We thank the reviewers for their insightful and constructive feedback. Our main contributions are (1) providing a principled model and learning-based method for building deformable templates, (2) extending this to learn *conditional* templates, and (3) demonstrating its usability in a series of experiments, with a focus application of neuroimaging.

Regarding Reviewer 1 and 3's note about *code release*: we **will release code**, model weights and atlases. Perhaps we misunderstood the reproducibility checklist, we meant that we did not provide code *at submission time*. This is because it may break anonymity since the code is deeply integrated with our registration library that we previously open sourced.

We agree with comments from Reviewers 1 and 2 about the importance of neuroimaging experiments. Due to space constraints, we condensed these experiments in the main text, but expanded them in *Supplemental Sec. 1.4: Neuroimaging Analysis*. Specifically, we computed a baseline template using a small dataset and state-of-the-art method (as Reviewer 1 suggests), and demonstrated significantly improved performance of our method compared to this baseline. We also showed example deformations, and segmentation maps overlayed on conditional and unconditional templates. We will expand the description of these experiments in the supplementary, and add detail to the main text.

Reviewer 1. We will release code, please see our explanation above.

We agree that our work is not novel in *image classification*, but that is not the topic of our work. Based on the suggested citations, we believe that perhaps the reviewer is referring to one of the possible downstream applications: learning image representations up to a deformation. However, we emphasize that (1) this is not what our main contribution, and (2) our goal there was simply to illustrate that learning with a deformation-based loss yields *different* representations than other reconstruction losses. We believe that neither of the suggested citations is related to this message.

In our paper, "image registration" is one of several related works sections, to which we devote half a page and 40 citations. Unfortunately, due to space limitations, we could not provide more background about this topic.

We agree that our main application focus is neuroimaging, but we envision several application domains. Because of this, we demonstrate several characteristics and uses of our method in the results section. Please see lines 7-12 above, about experimental results – while condensed in the main text, we expanded the analysis in the supplementary material.

Reviewer 2. We provide extended architecture details in the *Supplemental Section 1.1: Architectures*. Due to space limitations, and since the architecture is not our focus, we chose to not include it in the main text.

The MNIST experiments were included to demonstrate our method's potential, but we agree that neuroimaging experiments are more important. Please see lines 7-12 above about extended MRI results in supplementary material.

We will clarify in the paper that our conditional method variant does not provide *multiple* templates, but a single conditional template *function*. For a test subject and their attribute value, this function efficiently provides the appropriate template (e.g., 45-year old female brain), and no template selection is required. One utility of conditional templates is that the deformation fields are more informative, for example by eliminating variability from confounding attributes. The reviewer also asks about the *quality* of conditional templates. We found that compared to unconditional templates, conditional templates have similar texture characteristics, enable comparable registration accuracy (Dice score) in general, and improvements for some age groups. For this rebuttal we compared the data-matching loss term yielded via conditional and unconditional templates, and found no statistical difference (0.69 ± 0.03 for both).

We agree that understanding hyper-parameters is important. As we touch upon in *Supplemental Section 1.2*, the model hyper-parameters have intuitive effects on the sharpness of the templates, the spatial smoothness of the registration fields, and the quality of the alignments. We will clarify that they should be chosen based on the desired goal of a given task. For example, in neuroimaging, hyper-parameters could be determined by maximizing the highest anatomical overlap based on segmentations or landmarks in a validation set, as is often done in image registration.

Based on the review suggestion, we believe our model can also be used to estimate *unknown* attributes, which may require model selection. While outside the scope of the current paper, this is an interesting future direction. Along with emphasizing the expanded MRI results, we will make the minor corrections/clarifications, including enlarging Fig. 6.

Reviewer 3. We will release code, please see lines 4-6 above. Affine or rigid alignment (such as the rotations mentioned by the reviewer) are usually easy to solve in a pre-processing step, and are generally not of interest in downstream analysis (e.g., not anatomically meaningful). As described in the MRI pre-processing description, standard affine alignment was performed in all of our datasets before using our method.

As suggested, we will expand the motivation of the velocity field prior. For most applications, we often desire spatially smooth deformations to encourage anatomical consistency, leading to the Laplacian prior choice. In estimating templates that represent an anatomical *mean*, we expect the deformations to be generally small, leading to the first prior term.

We appreciate the suggestion to illustrate intermediate steps. While response space is limited, in the paper we will show intermediate features and velocity fields, which appear similar (but less smooth) to the displacement fields in *Sup. 1.4*.