Supplementary Material for
Estimating Generic 3D Room Structures from 2D Annotations

Denys Rozumnyi[2*]   **Stefan Popov**[1]   **Kevis-Kokitsi Maninis**[1]
Matthias Nießner[3]   **Vittorio Ferrari**[1]

[1]Google Research   [2]ETH Zurich   [3]Technical University of Munich

1   Additional details

Access details.   The code and the data of this submission can be accessed in the accompanying GitHub repository: https://github.com/google-research/cad-estate.

Dataset documentation.   The dataset is documented in the accompanying GitHub repository. In particular, the included Jupyter notebook src/room_structure_notebook.ipynb allows loading, visualizing, and interactively exploring room structures.

License.   The dataset is released under the CC-BY license. The accompanying code is released under the Apache License 2.0.

Author statement.   The authors of the dataset bear all responsibility in case of violation of rights.

Hosting, licensing, and maintenance plan.   The data is hosted on Google Cloud by Google Research. The authors will maintain and update the dataset to achieve long-term preservation.

Intended use.   The intended use of the dataset is to enable training, evaluation, and benchmarking of 3D room structure estimation from RGB videos.

Supplementary video.   We include additional qualitative results in the supplementary video.

Instructions to annotators.   The full instructions that were used to annotate 2D structural elements (Sec. 3.1 of main paper) can be found here: https://docs.google.com/presentation/d/1MbBOADHqsjev50eOcN3UHUotNL51ro3eK56rLdJACoo. For 2D visible parts – here: https://docs.google.com/presentation/d/1612YdcFHyjBCIASE7jYKjk6XXdEOdmVTICgPJNu5U24. Various aspects of the annotation process (*e.g.* how to achieve faster annotation without degrading the quality) were communicated with the annotators verbally over multiple iterations, and are not fully reflected in the above two links.

2   Failure cases

The proposed method occasionally fails in the presence of annotation mistakes, such as for example annotating the floor and a wall as the same structural element. For instance, Fig. A1 (Reprojection IoU 0.43) shows that two slanted walls (pink) were mistakenly labeled with the same label, and the green floor was labeled as a door and was assigned to the structural element with which it shared most of its boundary (green wall). Similarly, in Fig. A1 (0.58), the yellow ceiling was labeled as a window and was assigned to the yellow wall. Another type of mistake is a wrong plane orientation (Fig. A1, 0.72, pink ceiling), which rare occurs mostly due to an insufficient number of constraints. We filter scenes with such mistakes automatically, as they tend to have low Reprojection IoU after

---

*Work done while the author was at Google.

37th Conference on Neural Information Processing Systems (NeurIPS 2023) Track on Datasets and Benchmarks.
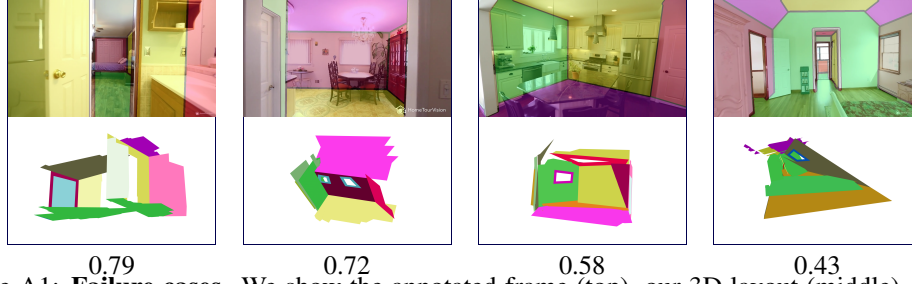
Figure A1: **Failure cases.** We show the annotated frame (top), our 3D layout (middle), and the Reprojection IoU (bottom).

reconstruction (Sec. 3.5, main paper). This however reduces the number of annotated scenes that end up in the dataset (Sec. 4, main paper).

In very rare cases, it can happen that some structural elements have slightly wrong orientation, again due to an insufficiently low number of constraints. For example, in Fig. A1 (0.79), a pink wall has a slightly wrong orientation since there are almost to points for tracking, and only a very small part of the wall is visible in just two frames.

We additionally conducted a small user study by randomly sampling 50 scenes that were filtered out by the quality control (IoU < 0.8) and identify why they failed:

- Around 30 % of failure cases failed because of annotation mistakes (wrong labels).
- Around 40 % failed because of wrong association between adjacent frames with point tracking, which leads to duplicate structural elements. This usually happens when the same structural element reappears after full occlusion.
- Another 30 % failed because of a poorly constrained scenes (eg. very short sequences with forward motion, almost no points for tracking due to clutter and furniture everywhere, small structural elements).

3   Qualitative comparison of different datasets

In Fig. A2, we also include side-by-side visualizations of layouts provided in various datasets as shown in Table 1 in the main paper.

2

| Replica | LayoutNet | MatterportLayout | Zillow Indoor |

| SceneCAD | Realtor360 | PanoContext | Ours |

Figure A2: **Dataset examples.** For each dataset we show examples of input data and the corresponding room layout annotation. The proposed dataset features more diverse type of rooms, while requiring only RGB videos as input and annotating only in the 2D domain.

3