Supplementary Material
Spectral Metric for Dataset Complexity Assessment

1. Instance selection

In the results section, we showed that STL-10 was a bit harder than CIFAR10 with 500 samples per class even though these datasets are visually very similar. To further our point, we show in Fig. 1 the MDS representation of both datasets. While the *car-truck* distance is smaller in CIFAR10, the six animal classes in STL-10 are overall closer together and thus a slightly more difficult to disentangle.

2. Comparison with other graph-based methods

We mentioned in the *Previous works* section that other graph-based methods have been proposed in the past. One could wonder how our measure differs from those. The main advantage of our approach compared to other graph-based methods is the fact that our graph embeds classes and not samples, thanks to the spectral clustering formalism. This leads to a $K \times K$ Laplacian matrix which is order of magnitude smaller that the $N \times N$ similarity matrix often required by other methods. This brings a huge advantage both memory and processing wise while allowing our method to naturally expand to the number of classes.

3. How to choose the auto-encoder architecture?

Empirical evidences show that any auto-encoder that properly reconstruct images can be used with our method. The key element for our method is to have low-dimensional space which correctly regroup images with similar content.

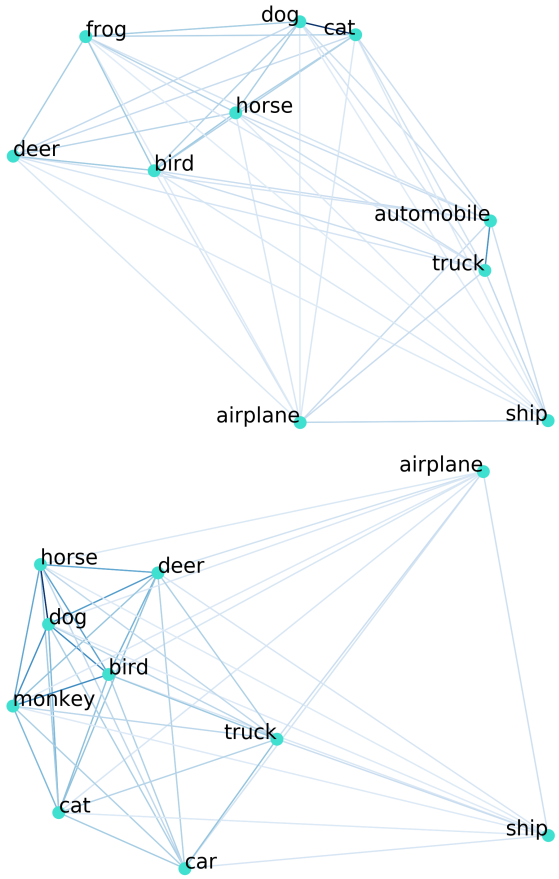

Figure 1: Comparison between the [Top] 2D plot of CIFAR10 with 500 samples per class and [Bottom] 2D plot of STL-10. As we see, the nodes in STL-10 seems closer than in CIFAR10.

CNN_{AE}	CSG	F1	F2	F3	F4	N1	N2	N3	N4	T1	T2	AlexNet Acc.	Resnet Acc.
compcars	6.421	0.025	2.10E-15	0.031	3.582	0.796	1.018	0.668	0.176	1	23.27	0.303	0.122
stl10	5.051	0.243	2.56E-15	0.056	4.915	0.869	1.075	0.78	0.609	1	19.531	0.314	0.367
cifar10	3.509	0.331	1.37E-14	0.022	0.87	0.712	1.002	0.618	0.468	1	195.312	0.821	0.807
svhn	3.986	0.008	1.28E-14	0.01	0.406	0.637	0.989	0.525	0.646	1	286.16	0.918	0.933
notMNIST	0.564	0.756	1.30E-11	0.133	6.915	0.226	0.599	0.155	0.042	1	73.141	0.951	0.96
mnist	0.202	0.723	2.44E-19	0.258	2.803	0.078	0.7	0.039	0.021	0.998	234.375	0.994	0.954

Table 1: Complexity estimation for all 10 class datasets for all methods using CNN_{AE} embedding.

CNN-TSNE	CSG	F1	F2	F3	F4	N1	N2	N3	N4	T1	T2	AlexNet Acc.	Resnet Acc.
compcars	2.928	0.014	35.578	0.031	0.084	0.845	1.633	0.632	0.874	0.993	1985.667	0.303	0.122
stl10	3.072	0.222	37.264	0.031	0.073	0.91	2.426	0.788	0.831	0.999	1666.667	0.314	0.367
cifar10	1.004	0.291	36.344	0.009	0.017	0.797	1.585	0.622	0.819	0.998	16666.666	0.821	0.807
svhn	1.154	0.001	42.921	0.002	0.004	0.748	1.408	0.546	0.89	0.999	24419	0.918	0.933
notMNIST	0.725	1.752	33.723	0.083	0.212	0.234	0.284	0.158	0.434	0.999	6241.333	0.951	0.96
mnist	0.112	6.193	26.706	0.229	0.522	0.055	0.112	0.03	0.104	0.982	20000	0.994	0.954

Table 2: Complexity estimation for all 10 class datasets for all methods using CNN_{AE} t-SNE embedding.

t-SNE	CSG	F1	F2	F3	F4	N1	N2	N3	N4	T1	T2	AlexNet Acc.	Resnet Acc.
compcars	2.079	0.018	30.066	0.03	0.092	0.857	1.518	0.675	0.867	0.99	1191.4	0.303	0.122
stl10	1.761	0.822	20.818	0.199	0.415	0.843	1.512	0.704	0.756	0.999	1000	0.314	0.367
cifar10	0.858	0.294	32.037	0.012	0.032	0.805	1.504	0.643	0.816	0.996	10000	0.821	0.807
svhn	0.967	0.007	38.829	0.009	0.02	0.712	1.265	0.518	0.88	1	14651.4	0.918	0.933
notMNIST	0.709	1.847	24.405	0.197	0.513	0.193	0.288	0.136	0.356	0.999	3744.8	0.951	0.96
mnist	0.151	4.721	16.875	0.393	1.112	0.051	0.105	0.029	0.083	0.998	12000	0.994	0.954

Table 3: Complexity estimation for all 10 class datasets for all methods using t-SNE embedding.

Raw	CSG	F1	F2	F3	F4	N1	N2	N3	N4	T1	T2	AlexNet Acc.	Resnet Acc.
compcars	6.700	0.031	2.10E-43	0.053	5.962	0.81	1.024	0.713	0.172	0.99	0.121	0.618	0.52
stl10	3.349	0.324	7.81E-06	0.243	16.97	0.795	1.037	0.732	0.418	0.99	0.181	0.994	0.954
cifar10	3.579	0.244	3.54E+01	0.008	0.975	0.742	1.012	0.649	0.602	0.99	16.276	0.951	0.96
svhn	3.889	0.005	2.45E-06	0.004	3	0.626	0.987	0.511	0.667	0.99	23.847	0.897	0.896
notMNIST	0.512	0.996	4.50E+01	0	0	0.198	0.577	0.137	0.038	1	7.961	0.918	0.933
mnist	0.084	0.735	0.00E+00	0.243	5.203	0.054	0.7	0.026	0.007	0.99	76.531	0.314	0.367

Table 4: Complexity estimation for all 10 class datasets for all methods using no embedding.

	Alexnet				ResNet				XceptionNet			
Method	Raw	CNN_{AE}	*t-SNE*	CNN_{AE} *t-SNE*	Raw	CNN_{AE}	*t-SNE*	CNN_{AE} *t-SNE*	Raw	CNN_{AE}	*t-SNE*	CNN_{AE} *t-SNE*
N4	0.141	0.558	0.763	0.780	0.021	0.407	0.746	0.745	0.067	0.558	0.763	0.780
F3	0.290	0.665	0.459	0.604	0.201	0.637	0.500	0.552	0.267	0.665	0.459	0.604
F1	0.483	0.895	0.667	0.673	0.519	0.922	0.648	0.626	0.458	0.895	0.669	0.673
F2	0.366	0.449	0.234	0.660	0.432	0.470	0.320	0.576	0.465	0.449	0.234	0.660
T1	0.642	0.519	0.505	0.332	0.642	0.461	0.677	0.214	0.642	0.519	0.505	0.332
T2	0.655	0.783	0.783	0.783	0.578	0.737	0.737	0.737	0.578	0.783	0.783	0.783
N2	0.677	0.812	0.794	0.854	0.619	0.775	0.761	0.760	0.564	0.816	0.794	0.854
F4	0.760	0.067	0.531	0.606	0.606	0.001	0.542	0.551	0.676	0.067	0.531	0.606
N1	0.767	0.861	0.817	0.833	0.700	0.809	0.784	0.783	0.651	0.861	0.817	0.833
N3	0.803	0.855	0.823	0.837	0.727	0.794	0.781	0.765	0.681	0.855	0.823	0.837
Our Method	0.696	0.823	0.903	**0.968**	0.712	0.838	0.932	**0.935**	0.718	0.804	0.931	**0.951**

Table 5: Correlation values between the accuracy of AlexNet, ResNet and XceptionNet and Ho & Basu c-measures with four different embeddings. These results complement Table 3, 4 and 5 in the original paper.