

Table 1: ID and OOD test grid accuracy with SEM (Standard Error of the Mean) as error bars for the *CompGen* experiments part of settings C1-C3 and for the ResNet and Vanilla-ViT models.

Setting	Experiment	ResNet		Vanilla-ViT	
		ID	OOD	ID	OOD
C1	experiment-1	0.1 ± 0.1	0.0 ± 0.0	23.3 ± 5.6	0.0 ± 0.0
	experiment-2	8.9 ± 3.1	0.0 ± 0.0	22.7 ± 3.1	0.0 ± 0.0
	experiment-3	1.0 ± 0.1	0.0 ± 0.0	5.4 ± 0.7	0.0 ± 0.0
	experiment-4	0.1 ± 0.0	0.0 ± 0.0	89.5 ± 2.1	76.6 ± 5.0
	experiment-5	0.6 ± 0.3	0.0 ± 0.0	14.4 ± 0.4	0.0 ± 0.0
C2	experiment-1	0.7 ± 0.2	0.0 ± 0.0	26.8 ± 0.5	0.0 ± 0.0
	experiment-2	12.2 ± 6.5	0.0 ± 0.0	14.6 ± 0.5	0.0 ± 0.0
	experiment-3	0.6 ± 0.2	0.0 ± 0.0	8.8 ± 1.2	0.0 ± 0.0
	experiment-4	0.4 ± 0.0	0.0 ± 0.0	88.6 ± 5.2	72.3 ± 10.8
	experiment-5	0.2 ± 0.1	0.0 ± 0.0	21.6 ± 1.5	0.0 ± 0.0
C3	experiment-1	0.0 ± 0.0	0.0 ± 0.0	13.8 ± 1.2	0.8 ± 0.1
	experiment-2	10.4 ± 3.2	3.2 ± 1.1	37.4 ± 5.8	5.5 ± 2.1
	experiment-3	3.0 ± 0.6	0.1 ± 0.0	12.6 ± 2.0	1.1 ± 0.3
	experiment-4	0.0 ± 0.0	0.0 ± 0.0	16.3 ± 1.6	0.4 ± 0.1
	experiment-5	0.1 ± 0.1	0.0 ± 0.0	15.1 ± 6.1	0.3 ± 0.1