

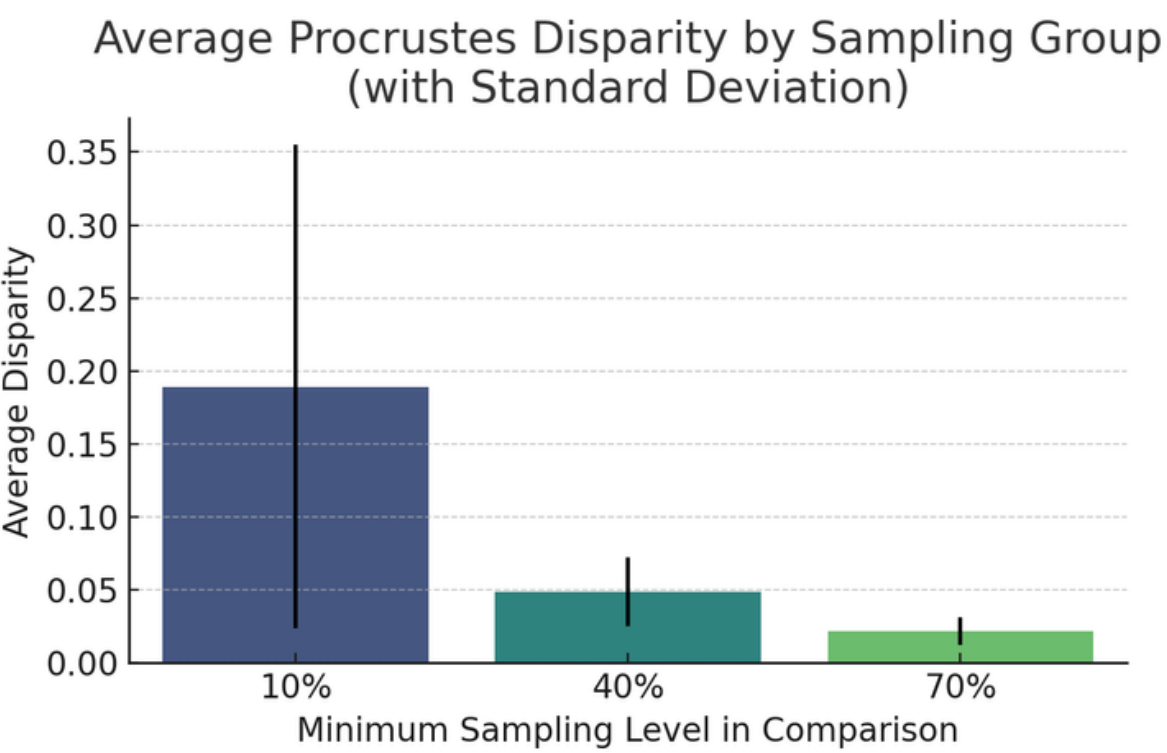
1. Introduction

t-SNE is an algorithm that visualizes high- dimensional data by giving each datapoint a location in a two or three-dimensional map [1]. It is slow and highly sensitive to perplexity, especially on large datasets. Choosing a good perplexity is challenging, and current evaluations rely heavily on visual inspection. There is no established quantitative method to validate sample-based t-SNE embeddings.

2. Objective

The goal of the research is to show that it is possible to find a quantitative method to compare the embeddings of two different visualizations and to provide users a medium to choose a suitable perplexity without exhaustive experiments.

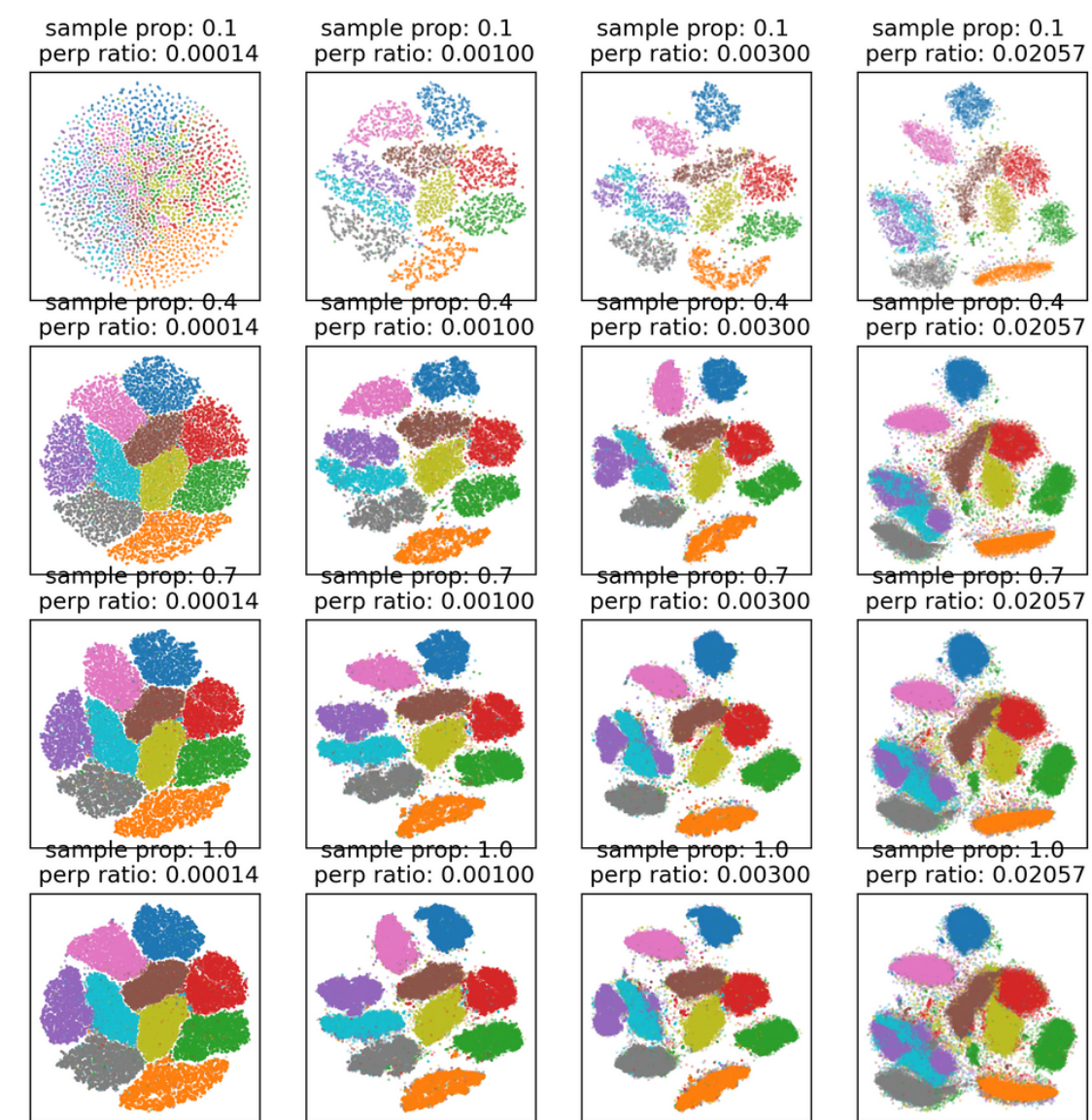
5. Analysis



- Sampling size has a stronger effect on embedding stability than perplexity
- Low sampling (10%) consistently yields high disparities, even with scaled perplexity (see MNIST col-wise comparison above)
- Embeddings are more stable column-wise (varying sample size) than row-wise (varying perplexity)
- Stability improves at ≥40-70% sampling with moderate perplexity, suggesting a sweet spot
- Results are consistent across datasets (MNIST, FMNIST, C. elegans) and random seeds
- Wasserstein distance confirmed Procrustes trends, supporting its use as an alternative similarity metric

3. Methodology

We use sample-based t-SNE to embed subsets of MNIST, FMNIST, and C. elegans data, varying both sampling proportions (10%, 40%, 70%, 100%) and perplexity (scaled per dataset) in a 4×4 grid. To compare embeddings, we apply Procrustes analysis on shared data points, aligning each pair and computing a disparity score that quantifies structural similarity.



The table above represents the grid we will be analyzing:

- Row-wise: fixed sample, varying perplexity
- Column-wise: fixed perplexity, varying sample size

Each setup is repeated with 3 random seeds for stability. This enables detecting robust embedding regions and guides perplexity selection without exhaustive tuning

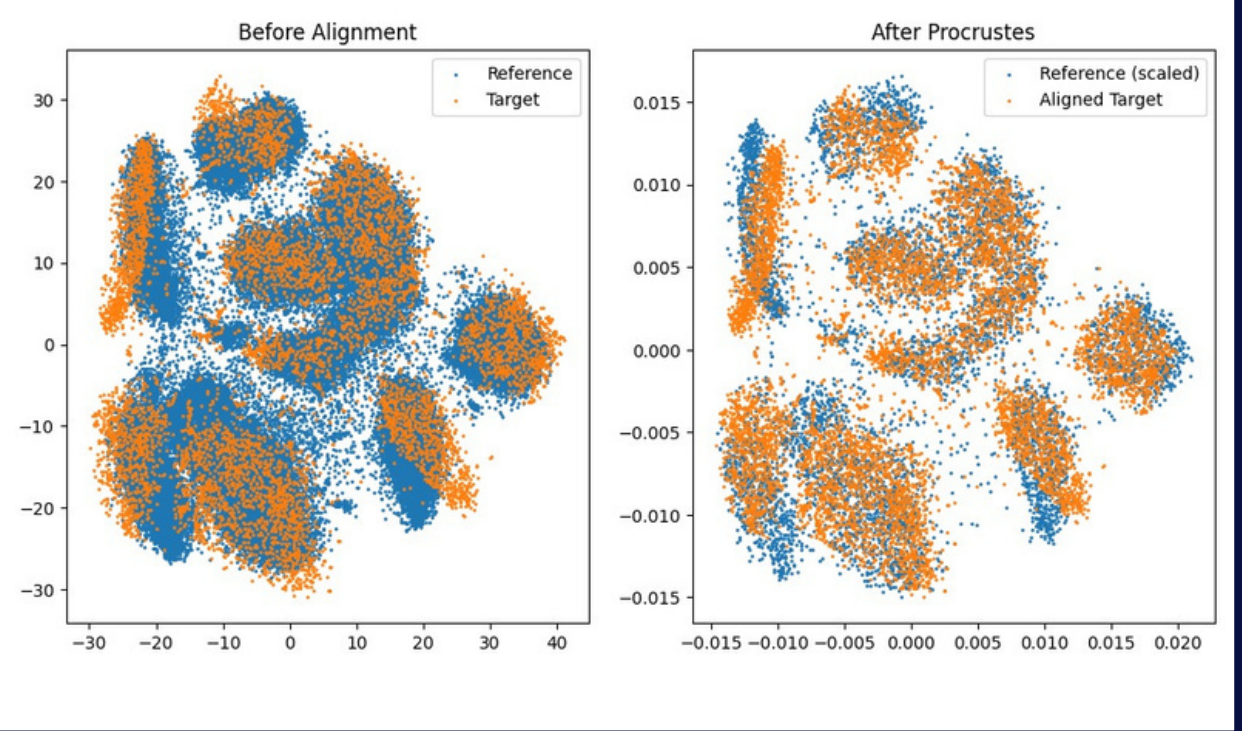
6. Future Work

Future work may include looking further into Wasserstein distance as an alternative metric.

4. Results

- A total of 48 Procrustes comparisons were performed:
 - 24 row-wise (fixed sample, varying perplexity)
 - 24 column-wise (fixed perplexity, varying sample size)
 - Each configuration was run with 3 random seeds (42, 100, 12)
 - For each pair, Procrustes disparity was computed on shared data points
 - Average disparity scores were collected in tables and heatmaps
 - Additional comparisons using Wasserstein distance showed consistent structural trends
- Example metrics:
- Row-wise (10%, low vs high perplexity): Disparity ≈ 0.4379
 - Column-wise (70% vs 100%, same perplexity): Disparity ≈ 0.0451

Below we can see an example of points before and after geometric alignment. (scaling, translation, rotation)



7. Conclusion

In summary, our work demonstrates that Procrustes analysis offers a quantitative approach to assess the structural similarity of embeddings, enabling us to identify suitable perplexity values without requiring exhaustive experimentation.