

1 We thank all the reviewers for the responses and detailed comments. After reading the feedback, we realized that some
2 parts of the proposed method might not be explained enough, which might make it difficult to appreciate some of the
3 motivations and novelties of this paper. We hope that this response may clarify any misunderstanding, and we will
4 revise the article accordingly.

5 **SRM versus SSTL:** The first difference between the earlier SRM versus our SSTL lies in defining the shared space.

6 SRM uses $\min \sum_{s=1}^{S_d} \|\mathbf{X}^{(d,s)} - \mathbf{R}^{(d,s)} \mathbf{G}^{(d,S_d)}\|$ as the objective function, where $\mathbf{G}^{(d,S_d)} = \sum_{s=1}^{S_d} \left(\mathbf{R}^{(d,s)}\right)^\top \mathbf{X}^{(d,s)}$

7 and $\left(\mathbf{R}^{(d,s)}\right)^\top \mathbf{R}^{(d,s)} = \mathbf{I}$. Instead, SSTL utilizes $\min \sum_{s=1}^{S_d} \|\mathbf{X}^{(d,s)} \mathbf{R}^{(d,s)} - \mathbf{G}^{(d,S_d)}\|$ as objective function, where

8 $\left(\mathbf{G}^{(d,S_d)}\right)^\top \mathbf{G}^{(d,S_d)} = \mathbf{I}$. Note that both the site-dependents ($\mathbf{G}^{(d,S_d)}$) and the global shared space (\mathbf{W}) in SSTL are

9 orthonormal; thus, transformation for each site ($\mathbf{W} \mathbf{G}^{(d,S_d)}$) is also orthonormal. Like [1], these transformation matrices

10 only apply standard rotations on the neural responses and will preserve the general shape of the data distribution during

11 the transformation procedure. Empirical studies in [3] also showed that the original forms of SRM and HA (*i.e.*, the

12 shared space) can rapidly reduce the performance of multi-site fMRI analysis. Moreover, SRM uses a probabilistic,

13 iterative optimization approach that may coverage to a different $\mathbf{G}^{(d,S_d)}$ in each independent run. We instead propose a

14 single-iteration optimization approach that our empirical studies demonstrate to be more time-efficient and robust.

15 **Stacking across subjects in equations (6-7):** Yes, this subject ordering can matter, but this is fairly standard — *i.e.*,

16 this is true for incremental PCA, SRM, or any stochastic algorithm. Further, our empirical results, over many datasets,

17 demonstrate that this approach works effectively.

18 **Learning procedure:** We used a scheme similar to the one proposed in [2–3] for evaluating all transfer learning

19 approaches described in this paper. In SSTL, we first compute the unsupervised site-dependent $\mathbf{G}^{(d,S_d)}$, from the data,

20 BUT NOT THE LABELS, for all sites. Note this is similar to the procedures used in learning any classical functional

21 alignment, such as SRM and HA. For classifying a subject in site d , we then use the labeled data from other $d - 1$ sites

22 to find the global shared space \mathbf{W} , then train the classifier — *n.b.*, using nothing from the d -th site. Hence, we never

23 use any labels from the d -th site, when computing the labels for those d -th site subjects. Like Westfall *et al.* 2017,

24 we also used the standard learning procedure, *i.e.*, using a shuffled form of neural responses for training the classifier

25 (not the temporally aligned version). This is currently in Sec 4 (lines 213–223); the actual algorithm appears in the

26 Supplementary Material. The revised version will explicitly summarize the entire training and performance processes.

27 **Reviewer 1:** Thank you for your insightful comments.

28 1) SSTL uses a two-step procedure for analyzing multi-site fMRI datasets. The primary objective functions are based

29 on equations (4) and (11) on pages 3 and 5. To define (4), we propose (1) as the appropriate form for generating the

30 site-dependent feature space $\mathbf{G}^{(d,S_d)}$. We learn this using the regularized projection matrix in (2–3). Lemma 1 proves

31 that we can calculate a regularized version of $\mathbf{G}^{(d,S_d)}$ by substituting (2) in (1). Note that (1–3) are all involved in

32 Lemma 1’s (4). In other words, the regularization is defined for (2), not (1), and appears in (4) after we use (2) to

33 estimate the site-dependent in (1). We will re-structure Sec 3.1, to better show this flow.

34 2) We NEVER said that *if $V \gg T_d$, then the scatter matrices are full rank*. Instead, we said that ‘scatter matrices

35 $\mathbf{X}^{(s,d)} \left(\mathbf{X}^{(s,d)}\right)^\top$ will be singular and non-invertible’, which means these matrices are NOT full rank. The papers [3–5,

36 19] explains that $V \gg T_d$ implies the singular scatter matrices.

37 3) Lines 172 and 174 show how $\bar{\mathbf{g}}_t$ use θ_2 , and \mathbf{q}_t , which in turn uses θ_1 .

38 4) Thanks for the references; the revised manuscript will cite those papers — *i.e.*, Mensch 2017 and Westfall 2017.

39 **Reviewer 2:** Thanks for the very useful feedback! The revised version will address all of your suggestions in the

40 "Additional feedback".

41 **Reviewer 3:** Thanks for the great comments! We will provide a better structure for presenting the proposed method

42 and the results in the camera-ready.

43 **Reviewer 4:** Thanks for the wonderful comments. The revision will address all of those comments. Indeed, SSTL

44 can be applied to multi-site fMRI datasets with any resolution. Further, we will provide the MNI results after feature

45 selection and address the Clarity and Weakness sections’ questions.

46 Thanks to all of your insightful comments, our paper now better shows that SSTL provides an effective way to analyze

47 multi-site fMRI data. We anticipate that this approach can be used in various mental health applications, and will

48 contribute to techniques that can help save people’s lives. We hope that the reviewers and Area Chair agree, and will

49 also support publishing this paper.