Neural Style Transfer-Based Framework for Bulk Multi-Omics Data Integration

Lihao Liu[1]
Washington University in St. Louis, 1 Brookings Dr. St. Louis, MO 63105-2215.

1 Introduction

We are now in the era of high-throughput sequencing technologies, which have enabled the synchronic measurement of molecular information across multiple biological levels including genome, transcriptome, proteome, and metabolome. Research in bioinformatics and computational systems biology increasingly regard methods to interpret multi-omics data and derive insights into the interrelationships in biological systems. Multi-omics integration provides unprecedented opportunities in the field. Different from horizontal harmonization tasks which focus on reducing the batch effects across datasets with identical features, multi-omics integration in this context is vertical and entails merging heterogeneous data from different modalities within partially or completely matched samples, thereby revealing potentially important patterns [1, 2].

Existing research in multi-omics integration has employed various machine learning approaches ranging from probabilistic graphical models to unsupervised neural networks. One key challenge arises from the high feature dimensionality of biological omics data, with a feature space that typically far exceeds the sample size. In RNA-Seq transcriptomics data, for example, feature count can range from 20 000 to 25 000, corresponding to the number of human genes [3]. As a subset of multi-omics integration tasks, single-cell data integration benefits from the large sample size (n) of single-cell data which can reach millions, vastly exceeding the feature count (p) ($n \gg p$) and enabling the use of complex machine learning approaches [4]. Conversely, bulk data integration usually concerns much smaller sample cohorts usually consisting of patients, and thereby grapples with the opposite scenario ($n \ll p$). In practical application of neural-network-based methods, this characteristic poses significant modeling challenges, including overfitting and poor generalization due to the curse of dimensionality.

Therefore, for bulk integration, most research focuses on Bayesian methods that utilize prior knowledge to impose regularization. Multi-Omics Factor Analysis (MOFA) [5], developed by Argelaguet *et al.* in 2018, models multi-omics data as linear combinations of latent factors. Bayesian priors are placed on these factors and their associated weights, whose posterior distribution is then approximated with variational inference. This method effectively addresses the dimensionality problem of bulk data and performs well in disentangling cross-modality semantics. Yet, the strengths of the model, including scalability and interpretability come at the expense of reduced robustness and complexity by assuming linear relationships within the data.

Style transfer refers initially to tasks in the field of computer vision, particularly image generation and editing. Its one goal is to disentangle the content and style of data, which enables the reconstruction of data that adopt the stylistic characteristics of one source and the semantic content of another. The neural style transfer approach proposed by Gatys *et al.* (2015) employs CNNs in the task [6], following which there has been increasing research on machine learning applications, and style transfer tasks have also

[1] lihaol@wustl.edu.

generalized to other domains including time-series data and natural language processing.

Researchers have achieved promising results in application of style transfer methods in the context of horizontal integration of biological data, which involves removing the batch effects or technical variations in the data collection process ("style") from the conserved homogenous semantic content across datasets. However, the framework's application in vertical data integration remains scarce, especially in bulk integration tasks, which suffer from the mismatch of dimensionality.

To address this gap, this project explores the possibility of applying a style transfer neural-network-based approach in vertical bulk integration. Neural network models generalize to non-linear functions and alleviate the scalability issues of Bayesian approaches. Style transfer models in particular focus on the separation of style and content semantics, which aligns with the goal of multi-omics integration and potentially reduces the risks of overfitting.

2 Related Work

2.1 CycleGAN

The CycleGAN model proposed by Zhu *et al.* in 2017 is a generative adversarial network (GAN) approach designed for unpaired image-to-image translation [7]. The model achieved transformation of style and content between two image domains by using two coupled GANs that map between them in opposite directions. A key innovation in CycleGAN is the cycle-consistency loss:

$$\mathcal{L}_{cyc}(G,F) = \mathbb{E}_{x\sim p_{data}(x)}[|F\big(G(x)\big) - x|_1] + \mathbb{E}_{y\sim p_{data}(y)}[|G\big(F(y)\big) - y|_1].$$

Where G and F are GANs mapping in directions $x \to y$ and $y \to x$ respectively.

CycleGAN has been widely applied in the field of computer vision in tasks such as artistic style transfer, photo editing, and even medical imaging. The model proposed by this work employs a similar framework of inverse mappings where each modality uses a distinct autoencoder. The cycle-consistency loss is also borrowed as an essential optimization objective.

2.2 Multi-Omics Factor Analysis (MOFA)

Introduced by Argelaguet *et al.* in 2018, MOFA is a computational framework that integrates multi-omics data by decomposing them into a small set of interpretable latent factors. Mathematically, MOFA uses a linear combination scheme:

$$\boldsymbol{Y}_m = \boldsymbol{Z}\boldsymbol{W}_m^T + \boldsymbol{\epsilon}_m$$

Where $\boldsymbol{Y}_m$ is the data in modality m, $\boldsymbol{Z}$ is the latent factor space, $\boldsymbol{W}_m$ is the modality-specific loadings, and $\boldsymbol{\epsilon}_m$ is the modality-specific prior noise. $\boldsymbol{Z}$ and $\boldsymbol{W}_m$ are both enforced with prior distributions, with the prior of $\boldsymbol{W}_m$ encouraging sparsity to achieve interpretable in weighting. Consequently, $\boldsymbol{Z}$ and $\boldsymbol{W}_m$ are optimized with variational Bayesian inference. This statistical approach provides high latent interpretability and computing efficiency and enables the model to be used on partially missing data, which is a common challenge in multi-omics datasets.

3 Methodology

This section introduces in detail the method and architecture of the model proposed by this work.

3.1 Problem Definition

To define the specific problem in bulk multi-omics integration, this project draws the unified latent space approach from MOFA. Storing disentangled cross-modality information in latent factors is a promising method that enables consistent and straightforward downstream evaluation. Therefore, we propose the problem as follows:

Given multi-omics datasets $\{\boldsymbol{X}_m \in \mathbb{R}^{N\times p_m}\}_{m=1}^M$ where m corresponds to a specific omics out of a total of M modalities, the objective is to find a **unified latent space** $\boldsymbol{Z} \in \mathbb{R}^{N\times k}$ that captures the shared cross-modality biological signals. The number of latent factors

is determined such that $k \ll p_m$ to enforce an extraction of features. For each m, define a mapping $f_m: \mathbb{R}^{p_m} \to \mathbb{R}^k$ such that $f_m(\boldsymbol{X}_m) \approx \boldsymbol{Z}$. The objective is then to optimize f_m based on various constraints and desired model properties.

3.2 Model Framework

Autoencoders (AEs) Autoencoders are used as mapping functions between the sample space and the latent space. A distinct encoder is used for each modality to generate a latent space carrying information shared across modalities: $\boldsymbol{Z}_s^m = E_s^m(\boldsymbol{X}^m)$, where $\boldsymbol{Z}_s^m$ is the shared latent representation for modality m. Simultaneously, a style representation is generated with another encoder for each modality: $\boldsymbol{Z}_t^m = E_t^m(\boldsymbol{X}^m)$. Conversely, to reconstruct the original data from the shared latent representation, modality-specific decoders are employed: $\widehat{\boldsymbol{X}}^m = D^m(\boldsymbol{Z}_s^m, \boldsymbol{Z}_t^m)$. Note that decoders are set to require style inputs.

Cross-Modality Attention Module In this model, we employ an attention-based algorithm in the steps of harmonizing the shared latent representations $\{\boldsymbol{Z}_s^m \in \mathbb{R}^{N\times k}\}_{m=1}^M$ into a unified latent space $\boldsymbol{Z} \in \mathbb{R}^{N\times k}$. First, each $\boldsymbol{Z}_s^m$ is projected to Queries ($\boldsymbol{Q}^m$), Keys ($\boldsymbol{K}^m$), and Values ($\boldsymbol{V}^m$) tensors using learnable weights ($\boldsymbol{W}_Q, \boldsymbol{W}_K, \boldsymbol{W}_V \in \mathbb{R}^{k\times k}$):

$$\boldsymbol{Q}^m = \boldsymbol{W}_Q \boldsymbol{Z}_s^m$$

$$\boldsymbol{K}^m = \boldsymbol{W}_K \boldsymbol{Z}_s^m$$

$$\boldsymbol{V}^m = \boldsymbol{W}_V \boldsymbol{Z}_s^m$$

Then, for each modality, the module computes the similarity between its query and the keys from all other modalities, generating pairwise attention scores. The attended latent representations are then the sum of the value vectors of other modalities weighted by these scores:

$$\text{Cross}-\text{attention}^{ml} = \text{softmax}\left(\frac{\boldsymbol{Q}^m(\boldsymbol{K}^l)^T}{\sqrt{k}}\right)$$

$$\boldsymbol{Z}_{attended}^m = \mathbb{E}_{m,l\sim M}[\text{Cross}-\text{attention}^{ml} \cdot \boldsymbol{V}^l]$$

Finally, the unified latent space is computed by a layer normalization of a combined latent representation of residual connection and the simple average of the attended vectors:

$$\boldsymbol{Z} = \text{LayerNorm}(\mathbb{E}_m \boldsymbol{Z}_{attended}^m + \mathbb{E}_m \boldsymbol{Z}_s^m)$$

In which the LayerNorm function operates across feature layers, scaling them for normalization within each sample.

3.3 Loss Implementation

Reconstruction Loss To ensure the Autoencoder effectively and meaningfully extract style and shared data, we impose the reconstruction loss as the mean squared error between reconstructed and original data.

$$\mathcal{L}_{rec} = \mathbb{E}_m \left\| D^m(\boldsymbol{Z}, \boldsymbol{Z}_t^m) - \boldsymbol{X}^m \right\|_2^2$$

Cycle-consistency Loss The cycle-consistency loss computes the reconstruction error by decoding the latent representation of a modality with the style of another then remapping the obtained data back to the original modality.

$$\mathcal{L}_{cyc} = \mathbb{E}_{i,j\sim M} \left\| D^m\left(E^j\left(D^j\left(E_s^i(\boldsymbol{X}^i), \boldsymbol{Z}_t^j\right)\right), \boldsymbol{Z}_t^i\right) - \boldsymbol{X}^i \right\|_2^2$$

Cross-Modality Loss Similar to the cycle-consistency loss, the cross-modality loss iteratively examines each pair of modalities and enforces the similarity between the shared latent representations.

$$\mathcal{L}_{cross} = \mathbb{E}_{i,j\sim M} \left\| \boldsymbol{Z}_s^i - \boldsymbol{Z}_s^j \right\|_2^2$$

Style Disentanglement Loss To ensure the disentanglement between style and content vectors, a loss is employed to their Frobenius distance.

$$\mathcal{L}_{dis} = \mathbb{E}_m \left\| \boldsymbol{Z}_t^m \circ \boldsymbol{Z}_s^m \right\|_2^2$$

Attention Sparsity Loss and Attention Entropy Loss To promote diversity and interpretability in the attention weights while preventing overfitting and overly deterministic distribution, a pair of counteracting losses is employed to regularize the attention weights spaces.

$$\mathcal{L}_{att-sparsity} = \mathbb{E}_{n\sim N}\left[\mathbb{E}_{i,j\sim M}\left\|\boldsymbol{A}_{ij} - \overline{\boldsymbol{A}_{ij}}\right\|_2^2\right]$$

$$\mathcal{L}_{att-entropy} = -\mathbb{E}_{n\sim N}\left[\mathbb{E}_{i,j\sim M}\left[\boldsymbol{A}_{ij}\log \boldsymbol{A}_{ij}\right]\right]$$

Where $\boldsymbol{A}_{ij} \in R^{N\times M\times M}$ is the attention weights, and $\overline{\boldsymbol{A}_{ij}} = \frac{1}{M}$ is the uniform tensor.

Latent Sparsity Loss Similar to the previous loss, the latent sparsity loss promotes sparsity in the final unified space.

$$\mathcal{L}_{lat-sparsity} = \mathbb{E}_m\left\|\boldsymbol{Z}_s^m\right\|_1$$

Variational Information Bottleneck (VIB) Loss To regularize the shared latent space to avoid biased distributions and overfitting, the last loss is chosen to be the VIB loss.

$$\mathcal{L}_{vib} = Var(\boldsymbol{Z}_s^m) = \mathbb{E}_{n\sim N}\left\|\boldsymbol{Z}_{s,n}^m - \mathbb{E}_{n\sim N}\boldsymbol{Z}_{s,n}^m\right\|_2^2$$

Integrated Loss Function

$$\begin{aligned}\mathcal{L}_{total} = \lambda_{rec}\mathcal{L}_{rec} + \lambda_{cross}\mathcal{L}_{cross} + \lambda_{cyc}\mathcal{L}_{cyc} + \lambda_{dis}\mathcal{L}_{dis} \\ + \lambda_{att-sparsity}\mathcal{L}_{att-sparsity} \\ + \lambda_{att-entropy}\mathcal{L}_{att-entropy} \\ + \lambda_{lat-sparsity}\mathcal{L}_{lat-sparsity} + \lambda_{vib}\mathcal{L}_{vib}\end{aligned}$$

3.4 Experiment Setup

Dataset and Preprocessing The experiments in this paper are conducted on the multi-omics datasets that originate from The Cancer Genome Atlas (TCGA) Breast Invasive Carcinoma (TCGA-BRCA) cohort, accessible online via the LinkedOmics database. The datasets employed include methylation, mutation, transcriptomics, proteomics and phenotype labels data, respectively with shapes (num_patients, num_features) = (315, 13807), (976, 7966), (1094, 20155), (106, 9733), and (1097, 20). The preprocessing pipeline of the datasets consists of the following steps: a) For data of each modality, features with any missing (NaN) values are removed for stability. b) The union and intersection of patient IDs across modalities are identified, and the patient IDs are re-ordered to ensure that the overlapping patient IDs are aligned. c) Log-transformation is applied to non-negative omics data (methylation and transcriptomics) to reduce skewness, and normalization is applied within each modality. Before being fed into the model, the preprocessed data is truncated to only include the overlapping patient IDs, which have data in all modalities available.

Experimental Environment Our experiment is based on the PyTorch framework and are conducted on a virtual environment using PyCharm 2024.2.1. The training function loops over 3000 epochs with a batch size of 32. For both the autoencoder framework and the cross-modality attention module, the learning rate is set to $lr = 3\times10^{-6}$, with the Xavier uniform initialization and the Adam optimizer applied. To stabilize training and prevent exploding gradients, gradient clipping is employed. $\lambda_{rec} = 10.1$, $\lambda_{cross} = 1.1$, $\lambda_{cyc} = 10$, $\lambda_{dis} = 0.1$, $\lambda_{att-sparsity} = 0.1$, $\lambda_{att-entropy} = 0.07$, $\lambda_{lat-sparsity} = 2$, and $\lambda_{vib} = 1.0$. Dynamic Weights Adjustment (DWA) is applied to adjust the loss weights across epochs. For the training function, we utilize a stage-wise method, where the epochs are equally divided into two stages: Training in stage 1 omits $\mathcal{L}_{cross}$ and $\mathcal{L}_{cyc}$, and aims to establish the effectiveness of the individual modality-specific autoencoders. The two losses are then integrated in step 2 to optimize cross-modality alignment and consistency. The attention mechanism also becomes more properly trained in stage 2. The final latent space uses 16 factors.

3.5 Evaluation Methods

The evaluation of the effectiveness of the unified latent space as a product of this multi-omics data integration method can simulate downstream analysis tasks. In this section, we briefly describe the metrics and analysis applied to the result in this experiment.

Variance Decomposition Each shared latent factor is separately fed into the decoders with the corresponding style latent space to reconstruct the original data. The proportion of variance explained (PVE) of a latent factor is then computed using the ratio of the reconstructed variance to the total original variance:

$$\mathrm{PVE}(m,k) = \frac{\mathrm{Var}\left(\widehat{\boldsymbol{X}}_k^m\right)}{\mathrm{Var}(\boldsymbol{X}_m)}$$

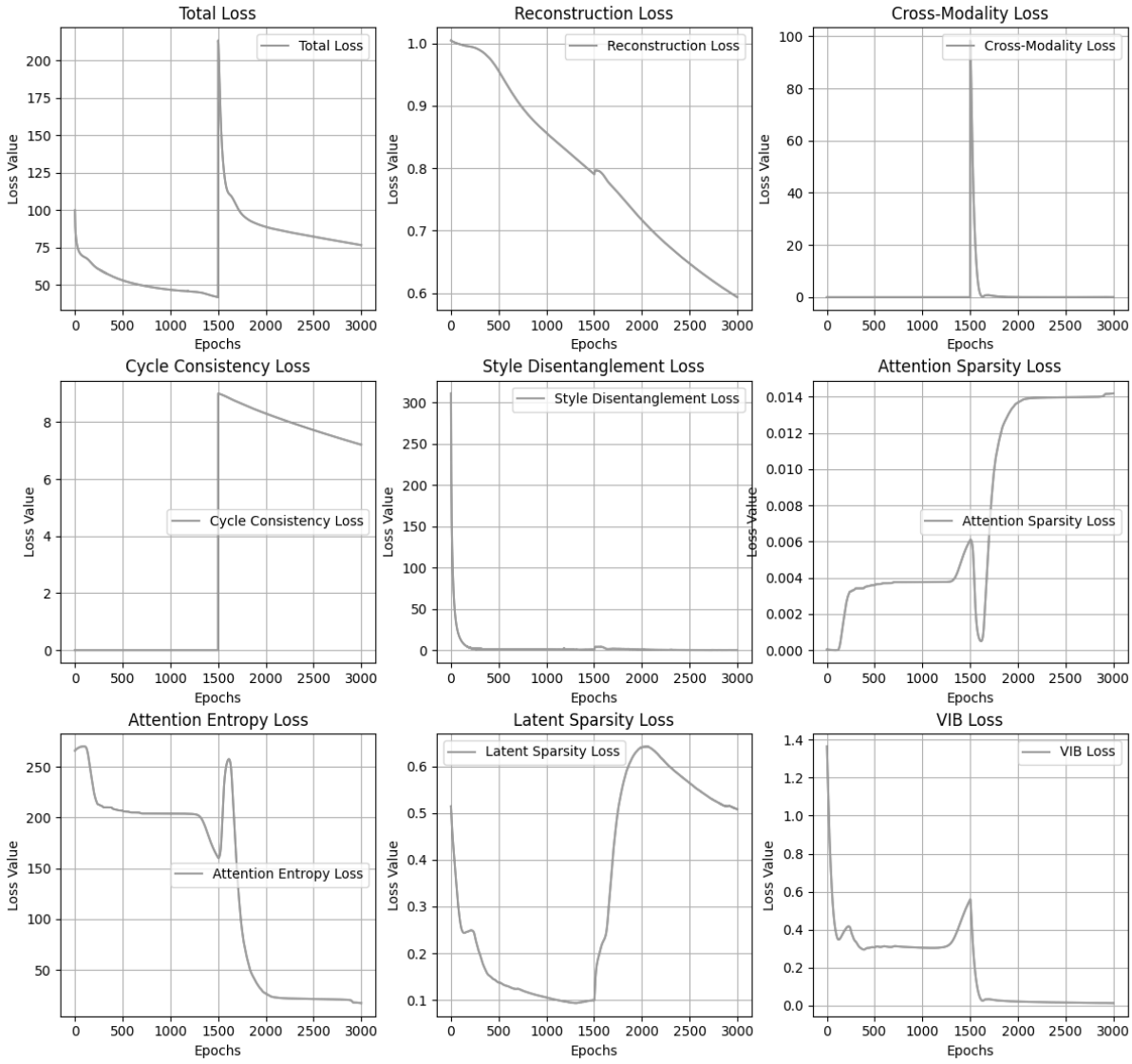

Fig. 1. *The behaviors of the total model loss and the 8 loss components in relation to training epochs. The cross-modality loss and cycle-consistency loss are activated at the start of stage 2 (*epoch *= 1500) to enforce cross-modality integration. In stage 2, the attention entropy loss is effectively further optimized. However, the attention sparsity loss, which seems to yield values exactly inversely related to the attention entropy loss, exhibit diverging behavior in stage 2.*

This method effectively identifies latent factors with varying explanatory power, thus providing an evaluation of the interpretability of the obtained latent space.

Latent Space t-SNE Visualization For a more general and deterministic evaluation, t-Distributed Stochastic Neighbor Embedding (t-SNE) analysis is applied to visualize sample clustering in the latent space and the relationship between the separation and biologically meaningful phenotype labels.

Biomarker Discovery As an important class of downstream analysis for multi-omics integration, biomarker discovery tasks regard the relationship between original features and the unified latent structure itself and predict the features that are most likely biomarkers in the data cohort with or without input of phenotype labels. This experiment conducts a brief simulation of the task on each of the modalities.

Phenotype Label Regression The integration of the unified latent space can be regarded as cross-modal dimensionality reduction since the obtained space has a far smaller feature count, which theoretically contributes to increased predictive capabilities. This analysis comprehensively evaluates the model's ability to encode phenotype-related information in the latent factors.

4 Results

Once our multi-omics integration model is trained, subsequent analysis and evaluation is conducted to reveal the capacity of the proposed framework to capture biologically meaningful information in the generated latent space. In this section, we demonstrate the results of the implementation of the four downstream tasks mentioned earlier.

While applying decomposition of variance to the model requires the trained decoder for reconstruction, the latent t-SNE visualization and phenotype label regression tasks can be conducted with solely the input of a unified latent space and the phenotype labels, which gives generalizability and enables comparison

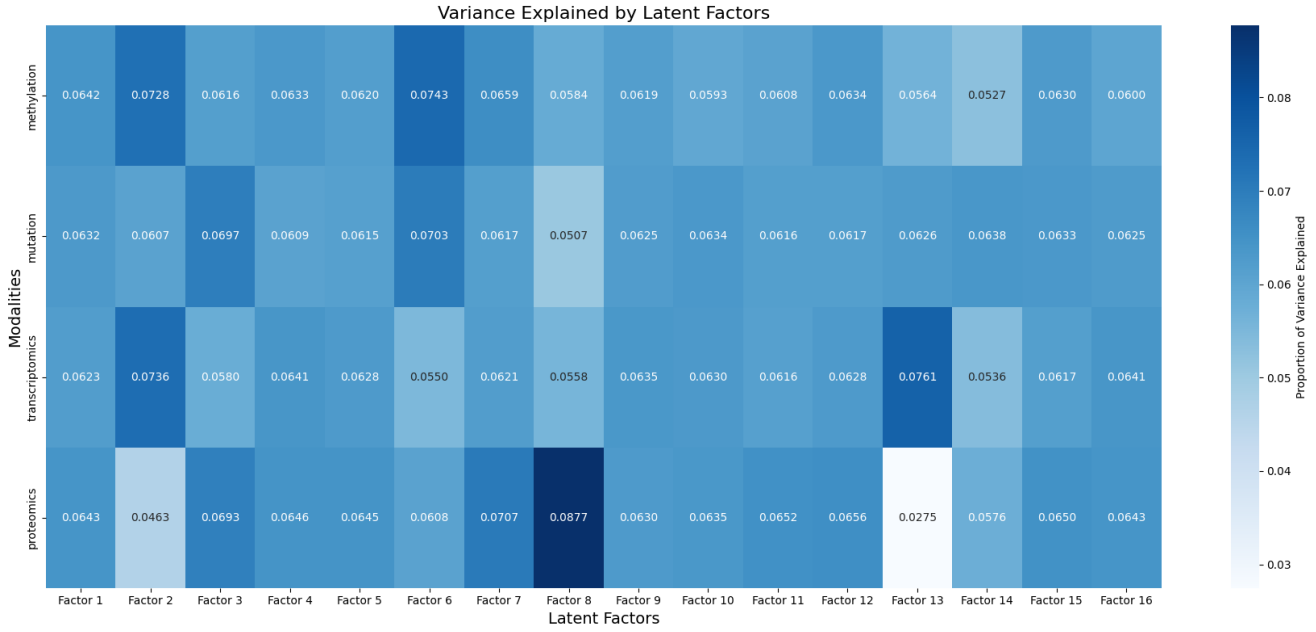

Fig. 2. *Proportions of variance in the original data explained by the 16 latent dimensions plotted as a heatmap. The relatively smooth baseline variance decomposition indicates a complicated and balanced latent representation, and the sparsity corresponds to disentanglement of the original data and enhanced interpretability.*

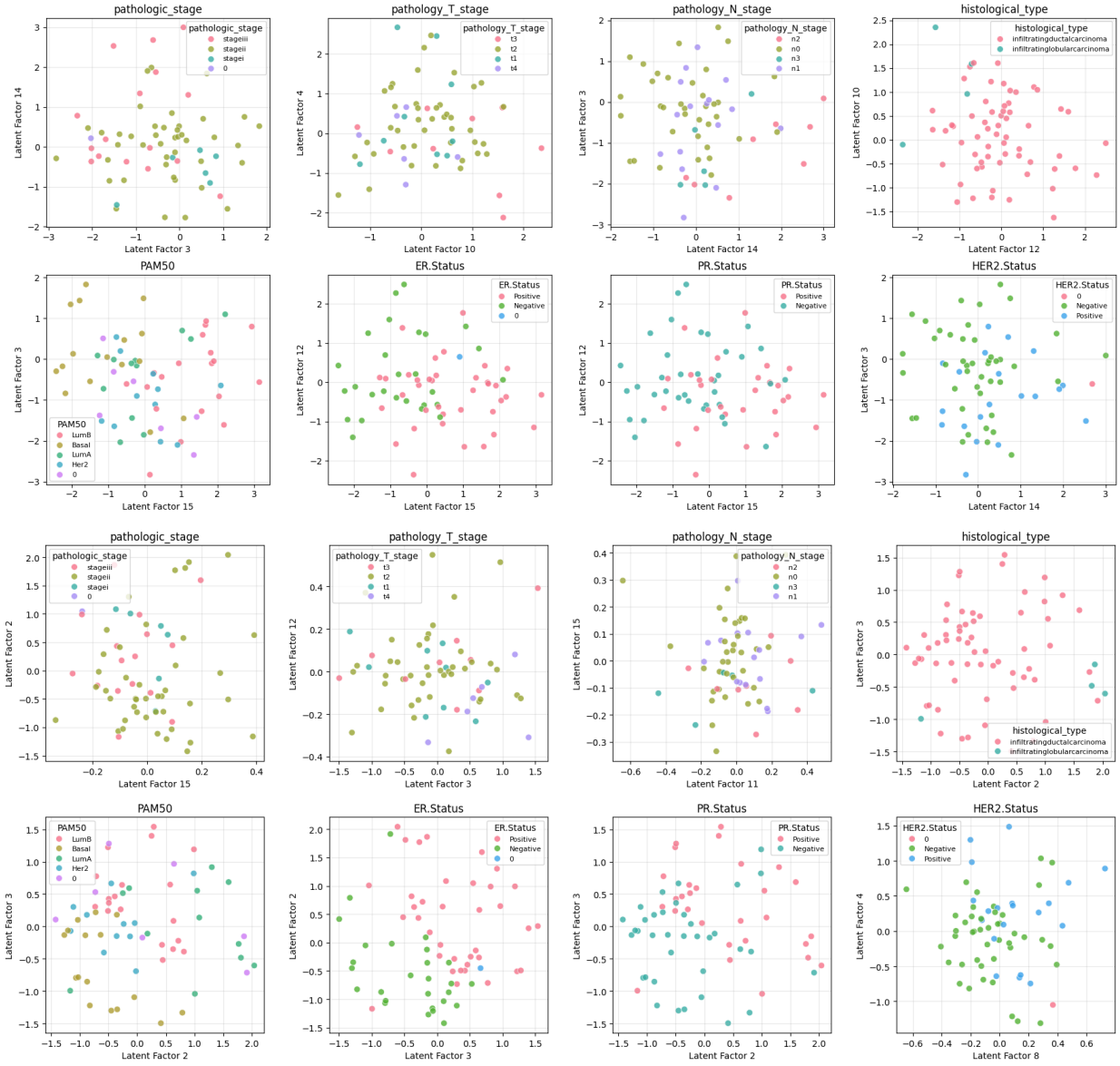

Fig. 3. *Separation of classes in 8 categorical phenotype features by two most principal latent factors, selected using ranks of Fisher scores, in the latent spaces generated in this model (rows 1-2) and the MOFA+ model (rows 3-4).*

experiments. Therefore, we compared this style-transfer-based model to the MOFA latent variable model for the two tasks.

The loss curves in Fig. 1 provide a general view of the system's optimization scheme. At the start of training stage 2 (epoch = 1500), with the activation of the cross-modality and cycle-consistency losses, certain losses sharply peak and thereafter quickly converge. At the stage transition, modality-specific losses including the reconstruction loss evolve with perturbations yet maintain their original converging trends, indicating their separation from cross-modality integration; the VIB loss and the attention entropy loss, on the other hand, converge to near-zero in stage 2, indicating that the incorporation of cross-modality losses alters their convergence potentials. The optimization of most losses

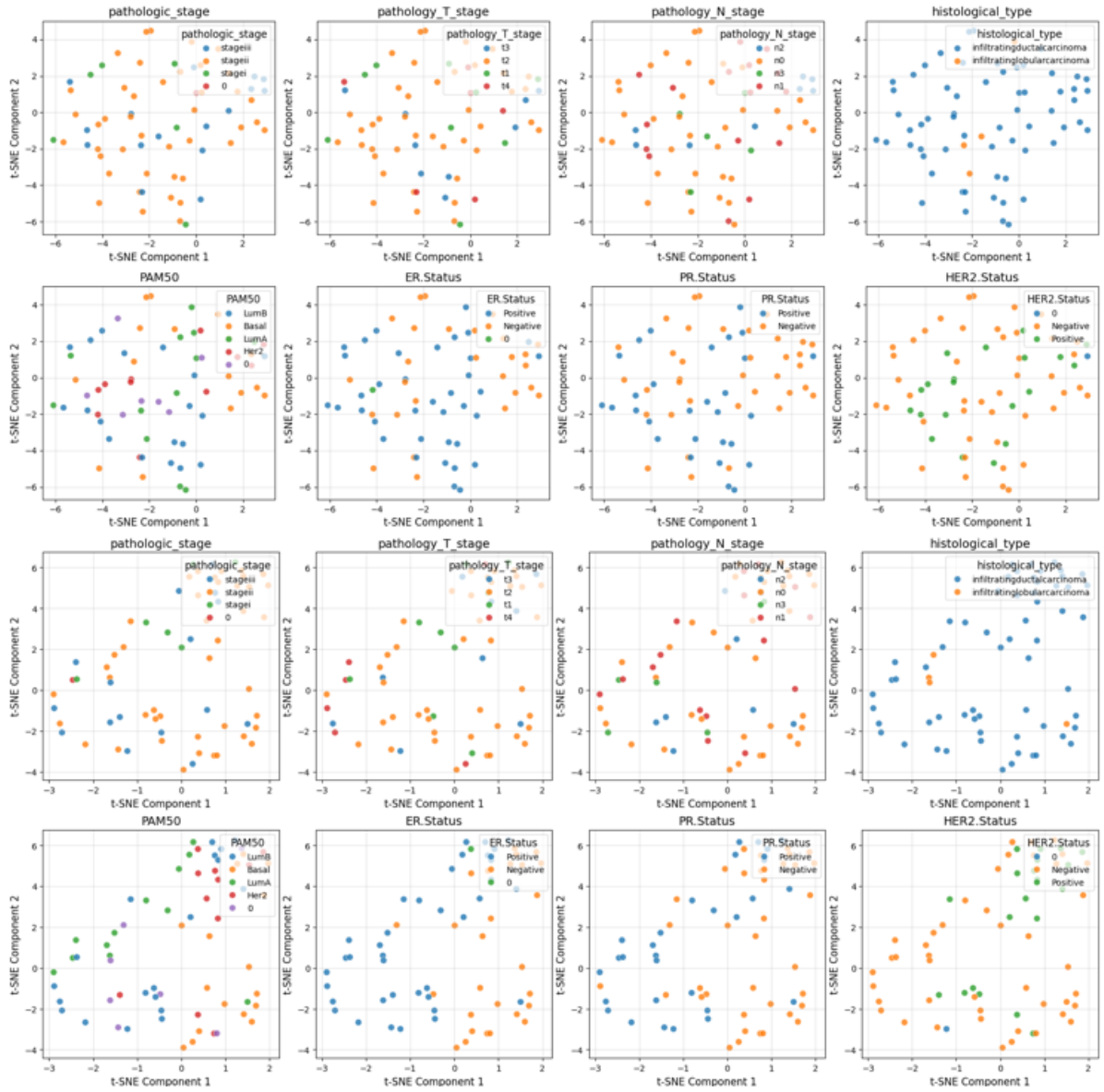

Fig. 4. *The labeled t-SNE distributions of samples in the latent spaces generated in our model (rows 1-2) and the MOFA model (rows 3-4) provide information on the correlation between the 8 categorical phenotype features and latent variance.*

reflects the model's robust training scheme. However, potential limitations can be identified. The cross-modality latent loss and the style disentanglement loss exhibit rapid early convergence, which is dangerous since they respectively mark the alignment and learning of across modality data and the separation of the orthogonal information in the content and style spaces. An overly rapid convergence of these crucial losses may indicate overfitting or failure of learning the underlying semantics in the data. Additionally, the unsatisfactory attention sparsity loss trend which is largely inverse to the entropy loss indicates failure of the two attention losses in imposing adversarial equilibrium. Finally, an observation of the end behaviors of several losses suggests that the number of training epochs can be even increased to further optimize the model, under which

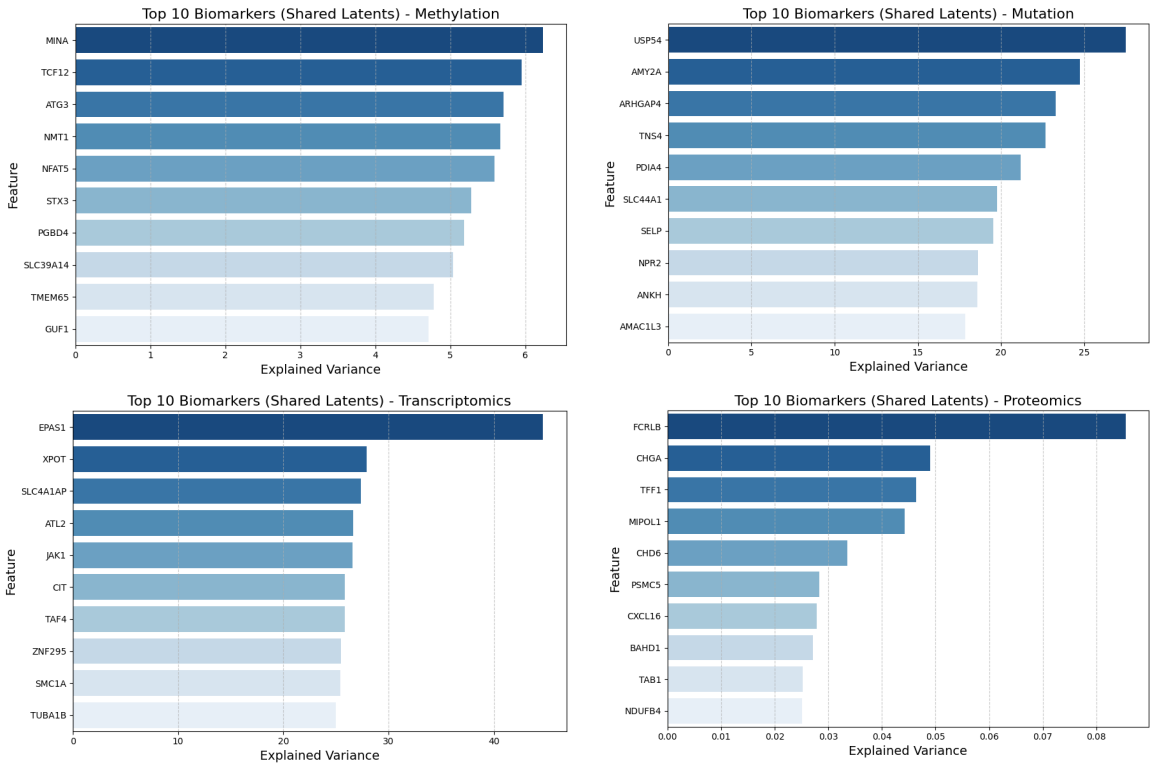

Fig. 5. *The top 10 biomarker candidates from each modality ranked by explained variance (unnormalized).*

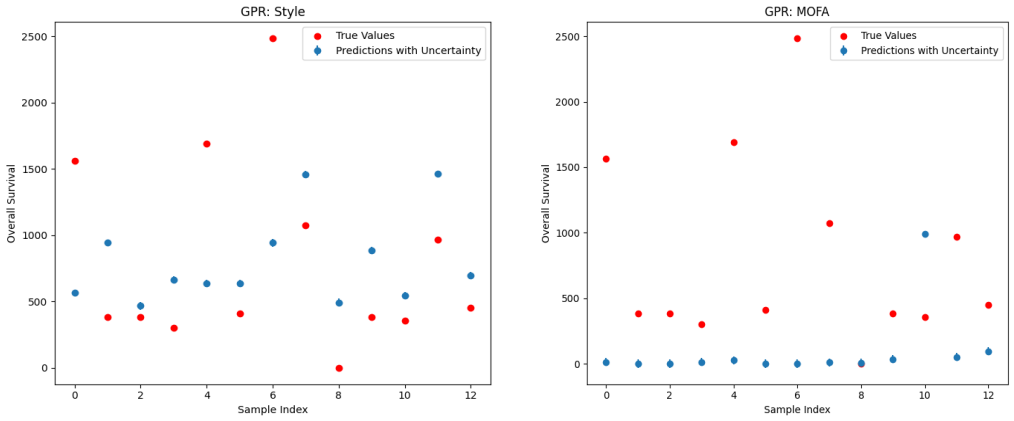

Fig. 6. *The predictive results on phenotype labels in both models using latent feature inputs and Gaussian Process regression.*

circumstance potential overfitting should however be addressed.

Variance decomposition of the obtained latent space reveals variability in the explainability of different latent factors (Fig. 2). Several factors deviate considerably from the average value of 0.0625, and no obvious degeneracy can be observed across factors and modalities. Notably, factor 8 shows a strong dominance in proteomics, and factor 18 a similar influence in transcriptomics, suggesting the capturing of key modality-specific patterns. However, most values in the

heatmap hover near the baseline, which suggests separation is likely insufficient. Note that the figures shown in this paper are constructed with results given by a hand-tweaked hyperparameter configuration that exhibits the most satisfactory performance among the ones tested. Other combinations, though not shown, provide certain insights. Interestingly, across experiments with altered $\lambda_{lat-sparsity}$, the final latent space sparsity, which can be directly inferred from the trend of $\mathcal{L}_{lat-sparsity}$ (Fig. 1), displays a negative correlation with respect to the sparsity in the explained variance map (Fig. 2). This suggests a complex interplay between the variance and the latent space that could potentially be explained by the reallocation of variance across latent factors. When higher sparsity is encouraged in the latent space, the model is constrained to use fewer active latent factors, resulting in the redistribution of variance towards other inactive factors to preserve explainability. This in theory indicates the presence of an optimal combination of both sparsity values to maximize the interpretability of the model. Therefore, further work could be done to investigate this interrelation.

Latent space visualizations in Fig. 3-4 directly evaluate the latent space's quality and have practical utility in identifying the biological meaning of latent factors. Additional to t-SNE analysis, direct latent factor separation is conducted (Fig. 3) on the 8 categorical phenotype labels in the phenotype dataset, including pathologic stage, T-stage, N-stage, histological type and several biomarker statuses. The latent representations resulting from the proposed method and the MOFA model present similar degrees of clustering in the corresponding labels, with MOFA achieving slightly more satisfactory separations between clusters ($p = 0.122$). Stronger clustering behaviors are observed in biomarker labels than pathologic stage labels. Among the biomarker statuses, the PR.status and ER.status labels are most clearly clustered and are most effectively explained by the same two latent factors in both models. The estrogen receptor (ER) and progesterone receptor (PR) are nuclear hormone receptors whose expression is closely linked. Estrogen signaling through ER directly induce the expression of PR, and both biomarkers are influenced by similar upstream factors, such as mutations or epigenetic changes in genes involved (e.g., ESR1 for estrogen receptor and related pathways). The successful results on this pair of biomarkers confirms a certain degree of capability of both models to learn interpretable patterns.

The t-SNE plots (Fig. 4) capture the general clusters between the datapoints using a Gaussian kernel with perplexity = 24. Unlike Fisher score analysis, t-SNE analysis takes solely the latent structure as its input and therefore qualitatively demonstrates clustering. Evidently, the MOFA model outperforms the proposed model in both the degree of clustering and the separation of phenotype labels in the neighbor space ($p = 0.060$).

A simulated **biomarker discovery** task is conducted with the latent space of the proposed model. For each modality, features are ranked by the reconstructability of their original values using the modality decoder, and the top 10 features identified are expected to have high correlation to other features in all modalities and are potential biomarkers for the cohort (Fig. 5). An advantage of this method is being able to run without phenotype labels, in which manner it is run in this experiment. The MINA and EPAS1 genes, ranked first in methylation and transcription omics data respectively, were previously confirmed to exhibit distinct expression patterns in various carcinomas. They are closely involved in carcinogenesis, responsible for histone demethylation and angiogenesis respectively. The biomarkers predicted for the other two modalities, FCRLB and USP54, are less understood by research yet were also shown to overexpress in several types of cancers. The USP54 deubiquitinase gene is possibly related to androgen receptor signaling and is a potential therapeutic target in castration-resistant prostate cancer. These results with biologically meaningful insights indicate a promising downstream application of the proposed model. However, quantitatively, the large gap of explained variance in transcriptomics and proteomics candidates is not directly interpretable and is likely instead stochastic, since repeated training of the model can lead to distinct results. Yet, some of the more prominent biomarkers in current knowledge, including the JAK1 gene (ranked fifth in transcriptomics in this

graph), are rarely detected as candidates. These concerns indicate the model's stability and interpretability could use further improvement.

To generalize evaluation to more indirect applications of the latent representation, the model performance in **phenotype prediction** of a continuous variable (overall_survival) in the phenotype dataset is examined (Fig. 6). A Gaussian Process regression model is used, and the train-test split ratio is set to 4:1. While both models' latent space exhibit unsatisfactory predictive performance, the MOFA latent space somehow does not encourage meaningful regression, even with altered hyperparameters. Conversely, the predictions based on the proposed model generally follows the mean values across samples although clustering within a narrower range, which is nevertheless more expected because the correlation between survival and omics data is usually of high variance and can contain outliers. The results in this continuous prediction task simulation show the promising robustness of the proposed model due to the introduction of nonlinearity and other complexities.

5 Conclusion

In this project, a novel method for bulk multi-omics integration inspired by style transfer, with the application of autoencoders and attention mechanisms, is presented. The proposed model demonstrated satisfactory performance in most downstream analysis tasks, indicating its high robustness and an overall high capability of latent representation.

From the implementation of the downstream tasks with the proposed method and the existing MOFA model, one can notice the high practicality and clinical utility of multi-omics analysis. Unlike traditional association studies which focus on the local cross-modality interactions, multi-omics integration offers a coherent global map and avoids the inherent biases in the selection of specific loci of interest as variation axes, and it is furthermore essential in circumstances where the axes of variation are not identified *a priori* [5]. The task therefore enables comprehensive analysis and discovery of novel biological patterns, shining light in fields such as precision medicine and systems biology. For example, biomarker discovery using the proposed model's latent representations provides opportunities for translational research in molecular or cell biological causes of diseases, and tasks including phenotype label prediction have great practicality as potential diagnostic and prognostic tools.

Simultaneously however, the model presents drawbacks in several aspects examined in this experiment. Firstly, the overall architecture lacks standardization and consolidation. The current incorporation of various machine learning mechanisms is neither fully organic nor optimized, resulting in a relatively complex model. This complicates the fine tuning of hyperparameters and general model schemes, leading to suboptimal training convergence. Secondly, the interpretability of the model latent representation is questionable. The model exhibits acceptable performance in tasks demanding downstream interpretation including biomarker discovery and phenotype prediction and yet is inferior to MOFA in latent space clustering and direct interpretability. This suggests the proposed model performs decently in dimensionality reduction and latent encoding yet insufficient cross-modality learning capabilities. This could be due to the void of assumption in the model as compared to the employment of Bayesian priors in MOFA or other explicit statistical assumptions in other existing models.

Future studies may explore the following: a) The model's application to integrating incompletely overlapping multi-omics data, also known as diagonal integration. b) The optimization and consolidation of the model, using more novel approaches to combat the curse of dimensionality. c) Combining style transfer methods with other neural-network-based approaches that enable the incorporation of prior knowledge, including graph neural networks (GNNs), which is commonly used in single-cell multi-omics integration.

6 Acknowledgement

We acknowledge LinkedOmics and TCGA for providing the data used in this study [8, 9]. The datasets and source code for the experiments are available in a repository on GitHub: https://github.com/Lihao-Liu2014/StyleOmics/ [10].

References

[1] M. S. Hossain, T. Joshi, and G. Stacey, System approaches to study root hairs as a single cell plant model: current status and future perspectives, *Front. Plant Sci.* **6**, (2015). https://doi.org/10.3389/fpls.2015.00363

[2] R. Argelaguet, A. S. E. Cuomo, O. Stegle, and J. C. Marioni, Computational principles and challenges in single-cell data integration, *Nat Biotechnol* **39**, 1202 (2021). https://doi.org/10.1038/s41587-021-00895-7

[3] Z. Wang, M. Gerstein, and M. Snyder, RNA-Seq: a revolutionary tool for transcriptomics, *Nat Rev Genet* **10**, 57 (2009). https://doi.org/10.1038/nrg2484

[4] Z. Miao, B. D. Humphreys, A. P. McMahon, and J. Kim, Multi-omics integration in the age of million single-cell data, *Nat Rev Nephrol* **17**, 710-724 (2021). https://doi.org/10.1038/s41581-021-00463-x

[5] R. Argelaguet, B. Velten, D. Arnol, S. Dietrich, T. Zenz, J. C. Marioni, F. Buettner, W. Huber, and O. Stegle, Multi-Omics Factor Analysis—a framework for unsupervised integration of multi-omics data sets, *Molecular Systems Biology* **14**, (2018). https://doi.org/10.15252/msb.20178124

[6] L. A. Gatys, A. S. Ecker, and M. Bethge, A Neural Algorithm of Artistic Style. https://doi.org/10.48550/arXiv.1508.06576

[7] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks. https://doi.org/10.48550/arXiv.1703.10593

[8] Vasaikar, S., Straub, P., Wang, J., & Zhang, B. (2018). LinkedOmics: analyzing multi-omics data within and across 32 cancer types. *Nucleic Acids Research*, **46**(D1), D956–D963. https://doi.org/10.1093/nar/gkx1090

[9] The Cancer Genome Atlas Research Network. (2013). Comprehensive molecular portraits of human breast tumours. *Nature*, **490**(7418), 61–70. https://doi.org/10.1038/nature11412

[10] Lihao Liu. (2024). StyleOmics. GitHub. Available at: https://github.com/Lihao-Liu2014/StyleOmics/