A Comprehensive Study of DeepLinc Model and Robustness Optimization by Spatial Transformer

Li Yujian	2130034019
Wang Haowei	2130034035
Wu Rui	2130034041
Xu Xinmu	2130034044

Abstract

In bioinformatics, cell clusters are used to study intercellular interactions. However, single-cell intercellular interactions networks also contain essential information. The construction of this network is challenging because it is difficult to obtain distal cell interactions from the residual of a priori knowledge and some existing models are susceptible to noise data. The proposed deepLinc model helps solve these problems and interpolate the incomplete interactions network. According to the authors, deepLinc applies deep generative model and graph convolution network to form a model consisting of three parts: an encoder, an adversarial network module and a decoder. Spatially encoded clustering is performed using latent space to cluster single cells into cell clusters with unsupervised learning and re-model the network between cell clusters. The report is an overview of the structure and algorithm of deepLinc and the idea of using spatial transformer to further improve the robustness of the model is presented and experimented.

Keywords: DeepLinc, Intercellular interactions network, Spatial transformer, Bioinformatics, Deep generative models.

1. Introduction

According to research findings, physiological functions in body depend on cell-cell communication, which manage single cell functions, homeostasis, and organismal development [1]. Disease results from improper cell communication or molecular message interpretation. Thus, identifying intercellular signaling networks is of great significance. Besides, research on cellular function increasingly needs to consider more the coordination of cellular activity, as multicellular organisms have different cell types and cell-cell interactions within tissues [2]. This further emphasizes to researchers the importance of the interaction between single-celled and multi-cellular environments when studying the intrinsic gene expression of single cells. In the relevant studies, spatially resolved transcriptome profile has provided insightful resources and progress for understanding cellular tissue patterns of various types of tissues and organs. However, the realization of simulating cellular interaction networks about distant cell-to-cell interaction is still a challenge at present. Deep learning is a compelling choice for describing and forecasting complicated biological phenomena for which there are no priori hypotheses and it contributes to dealing with

single-cell spatial transcriptome, since it enables a multilayer neural network to learn data representations with multiple levels of abstraction [3]. Therefore, DeepLinc based on deep generative models can be used for modeling high-dimensional scRNA-seq data using latent features.

2. Related Works

2.1 Graph Convolutional Networks

Graph Convolutional Network (GCN), first proposed in 2017, is a neural network architecture designed to process data that can be represented as a graph, which can be considered as a generalization of CNN to graph-structured data [4]. By representing data in the form of graphs, the connections between different elements can be captured and provide more insightful information about the dataset. While some embedding methods have been successful, many are limited by their shallow learning mechanisms and may fail to uncover more complex patterns within graphs [5]. Deep learning models have shown great success due to the ability of CNNs is to take advantage of the stationarity and compositionality properties of certain types of data. For example, the grid-like structure of images allows CNNs to use convolutional layers to extract hierarchical patterns and high-level features. CNNs achieve this by using trainable filters that scan each pixel and combine information from surrounding pixels [6].

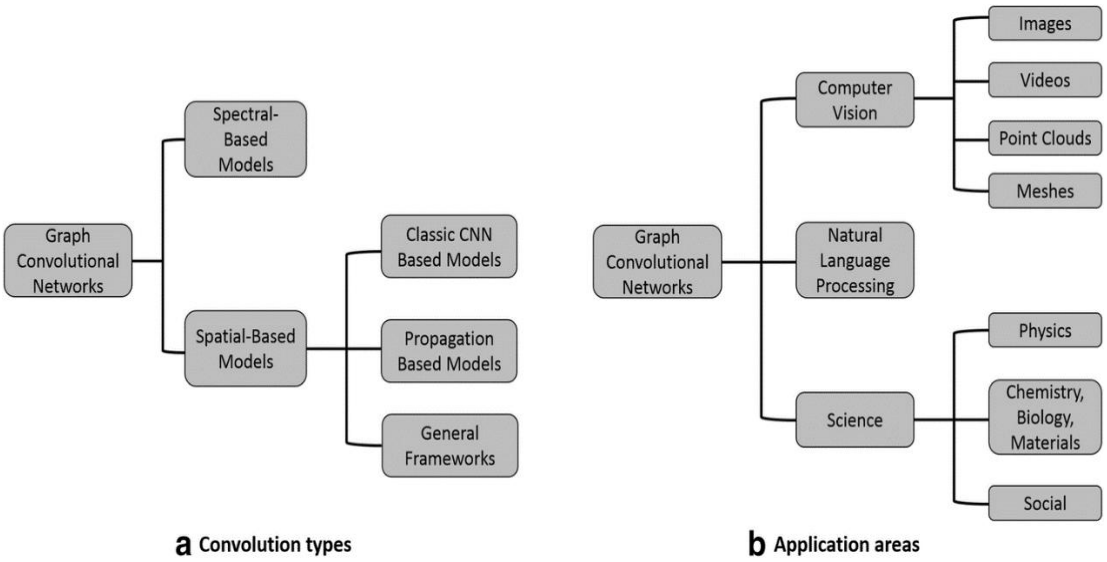

Figure 1. An overview of graph convolutional networks [4].

2.2 Deep Generative Models

Deep generative models (DGM) are neural networks with many hidden layers trained to approximate complicated, high-dimensional probability distributions using loads of samples [7]. It is widely used in many fields of the union of AI and Machine Learning. The latest developments in parameterizing these models using deep neural networks, combined with advancements in stochastic optimization methods, have enabled scalable modeling of complex high-dimensional data, including images, text, and speech. For common deep generative models, there are two types: GANs and VAEs.

Generative Adversarial Networks (GANs) are a powerful type of neural network used for unsupervised learning, which is inspired by game theory and involves two competing models, a generator, and a discriminator. Variational Autoencoder (VAE), is a powerful generative model that can describe the probability distribution of each latent attribute by specifying the model's encoder. It provides a probabilistic manner for describing an observation in latent space.

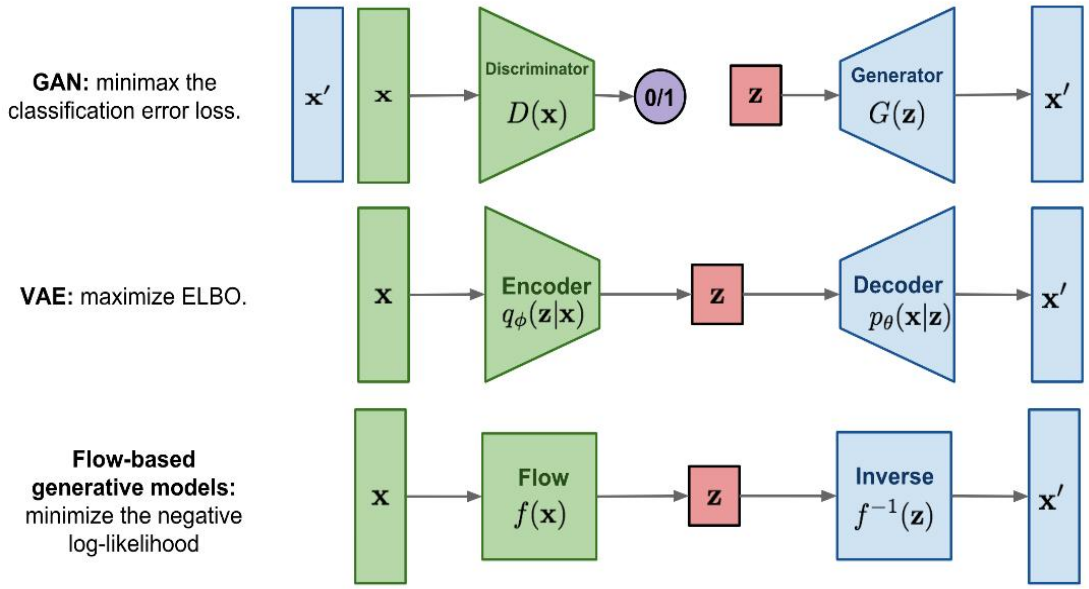

Figure 2. An overview of GAN and VAE [8].

2.3 Comparative analysis of cell–cell interaction by Scriabin

For many research scholars, they insist that the physiological functions of multicellular tissues are dominated by the cell composition, such as the pancreas which is constrained only by the two types of islet cells that make it up [9]. In contradiction to this, the physiological functions of multicellularity rely on the coordination within and between their tissue ecological niches to maintain homeostasis and to respond appropriately to internal and external disturbances [10]. In intercellular interactions, the intrinsic gene expression of each single cell is the result of a network of cellular interactions in the physiological environment. According to the biological knowledge, it is commonly acknowledged that the interactions have the ways, including electrical signals (between neurons), chemical signals and intercellular direct exposure [11].

DeepLinc and Scriabin are both methods used to analyze cell-cell interactions at single-cell resolution with GCN. DeepLinc exhibit its extraordinary capability of learning from imperfect and incomplete spatial transcriptome data, filtering spurious interactions, and inferring missing proximal and distal interactions [12]. As a matter of fact, apart from the interruption from the missing and noise data, the extremely redundant attention paid on the cell clusters also have similar drawbacks due to the ignorance of information in the single cell levels, although it seems to be reasonable to do this. Stanford indicated that they have developed a new approach, Scriabin, a flexible and extendable framework for the comparison analysis of the cell-cell

interactions with single cell resolution. With the assistance of Scriabin, it is more accurate to permutate the expected edges of intercellular communications and it is more feasible to recognize the communication concealed by the cellular clustering methods [10].

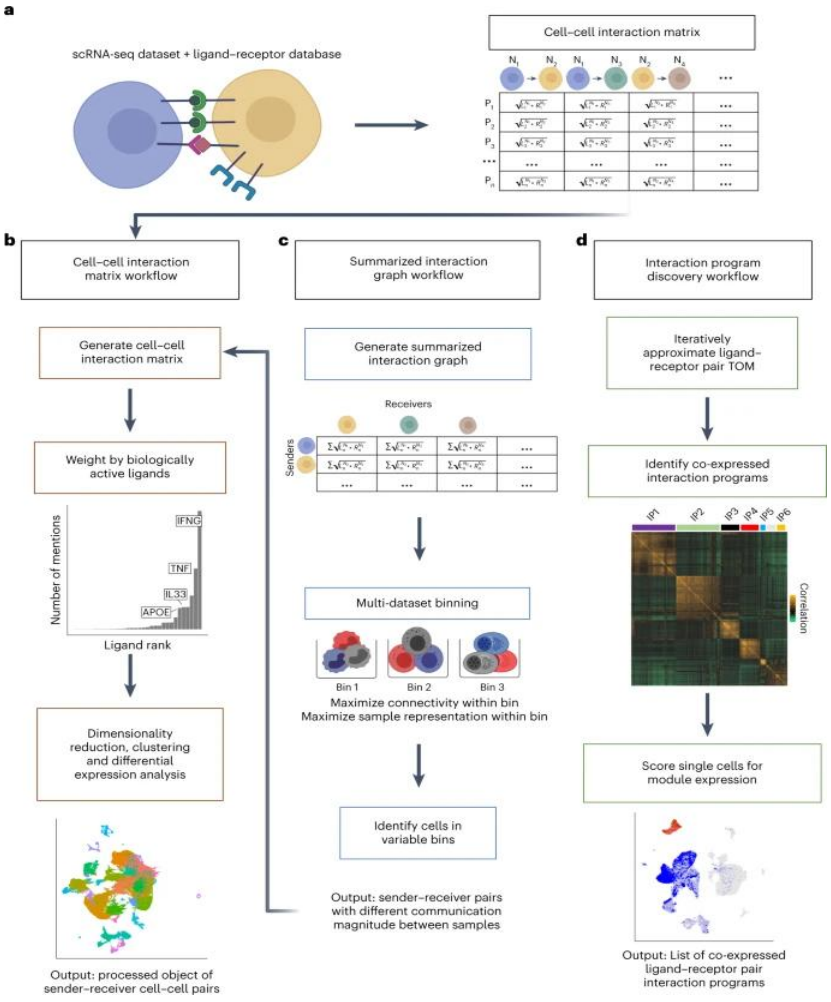

Figure 3. An overview of Scriabin Structure [10].

4. Methodology

4.1 DeepLinc Structure

DeepLinc has been specifically designed and optimized to denoise and infer missing interactions by efficiently learning from the incomplete sets of cell organization revealed by various methods of single-cell spatial transcript profiling. The database applied in DeepLinc is called seqFISH, it contains four csv files of mouse visual cortex cells. The four csv files are adj.csv, cell_type_1.csv, coord.csv, and counts.csv. These four csv files correspond to four variables: adj.csv contains an adjacent matrix as a predefined local interaction map, cell_type_1.csv contains the types of cells, coord.csv contains the coordinate of cells, and counts.csv contains the raw count matrix with cells in rows and genes in columns. These four variables are preprocessed by a preprocessing module to transform the general coordinate information from single-cell spatial transcriptome data into the adjacency matrix. The adjacency matrix, which contains the information about cells and genes, is the input

of the DeepLinc model. DeepLinc combines the VGAE and an adversarial network to learn from the single-cell spatial transcriptome profiles, which helps to generate a latent distribution that captures the intrinsic associations between cell-cell interactions and gene expression patterns of single cells. DeepLinc consists of three main components: an encoder, a decoder, and an adversarial regularization module in the VGAE framework.

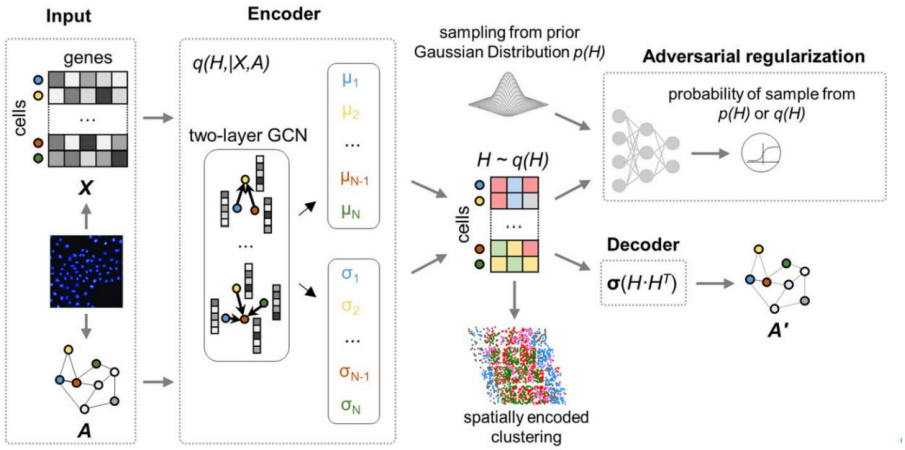

Figure 4. An overview of DeepLinc Structure [12].

The encoder is based on a two-layer graph convolutional network (GCN) to determine whether there is communication between cells. The graph-structured profiles and the preprocessed node features are fed into the encoder consisting of two graph convolutional layers. GCN evaluates the relationship between some entities, simply expressed as points and edges. The main purpose of GCN is to determine whether there is a relationship between each point (cell), these relationships are the features in deep learning. The output of this variational graph convolutional network (VGCN) encoder is the latent representation (H), which captures the characteristics of a single cell itself and its neighboring cells. An adversarial regularization module from a prior Gaussian distribution is also used to constrain H. The decoder is a Sigmond function for the dot product of latent variables, which performs a dot product operation on H to generate a new adjacency matrix to represent the reconstructed cell-cell interaction network. The vectors of H contain the latent information of cell interaction landscapes and gene expression profiles can visualize and cluster single cells.

4.2 Spatially Encoded Clustering

The main output of DeepLinc is to reconstruct the incomplete intercellular interactions network, however, it uses the latent information of interaction landscapes and gene expression profiles as well for the visualization and clustering of single cells. After the clustering, the single cell interaction network can be transformed into interaction network between cell clusters. In this process, our group use **Calinski-Harabasz score** to evaluate the performance of the clustering learner based on the variance analysis of the result. The higher the score, the better the clusters are. Calinski-Harabasz score (Variance Ratio Score) can be calculated as follows:

Step1: Calculate the inter-cluster dispersion B. The inter-cluster dispersion B is calculated as the sum of squared distances between the centroids of each cluster and the centroid of the whole dataset.

$$B = \sum_i |c_i|(m - m_i)^2$$

Where $|c_i|$ is the size of cluster i, m is the centroid for all clusters and m_i is the centroid for cluster c_i.

Step2: Calculate the intra-cluster dispersion W. The intra-cluster dispersion W is calculated as the sum of squared distances between each point and its cluster centroid.

$$W = \sum_i \sum_{x \epsilon c_i} (x - m_i)^2$$

Where m_i is the centroid for cluster c_i.

Step3: Calculate the Calinski-Harabasz Score CH using the following formula:

$$CH = \frac{\frac{B}{k-1}}{\frac{W}{n-k}} = \frac{n-k}{k-1} \times \frac{B}{W}$$

where k is the total number of clusters, and n is the total number of observations.

4.2 The algorithms of DeepLinc

4.2.1 The Choice of Direct Contacts.

DeepLinc uses neighboring cells with direct contacts as the positive set for learning the transcriptome features related to cell-cell interactions. For each cell, we only used the 3 nearest neighbors to define direct contacts. This is a balanced choice generating enough number of direct contacts for training the DeepLinc pipeline and ensuring few false positives to contaminate the positive training set.

4.2.2 The Optimization of VAE in the DeepLinc Pipeline

The VAE model maps the original graph to a Gaussian distribution in the latent space through an encoder, samples a random point from the distribution, and then maps the sampled point back to the original graph through a decoder. By minimizing the reconstruction error and the KL divergence in the latent space, the VAE model learns a low-dimensional representation of compressed data

KL divergence is used to measure the difference between this Gaussian distribution and the standard normal distribution, thus helping the VAE model learn a better representation of the latent space. During the training, the model parameters are updated using the Adam optimizer. The embedding vector and loss values are calculated after each update to evaluate the model performance. In the VAE model, the encoder maps the input data to a Gaussian distribution in the latent space, and the mean and variance of the distribution are generated by encoder. A point is randomly sampled from this Gaussian distribution and used as the embedding vector.

Eventually, the DeepLinc pipeline was optimized to maximize the log-likelihood of cell-cell interaction network. The objective of VGAE is to minimize its cost function by optimizing the evidence lower bound (ELBO):

$$\zeta_{ELBO} = E_{q(H|A)}[\log(A|H)] - D_{KL}(q(H|X,A)||p(H))$$

where A is the observed data expressed as adjacency matrix, X is the gene expression matrix, H is the latent variable, $p(H)$ is the prior distribution over the latent variable, $q(H|X,A)$ is the approximate posterior distribution over the latent variable given the data, $log(A|H)$ is the log-likelihood of the data given the latent variable, and $D_{KL}(q(H|X,A)||p(H))$ is the KL divergence between the approximate posterior distribution and the prior distribution over the latent variables.

$$D_{KL}(q(H|X,A)||p(H)) = E_{q(H|A)}\left[\frac{\log(q(H|X,A))}{p(H)}\right]$$

where $q(H|X,A)$ is the approximate posterior distribution over the latent variable given the data, $p(H)$ is the prior distribution over the latent variable, and $\log(q(H|X,A))$ is the logarithm of the ratio of the approximate posterior distribution and the prior distribution over the latent variable.

4.3 The Optimization of Robustness by Spatial Transformer

The nonlinear, high-dimensional, sparse, and multimodal features of single-cell spatial transcriptome data make it a suitable and feasible target for the deep learning strategy. However, without sophisticated methods to discover intercellular interactions, spatially resolved single-cell transcriptome profiles fail to vary degrees from data imperfections such as too many missing values, batch effects, biased and low coverage, and high noise.

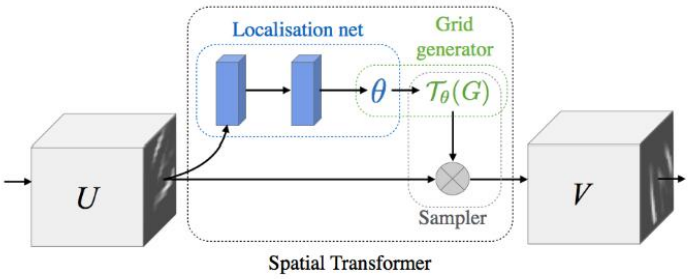

Figure 5. An overview of the spatial transformer structure.

Spatial Transformer is a deep learning network structure that improves the accuracy and robustness of the model by learning how to spatially transform the input image. The model can learn how to rotate, scale, or translate the target region in the input image to better fit the shape and scale of the target. Spatial Transformer consists of three components: Localization net, Grid generator, and Sampler. Localization net is a small convolutional neural network that accepts and input image array with a set of parameters θ that describe the transformation of the input, such as translation, rotation, and scaling. Grid generator receives the output parameters θ of Localization net and generates a grid that describes how pixels are sampled in the input image and

the grid generator maps each pixel of the input image to the corresponding location in the output image. Sampler receives the grid the output grids of Grid generator and sampling in the input image. Sampler sets the value of each output pixel to the pixel value at the corresponding position in the input image. The values can be sampled using different interpolation methods to handle pixel sampling at non-integer locations.

Through the combination of these three components, Spatial Transformer learns how to transform the input image and feeds the transformed image to the subsequent network for processing. During training, the network updates the parameters of the localization network (θ) by a back-propagation algorithm to minimize the loss function of the model. During testing, Spatial Transformer can transform the new input images to improve the accuracy of the model. With the combination of Spatial Transformer, we added one more feature to the original dataset data, which is whether there is a distal interaction between cell a and remote cell x. In this case, the encoder of DeepLinc will make the network more robust since considering more features will make the cell reconstruction more accurate.

5. Results

5.1 Experimental Study

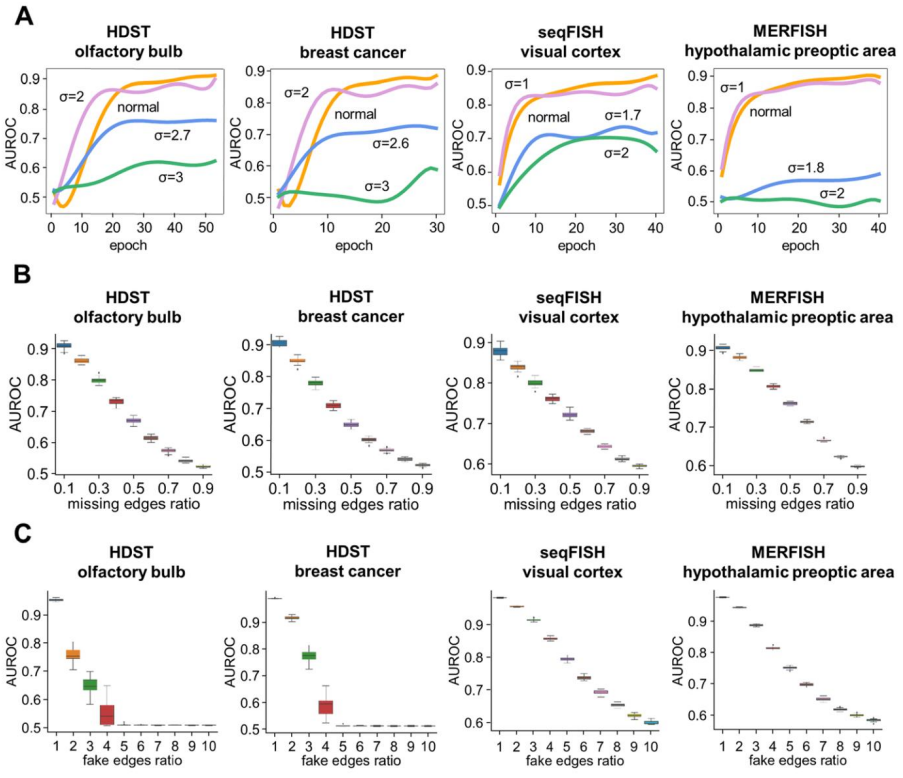

Figure 6. Result of the denoising performance of DeepLinc [12].

The original data or data with different levels of random noise in the gene expression profiles were used. Larger sigma values would result in higher levels of random noise in general. However, according to the result, it shows that our model is robust to the gene expression, even the sigma tends to become larger, the AUROC will not change a lot. Different proportions of the edges in the original cell adjacency

maps were randomly picked and discarded, the remaining edges were used to continue training. The result shows that our model can extract the uncomplete edges features or limited edges features to recover the deletion edges. The AUROCs between mixed-in fake edges and pre-existing real edges. Different numbers of randomly generated fake edges were added into the original cell adjacency map. This process was repeated 30 times for each level of fake edges to draw a boxplot. It shows that our model has a high tolerance and have a good denoising effect.

5.2 Result Analysis

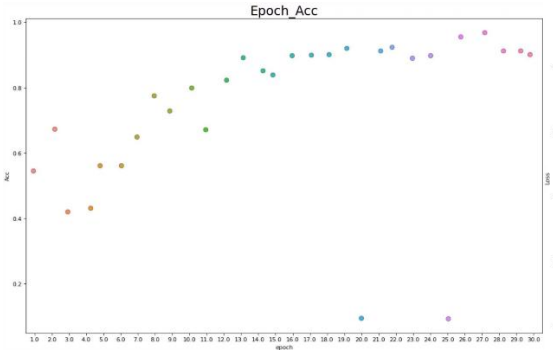

Figure 7. Accuracy in each epoch of 30 epochs.

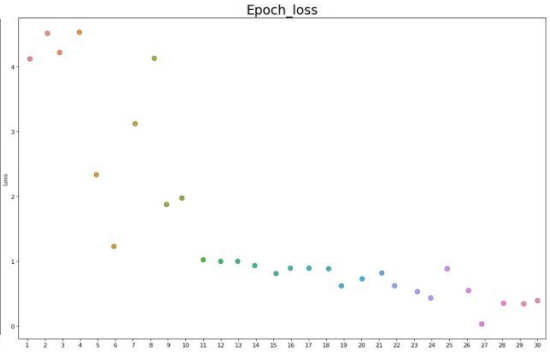

Figure 8. Loss in each epoch of 30 epochs.

When the model was trained, we recorded the accuracy and loss for the first 30 epochs of training. The high accuracy indicates that the design and implementation of the model is effective and able to make good predictions for the given dataset. This indicates that implementation of the model is excellent. While high accuracy indicates that implementation of the model is effective, it is significant whether there is poor generalization problem. This requires further deep analysis and tests of the model to confirm its performance and stability in different situations.

6. Conclusion and Future Work

6.1 Conclusion

In conclusion, our model has the similar performance to DeepLinc, it shows that the retrofit is successful and has no negative effect on the performance of the model. Multiple solutions are possible: The success of the retrofit shows that there are ways to achieve the same performance. This means that there are multiple options to choose from when designing and improving the model, not just one approach. Model complexity may not increase: The modified model is identical in performance to the original model, which may mean that the modified model does not increase the complexity of the model. If this is correct, it can be a good result because simpler models are likely to be easier to maintain and interpret.

Table 1. The comparison of average accuracy and loss between old and new model.

Datasets	Accuracy DeepLinc	Loss DeepLinc	Accuracy New_DeepLinc	Loss New_DeepLinc
seqFISH	0.845775	0.992435	0.79124	1.5346
MERFISH	0.80346	1.249603	0.802413	1.37302
HDST_ob	0.823368	0.8925860	0.823154	0.92345
HDST_cancer	0.856	0.790472897	0.781245	1.23563

6.2 Future Work

Our group have proposed to use spatial transformer to optimize DeepLinc. However, due to data absence and limited computing power of current computers, it is demanding for us to build a comprehensive network presently. Although the modified model is identical in performance to the original model, more evaluation is needed to confirm the performance and stability of the model in different situations. If we can continue to try the combination of DeepLinc and spatial transformer in the future, this model will make great progress in biomedical and biological breeding.

References

[1] "Cell Communication - an overview | ScienceDirect Topics," Cell Communication - an overview | ScienceDirect Topics. https://www.sciencedirect.com/topics/neuroscience/cell-communication

[2] H. Rouault and V. Hakim, "Different Cell Fates from Cell-Cell Interactions: Core Architectures of Two-Cell Bistable Networks," Biophysical Journal, vol. 102, no. 3, pp. 417–426, Feb. 2012, doi: 10.1016/j.bpj.2011.11.4022.

[3] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, May 2015, doi: 10.1038/nature14539.

[4] S. Zhang, H. Tong, J. Xu, and R. Maciejewski, "Graph convolutional networks: a comprehensive review," Computational Social Networks, vol. 6, no. 1, Nov. 2019, doi: 10.1186/s40649-019-0069-y.

[5] A. Grover and J. Leskovec, "node2vec: Scalable Feature Learning for Networks," arXiv.org, Jul. 03, 2016. https://arxiv.org/abs/1607.00653v1

[6] Z. Gao, Z. Lu, J. Wang, S. Ying, and J. Shi, "A Convolutional Neural Network and Graph Convolutional Network Based Framework for Classification of Breast Histopathological Images," IEEE Journal of Biomedical and Health Informatics, vol. 26, no. 7, pp. 3163–3173, Jul. 2022, doi: 10.1109/jbhi.2022.3153671.

[7] L. Ruthotto and E. Haber, "An introduction to deep generative modeling," GAMM-Mitteilungen, vol. 44, no. 2, May 2021, doi: 10.1002/gamm.202100008.

[8] L. Weng, "Flow-based Deep Generative Models," Flow-based Deep Generative Models | Lil'Log, Oct. 13, 2018.

[9] R. Farhadifar, J.-C. Röper, B. Aigouy, S. Eaton, and F. Jülicher, "The Influence of Cell Mechanics, Cell-Cell Interactions, and Proliferation on Epithelial Packing," Current Biology, vol. 17, no. 24, pp. 2095–2104, Dec. 2007, doi: 10.1016/j.cub.2007.11.049.

[10] A. J. Wilk, A. K. Shalek, S. Holmes, and C. A. Blish, "Comparative analysis of cell–cell communication at single-cell resolution," Nature Biotechnology, May 2023, Published, doi: 10.1038/s41587-023-01782-z.

[11] C. D. Buckley, C. Ospelt, S. Gay, and K. S. Midwood, "Location, location, location: how the tissue microenvironment affects inflammation in RA," Nature Reviews Rheumatology, vol. 17, no. 4, pp. 195–212, Feb. 2021, doi: 10.1038/s41584-020-00570-2.

[12] R. Li and X. Yang, "De novo reconstruction of cell interaction landscapes from single-cell spatial transcriptome data with DeepLinc," Genome Biology, vol. 23, no. 1, Jun. 2022, doi: 10.1186/s13059-022-02692-0.