A. Appendix

A.1. Dataset

We used 4 types of datasets: Mouse retina transcriptomes (Macosko et al., 2015; Poličar et al., 2019), Fashion MNIST (Xiao et al., 2017), EMNIST Letters (Cohen et al., 2017), and CIFAR-10 (Krizhevsky et al., 2009). The details on the datasets and the numbers used in training and evaluations are shown in Table 1. The Mouse retina transcriptomes dataset consists of PCA projections of single-cell transcriptome data collected from mouse retina. The Fashion MNIST dataset consists of 10 types of clothing such as shirts, sneakers, and ankle boots. The EMNIST Letters dataset is a set of handwritten alphabet characters of 26 classes. The CIFAR-10 dataset consists of real-world images in 10 classes such as airplanes, automobiles, birds and cats. Except for Mouse retina transcriptomes dataset, the distribution of classes is uniform. The class distribution of the Mouse retina transcriptomes dataset is shown in Table 2. The distribution of each dataset is visualized in Figure 1. For visualization, each dataset was projected into a 2-dimensional embedding space using Parametric UMAP.

At least 10,000 samples of each dataset were used as test dataset for the performance evaluation of MPART. We also used 30% of the training dataset to train the Parametric UMAP (Sainburg et al., 2020) for dimensionality reduction. MPART was trained using only 10,000 randomly sampled data from the remaining 70% of the training dataset.

Table 1. Datasets used to evaluate our proposed method and the number of samples used in training and testing.

Dataset	Num. of Classes	Dimensions	Parametric UMAP	MPART	
			(Train)	Train	Test
Mouse retina transcriptomes	12	50	10,442	24,366	10,000
Fashion MNIST	10	28x28	18,000	42,000	10,000
EMNIST Letters	26	28x28	37,440	87,360	20,800
CIFAR-10	10	32x32x3	15,000	35,000	10,000

Table 2. Distribution of classes in Mouse retina transcriptomes dataset.

Label	Class Name	Parametric UMAP	MPART		Total
		(Train)	Train	Test	
0	Cones	425	1,029	414	1,868
1	Horizontal cells	67	131	54	252
2	Pericytes	16	37	10	63
3	Amacrine cells	1,041	2,314	1,071	4,426
4	Retinal ganglion cells	113	228	91	432
5	Fibroblasts	21	50	14	85
6	Vascular endothelium	54	139	59	252
7	Muller glia	363	893	368	1,624
8	Astrocytes	15	29	10	54
9	Microglia	15	41	11	67
10	Rods	6,855	16,047	6,498	29,400
11	Bipolar cells	1,457	3,428	1,400	6,285
	Total	10,442	24,366	10,000	44,808

The datasets can be downloaded from the following links.

- Mouse retina transcriptomes: `http://file.biolab.si/opentsne/macosko_2015.pkl.gz`
- Fashion MNIST: `https://github.com/zalandoresearch/fashion-mnist`
- EMNIST Letters: `https://github.com/aurelienduarte/emnist`
- CIFAR-10: `https://www.cs.toronto.edu/~kriz/cifar.html`

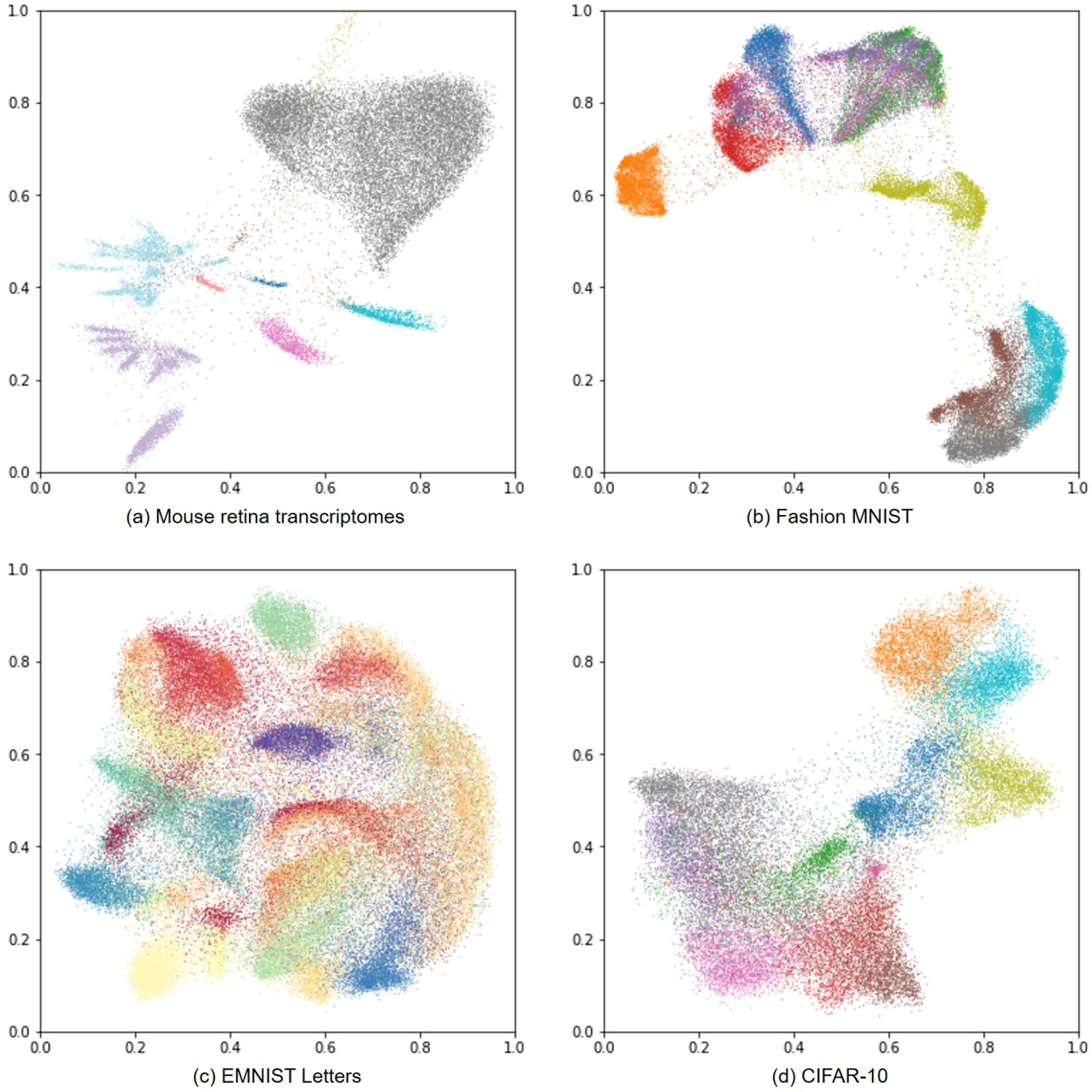


Figure 1. Visualization of the distribution of each dataset using Parametric UMAP.

A.2. Implementation Details

Parametric UMAP. In order for MPART to cluster input data and properly generate the topological graph, it is necessary to represent high-dimensional input data in a low-dimensional latent space. So, the performance of the dimensionality reduction can affect the overall performance. We extracted the 2048-dimensional feature vector of the CIFAR-10 dataset using the BYOL (Grill et al., 2020) model pretrained with the ImageNet (Deng et al., 2009) dataset. Then we projected the input data into a 4-dimensional embedding space using the pretrained Parametric UMAP for each dataset. The Parametric UMAP was trained using the parameters shown in Table 3, and the training was conducted for 50 epochs. A convolutional neural network (CNN) was used as an embedding algorithm of Parametric UMAP for Fashion MNIST and EMNIST Letters. For the Mouse retina transcriptome and CIFAR-10, a 3-layer multi-layer perceptron (MLP) was used with 100-neurons per layer. More information about Parametric UMAP can be found at the following link: `https://umap-learn.readthedocs.io/`.

Table 3. Summary of important UMAP parameters used in this study.

NAME	VALUE	DESCRIPTION
N_NEIGHBORS	15	CONTROLS HOW UMAP BALANCES LOCAL VERSUS GLOBAL STRUCTURE IN THE DATA
MIN_DIST	0.1	CONTROLS HOW TIGHTLY UMAP IS ALLOWED TO GROUP POINTS TOGETHER
N_COMPONENTS	4	CONTROLS THE DIMENSIONALITY OF THE REDUCED DIMENSION SPACE
METRIC	EUCLIDEAN	THE METHOD FOR DISTANCE MEASURE IN THE AMBIENT SPACE OF THE INPUT DATA

MPART. We chose the parameters of the MPART empirically. The parameters used in the model are shown in Table 5, and the same values were used for all datasets and experiments. The value of ρ should be determined by considering the number of dimensions of the input space and computational power. The larger ρ value, the more MPART nodes are created, which allows for finer clustering, but increases the required computational power. The value of k_e should be set according to how many queries are possible. A small value of k_e should be used in situations where enough queries are possible. Since we dealt with very few labeled data, we used a relatively large value of $k_e = 1.0$. The value of k_d should be set small enough according to the total amount of training data including unlabeled data. We also show in Table 5 the parameters used in the competitive models.

Table 4. Model parameters used in the MPART.

SYMBOL	VALUE	DESCRIPTION
α	0.01	CHOICE PARAMETER
β	0.5	LEARNING RATE FOR NODE WEIGHTS
ρ	0.95	VIGILANCE PARAMETER
δ	0.1	PROPAGATION RATE FOR MESSAGE PASSING
τ	0.7	WEIGHT FOR QUERY SELECTION SCORE
k_e	1.0	SENSITIVITY FOR EPISTEMIC UNCERTAINTY
k_d	0.01	SENSITIVITY FOR DENSITY-WEIGHTED QUERY SELECTION SCORE
L	3	NUMBER OF LAYERS FOR MESSAGE PASSING

Table 5. Model parameters used in LPART and A-SOINN.

MODEL	SYMBOL	VALUE	DESCRIPTION
LPART	α	0.01	CHOICE PARAMETER
	β	0.5	LEARNING RATE FOR NODE WEIGHTS
	ρ	0.95	VIGILANCE PARAMETER
	δ	1.0	PROPAGATION RATE
A-SOINN	λ	500	PERIOD FOR NODE REMOVAL AND CLUSTERING
	age_{max}	30	MAXIMUM AGE OF EDGE
	α	2.0	SMOOTHING PARAMETER FOR GROUPING

A.3. Parameter Search for δ and τ

The choice of values of the propagation rate δ in message passing and the weight τ for query selection score affects the performance only marginally, unless δ is set to 0. Table 6 shows the classification performance on CIFAR-10 according to the change of δ and τ. The 'Explorer' strategy and 1/500 query frequency were used. The mean and standard deviation were drawn from results of 10 trials.

Table 6. The classification accuracy (mean $\pm$ std) of our model with various δ and τ. For each L, **boldface** indicates the top three and the blue text indicates the bottom three.

NUMBER OF LAYERS	δ	τ						
		0.0	0.1	0.3	0.5	0.7	0.9	1.0
$L=0$	0.0	12.5±0.4	12.6±0.3	12.8±0.3	12.7±0.4	12.6±0.5	12.8±0.6	12.8±0.4
$L=1$	0.1	39.2±1.9	38.6±1.8	39.5±1.9	39.3±2.0	39.3±1.9	35.8±1.7	33.6±2.6
	0.3	39.2±1.9	38.7±2.2	39.7±2.2	39.2±1.8	38.4±2.4	38.4±1.9	34.2±1.6
	0.5	39.3±1.4	39.5±2.3	39.3±1.6	39.1±1.4	38.4±1.7	38.3±2.2	37.2±1.5
	0.7	38.2±2.4	38.7±1.7	**39.8**±3.1	37.8±2.3	**39.7**±1.4	37.8±1.8	38.7±2.8
	0.9	39.3±2.4	38.1±2.5	38.8±2.6	37.0±2.9	37.2±2.2	38.8±2.0	38.4±1.6
	1.0	39.1±3.3	38.4±2.4	37.1±3.3	**40.3**±2.0	38.1±2.0	36.7±3.9	38.2±2.3
$L=3$	0.1	61.4±3.8	61.1±1.7	**63.4**±2.4	59.3±3.6	58.9±2.9	56.7±3.4	55.2±3.5
	0.3	59.5±5.8	58.8±3.8	**61.7**±3.3	59.3±3.9	59.3±3.5	58.5±5.9	59.3±3.0
	0.5	57.9±5.5	58.8±4.0	59.8±4.9	60.5±3.4	60.5±2.8	59.2±3.6	57.5±4.6
	0.7	60.4±2.5	56.3±5.2	60.4±3.7	59.9±3.1	58.7±3.1	59.9±3.8	60.0±4.2
	0.9	58.0±4.8	56.4±5.3	58.4±5.9	58.9±3.1	60.0±3.8	60.0±3.0	**61.5**±3.9
	1.0	58.9±4.3	58.6±4.5	59.2±5.2	56.6±5.9	61.0±3.5	59.7±3.9	59.5±3.7
$L=5$	0.1	54.6±6.6	55.1±7.8	60.1±2.7	60.1±2.6	60.0±4.0	57.4±2.3	60.0±3.4
	0.3	50.6±4.4	51.2±6.6	54.3±3.8	**61.5**±2.1	**60.5**±4.5	60.3±4.3	**60.8**±3.7
	0.5	49.7±7.3	53.8±4.6	56.1±5.2	59.1±3.7	58.9±3.5	59.1±4.9	58.9±3.6
	0.7	48.2±4.9	52.0±2.9	53.8±8.0	57.5±5.3	58.2±3.8	60.4±4.4	60.5±2.2
	0.9	47.2±6.3	45.7±8.6	52.1±4.4	58.0±4.1	56.3±5.7	57.2±6.3	56.8±5.8
	1.0	49.1±4.3	51.0±4.0	51.3±4.2	53.5±7.9	56.2±10.4	54.0±7.4	58.9±2.1

A.4. Additional Experimental Results

Online Topology Learning. Figure 2 shows an example result of MPART's online topology learning. MPART continuously learns the distribution and topology structure of sequential input data without forgetting.

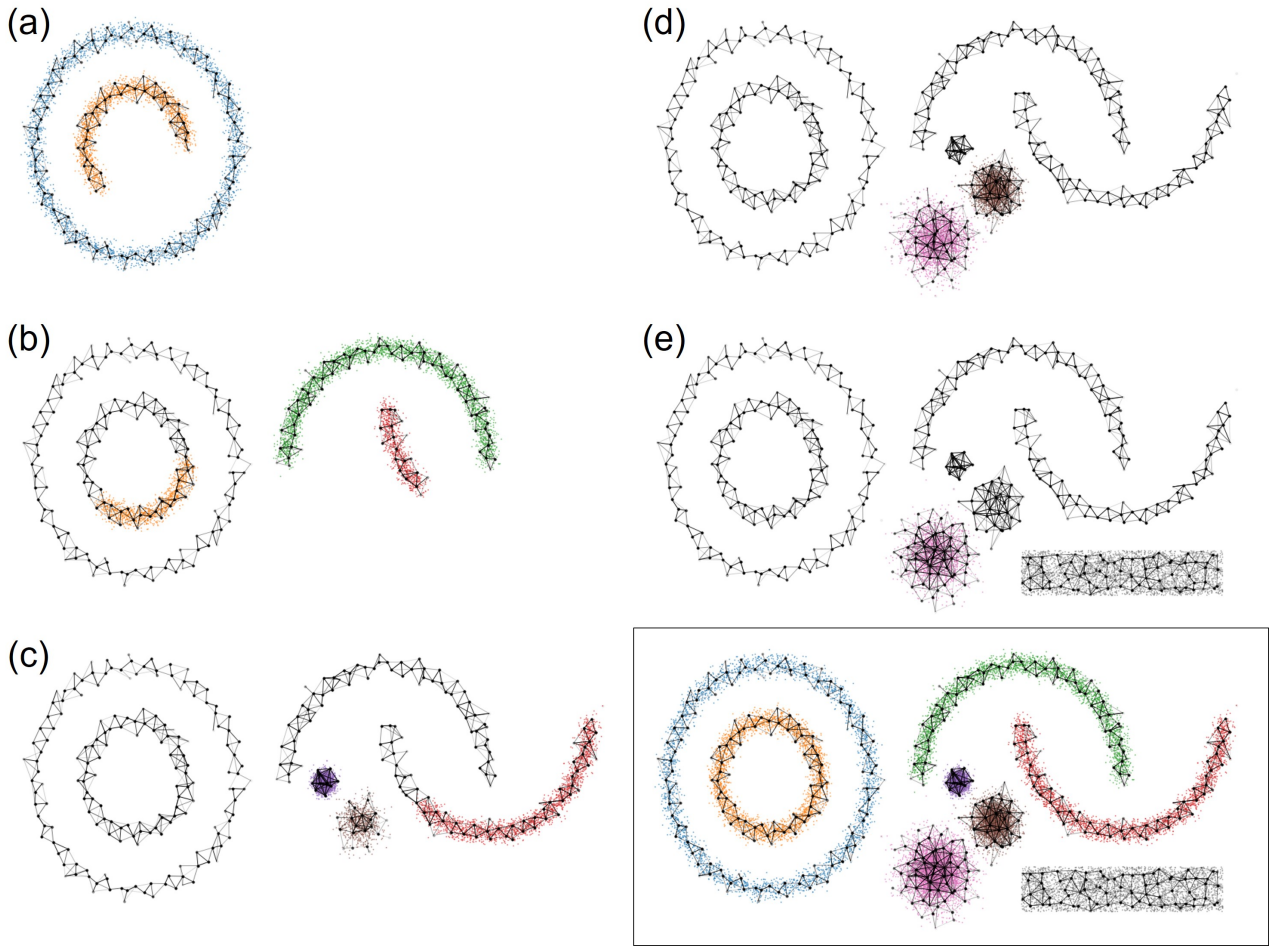


Figure 2. An example result of topology learning. (a) $\sim$ (e) is the process of topology learning, and the lower right figure is the final learning result. The colored scattered points refer to newly entered data and the intensity of the topological graph represents the density of nodes and edges. All data samples were input only once, one by one.

Analysis of the Generated Topological Graph. For a deeper analysis, it would be interesting to see some statistics of the topological graph. In Table 7, we show the numbers of generated nodes, co-activations for each task, and co-activated nodes per winning node and per sample. We also analyzed the average number of neighboring nodes per node, weights of edges, and the nodes that do not have any co-activated node, in which case will be indicated by nodes without edges.

Table 7. Detailed analysis of the trained topological graph. The mean and standard deviation are drawn from 10 trials for each dataset.

ITEM	MOUSE RETINA TRANSCRIPTOMES	FASHION MNIST	EMNIST LETTERS	CIFAR-10
TOTAL # OF NODES	290.3±5.3	256.2±4.3	1165.5±8.0	984.6±10.5
TOTAL # OF CO-ACTIVATIONS	15873.2±399.5	15533.1±404.4	9222.8±151.4	11230.6±210.6
AVG. # OF CO-ACTIVATIONS PER NODE	54.69±1.68	60.63±1.48	7.91±0.15	11.41±0.24
AVG. # OF CO-ACTIVATIONS PER SAMPLE	1.59±0.04	1.55±0.04	0.92±0.02	1.12±0.02
AVG. # OF NEIGHBORING NODES PER NODE	7.46±0.12	7.51±0.22	4.16±0.06	6.02±0.08
TOTAL # OF NODES W/O EDGES	11.00±3.43	6.70±2.11	34.20±6.07	19.50±3.03
AVG. # OF WEIGHTS OF EDGES	0.179±0.009	0.193±0.007	0.172±0.003	0.150±0.003

Computational Cost. We measured the run-time of our Python implementation on a 3.8 GHz CPU machine. Tables 8 and 9 show the number of nodes generated by MPART and the time required for training and inference according to the number of training samples.

Table 8. The number of nodes generated by MPART according to the query selection frequency and the number of training samples. The number of layer $L = 3$ and the 'Explorer' strategy were used. The mean and standard deviation are drawn from 10 trials for each dataset.

QUERY SELECTION FREQUENCY	NUMBER OF TRAINING DATA	MOUSE RETINA TRANSCRIPTOMES	FASHION MNIST	EMNIST LETTERS	CIFAR-10
1 / 1000	10,000	290.2±5.0	254.8±4.3	1170.6±9.4	970.7±7.4
	15,000	349.0±5.8	301.0±4.1	1504.4±9.6	1260.6±6.7
	20,000	396.1±4.4	338.1±4.8	1810.3±6.9	1505.7±7.8
1 / 500	10,000	290.5±4.5	254.3±5.3	1167.1±8.9	980.8±5.9
	15,000	348.6±6.0	301.6±3.7	1511.6±11.3	1262.1±3.8
	20,000	394.9±5.7	337.7±5.3	1810.6±8.5	1510.7±7.3
1 / 100	10,000	290.5±4.6	255.0±4.5	1163.5±8.2	976.4±8.5
	15,000	347.3±4.6	301.3±5.7	1510.7±10.8	1262.7±9.0
	20,000	398.1±5.7	337.8±4.2	1807.6±8.4	1508.6±7.6

Table 9. Comparison of time taken by MPART for training and inference according to query selection frequency and the number of training samples. The number of layer $L = 3$ and the 'Explorer' strategy were used. The mean and standard deviation are drawn from 10 trials for each dataset. (unit: sec)

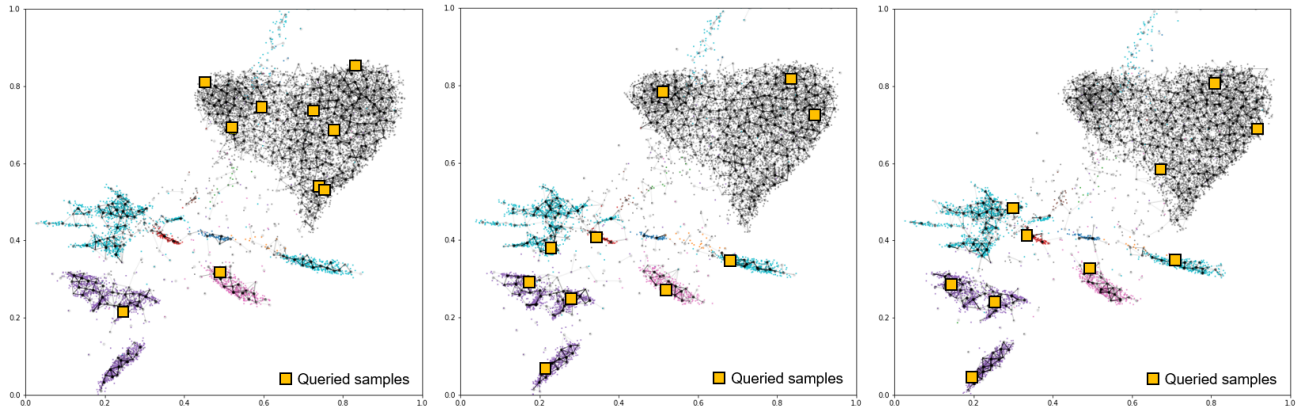
QUERY SELECTION FREQUENCY	NUMBER OF TRAINING DATA	MOUSE RETINA TRANSCRIPTOMES	FASHION MNIST	EMNIST LETTERS	CIFAR-10
1 / 1000	10,000	9.24±0.14	9.03±0.35	13.25±0.58	11.21±0.15
	15,000	14.46±0.14	13.98±0.11	25.54±0.17	20.70±0.23
	20,000	19.49±0.13	19.12±0.11	43.71±0.33	34.38±0.26
1 / 500	10,000	9.08±0.10	8.90±0.07	13.03±0.11	11.43±0.20
	15,000	14.55±0.11	13.93±0.05	25.84±0.32	20.90±0.21
	20,000	19.83±0.08	19.12±0.17	43.60±0.41	34.59±0.19
1 / 100	10,000	9.15±0.09	8.87±0.06	13.26±0.17	11.55±0.17
	15,000	13.93±0.05	13.68±0.07	25.93±0.26	21.11±0.16
	20,000	18.78±0.06	18.71±0.10	43.65±0.39	34.72±0.26

Various Query Frequencies. Table 10 summarizes the performance evaluation of MPART and competitive models on four datasets according to various query selection frequencies. Given the same total query budget, MPART shows similar performance regardless of query frequencies.

Table 10. Comparison of classification accuracy (mean ± std) between MPART and the competitive models. The number of layer $L = 3$ and the 'Explorer' strategy were used for MPART. The mean and standard deviation are drawn from 30 trials for each dataset. (unit : %)

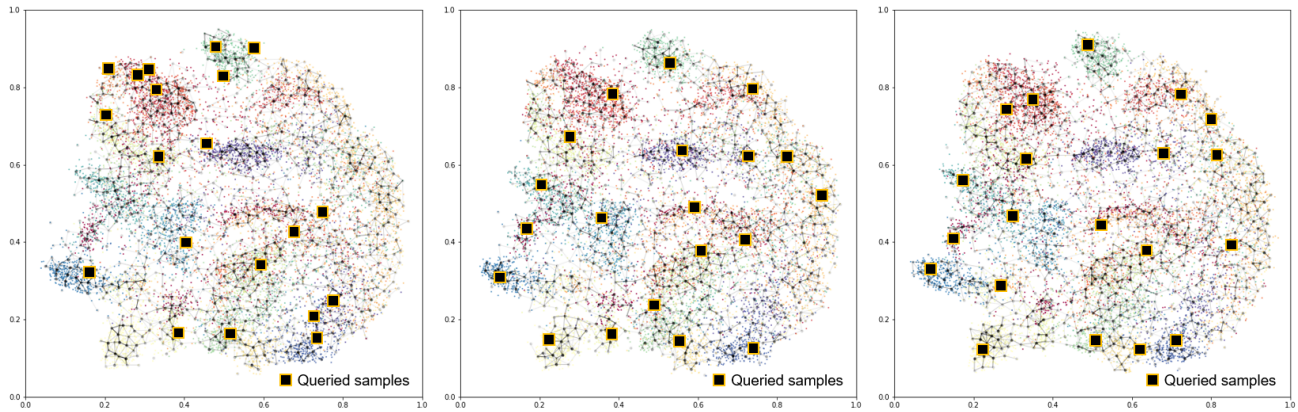
Total Query Budget	Query Selection Frequency	Model	Mouse retina transcriptomes	Fashion MNIST	EMNIST Letters	CIFAR-10
10	1 / 1000	LPART	77.1±7.4	41.5±4.4	13.3±3.2	27.4±5.0
		A-SOINN	88.1±13.0	45.2±7.0	16.9±4.1	37.7±6.8
		MPART	**91.2**±5.6	**56.4**±5.8	**26.4**±2.7	**52.6**±4.9
	2 / 2000	LPART	83.8±6.0	37.6±4.9	14.8±3.1	28.0±3.4
		A-SOINN	90.9±5.0	45.4±7.9	17.8±2.7	42.7±7.0
		MPART	**91.7**±4.0	**56.8**±3.7	**25.8**±1.9	**54.8**±4.0
	5 / 5000	LPART	84.3±9.9	41.6±7.4	15.4±2.8	30.1±3.7
		A-SOINN	**93.6**±2.6	49.8±5.2	21.5±2.5	46.2±6.2
		MPART	92.2±4.4	**55.0**±3.7	**26.2**±1.7	**54.4**±5.3
	10 / 10000	LPART	81.3±4.2	38.4±6.4	15.6±3.7	36.0±6.5
		A-SOINN	**94.4**±2.0	50.3±4.7	23.2±2.6	50.8±4.6
		MPART	91.8±4.7	**54.6**±4.6	**25.4**±2.1	**56.0**±4.3
20	1 / 500	LPART	85.4±4.3	49.3±7.4	21.2±2.7	42.0±5.7
		A-SOINN	91.0±5.4	50.3±6.0	25.6±4.8	44.2±7.0
		MPART	**93.5**±2.6	**61.4**±3.1	**35.9**±3.4	**59.2**±2.6
	2 / 1000	LPART	89.4±4.8	52.3±6.3	20.7±2.9	42.1±7.9
		A-SOINN	92.0±3.7	51.9±4.6	26.0±4.0	49.3±6.1
		MPART	**94.3**±1.8	**59.8**±4.1	**35.7**±3.4	**59.1**±3.1
	4 / 2000	LPART	86.5±5.5	54.8±4.0	22.9±3.6	42.8±6.9
		A-SOINN	93.3±3.2	55.0±5.1	28.4±3.9	51.9±4.9
		MPART	**93.5**±2.9	**60.8**±4.6	**35.7**±2.8	**60.0**±3.7
	10 / 5000	LPART	89.5±6.9	50.1±7.0	19.3±3.0	41.7±6.2
		A-SOINN	**94.2**±2.3	56.4±4.7	31.9±2.8	57.8±3.8
		MPART	93.9±1.8	**59.7**±4.5	**36.7**±2.1	**60.4**±3.2
	20 / 10000	LPART	89.8±4.5	49.4±4.4	23.0±3.0	45.4±6.8
		A-SOINN	**94.6**±1.8	59.8±4.0	36.2±3.1	**60.9**±4.0
		MPART	93.4±2.5	**60.1**±3.8	**36.9**±4.3	59.9±4.0
100	1 / 100	LPART	94.4±0.9	65.8±1.8	37.7±2.3	58.2±1.9
		A-SOINN	91.5±7.0	55.7±6.8	29.9±4.9	50.9±7.9
		MPART	**95.7**±0.8	**67.3**±1.7	**47.9**±1.8	**67.0**±1.4
	10 / 1000	LPART	94.5±1.3	64.4±1.8	37.8±2.4	59.8±2.5
		A-SOINN	93.9±1.6	63.0±3.8	39.5±3.0	60.3±5.8
		MPART	**95.6**±0.9	**67.5**±1.4	**47.7**±1.7	**67.4**±1.1
	20 / 2000	LPART	93.1±1.4	64.7±2.2	38.8±2.4	60.0±2.0
		A-SOINN	93.6±2.3	63.9±3.0	40.6±3.8	60.6±4.3
		MPART	**95.7**±0.9	**66.7**±1.9	**47.6**±1.3	**66.8**±1.5
	50 / 5000	LPART	93.9±1.6	63.4±2.4	37.2±2.6	59.4±1.7
		A-SOINN	93.6±2.4	64.3±2.8	42.2±3.3	60.5±3.9
		MPART	**95.4**±1.1	**67.3**±1.5	**47.8**±1.9	**66.7**±1.5
	100 / 10000	LPART	94.9±1.1	64.1±2.3	37.7±1.8	58.4±2.8
		A-SOINN	94.2±1.3	64.8±3.1	41.9±2.8	63.6±4.1
		MPART	**95.5**±0.7	**67.6**±1.2	**47.3**±1.3	**66.7**±1.3

Visualization of Training Results. In order to evaluate the capability of our model for online active semi-supervised learning, we visualized the generated topological graph and query distributions. The visualization results on the Mouse retina transcriptomes and EMNIST Letters datasets are shown in Figures 3 and 4. The visualization shows that the Memory and Explorer strategies perform better than the Random strategy, as implied by the evenly distributed queried samples.



(a) Random (b) Memory (c) Explorer

Figure 3. The visualization of training results on the Mouse retina transcriptomes dataset according to query selection strategy. The number of layer $L = 3$ and 1/1000 query frequency were used. The colored scattered points refer to input samples and the intensity of the topological graph represents the density of nodes and edges. The queried samples are indicated in yellow boxes. In the case of Random strategy, we can observe more queried samples in the class represented by the black dots. In contrast, the queried samples are more evenly distributed for each class when using the 'Memory' or 'Explorer' strategy. All data samples were input only once, one for each time step. The input data was projected into a 2-dimensional embedding space for visualization.



(a) Random (b) Memory (c) Explorer

Figure 4. The visualization of training results on the EMNIST Letters dataset according to query selection strategy. The number of layer $L = 3$ and 1/500 query frequency were used. The colored scattered points refer to input samples and the intensity of the topological graph represents the density of nodes and edges. The queried samples are indicated in black boxes. In the case of Random strategy, we can observe more queried samples in certain classes than others. In contrast, the queried samples are more evenly distributed for each class when using the 'Memory' or 'Explorer' strategy. All data samples were input only once, one for each time step. The input data was projected into a 2-dimensional embedding space for visualization.

References

Cohen, G., Afshar, S., Tapson, J., and Van Schaik, A. Emnist: Extending mnist to handwritten letters. In *2017 International Joint Conference on Neural Networks (IJCNN)*, pp. 2921–2926. IEEE, 2017.

Deng, J., Dong, W., Socher, R., Li, L.-J., Li, K., and Fei-Fei, L. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pp. 248–255. Ieee, 2009.

Grill, J.-B., Strub, F., Altché, F., Tallec, C., Richemond, P. H., Buchatskaya, E., Doersch, C., Pires, B. A., Guo, Z. D., Azar, M. G., et al. Bootstrap your own latent: A new approach to self-supervised learning. *arXiv preprint arXiv:2006.07733*, 2020.

Krizhevsky, A., Hinton, G., et al. Learning multiple layers of features from tiny images. Technical report, Citeseer, 2009.

Macosko, E. Z., Basu, A., Satija, R., Nemesh, J., Shekhar, K., Goldman, M., Tirosh, I., Bialas, A. R., Kamitaki, N., Martersteck, E. M., et al. Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets. *Cell*, 161(5):1202–1214, 2015.

Poličar, P. G., Stražar, M., and Zupan, B. opentsne: a modular python library for t-sne dimensionality reduction and embedding. *BioRxiv*, pp. 731877, 2019.

Sainburg, T., McInnes, L., and Gentner, T. Q. Parametric umap: learning embeddings with deep neural networks for representation and semi-supervised learning. *arXiv preprint arXiv:2009.12981*, 2020.

Xiao, H., Rasul, K., and Vollgraf, R. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv:1708.07747*, 2017.