Algorithms in Computational Biology - Scribes
Latent representation learning in biology and translational medicine

Bar Melinarskiy

June 4, 2023

1 Introduction

Latent representation learning refers to the process of discovering and learning hidden or abstract features that underlie the data in a system. In biology and translational medicine, this concept has been applied to a variety of contexts, including the analysis of genetic data, the identification of biomarkers, and the development of personalized medicine approaches.

Latent representation learning is strongly connected to disentanglement, which is a method developed originally for image processing, and uses latent representation learning and refers to the process of separating the factors that contribute to an image or video. This can be useful for various applications, such as image and video compression, image classification, and video analytics. Disentanglement can be achieved through various techniques, such as deep learning, feature extraction, and data compression algorithms. Disentanglement helps improve the efficiency and accuracy of image processing tasks by isolating the relevant features and reducing noise or redundancy in the data. It also allows for better interpretability of the results, as it separates the different factors that contribute to the image or video.

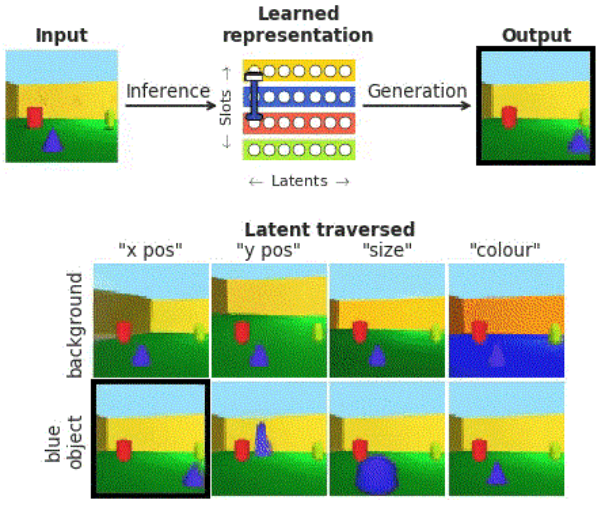

Figure 1: A toy example of disentanglement of an image. This is taken from DeepMind's latest neural network MONet. We can see how the interface perfectly separates each object from the background. Such skills are fully unlocked during unsupervised learning.

Fig 1. illustrates a toy example showcasing the disentanglement method applied to a simple input image. The image encompasses distinct shaping factors, such as a yellow background, green ground, and various shapes within it. These factors can be encoded individually using an encoder, which results in a learned latent space where each attribute is represented separately. Once we obtain the learned representation, we can manipulate individual or multiple attributes within this space. By utilizing a generator, we can then decode this matrix back to the image space and analyze the effects on the image. Through this process, we can enhance our understanding of how each shaping factor contributes to the overall topology of the image.

In medical imaging, disentanglement may be used to separate out different tissues or organs within an image, allowing doctors to more easily diagnose and treat patients. More generally, in computer vision, disentanglement may be used to separate out different objects within an image, allowing for better object recognition and classification. Overall, disentanglement is an important process in image processing, as it allows for more accurate and efficient analysis of images, leading to improved understanding and decision-making.

2 Generative models

Learning disentangled representations is regarded as a fundamental task for improving the generalization, robustness, and interpretability of generative models [1]. So in order for us to discuss this subject in depth we first need to give a brief reminder of what are Generative models. Generative models provide a well-established statistical framework for evaluating uncertainty and deriving conclusions from large data sets, especially in the presence of noise, sparsity, and bias. This learning approach was initially developed for computer vision and NLP tasks, including supervised learning tasks, such as assigning labels to images; unsupervised learning tasks, such as dimensionality reduction; and out-of-sample generation, such as de novo image synthesis [2]. As it often happens in science, a solution developed for one field is found to be applicable also for other fields that are not necessarily related to each other. Indeed the power of generative models is now being increasingly leveraged in molecular biology, with applications ranging from designing new molecules with properties of interest to identifying harmfulness mutations in our genomes and to dissecting transcriptional variability between single cells [2].

A classic example of a generative model is the one used to generate new digits, as can be seen in Fig 2. Using the dataset of handwritten digits, you could train a generative model to generate new digits. In the training phase, a loss function is incorporated to adjust the model's parameters to minimize a loss function and learn the probability distribution of the training set. Then, with the model trained, you could generate new samples. To output new samples, generative models usually consider a stochastic, or random, element that influences the samples generated by the model. The random samples used to drive the generator are obtained from a latent space in which the vectors represent a kind of compressed form of the generated samples.

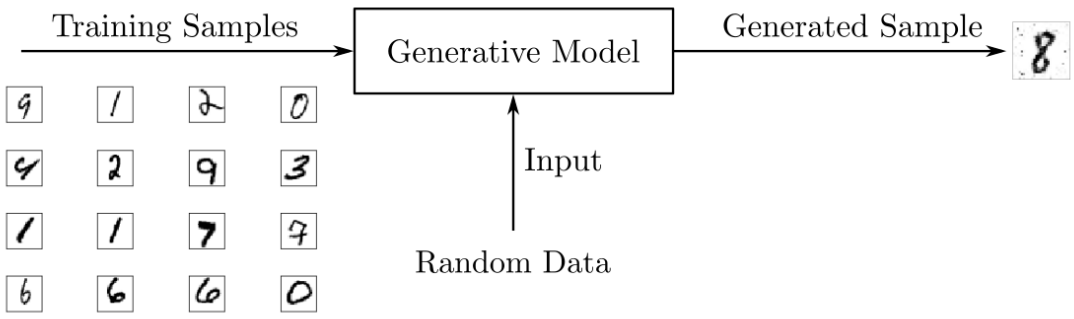

Figure 2: Using a data set of handwritten digits, you can train a generative model to generate new images.

Unlike discriminative models, generative models learn the probability $P(x)$ of the input data x, and by having the distribution of the input data, they're able to generate new data instances. Note: Generative models can also be used with labeled datasets. When they are, they're trained to learn the probability $P(X|Y)$ of the input x given the output y. They can also be used for classification tasks, but in general, discriminative models perform better when it comes to classification.

Generative models and disentanglement learning are closely connected in the field of machine learning. Generative models are used to create synthetic data that mimics the distribution of real data. Disentanglement learning, on the other hand, is a technique that aims to separate the underlying factors of variation in data.

By incorporating disentanglement learning into the training of generative models, we can create models that generate more meaningful and interpretable data. This is because disentanglement learning helps the generative model to capture the underlying factors of variation in the data, such as shape, color, texture, and pose, as separate and independent factors. This, in turn, allows us to control these factors and generate new data that is consistent with our desired changes. For example, we can change the pose of a generated object without changing its color or shape. Thus, the combination of generative models and disentanglement learning provides a powerful tool for creating realistic and controllable synthetic data for a variety of applications.

3 Disentanglement in image processing

Before diving into a review of a recent paper on a new disentanglement tool that incorporates a generative model let us first explore more the problem of disentanglement from an image-processing perspective. If we look at the world around us, it is easy to notice that each object is composed of various attributes that maintain complex relationships between them. Some of the features are permanent i.e. the class identity of the object, whereas others are transitory e.g. the pose of the object. Humans can often effectively separate between the class identity of the object, and the transitory pose of the object, even from a single observation. A key task for AI, often referred to as disentanglement, is to allow computers to mimic this ability and learn to separate between different attributes of observed data. There are multiple settings for disentanglement. The simplest is fully supervised - for each training image both the class and content are given as labels. A fully supervised scheme (e.g. deep encoders) may be trained to recover the class and content information from a single image. Conversely, a generative model can be trained to generate an image given input class and content information. on the other hand, fully unsupervised disentanglement takes as input a set of images with no further information. A successful unsupervised disentanglement algorithm will be able to learn a representation in which different factors of variation such as class and content will be represented separately. Fully

unsupervised disentanglement is highly ambitious and is work in progress, current methods typically do not produce consistently good results in this setting [3].

Let us review a recently published work from the lab of Yedid Hoshen here at the Hebrew University of Jerusalem. In their article, they revealed a new tool called LORD[4] which is a tool that deals with the class-supervised disentanglement task. The objective of the disentanglement task is to learn a representation containing all the information not available in the class label, denoted as content. In the case of faces, this content includes head pose, facial expression, etc. Firstly, it analyzes the information contained in the class and content representations.

A prominent approach in disentanglement research involves decomposing input images into latent variables that capture distinct properties such as geometry and style. This technique, known as Content-Style Disentanglement (CSD) [5], aims to represent images as domain-invariant "content" and domain-specific "style" representations [4]. In CSD, content is commonly encoded in spatial representations to preserve spatial correlations and exploit them for tasks like Image-to-Image translation and semantic segmentation. On the other hand, style information, governing image appearance such as color and intensity, is typically encoded in a vector. However, decomposing content from style is a nontrivial process, and relying solely on high-dimensional content representation is insufficient. Recent research introduces various design choices in terms of the model architecture and learning biases, encompassed as "building blocks," to achieve effective separation between content and style. These inductive biases are discussed further in the following section.

While current methods allow information to leak between the representations leading to imperfect disentanglement LORD ensures no information is leaked between the class and content representations. In order to achieve this goal, the training process requires optimization for every image (including at inference time). In the training set, a class embedding is shared across multiple images, which prevents the embedding from including content information. However, at inference time, a single image from an unknown class is observed. Optimizing over the latent codes for a single image leads to overfitting which results in entangled representations. Moreover, it requires iterative test-time inference since it does not perform amortized inference (Amortized inference is a technique in machine learning that uses a learned model to efficiently approximate complex inference tasks, reducing computational costs by sharing computations across multiple instances). This is obviously not ideal and resource-consuming so they have added a second stage which learns class and content encoders that directly infer class e_{yi} and content c_i representations from a single image x_i. The second stage effectively amortizes the results of the first stage and generalizes well to unseen classes and images. Meaning the second stage entails training the encoders $E_y : \mathrm{X} \to \mathrm{Y}$ and $E_c : \mathrm{X} \to \mathrm{C}$, which take as input an image x_i and output its class and content embeddings that were learned in the first stage. This is done while using reconstruction loss, to ensure the representations learned in the second stage must reconstruct the original image x_i. The general architecture of LORD can be seen in Fig 3.

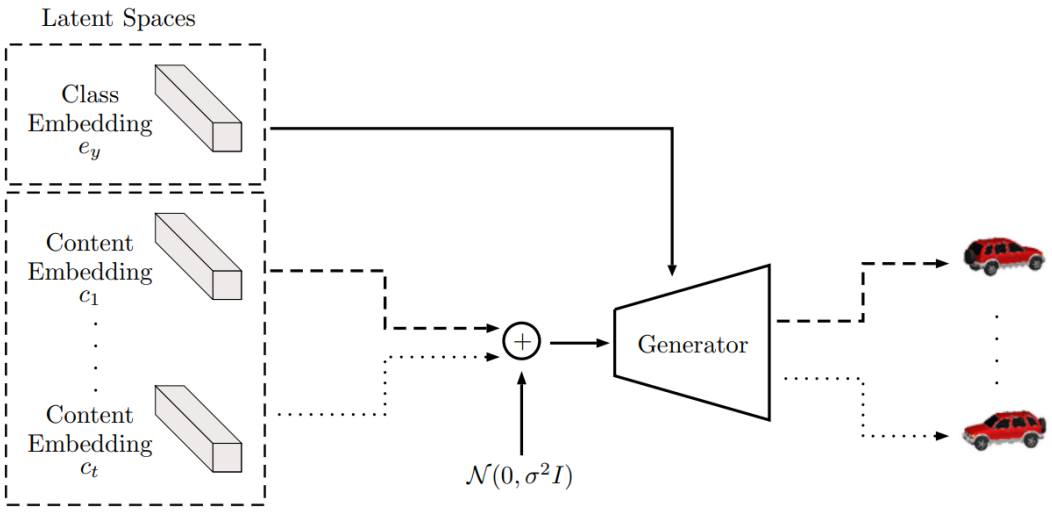

Figure 3: Figure taken from the work of Gabbay et al. [4] A sketch of the first stage: all class and content embeddings and the generator are jointly optimized. All images of the same class share a single class embedding. The content embeddings are regularized by Gaussian noise. By the end of this stage, the latent space of the training set is disentangled. Note that the second stage is not shown.

An example of a test case in which LORD achieved better results can be seen in Fig 4. The goal of this experiment is to achieve facial expression transfer. They compared their method against StarGAN [6] in the task of Multi-Domain translation. They followed the protocol in Choi et al. (2018) and compute the classification error of a facial expression classifier (trained on real images from RaFD) on synthesized images. They trained both models using the same training set and perform image translation on the same, unseen test set. As can be seen in Tab 1, LORD achieved lower classification error than StarGAN, indicating that it produces more realistic facial expressions without using adversarial training.

Table 1: Classification error (%)↓ on transferred facial expressions from RaFD.

StarGAN (Choi et al., 2018)	Ours	Real Images
2.2	**1.8**	0.8

Figure 4: Figure taken from the work of Gabbay et al. [4] A qualitative comparison between LORD (upper row) and StarGAN (bottom row) in facial expression transfer on RaFD.

Unlike previous approaches, they adopt a non-adversarial training strategy to achieve disentanglement between class and content. This approach offers significant optimization benefits and surprisingly achieves state-of-the-art performance without relying on adversarial constraints.

Their proposed approach utilizes shared latent optimization, asymmetric regularization, and a second amortization stage for single-shot generalization, resulting in effective class-supervised image disentanglement. Compared to both adversarial and non-adversarial disentanglement methods, they achieve state-of-the-art performance. Furthermore, they demonstrate the potential of extending their method by incorporating style clustering, enabling inter-class disentanglement for domain translation with promising results.

4 Applications of Latent representation learning in biology

One important application of latent representation learning in biology is in the analysis of genomic data. By identifying hidden patterns and structures in genetic data, researchers can gain insights into the underlying mechanisms of diseases and identify potential targets for therapeutic intervention. For example, latent representation learning techniques have been used to identify genetic markers that are associated with particular diseases or traits, such as the risk of developing a particular type of cancer.

Another area where latent representation learning has been applied in biology and translational medicine is in the identification of biomarkers. Biomarkers are measurable indicators of a particular biological process or disease that can be used to diagnose or monitor a condition. By learning the latent representations of different biomarkers, researchers can better understand the underlying mechanisms of diseases and identify potential targets for therapeutic intervention.

In addition to these applications, latent representation learning has also been used in the development of personalized medicine approaches. Personalized medicine involves tailoring treatment strategies to the individual needs and characteristics of each patient. By learning the latent representations of patient data, researchers can identify patterns and relationships that may be relevant to the development of personalized treatment plans.

In the following sections, I will dive deeper into some of these applications, providing a more comprehensive understanding of their significance and implications.

5 Out-of-distribution prediction with disentangled representations

Out-of-distribution prediction with disentangled representations is a technique used in machine learning to improve the ability of a model to predict the outcomes of input data that is outside of the training distribution. Disentangled representations refer to the separation of different features or variables in a dataset, allowing the model to better understand and predict the relationships between these variables. By training a model with disentangled representations, it is able to more accurately predict the outcomes of data that is outside of the training distribution, leading to better overall performance. This is particularly useful in real-world applications where the data encountered may vary significantly from the training data.

A recent article by Mohammad et al. [7] presents a study on disentangled representation learning for single-cell RNA sequencing (scRNA-seq) data. The authors employ a $\beta - VAE$ model (β variational autoencoder, a new state-of-the-art framework for automated discovery of interpretable factorised latent representations from raw image data in a completely unsupervised manner) [8] to learn disentangled representations and demonstrate its effectiveness in predicting gene expression for cell types not present in the training data. The $\beta - VAE$ model outperforms a state-of-the-art disentanglement method for scRNA-seq in both disentanglement performance and out-of-distribution (OOD) prediction. The authors also compare the $\beta - VAE$ model with a dHSIC-VAE (d-variable Hilbert-Schmidt Independence Criterion VAE) model [9] and show that the $\beta - VAE$ achieves better results. The study highlights the potential of disentangled representations for understanding biological variations and predicting cellular responses in scRNA-seq

data.

I add here an interesting short 5 minutes talk on of the paper's main researcher: Out-of-distribution prediction with disentangled representations for scRNA-seq data.

6 Diving into disentanglement in single-cell

My last example is taken from the very recently published article titled "Disentangling shared and group-specific variations in single-cell transcriptomics data with multiGroupVI"[10] that focuses on analyzing single-cell RNA sequencing (scRNA-seq) data, which is a technique that allows researchers to study the gene expression of individual cells. The goal of the study is to identify patterns in the data that are shared across different groups of cells, and patterns that are specific to each group. The authors propose a new method, called multiGroupVI, which is a deep generative model that can be used to analyze grouped scRNA-seq datasets. The method decomposes the data into shared and group-specific factors of variation. The authors validate the method on a simulated dataset and also apply it to an scRNA-seq dataset sampled from multiple regions of the mouse small intestine. They implemented multiGroupVI using the scvi-tools library and released it as open-source software.

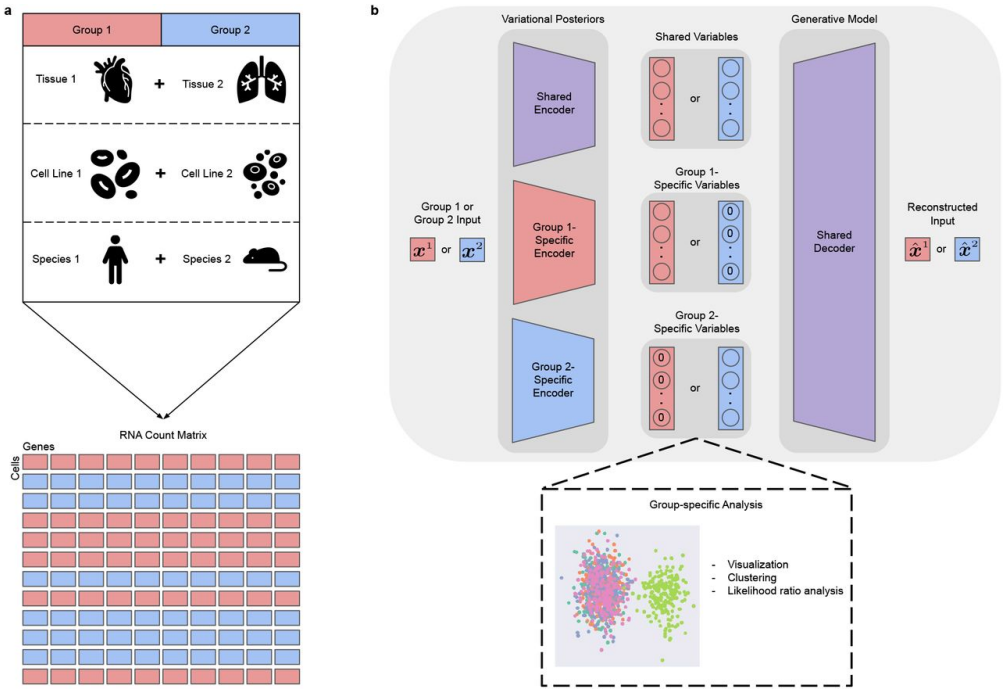

Figure 5: Figure taken from the work of Weinberger E. et al. [10] Overview of multiGroupVI. a, Given cells divided into non-overlapping groups of interest, multiGroupVI deconvolves the variations shared across groups versus those specific to individual groups. b, Schematic of the multiGroupVI model. A shared encoder network embeds cells, regardless of group membership, into the model's shared latent space, which captures variations shared across all groups. Group-specific encoders also embed cells into group-specific latent spaces, which capture variations particular to a given group. For a cell from a given group γ, the group-specific latent variables for other groups $\gamma' \neq \gamma$ are fixed to be zero vectors. Cells' latent representations are decoded back to the full gene expression space using a shared decoder. Here for simplicity we depict only two groups, though our model can easily be extended to handle more groups by adding additional group-specific encoders.

Existing methods for multi-group analysis have limitations, such as ad-hoc analysis or constraints that may not hold in practice. On the other hand, multiGroupVI explicitly decomposes scRNA-seq data into shared and group-specific factors of variation. The model incorporates parameter sharing and a novel regularization term based on optimal transport to encourage each set of latent variables to capture their corresponding variations. The proposed method offers greater flexibility and improved modeling compared to previous linear latent variable models.

In qualitative analysis, it is observed that multiGroupVI effectively distinguishes cells by cell type within its shared latent space (Fig 6d), displaying significant mixing across groups (Fig 6e). This indicates that the model successfully captures shared variations across the two groups in its shared latent space. Additionally, the separation of group-specific gene programs within its group-specific latent spaces is accurately achieved by multiGroupVI (Fig 6f-g). Collectively, these findings provide evidence that multiGroupVI demonstrates the ability to successfully disentangle shared and group-specific variations.

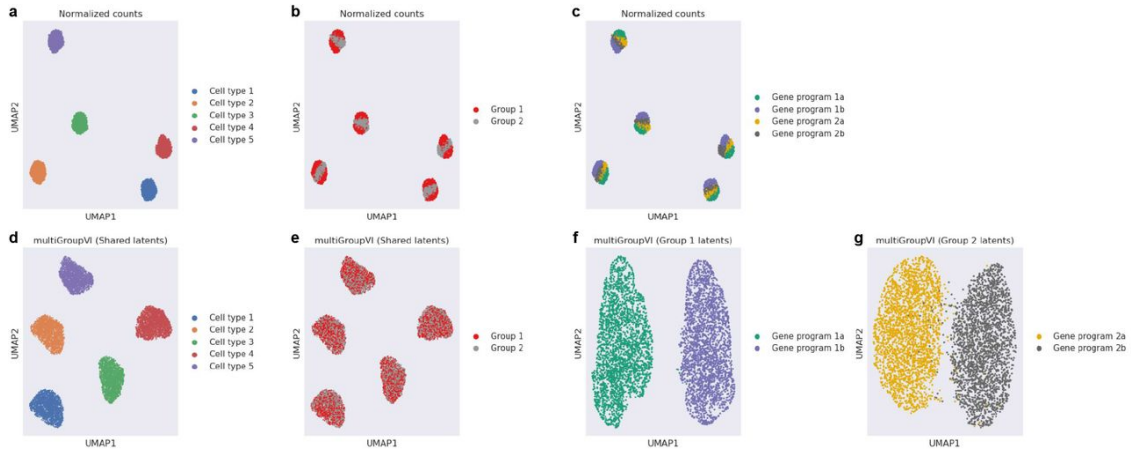

Figure 6: Figure taken from the work of Weinberger E. et al. [10] multiGroupVI correctly deconvolves shared and group-specific variations in a simulated dataset. a-c, Splatter [25] was used to generate a simulated scRNA-seq dataset. The dataset consisted of cells from five simulated cell types (a) divided into two groups (b) with each group having two group-specific gene programs (c). d-g, We find qualitatively that multiGroupVI correctly relegates cell-type-related variations to its shared latent space (d) with strong mixing across groups (e) while isolating group-specific variations in the group-specific latent spaces (f, g).

After they have proven the tool is effective on a simulated dataset they next explored its applications on a real-world scRNA-seq dataset from Haber et al. [11]. The dataset comprises 11,665 epithelial cells obtained from three distinct regions of the mouse small intestine: the duodenum, jejunum, and ileum. Although all regions contribute to the overall process of nutrient absorption, each region possesses specific functions. For instance, it is widely acknowledged [12, 13] that vitamin B12 absorption predominantly occurs in the ileum. To gain insights into these region-specific processes, it is essential to differentiate between variations in gene expression shared across regions and those specific to individual regions. In this study, multiGroupVI is trained using the region of origin as the group label for each cell, and the findings are presented in Fig 7. Initially, the cells exhibited distinct separation based on both region (Fig 7a) and cell type (Fig 7b). However, in the multiGroupVI shared latent space, there is observable mixing across regions (Fig 7c), with cells primarily separating based on cell type (Fig 7d). The region-specific multiGroupVI latent spaces were also examined (Fig 7e-g). It was observed that cell types tend to intermingle in the region-specific latent spaces, although notable exceptions exist. Notably, Paneth cells exhibited separation from other cell types in all region-specific latent spaces, indicating the presence of Paneth-cell-specific variations unique to each region, as also observed by [11]. Likewise, in the ileum and jejunum-specific latent spaces (Fig 7f), enterocytes displayed strong separation from other cell types, suggesting the existence of enterocyte-specific variations specific to those regions.

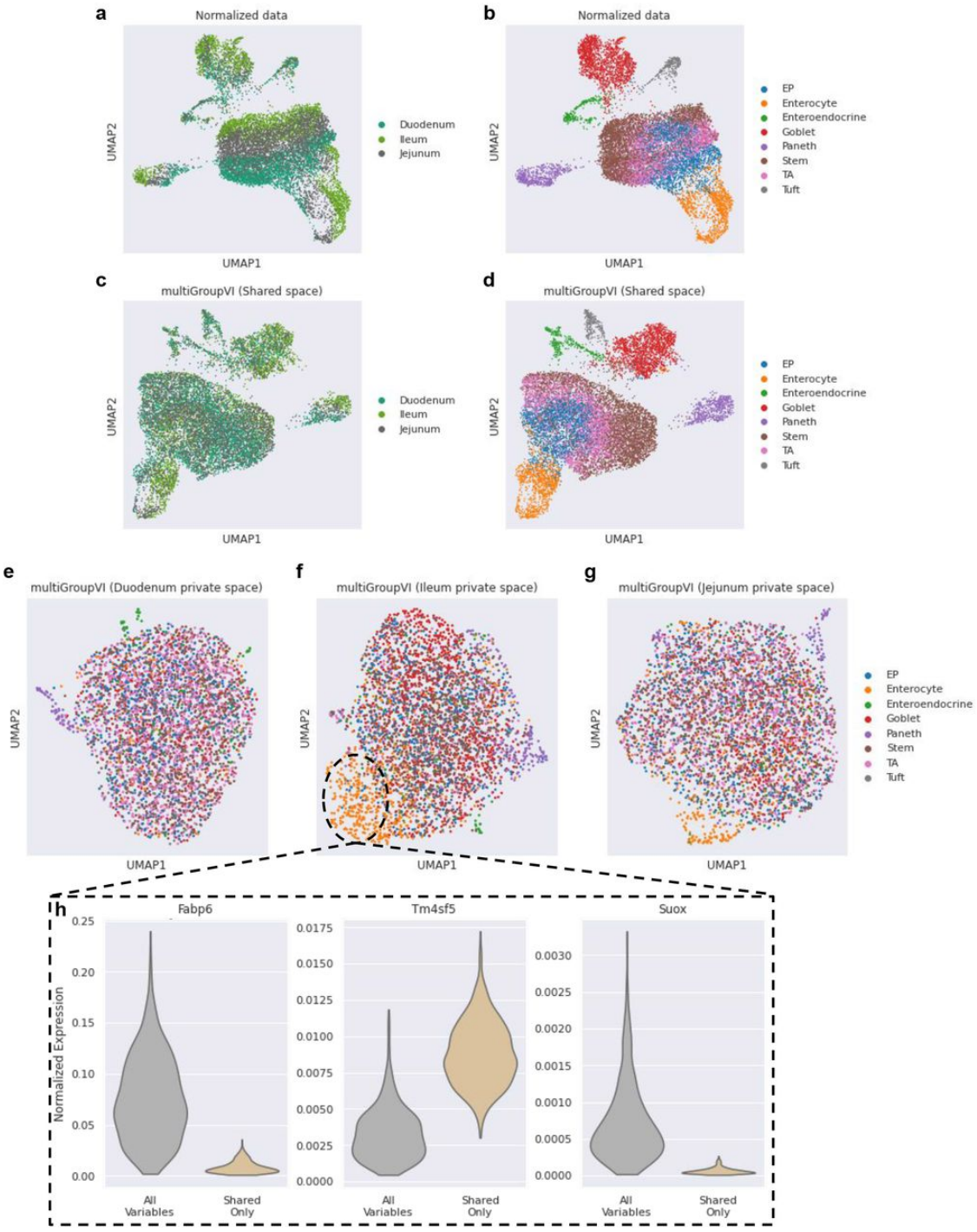

Figure 7: Figure taken from the work of Weinberger E. et al. [10] Exploring the multi-region mice intestine epithelial cell dataset from Haber et al. [11] with multiGroupVI. a-b, The original data (i.e., normalized and log-transformed counts) strongly separated by both region (a) and cell type (b). c-d, In the multiGroupVI shared latent space we observe stronger mixing across regions (c) while separation between cell types is preserved (d). e-g, In the region-specific multiGroupVI latent spaces we observe some separation between cell types, indicating that region-specific variations are dependent on cell type. For example, we find that enterocytes strongly separate from other cell types in the ileum-specific latent space. h, Distributions of normalized expression values for ileum enterocytes as computed by the multiGroupVI decoder when decoding using all latent variables versus only the shared variables for the top three genes ranked by likelihood ratio for ileum enterocytes.

The initial findings presented in their manuscript indicate that multiGroupVI can indeed effectively deconvolve shared and group-specific variations compared to previously proposed methods. However, interpreting the variations captured by the different multiGroupVI latent spaces remains a challenge. To address this, a procedure was introduced for identifying genes with strong group-specific effects, although limited to single genes rather than coordinated gene programs that may hold greater biological significance. Additionally, the procedure focused solely on identifying genes with group-specific effects and did not provide further insights into the trends captured in the model's shared latent space. They state that their future work aims to enhance the multi-group architecture by incorporating interpretable VAE models, such as BasisVAE, which can identify clusters of coordinated features represented by the learned latent variables while performing dimensionality reduction.

7 Summary

I tried to give an overall exploration of the applications of latent representation learning in the fields of biology and translational medicine. Latent representation learning involves uncovering hidden or abstract features within data systems. In the context of biology, this concept has found utility in various areas, including genetic data analysis, biomarker identification, and personalized medicine development. Specifically, I focused on the connection between latent representation learning and disentanglement. This approach enhances efficiency, accuracy, and interpretability by isolating relevant features and reducing noise. Overall, latent representation learning holds promise for advancing biological and medical research by revealing hidden patterns and facilitating a better understanding of complex data.

References

1. Sharon Zhou *et al.* Evaluating the Disentanglement of Deep Generative Models through Manifold Topology. *Published as a conference paper at ICLR 2021*. https://openreview.net/pdf?id=djwS0m4Ft_A (2021).
2. Lopez, R., Gayoso, A. & Yosef, N. Enhancing scientific discoveries in molecular biology with deep generative models. *Molecular Systems Biology* **16.** ISSN: 1744-4292 (Sept. 2020).
3. Locatello, F. *et al.* Challenging Common Assumptions in the Unsupervised Learning of Disentangled Representations (Nov. 2018).
4. Gabbay Aviv and Hoshen Yedid. Demystifying inter-class disentanglement. *International Conference on Learning Representations (ICLR)* (2020).
5. Gatys, L. A., Ecker, A. S. & Bethge, M. *Image Style Transfer Using Convolutional Neural Networks* in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, June 2016), 2414–2423. ISBN: 978-1-4673-8851-1.
6. Choi, Y. *et al. StarGAN: Unified Generative Adversarial Networks for Multi-domain Image-to-Image Translation* in *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition* (IEEE, June 2018), 8789–8797. ISBN: 978-1-5386-6420-9.
7. Mohammad Lotfollahi, e. a. Out-of-distribution prediction with disentangled representations for single-cell RNA sequencing data. *bioRxiv* **458535.** https://doi.org/10.1101/2021.09.01.458535 (Sept. 2021).
8. Higgins, I. *et al. beta-vae: Learning basic visual concepts with a constrained variational framework* in *International conference on learning representations* (2017).

9. Lopez, R., Regier, J., Jordan, M. I. & Yosef, N. *Information Constraints on Auto-Encoding Variational Bayes* in *Advances in Neural Information Processing Systems* (eds Bengio, S. *et al.*) **31** (Curran Associates, Inc., 2018). https://proceedings.neurips.cc/paper_files/paper/2018/file/9a96a2c73c0d477ff2a6da3bf538Paper.pdf.
10. Weinberger, E., Lopez, R., Hütter, J.-C. & Regev, A. Disentangling shared and group-specific variations in single-cell transcriptomics data with multiGroupVI. *bioRxiv,* 2022.12.13.520349. http://biorxiv.org/content/early/2022/12/15/2022.12.13.520349.abstract (Jan. 2022).
11. Haber, A. L. *et al.* A single-cell survey of the small intestinal epithelium. *Nature* **551,** 333–339. ISSN: 0028-0836 (Nov. 2017).
12. Booth, C. & Mollin, D. THE SITE OF ABSORPTION OF VITAMIN B12 IN MAN. *The Lancet* **273,** 18–21. ISSN: 01406736 (Jan. 1959).
13. Drapanas, T. Role of the Ileum in the Absorption of Vitamin B12 and Intrinsic Factor (NF). *JAMA: The Journal of the American Medical Association* **184,** 337. ISSN: 0098-7484 (May 1963).