IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 22 February 2023, accepted 6 March 2023, date of publication 15 March 2023, date of current version 31 March 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3257427

RESEARCH ARTICLE

SpectralMAP: Approximating Data Manifold With Spectral Decomposition

KOSHI WATANABE[1], (Graduate Student Member, IEEE), KEISUKE MAEDA[2], (Member, IEEE), TAKAHIRO OGAWA[2], (Senior Member, IEEE), AND MIKI HASEYAMA[2], (Senior Member, IEEE)

[1]Graduate School of Information Science and Technology, Hokkaido University, Sapporo 060-0814, Japan
[2]Faculty of Information Science and Technology, Hokkaido University, Sapporo 060-0814, Japan

Corresponding author: Miki Haseyama (mhaseyama@lmd.ist.hokudai.ac.jp)

This work was supported in part by the JSPS KAKENHI Grant Number JP21H03456, and in part by the AMED Grant Number JP22zf0127004h0002.

ABSTRACT Dimensionality reduction is widely used to visualize complex high-dimensional data. This study presents a novel method for effective data visualization. Previous methods depend on local distance measurements for data manifold approximation. This leads to unreliable results when a data manifold locally oscillates because of some undesirable effects, such as noise effects. In this study, we overcome this limitation by introducing a dual approximation of a data manifold. We roughly approximate a data manifold with a neighborhood graph and prune it with a global filter. This dual scheme results in local oscillation robustness and yields effective visualization with explicit global preservation. We consider a global filter based on principal component analysis frameworks and derive it with the spectral information of the original high-dimensional data. Finally, we experiment with multiple datasets to verify our method, compare its performance to that of state-of-the-art methods, and confirm the effectiveness of our novelty and results.

INDEX TERMS Data visualization, dimensionality reduction, spectral-based filtering.

I. INTRODUCTION

Real-world data are typically found in a complex high-dimensional data space [1]. One approach for analyzing these data is to visualize their intrinsic low-dimensional manifolds, which can be a guidepost for understanding them [2]. In recent research, data visualization techniques have been applied to the analysis of biological data [3] and neural networks that contain complex structures such as hierarchical structures [4], [5], time-series changes [6], [7], and non-convexity of loss functions [8] yielding many successful results.

Dimensionality reduction [9] is one of the basic approaches for reducing the dimensionality of high-dimensional data with some assumptions that motivate effective reduction. One of the principal approaches for the reduction is to minimize the reconstruction error from a low-dimensional representation to high-dimensional data. Principal component analysis (PCA) is a classical and well-known method used to find a subspace that minimizes the reconstruction discrepancy. PCA specifically searches for dominant axes based on spectral decomposition and regards them as the new axes of a low-dimensional subspace. Through this procedure, PCA finds the most explainable linear transformation in the spectral domain and provides a big picture of data. Gaussian process latent variable model (GP-LVM) [10], [11] is a nonlinear extension of the PCA framework that explicitly assumes a low-dimensional representation as latent variables. This modification prompts an optimal representation for its dimensionality with a nonlinear expression. Autoencoder (AE) [12], [13], [14] is another nonlinear approach parameterized by deep neural networks [15], which yields a transferable representation to downstream tasks [16]. However, these approaches use original high-dimensional data to measure

The associate editor coordinating the review of this manuscript and approving it for publication was Walter Didimo.

reconstruction accuracy and lack locality preservation useful for visualization. t-Distributed stochastic neighborhood embedding (t-SNE) is a long-standing dimensionality reduction technique for visualization purposes. t-SNE first estimates the local structure by constructing a neighborhood graph and embedding it in a low-dimensional space. The graph construction corresponds to data manifold approximation as a weighted graph and induces an understandable representation that effectively visualizes local structures. From the above, *graph embedding approaches* (e.g., t-SNE) provide a more interpretable representation than *data embedding approaches* (e.g., PCA and AEs for their locality preservation property and are typically used to visualize high-dimensional data.

However, the benefit of the locality preservation of graph embedding approaches is coincident with some of the problematic aspects caused by the lack of global preservation. One major problem is the absence of inter-cluster similarities, which degrades the interpretability of embedding and induces a misleading visualization. Many state-of-the-art studies have recently attempted to address this issue by introducing better initialization [3], [17], diffusion operation [18], and triplet interaction [19], [20] into graph embedding approaches. The other important problem is the sensitivity to local oscillation [21]. Conventional and recent methods only use the local distance measurement and not other global information, to approximate the data manifold. Under this condition, when the data manifold locally oscillates through some undesirable effects (e.g., noises), these methods do not effectively measure the local distance, thereby significantly degrading the manifold approximation and resulting visualization.

In this study, we propose *spectral-based manifold approximation and projection (SpectralMAP)* to overcome the limitations of graph embedding approaches. We then integrate UMAP [17], which solves the inter-cluster similarity problem by better initialization, and introduce filtering-based global preservation and a distance-independent approximation scheme. Our main idea is to roughly estimate data topology with a neighborhood graph and performs a novel global filter. This dual approximation allows explicit global preservation and effective visualization when the manifold locally oscillates, which is the main difference from previous methods. For efficient global filter computation, we explore a learning rule of a data embedding approach, GP-LVM. Consequently, we derive singular value decomposition (SVD)-based filter computation independent of the distance calculation and contains global data in the spectral domain. Our contributions are as follows:

- We introduce a novel approximation scheme of data manifolds into UMAP to roughly approximate the data manifold and prune it with a global filter. Through this dual procedure, we can explicitly consider the global structure and avoid misleading visualization under local oscillation.
- We explore the data embedding approach and derive a global representation of data based on its learning rule. Consequently, we calculate a global filter based on the SVD, which allows us to refer to a global structure in the data spectral domain independent of sensitive local distance measurements.

In Section II, we review previous dimensionality reduction techniques for data visualization. Furthermore, Section III presents SpectralMAP. In Section IV, we demonstrate the performance of SpectralMAP by comparing with baselines and state-of-the-art methods.

II. RELATED WORK

Dimensionality reduction techniques are broadly used for various purposes [22]. We categorized them into data and graph embedding approaches, which may be useful in presenting an overview of visualization-aided dimensionality reduction techniques.

A. DATA EMBEDDING APPROACHES

Data embedding approaches directly embed high-dimensional data in a low-dimensional space. PCA [23], [24] and its variants [25], [26], [27], [28] are well-known methods that search for dominant axes in a high-dimensional data space. The projection derived by PCA maximizes the variance in a low-dimensional subspace; thus, it provides a large representation of the given high-dimensional data. GP-LVM [10] is a nonlinear extension of the PCA framework, and Bayesian GP-LVM [11] performs a fully Bayesian dimensionality reduction that provides uncertainty to a low-dimensional representation. Those PCA-based approaches are often applied to multi-view data analysis [22], [29]. AEs [12] are a nonlinear data embedding approach parameterized by deep neural networks. SCVIS [14] is a current visualization method based on variational AEs [13]. It can effectively derive a low-dimensional representation of new input data by parametric mapping, which is difficult for non-parametric approaches, such as t-SNE and UMAP. Data embedding approaches estimate a low-dimensional representation by directly embedding high-dimensional data and typically succeed in obtaining a global representation. However, these methods do not consider the locality within high-dimensional data; thus, they tend to overlook potential clusters and do not derive a suitable representation for visualization.

B. GRAPH EMBEDDING APPROACHES

Graph embedding approaches focus on the locality preservation of high-dimensional data. They first approximate a data manifold by constructing a neighborhood graph and then embed it in a low-dimensional space. Laplacian eigenmaps (LE) [30], [31] are an early method based on a neighborhood graph. They derive the low-dimensional representation as the eigenvectors of the *graph Laplacian*. The graph Laplacian is the Laplace-Beltrami operator on a discrete manifold; thus,

LE can provide an optimal low-dimensional representation of data based on convex optimization. Stochastic neighbor embedding (SNE) [32] is a non-convex approach that calculates a neighborhood graph as conditional probabilities in both high- and low-dimensional spaces. SNE minimizes the differences in these graphs and derives a visible two-dimensional or three-dimensional representation. However, in several cases, representations derived by SNE tend to mix independent clusters. This is known as the *crowding problem*. t-SNE [33] addresses this issue by introducing a heavy-tailed distribution to calculate a graph in a low-dimensional space. The remarkable success of t-SNE has led to its many variants, such as the triplet loss function [34], parametrization by neural networks [35], and a scalable computation based on the Barnes-Hut algorithm [36]. Although graph embedding approaches can effectively reduce dimensions for visualization, they cannot consider the global structure of high-dimensional data.

UMAP [17] is a recent graph embedding approach that introduces better initialization based on LE, which partially solves the absence of the global structure. It derives a neighborhood graph as a fuzzy topological representation, which reduces computational costs by contrastive optimization based on negative sampling [37]. PHATE [18], [38] takes another approach to address the abovementioned issue and introduces the diffusion operation [39], which enforces the graph to propagate similarities between similar samples and recover the global structure of data. Anchor t-SNE (At-SNE) [40] introduces hierarchical optimization into t-SNE. At-SNE regards the centers of the K-means clustering as anchor points and projects them into the low-dimensional space. It then arranges the remaining data points around the anchor points, thereby enabling high-level global structure preservation. Pairwise controlled manifold approximation projection (PaCMAP) [20] extends the loss function of UMAP and explicitly defines mid-near points for global structure preservation. Although the abovementioned approaches achieve good results in several experimental settings, they fail to approximate the data manifold when the neighborhood structure oscillates highly (e.g., noise effects). Under this condition, even though the separation of each independent cluster is clear from the global perspective, they cannot be separated because their global information fully depends on the distance measurement in a noisy high-dimensional space. As a solution, herein, we introduce a spectral-based global reference independent of the local distance measurement, which approximates the manifold based on ideas of other data embedding approaches.

III. PROPOSED METHOD

In this section, we present SpectralMAP, a novel dimensionality reduction technique for visualization purposes. In III-A and III-B, we briefly review the UMAP algorithm, describe GP-LVM, and consider its learning rule, which is beneficial for obtaining a distance-independent global representation. In III-C, we derive a global data representation. Finally, in III-D, we derive the SpectralMAP algorithm by applying it to the UMAP algorithm enabling the explicit preservation of the global structure and motivating an effective low-dimensional representation.

A. UNIFORM MANIFOLD APPROXIMATION AND PROJECTION

Let $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_N]^\top \in \mathbb{R}^{N\times D}$ be the D-dimensional observed variables with N data points. We assume that $\mathbf{X}$ is centralized. We aim herein to extract an intrinsic low-dimensional representation $\mathbf{Y} = [\mathbf{y}_1, \mathbf{y}_2, \ldots, \mathbf{y}_N] \in \mathbb{R}^{N\times Q}$ (Q being the number of the dimensions of $\mathbf{Y}$). To achieve this, UMAP first approximates a data manifold by assuming that the data points are uniformly distributed on the manifold. This approximation is simplified to a k-nearest neighborhood graph computation. UMAP then calculates the weight matrix $\mathbf{W}_{n|m} \in \mathbb{R}^{N\times N}$ between each data point as follows:

$$[\mathbf{W}_{n|m}]_{ij} = \begin{cases} \exp\left[-\dfrac{\mathrm{d}(\mathbf{x}_i, \mathbf{x}_j) - \rho_i}{\sigma_i}\right] & (j \in \{i_1, i_2, \ldots, i_k\}) \\ 0 & (\text{otherwise}), \end{cases} \tag{1}$$

where $[\mathbf{W}_{n|m}]_{ij}$ represents the ij-th entry of $\mathbf{W}_{n|m}$, i_l ($l = 1, 2, \ldots, k$) represents an index of the l-nearest neighbor of $\mathbf{x}_i$ measured by a distance function $\mathrm{d}(\cdot,\cdot)$, $\rho_i = \mathrm{d}(\mathbf{x}_i, \mathbf{x}_{i_1})$ represents the distance from the nearest neighbor, and σ_i represents a parameter indicating local connectivity around $\mathbf{x}_i$. The hyperparameter in Eq. (1) is the number of neighbors k typically set between 20 and 100. UMAP then estimates the data distribution called *fuzzy topological representation*, as follows:

$$\mathbf{W} = \mathbf{W}_{n|m} + \mathbf{W}_{m|n} - \mathbf{W}_{n|m} \circ \mathbf{W}_{m|n}, \tag{2}$$

where $\circ$ represents the Hadamard product. Practically, Eq. (2) corresponds to the symmetrization of the first constructed graph. From Eqs. (1) and (2), we approximate the manifold in which the data point exists, and derive a locality-aware representation of $\mathbf{X}$.

With the fuzzy topological representation $\mathbf{W}$, UMAP arranges the low-dimensional representation $\mathbf{Y}$ with the basic concept of matching the local similarity of $\mathbf{Y}$ to the fuzzy topological representation $\mathbf{W}$. UMAP defines the local similarity v_{ij} between $\mathbf{y}_i$ and $\mathbf{y}_j$ similar to t-SNE as

$$v_{ij} = \left(1 + a||\mathbf{y}_i - \mathbf{y}_j||_2^{2b}\right)^{-1}, \tag{3}$$

where a and b represent the hyperparameters and are set as $a \approx 1.6$ and $b \approx 0.9$ [17], respectively. UMAP matches them by minimizing the cross-entropy between v_{ij} and w_{ij} (w_{ij} being the ij-th entry of $\mathbf{W}$). Omitting the constant terms, the UMAP loss function is written as follows:

$$\mathcal{L} = -\sum_{i=1}^{N}\sum_{j=1}^{N}\left\{w_{ij}\log v_{ij} + (1 - w_{ij})\log(1 - v_{ij})\right\}. \tag{4}$$

Equation (4) is regarded as a physical system [41]. That is, the first term of Eq. (4) is regarded as an *attraction force*, and the second one is a *repulsion force* between samples i and j. This motivates a stochastic optimization based on negative sampling [37] and enables UMAP to deal with large-scale datasets. However, UMAP does not explicitly consider the global structure in its algorithm and fully depends on the distance calculation in high- and low-dimensional spaces. This results in accuracy degradation during local oscillation, inducing a misleading visualization, where inter-cluster similarities are dismissed. Our aim is to overcome the limitations caused by the lack of explicit global preservation. In the following section, we explore a learning rule behind the data embedding approach, GP-LVM, and consider a global representation of the given data.

B. LEARNING RULE BEHIND GP-LVM

We first briefly describe GP-LVM here. GP-LVM assumes that the observed variables $\mathbf{X}$ are generated by Gaussian processes [42] with a low-dimensional representation $\mathbf{Y}$ as an input. GP-LVM estimates $\mathbf{Y}$ by maximizing the following marginal log-likelihood:

$$\log p(\mathbf{X}|\mathbf{Y}) = -\frac{ND}{2}\log(2\pi) - \frac{D}{2}\log|\mathbf{K}| - \frac{1}{2}(\mathbf{K}^{-1}\mathbf{X}\mathbf{X}), \tag{5}$$

where $\mathbf{K} \in \mathbb{R}^{N\times N}$ is a similarity matrix of $\mathbf{Y}$ on a feature map $\phi(\cdot)$ or known as a gram matrix whose ij-th entry is calculated using a kernel function $k(\mathbf{y}_i, \mathbf{y}_j) = \phi(\mathbf{y}_i)^\top\phi(\mathbf{y}_j)$. The kernel function corresponds to the UMAP similarity measurement in Eq. (3). Equation (5) contains all observed variables $\mathbf{X}$ in contrast to the UMAP loss function in Eq. (4) that contains a locally encoded representation $\mathbf{W}$ motivating the global preservation property of GP-LVM. The maximization problem of Eq. (5) is typically non-convex due to the nonlinearity of the kernel function and is solved using the quasi-Newton algorithm. However, although Eq. (5) is a non-convex function w.r.t. the low-dimensional representation $\mathbf{Y}$, it is a convex function w.r.t. the gram matrix $\mathbf{K}$. We can easily see this by differentiating Eq. (5) w.r.t. $\mathbf{K}$ as

$$\frac{\partial}{\partial\mathbf{K}}\log p(\mathbf{X}|\mathbf{Y}) = \frac{D}{2}\mathbf{K}^{-1} - \frac{1}{2}\mathbf{K}^{-1}\mathbf{X}\mathbf{X}^\top\mathbf{K}^{-1}. \tag{6}$$

At the saddle point, the differentiation in Eq. (6) is equal to 0 in every $\mathbf{K}$ entry It obtains the following optimal solution ignoring a trivial solution $\mathbf{K} = \mathbf{O}$:

$$\mathbf{K} = D^{-1}\mathbf{X}\mathbf{X}^\top. \tag{7}$$

From Eq. (7), the gram matrix gets closer to $\mathbf{X}\mathbf{X}^\top$ in each optimization step. Therefore, GP-LVM imitates the sample similarity matrix $\mathbf{X}\mathbf{X}^\top$ in its low-dimensional representation in contrast to UMAP, which imitates a fuzzy topological representation $\mathbf{W}$. The similarity matrix is computed from the original data matrix $\mathbf{X}$ without any distance calculation. Thus, $\mathbf{X}\mathbf{X}^\top$ is regarded as a central component for obtaining a distance-independent global representation of the given data.

C. GLOBAL REPRESENTATION OF THE GIVEN DATA

In this subsection, we define a global representation of the given data. Note that we aim to visualize the given high-dimensional data; thus, we require a scalable algorithm to obtain such a representation. Accordingly, we consider the global representation $\widetilde{\mathbf{X}}$ a linear combination of the original data with the equation $\widetilde{\mathbf{X}} = \mathbf{X}\mathbf{W}$, where $\mathbf{W} \in \mathbb{R}^{D\times D}$ represents a linear operator. We then constrain $\mathbf{W}$ to obtain a reasonable global representation. In the previous section, we observed the similarity matrix containing original global data and constrained the similarity matrix of $\widetilde{\mathbf{X}}$ into a specific form. We assume here that the similarity matrix $\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^\top$ is invariant to its power operation as follows:

$$(\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^\top)(\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^\top) = \widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^\top. \tag{8}$$

Therefore, we obtain

$$\begin{aligned}\widetilde{\mathbf{X}}^\top\widetilde{\mathbf{X}} &= \mathbf{W}^\top\mathbf{X}^\top\mathbf{X}\mathbf{W}\\ &= \mathbf{I}.\end{aligned} \tag{9}$$

We have two aims for this constraint. First, under this constraint, the similarity matrix $\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^\top$ is stationary to the matrix power operation; thus, we can obtain a representation that is more stable than the original similarity matrix $\mathbf{X}\mathbf{X}^\top$, which is not stationary to the power operation. This idea is motivated by the diffusion maps [39] or the limiting distribution of the Markov chain [43]. Second, the rank of $\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^\top$ does not further drop by the power operation; thus, we expect a more sparse representation compared with the original form $\mathbf{X}\mathbf{X}^\top$. Under the constraint in Eq. (9), we maximize the cross-covariance [44] between the original data and the global representation as $||\mathbf{X}^\top\widetilde{\mathbf{X}}||_{\mathrm{F}}^2 = ||\mathbf{X}^\top\mathbf{X}\mathbf{W}||_{\mathrm{F}}^2$. We obtain the following optimization problem by replacing the Frobenius norm with the trace expression:

$$\max_{\mathbf{W}} \mathrm{tr}\left[(\mathbf{X}^\top\mathbf{X}\mathbf{W})^\top(\mathbf{X}^\top\mathbf{X}\mathbf{W})\right] \quad \text{s.t. } \mathbf{W}^\top\mathbf{X}^\top\mathbf{X}\mathbf{W} = \mathbf{I}. \tag{10}$$

We can now solve Eq. (10) in a closed form. We first apply the SVD to $\mathbf{X}$ as $\mathbf{X} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top$, where $\mathbf{U} \in \mathbb{R}^{N\times N}$ and $\mathbf{V} \in \mathbb{R}^{D\times D}$ denote orthogonal matrices, and $\mathbf{\Sigma} \in \mathbb{R}^{N\times D}$ represents a rectangular diagonal matrix with a singular value of $\mathbf{X}$. Under the constraint function in Eq. (9), we restrict $\mathbf{W}$ to the following form:

$$\mathbf{W} = (\mathbf{X}^\top\mathbf{X})^{-\frac{1}{2}}\mathbf{Q}, \tag{11}$$

where $\mathbf{Q} \in \mathbb{R}^{D\times D}$ is an arbitrary orthogonal matrix, and $(\mathbf{X}^\top\mathbf{X})^{-\frac{1}{2}} = \mathbf{V}(\mathbf{\Sigma}^\top\mathbf{\Sigma})^{-\frac{1}{2}}\mathbf{V}^\top$. we obtain the following modified optimization problem by substituting Eq. (11) into Eq. (10):

$$\max_{\mathbf{Q}} \mathrm{tr}\left(\mathbf{Q}^\top\mathbf{X}^\top\mathbf{X}\mathbf{Q}\right) \quad \text{s.t.}\mathbf{Q}^\top\mathbf{Q} = \mathbf{I}. \tag{12}$$

We obtain the following eigenvalue problem by introducing the Lagrange multiplier method into Eq. (12):

$$\mathbf{X}^\top\mathbf{X}\mathbf{Q} = \mathbf{\Lambda}\mathbf{Q}, \tag{13}$$

Algorithm 1 SpectralMAP Algorithm

Require: High-dimensional data $\mathbf{X}$, number of neighbors k, dimension Q, threshold θ

Ensure: Low-dimensional representation $\mathbf{Y}$

1: Calculate a fuzzy topological representation $\mathbf{W}$.
2: Calculate C largest left singular vectors $\mathbf{U}$ by compact SVD of $\mathbf{X}$.
3: Derive a normalized global representation $\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^\top$ and a global filter $\mathbf{M}_\theta$ by thresholding $\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^\top$ by θ.
4: Filter $\mathbf{W}$ by $\mathbf{M}_\theta$ by Eq. (15).
5: Normalize each row of $\mathbf{W}_\theta$ by Eq. (16).
6: Minimize the cross entropy of Eq. (4) w.r.t. $\mathbf{Y}$.
7: **return** $\mathbf{Y}$

where $\mathbf{\Lambda}$ is a diagonal matrix with Lagrange multipliers. The SVD result of $\mathbf{X}$ shows that the eigenvalues and eigenvectors of $\mathbf{X}^\top\mathbf{X}$ are equal to $\mathbf{\Sigma}^\top\mathbf{\Sigma}$ and $\mathbf{V}$, respectively. Therefore, the optimal solution of $\mathbf{W}$ is $\mathbf{W} = (\mathbf{X}^\top\mathbf{X})^{-\frac{1}{2}}\mathbf{V}$, and the corresponding $\widetilde{\mathbf{X}}$ is

$$\begin{aligned}\widetilde{\mathbf{X}} &= \mathbf{X}\mathbf{W}\\ &= \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-\frac{1}{2}}\mathbf{V}\\ &= \mathbf{U}\mathbf{\Sigma}\mathbf{V}^\top\mathbf{V}(\mathbf{\Sigma}^\top\mathbf{\Sigma})^{-\frac{1}{2}}\mathbf{V}^\top\mathbf{V}\\ &= \mathbf{U}\mathbf{\Sigma}(\mathbf{\Sigma}^\top\mathbf{\Sigma})^{-\frac{1}{2}}. \end{aligned} \tag{14}$$

The result in Eq. (14) is equal to the left singular vectors of the *compact* SVD of $\mathbf{X}$. We then obtain an optimal global representation of $\mathbf{X}$ by spectral decomposition. Note that this global representation is independent of the distance measurement between the samples and is robust to the local oscillation of the data manifold. In practice, we calculate the first C largest singular vectors as $\mathbf{U} = [\mathbf{u}_{:,1}, \mathbf{u}_{:,2}, \ldots, \mathbf{u}_{:,C}] = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_N]^\top \in \mathbb{R}^{N\times C}$ and the global representation as $\widetilde{\mathbf{X}} = \mathbf{U} \in \mathbb{R}^{N\times N}$. We automatically select C by the elbow point detection [45] of the cumulative Von-Neumann entropy [46] and further explain it in Section IV. Note that this transformation is the same as *PCA whitening* [44]. We can obtain different $\widetilde{\mathbf{X}}$ expressions (e.g., Mahanolabis whitening) by modifying the objective function. Using this spectral data information, we compute the global structure of the given data and introduce it into the UMAP algorithm in the following section.

D. SPECTRAL-BASED MANIFOLD APPROXIMATION AND PROJECTION

In this subsection, we derive the SpectralMAP algorithm, which introduces spectral-based global preservation into the UMAP algorithm. We construct the neighborhood graph with a *spectral filtering scheme* to effectively use the global representation $\widetilde{\mathbf{X}}$. We first roughly approximate the data manifold by constructing a neighborhood graph and then prune it by filtering. Subsequently, we compute the filter by thresholding the similarity matrix $\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^\top$ as $\mathbf{M}_\theta$ and prune the fuzzy topological representation in Eq. (2) as follows:

$$\mathbf{W}_\theta = \mathbf{W} \circ \mathbf{M}_\theta, \tag{15}$$

where θ represents a threshold value. We normalize $\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^\top$ before thresholding. For a rough approximation of the data manifold before thresholding, we first set the number of neighbors k to a large value (e.g., $k = 100\ \sim$). Although the obtained $\mathbf{W}_\theta$ is a symmetry matrix, each row has a different maximum value that should be normalized. Therefore, we evenly set the maximum value as 1 and calculate the following spectral fuzzy topological representation $\mathbf{W}_\mathrm{S}$:

$$\mathbf{W}_\mathrm{S} = \mathbf{D}^{-1}\mathbf{W}_\theta + \mathbf{W}_\theta\mathbf{D}^{-1} - \mathbf{D}^{-1}\mathbf{W}_\theta \circ \mathbf{W}_\theta\mathbf{D}^{-1}, \tag{16}$$

where $\mathbf{D}$ represents a diagonal matrix with entries of the maximum values of each row of $\mathbf{W}_\theta$. We compute the low-dimensional representation by substituting $\mathbf{W}_\mathrm{S}$ into the objective function in Eq. (4). Algorithm 1 describes the SpectralMAP. Our algorithm needs the hyperparameters of UMAP, but we ignore them for simple notation. The main difference between our method and UMAP is the filtering process of Eqs. (15) and (16). This enables us to consider the global structure of the given data explicitly revealed in the spectral domain of $\mathbf{X}$ and allows the distance-independent approximation of the data manifold. If we obtain a similar manifold topology from $\mathbf{M}_\theta$ and $\mathbf{W}$, the obtained algorithm becomes closer to the UMAP algorithm. Conversely, when we obtain different manifold structures from them, the topological representation is pruned by the mask matrix and modified by the spectral structure. The compact SVD computation linearly scales to the number of samples N. Accordingly, we can compute the global representation in a computationally acceptable manner. We overcome the previous limitations with the SpectralMAP algorithm and verify this in the subsequent section.

IV. EXPERIMENT

In this section, we validate our method with qualitative and quantitative experiments. We implement our method with `umap-learn` [17] and run it and all comparative methods on an Intel Core i7-10700 CPU.

A. DATASET

We use multiple datasets from several domains (e.g., images, text, and biological datasets).

1) **SwissRoll**. SwissRoll is a well-known synthetic dataset used to verify visualization-aided dimensionality reduction techniques [47], [48], [49]. We use this dataset and regard its synthesis vectors as features.
2) **MNIST and Noisy MNIST.**[1] MNIST contains 60 K real images of handwritten digits and labels corresponding to each digit. We create a noisy version of MNIST (Noisy MNIST) by randomly losing the pixels of MNIST with 35%. We use their high-dimensional vectorized pixels as feature vectors.

[1] http://yann.lecun.com/exdb/mnist/

TABLE 1. Statistical details of the employed datasets.

Dataset	#Samples	#Dimensions	#Categories	Type	Features
SwissRoll	30,000	3	Continuous	Synthesis	Vectors
MNIST	60,000	784	10	Image	Pixels
Noisy MNIST	60,000	784	10	Image	Pixels
COIL-20	1,440	16,384	20	Image	Pixels
Noisy COIL-20	1,440	16,384	20	Image	Pixels
Fashon MNIST	60,000	784	10	Image	Pixels
DBPedia	100,000	100	14	Text	FastText
Tasic et al.	23,822	3,000	133	scRNA-seq	Genes

TABLE 2. Default hyperparameters of SpectralMAP.

Number of neighbors k	150
Dimension Q	2
Threshold θ	0.6
Number of singular vectors C	Automatically selected

3) **COIL-20 and Noisy COIL-20 [50].** COIL-20 contains 1,440 grayscale images with 20 object variations and rotation structures and labels corresponding to each object. Each object of COIL-20 is evenly captured in a single rotation across 72 images. We also create a noisy version of COIL-20 (Noisy COIL-20) by randomly losing the pixels with 35%. We employ their high-dimensional vectorized pixels as feature vectors.
4) **Fashion MNIST [51].** Fashion MNIST contains 60 K real images of fashion items and labels corresponding to each item. It has more complex patterns and clusters than MNIST. We also use their high-dimensional vectorized pixels as feature vectors.
5) **DBPedia.**[2] DBPedia contains 560 K Wikipedia articles with 14 categories and labels corresponding to each category. We extract 100-dimensional feature vectors with FastText following [40] and use 100 K samples for embedding.
6) **Tasic et al. [52]** The dataset of Tasic et al. contains single-cell RNA sequence (scRNA-seq) data consisting of adult mouse cortex cells and labels corresponding to each cell. It has strong hierarchical clusters along with the function of each cell [3]. We use this dataset to verify the applicability of our method to a complex dataset.

Table 1 summarizes statistical information about the datasets used here (e.g., the number of samples and dimensions).

B. EXPERIMENTAL SETTINGS

Here, we compare our method with three popular and three state-of-the-art methods.

1) **PCA [23].** PCA is a well-known dimensionality reduction method used for various purposes, including visualization and preprocessing. We use it in this study as a conventional method for visualization.
2) **t-SNE [33].** t-SNE is a baseline method for visualization-aided dimensionality reduction. In this study, we use Burnes-Hut t-SNE [36] of `sklearn` implementation [53], an efficient version of the original t-SNE algorithm.
3) **UMAP [17].** UMAP is a recent baseline method used for biological data visualization [54]. It is one of the main components of SpectralMAP. We used it here for the ablation study of our method. We use the original implementation provided in `umap-learn` [17].
4) **PHATE [18].** PHATE is a recent state-of-the-art method that introduces a diffusion operation into neighborhood graph construction. It is an early method for global preservation used in this study as a state-of-the-art method. We use the publicly available implementation provided in the original paper.
5) **TriMAP [19].** TriMAP also aims to preserve the global structure and introduces the triplet loss function into the UMAP algorithm. We use it as a state-of-the-art method and employ the open implementation provided in the original paper.
6) **PaCMAP [20].** PaCMAP is the most recent and comprehensive method of the UMAP frameworks. We use it as another state-of-the-art method and employ the open implementation provided in the original paper.

Data visualization is used in a wide range of fields without any specific knowledge; thus, we use the default parameters of these comparative methods. We also define the default parameters of our method. The details are described in IV-C.

We compared the visual effect of low-dimensional embedding and the classification accuracy of the k-nearest neighbor classifier in the experiment. We conducted qualitative experiments with several neighbor selections as $k \in [1, 3, 5, 10, 15, 20, 25, 30]$ and used the result of the best k selection as the accuracy of each method. The quantitative experiments were conducted on MNIST, Noisy MNIST, COIL-20, Noisy COIL-20, Fashion MNIST, and DBPedia which contain clear categorical information.

C. TRAINING PROCEDURE

A data visualization technique should not contain hyperparameter tuning because it is used in various fields without specific domain knowledge. Therefore, we set the default parameters of our method and used them across

[2]https://wiki.dbpedia.org/

TABLE 3. Quantitative results of k-nearest neighbor classifier on each dataset. The index of each result indicates the number of neighbors of each result and is the best selection in [1, 3, 5, 10, 15, 20, 25, 30]. The best result in each dataset is bold-faced, and the second best is underlined.

Dataset	PCA	t-SNE	UMAP	PHATE	TriMAP	PaCMAP	SpectralMAP
MNIST	$0.4514_{(30)}$	$\mathbf{0.9724}_{(5)}$	$0.9675_{(10)}$	$0.7930_{(30)}$	$0.9539_{(30)}$	$\underline{0.9696}_{(30)}$	$0.9571_{(25)}$
Noisy MNIST	$0.3983_{(30)}$	$0.7456_{(20)}$	$\underline{0.8014}_{(30)}$	$0.5520_{(30)}$	$0.7432_{(30)}$	$0.7470_{(30)}$	$\mathbf{0.8700}_{(20)}$
COIL-20	$0.7215_{(10)}$	$\mathbf{0.9986}_{(1)}$	$0.9028_{(1)}$	$0.8917_{(1)}$	$0.8722_{(1)}$	$0.8965_{(1)}$	$\underline{0.9111}_{(1)}$
Noisy COIL-20	$0.6896_{(15)}$	$0.5563_{(10)}$	$0.5951_{(15)}$	$\underline{0.7549}_{(10)}$	$0.7007_{(10)}$	$0.6694_{(20)}$	$\mathbf{0.9047}_{(1)}$
Fashion MNIST	$0.5538_{(30)}$	$\mathbf{0.8397}_{(10)}$	$\underline{0.7861}_{(25)}$	$0.7382_{(15)}$	$0.7593_{(30)}$	$0.7824_{(30)}$	$0.7821_{(30)}$
DBPedia	$0.6559_{(30)}$	$\underline{0.9995}_{(3)}$	$\underline{0.9995}_{(10)}$	$0.9976_{(5)}$	$0.9995_{(3)}$	$\mathbf{0.9996}_{(10)}$	$\mathbf{0.9996}_{(3)}$

FIGURE 1. Visualization results on SwissRoll with UMAP and SpectralMAP.

all experiments. Table 2 presents the details of the default parameters. Our method requires the left singular vectors $\mathbf{U}$ and their number C and is sensitive to C selection. Therefore, we determined it on the basis of the elbow detection of the cumulative Von Neumann entropy [46] given as follows:

$$\mathcal{H} = -\sum_{i=1}^{C} \rho_i \log \rho_i, \tag{17}$$

where $\rho_i = \frac{\lambda_i}{\mathrm{tr}(\mathbf{X}^\top \mathbf{X})}$, and λ_i represents i-th eigenvalue of $\mathbf{X}\mathbf{X}^\top$ obtained by the singular decomposition of $\mathbf{X}$. We calculated singular values from the largest to the 200th and detected the elbow point from the cumulative entropy graph. Through this procedure, we can robustly correlate the visualization results to the hyperparameters.

D. RESULTS AND DISCUSSION

1) ABLATION STUDIES

We first conducted an ablation study to verify our novelty. Figure 1 depicts the visualization results on SwissRoll. The results easily confirm that SpectralMAP produces a smoother representation than UMAP. The smooth representation was the global property of the data manifold; therefore, our novelty contributed to the revelation of an effective representation by considering the global structure of the given data. Figure 2 depicts the COIL-20 and Noisy COIL-20 results. In normal COIL-20,SpectralMAP and UMAP preserved the object variation (i.e., each color) and rotation structures. However, in the noisy setting, UMAP did not even preserve the object variation, and significant accuracy degradation was observed. This degradation was caused by noise oscillation. UMAP did not approximate the data manifold with the neighborhood graph in Eq. (2). In contrast, SpectralMAP effectively preserved

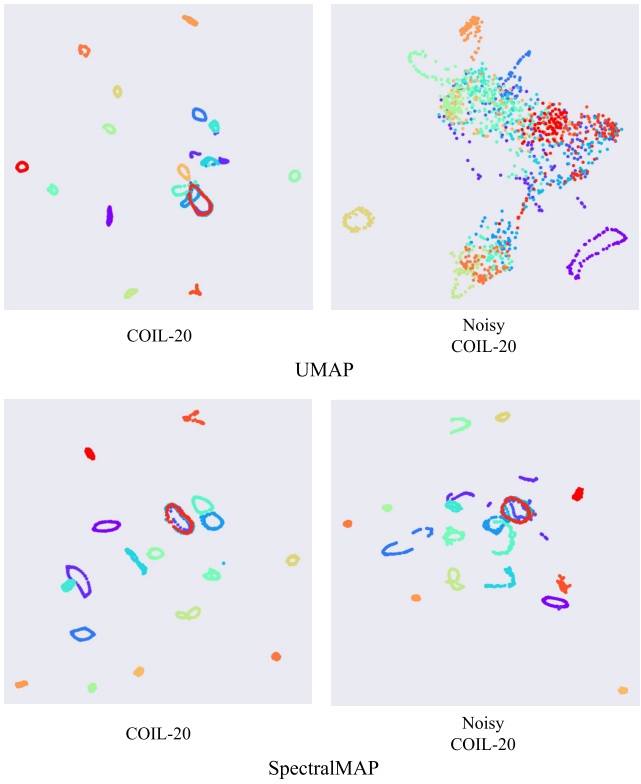


FIGURE 2. Visualization results on COIL-20 and Noisy COIL-20 with UMAP and SpectralMAP.

both object variation and local rotation well by the dual approximation of the data manifold. Therefore, these results confirm the effectiveness of our novelty.

2) QUANTITATIVE RESULTS

Table 3 presents the quantitative results. Previous approaches have shown a significant accuracy degradation on noisy datasets because of the mismatching between the local distance measurement and the locally oscillated data manifold. Although our method was also obliged to accuracy degradation, it robustly dealt with the oscillated data by considering the global structure with spectral information independent of the distance calculation. In the normal setting, SpectralMAP exhibited slight accuracy degradation compared with UMAP due to the global consideration of our method. However, the

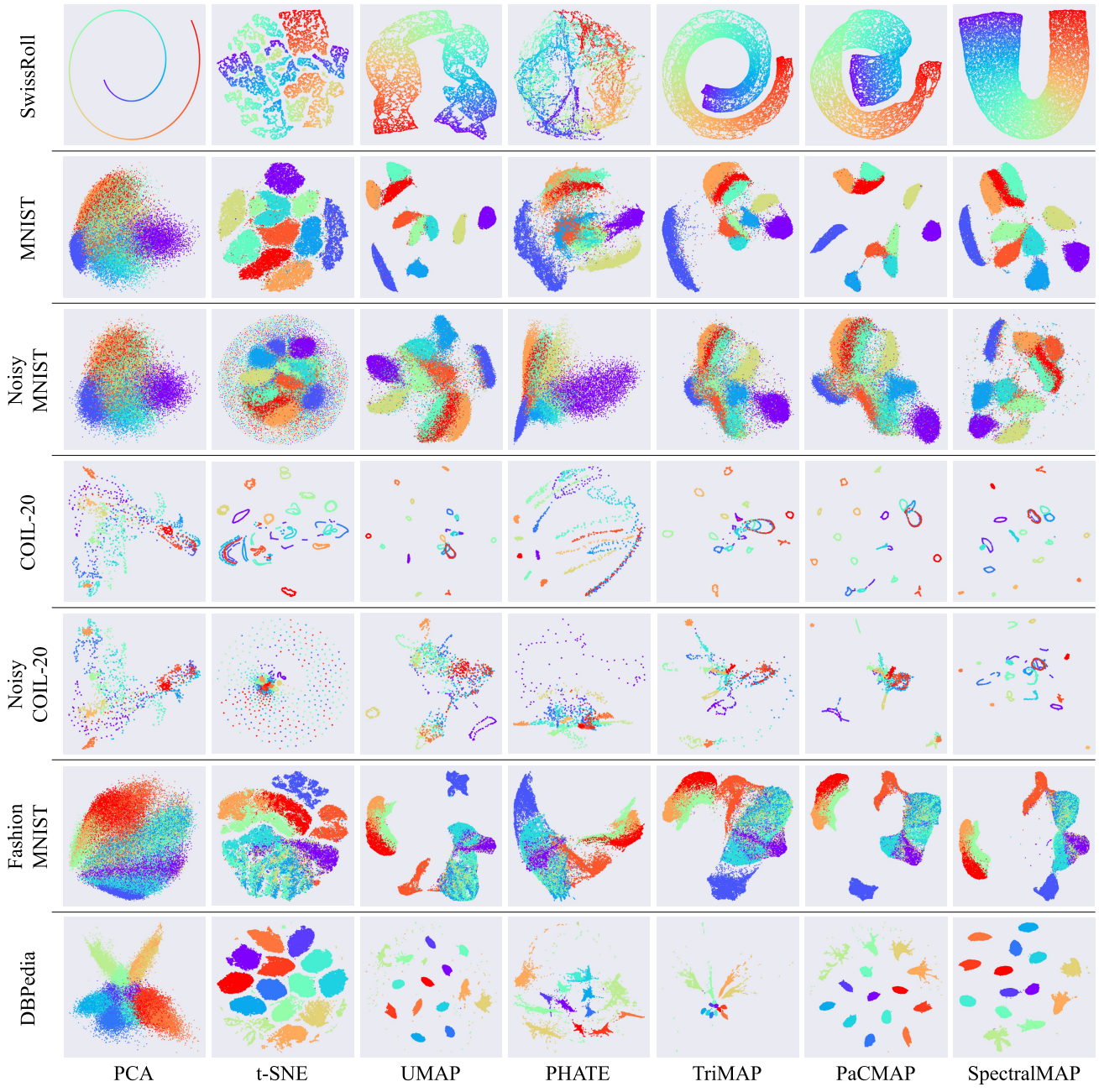


FIGURE 3. Visualization results on the synthesis, image, and text datasets. Each color indicates each category. SwissRoll dataset has continuous labels, and the other datasets have separatable classes; thus, the embeddings on SwissRoll should connect each color, and the embeddings on the other datasets should separate each color.

effectiveness of our novelty is verified by the noisy setting results, where our method significantly outperforms all comparison and state-of-the-art methods. Therefore, these results confirm the effectiveness of our method.

3) QUALITATIVE RESULTS

Figure 3 shows the experimental results on the synthesis, image, and text datasets. The SwissRoll results show that SpectralMAP embeds high-dimensional data in an unrolling smooth representation, unlike other methods. On the MNIST and COIL-20 datasets, SpectralMAP produces similar visualization compared with the current state-of-the-art methods. In the noisy settings (i.e., Noisy MNIST and Noisy COIL-20), the SpectralMAP embedding depicts more clear boundaries for each category and produces visible results compared with the other methods. On the fashion MNIST dataset, SpectralMAP yields a competitive representation to the other state-of-the-art methods. On DBPedia, the embedding of our method reveals fewer noise-like outliers and a more compact cluster representation than that of the other

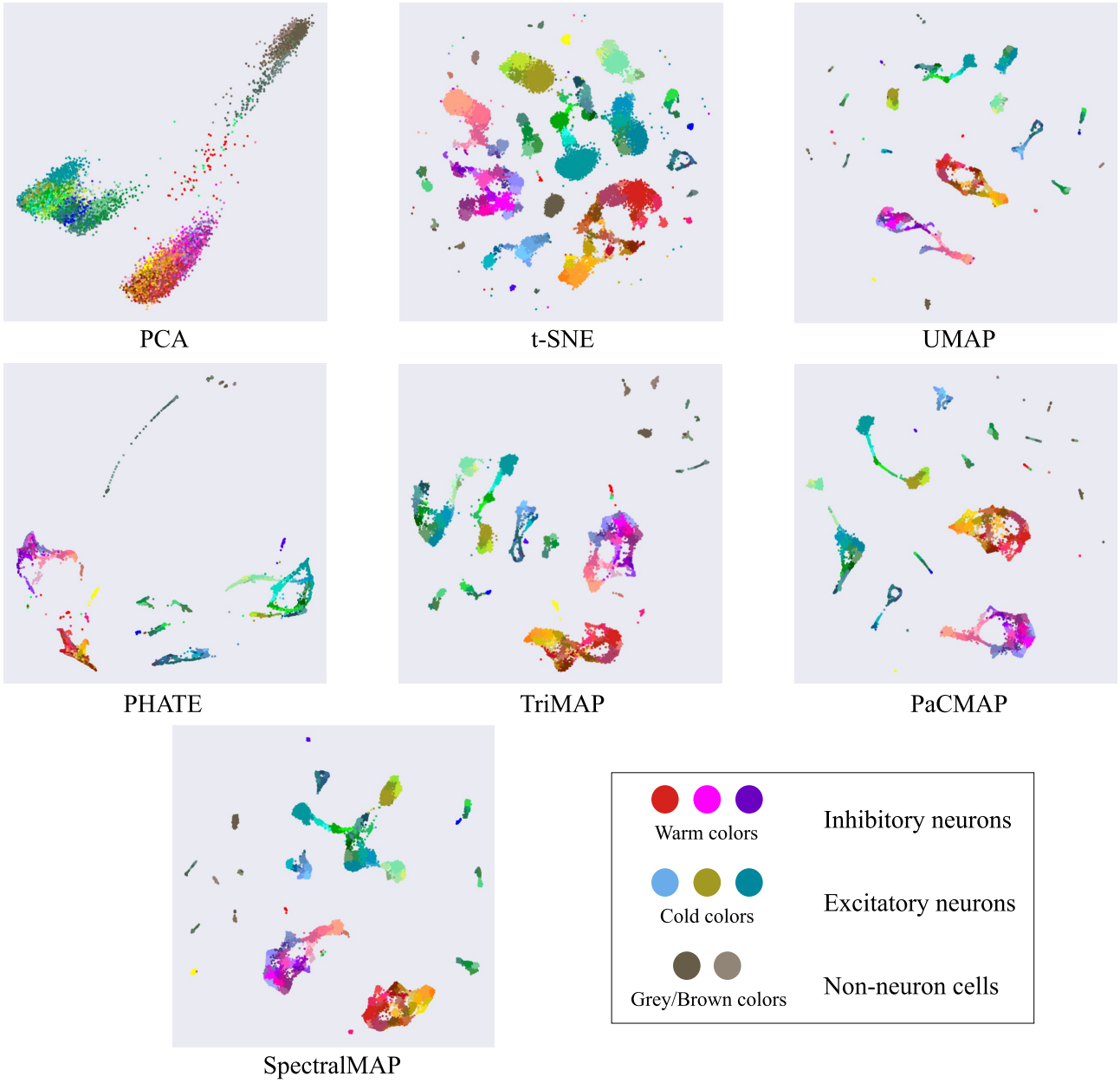


FIGURE 4. Visualization results on Tasic et al. dataset. Each color indicates each category.

methods. Figure 4 illustrates the results on the single-cell dataset. Despite confirming that this representation had clusters along the cell functions, the PCA visualization did not show the sub-clusters of each cluster. In contrast, the t-SNE result shows sub-clusters along the cell function. However, the sub-clusters of the inhibitory cells were split, and no inter-cluster similarity preservation was found between them. UMAP, PHATE, TriMAP, and PaCMAP exhibited a hierarchical structure, and their representations were unsplit within the same clusters. The embedding of SpectralMAP had the same tendency, but the clusters corresponding to the excitatory cells were not to be divided into many sub-clusters like UMAP and PaCMAP. Therefore, our embedding was more useful when we want to infer sub-cluster relationships. These results verify the effectiveness of our method in the visual effects of representations.

V. CONCLUSION

In this study, we proposed a novel visualization method with global structure preservation. We approximated a data manifold with a dual scheme, roughly approximating it with a neighborhood graph and then pruning it with a global filter. With this novelty, we can embed the given data with explicit global preservation when the data manifold locally oscillates. We verified our method on synthesis, image, text, and complex biological datasets and confirmed its effectiveness through qualitative and quantitative experiments.

REFERENCES

[1] L. Wasserman, "Topological data analysis," *Annu. Rev. Statist. Appl.*, vol. 5, pp. 501–532, May 2018.

[2] S. Liu, D. Maljovec, B. Wang, P.-T. Bremer, and V. Pascucci, "Visualizing high-dimensional data: Advances in the past decade," *IEEE Trans. Vis. Comput. Graph.*, vol. 23, no. 3, pp. 1249–1268, Mar. 2017.

[3] D. Kobak and P. Berens, "The art of using t-SNE for single-cell transcriptomics," *Nature Commun.*, vol. 10, no. 1, pp. 1–14, Nov. 2019.

[4] E. Z. Macosko, A. Basu, R. Satija, J. Nemesh, K. Shekhar, M. Goldman, I. Tirosh, A. R. Bialas, N. Kamitaki, E. M. Martersteck, J. J. Trombetta, D. A. Weitz, J. R. Sanes, A. K. Shalek, A. Regev, and S. A. McCarroll, "Highly parallel genome-wide expression profiling of individual cells using nanoliter droplets," *Cell*, vol. 161, no. 5, pp. 1202–1214, 2015.

[5] K. Shekhar, S. W. Lapan, I. E. Whitney, N. M. Tran, E. Z. Macosko, M. Kowalczyk, and X. Adiconis, "Comprehensive classification of retinal bipolar neurons by single-cell transcriptomics," *Cell*, vol. 166, no. 5, pp. 1308–1323, Aug. 2016.

[6] E. R. Zunder, E. Lujan, Y. Goltsev, M. Wernig, and G. P. Nolan, "A continuous molecular roadmap to iPSC reprogramming through progression analysis of single-cell mass cytometry," *Cell Stem Cell*, vol. 16, no. 3, pp. 323–337, Mar. 2015.

[7] G. Schiebinger, J. Shu, M. Tabaka, B. Cleary, V. Subramanian, A. Solomon, J. Gould, S. Liu, P. Berube, L. Lee, J. Chen, J. Brumbaugh, P. Rigollet, K. Hochedlinger, R. Jaenisch, A. Regev, and E. S. Lander, "Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming," *Cell*, vol. 176, no. 4, pp. 928–943, 2019.

[8] H. Li, Z. Xu, G. Taylor, C. Studer, and T. Goldstein, "Visualizing the loss landscape of neural nets," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 31, 2018, pp. 1–11.

[9] L. Van Der Maaten, E. Postma, and J. Van den Herik, "Dimensionality reduction: A comparative," Tilburg Univ., Tilburg, The Netherlands, Tech. Rep., TiCC-TR 2009-005, 2009.

[10] N. Lawrence, "Probabilistic non-linear principal component analysis with Gaussian process latent variable models," *J. Mach. Learn. Res.*, vol. 6, pp. 1783–1816, Nov. 2005.

[11] M. Titsias and N. D. Lawrence, "Bayesian Gaussian process latent variable model," in *Proc. Int. Conf. Artif. Intell. Statist. (AISTATS)*, 2010, pp. 844–851.

[12] G. E. Hinton and R. R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, vol. 313, no. 5786, pp. 504–507, 2006.

[13] D. P. Kingma and M. Welling, "Auto-encoding variational Bayes," 2013, *arXiv:1312.6114*.

[14] J. Ding, A. Condon, and S. P. Shah, "Interpretable dimensionality reduction of single cell transcriptome data with deep generative models," *Nature Commun.*, vol. 9, no. 1, pp. 1–13, May 2018.

[15] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, pp. 436–444, Sep. 2015.

[16] B. Poole, S. Lahiri, M. Raghu, J. Sohl-Dickstein, and S. Ganguli, "Exponential expressivity in deep neural networks through transient chaos," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2016, pp. 1–9.

[17] L. McInnes, J. Healy, and J. Melville, "UMAP: Uniform manifold approximation and projection for dimension reduction," 2018, *arXiv:1802.03426*.

[18] K. R. Moon, D. van Dijk, Z. Wang, S. Gigante, D. B. Burkhardt, W. S. Chen, K. Yim, A. van den Elzen, M. J. Hirn, R. Coifman, N. B. Ivanova, G. Wolf, and S. Krishnaswamy, "Visualizing structure and transitions in high-dimensional biological data," *Nature Biotechnol.*, vol. 37, no. 12, pp. 1482–1492, 2019.

[19] E. Amid and M. K. Warmuth, "TriMap: Large-scale dimensionality reduction using triplets," 2019, *arXiv:1910.00204*.

[20] Y. Wang, H. Huang, C. Rudin, and Y. Shaposhnik, "Understanding how dimension reduction tools work: An empirical approach to deciphering t-SNE, UMAP, TriMap, and PaCMAP for data visualization," *J. Mach. Learn. Res.*, vol. 22, no. 201, pp. 1–73, 2021.

[21] K. Watanabe, K. Maeda, T. Ogawa, and M. Haseyama, "Summarizing data structures with Gaussian process and robust neighborhood preservation," in *Proc. Eur. Conf. Mach. Learn. Princ. Pract. Knowl. Discovery Databases (ECMLPKDD)*, 2022, pp. 1–17.

[22] T. A. Abeo, X.-J. Shen, B.-K. Bao, Z.-J. Zha, and J. Fan, "A generalized multi-dictionary least squares framework regularized with multi-graph embeddings," *Pattern Recognit.*, vol. 90, pp. 1–11, Jan. 2019.

[23] H. Hotelling, "Analysis of a complex of statistical variables into principal components," *J. Educ. Psychol.*, vol. 24, no. 6, pp. 417–441, 1933.

[24] S. Wold, K. Esbensen, and P. Geladi, "Principal component analysis," *Chemometrics Intell. Lab. Syst.*, vol. 2, nos. 1–3, pp. 37–52, 1987.

[25] H. Zou, T. Hastie, and R. Tibshirani, "Sparse principal component analysis," *J. Comput. Graph. Statist.*, vol. 15, no. 2, pp. 265–286, Jan. 2012.

[26] E. J. Candès, X. Li, Y. Ma, and J. Wright, "Robust principal component analysis?" *J. ACM*, vol. 58, no. 1, pp. 1–37, 2009.

[27] E. D. Ganaa, X.-J. Shen, and T. A. Abeo, "Deflated manifold embedding PCA framework via multiple instance factorings," *Multimedia Tools Appl.*, vol. 80, no. 3, pp. 3809–3833, Jan. 2021.

[28] E. D. Ganaa, T. A. Abeo, L. Wang, H.-P. Song, and X.-J. Shen, "Robust deflated principal component analysis via multiple instance factorings for dimension reduction in remote sensing images," *J. Appl. Remote Sens.*, vol. 14, no. 3, 2020, Art. no. 032608.

[29] A. C. Damianou, C. H. Ek, M. K. Titsias, and N. D. Lawrence, "Manifold relevance determination," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2012, pp. 531–538.

[30] M. Belkin and P. Niyogi, "Laplacian eigenmaps and spectral techniques for embedding and clustering," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2001, pp. 1–7.

[31] M. Belkin and P. Niyogi, "Laplacian eigenmaps for dimensionality reduction and data representation," *Neural Comput.*, vol. 15, no. 6, pp. 1373–1396, 2003.

[32] G. E. Hinton and S. Roweis, "Stochastic neighbor embedding," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2002, pp. 1–8.

[33] L. van der Maaten and G. Hinton, "Visualizing data using t-SNE," *J. Mach. Learn. Res.*, vol. 9, pp. 2579–2605, Nov. 2008.

[34] L. van der Maaten and K. Weinberger, "Stochastic triplet embedding," in *Proc. IEEE Int. Workshop Mach. Learn. Signal Process.*, Sep. 2012, pp. 1–6.

[35] L. Van Der Maaten, "Learning a parametric embedding by preserving local structure," in *Proc. Int. Conf. Artif. Intell. Statist. (AISTATS)*, 2009, pp. 384–391.

[36] L. van der Maaten, "Accelerating t-SNE using tree-based algorithms," *J. Mach. Learn. Res.*, vol. 15, no. 1, pp. 3221–3245, Jan. 2014.

[37] T. Mikolov, I. Sutskever, K. Chen, G. S. Corrado, and J. Dean, "Distributed representations of words and phrases and their compositionality," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, 2013, pp. 1–9.

[38] M. Kuchroo et al., "Multiscale PHATE identifies multimodal signatures of COVID-19," *Nature Biotechnol.*, vol. 40, no. 5, pp. 681–691, 2022.

[39] R. R. Coifman and S. Lafon, "Diffusion maps," *Appl. Comput. Harmon. Anal.*, vol. 21, no. 1, pp. 5–30, Jul. 2006.

[40] C. Fu, Y. Zhang, D. Cai, and X. Ren, "AtSNE: Efficient and robust visualization on GPU through hierarchical optimization," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jul. 2019, pp. 176–186.

[41] J. N. Böhm, P. Berens, and D. Kobak, "Attraction-repulsion spectrum in neighbor embeddings," *J. Mach. Learn. Res.*, vol. 23, no. 95, pp. 1–32, 2022.

[42] C. E. Rasmussen and C. K. Williams, *Gaussian Processes for Machine Learning*. Cambridge, MA, USA: MIT Press, 2006.

[43] S. P. Meyn and R. L. Tweedie, *Markov Chains and Stochastic Stability*. Cham, Switzerland: Springer, 2012.

[44] A. Kessy, A. Lewin, and K. Strimmer, "Optimal whitening and decorrelation," *Amer. Statistician*, vol. 72, no. 4, pp. 309–314, Oct. 2018.

[45] V. Satopaa, J. Albrecht, D. Irwin, and B. Raghavan, "Finding a 'Kneedle' in a haystack: Detecting knee points in system behavior," in *Proc. 31st Int. Conf. Distrib. Comput. Syst. Workshops*, Jun. 2011, pp. 166–171.

[46] I. Bengtsson and K. Życzkowski, *Geometry of Quantum States: An Introduction to Quantum Entanglement*. Cambridge, U.K.: Cambridge Univ. Press, 2017.

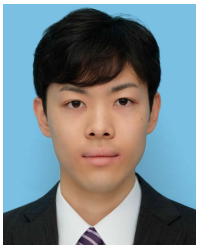
[47] S. T. Roweis and L. K. Saul, "Nonlinear dimensionality reduction by locally linear embedding," *Science*, vol. 290, no. 5500, pp. 2323–2326, Dec. 2000.

[48] M. Balasubramanian and E. L. Schwartz, "The Isomap algorithm and topological stability," *Science*, vol. 295, no. 5552, p. 7, 2002.

[49] X. Zu and Q. Tao, "SpaceMAP: Visualizing high-dimensional data by space expansion," in *Proc. Int. Conf. Mach. Learn. (ICML)*, 2022, pp. 27707–27723.

[50] S. A. Nene, S. K. Nayar, and H. Murase. (1996). *Columbia Object Image Library (COIL-20)*. [Online]. Available: https://www.cs.columbia.edu/CAVE/software/softlib/coil-20.php

[51] H. Xiao, K. Rasul, and R. Vollgraf, "Fashion-MNIST: A novel image dataset for benchmarking machine learning algorithms," 2017, *arXiv:1708.07747*.
[52] B. Tasic et al., "Shared and distinct transcriptomic cell types across neocortical areas," *Nature*, vol. 563, no. 7729, pp. 72–78, Nov. 2018.
[53] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, and M. Blondel, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, Jan. 2011.
[54] E. Becht, L. McInnes, J. Healy, C.-A. Dutertre, I. W. H. Kwok, L. G. Ng, F. Ginhoux, and E. W. Newell, "Dimensionality reduction for visualizing single-cell data using UMAP," *Nature Biotechnol.*, vol. 37, no. 1, pp. 38–44, Jan. 2019.



KOSHI WATANABE (Graduate Student Member, IEEE) received the B.S. degree in electronics and information engineering from Hokkaido University, Japan, in 2022, where he is currently pursuing the M.S. degree with the Graduate School of Information Science and Technology. His research interests include statistical machine learning and data visualization.

KEISUKE MAEDA (Member, IEEE) received the B.S., M.S., and Ph.D. degrees in electronics and information engineering from Hokkaido University, Japan, in 2015, 2017, and 2019, respectively. He is currently a specially appointed Assistant Professor with the Faculty of Information Science and Technology, Hokkaido University. His research interests include multimodal signal processing and machine learning and its applications. He is a member of IEICE.

TAKAHIRO OGAWA (Senior Member, IEEE) received the B.S., M.S., and Ph.D. degrees in electronics and information engineering from Hokkaido University, Japan, in 2003, 2005, and 2007, respectively. He joined the Graduate School of Information Science and Technology, Hokkaido University, in 2008. He is currently a Professor with the Faculty of Information Science and Technology, Hokkaido University. His research interests include artificial intelligence, the Internet of Things, and big data analysis for multimedia signal processing and its applications. He is a member of ACM, IEICE, and ITE. He has been the Special Session Chair of IEEE ISCE, in 2009, the Doctoral Symposium Chair of ACM ICMR, in 2018, the Organized Session Chair of IEEE GCCE, from 2017 to 2019, the TPC Vice Chair of IEEE GCCE, in 2018, and the Conference Chair of IEEE GCCE, in 2019. He has also been an Associate Editor of *ITE Transactions on Media Technology and Applications*.

MIKI HASEYAMA (Senior Member, IEEE) received the B.S., M.S., and Ph.D. degrees in electronics from Hokkaido University, Japan, in 1986, 1988, and 1993, respectively. She joined the Graduate School of Information Science and Technology, Hokkaido University, as an Associate Professor, in 1994. She was a visiting Associate Professor with Washington University, USA, from 1995 to 1996. She is currently a Professor with the Faculty of Information Science and Technology, Hokkaido University. Her research interests include image and video processing and its development into the semantic analysis. She is a member of IEICE and ASJ and a fellow of ITE. She has been the Vice President of the Institute of Image Information and Television Engineers, Japan (ITE), the Editor-in-Chief of *ITE Transactions on Media Technology and Applications*, and the Director of the International Coordination and Publicity, Institute of Electronics, Information and Communication Engineers (IEICE).