Groundtruth    Ours    Neural3DMM [7]    Spectral (Cheb) [10, 26]    MoNet [25]    FeaST [33]


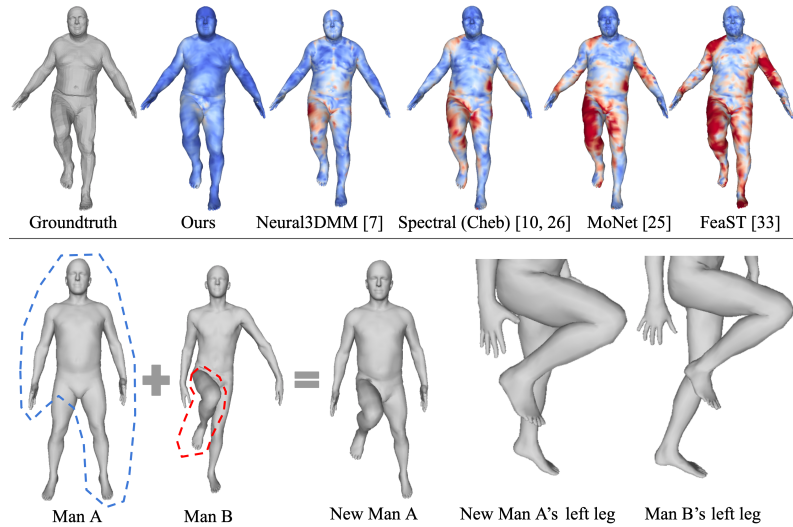Man A    Man B    New Man A    New Man A's left leg    Man B's left leg

We appreciate the reviewers' feedback. We will fix the typos and errors, add the ICCV Workshop 2019 paper (**R2**) within the related work and open-source our code. Below, we list reviewer concerns in bold and address them underneath.

**Concern over experimentation (R2 and R3)** For the concern of not enough comparisons with competitor methods, while it is true that we compare only with 2 other "entire" methods (Table 1.0), in fact we compare with 5 additional graph convolution networks in Table 1.2, resulting in a total of 7 comparisons. The distinction is because the methods in Table 1.2 come from publications which do not provide up and down-sampling schemes and therefore cannot perform the autoencoder task, so it is not possible to compare with them in an "entire" fashion. The only way to compare with them is to use their proposed convolution layers inside our pipeline, which is designed for the task. To alleviate the confusion for this, we will add citation links next to methods listed in Table 1.2, so that it is clear they are competitor methods. In addition, we will add error visualizations of these methods (Top half of added figure). Finally, we are also happy to add any additional comparisons with baselines or competitor methods the reviewers wish to see, as well as further interpolation, extrapolation, and analogy examples to the final appendix.

**"I would love to see some interpolation examples between humans..."(R3)** Bottom half of added figure shows identity of "Man A" while interpolating only the latent-node which corresponds to the foot area of "Man B". The result is Man A with a new foot pose, while keeping the rest of Man A the same. In particular, note that Man A's new leg pose is still "his" leg, and not "Man B's" leg. This shows that we can in fact perform reasonable pose interpolation even using different identities. We can add additional human interpolation to the final appendix if desired.

**"The vertices need to be ordered...it assumes to have fixed connectivity in input... [25] does not suffer from this problem...Can you think of any strategy to overcome these limits?" (R3)** It is true that the requirement of fixed-connectivity datasets is the main limitation of our method. However, using registration to get the same connectivity across dynamic meshes is well studied, and high accuracy correspondence can be achieved. Therefore, it is not necessary to solve the connectivity and shape problem simultaneously. While [25] gets around this, its accuracy and reconstruction fidelity is much worse than fixed-connectivity methods like ours or Neural3DMM [7] (see the added figure, top). It is our impression that there is a clear trade-off here between reconstruction quality and generalize-ability, and we are examining the other side of the tradeoff compared to [25]. As to strategies which allow our convolution method to extend to variant topology input while keeping the reconstruction fidelity high, we believe this is a quite difficult problem (middle of the tradeoff-curve) and will require several future research projects.

**"It looks like a min-min problem...and in general it is not directly optimizable." (R3)** Actually, it is not necessary to do any sort of alternating optimization. In fact, our formulation of convolution through the basis and coefficients is a fully differentiable operation, and so all parameters can be jointly optimized with standard SGD. We have also not encountered any issues with basis vectors going to 0 during training.

**"...[is] There is some advantage to use a linear combination of the learned basis? Also, have you ever had problems with the linear independence of the learned basis?" (R3)** The linear option is the straightforward choice. We have encountered no problems with linear independence of the learned basis, though it is possible to put in some type of regularization to prevent correlation between basis vectors. In early experiments, however, we have found adding regularization to the basis hurts performance.

**"...assumption that the connectivity is almost evenly distributed on all the graphs. How does the method perform in case of high-variance in the connectivity?" (R3)** The method has no requirement on the distribution of connectivity. The tree example (Figure 3) does not have an even distribution of connectivity, and the network is able to learn it well. We can add zoomed-in figures of the tree example to show it's uneven connectivity.

**Is the local neighborhood N defined based on Euclidean distance or geodesic distance? (R1)** It is defined on vertex connectivity neighborhood.

**"...should multiplications between W j and x ij be switched?" (R3)** Yes, thank you, it should be $W_j^T$