Clustering Plotted Data by Image Segmentation

Tarek Naous , Srinjay Sarkar , Abubakar Abid , James Zou

409410003 李文翔　　408410046 龔鈺閎

Outline

- Introduction
- Method
  - Core Algorithm
  - Optional processing
- Experimental results

Introduction

Visual Clustering

A new clustering method inspired by how humans cluster data

Faster

&

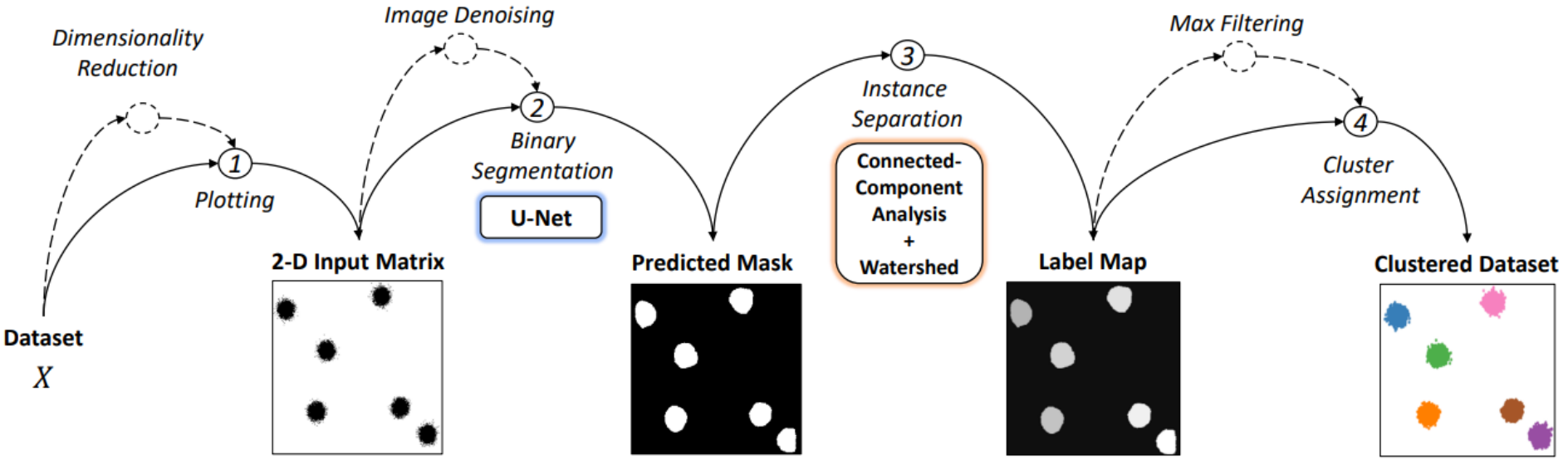
Human intuition

&

Neural Networks used

K means, GMM, DBSCAN

Method



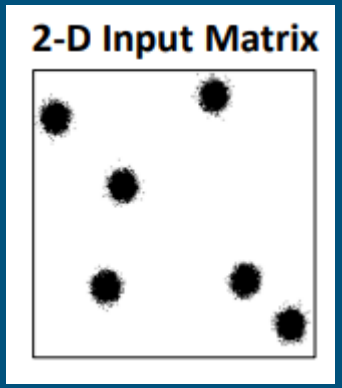
Diagrammatic view of our Visual Clustering algorithm

Core Algorithm

1. linearly shifting the values of both features to be 256*256

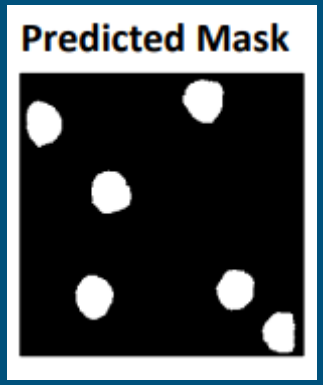
Core Algorithm

2. feed as an image input to the pre-trained binary segmentation model



Binary Segmentation

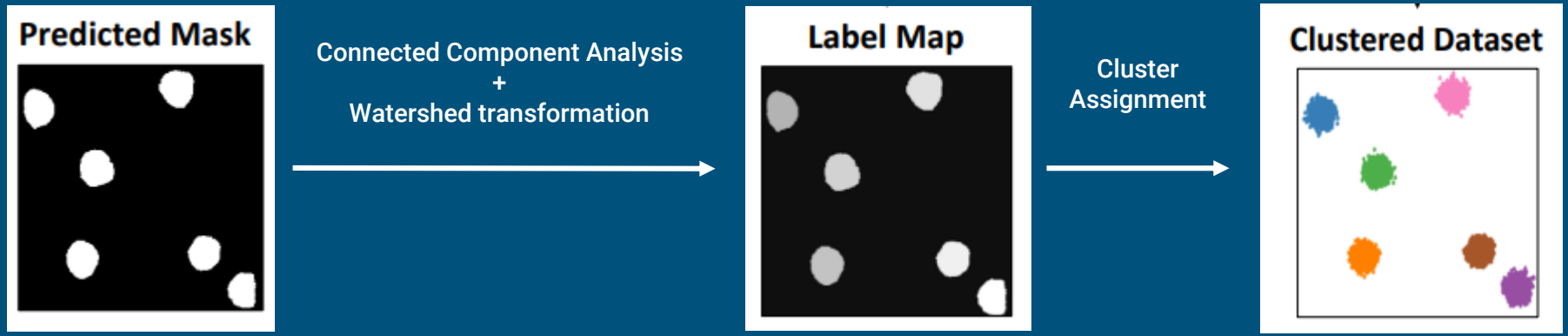
(U-Net)



Shows where cluster areas are located.

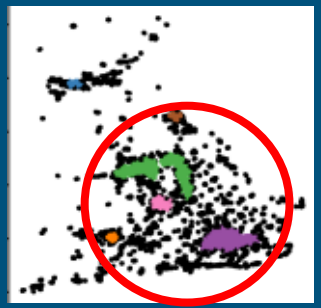
Core Algorithm

3. Separating the instances (or clusters) present in the binary mask.

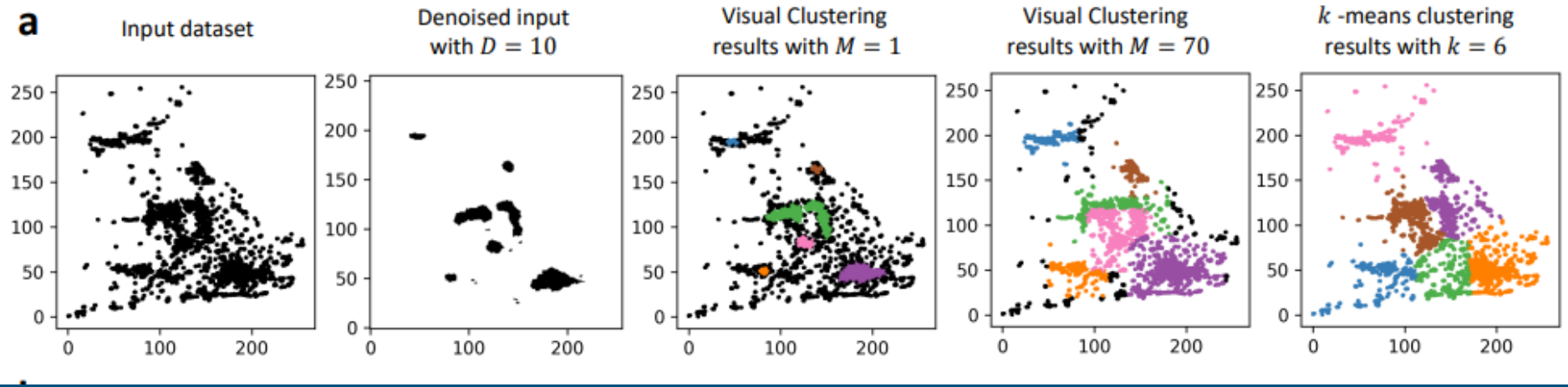


Optional Processing

- Inseparable Clusters
  - Close to real-world datasets
  - Median filter
  - Filter out low-density areas.
- Unassigned Points
  - Points fall in regions that are outside but near the cluster area
  - Ignored by binary segmentation model
  - Maximum filter
  - Increase the area of each cluster

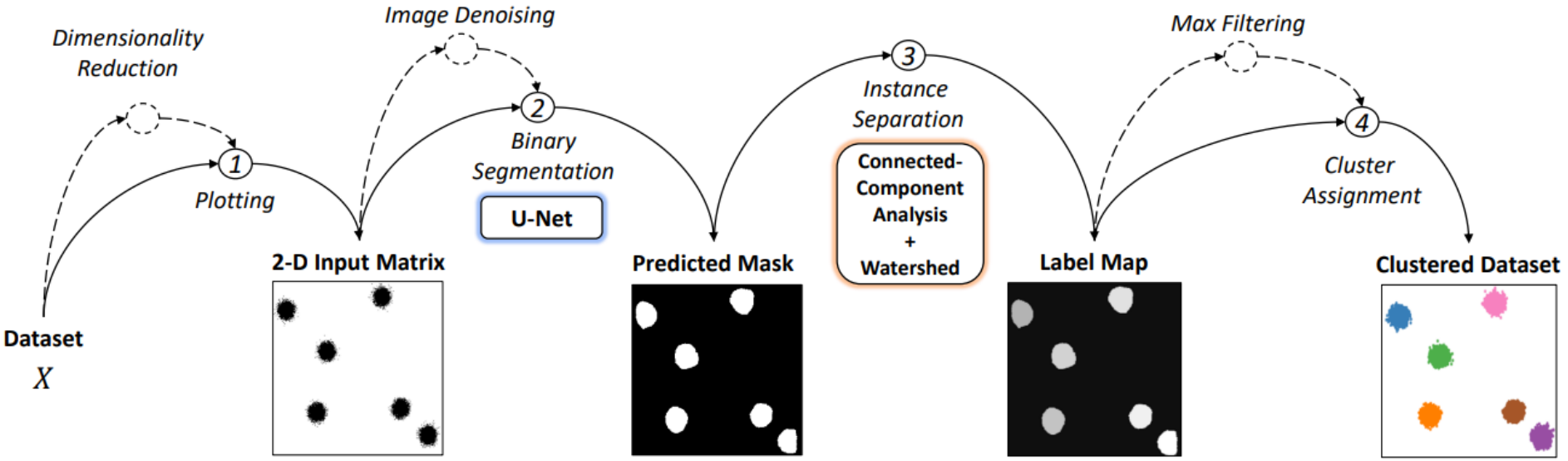


Optional Processing



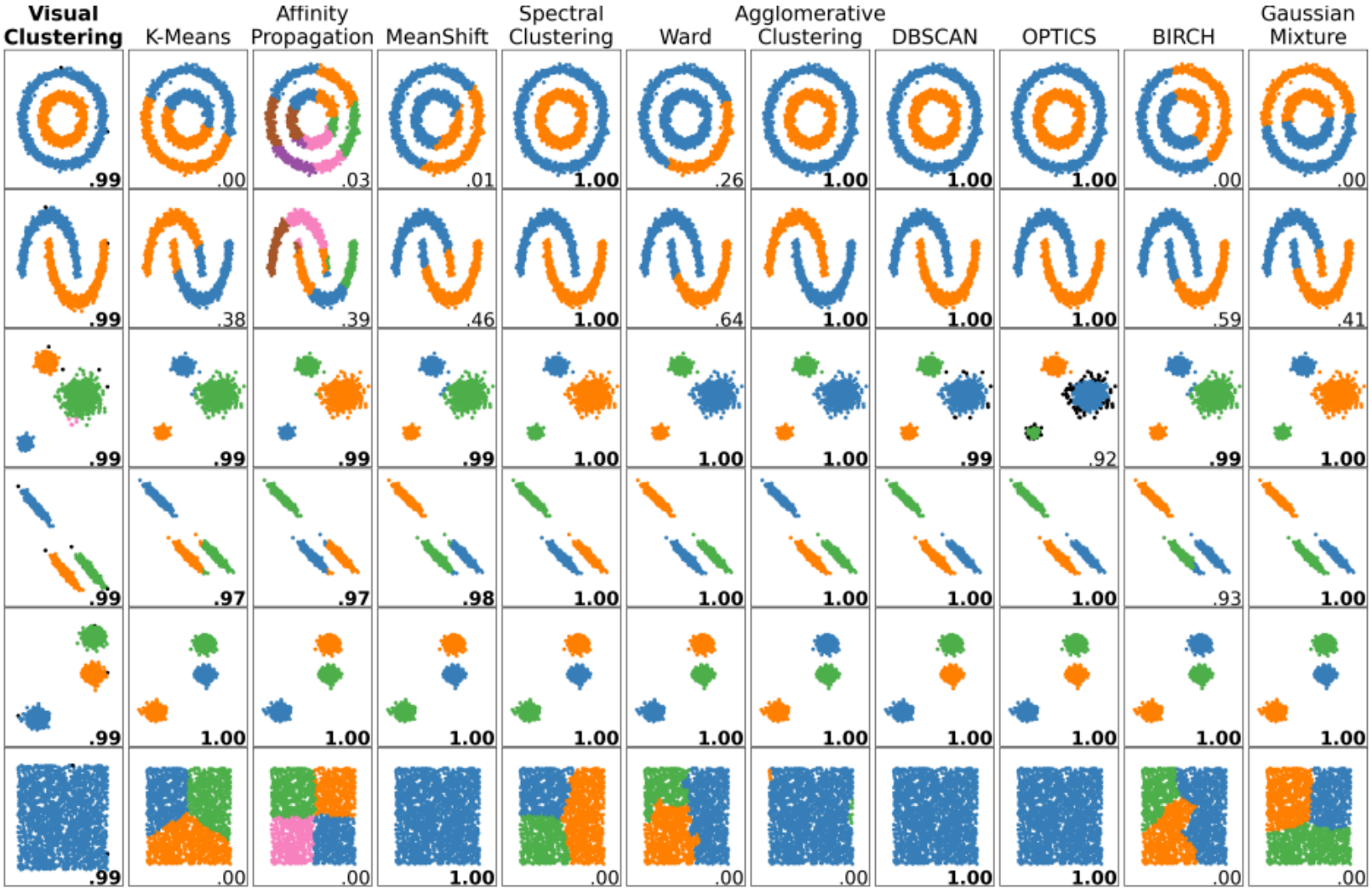
Apply Median Filter and Max Filter to input data

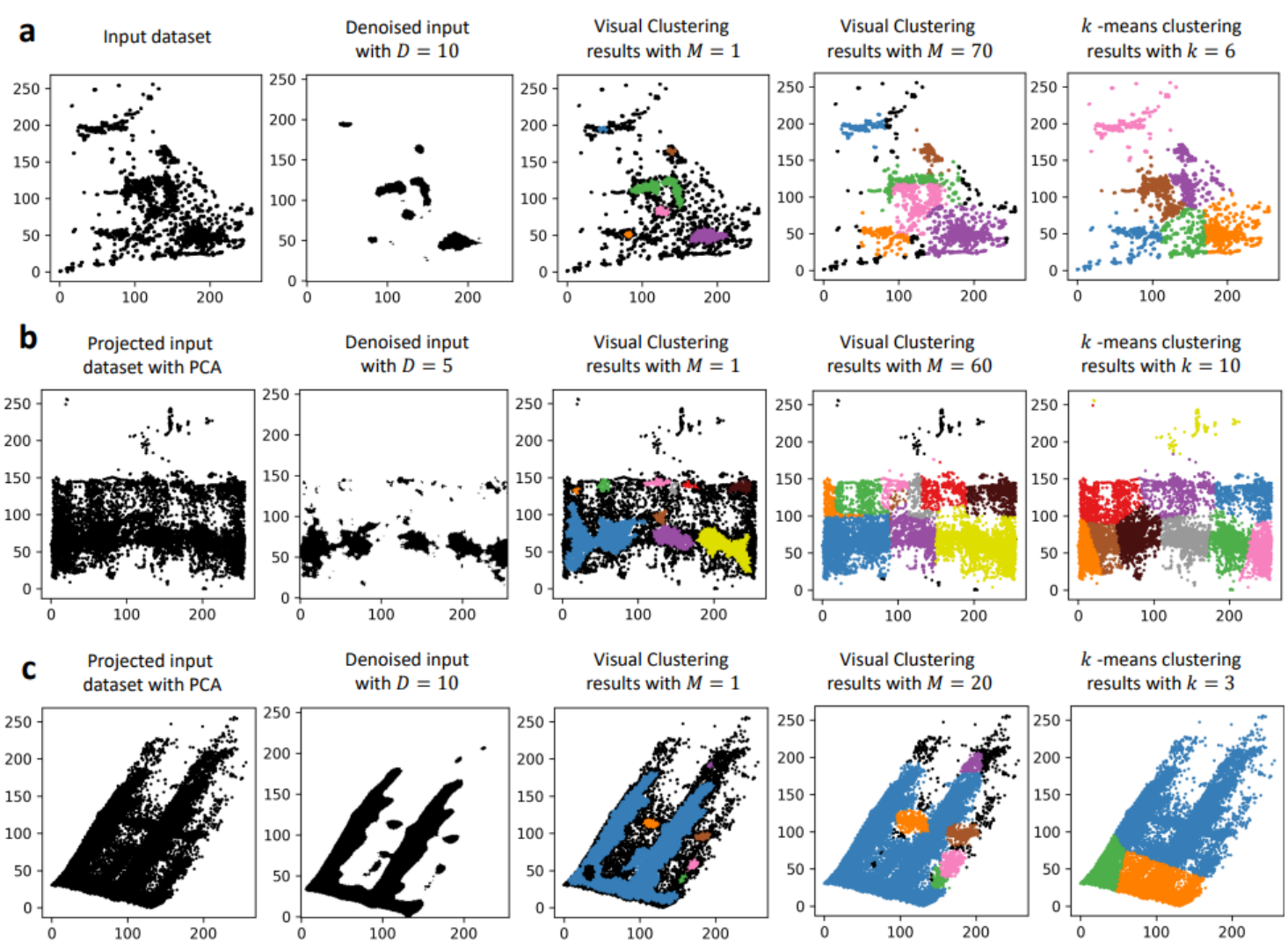
Method



Diagrammatic view of our Visual Clustering algorithm

Experimental Results


Visual Clustering
K-Means
Affinity Propagation
MeanShift
Spectral Clustering
Ward
Agglomerative Clustering
DBSCAN
OPTICS
BIRCH
Gaussian Mixture
.99 .00 .03 .01 1.00 .26 1.00 1.00 1.00 .00 .00
.99 .38 .39 .46 1.00 .64 1.00 1.00 1.00 .59 .41
.99 .99 .99 .99 1.00 1.00 1.00 .99 .92 .99 1.00
.99 .97 .97 .98 1.00 1.00 1.00 1.00 1.00 .93 1.00
.99 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00 1.00
.99 .00 .00 1.00 .00 .00 .00 1.00 1.00 .00 .00


a
Input dataset
Denoised input with $D = 10$
Visual Clustering results with $M = 1$
Visual Clustering results with $M = 70$
k -means clustering results with $k = 6$
b
Projected input dataset with PCA
Denoised input with $D = 5$
Visual Clustering results with $M = 1$
Visual Clustering results with $M = 60$
k -means clustering results with $k = 10$
c
Projected input dataset with PCA
Denoised input with $D = 10$
Visual Clustering results with $M = 1$
Visual Clustering results with $M = 20$
k -means clustering results with $k = 3$

Algorithm	Number of Samples					
	10K	50K	100K	500K	1M	2M
Visual Clustering	0.292	0.571	0.909	3.686	7.222	14.096
K-Means	0.155	0.541	1.103	5.470	9.519	18.959
Gaussian Mixture	0.089	0.358	0.726	3.047	5.949	11.962
DBSCAN	0.093	0.357	0.837	7.604	20.009	52.648
Affinity Propagation	175.35	∞	∞	∞	∞	∞
MeanShift	3.482	101.82	∞	∞	∞	∞
Spectral Clustering	0.052	0.509	0.796	7.455	20.001	53.559
Ward	1.994	27.965	93.564	∞	∞	∞
Agglomerative Clustering	1.177	12.154	39.886	∞	∞	∞
Optics	16.515	∞	∞	∞	∞	∞
BIRCH	1.298	7.390	14.320	∞	∞	∞

Thanks For Listening