JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

JacobiNeRF: NeRF Shaping with Mutual Information Gradients

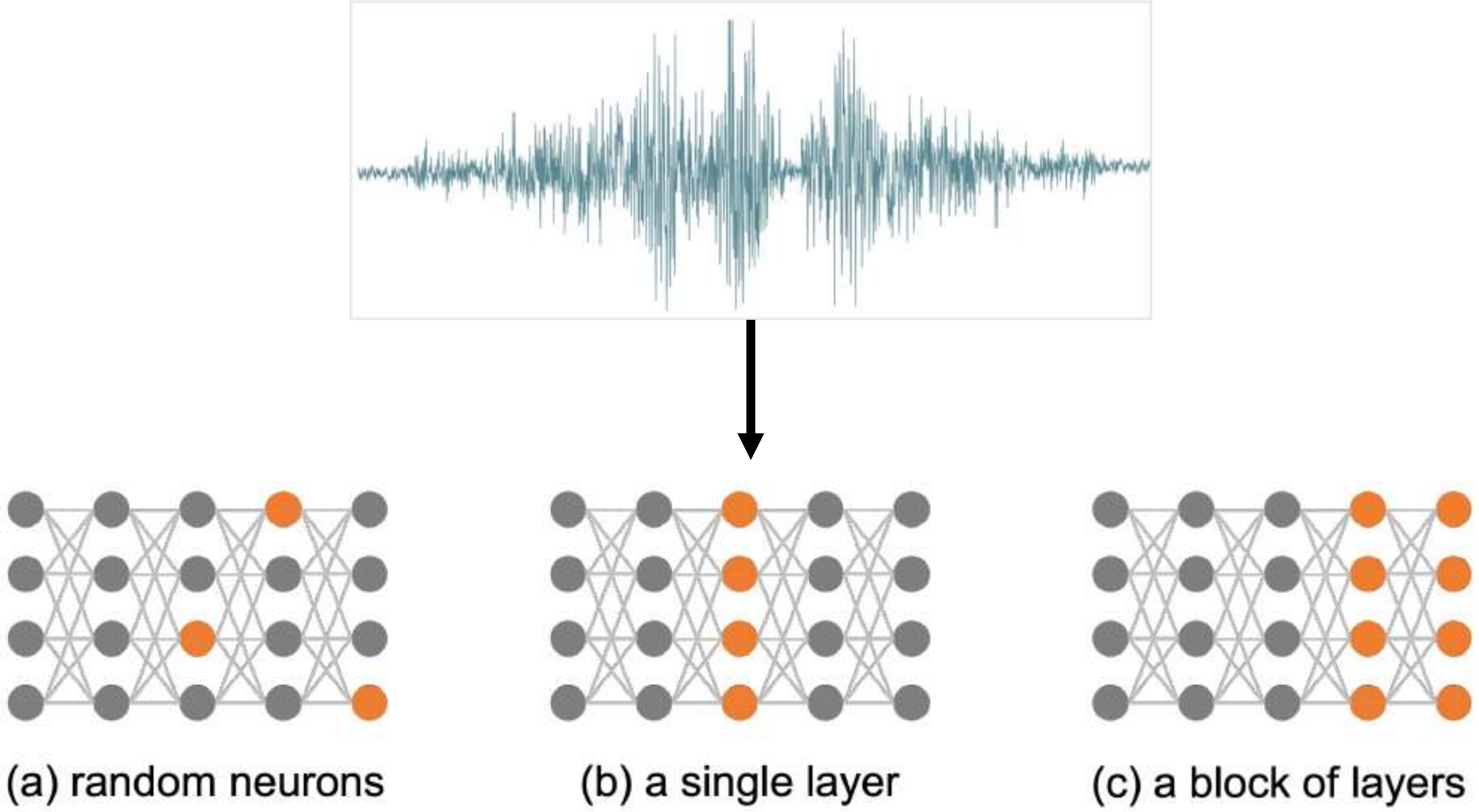
Xiaomeng Xu[1,*], Yanchao Yang[2,3,*,†], Kaichun Mo[3,4], Boxiao Pan[3], Li Yi[1,5,6], Leonidas Guibas[3,7]

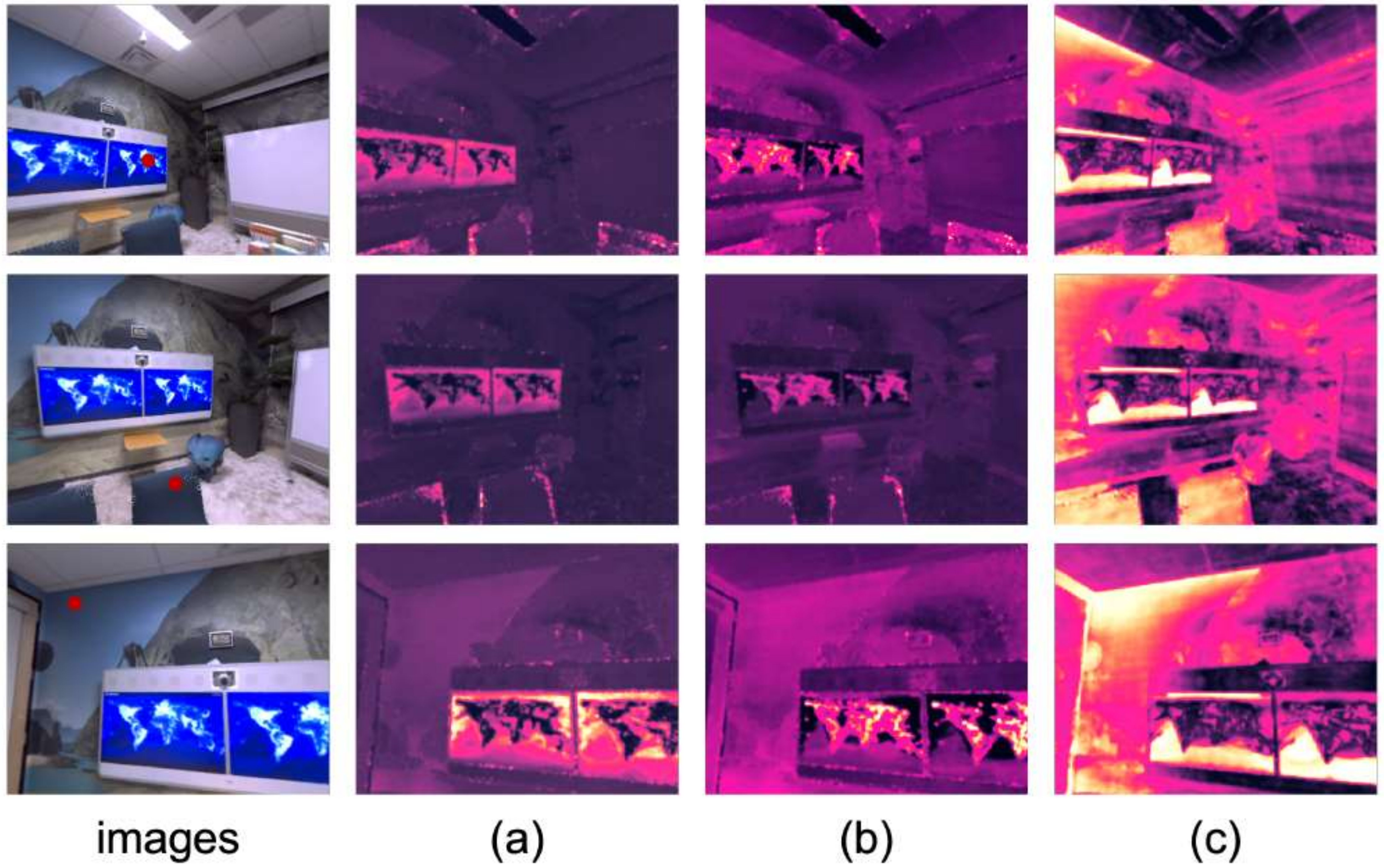
[1]Tsinghua University, [2]The University of Hong Kong, [3]Stanford University, [4]NVIDIA Research, [5]Shanghai AI Laboratory, [6]Shanghai Qizhi Institute, [7]Google Research

*Equal Contributions, †Corresponding Author

Paper Tag: THU-AM-002

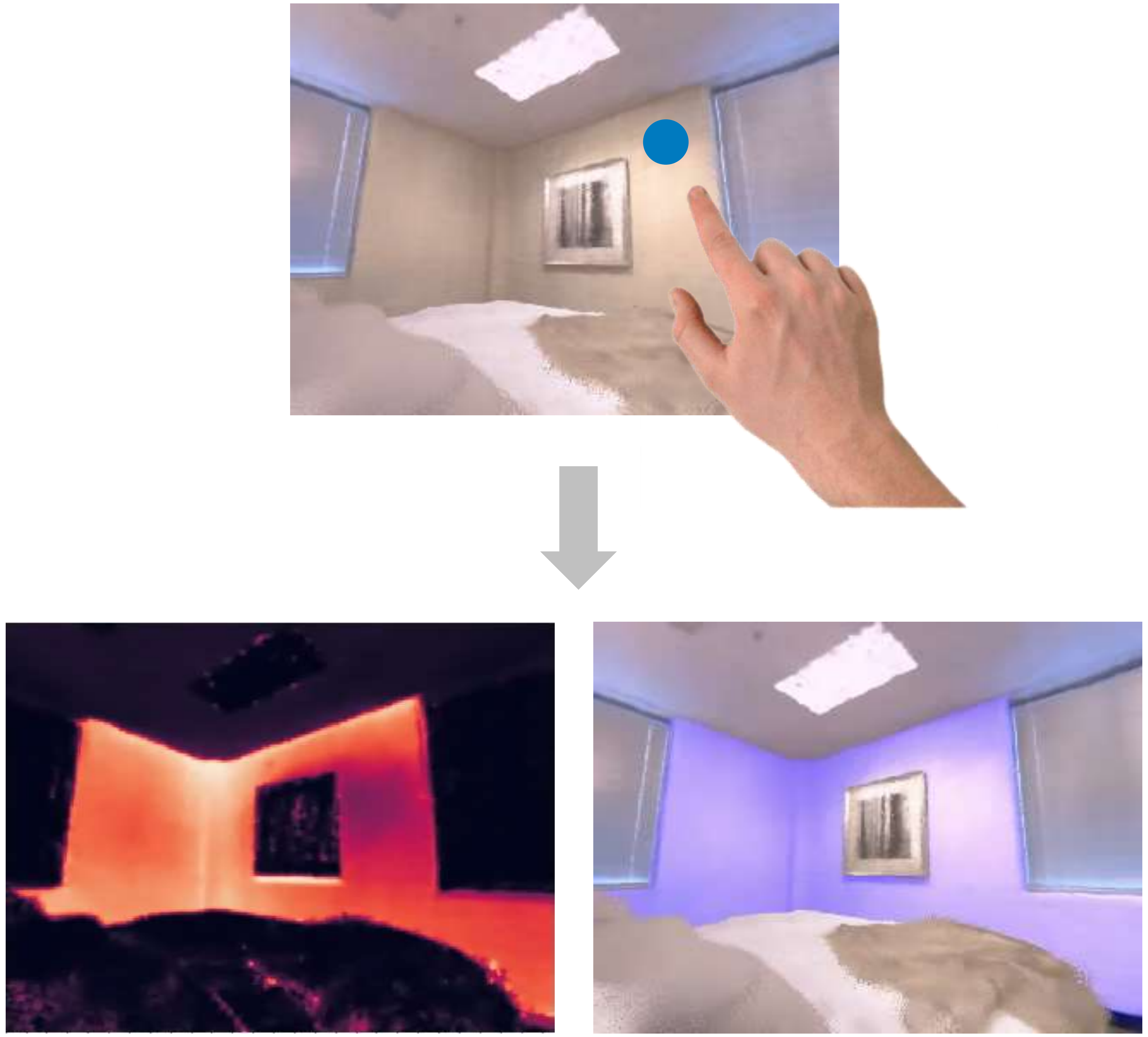
Preview: Plain NeRF Lacks Semantic Awareness





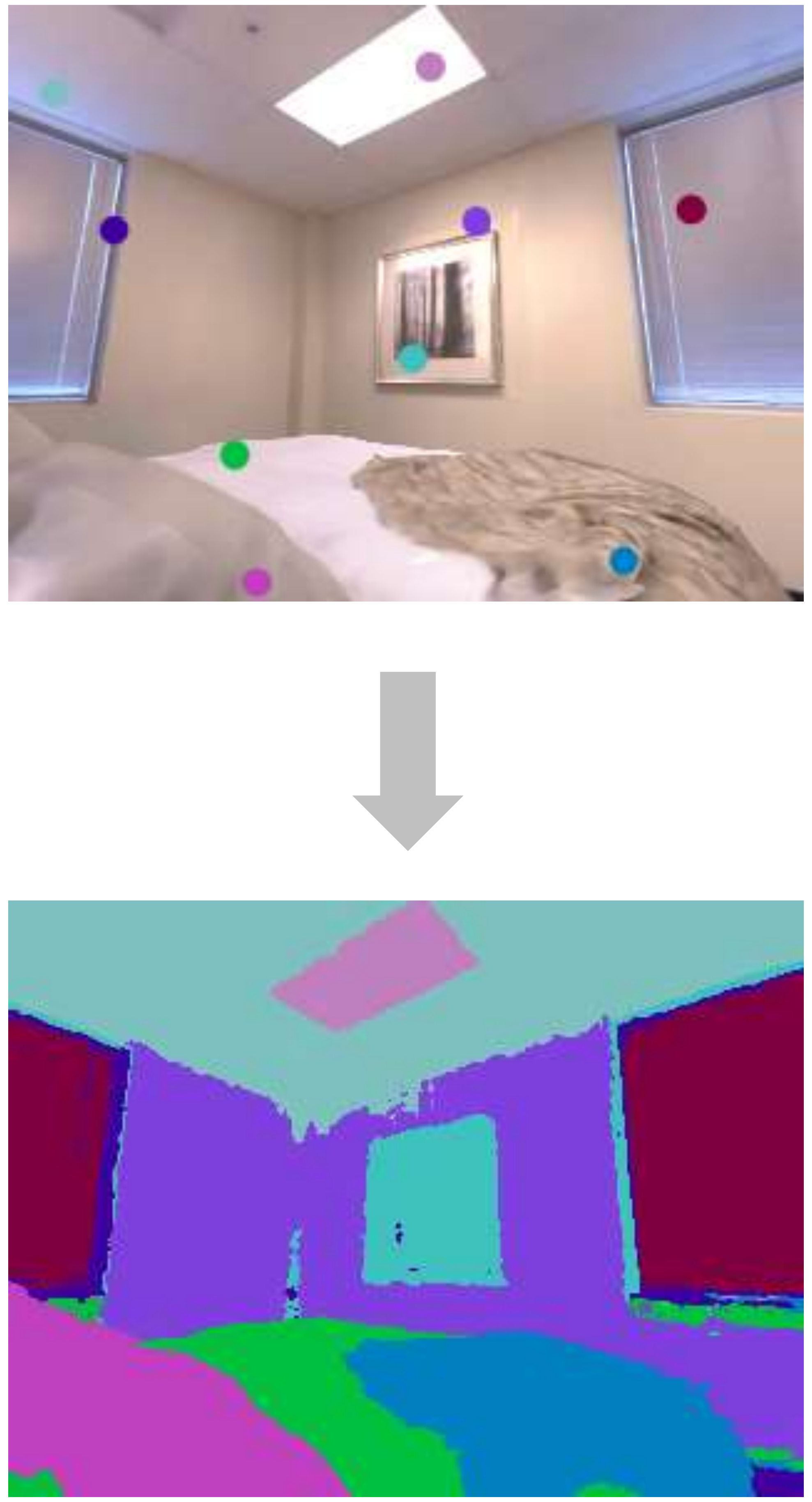
“Jiggling” the neuron weights of the NeRF results in diffuse perturbations that are not semantically aligned

Preview: User Interactions with 3D Scenes through 2D Views



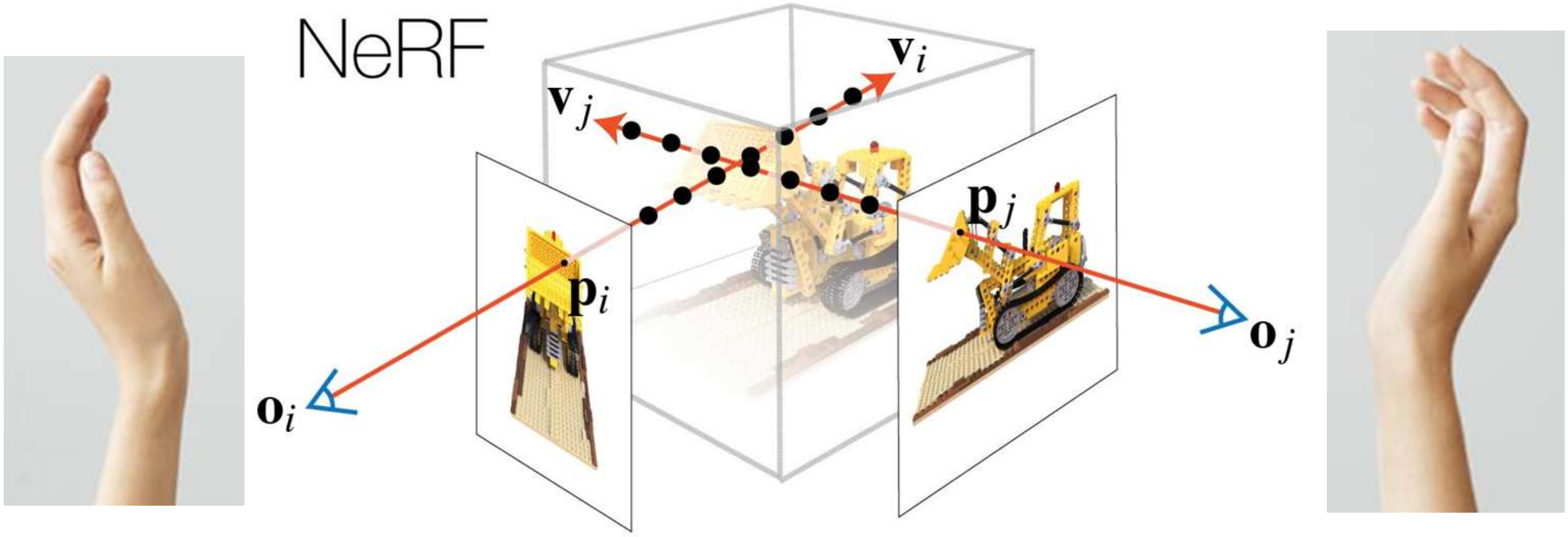
selecting

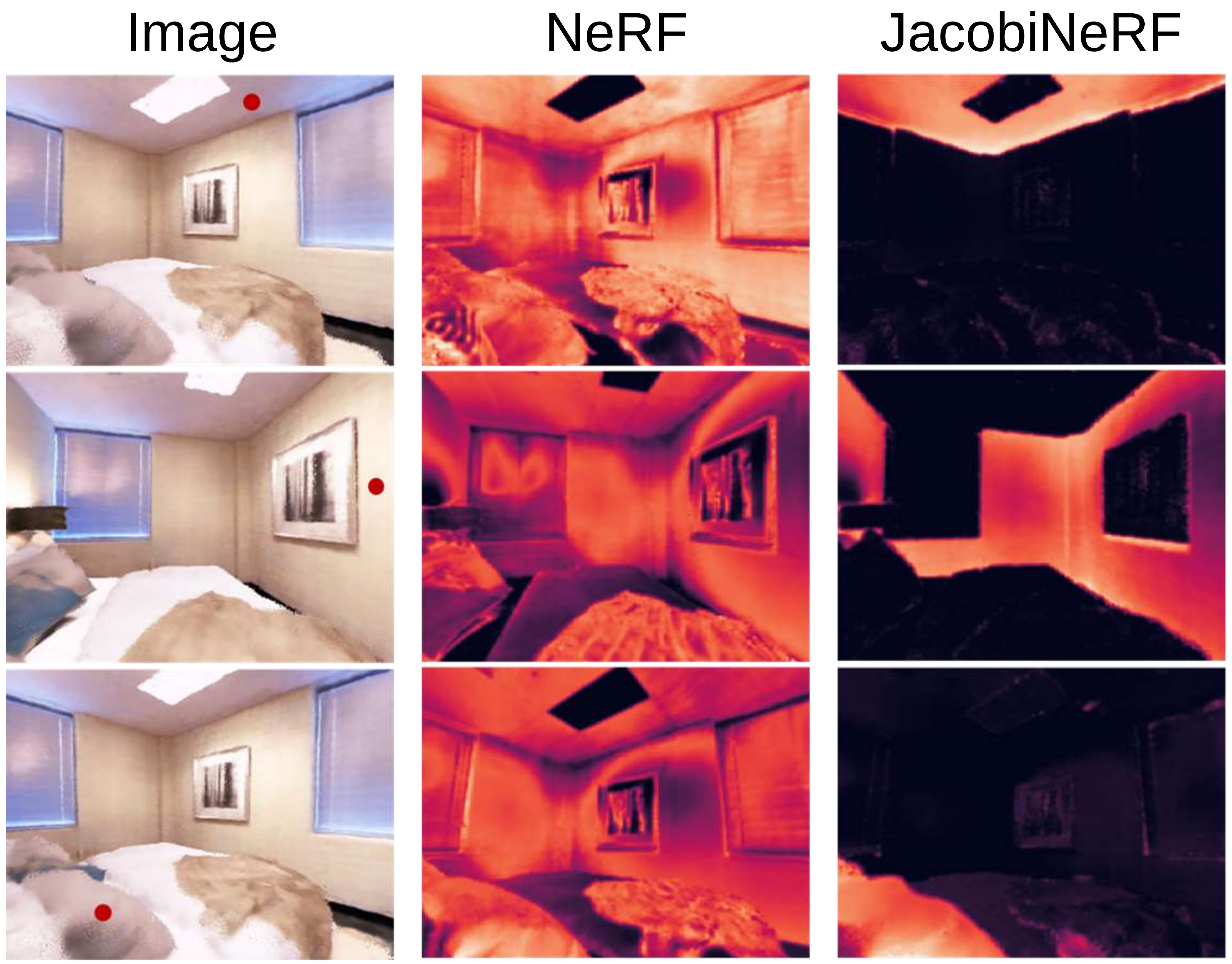
editing



label propagation

Preview: Encode Semantic Correlations into a NeRF



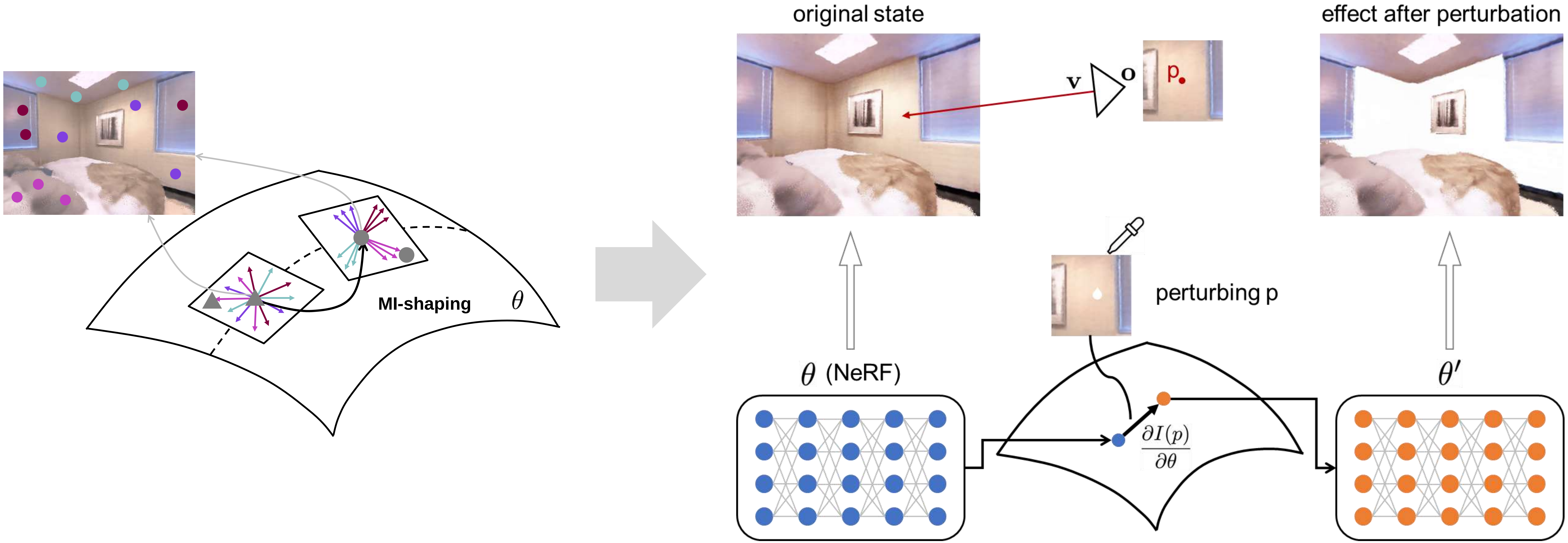


Image

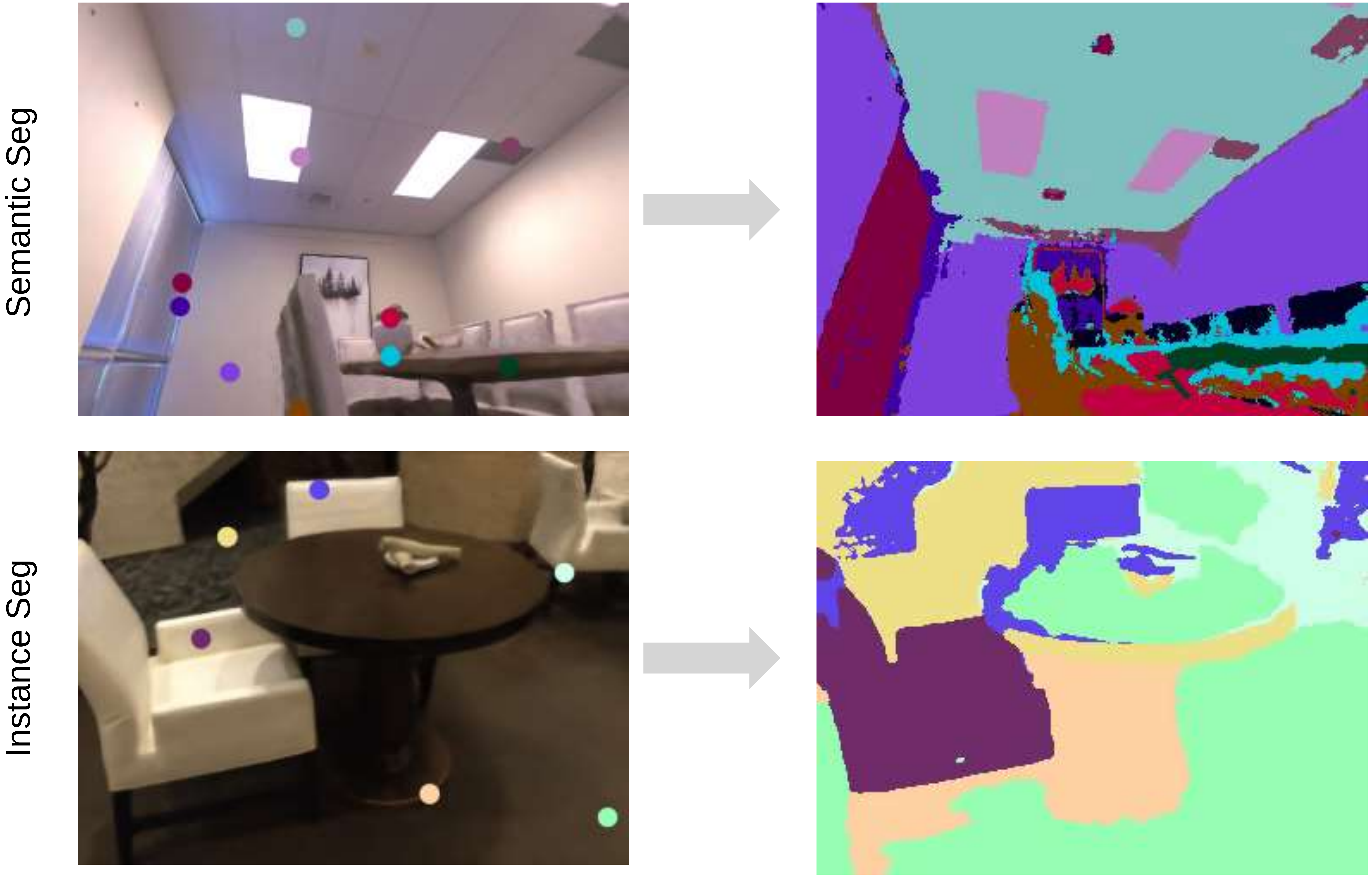
NeRF

JacobiNeRF

Preview: NeRF Shaping with Mutual Information Gradients

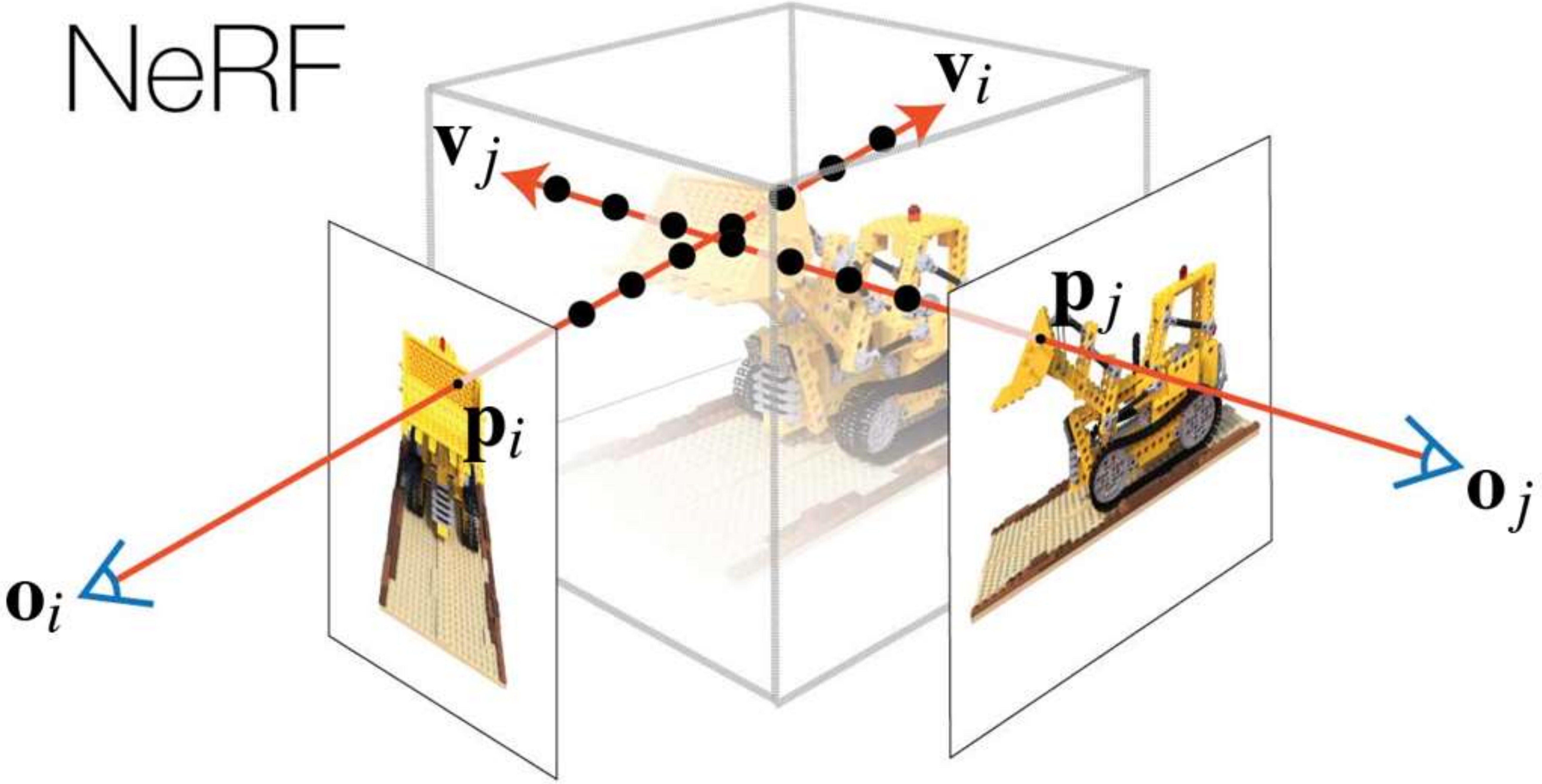


Preview: Label Propagation

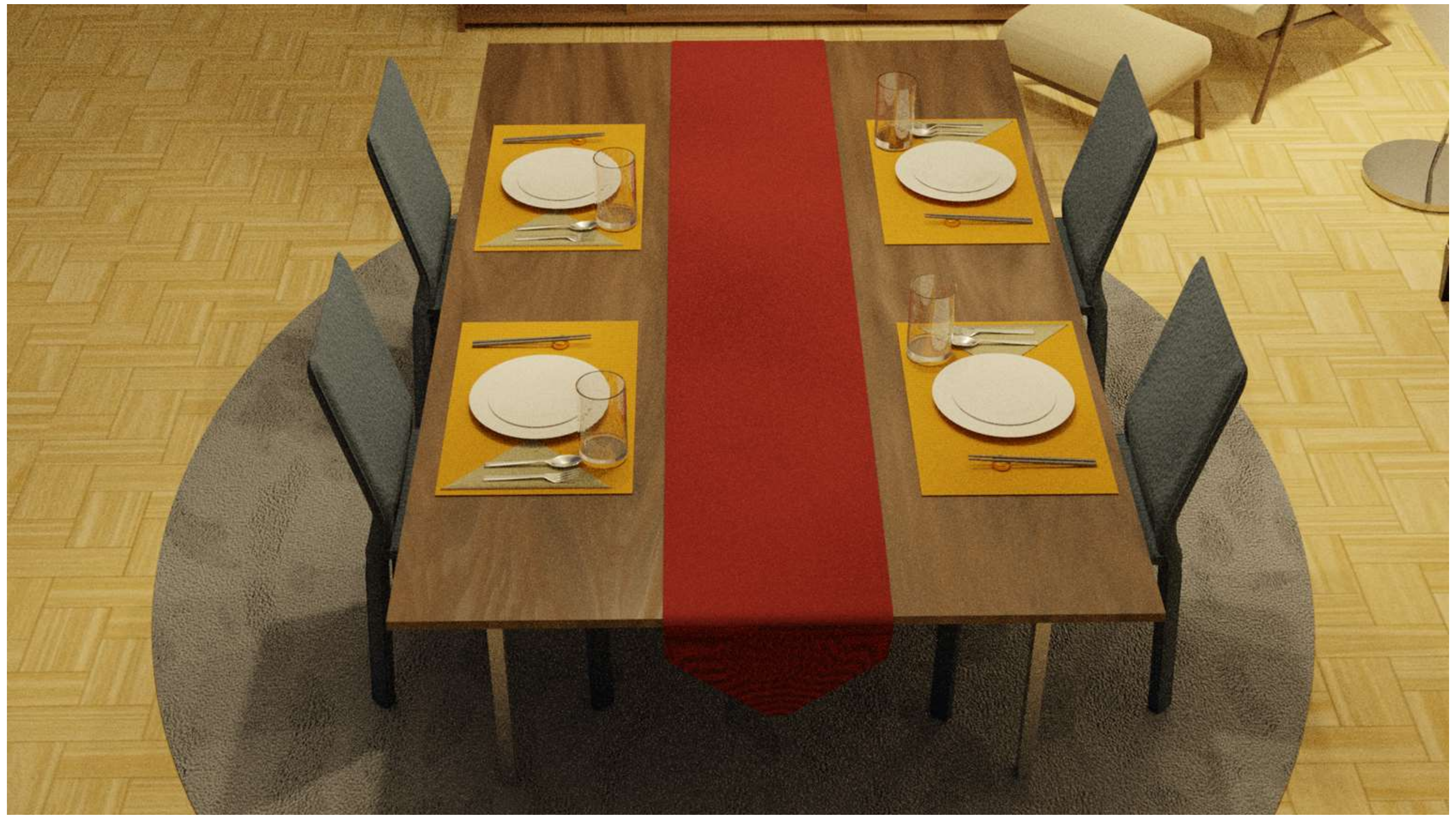


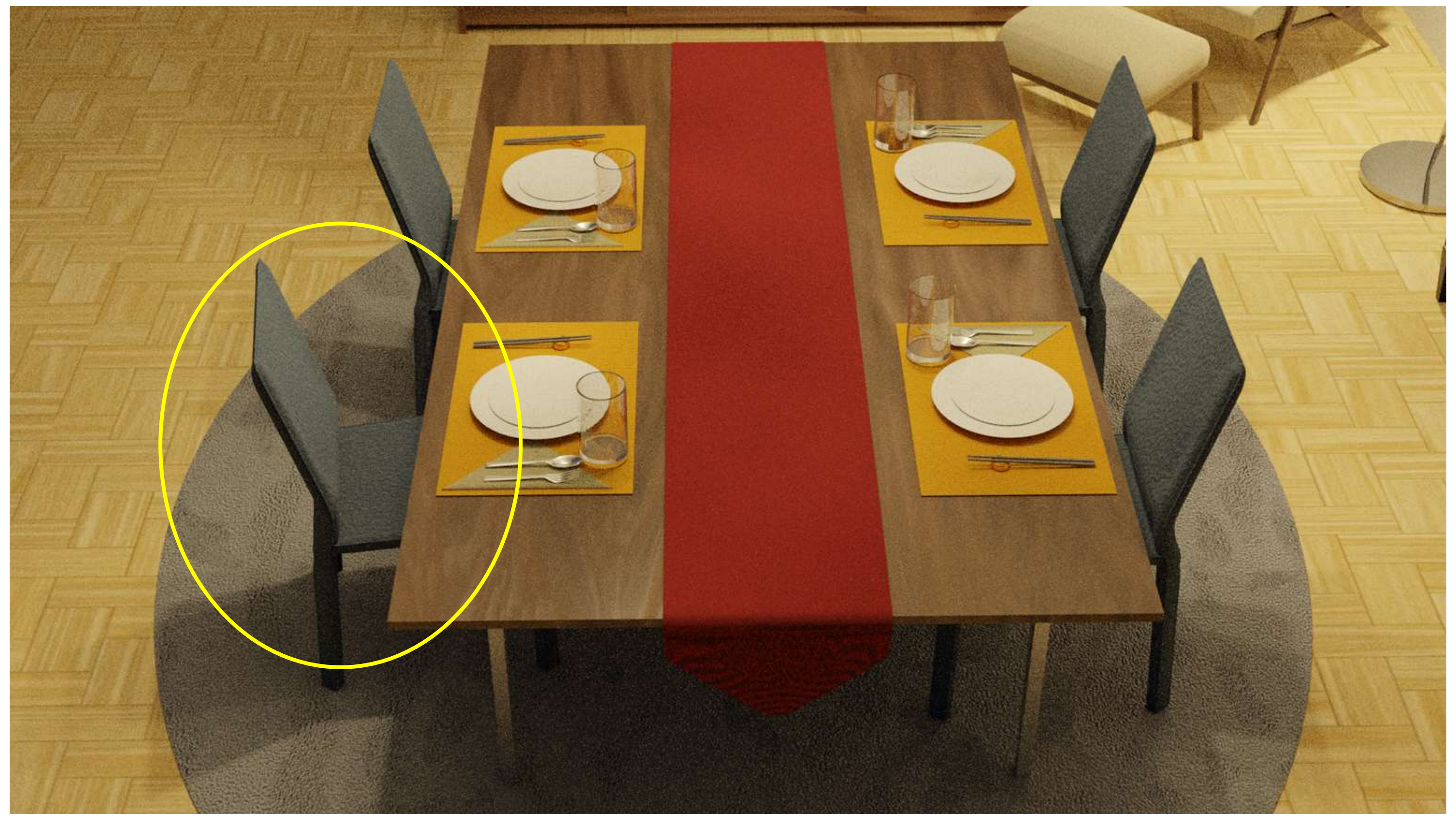
Example sparse annotation interaction

User Interactions with NeRFs

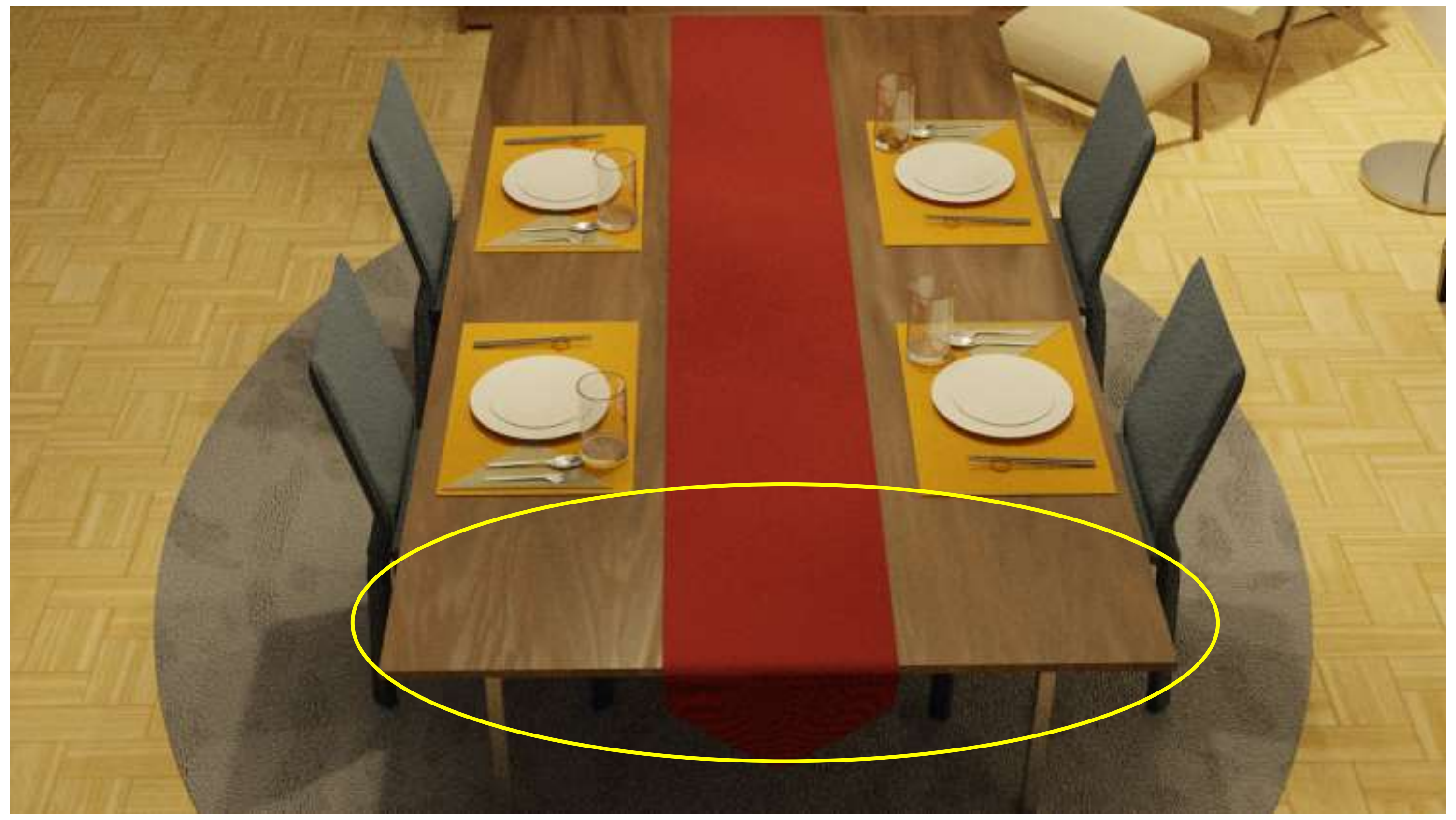


Semantic Structure of a Scene is Reflected in its Co-Variations

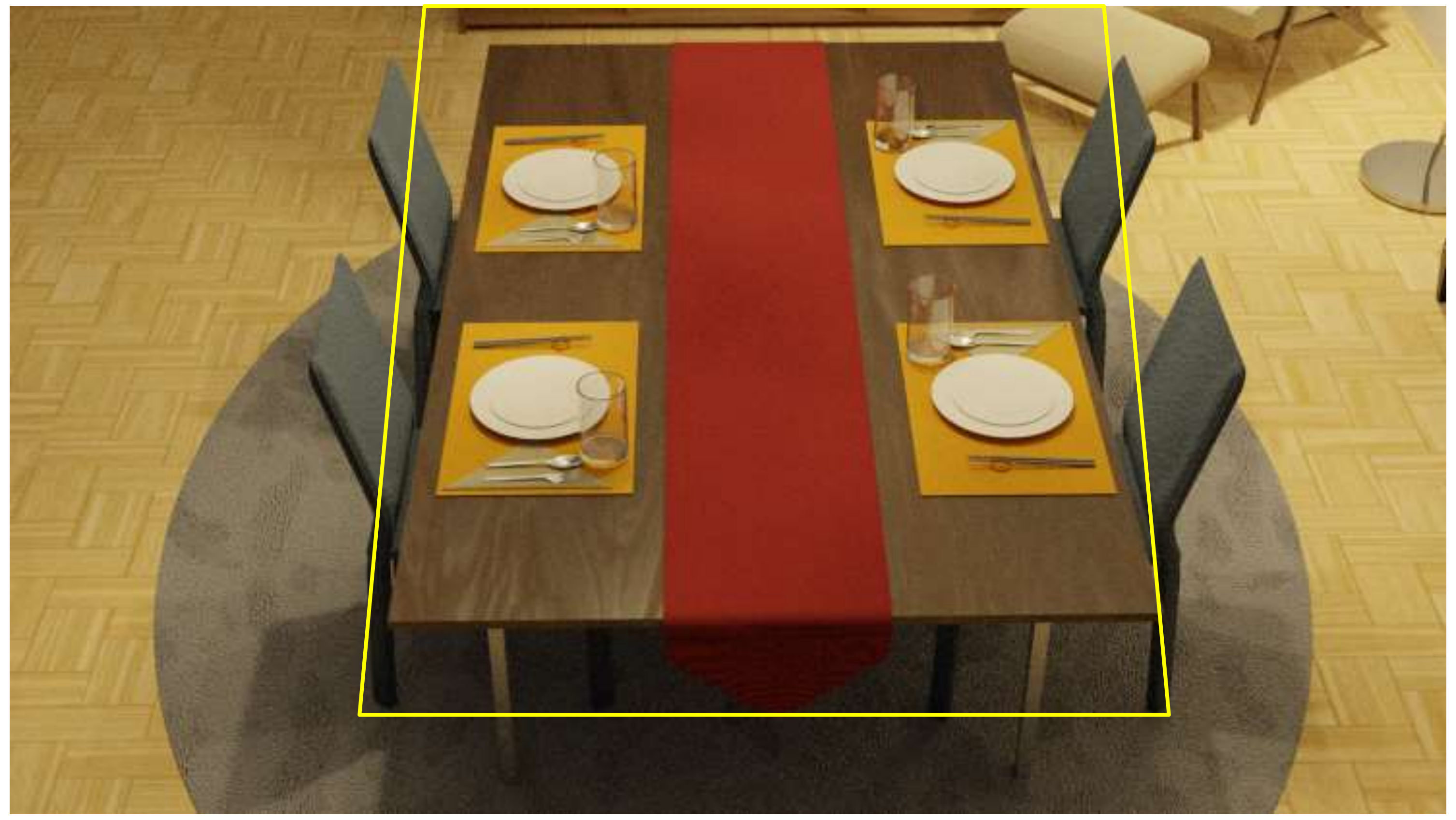




this chair has moved

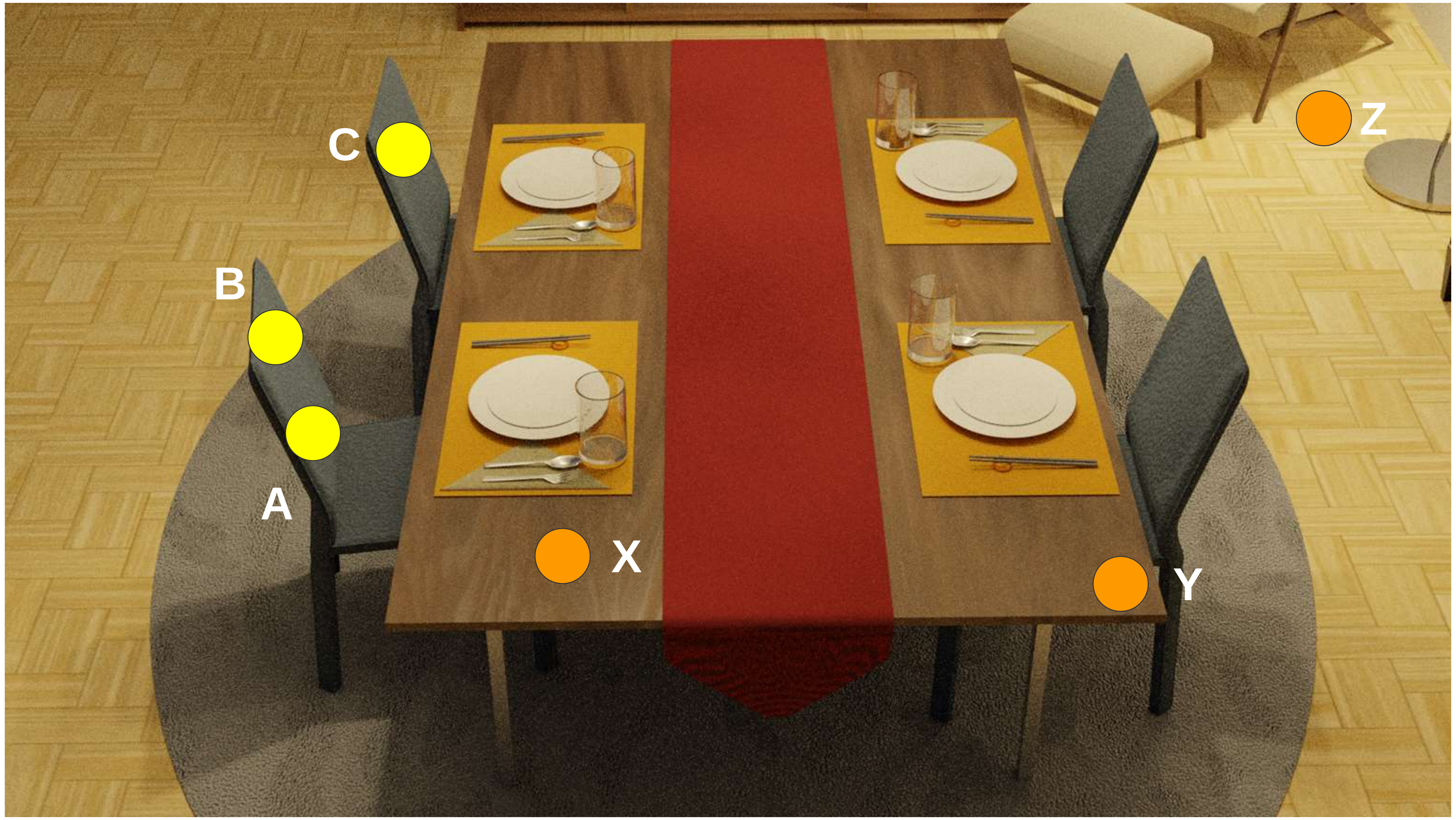


the table became longer



the table became darker

Mutual Information and 2nd Order Relationships



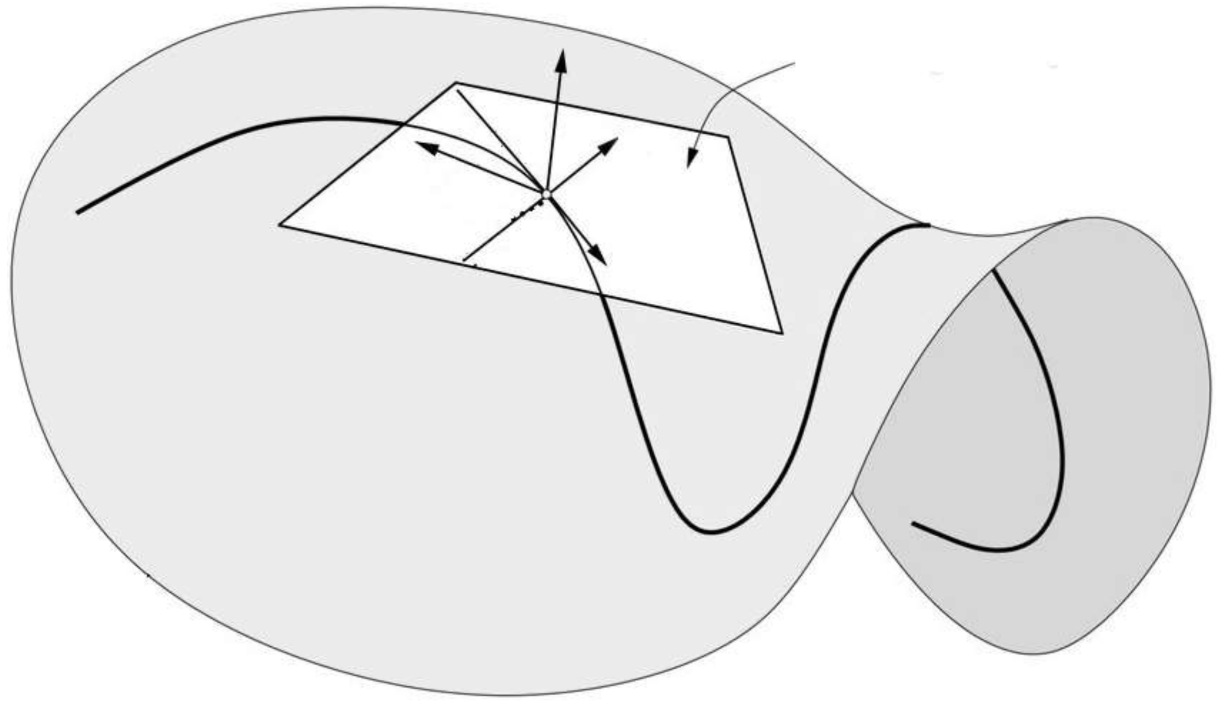
A is more correlated with B than with C $\mathbb{I}(A, B) > \mathbb{I}(A, C)$

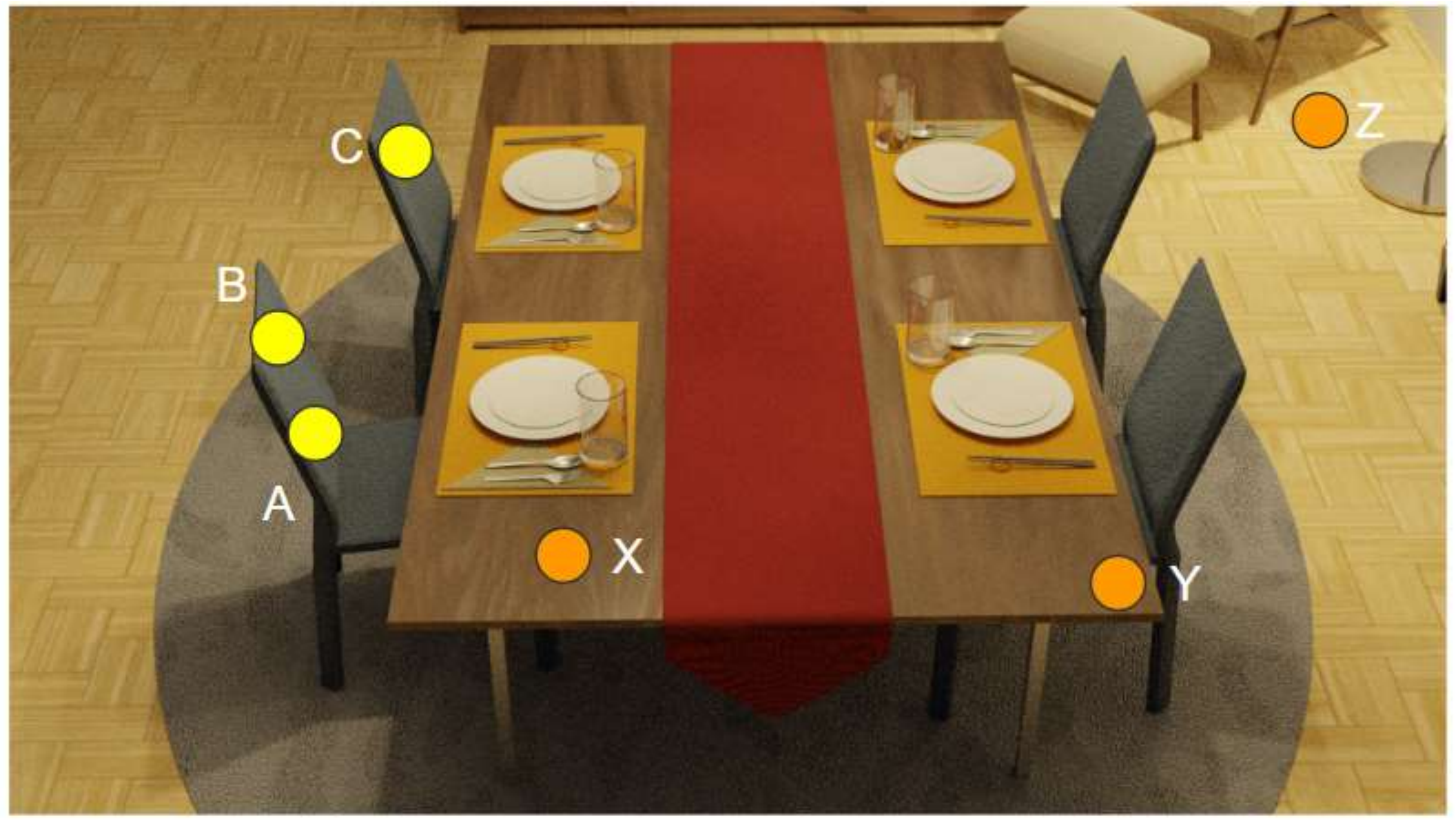
X is more correlated with Y than with Z $\mathbb{I}(X, Y) > \mathbb{I}(X, Z)$

Shaping Neural Representations

Can we shape NeRFs to better reflect mutual correlations in the scene?

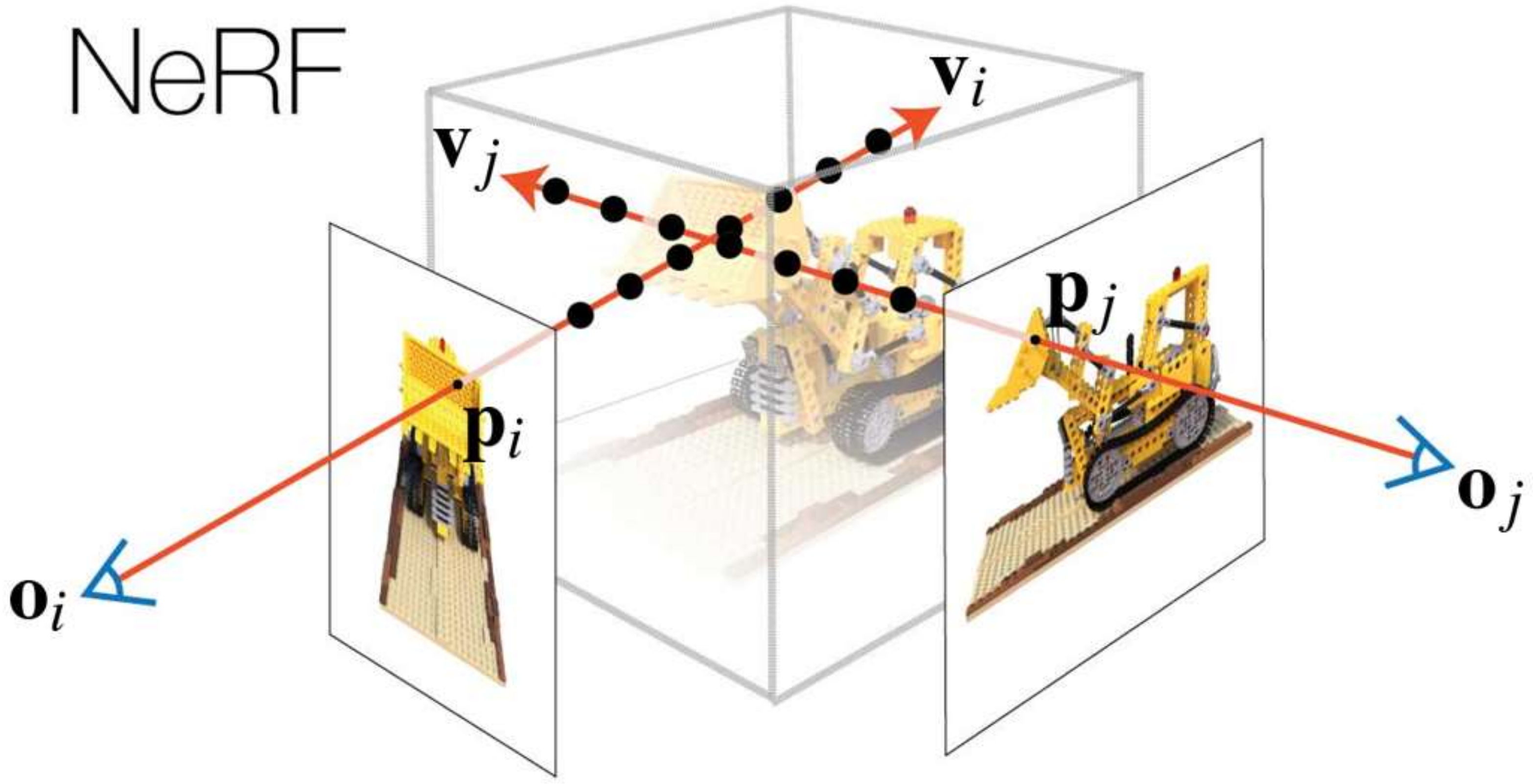
The scene tangent space





$\mathbb{I}(A,B) > \mathbb{I}(A,C)$
$\mathbb{I}(X,Y) > \mathbb{I}(X,Z)$

Mutual Information via NeRF Gradients



$$I(\mathrm{p}_i) = \Phi(\mathbf{o}_i, \mathbf{v}_i; \theta)$$
$$I(\mathrm{p}_j) = \Phi(\mathbf{o}_j, \mathbf{v}_j; \theta)$$

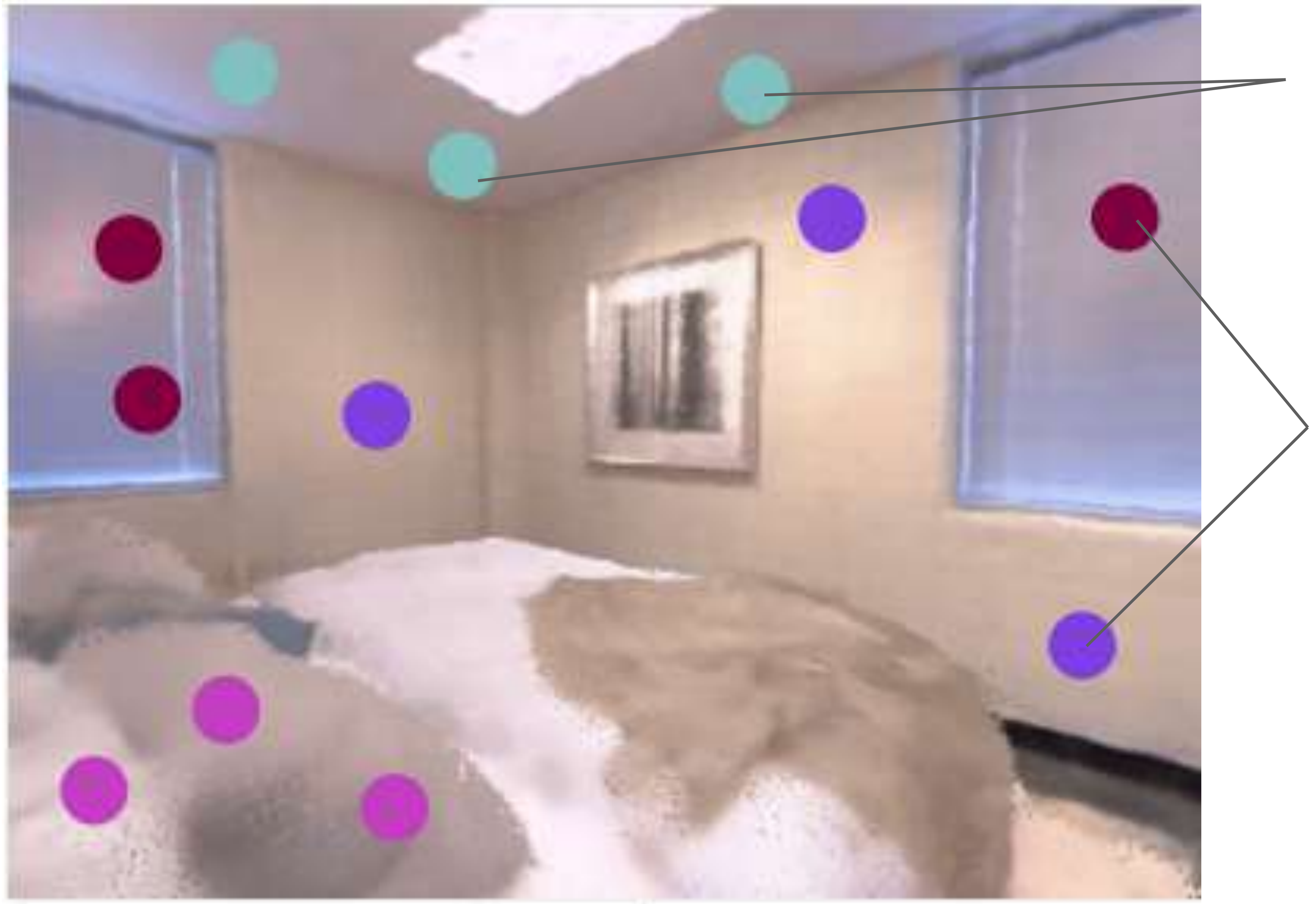
$$\hat{I}(\mathrm{p}_i) = \Phi(\mathbf{o}_i, \mathbf{v}_i; \theta^D + \mathbf{n})$$
$$\hat{I}(\mathrm{p}_j) = \Phi(\mathbf{o}_j, \mathbf{v}_j; \theta^D + \mathbf{n})$$

Mutual information $\mathbb{I}$

$$\boxed{\mathbb{I}(\hat{I}(\mathrm{p}_i), \hat{I}(\mathrm{p}_j)) \approx \left| \cos\left(\frac{\partial \Phi_i}{\partial \theta^D}, \frac{\partial \Phi_j}{\partial \theta^D} \right) \right|}$$

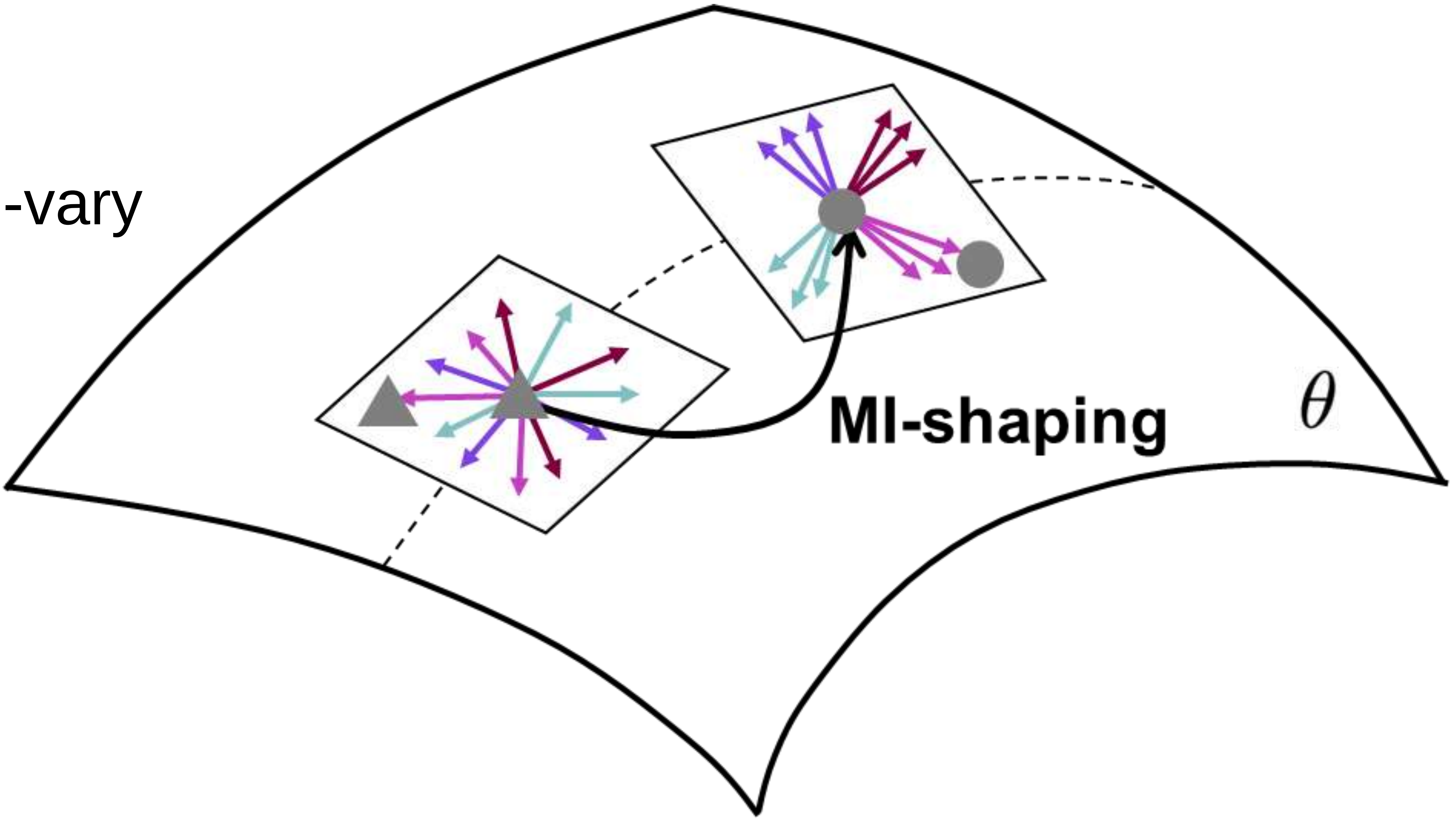
Inter-pixel correlations are captured by cosine similarity of the NeRF Jacobians

Setting up Semantic “Neuronal Resonances” via Aligning Gradients



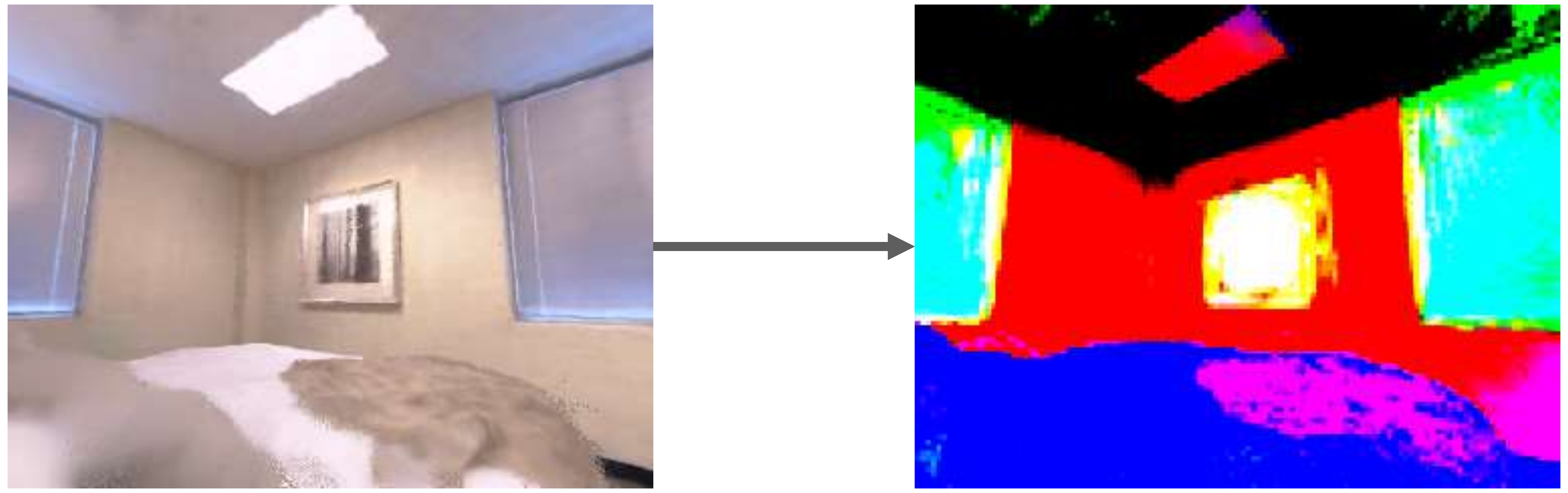
These pixels should co-vary

But these should not

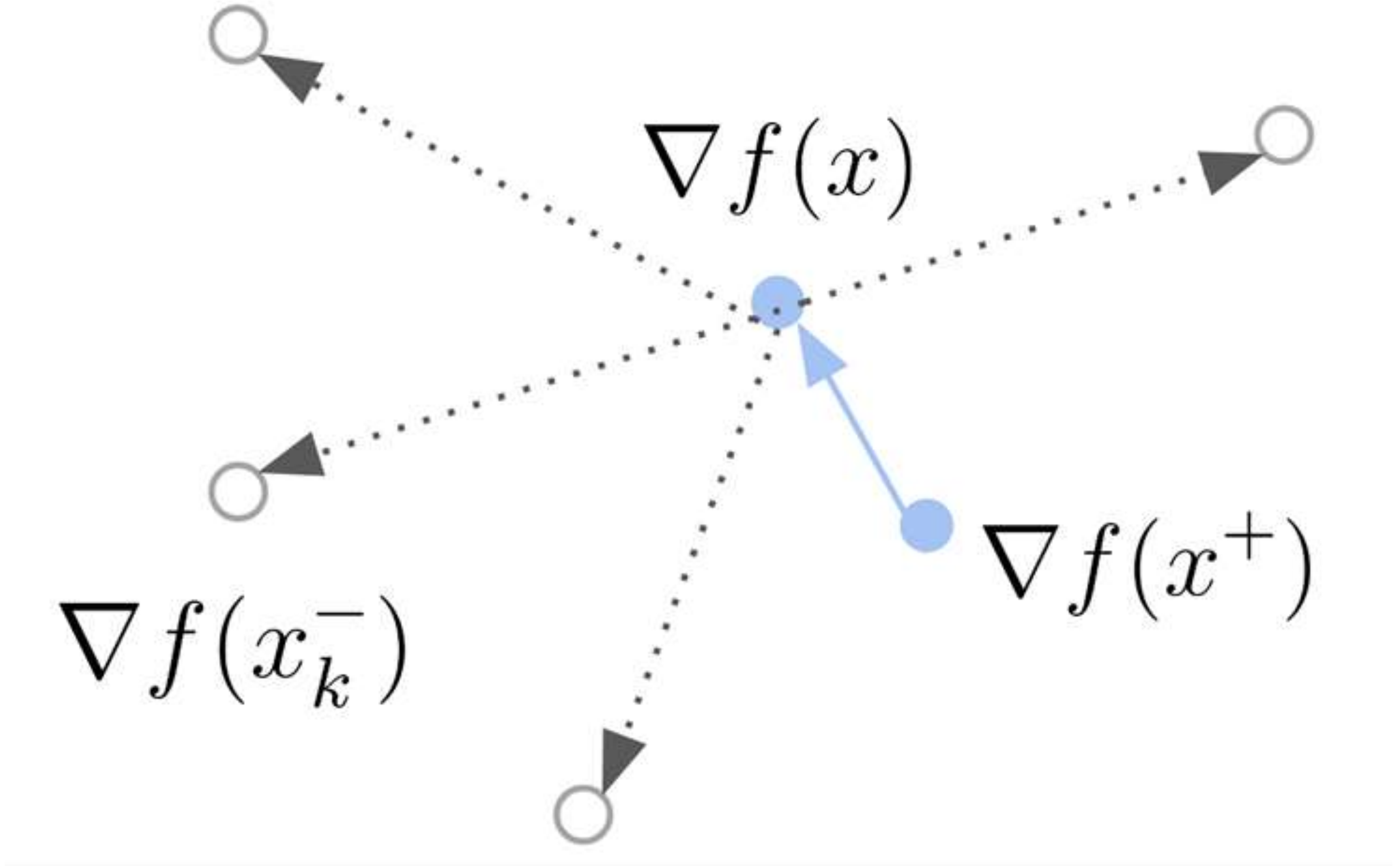


Shaping co-aligns gradients of correlated points (here points of the same semantic class)

NeRF MLP Shaping via Mutual Information Gradients



General purpose image feature: DINO

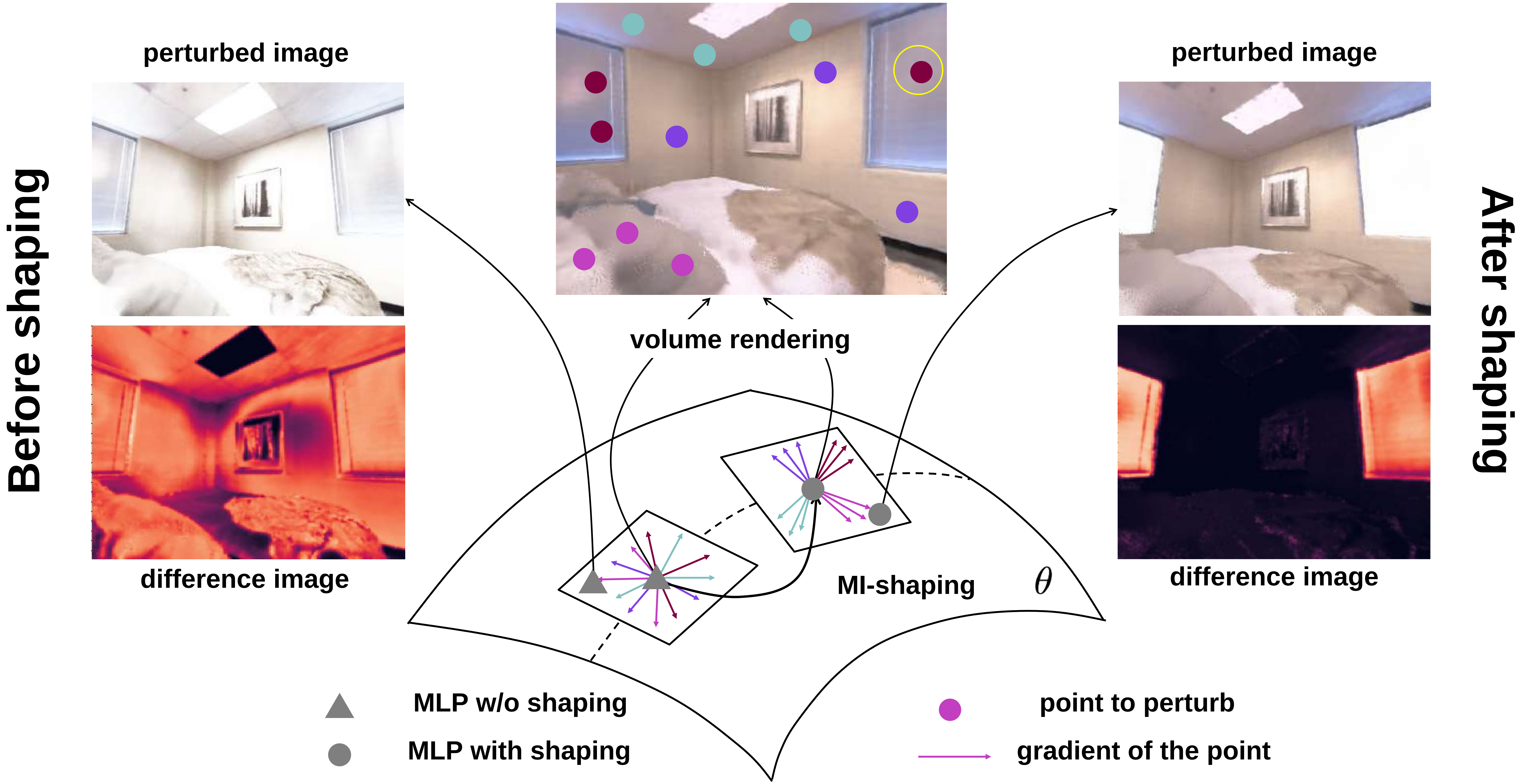


InfoNCE contrastive loss on gradients + reconstruction loss

Obtain some source of semantic affinity

Contrastive training aligning gradients

NeRF Shaping Causes Gradient Alignment



Before shaping

perturbed image

difference image

volume rendering

MI-shaping

θ

After shaping

perturbed image

difference image

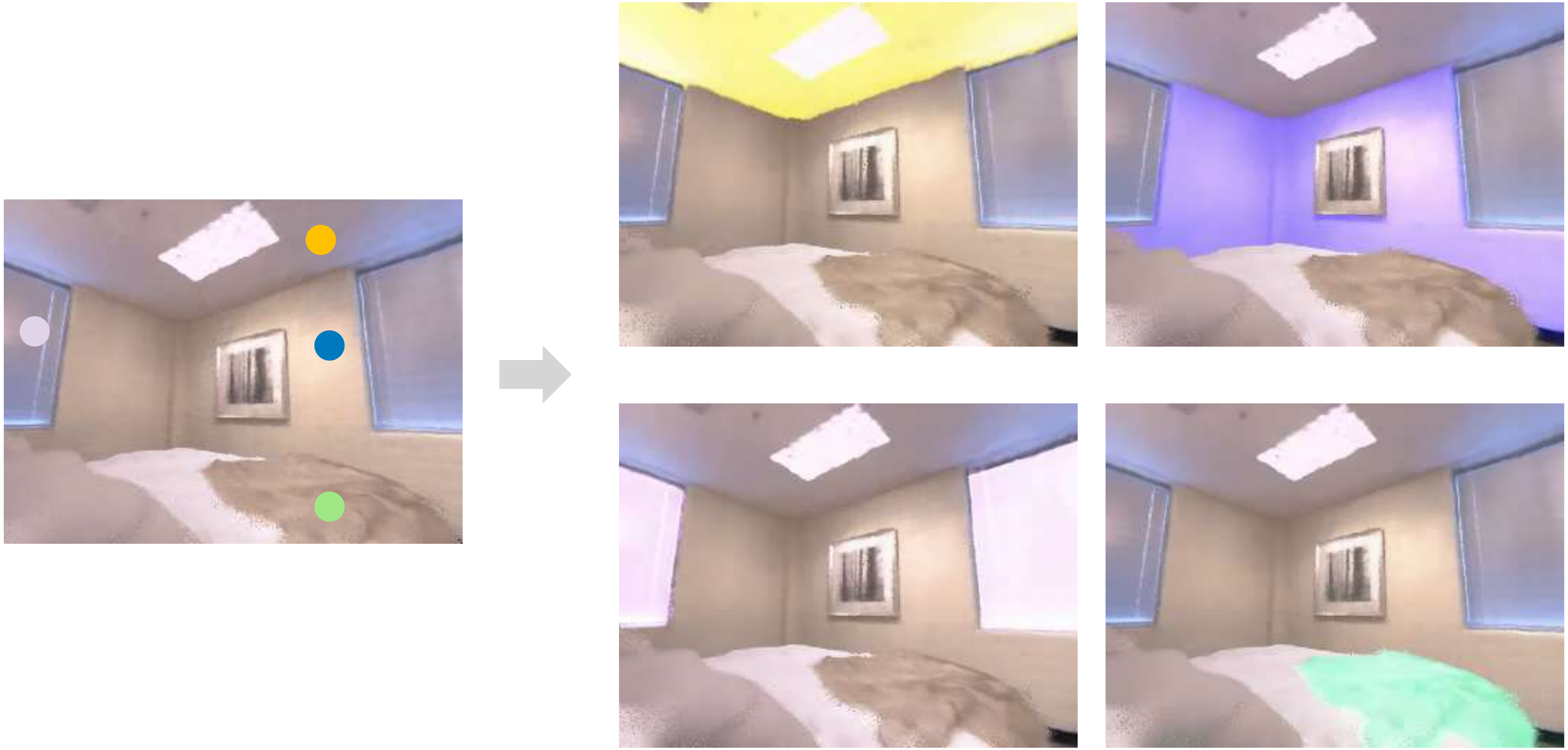
▲ MLP w/o shaping

● MLP with shaping

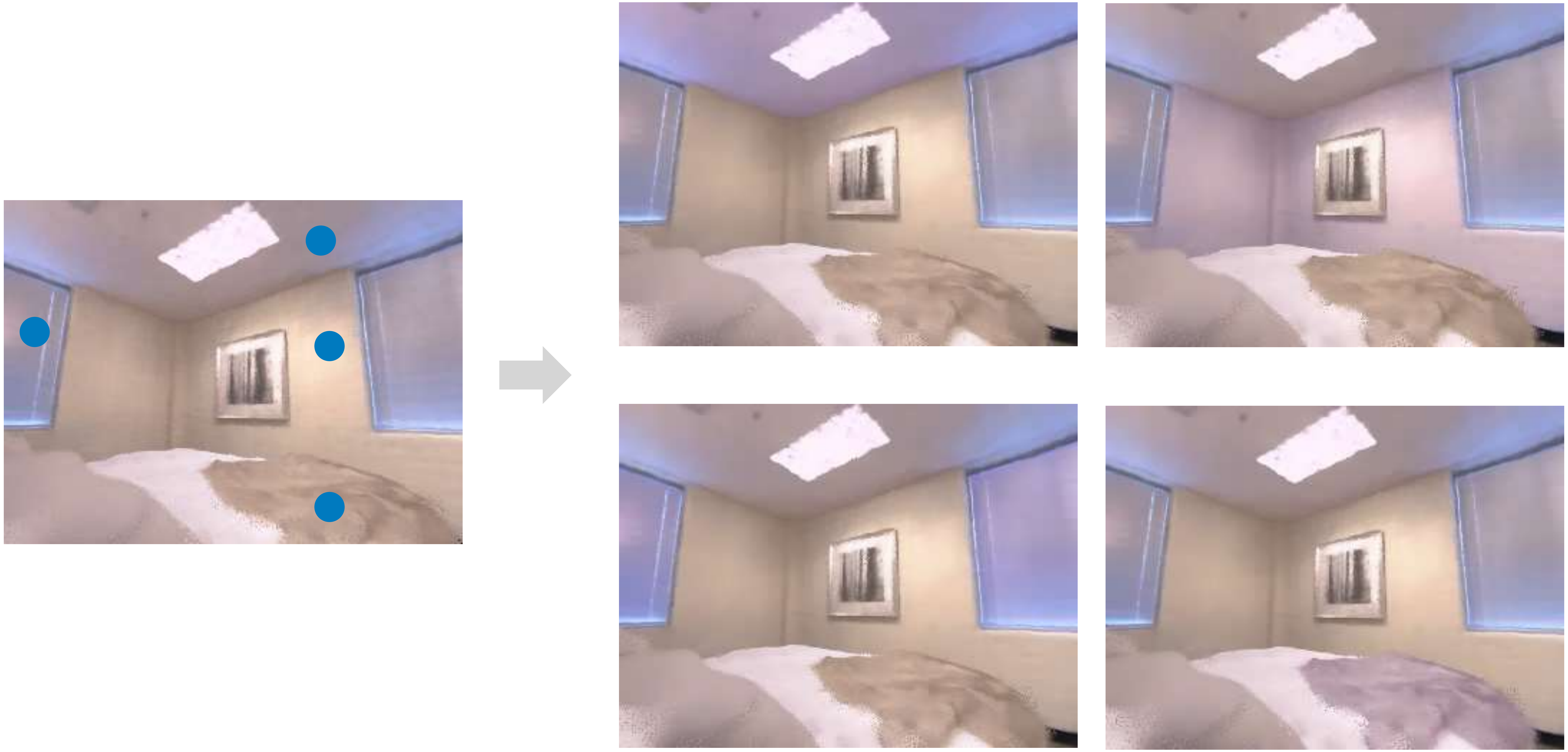
● point to perturb

→ gradient of the point

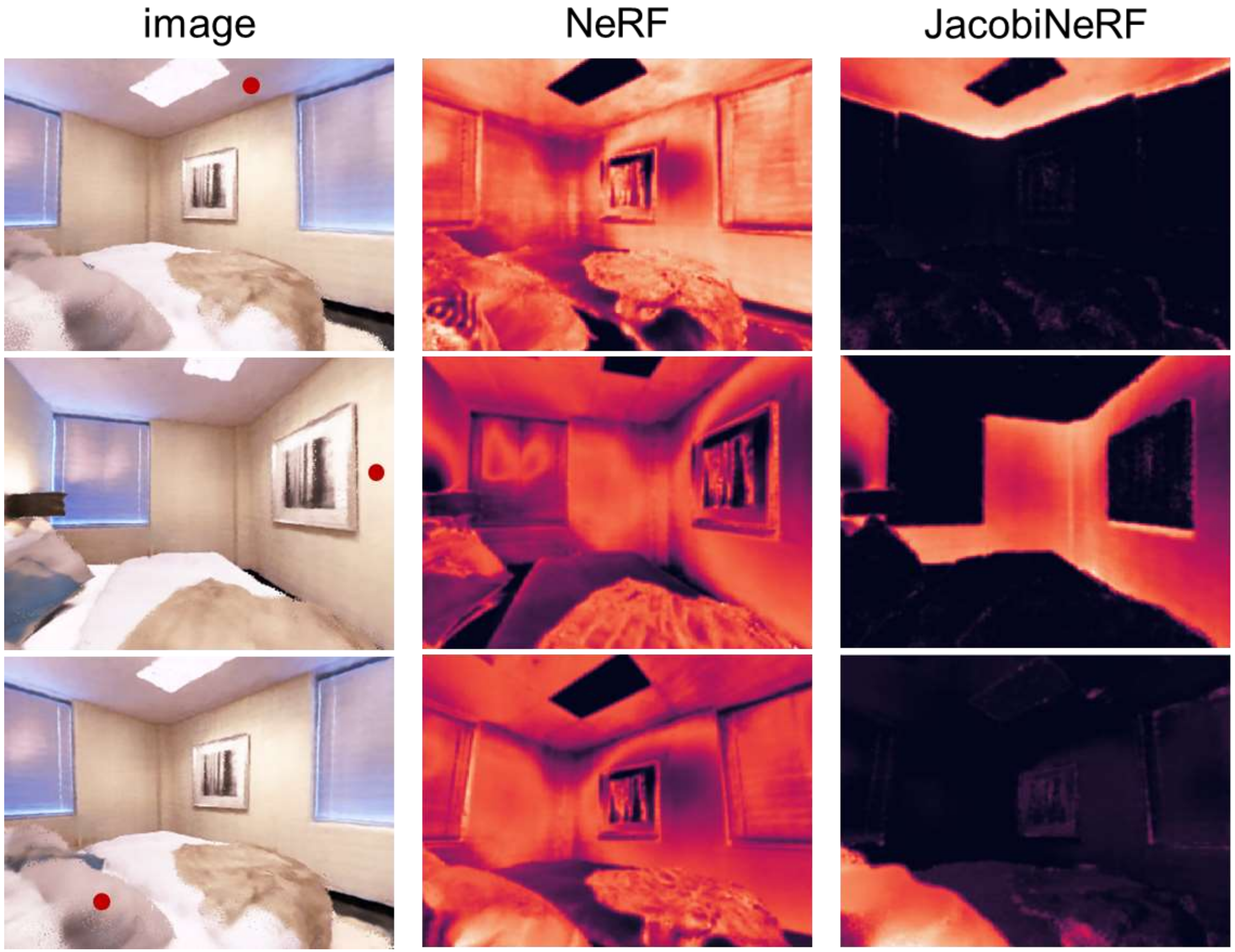
Application: Editing Appearance



Application: Editing Appearance



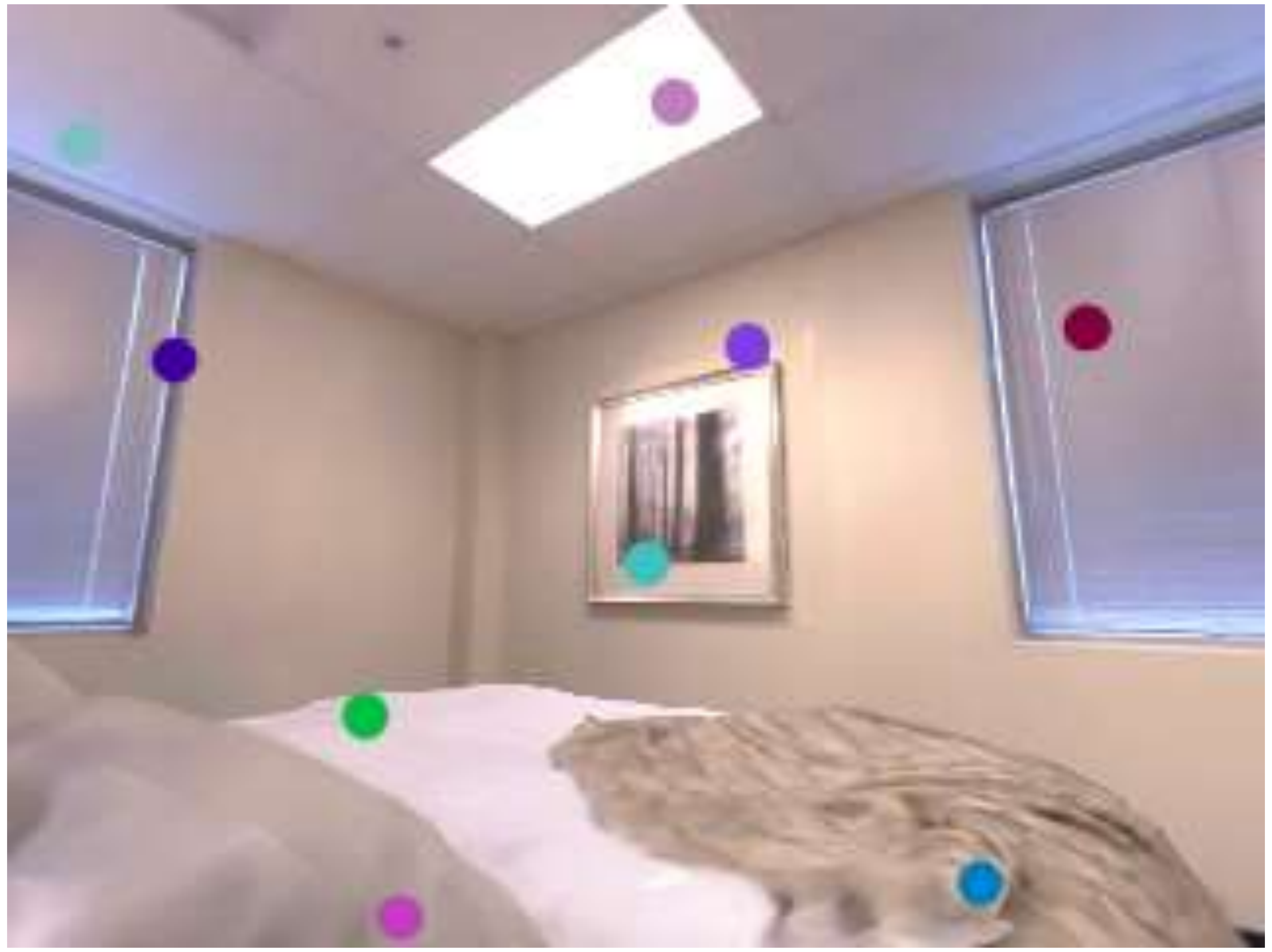
Application: Entity Selection



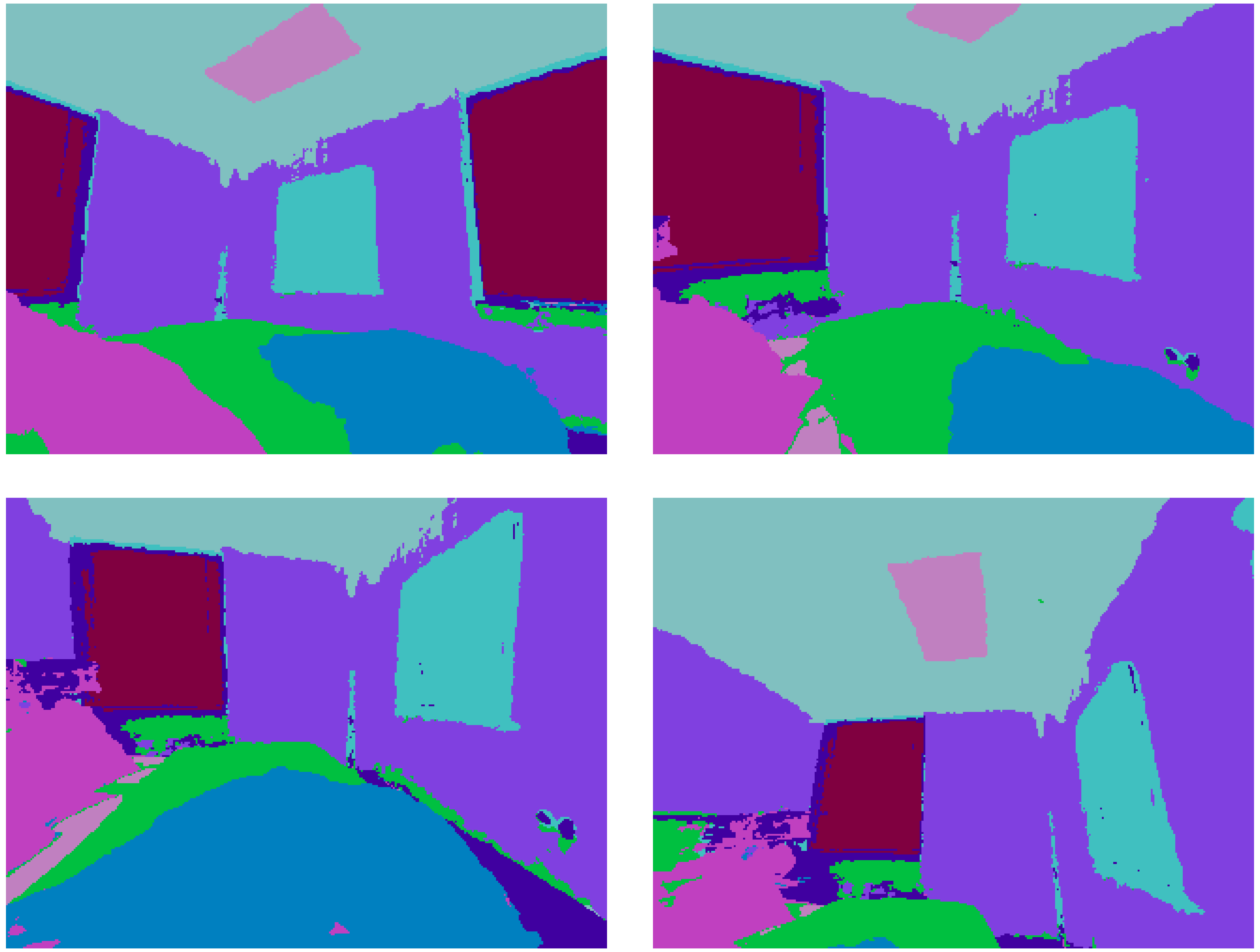
From a single point we can select an entire semantic entity.

Application: Label Propagation

Acquire dense labels of a scene given sparse annotations.



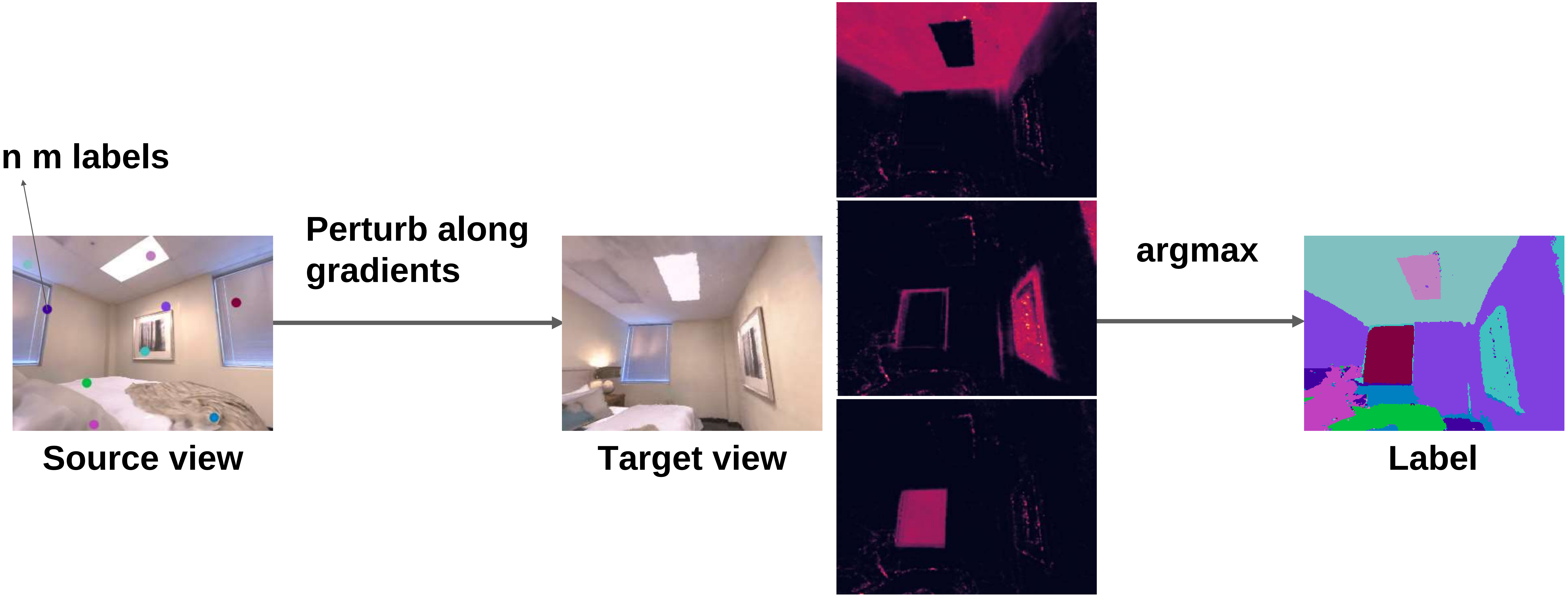
Label one pixel for each class from one view



Dense label for any view

Propagation Through Resonances

2D version (JacobiNeRF-2D)



Given m labels

Perturb along gradients

argmax

Source view

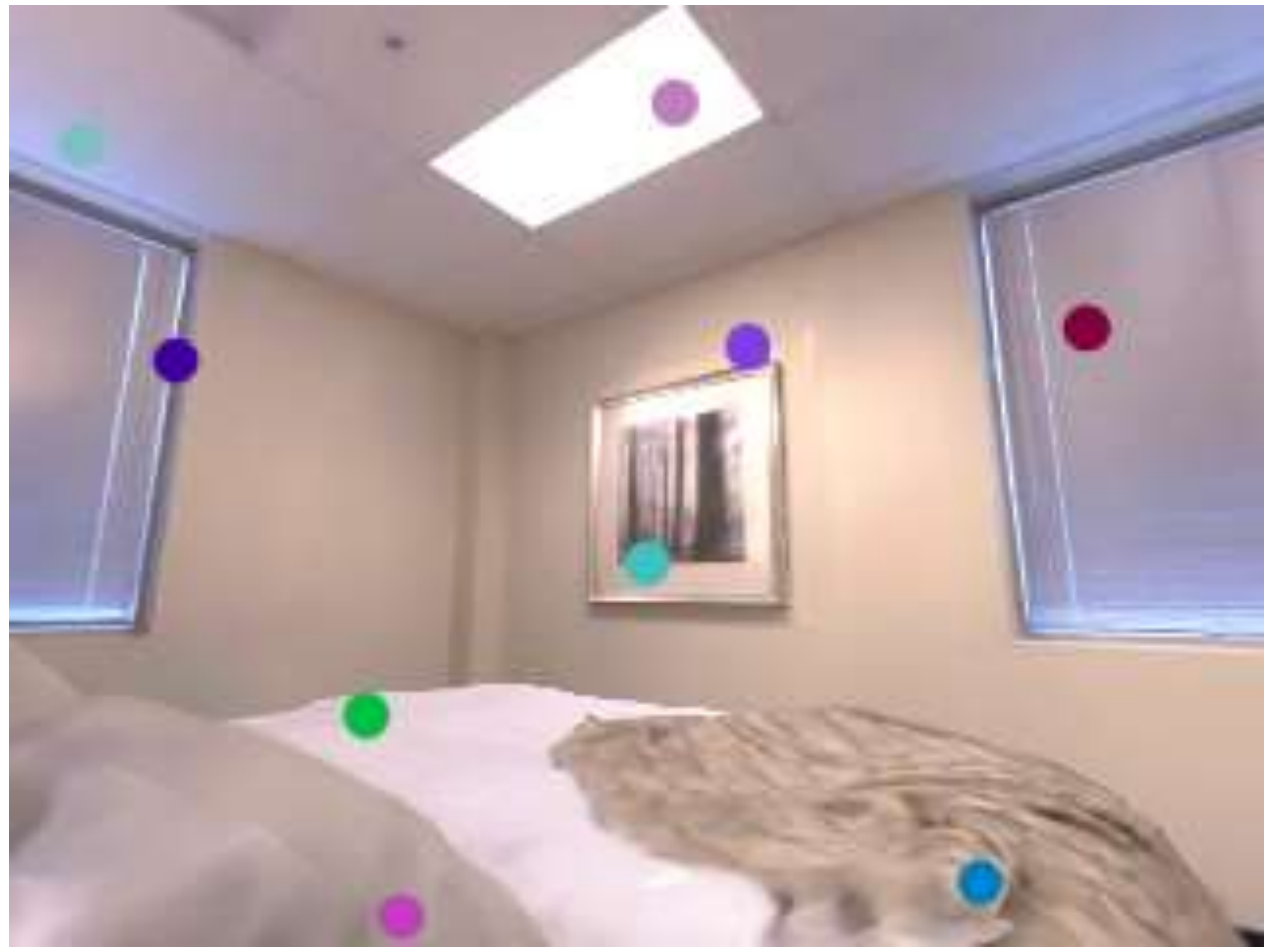
Target view

Label

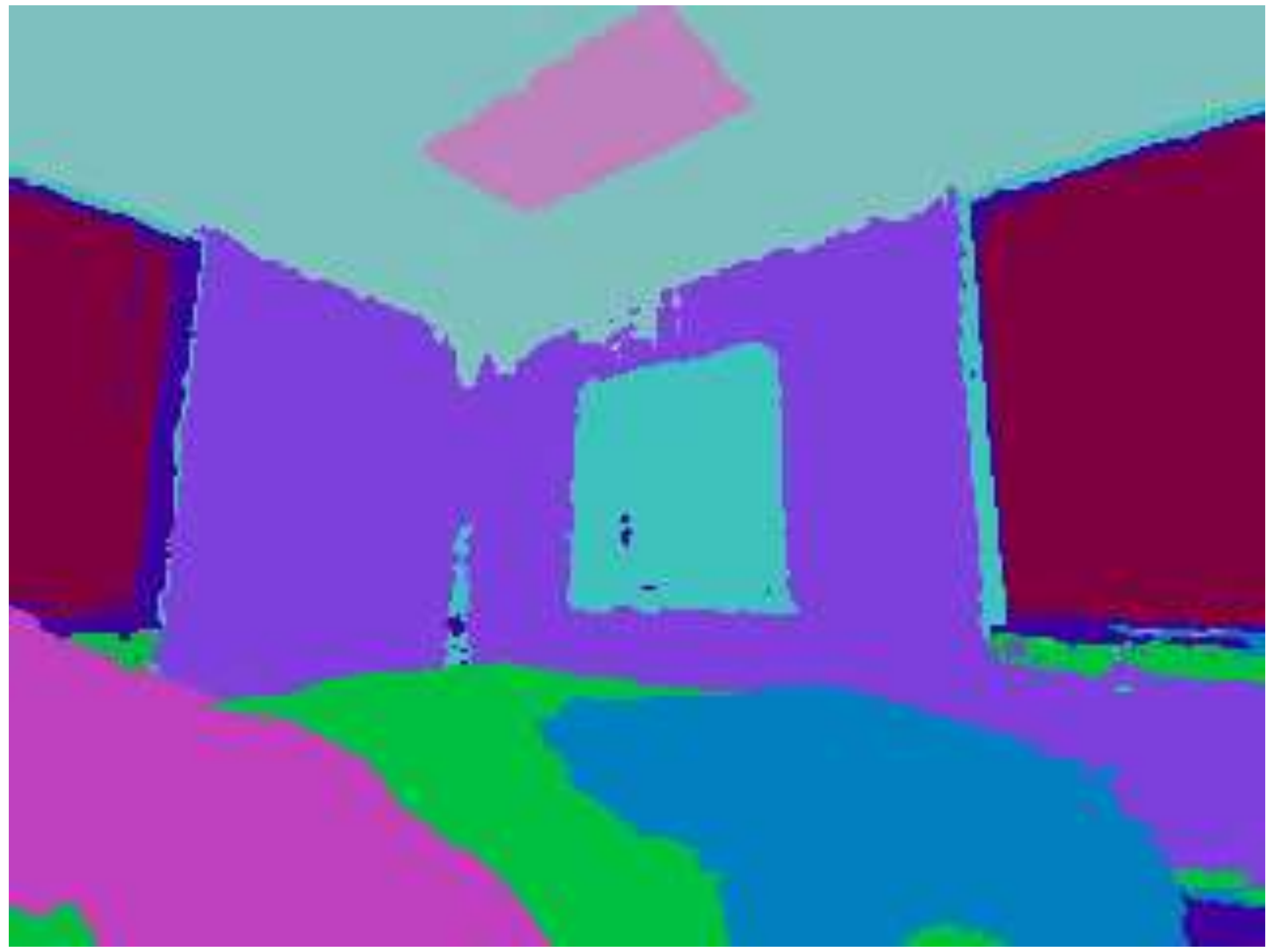
……

m responses

Semantic Segmentation (sparse 1pix/class, Replica)

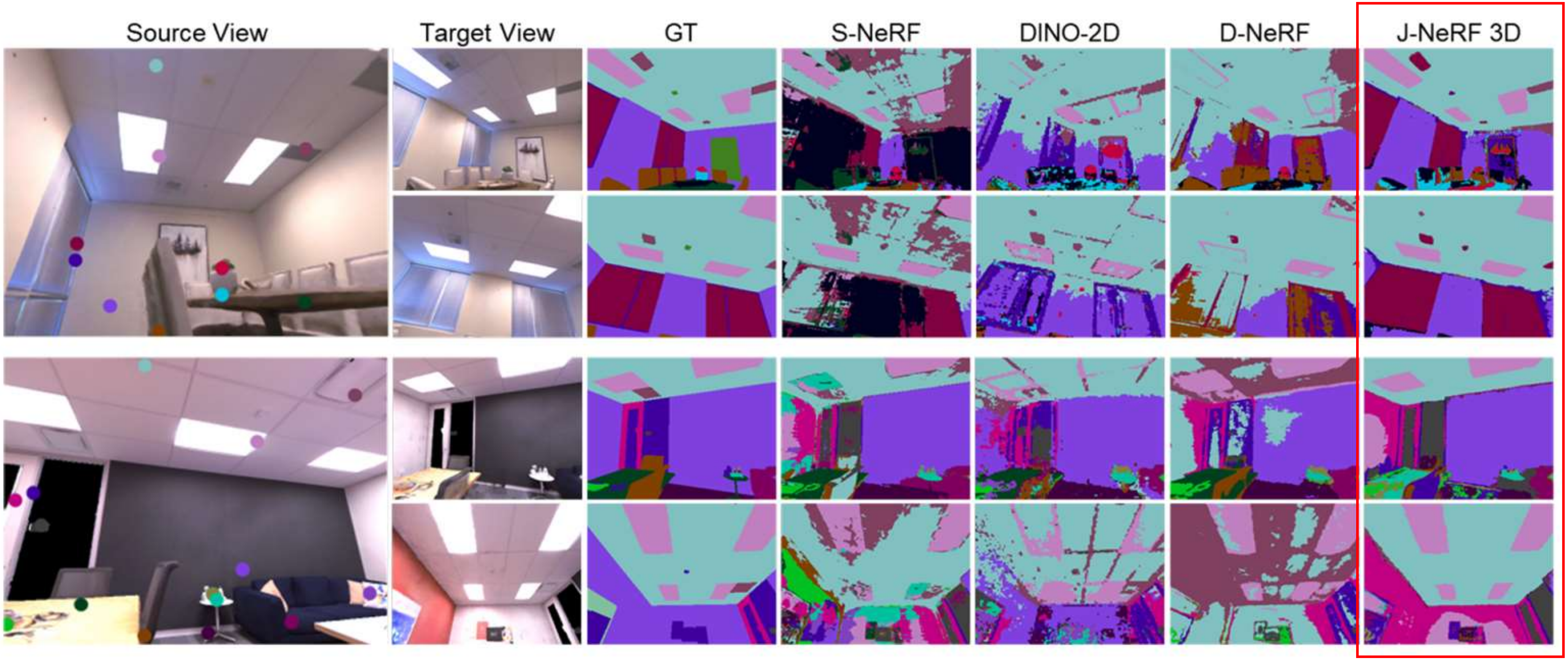


Given label

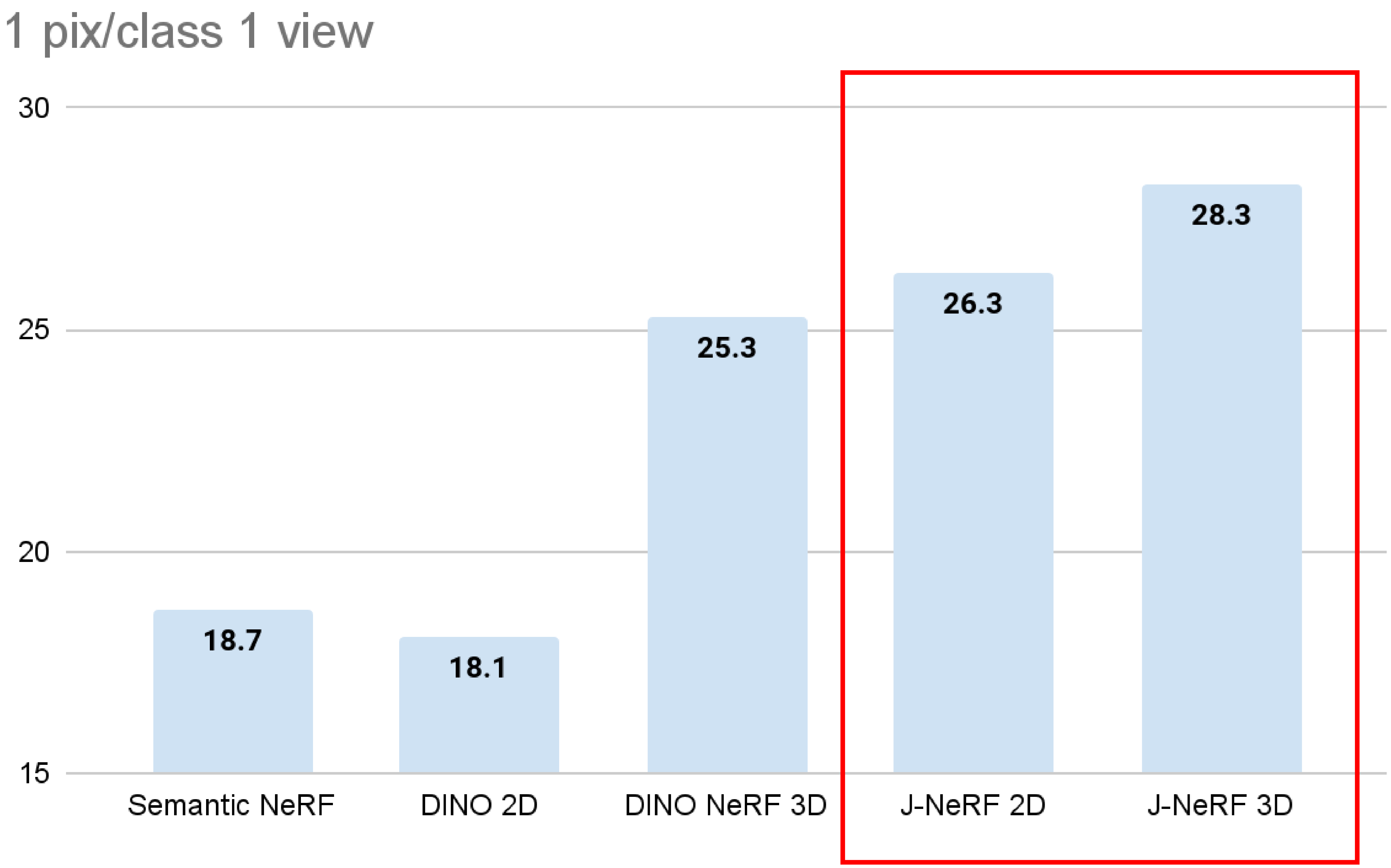


J-NeRF 3D

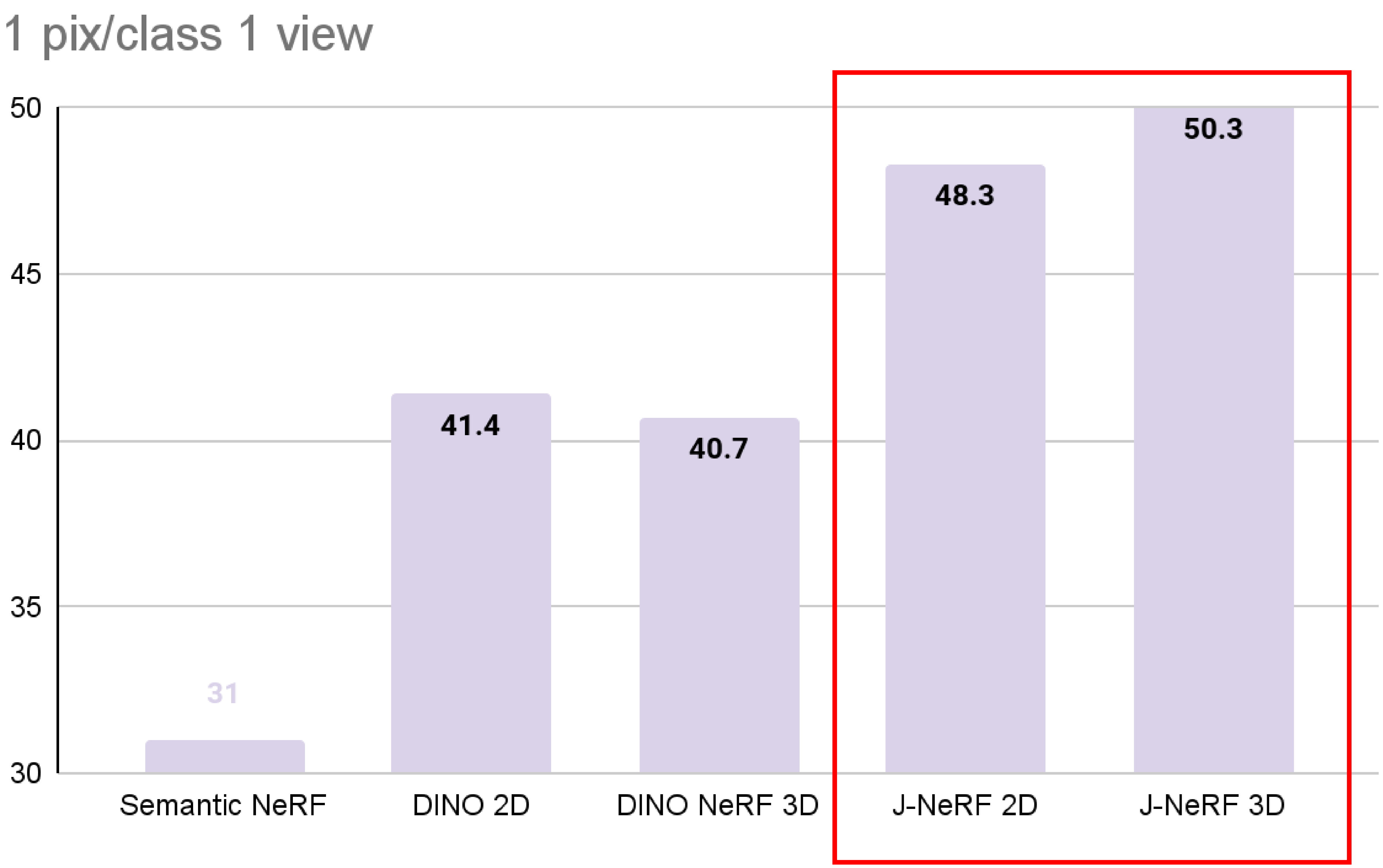
Semantic Segmentation (sparse 1pix/class, Replica)



Semantic Segmentation (sparse 1pix/class, Replica)



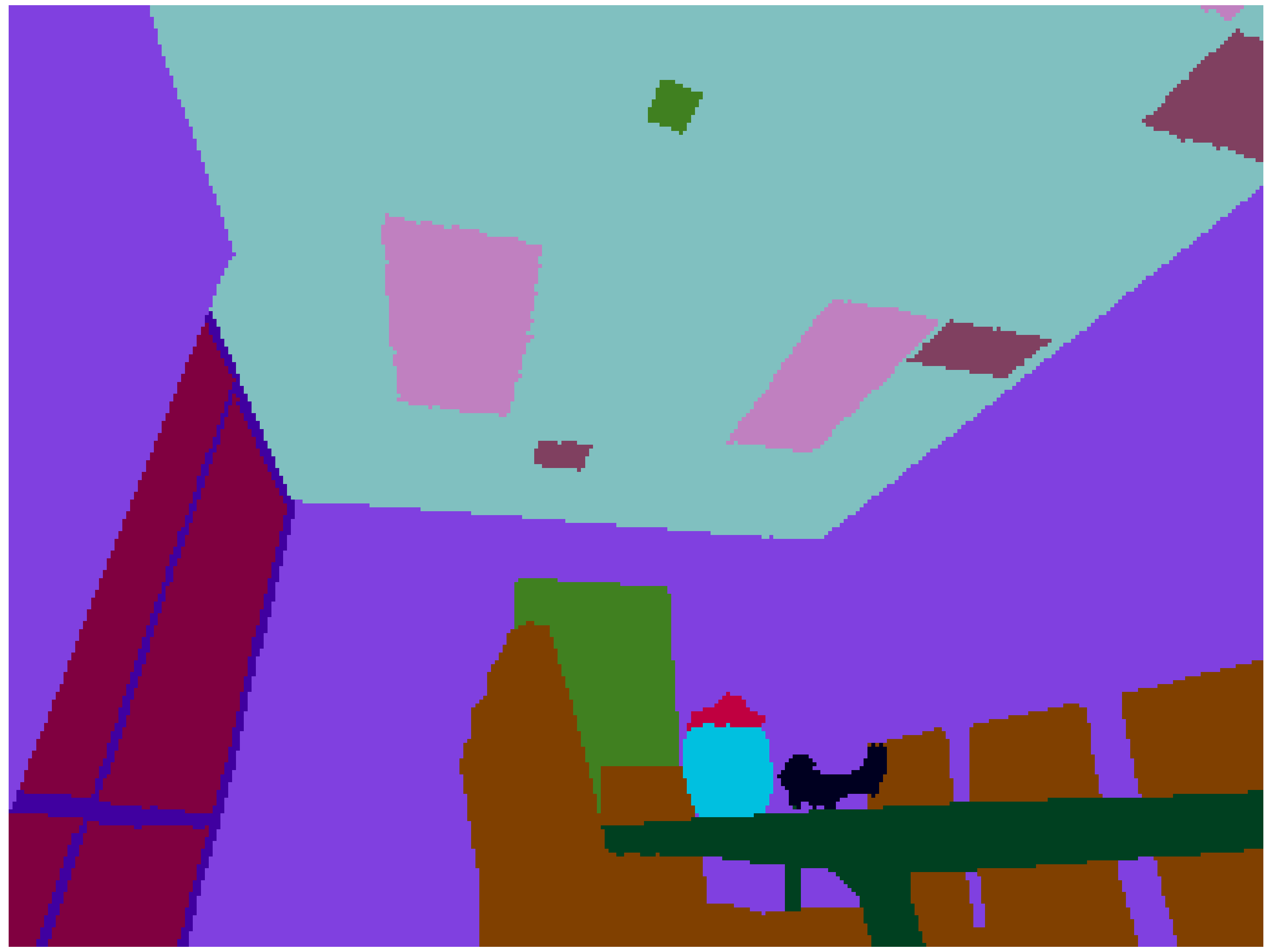
mIoU



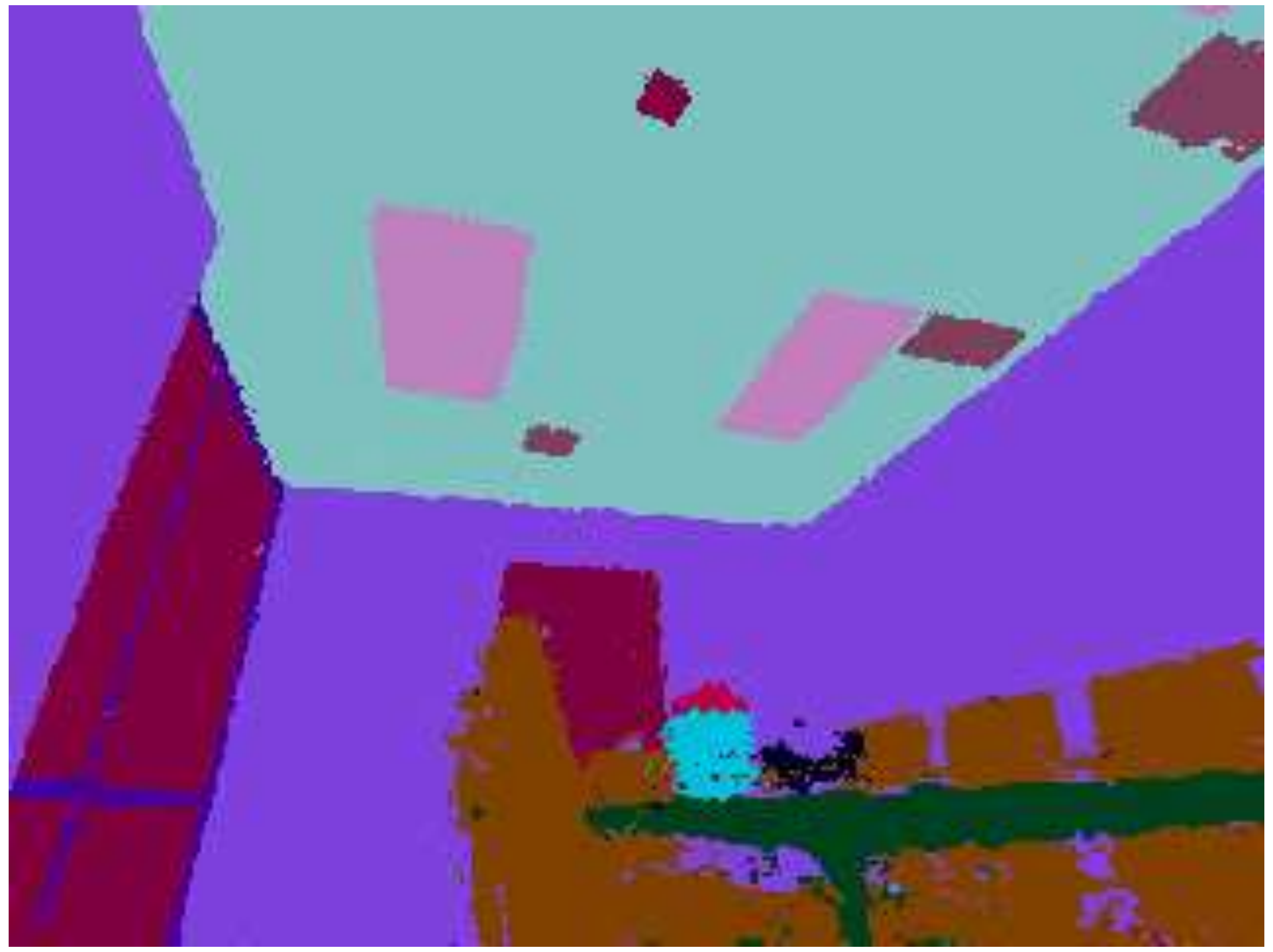
Acc

Average results on 7 scenes, 180 test views for each scene

Semantic Segmentation (dense 1view, Replica)

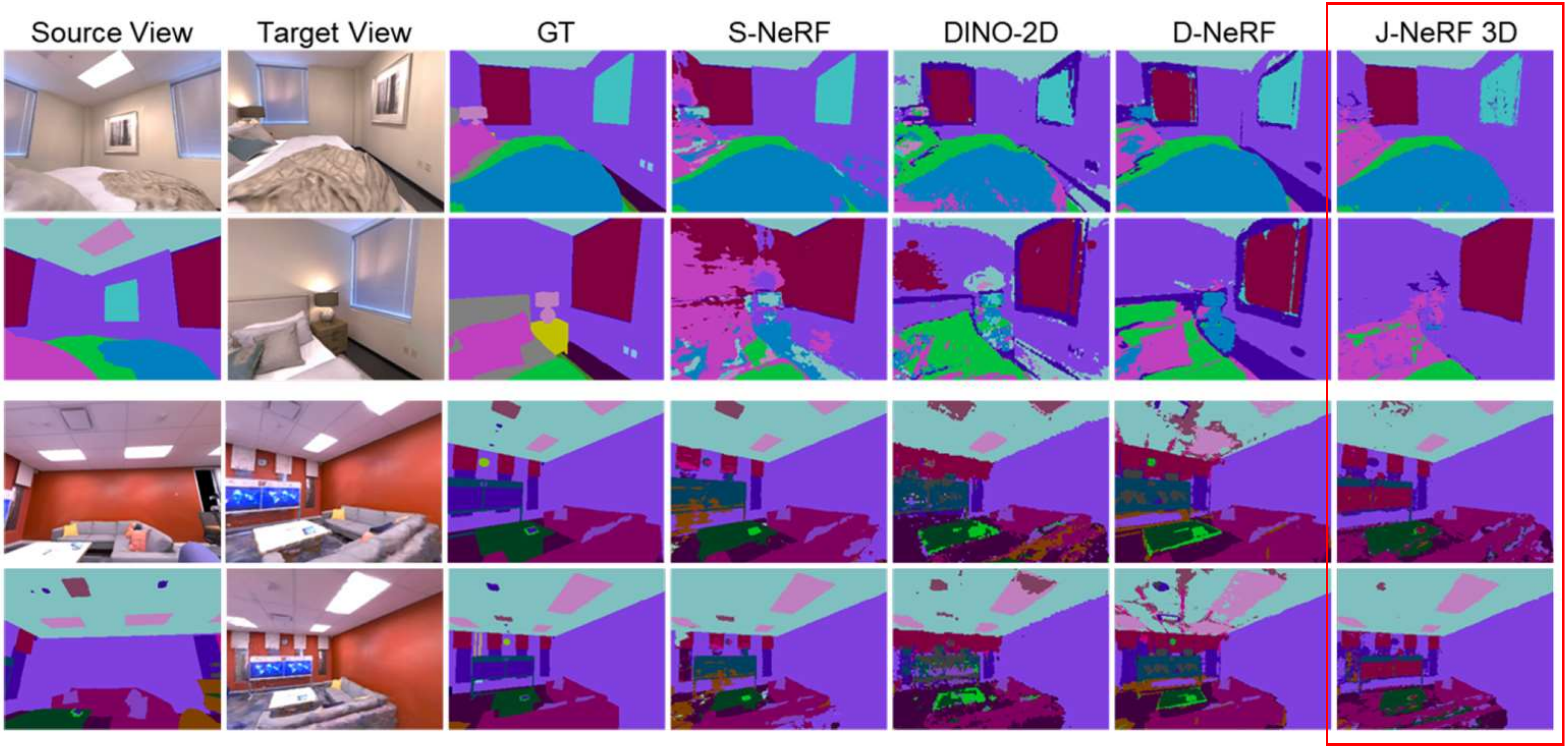


Given label

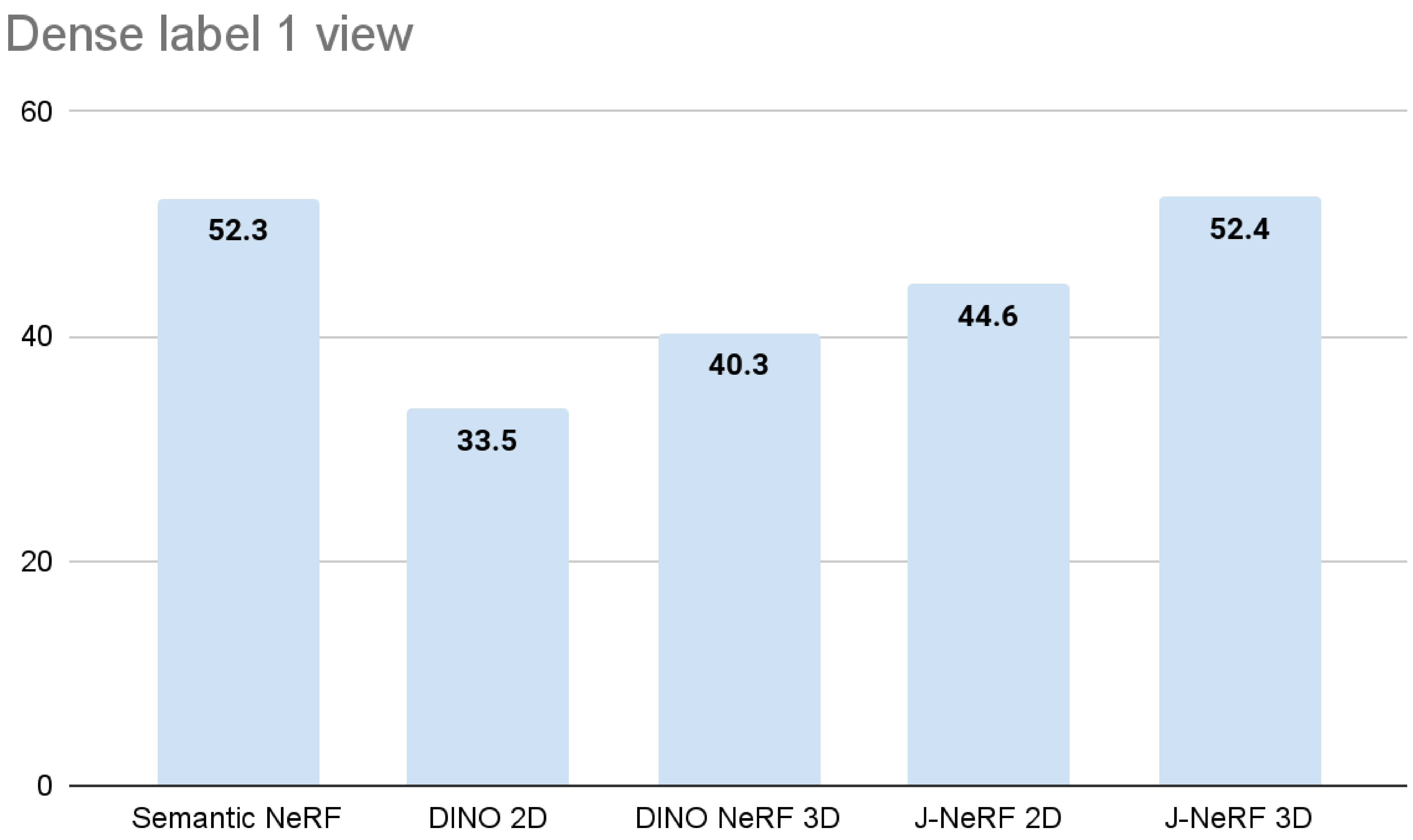


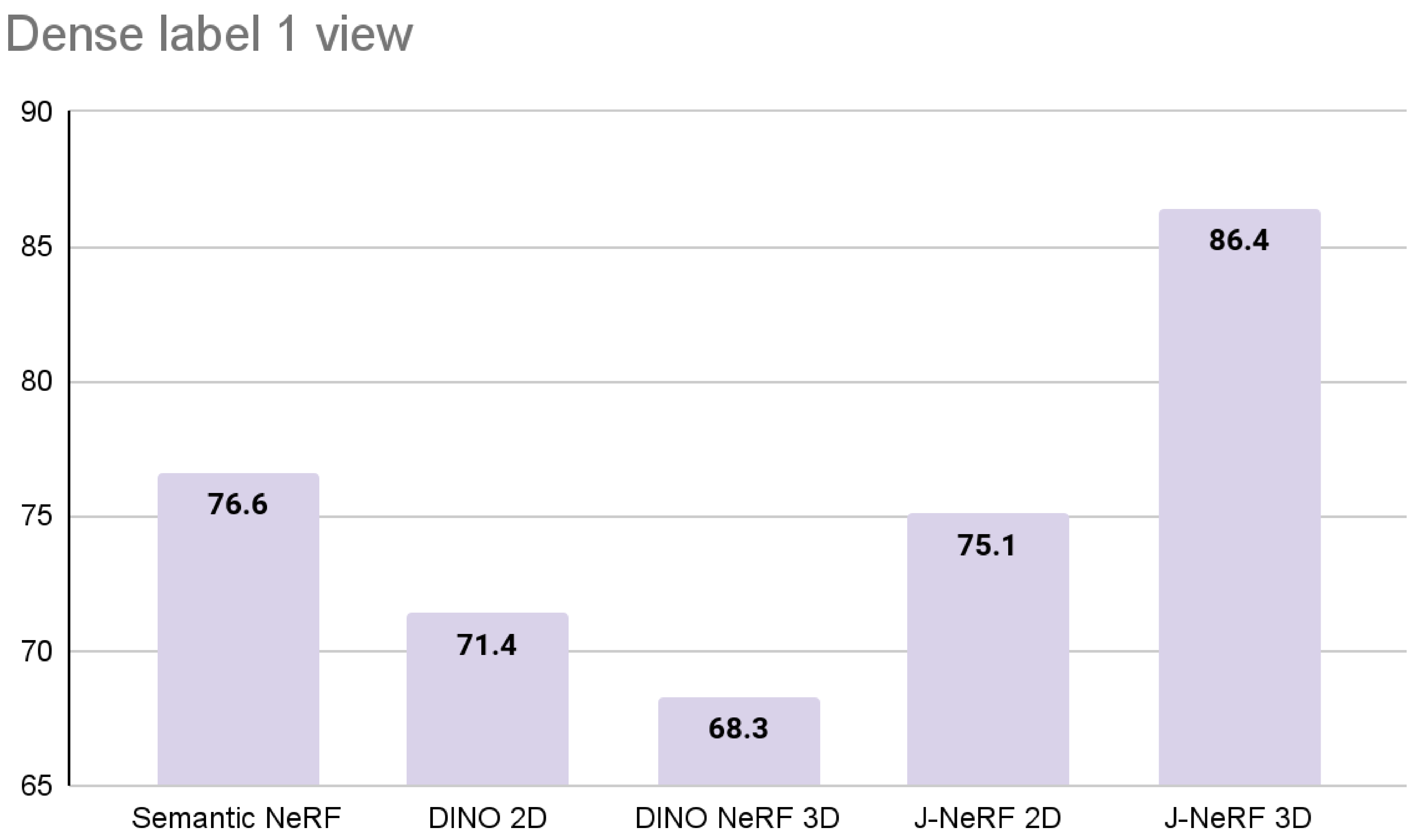
J-NeRF 3D

Semantic Segmentation (dense 1view, Replica)



Semantic Segmentation (dense 1view, Replica)



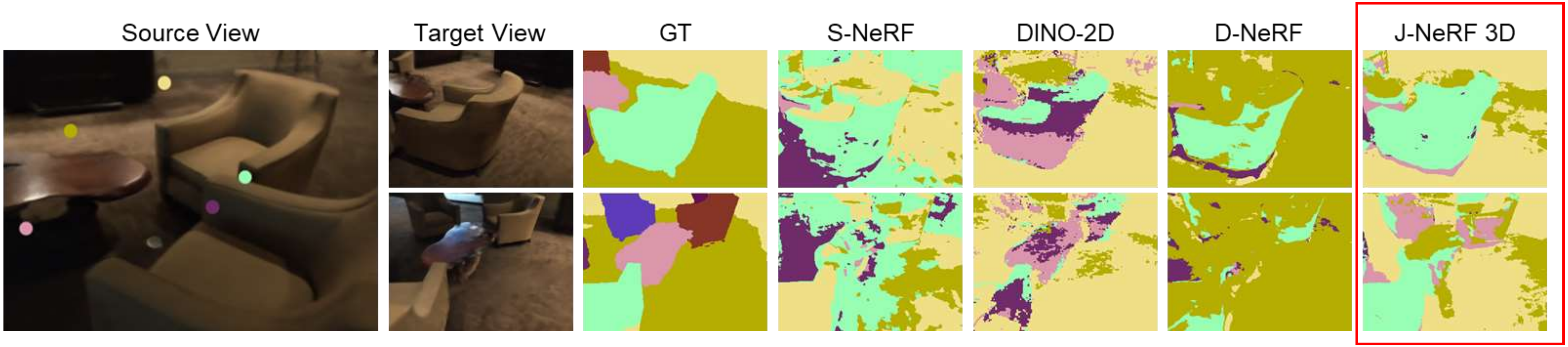


mIoU

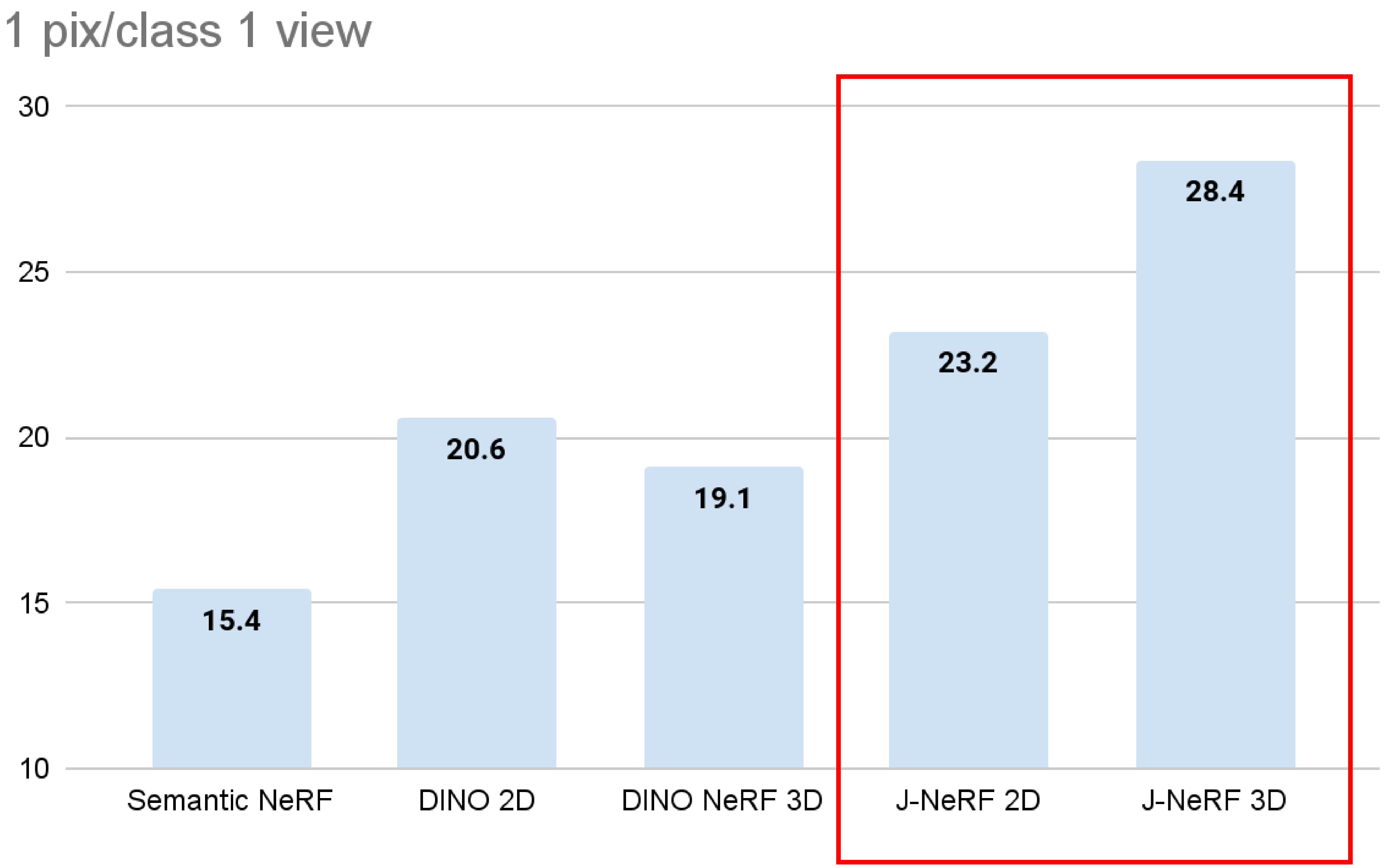
Acc

Average results on 7 scenes, 180 test views for each scene

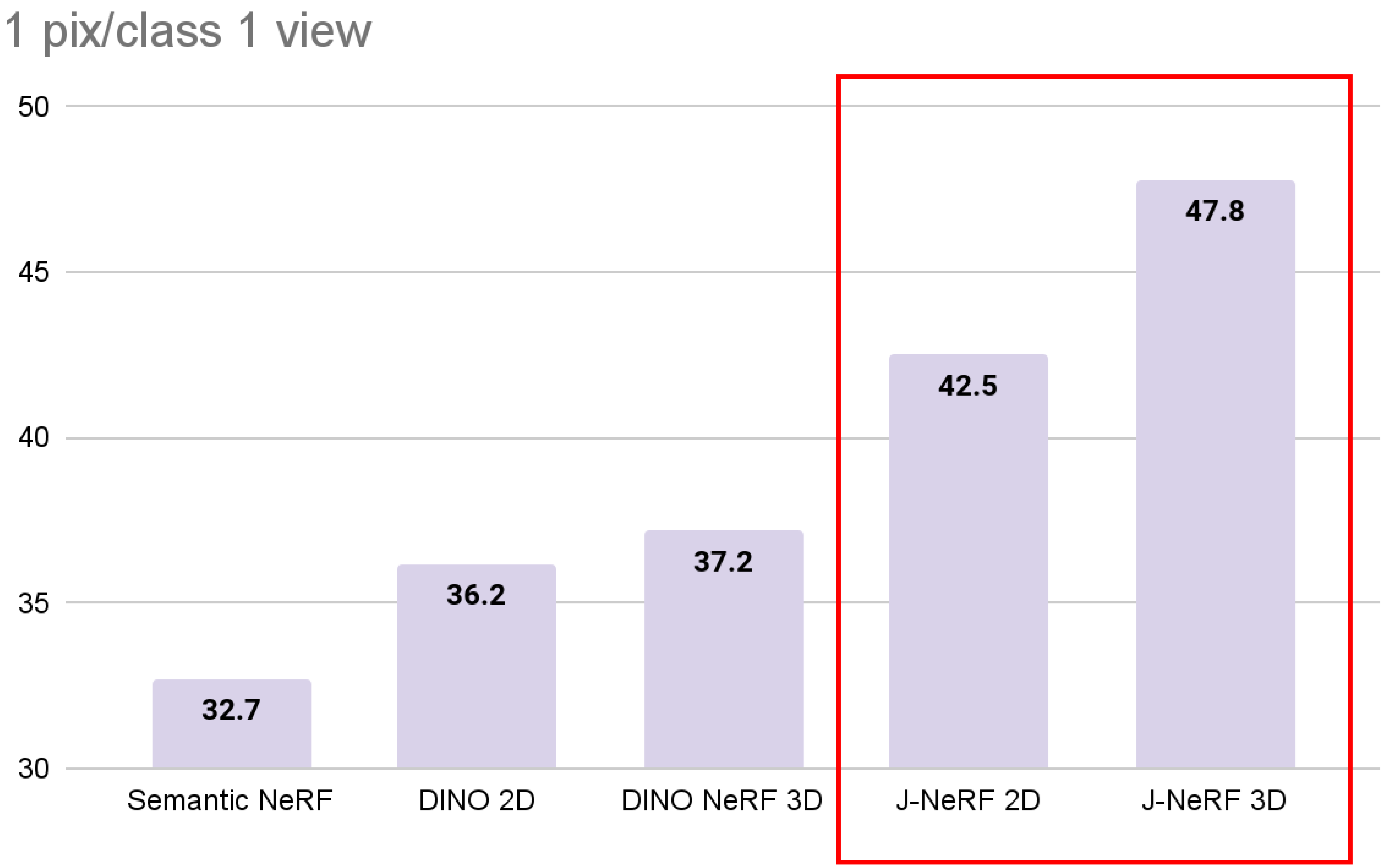
Instance segmentation (sparse, 1pix/instance, ScanNet)



Instance segmentation (sparse, 1pix/instance, ScanNet)



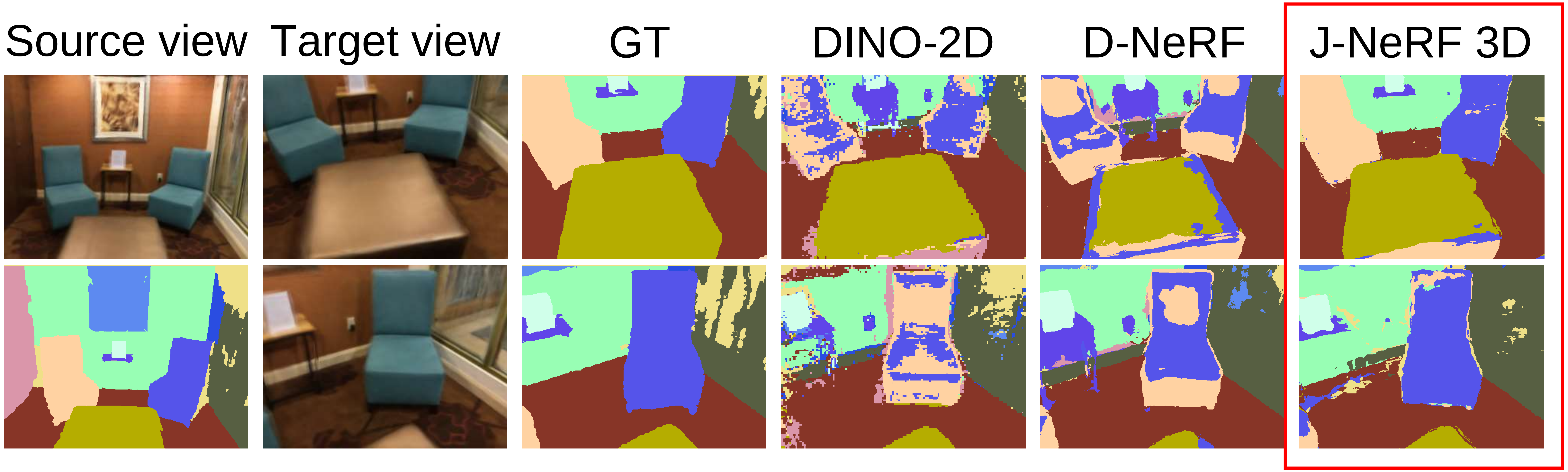
mIoU



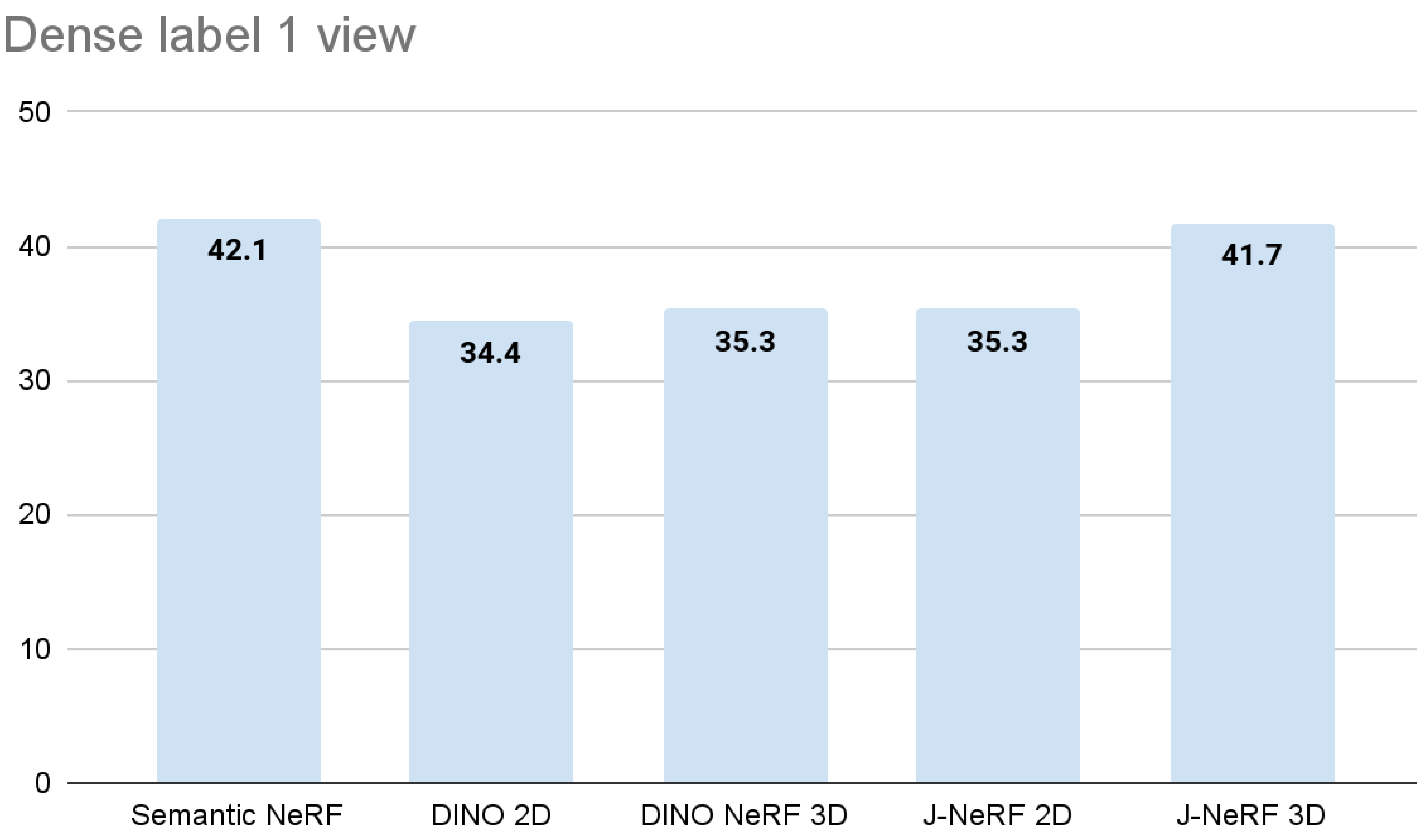
Acc

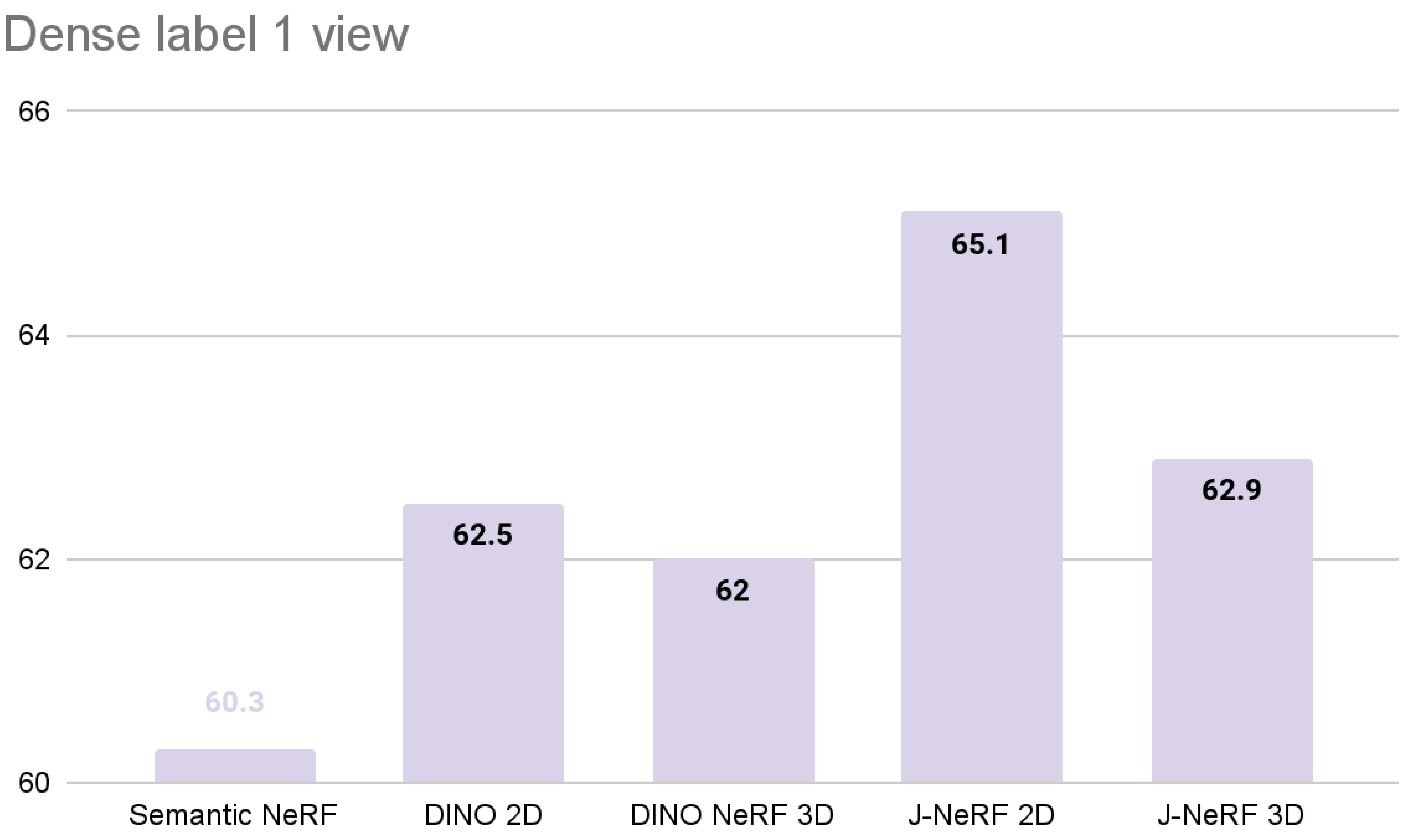
Average results on 4 scenes, ~180 test views for each scene

Instance segmentation (1 view, dense, ScanNet)



Instance segmentation (1 view, dense, ScanNet)





mIoU

Acc

Average results on 4 scenes, ~180 test views for each scene

GitHub Repo

https://github.com/xxm19/jacobinerf