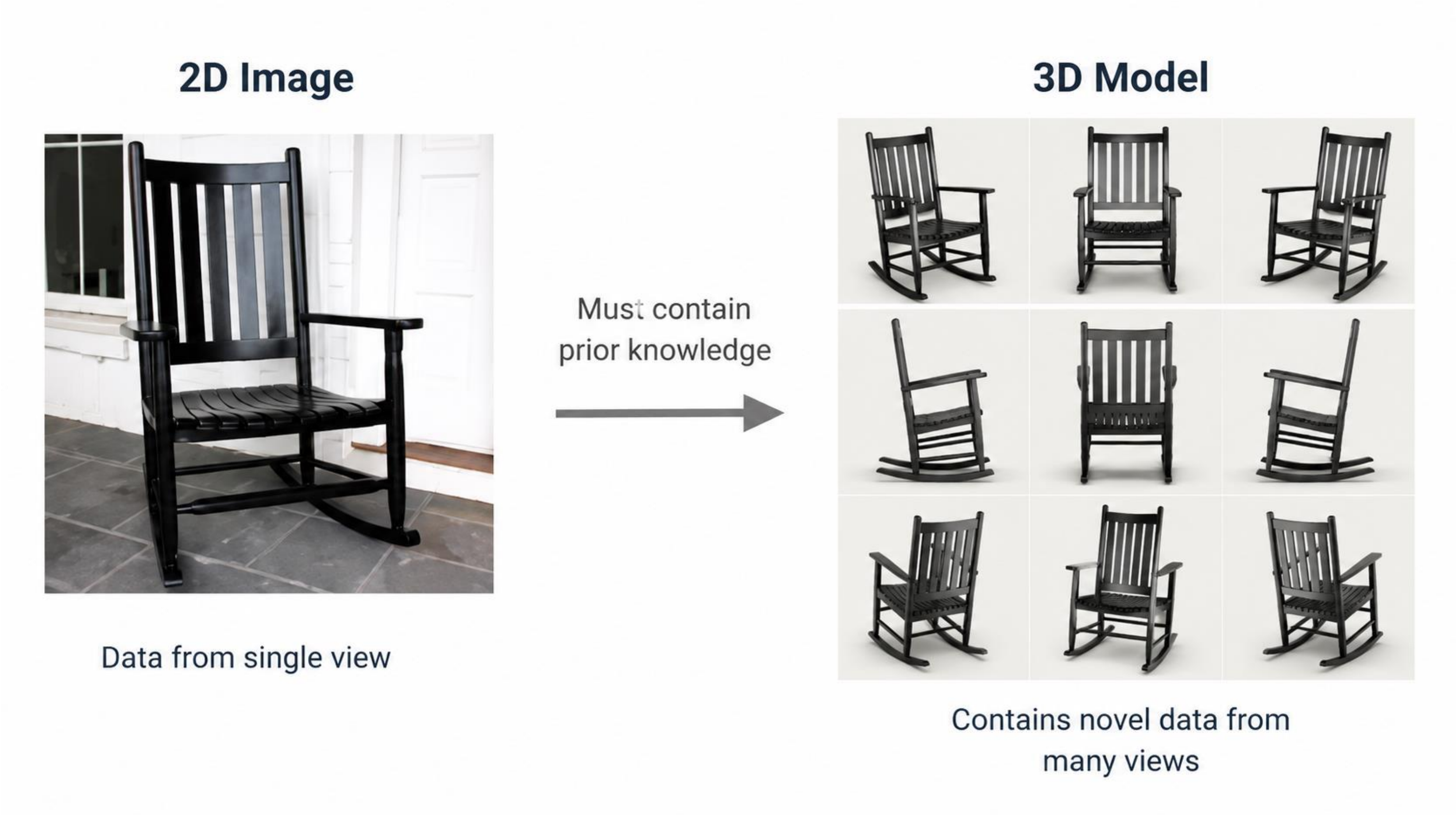
Continuous Viewpoint Adaptation for Single View 3D Object Reconstruction

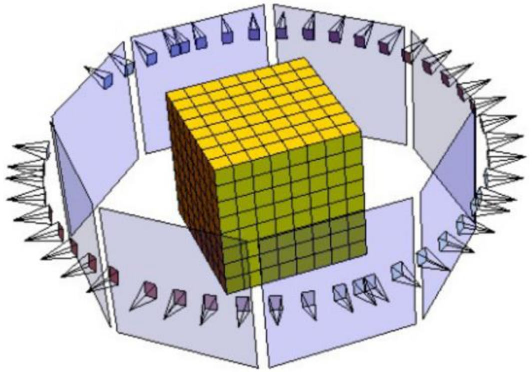
Seunghyun Hwang, Qiang Qiu

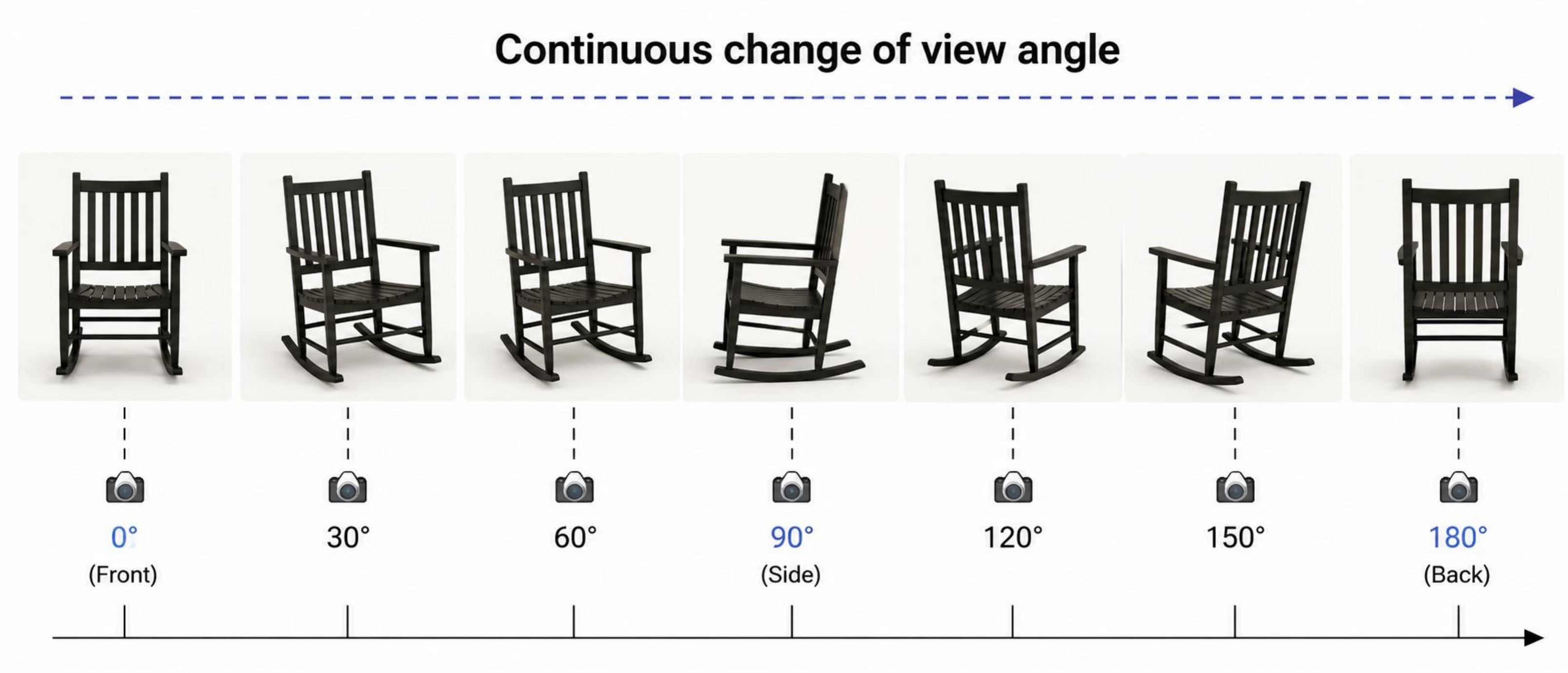
Purdue University

3D Novel View Synthesis

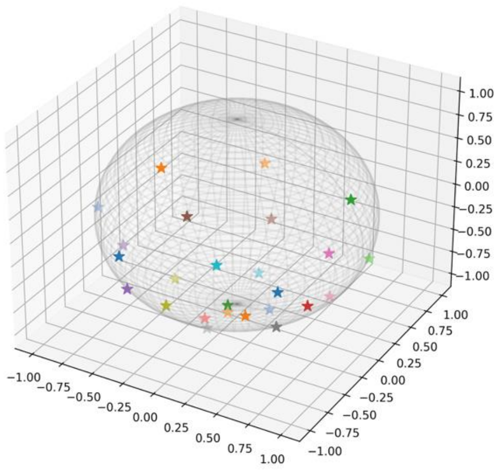


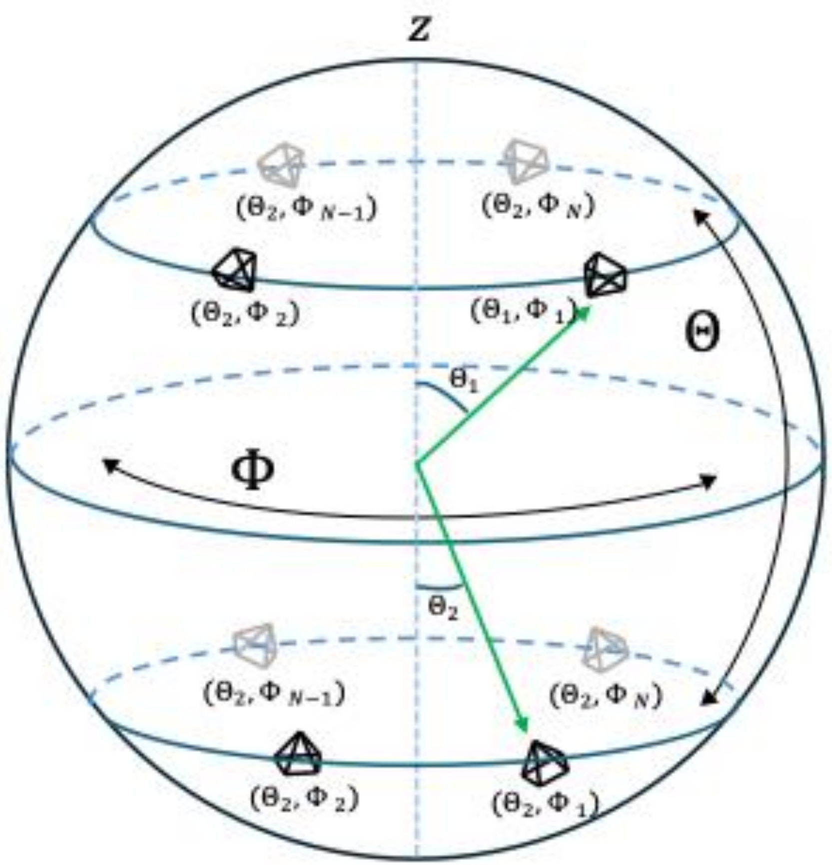
Continuous camera positions





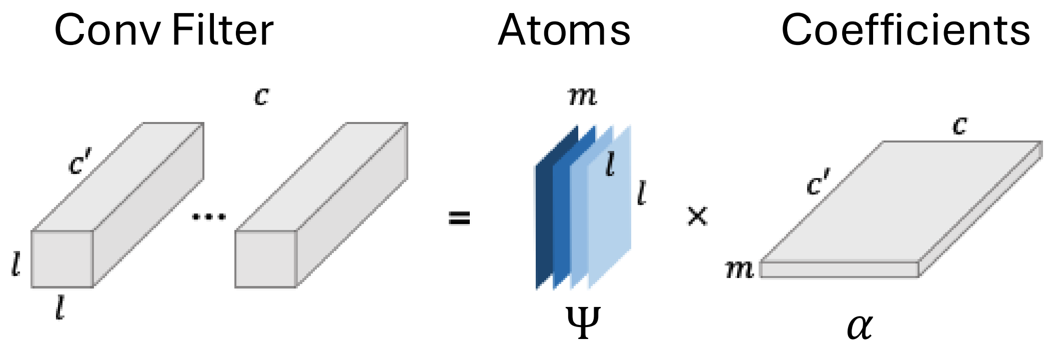
Filter subspace modeling of 3D Camera space





- Camera positions lie on a spherical world space.
- Spherical world space can be represented by azimuthal (Φ) and polar angle (Θ).
- For each polar angle (Θ), there can be infinite options for azimuthal (Φ) angles.

Filter subspace modeling of 3D Camera space



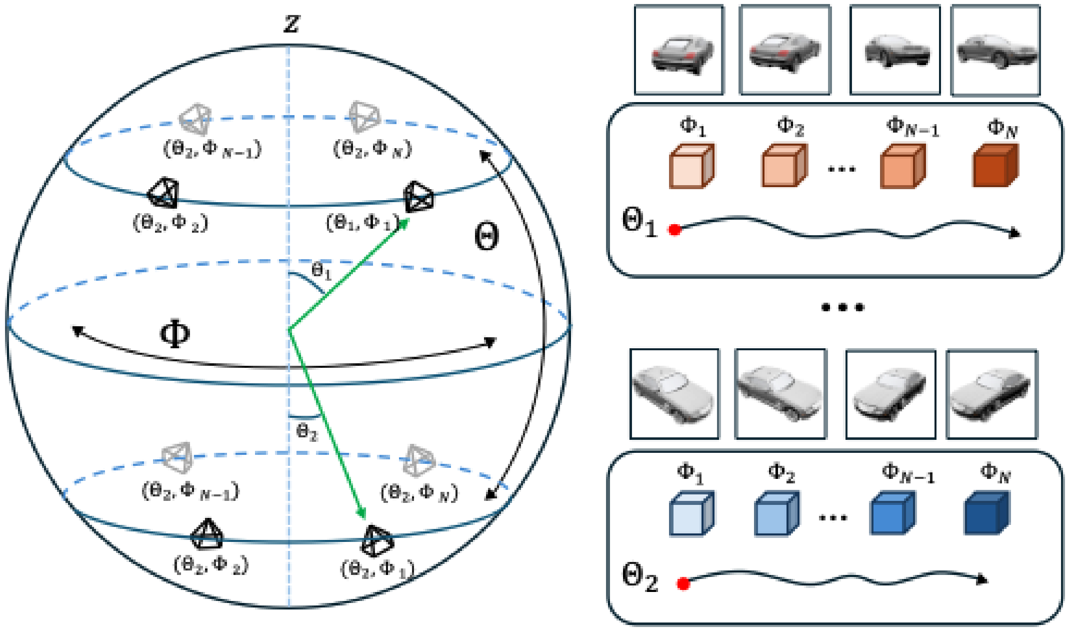
LxLxCxC' convolutional layer is decomposed into the product of ***m*** bases of size ***LxL***

Decompose filter weights W over filter atoms Ψ and coefficients α.

$$\mathbf{W}^i(\Theta, \Phi) = \mathbf{a}^i \mathbf{\Psi}^i(\Theta, \Phi),$$

Qiu, Qiang, Xiuyuan Cheng, and Guillermo Sapiro. "DCFNet: Deep neural network with decomposed convolutional filters." International Conference on Machine Learning. PMLR, 2018.

Filter subspace modeling of 3D Camera space



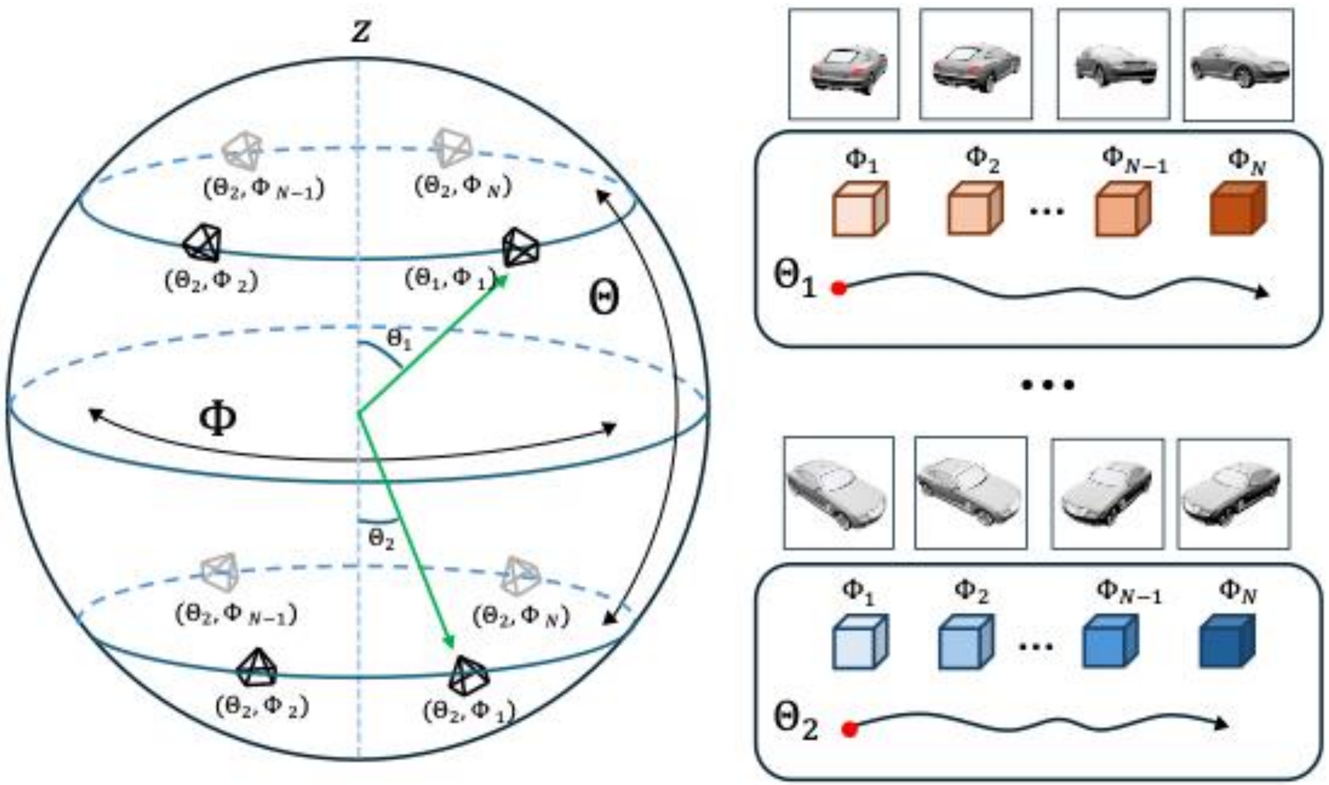
Embedding *spherical camera world space* into *filter subspace*

- Generate different sets of filter atoms according to ODE function f.

$$\Psi^i(\Phi_s) = \Psi^i(\Phi_0) + \int_{\Phi_0}^{\Phi_s} f(\Psi^i(\Phi), \Phi; \theta)\, d\Phi.$$

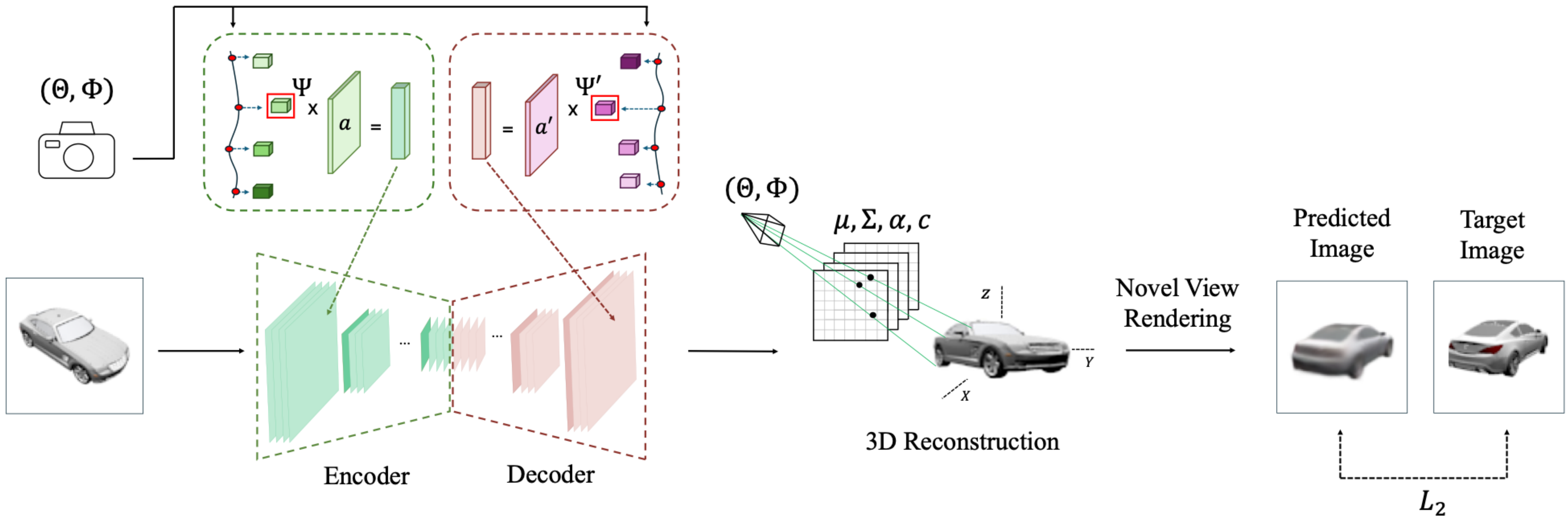
- Filter atoms can cover whole spherical camera space.

Filter subspace modeling of 3D Camera space



- Embedding *spherical camera world space* into *filter subspace*.
- Condition initial condition of **ODE** with polar angle (Θ), [0, 180°] .
- For each initial state, evolve ODE according to different azimuthal (Φ) angles [0, 360°] .
- Generate different sets of filter atoms according to Θ and Φ.

Overview



(Θ, Φ)
Ψ
x
a
=
=
a′
x
Ψ′
...
...
Encoder
Decoder
(Θ, Φ)
μ, Σ, α, c
Z
Y
X
3D Reconstruction
Novel View
Rendering
Predicted
Image
Target
Image
L_2

Filter subspace modeling of 3D Camera space

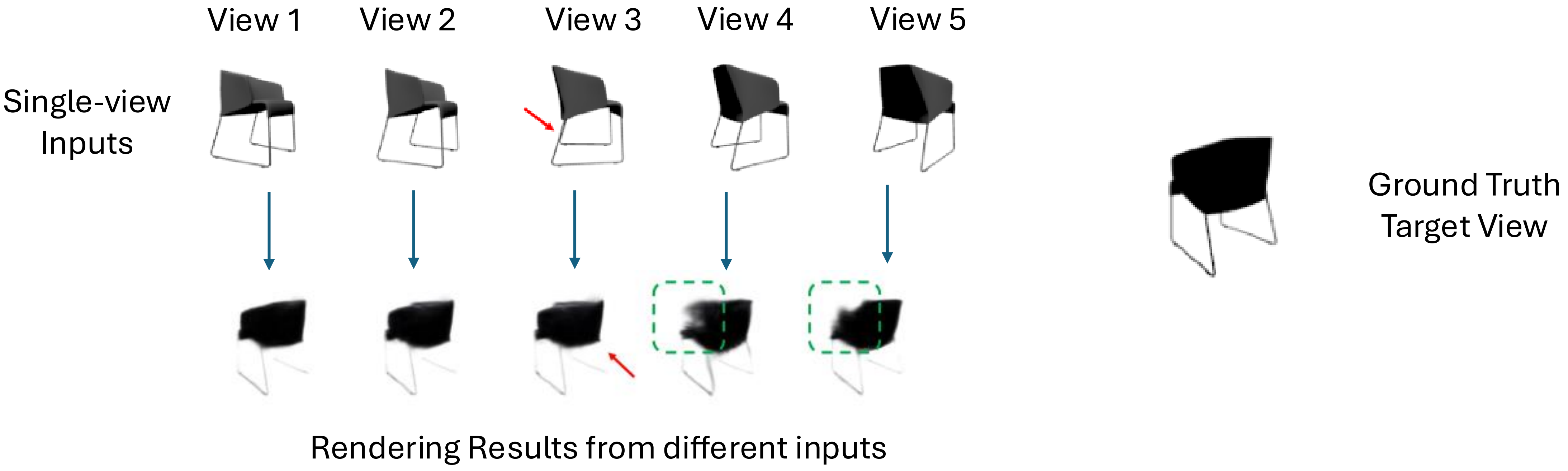
- **Continuity Propagation - Variations in the 3D reconstructions** results of our networks are upper bounded by the **variations in input view-points** through Lipshitz continuity

Proposition 4.1 (Continuity of View-Specific Features). *Let* $\mathbf{Z}_u^i$ *and* $\mathbf{Z}_v^i$ *be inputs to the* i*-th layer from adjacent views, and* $\mathbf{\Psi}_u^i$ *and* $\mathbf{\Psi}_v^i$ *represent two continuously generated sets of filter atoms corresponding to those views, which together with the common atom coefficients* $\mathbf{a}^i$ *form* i*-th layer of the convolutional layer. If the difference between* $\mathbf{Z}_u^i$ *and* $\mathbf{Z}_v^i$ *is small (i.e.,* $\|\mathbf{Z}_u^i - \mathbf{Z}_v^i\|_2 \leq \delta$*) for some* δ*, and for non-expansive activation function, e.g., SiLU, we can upper bound the changes in the corresponding outputs* $\mathbf{Z}_u^{i+1}$ *and* $\mathbf{Z}_v^{i+1}$ *in terms of the changes in the filter atoms as:*

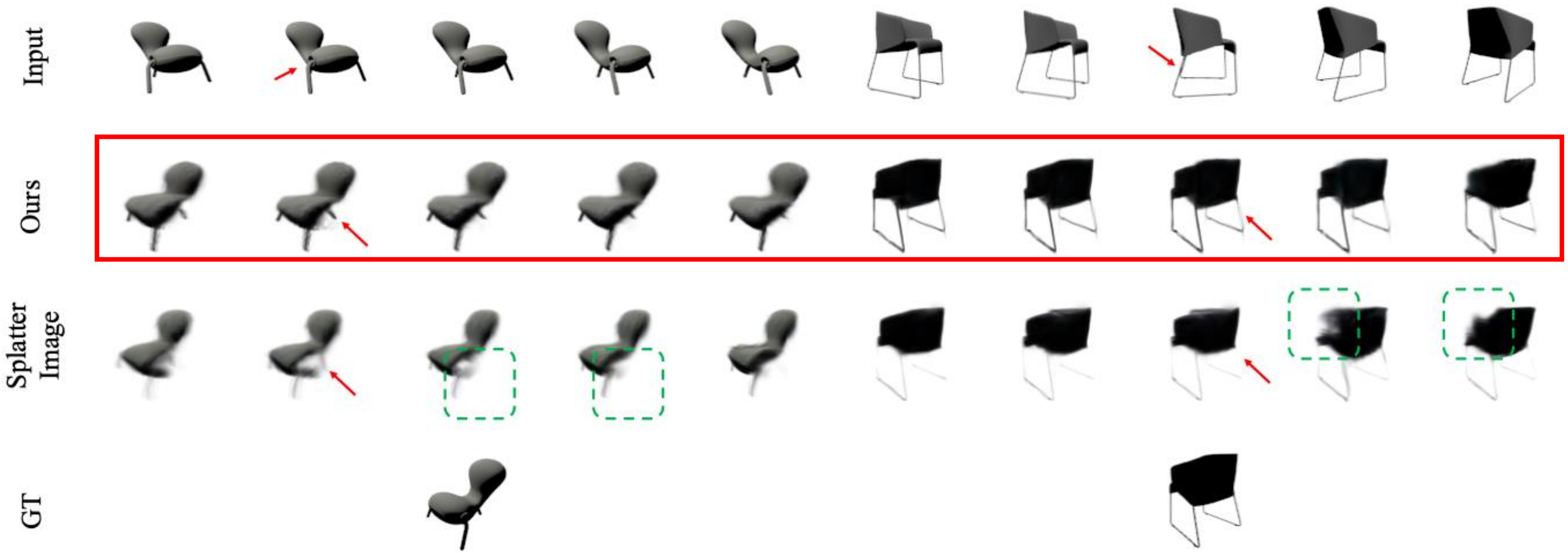
$$\|\mathbf{Z}_u^{i+1} - \mathbf{Z}_v^{i+1}\|_2 \leq (\|\mathbf{a}^i\|_2 \lambda \sqrt{|\mathcal{S}|} + \delta) \cdot \|\mathbf{\Psi}_u^i - \mathbf{\Psi}_v^i\|_2, \quad \text{with } \lambda = \sup_{s \in \mathcal{S}} \|\mathbf{Z}_u^i\|_{2,N_s}. \tag{11}$$

Remark 4.2. *According to Proposition 4.1, the difference between feature embeddings at each layer is upper bounded by the difference in the continuously generated filter atoms, for some small difference of* δ *between inputs from close views. Similarly, for the decoder, where each layer is also decomposed over continuously generated view-dependent filter atoms and cross-view shared coefficients, if the difference between the input features at* i*-th layer is small (i.e.,* $\|\mathbf{Z}_u^i - \mathbf{Z}_v^i\|_2 \leq \delta'$*), we can upper bound the variations in the corresponding outputs in terms of the changes in the filter atoms.*

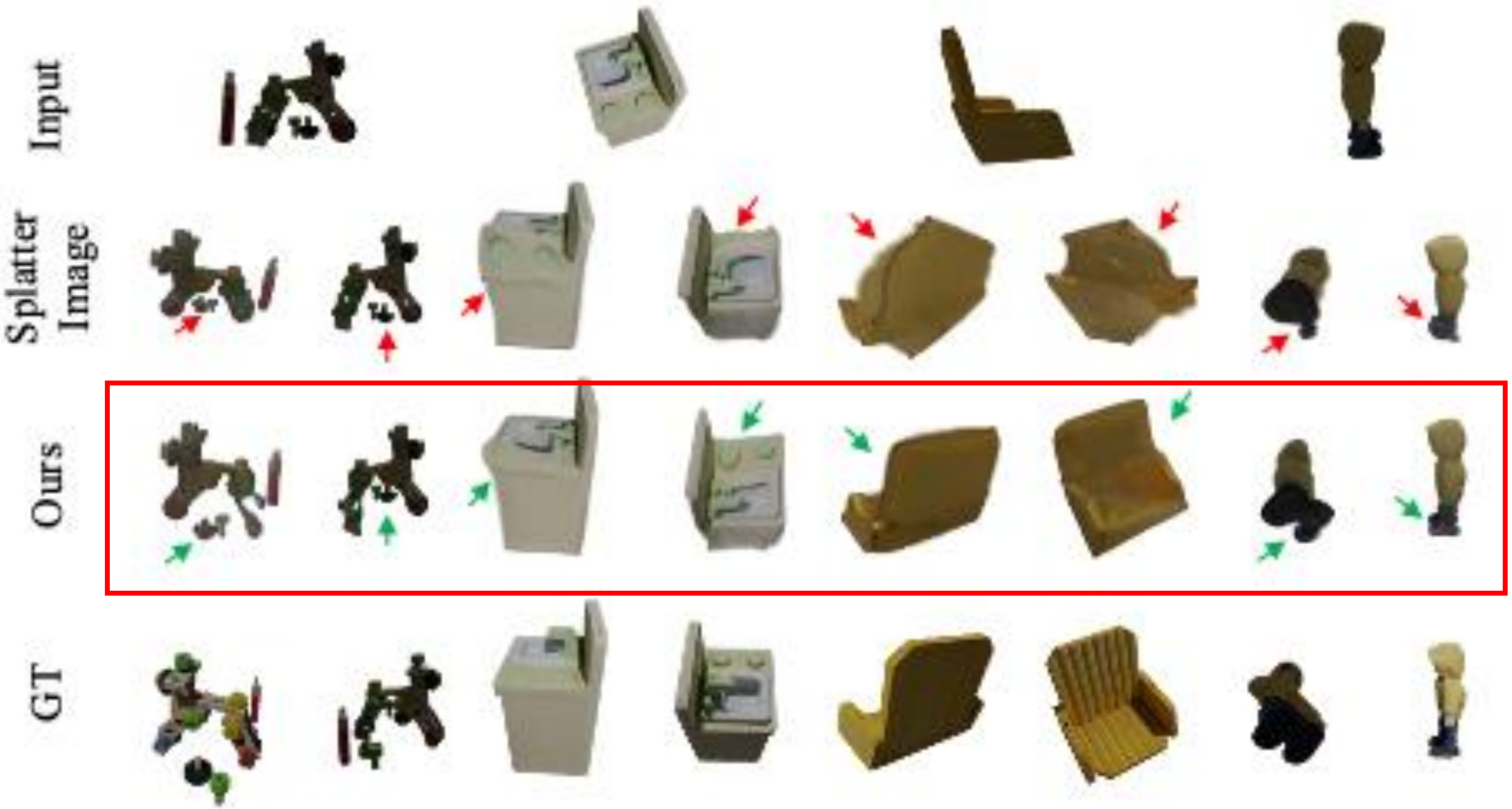
Inconsistency in 3D reconstruction across different input views



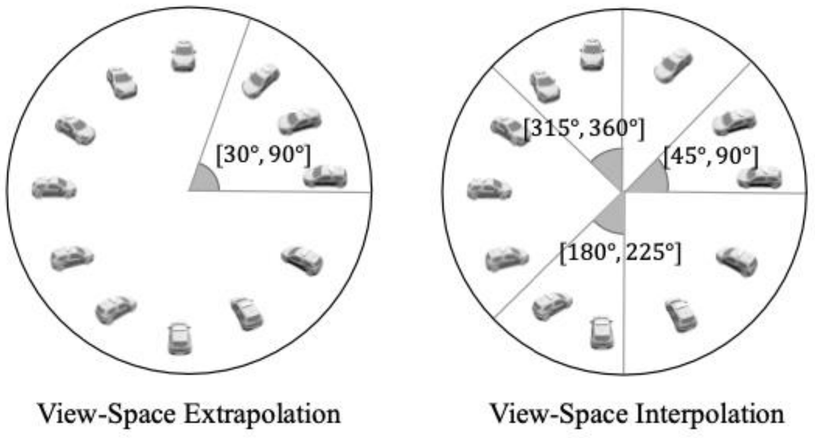
View consistency for different input single view images

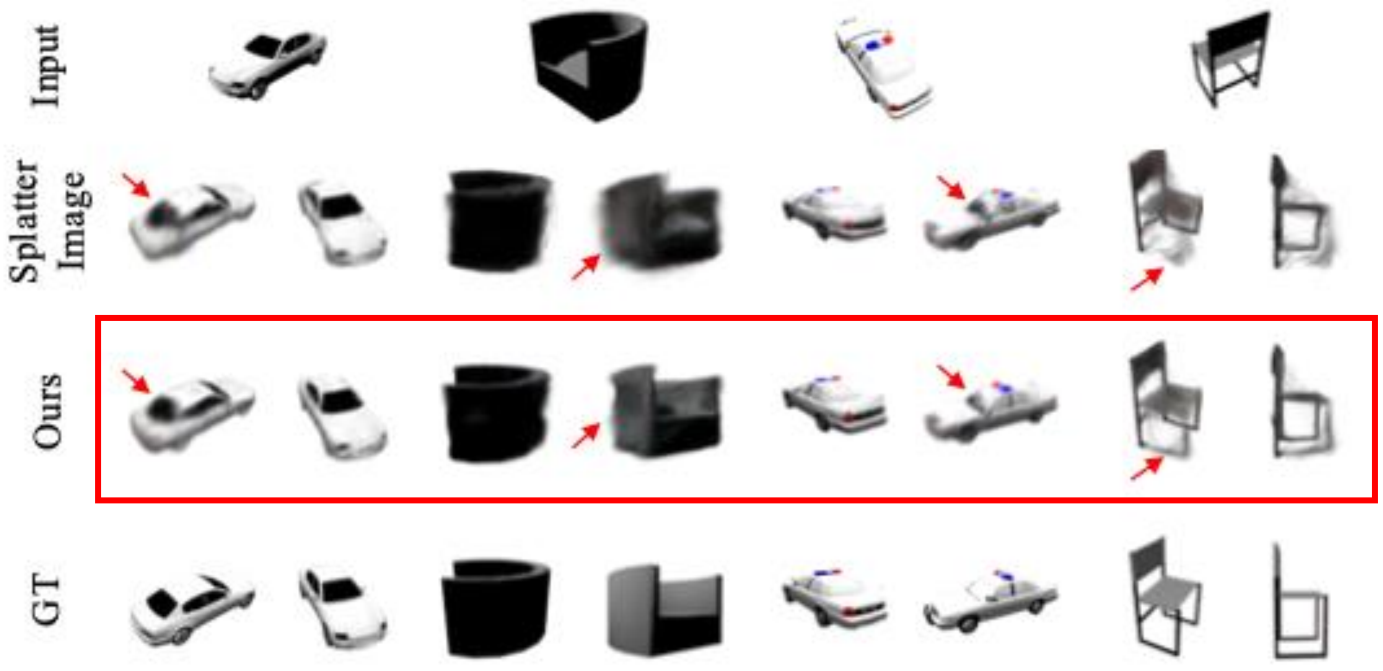


Results



Improved results in view extrapolation/interpolations

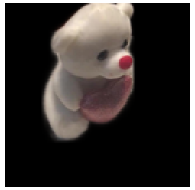




Results

Table 6. Ablation studies by adding different components.

	PSNR ↑	SSIM ↑	LPIPS ↓
Baseline	21.81	0.88	0.166
+ Camera	22.13	0.88	0.156
+ DCF	22.19	0.88	0.155
+ ODE(Φ)	22.64	0.89	0.143
+ ODE(Φ, Θ)	**23.12**	**0.90**	**0.129**



Single-view Input Image

Ours GT

Single-view Input Image

Ours GT

Takeaways