An Optimized Registration Workflow and Standard Geometric Space for Small Animal Brain Imaging

Horea-Ioan Ioanas[1] Markus Marks[2] Mehmet Fatih Yanik[2] Markus Rudin[1]

[1]Institute for Biomedical Engineering, ETH and University of Zurich

[2]Institute of Neuroinformatics, ETH and University of Zurich

Abstract — Given the need for comparability across subjects and studies, the quality of registration to a standard space is crucial for the reliability of Magnetic Resonance Imaging (MRI), and in particular functional MRI (fMRI). Small Animal MRI workflows commonly utilize scripts optimized for human data, and adapt the data to the analysis rather than vice-versa. Quality control (QC) is commonly performed by interactive operator inspection, making it infrequent, open to bias, slow, and unreproducible. In this paper we present a workflow and standard space optimized for mouse brain registration, as well as four separate metrics for automated QC, along with a visualization method to aid operator inspection. This novel workflow conserves variance across subjects while minimizing variance across sessions, and outperforms legacy practices **??**-fold in terms of volume conservation, and **??**-fold in terms of smoothness conservation. On account of superior performance and transparency, the "SAMRI Generic" workflow showcases best practices in the development of specialized small animal MRI analysis, and — given the release of all code needed to reproduce this publication from raw data — allows researchers to reuse our work for validated preprocessing and publication-ready QC.

Background

In order to make meaningful comparisons across subjects inside a study, it is imperative that the images lie in a standard reference frame. Because of positioning imprecision and anatomical animal variations, this is not the case for the original MR acquired images. To solve this issue, the images need to be projected into the reference frame via registration [**? ?**]. Intensities outside the brain region of a mouse MR image present high variations and bias the registration process. As a remedy, it would be useful to extract the region of interest and perform the registration on it. For this purpose, we propose a machine learning enabled brain extraction in an additional node to the workflow presented by Ioanas et al. [**?**]. The additional node creates a mask of the brain region with segmentation using a classifier and masks the image such that only the region of interest remains. **To do** (??)

Convolutional Neural Networks

In recent years it has been shown that convolutional neural network give the best results for semantic image segmentation in terms of precision and flexibility [**?**] [**?**]. Training a convolutional neural network into a classifier is a supervised method, meaning that the model needs to learn its parameters based on observations of data.

The training data set of a classifier is as important as the architecture of the model itself. To improve general-purpose application, training examples need to be drawn from a usually unknown probability distribution, that is expected to be representative of the space of occurrences. We define the space of occurrences as the space of which the data of interest is drawn from. In our case this consists of all the different mouse brain MRI data sets coming from multiple experiments, with their corresponding labels. Ideally experiment setups are uniform and the resulting data does not differ much, but small variations in the experiment setup and animal size are unavoidable. Based on an approximation of the occurrence space, the network has to build a general model that enables it to extrapolate and produce sufficiently accurate predictions in new cases. Manually creating annotations as required to train a deep-learning classifier for high-resolution data is often infeasible, as it requires manual expert segmentation of vast amounts of slices.

While our purpose was to create a workflow that generates better masks than the ones used for registration**To do** (??), we showed that the latter could be used as training data for the deep-learning model, by applying small changes to them.

Methods

The slice-wise predictions of the model are reconstructed to a 3D mask via the command *Nifit1Image* from the neuro-imaging python package nibabel [?]. This is done using the same affine space as the input image.

For the training of the classifier, the data are separated into a Training, Validation and Test set with the help of the function *train_test_split* from the package sklearn.model_selection [?].

Evaluation

A major challenge of registration QC is that a perfect mapping from the measured image to the template is undefined. Similarity metrics are ill-suited for QC because they are used internally by registration functions, whose mode of operation is based on maximizing them. Extreme similarity score maximization is not a desired outcome, particularly if nonlinear transformations are employed, as this may result in image distortions which should be penalized in QC. Moreover, similarity metrics are not independent, so the issues arising from similarity score maximization cannot be circumvented by maximizing a subset of metrics and performing QC via the remainder. To address this challenge, we developed four alternative evaluation metrics: volume conservation, smoothness conservation, functional analysis, and variance analysis. In order to mitigate possible differences arising from template features, we use these metrics for multifactorial analyses — including both a template and a workflow factor.

Volume Conservation

??

Volume conservation is based on the assumption that the total volume of the scanned segment of the brain should remain roughly constant after preprocessing. Beyond just size differences between the acquired data and the target template, a volume increase may indicate that the brain was stretched to fill in template brain space not covered by the scan, while a volume decrease might indicate that non-brain voxels were introduced into the template brain space. For this analysis we compute a Volume Conservation Factor (VCF), whereby volume conservation is highest for a VCF equal to 1.

As seen in **??**, we note that in the described dataset VCF is sensitive to the workflow (**??**), the template (**??**), but not the interaction thereof (**??**).

The performance of the Generic SAMRI workflow in conjunction with the Generic template is significantly different from that of the Legacy workflow in conjunction with the Legacy template, yielding a two-tailed p-value of **??**. Moreover, the root mean squared error ratio strongly favours the Generic workflow ($\mathrm{RMSE_L/RMSE_G} \simeq$**??**).

Descriptively, we observe that the Legacy level of the template variable introduces a notable volume loss (VCF of **??**), while the Legacy level of the workflow variable introduces a volume gain (VCF of **??**). Further, we note that there is a very strong variance increase in all conditions for the Legacy workflow (**??**-fold given the Legacy template, and **??**-fold given the Generic template).

With respect to the data break-up by contrast (CBV versus BOLD, **??**), we see no notable main effect for the contrast variable (VCF of **??**). We do, however, report a notable effect for the contrast-template interaction, with the Legacy workflow and CBV contrast interaction level introducing a volume loss (VCF of **??**).

Smoothness Conservation

A further aspect of preprocessing quality is the resulting image smoothness. Although controlled smoothing is a valuable preprocessing tool used to increase the signal-to-noise ratio (SNR), uncontrolled smoothness limits operator discretion in the trade-off between SNR and feature granularity. Uncontrolled smoothness can thus lead to undocumented and implicit loss of spatial resolution and is therefore associated with inferior anatomical alignment [?]. We employ a Smoothness Conservation Factor (SCF), expressing the ratio between the smoothness of the preprocessed images and the smoothness of the original images.

With respect to the data shown in **??**, we note that SCF is sensitive to the template (**??**), the workflow (**??**), and the interaction of the factors (**??**).

The performance of the Generic SAMRI workflow in conjunction with the Generic template is significantly different from that of the Legacy workflow in conjunction with the Legacy template, yielding a two-tailed p-value of **??**. In this comparison, the root mean squared error ratio favours the Generic workflow ($\mathrm{RMSE_L/RMSE_G} \simeq$**??**).

Descriptively, we observe that the Legacy level of the template variable introduces a smoothness reduction (SCF of **??**), while the Legacy level of the workflow variable introduces a smoothness gain (SCF of **??**). Further, we note that there is a strong variance increase for the Legacy workflow (**??**-fold given the Legacy template and **??**-fold given the Generic template).

Given the break-up by contrast shown in **??**, we see only very weak effect sizes for the contrast variable (SCF of **??**) and the contrast-template interaction (SCF of **??**).

Functional Analysis

Functional analysis is a frequently used avenue for preprocessing QC. Its viability derives from the fact that the metric being maximized in the registration process is not the same output metric as that used for QC. The method is however primarily suited to examine workflow effects in light of higher-level applications, and less suited for wide-spread QC (as it is computationally intensive and only applicable to stimulus-evoked functional data). Additionally, functional analysis significance is documented to be sensitive to data smoothness [?], and thus an increased score on account of uncontrolled smoothing can be expected. For this analysis we compute the Mean Significance (MS), expressing the significance detected across all voxels of a scan.

As seen in ??, MS is sensitive to the workflow (??), but not to the template (??), nor the interaction of both factors (??).

The performance of the SAMRI Generic workflow (with the Generic template) differs significantly from that of the Legacy workflow (with the Legacy template) in terms of MS, yielding a two-tailed p-value of ??.

Descriptively, we observe that the Legacy level of the workflow variable introduces a notable significance increase (MS of ??), while the Legacy level of the template variable (MS of ??), and the interaction of the Legacy template and Legacy workflow (MS of ??) introduce no significance change. Furthermore, we again note a variance increase in all conditions for the Legacy workflow (??-fold given the Legacy template, and ??-fold given the Generic template).

With respect to the data break-up by contrast (??), we see no notable main effect for the contrast variable (MS of ??) and no notable effect for the contrast-template interaction (MS of ??).

Functional analysis effects can further be inspected by visualizing the statistic maps. Second-level t-statistic maps depicting the CBV and BOLD omnibus contrasts (common to all subjects and sessions) provide a succinct overview capturing both amplitude and directionality of the signal (??). Crucial to the examination of registration quality and its effects on functional read-outs is the differential coverage. We note that the Legacy workflow induces coverage overflow, extending to the cerebellum (????????), as well as to more rostral areas when used in conjunction with the Legacy template (????). Separately from the Legacy workflow, the Legacy template causes acquisition slice misalignment (????????). Positive activation of the Raphe system, most clearly disambiguated from the surrounding tissue in the BOLD contrast, is notably displaced very far caudally by the joint effects of the Legacy workflow and the Legacy template (??). We note that processing with the Generic template and workflow (????), does not show issues with statistic coverage alignment and overflow.

??

Variance Analysis

??

An additional way to assess preprocessing quality focuses on the robustness to variability across repeated measurements, and whether this is attained without overfitting (i.e. compromising physiologically meaningful variability). The core assumption of this analysis of variance is that adult mouse brains in the absence of intervention retain size, shape, and implant position throughout the 8 week study period. Consequently, when examining similarity scores of preprocessed scans with respect to the target template, more variation should be found across levels of the subject variable rather than session variable. This comparison can be performed using a type 3 ANOVA, modelling both the subject and the session variables. For this assessment we select three metrics 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).

?? renders the similarity metric scores for both the SAMRI Generic and Legacy workflows (considering only the matching workflow-template combinations). The Legacy workflow produces results which show a higher F-statistic for the session than for the subject variable: CC (subject: ??, session: ??), GC (subject: ??, session: ??), and MI (subject: ??, session: ??).

The Generic SAMRI workflow shows a reversing trend. Resulting data F-statistics are consistently higher for the subject variable than for the session variable: CC (subject: ??, session: ??), GC (subject: ??, session: ??), and MI (subject: ??, session: ??).

Discussion

The workflow and template design presented herein offer significant advantages in terms of reducing coverage overestimation, uncontrolled smoothness, and guaranteeing session-to-session consistency. This is most clearly highlighted by Volume Conservation (??), Smoothness Conservation (??), and Variance Analysis (??), where the combined usage of the SAMRI Generic workflow and template outperforms all other combinations of the multi-factorial analysis. Increased region assignment validity is also revealed in a qualitative examination of higher-level functional maps (??), where only the combination of the Generic workflow and template provides accurate coverage of the sampled volume for both BOLD and CBV fMRI data. iThese benefits are robust to the functional contrast (??????), with the Generic workflow-template combination being less or equally susceptible to the contrast variable, when compared to the Legacy workflow-template combination. The performance of the Generic workflow is more consistent across all metrics, as demonstrated by notable reductions of the standard deviation for both VCS, SCF, as well as MS (??????).

Closer model inspection reveals that in addition to the processing factor, the template factor is also a strong source of variability. The Legacy template induces both a volume and a smoothness decrease beyond the original data values (????). This clearly indicates a whole-volume effect, whereby a target template smaller than the recoded brain size causes a contraction of the brain during registration, resulting both in a volume and a smoothness loss. This effect can also be observed qualitatively in ??. We thus highlight the importance of an appropriate template choice, and strongly recommend usage of the Generic template on account of its better scale similarity to data acquired in adult mice.

The volume conservation, smoothness conservation, and session-to-session consistency of the SAMRI Generic workflow and template combination are further augmented by numerous design benefits (????). These include increased transparency and parameterization of the workflow (which can more easily be inspected and further improved or customized), veracity of resulting data headers, and spatial coordinates more meaningful for surgery and histology.

Quality Control

A major contribution of this work is the implementation of multiple metrics providing simple, powerful and robust QC for registration performance (VCF, SCF, MS, and Variance Analysis) and the release of a dataset [?] suitable for such multifaceted benchmarking — including the analysis of session-wise and subject-wise variability.

The VCF and SCF provide good quantitative estimates of distortion prevalence. The analysis comparing subject-wise and session-wise variance is an elegant avenue allowing the operator to ascertain how much a registration workflow is potentially overfitting, by differentiating between meaningful (inter-subject) and confounding (inter-session) variability. These metrics are relevant to both preclinical and clinical MRI workflow improvements, and could themselves be further optimized.

Global statistical power is not a reliable metric for registration optimization. Regrettably, however, it may be the most prevalently used if results are only inspected at a higher level — and could bias analysis. This is exemplified by the positive main effect of the Legacy workflow seen in ??. In this particular case, optimizing for statistical power alone would give a misleading indication, as becomes evident when all other metrics are inspected.

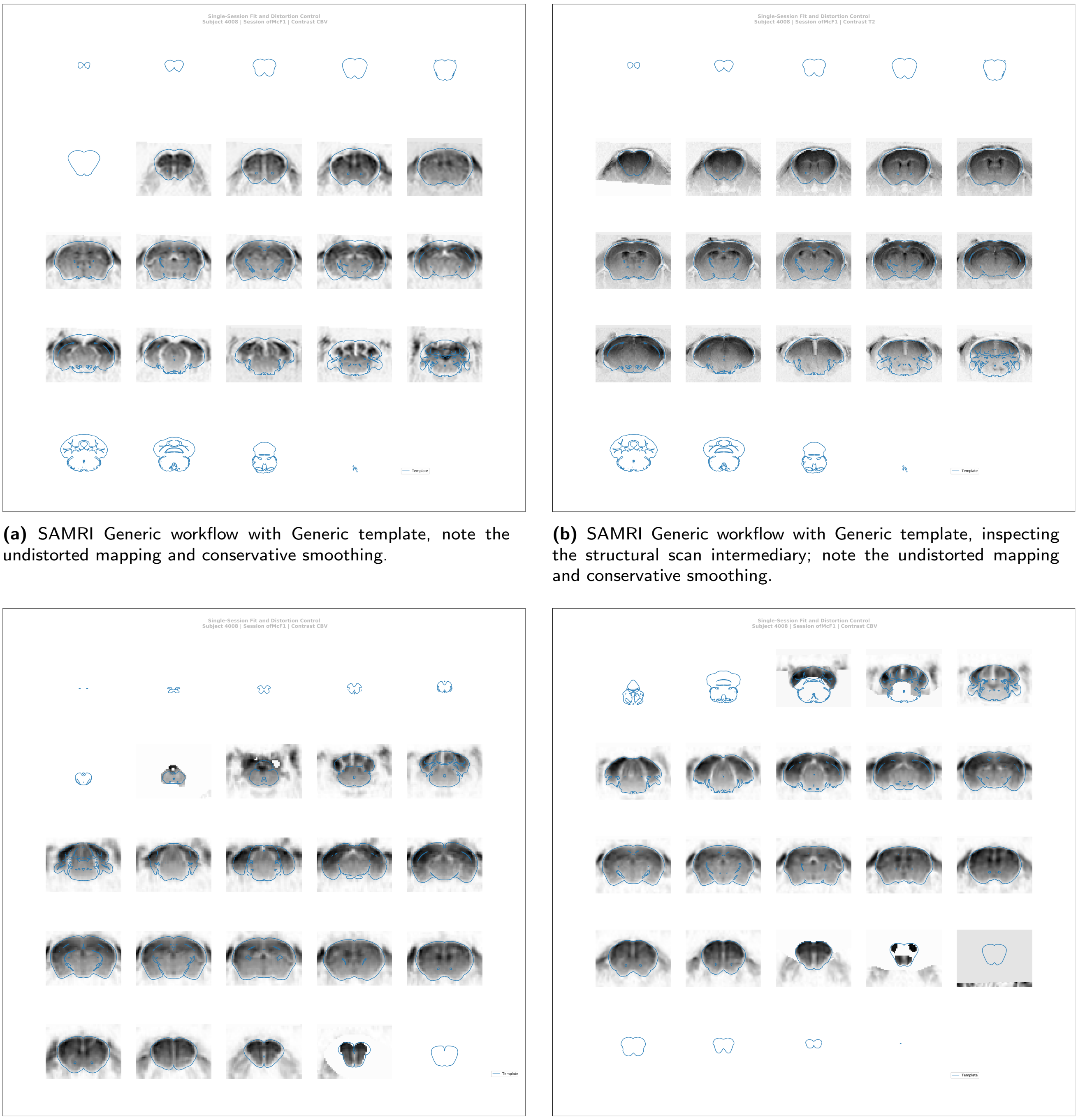
We suggest that a VCF, SCF and Variance based comparison, coupled with visual inspection of a small number of omnibus statistic maps is a feasible and sufficient tool for benchmarking workflows. We recommend reuse of the presented data for workflow benchmarking, as they include (a) multiple sources of variation (contrast, session, subjects), (b) functional activity with broad coverage but spatially distinct features, and (c) significant distortions due to implant properties — which are appropriate for testing workflow robustness. In addition to the workflow code [?], we openly release the re-executable source code [?] for all statistics and figures in this document. It is thus not just the novel method, but also the benchmarking process which is fully transparent and reusable with further data.

Conclusion

We present a novel registration workflow, entitled SAMRI Generic, which offers several advantages compared to the ad hoc approaches commonly used for small animal MRI. In depth multivariate comparison with a thoroughly documented Legacy pipeline revealed superior performance of the SAMRI Generic workflow in terms of volume and smoothness conservation, as well as variance structure across subjects and sessions. The metrics introduced for registration QC are not restricted to the processing of small animal fMRI data, but can be readily expanded to other brain imaging applications. The optimized registration parameters of the SAMRI Generic Workflow are easily accessible in the source code and transferable to any other workflows making use of the ANTs package. 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. Overall, we believe that using the SAMRI Generic workflow should facilitate and harmonize processing of mouse brain imaging data across studies and centers.

Supplementary Materials

??


Single-Session Fit and Distortion Control
Subject 4008 | Session ofMcF1 | Contrast CBV
Template
Single-Session Fit and Distortion Control
Subject 4008 | Session ofMcF1 | Contrast T2
Template
Single-Session Fit and Distortion Control
Subject 4008 | Session ofMcF1 | Contrast CBV
Template
Single-Session Fit and Distortion Control
Subject 4008 | Session ofMcF1 | Contrast CBV
Template

(a) SAMRI Generic workflow with Generic template, note the undistorted mapping and conservative smoothing.

(b) SAMRI Generic workflow with Generic template, inspecting the structural scan intermediary; note the undistorted mapping and conservative smoothing.

(c) SAMRI Legacy workflow with Legacy template, note the strong smoothing and the mapping distortion in the rostral and caudal areas.

(d) SAMRI Legacy workflow with Generic template, note the strong smoothing and the mapping distortion in the rostral and caudal areas.

Figure S1: The SAMRI Generic workflow induces less smoothness, and provides more accurate coverage. Depicted are automatically created operator overview graphics, which allow a slice-by-slice (spacing analogous to acquisition) inspection of the registration fit. This representation affords a particularly detailed view of the preprocessed MRI data, and highly accurate template contours.

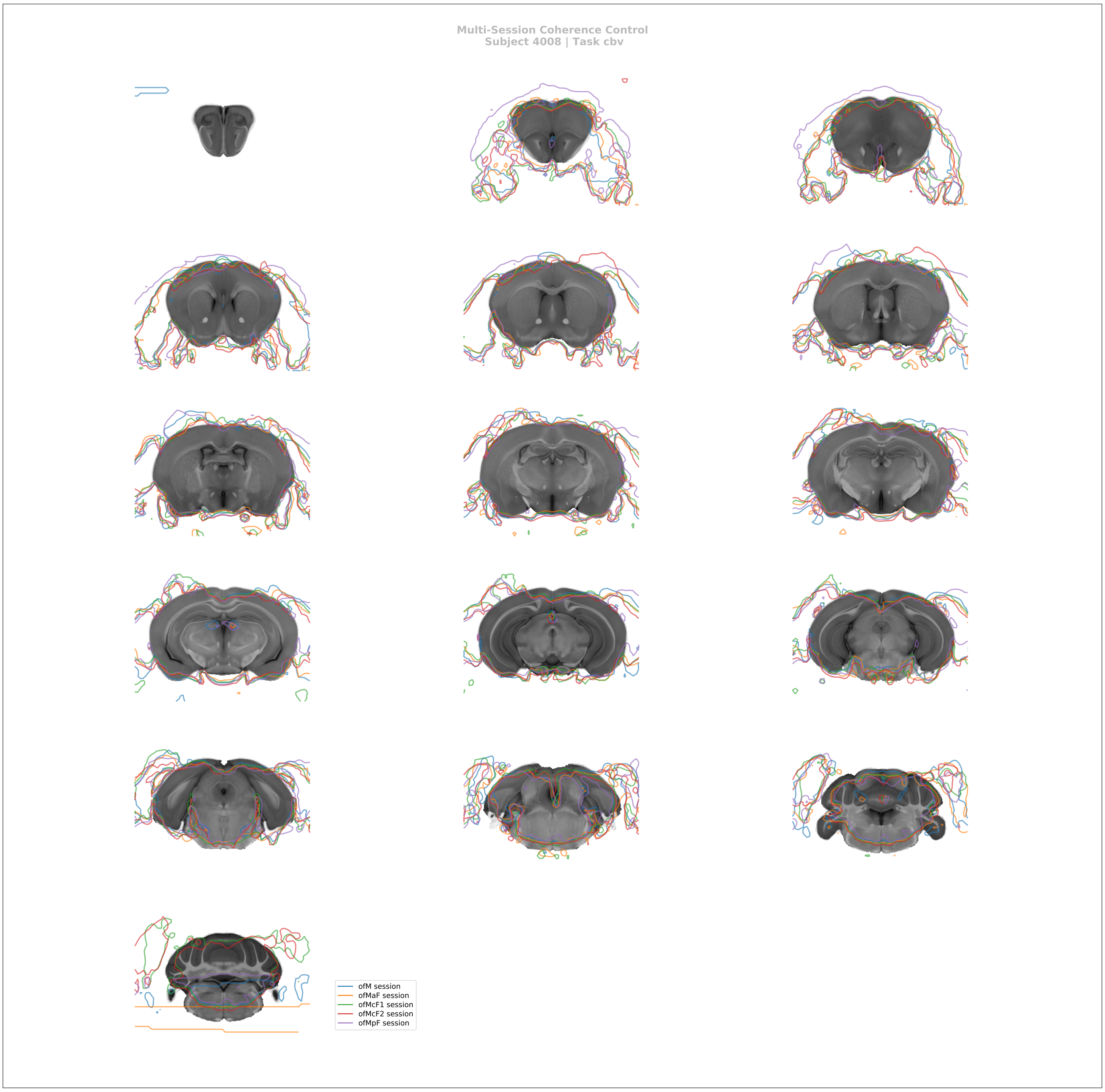
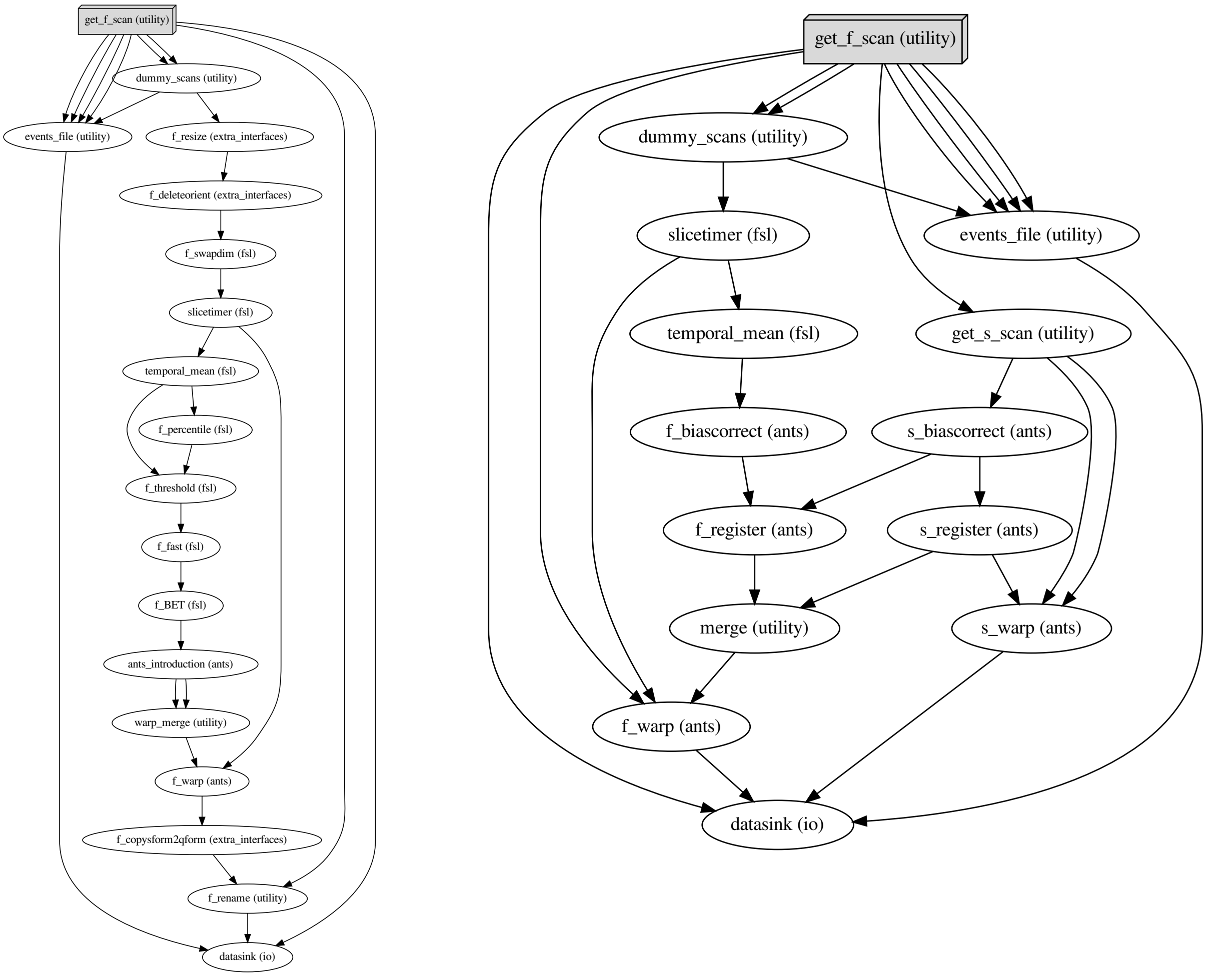


Figure S2: The SAMRI Generic workflow consistently maps high-salience features such as the implant site across sessions. Automatically created operator overview graphic, allowing a slice-by-slice (spacing analogous to acquisition) inspection of registration coherence. This representation permits a coarse assessment of registration consistency for multiple sessions — though at the cost of some clarity. Particularly, this visualization, allows an operator to track the position of high-amplitude fixed features across scans in order to ascertain coherence (similarly to what is automatically assessed by the Variance analysis of the session factor).



(a) "SAMRI Legacy" workflow, which is based on the `antsIntroduction.sh` function (and other functions with hard-coded parameters optimized for human brain registration), and also performs destructive affine manipulations.

(b) "SAMRI Generic" workflow, based on the `antsRegistration` function. The pipeline uses a higher-resolution structural scan intermediary for registration (note the two processing streams), which facilitates differential handling of anatomical variation and susceptibility artefacts. The function used is highly parameterized, and both of its instances — "s_register" and "f_register" — are supplied in the workflow with defaults optimized for mouse brain registration.

Figure S3: Directed acyclic graphs detailing the precise node names (as seen in the SAMRI source code) for the two alternate MRI registration workflows. The package correspondence of each processing node is appended in parentheses to the node name. The "utility" indication corresponds to nodes based on Python functions specific to the workflow, distributed alongside it, and dynamically wrapped via Nipype. The "extra_interfaces" indication corresponds to nodes using explicitly defined Nipype-style interfaces, which are specific to the workflow and distributed alongside it.