

Evaluating InteriorSim

**Mike Roberts
Intel Labs**

Simulator	Number of scenes	Photorealistic rendering	Realistic physics
iGibson 2.0	15 home-sized scenes (108 rooms) 12000 additional scenes from CubiCasa5K and 3D-FRONT	No	Yes
Habitat 2.0	6 scenes from ReplicaCAD 105 scenes after remixing and randomizing	No	Yes (except gripper)
ThreeDWorld	15 houses	Yes	Yes
SAPIEN	Unstated	Yes	Yes
AI2-THOR	120 rooms	No	Yes (except gripper)
InteriorSim (ours)	300 scenes	Yes	Yes

Existing simulators already provide large variety of scenes, realistic rendering, and realistic physics
 We can't motivate InteriorSim based on an enhanced set of features
 We need to empirically compare to existing simulators

Empirical results that would motivate InteriorSim

Empirical results that would motivate InteriorSim

"Navigation policies trained in InteriorSim transfer to the real world better than policies trained in other simulators"

Empirical results that would motivate InteriorSim

“Navigation policies trained in InteriorSim transfer to the real world better than policies trained in other simulators”

Train navigation policies in InteriorSim and other simulators

Deploy policies in the real world

Show that InteriorSim policies perform better than other simulators

Empirical results that would motivate InteriorSim

“Navigation policies trained in InteriorSim transfer to the real world better than policies trained in other simulators”

Train navigation policies in InteriorSim and other simulators

Deploy policies in the real world

Show that InteriorSim policies perform better than other simulators

“InteriorSim is more predictive of real-world navigation performance than other simulators”

Empirical results that would motivate InteriorSim

“Navigation policies trained in InteriorSim transfer to the real world better than policies trained in other simulators”

- Train navigation policies in InteriorSim and other simulators
- Deploy policies in the real world
- Show that InteriorSim policies perform better than other simulators

“InteriorSim is more predictive of real-world navigation performance than other simulators”

- Train a variety of navigation policies in any simulator
- Deploy policies in the real world
- Deploy policies in InteriorSim and other simulators
- Show that InteriorSim performance is more closely correlated with real-world performance

Empirical results that would motivate InteriorSim

“Navigation policies trained in InteriorSim transfer to the real world better than policies trained in other simulators”

- Train navigation policies in InteriorSim and other simulators
- Deploy policies in the real world
- Show that InteriorSim policies perform better than other simulators

“InteriorSim is more predictive of real-world navigation performance than other simulators”

- Train a variety of navigation policies in any simulator
- Deploy policies in the real world
- Deploy policies in InteriorSim and other simulators
- Show that InteriorSim performance is more closely correlated with real-world performance

Empirical results that would motivate InteriorSim

“Navigation policies trained in InteriorSim transfer to the real world better than policies trained in other simulators”

Train navigation policies in InteriorSim and other simulators

Deploy policies in the real world

Show that InteriorSim policies perform better than other simulators

“InteriorSim is more predictive of real-world navigation performance than other simulators”

Train a variety of navigation policies in any simulator*

Deploy policies in the real world*

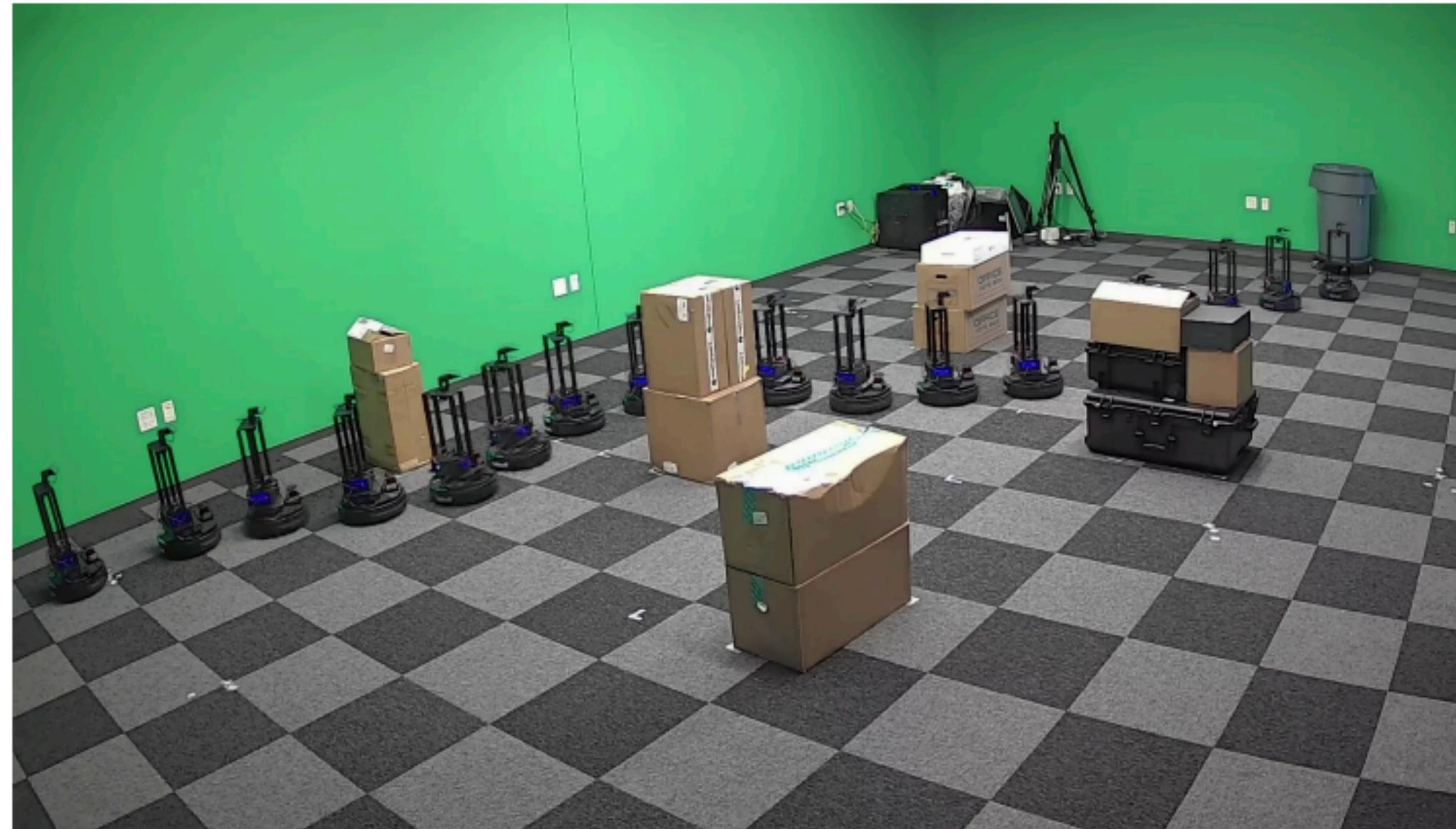
Deploy policies in InteriorSim and other simulators

Show that InteriorSim performance is more closely correlated with real-world performance

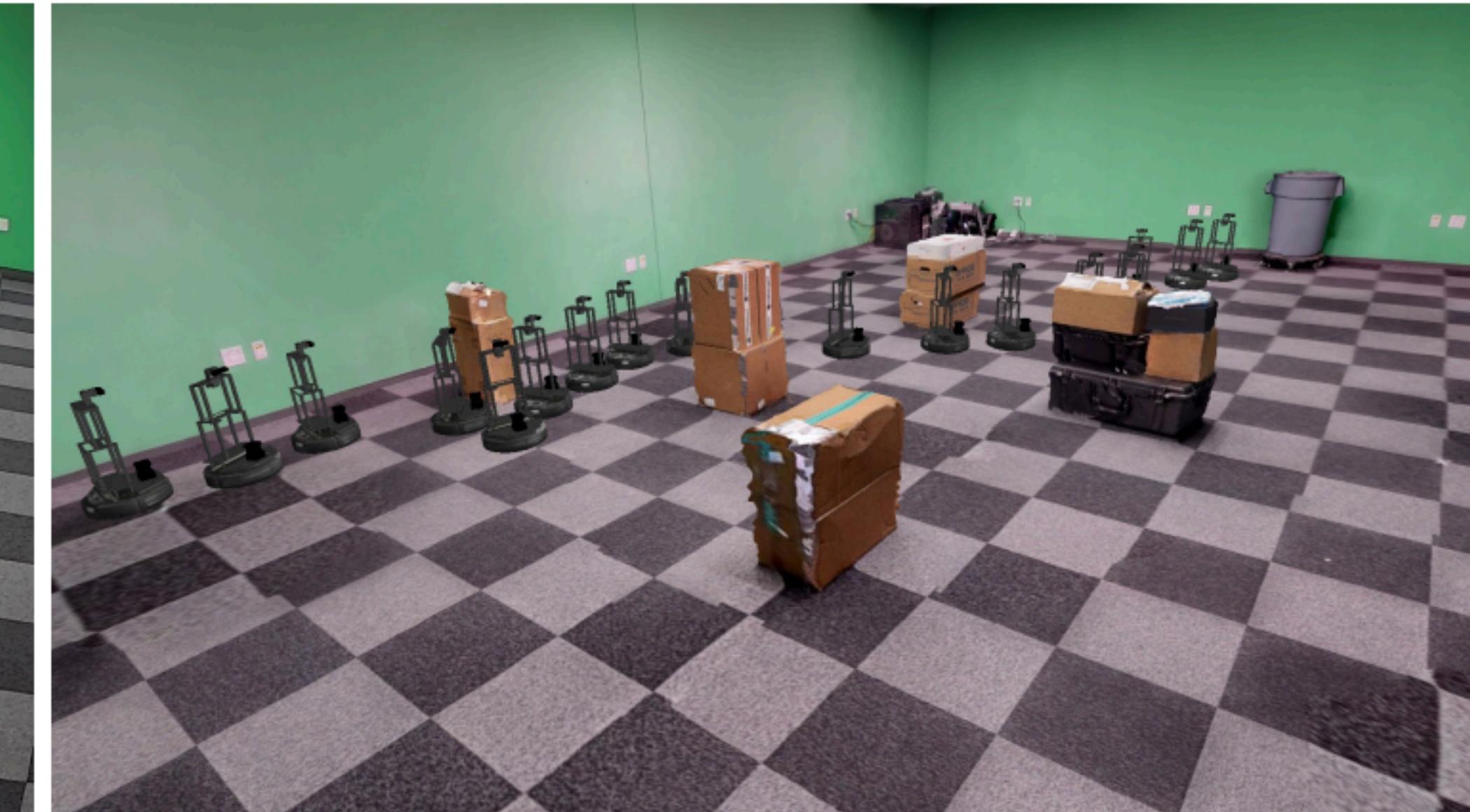
* Data available from previous work!

Sim2Real Predictivity: Does Evaluation in Simulation Predict Real-World Performance?

Abhishek Kadian^{1*}, Joanne Truong^{2*}, Aaron Gokaslan¹, Alexander Clegg¹, Erik Wijmans^{1,2}
Stefan Lee³, Manolis Savva^{1,4}, Sonia Chernova^{1,2} Dhruv Batra^{1,2}



Reality



Simulation

Sim2Real Predictivity: Does Evaluation in Simulation Predict Real-World Performance?

Abhishek Kadian^{1*}, Joanne Truong^{2*}, Aaron Gokaslan¹, Alexander Clegg¹, Erik Wijmans^{1,2}
Stefan Lee³, Manolis Savva^{1,4}, Sonia Chernova^{1,2} Dhruv Batra^{1,2}

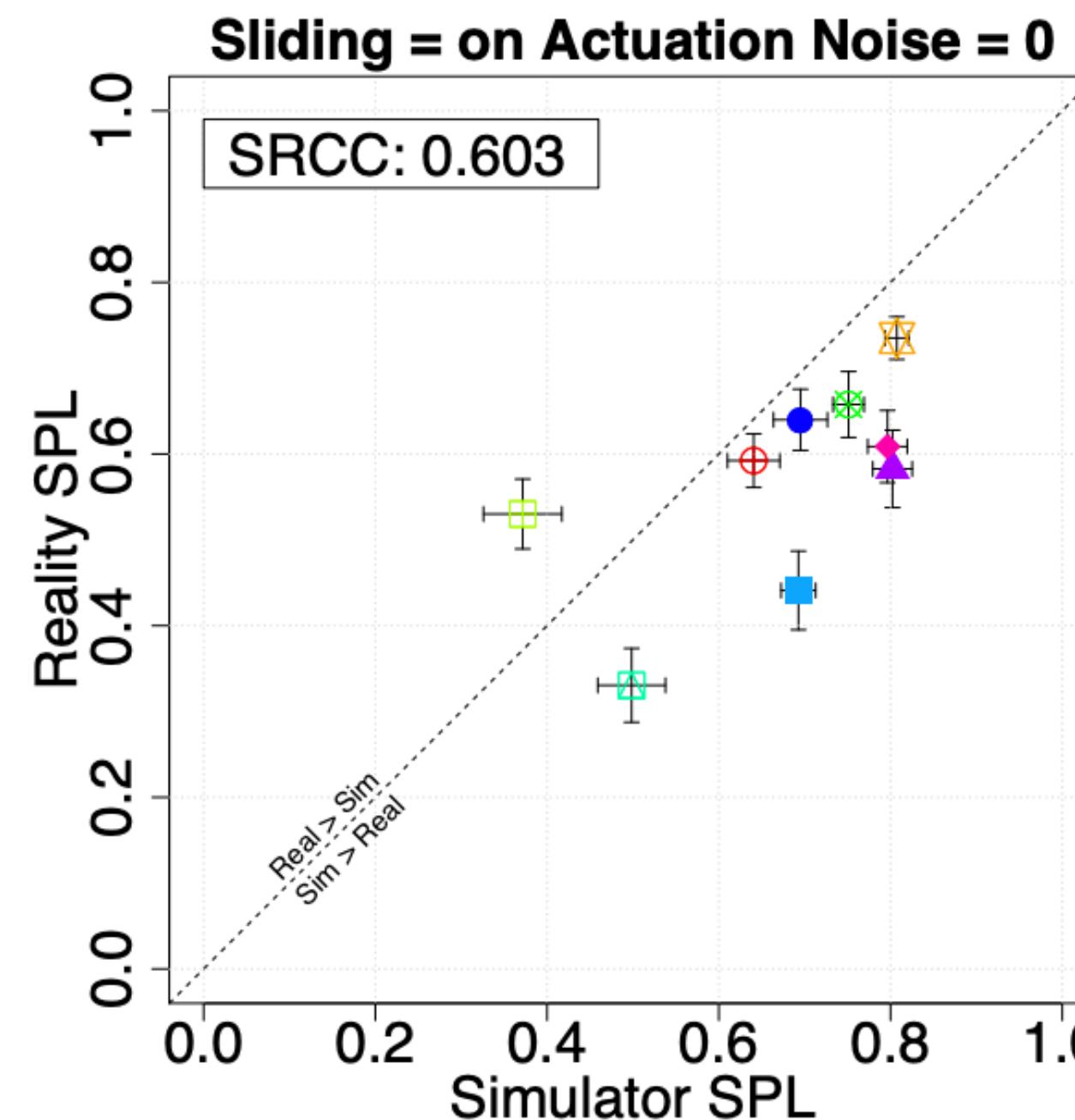


9 pre-trained navigation policies (trained in Habitat with various simulator settings)

All policies deployed in a real-world navigation task (real-world green room scene, with a digital twin, not observed during training)

Sim2Real Predictivity: Does Evaluation in Simulation Predict Real-World Performance?

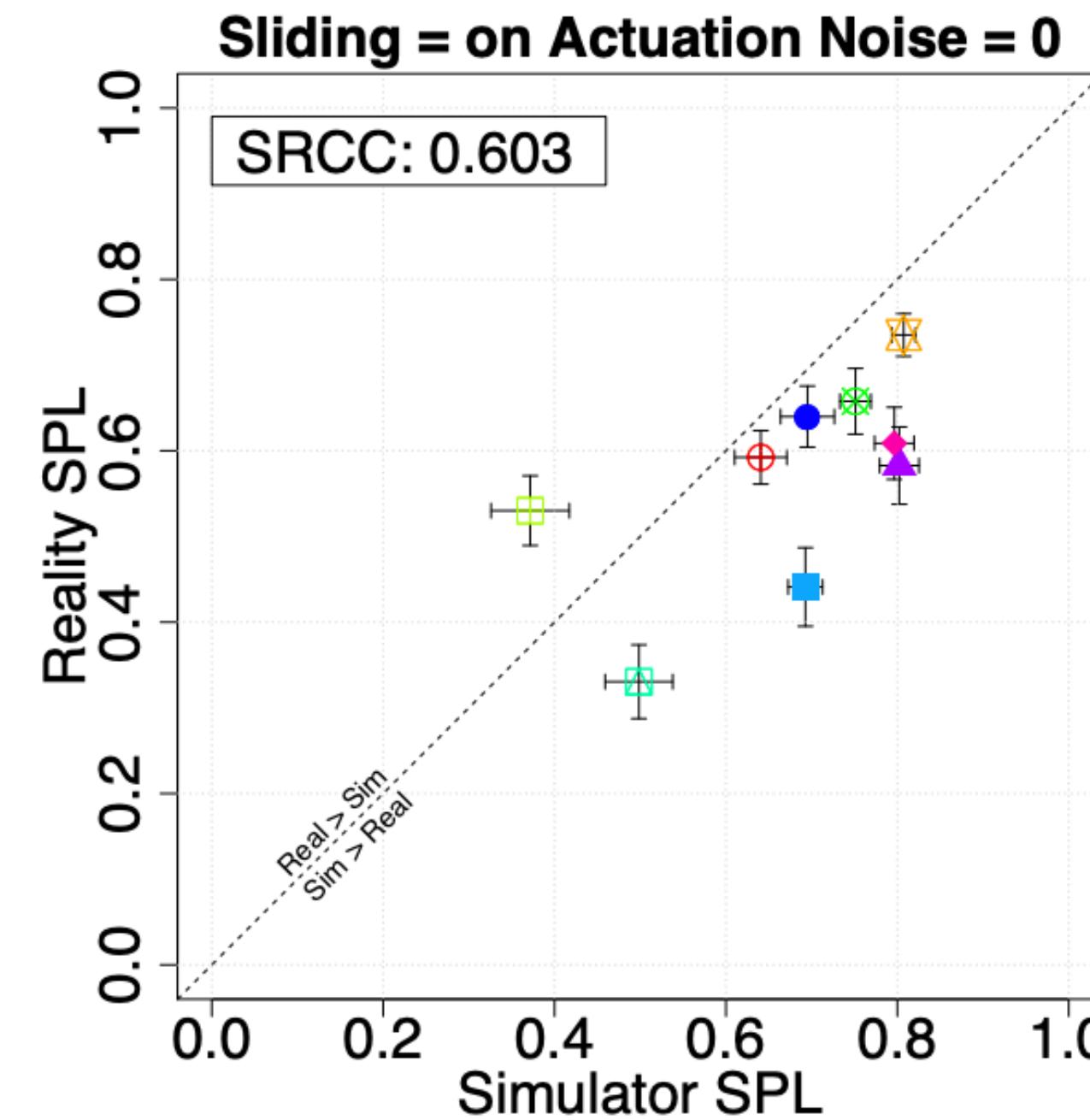
Abhishek Kadian^{1*}, Joanne Truong^{2*}, Aaron Gokaslan¹, Alexander Clegg¹, Erik Wijmans^{1,2}
Stefan Lee³, Manolis Savva^{1,4}, Sonia Chernova^{1,2} Dhruv Batra^{1,2}



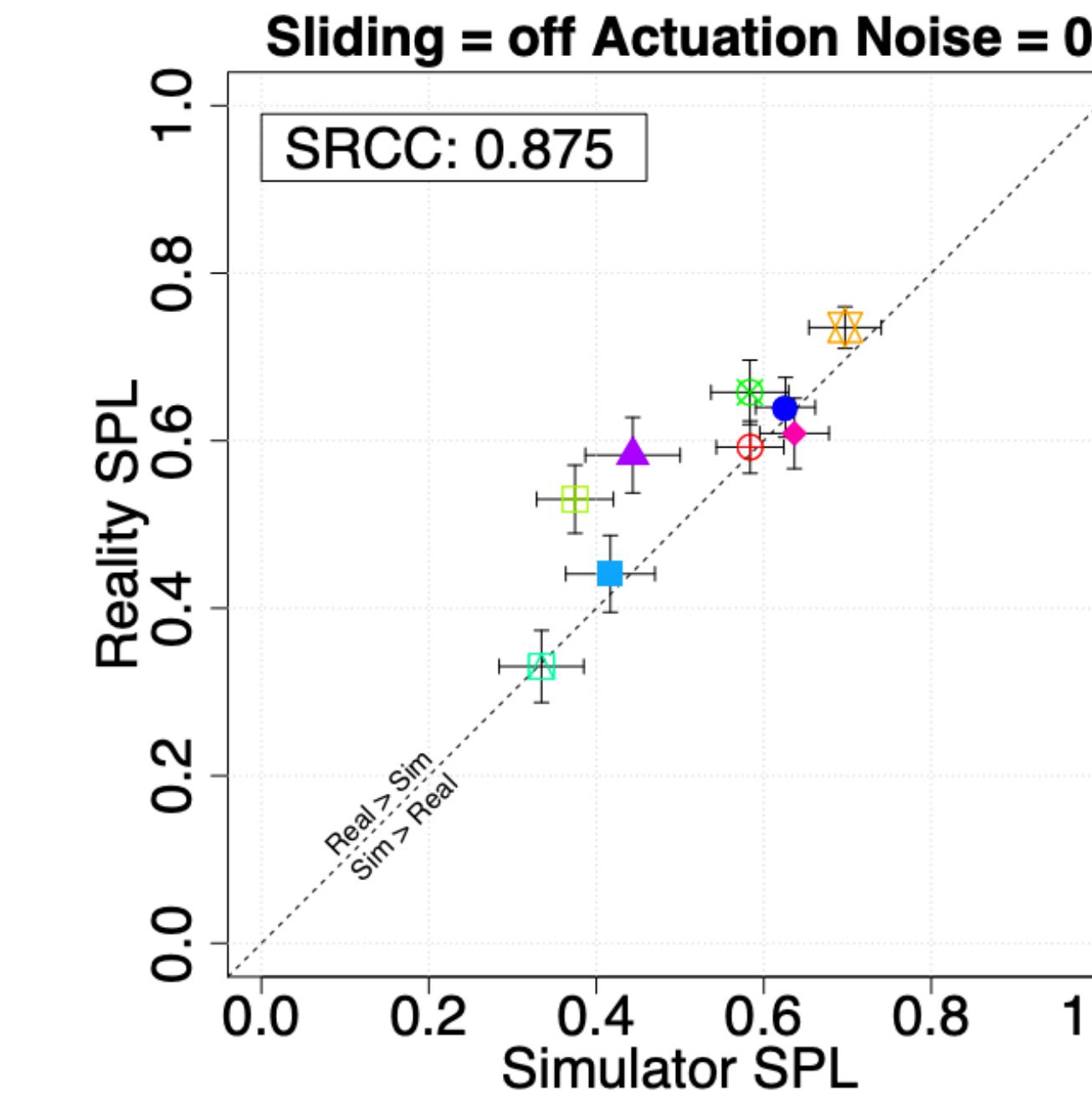
Habitat Challenge 2019 simulator settings
green room (real) vs green room (sim)
Low correlation (0.603)

Sim2Real Predictivity: Does Evaluation in Simulation Predict Real-World Performance?

Abhishek Kadian^{1*}, Joanne Truong^{2*}, Aaron Gokaslan¹, Alexander Clegg¹, Erik Wijmans^{1,2}
Stefan Lee³, Manolis Savva^{1,4}, Sonia Chernova^{1,2} Dhruv Batra^{1,2}



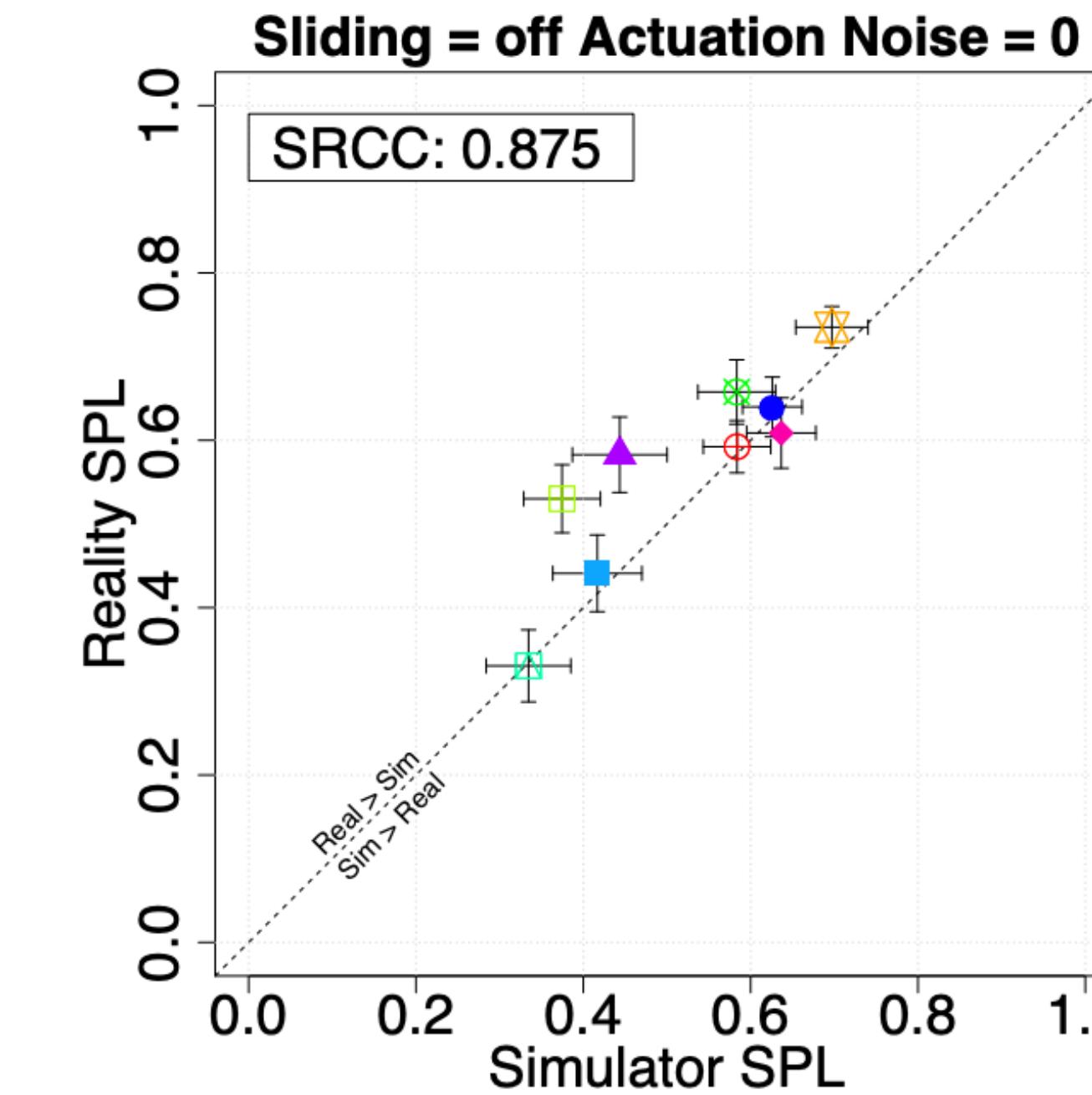
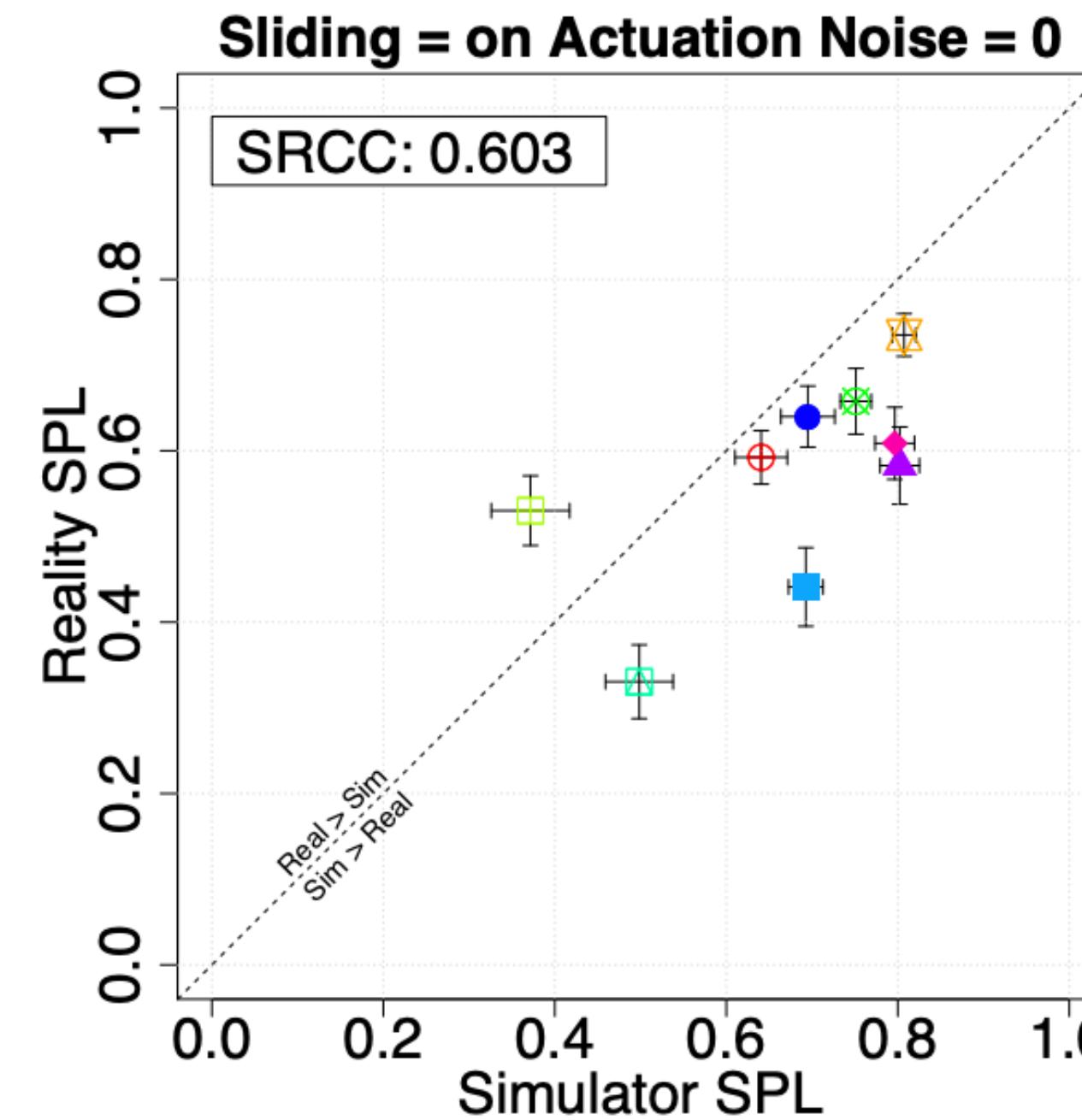
Habitat Challenge 2019 simulator settings
green room (real) vs green room (sim)
Low correlation (0.603)



Optimized simulator settings
green room (real) vs green room (sim)
High correlation (0.875)

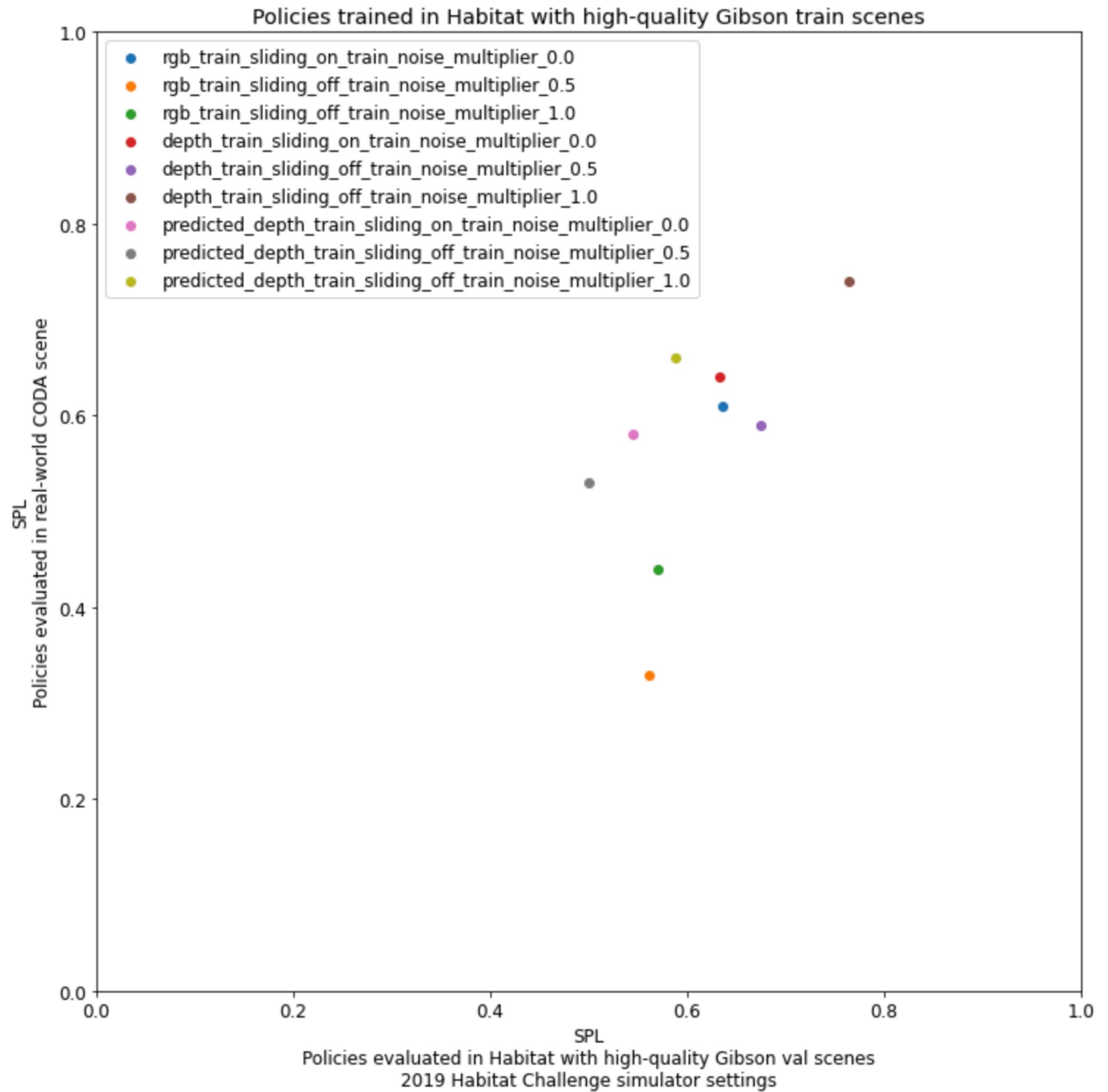
Sim2Real Predictivity: Does Evaluation in Simulation Predict Real-World Performance?

Abhishek Kadian^{1*}, Joanne Truong^{2*}, Aaron Gokaslan¹, Alexander Clegg¹, Erik Wijmans^{1,2}
Stefan Lee³, Manolis Savva^{1,4}, Sonia Chernova^{1,2} Dhruv Batra^{1,2}



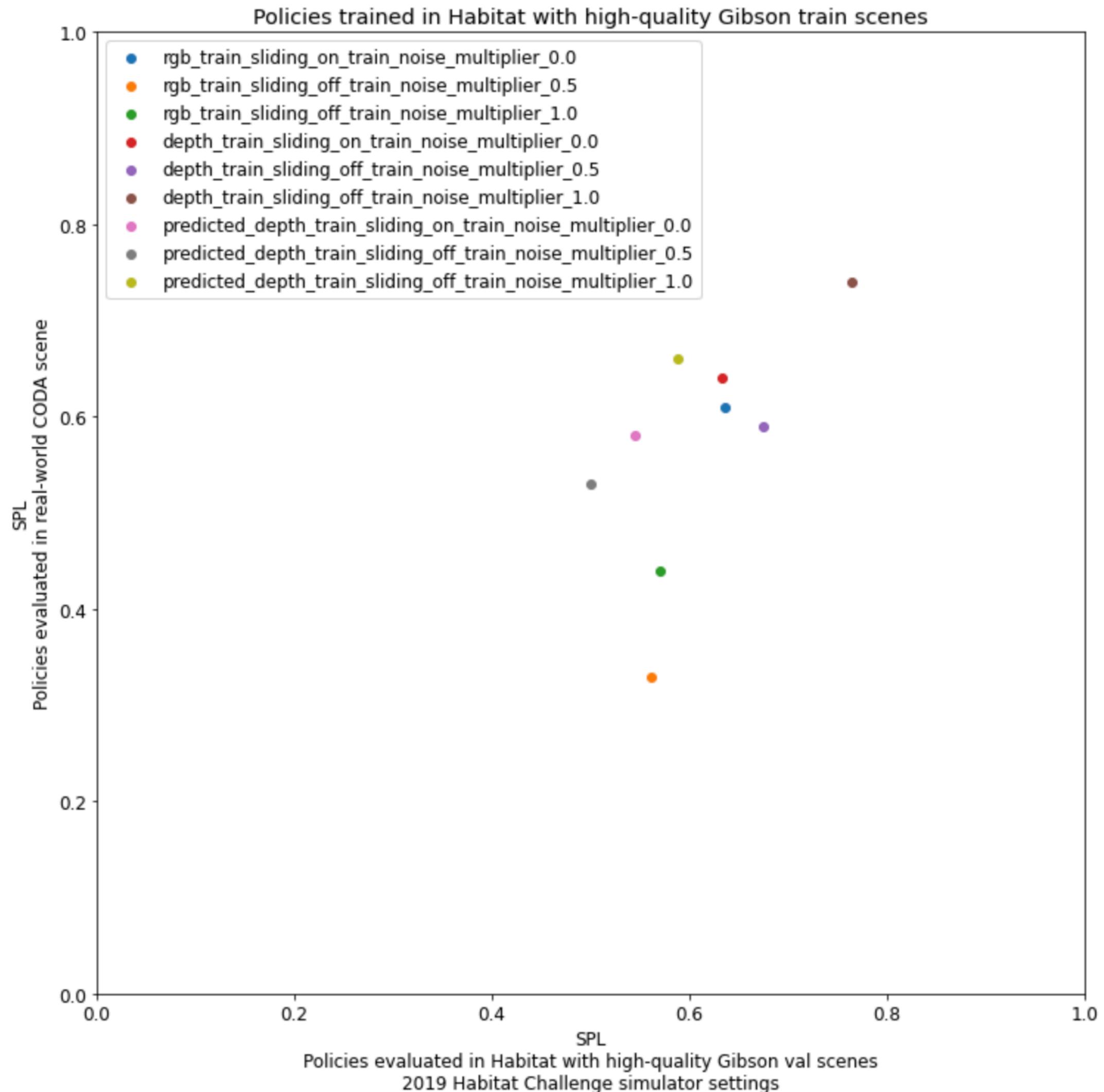
Can we compute correlation with **unpaired** real and sim scenes?

$r = 0.6314968310270521$



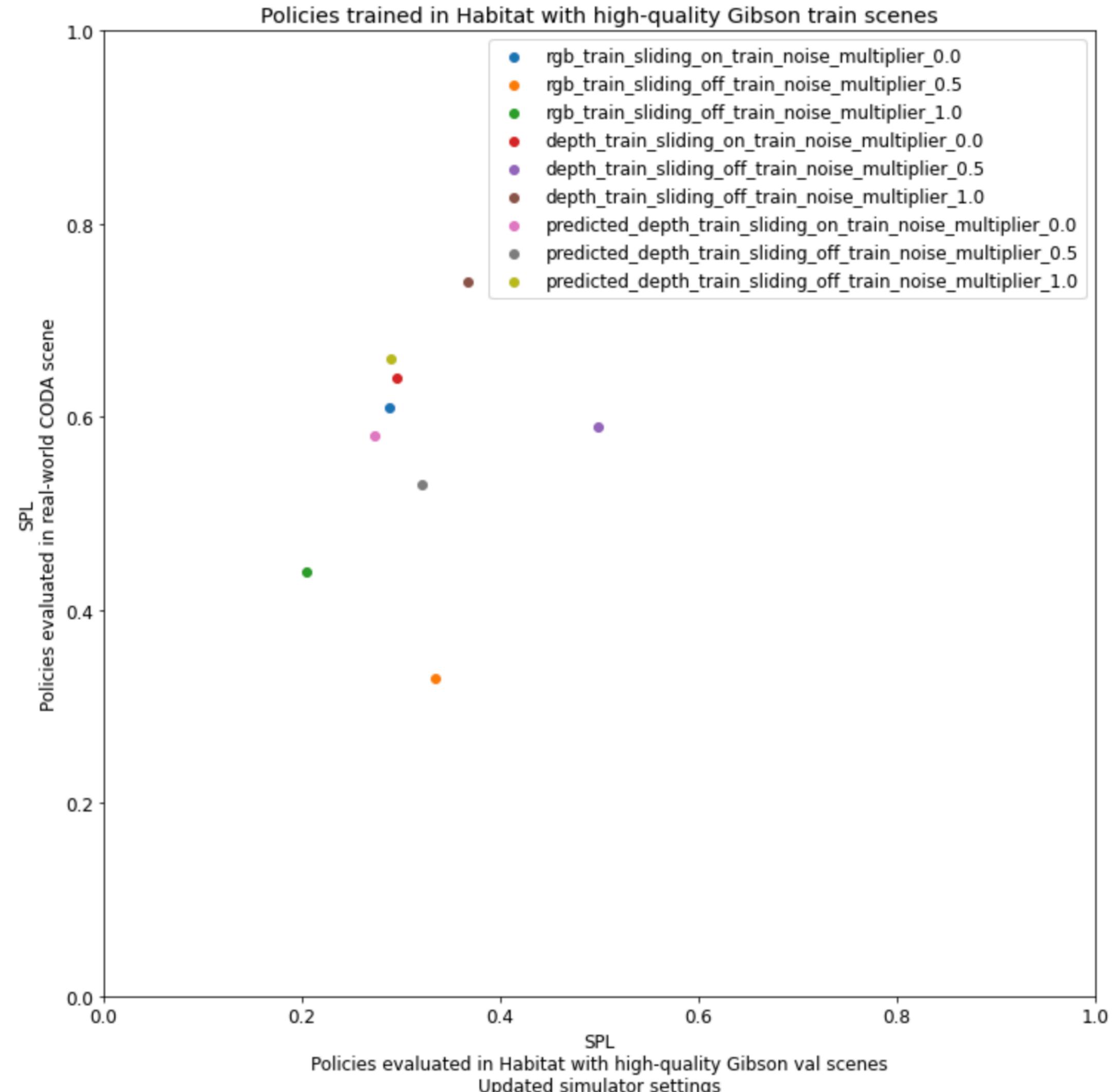
Habitat Challenge 2019 simulator settings
green room (real) vs Gibson val (sim)
Low correlation (0.631)

$r = 0.6314968310270521$



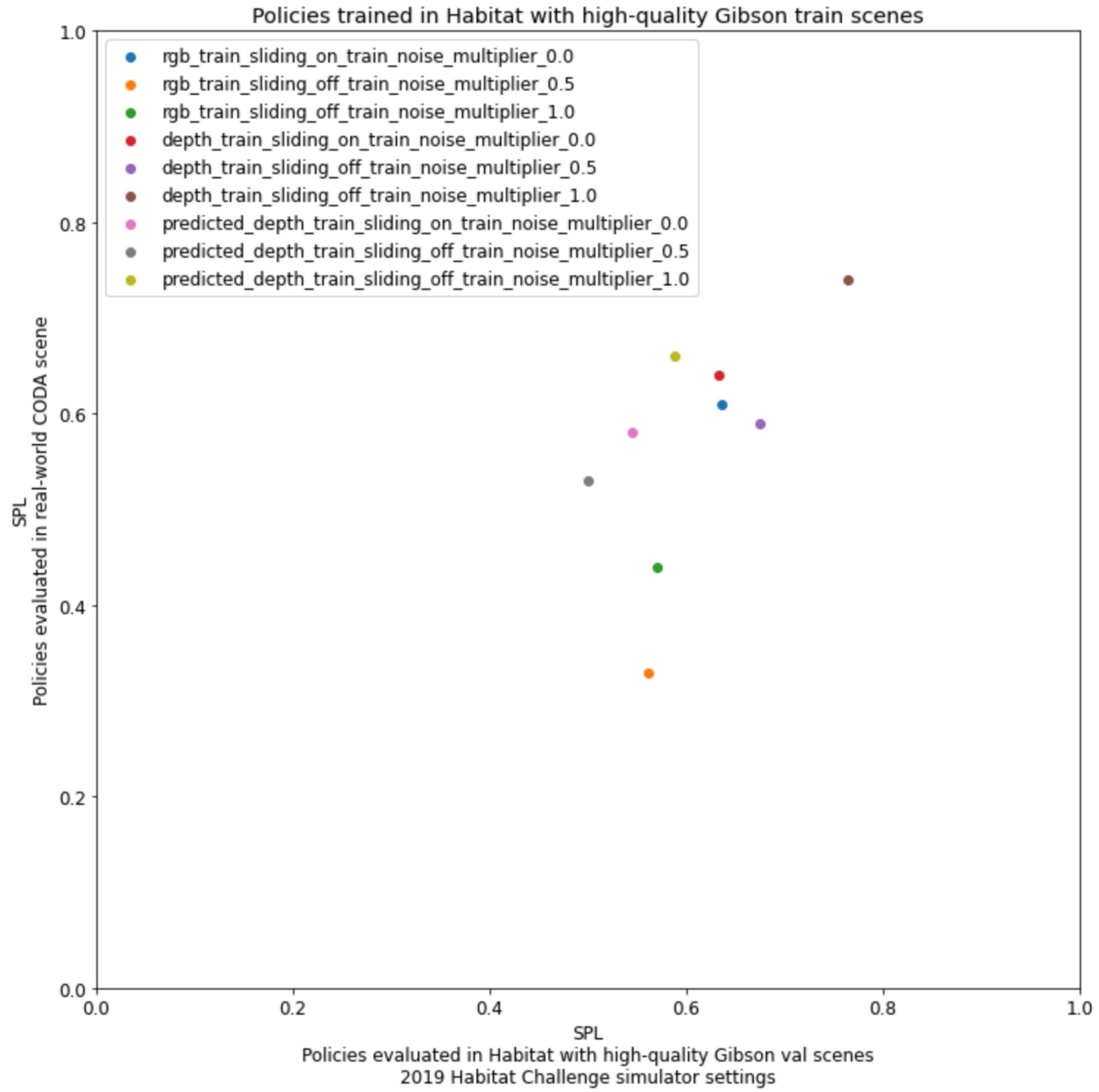
Habitat Challenge 2019 simulator settings
green room (real) vs Gibson val (sim)
Low correlation (0.631)

$r = 0.21552588589707233$

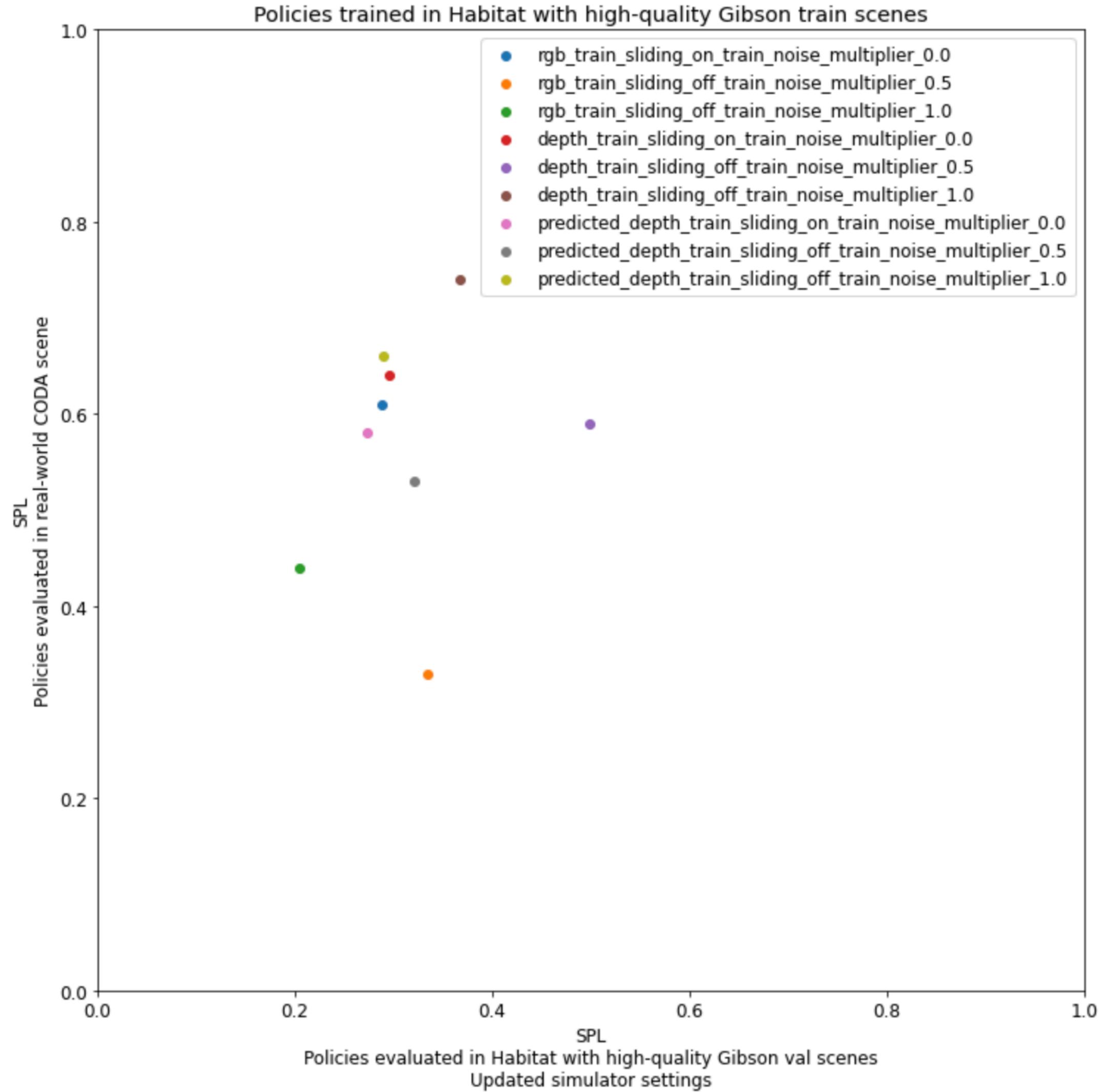


Optimized simulator settings
green room (real) vs Gibson val (sim)
Even lower correlation (0.216)

$r = 0.6314968310270521$

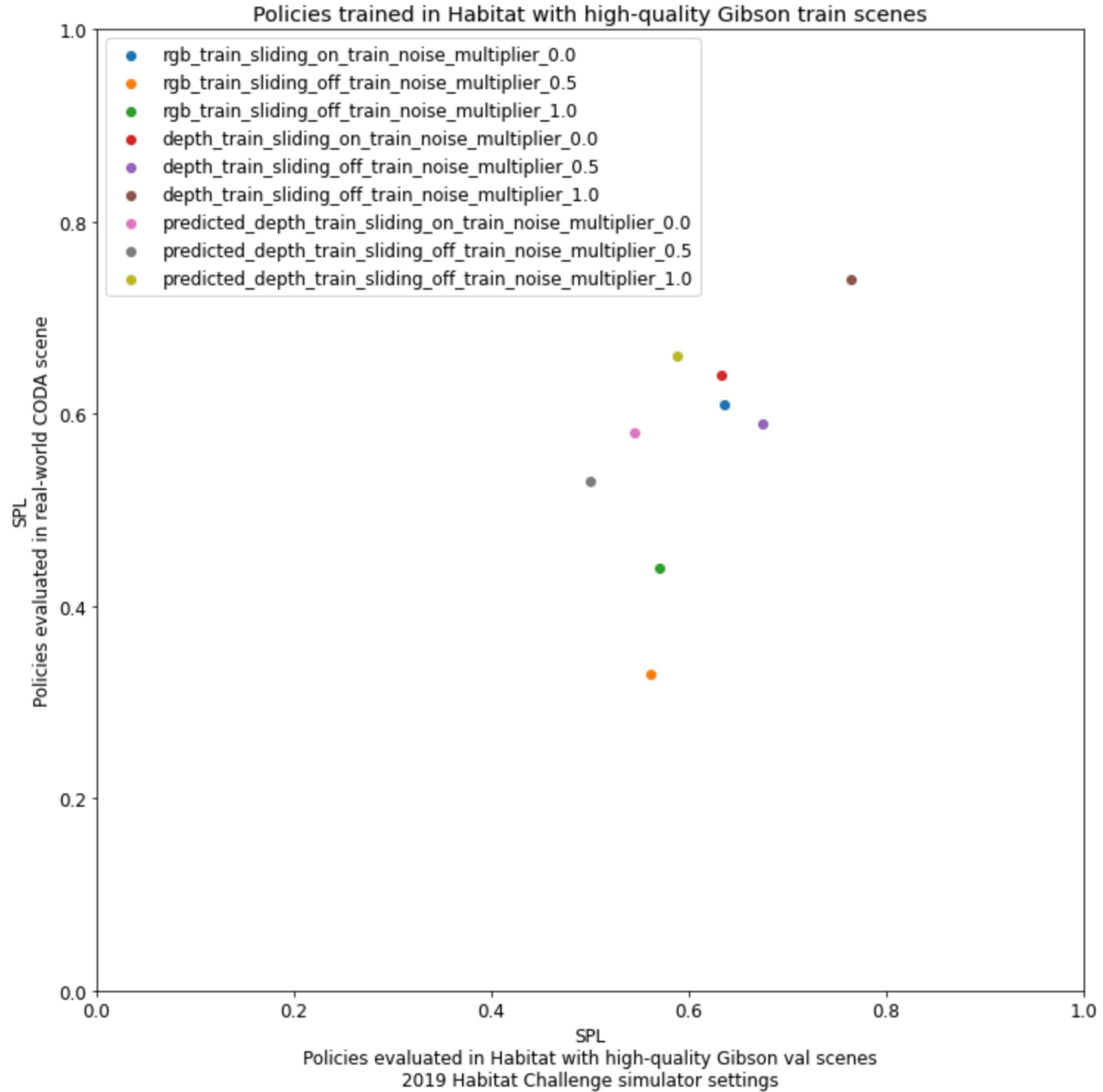


$r = 0.21552588589707233$

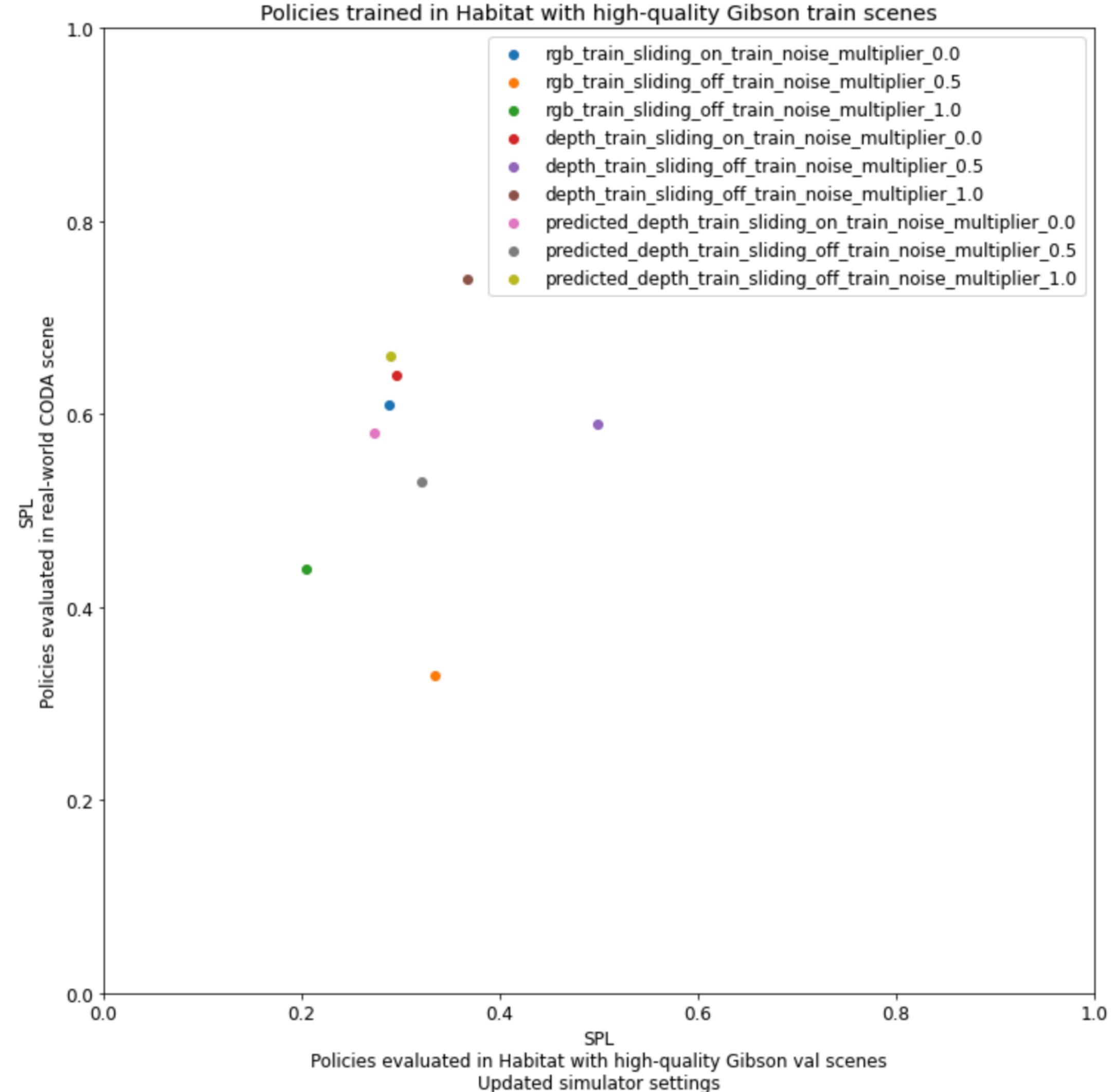


Conclusion: either the optimized simulator settings are actually less realistic than the original settings, or this methodology cannot be used with **unpaired** real and sim scenes.

$r = 0.6314968310270521$

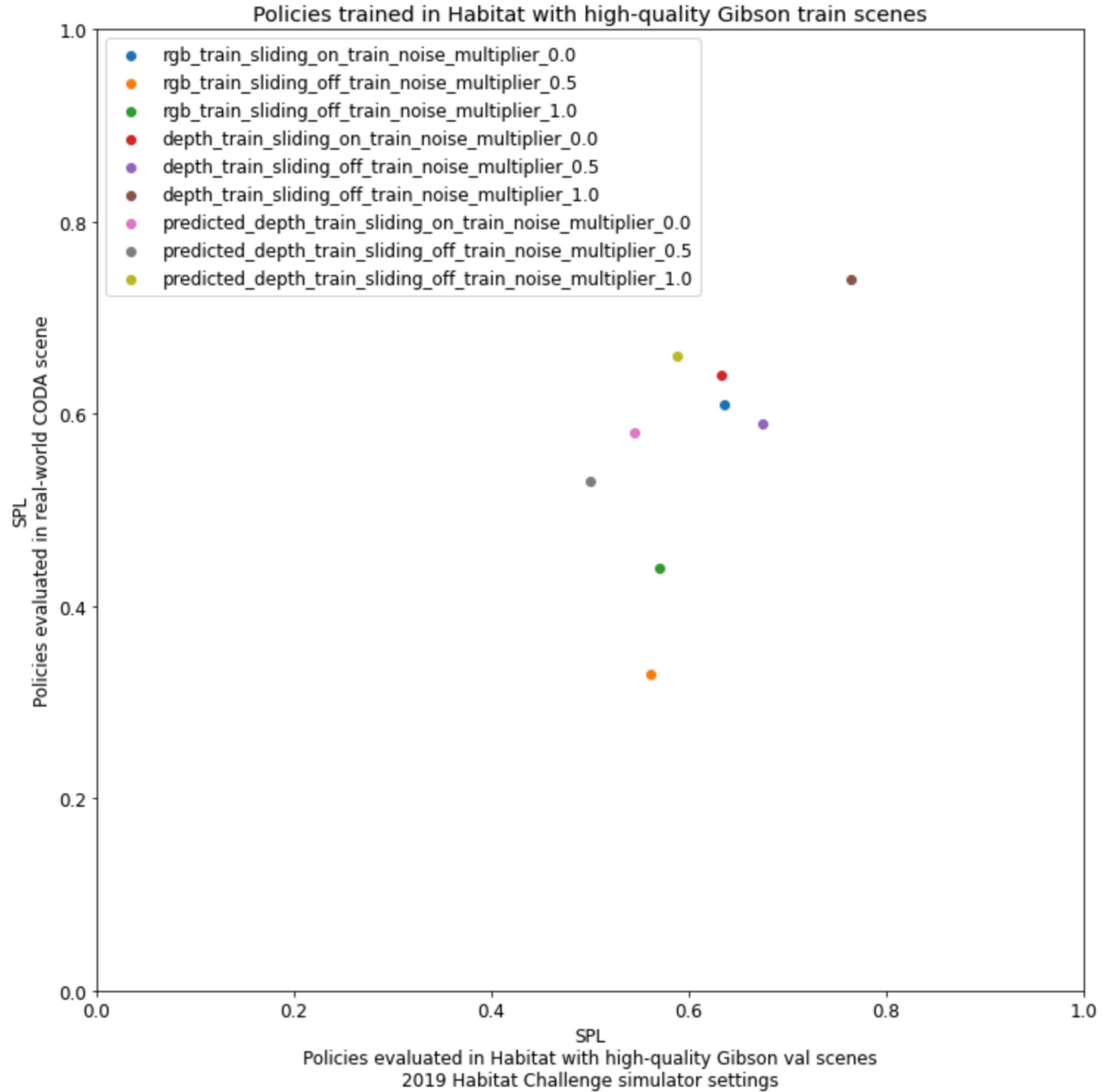


$r = 0.21552588589707233$

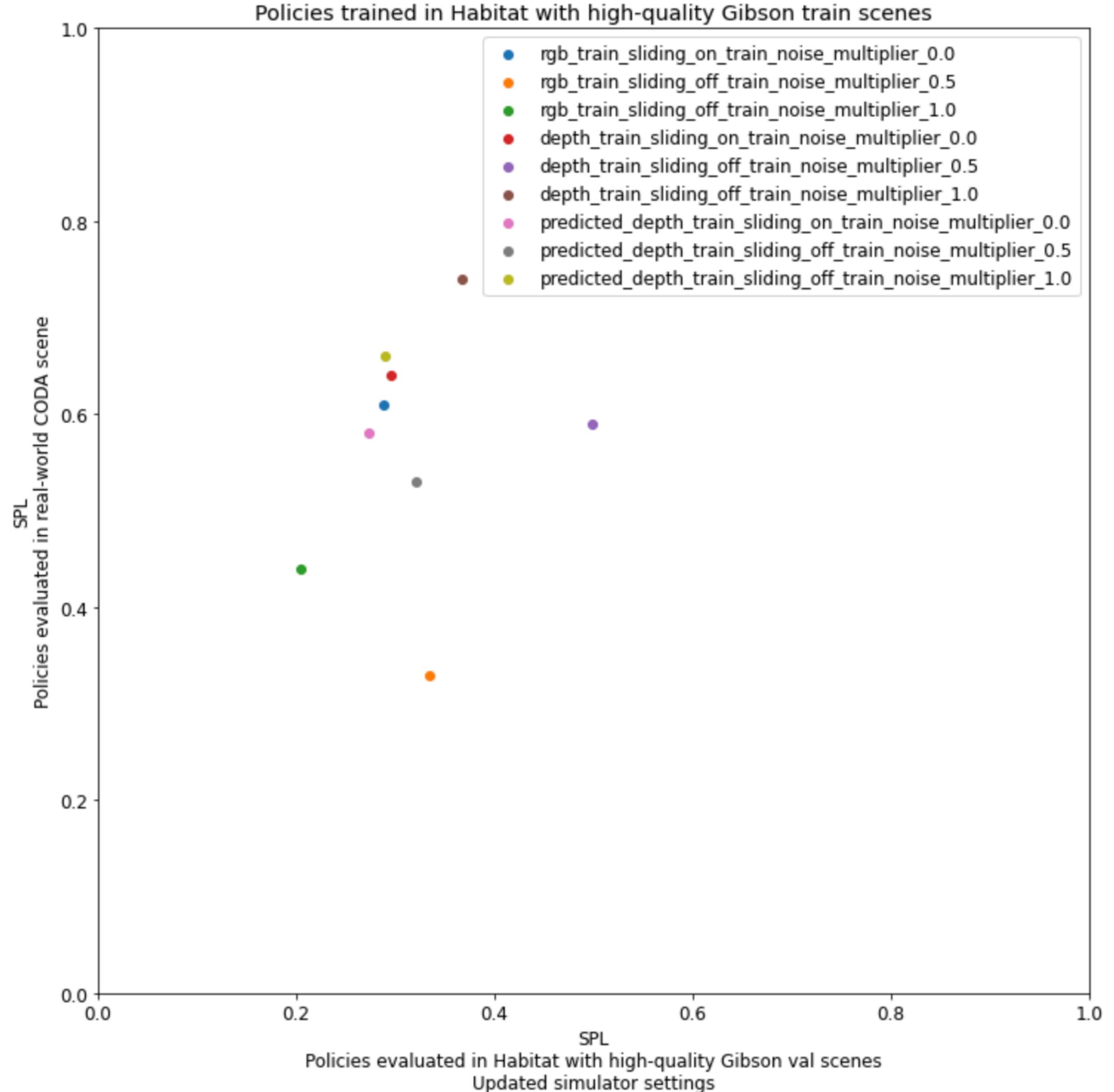


I believe the optimized simulator settings are actually more realistic than the original settings.
So why can't we use unpaired real and sim scenes?

$r = 0.6314968310270521$

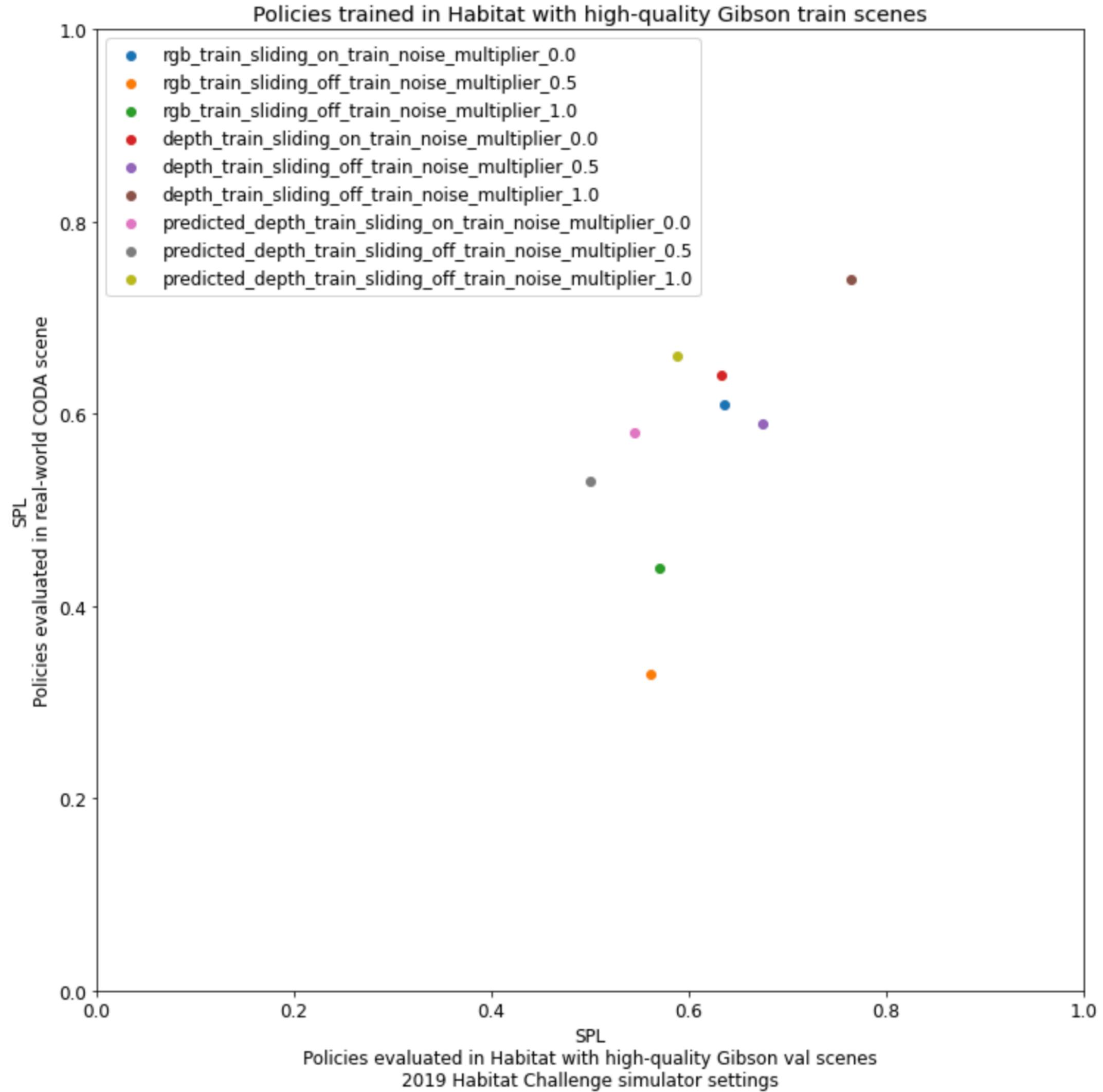


$r = 0.21552588589707233$

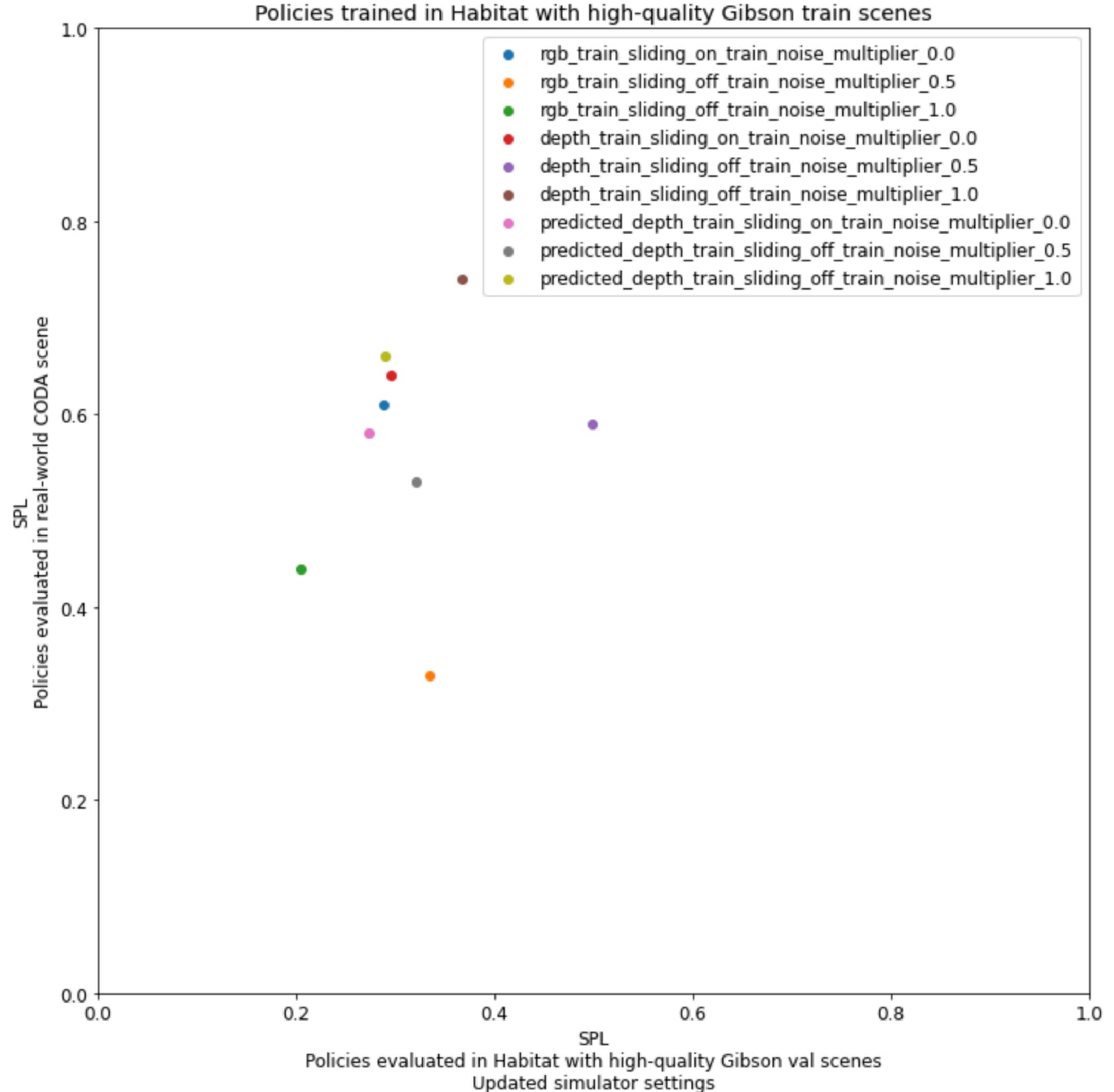


Hypothesis: The real-world scene we're using is **not representative** of an average Gibson scene, and the extent to which the scene is not representative is different for different simulators.

$r = 0.6314968310270521$

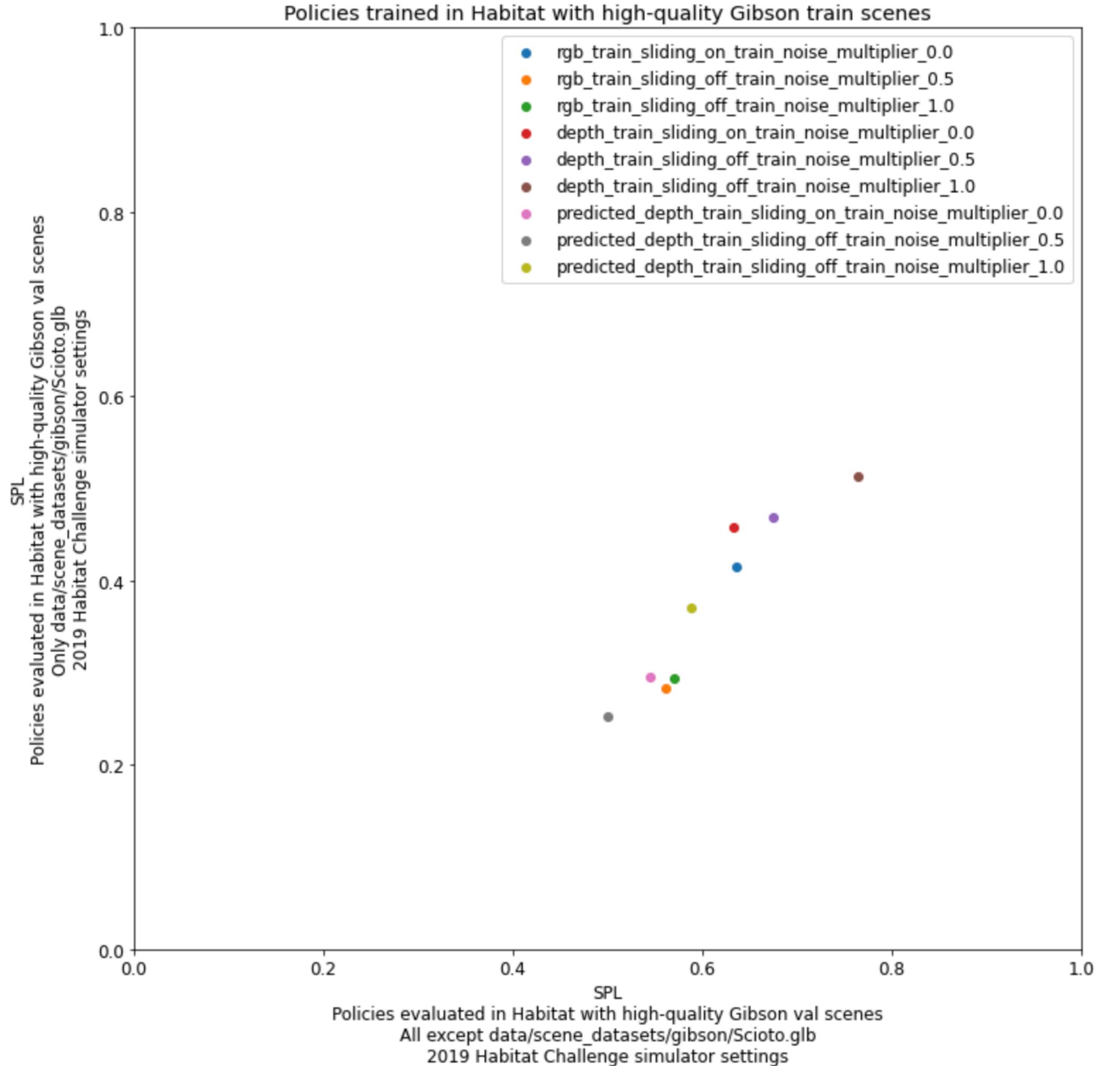


$r = 0.21552588589707233$



If this hypothesis is correct, we should be able to find Gibson scenes that behave similarly.
Can we find any such scenes?

$r = 0.9424696043851616$

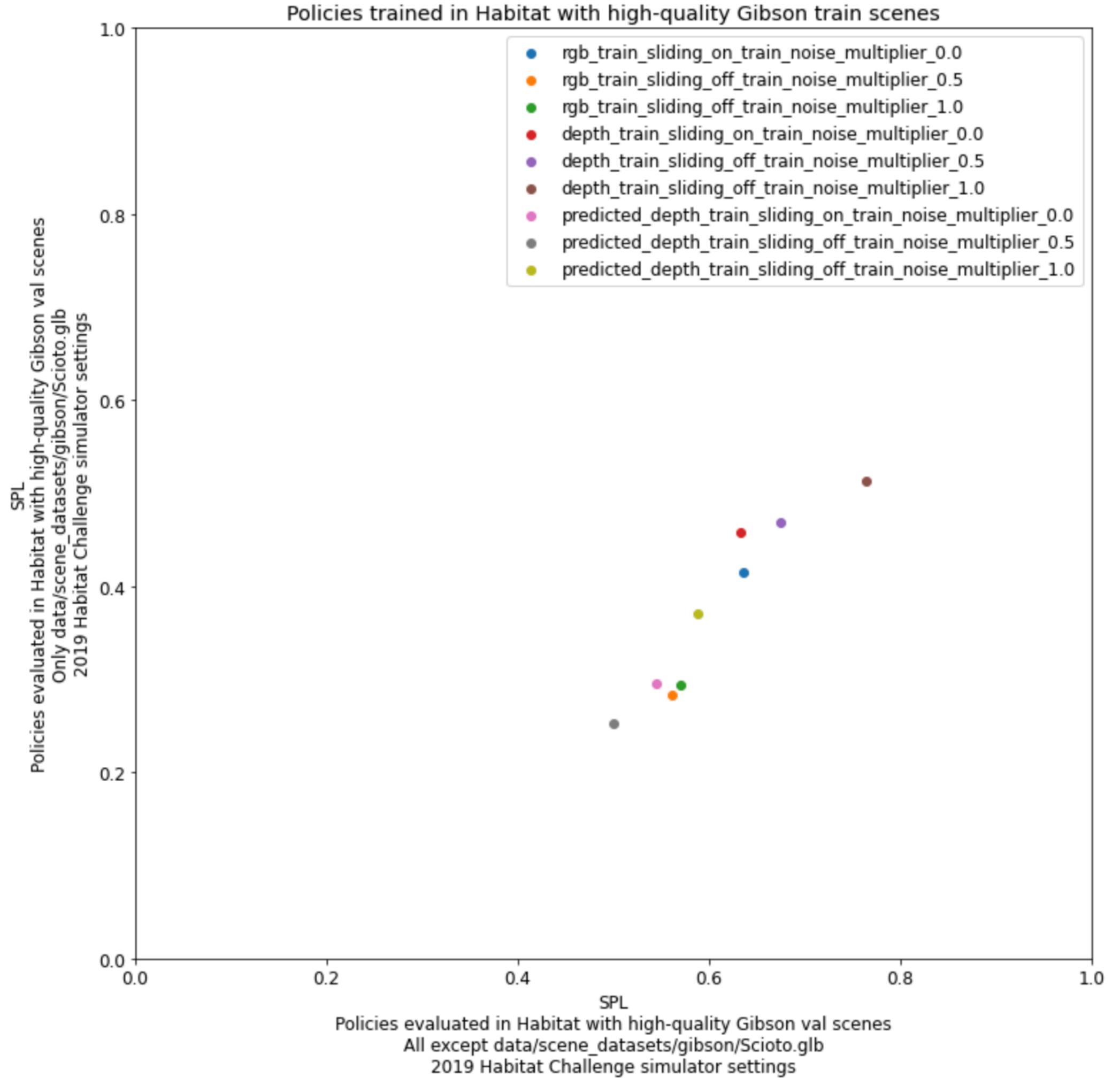


Habitat Challenge 2019 simulator settings

Scioto scene from Gibson val (sim) vs all other scenes from Gibson val (sim)

High correlation (0.942)

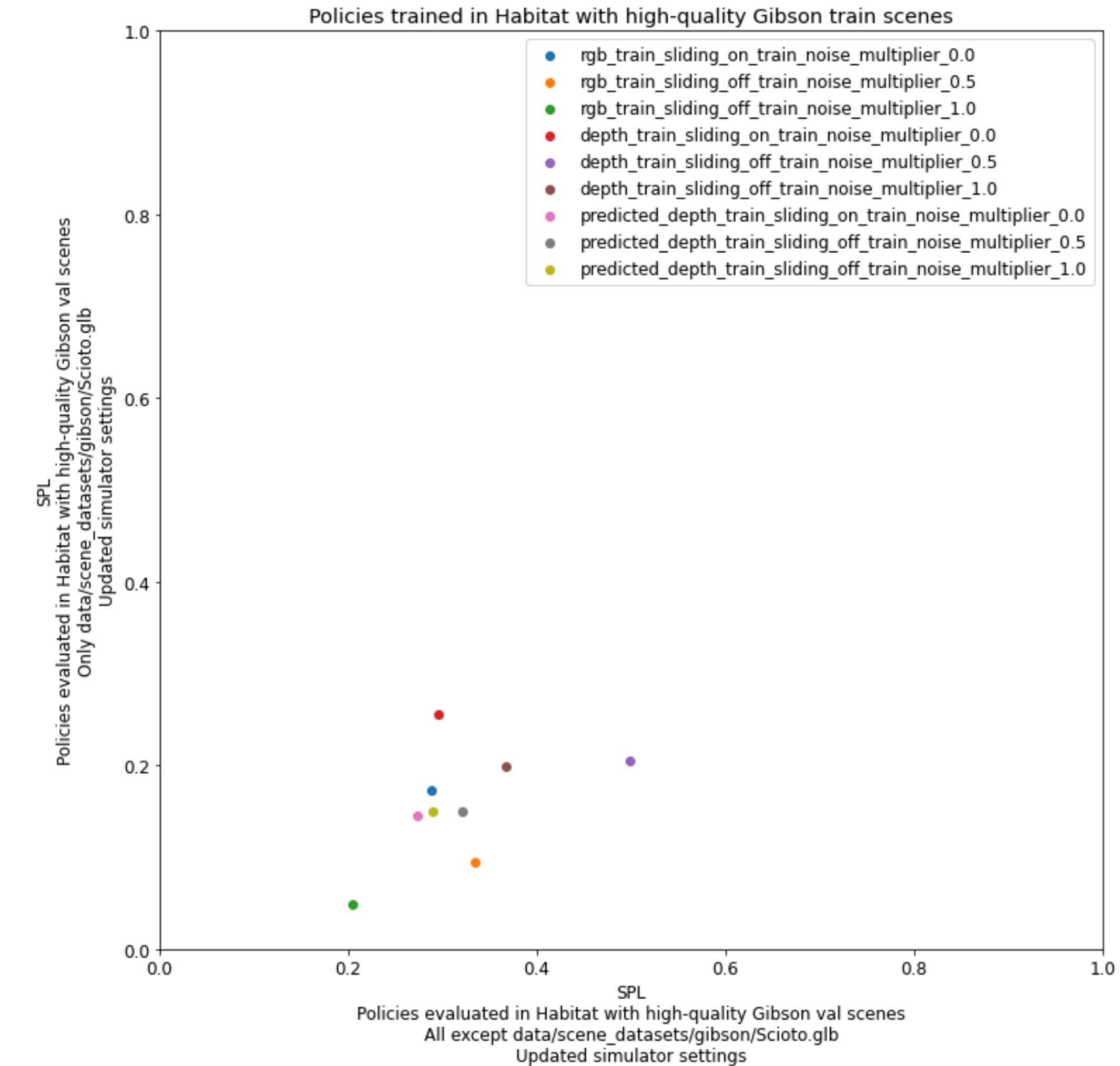
$r = 0.9424696043851616$



Habitat Challenge 2019 simulator settings

Scioto scene from Gibson val (sim) vs all other scenes from Gibson val (sim)
High correlation (0.942)

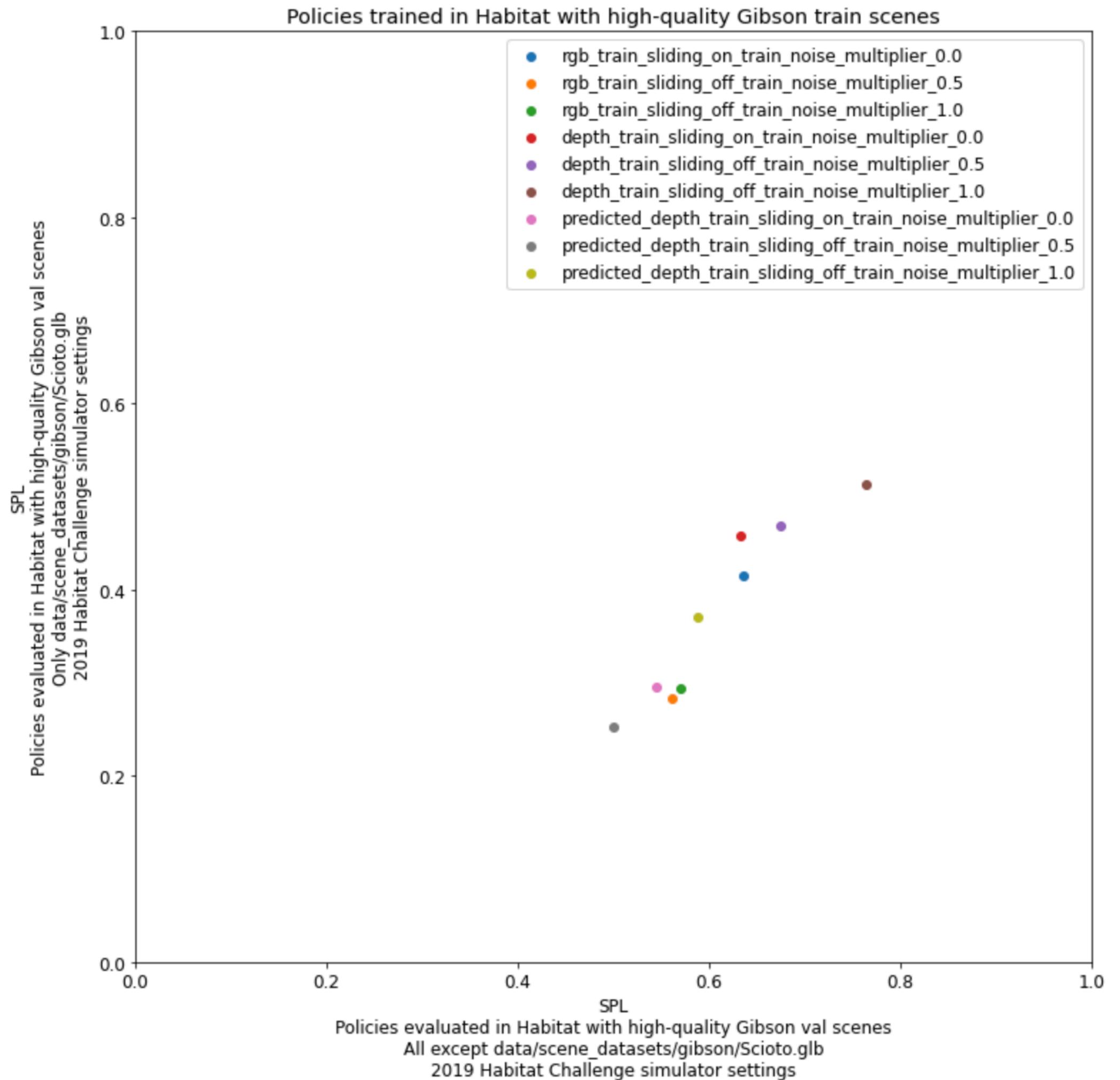
$r = 0.5103211546778714$



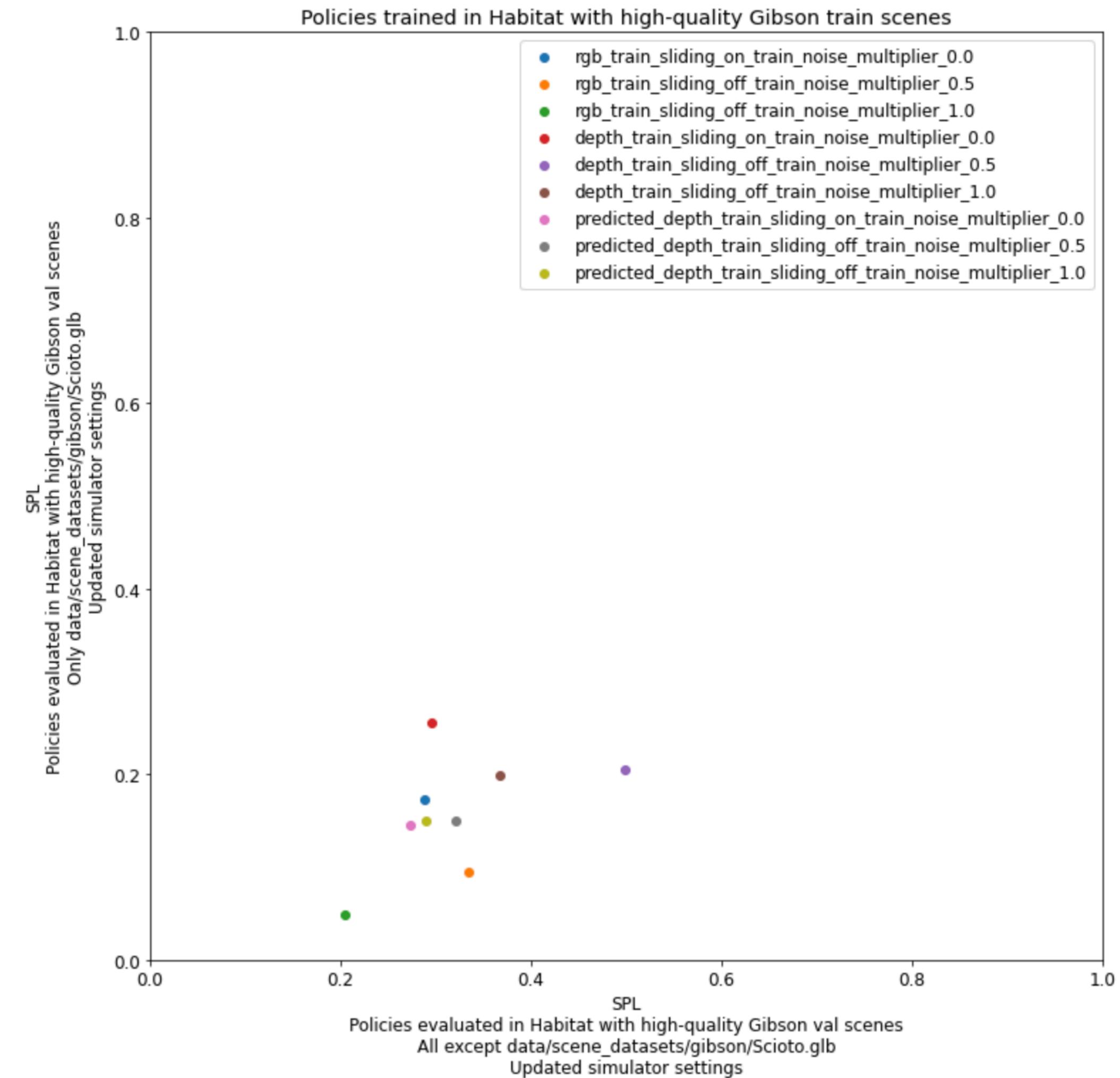
Optimized simulator settings

Scioto scene from Gibson val (sim) vs all other scenes from Gibson val (sim)
Low correlation (0.510)

$r = 0.9424696043851616$



$r = 0.5103211546778714$



Can we find a Gibson scene that behaves similarly to the real-world green room? Yes.



Scioto scene from Gibson val (sim)