Habitat-Matterport 3D Semantics Dataset
Supplemetary Materials

Karmesh Yadav[1*], Ram Ramrakhya[2*], Santhosh Kumar Ramakrishnan[3*], Theo Gervet[6], John Turner[1], Aaron Gokaslan[4], Noah Maestre[1], Angel Xuan Chang[5], Dhruv Batra[1,2], Manolis Savva[5], Alexander William Clegg[1†], Devendra Singh Chaplot[1†]

[1]Meta AI [2]Georgia Tech [3]UT Austin
[4]Cornell University [5]Simon Fraser University [6]Carnegie Mellon University

A. Appendix Annotation Inferences

A.1. Assumptions

To derive accurate and reasonable inferences for the HM3D dataset annotations, we make the following assumptions about annotations and layouts:

- Objects: Annotations accurately describe the object being annotated.
- Objects: Object annotations with qualifiers in their name, such as "bath xxx" or "kitchen xxx" infer strongly where these objects are found.(i.e. bathrooms or kitchens, respectively)
- Objects: Annotated objects are generally not staged to be "unnatural" in their configuration but rather the object layouts are natural reflections of human use (i.e. we would not expect to see the same region holding a toilet, a refrigerator, a stovetop and a bed).
- Regions: Region annotations are derived from reasonable estimates of room boundaries within the scenes, (i.e. two objects with the same region annotation could be said to be "in the same room")
- Scenes: Scenes are generally reasonable representations of actual human environments and are not staged in some unnatural way (scene full of bathrooms, for instance). We do allow for non-habitation scenes, such as office spaces or restaurants.

Using these assumptions we analyze the semantic annotation text files to learn about the nature of the underlying dataset. We can infer relationships between objects based on mutual regional membership, properties of the regions that contain these objects, and even gain some insight into the nature of the scenes themselves.

*Equal Contribution, Correspondence: karmeshyadav@meta.com
†Equal Contribution

Figure 1. **Visualizing instance segmentation.** In each row, we display an image sampled from ADE20k with overlayed masks denoting the ground-truth (column 1) and model predictions (columns 2-4). The HM3DSEM-trained model generalizes to ADE20k much better than those trained on Gibson and MP3D.

A.2. Instance Segmentation and Object Detection

	Object detection (mAP@0.5)				Instance segmentation (mAP@0.5)			
Training ↓	Gibson	MP3D	HM3D	ADE20k	Gibson	MP3D	HM3D	ADE20k
Gibson	24.4	2.1	3.1	1.1	23.5	0.7	1.3	0.6
MP3D	26.9	27.3	31.7	14.5	21.5	20.1	25.9	10.0
HM3DSEM	**39.6**	**33.7**	**54.0**	**31.8**	**35.2**	**28.2**	**50.1**	**26.4**

Table 1. **Benchmarking object detection, instance segmentation:** We learn Mask-RCNN models on each train dataset (column 1) and evaluate them on all val datasets (columns 2-9). We also test on real-world images from ADE20k. Training on HM3DSEM leads to the best generalization to new scenes and datasets.

We rendered instance segmentation annotations for 150k train, 10k val images from each of HM3DSEM, Gibson, and MP3D. For each dataset, we train a Mask-RCNN for 10 epochs to predict the 6 object classes used in ObjectNav: chair, bed, plant, toilet, tv/monitor, and sofa (similar to [1]). We then evaluate each model on all three val splits for object detection and instance segmentation. We additionally test on ∼500 real-world images of residential scenes from the ADE20k dataset. See Tab. 1. The model trained on HM3DSEM generalizes best across scenes and datasets by a large margin. This echoes our ObjectNav results and reaffirms the value of HM3DSEM for visual perception.

Training on Gibson leads to poorest performance due to annotation inaccuracies and sparsity. We visualize examples in fig. 1.

A.3. Analysis Method

To accomplish this analysis, we hand assigned a region proposal to 261 of the 1624 unique annotation tags provided by the annotators heuristically based on the annotation name. These region annotation proposals are not treated as absolutes, but rather suggestions - if an object found within some region's category tag is mapped to a specific proposal, this proposal serves only to suggest that the containing region might be described using this annotation. In this way, instead of expecting a direct labeling, the objects' region proposals are used as votes for their containing regions' possible annotations.

The possible region names chosen for this experiment (and the category tags mapped to each) are listed below:

- **bathroom**: bath, bath bar, bath cabinet, bath carpet, bath cosmetics, bath curtain, bath curtain bar, bath dial, bath door, bath door frame, bath faucet, bath floor, bath grab bar, bath hanger, bath mat, bath shelf, bath shower, bath side table, bath sink, bath tap, bath towel, bath towels, bath tub, bath utensil, bath wall, bathmat, bathrobe, bathroom accessory, bathroom art, bathroom cabinet, bathroom cabinet door, bathroom cabinet drawer, bathroom counter, bathroom floor, bathroom glass, bathroom mat, bathroom rug, bathroom shelf, bathroom stuff, bathroom towel, bathroom utensil, bathroom utensils, bathroom wall, bathroom window, bathtub, bathtub knob, bathtub platform, bathtub tap, bathtub utensil, bidet, shower, shower bar, shower base, shower battery, shower bench, shower cabin, shower cabinet, shower caddy, shower case, shower ceiling, shower ceiling lamp, shower cockpit, shower cosmetics, shower curtain, shower curtain bar, shower curtain rod, shower dial, shower door, shower door frame, shower door knob, shower floor, shower frame, shower glass, shower grab bar, shower handle, shower handrail, shower hanger, shower hose, shower hose/head, shower knob, shower mat, shower mirror, shower pipe, shower rail, shower rod, shower seat, shower shelf, shower soap shelf, shower stall, shower step, shower tap, shower tray, shower tub, shower utensil, shower valve, shower wall, shower wall cubby, shower window frame, shower-bath cabinet, showerhead, toilet, toilet brush, toilet brush holder, toilet cleaner, toilet paper, toilet paper dispenser, toilet seat, toothbrush, toothpaste, wall toilet paper
- **bedroom**: bed, bed base, bed cabinet, bed cabinet lamp, bed comforter, bed curtain, bed ladder, bed light, bed sheet, bed small, bed stand, bed table, bedding, bedframe, bedpost, bedroom ceiling, bedroom table, bedside cabinet, bedside cabinet door, bedside cabinet drawer, bedside lamp, bedside table, ceiling bedroom, dresser, jewelry box, nightstand, wardrobe
- **dining room**: dining chair, dining table, dinner chair, dinner table
- **garage**: garage door, garage door frame, garage door motor, garage door opener, garage door opener bar, garage door opener motor, garage door opener railing, garage door railing, garage light
- **hall/stairwell**: stair, stair frame, stair handle, stair step, stair wall, staircase, staircase handrail, staircase trim, staircase wall, stairs, stairs railing, stairs skirt, stairs trim, stairs wall, stairwell
- **kitchen**: cabinet kitchen, dish rack, dishwasher, fridge, kitchen appliance, kitchen cabinet, kitchen cabinet door, kitchen cabinet drawer, kitchen cabinet lower, kitchen ceiling, kitchen chair, kitchen counter, kitchen counter item, kitchen counter support, kitchen countertop item, kitchen countertop items, kitchen decoration, kitchen extractor, kitchen gloves, kitchen handle, kitchen island, kitchen knife set, kitchen lower cabinet, kitchen lower shelf, kitchen seating, kitchen shelf, kitchen sink, kitchen sink cabinet, kitchen table, kitchen top, kitchen towel, kitchen utensil, kitchen utensils, kitchen wall, kitchenware, knife holder, knife set, oven, oven and stove, refrigerator, refrigerator cabinet, stove, stovetop
- **laundry room**: washer-dryer, washing machine, washing machine and dryer, washing powder, washing stuff
- living room: circular sofa, coffee table, couch, l-shaped sofa, recliner, remote control, sofa, sofa chair, sofa seat, sofa set
- **office**: computer, computer chair, computer desk, computer equipment, computer mouse, computer tower, desk, desk cabinet, desk chair, desk clutter, desk door, desk lamp, desk organizer, laptop, office chair, office table, office wall
- **rec room**: barbell, exercise ball, exercise bike, exercise equipment, exercise ladder, exercise machine, exercise mat, exercise mat roll, exercising blocks, foosball game table, foosball table, gym equipment, gym mat, gym rope, gym stepper, pool stick, pool table, rack of weights, weight bench, yoga mat

We then record the category tag for every object instance in the dataset, along with region annotation proposals for all categories that have been assigned them, on a per tag (not per object instance) basis, and organize this data per region per scene.

A.4. Scene-level Statistics

Some observations about all 216 scenes were made using the region labeling heuristics and examining category presence as well as room annotations derived from proposal votes.

Using only proposal-tagged category presence as a guide, we found:

- 12 scenes lacked any objects containing tags with the “bedroom” proposal. These scenes were all visually verified to be commercial spaces, either offices, restaurants, or stores.
- 7 scenes lacked category tags with the “bathroom” proposal. These were also visually verified to be non-residential spaces. 1 scene had instances of “bedroom” categories but none of “bathroom”; this scene is a large house that has been converted to a museum.
- 25 scenes contained objects with proposed “garage” region annotations. These were visually verified to all contain garages.
- 9 scenes lacked any objects with proposed “kitchen” region annotations. 4 of these were commercial spaces that also lacked “bedroom” proposals, while 4 of the remaining 5 were hotel rooms or suites. The last remaining scene was found to actually contain a kitchen through visual inspection, which had a modern design, lacking most obvious appliances such as “stove” or “oven”; however, a refrigerator was present and visible but was mislabeled as a “cabinet”.

By aggregating votes per region of the number of category-derived region proposals, we derived potential room labels, which provided even more accurate suggestions of scene content, as shown in section 3.4.

A.5. Region Label Inference

It would be useful if the region proposal votes derived from each region’s constituent object categories could be used to yield labels for the region itself. By randomly picking scenes, the legitimacy of this data for region labeling could be investigated through visual verification in the Habitat engine. Figure 2 shows the aggregation results for a randomly chosen scene.

Using the highest vote counts per region/row to suggest that region’s proposed annotation, this scene’s 13 regions are proposed to be 3 bathrooms, 4 bedrooms, 3 hallway/stairwells, 1 kitchen, 1 living room, 1 of either bedroom or dining room. Visually inspecting this room yields the same count, with the confused room being the dining room. A serving buffet is mislabeled in this room as a dresser.

Even for larger scenes with many regions, the per-region proposal aggregations can provide useful insights. Figure 3 shows the results for a larger scene.

Note that 3 of the 21 regions in this scene lack specific proposal votes (regions 1,2,8) due to the categories of the objects found in these regions not having region proposals. Of the 18 regions with proposals, 5 bathrooms (6,11,12,15,17), 4 bedrooms(9,10,14,20), 1 living room(0), 2 offices(3 ,16), 1 rec room(19) and 2 kitchens(4,7), along with 3 ambiguous mappings of 1 bedroom or office (13), 1 hallway or bedroom (5), and 1 hallway or rec room (18).

Visually inspecting this scene yields very similar results: 2 rec rooms(regions 18,19), 2 kitchen(4,7), 1 dining room(1), 1 office(3), 3 stairs/hallways (2,5,8), 5 bathrooms (6, 11,12, 15,17), 6 bedrooms (9,10, 13, 14, 16, 20).

Ambiguity in the proposal assignments can be mitigated if certain categories, such as “bed”, received more votes, although this might miscategorize regions where beds were in storage. Figure 4 and Figure 5 show these same two scenes with "bed" category receiving 10 votes for "bedroom" proposal instead of 1.

Using these region proposals at the scene level gives a reasonable estimate for the room layout and count of each scene.

A.6. Object-Level Analysis and Files

Along with collecting and organizing object instance-based category data organized by scene and then region, we also aggregated the scene, region and per-region neighbor categories for each category present across the entire dataset, so that the categories of all objects that share a region are known to one another, as are the region annotation proposals for those categories that have them.

The statistics on category prevalence throughout all the regions in the dataset provides suggestions for possible region proposal labels for otherwise unmapped category tags based on the category “company they keep”.

We have provided the following files to assist users in conducting their own analyses of the semantic scenes.

- **HM3D_CountsOfObjectTypes.csv** : This file provides the name and number of occurrences of every category label in use across the entire dataset.
- **Per_Category_Counts_Uncommon.csv** : This file provides the number of occurrences of every category label not including common architectural components (excluding doors/walls/ceilings/etc)
- **Per_Scene_Neighborhood_Stats.csv** : This file contains scene-level category statistics (mean, variance, skew, kurtosis) describing number of regions per scene and the unique categories and object instances they contain.
- **Per_Scene_Region_Neighborhoods.csv** : This file contains per-scene-per-region unique category and instance counts, and all object instance labels (including common labels) within each region.
- **Per_Scene_Region_Votes.csv** : This file lists the per-scene-per-region votes for region/room label proposal based on the categories of the various object instances present with hand-annotated labels. Each object instance of a category with a region proposal gets 1 vote.
- **Per_Scene_Total_Votes.csv** : This file has the per-scene room label proposal counts (i.e. how many proposed bedrooms, bathrooms, etc. are present in a scene), built from the “Per_Scene_Region_Votes.csv” data.

Scene Name	Region #	Bathroom	Bedroom	Dining room	Garage	Hall/stairwell	Kitchen	Laundry room	Living room	Office	Rec room	Unknown room	Room Proposal
00546-nS8T59Aw3sf	1	1	5	0	0	0	0	0	0	0	0	21	Bedroom
00546-nS8T59Aw3sf	2	4	0	0	0	0	0	0	0	0	0	13	Bathroom
00546-nS8T59Aw3sf	3	0	0	0	0	2	0	0	0	0	0	6	Hall/stairwell
00546-nS8T59Aw3sf	4	1	4	0	0	0	0	0	0	2	0	23	Bedroom
00546-nS8T59Aw3sf	5	0	5	0	0	0	0	0	0	1	0	25	Bedroom
00546-nS8T59Aw3sf	6	0	2	0	0	0	0	0	0	0	0	15	Bedroom
00546-nS8T59Aw3sf	7	5	0	0	0	0	0	0	0	0	0	15	Bathroom
00546-nS8T59Aw3sf	8	0	1	0	0	0	0	0	2	0	0	26	Living room
00546-nS8T59Aw3sf	9	0	1	1	0	0	0	0	0	0	0	18	Tie: Bedroom & Dining room
00546-nS8T59Aw3sf	10	2	0	0	0	0	0	0	0	0	0	11	Bathroom
00546-nS8T59Aw3sf	11	0	0	0	0	0	4	0	0	0	0	30	Kitchen
00546-nS8T59Aw3sf	12	0	0	0	0	1	0	0	0	0	0	11	Hall/stairwell
00546-nS8T59Aw3sf	13	0	1	0	0	2	0	0	0	0	0	15	Hall/stairwell

Figure 2. Scene 00546-nS8T59Aw3sf Region label proposals based on category presence

Scene Name	Region #	Bathroom	Bedroom	Dining room	Garage	Hall/stairwell	Kitchen	Laundry room	Living room	Office	Rec room	Unknown room	Room Proposal
00064-gQgtJ9Stk5s	0	0	0	0	0	0	0	0	3	1	0	55	Living room
00064-gQgtJ9Stk5s	1	0	0	0	0	0	0	0	0	0	0	33	Unknown room
00064-gQgtJ9Stk5s	2	0	0	0	0	0	0	0	0	0	0	5	Unknown room
00064-gQgtJ9Stk5s	3	0	1	0	0	0	0	0	0	2	0	26	Office
00064-gQgtJ9Stk5s	4	0	0	0	0	1	9	0	0	0	0	26	Kitchen
00064-gQgtJ9Stk5s	5	0	1	0	0	1	0	0	0	0	0	15	Tie: Hall/stairwell & Bedroom
00064-gQgtJ9Stk5s	6	2	0	0	0	0	0	0	0	0	0	12	Bathroom
00064-gQgtJ9Stk5s	7	0	0	0	0	0	4	0	0	0	0	15	Kitchen
00064-gQgtJ9Stk5s	8	0	0	0	0	0	0	0	0	0	0	19	Unknown room
00064-gQgtJ9Stk5s	9	0	4	0	0	1	0	0	0	0	0	16	Bedroom
00064-gQgtJ9Stk5s	10	0	2	0	0	0	0	0	0	1	0	32	Bedroom
00064-gQgtJ9Stk5s	11	4	0	0	0	0	0	0	0	0	0	12	Bathroom
00064-gQgtJ9Stk5s	12	4	0	0	0	0	0	0	0	0	0	10	Bathroom
00064-gQgtJ9Stk5s	13	0	2	0	0	0	0	0	0	2	0	23	Tie: Office & Bedroom
00064-gQgtJ9Stk5s	14	0	3	0	0	0	0	0	0	0	0	26	Bedroom
00064-gQgtJ9Stk5s	15	8	0	0	0	0	0	0	0	0	0	10	Bathroom
00064-gQgtJ9Stk5s	16	0	1	0	0	0	0	0	0	3	0	39	Office
00064-gQgtJ9Stk5s	17	4	0	0	0	0	0	0	0	0	0	23	Bathroom
00064-gQgtJ9Stk5s	18	0	1	0	0	2	0	0	1	1	2	39	Tie: Hall/stairwell & Rec room
00064-gQgtJ9Stk5s	19	0	0	0	0	1	0	0	0	1	2	31	Rec room
00064-gQgtJ9Stk5s	20	0	2	0	0	1	0	0	0	1	0	42	Bedroom

Figure 3. Scene 00064-gQgtJ9Stk5s Region label proposals based on category presence

- **Per_Scene_Region_Weighted_Votes.csv** : This file also lists the per-scene-per-region votes for region/room label proposal based on the categories of the various object instances present with hand-annotated labels, except in this case, instances of the category "bed" receive 10 votes. All other object instances of categories with assigned region proposals still receive 1 vote.
- **Per_Scene_Total_Weighted_Votes.csv** : This file has the per-scene room label proposal counts (i.e. how many proposed bedrooms, bathrooms, etc. are present in a scene), built from the "Per_Scene_Region_Weighted_Votes.csv" data, where instances of the "bed" category received 10 votes instead of just 1.
- **Region_Tag_Mappings.csv** : This file lists per-scene-per-region count of categories present, and names of "uncommon" tags (excluding common architectural categories like wall, ceiling/etc)

The following files include the region proposal aggregate categories in their reporting. Each of these aggregate categories are formed by a union of all the categories that share the same hand-annotated region proposal.

- **Per_Category_Region_Neighbors.csv** : This file provides statistics for category and instance presence in scenes and regions. This includes the number of scenes and number of regions that instances of the category are present, as well as the number of instances total and the average number of instances per scene and per region when present. The total number of unique neighbor categories, where a neighbor is defined as sharing the same region, for each category is also listed as well as the categories and region counts of each neighbor.
- **Per_Category_Region_Per_Cat_Votes.csv** : This file lists the per-category hand-assigned region proposal tags (bed inferring bedroom, for example), if present, as well as the counts of other neighbor categories' hand-labeled region proposals. This is useful in determining the types of regions where instances of categories are most likely to be found. For example, the "air vent" category shares regions with 206 instances of "bathroom"-labeled categories, 49 instances of "bedroom"-labeled categories, 61 instances of "kitchen"-labeled categories, etc.
- **Per_Scene_Region_Cat_Presence.csv** : This file holds per-scene-per-region unique category presence and count of instances of each category.

Scene Name	Region #	Bathroom	Bedroom	Dining room	Garage	Hall/stairwell	Kitchen	Laundry room	Living room	Office	Rec room	Unknown room	Room Proposal
00546-nS8T59Aw3sf	1	1	14	0	0	0	0	0	0	0	0	21	Bedroom
00546-nS8T59Aw3sf	2	4	0	0	0	0	0	0	0	0	0	13	Bathroom
00546-nS8T59Aw3sf	3	0	0	0	0	2	0	0	0	0	0	6	Hall/stairwell
00546-nS8T59Aw3sf	4	1	13	0	0	0	0	0	0	2	0	23	Bedroom
00546-nS8T59Aw3sf	5	0	14	0	0	0	0	0	0	1	0	25	Bedroom
00546-nS8T59Aw3sf	6	0	11	0	0	0	0	0	0	0	0	15	Bedroom
00546-nS8T59Aw3sf	7	5	0	0	0	0	0	0	0	0	0	15	Bathroom
00546-nS8T59Aw3sf	8	0	1	0	0	0	0	0	2	0	0	26	Living room
00546-nS8T59Aw3sf	9	0	1	1	0	0	0	0	0	0	0	18	Tie: Bedroom & Dining room
00546-nS8T59Aw3sf	10	2	0	0	0	0	0	0	0	0	0	11	Bathroom
00546-nS8T59Aw3sf	11	0	0	0	0	0	4	0	0	0	0	30	Kitchen
00546-nS8T59Aw3sf	12	0	0	0	0	1	0	0	0	0	0	11	Hall/stairwell
00546-nS8T59Aw3sf	13	0	1	0	0	2	0	0	0	0	0	15	Hall/stairwell

Figure 4. Scene 00546-nS8T59Aw3sf Region label proposals based on category presence with weighting

Scene Name	Region #	Bathroom	Bedroom	Dining room	Garage	Hall/stairwell	Kitchen	Laundry room	Living room	Office	Rec room	Unknown r	Room Proposal
00064-gQgtJ9Stk5s	0	0	0	0	0	0	0	0	3	1	0	55	Living room
00064-gQgtJ9Stk5s	1	0	0	0	0	0	0	0	0	0	0	33	Unknown room
00064-gQgtJ9Stk5s	2	0	0	0	0	0	0	0	0	0	0	5	Unknown room
00064-gQgtJ9Stk5s	3	0	1	0	0	0	0	0	0	2	0	26	Office
00064-gQgtJ9Stk5s	4	0	0	0	0	1	9	0	0	0	0	26	Kitchen
00064-gQgtJ9Stk5s	5	0	1	0	0	1	0	0	0	0	0	15	Tie: Hall/stairwell & Bedroom
00064-gQgtJ9Stk5s	6	2	0	0	0	0	0	0	0	0	0	12	Bathroom
00064-gQgtJ9Stk5s	7	0	0	0	0	0	4	0	0	0	0	15	Kitchen
00064-gQgtJ9Stk5s	8	0	0	0	0	0	0	0	0	0	0	19	Unknown room
00064-gQgtJ9Stk5s	9	0	13	0	0	1	0	0	0	0	0	16	Bedroom
00064-gQgtJ9Stk5s	10	0	11	0	0	0	0	0	0	1	0	32	Bedroom
00064-gQgtJ9Stk5s	11	4	0	0	0	0	0	0	0	0	0	12	Bathroom
00064-gQgtJ9Stk5s	12	4	0	0	0	0	0	0	0	0	0	10	Bathroom
00064-gQgtJ9Stk5s	13	0	11	0	0	0	0	0	0	2	0	23	Bedroom
00064-gQgtJ9Stk5s	14	0	12	0	0	0	0	0	0	0	0	26	Bedroom
00064-gQgtJ9Stk5s	15	8	0	0	0	0	0	0	0	0	0	10	Bathroom
00064-gQgtJ9Stk5s	16	0	10	0	0	0	0	0	0	3	0	39	Bedroom
00064-gQgtJ9Stk5s	17	4	0	0	0	0	0	0	0	0	0	23	Bathroom
00064-gQgtJ9Stk5s	18	0	1	0	0	2	0	0	1	1	2	39	Tie: Rec room & Hall/stairwell
00064-gQgtJ9Stk5s	19	0	0	0	0	1	0	0	0	1	2	31	Rec room
00064-gQgtJ9Stk5s	20	0	11	0	0	1	0	0	0	1	0	42	Bedroom

Figure 5. Scene 00064-gQgtJ9Stk5s Region label proposals based on category presence with weighting

References

[1] Devendra Singh Chaplot, Dhiraj Prakashchand Gandhi, Abhinav Gupta, and Russ R Salakhutdinov. Object goal navigation using goal-oriented semantic exploration. *Advances in Neural Information Processing Systems*, 33:4247–4258, 2020. 1