We thank all the reviewers for their constructive and detailed feedback.

Goal: Automatically register scans of dressed and undressed 3D humans with the SMPL model.

Key novelty/ technical contributions. Our idea of implicitly representing SMPL as zero level set of a SDF is key for making the correspondences differentiable and is well received by the reviewers, [R1: "implicit... nice trick"], [R2: "several innovative ideas"], [R3: "interesting and inspiring."], [R4: "useful trick"]. Our second key contribution of formulating registration as a closed loop (scan-model-scan) is also appreciated by the reviewers, [R1: training without... labelled data... interesting contribution."], [R3: "closed loop... makes a lot of sense... implemented in a smart way.", "cycle idea is definitely useful"]. The reviewers also highlight that our approach is mathematically well grounded and sound [R2, R1: "solid mathematical framework"], and will have a positive impact outside this specific task [R2: "ideas are quite general... not limited to fitting of body scans only... adopted in other problems"]. We will release our code.

Results and writing. Reviewers acknowledge the superior performance of our method as compared to the existing methods. [R2: "results... quite impressive"], [R3: "accurate, stable registration... outperforming SOTA"]. We thank the reviewers for acknowledging our writing effort, [R1: "very well written.", "Reference to prior work is excellent."], [R2: "very well written... pleasure to read"], [R3: "appreciate rigorous mathematical notations"].

We address the reviewers concerns below:

Requirement of supervised warm-start. [R1, R2, R3, R4] We agree with the reviewers that our method requires a supervised warm-start and hence is not fully self-supervised (Abstract L20, Method L205, Experiments L238,267,271). We notice that the terms *self-* and *semi-* supervised learning are a bit confusing in the paper and we will correct this. However, to highlight that the requirements for supervised data are not very stringent, we show that the warm-start can be performed with *only* un-dressed scans, and we can self-supervise our method with dressed scans. Compared to a warm-start with labeled dressed scans, this results in only $0.24mm$ higher error, see Fig. 1. We note that acquiring supervised data for undressed shapes (in the form of SMPL meshes) is relatively easy, given existing datasets such as AMASS [Mahmood *et al.* ICCV'19], SURREAL [Varol *et al.* CVPR'17].

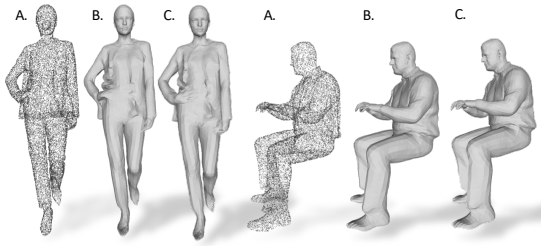

Figure 1: We show registration of subjects (A) from Fig. 4 (main paper) using undressed data for warm-start. We compare undressed warm-start (B) with dressed warm-start (C) and report negligible difference.

Robustness to noise and run-time. [R1] To show robustness to noise, we evaluate our method on FAUST [Bogo *et al.* CVPR'14] (scans containing noisy points, holes, self-intersections etc). We use our semi-supervised method (1K supervised and ∼1.6K unsupervised) and compare performance with [26] (fully supervised, trained on ∼200K meshes) in supp. mat. Table. 2. We outperform [26] and other competing methods. [R2] Correspondence prediction by NN takes < 1 sec., SMPL/SMPL+D fitting depends on the complexity of the pose/clothing etc. but can be done in under 15 mins. Since this step does not take up a lot of memory, we can fit ∼200 meshes in parallel (P100 GPU 22GB memory).

Clarifications. [R2, R3] We will better explain the diffusion of SMPL to $\mathbb{R}^3$. We will update Fig. 2 (main paper) and use a toy example to better convey the diffusion process and will update the text. [R2] We will better illustrate the problems with UV parametrization and add failure cases to our paper. [R4] We admit the paper is math heavy, but we found it essential to keep the notations rigorous. We will add more text around equations to intuitively explain the key ideas. [R4: "comparison in Fig 3"]. Our method and baselines [35, 6] require very different levels of 'good initialization'. We require small amount of supervised data at *training* time to warm-start our method, whereas [35, 6] require manual selection and 3D annotations for *every test input*. [R4: "Why predicting the latent parameters is difficult?"] This was our initial experiment as well (predict SMPL+D params. from scan) and as can be seen in Fig. 2, we could not get any reasonable results. There are no existing works that do this. We can add these results to supp. mat.

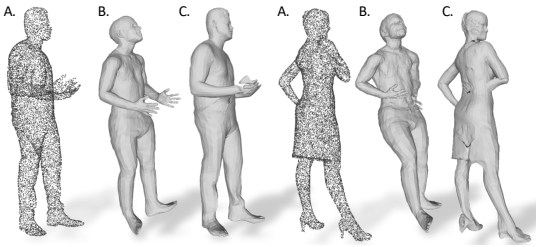

Figure 2: Given a dressed subject (A) we show that directly predicting SMPL+D parameters (B) performs significantly worse than our method (C).