Unsupervised Deep Embedded Fusion Representation of Single-Cell Transcriptomics

Yue Cheng[1*], **Yanchi Su**[1*], **Zhuohan Yu**[1], **Yanchun Liang**[2], **Ka-Chun Wong**[3], **Xiangtao Li**[1†]

[1]School of Artificial Intelligence, Jilin University, Jilin, China
[2]Zhuhai Laboratory of Key Laboratory of Symbol Computation and Knowledge Engineering of Ministry of Education, Zhuhai College of Science and Technology, Zhuhai 519041, China
[3]Department of Computer Science, City University of Hong Kong, Hong Kong SAR
chengyue22@mails.jlu.edu.cn, suyanchi@gmail.com, zhuohan20@mails.jlu.edu.cn
ycliang@jlu.edu.cn, kc.w@cityu.edu.hk, lixt314@jlu.edu.cn

Abstract

Cell clustering is a critical step in analyzing single-cell RNA sequencing (scRNA-seq) data that allows characterization of the cellular heterogeneity after transcriptional profiling at the single-cell level. Single-cell deep embedded representation models have gained popularity recently as they can learn feature representation and clustering simultaneously. However, the models still pose a variety of significant challenges, including the massive amount of data, pervasive dropout events, and complicated noise patterns in transcriptional profiling. Here, we propose a **S**ingle-**C**ell **D**eep **E**mbedding **F**usion **R**epresentation (scDEFR) model that produces a deep embedded fusion representation to learn the fused heterogeneous latent embedding containing both the gene-level transcriptome and cell topology information. We first fuse them layer by layer to obtain compressed representations of the intercellular relationships and transcriptome information. Then, we use the zero-inflated negative binomial model (ZINB)-based decoder to capture the global probabilistic structure of the data to reconstruct the final gene expression information. Finally, by simultaneously integrating the clustering loss, cross-entropy loss, ZINB loss, and cell graph reconstruction loss, scDEFR can optimize clustering performance and learn the latent representation from the fused information in a joint mutual supervised strategy. We conducted comprehensive experiments on 15 single-cell RNA-seq datasets from different sequencing platforms and demonstrated the superiority of scDEFR over a variety of state-of-the-art methods.

Introduction

Single cell RNA sequencing (scRNA-seq) technology allows analyzing the entire transcriptomes of very large numbers of individual cells, which is essential for characterizing the stochastic heterogeneity within cell populations and establishing distinct developmental trajectories, for example, immune cells (Papalexi and Satija 2018). Cell clustering is one of the most crucial tasks in the standard scRNA-seq pipeline that can group cells according to their gene expression patterns and thereby establish the intrinsic molecular profiles (Svensson et al. 2017). In previous studies, traditional clustering methods such as K-means (MacQueen 1967) and hierarchical clustering (Johnson 1967) were employed for unbiased cell type identification. However, due to the high heterogeneity of genome coverage and technical limitations, scRNA-seq data is very sparse, with 95% of measurements being zero (Grün, Kester, and Van Oudenaarden 2014). These problems prevent traditional clustering methods that rely on similarity metrics from being adequate. It is necessary to explore new computational approaches to learn the diverse characteristics in order to better uncover the specific patterns in scRNA-seq data. Deep embedded clustering algorithms tailored to scRNA-seq data have been developed to identify cell types. These methods usually utilize an autoencoder to learn the latent representation of the gene expression matrix and to optimize cluster assignment simultaneously; for instance, scDCC (Tian et al. 2021), scDeepCluster (Tian et al. 2019), scziDesk (Chen et al. 2020a). However, these methods only focus on learning the gene expression matrix itself and ignore the topology of the cells. Moreover, the similarity between cells is the crucial point for guiding clustering, and due to the high sparsity of scRNA-seq as mentioned above, higher order cellular structural information could also be investigated to reveal potential similarities in common neighboring cells. Recently, deep graph embedding clustering algorithms have been advanced to address the clustering problem; for instance, scTAG (Yu et al. 2022), scGNN (Wang et al. 2021), scGAE (Luo et al. 2021) and GraphSCC (Zeng et al. 2021), which usually use a graph

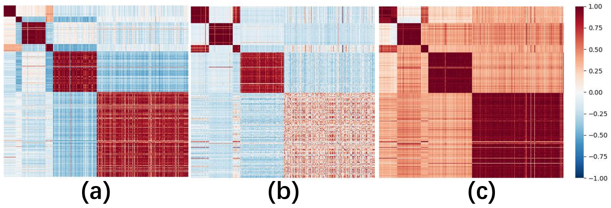

Figure 1: Heatmaps of cell similarity matrices in the latent space of (a) scDEFR, (b) scDEFR with only fusion cell topology encoder and, (c) scDEFR with only transcriptomics profile-based graph encoder

[*]These authors contributed equally.

[†]Corresponding author.

autoencoder to capture the cell structural information in a graph and then learn the latent representations for clustering. These methods are often prone to missing key patterns in the gene expression data, leading to the collapse of depth map clustering methods. It is natural, therefore, to couple the two methods to learn the latent representation from the gene expression matrix and cell graph simultaneously. More recently, a deep fusion clustering network (DFCN) was designed that integrates autoencoder and graph convolutional networks into a unified framework to process finely the attributes and structural information extracted from the autoencoder (AE) and graph autoencoder (GAE) (Tu et al. 2021). However, this research is still in its infancy stage and such a fusion framework has not yet been exploited for single-cell RNA-seq data analysis.

Motivated by the above observations, we propose scDEFR, a single-cell deep embedding fusion representation model, as depicted in Fig. 2. To effectively fuse the feature representations of the transcriptomic information and cell-cell topology information, scDEFR incorporates a fused cell topology encoder and a transcriptomics-based graph encoder via an interlayer fusion operation. Then, scDEFR uses the heterogeneous structural fusion mechanism to obtain the fused heterogeneous latent embeddings of the different encoders and a ZINB-based multimodal fusion decoder for subsequent clustering. Finally, we adopt a mutual supervised strategy that combines four kinds of training loss, including the clustering loss, cross-entropy loss, ZINB loss, and cell graph reconstruction loss. Thus, scDEFR can uniformly optimize the clustering process as well as the fused heterogeneous latent embedding to avoid representation collapse to accurately identify cell clusters and thereby improve clustering performance. Fig. 1 illustrates the phenomenon of representation collapse on the Qs_Limb_Muscle dataset with six cell clusters. We observe from the figure that the clustering performance is limited and some smaller-scale cell clusters are not easily distinguishable from other cell clusters.

The main contributions of our work are summarized below:

- We propose a single-cell deep embedding fusion representation model called scDEFR, combining a cell topology encoder with a transcriptomics profile-based graph encoder to capture the fusion-compressed representation of scRNA-seq data.
- scDEFR adopts a TAGCN to extract the topological information from the data and then augments fusion across each of the layers of the encoders during the encoding. In addition, scDEFR utilises a ZINB-based multimodal fusion decoder to capture the global probabilistic structure of the data, modelling the highly sparse and over-dispersed scRNA-seq data.
- To the best of our knowledge, this is the first architecture proposing fusion of heterogeneous structural information to address single-cell transcriptomics analysis.
- We compare our method with competitive state-of-the-art methods on 15 real scRNA-seq datasets. The results demonstrate that scDEFR outperforms all of the other baseline methods.

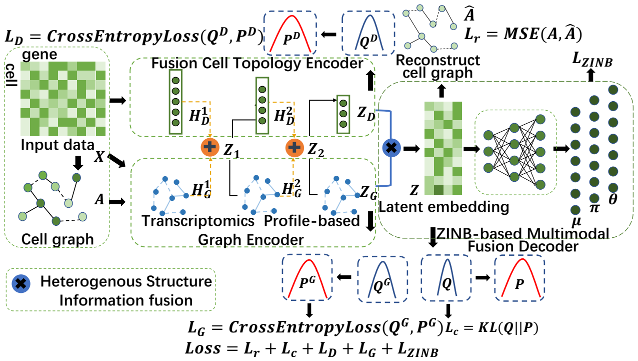
$L_D = CrossEntropyLoss(Q^D, P^D)$
P^D
Q^D
$\hat{A}$
$L_r = MSE(A, \hat{A})$
Reconstruct cell graph
L_{ZINB}
gene
cell
Input data
Fusion Cell Topology Encoder
H_D^1
H_D^2
Z_D
Z_1
Z_2
X
A
Cell graph
H_G^1
H_G^2
Z_G
Transcriptomics Profile-based Graph Encoder
Z
Latent embedding
μ
π
θ
ZINB-based Multimodal Fusion Decoder
Heterogenous Structure Information fusion
P^G
Q^G
Q
P
$L_G = CrossEntropyLoss(Q^G, P^G)$
$L_c = KL(Q||P)$
$Loss = L_r + L_c + L_D + L_G + L_{ZINB}$

Figure 2: The model architecture of scDEFR. scDEFR integrates a deep graph autoencoder into a ZINB-based deep autoencoder to learn the fusion latent representation, which contains the gene information and cell-cell topology representation, and adopts cross-entropy loss and Kullback–Leibler (KL) divergence to optimize the clustering performance.

Related Work

In this section, we present two related works: deep embedded clustering and graph embedded clustering in the single-cell RNA sequence field.

In single-cell RNA sequence analysis, deep clustering methods aim to learn the rich representations of gene expression matrices by optimizing clustering targets; for instance, scDeepCluster (Tian et al. 2019) offers a single-cell model-based deep embedded clustering approach that uses the loss function of Kullback-Leibler(KL) divergence to make the latent representation. scDMFK (Chen et al. 2020b) employs a multinomial distribution to characterize scRNA-seq data and a fuzzy weighted k-means clustering algorithm to cluster the cells in the latent space. scziDesk (Chen et al. 2020a) utilizes a negative log-likelihood function to capture the global probability structure of scRNA-seq in the latent space. DCA (Eraslan et al. 2019) is a deep count autoencoder network to denoise the scRNA-seq and utilizes a negative binomial noise model to capture the count distribution, overdispersion, and sparsity of the scRNA-seq data. However, these methods focus only on extracting the characteristics of individual cells, while the common structural information between cells is essentially ignored in the learning representation.

Besides, deep graph embedded clustering methods are well recognized in the single-cell field, focusing on the importance of the structural relationship between cells. For instance, scTAG (Yu et al. 2022) develops a topological adaptive graph convolution autoencoder to learn cell-cell topological representations. GraphSCC (Zeng et al. 2021) combines a graph convolutional network and a denoising autoencoder to integrate structural information of the scRNA-seq data, and a dual self-supervised module is employed to optimize the latent representations. scMGCA (Yu et al. 2023) builds a graph-embedding autoencoder to simultaneously learn cell-cell topology representation and cluster assignments. scGNN (Wang et al. 2021) is a hypothesis-free deep learning framework that combines a mixture Gaussian model with graph neural networks for scRNA-Seq analysis that utilizes graph neural networks to formulate the cell-cell relationships and combines a left-truncated mixture Gaussian model to capture the heterogeneous gene expression patterns.

Method

Fusion Cell Topology Encoder

Most previous deep embedded clustering always focused on learning the gene expression matrix and ignored the topology between cells. Therefore, in our study, we embed the cell-cell topology into an encoder that focuses on gene-level information analysis. To merge cell topology information into an effective feature representation at the gene level, we propose a fusion cell topology encoder with an inter-layer embedded operation. The fusion cell topology encoder consists of three fully connected layers. In contrast to the previous studies, we consider the compressed representation after fusion as the input of the fully connected layers and combine it with the cell-cell topological structural information at each layer. The input to layer l of the fusion cell topology encoder can be formulated as follows:

$$Z^l = f_\alpha(H_D^{l-1}, H_G^{l-1}) \tag{1}$$

where $f_\alpha(\cdot)$ is a linear combination of two different hidden information and α is the pre-defined hyper-parameter; H_D^{l-1} is the hidden information of the $i-1$ layers in the fusion cell topology encoder and H_G^{l-1} is the hidden information of the $l-1$ layers in the transcriptomics profile-based graph encoder. After that, the latent representation in the fusion cell topology encoder can be obtained as follows:

$$H_D^l = \phi(W_e Z^l + b_e) \tag{2}$$

where ϕ is ReLU(), the activation function of the fully connected layers; W_e and b_e are the learnable weight matrix and bias of the fusion cell topology encoder, respectively.

Transcriptomics Profile-based Graph Encoder

Existing single-cell graph embedding autoencoders lack a robust and synergistic approach to embed the cell topology and gene-level transcriptome information for feature representation learning. To capture the feature representation of gene expression data and the relationships between cells, we develop a transcriptomics profile-based graph encoder to integrate the gene-level transcriptome information into the graph encoder. Transcriptomics profile-based graph encoder of our model consists of three topology adaptive graph convolutional layers (TAGCN) (Du et al. 2017). A topology adaptive graph convolutional layer uses M graph convolution kernels to extract local structural features of different sizes, which could fully extract the graph information. The internal architecture of the polynomial convolution kernel is defined as:

$$G_{c,f}^l = \sum_{m=0}^{M} g_{c,f,m}^l (D^{-\frac{1}{2}} A D^{-\frac{1}{2}})^m \tag{3}$$

where $g_{c,f,m}^l$ denotes the polynomial coefficients; M is the number of graph convolution kernels; A is the adjacency matrix generated from the gene expression matrix by the KNN algorithm; and D is the degree matrix. After that, the $g_{c,f,m}^{(l)}$ is transmitted by the normalized adjacency matrix $\tilde{A} = D^{-\frac{1}{2}} A D^{-\frac{1}{2}}$. To embed the gene-level transcriptome information H_D^l from the fusion cell topology encoder into the transcriptomics profile-based graph encoder, we use the compressed representation Z^l as the input feature map of TAGCN on the l-th hidden layer, which is calculated from the formula:

$$Z^l = \alpha H_D^{l-1} + (1-\alpha) H_G^{l-1} \tag{4}$$

On this basis, the integrated representation can be used to generate a new representation to learn the feature representation of gene expression data and relationships between cells, which provides an approximate second-order graph regularization for the fused gene expression representation(Bo et al. 2020). In particular, we assume that before feature mapping each node has C_l features and c-th feature is $z_c^l \in R^N$, where $c = 1, 2, ..., C_l$. Then, the convolution operation process of TAGCN is defined as follows:

$$h_f^l = \phi(\sum_{c=1}^{C_l} G_{c,f}^l z_c^l + b_f \mathbf{1}_N) \tag{5}$$

where h_f^l represents the f-th output feature map, ϕ is ReLU(), b_f is a learnable bias, and $\mathbf{1}_N$ is the N_l dimension vector of all ones.

To simplify, we can provide the formulation of the output of the transcriptomics profile-based graph encoder as:

$$H_G^l = \phi(f_G(Z^l)) \tag{6}$$

where f_G is the graph convolution operation of the topology adaptive graph convolutional layer in the transcriptomics profile-based graph encoder; and Z^l is the input of the l-th layer that is calculated from Eq. 4.

Heterogeneous Structure Information Fusion

Inspired by the cross-modal fusion mechanism (Tu et al. 2021), we generate the fused heterogeneous latent represtation Z by incorporating the two latent embeddings (Z_D and Z_G) of the fusion cell topology encoder and the transcriptomics profile-based graph encoder. Indeed, our model takes both local and global cell correlations into account. After the cell graph has been processed by the transcriptomics profile-based graph encoder, the encoder's latent embedding will contain various orders of structural information. Then, we fuse them with the transcriptome gene level representation, thereby reducing the correlation between cells with multiple-hop relationships.

ZINB-based Multimodal Fusion Decoder

After integrating the compressed representations from the fusion cell topology encoder and a transcriptomics profile-based graph encoder, we get a fused heterogeneous latent embedding Z as the input of the decoder. In contrast to other fusion models, we used only a single decoder based on the zero-inflated negative binomial (ZINB) model to simultaneously connect the two encoder models, achieving the effect of reconstructing single-cell transcriptomic profiles and cell graphs. The ZINB distribution is used to capture the overall probability structure of the data and thus to model highly sparse and over-dispersed gene expression data. On this basis, we propose a ZINB-based multimodal fusion decoder to capture the characteristics of scRNA-seq data. At first, the decoder that reconstructs $\bar{X}$ can be described as:

$$\bar{X} = f_{dec}(W^{'}Z + b^{'}) \tag{7}$$

where $W^{'}$ and $b^{'}$ represent the weight matrix and the bias vecoters of the decoder, respectively; f_{dec} is a three-layer fully connected neural network; and Z is the fused latent embedding obtained by the heterogeneous structure information fusion mechanism. The reconstructed adjacency matrix A_r can be defined as the inner product between the latent embedding:

$$A_r = \sigma(Z^T Z) \tag{8}$$

And we define the reconstruction loss of A as follows:

$$L_r = \|A - A_r\|_2^2 \tag{9}$$

Subsequently, to capture the global probabilistic structure of the data, we also integrate the ZINB-based decoder into our decoder model, which connects three independent full connection layers with the last layer to estimate the parameters of ZINB: dropout rate π, dispersion degree θ and mean μ. The parameter matrices of the network output are defined as follows:

$$\Pi = sigmoid(W_{\pi}\bar{X}) \tag{10}$$

$$M = exp(W_{\mu}\bar{X}) \tag{11}$$

$$\Theta = exp(W_{\theta}\bar{X}) \tag{12}$$

where W represents the learned weights of the loss functions. The ZINB-based decoder reconstructs the scRNA-seq data as follows:

$$NB(X|\mu,\theta) = \frac{\Gamma(X+\theta)}{X!\Gamma(\theta)}(\frac{\theta}{\theta+\mu})^{\theta}(\frac{\mu}{\theta+\mu})^{X} \tag{13}$$

$$ZINB(X|\pi,\mu,\theta) = \pi\delta_o(X) + (1-\pi)NB(X) \tag{14}$$

Then, the reconstruction loss function of the original data X is defined as the negative log likelihood of the ZINB distribution:

$$L_{ZINB} = -log(ZINB(X|\pi,\mu,\theta)) \tag{15}$$

Joint Mutual Supervised Strategy

As the deep embedded clustering method is unsupervised, we apply a mutual supervised strategy, which unifies the fusion cell topology encoder, transcriptomics profile-based graph encoder and cluster module into a uniform optimization framework to effectively train the two modules for clustering. In our model, it consists of three different distribution blocks: the cluster distribution Q, the target distribution P and the fused heterogeneous latent embedding Z of the fusion model scDEFR; the cluster distribution Q^D, the target distribution P^D and the latent embedding Z_D of the fusion cell topology encoder; and the transcriptomics profile-based graph encoder with a clustering distribution Q^G, target distribution P^G and latent embedding Z_G. These three distributions in the same block are mutually supervised and united in a framework for learning and training. Considering the fusion model scDEFR, we define the soft label q_{iu} as follows:

$$q_{iu} = \frac{(1+\|z_i - \mu_u\|^2)^{-1}}{\sum_r (1+\|z_i - \mu_r\|^2)^{-1}} \tag{16}$$

This label represents the similarity between the latent embedding z_i and the cluster center μ_u, which is generated by spectral clustering or K-Means clustering after pre-training of the ZINB-based multimodal fusion decoder. In addition, on the basis of q_{iu}, we define the auxiliary target distribution p_{iu} as follows:

$$p_{iu} = \frac{q_{iu}^2/\sum_i q_{iu}}{\sum_r (q_{ir}^2/\sum_i q_{ir})} \tag{17}$$

Finally, we adopt the Kullback-Leibler (KL) divergence by minimizing the clustering target (Xie, Girshick, and Farhadi 2016), defined as follows:

$$L_C = KL(P||Q) = \sum_i \sum_u p_{iu} log\frac{p_{iu}}{q_{iu}} \tag{18}$$

It can be observed that distribution P supervises distribution Q learning, and target distribution P is calculated by distribution Q. Such a mutual supervision strategy contributes to learning a better representation of the data for the fusion model, resulting in higher quality clustering.

Moreover, to enable the latent representations generated by the different models to be as close as possible to the clustering centres of the original data and to avoid the collapse of the entire model, we use the binary cross-entropy as another objective function, using the distributions P^D and P^G to supervise distributions Q^D and Q^G, respectively. Note that they are calculated using Eq. 16 and 17 by replacing Z with Z_D and Z_G. Thus the target distributions can help the encoders learn better latent representations to achieve better clustering results. It is generated by a fusion cell topology encoder and a transcriptomics profile-based graph encoder, as described below:

$$L_D = -P^D log(Q^D) - (1-P^D)log(1-Q^D) \tag{19}$$

$$L_G = -P^G log(Q^G) - (1-P^G)log(1-Q^G) \tag{20}$$

Therefore, in the training process, the latent representations of the fusion cell topology encoder and the transcriptomics-based graph encoder and their fused representations are aligned with the robust target distribution simultaneously. Our scDEFR method has five optimization objectives:

$$L = \gamma_1 L_r + \gamma_2 L_{ZINB} + \gamma_3 L_C + \gamma_4(L_D + L_G) \tag{21}$$

where L_r is the reconstruction loss; L_{ZINB} is the ZINB loss; γ_1, γ_2, γ_3 and γ_4 are weight coefficients to control the balance of the total loss function.

Dataset	Platform	Cell	Gene	Group	Reference
Yan	Tang	90	20214	6	Yan et al.
Camp1	SMARTer	734	18927	6	Camp et al.
Camp2	SMARTer	777	19020	7	Camp et al.
QS_Diaphragm	Smart-seq2	870	23341	5	Consortium et al.
QS_Limb_Muscle	Smart-seq2	1090	23341	6	Consortium et al.
QS_Lung	Smart-seq2	1676	23341	11	Consortium et al.
Muraro	CEL-seq2	2122	19046	9	Muraro et al.
Adam	Drop-seq	3660	23797	8	Adam, Potter, and Potter
QX_Limb_Muscle	10x	3909	23341	6	Consortium et al.
QS_Heart	Smart-seq2	4365	23341	8	Consortium et al.
Young	10x	5685	33658	11	Young et al.
Plasschaert	inDrop	6977	28205	8	Plasschaert et al.
QX_Trachea	10x	11269	23341	5	Consortium et al.
QS_Trachea	Smart-seq2	1350	19992	4	Consortium et al.
QX_Bladder	10x	2500	23341	4	Consortium et al.

Table 1: Summary of the fifteen real scRNA-seq datasets.

Experiments

Data Sources and Preprocessing

To demonstrate the effectiveness of scDEFR, we applied our method to fifteen real scRNA-seq datasets collected from (Yu et al. 2022). These fifteen real datasets were generated from seven different representative sequencing platforms and originate from several species. The detailed information is described in Table 1. We first filter out genes that are expressed as non-zero in more than 1% of the cells and genes that are not expressed. Second, we normalized the data using the scanpy package. Third, we selected the top d highly variable genes based on the ranking of the normalized values. Finally, the KNN algorithm was employed to construct the cell graph, where each node in the graph represents a cell. For each cell, we identified its top-K similar neighbors and connected them via edges.

Baseline

We selected ten state-of-the-art methods for comparison

- Deep fusion clustering network (DFCN)(Tu et al. 2021): DFCN is a deep fusion clustering network with a fusion module for structure and attribute information based on interdependency learning for representation learning.
- ZINB-based graph embedded autoencoder (scTAG)(Yu et al. 2022): scTAG is a deep graph embedding clustering method that learns cell-cell topology representations and identifies cell clusters.
- Single-cell graph autoencoder (scGAE)(Luo et al. 2021): scGAE is a dimensionality reduction technique that builds a cell graph and employs a multitask-oriented graph autoencoder to preserve the topological structure and feature information in scRNA-seq data.
- Single-cell graph neural network (scGNN)(Wang et al. 2021): scGNN is a framework for hypothesis-free deep learning that formulates and aggregates cell-cell relationships using graph neural networks and then constructs heterogeneous gene expression patterns.
- Deep soft K-means clustering for scRNA-seq data (scziDesk)(Chen et al. 2020a): scziDesk combines the technique of deep learning with a denoising autoencoder to characterize scRNA-seq data in the latent space.
- Single-cell model-based deep embedded clustering (scDeepCluster)(Tian et al. 2019): scDeepCluster is a single-cell model-based deep embedded clustering method that clusters scRNA-seq data.
- Deep embedded clustering (DEC)(Xie, Girshick, and Farhadi 2016): DEC utilizes deep neural networks to discover feature representations and cluster assignments in a lower-dimensional feature space.
- Deep count autoencoder network (DCA)(Eraslan et al. 2019): DCA is a deep count autoencoder network that uses a negative binomial noise model to account for count distribution, overdispersion and sparsity data.

Implementation Details

In our study, we constructed the cell graph with the KNN algorithm with the nearest neighbor parameter at $K = 10$. In addition, we constructed the network using the combined fusion cell topology encoder and transcriptomics profile-based graph encoder, and the linear fusion parameter α was set to 0.1; each layer was configured with 1024, 128, and 24 nodes; and the layer of the fully connected decoder was configured with a symmetric encoder form. In particular, our algorithm consisted of pre-training and training, both of which were set to 250 epochs. The Adam algorithm was used as an optimizer, with a learning rate of 5e-5 for pre-training and 1e-7 for formal training. The weight coefficients for objective functions $\{\gamma_1, \gamma_2, \gamma_3, \gamma_4\}$ are respectively set to $\{0.3, 0.1, 0.5, 0.1\}$. The parameters of the baseline methods were set exactly as in the original publications. Finally, our experiments are conducted on an Ubuntu server with an NVIDIA Quadro RTX 6000 GPU and 24GB of memory.

Clustering Performance

To evaluate the performance of scDEFR, we employ two widely-used indices, the Normalised Mutual Information (NMI) and the Adjusted Rand Index (ARI). The higher the value of the index, the better the clustering performance.

The clustering performance of our method compared to the baseline methods on 15 scRNA-seq datasets is presented in Table 2. Each clustering method was run ten times to compute the average, and the values in red represent the best average index of clustering performance. On the 15 datasets, our method yields the 11 best NMI and ARI scores compared to other baseline algorithms, even reaching 0.9934 and 0.9976 on the 'Qx_Bladder', respectively. Furthermore, we observe that the deep embedding approach did not lead to a stable performance in clustering with a significant advantage. The main reason may be that gene expression information alone cannot fully capture the characteristics of the highly sparse scRNA-seq data. Compared to the deep graph embedding clustering methods, our proposed scDEFR is generally better, which demonstrates that the other deep graph embedding clustering methods may miss key patterns in the transcriptomic profiles. For DFCN containing fused representations, the clustering performance is better and more stable than the other three deep graph embedding

	Datasets	Ours	Deep Graph Embedded Methods				Deep Embedded Methods				Base Methods	
		scDEFR	DFCN	scTAG	scGNN	scGAE	scziDesk	scDeepCluster	DEC	DCA	K-Means	Spectral
NMI	Yan	**0.9057**	0.9056	0.682	0.8244	0.672	0.7656	0.7972	0.7205	0.8566	0.8069	0.6346
	Camp1	0.8037	**0.8749**	0.7716	0.6569	0.781	0.7356	0.8258	0.657	0.7006	0.7096	0.7881
	Camp2	**0.5606**	0.552	0.5085	0.3577	0.462	0.5087	0.4766	0.487	0.4667	0.3963	0.4712
	Qs_Diaphragm	0.9306	0.9402	0.8929	0.86	0.615	**0.9409**	0.0012	0.8253	0.8667	0.2537	0.3381
	Qs_Limb_Muscle	**0.9747**	0.9635	0.9327	0.8474	0.628	0.9519	0.0102	0.8523	0.8018	0.0937	0.3333
	Qs_Lung	**0.8133**	0.627	0.7845	0.7019	0.616	0.8094	0.0117	0.7474	0.712	0.2341	0.412
	Muraro	**0.8903**	0.785	0.8	0.7198	0.617	0.7759	0.5725	0.6592	0.8045	0.7018	0.8264
	Adam	**0.8781**	0.7129	0.8651	0.5587	0.633	0.8476	0.768	0.6688	0.4897	0.0722	0.0998
	Qx_Limb_Muscle	**0.953**	0.9412	0.9273	0.7166	0.567	0.9407	0.7359	0.8122	0.8129	0.4481	0.8443
	Qs_Heart	**0.9226**	0.8375	0.8689	0.7184	0.482	0.8448	0.1844	0.7991	0.86	0.2035	0.4648
	Young	0.762	0.6459	**0.8055**	0.5623	0.586	0.7665	0.2736	0.6002	0.5441	0.2544	0.3661
	Plasschaert	**0.8934**	0.6046	0.6561	0.5764	0.37	0.8582	0.5876	0.6406	0.6919	0.3965	0.5409
	Qx_Trachea	0.8356	0.5274	0.6615	0.4223	0.32	**0.8397**	0.5063	0.5541	0.5187	0.1	0.5585
	Qs_Trachea	**0.8619**	0.4647	0.7151	0.5262	0.489	0.7417	0.1201	0.6698	0.6632	0.1102	0.3222
	Qx_Bladder	**0.9934**	0.7994	0.8118	0.742	0.389	0.9511	0.5267	0.6193	0.661	0.5426	0.7545
ARI	Yan	**0.8955**	**0.8955**	0.4956	0.6822	0.55	0.5628	0.6911	0.5729	0.8029	0.7016	0.4461
	Camp1	0.6433	**0.7952**	0.6283	0.5046	0.547	0.5464	0.6303	0.4752	0.5717	0.5281	0.6214
	Camp2	0.4274	**0.5087**	0.3787	0.2306	0.229	0.4061	0.3617	0.3369	0.4091	0.332	0.3681
	Qs_Diaphragm	**0.9661**	0.9541	0.9137	0.8797	0.254	0.918	-0.0011	0.8645	0.7929	0.1617	0.28
	Qs_Limb_Muscle	**0.9873**	0.9751	0.956	0.8309	0.248	0.9729	-0.0023	0.8849	0.6802	0.0352	0.2398
	Qs_Lung	0.7002	0.4135	0.6252	0.5259	0.199	**0.7462**	0.0026	0.6314	0.6435	0.1453	0.2505
	Muraro	**0.9248**	0.789	0.8085	0.6031	0.202	0.6949	0.494	0.5445	0.8533	0.4959	0.6436
	Adam	**0.8851**	0.6348	0.8713	0.4125	0.249	0.8431	0.6241	0.5222	0.3647	0.0218	0.0368
	Qx_Limb_Muscle	**0.9744**	0.9516	0.95	0.6206	0.13	0.9127	0.4381	0.8253	0.7819	0.3775	0.8981
	Qs_Heart	**0.9684**	0.6947	0.9299	0.57	0.071	0.8307	-0.006	0.8778	0.9339	0.1147	0.3347
	Young	0.6344	0.5223	**0.7161**	0.4193	0.135	0.6831	0.1564	0.4505	0.3842	0.2008	0.225
	Plasschaert	**0.9353**	0.4966	0.6203	0.4394	0.037	0.8061	0.2974	0.5645	0.5934	0.3071	0.3537
	Qx_Trachea	**0.9463**	0.4159	0.6317	0.2478	0.1	0.9082	0.1568	0.4912	0.233	0.0652	0.4916
	Qs_Trachea	**0.9345**	0.3795	0.8152	0.4462	0.126	0.7626	0.0987	0.7639	0.4857	0.0043	0.3824
	Qx_Bladder	**0.9976**	0.7845	0.7538	0.7062	0.067	0.9612	0.2214	0.5476	0.6268	0.538	0.7385

Table 2: Performance of scDEFR and the other baseline methods on 15 scRNA-seq datasets. The red font indicates the best values among the compared methods.

methods, which also highlight the superiority of the fusion mechanism. In addition, we find that scDEFR had the 14 best NMI and 12 best ARI scores out of the 15 datasets compared with DFCN, respectively, indicating that the fusion suggestion in scDEFR improves the feature representation of scRNA-seq data. This is likely due to the fact that the ZINB model is capable of modeling scRNA-seq data effectively, and the latent embedding of the fused cell topology information can better capture the higher-order structural information of single-cell RNA-seq data. In addition, we also compared the running time of scDEFR with the other deep graph embedding clustering methods, including scGNN, scGAE, scTAG, and DFCN. As depicted in Fig. 3(A), scDEFR has the shortest running times on most scRNA-seq datasets. In addition, we applied t-SNE to visualize the latent embedding of the scDEFR and the other baseline methods in two-dimensional space, as depicted in Fig. 4. We can clearly observe that identical cells in the dataset can be well separated in the latent embedding representation of scDEFR. In summary, scDEFR can perform better than other methods.

Parameter Analysis

Impact of the Neighbor Parameter K: When constructing the cell graph with KNN, K is the number of edges between nodes. We ran our program with K parameters of 5, 10, 15, 20, and 25 to investigate the effect of K. Fig. 3(B) depicts the NMI and ARI values for the proposed model with various K values. As depicted in Fig. 3(B), the values of ARI and NMI increase rapidly from parameters 5 to 10, reaching an optimal value at K equals 10, then decreasing gradually from parameters 15 to 25. Therefore, we set the value of the neighbor parameter K to 10 in the model.

Impact of the Fusion Parameter α: In our study, α determines the fusion ratio between the transcriptomic profile information and the structural cell topology information. To explore the impact of the α, we ran our program with the α parameters of 0.1, 0.3, 0.5, 0.7, and 0.9. Fig. 3(C) shows the average NMI and ARI measures on the 15 datasets with different α values. We observe that when α goes from 0.1 to 0.3, both two indices decrease sharply. Although metrics improve as the value of the parameter increases from 0.3 to 0.5, the performance of α = 0.1 remains the best. Therefore, we set the α parameter as 0.1 in our model.

Different Numbers of Variable Genes Analysis: In the analysis of single-cell RNA data, highly variable genes play a significant role in determining the cell-type specificity and providing biological significance. To investigate the impact of the number of highly variable genes, we tested our method on the 15 datasets that contain varying numbers of differentially expressed genes. Fig. 3(D) depicts the average NMI and ARI values obtained on the 15 datasets that selected 300, 500, 1000, 1500, and 2000 highly variable genes. From the results, we conclude that the 2000 highly variable genes are the most effective for the proposed model. Therefore, we set the number of highly variable genes to 2000.

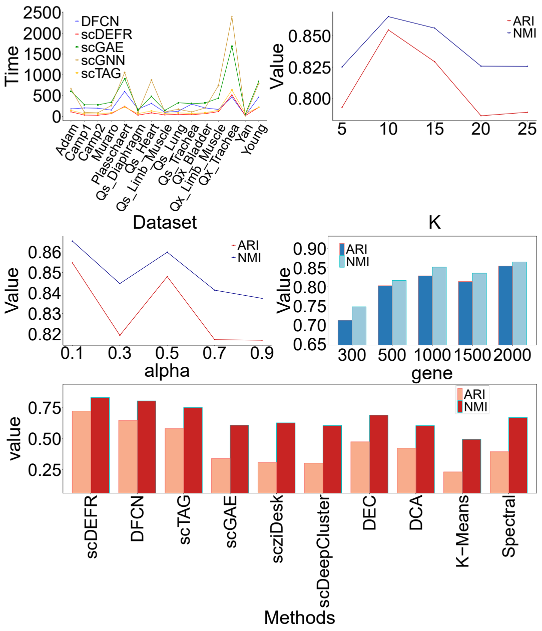

Figure 3: The analysis of (A) the running time comparison on the different datatsets. (B) the average NMI and ARI values with different neighbor parameters, K. (C) the average NMI and ARI values with different fusion parameter, α. (D) Comparison of the average NMI and ARI values with different numbers of genes. (E) NMI and ARI values for scDEFR and baseline methods on the 'Tabula Muris' dataset.

Methods	NMI	ARI
$scDEFR_{NAu}$	0.8566	0.8412
$scDEFR_{NGrA}$	0.8424	0.8087
$scDEFR_{NFir}$	0.8208	0.7716
$scDEFR_{NSec}$	0.8157	0.7579
$scDEFR_{NFcte}$	0.6646	0.5272
$scDEFR_{NTpge}$	0.8381	0.8069
$scDEFR_{NZinb}$	0.8564	0.8318
Ours	**0.8653**	**0.8547**

Table 3: Ablation study measured by NMI and ARI values

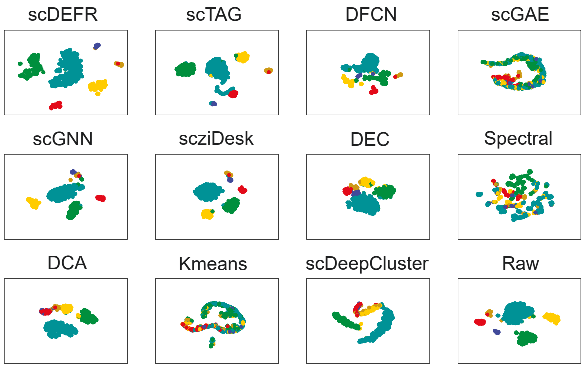

Figure 4: Comparison of clustering results with 2D visualization by t-SNE on the 'Qs_Limb_Muscle' dataset

Ablation Study

In this experiment, we analyzed the impact of each methodological component. We specifically ablated each component as described: 1) In the encoder, the fusion component of the fusion cell topology encoder was removed, resulting in the data feeding directly to the autoencoder, called $scDEFR_{NAu}$; 2) The transcriptomic profiles-based graphical encoder component, resulting in data going directly to the graph encoder , called $scDEFR_{NGrA}$; 3) The fusion component between the first and second layers of both encoders , called $scDEFR_{NFir}$; 4) The fusion elements between the second and third layers, called $scDEFR_{NSec}$; 5) without fusion cell topology encoder, called $scDEFR_{NFcte}$. 6) without a transcriptomic profiles-based graph encoder, called $scDEFR_{NTpge}$. 7) without ZINB, called $scDEFR_{NZinb}$. It is evident from Table 3 that combining the fusion cell topology encoder and the transcriptomic profiles-based graph encoder improves the clustering performance. Moreover, fusion cell topology encoder and ZINB had a substantial effect on the final clustering result, indicating the necessity to model scRNA-seq data based on ZINB distribution. In summary, each aspect of the scDEFR is reasonable and valid.

Scalability of scDEFR

To establish whether scDEFR can be used to analyze large datasets, we used it to cluster the model organism Mus musculus dataset called 'Tabula Muris' (Consortium et al. 2018). There are nearly 100,000 cells from 20 organs and tissues and 19179 genes with 55 cell types in the Tabula Muris. Fig. 3 (E) illustrates the clustering performance of scDEFR compared to the baseline methods on Tabula Muris. The scGNN clustering method was removed since it was unable to run on such a massive dataset. As depicted in Fig. 3 (E), scDEFR can provide excellent clustering performance on large datasets with the highest NMI and ARI values.

Conclusion

In this paper, we propose a single-cell model-based deep embedding fusion representation model for clustering. The compressed representations obtained from the fusion cell topology encoder and the transcriptomics profile-based graph encoder are incorporated into the model in a layer-by-layer fashion. Then, the heterogeneous structural information representations are produced after fusing the two aforementioned distinct kinds of data and employing the ZINB-based multimodal fusion decoder for reconstructing a cell graph and the transcriptomic information, respectively. Finally, a joint mutual supervision strategy is proposed to optimize both the embedded fusion representations and the clustering performance. Through experiments on 15 real scRNA-seq datasets, we demonstrate the superiority of the proposed scDEFR method over other state-of-the-art baseline methods. Moreover, evidence from ablation studies and scalability studies demonstrates that scDEFR is efficient, reliable, and extensible.

Acknowledgements

The work described in this paper was substantially supported by the National Natural Science Foundation of China under (Grant No. 62076109, 61972174) and the Jilin Province Outstanding Young Scientist Program (Grant NO. 20230508098RC), and also funded by 'the Fundamental Research Funds for the Central Universities, JLU'. This research was substantially sponsored by the research projects (Grant No. 32170654 and Grant No. 32000464) supported by the National Natural Science Foundation of China and was substantially supported by the Shenzhen Research Institute, City University of Hong Kong.

References

Adam, M.; Potter, A. S.; and Potter, S. S. 2017. Psychrophilic proteases dramatically reduce single-cell RNA-seq artifacts: a molecular atlas of kidney development. *Development*, 144(19): 3625–3632.

Bo, D.; Wang, X.; Shi, C.; Zhu, M.; Lu, E.; and Cui, P. 2020. Structural deep clustering network. In *Proceedings of The Web Conference 2020*, 1400–1410.

Camp, J. G.; Badsha, F.; Florio, M.; Kanton, S.; Gerber, T.; Wilsch-Bräuninger, M.; Lewitus, E.; Sykes, A.; Hevers, W.; Lancaster, M.; et al. 2015. Human cerebral organoids recapitulate gene expression programs of fetal neocortex development. *Proceedings of the National Academy of Sciences*, 112(51): 15672–15677.

Camp, J. G.; Sekine, K.; Gerber, T.; Loeffler-Wirth, H.; Binder, H.; Gac, M.; Kanton, S.; Kageyama, J.; Damm, G.; Seehofer, D.; et al. 2017. Multilineage communication regulates human liver bud development from pluripotency. *Nature*, 546(7659): 533–538.

Chen, L.; Wang, W.; Zhai, Y.; and Deng, M. 2020a. Deep soft K-means clustering with self-training for single-cell RNA sequence data. *NAR genomics and bioinformatics*, 2(2): lqaa039.

Chen, L.; Wang, W.; Zhai, Y.; and Deng, M. 2020b. Single-cell transcriptome data clustering via multinomial modeling and adaptive fuzzy k-means algorithm. *Frontiers in genetics*, 11: 295.

Consortium, T. M.; et al. 2018. Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature*, 562(7727): 367–372.

Du, J.; Zhang, S.; Wu, G.; Moura, J. M.; and Kar, S. 2017. Topology adaptive graph convolutional networks. *arXiv preprint arXiv:1710.10370*.

Eraslan, G.; Simon, L. M.; Mircea, M.; Mueller, N. S.; and Theis, F. J. 2019. Single-cell RNA-seq denoising using a deep count autoencoder. *Nature communications*, 10(1): 1–14.

Grün, D.; Kester, L.; and Van Oudenaarden, A. 2014. Validation of noise models for single-cell transcriptomics. *Nature methods*, 11(6): 637–640.

Johnson, S. C. 1967. Hierarchical clustering schemes. *Psychometrika*, 32(3): 241–254.

Luo, Z.; Xu, C.; Zhang, Z.; and Jin, W. 2021. A topology-preserving dimensionality reduction method for single-cell RNA-seq data using graph autoencoder. *Scientific reports*, 11(1): 1–8.

MacQueen, J. 1967. Classification and analysis of multivariate observations. In *5th Berkeley Symp. Math. Statist. Probability*, 281–297.

Muraro, M. J.; Dharmadhikari, G.; Grün, D.; Groen, N.; Dielen, T.; Jansen, E.; Van Gurp, L.; Engelse, M. A.; Carlotti, F.; De Koning, E. J.; et al. 2016. A single-cell transcriptome atlas of the human pancreas. *Cell systems*, 3(4): 385–394.

Papalexi, E.; and Satija, R. 2018. Single-cell RNA sequencing to explore immune cell heterogeneity. *Nature Reviews Immunology*, 18(1): 35–45.

Plasschaert, L. W.; Žilionis, R.; Choo-Wing, R.; Savova, V.; Knehr, J.; Roma, G.; Klein, A. M.; and Jaffe, A. B. 2018. A single-cell atlas of the airway epithelium reveals the CFTR-rich pulmonary ionocyte. *Nature*, 560(7718): 377–381.

Svensson, V.; Natarajan, K. N.; Ly, L.-H.; Miragaia, R. J.; Labalette, C.; Macaulay, I. C.; Cvejic, A.; and Teichmann, S. A. 2017. Power analysis of single-cell RNA-sequencing experiments. *Nature methods*, 14(4): 381–387.

Tian, T.; Wan, J.; Song, Q.; and Wei, Z. 2019. Clustering single-cell RNA-seq data with a model-based deep learning approach. *Nature Machine Intelligence*, 1(4): 191–198.

Tian, T.; Zhang, J.; Lin, X.; Wei, Z.; and Hakonarson, H. 2021. Model-based deep embedding for constrained clustering analysis of single cell RNA-seq data. *Nature communications*, 12(1): 1–12.

Tu, W.; Zhou, S.; Liu, X.; Guo, X.; Cai, Z.; Zhu, E.; and Cheng, J. 2021. Deep Fusion Clustering Network. In *Proceedings of The Thirty-Fifth AAAI Conference on Artificial Intelligence*, 9978–9987.

Wang, J.; Ma, A.; Chang, Y.; Gong, J.; Jiang, Y.; Qi, R.; Wang, C.; Fu, H.; Ma, Q.; and Xu, D. 2021. scGNN is a novel graph neural network framework for single-cell RNA-Seq analyses. *Nature communications*, 12(1): 1–11.

Xie, J.; Girshick, R.; and Farhadi, A. 2016. Unsupervised deep embedding for clustering analysis. In *International conference on machine learning*, 478–487. PMLR.

Yan, L.; Yang, M.; Guo, H.; Yang, L.; Wu, J.; Li, R.; Liu, P.; Lian, Y.; Zheng, X.; Yan, J.; et al. 2013. Single-cell RNA-Seq profiling of human preimplantation embryos and embryonic stem cells. *Nature structural & molecular biology*, 20(9): 1131–1139.

Young, M. D.; Mitchell, T. J.; Vieira Braga, F. A.; Tran, M. G.; Stewart, B. J.; Ferdinand, J. R.; Collord, G.; Botting, R. A.; Popescu, D.-M.; Loudon, K. W.; et al. 2018. Single-cell transcriptomes from human kidneys reveal the cellular identity of renal tumors. *science*, 361(6402): 594–599.

Yu, Z.; Lu, Y.; Wang, Y.; Tang, F.; chun Wong, K.; and Li, X. 2022. ZINB-Based Graph Embedding Autoencoder for Single-Cell RNA-Seq Interpretations. In *AAAI*.

Yu, Z.; Su, Y.; Lu, Y.; Yang, Y.; Wang, F.; Zhang, S.; Chang, Y.; Wong, K.-C.; and Li, X. 2023. Topological identification and interpretation for single-cell gene regulation elucidation across multiple platforms using scMGCA. *Nature Communications*, 14(1): 400.

Zeng, Y.; Lin, J.; Zhou, X.; Lu, Y.; and Yang, Y. 2021. Graph Convolutional Network-based Method for Clustering Single-cell RNA-seq Data. *bioRxiv*, 2020–09.