

Fmriprep-next: Preprocessing as a fit-transform model

Poster No:

1240

Submission Type:

Abstract Submission

Authors:

Christopher Markiewicz¹, Mathias Goncalves¹, Ma Feilong², Lea Waller³, John Kruper⁴, Robert Smith⁵, Russell Poldrack¹

Institutions:

¹Stanford University, Stanford, CA, ²Dartmouth College, Hanover, NH, ³Charité Universitätsmedizin Berlin, Berlin, Berlin, ⁴University of Washington, Seattle, WA, ⁵Florey Institute of Neuroscience and Mental Health, Melbourne, VIC

First Author:

Christopher Markiewicz, PhD

Stanford University
Stanford, CA

Co-Author(s):

Mathias Goncalves
Stanford University
Stanford, CA

Ma Feilong
Dartmouth College
Hanover, NH

Lea Waller
Charité Universitätsmedizin Berlin
Berlin, Berlin

John Kruper
University of Washington
Seattle, WA

Robert Smith
Florey Institute of Neuroscience and Mental Health
Melbourne, VIC

Russell Poldrack
Stanford University
Stanford, CA

Introduction:

The value of publicly shared neuroimaging data depends on the level of processing applied to the data. While raw data provide the greatest opportunity for asking novel questions, each step of processing left to secondary researchers is a potential source of analytical variation that can lead to conflicting results from the same source data [NARPS]. A researcher wishing to share the data they have collected can reduce sources of variability in downstream analyses by providing a canonical set of preprocessed data for reuse [HCPPipelines].

Publication of data that have been resampled into several spaces can enable different analyses while limiting analytical variability, but this requires significantly more storage and bandwidth. Generation of many derivatives may also be inefficient on shared, high-performance computing systems suited to computationally intensive tasks with relatively little storage use. It would thus be beneficial to calculate and distribute a compact set of preprocessing derivatives that permit the remaining derivatives to be generated cheaply and deterministically at the time of analysis.

Here we present recent changes in the architecture of fMRIPrep, a preprocessing pipeline for functional MRI, which split processing into "fit" and "transform" workflows. The "fit" workflows perform expensive, potentially nondeterministic steps, such as segmentation, registration and surface reconstruction, while the "transform" workflows combine raw files and derivative files to produce other derivatives, for example, a BOLD series resampled into MNI space.

The computationally expensive "fit" stage can be executed just once, with resulting derivatives being small and therefore easy to distribute. Pre-processed fMRI data in any desired target space can then be generated from the original fMRI data and the derivatives generated by the "fit" stage with minimal additional computational cost.

Methods:

The fit-transform architecture has been published in version 23.2.0a2. To test the impact of the described changes, fMRIPrep 23.2.0a2 and the prior release, 23.1.4, were run on two subjects from two different datasets.

Dataset A: 6 T1-weighted, 3 T2-weighted scans, 2 phase-difference fieldmaps, 4 single-echo BOLD runs with 195 volumes, and 1 single-band reference volume per BOLD series.

Dataset B: 2 T1-weighted scans, 6 spin-echo fieldmaps, and 8 single-echo BOLD runs with varying lengths, for a total of 4274 volumes.

The commands tested requested outputs registered to MNI152NLin2009cAsym volumetric template and the fsLR "grayordinate"

template. All processes were run on a single, 20-core Intel i9-10900 2.8GHz processor.

Results:

Table 1 compares runtimes and the sizes/counts of files written to disk by each version. Running fit-only workflows resulted in a decrease in runtime by 34-66% compared to the previous version. Scratch storage usage was reduced by 94-98% in total size and 84-92% in file count, and output storage usage was reduced by 73-89% in total size and 27-28% in file count.

Running fit and transform workflows resulted in a decrease in runtime of 25-52%. Scratch storage usage was reduced by 53-64% in total size and 72-87% in file count. Output usage increased by 177-188% in total usage and 17-22% in file count.

Dataset	Version/Mode	Runtime	Scratch Size	Scratch Files	Output Size	Output files
A	23.1.4	2h24m	54.8GB	36.8K	2.30GB	176
A	23.2.0a2 / fit	1h35m	2.91GB	5.89K	602MB	128
A	23.2.0a2 / fit+transform	1h47m	19.8GB	10.0K	6.37GB	206
B	23.1.4	4h25m	121GB	157K	5.10GB	286
B	23.2.0a2 / fit	1h29m	1.88GB	12.0K	543MB	206
B	23.2.0a2 / fit+transform	2h7m	56.5GB	19.8K	14.7GB	348

Table 1: Comparison of runtimes and disk usage between fMRIPrep version 23.1.4 and 23.2.0a2. "fit" mode included only fit workflows and their outputs; "fit+transform" includes all outputs, with rough parity with 23.1.4 outputs. All processes were completed independently, with no shared CPU, memory or files.

Conclusions:

Here we describe changes which result in a significant decrease in computational time and storage utilization for researchers interested in separating "fit" and "transform" portions of their workflows. We expect this to improve the efficiency of resource utilization for large scale data processors.

We also anticipate that these changes will simplify the process of resolving errors in preprocessing, as errors of fit and transformation can be addressed separately, and researchers will have the option of providing alternative fit results to be used in transformation. At the same time, the full workflow continues to provide the full range of derivatives that make fMRIPrep an attractive option.

Modeling and Analysis Methods:

Motion Correction and Preprocessing ²

Neuroinformatics and Data Sharing:

Workflows ¹

Keywords:

- Computing
- Data analysis
- FUNCTIONAL MRI
- Informatics
- Open-Source Software
- Workflows

¹²Indicates the priority used for review

Abstract Information

My abstract is being submitted as a Software Demonstration.

No

Please indicate below if your study was a "resting state" or "task-activation" study.

- Resting state
- Task-activation

Healthy subjects only or patients (note that patient studies may also involve healthy subjects):

Healthy subjects

Are you Internal Review Board (IRB) certified? Please note: Failure to have IRB, if applicable will lead to automatic rejection of abstract.

No

Was any human subjects research approved by the relevant Institutional Review Board or ethics panel? NOTE: Any

human subjects studies without IRB approval will be automatically rejected.

Not applicable

Was any animal research approved by the relevant IACUC or other animal research panel? NOTE: Any animal studies without IACUC approval will be automatically rejected.

Not applicable

Please indicate which methods were used in your research:

Functional MRI

Structural MRI

For human MRI, what field strength scanner do you use?

1.5T

3.0T

Which processing packages did you use for your study?

AFNI

FSL

Free Surfer

Other, Please list - ANTs, Nipype, nitransforms

Provide references using author date format

TBD

UNESCO Institute of Statistics and World Bank Waiver Form

I attest that I currently live, work, or study in a country on the UNESCO Institute of Statistics and World Bank List of Low and Lower-Middle Income Countries list provided.

No