Pattern Recognition 139 (2023) 109447

Contents lists available at ScienceDirect

Pattern Recognition

journal homepage: www.elsevier.com/locate/patcog

MEQA: Manifold embedding quality assessment via anisotropic scaling and Kolmogorov-Smirnov test

Subhadip Boral[a,b], Mainak Sarkar[c], Ashish Ghosh[d,*]

[a] Department of Computer Science & Engineering, University of Calcutta, India
[b] Technology Innovation Hub, Indian Statistical Institute, India
[c] Department of Electrical Engineering, Jadavpur University, India
[d] Machine Intelligence Unit, Indian Statistical Institute, 203 B. T. Road, Kolkata 700108, India

ARTICLE INFO

Article history:
Received 17 December 2021
Revised 19 November 2022
Accepted 16 February 2023
Available online 25 February 2023

Keywords:
Manifold learning
Anisotropic scaling
Gradient descent
Global scaling
Singular value decomposition
Kolmogorov-Smirnov test

ABSTRACT

Manifold learning methods unfold the manifold structures and embed them in a lower-dimensional space. The quality of such an embedding should be measured both qualitatively and quantitatively. The proposed manifold embedding quality assessment (MEQA) method does so by taking into account of local and global structure preservation as both are important traits of an embedding. To measure the local structure preservation MEQA uses two transformations. Initially anisotropic scaling, rotation and translation are incorporated to measure the closeness between the original and the embedded data points. In the next stage, rigid transformation is incorporated to quantify the previous transformation which involved anisotropic scaling. For quantifying the global structure preservation, the Kolmogorov-Smirnov test is applied in a distributed manner over each dimension and averaged over them. To establish the superiority of MEQA we conducted several studies over standard synthetic and real-life datasets across separate existing feature extraction techniques.

1. Introduction

In the field of machine learning, high-dimensional data is present in every branch and it incurs a high computational cost while processing and faces "the curse of dimensionality" [1]. Dimensionality reduction (DR) [2] is one of the most widely used remedies for this problem where a low-dimensional representation of a given high-dimensional data is generated while retaining most of the relevant information of the high-dimensional data. DR methods are broadly divided into two categories: Feature Selection [3] and Feature Extraction [4]. Since, it is almost impossible to know beforehand, which features are relevant for model construction, feature extraction should be considered instead of feature selection for DR.

Initial feature extraction methods [5,6] assume that data lie on a linear subspace in a high-dimensional ambient space, whereas in real life, data lies on a low-dimensional, non-linear manifold embedded in a higher-dimensional ambient space. Manifold learning refers to the unsupervised learning which obtain an optimal embedding of non-linear data in its intrinsic dimension. A plethora of manifold learning methods [6–12] have been proposed and developed over the years and many of them learn non-linear data very well. In this regard, it is also true that different manifold learning methods adopt different views for it's learning. With the advancement of manifold learning techniques, it has become mandatory to establish a quality measure that will be a reliable standard to quantitatively gauge the quality of an embedding. Major motivations behind this are lack of manifold learning method which achieves consistently good quality embedding for all types of data and hyperparameters associated with these learning techniques also determine the quality of embedding. So it is necessary to determine which manifold learning technique with the best hyperpaparameter setting provides the optimal quality embedding for a particular data.

The methods that quantitatively evaluate the quality of an embedding with respect to high-dimensional data are known as Embedding Quality Assessment (EQA) measures. The primary focus of these measures is to determine how close the structure of the original manifold is preserved in the embedding. Depending on the approach, EQA methods can be broadly classified into two categories: (a) Transformation based techniques and (b) Rank-preservation based techniques. Among the Transformation based techniques, the Procrustes measure [13,14] works well for data lying on a linear subspace but the Local Procrustes measure [15] can handle isomet-

* Corresponding author.
E-mail addresses: subhadipb_t@isical.ac.in (S. Boral), sarkar.mainak101@gmail.com (M. Sarkar), ash@isical.ac.in (A. Ghosh).

https://doi.org/10.1016/j.patcog.2023.109447
0031-3203/

ric embeddings too. Manifold learning techniques like LLE, HLLE and LE generate normalized embedding and incur anisotropic scaling. Here, the Local Procrustes measure quantifies incorrectly due to adoption of rigid transformation and conformal mappings but NIEQA [16] provides comparatively relevant assessment. Now, due to the inclusion of anisotropic scaling, NIEQA senses two embeddings of different quality as the same and fails to provide meticulous analysis. NIEQA quantifies the global structure preservation of the embedding too but involves a time complex method. Rank-preservation based techniques [17–20] are based on the concept of how much the distance-based ranks of neighbourhood points of each point are preserved in the embedded space. Similar to the rigid-transformation based techniques these rank-preservation based measures work well for isometric manifold learning methods but fail for normalized embedding.

From the discussion, it is evident that state of the art measures either fail to handle normalized embedding or involves uncontrolled flexibility and they also fail to apprehend global structure preservation information. The quality measure proposed in this article attempts to improve upon the deficiencies faced by the existing EQA methods by analyzing both local and global structure preservation of an embedding. The novelty of the proposed method, Manifold Embedding Quality Assessment via Anisotropic Scaling and Kolmogorov-Smirnov Test (MEQA), is two-fold:

i) Local Structure Preservation: The local neighbourhood preservation in the embedded data with respect to the original data is measured by anisotropic scaling with rotation and translation which handles both isometric and normalized embeddings. Irrespective of the embedding quality, the transformation imposed or effort given during this process makes the embedded data very close to the original. So, the amount of transformation imposed is measured by finding the gap between the embedding and the transformed structure of the embedding by the rigid transformation. For inferior quality embedding that effort or gap will be higher in comparison with the superior one. Applying these two ideas MEQA determines the local structure preservation.
ii) Global Structure Preservation: MEQA involves non-parametric Kolmogorov-Smirnov test statistics to evaluate the global topology preservation accurately. The test is performed between the original structure and the transformed structure of the embedding obtained by anisotropic scaling.

Performance by MEQA on the embeddings generated by well established manifold learning algorithms over five synthetic and one real-life datasets are studied to analyse its efficacy. The outcome of the study and comparison against existing EQA methods reveals that MEQA evaluates the quality of an embedding better than the state of the art quality measures and takes lesser time to evaluate.

The subsequent sections are as follows. Section 2 presents an extensive literature survey of EQA methods. MEQA is presented in Section 3 and performance analysis with experimental setup is presented in Section 4. At the end, in Section 5, concluding remarks along with the scope of improvements for future research are given.

2. Related work

The earliest EQA techniques date back to the introduction of Shepard's Diagram [21,22], Kruskal's stress [23] and Procrustes Measure [13,14]. Over the years, several EQA methods have been proposed which can be grouped into two major categories based on their approach.

2.1. Transformation based techniques

Transformation based techniques approach the task of evaluating an embedding by applying certain transformations on the data in the embedded space in order to best fit the data in the original space.

2.1.1. Procrustes measure

The Procrustes measure uses linear rigid transformation to optimally fit a shape into another shape. Here, the word shape references a global geometric structure represented by a number of points in a finite space. Linear rigid transformation include translation, rotation and scaling to match the shapes. The Procrustes measure works well for data lying on linear subspace but fails to work on non-linear manifolds as it processes the entire dataset at once.

2.1.2. Local procrustes measure [15]

The primary intuition that forms this method's basis is that for non-linear manifolds, a small neighborhood could be considered containing a set of points lying on a linear space. Considering this intuition, an optimal match can be obtained between the neighborhood of every data point in the original and embedded space, via rigid transformations. Then the Procrustes measures between the neighborhood of each point in high and low-dimensional space are obtained and the average of them is considered as the Local Procrustes measure.

2.1.3. Normalization independent embedding quality assessment (NIEQA) [16]

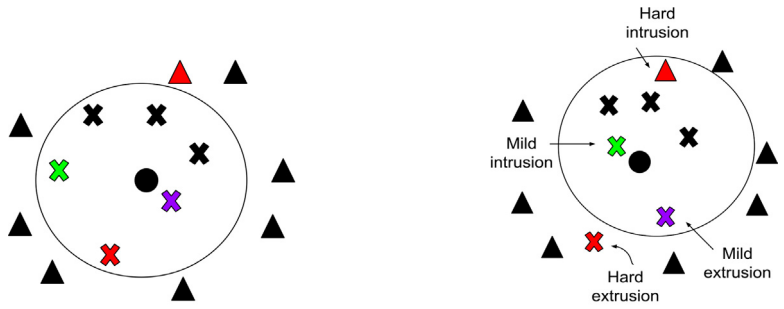
NIEQA measures embedding quality in the presence of both isometric and normalized embeddings and to do so it incorporates rotation, anisotropic scaling and translation. NIEQA also evaluates the quality of global structure preservation based on landmark points that represent the global skeleton of the manifold in the original dimension. The local measure between the landmark points and the corresponding embedded data points is considered as the global measure. After obtaining the local (Q_{local}) and the global measure (Q_{global}) the overall measure along with a compromise factor $\alpha \in [0, 1]$ is given as

$$Q_{overall} = \alpha Q_{local} + (1 - \alpha) Q_{global} \quad (1)$$

For both isometric and normalized embeddings, NIEQA gives a comparatively more accurate value. The determination of optimal orthogonal matrix for rotation involves an iterative and time-consuming process. The determination of all-pairs of shortest paths is also time-complex and threshold based landmark points selection is not sound.

2.2. Rank-preservation based techniques

On the other hand, rank-preservation based techniques keep the data in the embedded space intact. They comprehend the change of distance-based ranks of neighbourhood points for every point when high-dimensional data is embedded into a low-dimensional space. Rank errors arise out of two cases. One is intrusion of points in the neighborhood in the embedded space which are not present in the original space. Another is extrusion of points from neighborhood in embedded space which are present in the original space. These errors can be further sub-divided into two categories: Hard intrusion/ extrusion and Mild intrusion/ extrusion which are shown in Fig. 1. The notations used for following methods are given in Table 1.



(a) Ideal embedding of original space (b) Embedding obtained

Fig. 1. Ranking Errors: Hard Intrusion/ Extrusion and Hard Intrusion Extrusion.

Table 1
List of notations.

Abbreviations	Meaning
X	Original high-dimensional data of dimension D
Y	Lower dimensional embedding of X having dimension $d < D$
n	Number of data points in X and Y
k	Number of data points considered for neighborhood
x_i	ith data point in X
y_i	ith data point in Y
$\Phi(x_i)$	k-neighborhood data points of x_i
$\Phi(y_i)$	k-neighborhood data points of y_i
r_{ij}^X	Distance-based rank of x_j with respect to x_i
r_{ij}^Y	Distance-based rank of y_j with respect to y_i
x_{ij}	jth data-point in $\Phi(x_i)$
y_{ij}	jth data-point in $\Phi(y_i)$

2.2.1. Local continuity meta-criterion (LCMC) [17]

LCMC assess an embedding by measuring the overlap between neighborhoods in the original and the embedded space for every point. It is formulated as follows:

$$LCMC = 1 - \frac{1}{nk}\sum_{i=1}^{n}(|\Phi(x_i) \cap \Phi(y_i)| - \frac{k^2}{n-1}). \tag{2}$$

2.2.2. Trustworthiness and continuity (T&C) [18]

Trustworthiness checks whether there are intrusions and continuity checks whether there are extrusions in low-dimensional embedding.

$$T = 1 - \frac{2}{\rho}\sum_{i=1}^{n}\sum_{j\epsilon(\Phi(y_i)\setminus\Phi(x_i))}(r_{ij}^X - k); \tag{3}$$

$$C = 1 - \frac{2}{\rho}\sum_{i=1}^{n}\sum_{j\epsilon(\Phi(x_i)\setminus\Phi(y_i))}(r_{ij}^Y - k); \tag{4}$$

here $\rho = nk(2n - 3k - 1)$ is the normalizing constant and $\Phi(y_i)\setminus\Phi(x_i)$ denotes the set of points present in the neighborhood of embedded space but not present in the neighborhood of original space and $(\Phi(x_i)\setminus\Phi(y_i))$ denotes the reverse.

$$T\&C = \mu T + (1 - \mu)C \tag{5}$$

where $\mu \epsilon [0,1]$ is a compromise parameter which determines the effect of either metric on the overall result. *T&C* usually lies in the range [0,1] where a higher value represents a better quality of embedding.

2.2.3. Mean relative rank error (MRRE) [19]

MRRE consists of two quality evaluation metrics which calculates two distance-based ranking errors in lower and higher dimensional neighborhoods. Those are,

$$W_T = 1 - \frac{1}{H_k}\sum_{i=1}^{n}\sum_{j\epsilon\Phi(y_i)}\frac{|r_{ij}^X - r_{ij}^Y|}{r_{ij}^X}; \tag{6}$$

$$W_C = 1 - \frac{1}{H_k}\sum_{i=1}^{n}\sum_{j\epsilon\Phi(x_i)}\frac{|r_{ij}^X - r_{ij}^Y|}{r_{ij}^Y}; \tag{7}$$

where $H_k = n\sum_{i=1}^{n}\frac{|n-2i+1|}{i}$ is the normalizing constant. Unlike *T&C*, *MRRE* considers the points present in the embedded space but not in the original space for W_T and vice versa for W_C. *MRRE* gives an overall measure as follows:

$$MRRE = \beta W_T + (1 - \beta)W_C; \tag{8}$$

where β is a compromise parameter which lies in the range [0,1]. $MRRE \epsilon$ [0,1] where higher value indicates a good preservation of local neighborhood structure.

2.2.4. Co-ranking matrix (Q) [24]

The co-ranking matrix serves as a unifying framework for every possible rank error criterion. The cell Q_{pq} of $(n-1)\times(n-1)$ matrix Q represents the number of neighbourhood points for every point whose $r_{ij}^X = p$ and $r_{ij}^Y = q$. The diagonal cells of Q indicate the rank preserved points while extrusions and intrusions are represented by the upper diagonal cells and lower diagonal cells respectively. A scale-independent, co-ranking matrix based quality criterion $Q_{nX}(k)$ [20] is formulated as follows:

$$Q_{nX}(k) = \frac{1}{nk}\sum_{p=1}^{k}\sum_{q=1}^{k}Q_{pq} \tag{9}$$

Between these two genres of techniques, rank-based methods indicate the degree of pairwise distance preservation among neighbourhood points but fail to retrieve information about the transformation incorporated by the embedding algorithms. Whereas, transformation-based techniques can handle both but with higher computational power. In this context, the proposed method is a transformation based technique with significant improvement in novelty and complexity.

2.3. Statistical measures for global structure preservation

Lack of methods dedicated to measure global structure preservation in the embedded space has been motivational to explore statistical tools for that purpose. It can also be observed that data points in the original and embedded space are data distributed in D and d dimensional space respectively. Hence, the use of statistical tools will be prudent for measuring the global closeness between the original and embedded data points. There are two major categories of statistical tools that are useful for this purpose.

2.3.1. Divergence measures

Divergence is used to measure the similarity between two probability distributions. Now, two sets of data points can be represented as two distributions $p(x)$ and $q(x)$ and divergence measures can be applied to sense the similarity between them. Kullback-Leibler (KL) divergence [25] measures relative entropy or difference between two distributions and Pearson divergence [25] is defined as the squared-loss variant of the KL divergence. These non-negative divergence measures are zero if p and q are the same; otherwise their value increases with dissimilarity. Chernoff's measure of discriminatory information $C_\alpha(p, q)$ [25] also tries to discriminate two distributions.

2.3.2. Non-parametric statistical tests [26]

Non-parametric tests are preferred as they do not require prior knowledge about the distribution for comparison of two distributions. The non-parametric two-sample Kolmogorov Smirnov test determines the difference between two distributions using empirical cumulative distribution functions (CDF) of two data sets. The maximum difference between the two CDFs is taken as the test statistic. The Wilcoxon signed-rank test and Mann-Whitney U test are two rank based techniques for finding differences between two distributions. Both these methods are inclined to changes in median than they are to changes in shape of the distribution

The methods studied for global structure preservation are well established methods where each one has its own pros and cons and different application areas. In Section 3, an elaborate discussion is made on the applicability of the methods considered.

3. Proposed method

Dimensionality reduction methods analyse data points of original space in different ways, i.e., process data points locally using k-neighbourhood like LLE [7], process data points globally like t-SNE [10], or both like ISOMAP [8], to embed them in lower-dimensional space. To cope up with these diverse subjects, the proposed embedding quality evaluation method MEQA is robust in the sense that it can quantify the embedding quality in every possible way the higher dimensional points are functionally mapped. MEQA takes two different views of both the original and the embedded structure. It checks the local structure preservation initially and then analyses the global structure preservation and formulate the overall embedding quality assessment.

3.1. Local measure

The local measure captures how efficiently the neighbourhood structure has been preserved. To measure the neighbourhood preservation of Y with respect to X, Y should be sculpt in such a way that the difference with X is minimal. For this purpose, three transformations have been integrated by the proposed method between every $\Phi(x_i)$ and corresponding $\Phi(y_i)$. The transformations are translation, anisotropic scaling and rotation. Here, instead of isometric scaling, which fails to handle normalized embedding, anisotropic scaling has been incorporated. Anisotropic scaling scales a set of points by different scaling factors along each co-ordinate axis. Let us consider a d-dimensional vector $\vec{a}$ and a d-dimensional diagonal scaling matrix Q. Then, the diagonal elements of Q will be different for anisotropic scaling and they will be same for global scaling. In Fig. 2, the blue triangle is obtained by anisotropic scaling over the black triangle where $Q_{11} = 3$ and $Q_{22} = 0.5$. The red triangle is obtained by isometric scaling with $Q_{11} = Q_{22} = 2$.

The proposed algorithm considers, under ideal embedding, for every neighborhood $\Phi(x_i)$ in the original manifold, the embedded neighborhood $\Phi(y_i)$ should be obtained in such a way that, for all $x_{ij} \in \Phi(x_i)$ and $y_{ij} \in \Phi(y_i)$,

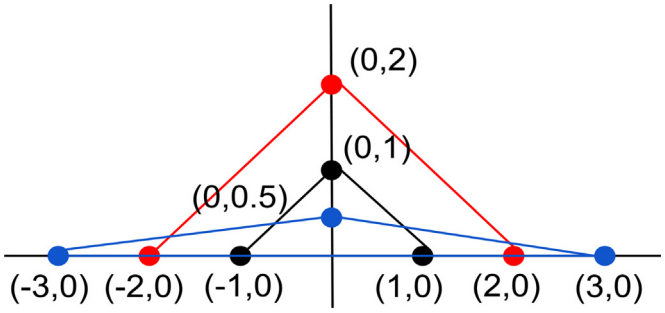


Fig. 2. Anisotropic scaling.

$$x_{ij} = P_i S_i y_{ij} + t_i; \qquad j = 1, 2, 3, \ldots, k \tag{10}$$

where P_i is a $D \times d$ orthogonal rotation matrix, S_i is a $d \times d$ diagonal matrix for anisotropic scaling and t_i is a D-dimensional translation vector. Therefore, the first task is to find out the optimal parameters P_i, S_i and t_i so that the best fit between $\Phi(x_i)$ and $\Phi(y_i)$ can be obtained. So, the following constrained optimization is performed

$$\min_{P_i, S_i, t_i} \sum_{j=1}^{k} ||x_{ij} - P_i S_i y_{ij} - t_i||^2 \tag{11}$$

such that $P_i^T P_i = I_d$ and $S_i \in Diag(d)$, where $Diag(d)$ denotes the set of all $d \times d$ diagonal matrices. Finally, the normalized sum of squared Euclidean distances between the data points in $\Phi(x_i)$ and the transformed data points of $\Phi(y_i)$ is obtained as follows. In other words, M_L is expressed as follows:

$$M_L(x_i, y_i) = \frac{||\Phi(x_i) - P_i S_i \Phi(y_i) - t_i e_k^T||_F^2}{||\Phi(x_i) H_k||_F^2} \tag{12}$$

where, e_k is a k-dimensional all-one vector and $H_k = I_k - \frac{1}{k} e_k e_k^T$. H_k, the centering matrix, translates the centroid to the origin. So, the four step determination of M_L will be determination of optimal translation vector t_i, optimal scaling matrix S_i, optimal rotation matrix P_i and ultimately determine $M_L(x_i, y_i)$. It is important to note that the optimization Eq. (11) does not admit a closed form solution. All $D \times d$ orthogonal matrices embedded in $\mathbb{R}^{Dd}$ form steifel manifold [27], which is a Riemannian submanifold [28]. Along with that $Diag(d)$ is closed for matrix addition, multiplication and scalar multiplication. Thus optimization of Eq. (11) can be solved by gradient-descent method and for that, numerator in the right hand side of Eq. (12) should be minimized. Hence, the problem statement can be restated as:

$$\min_{P_i, S_i, t_i} ||\Phi(x_i) - P_i S_i \Phi(y_i) - t_i e_k^T||_F^2. \tag{13}$$

3.1.1. Determination of translation vector

The objective function in Eq. (13), in terms of trace function $tr()$, will be :

$$F(P_i, S_i, t_i) = tr((\Phi(x_i) - P_i S_i \Phi(y_i) - t_i e_k^T)^T (\Phi(x_i) - P_i S_i \Phi(y_i) - t_i e_k^T)). \tag{14}$$

Partial derivative of this equation with respect to t_i gives

$$\frac{\delta F(P_i, S_i, t_i)}{\delta t_i} = 2kt_i - 2\Phi(x_i) e_k + 2P_i S_i \Phi(y_i) e_k. \tag{15}$$

Since, $F(P_i, S_i, t_i)$ is a strict convex function of t_i, therefore by equating both sides of expression (15) to zero the value of t_i is

$$t_i = \frac{1}{k}(\Phi(x_i) - P_i S_i \Phi(y_i)) e_k. \tag{16}$$

Therefore, substituting t_i in Eq. (13), the objective function takes the form:

$$\min_{P_i,S_i} ||\overline{\Phi}(x_i) - P_i S_i \overline{\Phi}(y_i)||_F^2 \tag{17}$$

where,

$$\overline{\Phi}(x_i) = \Phi(x_i)H_k, \tag{18}$$

$$\overline{\Phi}(y_i) = \Phi(y_i)H_k. \tag{19}$$

3.1.2. The δ operator

The δ operator over a m-dimensional vector $A = [a_1, a_2, \ldots, a_m]^T$ as $\delta(A)$, yields an m-dimensional diagonal matrix whose diagonal elements are elements of A.

$$\delta(A) = \begin{bmatrix} a_1 & 0 & 0 & \cdots & 0 \\ 0 & a_2 & 0 & \cdots & 0 \\ 0 & 0 & a_3 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & a_m \end{bmatrix}. \tag{20}$$

The δ operator over an m-dimensional square matrix V is an m-dimensional vector formed by the diagonal elements of V. Thus, $\delta(V)$ becomes

$$\delta(V) = \begin{bmatrix} v_{11} & v_{22} & v_{33} & \cdots & v_{mm} \end{bmatrix}^T. \tag{21}$$

The δ operator can also be compounded using the following equation.

$$\delta^2(A) = A. \tag{22}$$

Also, $\delta^2(V)$ is given as follows:

$$\delta^2(V) = \begin{bmatrix} v_{11} & 0 & 0 & \cdots & 0 \\ 0 & v_{22} & 0 & \cdots & 0 \\ 0 & 0 & v_{33} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & v_{mm} \end{bmatrix}. \tag{23}$$

3.1.3. Determination of diagonal matrix

To determine the optimal value of the diagonal scaling matrix S_i, Eq. (17) can be written as:

$$\begin{aligned} F(P_i, S_i) &= tr((\overline{\Phi}(x_i) - P_i S_i \overline{\Phi}(y_i))^T (\overline{\Phi}(x_i) - P_i S_i \overline{\Phi}(y_i))) \\ &= tr(S_i^2 \overline{\Phi}(y_i)\overline{\Phi}^T(y_i)) - 2 * tr(S_i P_i^T \overline{\Phi}(x_i)\overline{\Phi}^T(y_i)) \\ &\quad + tr(\overline{\Phi}(x_i)\overline{\Phi}^T(x_i)) \end{aligned} \tag{24}$$

Let $U_i = \overline{\Phi}(y_i)\overline{\Phi}^T(y_i)$ and $V_i = P_i^T \overline{\Phi}(x_i)\overline{\Phi}^T(y_i)$. Therefore

$$\begin{aligned} F(P_i, S_i) &= tr(S_i^2 U_i) - 2 * tr(S_i V_i) + tr(\overline{\Phi}(x_i)\overline{\Phi}^T(x_i)) \\ &= \textstyle\sum_{j=1}^{d} (s_{ij})^2 u_{ij} - 2 \sum_{j=1}^{d} s_{ij} v_{ij} + tr(\overline{\Phi}(x_i)\overline{\Phi}^T(x_i)) \end{aligned} \tag{25}$$

where u_{ij}, v_{ij} and s_{ij} denote the diagonal entries in the jth row of U_i, V_i and S_i respectively. Since $u_{ij} \geq 0$, $F(P_i, S_i)$ is a convex function of vector $\delta(S_i)$. Thus by taking partial derivative of Eq. (25) with respect to s_{ij} and equating the resulting derivative equation to zero, the optimal solution of s_{ij} becomes

$$s_{ij} = \frac{v_{ij}}{u_{ij}}. \tag{26}$$

Therefore, S_i is given as follows:

$$S_i = (\delta^2(U_i))^{-1} \delta^2(V_i). \tag{27}$$

Substitution of the value of S_i in Eq. (25) forms

$$F(P_i) = -\sum_{j=1}^{d} \frac{v_{ij}^2}{u_{ij}} + tr(\overline{\Phi}(x_i)\overline{\Phi}^T(x_i)). \tag{28}$$

Assuming $A_i = \overline{\Phi}(x_i)\overline{\Phi}^T(y_i)(\delta^2(U_i))^{-1/2}$, Eq. (28) can be written as

$$F(P_i) = -\sum_{j=1}^{d} ((P_i(j))^T A_i(j))^2 + tr(\overline{\Phi}(x_i)\overline{\Phi}^T(x_i)); \tag{29}$$

where $P_i(j)$, $A_i(j)$ denote the jth column of P_i and A_i respectively. Also, Eq. (29) can be written in terms of element-wise product (denoted by $\odot$). Now, minimization of $F(P_i)$ involves maximization of the first term of Eq. (29) as the second term is a constant. Therefore, the optimization equation is

$$\begin{aligned} &\max_{P_i} \quad \Psi(P_i) = tr((P_i^T A_i) \odot (P_i^T A_i)) \\ &such\ that \quad P_i^T P_i = I_d. \end{aligned} \tag{30}$$

3.1.4. Determination of orthogonal rotation matrix

As the orthogonal matrix P_i belongs to the Steifel manifold, $St(D, d)$, maximization of the function $\Psi()$ in Eq. (30) is done by applying gradient-descent over $St(D, d)$. Initially, the gradient of $\Psi()$ on $St(D, d)$ should be found to perform gradient-descent. Let this gradient denoted be $\nabla\Psi(P_i)$. To obtain the value of $\nabla\Psi(P_i)$, first the value of the gradient of $\Psi()$ in $\mathbb{R}^{Dd}$ at P_i, $\nabla\overline{\Psi}(P_i)$, should be found. By the proposition of Steifel manifold, $\nabla\Psi(P_i)$ is the projection of $\nabla\overline{\Psi}(P_i)$ onto the tangential space at P_i. Hence the following relation is valid:

$$\nabla\Psi(P_i) = \nabla\overline{\Psi}(P_i) - P_i \frac{P_i^T \nabla\overline{\Psi}(P_i) + (\nabla\overline{\Psi}(P_i))^T P_i}{2}. \tag{31}$$

By applying matrix calculus, the value of $\nabla\overline{\Psi}(P_i)$ can be determined to be

$$\nabla\overline{\Psi}(P_i) = 2A_i \delta^2(P_i^T A_i). \tag{32}$$

Therefore, substituting the value of $\nabla\overline{\Psi}(P_i)$ in Eq. (31) we get

$$\nabla\Psi(P_i) = 2A_i \delta^2(P_i^T A_i) - P_i P_i^T A_i \delta^2(P_i^T A_i) - P_i \delta^2(P_i^T A_i) A_i^T P_i. \tag{33}$$

3.1.5. Formulation of L_1

Therefore, $M_L(x_i, y_i)$ is calculated for every data point and the mean of them is found to quantify the local neighborhood structure preservation. So, the local measure L_1 is

$$L_1(X, Y) = \frac{1}{n} \sum_{i=1}^{n} M_L(x_i, y_i). \tag{34}$$

3.1.6. Obtaining the matrix Y'

While obtaining the local measure L_1, for every data point $M_L(x_i, y_i)$ is calculated and in the way, the optimal parameters P_i, S_i and t_i are obtained. Therefore, from Eq. (10) the transformed data point y_i' is obtained where

$$y_i' = P_i S_i y_i + t_i. \tag{35}$$

Hence, the transformed embedded matrix Y' in the original space becomes

$$Y' = [y_i']_{i=1,2,3,\ldots,n} \tag{36}$$

Now, the incorporated transformation is necessary to incur minimal difference which helped to deal with all types of embeddings but it yields some disadvantages. The incorporated transformation techniques negate the pairwise distance distortion in the neighborhood of embedded space to a great extent to reach as close as possible to the neighborhood of the original space. This trait makes it inefficient to discriminate between inferior and superior quality embedding. So, the desire to keep the advantageous trait and discard the detrimental one builds the basis for the next measurement.

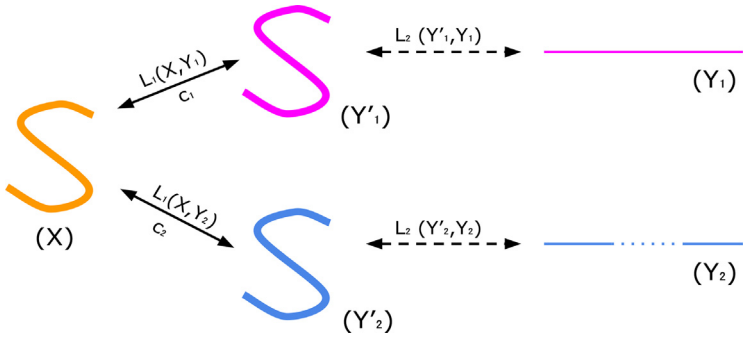


Fig. 3. The transformation pipeline for comparing a manifold and its embedding.

3.2. Error reduction of anisotropic scaling

During the calculation of L_1, a significant amount of flexibility has been imposed and due to that local neighbourhood structure in the embedded space gets distorted. In Fig. 3, X is the high dimensional datapoints which have two corresponding embeddings Y_1 and Y_2 where Y_1 maintains better neighbourhood structure compared to Y_2. Now, with the allowance of distortion, L_1 measure transforms unequal structures Y_1 and Y_2 to get similar structure like X, as shown by Y_1' and Y_2' and hence $L_1(X, Y_1)$ and $L_1(X, Y_2)$ will be same which ultimately depict a wrong fact.

So to rectify this error, the effort given by L_1 for transformation should be measured. Existing algorithms only measure how much neighbourhood structure has been preserved and thus fails to distinguish between embeddings which require varying amount of transformations to achieve the same reference neighbourhood structure. A good quality embedding requires a lesser effort for transformation to achieve the reference neighbourhood structure compared to an embedding which is subpar. So, MEQA measures how much flexibility was required by the data points in the lower dimensional space to get the high dimensional structure.

To quantify the transformation required for Y to reach Y', MEQA performs a rigid transformation of Y with respect to Y' and measures the difference. In Fig. 3, the differences measured through rigid transformation are represented by $L_2(Y_1', Y)$ and $L_2(Y_2', Y)$ for the embeddings Y_1 and Y_2 respectively. Rigid transformation is an efficient estimator of isometric embedding as it preserves pairwise distances and other geometries while achieving an optimal fit. For that reason how much transformation has taken place due to anisotropic scaling can be efficiently measured by rigid transformation. Proposed method performs rigid transformation and finds $M_R(y_i^{'}, y_i)$ to obtain a quantitative value of non-rigid distortion. Thus M_R serves as an interesting standpoint for adjusting the value of L_1.

$$M_R(y_i^{'}, y_i) = \frac{||\Phi(y_i^{'}) - \gamma_i P_i \Phi(y_i) - t_i e_k^T||_F^2}{||\Phi(y_i^{'}) H_k||_F^2} \tag{37}$$

where P_i, t_i and γ_i denote the optimal rotation matrix, translation vector and isometric scaling matrix respectively. Now, L_2 can be calculated as:

$$L_2 = \frac{1}{n}\sum_{i=1}^{n} M_R(y_i^{'}, y_i). \tag{38}$$

A high value of L_2 portrays that a higher neighbourhood transformation is imposed by the L_1 measure to obtain that optimal fit. With reference to Fig. 3, L_2 measure for Y_2 will be higher than Y_1 while L_1 will be same. Therefore, a combination of L_1 and L_2 will obtain a reliable measure of the local structure preservation of an embedding.

3.3. Global measure

Initially, it was mentioned that concentrating only on the local structure will be an insufficient and inappropriate exercise to properly measure the quality of an embedding. In an ideal embedding, the global structure is maintained in the embedded space and global measure L_3 in MEQA quantifies that preservation. So, Y', the structure achieved by L_1 is considered for measuring the global structure preservation with respect to X and finding the similarity. One advantage of Y' is that it has zero local neighbourhood distortion and for that L_3 will be able to capture the global structure preservation solely. Another advantage is both Y' and X are in the same dimension. Now, quantification of the global structure preservation is not the only criterion but also the measure should bond with the local measure to form the overall EQA. So, the global measure L_3 must be non-negative, bound and symmetric.

From the discussion in Section 2, it is clear that KL and PE divergence are non-negative but both are unbounded and non-symmetric whereas Chernoff distance is non-symmetric. So, instead of divergence measures, the non-parametric tests will be applicable as they satisfy all the criteria. From the perspective of global structure preservation, the distribution shape is the most useful and generic characteristic which will capture pertinent information about the distribution. Therefore, the global measure should be sensitive to the changes in the shape of the distribution. However, both Wilcoxon signed-rank test and Mann-Whitney U test are sensitive to the changes in the median as compared to changes in distribution shape but the KS test [29] is much more perceptive in that regard and also time-efficient while considering every point. Hence, the KS test will be a proper statistical tool for measuring global structure preservation.

3.3.1. The Kolmogorov-Smirnov(KS) test

The non-parametric KS test primarily computes the empirical Cumulative Distribution Function (CDF) of two given datasets to quantify the difference of two unknown data distributions. The empirical CDF, $F_n(a)$, of a dataset Z with n elements ($z_i \in Z$) is calculated as follows

$$F_n(a) = \frac{1}{n}\sum_{i=1}^{n} p(z_i \leq a); \tag{39}$$

where p is 1 if the argument is true else it is 0. On given two empirical CDFs, F_n^1 and F_n^2, the KS test computes the test statistic T_n as follows

$$T_n = \max_a |F_n^1(a) - F_n^2(a)|. \tag{40}$$

Hence, the KS test compares two empirical CDFs and finds the point of maximum difference. Suppose, Fig. 4a and b show the corresponding histograms of Data 1 and 2 respectively and Fig. 4c shows their empirical CDFs. The KS test detects the point

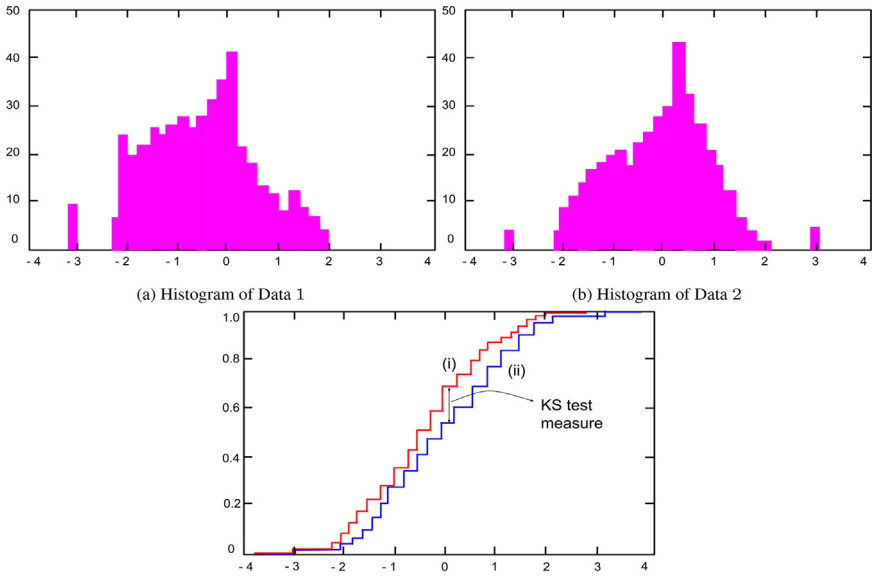


Fig. 4. The KS test on two sample data.

where the difference between the CDFs is maximum. Analysis of Eq. (40) shows that if the two distributions do not overlap, then T_n will be 1. On the contrary, if F_n^2 is the reference distribution from which x is drawn, then $T_n \to 0$ if $n \to \infty$. Therefore the smaller the value of T_n, the more closely the two sample distributions are and a value close to 1 exhibits considerable differences. Also, two distributions with the same sample mean but having different shapes will have a high value of T_n. Therefore the ability to measure the disparity between two dataset of the KS test will be the core of the proposed global measure.

3.3.2. Global measure

Y' is the closest structure Y can achieve and it is of the same dimension as X. So, the L_3 is computed between Y' and X. Now, the KS test is applicable for one-dimensional distributions only. Hence, the L_3 measure imposes a distributed KS test which is distributed over every feature space and by incorporating this approach the L_3 measure will be able to sense the differences in all the dimensions. Therefore, L_3 can be expressed as follows:

$$L_3 = \frac{1}{D}\sum_{j=1}^{D} T_n^{(j)} \tag{41}$$

where, $T_n^{(j)}$ is the KS test statistic evaluated for jth dimension. Now, for good quality embedding, the value of L_3 will be low otherwise it can be concluded that the global structure is less preserved.

3.4. Overall quality measure

The aim of the proposed method is to produce a single evaluation metric that represents the overall quality of an embedding. Now, there are three separate measures, L_1, L_2 and L_3, and all range in [0,1]. A bounded measure will be insightful for the judgement of an embedding and also for comparison. Therefore, a combination of these three measures is very important. Now, in another way L_2 and L_3 are dependent on the performance of L_1. So L_1 will carry more weightage for overall measurement. For L_2, which provides a flexibility check on the measure, should also have a significant effect on the overall quality measure but that has a biased stance towards isometric embedding. Therefore, L_2 will have lesser weight than L_1. Though good local structure preservation indicates good global structure preservation, nullification of every transformation of the DR method is practically impossible. Therefore, the significance of L_3 on the overall measure will be significant. So, keeping the above analysis in mind, every measure should be multiplied with weights, such that the overall measure $M_{overall}$ is in [0,1].

$$M_{overall} = \alpha_1 L_1 + \alpha_2 L_2 + \alpha_3 L_3 \tag{42}$$

where $\alpha_1, \alpha_2, \alpha_3 \in [0.1]$ and $\alpha_1 + \alpha_2 + \alpha_3 = 1.0$. Thus, from here onward, Eq. (42) will be used as the expression to generate the final evaluation metric of MEQA. The weights $\alpha_1, \alpha_2 \& \alpha_3$, associated with L_1, L_2 and L_3 respectively, can be seen as control parameters. So, α_1 will control the contribution of local structure preservation or L_1 in $M_{overall}$. α_2 will control the contribution of the penalization procedure or L_2 in $M_{overall}$. α_3 will control the contribution of the global structure preservation or L_3 in $M_{overall}$.

3.5. Algorithm of the proposed method

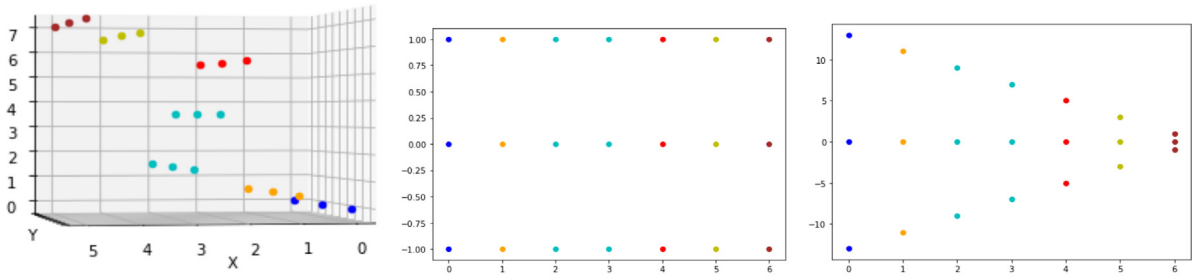
Algorithm 1 follows the intuition behind MEQA and also the three phase calculation of $M_{overall}$. Initially, the local measure L_1 and then the L_2 and L_3 measures are calculated and ultimately $M_{overall}$ is derived.

4. Experimental setup and performance analysis

This section elaborates and analyses the performance of the proposed method on a number of datasets and compares with EQA methods discussed in Section 2. The experiments are done on a 64-bit system with Intel core-i7 processor and 8GB RAM. Depending on the intuition and to give a balanced insight of embedding, for the experiment $\alpha_1 = 0.7, \alpha_2 = 0.2$ & $\alpha_3 = 0.1$. A detailed analysis is presented in Section 4.4.

4.1. Performance analysis on toy dataset

For the sake of presenting the proposed argument with more clarity, a toy 3D s-curve dataset, containing only 21 points, shown



(a) Toy dataset (b) Rectangular representation (c) Skewed trapezoidal representation

Fig. 5. Toy dataset and representations.

Algorithm 1: Manifold Embedding Quality Assessment Method MEQA.

Input: $D \times n$ matrix X, $d \times n$ matrix Y, neighborhood k, threshold ϵ

Result: $M_{overall}$, the embedding quality measure for matrices X and Y

for $i \leftarrow 1$ *to* n **do**

$\overline{\Phi}(x_i) = \Phi(x_i)(I_k - \frac{1}{k}e_k e_k^T)$

$\overline{\Phi}(y_i) = \Phi(y_i)(I_k - \frac{1}{k}e_k e_k^T)$

Calculate the thin SVD of $\overline{\Phi}(x_i)\overline{\Phi}^T(y_i) = LMN^T$

Initial value of $P_i = LN^T$

Use Equation 33 to calculate $\nabla\Psi(P_i)$

while $(||\nabla\Psi(P_i)||_F < \epsilon)$ **do**

$P_i \leftarrow P_i + \alpha\nabla\Psi(P_i)$

Compute the QR decomposition of $P_i = Q_i R_i$.

$P_i = Q_i$

Calculate S_i using Equation 27

Calculate t_i from Equation 16

Calculate $M_L(x_i, y_i)$ using Equation 12

$y_i' = P_i S_i y_i + t_i$

$L_1(X, Y) = \frac{1}{n}\sum_{i=1}^{n} M_L(x_i, y_i)$

for $j \leftarrow 1$ *to* n **do**

$\overline{\Phi}(y_j') = (I_k - \frac{1}{k}e_k e_k^T)\Phi^T(y_j')$

$\overline{\Phi}(y_j) = (I_k - \frac{1}{k}e_k e_k^T)\Phi^T(y_j)$

Concatenate $\overline{\Phi}(y_j)$ with $(D-d)$ columns of zeroes to obtain a $k \times D$ matrix $\overline{\Phi}(y_j'')$

$A = \overline{\Phi}^T(y_j')\overline{\Phi}(y_j'')$

Calculate the thin SVD of $A = USV^T$

Calculate $P_i = VU^T$

Calculate $\gamma_i = tr(\overline{\Phi}(y_j')P_i\overline{\Phi}(y_j'')/tr(\overline{\Phi}^T(y_i'')\overline{\Phi}(y_i''))$

Calculate $t_i = \frac{1}{k}(\Phi(y_i') - P_i\gamma_i\Phi(y_i''))e_k$

Calculate $M_R(y_i', y_i)$ using equation 37

$L_2 = \frac{1}{n}\sum_{j=1}^{n} M_R(y_i', y_i)$

for $j \leftarrow 1$ *to* D **do**

Empirical CDF for $X : F_n^{(1)}(a) = \frac{1}{n}\sum_{i=1}^{n} p(x_i \le a)$ [From Equation 39]

Empirical CDF for $Y' : F_n^{(2)}(a) = \frac{1}{n}\sum_{i=1}^{n} p(y_i' \le a)$ [From Equation 39]

Calculate $T_n^{(j)}$ from F_n^1 and F_n^2 using Equation 40.

$L_3 = \frac{1}{D}\sum_{j=1}^{D} T_n^{(j)}$

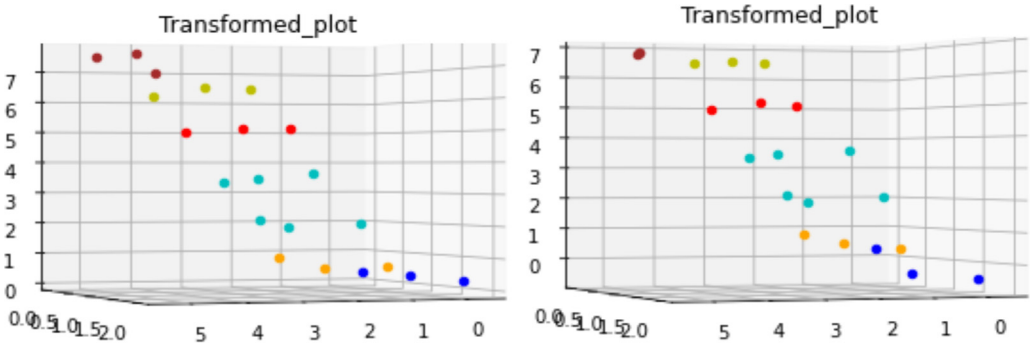
$M_{overall} = \alpha_1 L_1 + \alpha_2 L_2 + \alpha_3 L_3$

in Fig. 5a, have been created. Along with that two 2D representations of the toy dataset have been created. One has a rectangular shape (Fig. 5b), which is close to the ideal low-dimensional embedding. The other one has a skewed trapezoidal shape (Fig. 5c). Now, the L_1 measures for the rectangular and the skewed trapezoidal representations are 0.157 and 0.176 respectively. It is quite evident that due to the incorporation of anisotropic scaling, L_1 measure assigns nearly the same value; with a gap of 0.02. This is true that both the structures can achieve the same structure and this fact is analogous with the characteristic of L_1. Now, the L_2 measure between the rectangular and the skewed trapezoidal representations and their corresponding transformed plots (Fig. 6) are 0.06 and 0.38 respectively. Hence, the L_2 measure takes account of the effort or transformation provided by L_1 measure and compensates the deficiency of the L_1 measure. The L_3 measure between the toy dataset and the transformed plots depicts 0.19 and 0.19. As the transformed plots of both of these representations are similar (Fig. 6a and b), the global measures are also same.

Therefore, in cases like these, where the L_1 measure checks whether an embedding can achieve a structure but fails to provide a suitable discrimination, the L_2 measure serves to determine superiority of an embedding.

4.2. Dataset description and performance analysis

The experiment is done on five 3D benchmark synthetic datasets, i.e., S-Curve, Swiss Roll, Sphere, Helix and Twin Peaks with 3000 datapoints (Fig. 7) and one real life dataset COIL (Columbia Object Image Library) [30]. COIL consists 72 (128×128) grey-scale images for each of the 20 objects. For proper analysis of the performance by every EQA method, both original and lower dimensional structures of the dataset must be known. The synthetic datasets are well discussed in the literature and global structure of the COIL dataset is known, made them proper candidates for the discussion. In order to present a thorough comparative analysis, performance of MEQA is compared with every EQA methods discussed in Section 2. The embeddings (Y) generated by feature extraction methods [6–8,10,12,31] for each five datasets are shown in Figs. 8–12 along with corresponding transformed plots (Y') found by L_1 measure. Tables 2 to 6 show the results of MEQA against the results of the other EQA methods. The performance analysis will be both visual and intuitional due to the absence of ground truth or ideal embedding. First of all, whether the local structure of the original structure has been retained in the embedding should be checked. In the original space, the data points have been given color in a distinguishable manner. So, in an embedding, if two color bands overlap with each other, then there is local neighborhood distortion and the embedding is not proper. Next, for the global quality assessment, similarity between the transformed plots (Y') and the original manifold (X) should be



(a) Rectangular (b) Skewed trapezoidal

Fig. 6. Transformed plot of representations.

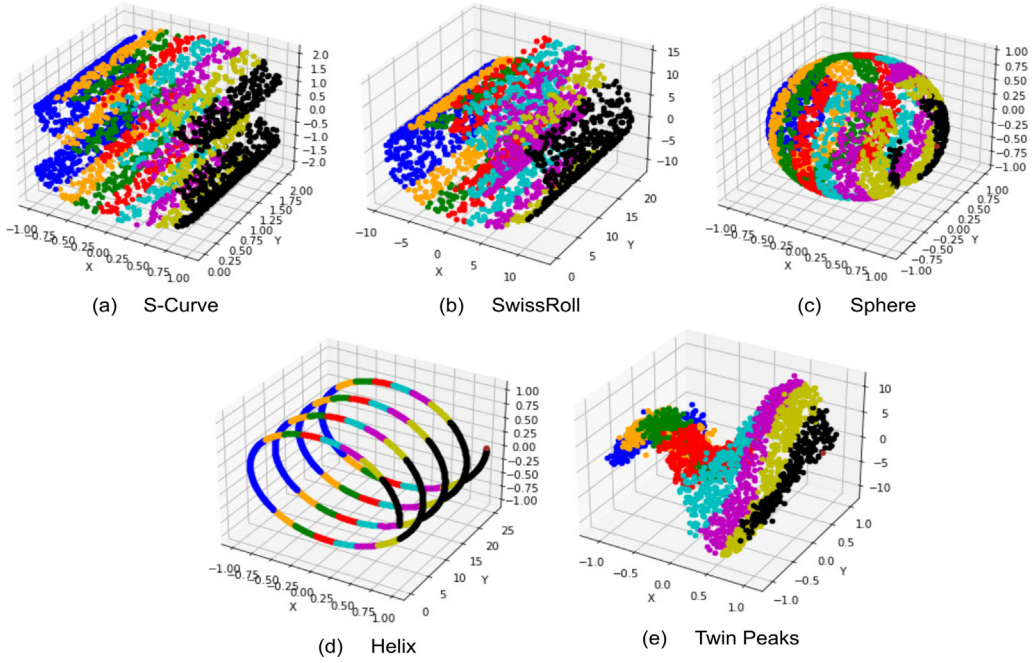


Fig. 7. Benchmark 3D synthetic datasets.

Table 2
Results of EQA methods for S-Curve dataset.

Methods	L_1	L_2	L_3	MEQA	P	P_L	*NIEQA*	*LCMC*	$Q_{nX}(k)$	*T& C*	*MRRE*
ISOMAP	0.00981	0.00323	0.01444	**0.00896**	0.2879	0.0029	0.2396	0.0331	0.932	0.999	0.999
LLE	0.00257	0.27209	0.01289	**0.0575**	0.6704	0.6433	0.2918	0.7257	0.571	0.996	0.997
t-SNE	0.08720	0.03157	0.01845	**0.0692**	0.6643	0.0453	0.2187	0.1723	0.837	0.998	0.998
UMAP	0.14849	0.07465	0.02234	**0.1211**	0.4901	0.5133	0.2940	0.5962	0.392	0.978	0.98
MDS	0.16708	0.88290	0.01322	**0.2949**	0.1258	0.8862	0.2578	0.5604	0.444	0.966	0.966
Auto-En	0.33566	0.50571	0.02289	**0.3384**	0.4931	0.1694	0.3336	0.2365	0.777	0.998	0.998

Table 3
Results of EQA methods for Swiss Roll dataset.

Methods	L_1	L_2	L_3	MEQA	P	P_L	*NIEQA*	*LCMC*	$Q_{nX}(k)$	*T& C*	*MRRE*
ISOMAP	0.16935	0.01417	0.01489	**0.1229**	0.8937	0.0029	0.4376	0.0309	0.937	0.999	0.999
LLE	0.16400	0.25186	0.01556	**0.1667**	0.7647	0.2308	0.4341	0.3768	0.612	0.997	0.997
t-SNE	0.23719	0.03763	0.01578	**0.1751**	0.8499	0.0444	0.3465	0.1688	0.838	0.998	0.998
UMAP	0.28119	0.07298	0.01911	**0.2133**	0.9253	0.1524	0.2796	0.228	0.767	0.997	0.998
MDS	0.30227	0.69565	0.00132	**0.3524**	0.2918	0.6972	0.3967	0.4905	0.515	0.961	0.963
Auto-En	0.59578	0.80203	0.02600	**0.5801**	0.3952	0.8022	0.4183	0.838	0.128	0.888	0.895

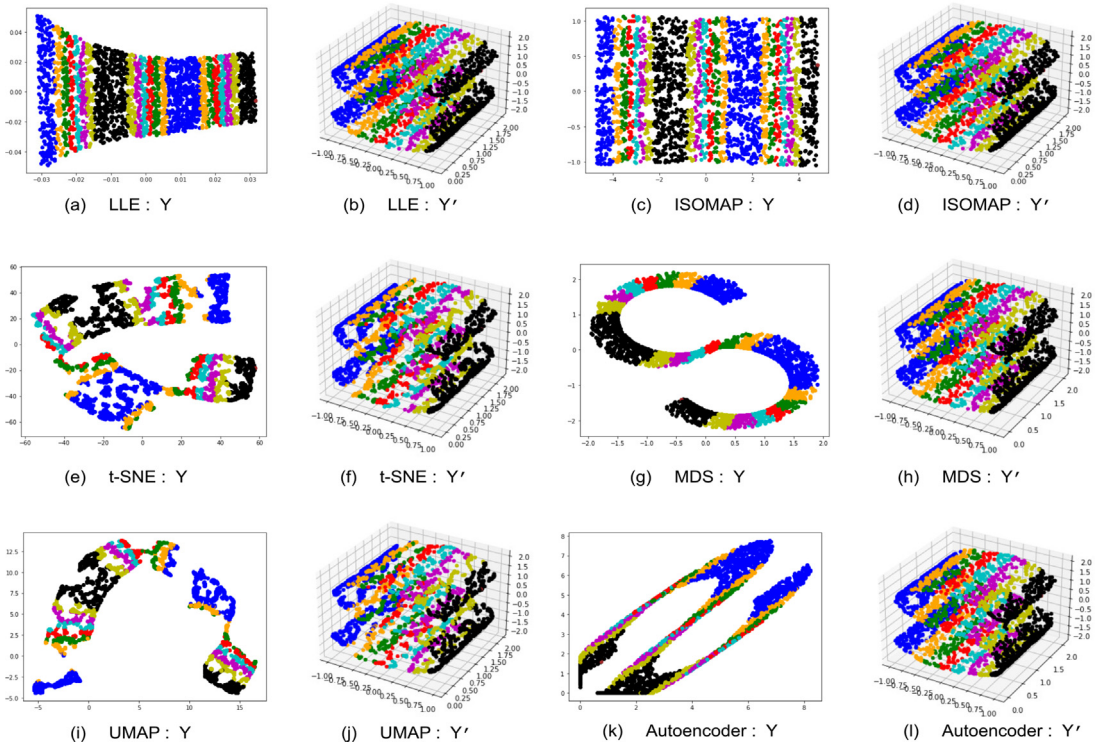


Fig. 8. S-Curve embeddings and transformed plots: The embeddings obtained by the DR methods are represented by Y and the transformed plots are represented by $Y^{'}$.

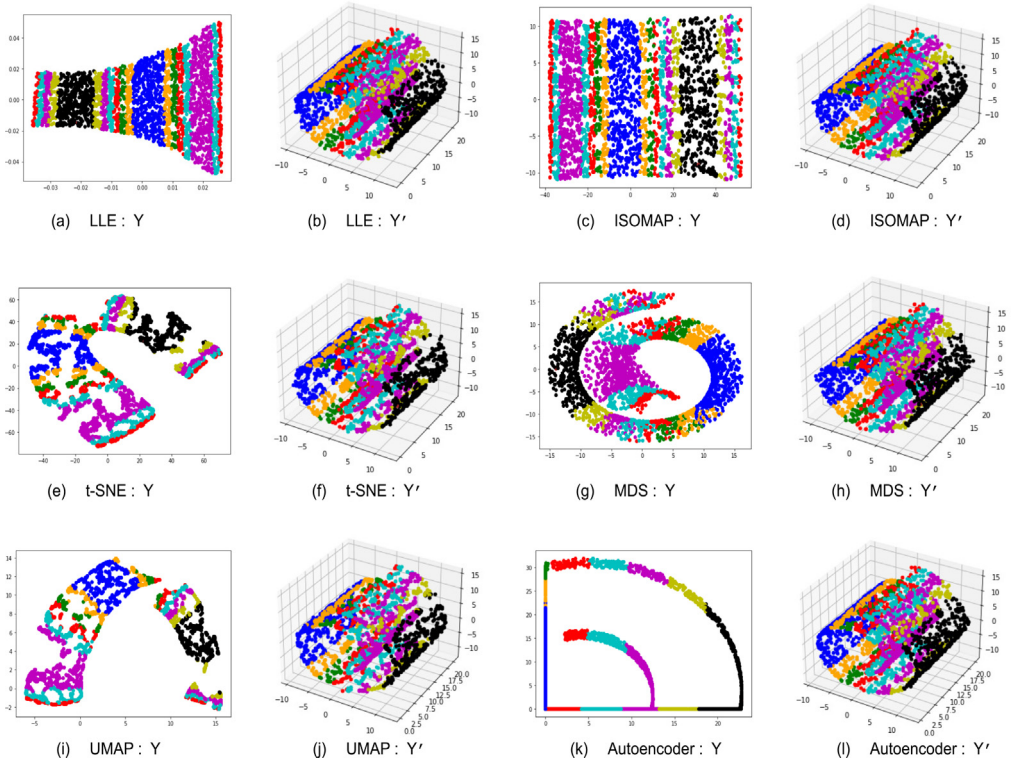


Fig. 9. Swiss Roll embeddings and transformed plots: The embeddings obtained by the DR methods are represented by the matrix Y and the transformed plots are represented by the matrix $Y^{'}$.

checked. As discussed earlier, X and Y' should perfectly coincide for an ideal embedding.

4.2.1. S-Curve Dataset

In Fig. 8b, ISOMAP preserves the inter-point distinction as well as the order of color bands and there is no discernible overlap between adjacent color bands, thus ISOMAP preserves the local structure. The transformed plot Y' (Fig. 8d) matches the orginal structure too. From Fig. 8a, it is visible that the color bands are distinct and there is no overlap between adjacent color bands; but the bands are not equally spaced which infers distortion in local neighborhood. Due to anisotropic scaling of L_1 measure both achieve similar structures alike (Fig. 8b and d), but the L_2 measures the effort and senses Fig. 8c to be better quality embedding than 8 a. However, from Fig. 8e, it is clear that although the distinction between the color bands is maintained and the overlap is minimal, the uniform spread of the original manifold is lost and it is not continuous. The transformed plot (Fig. 8f) shows distinguishable difference with the original structure. The local structure of UMAP embedding (Fig. 8j) is quite similar to Fig. 8e with further loss of uniformity and continuity. The embedding generated by MDS (Fig. 8g) has no overlap between the color bands but the color bands are of varying width, which indicates clear loss of intrinsic structure. This distortion in local neighborhood caused drop of quality in global scale. But it has an comparatively improved local structure in uniformity and continuity. The lowest quality embedding among all the algorithms considered is achieved by Autoencoder (Fig. 8k).

In Table 2, few transformation based techniques evaluated the embeddings partially correct but all failed to provide the correct order of rank. Due to the inability of judging normalized embeddings correctly, all rank preservation based techniques provide the least discrimination between the embeddings. But it can be seen that the proposed method analyses the quality of the embeddings with properly.

4.2.2. Swiss roll dataset

In Table 3, the Procrustes measure (P) and *NIEQA* erroneously quantifies ISOMAP and LLE as bad embeddings, where they preserve the manifold structure very well, whereas it senses embeddings generated by MDS and Auto-Encoder as superior. Local procrustes measure (P_L) evaluates LLE as bad quality embedding too. The remaining EQA methods also fail to provide reasonable measurement, and therefore do not adhere to the true ranking order of the embeddings. For the Swiss Roll dataset, MEQA is the only method that maintains the true order of the embeddings quality, thereby proving its effectiveness as a reliable EQA method.

4.2.3. Sphere dataset

Looking at Fig. 10a and c, the points in left and middle part of the LLE embedding are placed closer compared to the original structure. For the ISOMAP embedding, datapoints are placed sparsely. So from this perspective, the local structure deviation of these two embeddings are quantitatively similar. The Sphere whose top and bottom is hollowed and the front has no datapoints. Therefore, upon unfolding the structure, each side of the embedding should be convex. This feature is only present in the LLE embedding and the black color band in the right side is similar to the original structure, whereas these traits are missing in other embeddings.

Here too, as evident from Table 4 the proposed method is the only EQA method which maintains the true ranks of the algorithms based on the quality of their embeddings which proves its superiority.

4.2.4. Helix dataset

The ideal embedding of the Helix dataset would be a straight line. Fig. 11a shows that LLE achieves a good quality embedding with a small local neighborhood distortion in the centre, where the color bands are of disproportionate size compared to the original structure. ISOMAP embedding (Fig. 11c) maintains the overall shape of the embedding with a noticeable amount of local distortion as the color bands appear more compressed and not commensurate with the original manifold. UMAP (Fig. 11i) performs well in preserving the local structure but fails to preserve the overall structure. Similar to the UMAP, the overall structure of the MDS embedding (Fig. 11g) also deviates considerably from the linear structure and not uniformly spaced. The t-SNE embedding (Fig. 11e) not only fails to maintain local structure but also the global structure and the embedding achieved by the Autoencoder (Fig. 11k) have severe distortion compared to the original structure.

In case of Helix dataset as well, the analysis done by the proposed method (Table 5) conforms to the true ranking of the manifold learning algorithms.

4.2.5. Twin peaks dataset

In this instance as well, the proposed method is the only method that truly evaluates the manifold learning algorithms, thereby proving its effectiveness as an EQA.

4.2.6. COIL-20 dataset

In, COIL dataset each object was placed on a motorized turntable against a black background and images of the object were recorded with intervals of 5 degrees. So as shown in [32] the lower dimensional embedding should be in a circular shape. For the sake of demonstration, one object is considered and embeddings of different DR methods are shown in Fig. 13. The average of qualitative measures for 20 objects by MEQA along with other EQA methods are given in Table 7. Similar to synthetic datasets, MEQA is the only EQA measure which strictly abides by the true evaluation criteria and orders the manifold-learning algorithms perfectly. Hence, this analysis concludes that MEQA is not only applicable for synthetic datasets but can be reliably applied on real-life datasets too.

4.3. Quantitative study of the proposed method

To establish the effectiveness of *MEQA*, alongside the qualitative study, here a quantitative study is presented. For understanding the quantitative property of *MEQA*, its performance is studied over 3 datasets and is presented in Table 8. Using three DR methods: t-SNE, LLE and ISOMAP, the data points are embedded in different lower dimensional spaces which is depicted in second column of the Table 8 and *MEQA* is obtained. Now, in those lower dimensional spaces and the original space, the classification accuracy of three classification methods; K nearest neighbour(KNN), Support Vector Machine(SVM) and Random Forest(RF) on those datasets are obtained. Now, by intuition, it is understandable that if accuracy increases with dimension reduction then the manifolds are better unfolded in the lower dimension than the original space and *MEQA* should reflect that also and vice versa.

From the study, it is visible that in most of the cases with the reduction in dimensionality the classification accuracy is dropped monotonically and analogously *MEQA* measures increased, this scenario implies that *MEQA* senses that the manifold structures are not contained in the lower dimension which is also the reason behind low accuracy. Now, when LLE is applied on Wilt dataset, the classification accuracies over three classification algorithms do not change with a reduction in dimension and with that *MEQA* remains the same also. So, this shows the quantitative quality of the proposed method *MEQA*.

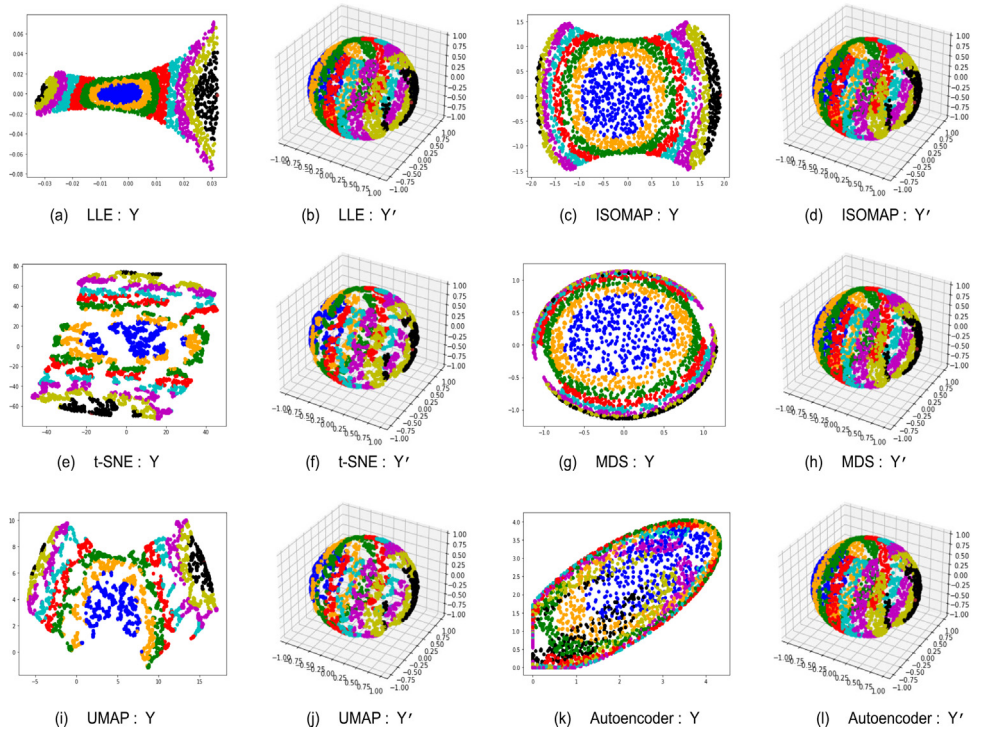


Fig. 10. Sphere embeddings and transformed plots: The embeddings obtained by the DR methods are represented by the matrix Y and the transformed plots are represented by the matrix $Y^{'}$.

Table 4
Results of EQA methods for Sphere dataset.

Methods	L_1	L_2	L_3	MEQA	P	P_L	$NIEQA$	$LCMC$	$Q_{nX}(k)$	$T\&C$	$MRRE$
LLE	0.03136	0.09791	0.00766	**0.0423**	0.3604	0.2653	0.2792	0.333	0.807	0.999	0.999
ISOMAP	0.05059	0.11828	0.00744	**0.0598**	0.3043	0.7414	0.2598	0.4891	0.775	0.999	0.998
t-SNE	0.07809	0.02942	0.00888	**0.0614**	0.4825	0.0396	0.4336	0.1787	0.826	0.998	0.999
UMAP	0.12828	0.04949	0.01255	**0.101**	0.5193	0.1163	0.3302	0.2062	0.79	0.998	0.998
MDS	0.17524	0.34681	0.00955	**0.193**	0.327	0.3561	0.2966	0.4154	0.591	0.992	0.993
Auto-En	0.23227	0.84588	0.0101	**0.3328**	0.4014	0.8519	0.3998	0.5724	0.392	0.931	0.93

Table 5
Results of EQA methods for Helix dataset.

Methods	L_1	L_2	L_3	MEQA	P	P_L	$NIEQA$	$LCMC$	$Q_{nX}(k)$	$T\&C$	$MRRE$
LLE	0.00015	0.00015	0.00866	**0.001**	0.5052	0.0002	0.2272	0.006	1	1	0.9998
ISOMAP	0.00019	0.00019	0.01022	**0.0012**	0.0183	0.0034	0.1999	0.006	1	1	0.9998
UMAP	0.00541	0.01447	0.04101	**0.0051**	0.7015	0.0074	0.1373	0.0315	0.9742	0.9999	0.9998
MDS	0.00378	0.00541	0.00466	**0.006**	0.0102	0.0045	0.1507	0.0073	0.9999	0.9998	0.9961
t-SNE	0.01268	0.00268	0.00755	**0.0102**	0.9141	0.0017	0.2779	0.0239	0.9941	0.9998	0.9997
Auto-En	0.01334	0.88014	0.00955	**0.1863**	0.9571	0.8802	0.418	0.7593	0.247	0.7724	0.7719

Table 6
Results of EQA methods for Twin Peaks dataset.

Methods	L_1	L_2	L_3	MEQA	P	P_L	$NIEQA$	$LCMC$	$Q_{nX}(k)$	$T\&C$	$MRRE$
t-SNE	0.20568	0.08596	0.02257	**0.1634**	0.6763	0.1676	0.5151	0.2039	0.801	0.997	0.998
MDS	0.16072	0.35285	0.01712	**0.1848**	0.0122	0.4843	0.2622	0.3096	0.695	0.997	0.996
UMAP	0.28060	0.19998	0.02901	**0.2393**	0.3917	0.4692	0.1871	0.3486	0.711	0.997	0.997
ISOMAP	0.24459	0.66401	0.00226	**0.3063**	0.0538	0.6018	0.1688	0.4334	0.507	0.991	0.99
LLE	0.34660	0.46550	0.01489	**0.3372**	0.5537	0.7476	0.662	0.6252	0.495	0.9888	0.99
Auto-En	0.88084	0.93654	0.03079	**0.807**	0.1838	0.9357	0.3227	0.6146	0.407	0.966	0.967

4.4. The proposed method's performance under different parameter settings

The performance of *MEQA* is studied across several parameter settings to assess not only the method's stability but also the importance of each individual measure. The performance of *MEQA* is investigated for each of the 6 datasets and 6 feature extraction methods using 10 parameter settings.

Table 9 shows that the proper order of assessment, as discussed in Section 4.2.1, is achieved for most parameter settings. In every case, ISOMAP is perceived as the best method for preserving the original structure in lower dimensional space, and other methods

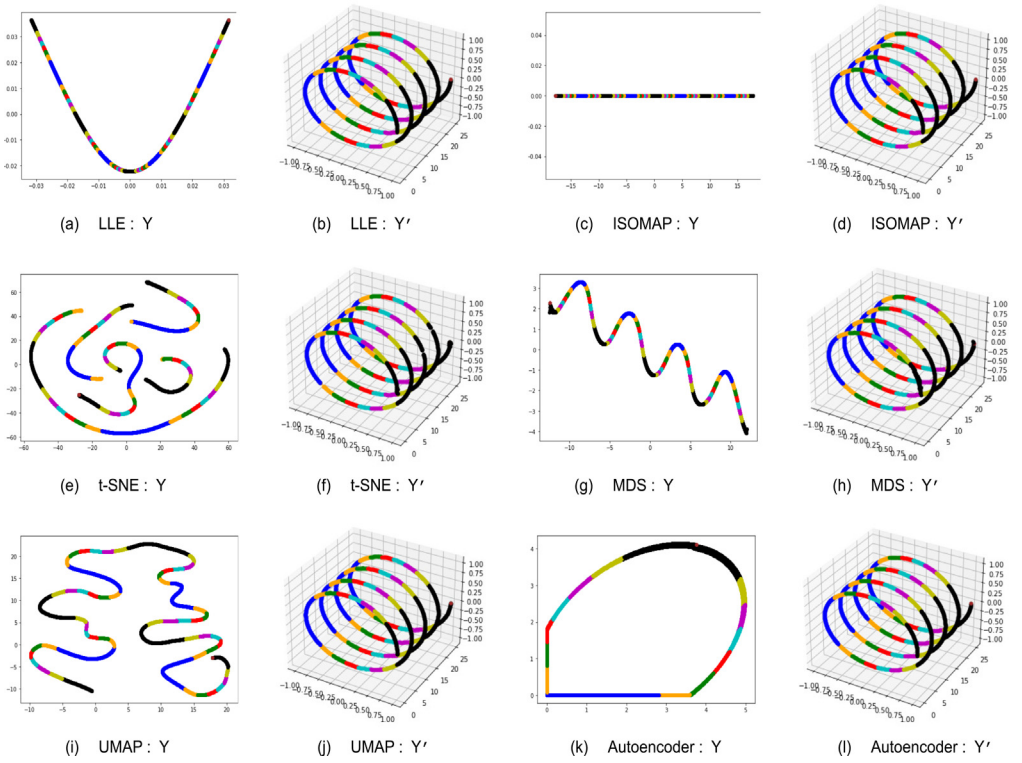


Fig. 11. Helix embeddings and transformed plots: The embeddings obtained by the DR methods are represented by the matrix Y and the transformed plots are represented by the matrix $Y^{'}$.

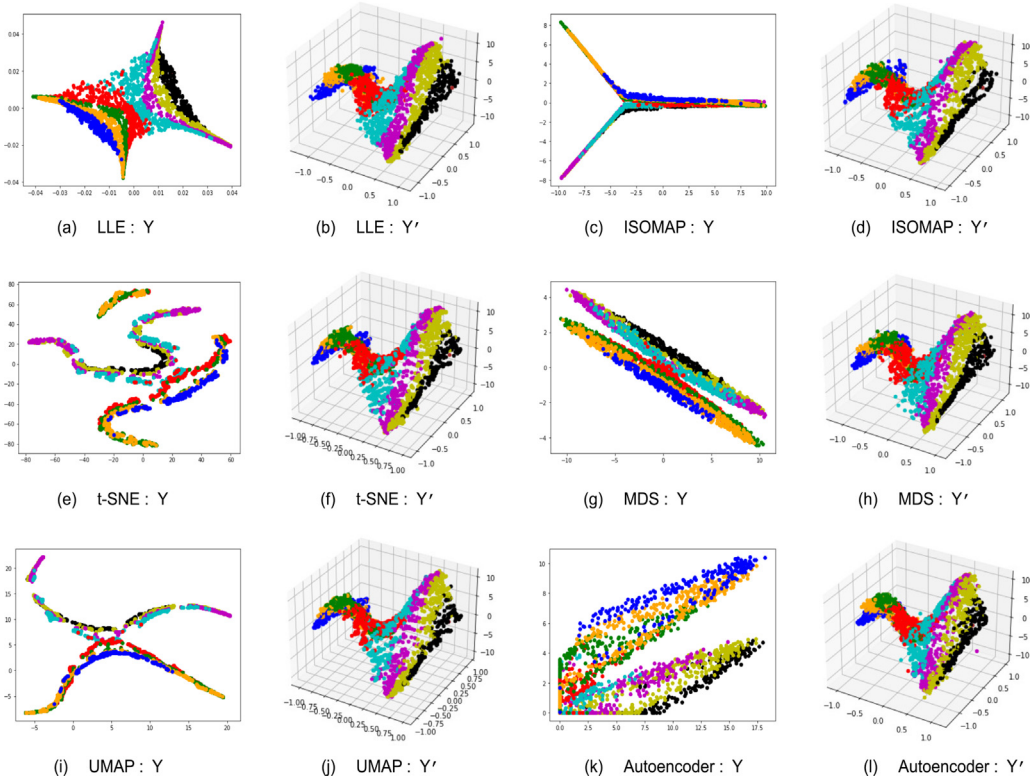


Fig. 12. Twin Peaks embeddings and transformed plots: The embeddings obtained by the DR methods are represented by the matrix Y and the transformed plots are represented by the matrix $Y^{'}$.

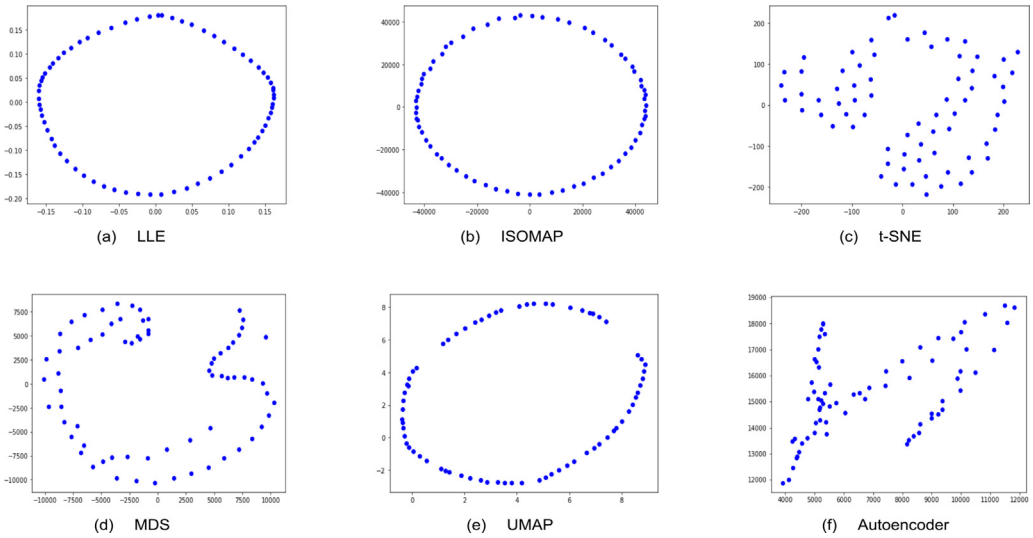


Fig. 13. COIL-20 Embeddings: The embeddings are generated by different DR methods.

Table 7
Results of EQA methods for COIL-20 dataset.

Methods	L_1	L_2	L_3	MEQA	P	P_L	$NIEQA$	$LCMC$	$Q_{nX}(k)$	$T\&C$	$MRRE$
LLE	0.31601	0.10521	0.15020	**0.2573**	0.5133	0.5395	0.3952	0.4598	0.7973	0.9619	0.8798
ISOMAP	0.31953	0.15696	0.14032	**0.2691**	0.4946	0.5908	0.3621	0.4512	0.806	0.971	0.854
UMAP	0.38264	0.13443	0.13602	**0.3083**	0.6866	0.3738	0.5996	0.3816	0.8764	0.9837	0.9494
MDS	0.40154	0.25142	0.13208	**0.3446**	0.5069	0.3698	0.4421	0.4363	0.8568	0.972	0.9178
t-SNE	0.48937	0.19602	0.13919	**0.3957**	0.5839	0.4989	0.4063	0.4426	0.8192	0.9411	0.8917
Auto-En	0.45463	0.44726	0.13636	**0.4213**	0.6979	0.667	0.523	0.6742	0.5829	0.8329	0.7502

Table 8
Quantitative study of *MEQA* over three classification algorithms with three DR methods.

Dataset		t-SNE				LLE				ISOMAP			
Name	Dimension	KNN	SVM	RF	*MEQA*	KNN	SVM	RF	*MEQA*	KNN	SVM	RF	*MEQA*
Wilt	4839 x 5, 2 class	96.28	80.68	80.68	0.0345	96.28	80.68	80.68	0.052	96.28	80.68	80.68	0.0544
	4839 x 3, 2 class	71.38	80.68	80.68	0.5070	79.33	80.68	80.68	0.5901	80.37	80.68	80.68	0.5766
	4839 x 2, 2 class	71.69	80.68	80.68	0.5599	80.47	80.68	80.68	0.599	78.30	80.68	80.68	0.619
Yeast	1484 x 8, 10 class	70.03	63.63	63.29	0.1285	70.03	63.63	63.29	0.0768	70.03	63.63	63.29	0.0815
	1484 x 4, 10 class	16.49	34.34	13.13	0.2289	21.54	25.92	24.24	0.2537	40.40	51.51	45.11	0.2654
	1484 x 2, 10 class	26.26	32.65	32.65	0.4940	21.54	20.53	21.21	0.5594	25.25	35.01	41.41	0.5431
Spectrometer	531 x 100, 4 class	71.96	67.28	66.35	0.023	71.96	67.28	66.35	0.0289	71.96	67.28	66.35	0.1899
	531 x 50, 4 class	36.44	50.46	45.79	0.6933	40.18	57.94	46.72	0.5639	48.59	49.53	57.94	0.5046
	531 x 25, 4 class	28.03	44.85	42.99	0.6868	36.44	54.20	43.92	0.5452	41.12	48.59	62.61	0.5051
	531 x 2, 4 class	41.12	37.38	42.99	0.5359	28.03	28.97	39.25	0.5415	37.38	45.79	45.79	0.5857

Table 9
MEQA measure on S-Curve dataset with different parameter settings.

α_1	α_2	α_3	LLE	ISOMAP	t-SNE	MDS	UMAP	Auto-En
0.3	0.7	0	0.1912	0.0052	0.0482	0.6681	0.0968	0.4545
0.5	0.5	0	0.1373	0.0065	0.0593	0.5249	0.1115	0.4206
0.7	0.3	0	0.0834	0.0078	0.0705	0.3818	0.1263	0.3866
0.2	0.7	0.1	0.1922	0.0056	0.0413	0.6527	0.0841	0.4234
0.45	0.45	0.1	0.1248	0.0073	0.0552	0.4738	0.1026	0.3809
0.7	0.2	0.1	0.0575	0.0089	0.0692	0.2948	0.1211	0.3383
0.2	0.6	0.2	0.0585	0.0094	0.0623	0.2794	0.1084	0.3071
0.4	0.4	0.2	0.1124	0.0081	0.0512	0.4226	0.0937	0.3411
0.6	0.2	0.2	0.1663	0.0067	0.0400	0.5658	0.0789	0.3751
0.33	0.33	0.33	0.0948	0.0090	0.0452	0.3508	0.0810	0.2852

are evaluated accordingly. When analysing each measure's importance, *MEQA* did not measure the global structure where $\alpha_3 = 0$ and instead assessed t-SNE and UMAP as performing better than LLE based on the local structure, which is not the case in real life. As a result, it denotes the significance of α_3. The importance of α_1 and α_2 are then studied while keeping the value of α_3 constant. In parameter settings where α_1 is greater than α_2, *MEQA* evaluates each method in the proper order and also quantifies with the proper value. Where α_2 is greater than α_1, *MEQA* assesses each method in the correct order, but the quantitative gap is not maintained properly. This supports the intuition of setting the value of α_1 greater than α_2. The analysis shows that methods that pre-

Table 10
MEQA measure on Swiss Roll dataset with different parameter settings.

α_1	α_2	α_3	LLE	ISOMAP	t-SNE	MDS	UMAP	Auto-En
0.3	0.7	0	0.2255	0.0607	0.0975	0.5776	0.1354	0.7401
0.5	0.5	0	0.2079	0.0917	0.1374	0.4989	0.1770	0.6989
0.7	0.3	0	0.1903	0.1227	0.1773	0.4202	0.2187	0.6576
0.2	0.7	0.1	0.0015	0.0014	0.0015	0.0013	0.0019	0.0026
0.45	0.45	0.1	0.1886	0.0840	0.1252	0.4507	0.1612	0.6316
0.7	0.2	0.1	0.1667	0.1228	0.1751	0.3524	0.2133	0.5800
0.2	0.6	0.2	0.1518	0.1074	0.1530	0.3239	0.1871	0.5230
0.4	0.4	0.2	0.1694	0.0763	0.1130	0.4026	0.1454	0.5643
0.6	0.2	0.2	0.1870	0.0453	0.0731	0.4812	0.1038	0.6055
0.33	0.33	0.33	0.1423	0.0654	0.0959	0.3350	0.1231	0.4698

Table 11
MEQA measure on Sphere dataset with different parameter settings.

α_1	α_2	α_3	LLE	ISOMAP	t-SNE	MDS	UMAP	Auto-En
0.3	0.7	0	0.0779	0.0979	0.0440	0.2953	0.0731	0.6618
0.5	0.5	0	0.0646	0.0844	0.0537	0.2610	0.0888	0.5390
0.7	0.3	0	0.0513	0.0709	0.0634	0.2267	0.1046	0.4163
0.2	0.7	0.1	0.0755	0.0936	0.0371	0.2787	0.0615	0.6395
0.45	0.45	0.1	0.0589	0.0767	0.0492	0.2358	0.0812	0.4861
0.7	0.2	0.1	0.0423	0.0598	0.0614	0.1929	0.1009	0.3327
0.2	0.6	0.2	0.0399	0.0555	0.0545	0.1764	0.0893	0.3105
0.4	0.4	0.2	0.0532	0.0690	0.0447	0.2107	0.0736	0.4332
0.6	0.2	0.2	0.0665	0.0825	0.0350	0.2450	0.0578	0.5560
0.33	0.33	0.33	0.0451	0.0581	0.0384	0.1754	0.0628	0.3591

Table 12
MEQA measure on Helix dataset with different parameter settings.

α_1	α_2	α_3	LLE	ISOMAP	t-SNE	MDS	UMAP	Auto-En
0.3	0.7	0	0.0001	0.0001	0.0056	0.0112	0.0029	0.6201
0.5	0.5	0	0.0001	0.0001	0.0076	0.0091	0.0036	0.4467
0.7	0.3	0	0.0001	0.0001	0.0096	0.0069	0.0043	0.2733
0.2	0.7	0.1	0.0010	0.0011	0.0051	0.0113	0.0033	0.6197
0.45	0.45	0.1	0.0010	0.0011	0.0076	0.0086	0.0042	0.4030
0.7	0.2	0.1	0.0010	0.0011	0.0101	0.0060	0.0050	0.1863
0.2	0.6	0.2	0.0018	0.0022	0.0096	0.0061	0.0054	0.1859
0.4	0.4	0.2	0.0018	0.0021	0.0076	0.0082	0.0047	0.3593
0.6	0.2	0.2	0.0018	0.0021	0.0056	0.0103	0.0040	0.5326
0.33	0.33	0.33	0.0029	0.0035	0.0075	0.0075	0.0054	0.2980

serve the high-dimensional structure in a similar way have similar assessments, while methods whose performance is quite different have dissimilar assessments. In comparison to other EQA methods, none of them achieves not only the proper order of assessment but also the quantitative value.

According to the discussion in Section 4.2.2, Table 10 shows that *MEQA* identifies ISOMAP as the best method for every parameter setting and Autoencoder as the worst method for the Swiss Roll dataset. The study also backs up the intuition that $\alpha 1$ should be greater than $\alpha 2$. In every parameter setting, *MEQA* outperforms other EQA methods in achieving proper assessment order and quantitative analysis.

According to the discussion in Section 4.2.3, LLE, ISOMAP, t-SNE, and UMAP have achieved nearly similar lower dimension structures for the Sphere dataset, and Table 11 highlights the exact situation across all parameter settings. *MEQA* obtained the proper rank for each extraction method in every setting, whereas other EQA methods did not.

MEQA achieves proper ranking and quantitative value across all parameter settings for the Helix dataset, and its performance is shown in Table 12.

For the Twin Peaks dataset *MEQA* achieves the proper rank for every feature extraction method over all parameter settings and maintains stability which is shown in Table 13.

Table 14 contains the real-world COIL-20 dataset. *MEQA* achieves the correct assessment order across all parameter settings, identifies LLE as the best method for feature extraction, and evaluates other methods accordingly. *MEQA* outperforms other EQA methods by achieving proper quantitative value across all parameter settings.

The performance analysis of *MEQA* over various parameter settings demonstrates its consistent performance and high potential in comparison to other existing EQA methods.

4.5. Time-complexity

Apart from generating more reliable and accurate embedding quality estimates, MEQA also makes an important stride towards achieving lower execution time. This is further corroborated in Table 15, where the execution times of other EQA methods have been listed along with MEQA. While Procrustes measure (P) incurs the lowest execution time, *NIEQA* requires the most. The remaining EQA methods: *LCMC*, Q_c, *T&C* and *MRRE*, focus mainly on local structure preservation measures at a slightly faster rate compared to MEQA. However, as seen from the experiments that all the EQA methods listed in Table 15 except MEQA, fail to draw the true picture of the preservation of the original manifold in its embedding.

Table 13
MEQA measure on Twin Peaks dataset with different parameter settings.

α_1	α_2	α_3	LLE	ISOMAP	t-SNE	MDS	UMAP	Auto-En
0.3	0.7	0	0.4298	0.5381	0.1218	0.2952	0.2241	0.9198
0.5	0.5	0	0.4060	0.4543	0.1458	0.2567	0.2402	0.9086
0.7	0.3	0	0.3822	0.3704	0.1697	0.2183	0.2564	0.8975
0.2	0.7	0.1	0.3966	0.5159	0.1035	0.2808	0.1990	0.8348
0.45	0.45	0.1	0.3669	0.4111	0.1334	0.2328	0.2191	0.8209
0.7	0.2	0.1	0.3372	0.3062	0.1634	0.1847	0.2393	0.8069
0.2	0.6	0.2	0.3040	0.2840	0.1451	0.1704	0.2141	0.7219
0.4	0.4	0.2	0.3278	0.3679	0.1211	0.2088	0.1980	0.7331
0.6	0.2	0.2	0.3516	0.4518	0.0972	0.2472	0.1819	0.7442
0.33	0.33	0.33	0.2729	0.3073	0.1036	0.1751	0.1681	0.6099

Table 14
MEQA measure on COIL-20 dataset with different parameter settings.

α_1	α_2	α_3	LLE	ISOMAP	t-SNE	MDS	UMAP	Auto-En
0.3	0.7	0	0.1648	0.2057	0.2840	0.2964	0.2088	0.4494
0.5	0.5	0	0.2106	0.2382	0.3427	0.3264	0.2585	0.4509
0.7	0.3	0	0.2527	0.2707	0.4013	0.3565	0.3081	0.4524
0.2	0.7	0.1	0.1518	0.1878	0.2490	0.2695	0.1842	0.4176
0.45	0.45	0.1	0.2045	0.2284	0.3223	0.3070	0.2462	0.4194
0.7	0.2	0.1	0.2572	0.2690	0.3956	0.3445	0.3083	0.4213
0.2	0.6	0.2	0.2406	0.2511	0.3606	0.3176	0.2836	0.3895
0.4	0.4	0.2	0.1985	0.2186	0.3020	0.2876	0.2340	0.3880
0.6	0.2	0.2	0.1563	0.1861	0.2433	0.2575	0.1843	0.3865
0.33	0.33	0.33	0.1885	0.2035	0.2721	0.2590	0.2155	0.3426

Table 15
Execution Time (in minutes) of MEQA and Other EQA Methods.

Methods	MEQA	*P*	P_L	*NIEQA*	*LCMC*	$Q_{nX}(k)$	*T*& *C*	*MRRE*
Execution Time	**1.94**	0.000017	0.0076	800.4	0.0016	0.5906	0.1653	0.1665

MEQA eclipses the performance of others by achieving more accurate quality estimates in less time.

5. Conclusions

MEQA quantifies the quality of an embedding by measuring both local and global structure preservation and this approach achieves superior results in comparison to the existing EQA methods both in terms of accuracy and time complexity. Now, the value of the control parameters or respective contributions will be dependent on the application *MEQA* is used. For, applications where global structure preservation does not play a significant role α_3 could be 0. Applications where closeness with respect to the original structure is more important α_2 will carry higher values than α_1 and α_3. The contribution of L_1, L_2 and L_3 will be determined by the application's emphasis. This control over what an application exactly wants to measure increases the strength and applicability of *MEQA*. In the field of pattern recognition with an application in real life, methods always deal with high-dimensional data and their dimensional reduction. In dimensionality reduction, the target is to achieve structural preservation where the contribution of *MEQA* will be higher. Along with this flexible applicability, *MEQA* is capable to address the majority of fields of pattern recognition. Due to the lack of a high-dimensional data set with an ideal or certified lower-dimensional representation, the dataset was mostly chosen synthetically. As a result, we must rely on visual inspection. As seen in Section 3.1, in $\mathbb{R}^{Dd}$, determining the value of $\nabla\overline{\Psi}(P_i)$ is an iterative process, which contributes significantly to the overall time complexity of the method. We hope to work in this direction in future. So, along with the numerous advantages of MEQA, there are two areas for improvement. The performance of *MEQA* is studied and presented for the classification problem here and other application domains will be explored in future.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

The authors do not have permission to share data.

Acknowledgment

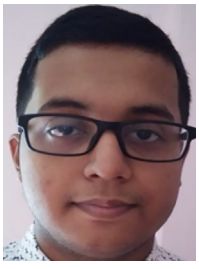
This work was supported by Technology Innovation Hub on Data Science, Big Data Analytics and Data Curation under Grant NMICPS/006/MD/2020-21 dt 16.10.2020.

References

[1] D.L. Donoho, High-dimensional data analysis: the curses and blessings of dimensionality, in: AMS Conference on Math Challanges of The 21st Century, 2000.
[2] L. Van Der Maaten, E. Postma, J. Van den Herik, Dimensionality reduction: a comparative review, J. Mach. Learn. Res. 10 (1–41) (2009) 66–71.
[3] I. Guyon, A. Elisseeff, An introduction to variable and feature selection, J. Mach. Learn. Res. 3 (Mar) (2003) 1157–1182.
[4] P. Pudil, J. Novovičová, Novel Methods for Feature Subset Selection with Respect to Problem Knowledge, Springer US, Boston, MA, 1998, pp. 101–116.
[5] H. Hotelling, Analysis of a complex of statistical variables into principal components, J. Educ. Psychol. 24 (6) (1933) 417–441, doi:10.1037/h0071325.
[6] W. Torgerson, Multidimensional scaling: I. Theory and method, Psychometrika 17 (4) (1952) 401–419, doi:10.1007/BF02288916.
[7] S.T. Roweis, L.K. Saul, Nonlinear dimensionality reduction by locally linear embedding, Science 290 (5500) (2000) 2323–2326, doi:10.1126/science.290.5500.2323.
[8] J.B. Tenenbaum, V. de Silva, J.C. Langford, A global geometric framework for nonlinear dimensionality reduction, Science 290 (5500) (2000) 2319–2323, doi:10.1126/science.290.5500.2319.

[9] M. Belkin, P. Niyogi, Laplacian eigenmaps for dimensionality reduction and data representation, Neural Comput. 15 (6) (2003) 1373–1396, doi:10.1162/089976603321780317.
[10] L. van der Maaten, G. Hinton, Visualizing data using t-SNE, J. Mach. Learn. Res. 9 (86) (2008) 2579–2605.
[11] D.L. Donoho, C. Grimes, Hessian eigenmaps: locally linear embedding techniques for high-dimensional data, Proc. Natl. Acad. Sci. 100 (10) (2003) 5591–5596, doi:10.1073/pnas.1031596100.
[12] L. McInnes, J. Healy, J. Melville, UMAP: uniform manifold approximation and projection for dimension reduction, 2020. 1802.03426
[13] R. Sibson, Studies in the robustness of multidimensional scaling: procrustes statistics, J. R. Stat. Soc. Ser. B (Methodological) 40 (2) (1978) 234–238.
[14] R. Sibson, Studies in the robustness of multidimensional scaling: perturbational analysis of classical scaling, J. R. Stat. Soc. Ser. B (Methodological) 41 (2) (1979) 217–229.
[15] Y. Goldberg, Y. Ritov, Local procrustes for manifold embedding: a measure of embedding quality and embedding algorithms, Mach. Learn. 77 (2009) 1–25.
[16] P. Zhang, Y. Ren, B. Zhang, A new embedding quality assessment method for manifold learning, Neurocomputing 97 (2012) 251–266, doi:10.1016/j.neucom.2012.05.013.
[17] L. Chen, A. Buja, Local multidimensional scaling for nonlinear dimension reduction, graph drawing, and proximity analysis, J. Am. Stat. Assoc. 104 (485) (2009) 209–219, doi:10.1198/jasa.2009.0111.
[18] J. Venna, S. Kaski, Local multidimensional scaling with controlled tradeoff between trustworthiness and continuity, in: WSOM'05, Paris, France September 5–8, 2005, 2005, pp. 695–702.
[19] J.A. Lee, M. Verleysen, Nonlinear Dimensionality Reduction, Springer Science & Business Media, 2007.
[20] J.A. Lee, M. Verleysen, Scale-independent quality criteria for dimensionality reduction, Pattern Recognit. Lett. 31 (14) (2010) 2248–2257, doi:10.1016/j.patrec.2010.04.013.
[21] R.N. Shepard, The analysis of proximities: multidimensional scaling with an unknown distance function. I, Psychometrika 27 (2) (1962) 125–140, doi:10.1007/BF02289630.
[22] R.N. Shepard, The analysis of proximities: multidimensional scaling with an unknown distance function. II, Psychometrika 27 (3) (1962) 219–246, doi:10.1007/BF02289621.
[23] J.B. Kruskal, Nonmetric multidimensional scaling: a numerical method, Psychometrika 29 (2) (1964) 115–129, doi:10.1007/BF02289694.
[24] J.A. Lee, M. Verleysen, et al., Rank-based quality assessment of nonlinear dimensionality reduction, in: ESANN, 2008, pp. 49–54.
[25] S.M. Ali, S.D. Silvey, A general class of coefficients of divergence of one distribution from another, J. R. Stat. Soc. Ser. B (Methodological) 28 (1) (1966) 131–142.
[26] G.W. Corder, D.I. Foreman, Nonparametric Statistics: A Step-by-Step Approach, John Wiley & Sons, 2014.
[27] I.M. James, The Topology of Stiefel Manifolds, Vol. 24, Cambridge University Press, Cambridge, London Mathematical Society, London, 1976.
[28] T. Lin, H. Zha, Riemannian manifold learning, IEEE Trans. Pattern Anal. Mach. Intell. 30 (5) (2008) 796–809, doi:10.1109/TPAMI.2007.70735.
[29] F.J. Massey, The Kolmogorov-Smirnov test for goodness of fit, J. Am. Stat. Assoc. 46 (253) (1951) 68–78.
[30] S.J. Nayar, Columbia object image library (COIL100), 1996.
[31] Y. Bengio, P. Lamblin, D. Popovici, H. Larochelle, et al., Greedy layer-wise training of deep networks, Adv. Neural Inf. Process. Syst. 19 (2007) 153.
[32] S. Boral, S. Dhar, A. Ghosh, Unsupervised segmentation of non-intersecting manifolds, in: The 12th International Conference on Advances in Information Technology, in: IAIT2021, Association for Computing Machinery, New York, NY, USA, 2021, doi:10.1145/3468784.3470467.

Subhadip Boral received the MSc in Computer Science from West Bengal State University in 2016. He is presently working as a Research Fellow at Technology Innovation Hub in Indian Statistical Institute. His current research interest includes Unsupervised Learning, Dimensionality Reduction, Streaming Data and Anomaly Detection.



Mainak Sarkar completed his BE degree in Electrical Engineering from Jadavpur University, Kolkata,West Bengal, India in June 2021. He is currently a Management Trainee at Haldia Energy Limited in Haldia, West Bengal, India. His current research interests include Data Science, Computer Vision and Machine Learning.

Ashish Ghosh is a Professor at the Indian Statistical Institute. He has published more than 270 research papers in international journals and conferences, and acting as the Principal Investigator of several funded projects. His current research interests include Machine and Deep Learning, Data Science, Image/Video Analysis, and Computational Intelligence.