

		Flat				Hierarchical		# leaves	Inference on test set
		NMI ($\uparrow$)	ARI ($\uparrow$)	ACC ($\uparrow$)	LP ($\uparrow$)	DP ($\uparrow$)	LHD ($\downarrow$)		
CIFAR-10	Agglomerative	0.074	0.038	0.211	0.246	0.121	0.549	10	$\times$
	HypHC	0.019	0.009	0.134	0.359	0.104	0.569	10	$\times$
	DeepECT	0.006	0.002	0.110	0.110	0.101	<u>0.369</u>	2-3	$\checkmark$
	TreeVAE	0.414	0.313	0.497	0.523	0.341	0.410	10	$\checkmark$
L2H-TCL		0.785	0.744	0.868	0.877	0.733	0.398	10	$\checkmark$
L2H-TEMI		0.907	0.906	0.956	0.957	0.902	0.348	10	$\checkmark$
L2H-Turtle		0.985	0.989	0.995	0.995	0.988	0.277	10	$\checkmark$
CIFAR-100	Agglomerative	0.223	0.020	0.090	0.131	0.019	0.428	100	$\times$
	HypHC	0.072	0.004	0.031	0.560	0.011	0.499	100	$\times$
	DeepECT	0.016	0.005	0.070	0.070	0.052	<u>0.121</u>	2-3	$\checkmark$
	TreeVAE	0.199	0.098	0.228	0.242	0.103	0.484	20	$\checkmark$
L2H-TCL		0.547	0.215	0.343	0.437	0.218	0.351	100	$\checkmark$
L2H-TEMI		0.778	0.565	0.682	0.701	0.502	0.298	100	$\checkmark$
L2H-Turtle		0.917	0.831	0.896	0.897	0.803	0.235	100	$\checkmark$
Food-101	Agglomerative	0.082	0.004	0.039	0.045	0.011	0.438	101	$\times$
	HypHC	0.035	0.002	0.022	0.630	0.011	0.573	101	$\times$
	DeepECT	0.003	0.000	0.011	0.011	0.010	<u>0.333</u>	2-3	$\checkmark$
	TreeVAE	0.114	0.017	0.057	0.058	0.016	<u>0.483</u>	20	$\checkmark$
L2H-TCL		0.455	0.168	0.279	0.348	0.117	0.396	101	$\checkmark$
L2H-TEMI		0.917	0.841	0.904	0.913	0.801	0.270	101	$\checkmark$
L2H-Turtle		0.894	0.800	0.876	0.877	0.758	0.297	101	$\checkmark$

Table: Results, highlighted in yellow, for our method implemented on top of TCL (L2H-TCL) on CIFAR-10, CIFAR-100, Food-101 datasets. These results complement the ones in Table 1 from the manuscript as shown here.