

# MAC-Ego3D: Multi-Agent Gaussian Consensus for Real-Time Collaborative Ego-Motion and Photorealistic 3D Reconstruction

## Supplementary Material

### A. More Qualitative Results

We present a detailed qualitative comparison between our approach, MAC-Ego3D, and the previous SOTA multi-agent dense SLAM method, CP-SLAM [4]. To showcase the generalization capabilities of MAC-Ego3D, we provide novel-view RGB and depth rendering results derived from its reconstructed 3D representations. Additionally, we visualize the Gaussian splat representation optimized by our multi-agent Gaussian consensus, showcasing the model’s ability to achieve photorealistic 3D reconstruction in real time.

To comprehensively demonstrate the spatial coherence of the rendered RGB-D outputs from the reconstructed 3D map of MAC-Ego3D and to compare our method with the previous state-of-the-art, CP-SLAM, we provide an immersive **video demo in the supplementary material**. This visualization illustrates that **our MAC-Ego3D model generates high-fidelity, temporally consistent renderings of RGB-D frames** from the reconstructed 3D map, while CP-SLAM’s predictions often suffer from missing regions, flickering effects, and spatial and temporal inconsistencies. Readers are encouraged to refer to the video for more insights.

#### A.1. More RGB-D Rendering Results

Figures A, B, C, and D provide a detailed comparison of RGB and depth rendering results between MAC-Ego3D and CP-SLAM [4]. Across both synthetic (*Multi-agent Replica*) and real-world (*7Scenes* [1]) datasets, MAC-Ego3D consistently delivers superior fidelity, continuity, and robustness in a variety of reconstruction scenarios.

**RGB rendering fidelity.** In RGB rendering, MAC-Ego3D generates continuous, photorealistic reconstructions with sharper texture details, as shown in Figures A, B, and C. In contrast, CP-SLAM often struggles in high-frequency textured areas, producing fragmented and visually inconsistent outputs, which expose its limitations in capturing complex geometries during 3D reconstruction. MAC-Ego3D’s unified Gaussian splat representation, optimized through multi-agent Gaussian consensus, ensures smoother transitions and enhanced fidelity, particularly in regions with high-frequency details. Moreover, it effectively handles RGB-D video inputs captured in challenging environments featuring rich semantic objects and dynamic motion. Even in scenarios where CP-SLAM completely fails to reconstruct details, MAC-Ego3D demonstrates the ability to capture fine details with remarkable pose and RGB-D precision.

**Depth rendering fidelity.** Depth rendering comparisons, il-

lustrated in Figure D, underscore MAC-Ego3D’s superiority in generating geometrically accurate and consistent depth maps. In contrast to CP-SLAM, which often exhibits abrupt discontinuities and fragmented geometry—especially in regions with sparse observations or intricate structures—MAC-Ego3D preserves depth continuity and structural integrity across diverse scenes. While minor blurring may occur in highly textured areas, the Gaussian splat representation optimized via multi-agent Gaussian consensus mitigates depth overfitting, a common issue with CP-SLAM, and enhances robustness across a variety of challenging scenarios.

#### A.2. Novel-View RGB Rendering Results

The novel-view RGB rendering capabilities of MAC-Ego3D are demonstrated in Figure E and the left columns of Figures F and G. MAC-Ego3D reliably synthesizes RGB images from unobserved viewpoints, reconstructing fine details with remarkable spatial coherence, particularly on the *Multi-agent Replica* dataset, which benefits from clean depth observations. Its ability to seamlessly interpolate between observed views highlights the strength of the Gaussian splat-based representation in capturing and generalizing scene details beyond the training data. This capability establishes a new benchmark for generalization and photorealistic reconstruction in multi-agent SLAM systems.

#### A.3. Novel-View Depth Rendering Results

The right columns of Figures F and G highlight MAC-Ego3D’s depth synthesis performance for novel viewpoints. The results demonstrate the model’s capability to accurately interpolate depth while maintaining geometric consistency across a range of challenging scenarios, including complex structures and sparsely observed regions. Notably, MAC-Ego3D achieves smooth and coherent transitions between viewpoints, underscoring its robustness in handling diverse and incomplete depth observations.

#### A.4. 3D Gaussian Splat Representation

Figures H, I, and J illustrate the reconstructed 3D Gaussian splat maps, which encode both RGB color and geometry across various scenes. These visualizations highlight the spatial coherence and expressive power of the representation. By effectively balancing computational efficiency and visual accuracy, MAC-Ego3D enables real-time collaborative SLAM with high photorealistic fidelity.

Figure K visualizes the 3D Gaussian splat representation before and after the pruning process. The raw Gaussian

Table A. Comparison on single-agent Replica dataset.

Methods	PSNR [dB] $\uparrow$	ATE [cm] $\downarrow$
Single-Agent 3DGS SLAM		
GS-SLAM [10] [CVPR'24]	34.27	0.50
SplaTAM [5] [CVPR'24]	33.89	0.36
MonoGS [6] [CVPR'24]	38.94	0.32
RTG-SLAM [8] [SIGGRAPH'24]	35.43	0.18
GS-ICP-SLAM [2] [ECCV'24]	38.83	0.16
Multi-Agent 3DGS SLAM (for Single-Agent Evaluation)		
<b>MAC-Ego3D<sup>†</sup> (Ours)</b>	<b>39.02</b>	<b>0.15</b>

<sup>†</sup> denotes *incomplete* methods *without multi-agent optimization*.

splats (left panels in Figure K) often include redundant and elongated primitives that adversely affect rendering quality. Our pruning method, outlined in Algorithm 1, effectively removes these artifacts, resulting in a compact and visually accurate 3D representation (right panels in Figure K). This pruning process enhances rendering fidelity by addressing over-redundancy and aligning splats more precisely.

## B. Additional Experiments on Single-Agent Replica Dataset

While the true novelty and focus of our proposed method MAC-Ego3D is on the multi-agent collaborative setting, we have conducted additional evaluations using the single-agent Replica dataset [9] (Table A) to provide more results under the single-agent setting.

By disabling the collaborative multi-agent optimization within MAC-Ego3D, we find that **the system still achieves on-par performance or superior performance than prior SOTA single-agent SLAM methods in both reconstruction and tracking**. This highlights the effectiveness of our 3DGS representation, particularly through the proposed Intra-Agent Gaussian Consensus. This approach remains highly capable even without multi-agent collaboration.

## C. Hyperparameter Sensitivity Analysis

### C.1. Similarity Threshold $\tau$

Figures L and M illustrate the impact of the similarity threshold  $\tau$  on MAC-Ego3D's performance for inter-agent overlap detection on the *Multi-agent Replica* and *7Scenes* datasets, respectively. The results indicate that MAC-Ego3D is robust to variations in  $\tau$  within a local range of the default value.

**Trajectory estimation quality, measured by ATE.** On both datasets, larger  $\tau$  values result in more aggressive overlap detection, slightly increasing trajectory errors due to reduced multi-agent collaborative map and pose optimization. Overly restrictive thresholds may miss valid overlaps, causing minor degradations in accuracy. Conversely, smaller  $\tau$  values can increase false positives, occasionally leading to minor improvements in ATE. The default  $\tau$  strikes a balance between

these competing effects, ensuring stable performance.

**Mapping fidelity, measured by PSNR, SSIM, and LPIPS.** Metrics for mapping fidelity remain largely unaffected by variations in  $\tau$ , reflecting MAC-Ego3D's ability to produce visually coherent outputs despite differences in inter-agent overlap quality. These findings highlight the model's robustness, as small variations in  $\tau$  do not significantly impact rendering quality or mapping accuracy.

### C.2. Communication Interval $T_{Comm}$

Figures N and O illustrate the sensitivity of MAC-Ego3D's performance to the communication interval  $T_{Comm}$  on the *Multi-agent Replica* and *7Scenes* datasets. The results demonstrate that MAC-Ego3D maintains robust and consistent performance across a local range of the default  $T_{Comm}$ .

**Trajectory estimation quality, measured by ATE.** Shorter communication intervals (*i.e.*, more frequent collaboration between agents) improve trajectory accuracy by enabling more frequent sharing and updates of the 3D map. This synchronization reduces inter-agent drift and ensures consistent alignment of Gaussian splats. In contrast, larger intervals, while more computationally efficient, introduce slight increases in trajectory error due to reduced inter-agent information sharing. The default  $T_{Comm}$  achieves an optimal balance between computational efficiency and ATE performance, making MAC-Ego3D suitable for both real-time and high-accuracy ego-motion estimation and 3D reconstruction.

**Mapping fidelity, measured by PSNR, SSIM, and LPIPS.** Metrics such as PSNR and SSIM remain largely stable across different  $T_{Comm}$  values, with marginal improvements observed at shorter intervals. Similarly, LPIPS remains consistently low, indicating high visual quality. These findings suggest that inter-agent communication frequency has minimal impact on rendering fidelity, further emphasizing the robustness of MAC-Ego3D's multi-agent collaborative ego-motion estimation and 3D mapping framework.

### C.3. General Hyperparameter Robustness

The sensitivity analyses confirm the robustness of MAC-Ego3D's key hyperparameters, such as the similarity threshold  $\tau$  and communication interval  $T_{Comm}$ , in multi-agent consensus optimization. Across a range of values, the model demonstrates stable performance without requiring precise tuning, underscoring its adaptability to varied cases.

Notably, the default hyperparameter settings strike a balance between computational efficiency and performance across both trajectory estimation and mapping fidelity metrics. The stability of  $\tau$  ensures reliable overlap detection for multi-agent collaborative mapping, even in challenging scenarios with sparse or noisy observations. Similarly,  $T_{Comm}$  balances the trade-off between frequent inter-agent synchronization for improved trajectory accuracy and computational

---

**Algorithm 1 Gaussian Splat Pruning**

---

**Input:** Opacity threshold  $\tau_o$ , scale threshold  $\tau_s$ , elongation threshold  $\tau_e$ , original Gaussian set  $\mathcal{M}$

**Output:** Pruned Gaussian set  $\mathcal{M}'$

```
1: Initialize:  $\mathcal{M}_{\text{prune}} \leftarrow \emptyset$ 
2: Opacity Pruning:  $\mathcal{M}_{\text{opacity}} \leftarrow \{\mathbf{G}_i \mid \lambda_i < \tau_o\}$ 
3:  $\mathcal{M}_{\text{prune}} \leftarrow \mathcal{M}_{\text{prune}} \cup \mathcal{M}_{\text{opacity}}$ 
4: if  $\tau_s \neq \emptyset$  then
5:   Scale Pruning:  $\mathcal{M}_{\text{large}} \leftarrow \{\mathbf{G}_i \mid \max(\mathbf{s}_i) > \tau_s\}$ 
6:    $\mathcal{M}_{\text{prune}} \leftarrow \mathcal{M}_{\text{prune}} \cup \mathcal{M}_{\text{large}}$ 
7:   Elongation Pruning:  $\mathcal{M}_{\text{elongated}} \leftarrow \{\mathbf{G}_i \mid$ 
    $\max(\mathbf{s}_i) > \tau_e \cdot (\sum(\mathbf{s}_i) - \max(\mathbf{s}_i))\}$ 
8:    $\mathcal{M}_{\text{prune}} \leftarrow \mathcal{M}_{\text{prune}} \cup \mathcal{M}_{\text{elongated}}$ 
9: end if
10: Remove Gaussians:  $\mathcal{M}' \leftarrow \mathcal{M} \setminus \mathcal{M}_{\text{prune}}$ 
11: Return:  $\mathcal{M}'$ 
```

---

efficiency required for real-time applications.

## D. More Implementation Details

### D.1. Gaussian Splat Pruning for Mapping

In the MAC-Ego3D framework, we enhance the 3D Gaussian map representation by introducing a pruning process to remove Gaussians that contribute minimally or negatively to the scene representation. Each Gaussian  $\mathbf{G}_i$  is characterized by its mean position  $\mathbf{x}_i$ , covariance matrix  $\Sigma_i$ , opacity  $\lambda_i$ , and scale dimensions  $\mathbf{s}_i = \{s_{i1}, s_{i2}, s_{i3}\}$ . The pruning process focuses on three key objectives: eliminating Gaussians with low opacity, excessively large Gaussians, and elongated Gaussians that can introduce artifacts in the final rendering.

**Opacity pruning.** The pruning process begins by evaluating the opacity  $\lambda_i$  of each Gaussian. Gaussians with opacity below the threshold  $\tau_o$  are considered negligible and removed:

$$\mathcal{M}_{\text{opacity}} = \{\mathbf{G}_i \mid \lambda_i < \tau_o\}. \quad (1)$$

By default,  $\tau_o$  is set to 0.005.

**Scale pruning.** To address excessively large Gaussians, the maximum scale dimension  $\max(\mathbf{s}_i)$  is examined. Gaussians exceeding a fraction of the scale threshold  $\tau_s$  are flagged for removal:

$$\mathcal{M}_{\text{large}} = \{\mathbf{G}_i \mid \max(\mathbf{s}_i) > \tau_s\}. \quad (2)$$

Here,  $\tau_s$  is set to 0.25 by default.

**Elongation pruning.** Rendering artifacts caused by elongated Gaussians are mitigated by introducing an elongation criterion. A Gaussian is classified as elongated if its largest scale dimension exceeds  $\tau_e$  times the sum of the other two dimensions:

$$\mathcal{M}_{\text{elongated}} = \{\mathbf{G}_i \mid \max(\mathbf{s}_i) > \tau_e \cdot (\sum(\mathbf{s}_i) - \max(\mathbf{s}_i))\}. \quad (3)$$

In our implementation,  $\tau_e$  is set to 10 to target needle-like artifacts. As illustrated in Fig. K, this step effectively removes elongated Gaussians, significantly improving the fidelity of the rendered images.

**Combined pruning.** The final pruning mask combines all three criteria:

$$\mathcal{M}_{\text{prune}} = \mathcal{M}_{\text{opacity}} \cup \mathcal{M}_{\text{large}} \cup \mathcal{M}_{\text{elongated}}. \quad (4)$$

The flagged Gaussians are removed from the map:

$$\mathcal{M}' \leftarrow \mathcal{M} \setminus \mathcal{M}_{\text{prune}}. \quad (5)$$

This pruning process ensures a compact and efficient 3D map representation while maintaining high visual fidelity in the rendered scenes. The full procedure is detailed in Algorithm 1.

### D.2. Keyframe Sampling for Pose Tracking

Building on prior work in Gaussian splatting SLAM [3, 7], we employ an adaptive keyframe sampling strategy to enhance pose tracking efficiency. This method leverages geometric correspondences derived from G-ICP to selectively sample keyframes based on a correspondence threshold, ensuring consistent tracking performance while maintaining a uniform Gaussian density within the map.

Unlike the selective approach used for tracking, every one of ten frames are utilized during mapping to fully exploit the RGB-D observations collected across all agents. This dual strategy enhances robustness by incorporating only non-overlapping Gaussians during pose tracking, thereby reducing accumulated tracking errors. It also achieves an optimal balance between tracking efficiency and mapping completeness, maximizing overall reconstruction accuracy. The geometric correspondence threshold for triggering pose tracking follows the setup described in [3].

### D.3. Construction of Testing Cases on the Real-World 7Scenes Dataset

To evaluate the performance of our method under real-world conditions, we construct testing cases using the 7Scenes dataset, which features diverse indoor environments with complex visual and geometric structures. The testing setup is carefully designed to benchmark our approach through comparisons with prior methods and ablation studies, assessing its adaptability and robustness.

**Benchmarking performance against CP-SLAM.** Since CP-SLAM supports only two-agent collaboration, we configure testing cases by selecting two sequences from each scene in the 7Scenes dataset. For most scenes, sequences seq-01 and seq-02 are used, as they provide sufficient spatial overlap to enable collaboration. However, for the *redkitchen* scene, we use sequences seq-01 and seq-03, as these sequences cover complementary areas, ensuring

Table B. Multi-agent testing case configurations for the ablation studies on the *7Scenes* dataset.

Scene	Testing Case	Sequences
Chess	1.1	01, 02, 03
	1.2	04, 05, 06
	1.3	01, 03, 05
Fire	2.1	01, 02, 03, 04
Heads	3.1	01, 02
Office	4.1	01, 02, 03
	4.2	04, 05, 06
	4.3	07, 08, 09, 10
	4.4	01, 02, 04
Redkitchen	5.1	01, 02, 03, 04
	5.2	05, 06, 07, 08
	5.3	11, 12, 13, 14
	5.4	01, 03, 04
Stairs	6.1	01, 02, 03
	6.2	04, 05, 06
	6.3	01, 02, 04

diverse and representative results. This setup allows for fair and consistent performance comparisons with CP-SLAM.

**Testing cases for multi-agent collaboration in ablation studies.** To evaluate our method’s flexibility in supporting multi-agent collaboration with varying numbers of agents, we select sequences from the same scene with overlapping spatial regions to simulate realistic collaboration scenarios. Sequences are grouped based on their serial numbers to ensure even distribution and adequate overlap for inter-agent collaboration. Further, we select the best-performing sequence groups to conduct additional experiments, providing deeper insights into our method’s adaptability. The configurations for all testing cases are detailed in Table B.

## E. More Details on Evaluation Metrics

Following prior works in dense Neural SLAM [3–5], we evaluate the proposed system using metrics including Absolute Trajectory Error (ATE), Peak Signal-to-Noise Ratio (PSNR), Structural Similarity Index Measure (SSIM), Learned Perceptual Image Patch Similarity (LPIPS), and Depth L1 Error. These metrics assess the fidelity of predictions relative to clean ground truth under two input conditions: (i) clean RGB-D observations  $\mathbf{Z}_{1:t}^{a_i} = \{(\mathbf{I}_k^{a_i}, \mathbf{D}_k^{a_i})\}_{k=1}^t$  for agent  $a_i$ , under the standard evaluation settings and (ii) partially perturbed observations in either the RGB  $\mathbf{I}_k^{a_i}$  or depth  $\mathbf{D}_k^{a_i}$  modalities under the robustness evaluation settings.

Specifically, ATE evaluates the spatial alignment between the estimated trajectory  $\mathbf{T}_{1:t}^{a_i}$  and the ground truth trajectory

$\mathbf{T}_{1:t}^{\text{GT}}$  as the root mean square error:

$$\text{ATE} = \sqrt{\frac{1}{N} \sum_{k=1}^N \|\mathbf{T}_k^{a_i} - \mathbf{T}_k^{\text{GT}}\|^2}. \quad (6)$$

PSNR quantifies the pixel-level fidelity of the reconstructed RGB image  $\mathbf{I}_k^{a_i}$  relative to the ground truth  $\mathbf{I}_k^{\text{GT}}$ :

$$\text{PSNR} = 10 \cdot \log_{10} \left( \frac{\text{MAX}^2}{\text{MSE}} \right) [dB], \quad (7)$$

where MAX is the maximum pixel intensity and MSE is the mean squared error.

SSIM evaluates structural similarity, incorporating luminance, contrast, and structural information:

$$\text{SSIM} = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)}, \quad (8)$$

where  $\mu_x, \mu_y, \sigma_x$ , and  $\sigma_y$  denote the means and variances of  $\mathbf{I}_k^{a_i}$  and  $\mathbf{I}_k^{\text{GT}}$ , while  $\sigma_{xy}$  represents their covariance.

LPIPS [11] measures perceptual similarity between the predicted RGB image  $\mathbf{I}_k^{a_i}$  and the ground truth  $\mathbf{I}_k^{\text{GT}}$ , based on feature-level differences from a pre-trained neural network. The LPIPS score is:

$$\text{LPIPS} = \sum_l w_l \cdot \|\phi_l(\mathbf{I}_k^{a_i}) - \phi_l(\mathbf{I}_k^{\text{GT}})\|_2^2, \quad (9)$$

where  $\phi_l(\cdot)$  denotes feature maps from the  $l$ -th layer of a pre-trained network, with  $w_l$  as learned layer weights.

Finally, Depth L1 Error quantifies the geometric accuracy of the predicted depth  $\mathbf{D}_k^{a_i}$  compared to the ground truth depth  $\mathbf{D}_k^{\text{GT}}$ :

$$\text{Depth L1} = \frac{1}{N} \sum_{k=1}^N \|\mathbf{D}_k^{a_i} - \mathbf{D}_k^{\text{GT}}\|. \quad (10)$$

## References

- [1] Ben Glocker, Shahram Izadi, Jamie Shotton, and Antonio Criminisi. Real-time rgb-d camera relocalization. In *International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 2013. 1
- [2] Seongbo Ha, Jiung Yeon, and Hyeonwoo Yu. Rgb-d gs-icp slam. *arXiv preprint arXiv:2403.12550*, 2024. 2
- [3] Seongbo Ha, Jiung Yeon, and Hyeonwoo Yu. Rgb-d gs-icp slam. In *European Conference on Computer Vision*, pages 180–197. Springer, 2025. 3, 4
- [4] Jiarui Hu, Mao Mao, Hujun Bao, Guofeng Zhang, and Zhaopeng Cui. Cp-slam: Collaborative neural point-based slam system. In *Advances in Neural Information Processing Systems*, pages 39429–39442. Curran Associates, Inc., 2023. 1



- [5] Nikhil Keetha, Jay Karhade, Krishna Murthy Jatavallabhula, Gengshan Yang, Sebastian Scherer, Deva Ramanan, and Jonathon Luiten. Splatam: Splat, track map 3d gaussians for dense rgb-d slam. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2024. [2](#), [4](#)
- [6] Hidenobu Matsuki, Riku Murai, Paul HJ Kelly, and Andrew J Davison. Gaussian splatting slam. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18039–18048, 2024. [2](#)
- [7] Raul Mur-Artal, Jose Maria Martinez Montiel, and Juan D Tardos. Orb-slam: a versatile and accurate monocular slam system. *IEEE transactions on robotics*, 31(5):1147–1163, 2015. [3](#)
- [8] Zhexi Peng, Tianjia Shao, Yong Liu, Jingke Zhou, Yin Yang, Jingdong Wang, and Kun Zhou. Rtg-slam: Real-time 3d reconstruction at scale using gaussian splatting. In *ACM SIGGRAPH 2024 Conference Papers*, pages 1–11, 2024. [2](#)
- [9] Julian Straub, Thomas Whelan, Lingni Ma, Yufan Chen, Erik Wijmans, Simon Green, Jakob J Engel, Raul Mur-Artal, Carl Ren, Shobhit Verma, et al. The replica dataset: A digital replica of indoor spaces. *arXiv preprint arXiv:1906.05797*, 2019. [2](#)
- [10] Chi Yan, Delin Qu, Dan Xu, Bin Zhao, Zhigang Wang, Dong Wang, and Xuelong Li. Gs-slam: Dense visual slam with 3d gaussian splatting. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 19595–19604, 2024. [2](#)
- [11] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 586–595, 2018. [4](#)

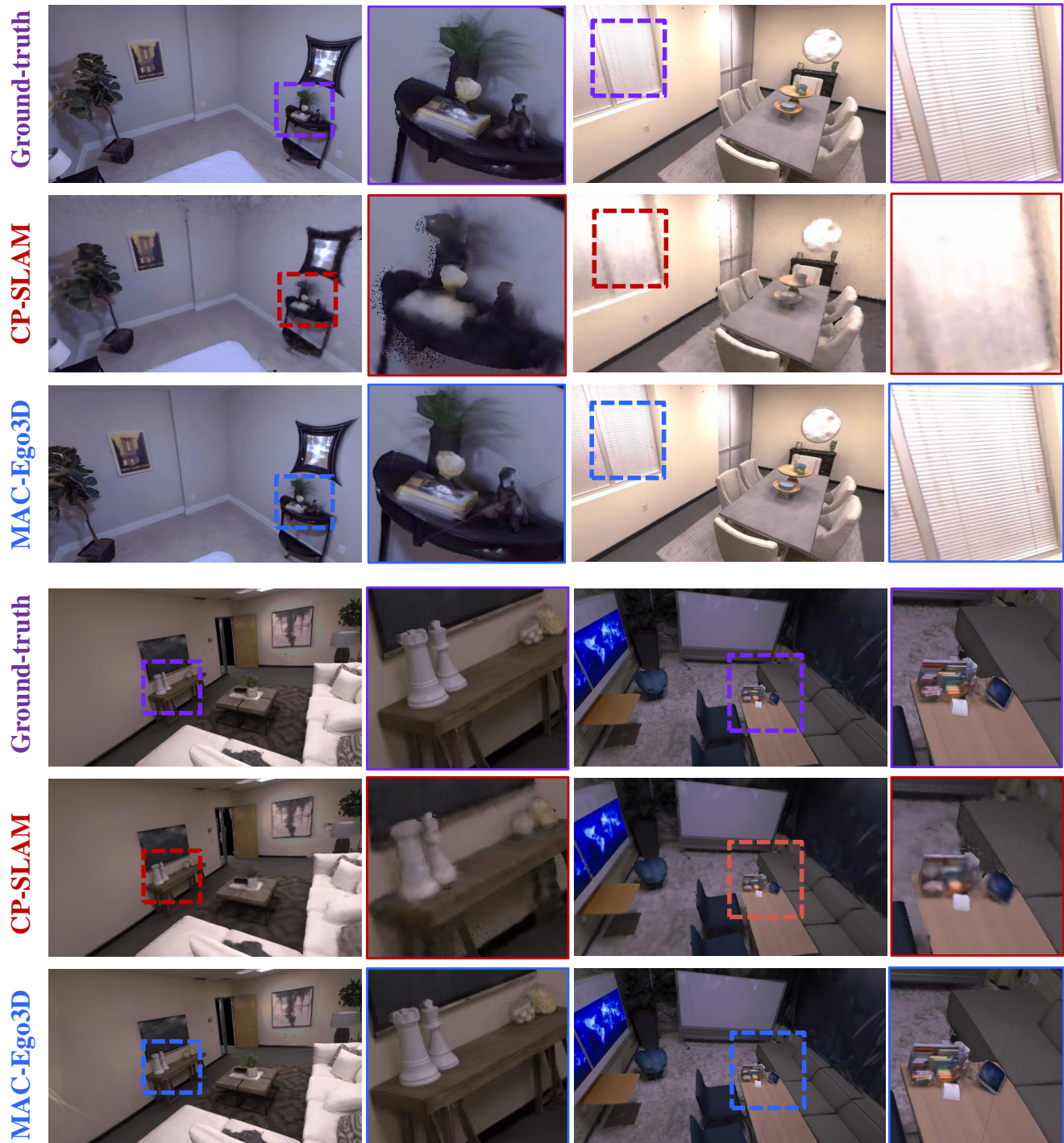


Figure A. **Qualitative RGB image rendering quality comparison between multi-agent SLAM models with dense reconstruction capability, i.e., CP-SLAM and our MAC-Ego3D, on the *Multi-agent Replica* dataset.**



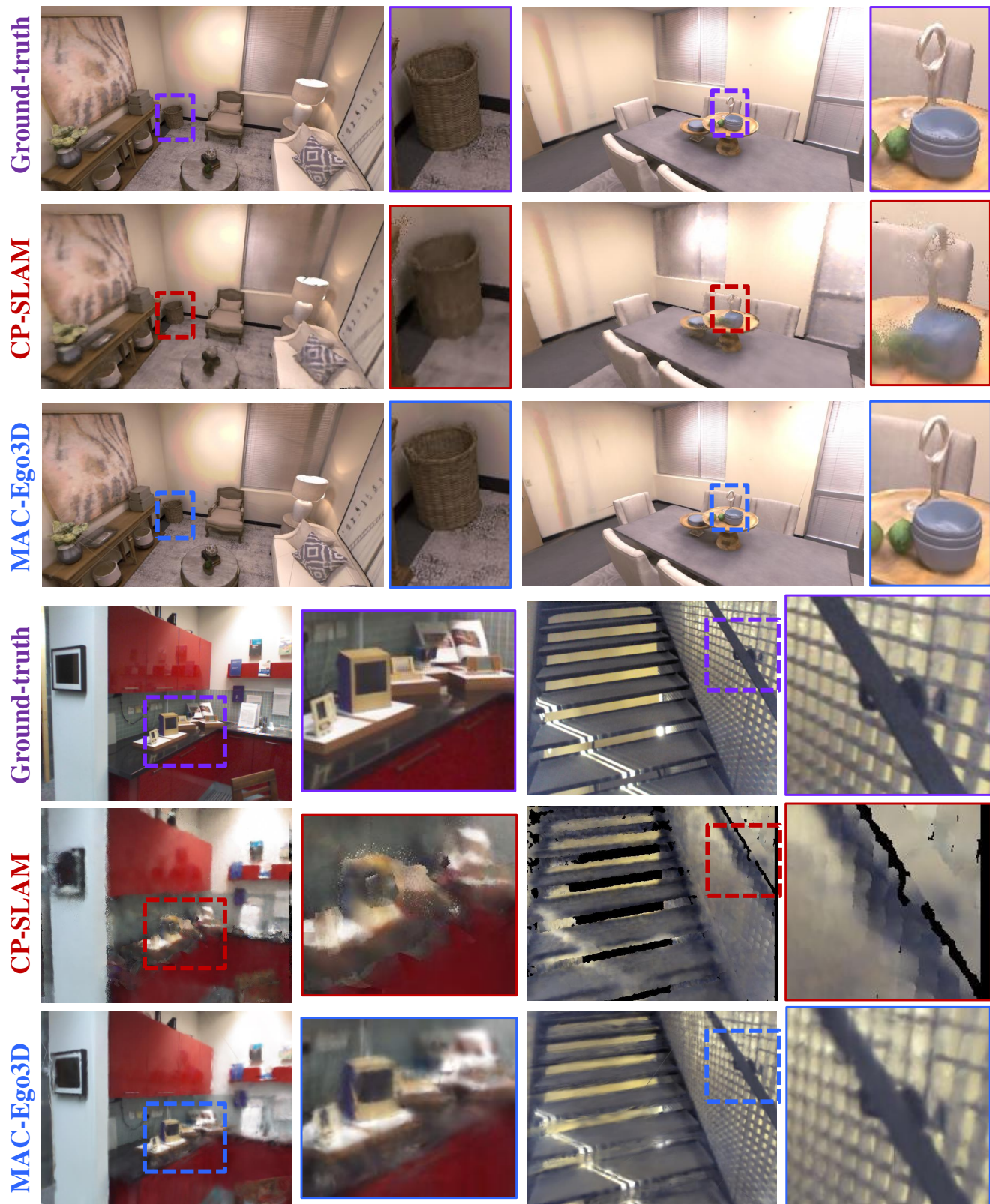


Figure B. **Qualitative RGB image rendering quality comparison between multi-agent SLAM models with dense reconstruction capability, i.e., CP-SLAM and our MAC-Ego3D, on the *Multi-agent Replica* and *7-Scenes* dataset.**



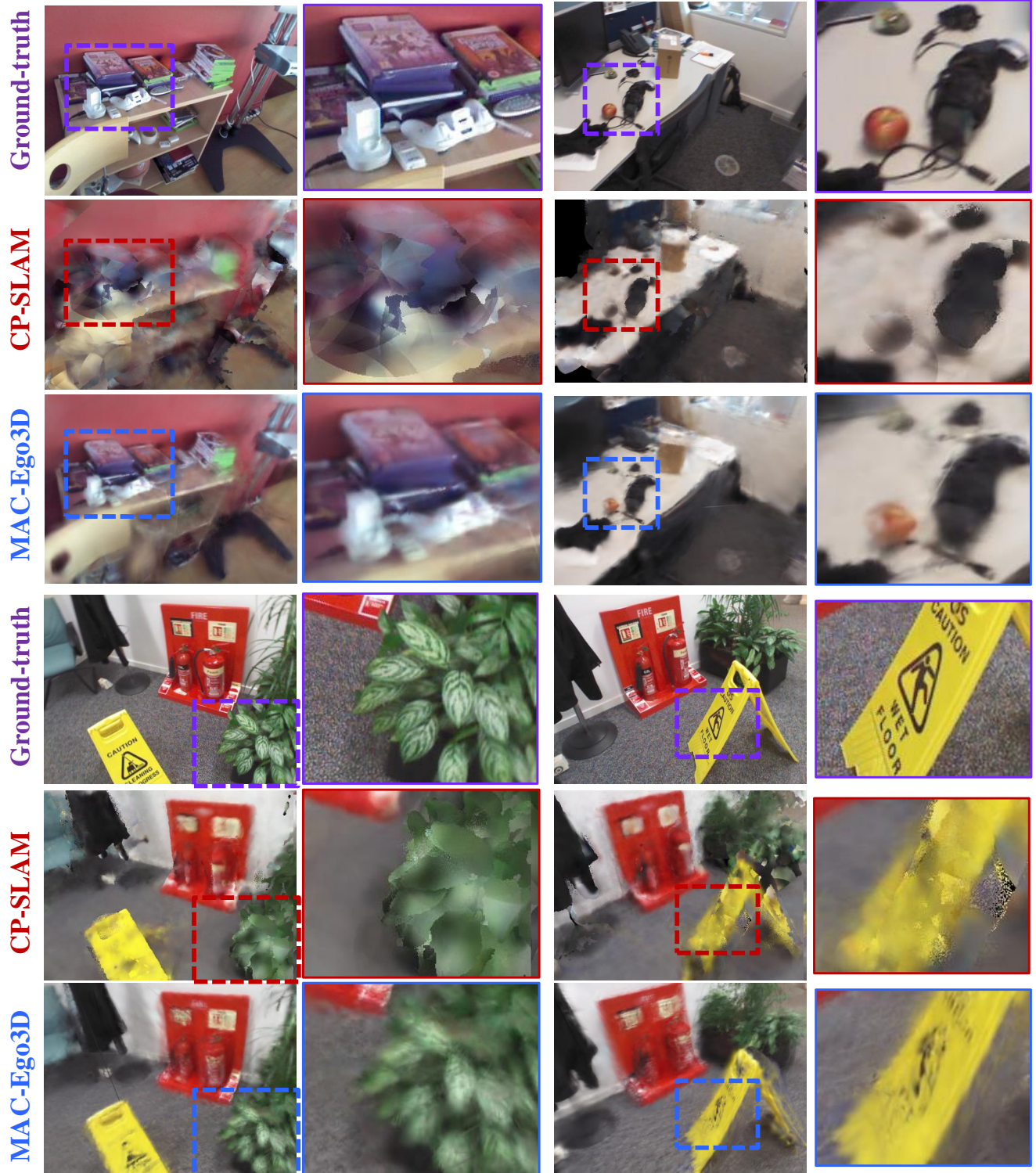


Figure C. Qualitative RGB image rendering quality comparison between multi-agent SLAM models with dense reconstruction capability, *i.e.*, CP-SLAM and our MAC-Ego3D, on the 7-Scenes dataset.



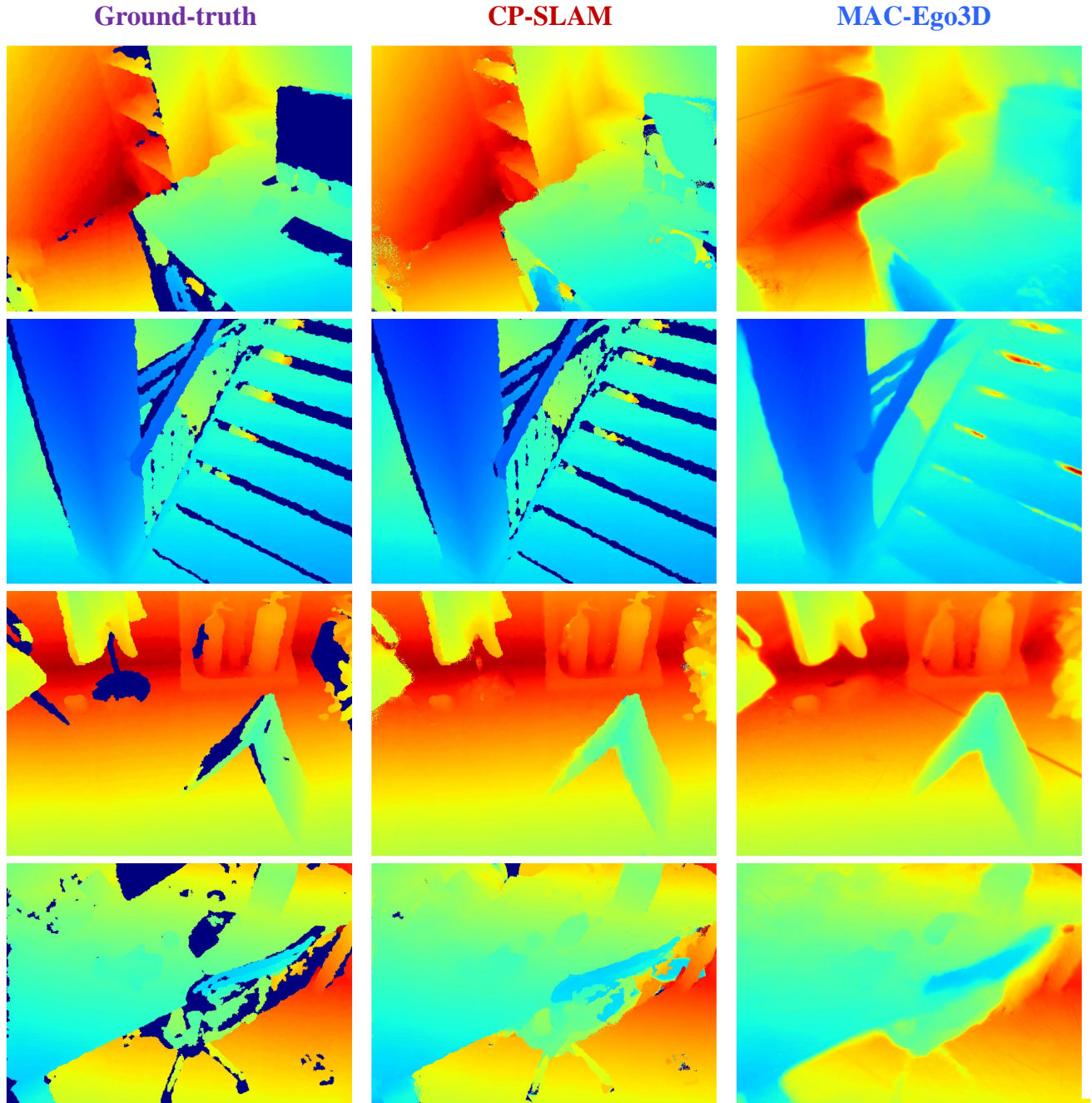


Figure D. Qualitative depth image rendering quality comparison between multi-agent SLAM models with dense reconstruction capability, *i.e.*, CP-SLAM and our MAC-Ego3D, on *Multi-agent Replica* (Left) and *7-Scenes* (Right) datasets.

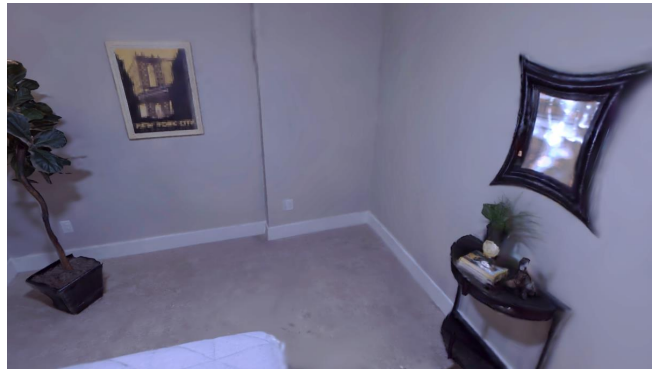


Figure E. Novel-view RGB image synthesis results via the proposed MAC-Ego3D model.



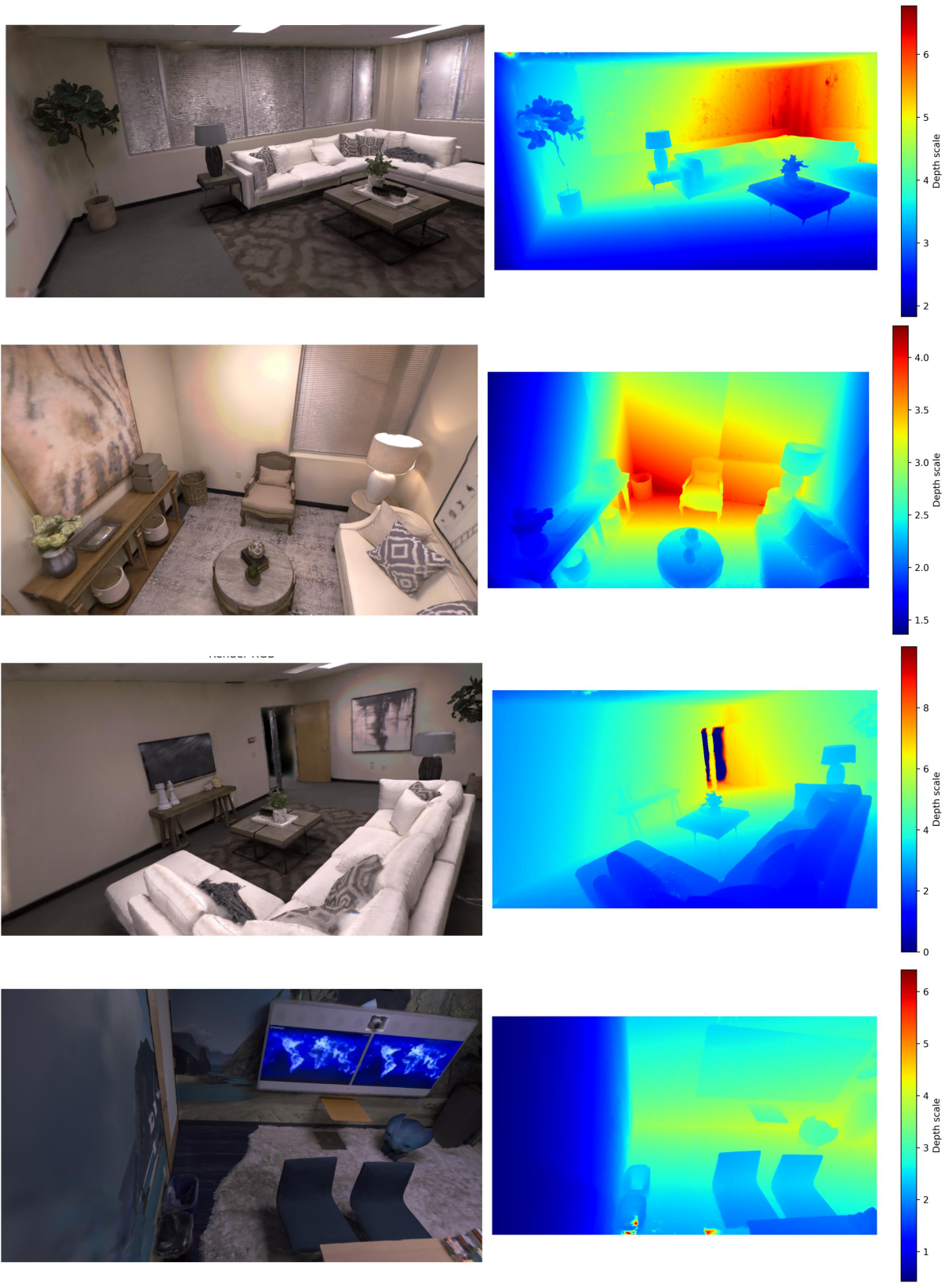


Figure F. Novel-view RGB and depth image synthesis results via the proposed MAC-Ego3D model.

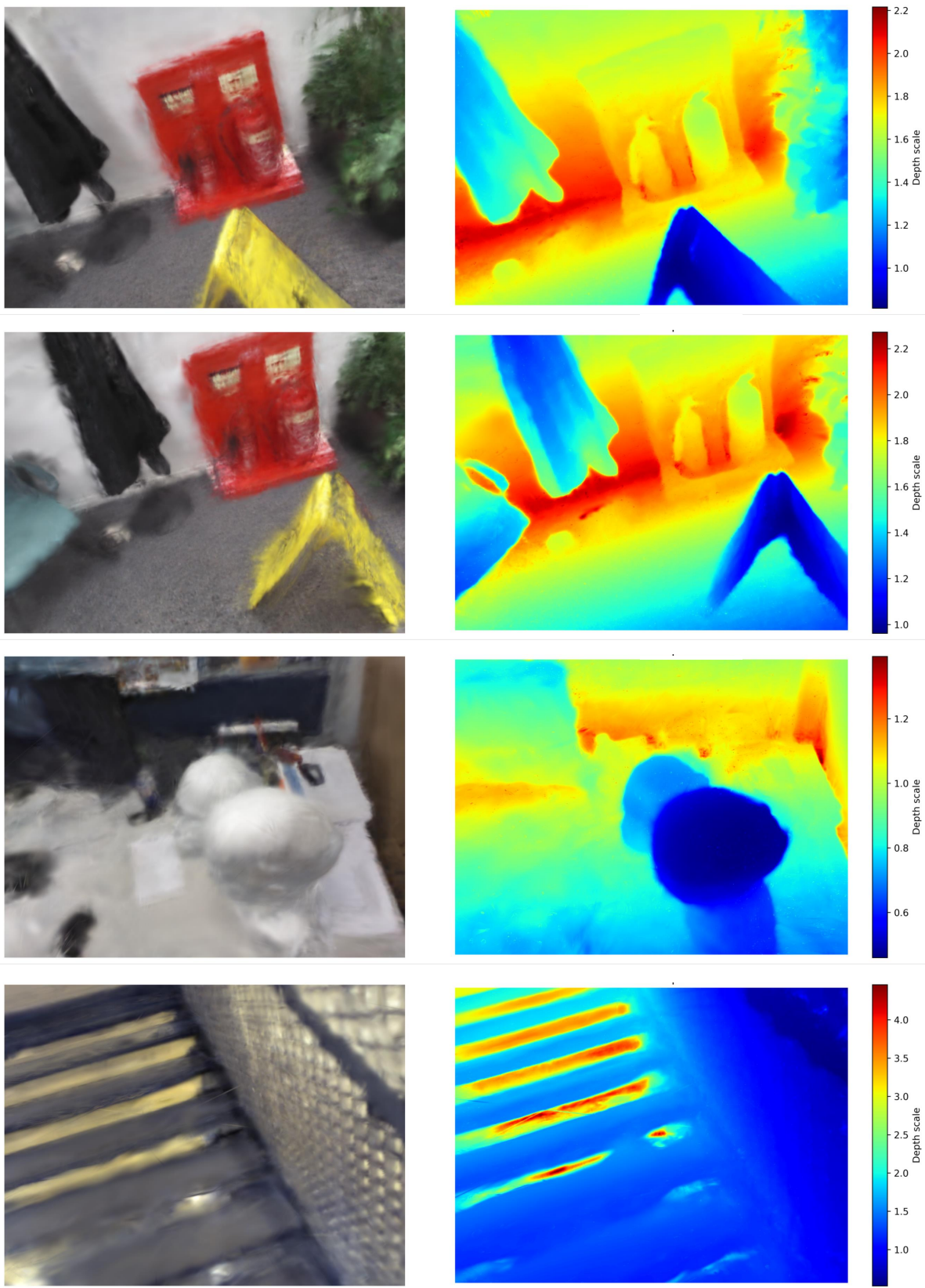


Figure G. Novel-view RGB and depth image synthesis results via the proposed MAC-Ego3D model.



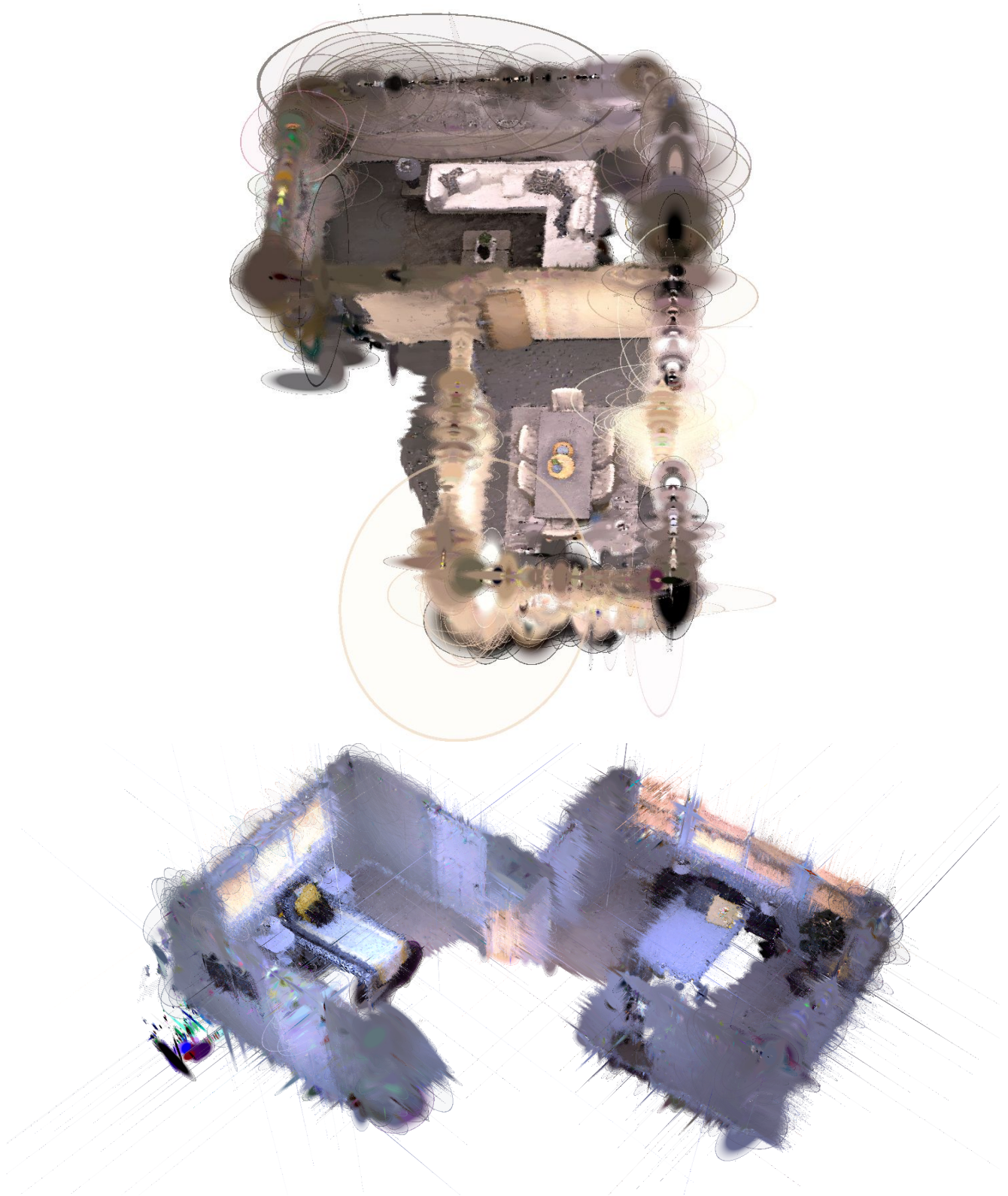


Figure H. Visualization of reconstructed 3D map with Gaussian splat representation.

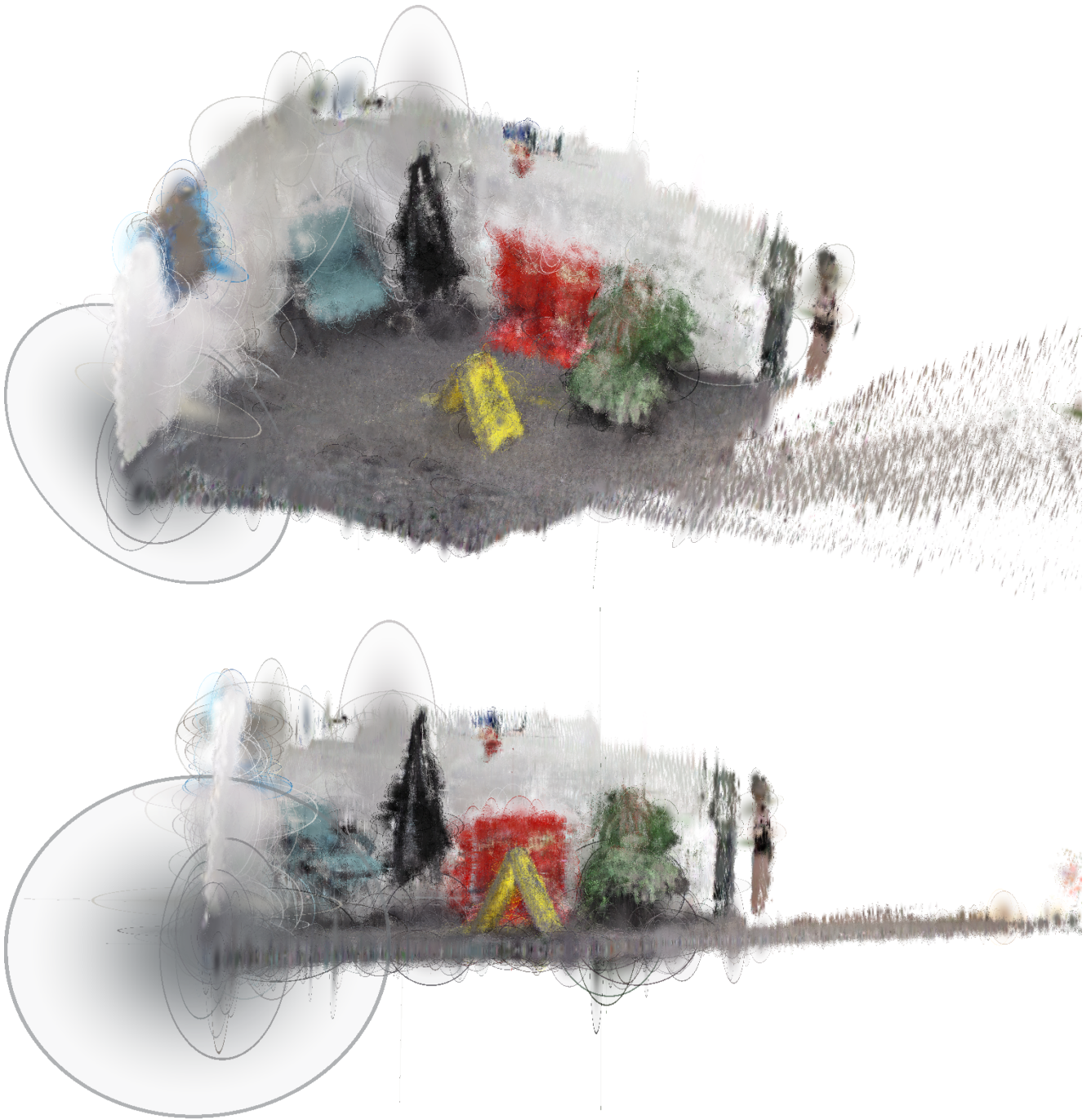


Figure I. Visualization of reconstructed 3D map with Gaussian splat representation.



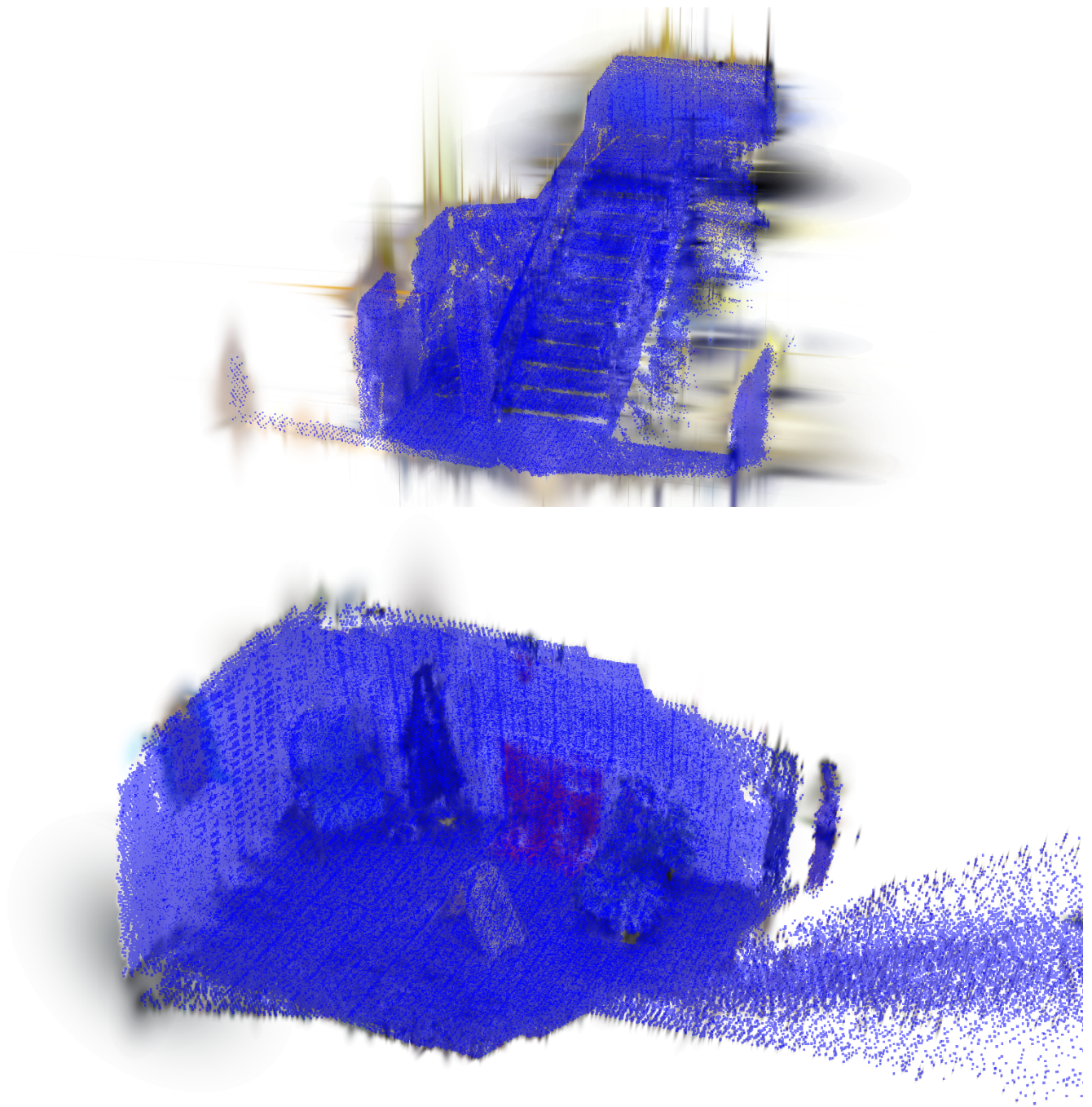


Figure J. Visualization of the coordinate centers (shown in blue points) of reconstructed 3D map with Gaussian splat representation.

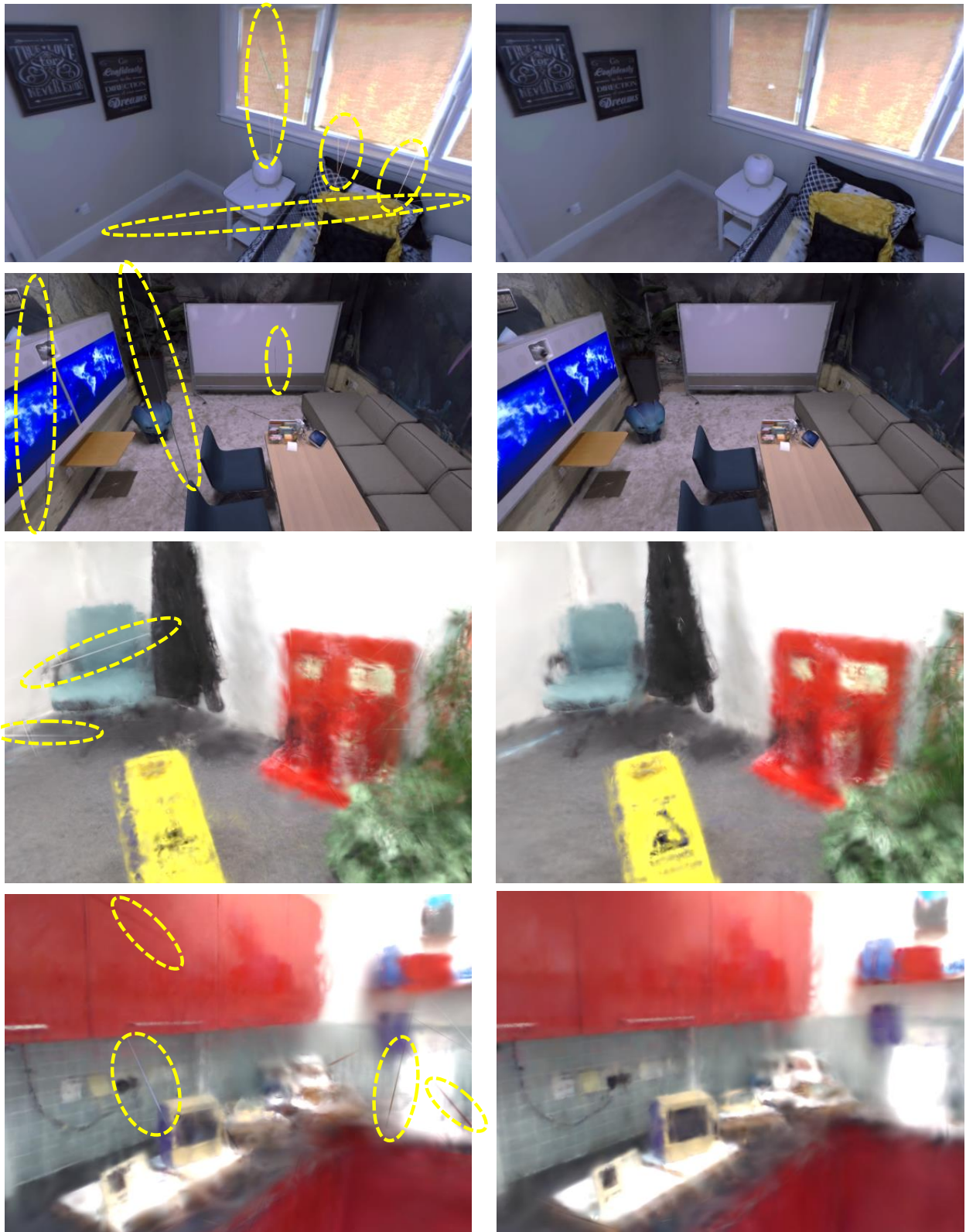


Figure K. **Qualitative comparison of Gaussian splat pruning.** The left column shows the raw Gaussian splats, while the right column presents the results after pruning. The pruning process effectively removes redundant or elongated splats, enhancing the rendering quality.



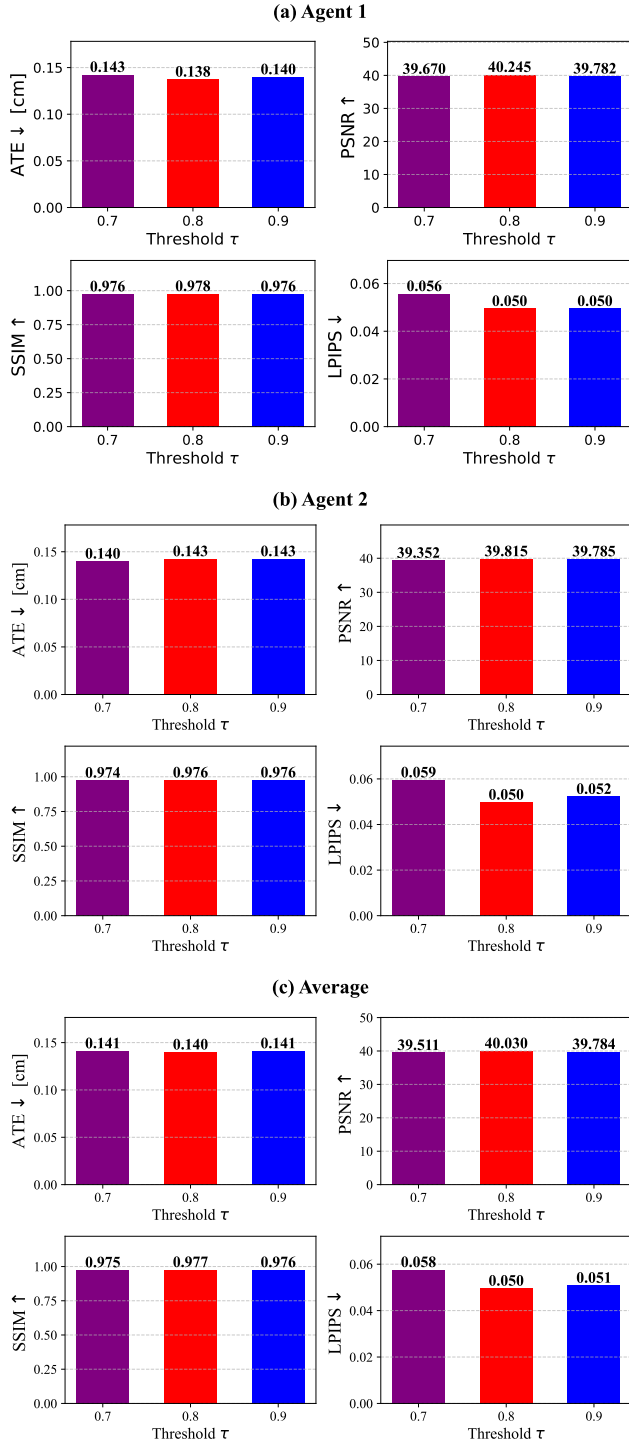


Figure L. Hyperparameter sensitivity analysis of the threshold  $\tau$  for inter-agent overlap detection on *Multi-agent Replica*.

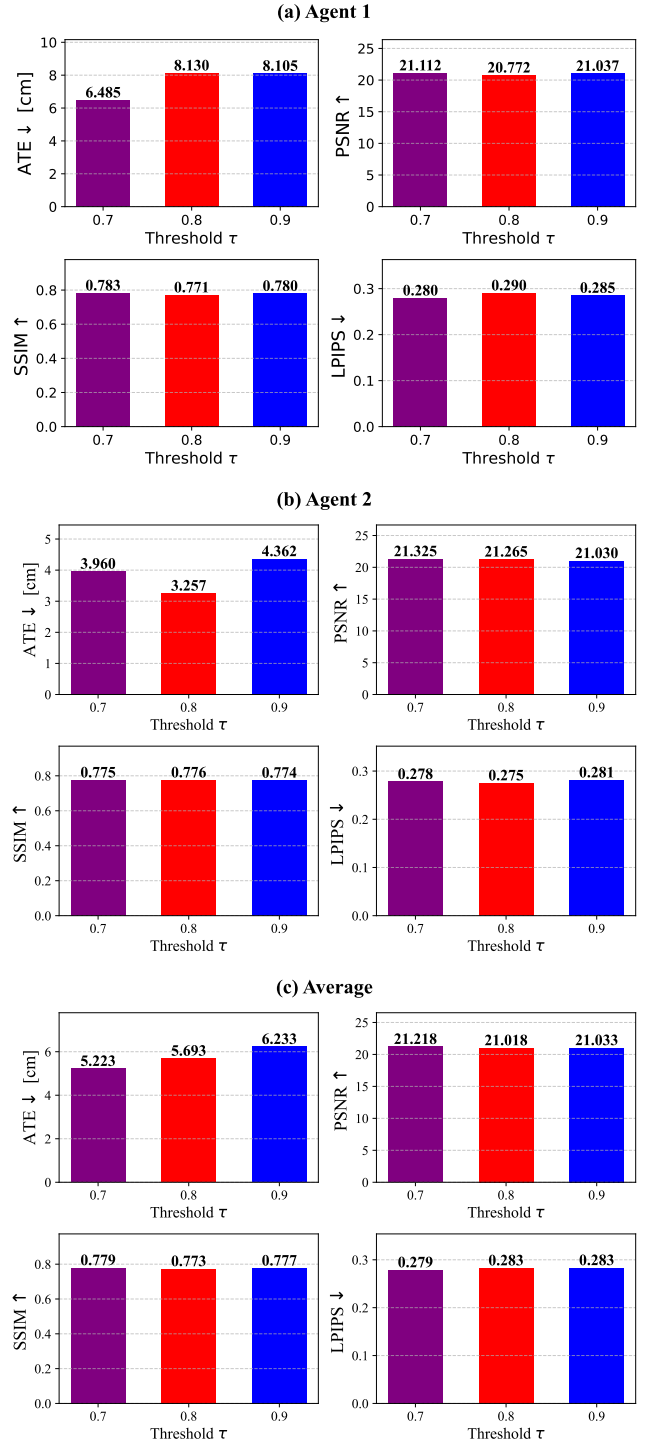


Figure M. Hyperparameter sensitivity analysis of the similarity threshold  $\tau$  for inter-agent overlap detection on *7Scenes*.

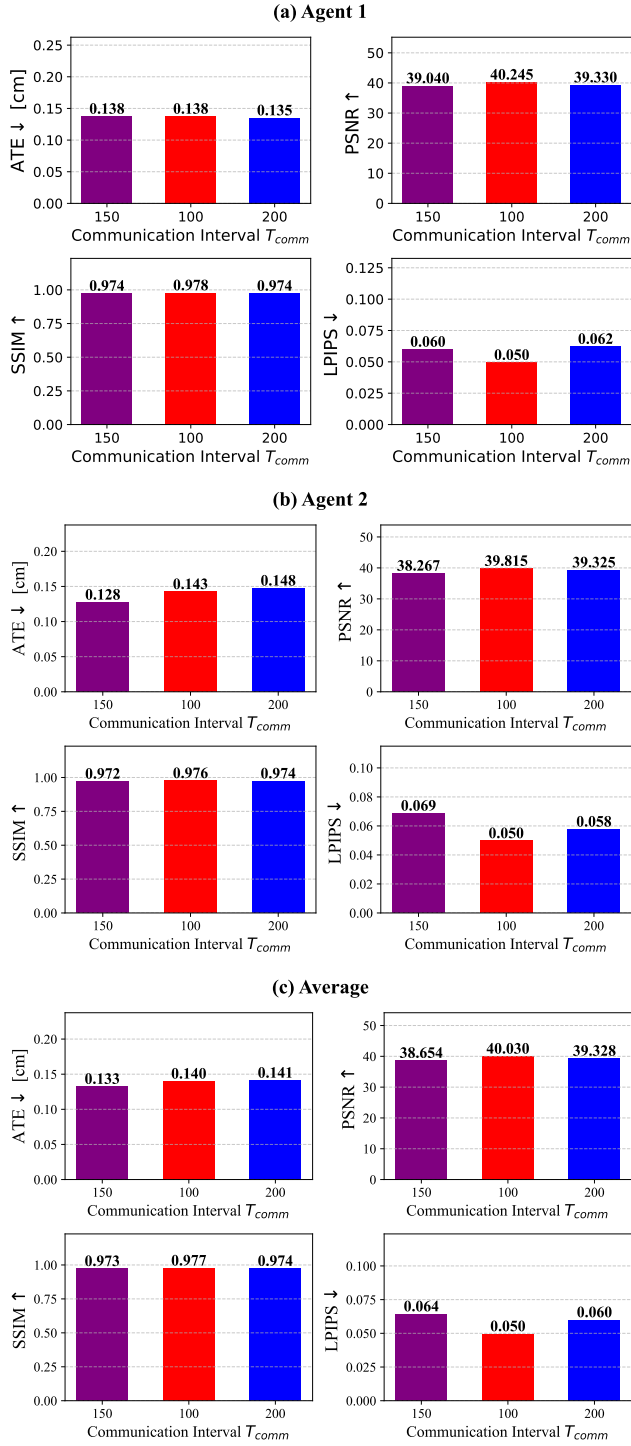


Figure N. Hyperparameter sensitivity analysis of the inter-agent communication interval  $T_{Comm}$  on *Multi-agent Replica*.

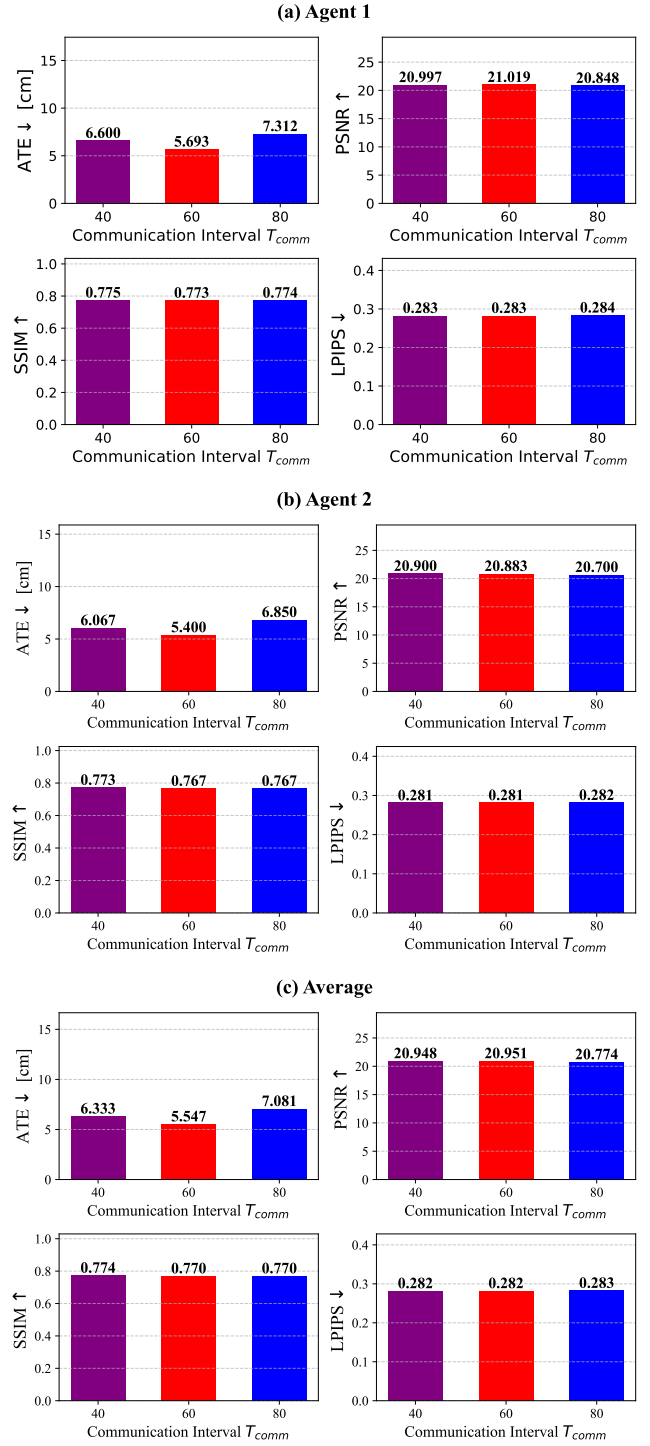


Figure O. Hyperparameter sensitivity analysis of the inter-agent communication interval  $T_{Comm}$  on the *7Scenes* dataset.