Variational autoencoders with flexible priors enable robust distribution learning on single-cell RNA sequencing data

Leander Dony [1 2 3 4] **Martin König** [5] **David S. Fischer** [1 3] **Fabian J. Theis** [1 4 5]

Abstract

Generative modeling in single cell transcriptomics allows the efficient construction of latent spaces for denoising, batch-effect removal and prediction of experimental perturbations. To obtain biologically informative latent representations, recently established methods, however, rely on adapting the variational autoencoder (VAE) loss through down-weighting the Kullback-Leibler divergence term. These adaptations can limit the model's ability to learn the underlying data distribution. Here, we adapt two enhanced VAE architectures to the scRNA-seq setting which do not require tuning the loss: (i) a VAE with inverse autoregressive flow (IAF) and (ii) a VAE with a Variational Mixture of Posteriors (VAMP) prior. We assess the models' ability to learn biologically informative embeddings using four metrics in a large-scale comparison on 16 public scRNA-seq datasets from 9 tissues with over 700,000 cells. We find that in particular the VAE with a VAMP prior is capable of learning biologically informative embeddings without compromising on generative properties. This suggests that the VAE-VAMP is a useful starting point for improved generative modelling of scRNA-seq data.

1. Introduction

Generative modelling tools are becoming increasingly popular for a range of tasks in the analysis of scRNA-seq data. This includes visualisation, clustering (Grønbech et al., 2020), batch-correction, differential expression analysis (Lopez et al., 2018) as well as modelling drug perturbation and out-of-sample prediction (Lotfollahi et al., 2019b;a). Such models generally employ a VAE-based generative framework to learn the latent distribution of the data (Kingma & Welling, 2014). Compared to non-generative autoencoder (AE)-based modelling of scRNA-seq data (Eraslan et al., 2019), VAE-based models face difficulties with generating biologically meaningful embeddings. This is likely caused by the poor match of the unimodal prior to inherently mulitmodal scRNA-seq data. To overcome these problems, all previously mentioned generative tools use a modified loss-function during training. Either the KL-term of the loss function is scaled down by a constant factor between $5e{-}5$ and $5e{-}7$ (Lotfollahi et al., 2019b;a), or the scaling constant is linearly increased over training, starting from a default value of $2.5e{-}3$ (Lopez et al., 2018). In our study, the contribution of the unscaled KL-term to the total loss was 0.5 – 1.0 %, and reducing this further can lead to unwanted effects. In particular, with too little contribution of the KL-term, regularisation through the prior is not enforced anymore. One would therefore no longer sample from the learned data distribution when sampling the prior. A generative model which is able to learn a good posterior while conserving biological information in its latent space would hence be of great help for modelling scRNA-seq data.

Many approaches to improving the VAE-framework have been previously suggested. Here, we evaluated two of them for their ability to produce biologically informative latent representations. The first VAE-adaptation uses inverse autoregressive flows (IAF) to learn more flexible posterior distributions (VAE-IAF) (Kingma et al., 2016; Boyeau et al., 2019). The second model introduces a "Variational Mixture of Posteriors" (VAMP) prior for learning richer latent representations of the data (VAE-VAMP) (Tomczak & Welling, 2018). We adapted both models to fit the negative-binomial noise distribution found in droplet-based scRNA-seq data (Svensson, 2020). We evaluated both models for four different properties: (i) solving the KL over-regularisation problem, (ii) conserving biological variation in the latent representation, (iii) goodness of fit to the data and (iv) learning compact latent representations. We based our evaluation on data from 9 tissues with a total of 720 thousand cells.

[1]Institute of Computational Biology, Helmholtz Center Munich, Neuherberg, Germany [2]Department of Translational Psychiatry, Max Planck Institute of Psychiatry, Munich, Germany [3]School of Life Sciences Weihenstephan, Technical University of Munich, Munich, Germany [4]International Max Planck Research School for Translational Psychiatry, Max Planck Institute of Psychiatry, Munich, Germany [5]Department of Mathematics, Technical University of Munich, Munich, Germany. Correspondence to: Fabian Theis <fabian.theis@helmholtz-muenchen.de>.

Presented at the ICML 2020 Workshop on Computational Biology (WCB).

2. Methods

We adapt three generative neural-network architectures in this study: a vanilla VAE, (Kingma & Welling, 2014), a VAE with inverse autoregressive flow (Kingma et al., 2016) and a VAE with a VAMP prior (Tomczak & Welling, 2018). As a non-generative reference model, we employ an AE with a negative-binomial noise model (Eraslan et al., 2019). We use the same negative-binomial reconstruction objective to adapt the three generative models to better fit scRNA-seq data.

The reconstruction loss (RL) of all our models therefore corresponds to the sum of the negative log-likelihoods of the negative-binomial distribution ($-\mathcal{L}_{NB}$), over n cells and p genes, parameterized by the learned mean and dispersion parameters (μ, θ), given the input data $\mathbf{X}$:

$$RL\left(\mathbf{X}, \mathbf{M}, \mathbf{\Theta}\right) = \sum_{i=0}^{n} \sum_{j=0}^{p} -\mathcal{L}_{NB}\left(\mu_{i,j}, \theta_{i,j} | x_{i,j}\right) \quad (1)$$

While training of the AE optimises RL, training of the VAE optimises the Evidence Lower Bound (ELBO) (Kingma & Welling, 2014). The VAE training objective is hence defined as: $\hat{\mathbf{M}}, \hat{\mathbf{\Theta}} = \arg\min_{\mathbf{M},\mathbf{\Theta}} \left(RL\left(\mathbf{X}, \mathbf{M}, \mathbf{\Theta}\right) + D_{\mathrm{KL}}(q(\mathbf{z}|\mathbf{x})||p(\mathbf{z})\right)$

The VAE-VAMP introduces a more flexible formulation of the prior compared to the Gaussian prior used in standard VAE models (Tomczak & Welling, 2018). The prior uses an adapted form of the aggregated posterior and is defined based on K learnable pseudo-inputs: $p(\mathbf{z}) = \frac{1}{K} \sum_{k=1}^{K} q(\mathbf{z}|\mathbf{u}_k)$. $K = 500$ was used in this work.

For all models, we mapped each dataset input gene space to the GRCh38 Ensembl97 human genome (protein-coding only), replacing any missing expression values with zeros. We also introduced an additional layer at the input of each model which normalised all counts to 10000, followed by a $\ln(1 + x)$ operation.

We trained all models using Tensorflow 2.0 with the following model architecture: (512, 256, 128, 256, 512). We used a $tanh$ activation after every dense layer except the bottleneck and the last decoder layer, scaled and centered batch-normalisation and dropout (rate: 0.2) after each non-linearity. We used L1 and L2 regularization (5e-4 each), a learning rate of 5e-5 (Adam optimizer) with 50 % learning rate decay after 10 epochs without improvement of the validation loss. We used early stopping with a patience of 100 epochs and held-out a random selection of 10 % validation data, and 10 % test data from training. We used 5 flow layers in the VAE-IAF model. VAE-IAF and VAE-VAMP training took roughly twice as long as AE and VAE training.

For data points $\mathbf{x}^{(n)}$ and a latent space $\mathbf{z}^{(n)} \in \mathbb{R}^D$, we use the activity metric introduced by Burda et al. (2016), to quantify the empirical variance of the expected latent space:

$$A_{\mathbf{z}} = \mathrm{diag}\left(\mathrm{Cov}_{\mathbf{x}}\left(\mathbb{E}_{q(\mathbf{z}|\mathbf{x})}[\mathbf{z}]\right)\right) \quad (2)$$

A latent unit i is considered active if $A_{z_i} > 0.01$.

We use the average silhouette width (ASW) (Rousseeuw, 1987) to measure the how well cells of the same cell-type group together in our embeddings. For the mean intra-cell-type distance a and the mean nearest-cell-type distance b for each sample: $ASW = (b - a) / (\max(a, b))$.

Table 1. Comparison metric results computed on full dataset: First value: number of active units in the model bottleneck (Eq. 2); Second value: minimum number of principal components (PCs) required to explain 95 % of the variance in the latent space PCA. Bold: best performance across the generative models (VAE, IAF, VAMP). Abbreviations as in Fig. 1

ORGAN	AE	VAE	IAF	VAMP
BLOOD	128; 36	12; 11	**128**; 72	90; **10**
BONE	128; 39	15; 13	**128**; 61	80; **11**
COLON	128; 09	**128**; 53	**128**; 22	**128**; **9**
ESOPHAGUS	128; 35	14; 13	**128**; 47	**128**; **11**
KIDNEY	127; 39	23; 19	**128**; 31	108; **17**
LIVER	128; 37	21; 15	**128**; 40	123; **11**
PANCREAS	128; 41	**128**; 23	**128**; 29	**128**; **11**
PLACENTA	128; 37	**128**; 17	**128**; 34	**128**; **9**
SPLEEN	128; 71	15; 14	**128**; 57	115; **11**

3. Results

3.1. VAE-VAMP and VAE-IAF models alleviate the inactive latent unit problem of vanilla VAEs for scRNA-seq data

A common problem with using vanilla VAEs for representation learning is the inactivity of a large number of latent space units. This can dramatically reduce the biological variation captured by a latent space embedding generated with such a model (cp. Spleen embedding Fig. 1). VAE-VAMP models have previously been shown to alleviate this problem by regularising the latent space with a richer prior than the standard Gaussian one used in vanilla VAEs (Tomczak & Welling, 2018). To investigate whether this also holds for our adapted VAE-VAMP model on scRNA-seq data, we computed the empirical variance of the expected latent space (Eq. 2) for each model. As previously observed by Burda et al. (2016), we found that in vanilla VAE models, less than 20 % of the latent units are active in more than half the models we trained (Table 1). In the VAE-VAMP model this problem is dramatically reduced while it does not occur at all for the VAE-IAF model. Additionally, the covariance of the latent spaces (Fig. 2) is higher in the VAE-VAMP model. These results suggest, that the VAE-IAF and VAE-VAMP model alleviate the issue of inactive

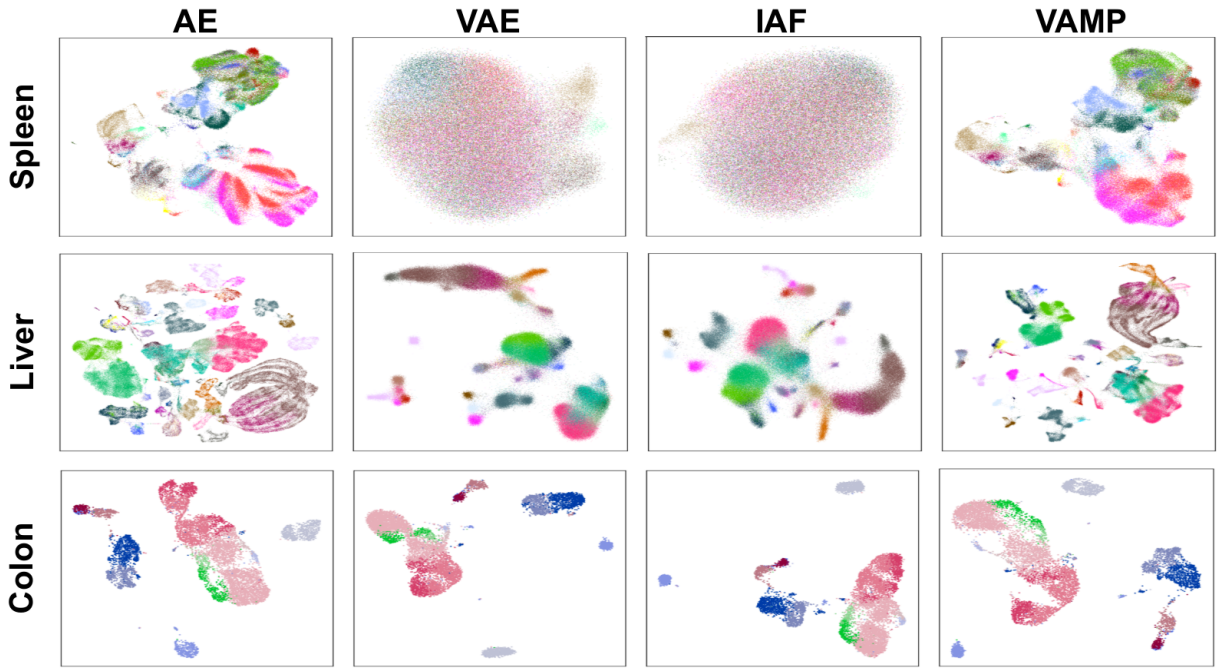

Figure 1. UMAP (McInnes et al., 2018) visualisations of the model bottleneck. Colours represent biological cell-type. AE: autoencoder, VAE: variational autoencoder, IAF: VAE with inverse autoregressive flow, VAMP: VAE with a Variational Mixture of Posteriors prior.

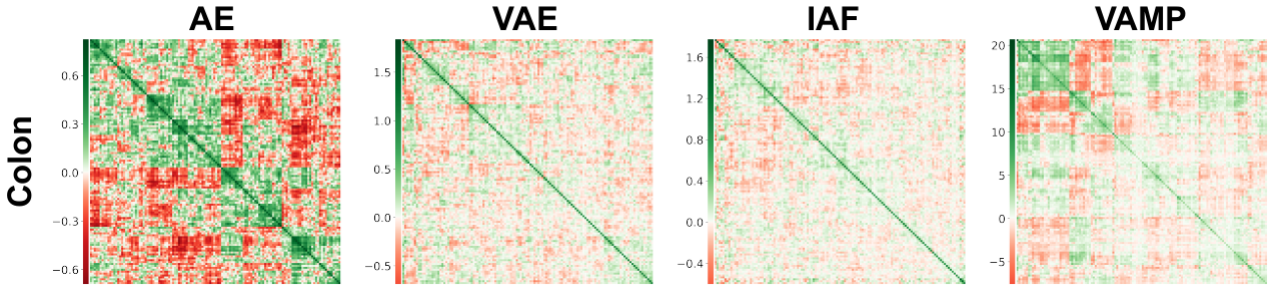

Figure 2. Latent-space covariance matrices for Colon data. A similar trend was observer for data from other organs. Note the different colour scales. Abbreviations as in Fig. 1

units. Furthermore, the VAE-VAMP model is the only generative model that is allows for covariance between its latent units which might enable the VAE-VAMP model to learn meaningful embeddings more efficiently.

3.2. VAE-VAMP models robustly generate more biologically informative latent space embeddings from scRNA-seq data

One main purpose of generating latent-space representations of high-dimensional scRNA-seq data is to distil biologically relevant features from a very large input space. A criterion for a useful latent representation would hence be its ability to separate biologically distinct cell-types while maintaining proximity between cells from the same biological cell-types. To assess this ability in the studied models, we computed the average silhouette width (ASW) with respect to the cell-type label for each embedding. A high ASW indicates an embedding where cells of each cell-type are tightly clustered together while being well separated from cells of other cell-types. We also visually assessed UMAP (McInnes et al., 2018) visualisations of the learned latent representations. We found that while for some organs all models are able to conserve biological variation in the embedding (cp. Colon embedding Fig. 1), for other organs the VAE-VAMP model is the only generative one that is able to produce finely-resolved embeddings (cp. Spleen and Liver embedding Fig. 1). This is also reflected in the ASW (Table 2). These results indicate that our adapted VAMP-VAE model is able to learn latent representations which are competitive with embeddings learned by conventional AE-based embedding models while retaining the generative properties of VAEs.

Table 2. Comparison metric results for all four evaluated models: (i) Reconstruction loss (Eq. 1) of the test-data; (ii) Average Silhouette Width (ASW) over all samples from full dataset with respect to provided cell-type labels, no cell-type labels were available for Blood and Bone. Bold: best performance across the generative models (VAE, IAF, VAMP). Italics: AE shows better performance than all generative models. Abbreviations as in Fig. 1

	Reconstruction Loss				ASW for Cell-Type			
Organ	AE	VAE	IAF	VAMP	AE	VAE	IAF	VAMP
Blood	*0.125065*	0.126222	**0.125696**	0.125976	–	–	–	–
Bone	*0.139278*	0.140641	0.141112	**0.139852**	–	–	–	–
Colon	*0.309301*	0.318625	0.314449	**0.313169**	*0.120995*	**0.103665**	0.092223	0.082633
Esophagus	*0.287229*	**0.288587**	0.288752	0.289268	*0.103518*	-0.011824	-0.010275	**0.090486**
Kidney	0.258964	0.260199	**0.257875**	0.260076	*0.004963*	-0.019341	-0.027443	**-0.010113**
Liver	*0.313488*	0.315088	**0.314222**	0.314506	*0.146179*	-0.002620	-0.006442	**0.064773**
Pancreas	1.675346	1.688566	1.694367	**1.647528**	*0.075400*	**0.070587**	0.032181	0.063263
Placenta	0.408710	0.414898	**0.407160**	0.413223	*0.193912*	0.051128	0.018597	**0.116071**
Spleen	*0.229657*	0.231963	**0.230486**	0.233184	0.021080	-0.009836	-0.017153	**0.027596**

3.3. VAE-VAMP and VAE-IAF models are able to fit scRNA-seq datasets better than vanilla VAEs

Besides various regularisation terms, the reconstruction loss (Eq. 1) is the core part of the loss function in any AE-type learning framework. The lower the reconstruction loss, the better the model has fitted the data. Generally, AEs tend to fit the data better due to the absence of competing regularisation terms in the loss function as present in any VAE-based model. We compared the reconstruction loss of all four models (Table 2) and found that, as expected, in most cases the AE model was able to achieve the lowest reconstruction loss on the test-set. Among the VAE-based models, the VAE-IAF and VAE-VAMP outperformed the vanilla VAE for data from all but one human organ. This indicates that both approaches of improving the VAE - more flexible priors and more flexible posteriors - allow the VAE model to better fit the data compared to vanilla VAEs.

3.4. VAE-VAMP models learn more compact latent representations than VAE-IAF, VAE and AE

A key feature of an AE-type embedding model is the ability to capture variability in the data with a number of latent units, much smaller than the dimensionality of the input space. While different datasets require different latent-space complexity, the number of latent units in an autoencoder is typically unchanged for different input data. We therefore assess the ability of the models to learn a compact representation, even when the number of latent units is higher than needed. To assess the latent-space compactness of the different models, we compared the number of principle components (PCs) required to capture 95% of the latent space variation for each model (Table 1). We consistently found the VAE-VAMP model to require fewer PCs for this than the other models. Combined with the higher covariance of the VAE-VAMP latent spaces (Fig. 2), this result suggests, that the VAE-VAMP model is the only model in our comparison that is able to learn a compact representation of the data by allowing for co-varying latent units. This makes the VAE-VAMP model particularly suitable for scenarios where the embedding model is trained with very diverse data such as when assembling complex single-cell atlases.

4. Conclusion

In this work, we adapted three VAE-based generative models to fit scRNA-seq data. We evaluated them for their ability to improve the quality of the learned latent representation of the data. We found that both the VAE-IAF model and VAE-VAMP model alleviate the inactive unit problem during VAE training and produce better model fits. Based on the ASW as well as visual inspection of the generated embeddings, the VAE-VAMP model outperforms the other generative models in preserving biological information in the latent representation, which we also found to be more compact.

An interesting next step would be to quantify the reduction in generative performance caused by the established down-scaling of the of the KL loss. Moreover, we see scope for further improving the embedding quality of generative models by combining the complimentary approaches taken by the VAE-IAF and VAE-VAMP in a single model.

In conclusion, the richer prior of the adapted VAE-VAMP model make it particularly suitable for distribution learning on highly multimodal data as found in scRNA-seq experiments. The VAMP prior allows the model to capture biological variation in the latent representation nearly as well as standard AE-based learning methods while fully retaining generative capabilities. We also expect the richer latent representation of the VAE-VAMP model to provide better reconstruction performance with a linear decoder network, therefore enabling the use of more interpretable models. We expect this model to be a useful starting point for future atlas-scale distribution-learning tasks in single-cell genomics.

References

Boyeau, P., Lopez, R., Regier, J., Gayoso, A., Jordan, M. I., and Yosef, N. Deep generative models for detecting differential expression in single cells. *bioRxiv*, 2019. doi: 10.1101/794289. URL https://www.biorxiv.org/content/early/2019/10/04/794289.

Burda, Y., Grosse, R. B., and Salakhutdinov, R. Importance weighted autoencoders. In Bengio, Y. and LeCun, Y. (eds.), *4th International Conference on Learning Representations, ICLR 2016, Conference Track Proceedings*, San Juan, Puerto Rico, 2–4 May 2016. URL http://arxiv.org/abs/1509.00519.

Eraslan, G., Simon, L. M., Mircea, M., Mueller, N. S., and Theis, F. J. Single-cell RNA-seq denoising using a deep count autoencoder. *Nature Communications*, 10(1), December 2019. ISSN 2041-1723. doi: 10.1038/s41467-018-07931-2. URL http://www.nature.com/articles/s41467-018-07931-2.

Grønbech, C. H., Vording, M. F., Timshel, P., Sønderby, C. K., Pers, T. H., and Winther, O. scVAE: Variational auto-encoders for single-cell gene expression data. *Bioinformatics*, 05 2020. ISSN 1367-4803. doi: 10.1093/bioinformatics/btaa293. URL https://doi.org/10.1093/bioinformatics/btaa293.

Kingma, D. P. and Welling, M. Auto-encoding variational bayes. In Bengio, Y. and LeCun, Y. (eds.), *2nd International Conference on Learning Representations, ICLR 2014, Conference Track Proceedings*, Banff, Canada, 14–16 April 2014. URL https://arxiv.org/abs/1312.6114.

Kingma, D. P., Salimans, T., Jozefowicz, R., Chen, X., Sutskever, I., and Welling, M. Improved Variational Inference with Inverse Autoregressive Flow. In Lee, D. D., Sugiyama, M., Luxburg, U. V., Guyon, I., and Garnett, R. (eds.), *Advances in Neural Information Processing Systems 29*, pp. 4743–4751. Curran Associates, Inc., 2016. URL https://papers.nips.cc/paper/6581-improved-variational-inference-with-inverse-autoregressive-flow.

Lopez, R., Regier, J., Cole, M. B., Jordan, M. I., and Yosef, N. Deep generative modeling for single-cell transcriptomics. *Nature Methods*, 15 (12):1053–1058, December 2018. ISSN 1548-7091, 1548-7105. doi: 10.1038/s41592-018-0229-2. URL http://www.nature.com/articles/s41592-018-0229-2.

Lotfollahi, M., Naghipourfar, M., Theis, F. J., and Wolf, F. A. Conditional out-of-sample generation for unpaired data using trVAE. *arXiv preprint*, October 2019a. URL http://arxiv.org/abs/1910.01791. arXiv: 1910.01791.

Lotfollahi, M., Wolf, F. A., and Theis, F. J. scGen predicts single-cell perturbation responses. *Nature Methods*, 16(8):715–721, August 2019b. ISSN 1548-7091, 1548-7105. doi: 10.1038/s41592-019-0494-8. URL http://www.nature.com/articles/s41592-019-0494-8.

McInnes, L., Healy, J., and Melville, J. Umap: Uniform manifold approximation and projection for dimension reduction. *arXiv preprint*, February 2018. URL http://arxiv.org/abs/1802.03426. arXiv:1802.03426.

Rousseeuw, P. J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20:53 – 65, 1987. ISSN 0377-0427. doi: 10.1016/0377-0427(87)90125-7. URL https://www.sciencedirect.com/science/article/pii/0377042787901257.

Svensson, V. Droplet scRNA-seq is not zero-inflated. *Nature Biotechnology*, 38(2):147–150, February 2020. ISSN 1546-1696. doi: 10.1038/s41587-019-0379-5. URL https://www.nature.com/articles/s41587-019-0379-5.

Tomczak, J. and Welling, M. Vae with a vampprior. In Storkey, A. and Perez-Cruz, F. (eds.), *Proceedings of the Twenty-First International Conference on Artificial Intelligence and Statistics*, volume 84 of *Proceedings of Machine Learning Research*, pp. 1214–1223, Playa Blanca, Lanzarote, Canary Islands, 09–11 April 2018. PMLR. URL http://proceedings.mlr.press/v84/tomczak18a.html.