Comparative Analysis of Manifold Learning Methods for Face Pose Ordering

Ding Ding
Department of Mathematics
The Hong Kong University of Science and Technology
ddingab@ust.hk

Zihao Zhang
Department of Mathematics
The Hong Kong University of Science and Technology
zzhangjn@ust.hk

Bingsong Gao
Department of Mathematics
The Hong Kong University of Science and Technology
bgaoaj@ust.hk

Abstract

Recovering the underlying low-dimensional manifold structure from high-dimensional data, such as images, is a fundamental task in machine learning and data analysis. This project investigates the effectiveness of various dimensionality reduction techniques in recovering the 1D rotational pose manifold from a dataset of 33 grayscale face images of the same individual viewed from different angles. We applied nine methods: Principal Component Analysis (PCA), Multidimensional Scaling (MDS), Isomap, Locally Linear Embedding (LLE), Local Tangent Space Alignment (LTSA), Modified LLE (MLLE), Hessian LLE (HLLE), Spectral Embedding, and t-Distributed Stochastic Neighbor Embedding (t-SNE). Performance was evaluated based on the accuracy of pose ordering using Total Absolute Error (TAE) relative to a visually refined ground truth sequence, as well as standard metrics for local structure preservation, namely Trustworthiness (T) and Continuity (C). Key hyperparameters, specifically the number of neighbors (k) for neighbor-based methods and perplexity (p) for t-SNE, were systematically tuned. Our results indicate that Isomap (k=5) achieved the best ordering performance (TAE=6), closely followed by LLE (k=5, TAE=14), Spectral Embedding (k=5, TAE=14), HLLE (k=10, TAE=16), and LTSA (k=10, TAE=16). Most methods exhibited high Trustworthiness and Continuity, although some trade-offs were observed. The analysis highlights the importance of parameter selection and the influence of the ground truth definition on quantitative ordering metrics. t-SNE and MDS performed poorly in recovering the global pose order for this specific task and ground truth. The project code is available at https://github.com/dingding-ust/manifold-learning-face-pose/tree/main. A short video presentation summarizing the key findings is available on YouTube: https://youtu.be/97eKQAUA2nA.

33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.

1 Introduction

Dimensionality reduction is a crucial preprocessing step when dealing with high-dimensional datasets prevalent in fields like computer vision, bioinformatics, and natural language processing. These datasets often exhibit intrinsic low-dimensional structures, or manifolds, hidden within the high-dimensional observation space. Manifold learning algorithms aim to uncover these underlying structures, providing compact representations that facilitate visualization, clustering, and classification tasks [Van der Maaten et al., 2009]. Applications range from visualizing complex biological data to estimating object pose from images [Tenenbaum et al., 2000, Roweis and Saul, 2000].

This project focuses on a classic manifold learning problem: recovering the underlying pose manifold from a collection of face images. We utilize a well-known dataset comprising 33 grayscale images of a single individual captured under varying horizontal viewpoints [Tenenbaum et al., 2000]. The intrinsic structure is expected to be a one-dimensional manifold corresponding to the continuous head rotation. The primary goal is to evaluate and compare the ability of various linear and non-linear dimensionality reduction techniques to capture this 1D manifold and accurately order the face images according to pose.

We employ a suite of nine dimensionality reduction methods, including the linear baseline methods Principal Component Analysis (PCA) and classical Multidimensional Scaling (MDS), as well as prominent non-linear manifold learning algorithms: Isomap [Tenenbaum et al., 2000], Locally Linear Embedding (LLE) [Roweis and Saul, 2000], Local Tangent Space Alignment (LTSA) [Zhang and Zha, 2004], Modified LLE (MLLE) [Zhang and Wang, 2007], Hessian LLE (HLLE) [Donoho and Grimes, 2003], Spectral Embedding [Belkin and Niyogi, 2003], and t-Distributed Stochastic Neighbor Embedding (t-SNE) [Van der Maaten and Hinton, 2008].

To quantitatively assess the performance, particularly for the ordering task, we established a ground truth sequence based on visual inspection and refinement. We evaluate the methods using the Total Absolute Error (TAE) metric, which measures the deviation of the algorithm-induced order (derived from the primary embedding dimension) from the ground truth. Additionally, we employ Trustworthiness (T) and Continuity (C) metrics [Venna and Kaski, 2001] to evaluate the preservation of local neighborhood structures in the 2D embeddings. Recognizing the sensitivity of manifold algorithms to hyperparameters, we perform systematic parameter tuning for the number of neighbors (k) in neighbor-based methods and perplexity (p) for t-SNE, analyzing their impact on the evaluation metrics through visualizations.

This report is structured as follows: Section 2 describes the dataset and ground truth establishment. Section 3 reviews the dimensionality reduction methods employed in detail. Section 4 details the experimental setup and evaluation metrics. Section 5 presents and discusses the results. Finally, Section 6 concludes the report.

2 Dataset and Ground Truth

The primary dataset used consists of 33 grayscale images ($N = 33$) of a single individual's face, captured from varying horizontal viewpoints. Each image is 92×112 pixels (Height $\times$ Width). Flattened, each image $\mathbf{x}_i$ is a point in $\mathbb{R}^D$ where $D = 10304$. The data lies on an approximate 1D manifold corresponding to head rotation [Tenenbaum et al., 2000]. Figure 1 shows sample images.

A ground truth sequence for pose ordering was established via visual inspection and refinement: [9, 20, 22, 14, 10, 4, 7, 0, 32, 13, 19, 12, 1, 30, 26, 11, 8, 5, 2, 29, 15, 16, 25, 21, 31, 17, 23, 6, 24, 27, 28, 3, 18] (0-based indices). The subjectivity of visual ordering is acknowledged.

3 Methods

We employed nine dimensionality reduction techniques mapping $\mathbf{X} = \{\mathbf{x}_i \in \mathbb{R}^D\}_{i=1}^N$ to $\mathbf{Y} = \{\mathbf{y}_i \in \mathbb{R}^d\}_{i=1}^N$ ($d = 2$), using Scikit-learn [Pedregosa et al., 2011]. Let X be the $N \times D$ data matrix.

Figure 1: Sample images (Indices 0, 16, 32) from the face pose dataset.

3.1 Linear Methods

3.1.1 Principal Component Analysis (PCA)

PCA identifies principal components maximizing data variance via linear projection [Hotelling, 1933]. Assuming centered data X_c, it computes the covariance matrix $\mathbf{C} \propto X_c^T X_c$. The principal components are the eigenvectors $\mathbf{v}_j$ of $\mathbf{C}$ satisfying $\mathbf{C}\mathbf{v}_j = \lambda_j \mathbf{v}_j$. The embedding $\mathbf{Y}$ is obtained by projecting onto the top d eigenvectors $\mathbf{V}_d = [\mathbf{v}_1, ..., \mathbf{v}_d]$:

$$\mathbf{Y} = X_c \mathbf{V}_d$$

3.1.2 Multidimensional Scaling (MDS)

Classical MDS [Young and Householder, 1941] reconstructs coordinates from pairwise dissimilarities δ_{ij}. Using squared Euclidean distances $D_{ij} = \delta_{ij}^2 = ||\mathbf{x}_i - \mathbf{x}_j||^2$, the centered Gram matrix $\mathbf{B}$ is obtained via double centering:

$$\mathbf{B} = -\frac{1}{2}\mathbf{H}\mathbf{D}^{(2)}\mathbf{H}, \quad \text{where } \mathbf{H} = \mathbf{I} - \frac{1}{N}\mathbf{1}\mathbf{1}^T$$

From the eigendecomposition $\mathbf{B} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^T$, the embedding coordinates are $\mathbf{Y} = \mathbf{U}_d \mathbf{\Lambda}_d^{1/2}$, using the top d positive eigenvalues $\mathbf{\Lambda}_d$ and corresponding eigenvectors $\mathbf{U}_d$.

3.2 Non-Linear Manifold Learning Methods

3.2.1 Isomap

Isomap [Tenenbaum et al., 2000] estimates geodesic distances using graph shortest paths. First, a k-nearest neighbor (k-NN) graph G is constructed with edge weights $d(\mathbf{x}_i, \mathbf{x}_j)$. Second, all-pairs shortest path distances $\delta_G(i,j)$ are computed on G, forming the matrix $\mathbf{D}_G^{(2)} = [\delta_G(i,j)^2]$. Third, classical MDS is applied to $\mathbf{D}_G^{(2)}$ (using $\mathbf{B}_G = -\frac{1}{2}\mathbf{H}\mathbf{D}_G^{(2)}\mathbf{H}$) to obtain the embedding $\mathbf{Y}$. Depends on k.

3.2.2 Locally Linear Embedding (LLE)

LLE [Roweis and Saul, 2000] preserves local linear reconstructions. First, for each $\mathbf{x}_i$, its k nearest neighbors $\mathcal{N}_k(i)$ are identified. Second, weights W_{ij} minimizing the local reconstruction error are found:

$$\min_W \sum_{i=1}^{N} ||\mathbf{x}_i - \sum_{j \in \mathcal{N}_k(i)} W_{ij}\mathbf{x}_j||^2 \quad \text{s.t.} \sum_{j \in \mathcal{N}_k(i)} W_{ij} = 1$$

This involves solving a constrained least-squares problem for each i. Third, the embedding $\mathbf{Y}$ is found by minimizing the cost $\Phi(Y) = \sum_i ||\mathbf{y}_i - \sum_j W_{ij}\mathbf{y}_j||^2 = \text{Tr}(\mathbf{Y}^T\mathbf{M}\mathbf{Y})$, where $\mathbf{M} = (\mathbf{I} - \mathbf{W})^T(\mathbf{I} - \mathbf{W})$. The solution uses the bottom d non-zero eigenvectors of $\mathbf{M}$. Depends on k.

3.2.3 Local Tangent Space Alignment (LTSA)

LTSA [Zhang and Zha, 2004] aligns local tangent spaces. For each neighborhood $\mathbf{X}_i$, it finds the tangent basis $\mathbf{V}_i$ via local PCA and computes local coordinates $\mathbf{\Theta}_i = \mathbf{V}_i^T(\mathbf{X}_i - \bar{\mathbf{x}}_i \mathbf{1}_k^T)$. It then seeks the global embedding $\mathbf{Y}$ by minimizing the alignment error:

$$\sum_i ||(\mathbf{Y}_i - \mathbf{y}_i \mathbf{1}_k^T) - \mathbf{L}_i \mathbf{\Theta}_i||_F^2$$

where $\mathbf{L}_i$ is a local affine transformation. This involves an eigenvalue problem on a global alignment matrix. Depends on k.

3.2.4 Modified LLE (MLLE)

MLLE [Zhang and Wang, 2007] stabilizes LLE using multiple weight vectors W_i per neighborhood, based on the local covariance null space. The embedding $\mathbf{Y}$ minimizes the sum of reconstruction errors using these weights:

$$\min_Y \sum_i \text{Tr}((\mathbf{Y}_i - \mathbf{y}_i \mathbf{1}_k^T)^T (\mathbf{Y}_i - \mathbf{y}_i \mathbf{1}_k^T) \mathbf{W}_i \mathbf{W}_i^T)$$

Solved via eigendecomposition. Depends on k.

3.2.5 Hessian LLE (HLLE)

HLLE [Donoho and Grimes, 2003] uses local Hessian estimates $\mathcal{H}_i$. The embedding $\mathbf{Y}$ minimizes the integrated squared Hessian norm:

$$\min_Y \sum_i ||\mathcal{H}_i \mathbf{Y}||^2 \approx \min_Y \text{Tr}(\mathbf{Y}^T \mathbf{K} \mathbf{Y})$$

solved via eigenvectors of the kernel matrix $\mathbf{K}$ built from Hessian estimates. Requires $k > d(d+3)/2$. Depends on k.

3.2.6 Spectral Embedding (Laplacian Eigenmaps)

Based on graph spectral theory [Belkin and Niyogi, 2003]. A k-NN graph with adjacency $\mathbf{W}$ is built. The graph Laplacian $\mathbf{L} = \mathbf{D} - \mathbf{W}$ is computed. The embedding $\mathbf{Y} = [\mathbf{v}_1, ..., \mathbf{v}_d]$ consists of the eigenvectors corresponding to the d smallest non-zero eigenvalues of the generalized eigenproblem $L\mathbf{v} = \lambda D\mathbf{v}$. Depends on k.

3.2.7 t-Distributed Stochastic Neighbor Embedding (t-SNE)

Visualizes similarity [Van der Maaten and Hinton, 2008]. It computes high-dimensional (p_{ij}) and low-dimensional (q_{ij}) joint probabilities representing pairwise similarities, using Gaussian and t-distributions respectively. p_{ij} depends on the 'perplexity' p.

$$q_{ij} = \frac{(1 + ||\mathbf{y}_i - \mathbf{y}_j||^2)^{-1}}{Z}$$

It minimizes the KL divergence $KL(P||Q) = \sum_{i<j} p_{ij} \log \frac{p_{ij}}{q_{ij}}$ via gradient descent with gradient:

$$\frac{\partial KL}{\partial \mathbf{y}_i} = 4 \sum_{j \neq i} (p_{ij} - q_{ij}) q_{ij} Z (\mathbf{y}_i - \mathbf{y}_j)$$

Depends on p.

4 Experimental Setup

Experiments used Python 3.9, Scikit-learn (1.3.2), Pandas, Matplotlib/Seaborn. Data was standardized for relevant methods. All embeddings were $d = 2$.

Parameter tuning explored $k \in \{3, 4, 5, 6, 7, 8, 9, 10, 12\}$ for neighbor-based methods (HLLE for $k \geq 7$) and $p \in \{5, 8, 10, 13, 15, 18, 20, 25, 30\}$ for t-SNE. A fixed 'random_state=42' was used.

Performance metrics were:

TAE (min): Total Absolute Error between computed order π_{comp} (from sorting $y^{(1)}$ or $-y^{(1)}$) and ground truth π_{true}. Let $pos(i, \pi)$ be the rank of sample i in permutation π.

$$\text{TAE}(\pi) = \sum_{i=1}^{N} |pos(\pi_{comp}(i), \pi_{true}) - pos(i, \pi_{true})|$$

We report TAE(min) $= \min(\text{TAE}(\pi_{comp}), \text{TAE}(\pi_{comp}^{rev}))$. Lower is better.

Trustworthiness (T): Measures preservation of local structure (few false neighbors) [Venna and Kaski, 2001]. Let $U_k(i) = \{j | j \in \hat{\mathcal{N}}_k(i) \setminus \mathcal{N}_k(i)\}$ be the set of "intruders" in the k-neighborhood of i in the embedding $\mathbf{Y}$.

$$T(k) = 1 - \frac{2}{Nk(2N - 3k - 1)} \sum_{i=1}^{N} \sum_{j \in U_k(i)} (rank_{high}(i, j) - k)$$

Higher is better (max 1.0). Calculated with $k = 5$.

Continuity (C): Measures preservation of local structure (few missing neighbors) [Venna and Kaski, 2001]. Let $V_k(i) = \{j | j \in \mathcal{N}_k(i) \setminus \hat{\mathcal{N}}_k(i)\}$ be the set of "extruders" from the k-neighborhood of i in $\mathbf{X}$.

$$C(k) = 1 - \frac{2}{Nk(2N - 3k - 1)} \sum_{i=1}^{N} \sum_{j \in V_k(i)} (rank_{low}(i, j) - k)$$

Higher is better (max 1.0). Calculated with $k = 5$.

5 Results and Discussion

We evaluated the performance of the nine methods both quantitatively using the defined metrics and qualitatively through visualizations. Parameter sensitivity was also analyzed.

5.1 Baseline Methods: PCA and MDS

Figure 2 (top row) shows the embeddings. Both methods reveal the 'U' shape. PCA achieved TAE=30, while MDS had a high TAE=350 (Table 1).

5.2 Manifold Learning Methods: Performance with Optimal Parameters

Table 1 summarizes the best performance achieved by each method after parameter tuning, focusing on TAE(min). Isomap (k=5) yields the lowest TAE, followed closely by LLE, Spectral, HLLE, and LTSA with specific optimal k.

Table 1: Final Evaluation Metrics using Optimal Parameters (T&C calculated with k=5).

Method	Optimal Param	TAE (min)	Trustworthiness	Continuity
Isomap	k=5	6	0.9818	0.9898
LLE	k=5	14	0.9624	0.9728
Spectral	k=5	14	0.9554	0.9644
LTSA	k=10	16	0.9539	0.9396
HLLE	k=10	16	0.9539	0.9396
MLLE	k=12	22	0.9467	0.9619
PCA	N/A	30	0.9782	0.9908
t-SNE	p=20	90	0.9787	0.9908
MDS	N/A	350	0.9864	0.9818

The scatter plots (Figure 2) and image plots (Figure 3) using these optimal parameters visually confirm the ordering quality.

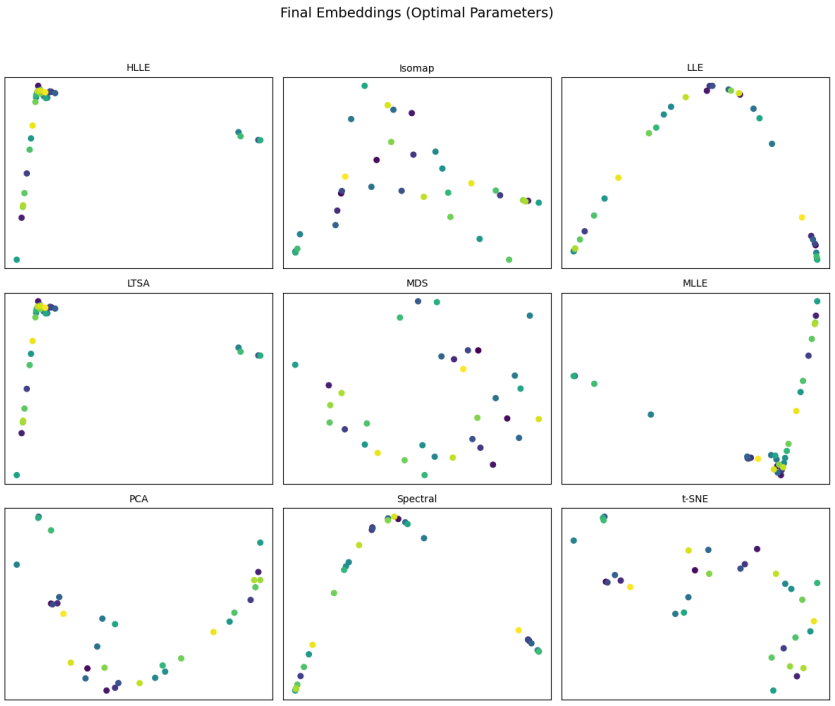

Figure 2: Final 2D Embeddings (Scatter Plots) using optimal parameters.

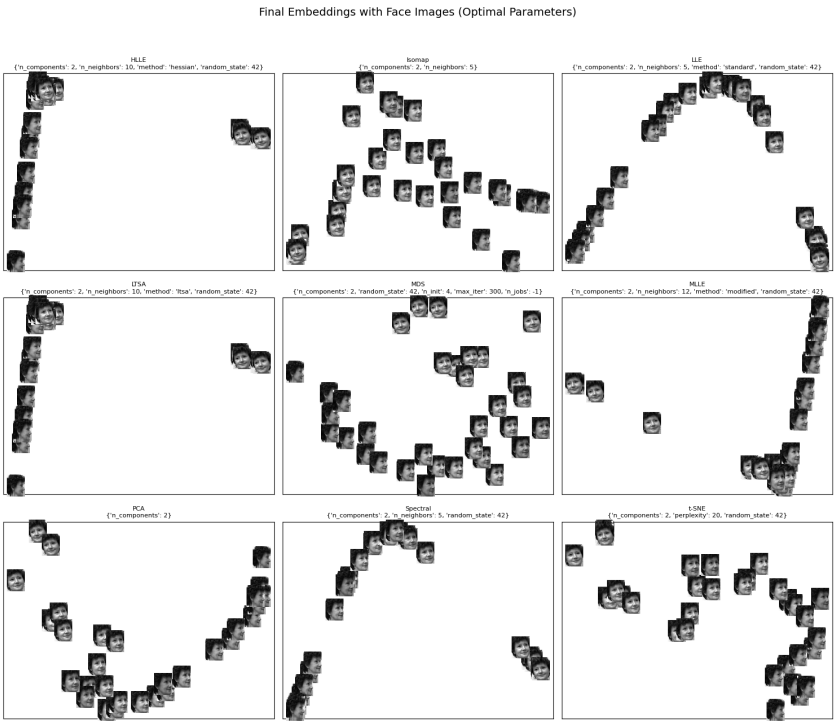

Figure 3: Final 2D Embeddings with Face Images using optimal parameters.

5.3 Parameter Tuning Analysis

The influence of the number of neighbors k is shown in Figure 4. Optimal TAE was mostly found for $k \in [5, 10]$. T&C generally plateaued for $k \geq 5$.

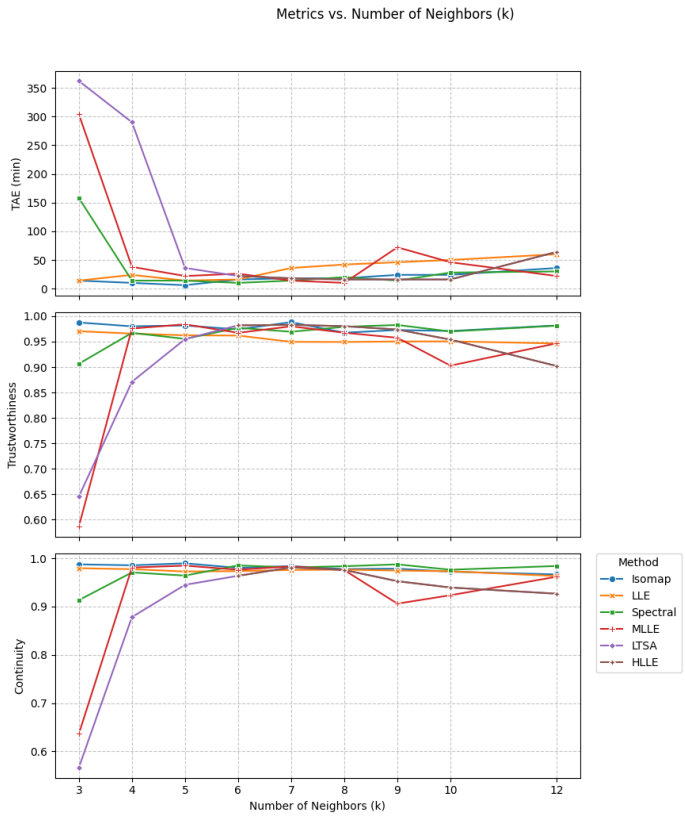

Figure 4: Evaluation metrics versus number of neighbors (k).

The effect of perplexity p on t-SNE (Figure 5) showed optimal TAE at $p = 20$, while T&C were less sensitive.

5.4 Discussion

Our results demonstrate the effectiveness of several manifold learning algorithms, particularly Isomap, LLE, Spectral, HLLE, and LTSA, in recovering the 1D rotational manifold. The optimal parameters ($k \in [5, 10]$, $p = 20$) were crucial for performance. The main discrepancy with the reference work [Liu et al., 2023], where LLE was solely optimal, likely stems from the differing ground truth sequences used, highlighting the sensitivity of TAE to this subjective choice. [Your further discussion points...]

6 Conclusion

This project comparatively analyzed nine dimensionality reduction techniques for face pose ordering. Isomap (k=5) provided the best ordering accuracy based on our visually refined ground truth, closely followed by LLE (k=5), Spectral (k=5), HLLE (k=10), and LTSA (k=10). Parameter tuning confirmed optimal performance typically occurs for $k \in [5, 10]$. The study underscores the utility of manifold learning and the importance of parameter selection and ground truth definition.

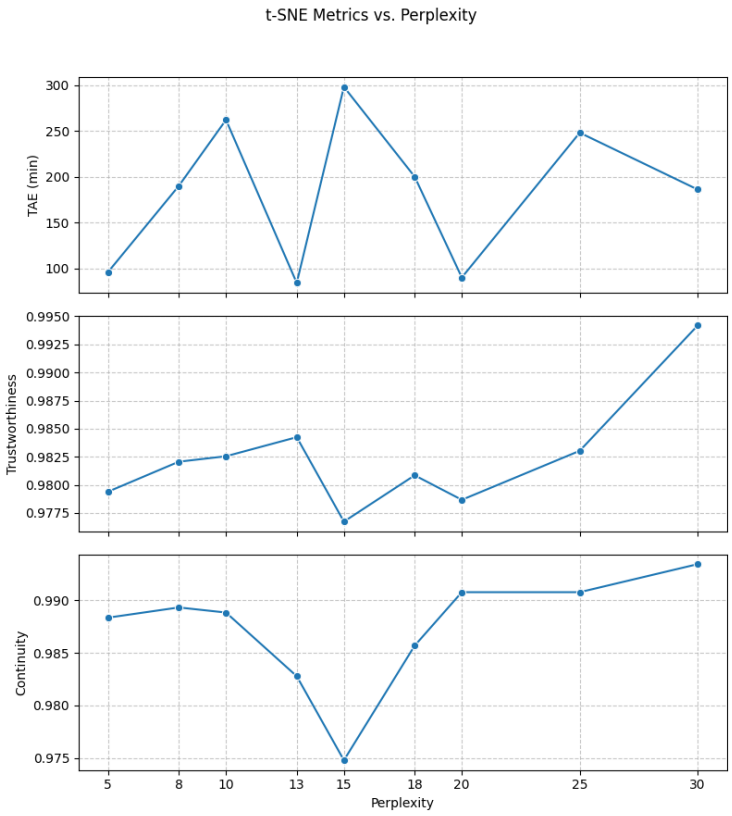

Figure 5: t-SNE evaluation metrics versus perplexity (p).

Contributions

Ding Ding: Developed the core code structure for data processing, implementing the embedding algorithms, and defining evaluation metrics. Took the lead on drafting and composing the final report manuscript.

Zihao Zhang: Focused on refining the experimental setup, systematically tuning hyperparameters (k and perplexity) for the manifold learning methods, and contributed to the generation of key visualizations for the report. Also assisted in initial data preprocessing steps.

Bingsong Gao: Contributed to the optimization of the codebase and performed validation of the evaluation metric implementations. Involved in the final refinement and proofreading of the report, and played a significant role in designing and preparing the presentation materials.

References

Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation*, 15(6):1373–1396, 2003.

David L Donoho and Carrie Grimes. Hessian eigenmaps: Locally linear embedding techniques for high-dimensional data. *Proceedings of the National Academy of Sciences*, 100(10):5591–5596, 2003.

Harold Hotelling. Analysis of a complex of statistical variables into principal components. *Journal of educational psychology*, 24(6):417, 1933.

D. Liu, M. Wang, and X. Han. Csic 5011 final project: Order the faces via manifold learning. Course project report, HKUST, 2023. unpublished.

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, et al. Scikit-learn: Machine learning in python. *Journal of machine learning research*, 12 (Oct):2825–2830, 2011.

Sam T Roweis and Lawrence K Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000.

Joshua B Tenenbaum, Vin De Silva, and John C Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.

Laurens Van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(11), 2008.

Laurens Van der Maaten, Eric Postma, and Jaap Van den Herik. Dimensionality reduction: a comparative review. *Journal of Machine Learning Research*, 10(Nov):66–71, 2009.

Jarkko Venna and Samuel Kaski. Neighborhood preservation in nonlinear projection methods: An experimental study. In *International Conference on Artificial Neural Networks*, pages 485–491, Berlin, Heidelberg, 2001. Springer.

Gale Young and Alston S Householder. Discussion of a set of points in terms of their mutual distances. *Psychometrika*, 6(3):19–22, 1941.

Zhenyue Zhang and Jing Wang. Mlle: Modified locally linear embedding using multiple weights. In *Advances in neural information processing systems*, volume 19, 2007.

Zhenyue Zhang and Hongyuan Zha. Principal manifolds and nonlinear dimensionality reduction via tangent space alignment. *SIAM journal on scientific computing*, 26(1):313–338, 2004.