Appendix of ”Deep Self-paced Active Learning for Image Clustering”

Appendix A: ACC Results of All Methods

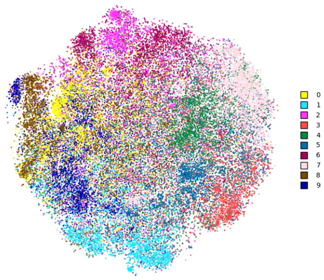
Table 1 shows the ACC results of all methods on all data sets. We can see that DSAC can achieve better performance with much fewer human annotations. Although TOD outperforms ours on CIFAR-100-20, it needs 5000 annotations but ours only needs 100 annotations. Moreover, with 100 annotations, our performance is comparable with TOD with 4000 annotations on ACC.

Appendix B: t-sne Results on CIFAR-10

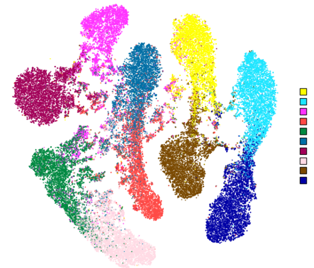
We show the t-sne [8] results of the embedding learned from our network with different numbers of annotations on the CIFAR-10 data set. Figure 1 shows the results. Figure 1(a) shows the embedding obtained by our networks without any annotations. Figure 1(b)-(f) show the results with $20, 40, \cdots, 100$ annotations, respectively. It can be seen that the embeddings entangle when there are no annotations and have a clearer clustering structure when having 20 annotations. Moreover, just with 40 annotations, our network can learn a much better embedding for data.

Table 1: ACC results with different numbers of selection annotations on all the data sets.

Data sets	Number of annotations	SDEC [1]	AutoEmbedder [2]	SRAAL [3]	TOD [4]	CoreGCN [5]	ADC [6]	MCDAL [7]	Number of annotations	DSAC-R	DSAC
STL-10	200	0.1957	0.5750	0.2883	0.3269	0.1193	0.1980	0.2695	20	0.3774	0.8725
	400	0.1958	0.7110	0.3486	0.3870	0.2180	0.2070	0.5301	40	0.5033	0.9020
	600	0.1963	0.7690	0.3819	0.4212	0.2076	0.2430	0.6627	60	0.6308	0.9023
	800	0.1951	0.7920	0.3929	0.4432	0.3442	0.2577	0.7327	80	0.7587	0.9051
	1000	0.1958	0.7730	0.4210	0.4605	0.2873	0.2621	0.7262	100	0.7931	**0.9103**
CIFAR-10	1000	0.2259	0.4910	0.4037	0.4876	0.3266	0.2315	0.5535	20	0.4786	0.5800
	2000	0.2290	0.5860	0.4647	0.5906	0.3816	0.2423	0.6352	40	0.5478	0.8754
	3000	0.2209	0.5820	0.5781	0.7053	0.4210	0.2575	0.7161	60	0.5543	0.8811
	4000	0.2262	0.6220	0.5964	0.7944	0.4899	0.2573	0.7643	80	0.5345	**0.8829**
	5000	0.2239	0.6820	0.6078	0.8366	0.5268	0.2604	0.7941	100	0.5437	0.8818
CIFAR-100-20	1000	0.1363	0.2880	0.2595	0.2884	0.2088	0.1053	0.3198	20	0.3101	0.3796
	2000	0.1370	0.3330	0.3396	0.3453	0.2480	0.1024	0.3631	40	0.3183	0.4903
	3000	0.1368	0.3240	0.3733	0.4173	0.2666	0.1098	0.4264	60	0.3263	0.5024
	4000	0.1367	0.3110	0.4255	0.4910	0.3204	0.1122	0.4627	80	0.3233	0.5016
	5000	0.1361	0.3880	0.4601	**0.5747**	0.3160	0.1095	0.5253	100	0.3239	0.5086



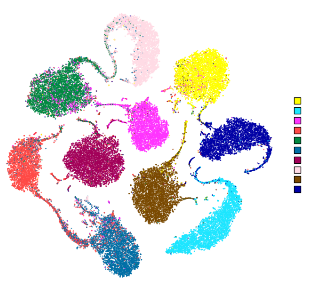
(a) The embedding of original data



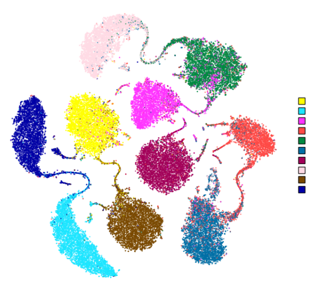
(b) The embedding with 20 annotations

(c) The embedding with 40 annotations

(d) The embedding with 60 annotations



(e) The embedding with 80 annotations



(f) The embedding with 100 annotations

Figure 1: t-sne of the embedding with different numbers of annotations of DSAC on CIFAR-10.

Table 2: Accuracy of the pseudo-labels.

Iteration	1	2	3	4	5
STL-10	0.9864	1.0000	0.9984	0.9971	0.9974
CIFAR-10	0.9926	0.9898	0.9945	0.9967	0.9982
CIFAR-100-20	0.8123	0.8190	0.8606	0.8753	0.8769

Appendix C: Effects of Pseudo-labels

We show the accuracy of the pseudo-labels, which means whether the pseudo-labels generated by our method are correct. The accuracy of the pseudo-labels within the 5 iterations is shown in Table 2. From Table 2, we can find that on STL-10 and CIFAR-10 data sets, the pseudo-labels generated by our method are very accurate, and thus can effectively enlarge the labeled set $\mathcal{L}$ and are helpful for clustering. On the CIFAR-100-20 data set, the pseudo-labels are also good enough for clustering. Moreover, on CIFAR-100-20 we find that the pseudo-labels become more accurate with the iterations. It is because that with the iterations, the quality of the learned embeddings improves and thus it is easier to generate the correct pseudo-labels. It well demonstrates that the embeddings obtained from our neural network have an increasingly clear clustering structure.

Appendix D: Hyperparameter Study

We show the sensitivity of the hyperparameter γ in Figure 2. Figure 2 shows ACC, NMI, and ARI of DSAC on all data sets with γ in the range $[10^{-4}, 10^{3}]$. It can be seen that the performance of DSAC is relatively stable in a wide range of γ.

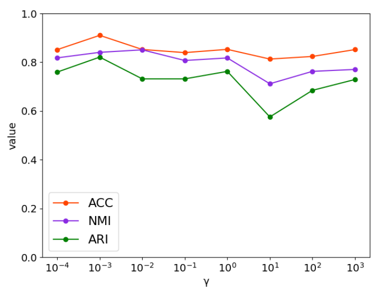
References

[1] Yazhou Ren, Kangrong Hu, Xinyi Dai, Lili Pan, Steven C. H. Hoi, and Zenglin Xu, “Semi-supervised deep embedded clustering,” *Neurocomputing*, vol. 325, pp. 121–130, 2019.

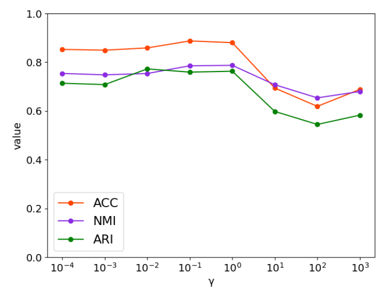
[2] Abu Quwsar Ohi, Muhammad F. Mridha, Farisa Benta Safir, Md. Abdul Hamid, and Muhammad Mostafa Monowar, “Autoembedder: A semi-supervised DNN embedding system for clustering,” *Knowl. Based Syst.*, vol. 204, pp. 106190, 2020.

[3] Beichen Zhang, Liang Li, Shijie Yang, Shuhui Wang, Zheng-Jun Zha, and Qingming Huang, “State-relabeling adversarial active learning,” in *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020, pp. 8753–8762.

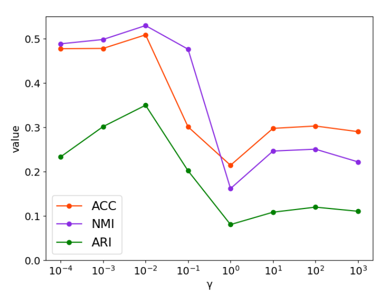
[4] Siyu Huang, Tianyang Wang, Haoyi Xiong, Jun Huan, and Dejing Dou, “Semi-supervised active learning with temporal output discrepancy,” in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021, pp. 3427–3436.



(a) STL-10



(b) CIFAR-10



(c) CIFAR-100-20

Figure 2: ACC, NMI, and ARI on all data sets with different γ.

[5] Razvan Caramalau, Binod Bhattarai, and Tae-Kyun Kim, "Sequential graph convolutional network for active learning," in *IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2021, virtual, June 19-25, 2021*. 2021, pp. 9583–9592, Computer Vision Foundation / IEEE.

[6] Bicheng Sun, Peng Zhou, Liang Du, and Xuejun Li, "Active deep image clustering," *Knowl. Based Syst.*, vol. 252, pp. 109346, 2022.

[7] Jae Won Cho, Dong-Jin Kim, Yunjae Jung, and In So Kweon, "Mcdal: Maximum classifier discrepancy for active learning," *IEEE Transactions on Neural Networks and Learning Systems*, pp. 1–11, 2022.

[8] Laurens van der Maaten and Geoffrey Hinton, "Visualizing data using t-sne," *Journal of Machine Learning Research*, vol. 9, no. 86, pp. 2579–2605, 2008.