Appendix A: Choice of Distance Measure

As described in Section 3.1, we chose a weighted geodesic distance measure for determining the influence radius of a control handle. Empirically, this approach is better than straight-line Euclidean distances or unweighted geodesic distances because it better preserves both the spatial and topological relationships of the mesh. Especially, the topological relationship can be crucial in regions such as the face, where the upper and lower lip might be close together in space but should not move together for a small influence radius.

Additionally, we tested another way of determining the influence radius, based on a correlation prior. To do so, we precomputed the correlation $corr(v_i, v_j)$ between each vertex pair v_i and v_j. When integrating the weighted geodesic distance along the edges, we multiply this with an additional scaling of $1 - |corr(v_i, v_j)|$. With this scaling, not only are the topological and spatial relationships taken into account, but also their prior correlation based on the data. This results in more correlated vertices being covered first when increasing the radius of a single handle. Figure 16 illustrates this effect, showing how the influence region initially covers the forehead rather than the upper eyebrows, unlike the weighted geodesic approach in Figure 4. Also note how the nose is not covered at all since that deformation was very uncorrelated in the data and can only be moved by selecting a handle on the nose.

While the precomputed correlation could potentially simplify the task the model has to learn, we did not observe any improvements (or deficits) in performance. However, from a user perspective, this way of controlling the influence region offers an interesting property by more clearly visualizing which regions can or cannot be moved together. From a UI-perspective, the visualization of the attention weights can offer a similar feedback to the user to understand the influence of a control handle on the vertices. Nevertheless, since the computation of the correlation matrix is challenging for high-resolution meshes, we defaulted to the simpler weighted geodesic distance measure in CANRig. Yet, one can also expose this way of shaping the influence region as an additional option to the user at runtime, since CANRig is not limited to the masks it was trained with.

Appendix B: Detailed Equations

In this section, we present the full mathematical formulations used throughout this paper. We begin with the equations for the non-iterative model in Section 3.1, followed by the formulations for the iterative approach in Section 3.2.

$$q = \Phi_{\text{query}}(S_0) \qquad \in \mathbb{R}^{K\times N\times F_i} \tag{9a}$$
$$k = \Phi_{\text{key}}(p_h) \qquad \in \mathbb{R}^{K\times H\times F_i} \tag{9b}$$
$$v = \Phi^*_{\text{value}}(\Delta_h) \qquad \in \mathbb{R}^{K\times H\times F_o} \tag{9c}$$
$$A = q * k^T \qquad \in \mathbb{R}^{K\times N\times H} \tag{9d}$$
$$A_M = \sigma_M(A) = \frac{\exp(A)\odot M}{\sum_{l=1}^{L}\exp(A_h)M_h} \qquad \in \mathbb{R}^{K\times N\times H} \tag{9e}$$
$$z = A_M * v \qquad \in \mathbb{R}^{K\times N\times F_o} \tag{9f}$$
$$S_{pred} = S_0 + \Phi^*_{\text{output}}(z) \qquad \in \mathbb{R}^{N\times D} \tag{9g}$$

Figure 16: *Deformations using a single control handle (blue) with two varying user-defined editing regions, indicated by 0 1, that are additionally weighted based on a correlation prior. ©Disney/Pixar*

Here, F_i, F_o are the attention input and output latent dimensions, and K is the number of attention heads.

Similarly, the equations for the shape-preserving model are given by:

$$q = \Phi_{\text{query}}(S_0) \qquad \in \mathbb{R}^{K\times N\times F_i} \tag{10a}$$
$$k = \Phi_{\text{key}}(p_h) \qquad \in \mathbb{R}^{K\times H\times F_i} \tag{10b}$$
$$v = \Phi^*_{\text{value}}(\Delta_h, b_h) \qquad \in \mathbb{R}^{K\times H\times F_o} \tag{10c}$$
$$A = q * k^T \qquad \in \mathbb{R}^{K\times N\times H} \tag{10d}$$
$$A_{M,\beta} = \sigma_M(A;\beta) = \left[\frac{\exp(A)\odot M}{\sum_{l=1}^{L}\exp(A_h)M_h};\beta\right] \qquad \in \mathbb{R}^{K\times N\times H+1} \tag{10e}$$
$$z = A_{M,\beta} * [v;0] \qquad \in \mathbb{R}^{K\times N\times F_o} \tag{10f}$$
$$S_{pred} = S_B + \Phi^*_{\text{output}}(z, S_B) \qquad \in \mathbb{R}^{N\times D} \tag{10g}$$

Appendix C: Additional Results

Pose A to B First, we present additional visual results of the A to B experiment for our animated characters (see Figure 17). In all cases, the target pose B can be hit accurately. In some cases such as the one if fifth row, where the target pose is more extreme, the error might be higher. However, the residual error can be further improved by our iterative editing capabilities, as shown in the next section.

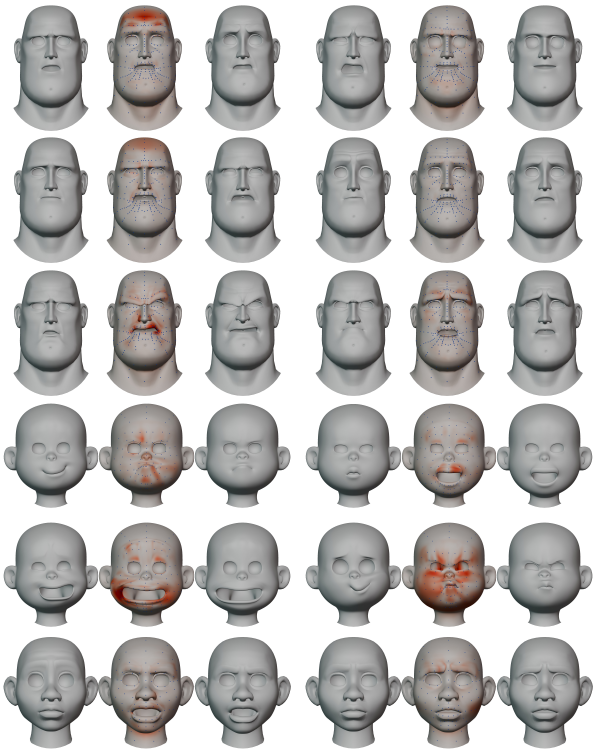

Figure 17: *Additional examples of the A to B test. The base shape is given in the first and fourth columns, the target pose in the third and sixth columns, and the reconstruction is given in the second and fifth columns. The error scale is from 0mm*  *10mm. ©Disney/Pixar*

Iterative Editing Secondly, we give additional visual results of the shape-preserving iterative editing in Figure 18). In all cases, we start from the neutral pose and reconstruct the target iteratively. By updating the base shape after each editing layer and bringing it closer to the target shape, the reconstruction error is reduced in each step. This is especially prominent for the second and fifth row that show more extreme target poses.

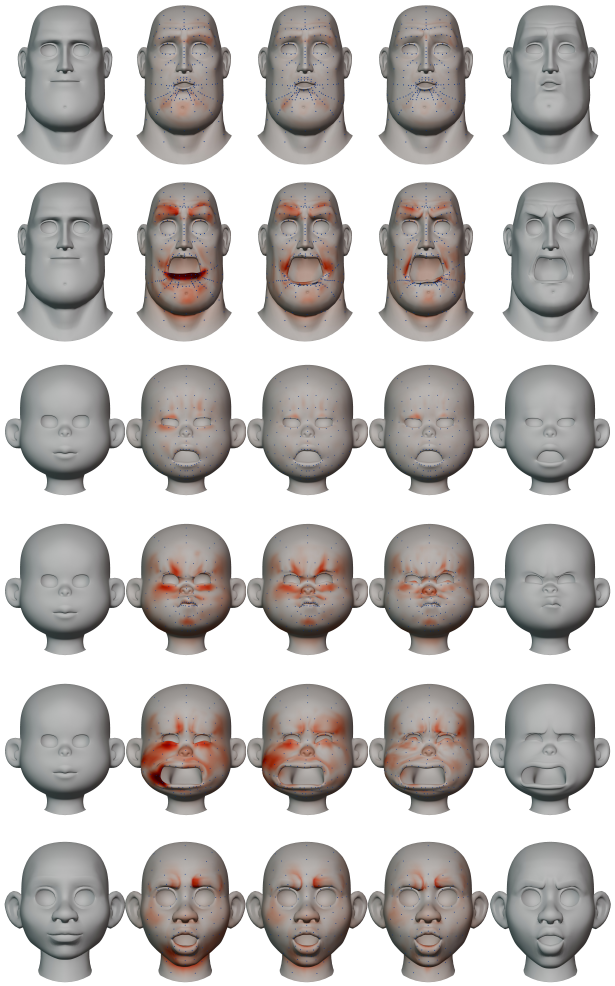

Figure 18: *Additional examples of iterative editing. Starting from the neutral face (column 1), our system predicts the target face (column 5). Column 2 shows the prediction with a single edit. By updating the base shape after each layer, additional edits can be applied, leading to a better reconstruction of the target face. The Euclidean reconstruction error relative to the target face is represented by color, using a scale from 0mm*  *10mm. ©Disney/Pixar*