	NMI ($\uparrow$)	ARI ($\uparrow$)	ACC ($\uparrow$)	LP ($\uparrow$)	DP ($\uparrow$)	LHD ($\downarrow$)
TreeVAE+CLIP	0.665	0.285	0.255	0.255	0.181	0.285
L2H-TEMI	0.778	0.565	0.682	0.701	0.502	0.298
L2H-TURTLE	0.917	0.831	0.896	0.897	0.803	0.235

Table: Performance on the CIFAR-100 dataset between TreeVAE trained on CLIP embeddings, and L2H-TURTLE, L2H-TEMI. The performance of TreeVAE improves if trained on CLIP embeddings, but it is still far from the performance of our approach (L2H-TURTLE, L2H-TEMI). Notably TreeVAE takes more than two hours on a GPU to train while e.g. L2H-TURTLE has a total runtime of less than two minutes.