

Alignment quality ( $\ell_2$ -distance)

2.0  
1.5  
1.0  
0.5  
0.0

Report-to-image Image-to-report

LoVT (Ours)

no global loss

