Supplementary Materials

Fengyuan Sun   Sezer Karaoglu   Theo Gevers
University of Amsterdam & 3DUniversum
Amsterdam, The Netherlands
fengyuansun2000@gmail.com s.karaoglu@3duniversum.com Th.Gevers@uva.nl

1. Additional qualitative examples

To further evaluate the proposed method, visualizations of multi-view inconsistent predictions and their resulting segmentation are provided in Figures 1 and 2. Our method predicts segmentation that is consistent with respect to each view by leveraging spatial and multi-view information. Figure 3 displays examples of temporal inconsistency, and Figure 4 further shows examples of the predictions in 3D.

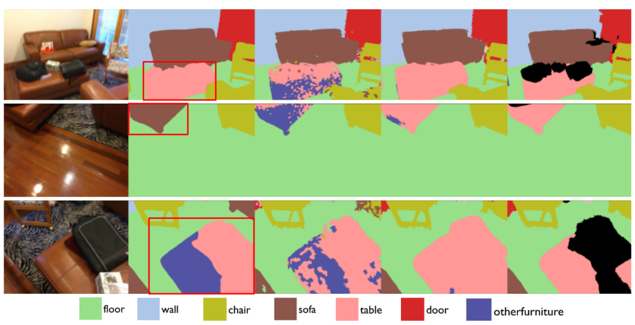

Figure 1. Visualization of cross-view inconsistency. From left to right: rgb image, ViT-Adapter, ViT-weighted averaging, ours, ground-truth. The initial segmentation is inconsistent between views. Consequently, weighted averaging fails to create a unified prediction. In contrast, the proposed method predicts a correct and consistent segmentation.

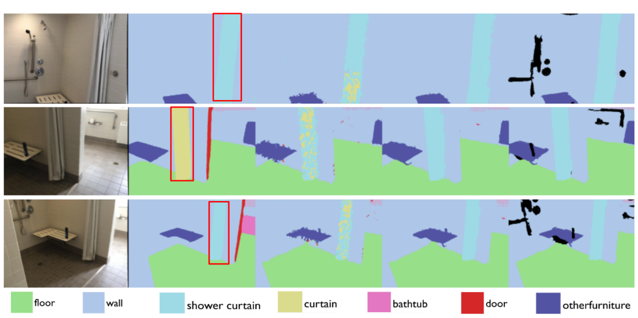

Figure 2. Visualization of an ambiguous class in 2D. From left to right: rgb image, ViT-Adapter, ViT-weighted averaging, ours, ground-truth. Some views in the initial segmentation confuse curtain with shower curtain.

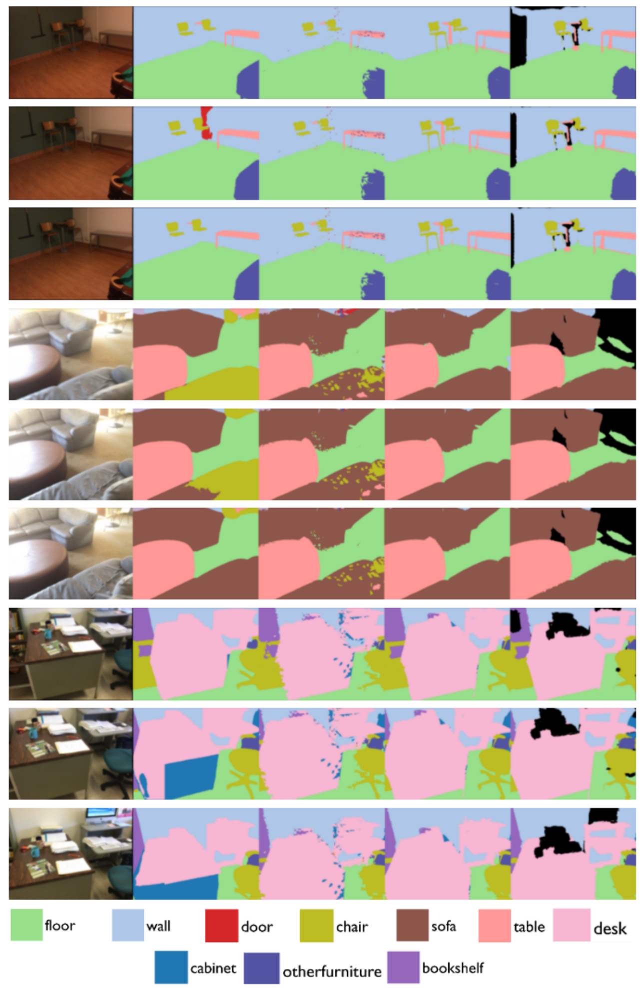

Figure 3. Visualization of temporal inconsistency. From left to right: rgb image, ViT-Adapter segmentation, ViT-weighted averaging segmentation, our segmentation, ground-truth segmentation. Minimal changes in viewpoint can result in significant changes in appearance, global context information and occlusions. This causes flickering in the video segmentation. By leveraging multi-view information, the proposed method is able to create predictions that are consistent in time.

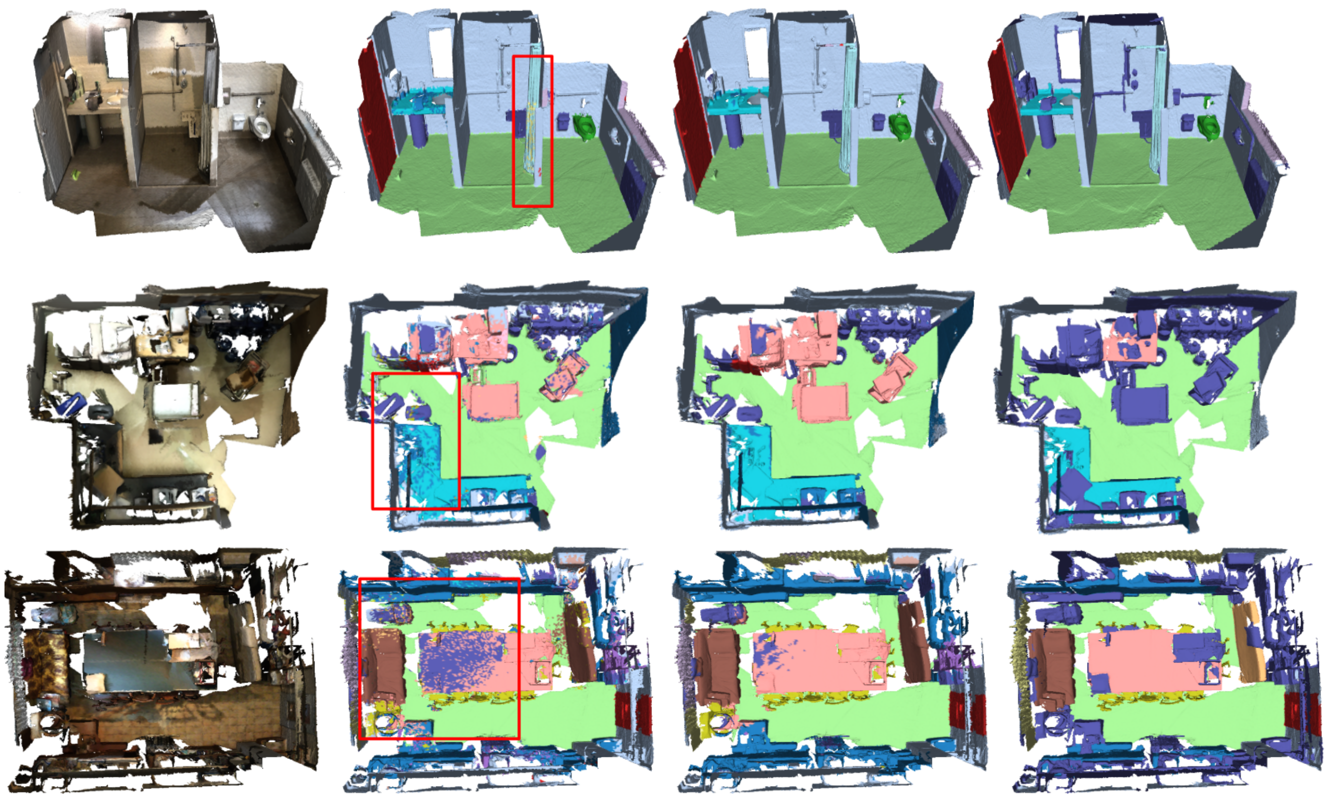

Figure 4. Visualization of the resulting semantic mesh. From left to right: colored mesh, weighted averaging segmentation, our segmentation, ground-truth segmentation. In contrast to weighted averaging, the proposed method is able to increase segmentation coherency and recover ambiguous predictions.