Machine Learning Enabled Brain Segmentation for Small Animal Neuroimaging Registration

Hendrik J. Klug[1] Berkan Lafci[2,3] Markus Rudin[4] Daniel Razansky[2,3] Horea-Ioan Ioanas[5]

[1]ETHZ, Department of Information Technology and Electrical Engineering, Zurich, Switzerland

[2]UZH, Institute of Pharmacology and Toxicology and Institute for Biomedical Engineering, Faculty of Medicine, Zurich, Switzerland

[3]ETHZ, Institute for Biomedical Engineering, Department of Information Technology and Electrical Engineering, Zurich, Switzerland

[4]ETHZ, Center for Imaging Science and Technology, Zurich, Switzerland

[5]Massachusetts Institute of Technology, Department of Biological Engineering, Cambridge, United States of America

Abstract

In biomedical imaging, between-subject comparability is attained at the voxel level via image registration. This process requires several data preparation steps, of which brain extraction is particularly problematic in preclinical applications. Current solutions rely on human brain extraction library adaptations, or full image processing — with introduce artefacts via rostrocaudal cropping and peripheral hyperintensities, respectively. We present a deep learning framework for multi-contrast MRI brain tissue segmentaion, and benchmark its performance with respect to novel workflow advances.

Introduction

Functional magnetic resonance imaging (fMRI) gives an indirect measurement of brain activity by being sensitive to the change of blood flow. It is one of the most prominent neuroimaging tool for many applications, such as drug discovery (Borsook, Becerra and Hargreaves, 2006) and neuromodeling (Friston, Harrison and Penny, 2003). For fMRI studies, it is necessary that all scans lie in a standard reference frame in order to make meaningful comparisons across the subjects. The common coordinate system enables a statistical evaluation of the likelihood of consistent activation across a group or, in other contexts, the differences in anatomy between two groups. Because of variability both in animal anatomy and in animal preparation, the original MR acquired images are not defined in a common template space. To solve this issue, scans need to be remapped to a reference frame via registration (Maintz and Viergever, n.d.; Sotiras, Davatzikos and Paragios, 2013). As reported by Ioanas et al., 2019, the legacy approach for mouse-brain image registration is to modify the data in order to conform to pre-existing functions, designed and optimized for human brain imaging. This requires the mouse-data to be adapted to the processing function instead of vice-versa. Ioanas et al., 2019 establishes a novel workflow defined as generic, specifically designed for mouse brain imaging, and benchmarks it against the legacy procedure. While the reported performance increase is considerable, registration is nonetheless influenced by intensity variations outside the brain region. In-vivo as well as ex-vivo MRI head scans, present higher variability in the viscerocranial and extracranial tissue than in the neurocranium and the brain region of interest. Usage of unmasked (i.e. non brain extracted) data as done by the generic method, can thus lead to stretching or skewing of the brain during the registration process. Computing the transformation solely on the brain volume removes disturbances induced by intensity variations outside the brain region and further improves registration quality.

In recent years it has been shown that convolutional neural networks give the best results for semantic image segmentation in terms of precision and flexibility (Geng, Zhou and Cao, 2018). Especially the U-Net architecture from (Ronneberger, Fischer and Brox, 2015) is to this day one of the most popular in the field of biomedical image segmentation. Training a neural network into a classifier is a supervised method. This means that the model needs to learn its parameters based on observations of labeled data. Manually creating annotations as required to train a deep-learning classifier for high-resolution data is often infeasible, since it requires manual expert segmentation of vast amounts of slices. In the medical domain especially, human labeled data is expensive to acquire and thus very scarce. A much more widely applicable approach is to train the network using the template mask as label together with registered scans. Registration is not as precise as human labeling, but it is automatic and does not depend on expert input. Tajbakhsh et al., 2020 show that deep learning methods can indeed show satisfiable results when trained with imperfect training data. While our purpose was to create a workflow that generates better masks than the one from the template space, we show that the latter can be used as training data for the deep-learning model, by applying small changes to it.

In this study we investigate whether and in how far reliable classification can be obtained from imperfect training data and whether preclinical image masking improves an optimized registration workflow. We provide the methods as a free and open source softare (FOSS) package (Klug, 2020, MLEBE) as well as the functions needed for the data analysis in this article as a RepSeP document (Ioanas and Rudin, 2018). We evaluate the effects of our classifier on a full-fledged registration workflow via the benchmarking algorithms from Ioanas et al., 2019.

Classifier Implementation

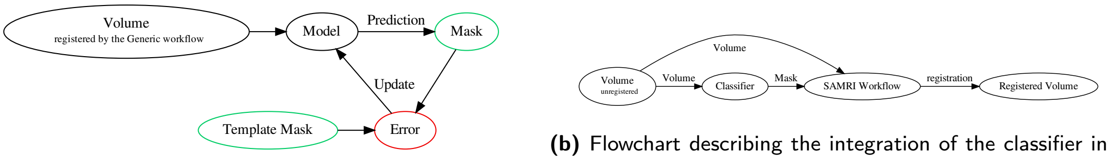
We lay out a preparatory step to improve brain registration by specifically extracting the brain volume from the MRI scans. Our solution utilises a machine learning enabled brain tissue classifier, and the software implementation is formulated to integrate with the SAMRI Generic workflow (Ioanas et al., 2019), in order to ensure broader usability and reproducible benchmarking. It creates a mask of the brain region using a classifier, which is then used to extract the region of interest. Two classifiers were trained, one for scans acquired with RARE sequences yielding T_2-weighted contrast and one scans acquired with gradient-echo EPI sequences yielding either BOLD (Ogawa et al., 1990) and CBV (Marota et al., 1999) contrasts (see section 3.2). The assignment of “brain” and “not brain” annotation to each voxel in the scans is performed via a trained U-Net, a popular neural network for medical image segmentation first introduced in Ronneberger, Fischer and Brox, 2015.

Workflow integration

The brain extraction nodes of the workflow return both the masked input and the binary mask. The latter is used to constrain image similarity metric estimation on the relevant region of interest (ROI), while the extracted brain volume is used to prevent drifting of extracranial hyperintensities into the ROI. The registration transformation is applied to the unmasked data to make the process minimally destructive. Figure 1b shows the integration of the classifier into the SAMRI workflow in a simplified manner.

Training Data

To improve general-purpose application, training examples need to be drawn from a usually unknown probability distribution, which is expected to be representative of the space of occurrences. We set up an occurence space from which the data of interest is drawn, consisting of all the different mouse brain MRI data sets coming from multiple experiments, with their corresponding labels. Based on an approximation of the occurrence space, the network builds a general model that enables it to extrapolate and produce sufficiently accurate predictions in new cases. As a training dataset, we use scans which were preprocessed with the *SAMRI* 2019, SAMRI Generic workflow. This data thus contains scans mapped onto a bregma-centered standard Ioanas et al., 2019 space derived from the Toronto Hospital for Sick Children Mouse Imaging Center brain template Dorr et al., 2008. A template-based mask is available in the same reference space, and constitutes a ground truth estimation. As registration in the absence of brain extraction is prone to imperfections, the mask does not always align perfectly with the brain region of every slice and some scans had to be removed manually. Figure 1a depicts the training workflow of the classifier.



(a) Flowchart describing the training process of the classifier.

(b) Flowchart describing the integration of the classifier into the SAMRI Generic workflow in a simplified manner.

Figure 1: Flowcharts describing the training and the integration of the classifier in a simplified manner.

Methods

For the benchmarking of the two workflows, the same methods that are described in the original paper have been applied in this work. A more detailed description can be found there.

Model

As the architecture of the classifier, the U-Net from Ronneberger, Fischer and Brox, 2015 was chosen based on its high performance in the field of biomedical image segmentation. This is a convolutional neural network that consists of a contracting path that captures context in addition to a symmetric expanding path that enables precise localization. Localization in this context means that a class label is assigned to each pixel. We used the attention gated U-Net implementation from Oktay et al., n.d. for which the code is publicly available (Oktay, 2020). Additionally to the original U-Net structure, their implementation has attention gates in the expanding part which weight the information coming from the symmetric counterpart. The additional parameters in these attention gates allow the model to learn which region of the image is important for specific tasks and to suppress irrelevant regions. In our use case this implementation helped reduce false positive classifications of high intensities, outside of the mouse brain region.

The model was trained using the Dice loss, which is computed from the Dice score. It calculates the similarity of two binary samples X and Y with

$$D_{coef} = \frac{2|X \cap Y| + \epsilon}{|X| + |Y| + \epsilon} \tag{1}$$

where a smoothing factor ϵ of 0.01 is used.

It is a quantity ranging from 0 to 1 that is to be maximized. The parameters of the model are updated such that the objective $1 - D_{coef}$ is minimized.

The mask consists of much more background than foreground which constitutes a class imbalance problem. Using the Dice coefficient as a loss function for training should make it invariant to this class imbalance (Milletari, Navab and Ahmadi, 2016).

Data Set

The data set consists of 3D MR images taken from an aggregation of three studies: (Ioanas and Rudin, 2019, irsabi), (Ioanas, Saab and Rudin, n.d.[a], opfvta), (Ioanas, Saab and Rudin, n.d.[b], drlfom) and other unpublished data, acquired with similar parameters.

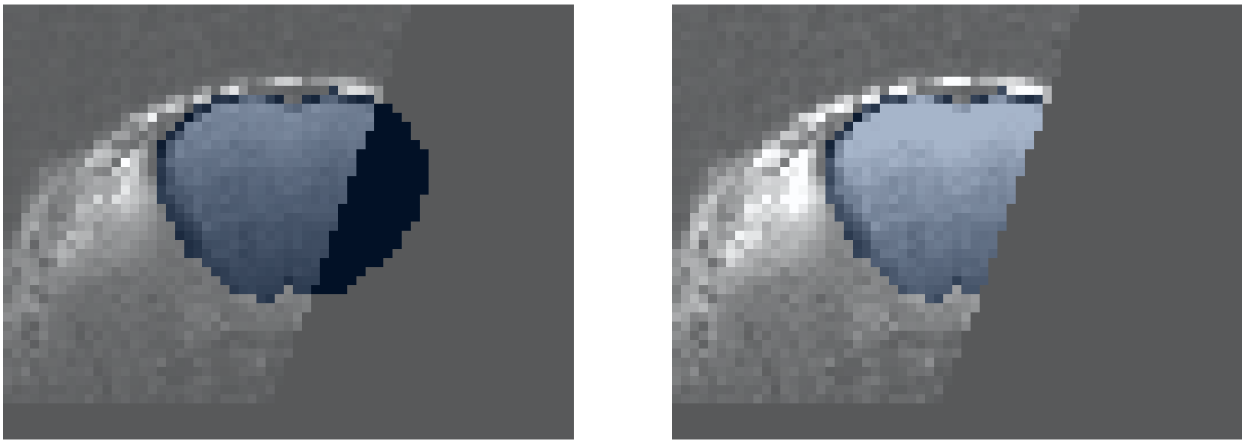
The irsabi data set consists of 102 scans coming from 11 adult animals, each scanned in up to 5 sessions with a 7T Bruker PharmaScan. The sessions were repeated at 14 days intervals, each containing one anatomical (echo-time: 21ms, inter-echo spacing: 7ms, repetition time (TR): 2500ms) and two functional (CBV and BOLD with a flip angle of 60°) scans. The functional scans were sampled at $\Delta x(\nu) = 312.5\,\mu m$, $\Delta y(\phi) = 281.25\,\mu m$, and $\Delta z(t) = 650\,\mu m$ (slice thickness of 500 µm).

The opfvta data set consists of 106 scans coming from 32 adult animals, each scanned in up to 8 sessions with a 7T Bruker PharmaScan. The sessions were repeated at ??? days intervals, each containing one anatomical (echo-time: 30ms, inter-echo spacing: 10ms, repetition time (TR): 2950ms) and a functional (CBV with a flip angle of 60°) scan. The functional scans were sampled at $\Delta x(\nu) = y(\phi) = 75\,\mu m$ and a slice thickness of $\Delta z(t) = 450\,\mu m$.

The drlfom data set consists of 306 scans coming from 39 adult animals, each scanned in up to 10 sessions with a 7T Bruker PharmaScan. The sessions were repeated at ??? days intervals, each containing one anatomical (echo-time: 30ms, inter-echo spacing: 10ms, repetition time (TR): 2950ms) and a functional (CBV with a flip angle of 60°) scan. The functional scans were sampled at $\Delta x(\nu) = y(\phi) = 225\,\mu m$, and a slice thickness $\Delta z(t) = 450\,\mu m$.

The measured animals were fitted with an optic fiber implant ($l = 3.2\,mm$ $d = 400\,\mu m$) targeting the Dorsal Raphe (DR) nucleus in the brain stem. Using this dataset shows that the classifier is robust to these types of experiment setups. Images from the irsabi study are only used for quality control of the registration and are thus unknown to the classifier. It is the same dataset that was used to benchmark the Generic workflow in the original paper and thus allows for a better estimation of the general performance of our improved pipeline.

The images are transformed into a standard space using a template mask via (*SAMRI* 2019, SAMRI) and are thus defined in the same affine space. SAMRI is a data analysis package of the ETH/UZH Institute for Biomedical Engineering. It is equipped with an optimized registration workflow and standard geomet-



(a) Example of an unpreprocessed slice.

(b) Example of a preprocessed slice.

Figure 2: The preprocessing removes the mask there, where the image-pixelvalues are 0. Plots of the same image, superposed with the template mask, with and without preprocessing.

ric space for small animal brain imaging (Ioanas et al., 2019).

Because of variance in mouse brain anatomy and in the experiment setup, some of the transformed data do not overlap perfectly with the reference template. To filter these images out, most of the incongruent volumes were removed manually from the data set.

For the registration of the images, a padding was needed to make the originally not affine space affine. As a result, the 3D volumes present many zero-valued slices, some of them overlapping with the mask.

Since it is not wanted for the model to predict a mask on black slices, the mask is set to zero where the image is zero-valued. This has also the advantage of bringing variance into the template. Because some pixels representing the brain tissue are zero-valued, holes result from this operation. To patch these, the function *binary_fill_holes* from the scipy package (Virtanen et al., 2020) is used. An example of the preprocessing can be seen in fig. 2.

Each volume of the transformed data is originally of shape (63, 96, 48), matching the reference space resolution of 200 µm. The volume is then reshaped into (64, 64, 96) by first zero-padding the smaller x-y dimension to the same size as the bigger one. This is done to conserve the ratio of the image. The z-axis is zero-padded to 96. The scan is then reshaped into (64, 64, 96) using the function *cv2.resize* from the opencv python package (Bradski, 2000).

Finally, the images are normalized by first clipping them from the minimum to the 99th percentile of the data to remove outliers and then divided by the maximum.

The data set is separated into Training, Validation and Test sets such that 90% of the total data are used for training and validation while 10% are used for testing. The Validation set is used for the optimization of hyperparameters while the Test set is used as a measure of extrapolation capability. The irsabi data was additionally added to the test set.

Data Augmentation

Because of diverse settings in the experiment setup, including animal manipulations causing artifacts, MR image quality can differ substantially between labs and even individual study populations. To account for these variations, we apply an extensive set of transformations to our data. This includes rotations of up to 20°, a zoom range of -0.2 to +0.1, a random bias field added on the images and horizontal as well as vertical flips. Additionally a gaussian noise is added to the images.

This not only increases the data set size but also makes it more representative of the general data distribution of mice brain MR images and results in a model with a better generalization capability.

Masking

To improve the SAMRI registration workflow, an additional node is implemented where the images are masked, such that only the brain region remains. The image is first resampled into the resolution of the template space, which has a voxel size of $0.2 \times 0.2 \times 0.2$. This is done with the *Resample* command from the FSL library which is an analysis tool for FMRI, MRI and DTI brain imaging data (Jenkinson et al., 2012). Then, the image is preprocessed using the operations described in section 3.2. Since the classifier was trained to predict on images of shape (64, 64, 96), the input needs to be reshaped. The predictions of the model are reconstructed to a 3D mask via the command *Nifit1Image* from the neuroimaging python package nibabel (*Neuroimaging in Python — NiBabel 2.5.0 documentation* n.d.). This is done using the same affine space as the input image. The latter is then reshaped into the original shape inverting the preprocessing step, either with the opencv resize method or by cropping. Additionally, the binary mask is resampled into its original affine space, before being multiplied with the brain image to extract the ROI.

Metrics

The VCF uses the 66th voxel intensity percentile of the raw scan before any processing as definition of the brain volume. The VCF is then obtained with eq. (2), where v is the voxel volume in the original space, v' the voxel volume in the transformed space, n the number of voxels in the original space, m the number of voxels in the transformed space, s a voxel value sampled from the vector S containing all values in the original data, and s' a voxel value sampled from the transformed data.

$$VCF = \frac{v' \sum_{i=1}^{m} [s'_i \geq P_{66}(S)]}{v \sum_{i=1}^{n} [s_i \geq P_{66}(S)]} = \frac{v' \sum_{i=1}^{m} [s'_i \geq P_{66}(S)]}{v \lceil 0.66n \rceil} \tag{2}$$

The SCF metric is based on the ratio of smoothness before and after processing. It is obtained by taking the full-width at half-maximum of the signal amplitude spatial autocorrelation function (ACF Eklund, Nichols and Knutsson, 2016). In eq. (3), r is the distance of two amplitude distribution samples, a is the relative weight of the Gaussian term in the model, b is the width of the Gaussian and c the decay of the mono-exponential term Cox et al., 2017.

$$ACF(r) = a * e^{-r^2/(2*b^2)} + (1-a) + e^{-r/c} \tag{3}$$

The for the MS relevant statistical power is obtained via the negative logarithm of first-level p-value maps. Voxelwise statistical estimates for the probability that a time course could — by chance alone — be at least as well correlated with the stimulation regressor as the voxel time course measured are averaged via eq. (4), where n represents the number of statistical estimates in the scan, and p is a p-value.

$$MS = \frac{\sum_{i=1}^{n} -log(p_i)}{n} \tag{4}$$

Statistics

In the results section, all statistics are presented with respect to the distributions of the absolute distances to 1, i.e. |1 - Metric|. Based on a Likelihood Ratio Test, we chose models that do not examine the Workflow- Contrast interaction. The full summaries of the analysis can be seen in tables table S1, table S2, table S3 and table S4.

Results

For the quality control of the workflow, we first evaluate the classification process, followed by a benchmark between the Generic and the improved "Masked" workflow.

Classification

Quality control of our classifier is difficult in the sense that the template mask does not always overlap perfectly with the brain region, such that small deviances of the predictions compared to the template could actually be caused by the prediction being more accurate than the template. Nevertheless, it is useful to verify whether the output is similar to the template, as it should be. As a similarity metric between the template mask and the classifier output we have used the Dice score (see eq. (1)). The average Dice score on the test data set is $D_{coef} = 0.982$, indicating that classifier output has only minor changes in comparison with the template.

Workflow

We use an established palette of workflow evaluation metrics — inspecting volume and smoothness conservation, as well as downstream effects on basic functional analysis (Ioanas et al., 2019) — to benchmark the novel SAMRI Masked workflow against the SAMRI Generic workflow. Statistics for the Volume Conservation and the Smoothness Conservation are presented with respect to the distributions of the absolute distances to the optimal value 1.

A qualitative evaluation of the registered volume shows that the classifier reduces the shifting of outer brain regions into the brain region and improves the quality of the registration. This can be seen in fig. 4, comparing slices of three different registered volumes with and without the help of the classifier.

Volume Conservation

Volume Conservation Factor (VCF) measures the registration induced deformation of the scanned brain, by computing the ratio of the brain volume before and after preprocessing. A positive ratio indicates that the brain was stretched to fill the template space, while a negative ratio indicates that non-brain voxels were introduced in the template brain space. Volume conservation is highest for a VCF equal to 1, indicating that the preprocessing has no influence on the brain volume of the scans.

As seen in fig. 5c, we note that in the described dataset the absolute distance of the VCF to 1 is sensitive to the workflow ($F_{1,133} = 4.529$, $p = 0.035$). The performance of the Generic SAMRI workflow is different from that of the Masked, yielding a two-tailed p-value of 0.019. With respect to the data break-up by contrast (CBV versus BOLD, fig. 5a), we see no notable main effect for the contrast variable (VCF of 0.01, 95%CI: 0.00 to 0.03 1).

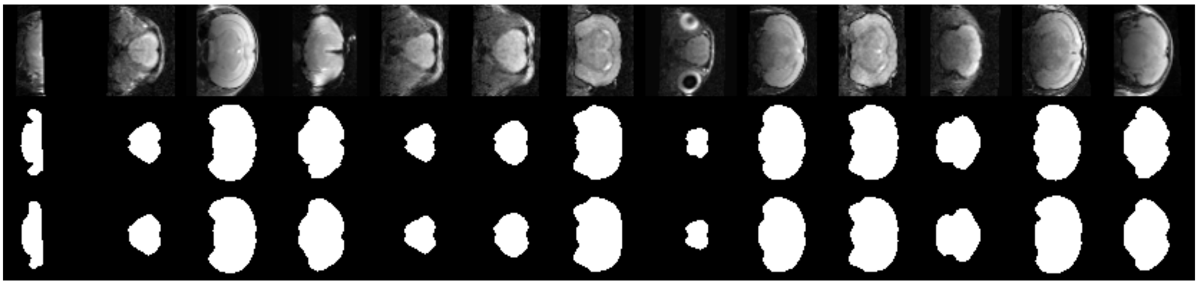


Figure 3: The Classifier predicts a similar mask to the ground truth. Randomly picked plots from the test set illustrate the predictions of the classifier. The first row presents the input image, the second the ground truth and the third row shows the predictions of the classifier.

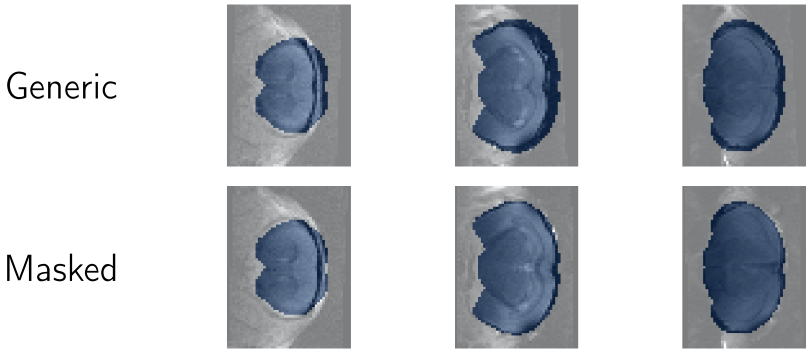


Figure 4: The Masked workflow prevents the shifting of outer-brain region voxels into the template-brain region (in blue). Comparison of slices from 3 different volumes, registered with the Generic (first row) and the Masked (second row) workflow.

We note that there is a significant variance decrease in all conditions for the Masked workflow (0.44-fold). Further, we note that the root mean squared error ratio favours the Masked workflow ($\mathrm{RMSE_M}/\mathrm{RMSE_G} \simeq 0.66$).

Smoothness Conservation

Smoothing is a popular tool employed by many preprocessing functions to increase the signal-to-noise ratio. Image smoothness comes at the cost of image contrast as well as feature saliency and has been shown to result in inferior anatomical alignment due to the loss of spatial resolution (Esteban et al., 2019). As an indicator of image smothness induced by the workflow, the Smoothness Conservation Factor (SCF) expresses the ratio between the smoothness of the preprocessed images and the smoothness of the original images. Smoothess Conservation is highest for a SCF equal to 1, indicating that the preprocessing does not influence image smoothness.

While the performance of the Generic SAMRI workflow is only slightly different from that of the Masked workflow, the root mean squared error ratio favors the Masked workflow ($\mathrm{RMSE_M}/\mathrm{RMSE_G} \simeq 0.96$).

Descriptively, we observe that neither the Generic nor the Masked workflow introduce a strong smoothing (SCF of 0.00, 95%CI: −0.01 to 0.01 1).

Further, we note that there is a slight variance decrease for the Masked workflow (0.92 -fold).

Given the break-up by contrast shown in fig. 5b, we see no effect for the contrast variable (SCF of 0.03, 95%CI: 0.01 to 0.05 1).

Functional Analysis

Functional Analysis expresses the significance detected across all voxels of a scan by computing the Mean Significance (MS) Ioanas et al., 2019.

We observe that the Masked level of the workflow variable does not introduce a notable significance loss (MS of −0.01, 95%CI: −0.03 to 0.01 1). Furthermore, we note a slight variance decrease in all conditions for the Masked workflow (0.95-fold).

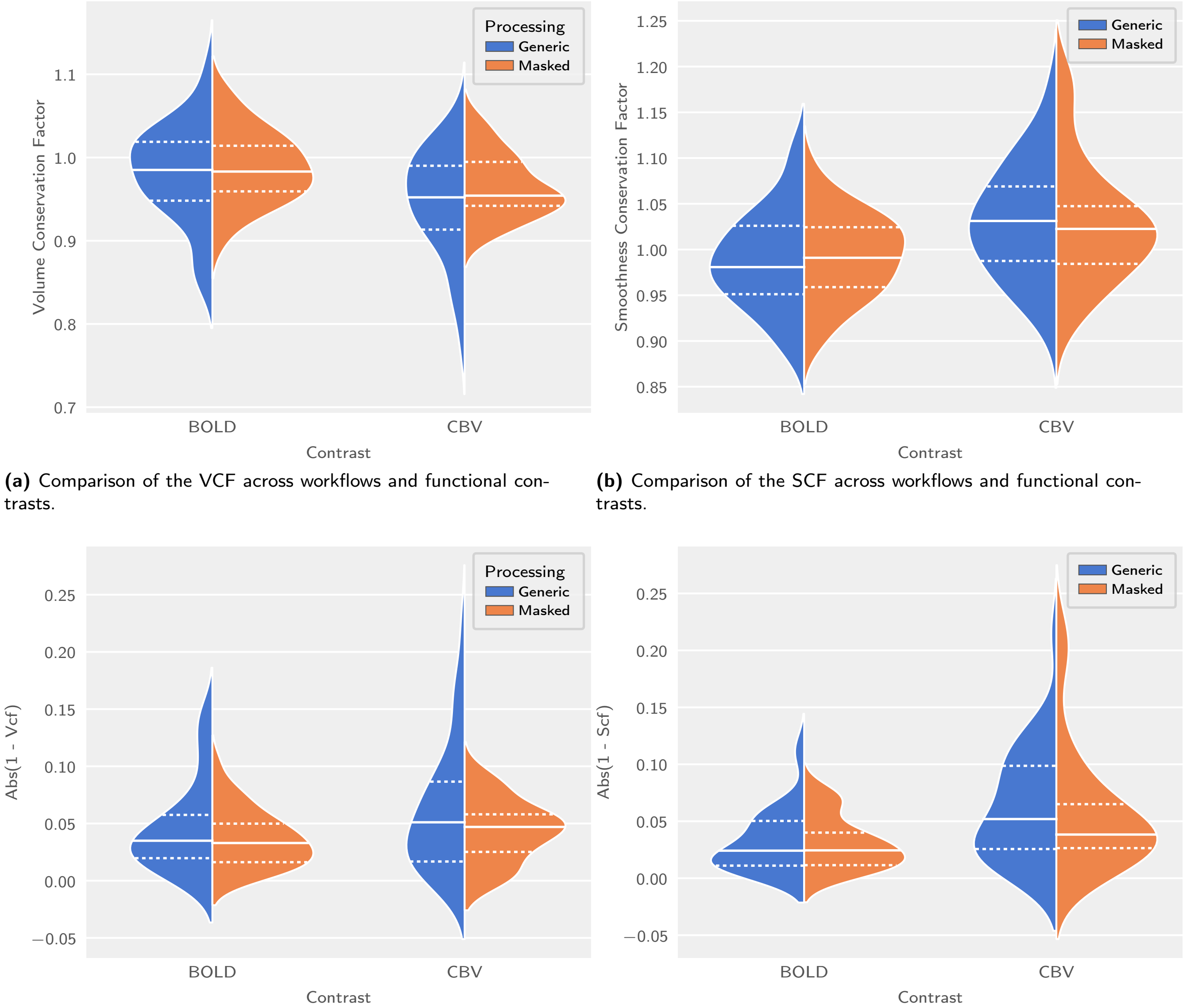
With respect to the data break-up by contrast (fig. S1), we see no notable main effect for the contrast variable (MS of −0.09, 95%CI: −0.87 to 0.69 1).

Variance Analysis

As an additional metric for the comparison between workflows, we evaluate if physiological meaningfull variability is retained across repeated measurements. It is based on the assumption that adult mouse brains retain size, shape, and implant position in the absence of intervention, throughout the 8 week study period Ioanas et al., 2019. Examining the similarity between the template and preprocessed scans, session-wise variability should be smaller than subject-wise variability. This comparison is performed using a type 3 ANOVA, modeling both the subject and the session variables. For this assessment three metrics are used, with maximal sensitivity to different features: Neighborhood Cross Correlation (CC, sensitive to localized correlation), Global Correlation (GC, sensitive to whole-image correlation), and Mutual Information (MI, sensitive to whole-image information similarity).

Both, the Generic and the Masked workflow produce results which show a higher F-statistic for the subject than for the session variable. For the Masked workflow, F-statistics show: CC (subject: $F_{10,19} = 9.035$, $p = 2.58 \times 10^{-5}$, session: $F_{4,19} = 6.127$, $p = 0.0024$), GC (subject: $F_{10,19} = 3.291$, $p = 0.012$, session: $F_{4,19} = 2.021$, $p = 0.13$), and MI (subject: $F_{10,19} = 1.31$, $p = 0.29$, session: $F_{4,19} = 1.392$, $p = 0.27$).

For the Generic SAMRI workflow, resulting data F-statistics show: CC (subject: $F_{10,19} = 3.662$, $p = 0.0072$, session: $F_{4,19}=3.09$, $p=0.041$), GC (subject: $F_{10,19} = 2.053$, $p = 0.085$, session: $F_{4,19} = 1.432$, $p = 0.26$), and MI (subject: $F_{10,19}=1.331$, $p=0.28$, session: $F_{4,19}=2.196$, $p=0.11$).



(a) Comparison of the VCF across workflows and functional contrasts.

(b) Comparison of the SCF across workflows and functional contrasts.

(c) Comparison of the distributions of the absolute VCF errors, across workflows and functional contrasts.

(d) Comparison of the distributions of the absolute SCF errors, across workflows and functional contrasts.

Figure 5: Both the SAMRI Generic and the Masked workflow optimally and reliably conserve volume and smoothness, the latter showing values that are closely distributed to 1. Plots showing the distribution of two target metrics in the first row, together with the respective distributions of the absolute distances to 1 in the second row. Solid lines in the colored distribution densities indicate the sample mean and dashed lines the inner quartiles.

Discussion

The classifier improves the volume conservation, smoothness conservation, and session-to-session consistency of the SAMRI Generic workflow in terms of precision while conserving accuracy.

Visual inspection of registration quality reveals that the classifier successfully reduces the shifting of outer brain region voxels into the template space.

These benefits of the classifier are robust to the functional contrast (figs. 5a and 5b), with the Generic Masking workflow being less or equally susceptible to the contrast variable, when compared to the Generic workflow.

The classifier improves the performance of the SAMRI Generic workflow, making these accessible in the same interface with the same advantages in terms of transparency, parametrization, ease of package management, and non-destructive metadata management.

Our workflow has the advantage that the performance of a Neural Network can increase when trained further with new data. The FOSS distribution model for both the classifier and workflow, as well as the article, allows users to easily take advantage of the classifier extendability and recreate the steps described herein. Registering new data with the Generic Maksed workflow can increase the size of the training data set of the classifier. After removing possibly bad registrations, the latter can be trained again, which will improve its generalisation capability. Another advantage of the trainability of the classifier and the openly published code is that this workflow can be adapted to a wast variety of data types.

The complete workflow of this report is fully reproducible and thus easily verifiable. We make public the functions used for the masking in the workflow as well as those used to train the classifier, through the *mlebe* python package (Klug, 2020).

Conclusion

We present a brain labeling classifier, that when used as a ROI extraction in an extention of the SAMRI Generic registration workflow, significantly improves the quality of the latter. The extended Generic Masked workflow offers several advantages summarized by established metrics for data features commonly biased by registration. Comparison with the original SAMRI Generic workflow revealed superior performance of the SAMRI Generic Masked workflow in terms of volume and smoothness conservation, as well as variance structure across subjects and sessions. The easily accessible, optimized registration parameters of the SAMRI Generic Workflow as well as the open source code to the classifier training functions make the pipeline transferable to any other imaging applications. The open source software choices in both the workflow and this article's source code empower users to better verify, understand, remix, and reuse our work.

References

Borsook, David, Lino Becerra and Richard Hargreaves (May 2006). “A role for fMRI in optimizing CNS drug development”. In: *Nature Reviews Drug Discovery* 5.5. Number: 5 Publisher: Nature Publishing Group, pp. 411–425. ISSN: 1474-1784. DOI: 10.1038/nrd2027. URL: https://www.nature.com/articles/nrd2027 (visited on 01/05/2020).

Bradski, G. (2000). “The OpenCV Library”. In: *Dr. Dobb's Journal of Software Tools*.

Cox, Robert W et al. (Apr. 2017). “FMRI clustering in AFNI: false-positive rates redux”. In: *Brain connectivity* 7.3, pp. 152–171. DOI: 10.1089/brain.2016.0475. URL: https://doi.org/10.1089/brain.2016.0475.

Dorr, A.E. et al. (Aug. 2008). “High resolution three-dimensional brain atlas using an average magnetic resonance image of 40 adult C57Bl/6J mice”. In: *NeuroImage* 42.1, pp. 60–69. DOI: 10.1016/j.neuroimage.2008.03.037. URL: https://doi.org/10.1016/j.neuroimage.2008.03.037.

Eklund, Anders, Thomas E Nichols and Hans Knutsson (June 2016). “Cluster failure: why fMRI inferences for spatial extent have inflated false-positive rates”. In: *Proceedings of the National Academy of Sciences*, p. 201602413. DOI: 10.1073/pnas.1602413113. URL: https://doi.org/10.1073/pnas.1602413113.

Esteban, Oscar et al. (Dec. 2019). “FMRIPrep: a robust preprocessing pipeline for functional MRI”. In: *Nature Methods*, 111–116. DOI: 10.1038/s41592-018-0235-4. URL: https://doi.org/10.1038/s41592-018-0235-4.

Friston, K. J., L. Harrison and W. Penny (1st Aug. 2003). “Dynamic causal modelling”. In: *NeuroImage* 19.4, pp. 1273–1302. ISSN: 1053-8119. DOI: 10.1016/S1053-8119(03)00202-7. URL: http://www.sciencedirect.com/science/article/pii/S1053811903002027 (visited on 01/05/2020).

Geng, Qichuan, Zhong Zhou and Xiaochun Cao (May 2018). “Survey of recent progress in semantic image segmentation with CNNs”. en. In: *Science China Information Sciences* 61.5, p. 051101. ISSN: 1674-733X, 1869-1919. DOI: 10.1007/s11432-017-9189-6. URL: http://link.springer.com/10.1007/s11432-017-9189-6 (visited on 17/01/2020).

Ioanas, Horea-Ioan and Markus Rudin (Aug. 2018). “Reproducible Self-Publishing for Python-Based Research”. In: EuroSciPy. DOI: 10.6084/m9.figshare.7247339.v1. URL: https://figshare.com/articles/Reproducible_Self-Publishing_for_Python-Based_Research/7247339.

Ioanas, Horea-Ioan and Markus Rudin (Apr. 2019). *BIDS Data for "An Optimized Registration Workflow and Standard Geometric Space for Small Animal Brain Imaging"*. DOI: 10.5281/zenodo.3445428. URL: https://doi.org/10.5281/zenodo.2651640.

Ioanas, Horea-Ioan, Bechara John Saab and Markus Rudin (n.d.[a]). “A Whole-Brain Map and Assay Parameter Analysis of Mouse VTA Dopaminergic Activation”. en. In: (), p. 19. URL: https://www.biorxiv.org/content/10.1101/2020.04.03.023648v1.

– (n.d.[b]). “Effects of Acute and Chronic Reuptake Inhibition on Optogenetically Induced Serotonergic Activity”. en. In: (), p. 20.

Ioanas, Horea-Ioan et al. (Apr. 2019). *An Optimized Registration Workflow and Standard Geometric Space for Small Animal Brain Imaging*. en. preprint. Neuroscience. DOI: 10.1101/619650. URL: http://biorxiv.org/lookup/doi/10.1101/619650 (visited on 05/01/2020).

Jenkinson, Mark et al. (2012). “Fsl”. In: *Neuroimage* 62.2, pp. 782–790.

Klug, Hendrik (18th Mar. 2020). *Jimmy2027/MLEBE*. original-date: 2019-10-13T09:57:20Z. URL: https://github.com/Jimmy2027/MLEBE (visited on 23/03/2020).

Maintz, J B Antoine and Max A Viergever (n.d.). “An Overview of Medical Image Registration Methods”. en. In: (), p. 22.

Marota, John J.A. et al. (Feb. 1999). “Investigation of the early response to rat forepaw stimulation”. In: *Magnetic Resonance in Medicine* 41.2, pp. 247–252. DOI: 10.1002/(sici)1522-2594(199902)41:2<247::aid-mrm6>3.0.co;2-u. URL: https://doi.org/10.1002/(sici)1522-2594(199902)41:2<247::aid-mrm6>3.0.co;2-u.

Milletari, Fausto, Nassir Navab and Seyed-Ahmad Ahmadi (June 2016). “V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation”. In: *arXiv:1606.04797 [cs]*. arXiv: 1606.04797. URL: http://arxiv.org/abs/1606.04797 (visited on 05/01/2020).

Neuroimaging in Python — NiBabel 2.5.0 documentation (n.d.). URL: https://nipy.org/nibabel/ (visited on 07/01/2020).

Ogawa, Seiji et al. (Apr. 1990). “Oxygenation-sensitive contrast in magnetic resonance image of rodent brain at high magnetic fields”. In: *Magnetic Resonance in Medicine* 14.1, pp. 68–78. DOI: 10.1002/mrm.1910140108. URL: https://doi.org/10.1002/mrm.1910140108.

Oktay, Ozan (22nd July 2020). *ozan-oktay/Attention-Gated-Networks*. original-date: 2018-04-02T21:50:07Z. URL: https://github.com/ozan-oktay/Attention-Gated-Networks (visited on 24/07/2020).

Oktay, Ozan et al. (n.d.). “Attention U-Net: Learning Where to Look for the Pancreas”. In: (), p. 10.

Ronneberger, Olaf, Philipp Fischer and Thomas Brox (May 2015). “U-Net: Convolutional Networks for Biomedical Image Segmentation”. In: *arXiv:1505.04597 [cs]*. arXiv: 1505.04597 version: 1. URL: http://arxiv.org/abs/1505.04597 (visited on 05/01/2020).

SAMRI (Dec. 2019). original-date: 2015-04-27T00:26:08Z. URL: https://github.com/IBT-FMI/SAMRI (visited on 05/01/2020).

Sotiras, Aristeidis, Christos Davatzikos and Nikos Paragios (July 2013). “Deformable Medical Image Registration: A Survey”. In: *IEEE Transactions on Medical Imaging* 32.7, pp. 1153–1190. ISSN: 1558-254X. DOI: 10.1109/TMI.2013.2265603.

Tajbakhsh, Nima et al. (3rd Apr. 2020). “Embracing Imperfect Datasets: A Review of Deep Learning Solutions for Medical Image Segmentation”. In: *Medical Image Analysis*, p. 101693. ISSN: 1361-8415. DOI: 10.1016/j.media.2020.101693. URL: http://www.sciencedirect.com/science/article/pii/S136184152030058X (visited on 10/04/2020).

Virtanen, Pauli et al. (2020). “SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python”. In: *Nature Methods* 17, pp. 261–272. DOI: https://doi.org/10.1038/s41592-019-0686-2.

Supplementary Materials

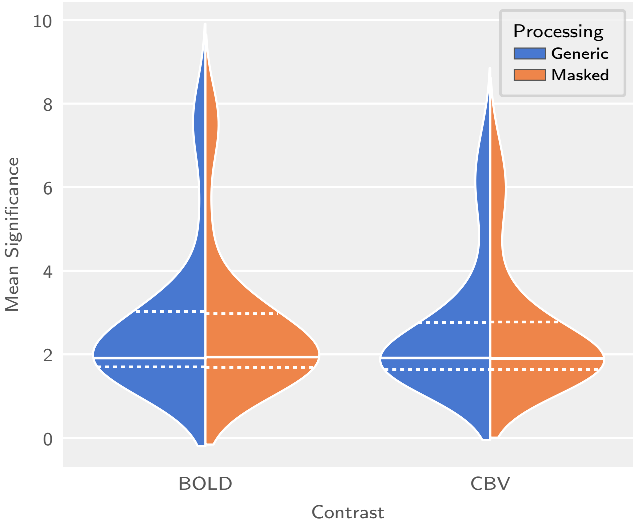
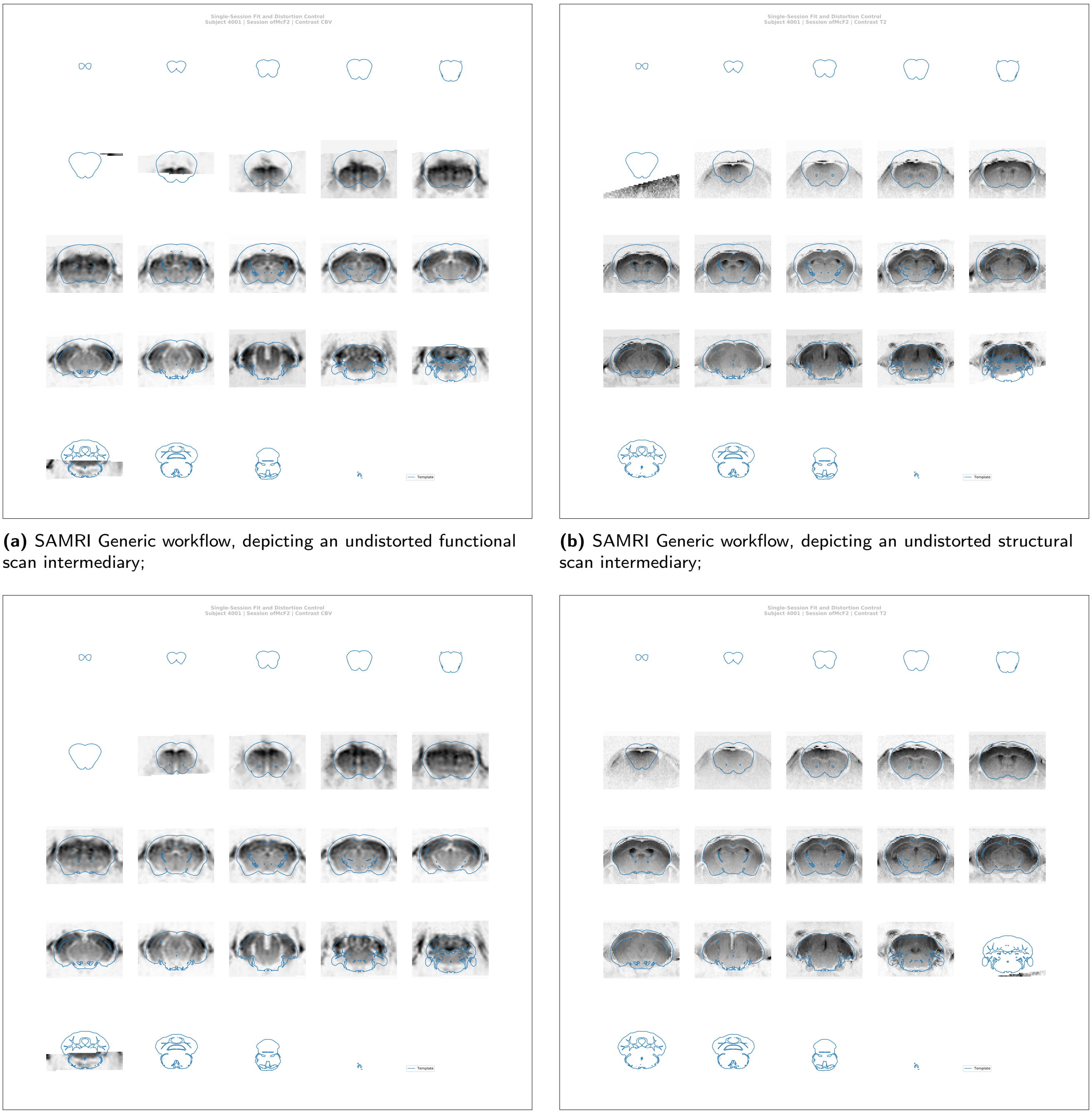


Figure S1: The Generic Masked workflow does not introduce a loss of significance. Comparison across workflows and functional contrasts.


Single-Session Fit and Distortion Control
Subject 4001 | Session ofMcF2 | Contrast CBV
Template
Single-Session Fit and Distortion Control
Subject 4001 | Session ofMcF2 | Contrast T2
Template
Single-Session Fit and Distortion Control
Subject 4001 | Session ofMcF2 | Contrast CBV
Template
Single-Session Fit and Distortion Control
Subject 4001 | Session ofMcF2 | Contrast T2
Template

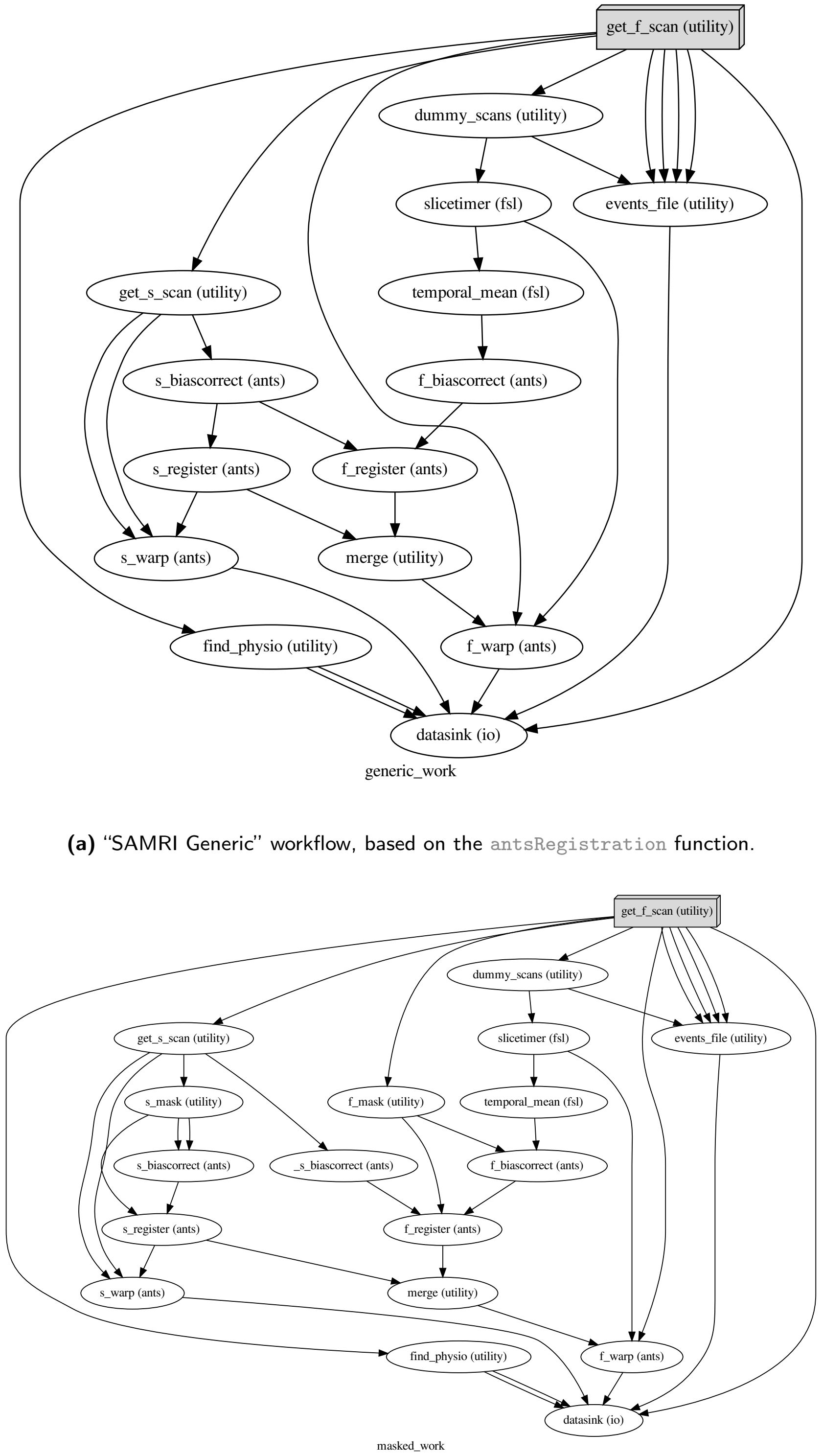
(a) SAMRI Generic workflow, depicting an undistorted functional scan intermediary;

(b) SAMRI Generic workflow, depicting an undistorted structural scan intermediary;

(c) SAMRI Generic Masked workflow, depicting an undistorted functional scan intermediary;

(d) SAMRI Generic Masked workflow, depicting an undistorted structural scan intermediary;

Figure S2: The SAMRI Generic Masked provides a more accurate coverage of the template space. Depicted are slice-by-slice inspections of the registration fit, with a spacing that is analogous to acquisition.



(a) "SAMRI Generic" workflow, based on the `antsRegistration` function.

(b) "SAMRI Generic Masked" workflow, which is based on the `antsRegistration` function.
Two additional nodes provide the workflow with both the masked image and the binary mask.

Figure S3: Directed acyclic graphs visualising the two registration workflows. Each node name is depicted together with its corresponding package name in paranthesis. The "utility" indication corresponds to nodes based on Python functions specific to the workflow, distributed alongside it, and dynamically wrapped via Nipype.

Table S1: Mixed Linear Model Regression Results – With Processing/Contrast Interaction for the Volume Conservation Factor

Model:	MixedLM	Dependent Variable:	Q("Abs(1 - Vcf)")
No. Observations:	136	Method:	REML
No. Groups:	68	Scale:	0.0008
Min. group size:	2	Log-Likelihood:	243.2700
Max. group size:	2	Converged:	Yes
Mean group size:	2.0		

	Coef.	Std.Err.	z	P> |z|	[0.025	0.975]
Intercept	0.046	0.007	7.081	0.000	0.034	0.059
Processing[T.Masked]	-0.010	0.007	-1.406	0.160	-0.024	0.004
Contrast[T.CBV]	0.015	0.009	1.619	0.105	-0.003	0.033
Processing[T.Masked]:Contrast[T.CBV]	-0.008	0.010	-0.813	0.416	-0.028	0.011
Uid Var	0.001	0.009				

Table S2: Mixed Linear Model Regression Results – Without Processing/Contrast Interaction for the Volume Conservation Factor

Model:	MixedLM	Dependent Variable:	Q("Abs(1 - Vcf)")
No. Observations:	136	Method:	REML
No. Groups:	68	Scale:	0.0008
Min. group size:	2	Log-Likelihood:	246.6337
Max. group size:	2	Converged:	Yes
Mean group size:	2.0		

	Coef.	Std.Err.	z	P> |z|	[0.025	0.975]
Intercept	0.048	0.006	7.988	0.000	0.037	0.060
Processing[T.Masked]	-0.014	0.005	-2.808	0.005	-0.024	-0.004
Contrast[T.CBV]	0.011	0.008	1.402	0.161	-0.004	0.026
Uid Var	0.001	0.009				

Table S3: Mixed Linear Model Regression Results – With Processing/Contrast Interaction for the Smoothness Conservation Factor

Model:	MixedLM	Dependent Variable:	Q("Abs(1 - Scf)")
No. Observations:	136	Method:	REML
No. Groups:	68	Scale:	0.0005
Min. group size:	2	Log-Likelihood:	250.2832
Max. group size:	2	Converged:	Yes
Mean group size:	2.0		

	Coef.	Std.Err.	z	P> |z|	[0.025	0.975]
Intercept	0.032	0.007	4.659	0.000	0.019	0.046
Processing[T.Masked]	-0.001	0.005	-0.261	0.794	-0.012	0.009
Contrast[T.CBV]	0.030	0.010	3.008	0.003	0.010	0.049
Processing[T.Masked]:Contrast[T.CBV]	-0.003	0.008	-0.341	0.733	-0.018	0.012
Uid Var	0.001	0.015				

Table S4: Mixed Linear Model Regression Results – Without Processing/Contrast Interaction for the Smoothness Conservation Factor

Model:	MixedLM	Dependent Variable:	Q("Abs(1 - Scf)")
No. Observations:	136	Method:	REML
No. Groups:	68	Scale:	0.0005
Min. group size:	2	Log-Likelihood:	254.1779
Max. group size:	2	Converged:	Yes
Mean group size:	2.0		

	Coef.	Std.Err.	z	P> |z|	[0.025	0.975]
Intercept	0.033	0.007	4.946	0.000	0.020	0.046
Processing[T.Masked]	-0.003	0.004	-0.715	0.475	-0.010	0.005
Contrast[T.CBV]	0.028	0.009	3.122	0.002	0.011	0.046
Uid Var	0.001	0.015				