Supplementary Material For NeurIPS 2021

Anonymous Author(s)
Affiliation
Address
email

1 Information of datasets

All benchmark datasets involved in this paper are available in this supplementary material (see ./data/), and the statistics of these datasets are summarized in Table 1.

Table 1: Description of datasets

Datasets	CALFW	FaceV5	CFPW	CMU	Colon	CPLFW	Olivetti	Umist
# Samples	12174	2500	7000	2856	62	11652	400	575
# Features	256	256	256	1024	2000	256	4096	644
# Subjects	4025	500	500	68	2	3930	40	20
Datasets	Dexter	FERET	GTDB	LFW	Madelon	Mpeg7	MUCT	Yale
# Samples	600	11338	750	13233	2600	1400	3755	165
# Features	20000	256	21600	256	500	6000	256	1024
# Subjects	2	994	50	5749	2	70	276	15

2 Fast spectral clustering

We also conducted the experiment for the instance of ratio-cut descripted in Section 3.1. That is to say, we optimized the following problem by the algorithm proposed in this paper.

$$\min_{\mathbf{Y}\in\Phi^{n\times c}} Tr\left((\mathbf{Y}^T\mathbf{Y})^{-p}\mathbf{Y}^T\mathbf{G}^{(k)}\mathbf{Y}\right), \tag{1}$$

with $p = 1$ and

$$\mathbf{g}_{ij}^{(k)} = \begin{cases} \sum_{j=1}^{n} w_{ij} & i = j \\ -w_{ij} & i \neq j, \text{ and } \mathbf{x}_i \in \mathcal{N}_k(\mathbf{x}_j) \\ 0 & \text{Otherwise} \end{cases}, \tag{2}$$

where

$$w_{ij} = \begin{cases} e^{-\frac{\|\mathbf{x}_i - \mathbf{x}_j\|}{t}} & \text{if } \mathbf{x}_i \in \mathcal{N}_k(\mathbf{x}_j) \\ 0 & \text{otherwise} \end{cases}, \tag{3}$$

where t is a parameter, and is setted as the mean value of the $\{\|\mathbf{x}_i - \mathbf{x}_j\| \mid \mathbf{x}_i \in \mathcal{N}_k(\mathbf{x}_j), i, j = 1, \cdots, n\}$. We call this fast implementation Fast spectral clustering (FSC) and compare it with the traditional spectral clustering optimized by eigendecomposition.

The experimental results are shown in Table 2. The k-NN graph is constructed by the 'Brute-force' algorithm and the time it consumed is shown in the column named "Graph".

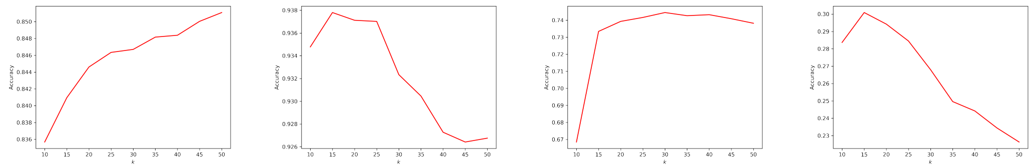
From the experimental results, we can see that FSC has achieved a great improvement in performance and time, compared to SC, which verifies the effectiveness and efficiency of the proposed algorithm.

Table 2: Ratio-cut optimized by the proposed algorithm vs. SC derived from it

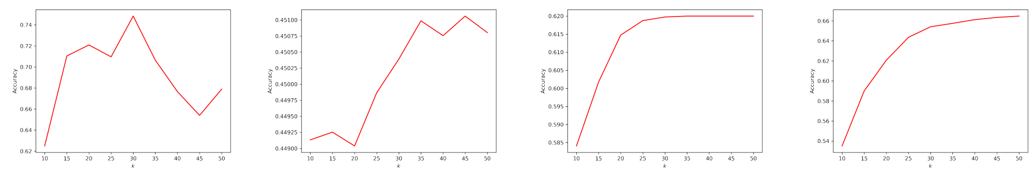
Datasets	Performance (± std)			Time (s)		
	Metrics	SC	FSC	Graph	SC	FSC
CALFW	ACC NMI ARI	0.562(±6.2e-03) 0.760(±2.3e-02) 0.006(±3.7e-03)	0.791(±2.7e-03) 0.959(±6.4e-04) 0.674(±5.5e-03)	9.3E+00	1.6E+03	1.0E+00
FaceV5	ACC NMI ARI	0.616(±1.5e-02) 0.809(±1.3e-02) 0.069(±2.2e-02)	0.771(±1.2e-02) 0.931(±3.7e-03) 0.673(±1.5e-02)	3.3E-01	5.4E+00	1.8E-01
CFPW	ACC NMI ARI	0.533(±1.1e-02) 0.732(±1.0e-02) 0.082(±1.6e-02)	0.800(±3.1e-03) 0.882(±2.3e-03) 0.356(±4.0e-02)	2.8E+00	3.0E+01	5.1E-01
CMU	ACC NMI ARI	0.289(±9.5e-03) 0.552(±7.1e-03) 0.173(±7.2e-03)	0.291(±9.0e-03) 0.565(±8.4e-03) 0.192(±7.9e-03)	4.8E-01	5.8E-01	1.6E-01
Colon	ACC NMI ARI	0.644(±3.4e-02) 0.065(±1.7e-02) 0.072(±4.0e-02)	0.690(±1.4e-01) 0.187(±1.8e-01) 0.214(±2.2e-01)	4.7E-04	4.3E-02	5.8E-03
CPLFW	ACC NMI ARI	0.331(±2.5e-03) 0.785(±3.8e-03) 0.019(±1.9e-03)	0.391(±2.5e-03) 0.815(±1.6e-03) 0.024(±1.6e-03)	7.0E+00	3.8E+03	9.3E-01
Dexter	ACC NMI ARI	0.600(±1.1e-16) 0.095(±4.4e-03) 0.039(±1.0e-05)	0.591(±3.8e-02) 0.050(±4.8e-02) 0.038(±2.5e-02)	1.8E-02	6.0E-02	2.1E-02
FERET	ACC NMI ARI	0.460(±7.3e-03) 0.737(±1.1e-02) 0.038(±1.0e-02)	0.633(±3.7e-03) 0.859(±6.1e-04) 0.531(±5.3e-03)	7.7E+00	1.5E+02	8.8E-01
GTdb	ACC NMI ARI	0.492(±3.3e-02) 0.670(±1.8e-02) 0.318(±4.0e-02)	0.516(±1.2e-02) 0.683(±4.3e-03) 0.373(±8.6e-03)	2.2E-02	9.6E-02	4.3E-02
LFW	ACC NMI ARI	0.419(±4.5e-03) 0.697(±1.5e-02) 0.010(±1.7e-03)	0.607(±2.4e-03) 0.895(±5.7e-04) 0.142(±1.2e-03)	1.1E+01	3.3E+03	9.5E-01
Madelon	ACC NMI ARI	0.509(±8.4e-04) 0.000(±4.0e-05) -0.000(±5.5e-05)	0.535(±3.6e-02) 0.007(±1.2e-02) 0.010(±1.6e-02)	3.7E-01	2.7E-01	9.4E-02
Mpeg7	ACC NMI ARI	0.464(±1.9e-02) 0.663(±1.4e-02) 0.235(±3.1e-02)	0.543(±8.0e-03) 0.716(±5.3e-03) 0.377(±1.7e-02)	8.8E-02	1.8E-01	6.2E-02
MUCT	ACC NMI ARI	0.625(±2.3e-02) 0.787(±2.9e-02) 0.087(±3.7e-02)	0.976(±4.7e-03) 0.992(±1.5e-03) 0.974(±4.7e-03)	7.7E-01	2.4E+00	1.0E-01
Olivetti	ACC NMI ARI	0.523(±2.5e-02) 0.723(±1.7e-02) 0.367(±3.1e-02)	0.522(±1.3e-02) 0.724(±8.1e-03) 0.388(±1.5e-02)	6.0E-03	6.1E-02	1.9E-02
Umist	ACC NMI ARI	0.443(±1.8e-02) 0.643(±1.9e-02) 0.335(±2.9e-02)	0.486(±2.2e-02) 0.663(±1.4e-02) 0.397(±2.3e-02)	1.5E-02	5.8E-02	1.8E-02
Yale	ACC NMI ARI	0.415(±3.4e-02) 0.478(±2.5e-02) 0.214(±2.9e-02)	0.427(±2.4e-02) 0.465(±1.8e-02) 0.204(±2.1e-02)	1.6E-03	3.8E-02	6.6E-03

3 Sensitivity analysis of the parameter

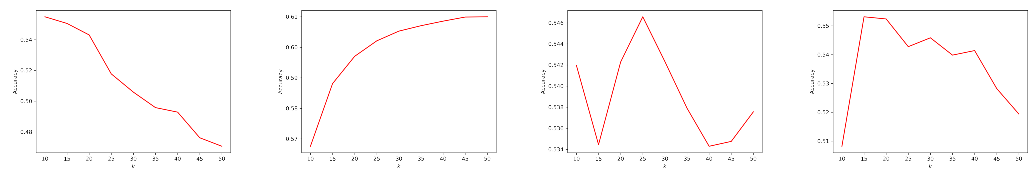
There are only one parameter in LKM, i.e., the number of nearest neighbors. It can be seen from Figure 1 that the clustering performance is largely affected by the value of k, which is a common phenomenon in graph-based clustering methods. How to find an appropriate value of k is a problem we will study in the future.



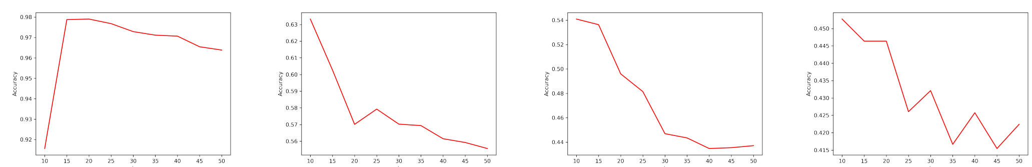
(a) CALFW (b) FaceV5 (c) CFPW (d) CMU



(e) Colon (f) CPLFW (g) Dexter (h) FERET



(i) GTdb (j) LFW (k) Madelon (l) Mpeg7



(m) MUCT (n) Olivetti (o) Umist (p) Yale

Figure 1: Performance vs. the number of nearest neighbors.

Table 3: Performance of LKM

	CALFW	FaceV5	CFPW	CMU	Colon	CPLFW	Dexter	FERET
k_{max}	50	15	30	15	30	45	35	50
Acc_{min}	0.836	0.926	0.668	0.226	0.625	0.449	0.584	0.535
Acc_{max}	0.851	0.938	0.745	0.301	0.748	0.451	0.62	0.665
std	0.005	0.004	0.023	0.026	0.035	0.001	0.012	0.041
	GTdb	LFW	Madelon	Mpeg7	MUCT	Olivetti	Umist	Yale
k_{max}	10	50	25	15	20	10	10	10
Acc_{min}	0.471	0.568	0.534	0.508	0.916	0.556	0.435	0.415
Acc_{max}	0.555	0.61	0.547	0.553	0.979	0.633	0.541	0.453
std	0.03	0.013	0.004	0.014	0.019	0.024	0.041	0.013

4 The standard deviations of the performance

Only one partition would be produced in RCC, if the parameter is fixed. Although multiple partitions would be produced in FINCH, only one is selected as the final clustering result. The number of clusters in the selected partition is nearest to that of the ground truth partition. Therefore, the standard deviations of the performance of RCC and FINCH are always zeros.

It can be seen from the results shown in Table 4 that the standard deviation of LKM is the smallest on most data sets, which verifies the robustness of LKM to initialization.

Table 4: The standard deviations of the performance

Datasets	Metrics	AGCI	FCDMF	Kmeans	KSUMS	SC	LKM
CALFW	ACC	±4.3e-03	±4.4e-03	±5.7e-03	±2.7e-03	±5.5e-03	±2.4e-03
	NMI	±2.0e-03	±1.3e-03	±2.0e-03	±5.3e-04	±1.9e-02	±5.5e-04
	ARI	±8.9e-03	±5.1e-03	±7.7e-03	±3.6e-03	±2.1e-03	±1.5e-02
CASIA	ACC	±1.4e-02	±7.9e-03	±1.3e-02	±1.1e-02	±1.4e-02	±5.1e-03
	NMI	±4.0e-03	±4.1e-03	±3.3e-03	±3.0e-03	±2.0e-02	±1.1e-03
	ARI	±2.5e-02	±1.7e-02	±1.8e-02	±1.4e-02	±2.3e-02	±6.0e-03
CFPW	ACC	±1.4e-02	±7.7e-03	±1.6e-02	±6.3e-03	±1.3e-02	±5.4e-03
	NMI	±7.5e-03	±3.6e-03	±7.7e-03	±2.1e-03	±1.2e-02	±3.5e-03
	ARI	±1.8e-02	±5.0e-03	±1.8e-02	±7.2e-03	±1.9e-02	±4.4e-02
CMU	ACC	±8.7e-03	±8.1e-03	±9.0e-03	±1.2e-02	±1.1e-02	±9.5e-03
	NMI	±7.2e-03	±9.6e-03	±7.9e-03	±8.1e-03	±6.1e-03	±9.6e-03
	ARI	±5.6e-03	±6.4e-03	±4.9e-03	±9.4e-03	±6.5e-03	±1.2e-02
Colon	ACC	±1.4e-01	±2.8e-02	±1.1e-01	±1.1e-01	±9.0e-03	±1.4e-01
	NMI	±1.8e-01	±4.6e-03	±1.3e-01	±1.5e-01	±1.6e-02	±2.0e-01
	ARI	±2.3e-01	±1.3e-02	±1.5e-01	±1.7e-01	±1.8e-02	±2.6e-01
CPLFW	ACC	±2.0e-03	±3.8e-03	±1.8e-03	±1.8e-03	±2.2e-03	±2.2e-03
	NMI	±2.1e-03	±3.9e-03	±1.5e-03	±3.3e-04	±4.2e-03	±1.8e-03
	ARI	±2.0e-05	±1.7e-04	±2.2e-05	±2.0e-03	±2.6e-03	±2.6e-03
Dexter	ACC	±5.1e-02	±0.0e+00	±3.8e-02	±3.2e-02	±6.7e-04	±1.2e-02
	NMI	±6.3e-02	±0.0e+00	±5.9e-02	±1.4e-02	±3.1e-04	±1.4e-02
	ARI	±2.7e-02	±1.4e-17	±2.3e-02	±1.9e-02	±3.6e-04	±8.4e-03
FERET	ACC	±5.8e-03	±5.1e-03	±8.9e-03	±2.1e-03	±7.5e-03	±3.3e-03
	NMI	±2.6e-03	±2.8e-03	±2.7e-03	±9.2e-04	±1.1e-02	±5.6e-04
	ARI	±9.6e-03	±9.4e-03	±1.2e-02	±2.3e-03	±1.0e-02	±5.2e-03
GTdb	ACC	±2.7e-02	±1.6e-02	±2.8e-02	±1.4e-02	±3.0e-02	±1.8e-02
	NMI	±1.5e-02	±8.7e-03	±1.3e-02	±7.9e-03	±2.2e-02	±8.6e-03
	ARI	±2.2e-02	±1.6e-02	±2.0e-02	±1.4e-02	±4.1e-02	±1.7e-02
LFW	ACC	±3.7e-03	±3.9e-03	±3.5e-03	±2.0e-03	±7.7e-03	±2.8e-03
	NMI	±6.1e-04	±8.7e-04	±7.4e-04	±4.8e-04	±2.1e-02	±5.0e-04
	ARI	±1.2e-03	±2.8e-03	±1.5e-03	±3.9e-04	±2.1e-03	±1.3e-03
Madelon	ACC	±2.9e-02	±1.1e-16	±3.4e-02	±2.8e-02	±2.8e-04	±2.1e-02
	NMI	±7.5e-03	±1.1e-19	±9.0e-03	±6.3e-03	±1.1e-05	±4.5e-03
	ARI	±1.0e-02	±5.4e-20	±1.2e-02	±8.8e-03	±1.5e-05	±6.2e-03
Mpeg7	ACC	±1.6e-02	±1.7e-02	±1.9e-02	±9.7e-03	±2.9e-02	±6.4e-03
	NMI	±9.3e-03	±7.8e-03	±1.4e-02	±6.7e-03	±2.9e-02	±5.8e-03
	ARI	±2.2e-02	±1.6e-02	±2.5e-02	±9.4e-03	±5.9e-02	±3.3e-02
MUCT	ACC	±2.6e-02	±1.5e-02	±3.3e-02	±2.3e-03	±2.4e-02	±3.3e-03
	NMI	±1.6e-02	±4.3e-03	±2.0e-02	±8.6e-04	±2.5e-02	±7.6e-04
	ARI	±1.4e-01	±1.7e-02	±1.5e-01	±2.6e-03	±3.2e-02	±3.1e-03
Olivetti	ACC	±2.9e-02	±1.7e-02	±3.2e-02	±2.6e-02	±2.4e-02	±1.4e-02
	NMI	±1.5e-02	±1.0e-02	±1.7e-02	±1.1e-02	±1.3e-02	±7.3e-03
	ARI	±2.7e-02	±1.7e-02	±3.2e-02	±2.9e-02	±2.6e-02	±1.6e-02
Umist	ACC	±1.9e-02	±2.5e-02	±2.2e-02	±3.1e-02	±2.3e-02	±3.1e-02
	NMI	±2.1e-02	±1.9e-02	±2.4e-02	±2.0e-02	±2.2e-02	±2.3e-02
	ARI	±1.9e-02	±2.5e-02	±2.7e-02	±3.2e-02	±2.7e-02	±4.4e-02
Yale	ACC	±2.4e-02	±1.8e-02	±3.1e-02	±3.5e-02	±3.6e-02	±2.5e-02
	NMI	±2.9e-02	±1.3e-02	±2.7e-02	±2.8e-02	±3.3e-02	±2.0e-02
	ARI	±2.7e-02	±1.2e-02	±2.8e-02	±3.4e-02	±3.3e-02	±2.3e-02