

For samples from the RE10K dataset, the three columns represent the decoded video, ground truth video, and rendered video, respectively.

For out-of-distribution samples, the four columns represent the decoded video, conditioned image, rendered video, and camera trajectory, respectively.

For samples from Worldscore, since the camera trajectory is fixed, we only show the generated decoded video.

We also present an example with a varying translation scale ranging from 1 to 4.

We present some examples with the same image but different camera conditions in the ablation folder, where the results without ReFL and with ReFL are compared.