FlowNet: A Deep Learning Framework for Clustering and Selection of Streamlines and Stream Surfaces

Jun Han, Jun Tao, *Member, IEEE*,

and Chaoli Wang, *Senior Member, IEEE*

Motivation

- Applications that study the aero- and hydro-dynamical systems generate large amount of vector field data that need to be analyzed and visualized
- Most fluids (air, water, etc.) are transparent have invisible flow patterns
- Flow visualization is needed to make the flow patterns visible for *qualitative* and *quantitative* analysis
- This paper focus on integration-based flow visualization

- Challenges:
    - generating representative flow lines or surfaces
    - visually exploring a large collection of flow lines or surfaces

    - seeding and selection of streamlines: make use of handcrafted features (e.g., entropy, curvature, torsion, saliency, critical points, separation lines, vortex cores)

Contribution

- a single deep learning approach for streamline and stream surface clustering, filtering, and selection
  - autoencoder that automatically learns line or surface feature descriptors

- develop a visual interface to enable users to effectively explore the underlying set of streamlines and stream surfaces

Method

- Goal: identify a subset that best captures the underlying flow features and patterns given a large set of streamlines or stream surfaces
    - partition the input set into clusters
    - select one from each cluster to form the representatives

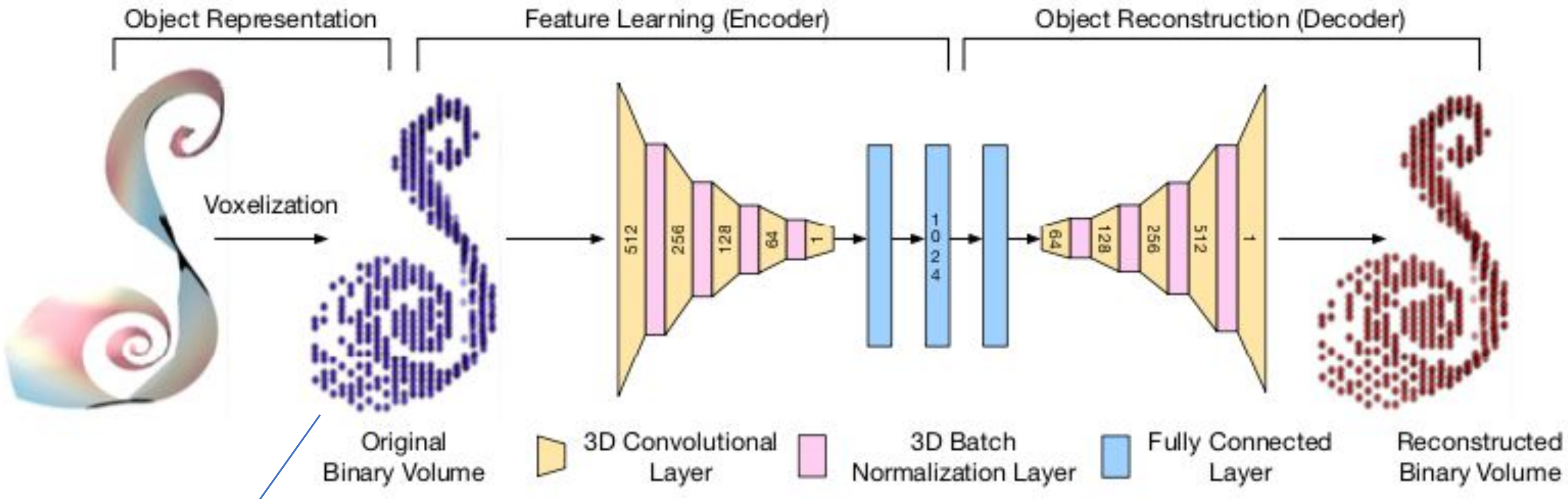
- How to learn the feature descriptor for a line or surface?
- Solution: *autoencoder*
    - voxelize and downsample each *object* (line or surface) into a 3D binary volume (input of autoencoder)
    - autoencoder learns feature descriptors automatically
    - t-SNE for dimensionality reduction
    - interactive clustering using DBSCAN to identify the representatives

Method

- **Object voxelization:** transfer sequence representation into voxel representation
  - Streamlines: (*sequence)* a 1D vector $s = \{x_1,y_1,z_1,\cdots ,x_n,y_n,z_n\}$
    - (x_i,y_i,z_i): a point on the object
    - n: #points
  - Stream surfaces: (*voxel) a* volume V with size L×W ×H
    - Stores sequences line by line
    - each seq: corresponding points following the streamline or timeline direction
  - Map each point on an object to its nearest voxel: $V[x_i, y_i, z_i] = 1$, (for i = 1 to n)
    - If the voxel $V[l_i , w_j , h_k]$ is occupied, then value of this voxel is 1 otherwise is 0
- **Downsample**: downsamping ratio $x_r = L / L'$,
  - $V[x_i, y_i, z_i]$ □ $V'[x_i',y_i',z_i']$
  - set $V'[x_i',y_i',z_i'] = 1$ where $x_i' = x_i/xr$, $y_i' = y_i/yr$, $z_i' = z_i/zr$

Method

- **FlowNet architecture**
  - *feature learning* (encoder) and *object reconstruction* (decoder)



(a)

  - AE Input: voxelized representations

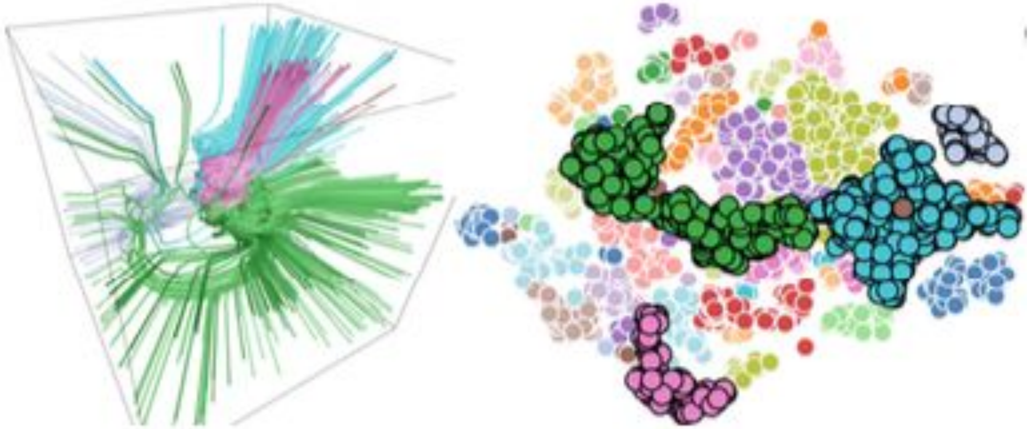
Method

- **Dimensionality Reduction (**t-SNE**)**
  - Map feature descriptors to a low-dimensional (2D) space for visual exploration
  - Input: distance matrix recording Euclidean distances between feature descriptors

- **Object Clustering:** DBSCAN

Interface

- two views: volume view and projection view
- **Clustering:** allows users to interactively tune the parameters of DBSCAN

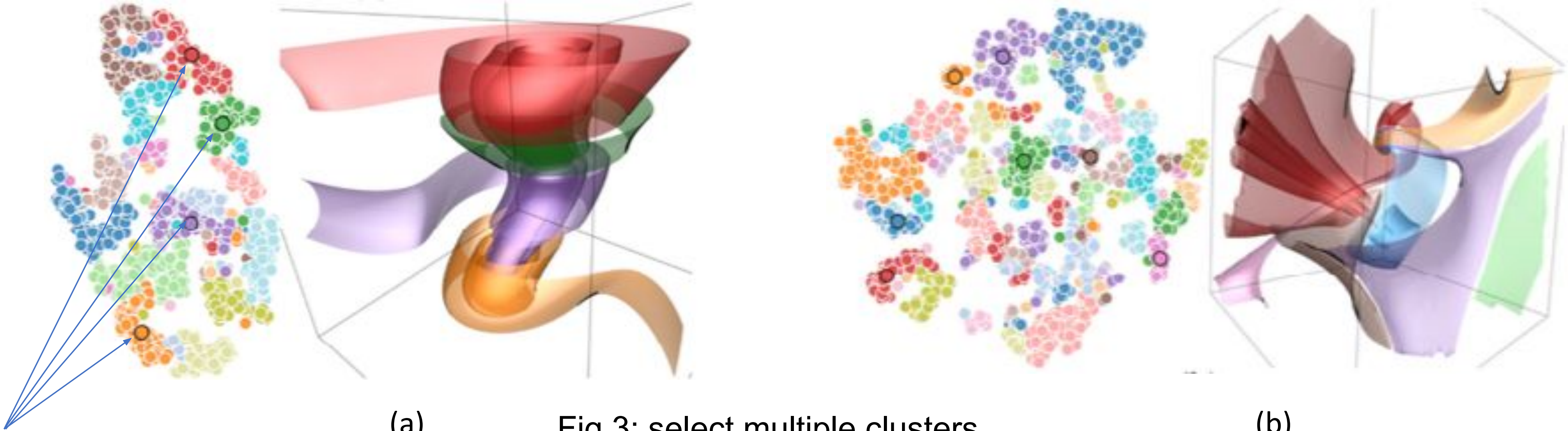


(a)

Fig 2: choose multiple clusters simultaneously

Interface

- two views: volume view and projection view
- **Clustering:** allows users to interactively tune the parameters of DBSCAN
- **Representatives:**
  - cluster’s center: sum of Euclidean distances from this point to all the other points in the same cluster is <u>minimum</u>



Clusters’ representatives

(a) (b)

Fig 3: select multiple clusters simultaneously

Interface

- two views: volume view and projection view
- **Clustering:** allows users to interactively tune the parameters of DBSCAN
- **Representatives:**
    - cluster’s center: sum of Euclidean distances from this point to all the other points in the same cluster is minimum
- **Neighborhood:**
    - the distance between the centers of two clusters as their inter-cluster distance

Interface

- two views: volume view and projection view
- **Clustering:** allows users to interactively tune the parameters of DBSCAN
- **Representatives:**
  - cluster's center: sum of Euclidean distances from this point to all the other points in the same cluster is minimum
- **Neighborhood:**

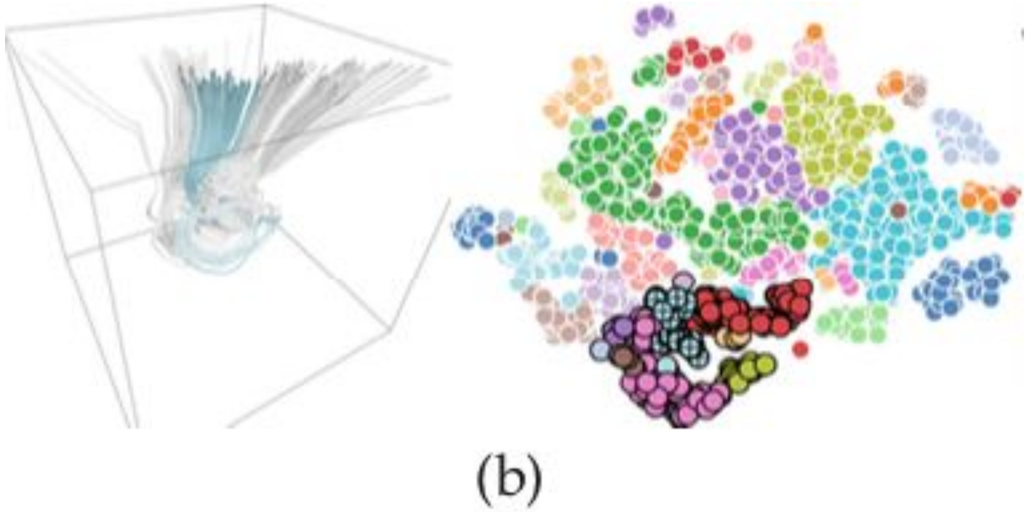


Fig 4: expanding from one selected cluster to its neighboring clusters

Validation

To justify the need for deriving feature descriptors from streamlines

- **Feature descriptor:** compare the results of using feature descriptors with binary volumes
  - project binary volumes directly <u>does not help to reveal useful potential structures</u>

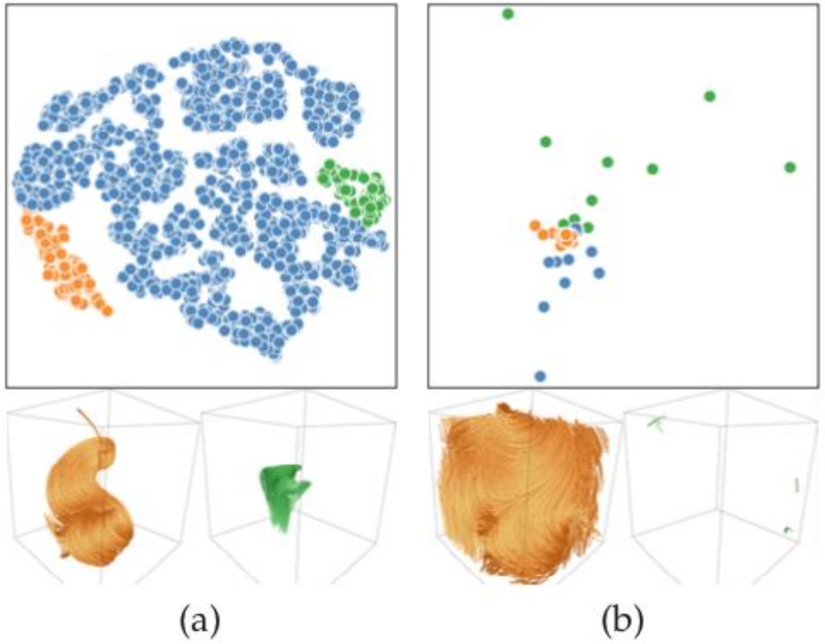


Fig 5: Comparison of t-SNE projections of (a) feature descriptors and (b) binary volumes.

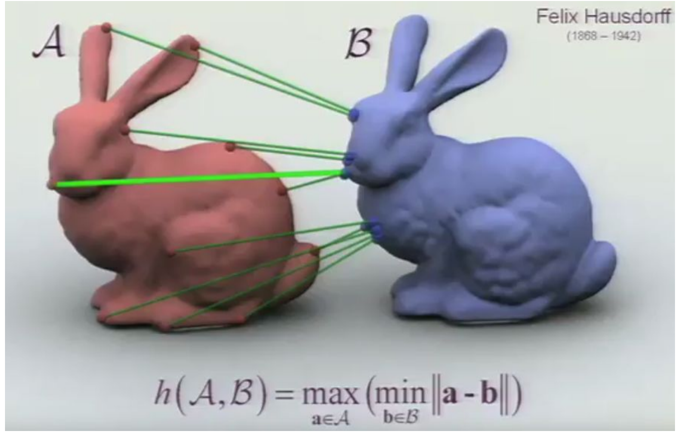
Comparison of t-SNE projections of (a) feature descriptors and (b) binary volumes.

Two point groups (orange and green) are selected

Conclusion: by extracting feature descriptors, FlowNet can preserve structure exhibited by the streamline.

Validation

- **Distance measure:** verify the effectiveness of using the Euclidean distance
  - compare against the mean of the closest point (MCP) distances and Hausdorff distance between streamlines
  - mean of the closest point (MCP)
  - Hausdorff distance



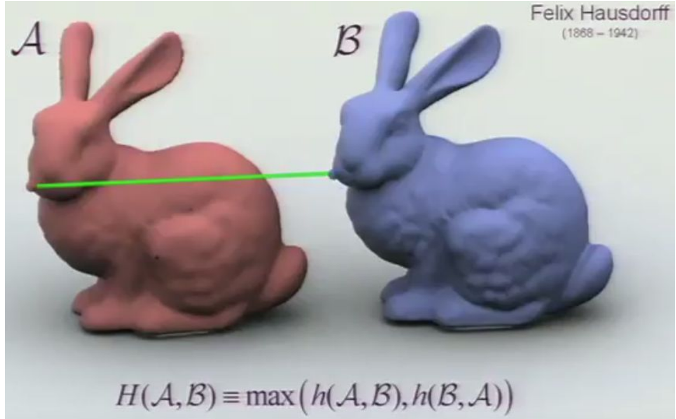


Image Credit: Min Tang , SIGGRAPH 2009 Presentation

Applications

Shape matching

Mesh simplification

Geometric modeling

Model rendering

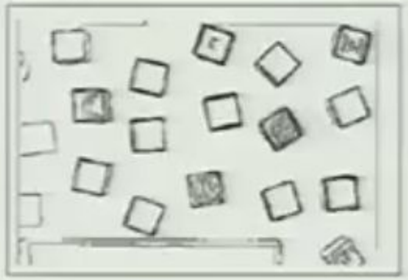


Image registration
and recognition

Face detection

Validation

- **Distance measure:** verify the effectiveness of using the Euclidean distance
    - compare against the mean of the closest point (MCP) distances and Hausdorff distance between streamlines

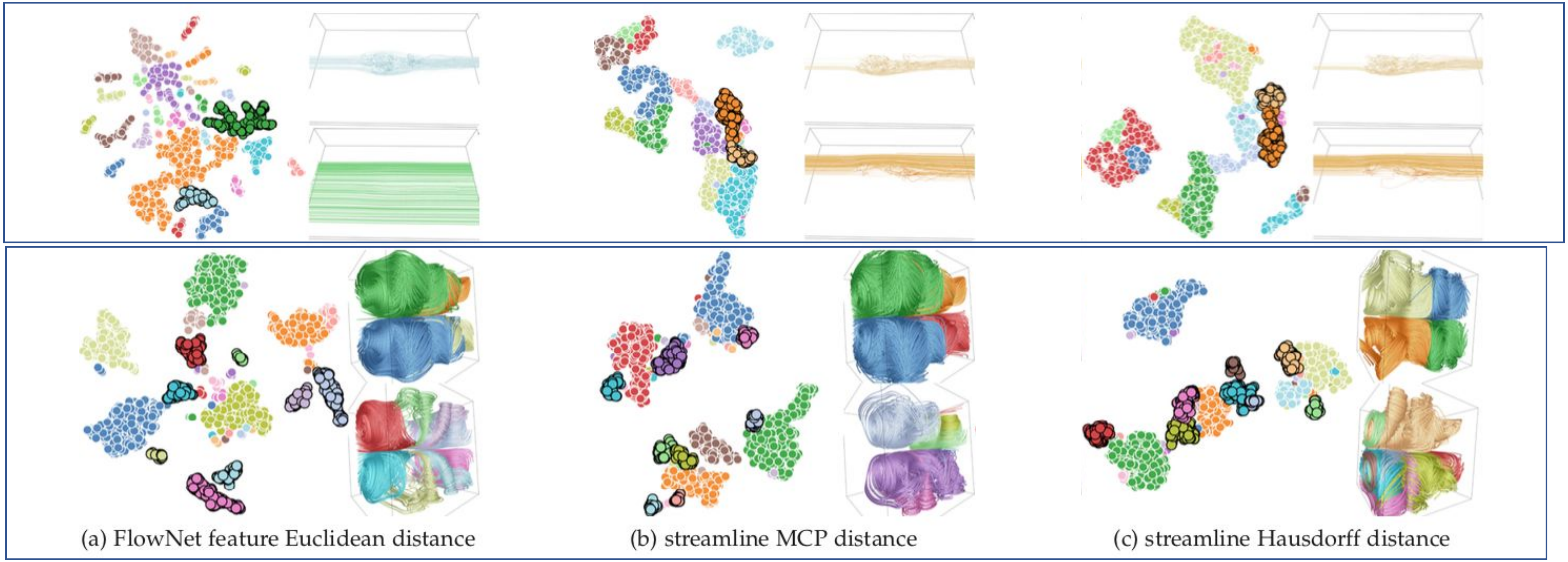


Fig 7: Comparison of different distance

Validation

$$F_1 = \frac{2}{\frac{1}{\text{recall}} + \frac{1}{\text{precision}}} = \frac{2}{\frac{\text{FNs}+\text{TPs}}{\text{TPs}} + \frac{\text{FPs}+\text{TPs}}{\text{TPs}}},$$

- **Underfitting and overfitting:** report F1 for training data and test data
  - TP: ground truth is 1, possibility predicted by FlowNet > 0.5
  - FP: ground truth is 0, possibility predicted by FlowNet > 0.5
  - FN: ground truth is 0, possibility predicted by FlowNet < 0.5

Data Set	Original Dimension	Downsampled Dimension	Kernel Size	Training # Lines	Training F_1 Score	Testing # Lines	Testing F_1 Score	Training # Surfaces	Training F_1 Score	Testing # Surfaces	Testing F_1 Score
ABC	$51 \times 51 \times 51$	$51 \times 51 \times 51$	$3 \times 3 \times 3$	3,000	0.91	3,000	0.82	2,000	0.84	2,000	0.71
Bénard flow	$128 \times 32 \times 64$	$64 \times 16 \times 32$	$4 \times 1 \times 2$	3,000	0.87	3,000	0.80	2,000	0.84	2,000	0.79
car flow	$368 \times 234 \times 60$	$92 \times 59 \times 15$	$6 \times 4 \times 1$	3,000	0.81	3,000	0.69				
computer room	$417 \times 345 \times 60$	$105 \times 87 \times 15$	$8 \times 6 \times 2$	3,000	0.74	3,000	0.68	2,000	0.83	2,000	0.59
crayfish	$322 \times 162 \times 119$	$81 \times 40 \times 30$	$4 \times 2 \times 2$	3,000	0.86	3,000	0.78				
five critical pts	$51 \times 51 \times 51$	$51 \times 51 \times 51$	$3 \times 3 \times 3$	3,000	0.96	3,000	0.72	2,000	0.72	2,000	0.57
solar plume	$126 \times 126 \times 512$	$32 \times 32 \times 128$	$2 \times 2 \times 8$	4,000	0.83	4,000	0.76	1,000	0.84	1,000	0.57
square cylinder	$192 \times 64 \times 48$	$96 \times 32 \times 24$	$8 \times 3 \times 2$	3,000	0.84	3,000	0.72	2,000	0.91	2,000	0.86
supernova	$100 \times 100 \times 100$	$50 \times 50 \times 50$	$2 \times 2 \times 2$	3,000	0.86	3,000	0.75				
tornado	$64 \times 64 \times 64$	$50 \times 50 \times 50$	$3 \times 3 \times 3$	3,000	0.91	3,000	0.76	2,000	0.88	2,000	0.78
two swirls	$64 \times 64 \times 64$	$32 \times 32 \times 32$	$4 \times 4 \times 4$	3,000	0.88	3,000	0.75	2,000	0.91	2,000	0.81

the training and test set sizes

respective F1 scores for streamlines / stream surfaces

Validation

- **Underfitting and overfitting:** report F1 for training data and test data
  - Visually demonstrate that FlowNet is not overfitting

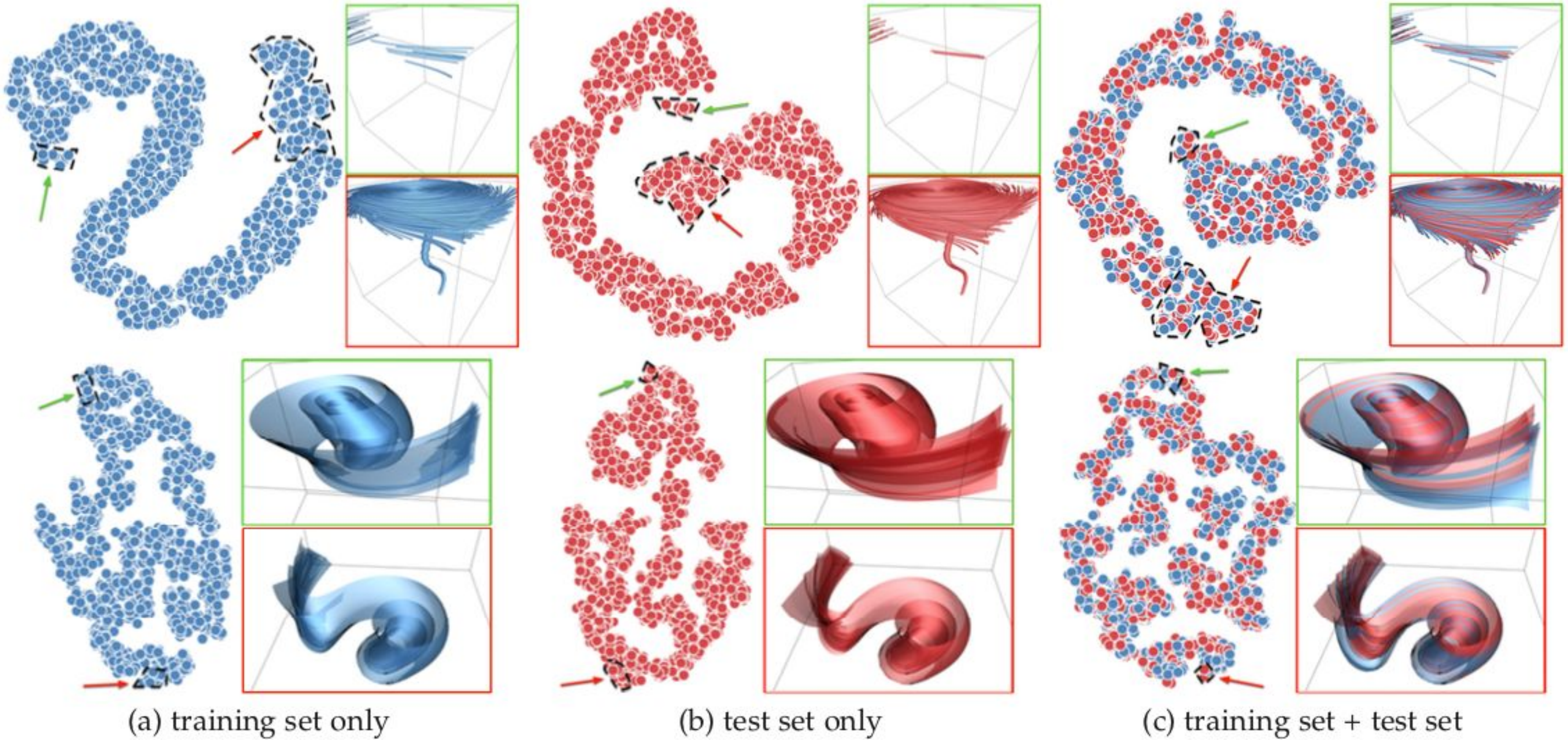


Fig 8: the tornado streamline and stream surface

Datasets

Data Set	Original Dimension	Downsampled Dimension	Kernel Size
ABC	$51 \times 51 \times 51$	$51 \times 51 \times 51$	$3 \times 3 \times 3$
Bénard flow	$128 \times 32 \times 64$	$64 \times 16 \times 32$	$4 \times 1 \times 2$
car flow	$368 \times 234 \times 60$	$92 \times 59 \times 15$	$6 \times 4 \times 1$
computer room	$417 \times 345 \times 60$	$105 \times 87 \times 15$	$8 \times 6 \times 2$
crayfish	$322 \times 162 \times 119$	$81 \times 40 \times 30$	$4 \times 2 \times 2$
five critical pts	$51 \times 51 \times 51$	$51 \times 51 \times 51$	$3 \times 3 \times 3$
solar plume	$126 \times 126 \times 512$	$32 \times 32 \times 128$	$2 \times 2 \times 8$
square cylinder	$192 \times 64 \times 48$	$96 \times 32 \times 24$	$8 \times 3 \times 2$
supernova	$100 \times 100 \times 100$	$50 \times 50 \times 50$	$2 \times 2 \times 2$
tornado	$64 \times 64 \times 64$	$50 \times 50 \times 50$	$3 \times 3 \times 3$
two swirls	$64 \times 64 \times 64$	$32 \times 32 \times 32$	$4 \times 4 \times 4$

Results

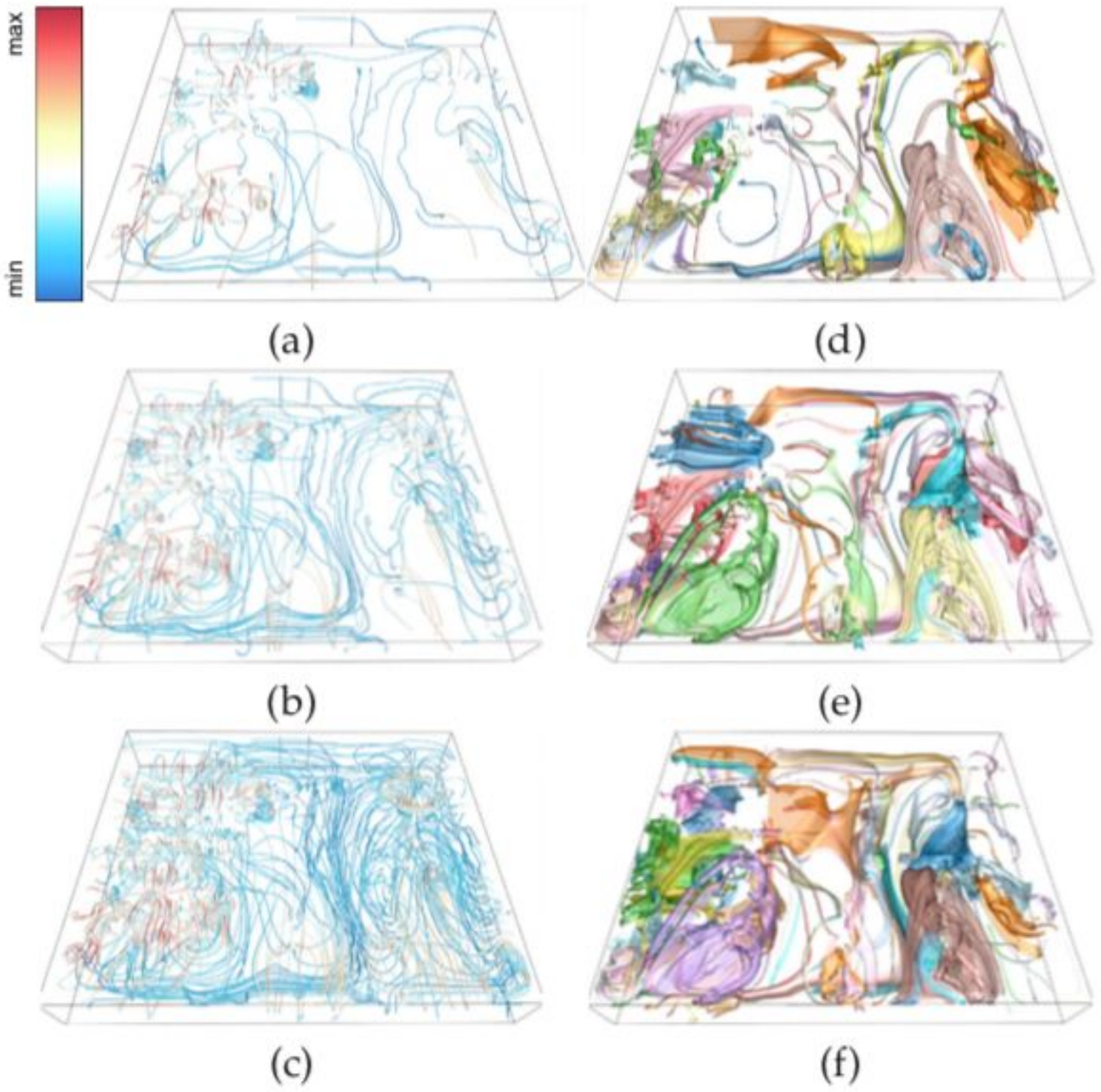


Fig 9: Representative streamlines and stream surfaces of the computer room data set.

Results

Comparison against Existing Methods: Stream surface selection

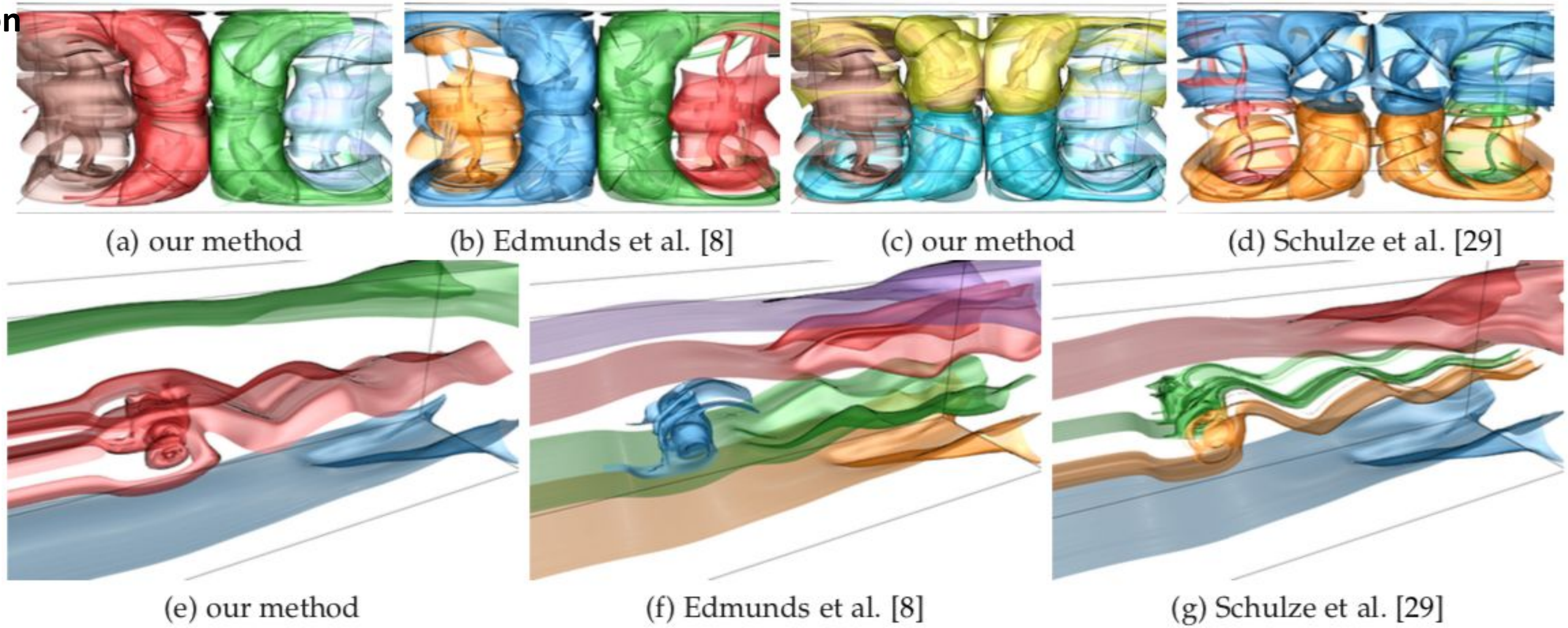


Fig 10: Top to bottom: comparison of surface selection results of the Be´ nard flow and square cylinder data sets using different methods.

Results

Comparison against Existing Methods:

Streamline selection

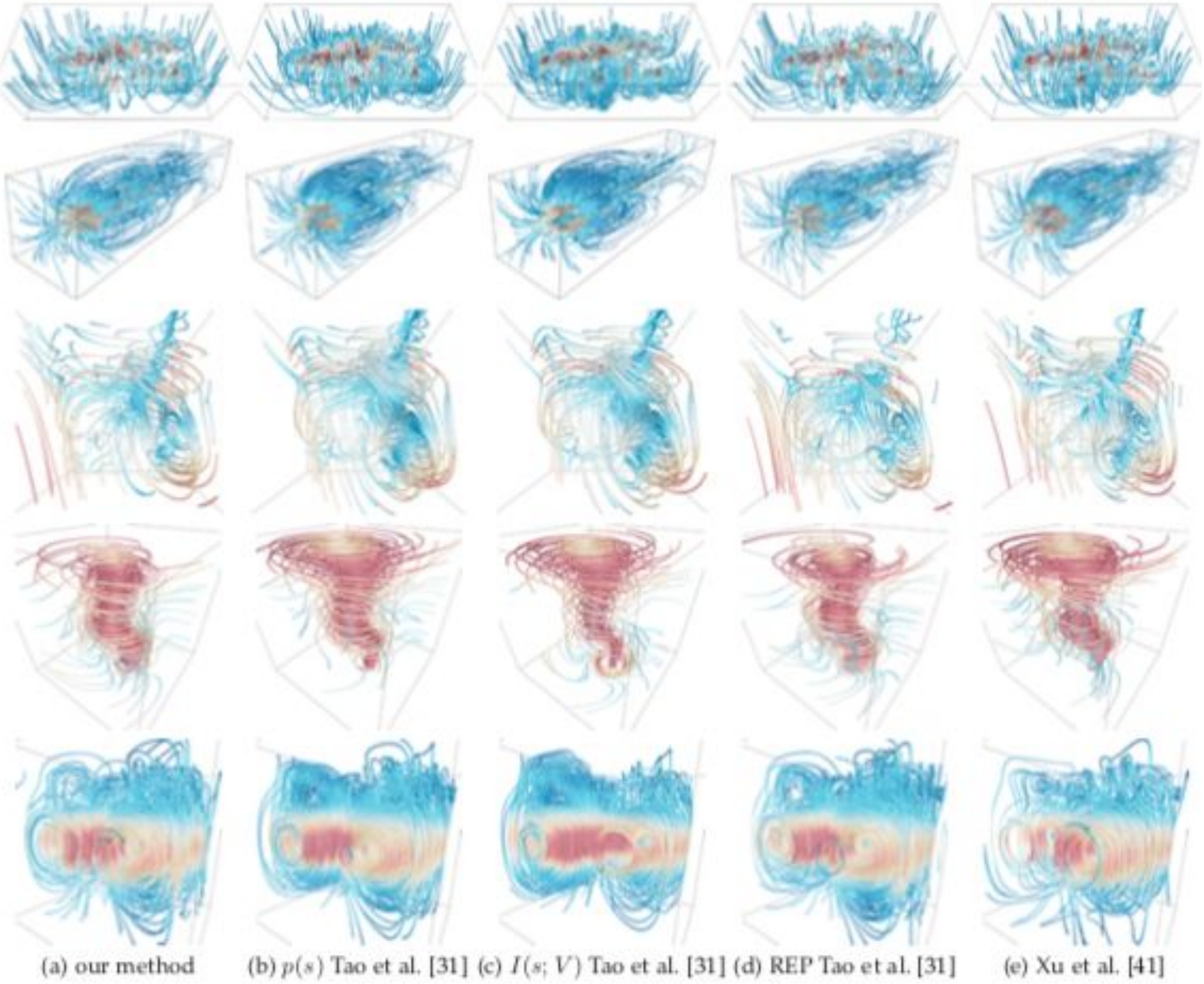


Fig 10: Top to bottom: comparison of surface selection results of the Be´ nard flow and square cylinder data sets using different methods.

Results

Comparison against Existing Methods:

Streamline selection

Data Set	# Lines	PSNR (db) Ours	$p(s)$	$I(s;V)$	REP	Xu's	AAD Ours	$p(s)$	$I(s;V)$	REP	Xu's
crayfish	70	**30.94**	30.84	30.91	30.02	28.97	**0.102**	0.105	0.103	0.116	0.144
solar plume	100	30.68	30.37	30.07	**30.75**	15.78	0.283	0.309	0.286	**0.280**	0.303
five critical pts	140	**26.25**	21.23	21.13	25.50	20.16	**0.023**	0.031	0.036	0.026	0.031
tornado	60	**29.74**	28.12	29.44	29.13	28.30	**0.080**	0.167	0.116	0.105	0.101
two swirls	80	**36.35**	36.30	34.55	36.21	27.72	**0.065**	0.066	0.079	0.070	0.071

Comparison of PSNR and AAD of reconstructed vector fields under different streamline selection methods. For each data set, the largest PSNR and the smallest AAD are highlighted in bold.

Conclusions and future work

- FlowNet, a novel approach for clustering and selection of streamlines and stream surfaces
  - learn latent features of streamlines and stream surfaces within a single framework in an unsupervised manner
  - The line and surface clusters generated from FlowNet capture both spatial proximity and/or geometric similarity.

- Future work
  - a more efficient way that maps these 3D volumes to 2D, e.g., spectrum space
  - a visual analytics interface that helps to make FlowNet explainable
    - identifying during training, what leads to clusters that can be explained with physical characteristics
  - design a deep neural net that learns the intricate relationships between the input and the output
    - (input: streamlines, output: stream surfaces)
    - (input: a seeding curve, output: the quality of the corresponding stream surface)