Appendix A
Proof of the inference margin

Proposition A.1. *The expected margin between the true log-likelihood* $\mathbb{E}_{p_{emp}(\mathbf{X})}\log p(\mathbf{X})$ *and* $\mathbb{E}_{p_{emp}(\mathbf{X})}ELBO$ *is*

$$\mathbb{E}_{p_{emp}(\mathbf{X})}[\log p(\mathbf{X})-ELBO] = \mathbb{E}_{p_{emp}(\mathbf{X})}[D_{\mathrm{KL}}(q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{X})\|p(\mathbf{z}|\mathbf{X}))$$

proof: We just need to prove the following equation

$$\log p(\mathbf{X}) - \mathrm{ELBO} = D_{\mathrm{KL}}(q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{X})\|p(\mathbf{z}|\mathbf{X}))$$

and the proof is:

$$\begin{aligned}\log p(\mathbf{X}) &= \int_{\mathbf{z}} q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{X})\log p(\mathbf{X})d\mathbf{z} \\ &= \int_{\mathbf{z}} q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{X})\log\frac{p(\mathbf{X},\mathbf{z})}{p(\mathbf{z}|\mathbf{X})}d\mathbf{z} \\ &= \mathbb{E}_{q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{X})}\log\frac{p(\mathbf{X},\mathbf{z})}{q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{X})} + \int_{\mathbf{z}} q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{X})\log\frac{q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{X})}{p(\mathbf{z}|\mathbf{X})} \\ &= \mathrm{ELBO} + D_{\mathrm{KL}}(q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{X})\|p(\mathbf{z}|\mathbf{X}))\end{aligned} \quad \square$$

Appendix B
Proof of the optimal transport scheme

Proposition B.1. *The shortest-path derived from the optimal transport scheme between* $\mathbf{z}_1 \sim \mathcal{N}(\boldsymbol{\mu}_1, \boldsymbol{\sigma}_1^2)$ *and* $\mathbf{z}_2 \sim \mathcal{N}(\boldsymbol{\mu}_2, \boldsymbol{\sigma}_2^2)$ *with* $\lambda \in [0,1]$ *is*

$$\begin{aligned}\tilde{\boldsymbol{\mu}} &= \lambda\boldsymbol{\mu}_1 + (1-\lambda)\boldsymbol{\mu}_2 \\ \tilde{\boldsymbol{\sigma}} &= \lambda\boldsymbol{\sigma}_1 + (1-\lambda)\boldsymbol{\sigma}_2\end{aligned}$$

proof: Utilizing the conclusions in [1], the closed form of optimal transport from one multi-normal distribution $\mathcal{N}(\mathbf{z}_1;\boldsymbol{\mu}_1,\boldsymbol{\Sigma}_1)$ to another normal distribution $\mathcal{N}(\mathbf{z}_2;\boldsymbol{\mu}_2,\boldsymbol{\Sigma}_2)$ is

$$\mathbf{z} \rightarrow \mathcal{T}(\mathbf{z}) = \boldsymbol{\mu}_2 + \mathbf{T}(\mathbf{z}-\boldsymbol{\mu}_1); \qquad \mathbf{T} = \boldsymbol{\Sigma}_1^{-\frac{1}{2}}(\boldsymbol{\Sigma}_1^{\frac{1}{2}}\boldsymbol{\Sigma}_2\boldsymbol{\Sigma}_1^{\frac{1}{2}})^{\frac{1}{2}}\boldsymbol{\Sigma}_1^{-\frac{1}{2}}$$

In our problem, the optimal transport scheme with λ is

$$\begin{aligned}\mathbf{z}_\lambda &= (1-\lambda)\mathbf{z}_1 + \lambda\mathcal{T}(\mathbf{z}_1) \\ \mathbf{T} &= diag(\boldsymbol{\sigma}_1/\boldsymbol{\sigma}_2)\end{aligned} \tag{1}$$

Utilize Equation 1, we can get

$$\begin{aligned}\mathbf{z}_\lambda &\sim \mathcal{N}(\tilde{\boldsymbol{\mu}}, diag(\tilde{\boldsymbol{\sigma}}^2)) \\ \tilde{\boldsymbol{\mu}} &= \lambda\boldsymbol{\mu}_1 + (1-\lambda)\boldsymbol{\mu}_2 \quad \square \\ \tilde{\boldsymbol{\sigma}} &= \lambda\boldsymbol{\sigma}_1 + (1-\lambda)\boldsymbol{\sigma}_2\end{aligned}$$

Appendix C
The Grad-FAM

Given an instance $\mathbf{X}$, its factors $\mathbf{z} = (z_1, \ldots, z_k)$ and a factor subset $\mathbf{z}_f = (z_{f_1}, \ldots, z_{f_p}), z_{f_i} \in \mathbf{z}$, the Grad-FAM is derived in two steps (Figure 1). First, we fix the factors in $\mathbf{z} - \mathbf{z}_f$ and compute the negative gradient of the feature maps $\mathbf{A}^q \in \mathbb{R}^{h*w}, q = 1, \ldots, C$ in the final convolutional layer. Therefore, we obtain the coarse-grained activation map $\mathbf{M}_{\mathrm{c}}^{\mathbf{z}_f}$ as:

$$\mathbf{M}_{\mathrm{coarse}}^{\mathbf{z}_f} = \mathrm{ReLU}(\sum_{q=1}^{c}\mathbf{A}^q \odot [\frac{\partial\mathbb{E}_{q_{\boldsymbol{\phi}}(\mathbf{z}_f|\mathbf{X})}\log p_{\boldsymbol{\theta}}(\mathbf{X}|\mathbf{z}_f)}{\partial\mathbf{A}^q} - \frac{\partial D_{\mathrm{KL}}(q_{\boldsymbol{\phi}}(\mathbf{z}_f|\mathbf{X})\|p(\mathbf{z}_f))}{\partial\mathbf{A}^q}]) \tag{2}$$

where $\mathrm{ReLU}(x) = \begin{cases}0 & \text{if } x \leq 0,\\ x & \text{else}\end{cases}$ is the rectified function. Combined with $\mathbf{M}_{\mathrm{c}}^{\mathbf{z}_f}$, we derive the fine-grained activation map $\mathbf{M}_{\mathrm{f}}^{\mathbf{z}_f}$ as

$$\mathbf{M}_{\mathrm{fine}}^{\mathbf{z}_f} = \mathrm{Interpolation}(\mathbf{M}_{\mathrm{coarse}}^{\mathbf{z}_f}) \odot \mathrm{ReLU}(\frac{\partial\mathbb{E}_{q_{\boldsymbol{\phi}}(\mathbf{z}_f|\mathbf{X})}\log p_{\boldsymbol{\theta}}(\mathbf{X}|\mathbf{z}_f) - D_{\mathrm{KL}}(q_{\boldsymbol{\phi}}(\mathbf{z}_f|\mathbf{X})\|p(\mathbf{z}_f))}{\partial\mathbf{X}}) \tag{3}$$

where the 'Interpolation' function matches the size of $\mathbf{M}_{\mathrm{c}}^{\mathbf{z}_f}$ to $\mathbf{X}$.

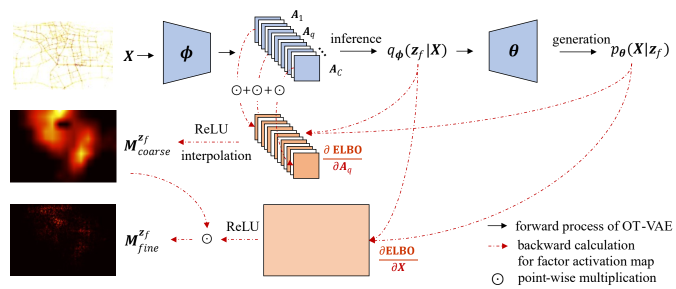


Fig. 1. The construction of Grad-FAM.

Appendix D
Implementation Details of OT-VAE

OT-VAE is trained in two steps. First, we train the basic VAE model until its ELBO threshold is reached. Second, we utilize the optimal transport estimation to obtain a tighter ELBO. Details of the training process are explained in Algorithm 1. During the training, we use the convolutional neural network (CNN) as the backbone model. We use SGD and adopt the reparameterization trick in [2], [3]. Moreover, we introduce $w_{M_{\mathbf{z}}}^{(t)}$ to balance the ELBO loss and the optimal transport estimation process. It is gradually increased from 0 to 1 with the exponential scheduler

$$w_{M_{\mathbf{z}}}^{(t)} = \exp(-5*(1-t/t_{\max})^2). \tag{4}$$

Algorithm 1 The inference and training process of OT-VAE in epoch t.

Input:
A batch of data $\mathbf{X}$ sampled from $p_{emp}(\mathbf{X})$;
Optimal transport estimation weights: $w_{M_{\mathbf{z}}}^{(t)}$;
Model parameters: $\boldsymbol{\theta}^{(t-1)}, \boldsymbol{\phi}^{(t-1)}$;
Model optimizer: SGD

Output:
Factors $q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{X}) = \mathcal{N}(\mathbf{z};\boldsymbol{\mu}(\mathbf{X};\boldsymbol{\phi}^{(t-1)}), \boldsymbol{\sigma}^2(\mathbf{X};\boldsymbol{\phi}^{(t-1)}))$;
Inferred data $p_{\boldsymbol{\theta}}(\mathbf{X}|\mathbf{z}) = \mathcal{N}(\mathbf{X}; f(\mathbf{z};\boldsymbol{\theta}^{(t-1)}), \boldsymbol{\sigma}^2)$;
Updated parameters: $\boldsymbol{\theta}^{(t)}, \boldsymbol{\phi}^{(t)}$

$q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{X}), p_{\boldsymbol{\theta}}(\mathbf{X}|\mathbf{z}) = \mathrm{VAE}(\mathbf{X};\boldsymbol{\theta}^{(t-1)},\boldsymbol{\phi}^{(t-1)})$
$L_{\mathrm{ELBO}} = \mathbb{E}_{q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{X})}[-\log p_{\boldsymbol{\theta}}(\mathbf{X}|\mathbf{z})] + D_{\mathrm{KL}}(q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{X})\|p(\mathbf{z}))$
$L_{M_{\mathbf{z}}} = \mathrm{OptimalTransportEstimation}(\mathbf{X}, q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{X}))$
$L = L_{\mathrm{ELBO}} + w_{M_{\mathbf{z}}}^{(t)} * L_{M_{\mathbf{z}}}$
$\boldsymbol{\theta}^{(t)}, \boldsymbol{\phi}^{(t)} = \mathrm{SGD}(\boldsymbol{\theta}^{(t-1)}, \boldsymbol{\phi}^{(t-1)}, \frac{\partial L}{\partial\boldsymbol{\theta}}, \frac{\partial L}{\partial\boldsymbol{\phi}})$
return $q_{\boldsymbol{\phi}}(\mathbf{z}|\mathbf{X}), p_{\boldsymbol{\theta}}(\mathbf{X}|\mathbf{z}), \boldsymbol{\theta}^{(t)}, \boldsymbol{\phi}^{(t)}$

APPENDIX E
MORE EXAMPLES OF THE FACTOR OPERATIONS

We provide more examples of the clustering operations and the factor operations to demonstrate that these operations can support effective pattern identification with the help of visualization. We choose the experimental settings of the OT-VAE model that reach the best ELBO to generate the results.

Clustering. The optimal transport based clustering provides the cluster label as well as the centroid for each cluster. Our OT-VAE model use the centroid as the barycenter of the optimal transport scheme in the factor space. Figure 2 shows that the derived centroid is representative of its corresponding cluster.

Factor Operations. We provide further examples of the Linear interpolation & transformation operations. For each dataset, we choose two groups of charts. Each group contains four charts. Charts in the first group are relatively similar to each other and presents the smooth transition, while those in the second group is distinct and presents the sharp transition (Figure 3).

APPENDIX F
GENERATION PERFORMANCE OF THE OT-VAE MODEL

Figure 4 compares the generation performance of OT—VAE with the VAE and Ladder-VAE model. It shows that OT-VAE achieves better generation performance with significantly small residual values compared with other baseline models

APPENDIX G
EXPERT INTERVIEW QUESTIONS

The following questions is asked at the end of each interview.

Workflow:

1) Does the ChartNavigator framework achieve effective and efficient pattern identification and annotation in visualization charts?
2) How can the OT-VAE model help in this process?
3) How can the ChartNavigator framework help with your daily work?

Visualization:

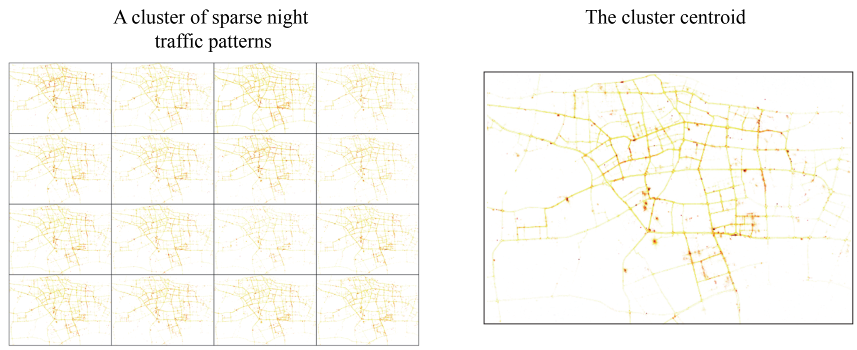
4) Does the ChartNavigator interface well-present the core information?
5) Do there exist confusing designs that need to be improved?
6) Is it intuitive to use the factor operations to define a pattern?
7) Is the pattern identification strategy useful in understanding the semantics of factors and identifying patterns?
8) Are there any findings with the ChartNavigator interface that can hardly be discovered with tools that identify patterns from the raw data?

REFERENCES

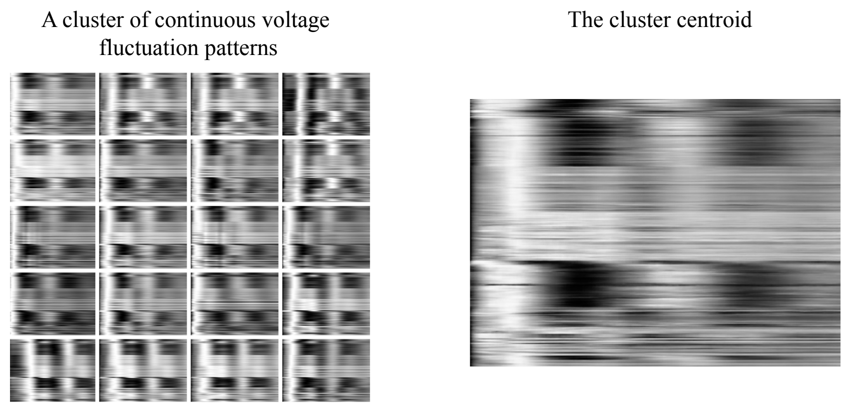
[1] Kuang, Max, and Esteban G. Tabak. ”Preconditioning of optimal transport.” SIAM Journal on Scientific Computing 39.4 (2017): A1793-A1810.

[2] Rezende, Danilo Jimenez, Shakir Mohamed, and Daan Wierstra. ”Stochastic backpropagation and approximate inference in deep generative models.” arXiv preprint arXiv:1401.4082 (2014).

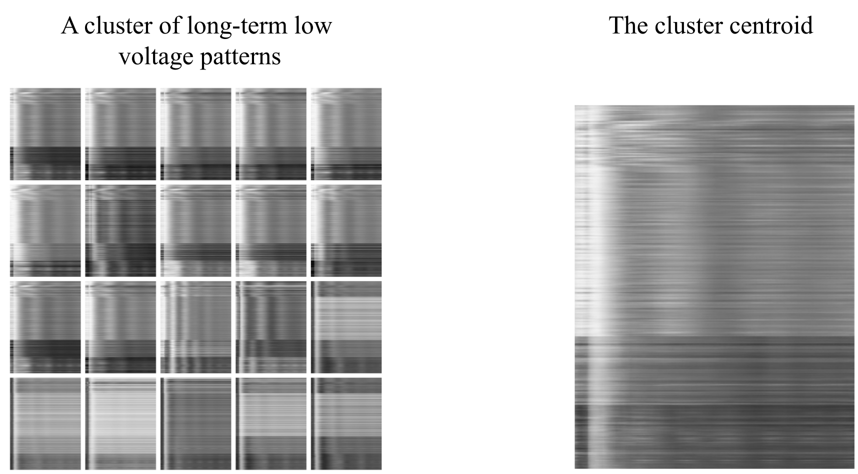
[3] Jang, Eric, Shixiang Gu, and Ben Poole. ”Categorical reparameterization with gumbel-softmax.” arXiv preprint arXiv:1611.01144 (2016).



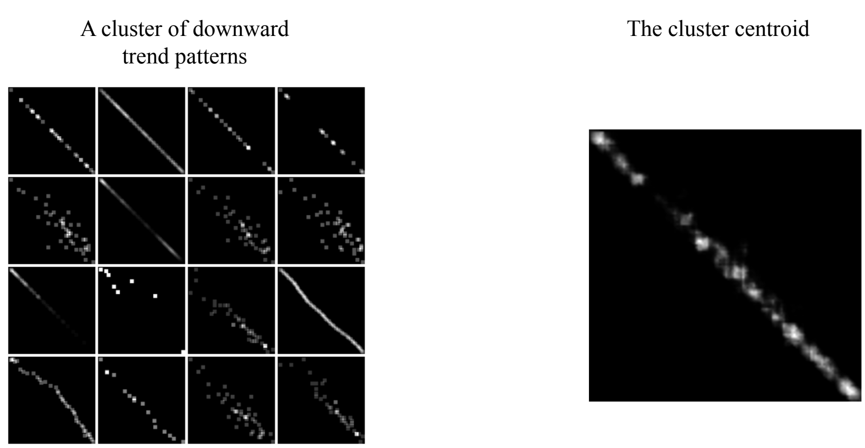
(a) Dataset:TR



(b) Dataset:PG-S

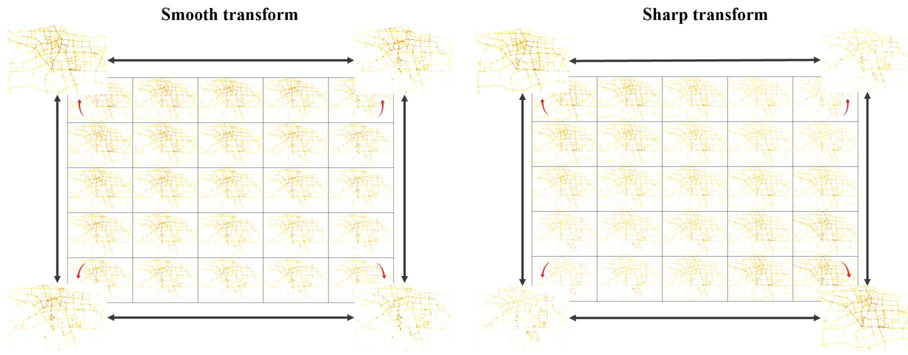


(c) Dataset:PG-L

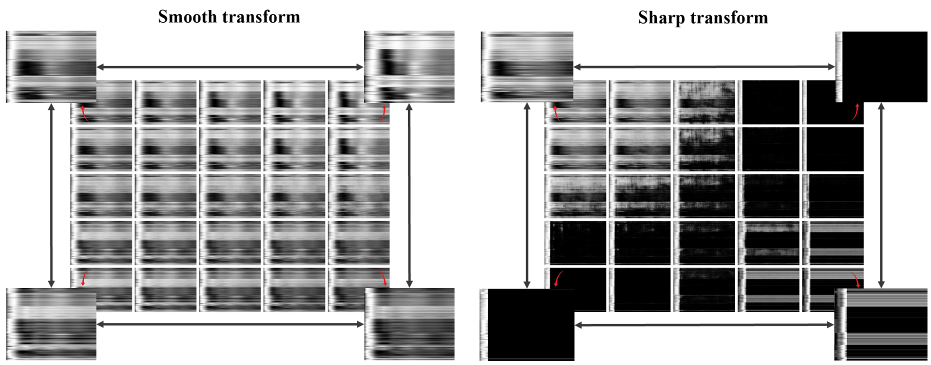


(d) Dataset:SP

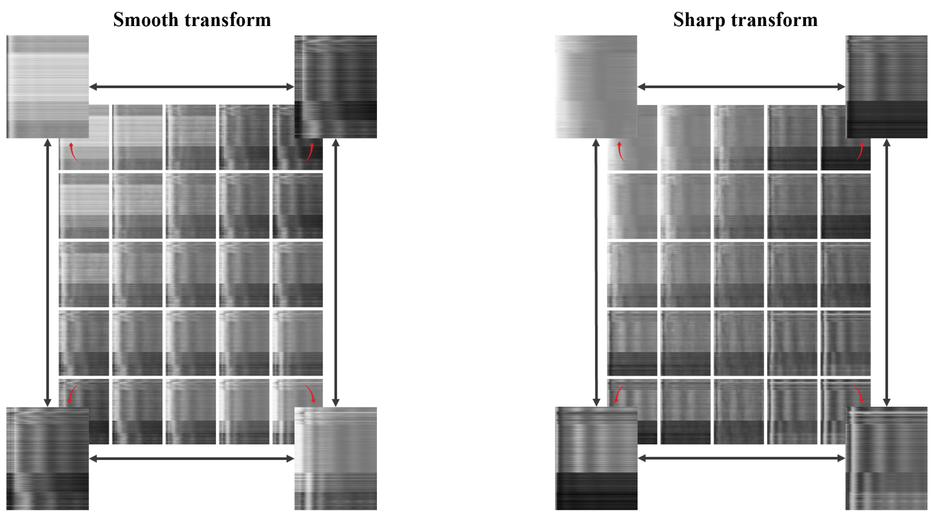
Fig. 2. The clustering performance of OT-VAE on four benchmark datasets. Left: the data instances belong to one cluster. Right: the corresponding cluster centroid.



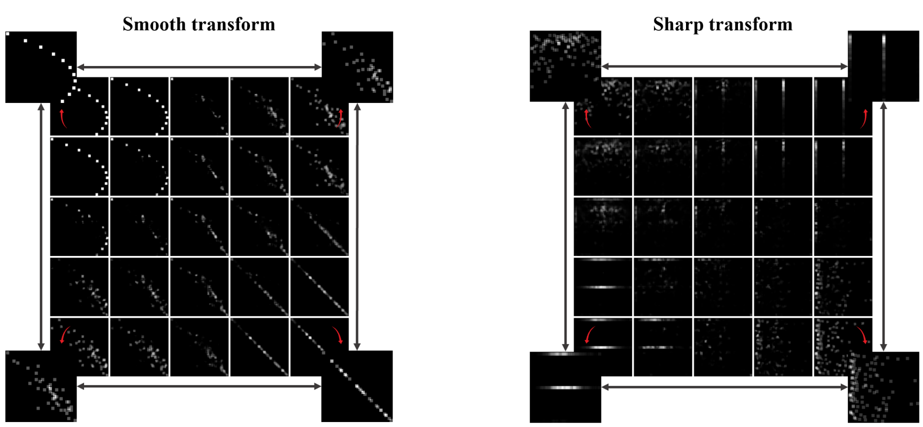
(a) Dataset:TR



(b) Dataset:PG-S

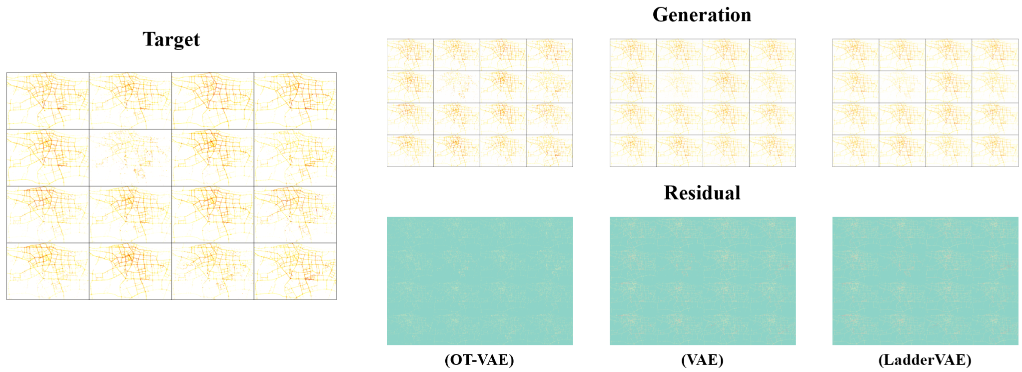


(c) Dataset:PG-L

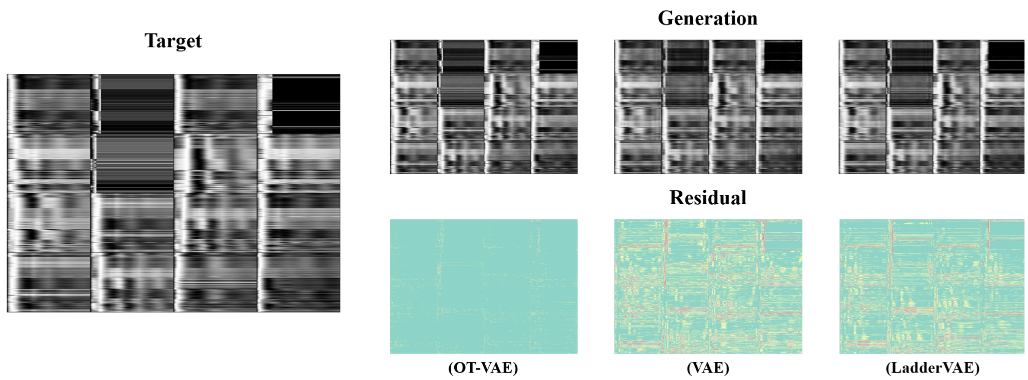


(d) Dataset:SP

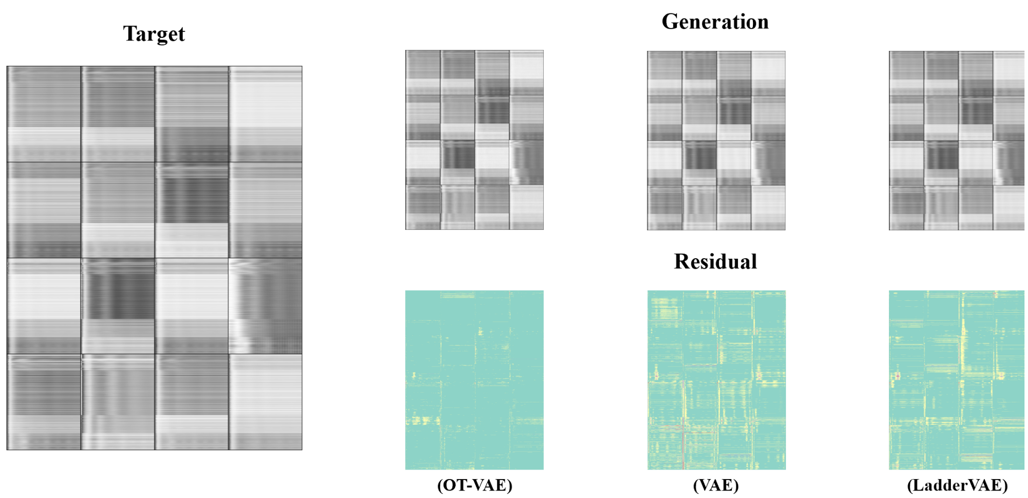
Fig. 3. The #-diagram of OT-VAE on four benchmark datasets with the interpolation and transformation algorithm. Left: the transform process for smooth changes. Right: the transform process for sharp changes.



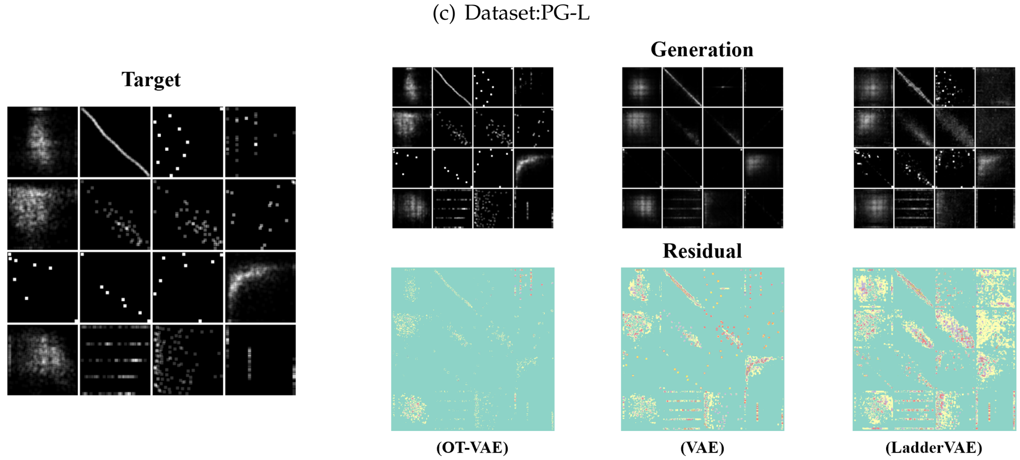
(a) Dataset:TR



(b) Dataset:PG-S



(c) Dataset:PG-L



(d) Dataset:SP

Fig. 4. The generation performance of OT—VAE on three benchmark datasets compared with two baseline models: VAE and LadderVAE. Target: the target instances $\mathbf{X}$ we want to reconstruct with the inferred factors $q_{\phi}(\mathbf{z}|\mathbf{X})$. Generation: the generation results $\tilde{\mathbf{X}}$ with $p_{\theta}(\mathbf{X}|\mathbf{z})$. Residual: the point-wise error between $\mathbf{X}$ and $\tilde{\mathbf{X}}$ as $|\mathbf{X} - \tilde{\mathbf{X}}|$.