Table 1: DCI evaluation. Note the re-ordering of C and I.

model	DCI := (D, I, C) ↑ Shapes3D	MPI3D	Falcor3D	Isaac3D
AE	(0.11 0.83 0.08)	(0.15 0.82 0.14)	(0.08 0.76 0.07)	(0.13 0.85 0.11)
β-VAE	(0.61 **1.00** 0.47)	(0.31 0.82 0.27)	(0.32 0.85 0.28)	(0.22 0.88 0.19)
β-TCVAE	(0.47 0.99 0.39)	(0.23 0.77 0.21)	(0.36 0.90 0.32)	(0.19 0.83 0.17)
BioAE	(0.32 0.94 0.25)	(0.23 0.79 0.19)	(0.22 0.81 0.17)	(0.38 0.91 0.31)
VQ-VAE	(0.40 0.83 0.34)	(0.10 0.63 0.14)	(0.30 0.80 0.28)	(0.33 0.89 0.31)
QLAE	(**0.80** 1.00 **0.61**)	(**0.36** **0.85** 0.35)	(**0.50** **0.97** 0.39)	(**0.70** **1.00** 0.55)
InfoWGAN-GP	(0.23 0.80 0.19)	(0.09 0.64 0.08)	(0.10 **0.74** 0.08)	(0.13 0.71 0.11)
QLInfoWGAN-GP	(**0.38** **0.84** 0.30)	(**0.23** **0.71** 0.25)	(**0.20** 0.72 0.24)	(**0.24** **0.79** 0.25)

Table 2: InfoMEC evaluation.

model	InfoMEC := (InfoM InfoE InfoC) ↑ Shapes3D	MPI3D	Falcor3D	Isaac3D
AE	(0.43 1.00 0.27)	(0.38 **0.72** 0.35)	(0.38 0.75 0.21)	(0.42 0.82 0.21)
β-VAE	(0.58 **1.00** 0.49)	(0.46 0.71 0.50)	(0.71 0.73 0.70)	(0.60 0.81 **0.52**)
β-TCVAE	(0.62 0.83 **0.63**)	(0.52 0.61 **0.57**)	(0.64 0.74 **0.71**)	(0.53 0.71 0.47)
BioAE	(0.56 0.98 0.44)	(0.45 0.66 0.36)	(0.54 0.72 0.31)	(0.62 0.66 0.35)
VQ-VAE	(0.71 0.98 0.46)	(0.45 0.57 0.23)	(0.61 **0.83** 0.42)	(0.58 0.87 0.44)
QLAE	(**0.94** **1.00** 0.55)	(**0.61** 0.63 0.50)	(**0.71** 0.77 0.44)	(**0.78** **0.97** 0.50)
InfoWGAN-GP	(0.61 **0.77** 0.41)	(0.44 0.39 0.20)	(0.45 **0.60** 0.30)	(0.52 0.51 0.24)
QLInfoWGAN-GP	(**0.74** 0.74 0.47)	(**0.60** **0.52** 0.37)	(**0.56** 0.54 0.56)	(**0.63** **0.58** 0.49)

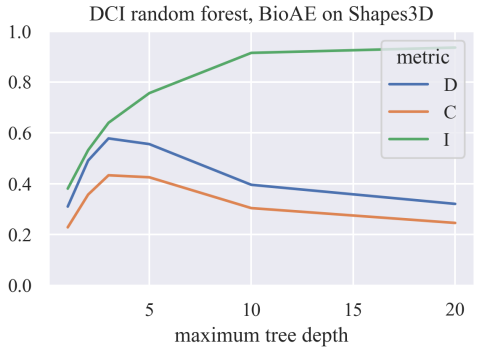

Figure 1: Computing random forest DCI is highly sensitive to the maximum tree depth hyperparameter, and it is easy to overestimate D and C if this is not tuned w.r.t. I.

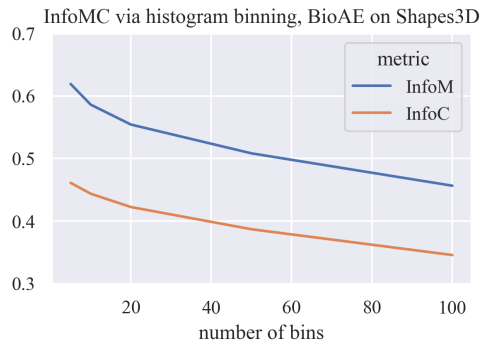

Figure 2: Computing metrics for a model with continuous latents with mutual information estimated via histogram binning results in high sensitivity to the binning strategy.

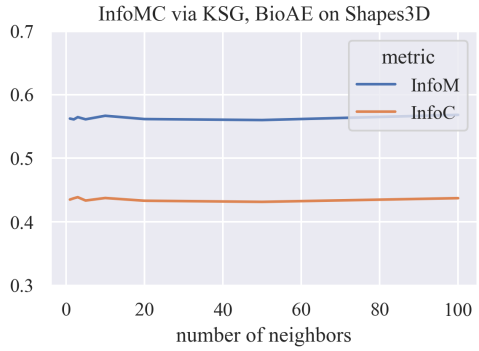

Figure 3: Computing metrics for a model with continuous latents with mutual information estimated via k-neighbors based KSG results in high robustness to k.

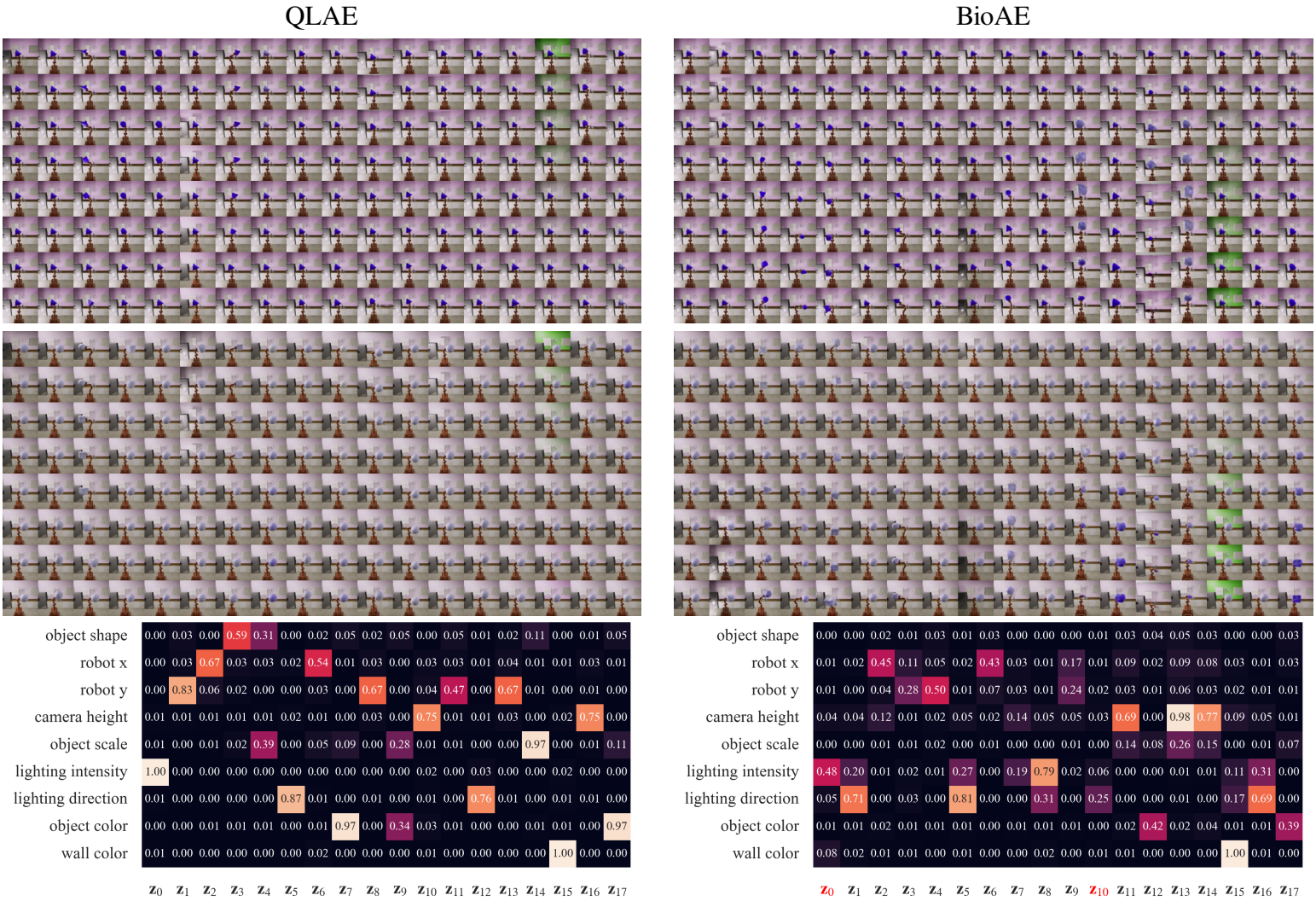

Figure 4: Decoded Isaac3D latent interventions and NMI heatmaps for QLAE and the prior method with highest InfoM. In each image block, a single data sample x is encoded into its latent representation z. To form each block's j-th column, z_j is intervened on with a linear interpolation of its range across the dataset. The resulting latent representations are then decoded. NMI heatmaps are annotated with source names, and inactive latents (as determined by range) have red font. In active latents, the visual variation in the generations caused by each latent tightly corresponds to sources that have significant NMI values in that latent's column.