

Habitat-Matterport 3D Semantics Dataset



Karmesh Yadav^{1*}



Ram Ramrakhya^{2*}



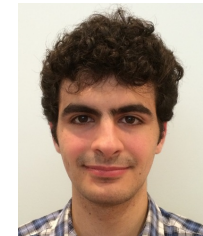
Santhosh Kumar
Ramakrishnan^{3*}



Theophile Gervet⁶



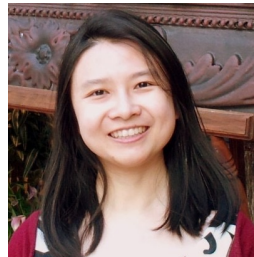
John M Turner¹



Aaron Gokaslan⁴



Noah D Maestre¹



Angel X Chang⁵



Dhruv Batra^{1,2}



Manolis Savva⁵



Alexander Clegg^{1†}



Devendra Singh
Chaplot^{1†}

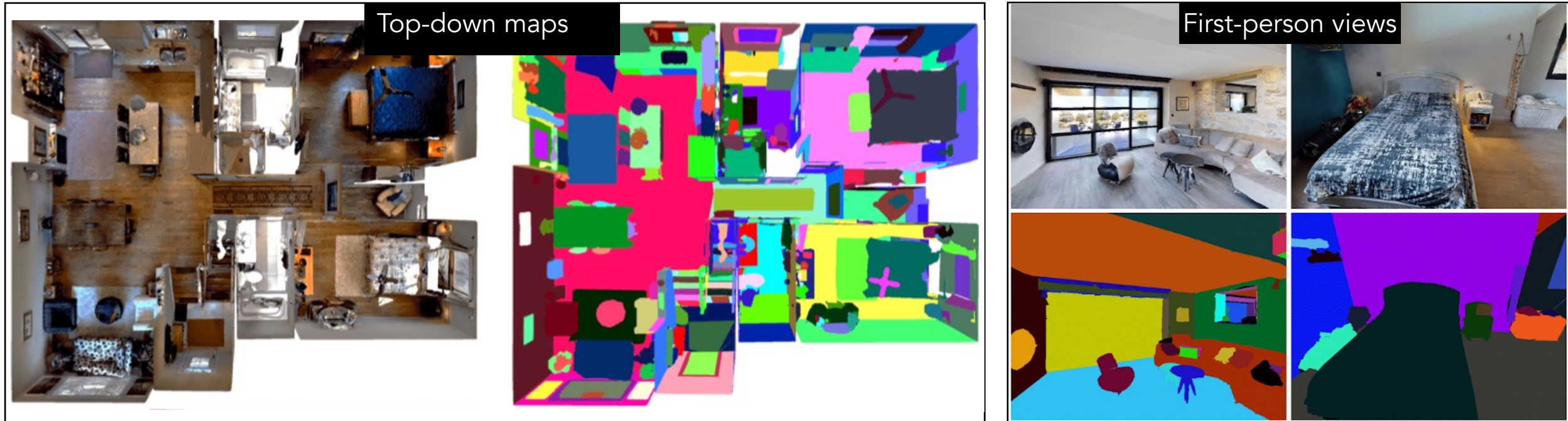
Poster session: TUE-PM-075

Project page: <https://aihabitat.org/datasets/hm3d-semantics/>



Overview

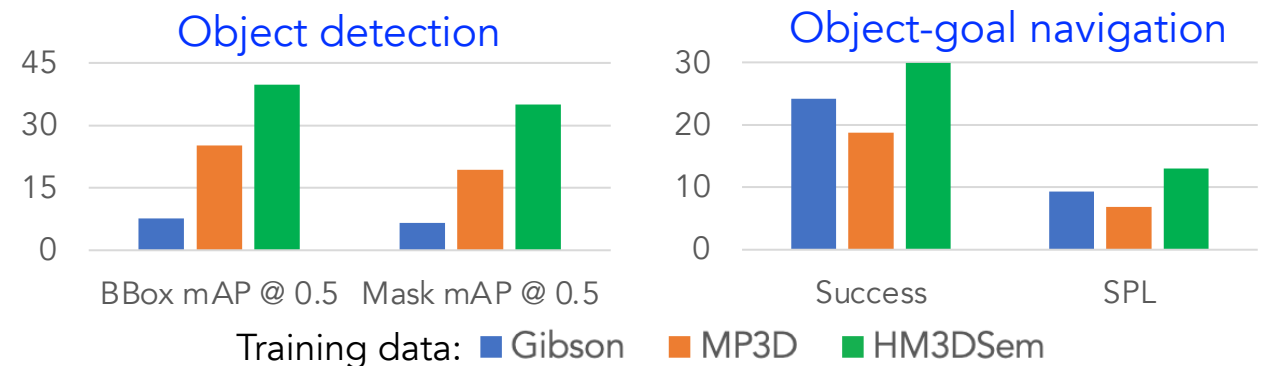
We introduce HM3DSem, the largest ever open dataset of 3D real-world spaces with densely annotated semantics.



Highlights of HM3DSem:

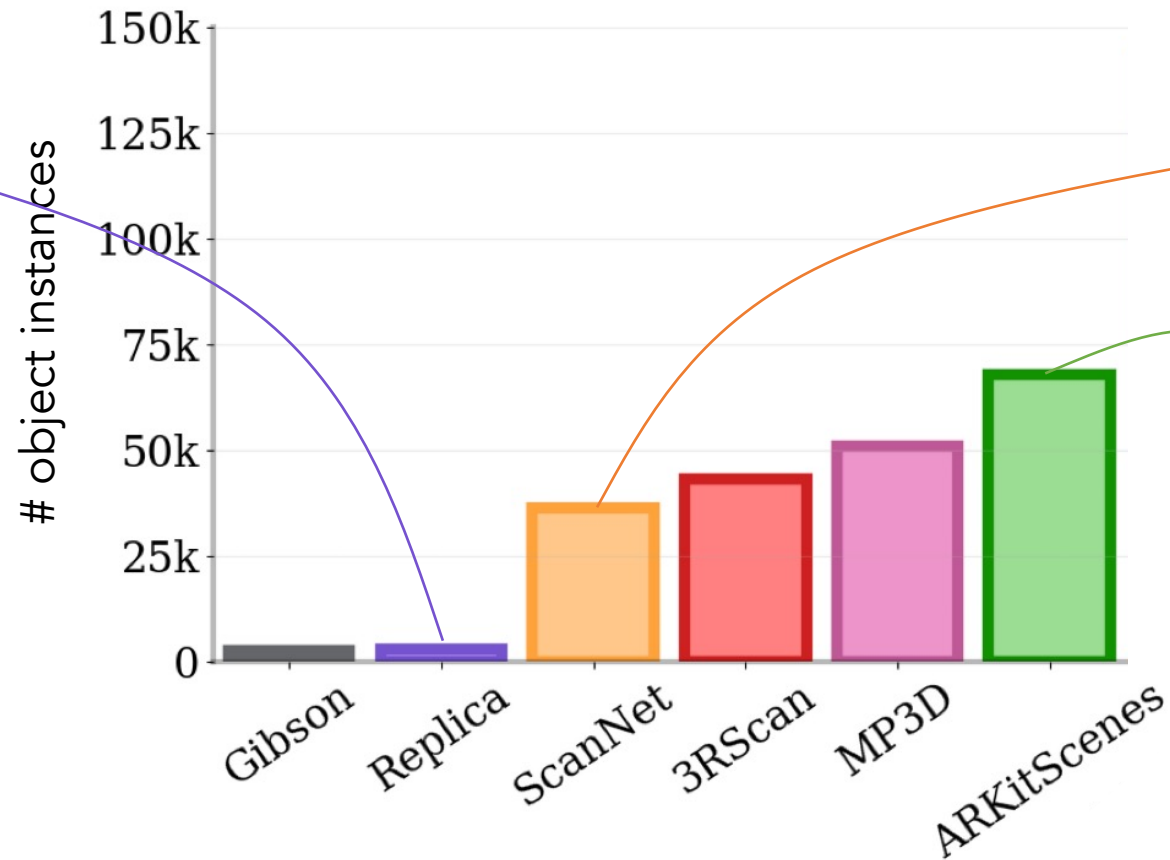
- 216 scenes with texture annotations
- Pixel-level alignment between semantic & scene meshes
- 14,200+ hours of human effort from 20+ annotators
- Room annotations to enable smart dataset subsampling

Models trained on HM3DSem generalize best across datasets



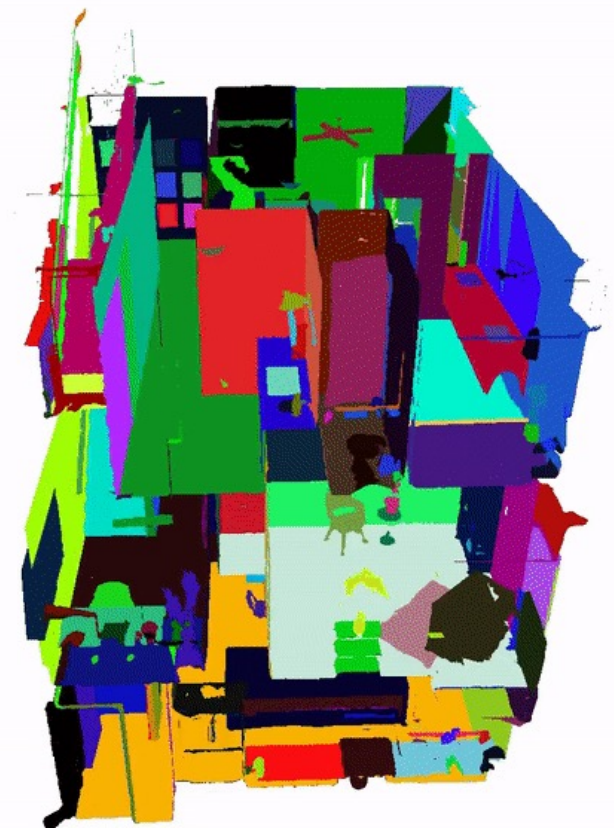
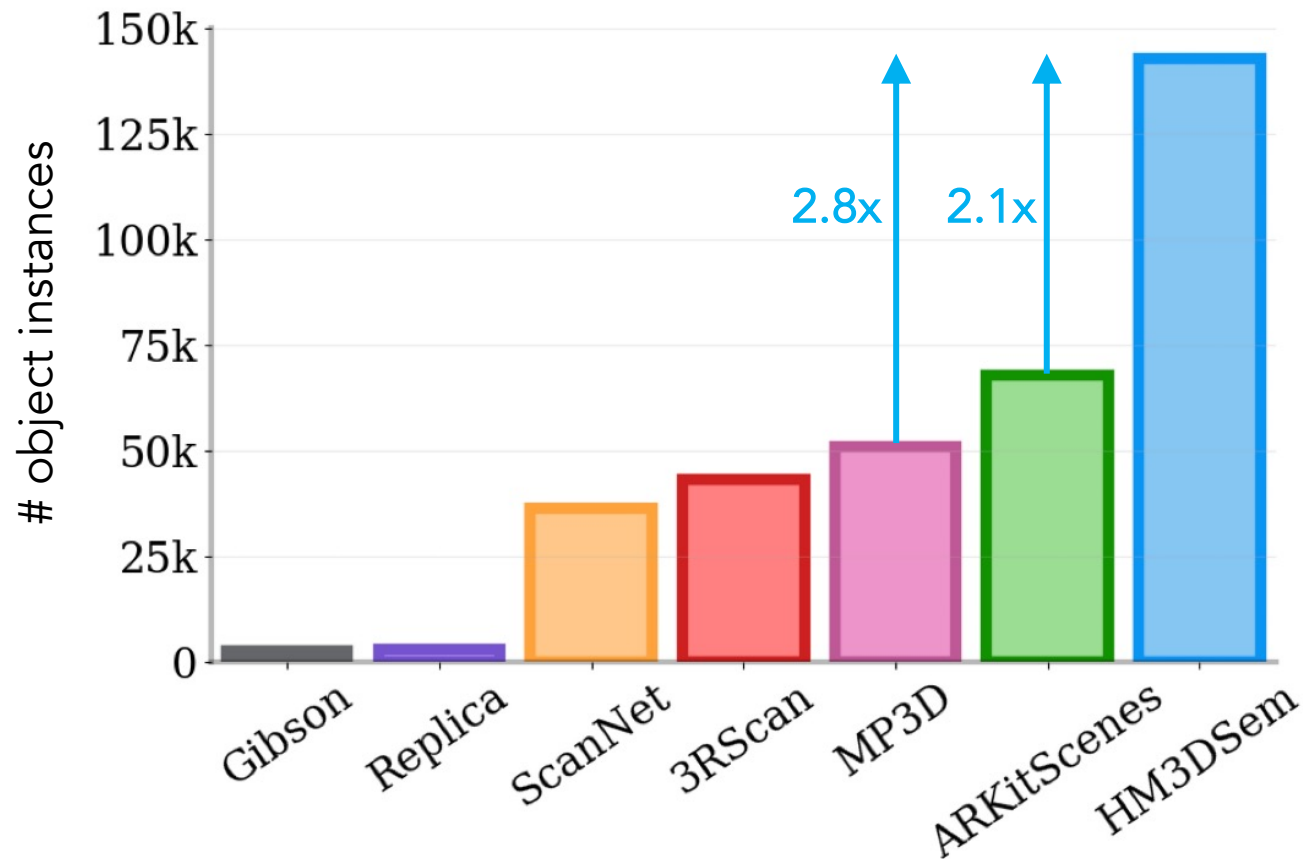
Motivation

Prior semantically annotated datasets are limited by size / quality.



Habitat-Matterport 3D Semantics

HM3DSem is the largest ever open dataset of 3D real-world spaces with densely annotated semantics.



Annotation pipeline

Artists perform instance-level annotations (~13.5k hours)

- Annotations for architectural elements, objects, wall decorations, etc
- Artists paint unique colors per instance on 3D textures
- Each unique color is mapped to object name + room id

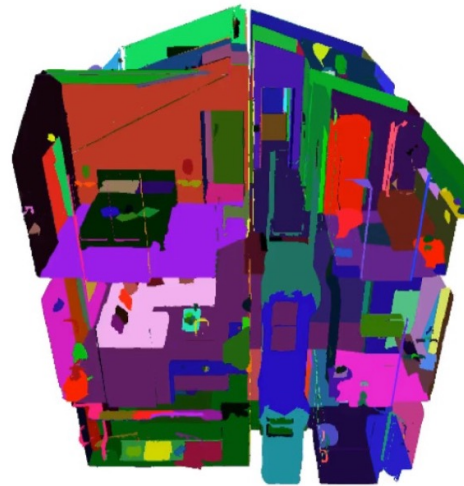


HM3D scene
3D mesh reconstruction

Paint 3D mesh



Unique color
per instance



Instance-level
texture annotation
glTF basis (.glb) format



Map colors
to objects
and rooms

Object id	color	object name	room id
1	E83FC9	floor	0
2	AC11D9	wall	0
.	.	.	.
17	109A8E	chair	0
18	9C19AC	painting	0
.	.	.	.
48	A3192E	wall	1
49	BEB769	ceiling	1
.	.	.	.
55	6DB1F8	desk	1
.	.	.	.

Object name annotation
Semantic .txt file

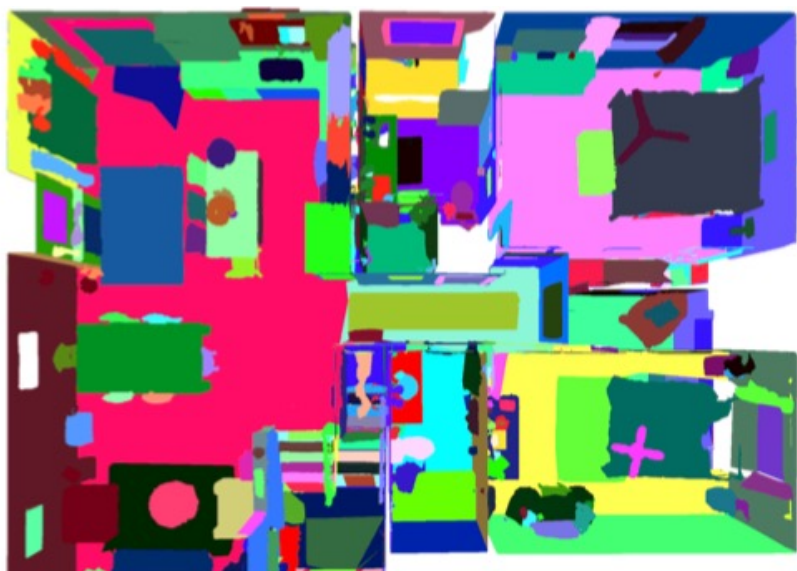
Verification process

Annotators iteratively verify & correct annotations (~640 hours)

- Automated checks
- Manual verification
 - Correct typos and group synonyms
 - Identify painting annotation errors

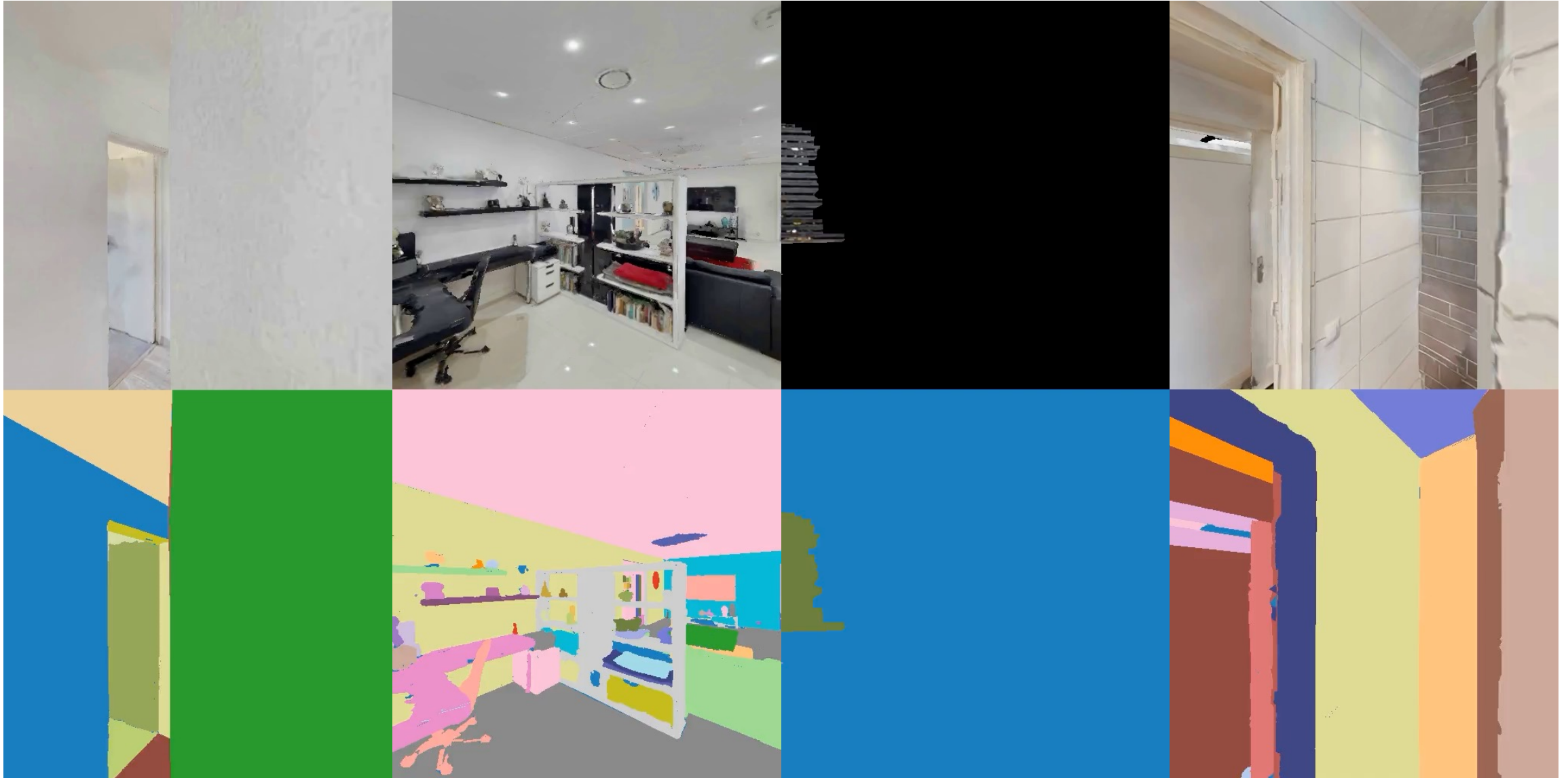
HM3DSem

Annotations are dense, high-quality, perfectly aligned with 3D meshes (unlike prior work).



HM3DSem

Annotations are dense, high-quality, perfectly aligned with 3D meshes (unlike prior work).



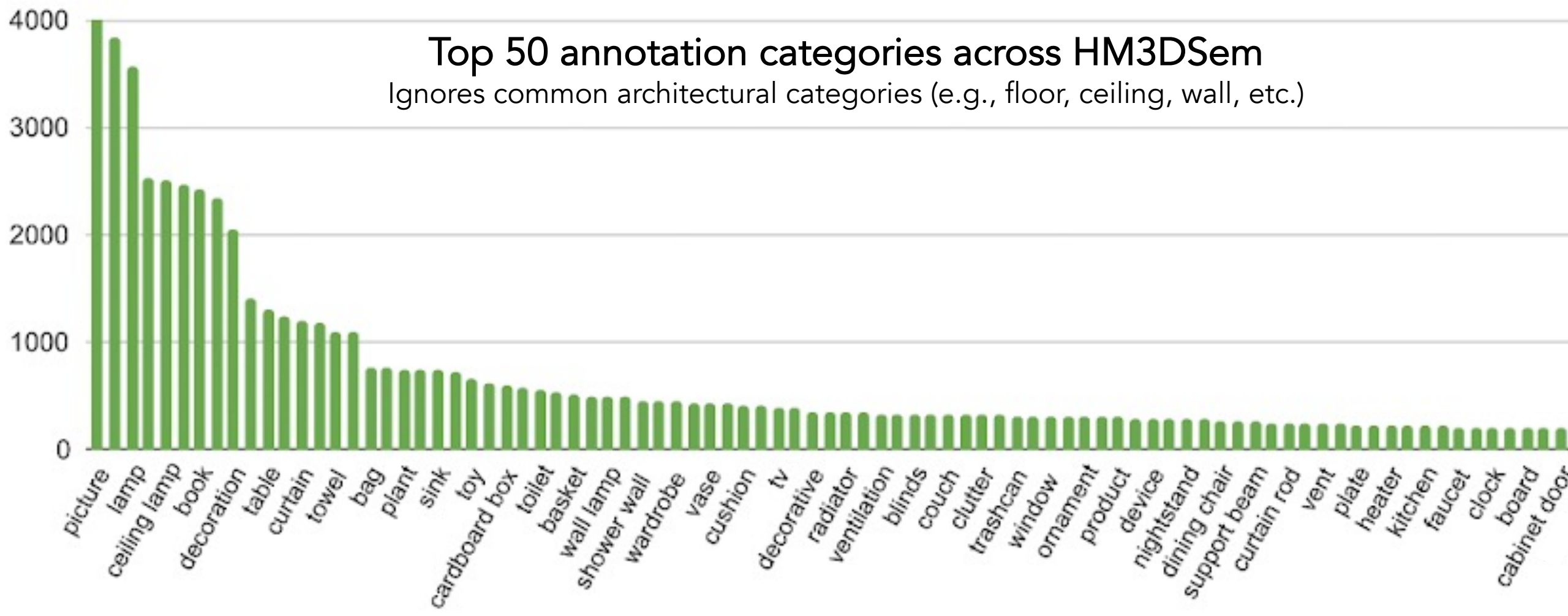
HM3DSem statistics

142,646 object instances

1,625 category tags

3,100 regions

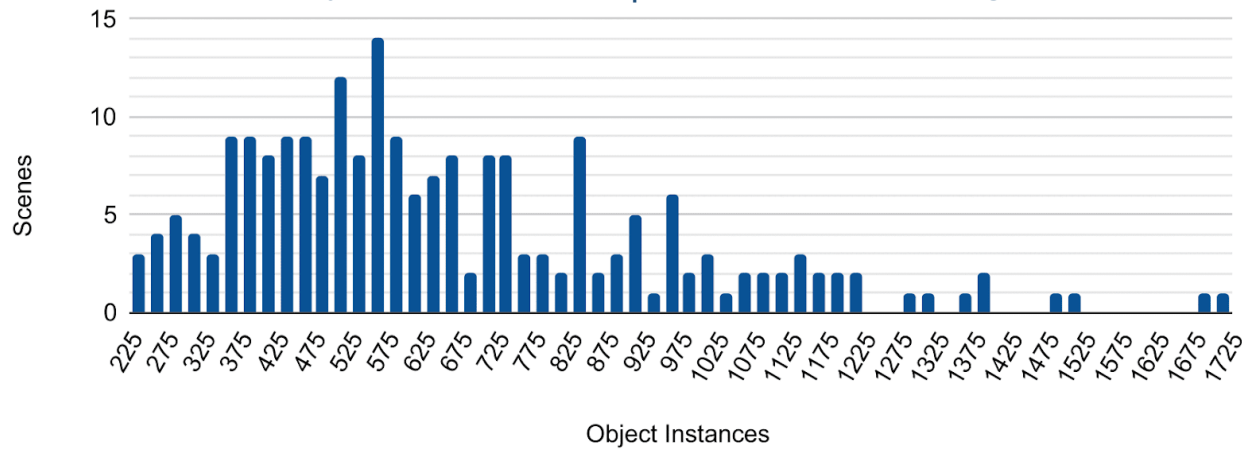
Top 50 annotation categories across HM3DSem
Ignores common architectural categories (e.g., floor, ceiling, wall, etc.)



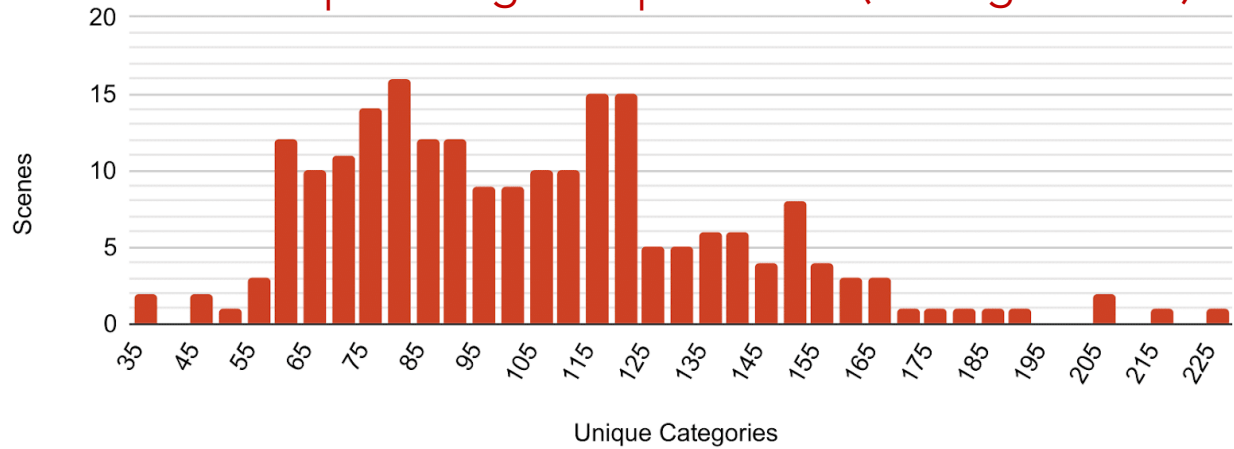
HM3DSem statistics

Diverse distributions over **objects**, **categories**, and **regions** per scene

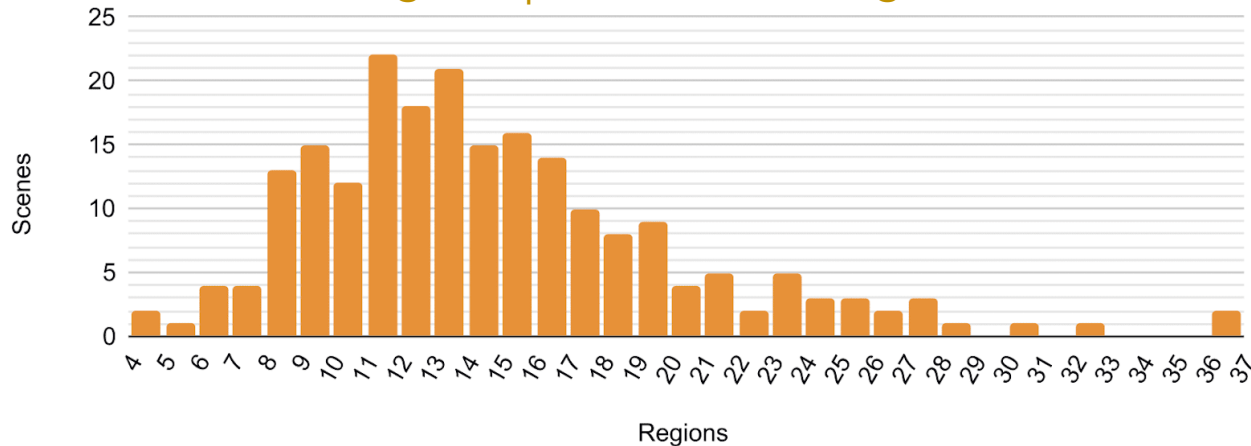
object instances per scene (average = 660)



unique categories per scene (average = 106)



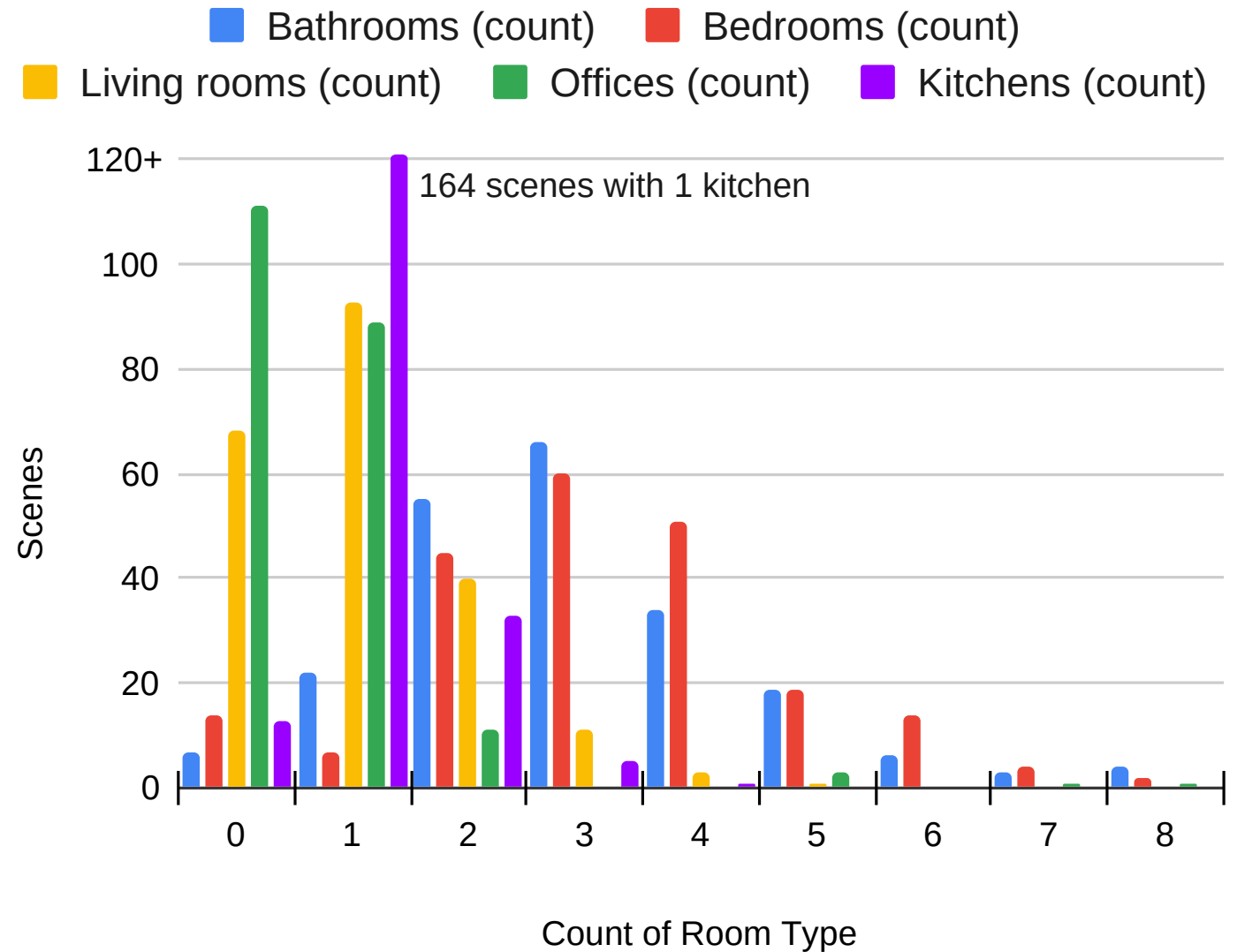
regions per scene (average = 14)



HM3DSem statistics

Room labels can be derived from the presence of certain objects.

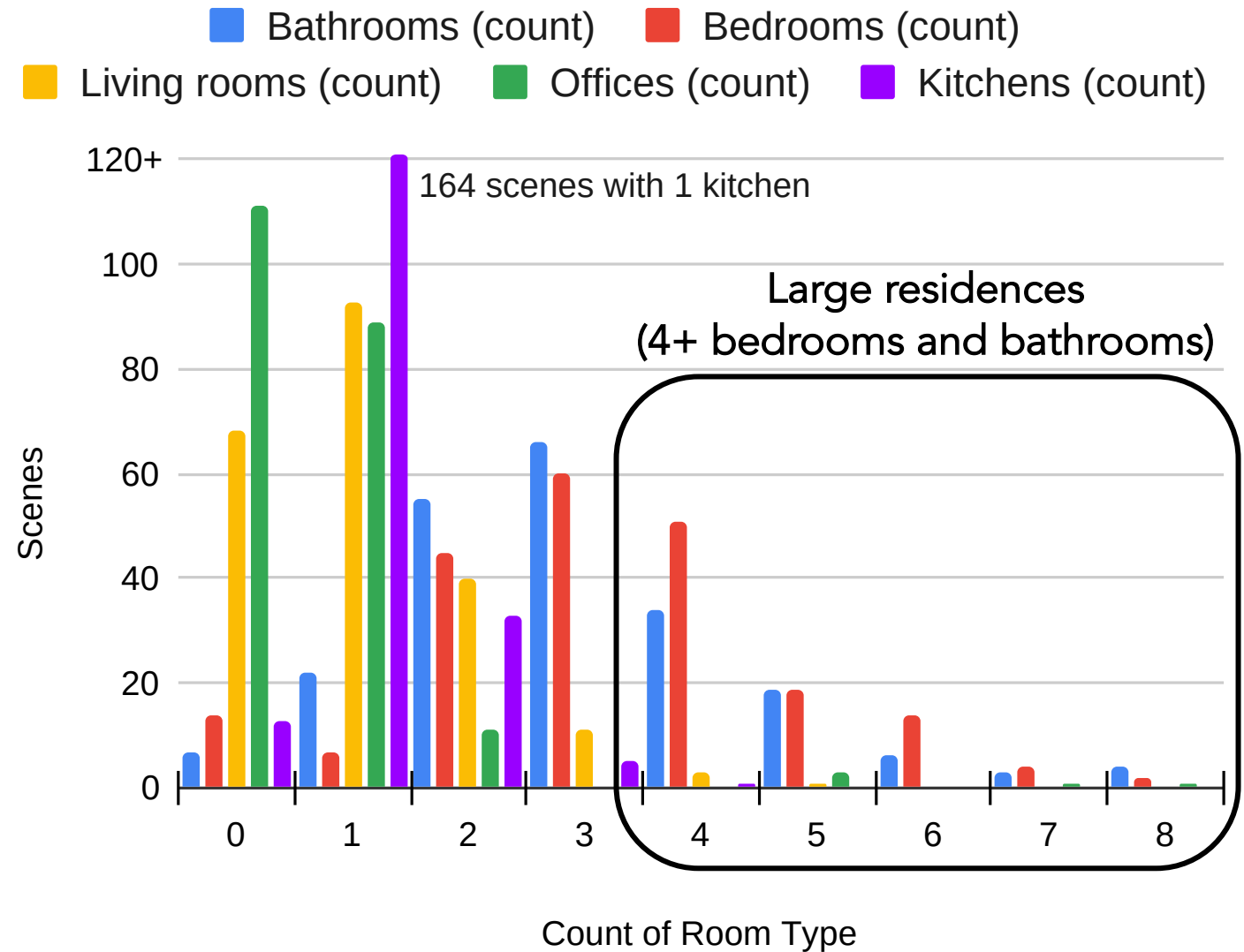
e.g., bed must be in a bedroom, toilet must be in a bathroom, etc.



HM3DSem statistics

Room labels can be derived from the presence of certain objects.

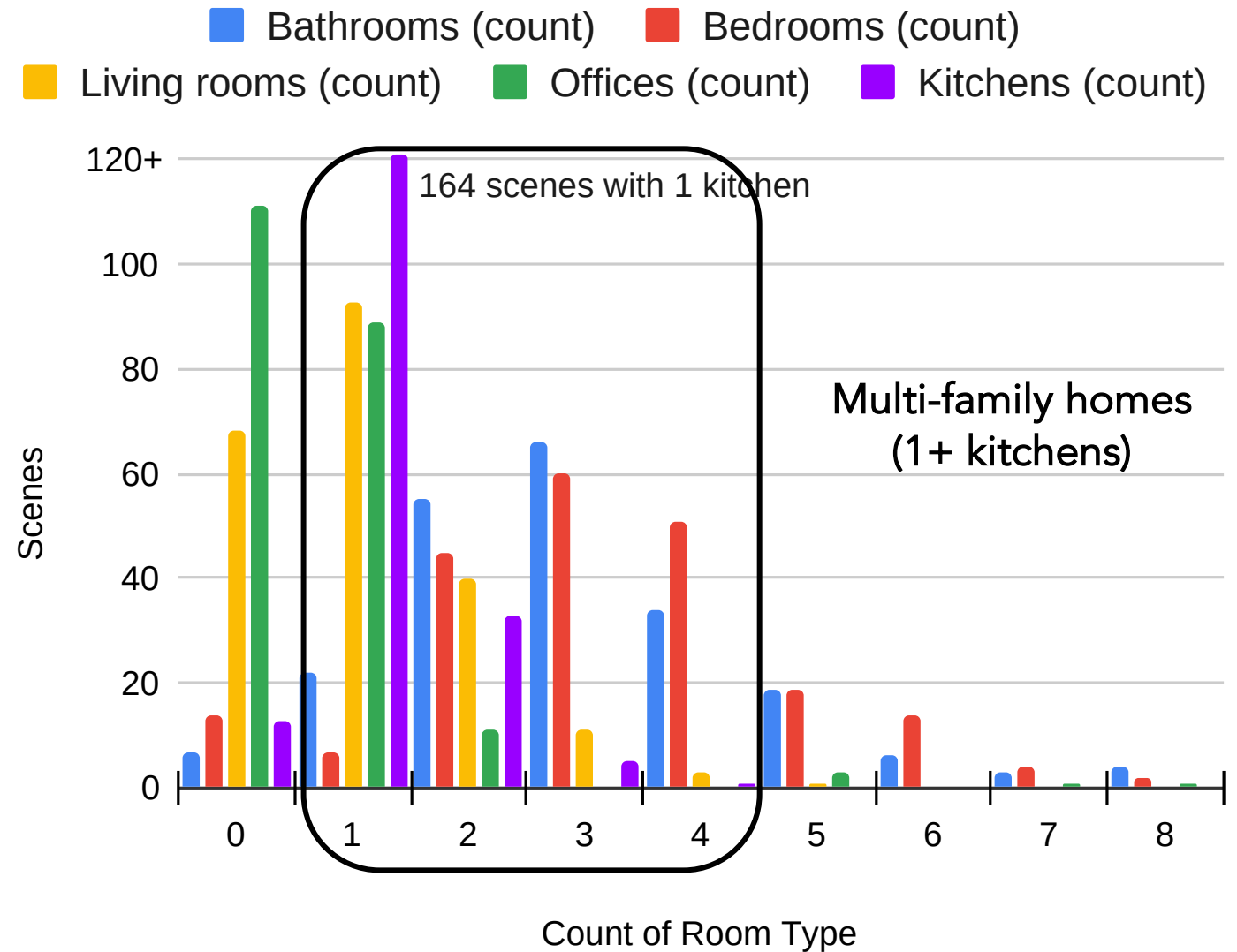
e.g., bed must be in a bedroom, toilet must be in a bathroom, etc.



HM3DSem statistics

Room labels can be derived from the presence of certain objects.

e.g., bed must be in a bedroom, toilet must be in a bathroom, etc.



HM3DSem statistics

Room labels can be derived from the presence of certain objects.

e.g., bed must be in a bedroom, toilet must be in a bathroom

Annotations can be used to intelligently subsample scenes for various applications

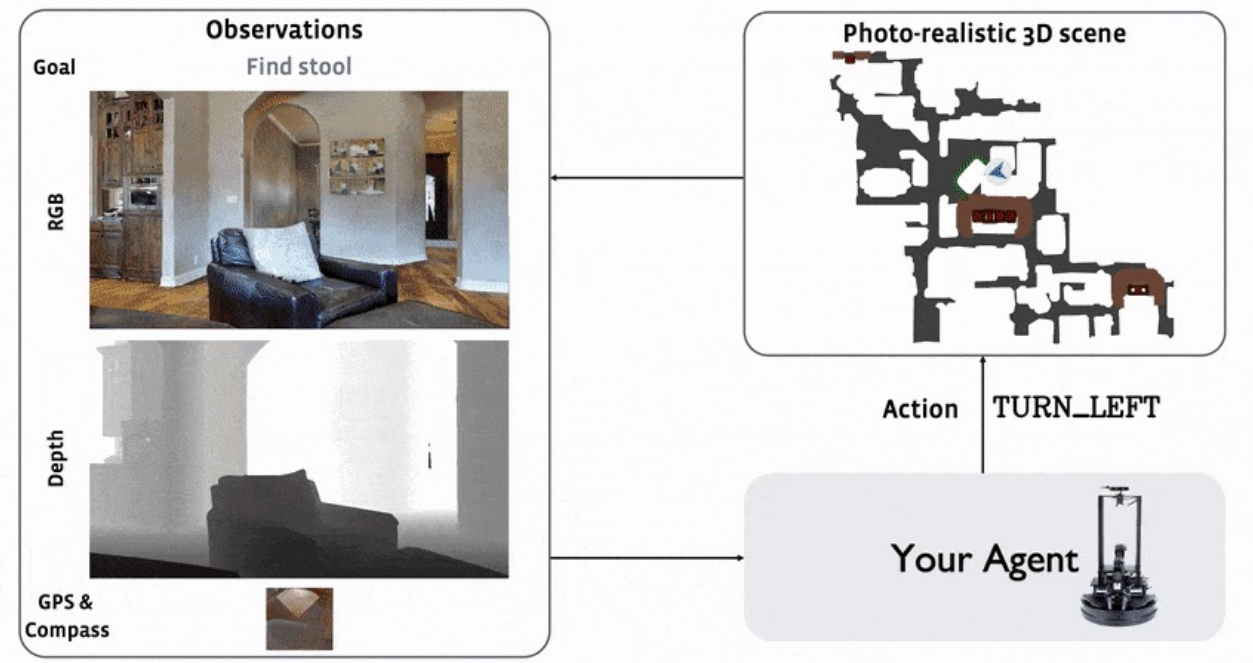
<u>Scene type</u>	<u>Count</u>
Commercial spaces No bedrooms	12
(Museum)	7
	25
Kitchen-free spaces Commercial / hotel rooms	8

Experiments

Task 1: Object detection



Task 2: ObjectGoal Navigation



Task 1: Object detection

Goal: Segment all instances of a pre-defined set of objects



Object categories: [chair, bed, plant, toilet, tv/monitor, sofa]

Task 1: Object detection

Experiment setup

Segmentation dataset

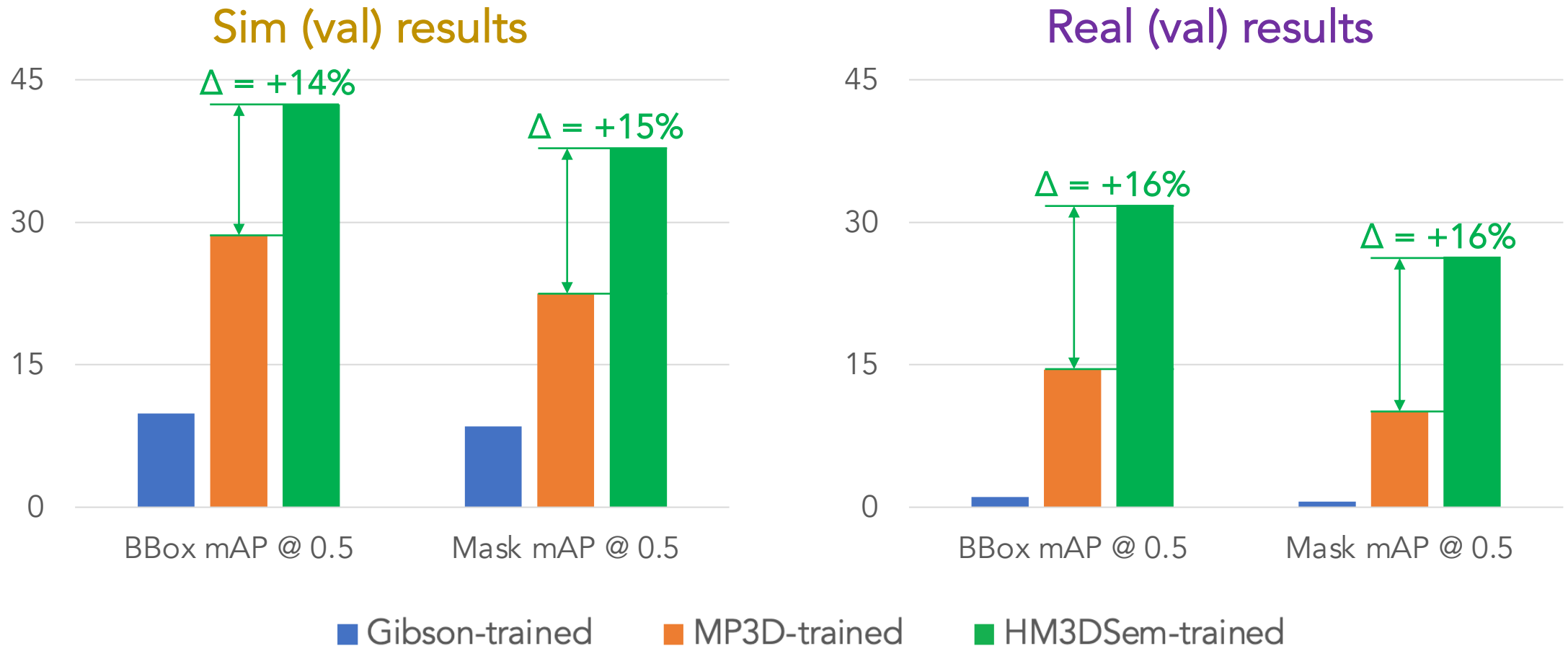


*Train one Mask-RCNN
model per dataset*



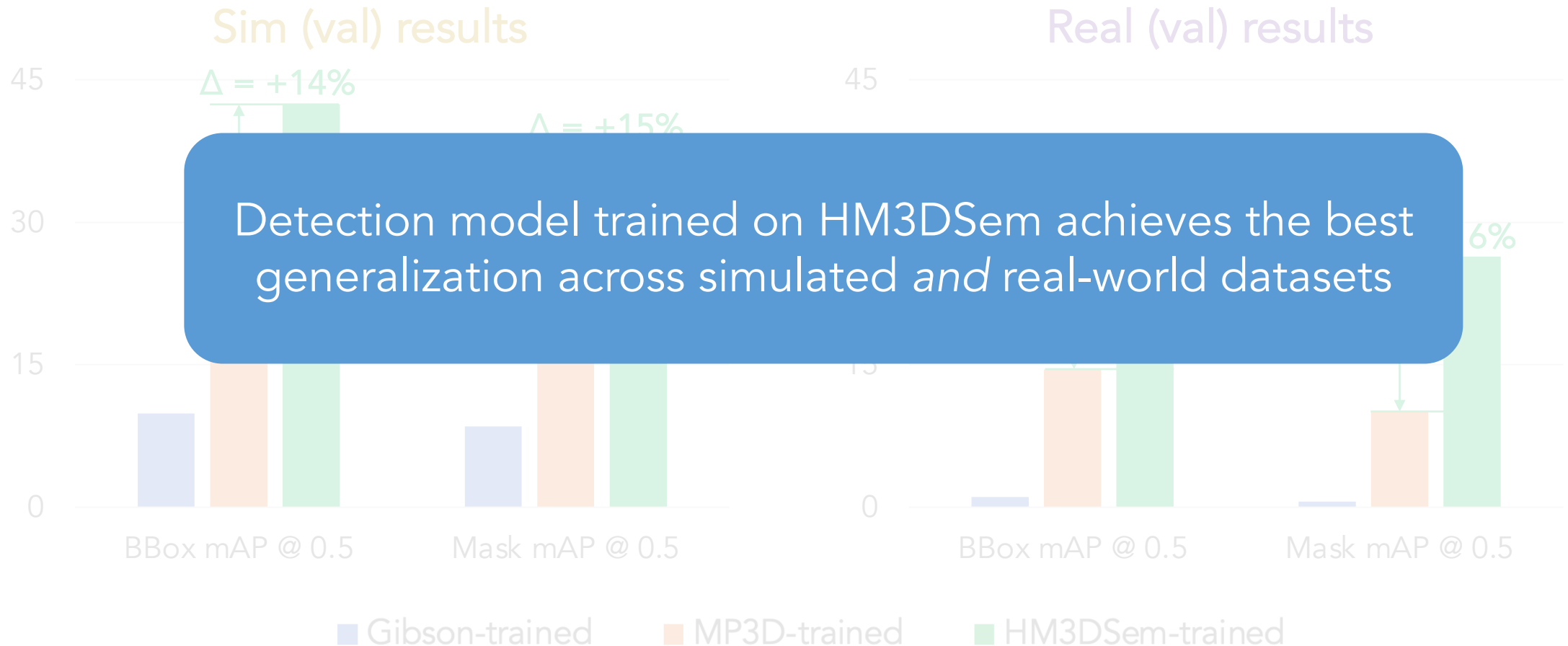
Task 1: Object detection

Training on HM3DSem leads to the best detector



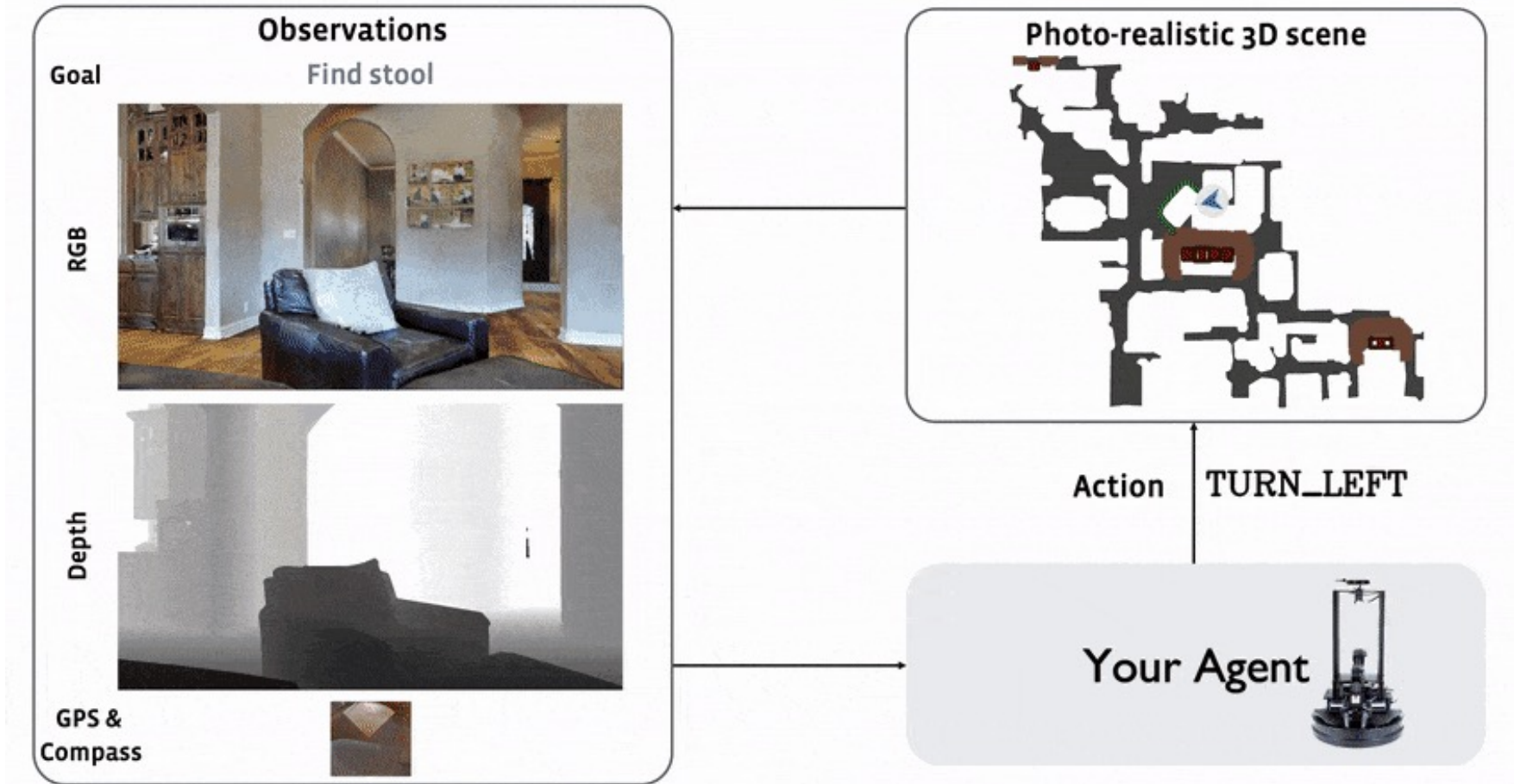
Task 1: Object detection

Training on HM3DSem leads to the best detector



Task 2: ObjectNav

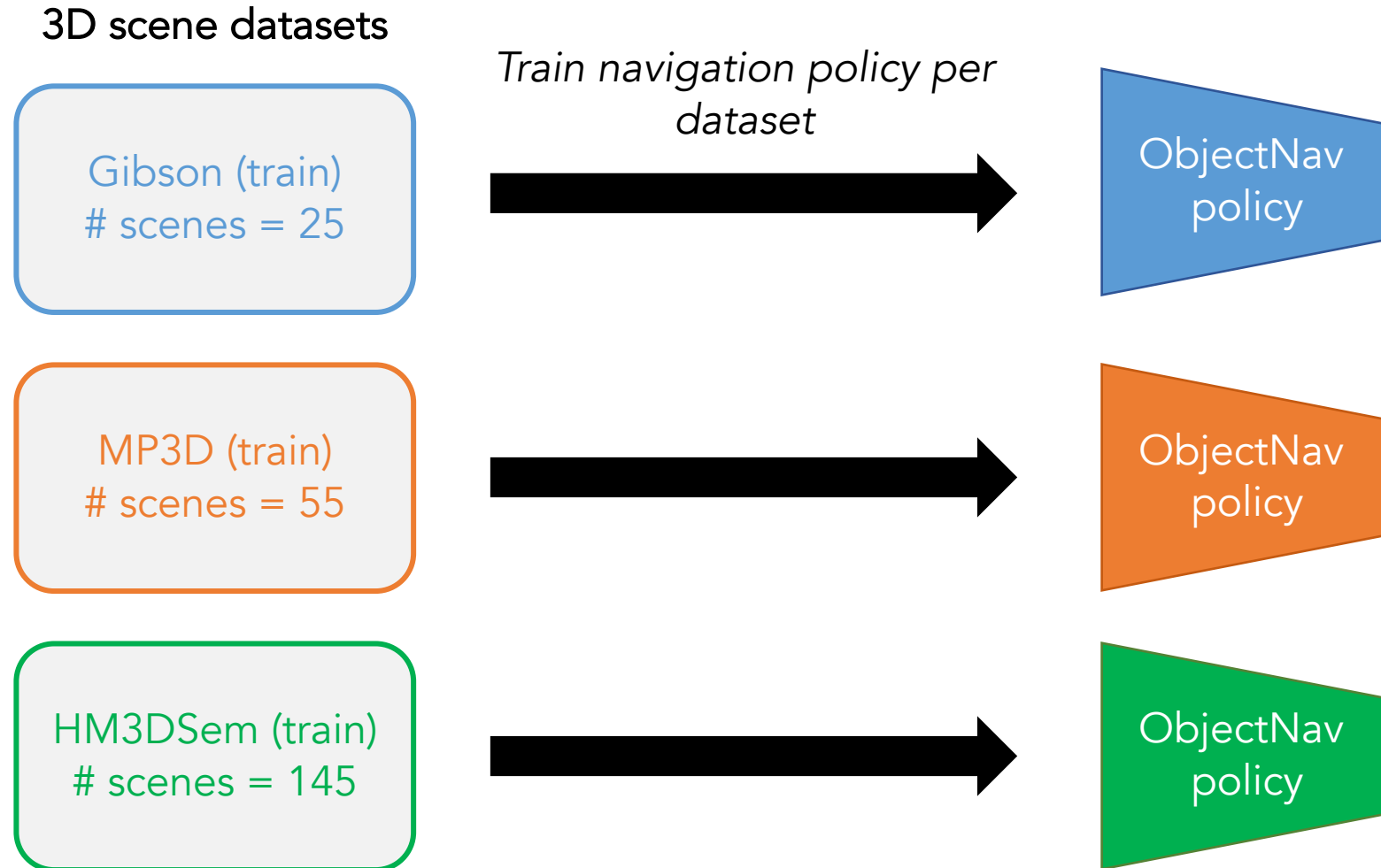
Goal: Find an object in a novel environment.



Goal categories: [chair, bed, plant, toilet, tv/monitor, sofa]

Task 2: ObjectNav

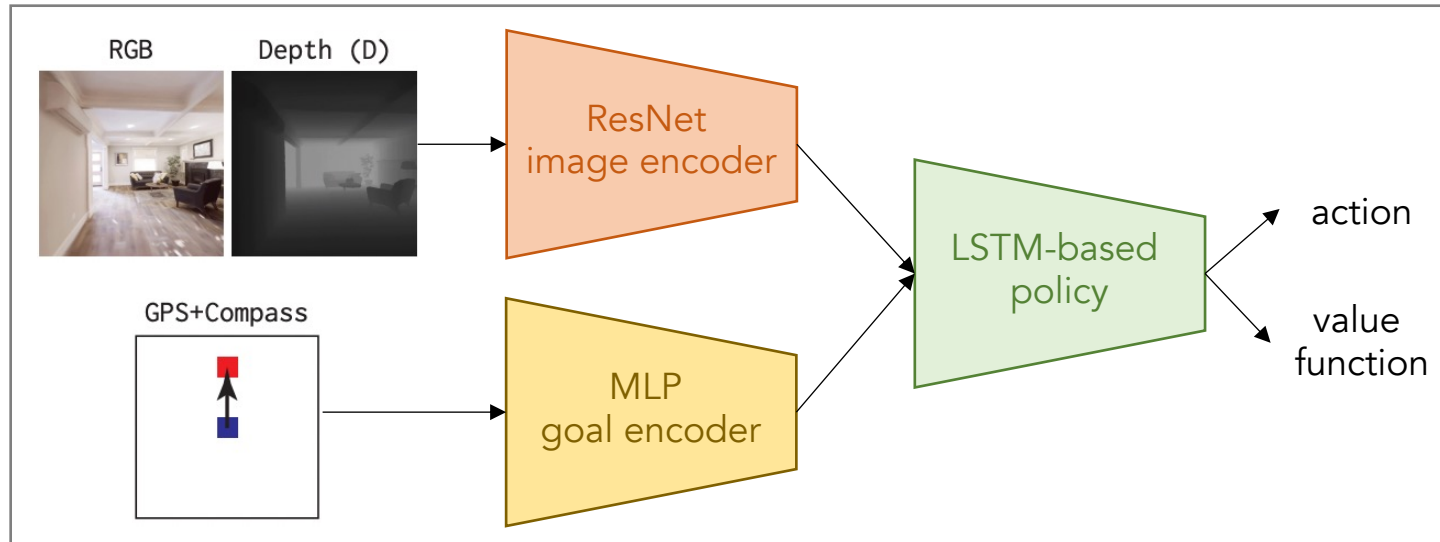
Experiment setup



Task 2: ObjectNav

Baseline policies for performing ObjectNav

ResNet + LSTM policy



Reinforcement Learning

DD-PPO, Wijmans et al., ICLR 2020

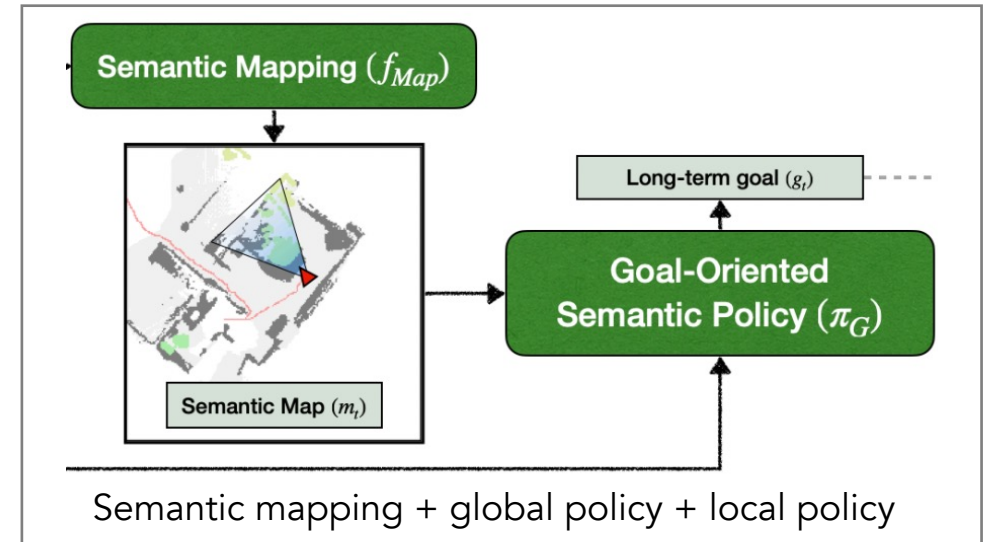
- End-to-end training using RL

Imitation Learning

Habitat-Web, Ramrakhya et al., CVPR 2022

- End-to-end training on human demonstrations

Hierarchical policy



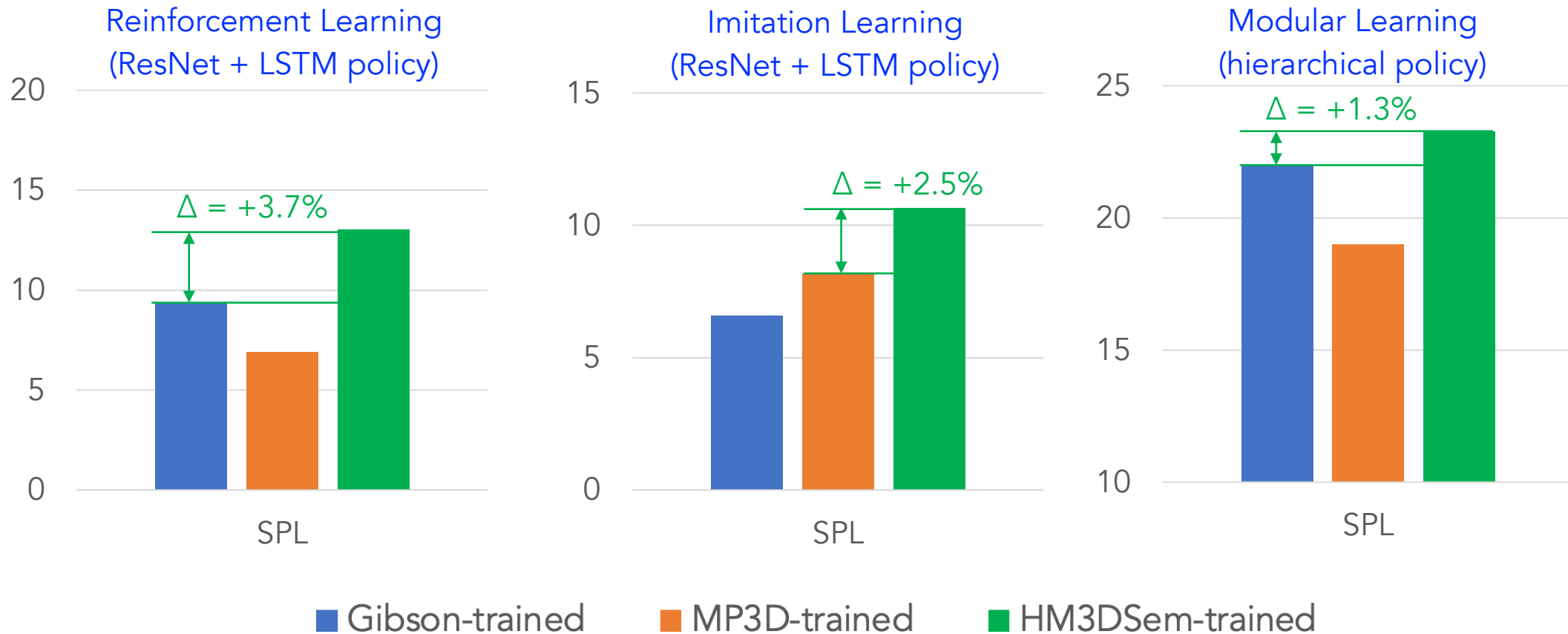
Modular Learning

Semantic Exploration, Chaplot et al., ICLR 2020

- Global policy is trained using RL

Task 2: ObjectNav

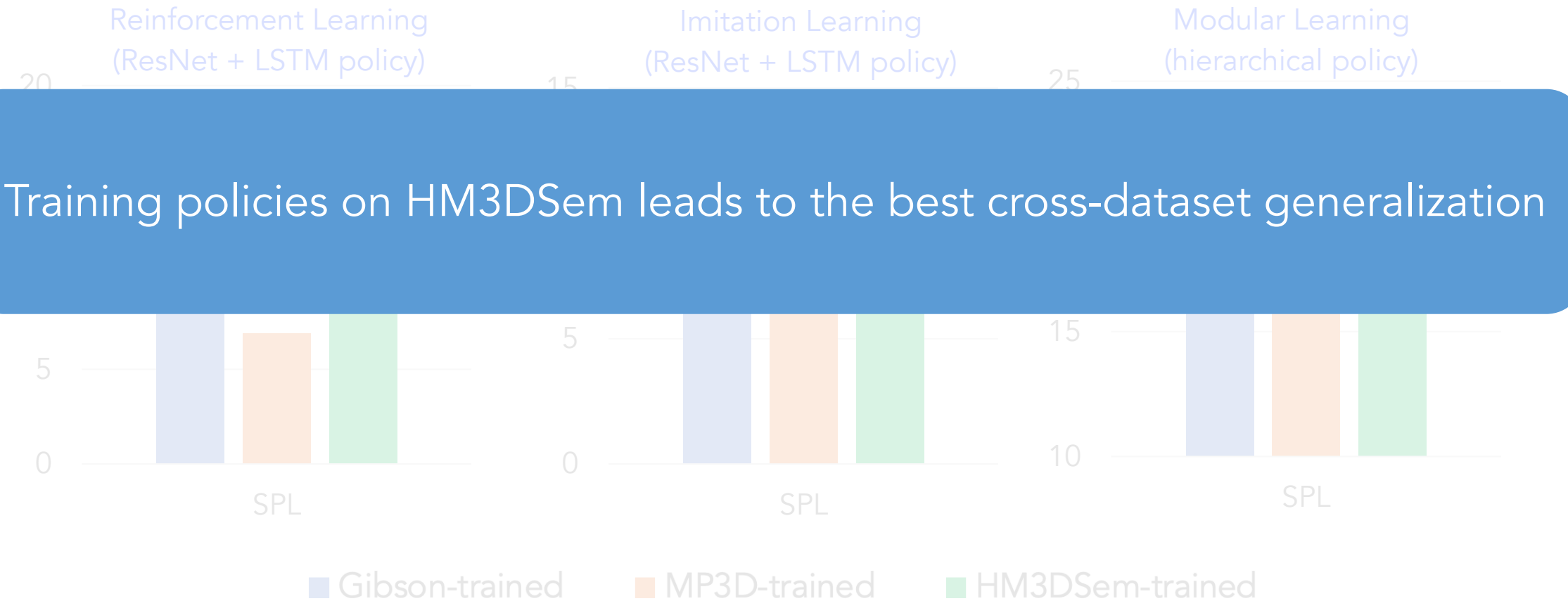
HM3DSem-trained agents generalize well across datasets



Each agent is evaluated on validation scenes of Gibson + MP3D + HM3D

Task 2: ObjectNav

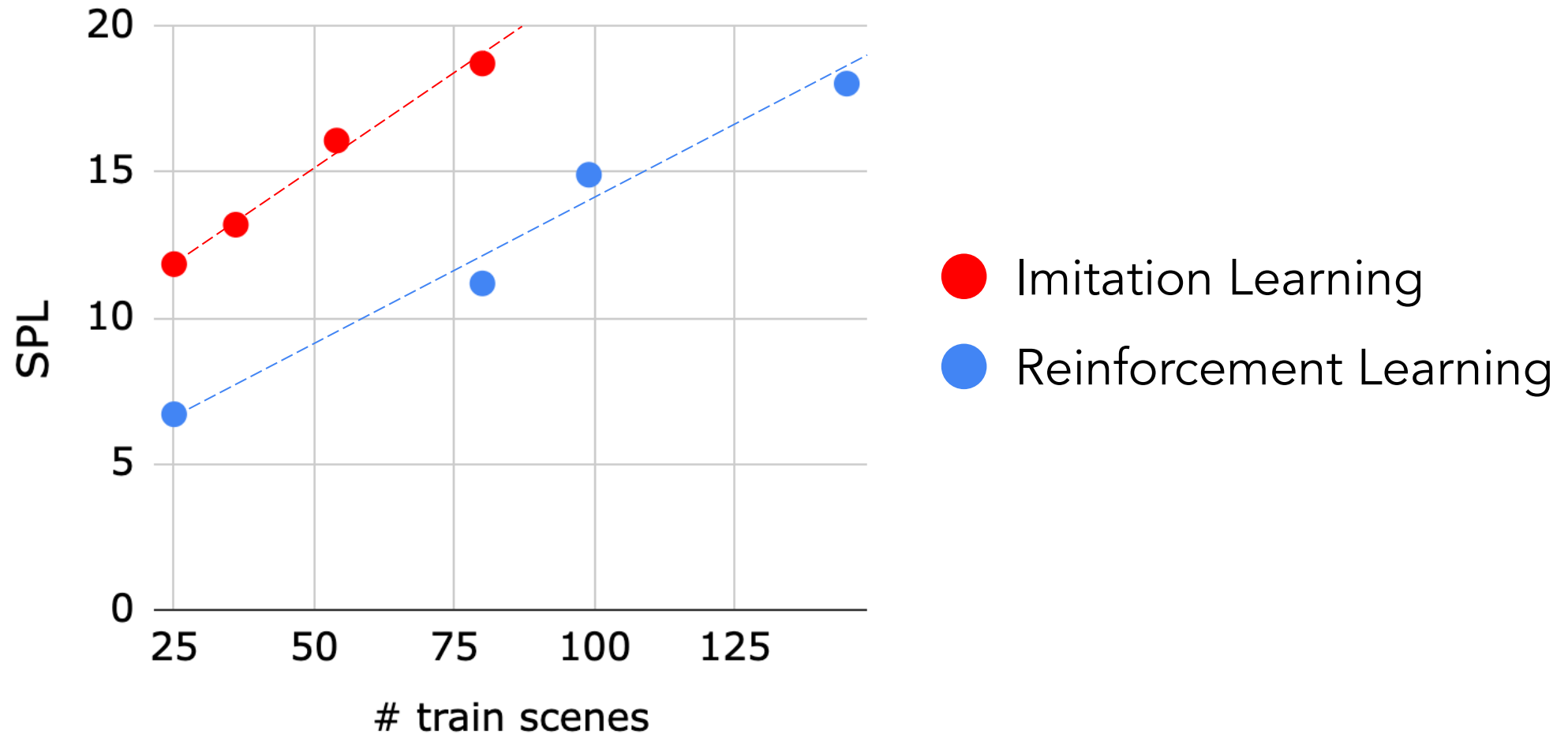
HM3DSem-trained agents generalize well across datasets



Each agent is evaluated on validation scenes of Gibson + MP3D + HM3D

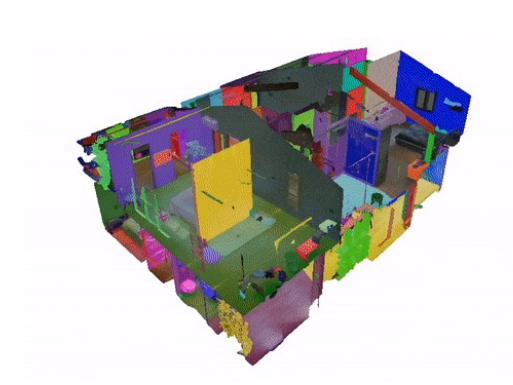
Task 2: ObjectNav

ObjectNav performance scales linearly with dataset size



Habitat Challenge 2022

- Uses HM3DSem v0.1 --- 80 train / 20 val / 20 test scenes
- 6 goal categories
 - chair, couch, plant, bed, toilet, tv
- March 18th, 2022 to August 21st, 2022
- 1022 submissions from 54 teams (150% increase from 2021)



Habitat Challenge 2022

Rank	Participant team	SPL (↑)	SOFT_SPL (↑)	DISTANCE_TO_GOAL (↑)	SUCCESS (↑)
1	ByteBOT	0.37	0.40	2.61	0.68
2	Stretch (Semantic Map + Frontier)	0.34	0.38	3.46	0.60
3	Finding NIMO (PEANUT)	0.33	0.36	3.00	0.64
4	BadSeed (PIRLNav)	0.33	0.37	2.22	0.65
5	Populus A. (0829-1-205600)	0.32	0.37	2.16	0.66
6	GoodSeed (ProcTHOR - Large)	0.32	0.38	2.58	0.54
7	不怕麻煩的兔女郎 (🐰)	0.31	0.35	3.40	0.57
8	SkillFusion (AIRI)	0.29	0.35	3.05	0.55
9	Walle (0825-5-21000)	0.27	0.33	2.78	0.56
10	OVRL (OVRL v1)	0.27	0.31	2.49	0.60
11	RUG-Windmill	0.25	0.29	3.75	0.54
12	Habitat-Web (IL-HD)	0.22	0.26	3.15	0.55
13	EpsGreedy	0.17	0.22	3.90	0.44
14	RLNJU	0.17	0.28	4.64	0.30
15	Host_74441_Team (DDPPO Baseline)	0.12	0.22	4.31	0.26

Challenge winner

161% gain in success!

Baseline

Conclusions

- HM3DSem is the largest dataset of 3D real-world spaces with densely annotated semantics
- Models trained on HM3DSem generalize well across datasets
- Model performance scales linearly with dataset size
- Check out the HM3DSem ObjectNav challenge!

Habitat-Matterport 3D Semantics Dataset



Poster session: TUE-PM-075

Project page: <https://aihabitat.org/datasets/hm3d-semantics/>

