LIM: Large Interpolator Model for Dynamic Reconstruction

Supplementary Material

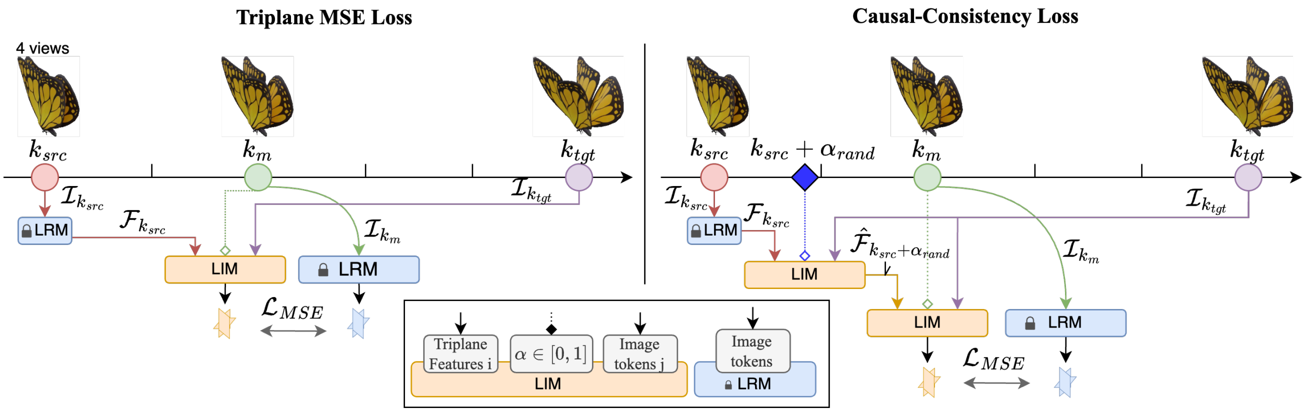

Figure 7. **LIM training losses**. (Left) The triplane MSE loss $\mathcal{L}_{\mathcal{T}}$ only supervises LIM on keyframes k_m. (Right) The causal consistency loss $\mathcal{L}_{\text{causal}}$ samples in-between keyframes with an additional forward-pass to LIM. Note that the second pass of LIM takes as input the intermediate features from LIM instead of the intermediate features from LRM.

A. Additional Evaluations

We recommend looking at the webpage in supplemental to see the video results. In particular, the webpage contains video result of RGB interpolation, XYZ canonical tracking, monocular reconstruction and mesh reconstruction.

B. Additional Method Insights

Weight Initialization. The composition of blocks in LIM and LRM is presented in Fig. 2. We initialize LIM with LRM to take advantage of the learned 3D intermediate representation. More specifically, the intermediate-features cross-attention layers are derived from the self-attention layers from LRM. Furthermore, the image cross-attention layers are initialized using the image cross-attention layers from LRM, and the self-attention layers are initialized from the self-attention layer of LRM.

Model size. We ablate the choice of the number of layers in Tab. 5. We observe that LIM accuracy is proportional to the number of blocks in the architecture. However, adding more blocks in LIM slows down the interpolation. We set $N_{layer} = 6$ as a good trade-off between speed and accuracy.

	PSNR ↑	PSNR$_{\text{FG}}$ ↑	LPIPS ↓
LIM- 3 layers	22.90	14.98	0.086
LIM- 8 layers	23.62	16.51	0.083
LIM	**23.58**	**16.44**	**0.083**

Table 5. **Performance as a function of # layers** reporting interpolation accuracy of LIM while varying the number of transformer blocks in the architecture.

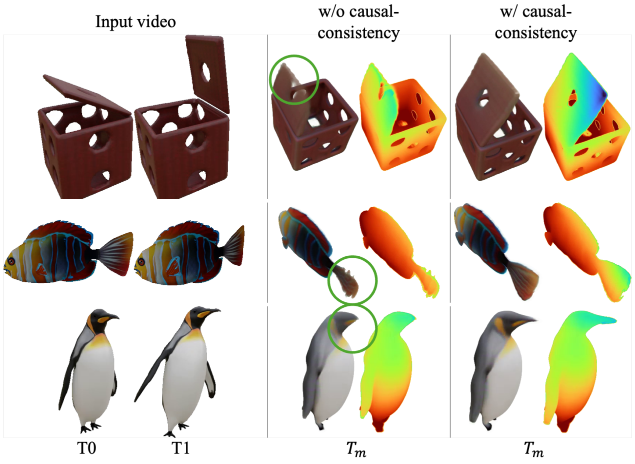

Figure 8. **Causal-loss ablation.** We show triplane interpolation result from LIM models trained either with the triplane MSE loss $\mathcal{L}_{\mathcal{T}}$ only, or with both $\mathcal{L}_{\mathcal{T}}$ and the causal-consistency loss $\mathcal{L}_{\text{causal}}$.

Causal consistency loss. We illustrate in Fig. 7 the behavior of the triplane MSE loss $\mathcal{L}_{\mathcal{T}}$ and the causal-consistency loss $\mathcal{L}_{\text{causal}}$ (see Sec. 3). $\mathcal{L}_{\mathcal{T}}$ involves a single pass of LIM and two passes of LRM, while $\mathcal{L}_{\text{causal}}$ involves 2 passes of LRM and 2 passes of LIM. Note that during LIM training, the weights of LRM are frozen. In practice, we discovered that the causal consistency loss was essential to achieve precise and accurate interpolation over a range of shapes and motions. We show interpolation results (in the same setting as Sec. 4.1) in Fig. 8, with a LIM model trained either with $\mathcal{L}_{\text{causal}}$ activated or deactivated.