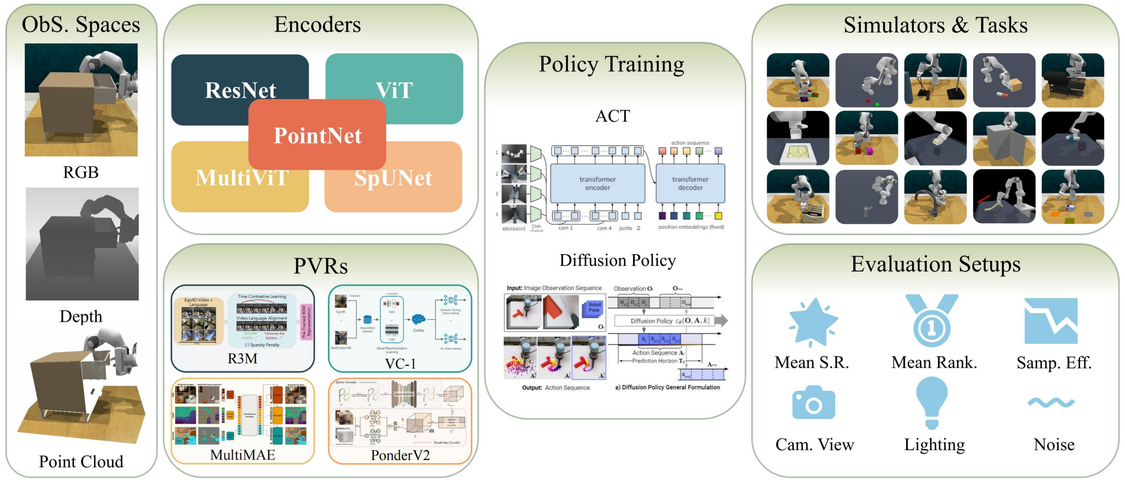
Point Cloud Matters: Rethinking the Impact of Different Observation Spaces on Robot Learning

Haoyi Zhu, Yating Wang, Di Huang, Weicai Ye, Wanli Ouyang, Tong He*

上海人工智能实验室
Shanghai Artificial Intelligence Laboratory

Website Github

OBSBench



We examine the impact of various observation spaces, specifically ***RGB, RGB-D, and point clouds***, on robot learning.

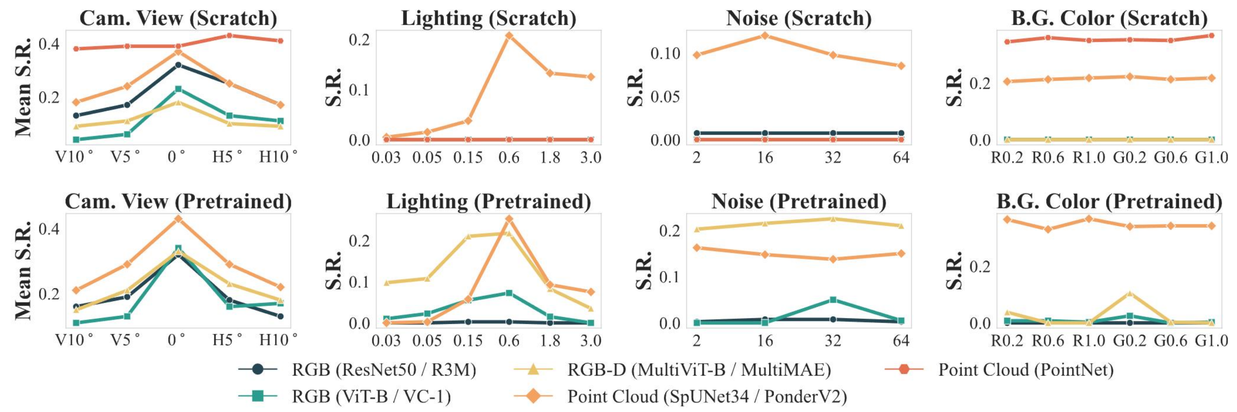
Study on Performance

- ✓ Point cloud methods consistently outperform other modalities.
- ✓ The depth modality generally degrades performance across all settings.
- ✓ Using PVRs can lead to better performance on average, though not for all individual tasks.

Tasks	ACT Policy								
	RGB		RGB-D			Point Cloud		Depth Only	
	● ResNet	● ViT	● ResNet	● ViT	● MultiViT	● SpUNet	● PointNet	● ResNet	● ViT
PickCube	0.60	0.14	0.75	0.03	0.04	0.74	**0.84**	0.05	0.01
StackCube	0.32	0.00	0.17	0.00	0.00	0.22	**0.35**	0.00	0.00
TurnFaucet	**0.49**	0.27	0.00	0.06	0.35	0.39	0.00	0.41	0.00
Peg-Insertion-Side (Grasp)	0.73	0.36	0.73	0.03	0.16	**0.81**	0.77	0.07	0.01
Peg-Insertion-Side (Align)	0.18	0.02	0.06	0.00	0.01	0.28	**0.40**	0.00	0.00
Peg-Insertion-Side (Insert)	**0.01**	0.00	0.00	0.00	0.00	**0.01**	**0.01**	0.00	0.00
Excavate	0.02	0.00	0.02	0.14	0.00	0.03	0.27	**0.29**	0.00
Hang	**0.86**	0.80	0.81	0.00	0.84	0.84	0.83	0.79	0.41
Pour	0.07	0.00	0.01	0.00	0.00	0.10	**0.14**	0.00	0.00
Fill	0.79	0.30	0.60	0.79	0.76	0.66	**0.91**	0.51	0.00
open drawer	0.00	0.16	0.08	0.00	0.20	**0.44**	0.00	-	-
sweep to	0.72	0.80	**1.00**	0.92	0.68	0.90	**1.00**	-	-
meat off grill	0.24	0.16	0.36	0.08	0.00	**0.72**	0.44	-	-
turn tap	0.00	0.00	0.00	0.00	0.00	0.00	**0.04**	-	-
reach and drag	0.32	0.28	**0.60**	**0.60**	0.04	0.20	**0.60**	-	-
put money	0.60	0.76	**0.84**	0.04	0.28	0.60	0.32	-	-
push buttons	0.12	0.40	0.28	0.08	0.14	0.00	**0.52**	-	-
close jar	0.04	0.00	**0.16**	0.00	0.00	0.04	0.00	-	-
place wine	0.00	0.00	0.00	0.00	0.00	0.00	0.00	-	-
Mean S.R. ↑	0.32	0.23	0.34	0.15	0.18	0.37	**0.39**	-	-
Mean Rank ↓	3.05	4.35	3.15	4.75	4.70	2.65	**2.15**	-	-

Study on Zero-Shot Generalization

Point cloud methods are better on ***camera view*** and ***visual changes***.



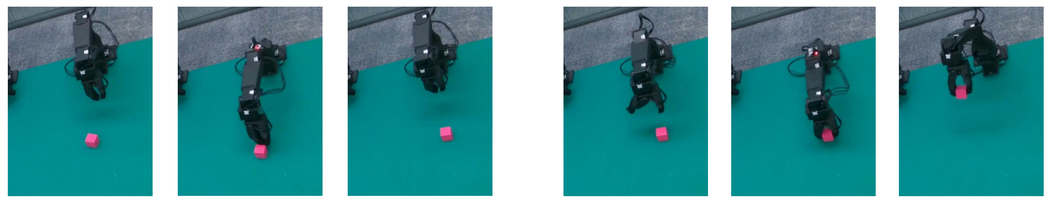
Study on Point Cloud Design Decisions

- ✓ ***Post-sampling*** can significantly enhance the performance.
- ✓ Coordinate > color information, but using both is the best.
- ✓ Alternatives like pointmap still lags behind point clouds.
- ✓ ***EE frame*** can enhance point cloud performance in many cases.

Study on Real-World Experiments

Real-world results align with our simulated experiments.

Task	Reach Cube	Pick Cube	Fold Cloth
RGB	0.60	0.05	0.65
RGB-D	0.30	0.20	0.50
Point Cloud	**0.80**	**0.40**	**0.80**



(a) Reach Cube (b) Pick Cube

(c) Fold Cloth