JRDB-PanoTrack: An Open-world Panoptic Segmentation and Tracking Robotic Dataset in Crowded Human Environments

Supplementary Material

6. More Experiments

Results of Various Mask Scales. The OSPA results for small, medium, and large objects in JRDB-Panotrack (defined in Fig. 5) reveal significant performance differences across scales. The results for CW and OW are in Tab. 9 and Tab. 10, respectively. Current methods excel with large-scale objects but struggle with smaller ones, highlighting the challenge of dynamic scene interpretation in JRDB-Panotrack.

7. More Statistics and information

Visualization. Our annotations are visualized in Fig. 7, displaying a rich diversity of settings that span both indoor and outdoor environments. These images illustrate the versatility of JRDB-PanoTrack, featuring a variety of lighting conditions, varying crowd densities, and diverse human activities. Additionally, Fig. 7 highlights our labeling approach, marking different floors and objects behind glass or on walls with striped areas. This detail is crucial for improving human-robot interaction and decision-making in robotics.

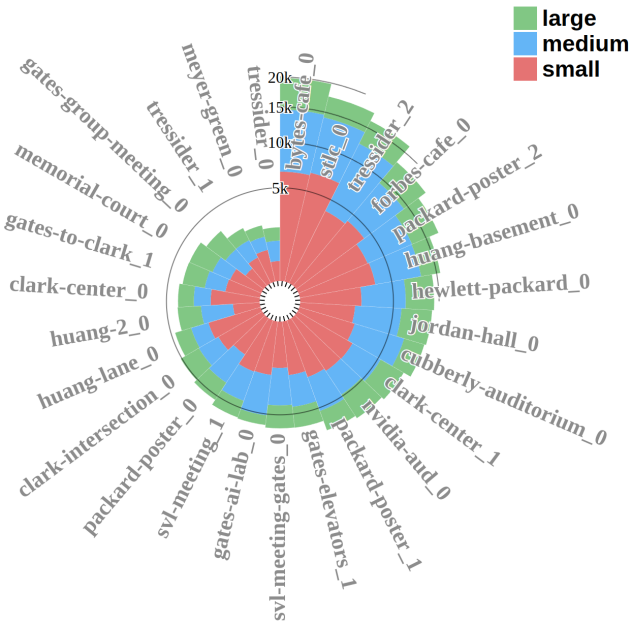
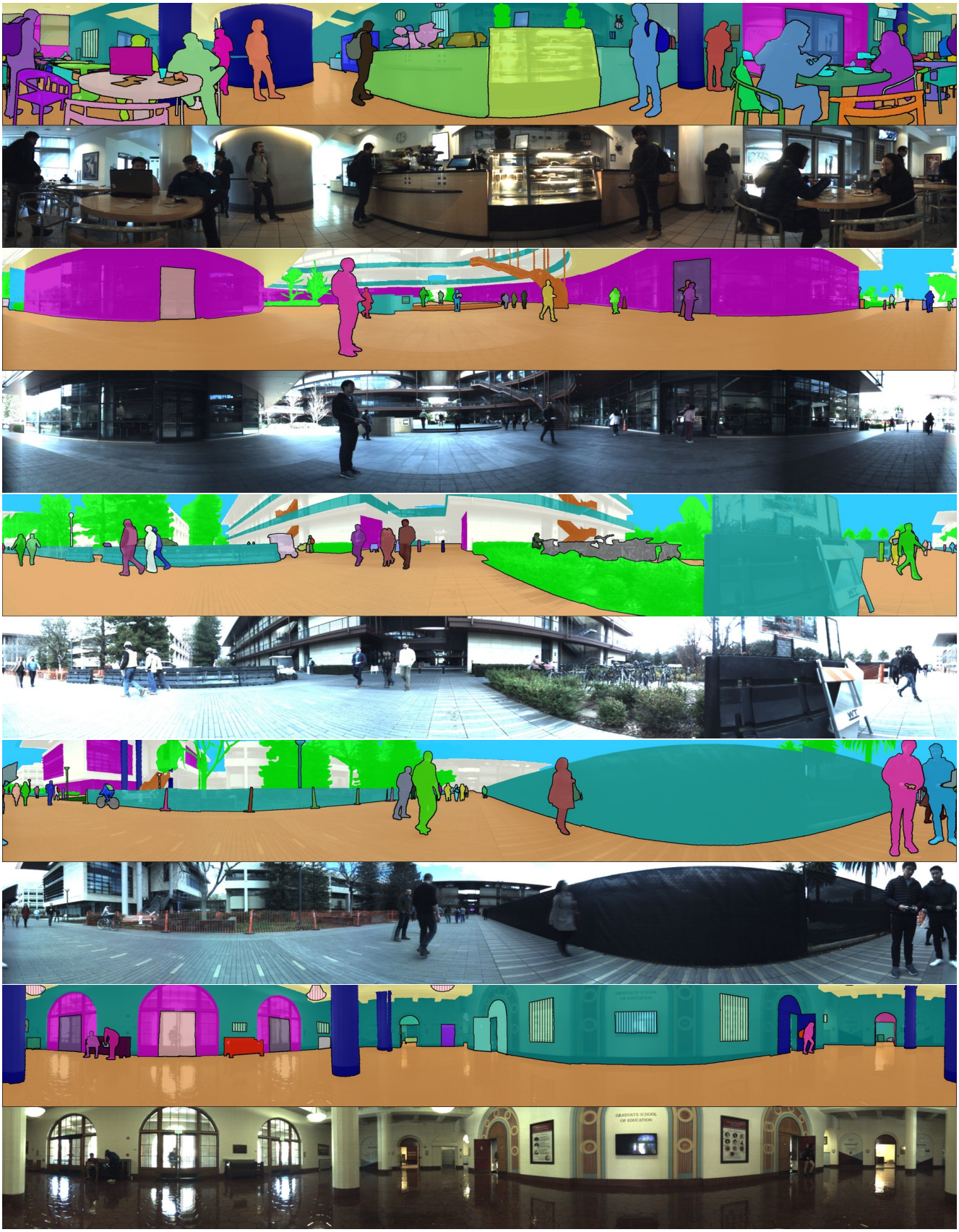
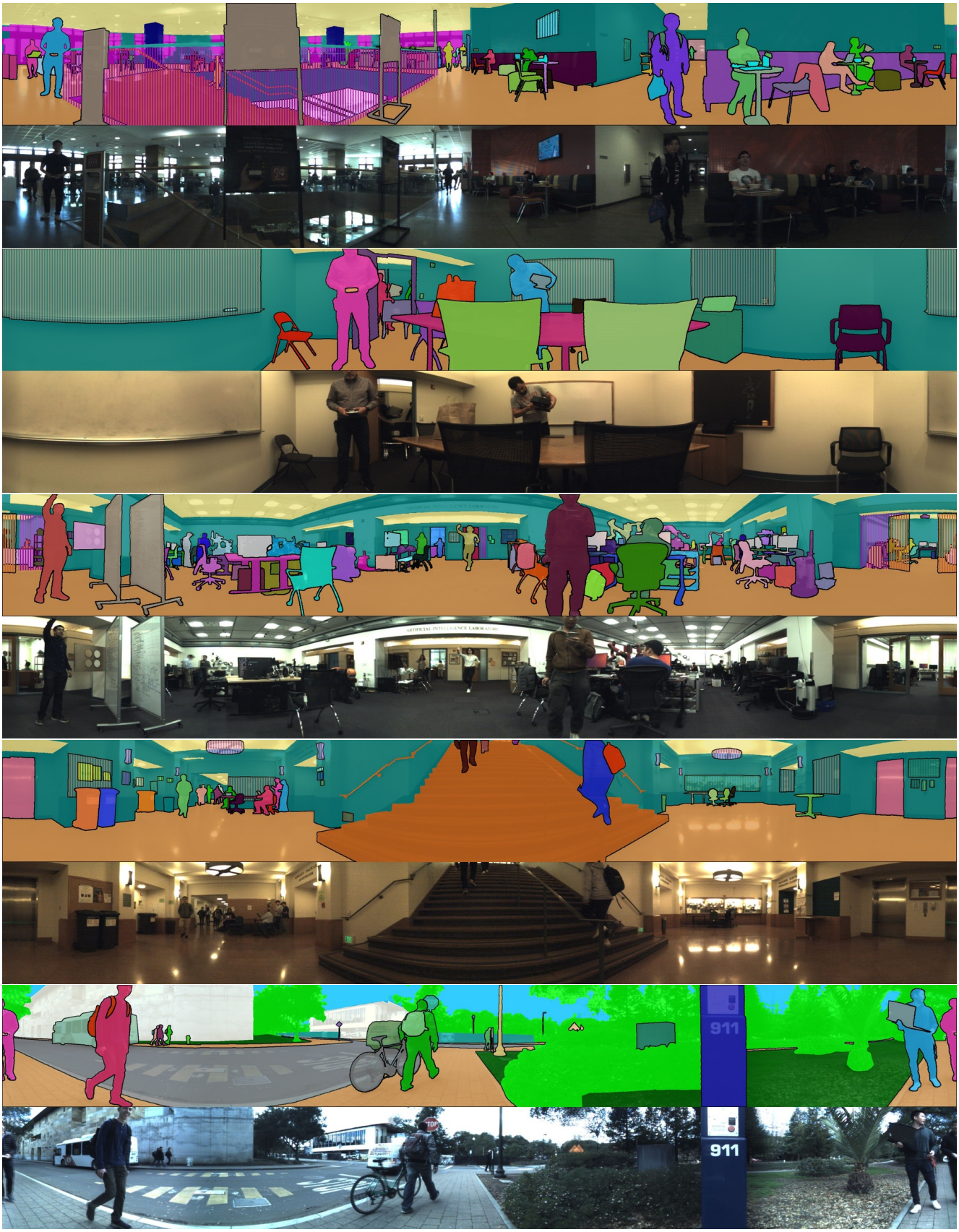


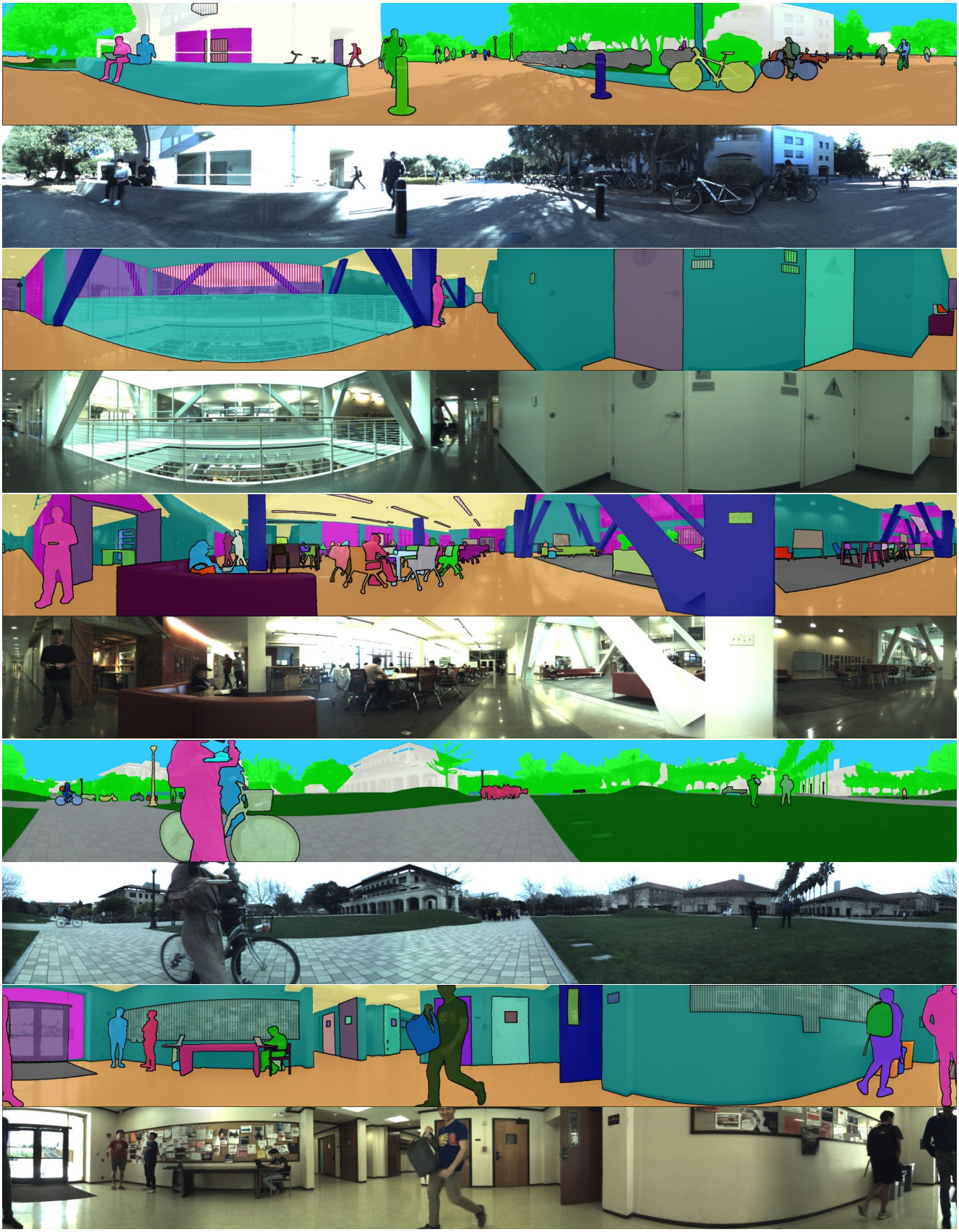
Figure 6. The distribution of mask sizes per sequence in the JRDB-PanoTrack training set.

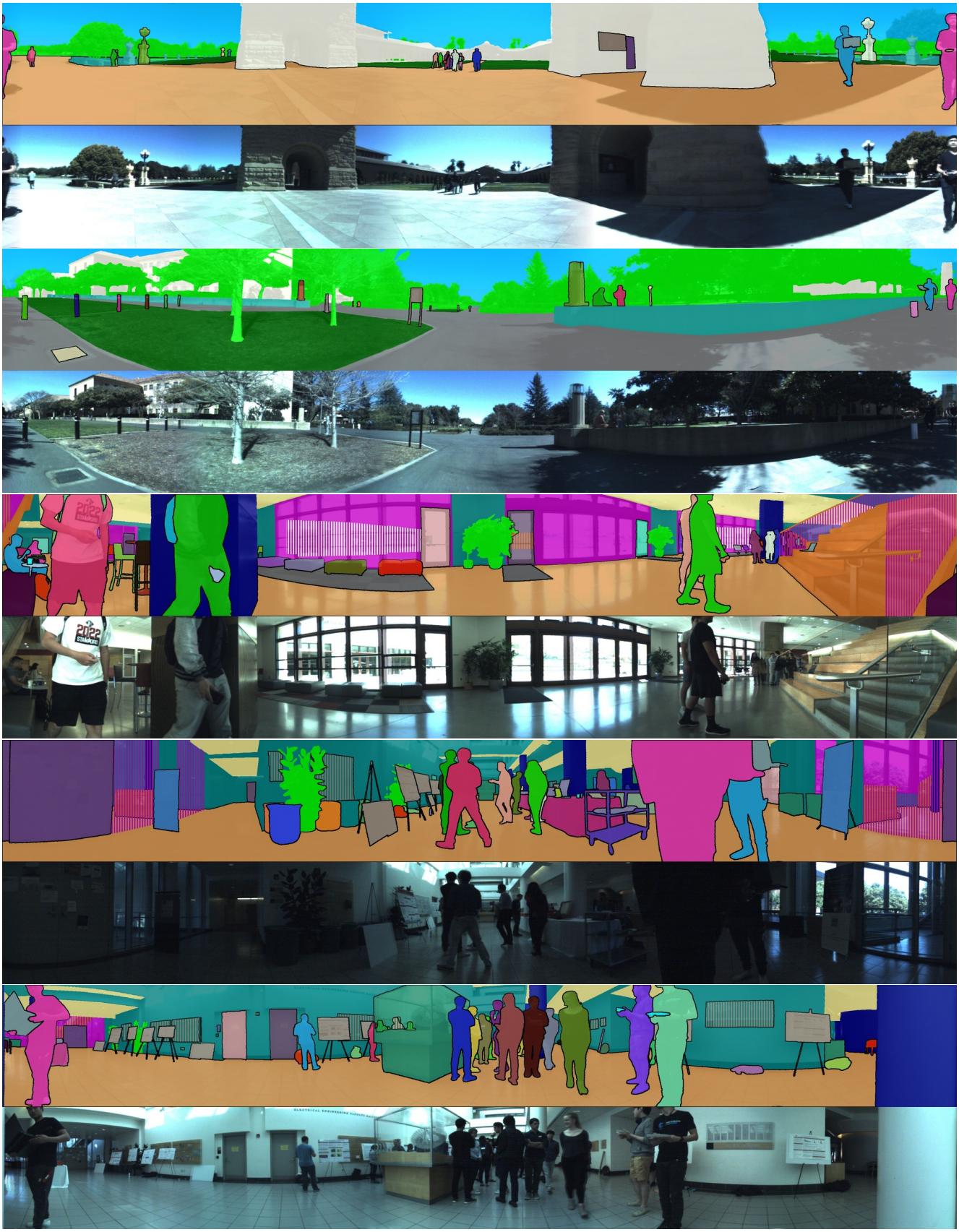
Detailed Mask size. Beyond the general mask size trends in JRDB-PanoTrack (Fig. 5), we also examine the distribution per sequence (Fig. 6). Masks are categorized by area M: small ($< 32^2$), medium ($32^2 < M \leq 96^2$), and large ($> 96^2$). Fig. 6 shows that the mask size distribution varies notably between indoor and outdoor sequences. Indoor sequences feature a larger quantity of masks, predominantly smaller in size, due to the crowded nature and presence of numerous small objects like chairs, tables, and people. In contrast, outdoor sequences have fewer, but larger masks, reflecting the less crowded spaces with larger 'stuff' classes, such as trees and buildings.

Behind glass. Traditional datasets often assume that a pixel only belongs to a single object. However, as shown in Fig. 7, in real-world environments, objects behind glass or being hung on wall also can be seen and are important for human-robot interaction, robot navigation and decision. Therefore, JRDB-PanoTrack provides annotations for such objects. Fig. 3 presents a word cloud of classes predominantly viewed through glass in JRDB-PanoTrack, emphasizing the prevalence of objects like *board, stair, bag, machines,* and *sign*.


GRADUATE SCHOOL
OF EDUCATION


911
911





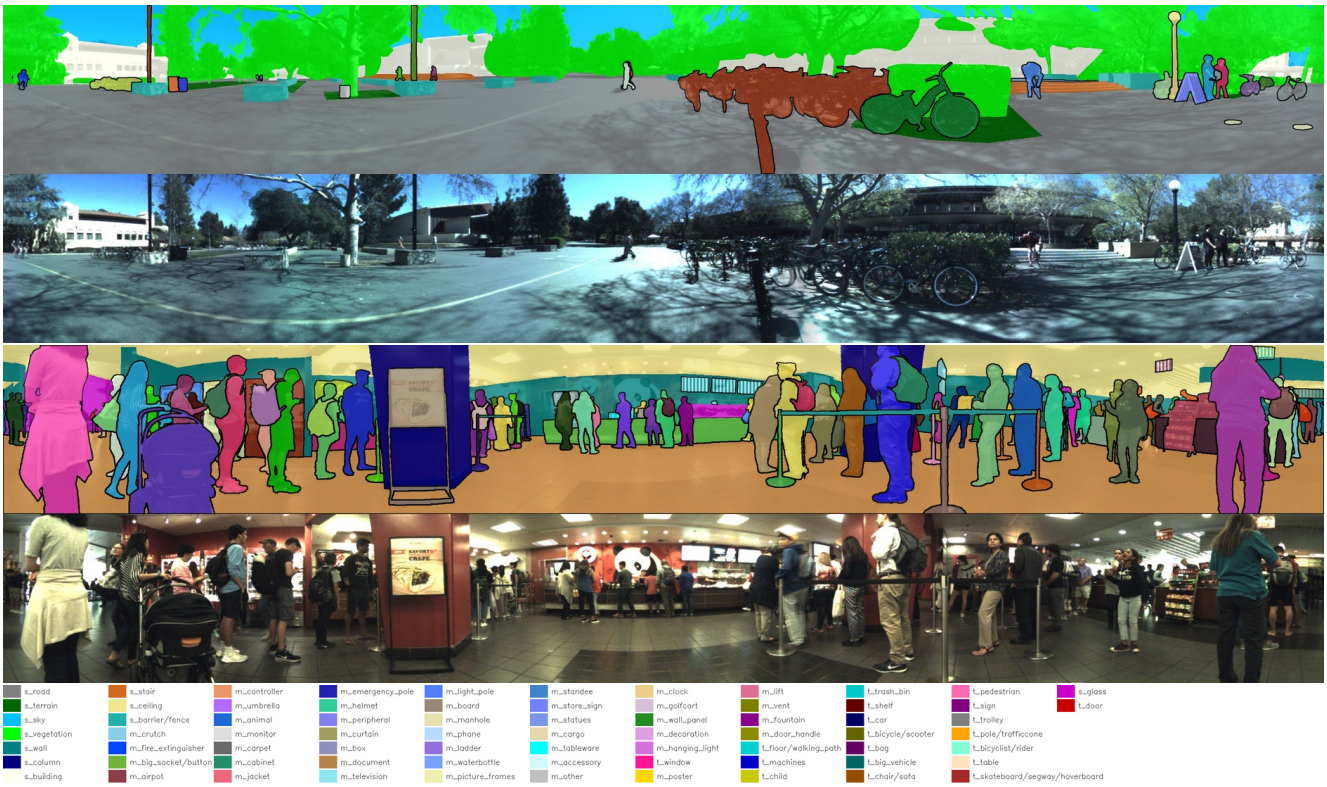


Figure 7. Visualization of JRDB-PanoTrack panoptic segmentation annotations. The images showcase diverse settings, ranging from indoor to outdoor environments, characterized by different lighting conditions, varied levels of human clutter, and dynamic activities. Featured locations include a variety of indoor and outdoor spaces, such as roads, malls, sidewalks, restaurants, plazas, parks, halls, classrooms, laboratories, and office buildings. Meanwhile, we label different floors and objects behind the glass (striped areas) different floors for better real-world robot navigation, human-robot interaction and decision.