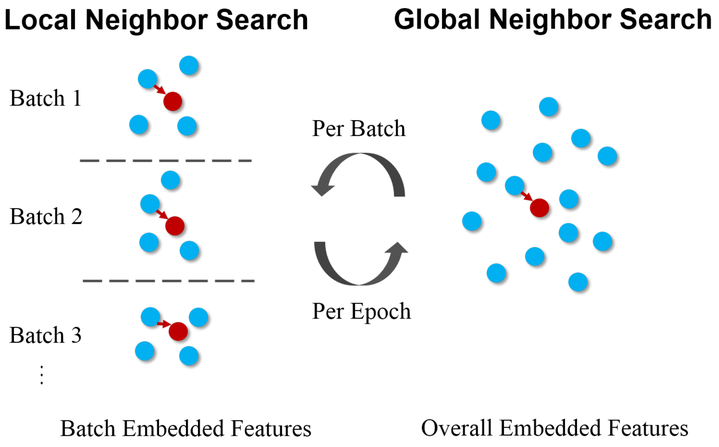
Nearest Neighbor Matching for Deep Clustering

Zhiyuan Dang, Cheng Deng*, Xu Yang, Kun Wei, Heng Huang

CVPR VIRTUAL JUNE 19-25

Overview

Deep Clustering methods have attracted more attention, which goal is that grouping the samples into clusters, such that similar samples into the same cluster while dissimilar ones into different clusters.

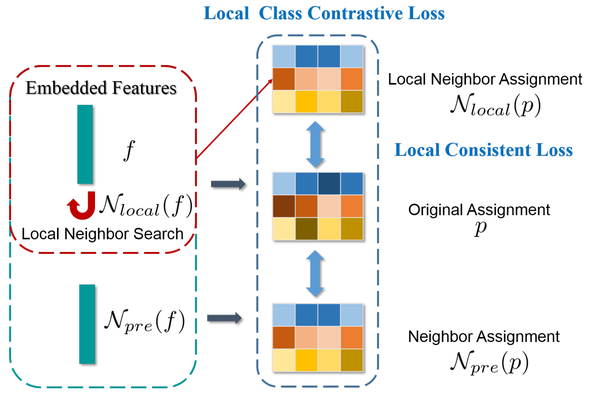


We propose NNM to match more semantically nearest neighbors from between **local (batch)** and **global (overall)** level.

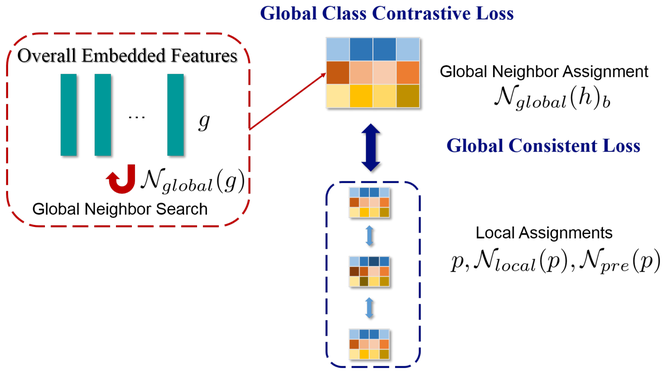
Benefit from the dynamic updated deep features with iteration and epoch increases, we can construct more and more semantically confident sample pairs from samples and its neighbors.

Proposed Method

Local Nearest Neighbor Matching



Global Nearest Neighbor Matching



Experimental Results

Datasets	CIFAR-10			CIFAR-100			STL-10		
Metrics	NMI	ACC	ARI	NMI	ACC	ARI	NMI	ACC	ARI
K-Means [28]	0.087	0.229	0.049	0.084	0.130	0.028	0.125	0.192	0.061
SC [43]	0.103	0.247	0.085	0.090	0.136	0.022	0.098	0.159	0.048
AC [11]	0.105	0.228	0.065	0.098	0.138	0.034	0.239	0.332	0.140
NMF [2]	0.081	0.190	0.034	0.079	0.118	0.026	0.096	0.180	0.046
AE [1]	0.239	0.314	0.169	0.100	0.165	0.048	0.250	0.303	0.161
DAE [35]	0.251	0.297	0.163	0.111	0.151	0.046	0.224	0.302	0.152
DCGAN [32]	0.265	0.315	0.176	0.120	0.151	0.045	0.210	0.298	0.139
DeCNN [42]	0.240	0.282	0.174	0.092	0.133	0.038	0.227	0.299	0.162
VAE [24]	0.245	0.291	0.167	0.108	0.152	0.040	0.200	0.282	0.146
JULE [39]	0.192	0.272	0.138	0.103	0.137	0.033	0.182	0.277	0.164
DEC [38]	0.257	0.301	0.161	0.136	0.185	0.050	0.276	0.359	0.186
DAC [5]	0.396	0.522	0.306	0.185	0.238	0.088	0.366	0.470	0.257
ADC [14]	-	0.325	-	-	0.160	-	-	0.530	-
DDC [4]	0.424	0.524	0.329	-	-	-	0.371	0.489	0.267
DCCM [36]	0.496	0.623	0.408	0.285	0.327	0.173	0.376	0.482	0.262
IIC [21] (Best)	-	0.617	-	-	0.257	-	-	0.610	-
PICA [20] (Best)	0.591	0.696	0.512	0.310	0.337	0.171	0.611	0.713	0.531
DCDC [10]	0.585	0.699	0.506	0.310	0.349	0.179	0.621	0.734	0.547
Supervised	0.862	0.938	0.870	0.680	0.800	0.632	0.659	0.806	0.631
Pretext [6] + K-means	0.598	0.659	0.509	0.402	0.395	0.239	0.604	0.658	0.506
SCAN* [34] (Best)	0.715	0.816	0.665	0.449	0.440	0.283	0.673	0.792	0.618
NNM*	**0.748**	**0.843**	**0.709**	**0.484**	**0.477**	**0.316**	**0.694**	**0.808**	**0.650**

Clustering performance

Confident samples in Cifar-10 dataset

Our code will be available at https://github.com/ZhiyuanDang/NNM