

# VENI: Variational Encoder for Natural Illumination

## Supplementary Material

### A. Proof of SO(2)-rotation-equivariance

In section 3, we introduced an SO(2)-rotation-equivariant fully connected layer. Here we prove the SO(2)-rotation-equivariance of this layer.

Let  $R \in SO(2)$  be a  $2 \times 2$  rotation matrix and:

$$R' = \begin{pmatrix} R & 0 \\ 0 & \mathbf{I}_{c_{inv}} \end{pmatrix},$$

where  $\mathbf{I}_{c_{inv}}$  is the identity matrix of size  $c_{inv} \times c_{inv}$ . From the definition of  $R'$ , it follows that

$$R' \mathbf{X} = [R \mathbf{X}_{eq}, \mathbf{X}_{inv}] . \quad (12)$$

The intermediate values  $\mathbf{T}_{inv}$  and  $\mathbf{T}'_{inv}$  are invariant to the rotation:

$$\mathbf{T}_{inv} = [\mathbf{X}_{inv}, \mathbf{1}] \quad (13)$$

$$\mathbf{T}'_{inv} = [\mathbf{X}_{inv}, \|R \mathbf{X}_{eq}\|] = [\mathbf{X}_{inv}, \|\mathbf{X}_{eq}\|] \quad (14)$$

Therefore,  $Y_{inv}$  is also invariant to the rotation:

$$\mathbf{Y}_{inv,o,v} = \sum_{i=1}^{d_{in}} \sum_{k=1}^{c_{inv}+1} \mathbf{W}_{inv,o,v,i,k} \cdot \mathbf{T}'_{inv,i,k} + \mathbf{B}_{inv,o,v} \quad (15)$$

For the  $Y_{eq}$  the rotation commutes:

$$\mathbf{Y}'_{eq,o,v} = \sum_{i=1}^{d_{in}} \sum_{k=1}^{c_{inv}+1} \mathbf{W}_{eq,o,k,i} \cdot \mathbf{T}_{inv,i,k} \cdot (R \mathbf{X}_{eq})_{i,v} \quad (16)$$

$$= \sum_{i=1}^{d_{in}} \sum_{k=1}^{c_{inv}+1} \mathbf{W}_{eq,o,k,i} \cdot \mathbf{T}_{inv,i,k} \cdot \left( \sum_{t=1}^2 R_{v,t} \mathbf{X}_{eq,i,t} \right) \quad (17)$$

$$= \sum_{t=1}^2 R_{v,t} \sum_{i=1}^{d_{in}} \sum_{k=1}^{c_{inv}+1} \mathbf{W}_{eq,o,k,i} \cdot \mathbf{T}_{inv,i,k} \cdot \mathbf{X}_{eq,i,t} \quad (18)$$

$$= \sum_{t=1}^2 R_{v,t} \mathbf{Y}_{eq,o,t} \quad (19)$$

$$= (R \mathbf{Y}_{eq})_{o,v} \quad (20)$$

We have shown that:

$$f([R \mathbf{X}_{eq}, \mathbf{X}_{inv}], \mathbf{W}_{eq}, \mathbf{W}_{inv}, \mathbf{B}_{inv}) = [R \mathbf{Y}_{eq}, \mathbf{Y}_{inv}]$$

$$f(R' \mathbf{X}, \mathbf{W}_{eq}, \mathbf{W}_{inv}, \mathbf{B}_{inv}) = R' f(\mathbf{X}, \mathbf{W}_{eq}, \mathbf{W}_{inv}, \mathbf{B}_{inv})$$

The rotation matrix commutes with this SO(2)-equivariant fully connected layer and thus, we can conclude that our SO(2)-equivariant fully-connected layer has the desired SO(2)-equivariance and invariance properties.

### B. SO(2)-rotation-equivariant transformer

In section 4, we use an SO(2)-equivariant transformer for ablations. This transformer is built by replacing every Vector Neuron component of the VN-Transformer [1] with SO(2) equivariant layers. The VN-Linear layers are replaced with our SO(2)-equivariant fully connected layer, and the VN-ReLU layers, are replaced with our SO(2)-ReLU version, in which, the Vector Neuron version is applied in the equivariant dimensions  $\mathbf{X}_{eq}$  and the standard version in the invariant dimensions  $\mathbf{X}_{inv}$ :

$$SO(2)\text{-ReLU}([\mathbf{X}_{eq}, \mathbf{X}_{inv}]) = [\text{VN-ReLU}(\mathbf{X}_{eq}), \text{ReLU}(\mathbf{X}_{inv})] \quad (21)$$

### C. Additional Results

We present additional qualitative results for our model. Figures 7 and 8 demonstrate interpolations using latent optimization and the autoencoder pass for the reconstructions. Figure 9 shows a comparison of the quality of reconstructions between RENI++ and our model. For our model, we present reconstructions obtained using the autoencoder pass and direct latent optimization. This figure corresponds to Table 1. We also include an application of our model in Figure 10, which shows an inverse rendering optimization over an increasing Blinn-Phong specular term  $K_s$  from 0 to 1. Figures 11, 12 and 13 show image reconstructions using the autoencoder pass for ablations of our model, corresponding to Table 4. Figure 14 includes additional examples following Figure 5, showing interpolations between latent codes optimized to the same GT image.

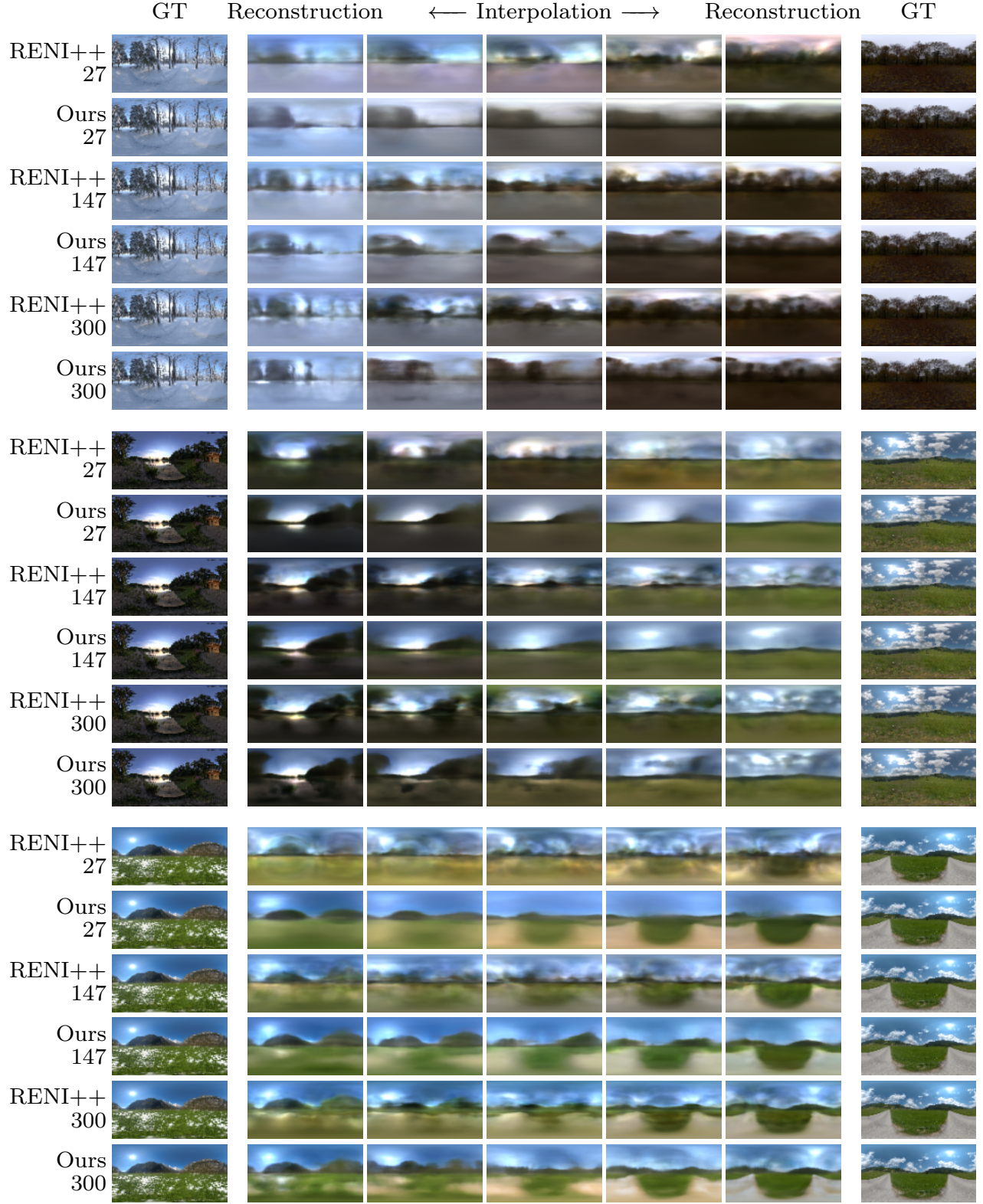


Figure 7. Interpolations using our model (with direct latent optimization) and RENI++ with different latent sizes. Three image pairs are interpolated: snow-forest, lake-field, mountain-road.

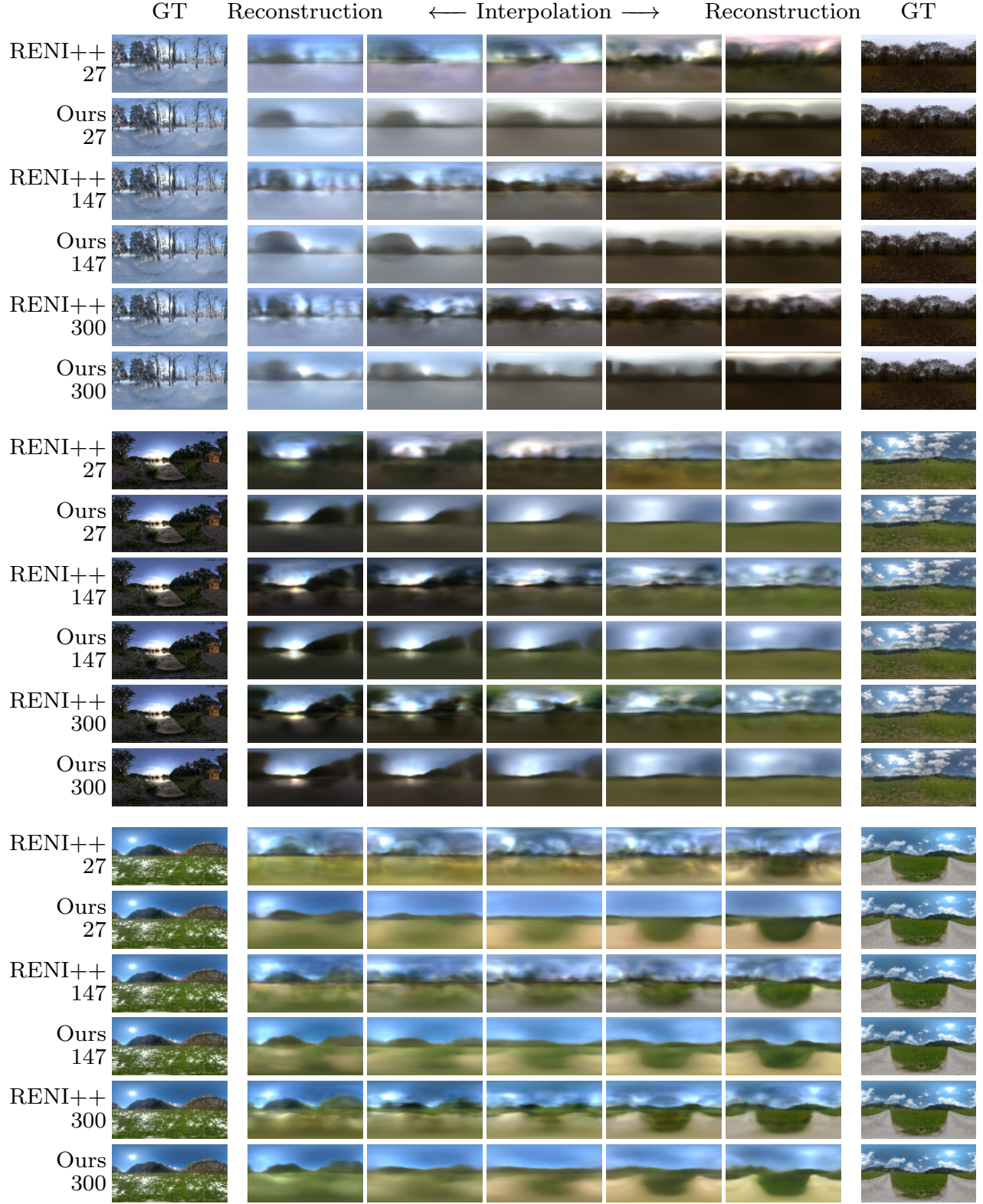


Figure 8. Interpolations using our model (autoencoder pass) and RENI++ with different latent sizes. Three image pairs are interpolated: snow-forest, lake-field, mountain-road.



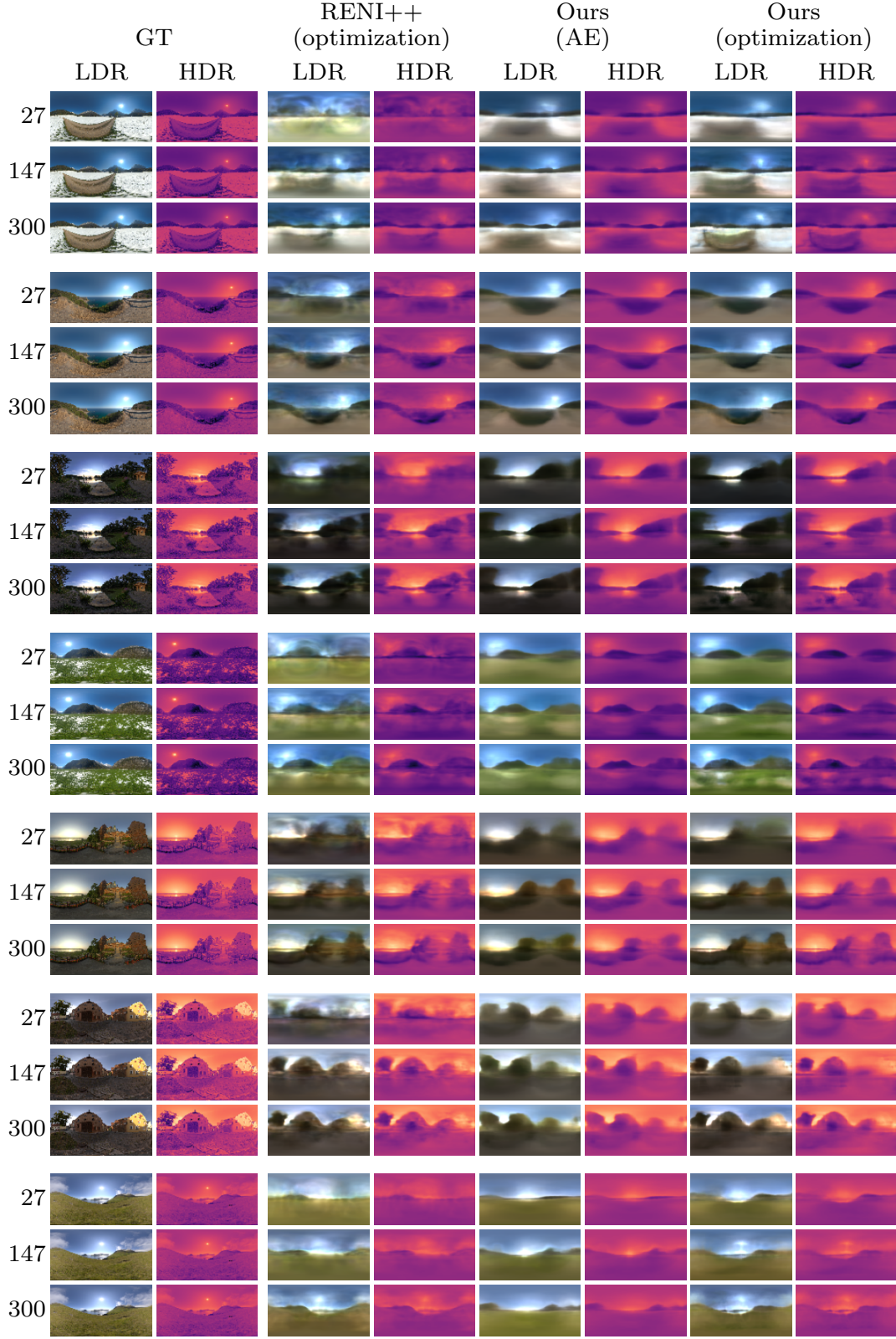


Figure 9. Corresponding figure to Table 1 comparing RENI++ and our model (autoencoder pass and direct latent code optimization), across varying latent dimensions  $D$  and target images, with reconstructions shown in LDR tone-mapped space and in log HDR space as heatmaps



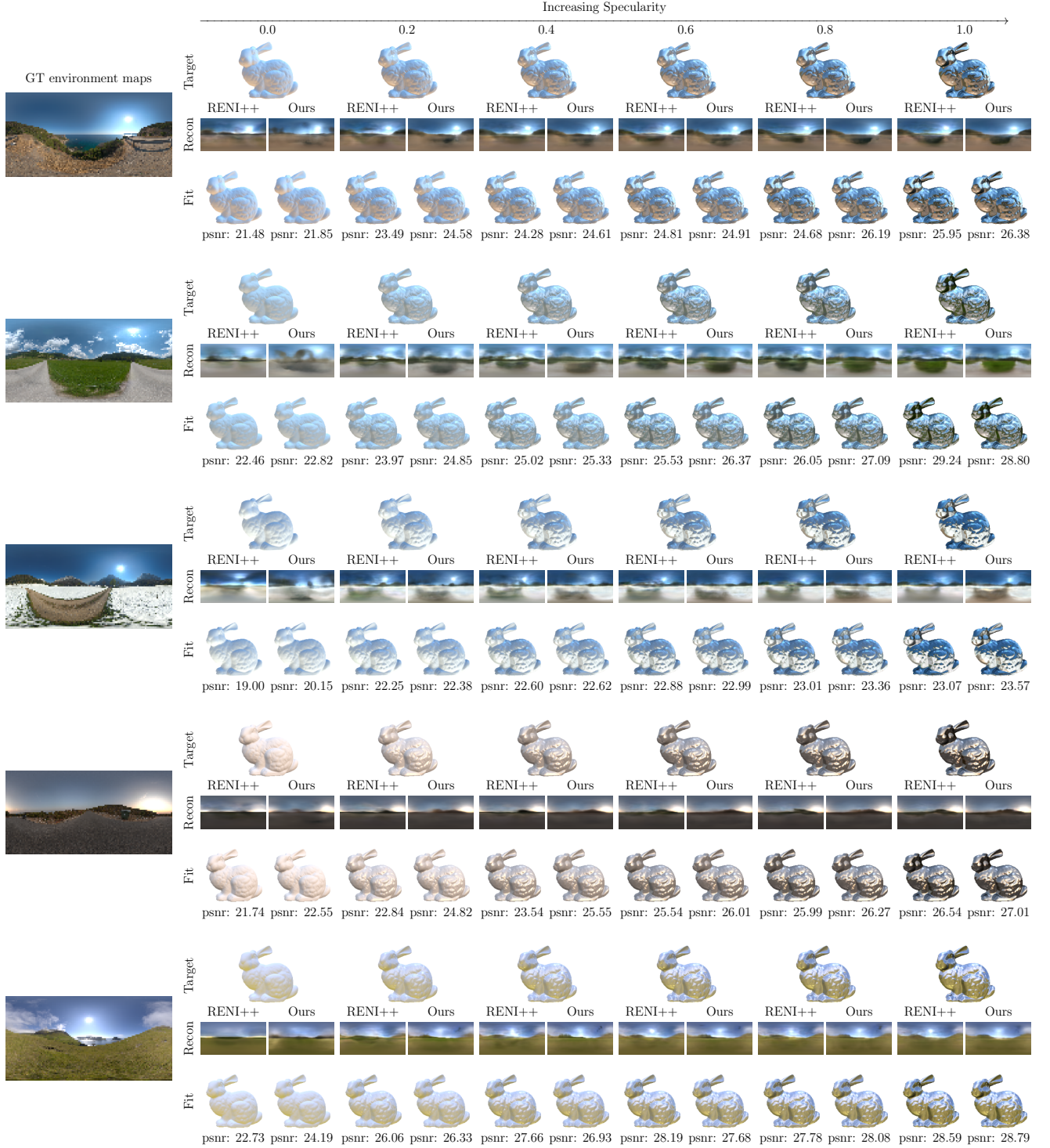


Figure 10. Further results of inverse rendering optimization using Blinn-Phong shading. We compare our model to RENI++, increasing the specular Blinn-Phong term  $K_s$  from 0 to 1. Both RENI++ and our model have a latent dimensionality of  $D = 300$ . Our model generally outperforms RENI++

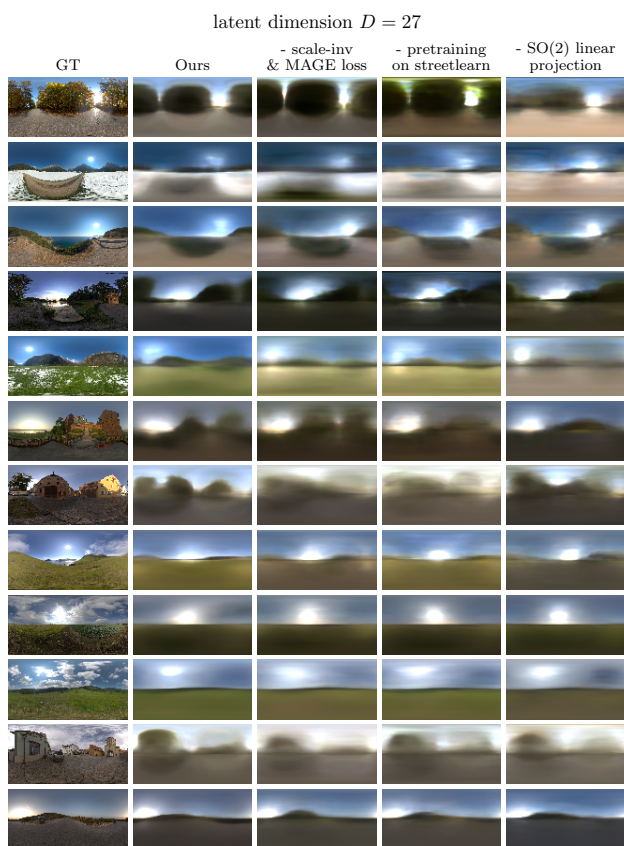


Figure 11. Corresponding figure to Table 4 comparing various ablations of our model, here with latent dimension  $D = 27$ .

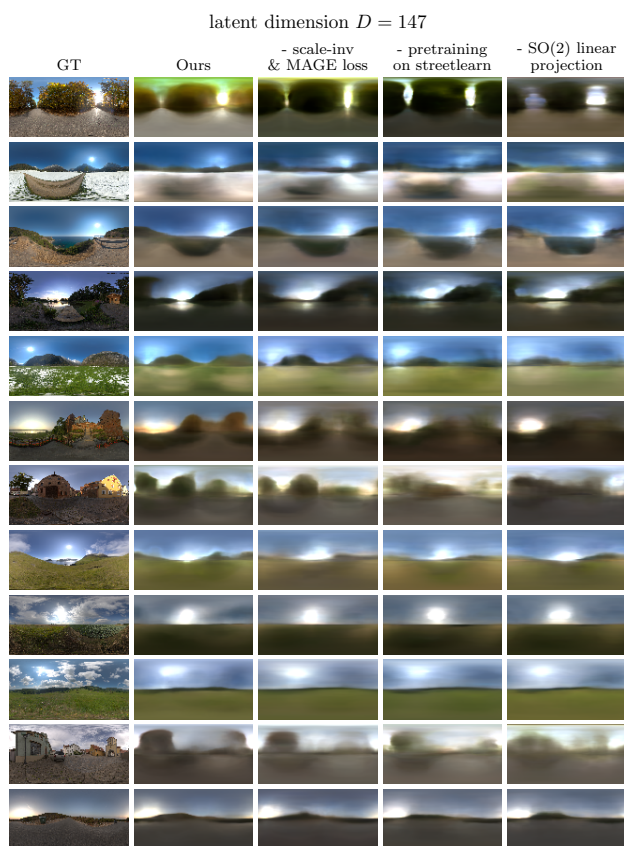


Figure 12. Corresponding figure to Table 4 comparing various ablations of our model, here with latent dimension  $D = 147$ .



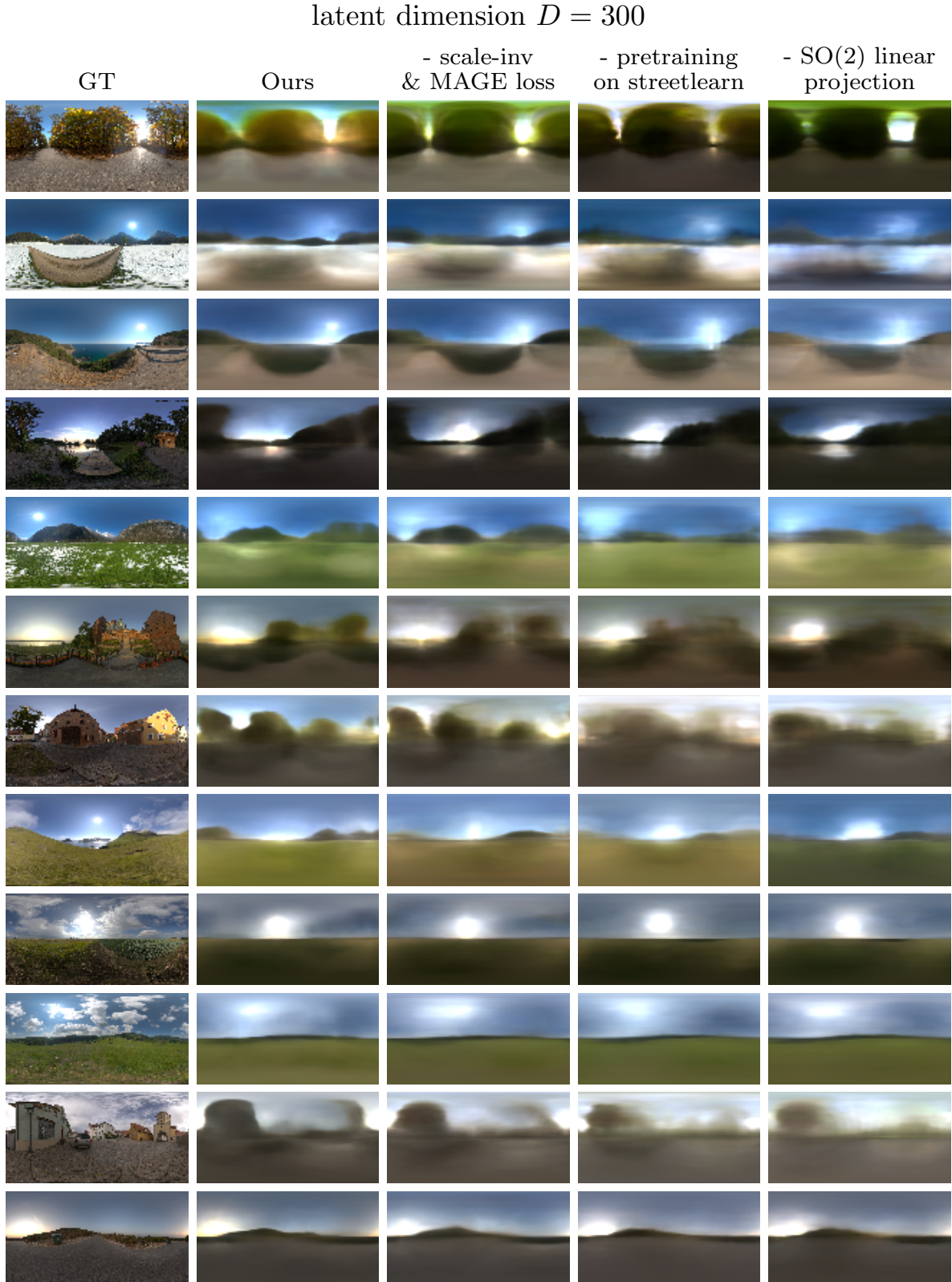


Figure 13. Corresponding figure to Table 4 comparing various ablations of our model, here with latent dimension  $D = 300$ .

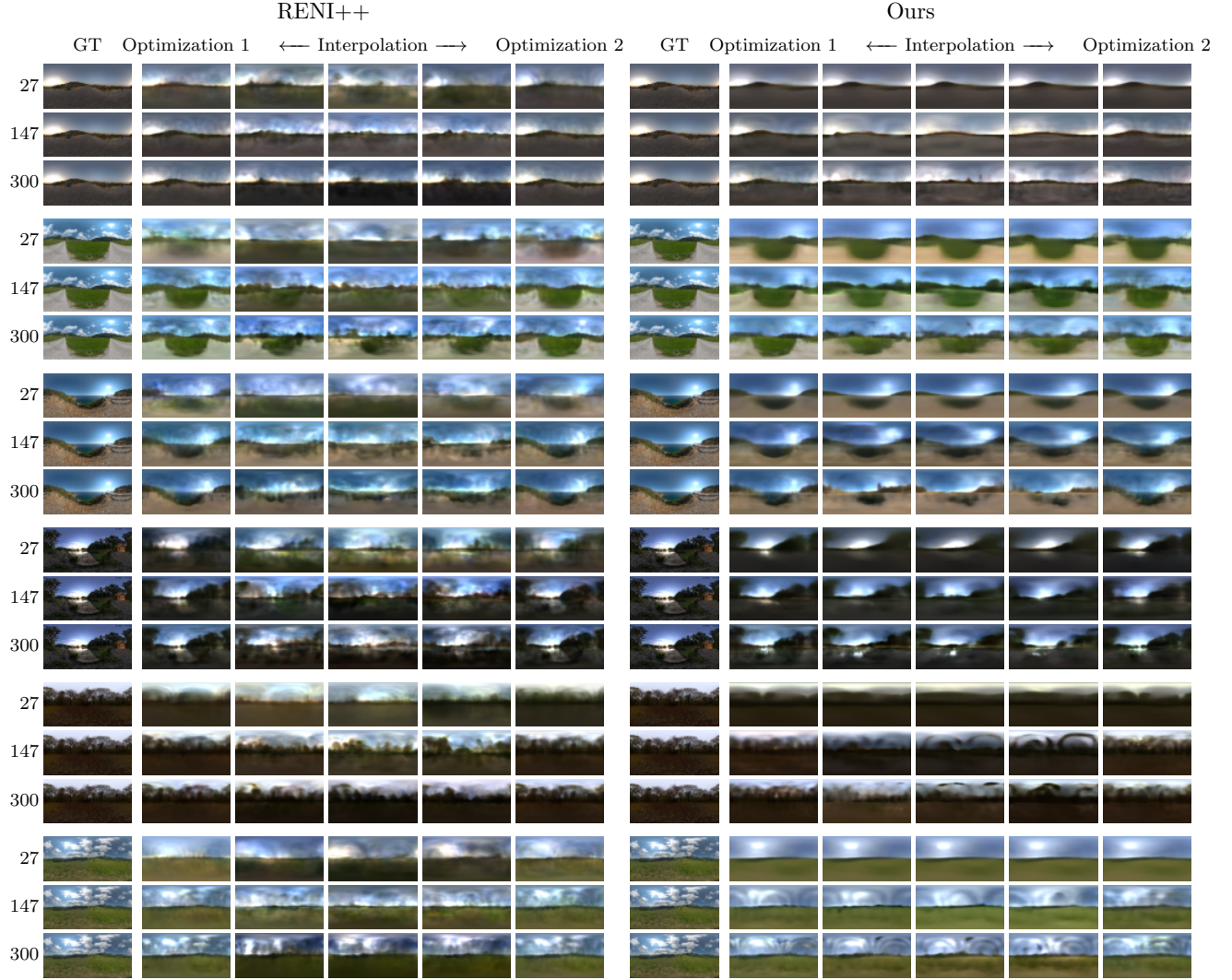


Figure 14. More examples following Figure 5. Comparing the uniqueness of RENI++ and our model across varying latent dimensions  $D$  by interpolating between two latent codes optimized to the same target image. With our model, the output images remain relatively consistent during interpolation, suggesting that the latent space is unique. With RENI++ the output images diverge heavily from the optimized images during interpolation. There are color, luminance and structural shifts visible in the interpolation for RENI++, suggesting that the latent space is less unique.