

An Exploration of PaCMAP

Abdelrahman ElJamal
Emory Salaun
Ayman Sandouk

November 2025

1 Problem Definition

The goal of this work is to deepen our understanding of dimensionality reduction (DR) techniques, with a particular focus on PaCMAP (Pairwise Controlled Manifold Approximation Projection). We aim to study PaCMAP in detail, understand how it can be implemented on a dataset, apply it to real data and use appropriate evaluation metrics to assess its effectiveness. Through this, we hope to gain insight into how DR methods are chosen for a particular dataset, how they are applied and how their performance is evaluated

2 Introduction

Real-world data is often high dimensional, which can make it difficult to process, visualize, and interpret. In order to handle it adequately, its dimensionality needs to be reduced. Dimensionality reduction is the transformation of high-dimensional data into a meaningful representation with fewer dimensions. Ideally, the reduced representation should match the intrinsic dimensionality of the original data while preserving important structure and relationships.

Dimension reduction (DR) techniques have demonstrated impressive visualization performance on many complex datasets. However, they often involve trade-offs, such as losing aspects of either local structure or global structure.

With this work, we intend to answer the following questions:

- How can we decide which datasets can benefit from a particular DR technique?
 - How can we correctly load and prepare a dataset for a DR technique?
- How can PaCMAP and other DR techniques be applied to different datasets?
- How can we effectively evaluate a particular DR technique?
 - How can we evaluate the local structure preservation of a DR technique?
 - How can we evaluate the global structure preservation of it?
- How can we effectively compare between DR techniques?

In this work, we refer to Wang, Yingfan, et al.'s paper, which explores how dimensionality reduction methods work and deciphers some of them, such as PaCMAP, for data visualization. The work is published as a GitHub repository with clear instructions on how to implement this method using the PaCMAP Python library that can be found in the repository.

3 Dataset

To evaluate PaCMAP on data of varying structure, dimensionality, and modality, we conducted experiments on four datasets: COIL-20, MNIST, Olivetti Faces, and 20 Newsgroups. While the original PaCMAP paper primarily evaluates image datasets, our study extends the scope by including a text dataset, allowing assessment of PaCMAP’s robustness across heterogeneous data types.

Table 1: Summary of datasets used in the experiments

Dataset	Type	Description	Size
COIL-20	Image (object recognition)	The Columbia Object Image Library (COIL-20) contains 20 household objects photographed from 72 views at 5° intervals. Images are grayscale and captured under controlled lighting, making it suitable for clustering and manifold learning.	1,440 images
MNIST	Image (digit classification)	MNIST consists of 28×28 grayscale images of handwritten digits (0–9). It is widely used for dimensionality reduction due to its clear digit clusters and moderate size.	70,000 images
Olivetti Faces	Image (face recognition)	Contains 64×64 grayscale images of 40 individuals, with 10 images each showing variations in pose, lighting, and expression. It is more challenging due to subtle inter-class differences.	400 images
20 Newsgroups	Text (topic classification)	Contains documents from 20 internet discussion groups on topics such as science, politics, and sports. Documents are converted into TF-IDF vectors, yielding high-dimensional sparse data suitable for evaluating manifold learning on text.	~18,000 documents

3.1 Dataset Loading and Preprocessing

To ensure consistent handling of different datasets, we implemented a unified dataset loading framework. Rather than relying on the original PaCMAP codebase, all datasets in our evaluation were processed using a custom `DatasetLoader` class, which provides a standardized interface for image, NumPy-based, and text datasets. This design enables reproducible preprocessing across modalities and simplifies the integration of additional datasets.

The loader supports four dataset types: COIL-20, MNIST, generic NumPy datasets (`.npy` or `.npz`), and the 20 Newsgroups text corpus. Each dataset type uses a dedicated internal loading routine, while routine tasks such as label encoding, reshaping, and train-test splitting are handled in a unified manner.

COIL-20 Loader. COIL-20 images are loaded from a directory structure where each class has its own folder. All images are converted to grayscale and stored as two-dimensional NumPy arrays. Labels are

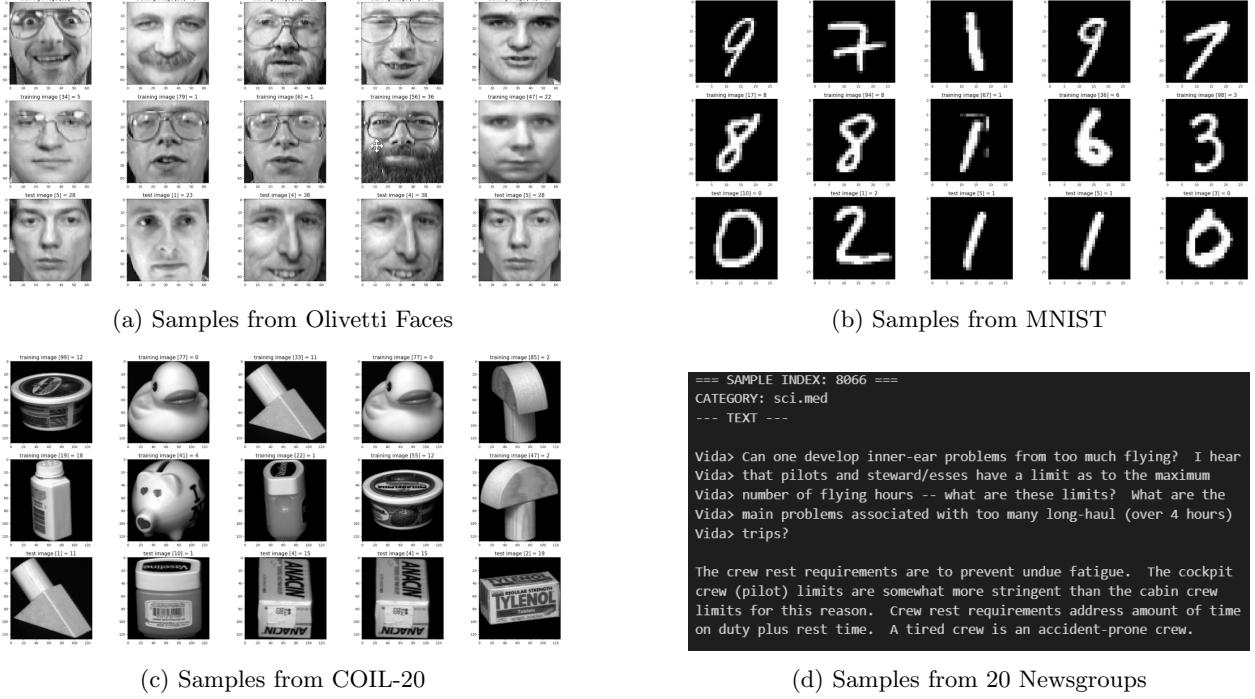


Figure 1: Sample images and representations from all datasets used in this study.

extracted from the folder names and encoded using `LabelEncoder`. The loader optionally applies an 80/20 stratified train–test split.

MNIST Loader. MNIST is read directly from its raw IDX binary files. Image and label files are parsed manually using structured binary reading, ensuring independence from external MNIST utilities. Each image is reshaped to 28×28 format. The loader can either return the original train–test partitions or merge them into a single dataset when no split is requested.

Generic NumPy Datasets. For datasets stored in `.npy` or `.npz` format, the loader automatically detects whether the file contains separated or combined data label arrays. Multi-dimensional inputs are flattened when necessary to produce a two-dimensional feature matrix. If the dataset does not include labels, dummy labels are generated to maintain compatibility with downstream methods.

20 Newsgroups Loader. To include a text modality, we added support for the 20 Newsgroups dataset. The raw documents are loaded using `fetch_20newsgroups`, and each document is transformed into a high-dimensional TF-IDF vector using the `TfidfVectorizer`. To maintain consistency across datasets, the loader exposes the text data in the same format as the image datasets: a feature matrix $X \in R^{n \times d}$ and a corresponding label vector

Unified Interface. All dataset types are accessed through a single `load_data(split)` interface, where `split` determines whether the loader returns the full dataset or an automatically generated train–test split. This unified design ensures consistent preprocessing, reduces boilerplate code, and facilitates experimentation across datasets of different modalities (images and text).

This unified loading pipeline ensures consistent preprocessing and facilitates straightforward integration of additional datasets in future work.

4 Methodology

5 Evaluation Techniques

In order to evaluate how well PaCMAP worked on our datasets, we chose four separate metrics to cover a variety of aspects of how well the low-dimensional embedding preserves the structure of the original data. Our metrics include k-Nearest-Neighbors (kNN) accuracy, trustworthiness, continuity, and Mean Relative Rank Error (MRRE)

5.1 kNN Classification Accuracy

This metric measures how well class information is preserved after the DR technique is applied. We train a k-nearest neighbor classifier on the original dataset, record the accuracy on how well it predicts labels in the test set. We then apply the dimensionality-reduction method, train a new classifier on the reduced data, and record its accuracy as well. A new accuracy value that is close to the original accuracy indicates that the dimensionality-reduction method successfully preserved the class-related structure of the data.

5.2 Trustworthiness

Trustworthiness evaluates whether the generated low-dimensional space introduces any false neighbors, which are points that look close in the reduced embedding, but were not close in the original dataset. Trustworthiness has a scale from 0-1, with a high score meaning the method did not create many false neighbors, therefore better preserving the local structure.

5.3 Continuity

Continuity checks for the opposite of what trustworthiness was looking for. Instead of looking for false neighbors, it checks for missing neighbors, points that were close in the original space but became far apart after dimensional reduction. Continuity also has a scale from 0-1, with a high score meaning the DR method keeps true neighbors together, even after projecting the data into fewer dimensions.

5.4 Mean Relative Rank Error (MRRE)

MRRE measures how much the ranking of neighbors changes after DR. Unlike kNN accuracy, instead of looking at whether a point is inside or outside the k-nearest neighborhood, MRRE looks at how far its ranking shifts, where a point's rank refers to how close it is to a given sample compared to all other points (the closest point has rank 1, the second-closest has rank 2, ...). Lower MRRE values indicate that the reduced embedding preserves the original neighborhood ordering more accurately.

5.5 Qualitative Analysis of 2D Embeddings

In addition to the quantitative metrics described above, we also include a qualitative evaluation step by visually examining the two-dimensional PaCMAP embeddings for each dataset. While numerical scores capture specific aspects of neighborhood preservation, the visual layout of the embedded points provides an interpretable assessment of how well the method organizes the data.

By plotting each dataset in the reduced 2D space and coloring points by class, we can directly observe cluster separability, overlap between classes, and global geometric structure. Datasets with well-preserved structure typically exhibit distinct, compact clusters that reflect the original class organization, whereas heavy overlap or diffuse regions indicate structural distortions introduced during the dimensionality-reduction process.

This qualitative inspection does not replace quantitative metrics, but it complements them by offering a interpretable view of the embedding's behavior. It allows us to visually confirm patterns suggested by kNN accuracy, trustworthiness, continuity, and MRRE, and it often reveals structure, such as cluster shape or global arrangement, that is not fully captured by numerical measures alone.

6 Results

6.1 kNN Classification Accuracy

Table 2: KNN classification accuracy before and after PaCMAP embedding for each dataset.

Dataset	Stage	K=1	K=3	K=5	K=7	K=9
Coil20	Before	0.9931	0.9722	0.9340	0.8958	0.8889
	After	0.8438	0.8333	0.8264	0.8229	0.8333
MNIST	Before	0.9691	0.9705	0.9688	0.9694	0.9659
	After	0.9429	0.9598	0.9609	0.9612	0.9612
Olivetti	Before	0.9375	0.8750	0.8625	0.8250	0.7500
	After	0.4750	0.4500	0.5250	0.4750	0.4750
20NewsGroups	Before	0.8021	0.7637	0.7623	0.7517	0.7493
	After	0.3435	0.3676	0.3918	0.4106	0.4172

As shown in Table 2, the MNIST dataset exhibits the smallest performance reduction, maintaining more than 94% accuracy for all k values in the embedded space. This indicates that MNIST’s digit classes remain well-separated even after dimensionality reduction, suggesting strong preservation of both global and local structure by PaCMAP.

Coil20 demonstrates a moderate accuracy drop (from 99.31% to 84.38% at $k = 1$). While the general class structure is preserved, fine-grained distinctions between visually similar objects become less separable in two dimensions.

In contrast, the Olivetti Faces dataset shows the most substantial degradation, with accuracy dropping from 93.75% to 47.50% at $k = 1$. Facial identity datasets require high-dimensional feature information to distinguish between individuals, and this structure is not well preserved when compressed to just two dimensions.

The 20Newsgroups dataset shows the largest performance drop among all evaluated datasets, with kNN accuracy decreasing from 80.21% to 34.35% at $k = 1$. Unlike the image datasets, 20Newsgroups consists of high-dimensional sparse text representations with subtle semantic differences between classes, making them difficult to preserve in a two-dimensional space. The gradual increase in accuracy for higher k values suggests that broader neighborhood structure is retained to some extent, but fine-grained class distinctions are largely lost after dimensionality reduction. This substantial degradation indicates that PaCMAP struggles to maintain textual structure when projecting the data into only two dimensions.

Overall, the results in Table 2 show that PaCMAP performs very well on datasets with naturally well-separated cluster structure (such as MNIST), and reasonably well on structured object datasets with consistent visual appearance (such as Coil20). Its performance declines substantially on datasets with subtle or high-dimensional class boundaries, such as the Olivetti Faces dataset, where visual similarities between individuals are difficult to represent in only two dimensions. The most severe degradation occurs on the 20Newsgroups dataset, where sparse text representations and semantic distinctions lead to major losses in class separability after dimensionality reduction. These findings collectively reinforce that the effectiveness of PaCMAP strongly depends on the structure and dimensionality of the dataset, and that DR method selection should consider the complexity and nature of the data being analyzed.

6.2 Trustworthiness

Table 3 reports the Trustworthiness values for Coil20, MNIST, Olivetti, and 20Newsgroups across neighborhood sizes K . Trustworthiness captures how many *false neighbors* appear in the low-dimensional embedding—points that were far apart in the original space but incorrectly become close after dimensionality reduction. Higher values therefore indicate better preservation of local structure.

Coil20 achieves the highest Trustworthiness values, remaining above 0.95 for most K values. This indicates that PaCMAP introduces very few false neighbors for this dataset, consistent with the clean, well-

Table 3: Trustworthiness values for each dataset at different neighborhood sizes K . Higher values indicate better preservation of local neighborhood structure.

Dataset	K=1	K=3	K=5	K=7	K=9
Coil20	0.9849	0.9836	0.9764	0.9694	0.9587
MNIST	0.9528	0.9524	0.9520	0.9516	0.9514
Olivetti	0.9170	0.8758	0.8540	0.8327	0.8238
20NewsGroups	0.7853	0.7817	0.7772	0.7729	0.7679

separated clusters visible in the embedding and the strong kNN performance. MNIST also shows high Trustworthiness (around 0.95 across all K), suggesting that the local neighborhood relationships between digit images are largely maintained even after compression to two dimensions.

Olivetti exhibits moderately lower Trustworthiness values, decreasing from 0.917 at $K = 1$ to 0.824 at $K = 9$. This decline reflects the difficulty of preserving the subtle, high-dimensional local structure necessary to distinguish between individual faces. The increased introduction of false neighbors aligns with the more diffuse clusters observed in the qualitative embedding and the reduced kNN accuracy.

20Newsgroups shows the lowest Trustworthiness values, ranging from 0.785 at $K = 1$ to 0.768 at $K = 9$. This pattern is expected for high-dimensional sparse text data, where semantic neighborhoods are difficult to preserve in a 2D embedding. The substantial number of false neighbors introduced by PaCMAP contributes to the sharp drop in kNN accuracy and the large MRRE values observed for this dataset.

Overall, the Trustworthiness results show that PaCMAP preserves local structure effectively for Coil20 and MNIST, performs moderately on Olivetti, and struggles substantially with the complex and sparsely represented relationships in 20Newsgroups.

6.3 Continuity

Table 4: Continuity values for each dataset at different neighborhood sizes K . Higher values indicate better preservation of global neighborhood structure.

Dataset	K=1	K=3	K=5	K=7	K=9
Coil20	0.9968	0.9938	0.9936	0.9928	0.9919
MNIST	0.9901	0.9874	0.9854	0.9837	0.9822
Olivetti	0.9740	0.9638	0.9448	0.9293	0.9180
20NewsGroups	0.8598	0.8465	0.8379	0.8316	0.8261

Table 4 shows the Continuity values for Coil20, MNIST, Olivetti, and 20Newsgroups across different neighborhood sizes K . All four datasets exhibit relatively high Continuity scores, indicating that the PaCMAP embedding preserves much of the global neighborhood structure. As expected, Continuity decreases slightly as K increases, since larger neighborhoods are more sensitive to global distortions introduced by the embedding.

Coil20 achieves the highest Continuity values, remaining above 0.99 for all K , reflecting PaCMAP’s strong ability to preserve global relationships in structured object data. MNIST also maintains high Continuity (above 0.98), suggesting that the global organization of digit classes is well preserved. Olivetti exhibits lower scores (0.918 - 0.974), consistent with the greater difficulty of maintaining global facial identity structure in two dimensions.

Olivetti shows a low Continuity score, although the values remain relatively high overall (0.918 - 0.974). This reduction is consistent with the increased difficulty of organizing facial identity data in a low-dimensional space, where global relationships between individuals are more complex and less well separated.

20Newsgroups shows the lowest Continuity values (0.826–0.860), which is expected given the high-dimensional, sparse nature of text data and the subtle semantic differences between topics. These characteristics make global relationships more challenging to preserve during dimensionality reduction.

Overall, PaCMAP most effectively maintains global structure in Coil20 and MNIST, performs moderately on Olivetti, and struggles the most with the high complexity of 20Newsgroups.

6.4 Mean Relative Rank Error (MRRE)

Table 5 presents the MRRE values for Coil20, MNIST, Olivetti, and 20Newsgroups. As expected, the absolute magnitude of MRRE varies substantially across datasets due to differences in dataset size and structure. MNIST, with nearly 70,000 samples, produces large MRRE values because even small rank shifts correspond to changes across thousands of possible neighbor positions. Coil20 (1,440 samples) and Olivetti (400 samples) yield much smaller MRRE values simply because their ranking spaces are far more limited.

20Newsgroups exhibits the highest MRRE values by a significant margin, even exceeding MNIST. This results from its extremely high-dimensional and sparse text representations, combined with complex semantic relationships that are difficult to preserve in a two-dimensional space. As a result, the relative rank ordering of neighbors changes dramatically after dimensionality reduction, leading to large MRRE scores.

These results reinforce that MRRE is most informative when used to *compare different dimensionality reduction techniques on the same dataset*, rather than for direct comparison across datasets of different sizes or modalities. Within each dataset, the decreasing trend in MRRE as K increases follows the expected pattern: smaller neighborhoods are more sensitive to local distortions, while larger neighborhoods are more tolerant of rank shifts.

Table 5: MRRE values for each dataset at different neighborhood sizes K . Lower values indicate better preservation of neighbor rank structure.

Dataset	K=1	K=3	K=5	K=7	K=9
MNIST	591.47	427.09	350.31	305.13	273.48
Coil20	3.70	3.98	3.32	2.98	2.75
Olivetti	8.28	6.75	6.63	6.41	6.09
20NewsGroups	2113.20	1371.72	1060.66	880.29	761.26

6.5 Qualitative Analysis of 2D Embeddings

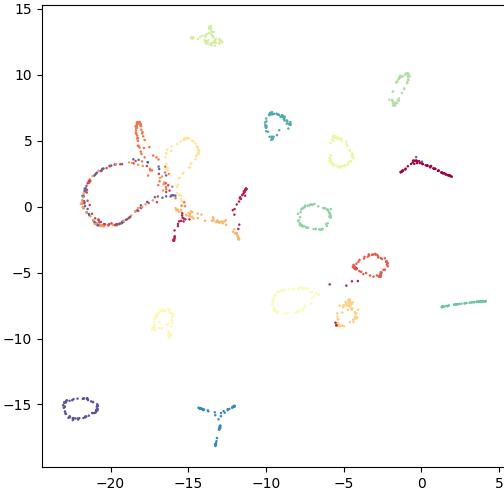
Figure 2a, Figure 2b, Figure 2c, and Figure 2d present the two-dimensional PaCMAP embeddings for Coil20, MNIST, Olivetti, and 20Newsgroups respectively. These visualizations provide an intuitive understanding of how well PaCMAP preserves structure beyond what quantitative metrics alone can show, allowing direct inspection of cluster separation, overlap, and the overall organization of each dataset in the reduced space.

The Coil20 embedding exhibits clearly separated, compact clusters, each corresponding to a different object category. The clusters form smooth, continuous loops that reflect the rotational nature of the dataset, and minimal overlap is observed between classes. This agrees with the high continuity values and the moderate post-embedding kNN performance, indicating that global and local structure are both well preserved.

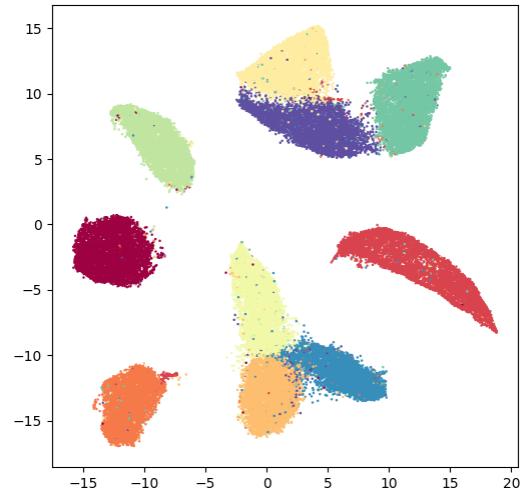
The MNIST embedding also shows strong cluster structure, with distinct regions corresponding to each digit class. While some visually similar digits (such as 4 and 9, or 3 and 8) exhibit partial overlap, most clusters are dense and well formed. This visual clarity aligns with MNIST’s high trustworthiness and continuity scores, as well as its relatively small drop in kNN accuracy after dimensionality reduction.

In contrast, the Olivetti embedding appears far more diffuse, with clusters that are less compact and more interwoven. Individual identities do not form tight groups, and many points lie between clusters or in areas of overlap. This visual fragmentation is consistent with the lower kNN accuracy and higher MRRE values for Olivetti, reflecting the difficulty of preserving facial identity information in only two dimensions.

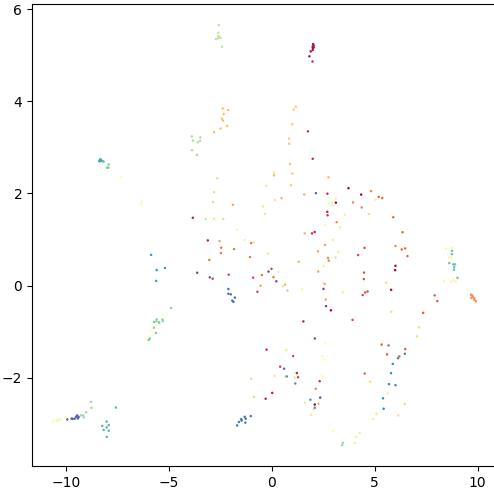
The 20Newsgroups embedding appears significantly more diffuse than the image datasets, with broad, elongated regions and many points scattered across the embedding space. Clusters corresponding to different



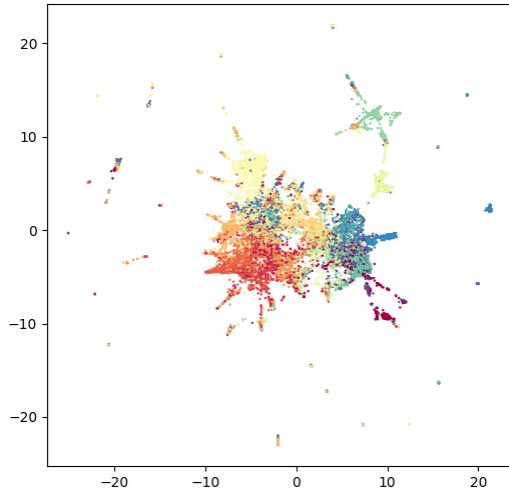
(a) Coil20



(b) MNIST



(c) Olivetti



(d) 20Newsgroups

Figure 2: PaCMAP embeddings for Coil20, MNIST, Olivetti, and 20Newsgroups datasets.

topics show partial structure but lack clear boundaries, and substantial mixing occurs between semantically related categories. This diffuse organization is consistent with the low kNN accuracy and high MRRE values observed for the dataset, reflecting the inherent difficulty of preserving high-dimensional, sparse text relationships within a two-dimensional representation.

Overall, the qualitative embeddings reinforce the quantitative evaluation: PaCMAP preserves structure effectively for Coil20 and MNIST, while Olivetti remains challenging due to its high intrinsic dimensionality and subtle class distinctions. The 20Newsgroups embedding further illustrates PaCMAP’s limitations on high-dimensional, sparse text data, where topic clusters become heavily intermixed. Taken together, these visual and numerical results provide a comprehensive assessment of PaCMAP’s performance across datasets.

of varying complexity and modalities, highlighting where the method excels and where it struggles to preserve meaningful structure in a two-dimensional space.

7 PaCMAP in Three Dimensions

To evaluate whether additional embedding capacity improves structure preservation, we repeated all experiments using a three-dimensional PaCMAP embedding. Since PaCMAP is primarily optimized for two-dimensional visualization, these experiments illustrate how the method behaves when given an additional degree of freedom. Overall, increasing the embedding dimension from 2D to 3D produced consistent but modest improvements across all datasets, with the largest gains appearing in the high-dimensional text data.

20Newsgroups showed the most notable improvement. kNN accuracy at $k = 1$ rose from 0.3435 in 2D to 0.4069 in 3D, and Trustworthiness at $k = 1$ increased from 0.7853 to 0.8244. MRRE also decreased (e.g., from 2113 at $K = 1$ to 2009), indicating that rank distortions were slightly reduced. These results suggest that the additional dimension provides meaningful extra capacity to represent sparse, high-dimensional semantic structure.

Coil20 experienced small but consistent improvements. For Coil20, kNN accuracy at $k = 1$ increased from 0.8438 to 0.8611, while Trustworthiness at $k = 1$ rose from 0.9849 to 0.9896, reflecting that its structure was already well captured in 2D.

Olivetti benefited slightly from the additional dimension. kNN accuracy at $k = 1$ increased from 0.4750 in 2D to 0.5625 in 3D, and Trustworthiness at $k = 1$ improved from 0.9170 to 0.9212. MRRE values decreased moderately as well. These modest improvements indicate that while 3D helps partially recover local facial structure, distinguishing identities remains difficult.

Across all datasets, the shift from 2D to 3D generally improves local and global structure preservation, but the degree of improvement varies substantially. Datasets with relatively simple structure, such as Coil20, show only marginal improvement when expanded to three dimensions. In contrast, more complex and highly dimensional datasets, most notably 20Newsgroups, benefit substantially from the additional degree of freedom. These findings suggest that while 2D embeddings are ideal for visualization, 3D embeddings can provide a useful balance between interpretability and compression for more challenging datasets.

Table 6: Changes in key evaluation metrics when increasing PaCMAP from 2D to 3D. Positive values indicate improvement.

Dataset	Δ kNN (k=1)	Δ Trust. (k=1)	Δ Cont. (K=1)	Δ MRRE (K=1)
Coil20	+0.0173	+0.0047	+0.0003	-0.3707
Olivetti	+0.0875	+0.0042	+0.0069	-2.19
20Newsgroups	+0.0634	+0.0391	+0.0070	-104.52

8 Conclusion

In this work, we evaluated PaCMAP across four datasets—Coil20, MNIST, Olivetti, and 20Newsgroups—using quantitative metrics (kNN accuracy, Trustworthiness, Continuity, and MRRE) and qualitative 2D visualizations. Our results show that PaCMAP performs strongly on datasets with well-separated structure such as Coil20 and MNIST, but degrades on more complex datasets like Olivetti and especially 20Newsgroups, where high dimensionality and sparse semantics make structure difficult to preserve in 2D.

Extending PaCMAP to three dimensions produced modest but consistent improvements, with the largest gains appearing in the most complex dataset (20Newsgroups). This suggests that while 2D embeddings are ideal for visualization, higher-dimensional embeddings can provide meaningful benefits for preserving structure in challenging datasets.

Overall, the effectiveness of PaCMAP depends heavily on dataset characteristics: it excels on clean visual data, performs moderately on subtle high-dimensional data, and struggles with text data. These findings

highlight the importance of matching dimensionality-reduction techniques—and embedding dimensionality—to the underlying structure of the data.