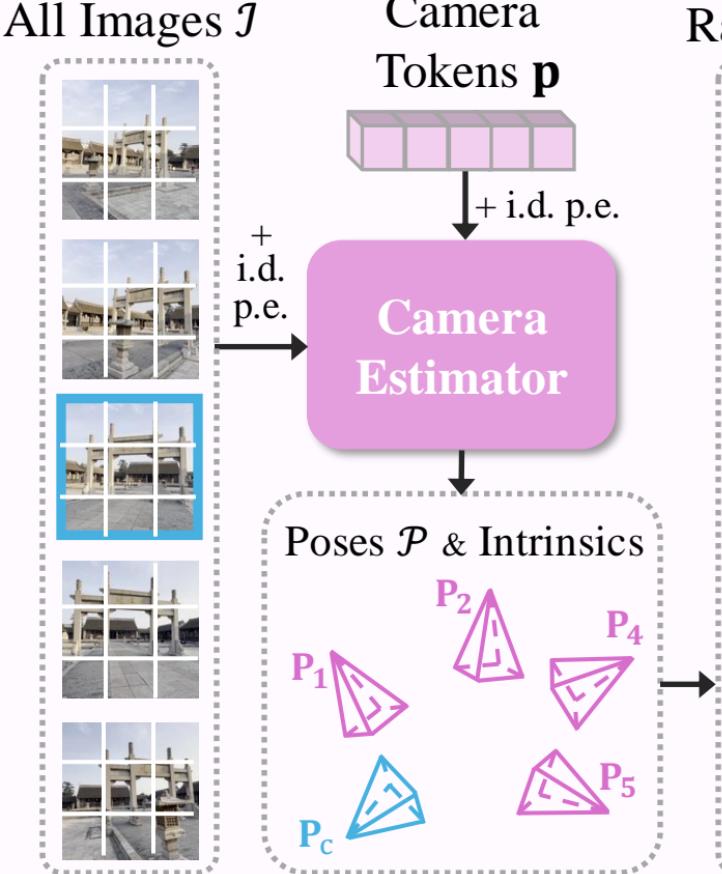
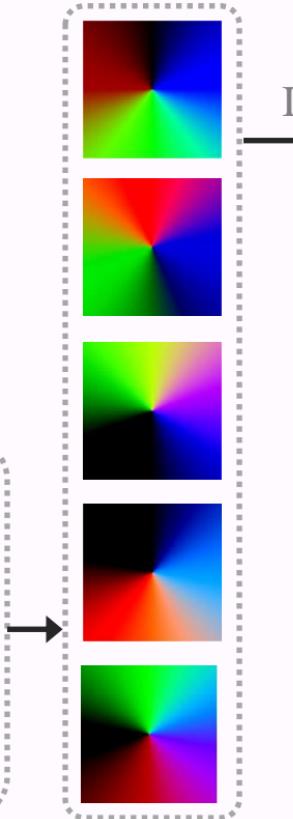


Camera Estimation



Predicted Plücker Ray Maps \mathcal{R}



Latent Scene Reconstruction

Predicted $\mathcal{R}_{\mathcal{A}}$ & Input $\mathcal{I}_{\mathcal{A}}$ (Indexed from \mathcal{R} and \mathcal{I})



Scene Tokens \mathbf{z}



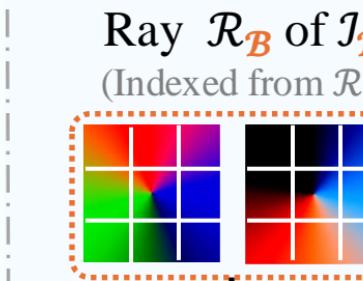
Updated Scene Tokens \mathbf{z}^*

Scene Reconstructor

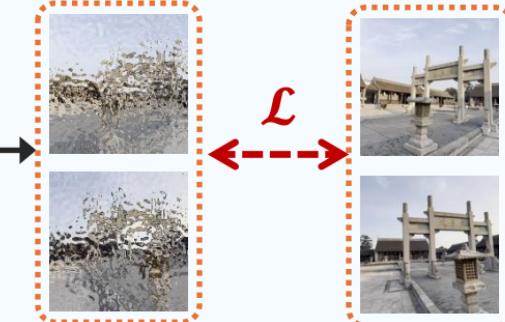
Updated Scene Tokens \mathbf{z}^*

Rendering (Training)

Predicted Plücker Ray $\mathcal{R}_{\mathcal{B}}$ of $\mathcal{I}_{\mathcal{B}}$ (Indexed from \mathcal{R})



Predicted Images $\hat{\mathcal{I}}_{\mathcal{B}}$
(Indexed from \mathcal{I})



$$\mathcal{L} = \frac{1}{|\mathcal{I}_{\mathcal{B}}|} \sum_{\hat{\mathcal{I}} \in \hat{\mathcal{I}}_{\mathcal{B}}} (\text{MSE}(\mathcal{I}, \hat{\mathcal{I}}) + \lambda \cdot \text{Percep}(\mathcal{I}, \hat{\mathcal{I}}))$$