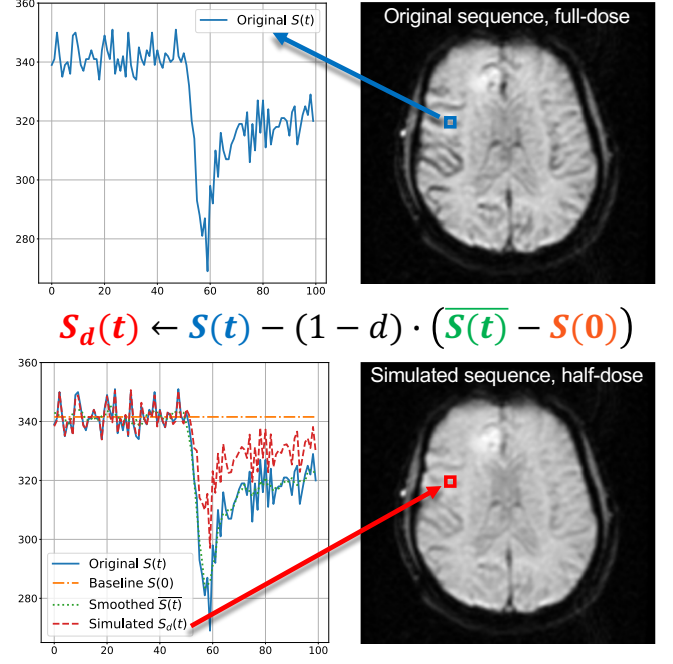
**Contributions.** In this paper, we introduce the spatio-temporal U-Net (stU-Net) architecture, designed to predict quantitative maps from raw perfusion sequences. This CNN notably differs from previous works in its original combination of a 2D decoder with a 3D encoder able to exploit the intrinsic spatio-temporal regularities of perfusion data by learning convolution filters that jointly act on the spatial and temporal domains. This core “(2+1)D” computational unit generalizes the architecture of [8], to which the performance of our method will be compared, on a public data set of 49 patients with brain tumours. We also propose a simple and lightweight method to simulate low-dose acquisitions by artificially degrading reference perfusion sequences, allowing the evaluation of the stU-Net on sequences acquired with reduced amount of contrast agent.

## 2. SIMULATING LOW-DOSE PERFUSION

According to the classical DSC-perfusion theory, the temporal signal  $S(t)$  in each voxel of the raw sequence varies linearly with the concentration  $C(t)$  of GBCA this voxel contains at any time  $t$  [4]. In practice, only a noisy version of the true signal can be observed at a discrete set of regularly spaced time points  $t \in t_0, t_1, \dots, t_T$ , typically every second:

$$S(t) = S(0) + \kappa \cdot C(t) + \epsilon(t) \quad \text{where} \quad \epsilon(t) \stackrel{\text{iid}}{\sim} \mathcal{N}(0, \sigma_\epsilon^2).$$

Assuming that the temporal dynamics of  $C(t)$  are slow with respect to the temporal sampling interval  $\Delta t = t_{j+1} - t_j$ , we can write  $\overline{C(t)} \approx C(t)$  where the operator  $\overline{(\cdot)}$  denotes a local average. Note that this hypothesis is reasonable since it actually corresponds to a desired behavior when the operator selects the sampling interval  $\Delta t$ . Since the noise distribution is assumed independent of time, we can finally write  $\overline{\epsilon(t)} \approx 0$  and therefore  $\epsilon(t) \approx S(t) - \overline{S(t)}$ .



**Fig. 2.** Artificial degradation method for the simulation of low-dose DSC perfusion sequences. Based on an original perfusion sequence (top right image), a synthetic 50%-dose sequence is simulated (bottom right image).

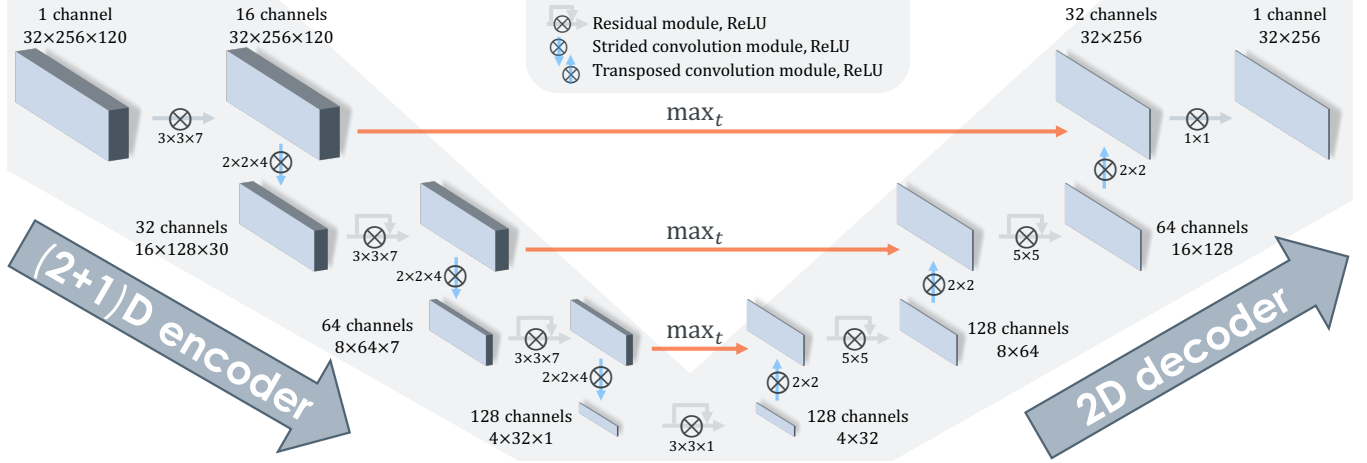
Introducing now the artificial dose reduction factor  $0\% \leq d \leq 100\%$ , we argue that the signal  $S_d$  that would have been observed instead of the reference signal  $S$  if the injected GBCA dose had been reduced by a factor  $d$ , writes:

$$\begin{aligned} S_d(t) &= S(0) + \kappa \cdot C(t) \times d + \epsilon(t) \\ &\approx S(t) - (1 - d) \cdot [\overline{S(t)} - S(0)]. \end{aligned} \quad (1)$$

This synthetic signal  $S_d$  can be easily computed for any reference signal  $S$  or artificial dose reduction factor  $d$ . Figure 2 displays an example of the proposed approach.

## 3. THE stU-NET ARCHITECTURE

Figure 3 summarizes the proposed stU-Net architecture, designed to predict perfusion quantitative maps from raw perfusion sequences. Inspired by the classical 2D U-Net architecture for biomedical segmentation [10], the stU-Net is a hybrid 2/3D CNN that performs 3D convolutions with or without striding in its encoding part, and alternates between 2D transposed convolutions and regular 2D convolution in its decoding part. The encoder is fed with temporal sequences of 2D “strips”, i.e. rectangular 2D patches, extracted from the raw perfusion data. The successive frames are stacked as a third dimension, homologous to the two spatial ones, allowing the network to naturally perform spatio-temporal convo-



**Fig. 3.** Architecture of the stU-Net, designed to predict perfusion quantitative maps from the corresponding raw sequences.

lutions, using anisotropic kernels of size  $3 \times 3 \times 7$  with stride 1 or size  $2 \times 2 \times 4$  with stride 4. The decoder predicts 2D quantitative map strips by chaining purely spatial convolution-like kernels of size  $5 \times 5$  with stride 1 or size  $2 \times 2$  with stride 2. To ensure the dimensional consistency, the skip connections perform a max-pooling reduction in the temporal dimension of the intermediate (2+1)D feature maps computed by encoder.

The stU-Net is trained for 200 epochs by the Adam optimizer with default hyper-parameters to minimize the L2 loss between its predictions and target perfusion maps, pre-computed from the raw sequences using a third-party software of reference. During training, data strips are sampled regularly from the original images with a stride of 8. At test time, strips of raw perfusion sequences are sampled with a stride of 1 and all the predictions are assembled by averaging.

## 4. EXPERIMENTS AND RESULTS

### 4.1. Validation of the low-dose simulation method

We validate the artificial degradation method proposed in Section 2 using DSC-perfusion data acquired from a rat. The animal is injected in a first session with a standard 0.10 mmol/kg dose of GBCA, and in a second session the following day with a reduced dose at 0.05 mmol/kg. The full-dose perfusion sequence is artificially degraded into a simulated half-dose sequence, as devised by Equation 1 with  $d = 0.5$ . The average brain signals are then extracted using manual segmentation, and visually compared to each other. We can finally verify on Figure 4 the good agreement between the simulated and target half-dose signals, as theoretically expected.

### 4.2. Evaluation of the tUNet predictive performance

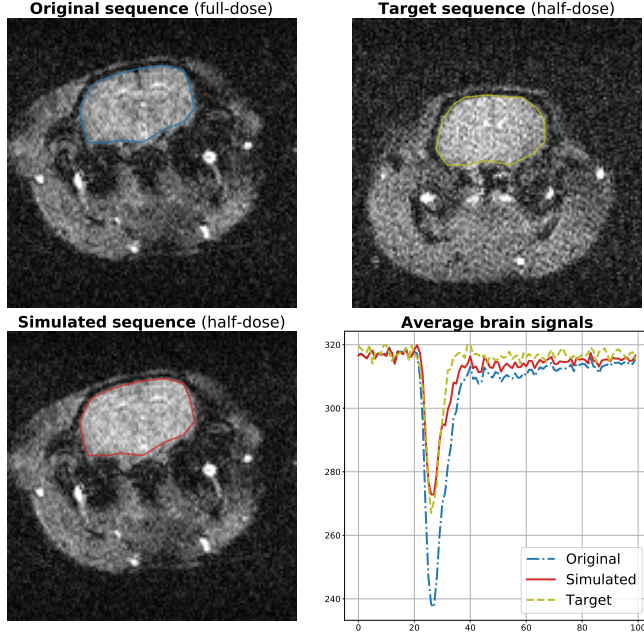
*Data and pre-processing.* We evaluate the predictive performance of the stU-Net on a public data set containing 49 pa-

tients with brain tumors [11, 12], that we randomly split into training, validation and test subsets respectively composed of 25, 12, and 12 subjects. Raw brain DSC-perfusion sequences are readily available for all subjects, along with individual segmentation maps large vessels from which we estimate the AIF. We generate reference quantitative CBV maps using the SVD pipeline of Verbena v4.0, freely distributed as part of the FSL v6.0 software suite [6, 13, 14]. Corresponding pairs of raw DSC sequences and CBV maps are individually cropped on square prisms tightly fitted on the brain voxels, and each slice is then resampled to a  $256 \times 256$  size. The DSC and CBV signals are then standardized using the brain content statistics, and clipped to the  $[-5, 5]$  interval in order to remove extreme values. All DSC frames are normalized similarly, according to the signal statistics of the first five frames only. In addition to this reference “full-dose” data set, we synthesize corresponding “half-dose” ( $d = 50\%$ ) and “quarter-dose” ( $d = 25\%$ ) perfusion sequences, following the simulation method presented in Section 2. The same previously-described preprocessing steps are similarly applied.

*Results.* The performance of the proposed stU-Net method is evaluated in three scenarios: reproducing the reference Verbena-computed CBV maps from the original perfusion sequences, from the synthetic half-dose sequences, and from the synthetic quarter-dose sequences. This performance is compared to the state-of-the-art method of Hess et al. [8], and to the SVD pipeline of Verbena for the two low-dose scenarios.

Figure 1 qualitatively plots the obtained results, for an arbitrarily-picked test case. The CBV map computed by Verbena [6] from the original perfusion sequence (top-left figure) plays the role of ground truth, to which all others maps should be compared. We see that the Verbena maps are significantly degraded in the lower-dose scenarios, and especially in the quarter-dose case where the tumour region is catastroph-





**Fig. 4.** Representative snapshots from DSC-perfusion sequences acquired from a rat (top row figures) or simulated by artificial degradation (bottom-left figure). Brains are manually delineated (blue, yellow and red line), and their respective average signals are plotted on the bottom-right figure.

ically misestimated. By contrast, the two deep learning approaches appear to produce more consistent results across the different doses regimes. The method of Hess et al. [8] seems however unable to reconstruct the small-scale details of the ground truth and to over-estimate the CBV values in the tumour region. On the other hand, the stU-Net predicts the most resembling CBV maps, as quantitatively confirmed by the L1 metric values (averaged across all slices).

Table 1 reports quantitative results, in terms of the L1, tumor-averaged L1 (t-L1), L2, tumor-averaged L2 (t-L2) and structural similarity (SSIM, see [15]) performance metrics. We first note that in the full-dose scenario, the method of Hess obtains a performance of  $34.01 \times 10^{-2}$  for the L1 metric i.e. approximatively 3% of the total signal variation range [-5, 5], in line with their published results on a different data set of stroke patients (clipped-L1 performance of 0.524 versus a signal range of [0, 20], i.e. 3% as well). However small, this discrepancy remains superior to the L1 performance of  $22.69 \times 10^{-2}$  obtained by the deconvolution-based Verbena software in the harder half-dose scenario, suggesting that the performance of Hess’ method is not optimal. In the quarter-dose regime, Verbena is although outperformed, as all its performance metric drop sharply (L1 metric of  $46.16 \times 10^{-2}$ ). Finally, we can read that in all scenarii and for all metrics, the stU-Net outperforms on average both the method of Hess and

			Verbena [6]	Hess [8]	stU-Net
full-dose	L2	( $10^{-2}$ )	(n/a)	<u>38.71</u> (15.06)	<b>8.38</b> (3.52)
	t-L2	( $10^{-2}$ )	(n/a)	<u>26.30</u> (67.29)	<b>5.93</b> (6.44)
	L1	( $10^{-2}$ )	(n/a)	<u>34.01</u> (5.49)	<b>17.58</b> (4.05)
	t-L1	( $10^{-2}$ )	(n/a)	<u>26.89</u> (21.51)	<b>16.52</b> (9.37)
	SSIM	(%)	(n/a)	<u>63.16</u> (5.71)	<b>82.08</b> (5.02)
$\frac{1}{2}$ -dose	L2	( $10^{-2}$ )	<u>15.97</u> (3.70)	39.00 (15.95)	<b>8.86</b> (2.98)
	t-L2	( $10^{-2}$ )	<u>11.24</u> (15.50)	28.09 (72.44)	<b>3.95</b> (4.17)
	L1	( $10^{-2}$ )	<u>22.69</u> (3.70)	33.87 (6.04)	<b>18.97</b> (3.76)
	t-L1	( $10^{-2}$ )	<u>20.62</u> (14.59)	27.31 (22.47)	<b>13.96</b> (7.07)
	SSIM	(%)	<u>77.90</u> (4.22)	63.67 (5.47)	<b>79.19</b> (5.17)
$\frac{1}{4}$ -dose	L2	( $10^{-2}$ )	70.90 (74.47)	<u>39.82</u> (15.86)	<b>9.68</b> (2.72)
	t-L2	( $10^{-2}$ )	38.67 (43.45)	<u>30.06</u> (70.35)	<b>6.26</b> (7.48)
	L1	( $10^{-2}$ )	46.16 (19.11)	<u>34.72</u> (6.09)	<b>18.22</b> (2.35)
	t-L1	( $10^{-2}$ )	41.66 (24.32)	<u>30.07</u> (22.32)	<b>17.53</b> (10.89)
	SSIM	(%)	59.08 (6.57)	<u>62.78</u> (5.37)	<b>81.36</b> (3.91)

**Table 1.** Means and standard deviations of the L2, t-L2, L1, t-L1, and SSIM performance metrics achieved in the full-dose, half-dose and quarter-dose scenarii by the SVD pipeline of the Verbena software [6], the state-of-the art method of Hess et al. [8], and our proposed stU-Net approach. Verbena generated the reference full-dose CBV maps, and therefore cannot be evaluated in this scenario (indicated as *n/a* in the Table). Metrics are only computed on the test cases. The best metric of each row is indicated in bold, the second best is underlined.

Verbena, achieving a minimal L1 discrepancy of  $17.58 \times 10^{-2}$  in the full-dose scenario.

## 5. DISCUSSION AND CONCLUSION

We proposed a deep learning method able to accurately predict quantitative perfusion maps from raw perfusion sequences, even in low-dose regimes of GBCAs. The spatio-temporal U-Net architecture, or stU-Net, combines a 2D decoder with a 3D encoder able to learn and leverage the spatio-temporal regularities of perfusion sequences, and showed superior performance to the current state-of-the-art on a public data set of patients with brain tumors. We also proposed a simulation approach to simulate synthetic low-dose perfusion sequences, and used it to benchmark the robustness profile of our approach as well as relevant related algorithms.

Basing our experiments on the publicly-available Verbena software to generate reference quantitative maps has the advantage of fostering reproducibility. However, this also represents a limit of our work, as Verbena is not a clinical standard. In future work, we may explore the opportunity of using multiple reference quantitative maps generated from a collection of software solutions when training our deep learning method.

## 6. COMPLIANCE WITH ETHICAL STANDARDS

This research study was conducted retrospectively using human data made available in open access by The Cancer Imaging Archive [11, 12], and animal data acquired in line with the principles of the Declaration of Helsinki. Ethical approval was not required in the first case, and was granted by the Ethics Committee of Guerbet Research in the second.

## 7. ACKNOWLEDGEMENTS

Study funded and conducted by Guerbet Research.

## 8. REFERENCES

- [1] Elliot Dickerson and Ashok Srinivasan, “Multicenter survey of current practice patterns in perfusion mri in neuroradiology: why, when, and how is it performed?,” *American Journal of Roentgenology*, vol. 207, no. 2, pp. 406–410, 2016.
- [2] 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.
- [3] Kirk Welker, J Boxerman, A Kalnin, T Kaufmann, M Shiroishi, MASFNR Wintermark, et al., “Asfnr recommendations for clinical performance of mr dynamic susceptibility contrast perfusion imaging of the brain,” *American Journal of Neuroradiology*, vol. 36, no. 6, pp. E41–E51, 2015.
- [4] Lisa Willats and Fernando Calamante, “The 39 steps: evading error and deciphering the secrets for accurate dynamic susceptibility contrast mri,” *NMR in Biomedicine*, vol. 26, no. 8, pp. 913–931, 2013.
- [5] Kohsuke Kudo, Ikuko Uwano, Toshinori Hirai, Ryuji Murakami, Hideo Nakamura, Noriyuki Fujima, Fumio Yamashita, Jonathan Goodwin, Satomi Higuchi, and Makoto Sasaki, “Comparison of different post-processing algorithms for dynamic susceptibility contrast perfusion imaging of cerebral gliomas,” *Magnetic Resonance in Medical Sciences*, vol. 16, no. 2, pp. 129, 2017.
- [6] Kim Mouridsen, Karl Friston, Niels Hjort, Louise Gyldensted, Leif Østergaard, and Stefan Kiebel, “Bayesian estimation of cerebral perfusion using a physiological model of microvasculature,” *Neuroimage*, vol. 33, no. 2, pp. 570–579, 2006.
- [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] Andreas Hess, Raphael Meier, Johannes Kaesmacher, Simon Jung, Fabien Scalzo, David Liebeskind, Roland Wiest, and Richard McKinley, “Synthetic perfusion maps: imaging perfusion deficits in dsc-mri with deep learning,” in *International MICCAI brainlesion workshop*. Springer, 2018, pp. 447–455.
- [9] Cagdas Ulas, Dhritiman Das, Michael J Thrippleton, Maria del C Valdes Hernandez, Paul A Armitage, Stephen D Makin, Joanna M Wardlaw, and Bjoern H Menze, “Convolutional neural networks for direct inference of pharmacokinetic parameters: Application to stroke dynamic contrast-enhanced mri,” *Frontiers in Neurology*, vol. 9, pp. 1147, 2019.
- [10] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, “U-net: Convolutional networks for biomedical image segmentation,” in *International Conference on Medical image computing and computer-assisted intervention*. Springer, 2015, pp. 234–241.
- [11] Kenneth Clark, Bruce Vendt, Kirk Smith, John Freymann, Justin Kirby, Paul Koppel, Stephen Moore, Stanley Phillips, David Maffitt, Michael Pringle, et al., “The cancer imaging archive (tcia): maintaining and operating a public information repository,” *Journal of digital imaging*, vol. 26, no. 6, pp. 1045–1057, 2013.
- [12] KM Schmainda, MA Prah, JM Connelly, and SD Rand, “Glioma dsc-mri perfusion data with standard imaging and rois,” *The Cancer Imaging Archive*. <http://doi.org/10.7937/K>, vol. 9, 2016.
- [13] Michael A Chappell, Amit Mehndiratta, and Fernando Calamante, “Correcting for large vessel contamination in dynamic susceptibility contrast perfusion mri by extension to a physiological model of the vasculature,” *Magnetic resonance in medicine*, vol. 74, no. 1, pp. 280–290, 2015.
- [14] Mark Jenkinson, Christian F Beckmann, Timothy EJ Behrens, Mark W Woolrich, and Stephen M Smith, “Fsl,” *Neuroimage*, vol. 62, no. 2, pp. 782–790, 2012.
- [15] 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.