Implicit Multidimensional Projection of Local Subspaces—Supplemental Material

Category: Research

Paper Type: algorithm/technique

Abstract—Contents that are left out from the paper for conciseness are documented in this supplemental material. Derivatives used for computing the implicit differentiation—for MDS and t-SNE—are reported. We show results of all examples used in the evaluation. Finally, further examples comparing our method and traditional point-based visualization are presented.

1 Partial Derivatives of MDS

The following results are derived from the SMACOF version of MDS [1].

1.1 Objective Function

$$f = F = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\left(\left\|x_i - x_j\right\| - \left\|y_i - y_j\right\|\right)^2$$

1.2 Partial Derivative of F w.r.t. Y

$$\frac{\partial F}{\partial y_i} = 2\sum_{j\neq i}\left(1 - \frac{\left\|x_i - x_j\right\|}{\left\|y_i - y_j\right\|}\right)\left(y_i - y_j\right)$$

1.3 Second Derivatives of the Objective Function w.r.t. Y

The following equations are referenced in Section 4.4—MDS in our paper.

1.3.1 $\frac{\partial^2 F}{\partial y_i \partial y_j}$

$$\frac{\partial^2 F}{\partial y_i \partial y_j} = \begin{cases} 2\sum_{k\neq i}\left(\left(1 - \frac{\|x_i - x_k\|}{\|y_i - y_k\|}\right)I + \frac{\|x_i - x_k\|}{\|y_i - y_k\|^3}\left(y_i - y_k\right)\left(y_i - y_k\right)^T\right) & \text{if} \quad i = j \\ 2\left(\frac{\|x_i - x_j\|}{\|y_i - y_j\|} - 1\right)I - \frac{2\|x_i - x_j\|}{\|y_i - y_j\|^3}\left(y_i - y_j\right)\left(y_i - y_j\right)^T & \text{otherwise} \end{cases} \tag{1}$$

1.3.2 $\frac{\partial^2 F}{\partial y_i \partial x_j}$

$$\frac{\partial^2 F}{\partial y_i \partial x_j} = \begin{cases} \sum_{k\neq i} -2\frac{(y_i - y_k)}{\|y_i - y_k\|}\frac{(x_i - x_k)^T}{\|x_i - x_k\|}, & \text{if} \quad i = j \\ 2\frac{(y_i - y_j)}{\|y_i - y_j\|}\frac{(x_i - x_j)^T}{\|x_i - x_j\|}, & \text{otherwise}. \end{cases} \tag{2}$$

2 Partial Derivatives of t-SNE

Our method is based on the original t-SNE [5].

2.1 Objective Function

$$f = C = KL(P\|Q) = \sum_i\sum_j p_{ij}\log\frac{p_{ij}}{q_{ij}},$$

where

$$p_{ij} = \frac{p_{j|i} + p_{i|j}}{2n},$$

$$q_{ij} = \frac{\left(1 + \left\|y_i - y_j\right\|^2\right)^{-1}}{\sum_{k\neq l}\left(1 + \|y_k - y_l\|^2\right)^{-1}},$$

$$p_{j|i} = \frac{\exp\left(-\left\|x_i - x_j\right\|^2 / 2\sigma_i^2\right)}{\sum_{k\neq i}\exp\left(-\|x_i - x_k\|^2 / 2\sigma_i^2\right)},$$

and σ_i is determined by the perplexity of t-SNE.

2.2 Partial Derivative of Objective Function w.r.t. Y

$$\frac{\partial C}{\partial y_i} = 4\sum_j\left(p_{ij} - q_{ij}\right)\left(1 + \left\|y_i - y_j\right\|^2\right)^{-1}\left(y_i - y_j\right)$$

2.3 Second Derivatives of Objective Function w.r.t. Y

Equations that are referenced in Section 4.4—t-SNE in our paper are shown as follows.

2.3.1 $\frac{\partial^2 C}{\partial y_i \partial y_j}$

$$\frac{\partial^2 C}{\partial y_i \partial y_j} = \begin{cases} 4\left(M_1 + M_2 + M_3\right), & \text{if } i = j \\ 4\left(A_1 + A_2 + A_3\right), & \text{otherwise} \end{cases} \tag{3}$$

where

$$M_1 = \sum_{k\neq i} -\left(1 + \|y_i - y_k\|^2\right)^{-1}\left(y_i - y_k\right)\left(-2q_{ik}\left(1 + \|y_i - y_k\|^2\right)^{-1}\left(y_i - y_k\right)^T + 4q_{ik}\sum_{l\neq a} q_{il}\left(1 + \|y_i - y_l\|^2\right)^{-1}\left(y_i - y_l\right)^T\right),$$

$$M_2 = -2\sum_{k\neq i}\left(p_{ik} - q_{ik}\right)\left(1 + \|y_i - y_k\|^2\right)^{-2}\left(y_i - y_k\right)\left(y_i - y_k\right)^T,$$

$$M_3 = \sum_{k\neq i}\left(p_{ik} - q_{ik}\right)\left(1 + \|y_i - y_k\|^2\right)^{-1} I,$$

$$A_1 = 2\left(p_{ij} - 2q_{ij}\right)\left(1 + \left\|y_i - y_j\right\|^2\right)^{-2}\left(y_i - y_j\right)\left(y_i - y_j\right)^T,$$

$$A_2 = -\left(p_{ij} - q_{ij}\right)\left(1 + \left\|y_i - y_j\right\|^2\right)^{-1} I,$$

$$A_3 = \left(\sum_k -4q_{ik}^2\left(y_i - y_k\right)\right)\left(\sum_{k\neq j}\left(1 + \left\|y_j - y_k\right\|^2\right)^{-2}\left(y_j - y_k\right)^T\right).$$

2.3.2 $\frac{\partial^2 C}{\partial y_i \partial x_j}$

$$\frac{\partial^2 C}{\partial y_i \partial x_j} = \frac{2}{n}\sum_{k\neq i}\left(1 + \|y_i - y_k\|^2\right)^{-1}\left(y_i - y_k\right)\left(\frac{\partial p_{i|k}}{\partial x_j} + \frac{\partial p_{k|i}}{\partial x_j}\right) \tag{4}$$

where

$$\frac{\partial p_{i|k}}{\partial x_j} = \begin{cases} \frac{x_i - x_k}{\sigma_k^2}\left(p_{i|k} - 1\right)p_{i|k} & j = i \\ \frac{p_{i|j}}{\sigma_j^2}\left(x_i - \sum_{l\neq j} p_{l|j} x_l\right) & j = k \\ p_{i|k} p_{j|k}\frac{x_j - x_k}{\sigma_k^2} & j \neq i, k \end{cases} \tag{5}$$

3 Evaluation

In our paper, only visualizations of the Iris data (Fig. 3) are included due to the page limit. Here, we show results that are left out from the paper.

Table 1: Statistics of transformed basis vectors of non-border data points of the planar data

	mean length of the first basis vector	mean length of the second basis vector	mean angle	std of angles
Our	0.10068096	0.10068095	89.92020374	0.00502165
Rand [3]	0.10005878	0.09957406	89.03798816	0.50298048

Table 2: Computation time (in seconds) of transformation methods evaluated on test datasets.

	Dataset	Ours	Rand [3]
MDS	Iris	0.55	3.13
	Wine	0.86	3.30
	Digits40	59.69	131.67
t-SNE	Iris	2.11	21.17
	Wine	3.58	36.23
	Digits40	49.60	332.36

3.1 Synthetic Data

In order to verify the correctness of our implicit function method. A synthetic planar data sampled on a 3D regular grid is generated and projected to 2D using MDS. With a neighborhood of $k = 8$, the transformed basis vectors of all data points that are not on the border of the plane should be, theoretically, orthogonal, i.e., having an angle of 90 degrees, and having the same magnitude.

It can be seen in Fig. 1 that glyph of transformed basis vectors with our method appear identical(Fig. 1 (a)), whereas glyph of different shapes are clearly visible in the result of the random approach (Fig. 1 (b)). Quantitatively, Table 1 summarizes basic statistics of transformed basis vectors of non-border data points of both methods. Distributions of the lengths of the two transformed basis vectors can be seen in Fig. 2 (a, b), and distributions of angles between basis vectors are seen (Fig. 2 (c, d)). These are evidence showing that our method is more accurate compared to the random method.

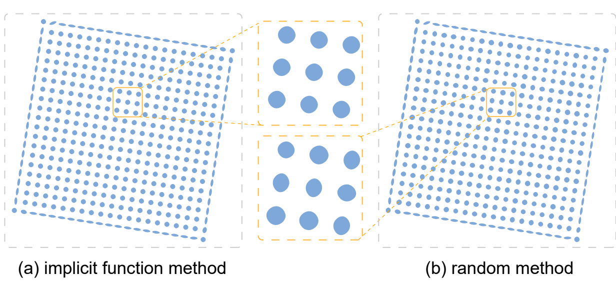


Fig. 1: Transformed basis vectors visualized as glyph by (a) our implicit function method and (b) the random method [3]. The data points are on a plane in 3D sampled using a regular grid, and projected to 2D with MDS.

3.2 Real-World Data

We show results of the wine data (Fig. 4), and digits data (Fig. 5) that are left out from the paper. It can be seen that compared to our implicit function-based vector transformation, the random perturbation method [3] generates more abnormal glyphs that are excessively long and thin.

We measure computation times for the two methods for transforming the top five (four for the Iris data) basis vectors, and visualize projected basis vectors using our glyph drawing technique. Table 2 summarizes the computation times for each method.

4 Further Examples of Our Method

In this section, we include interesting examples that are left out from the paper. For each dataset, we compare results of our method to

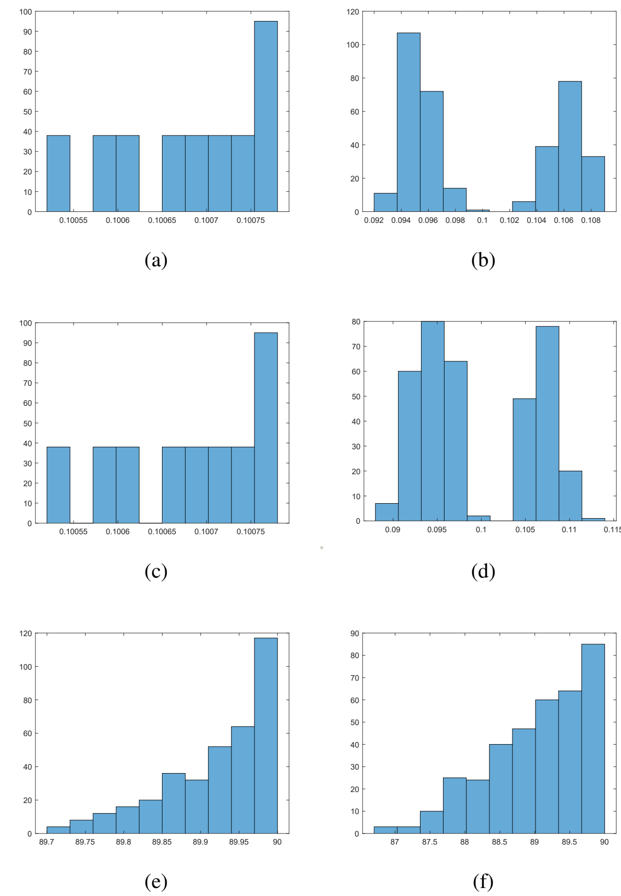
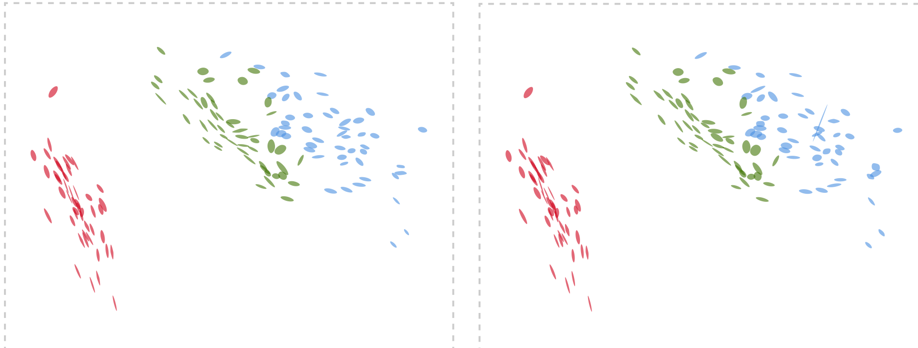
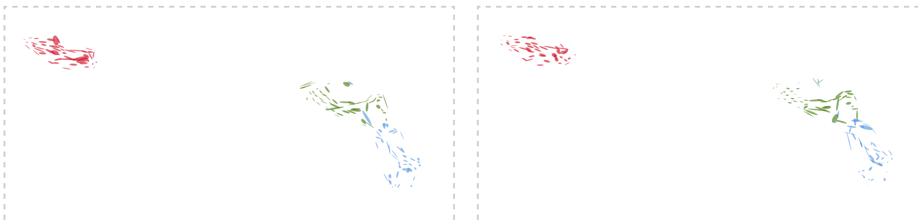


Fig. 2: Histograms of the lengths of first transformed basis vectors with our method (a) and the random method (b). Second transformed basis vectors: (c) our method, and (d) random method. Distributions of angles of the two basis vectors are shown as histograms: (e) our method, and (f) the random method.

the point-based visualization. Local features of interest in computer vision data are shown with zoom-ins and compared with the image data. Specifically, results of multidimensional datasets including “Iris”, “olive”, and “Shuttle” (all obtained from the UCI machine learning repository [2]). The “pendigits” data of 7494 data points of 16-D Computer vision datasets “digit1s”, “fashion” are shown in Fig. 6 and Fig. 7, respectively. The “digit1s” is a subset of the “digits” dataset [4] with handwritten 1s, and the “fashion” data is a subset of the “fashionMNIST” dataset [6] with three classes. The pendigits data (obtained from the UCI machine learning repository [2]) is the largest dataset we have experimented on our machine, it contains 7494 data points of 16-D, and is shown in Fig. 18.

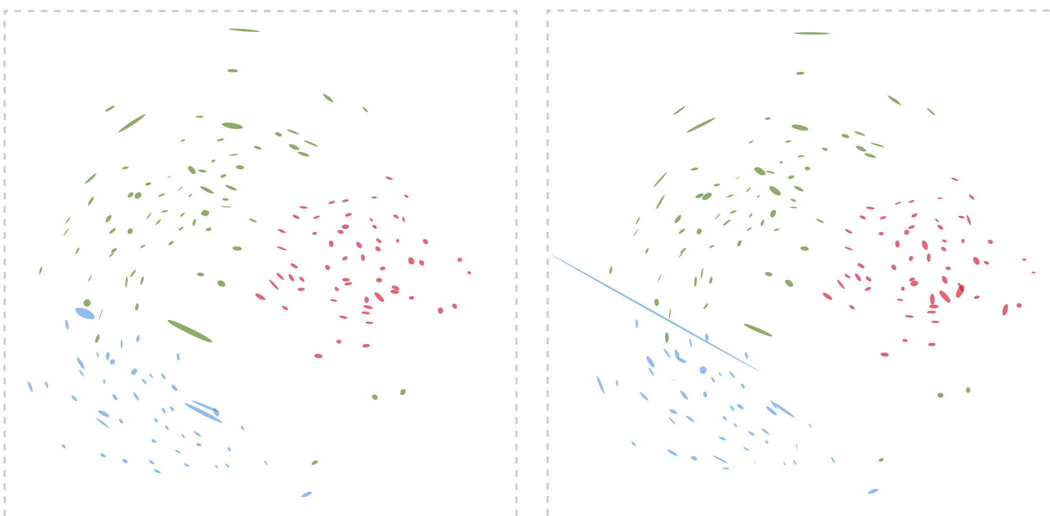


(a) MDS

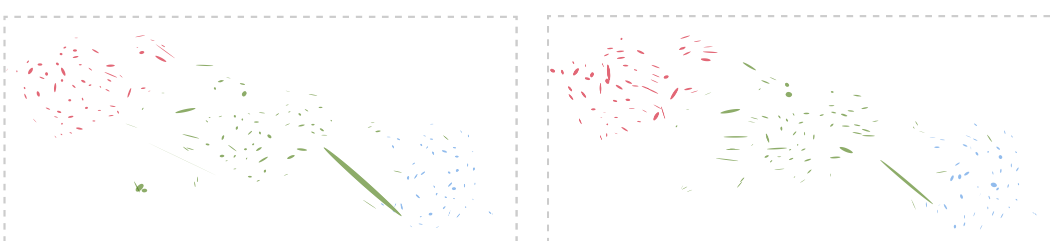


(b) t-SNE

Fig. 3: Glyph visualizations comparing basis vectors transformed by our implicit function method (left) and the random method [3] (right) on the Iris data.

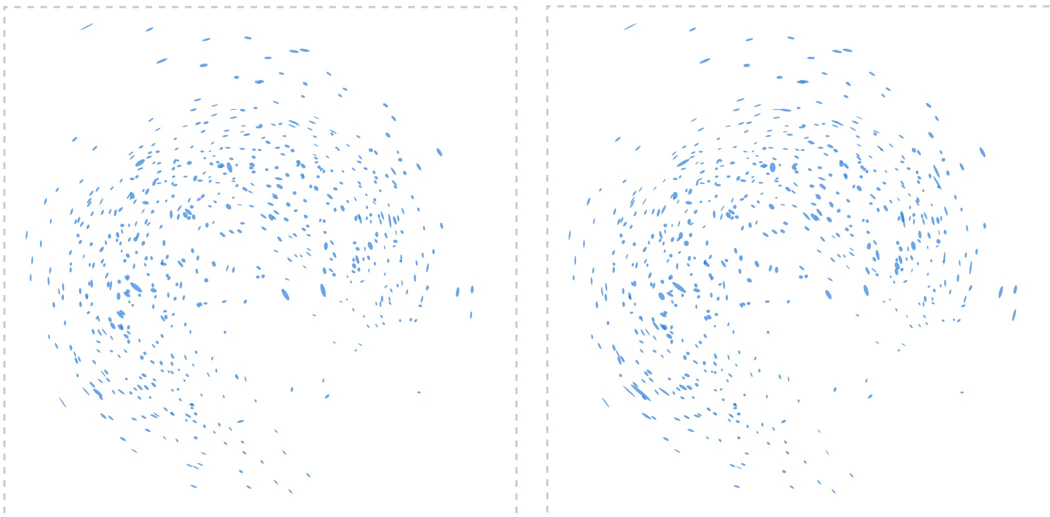


(a) MDS

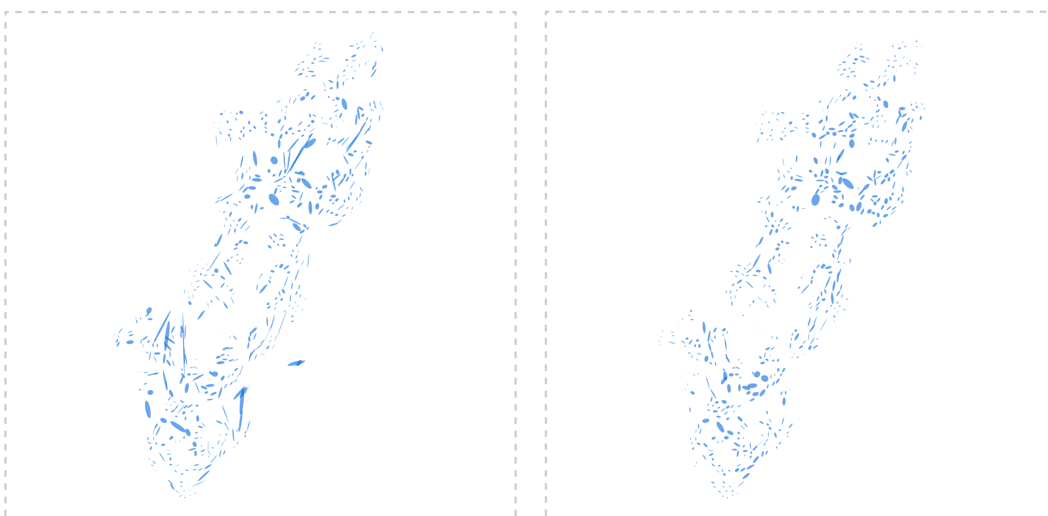


(b) t-SNE

Fig. 4: Glyph visualization with basis vectors transformed by our implicit function method (left) and the random method [3] (right) on the wine data.

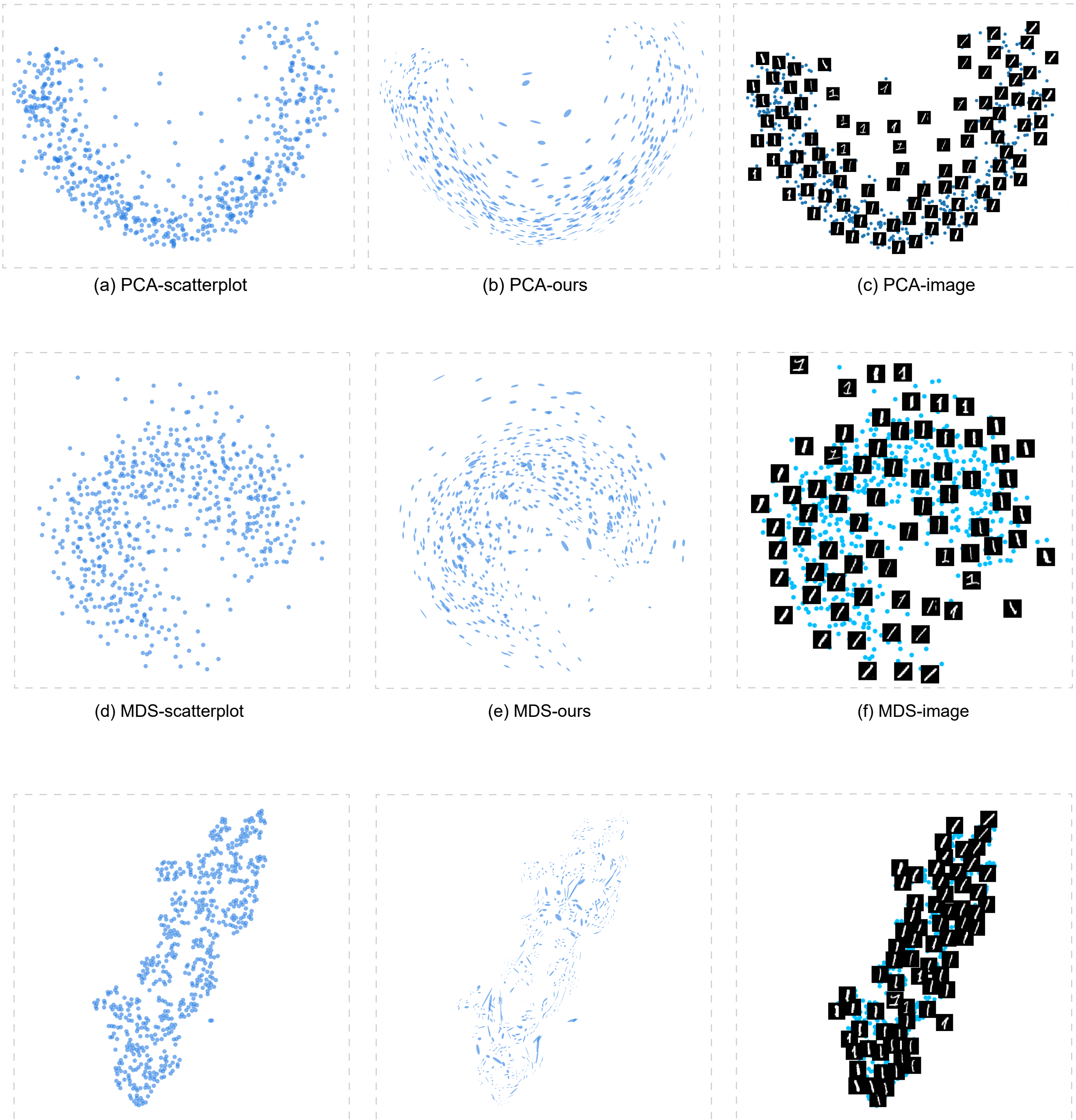


(a) MDS



(b) t-SNE

Fig. 5: Glyph visualization with basis vectors transformed by our implicit function method (left) and the random method [3] (right) on the digits data.



(a) PCA-scatterplot (b) PCA-ours (c) PCA-image

(d) MDS-scatterplot (e) MDS-ours (f) MDS-image

(g) t-SNE-scatterplot (h) t-SNE-ours (i) t-SNE-image

Fig. 6: Visualizations of a subset of the digits data with hand-written digit 1s, projected with (a, b, c) PCA, (d, e, f) MDS, and (g, h, i) t-SNE.

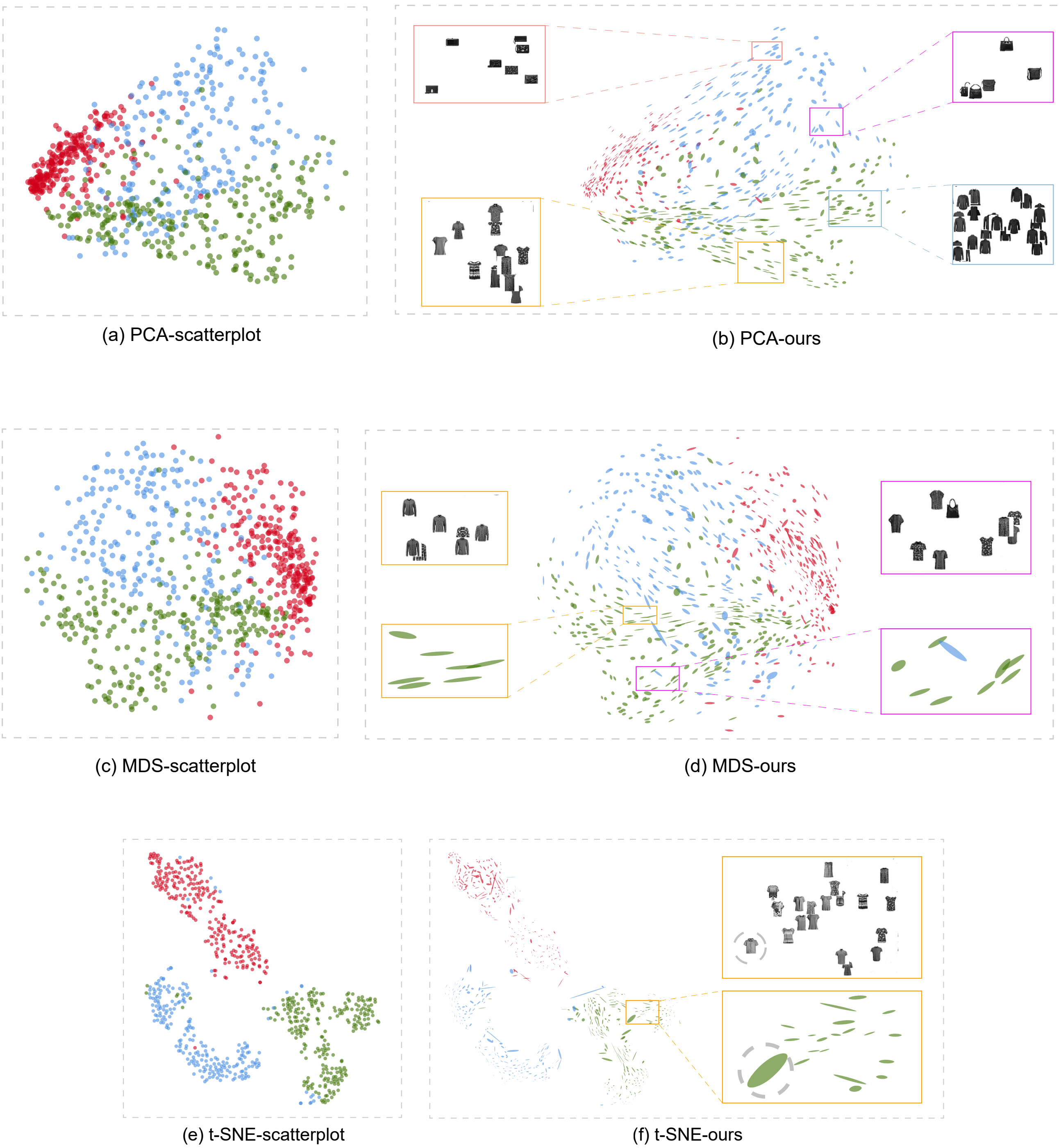
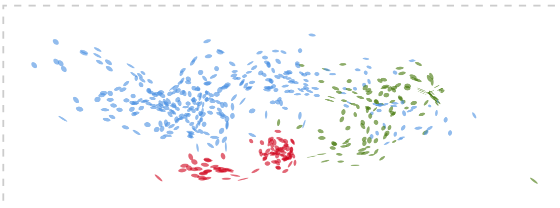
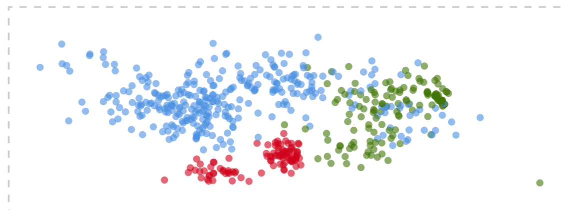


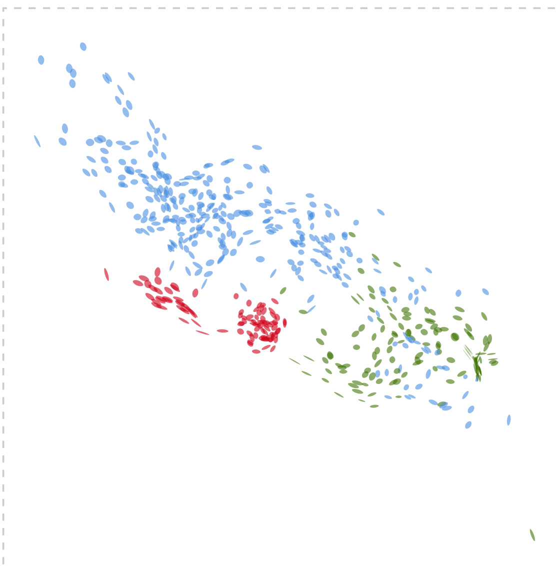
Fig. 7: Visualizations of the fashion data projected with (a, b) PCA, (c, d) MDS, and (e, f) t-SNE.



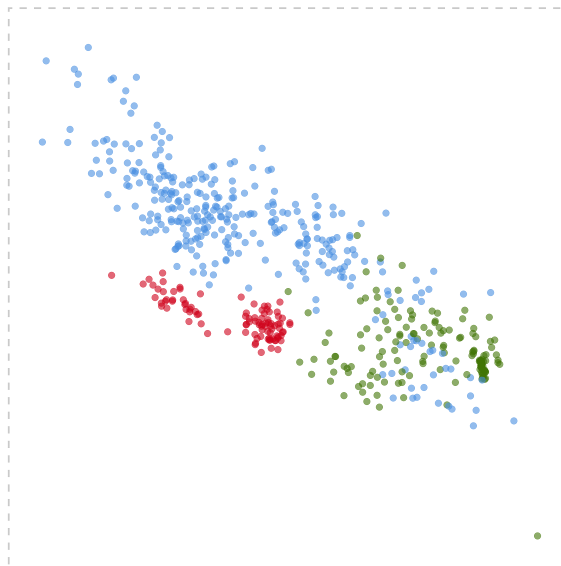
(a) PCA-ours



(b) PCA-scatterplot

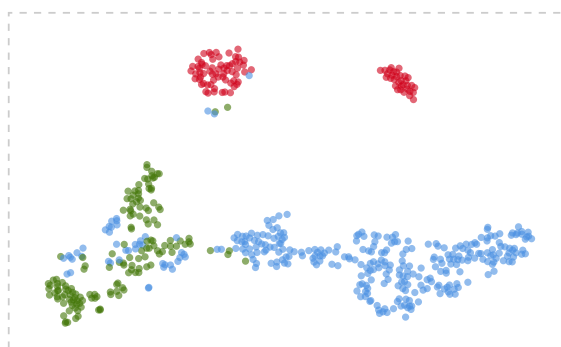


(c) MDS-ours



(d) MDS-scatterplot

(e) t-SNE-ours



(f) t-SNE-scatterplot

Fig. 8: Visualizations of the olive data projected with (a, b) PCA, (c, d) MDS, and (e, f) t-SNE.

(a) PCA-ours

(b) PCA-scatterplot

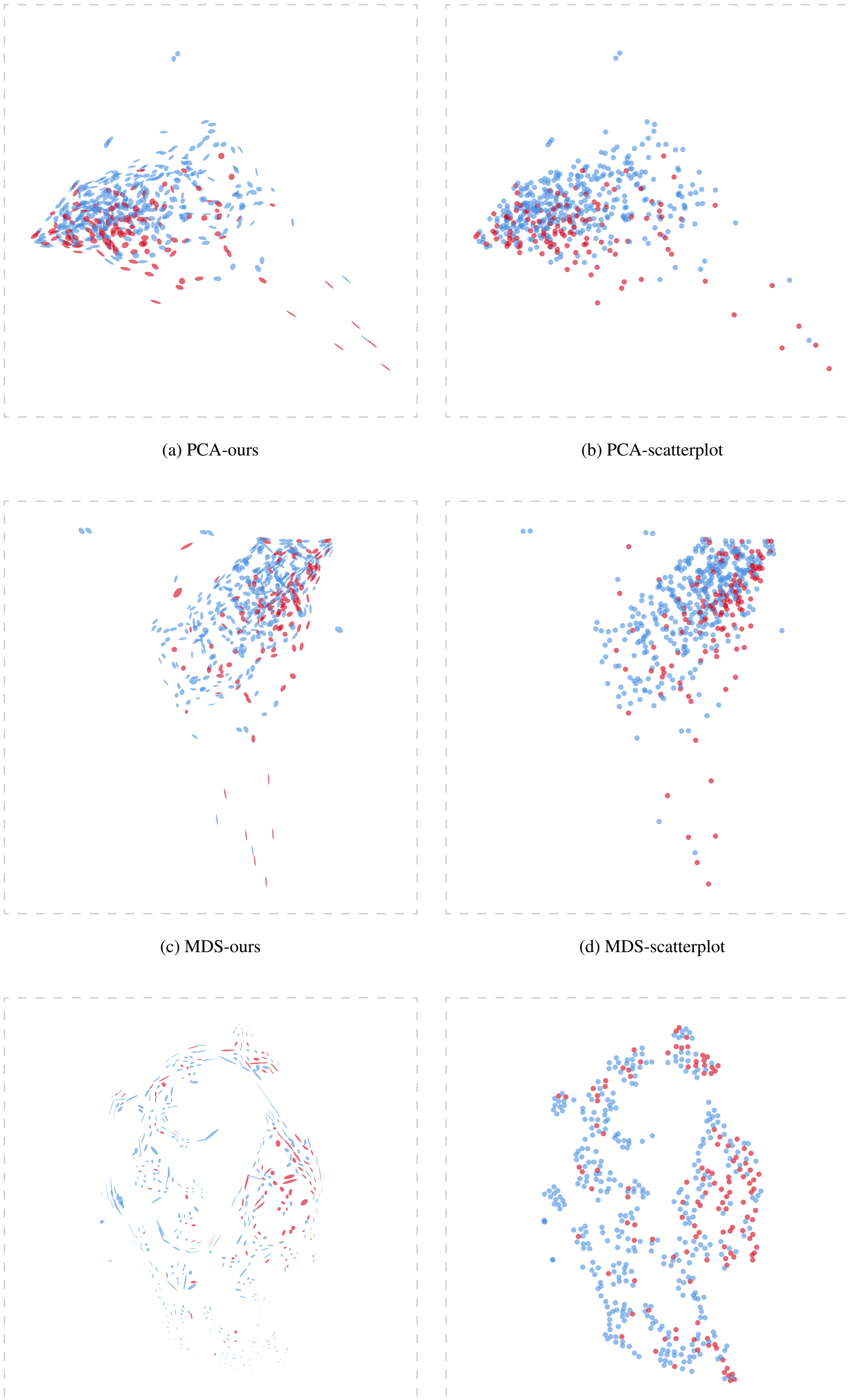
(c) MDS-ours

(d) MDS-scatterplot

(e) t-SNE-ours

(f) t-SNE-scatterplot

Fig. 9: Visualizations of the Iris data projected with (a, b) PCA, (c, d) MDS, and (e, f) t-SNE.



(a) PCA-ours

(b) PCA-scatterplot

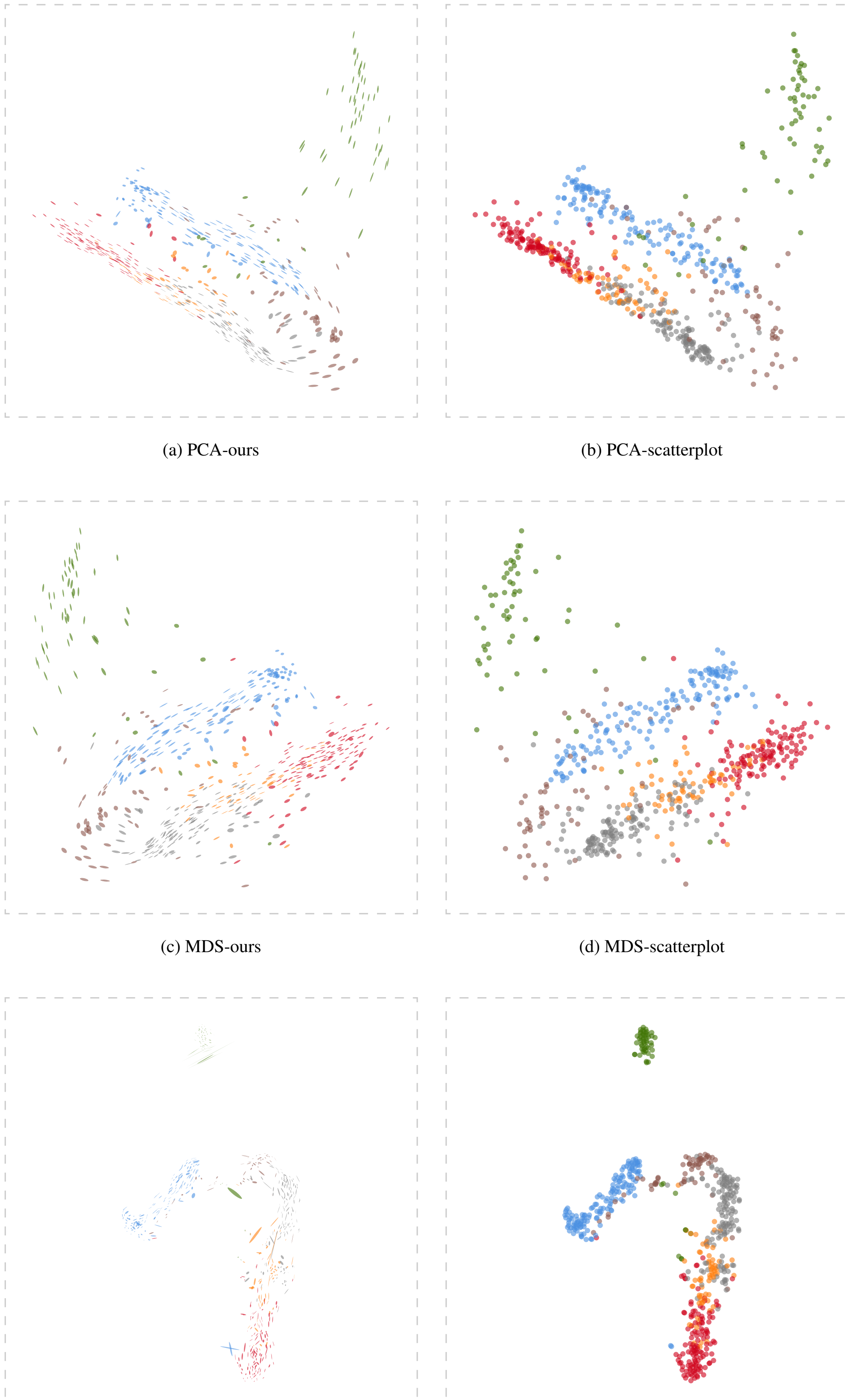
(c) MDS-ours

(d) MDS-scatterplot

(e) tsne-ours

(f) tsne-scatterplot

Fig. 10: Visualizations of Blood Transfusion Service Center data projected with (a, b) PCA, (c, d) MDS and (e, f) t-SNE.

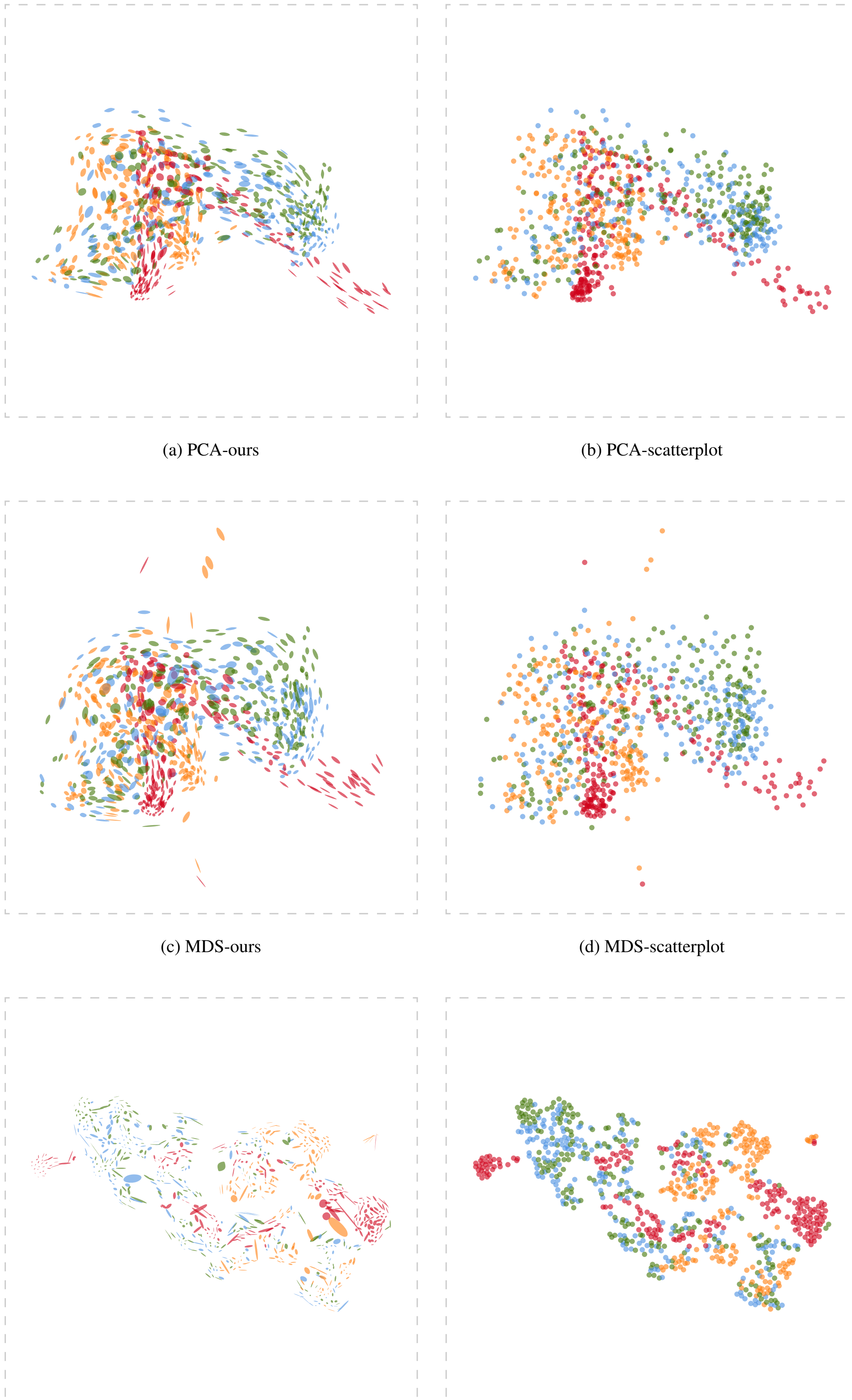


(a) PCA-ours
(b) PCA-scatterplot
(c) MDS-ours
(d) MDS-scatterplot
(e) t-SNE-ours
(f) t-SNE-scatterplot

Fig. 11: Visualizations of Statlog(Landsat Satellite) data projected with (a, b) PCA, (c, d) MDS and (e, f) t-SNE.

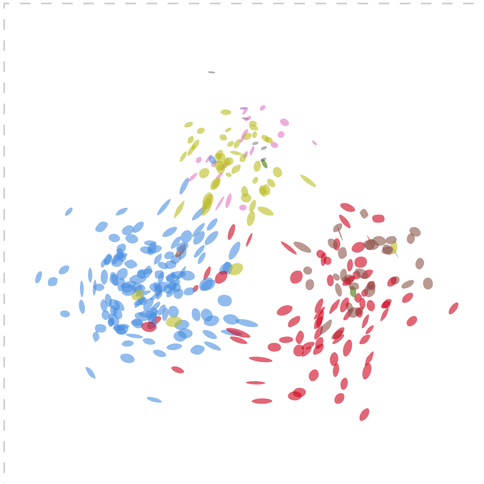
(a) PCA-ours
(b) PCA-scatterplot
(c) MDS-ours
(d) MDS-scatterplot
(e) t-SNE-ours
(f) t-SNE-scatterplot

Fig. 12: Visualizations of Statlog(Shuttle) data projected with (a, b) PCA, (c, d) MDS and (e, f) t-SNE.

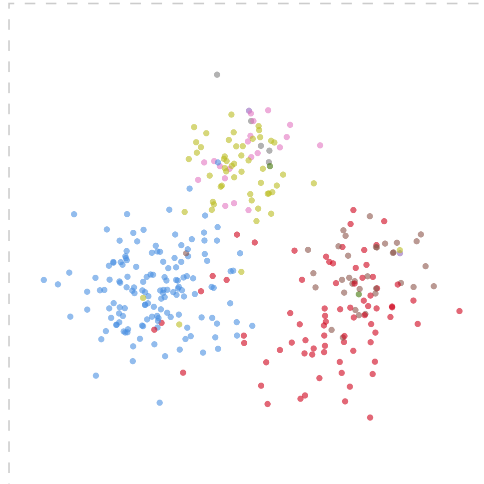


(a) PCA-ours
(b) PCA-scatterplot
(c) MDS-ours
(d) MDS-scatterplot
(e) t-SNE-ours
(f) t-SNE-scatterplot

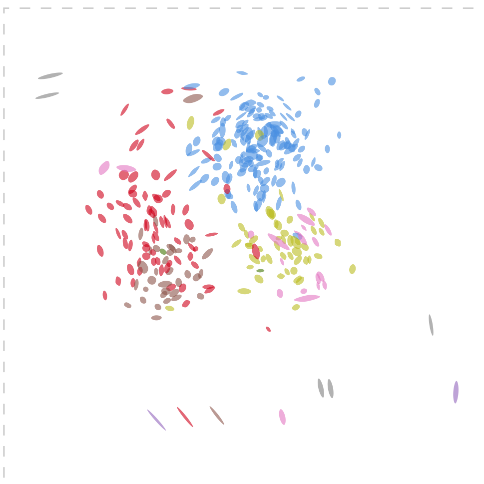
Fig. 13: Visualizations of Statlog(Vehicle Silhousettes) data projected with (a, b) PCA, (c, d) MDS and (e, f) t-SNE.



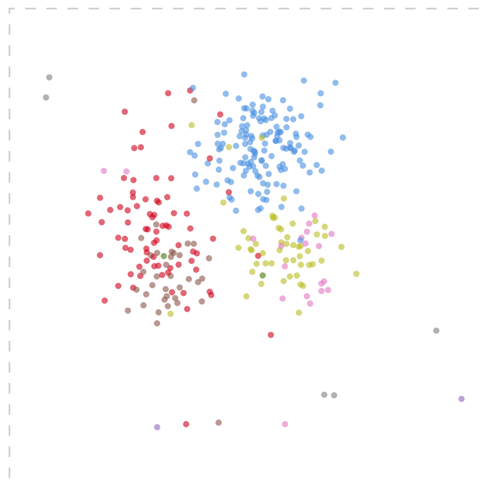
(a) PCA-ours



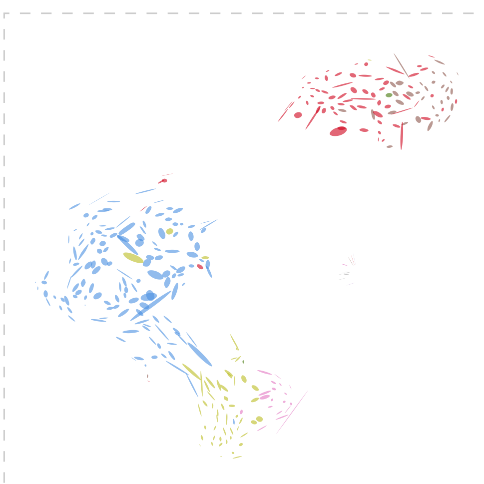
(b) PCA-scatterplot



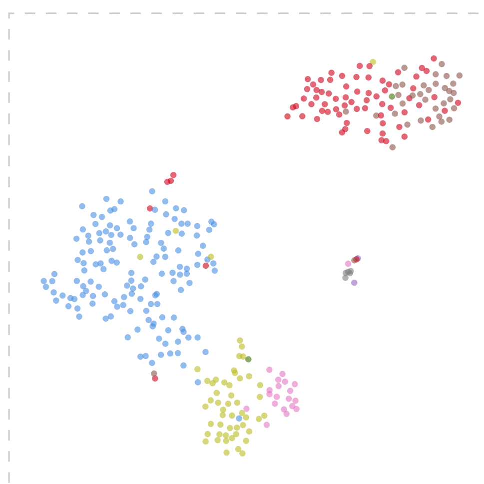
(c) MDS-ours



(d) MDS-scatterplot

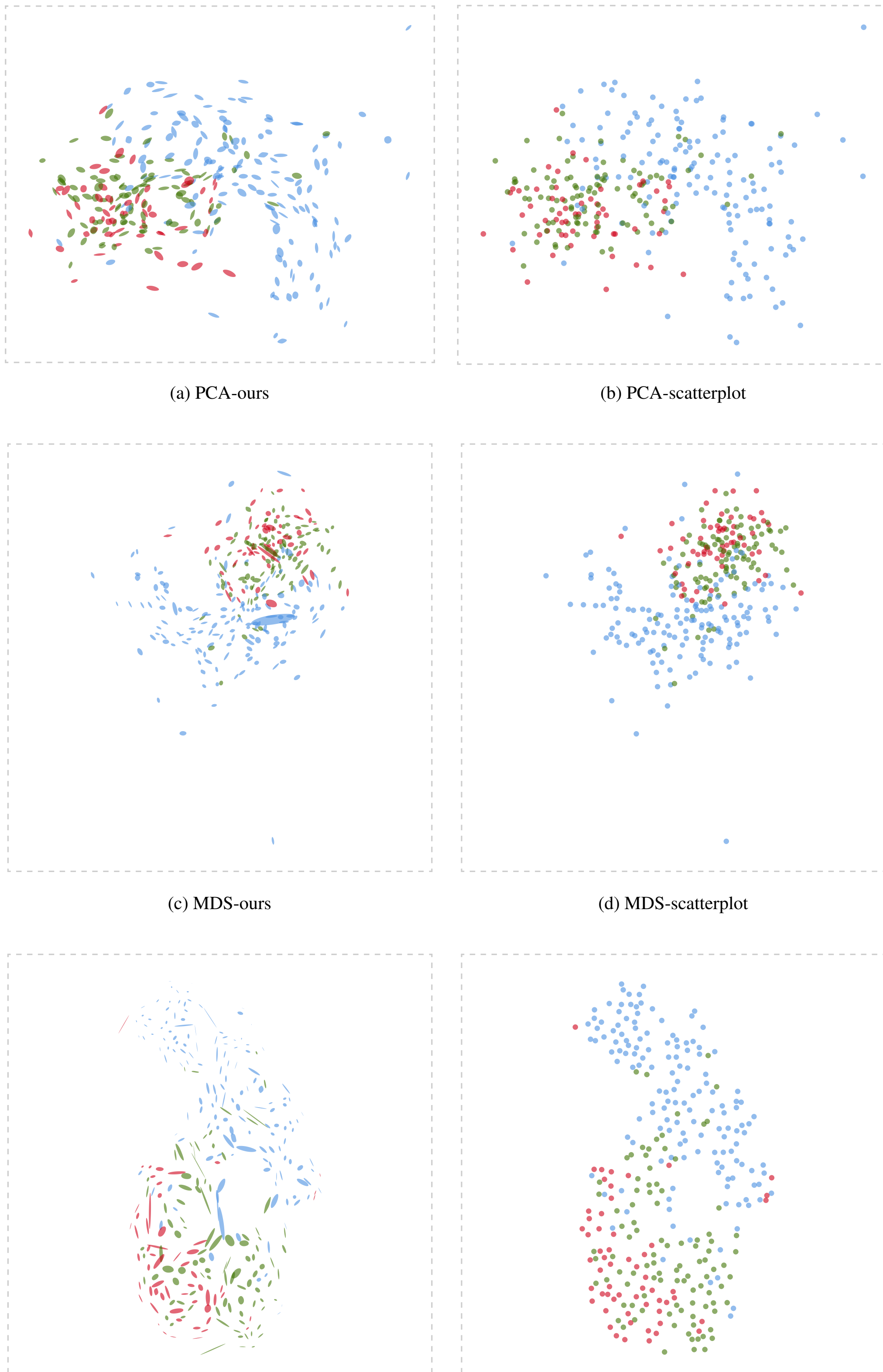


(e) t-SNE-ours



(f) t-SNE-scatterplot

Fig. 14: Visualizations of ecoli data projected with (a, b) PCA, (c, d) MDS and (e, f)t-SNE.

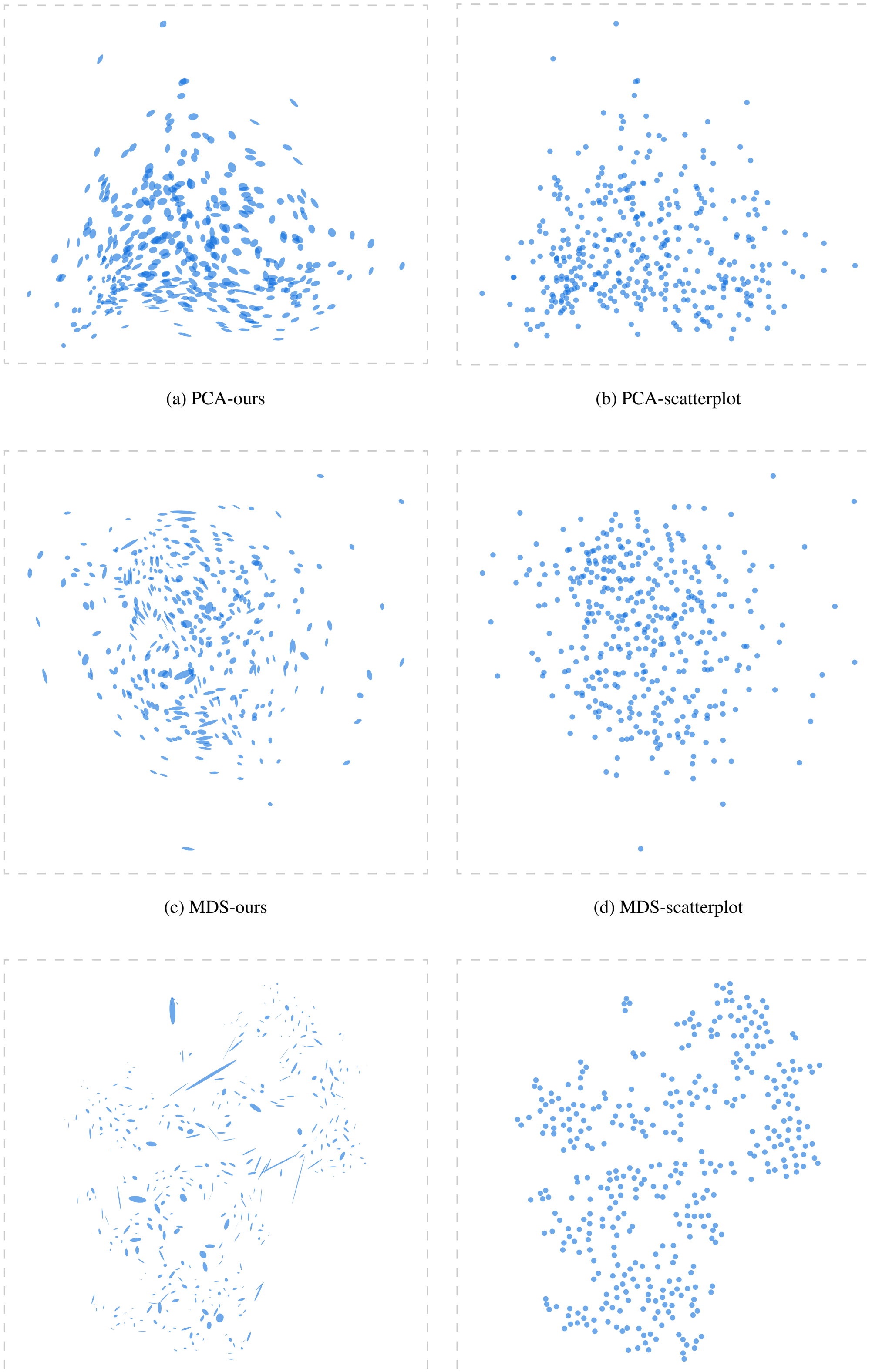


(a) PCA-ours
(b) PCA-scatterplot

(c) MDS-ours
(d) MDS-scatterplot

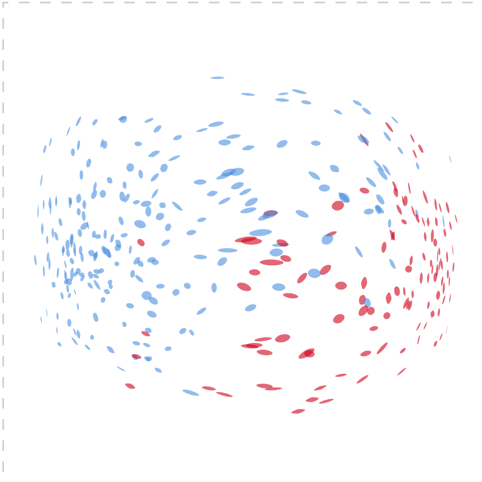
(e) t-SNE-ours
(f) t-SNE-scatterplot

Fig. 15: Visualizations of the Vertebral Column data projected with (a, b) PCA, (c, d) MDS, and (e, f) t-SNE.

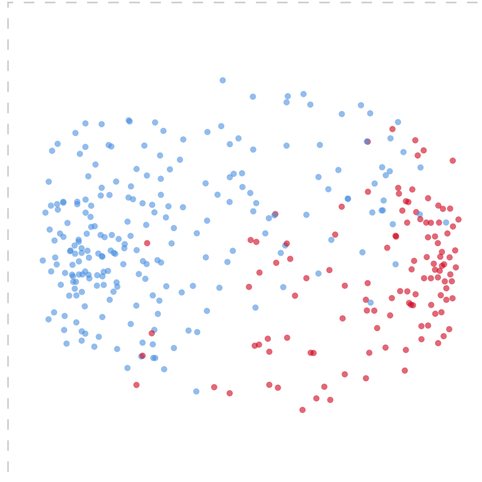


(a) PCA-ours (b) PCA-scatterplot (c) MDS-ours (d) MDS-scatterplot (e) t-SNE-ours (f) t-SNE-scatterplot

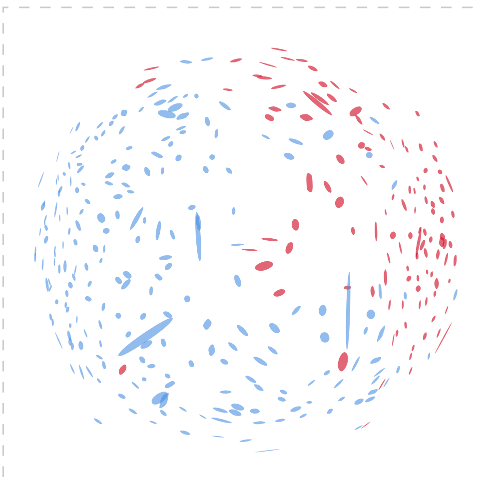
Fig. 16: Visualizations of WineQuality-Red data projected with (a, b) PCA, (c, d) MDS and (e, f) t-SNE.



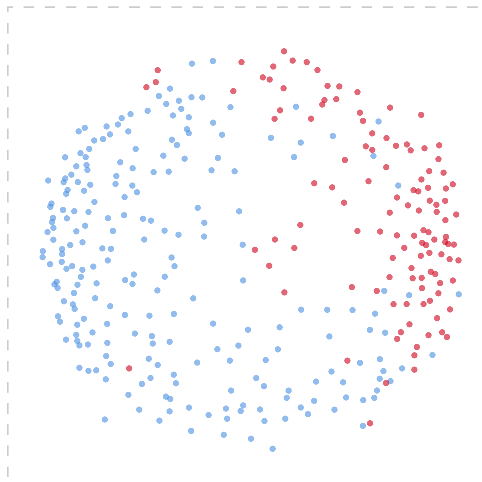
(a) PCA-ours



(b) PCA-scatterplot

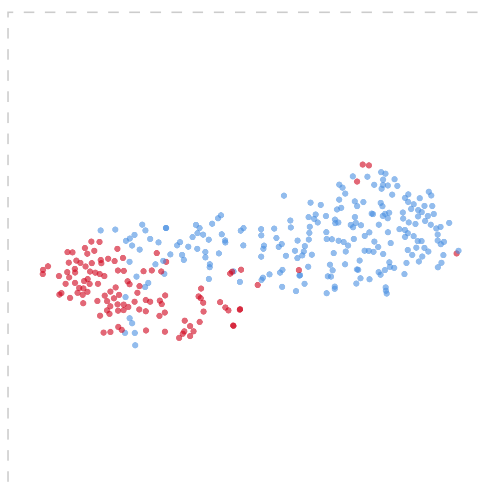


(c) MDS-ours



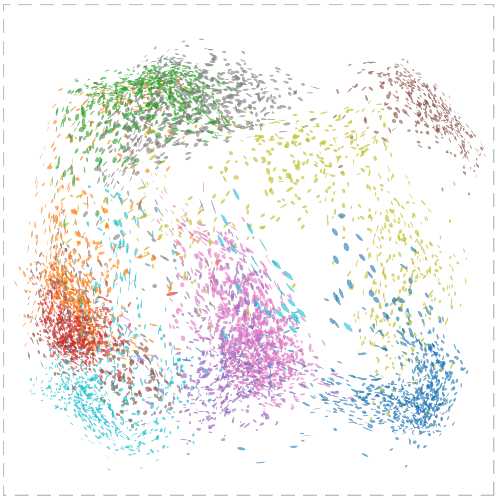
(d) MDS-scatterplot

(e) t-SNE-ours

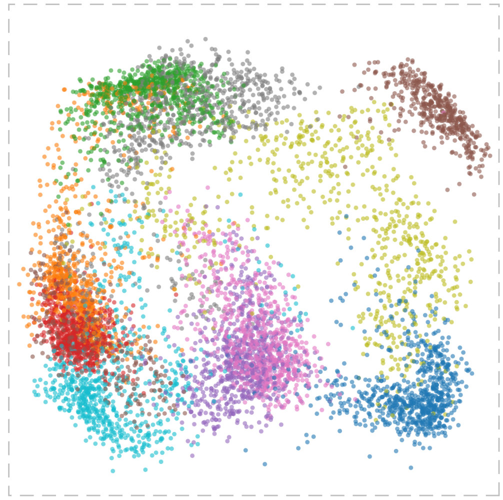


(f) t-SNE-scatterplot

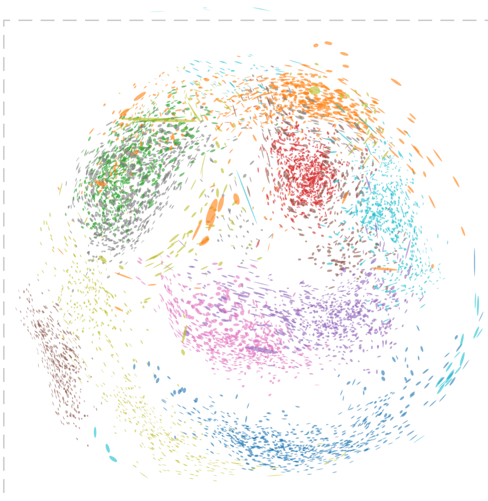
Fig. 17: Visualizations of Congressional Voting Records data projected with (a, b) PCA, (c, d) MDS and (e, f) t-SNE.



(a) PCA-ours

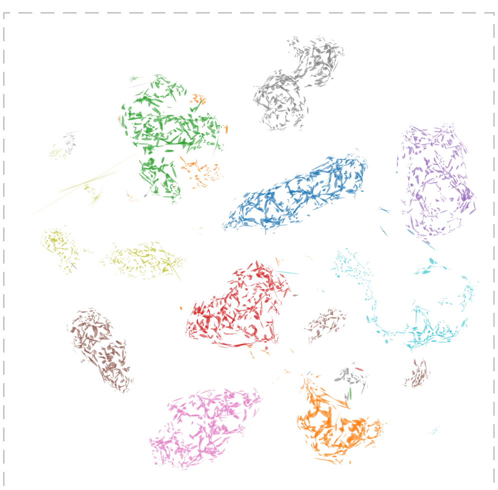


(b) PCA-scatterplot

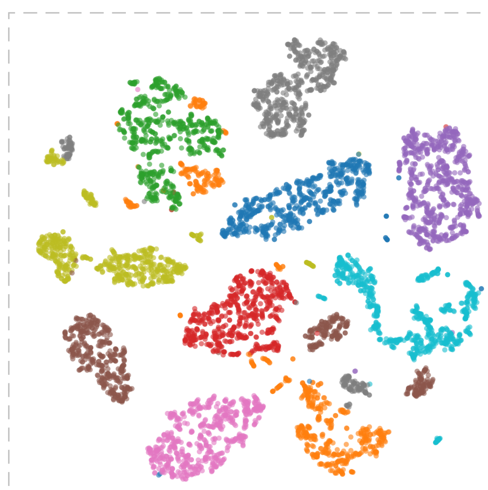


(c) MDS-ours

(d) MDS-scatterplot



(e) t-SNE-ours



(f) t-SNE-scatterplot

Fig. 18: Visualizations of pendigits data (7494 data points of 16-D) projected with (a, b) PCA, (c, d) MDS and (e, f) t-SNE.

References

[1] I. Borg and P. J. Groenen. *Modern Multidimensional Scaling: Theory and Applications*. Springer, 2005. doi: 10.1111/j.1745-3984.2003.tb01108.x

[2] D. Dua and C. Graff. UCI machine learning repository, 2017.

[3] R. Faust, D. Glickenstein, and C. Scheidegger. DimReader: axis lines that explain non-linear projections. *IEEE Transactions on Visualization and Computer Graphics*, 25(1):481–490, 2019. doi: 10.1109/TVCG.2018.2865194

[4] Y. LeCun, C. Cortes, and C. J. Burges. The MNIST database of handwritten digits, 1998.

[5] L. Van der Maaten and G. Hinton. Visualizing data using t-SNE. *Journal of Machine Learning Research*, 9(2579-2605):85, 2008.

[6] H. Xiao, K. Rasul, and R. Vollgraf. Fashion-mnist: a novel image dataset for benchmarking machine learning algorithms, 2017.