TENSORISED MODULAR ARCHITECTURES FOR MULTI-OMICS GENERATION

Arinbjörn Kolbeinsson
K01
Reykjavík, Iceland
arinbjorn@k01.is

Benedikt Kolbeinsson
K01
Reykjavík, Iceland
benedikt@k01.is

ABSTRACT

Multi-omics data has rich cross-modal biological structure. Standard flat architectures ignore this and treat the feature vector as flat and unstructured. We investigate a modular architecture approach where genes and proteins are grouped into biological modules with specific encoders. A coupling layer models higher-order inter-module interactions. We compare two coupling layer approaches: a dense matrix and a Tensor-Train decomposition, which naturally represents higher-order relationships between modules. Our findings show that modular architecture improves correlation metrics substantially over flat baselines at comparable parameter budgets. Further, our preliminary parameter efficiency experiments indicate that Tensor-Train reaches comparable performance with fewer parameters, a promising direction for capturing multi-omics relationships.

1 INTRODUCTION

Single-cell multi-omics data contains joint RNA and protein measurements for the same cells. A key challenge in modelling these data is to preserve the cross-modal biological relationships, such as RNA-protein correlates which are conditioned on cell type and specific interaction patterns between modalities. These multi-modal relationships are particularly important for generative models as they have applications in data augmentation, imputation and *in silico* perturbation where complete/holistic realism is critical. However, current approaches (e.g. VAEs like scVI (Lopez et al., 2018) and totalVI (Gayoso et al., 2021)) treat the data as unstructured: they concatenate feature vectors linearly.

Biological data has natural higher-order structure. Genes are grouped into co-expression modules (Langfelder & Horvath, 2008), while proteins form functional clusters (Barabási & Oltvai, 2004). The interactions between these modules are multi-way relationships that cannot be captured easily using a concatenated vector; they are naturally represented using a higher-order tensor. Tensor decomposition methods, such as CP, Tucker (Kolda & Bader, 2009), and Tensor-Train (Oseledets, 2011), offer a principled way of representing the different modalities as distinct factors, using far fewer parameters than equivalent dense layers (Novikov et al., 2015; Kossaifi et al., 2020). This setup allows for two key features: 1) a modular architecture that aligns with biological structure and 2) parameter-efficient decomposed tensor layers.

Our contributions are threefold. First, we encode each biological module separately, grouped by hierarchical co-expression, inspired by biologically structured NNs Ma et al. (2018); Seninge et al. (2021) but with an explicit tensorised coupling layer Novikov et al. (2015); Kossaifi et al. (2019). Second, our main finding is that the modular architecture is key as the biologically inspired modules are far better at capturing conditional correlation than flat architectures at comparable parameter budgets. Finally, tensor coupling layers show preliminary efficiency advantages over dense coupling, suggesting tensor methods merit further investigation for multi-omics.

2 METHOD

We describe a conditional generative architecture that decomposes the joint RNA–protein space into biologically meaningful modules and couples them via a structured tensor layer.

Modular architecture. The input to our baseline models is a feature vector of concatenated RNA (2000 HVGs) and ADT (Antibody-Derived Tags, 217 proteins). For biological module construction, we use hierarchical clustering on training data based on Ward linkage (Ward Jr, 1963) and correlation distance, inspired by co-expression module detection Langfelder & Horvath (2008). We have a per-module encoder that maps features to a fixed-dimensional embedding. The motivation here is that each encoder acts as a specialised module expert Jacobs et al. (1991) and learns specialised feature embeddings. This produces a $(K{+}L) \times d_e$ matrix of module embeddings.

Coupling layer. We introduce a coupling layer to model interactions between module embeddings. Here we test two different configurations: A) dense coupling, where embeddings are flattened to a vector, linear map applied and reshaped, resulting in $\mathcal{O}((K{+}L)^2 \cdot d_e^2)$ parameters. B) Tensor coupling, where instead of flattening the map is parameterised as a tensor-train decomposition Oseledets (2011) via BlockTT factorisation on the $(K{+}L, d_e)$ tensor shape Novikov et al. (2015); Kossaifi et al. (2019). It is important to note here that the parameters are initialised in the decomposed form, there is no operation to decompose a tensor. This tensor approach natively captures the higher-order module interactions using far fewer parameters.

Architecture. We use a dense autoencoder on a concatenated RNA+ADT input, with one hidden layer of width 22 and a latent bottleneck of size 32. The encoder uses a ReLU nonlinearity and a symmetric dense decoder reconstructs the original features. We use 50 gene clusters and 20 protein clusters, giving 70 modules, where each module gets its own linear encoder. Embeddings are then stacked into a $70 \times d_e$ representation. Coupling is then either dense or BlockTT, as described above, followed by a bottleneck of size 32 and symmetric per-group decoders to reconstruct the original features. In Figure 1, the modular dense model uses $d_e{=}2$ and the modular TT model uses $d_e{=}4, r{=}4$.

Training. We train all models using a modality-weighted mean squared error (MSE) loss: $(\text{MSE}_{\text{RNA}} + \text{MSE}_{\text{ADT}})/2$ to balance modalities and account for dimension mismatch (2000 vs 217). This is optimised with an Adam optimiser (Kingma & Ba, 2015), lr $= 10^{-3}$, batch size 256, early stopping (patience 20 epochs). 3 random seeds are reported for all results.

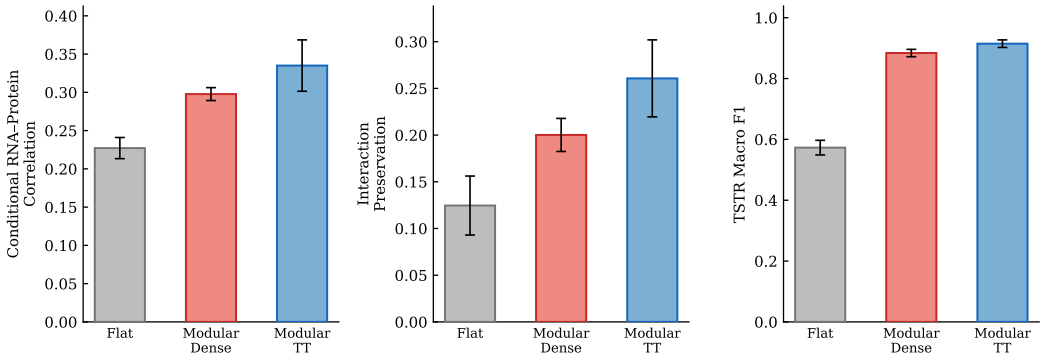

Figure 1: Modular vs flat architecture at comparable parameter budgets (3 seeds, 70 biological modules). Both modular variants dramatically outperform the flat baseline on all cross-modal metrics, with the modular TT model achieving +48% conditional correlation using fewer parameters.

3 EXPERIMENTS

Data and setup We test on the CITE-seq PBMC dataset Stoeckius et al. (2017) using 10,000 cells, 2000 RNA HVGs + 217 ADT surface proteins, 30 cell types (using Seurat v4 reference labels Hao et al. (2021)). Preprocessing consists of log1p RNA and CLR-normalisation of ADT. We first construct a stratified 80/20 train/test split, and for each training seed hold out 10% of the training partition for validation, yielding an effective 72/8/20 train/validation/test split. Results are reported on the test split. For cross-modal evaluation, we use 28 predefined RNA-protein pairs obtained by mapping ADT names to gene symbols and retaining pairs whose RNA gene is expressed in at least 2% of cells. After preprocessing, we end with K=50 gene modules and L=20 protein clusters, giving 70 biological groups.

Metrics. We select three metrics that capture correlation and conditional relationships between features. A) Conditional RNA-protein correlation (cond_r) measures how well the model preserves matched RNA-protein correlations across a predefined set of RNA-protein pairs, computed across cells within each cell type in synthetic data generated by the model compared with real data. This measures whether the model preserves biological signal conditioned on cell type. B) Interaction preservation (interact_r) measures correlation of pairwise cell-type contrast patterns between the modalities. This measures whether inter-cell-type interaction structure is preserved in the model. C) Train on synthetic test on real (TSTR) (Esteban et al., 2017) macro F1 score. This measures cell-type fidelity.

Modular architecture substantially improves key representation metrics. As shown in Figure 1, modular architecture dramatically outperforms the flat (dense) approach across all metrics. Parameter budgets were broadly comparable for this comparison (∼60–100K parameters). On conditional correlation, the modular TT architecture is +48% higher than the flat approach. On TSTR F1, the flat is substantially worse than either modular approach on preserving cell-type structure.

Tensor parameter efficiency. Our second set of results shows improved parameter efficiency of the tensor approach vs a modular dense approach. Tensor-train coupling consistently exceeds (or matches) dense coupling, Figure 2. These results are compared at similar parameter budgets, for example, at ∼195K params: TT achieves 0.380 vs Dense 0.348 on the conditional correlation metric. However, with only 3 repeated runs, the confidence bands overlap at several points. Both architectures preserve cell-type structure equally well, with $F1 > 0.90$ throughout the tested parameter range.

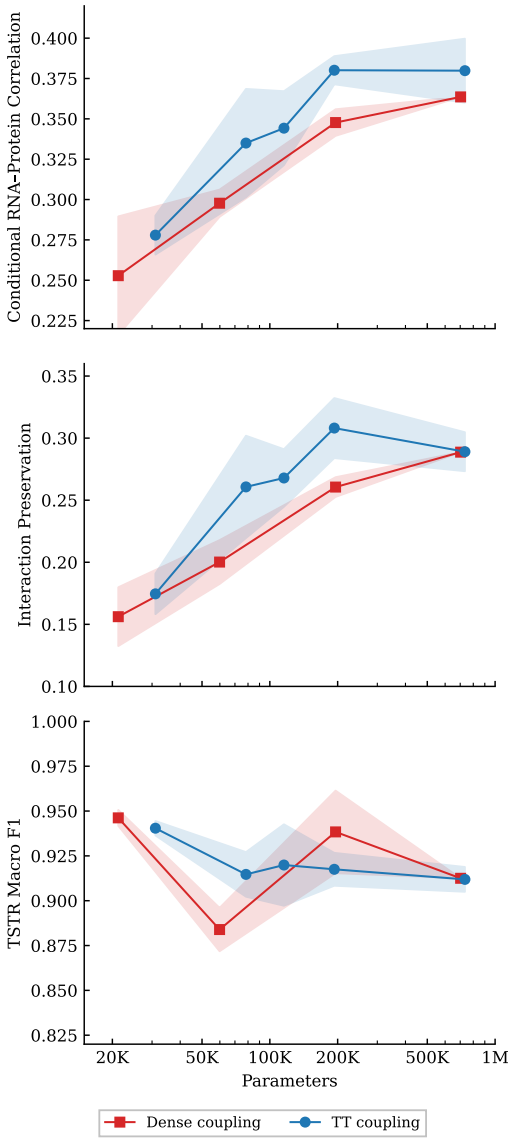

Figure 2: Parameter efficiency of TT vs dense coupling (70 biological modules). Shaded regions show ±1 std across available seeds. TT coupling matches or exceeds dense coupling across parameter budgets, with the clearest advantage in the 78K–193K range.

4 DISCUSSION

The key driver for cross-modal fidelity, in our findings, is to align the model architecture with biological module structure. This is consistent with prior work (Ma et al., 2018; Elmarakeby et al., 2021; Seninge et al., 2021). This can be viewed as an inductive bias that allows the model to learn module-specific representations and then natively learn higher-order cross-module interactions, rather than losing all the structure in a flat bottleneck. We find the result to be robust, holding for both coupling types tested.

Our second set of experiments show encouraging parameter efficiency and performance for the tensor coupling. The advantage is modest on this dataset and we hypothesise that the tensor structure would show larger gains on whole-genome-scale data, as the coupling matrix grows quadratically with the number of modes (Kolda & Bader, 2009). The tensor framing naturally allows for extension to even higher-order multi-omics (e.g. RNA + protein + chromatin accessibility) where interactions are even more acutely multi-way.

Although our results are promising, they are also limited by assumptions and computational resources devoted to the project. First, we test on only a single dataset and cannot confirm how generalisable our findings are to other datasets. Second, our module construction is only one way to perform biologically inspired clustering. This is a design choice and other options, such as different biological priors (pathway databases, gene ontologies Ma et al. (2018) or curated gene sets (Seninge et al., 2021; Lotfollahi et al., 2023)) would be an interesting alternative to try. We do not test random-group ablation, which would help determine whether gains arise from biological structure rather than modular decomposition alone. Our experiments only used 3 random seeds and we do not have a VEGA-style structured baseline. We have not performed complete hyperparameter optimisation of TT rank and module granularity. It is possible that further performance gains exist. Finally, our evaluation metrics only capture correlation, not causation.

5 CONCLUSION

Our findings demonstrate that encoding known biological module structure into generative model architecture substantially improves cross-modal fidelity, even at modest parameter budgets. This opens a path toward biologically grounded generative models for multi-omics that are both more faithful and more parameter-efficient. As single-cell atlases grow to encompass more modalities and cell types, we believe tensorised, modular architectures will become increasingly important for preserving the higher-order relationships that make these data biologically meaningful.

LLM DISCLOSURE.

An LLM tool was used as a coding assist to create boilerplate experimental scripts, configs and hyperparameter sweeps. All hypothesis generation, experimental decisions and interpretation of results was done by the authors. The paper was written by the authors without LLM assistance.

REFERENCES

Albert-László Barabási and Zoltán N Oltvai. Network biology: understanding the cell's functional organization. *Nature Reviews Genetics*, 5:101–113, 2004. doi: 10.1038/nrg1272.

Haitham A Elmarakeby, Jessica Hwang, Rand Arafeh, et al. Biologically informed deep neural network for prostate cancer discovery. *Nature*, 598:348–352, 2021. doi: 10.1038/s41586-021-03922-4.

Cristóbal Esteban, Stephanie L Hyland, and Gunnar Rätsch. Real-valued (medical) time series generation with recurrent conditional GANs. In *NIPS Workshop on Machine Learning for Health*, 2017.

Adam Gayoso, Zoë Steier, Romain Lopez, Jeffrey Regier, Kristopher L Nazor, Aaron Streets, and Nir Yosef. Joint probabilistic modeling of single-cell multi-omic data with totalVI. *Nature Methods*, 18:272–282, 2021. doi: 10.1038/s41592-020-01050-x.

Yuhan Hao, Stephanie Hao, Erica Andersen-Nissen, William M Mauck, Shiwei Zheng, Andrew Butler, Maddie J Lee, Aaron J Wilk, Charlotte Darby, Michael Zager, et al. Integrated analysis of multimodal single-cell data. *Cell*, 184(13):3573–3587, 2021. doi: 10.1016/j.cell.2021.04.048.

Robert A Jacobs, Michael I Jordan, Steven J Nowlan, and Geoffrey E Hinton. Adaptive mixtures of local experts. *Neural Computation*, 3(1):79–87, 1991. doi: 10.1162/neco.1991.3.1.79.

Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. In *International Conference on Learning Representations*, 2015.

Tamara G Kolda and Brett W Bader. Tensor decompositions and applications. *SIAM Review*, 51(3): 455–500, 2009.

Jean Kossaifi, Yannis Panagakis, Anima Anandkumar, and Maja Pantic. TensorLy: Tensor learning in Python. *Journal of Machine Learning Research*, 20(26):1–6, 2019.

Jean Kossaifi, Zachary C Lipton, Arinbjörn Kolbeinsson, Aran Khanna, Tommaso Furlanello, and Anima Anandkumar. Tensor regression networks. *Journal of Machine Learning Research*, 21 (123):1–21, 2020.

Peter Langfelder and Steve Horvath. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics*, 9:559, 2008. doi: 10.1186/1471-2105-9-559.

Romain Lopez, Jeffrey Regier, Michael B Cole, Michael I Jordan, and Nir Yosef. Deep generative modeling for single-cell transcriptomics. *Nature Methods*, 15:1053–1058, 2018. doi: 10.1038/s41592-018-0229-2.

Mohammad Lotfollahi, Sergei Rybakov, Karin Hrovatin, Soroor Hediyeh-Zadeh, Carlos Talavera-Lopez, Alexander V Misharin, and Fabian J Theis. Biologically informed deep learning to query gene programs in single-cell atlases. *Nature Cell Biology*, 25:337–350, 2023. doi: 10.1038/s41556-022-01072-x.

Jianzhu Ma, Michael K Yu, Samson Fong, Keiichiro Ono, Eric Sage, Barry Demchak, Roded Sharan, and Trey Ideker. Using deep learning to model the hierarchical structure and function of a cell. *Nature Methods*, 15:290–298, 2018. doi: 10.1038/nmeth.4627.

Alexander Novikov, Dmitry Podoprikhin, Anton Osokin, and Dmitry P Vetrov. Tensorizing neural networks. In *Advances in Neural Information Processing Systems*, volume 28, 2015.

Ivan V Oseledets. Tensor-train decomposition. *SIAM Journal on Scientific Computing*, 33(5):2295–2317, 2011. doi: 10.1137/090752286.

Lucas Seninge, Ioannis Anastopoulos, Hani Ding, and Joshua Stuart. VEGA is an interpretable generative model for inferring biological network activity in single-cell transcriptomics. *Nature Communications*, 12:5684, 2021. doi: 10.1038/s41467-021-26017-0.

Marlon Stoeckius, Christoph Hafemeister, William Stephenson, Brian Houck-Loomis, Pratip K Chattopadhyay, Harold Swerdlow, Rahul Satija, and Peter Smibert. Simultaneous epitope and transcriptome measurement in single cells. *Nature Methods*, 14:865–868, 2017. doi: 10.1038/nmeth.4380.

Joe H Ward Jr. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301):236–244, 1963.