Predictive Coding with Topographic Variational Autoencoders

T. Anderson Keller and Max Welling

UvA-Bosch Delta Lab, University of Amsterdam, Netherlands

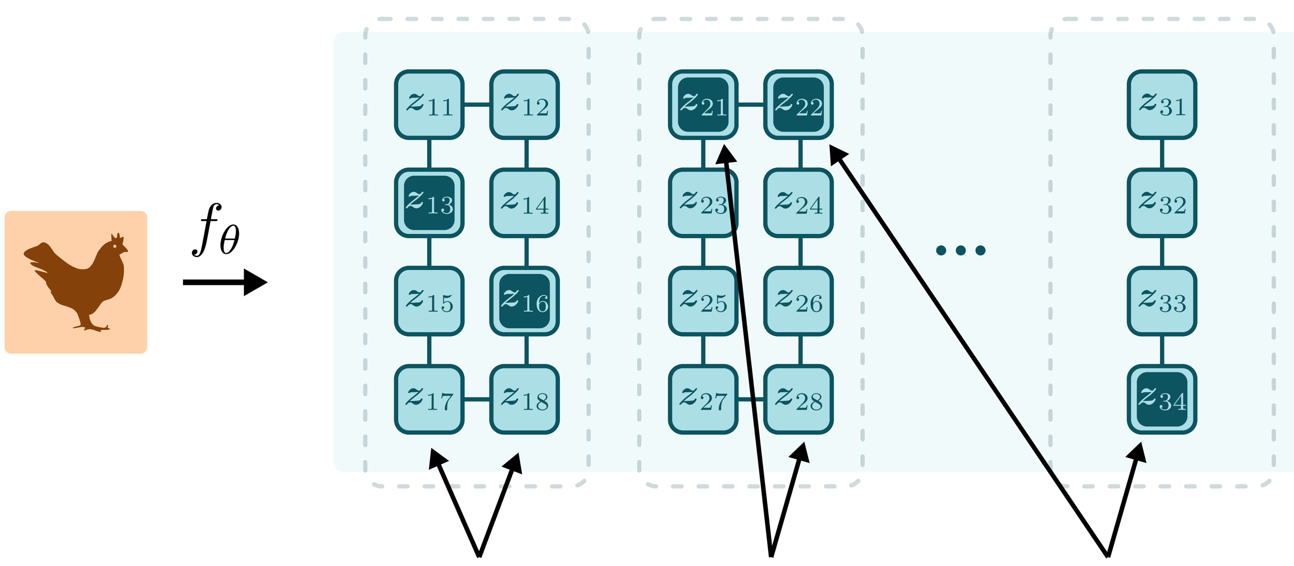
UNIVERSITEIT VAN AMSTERDAM

UvA - BOSCH DELTA LAB

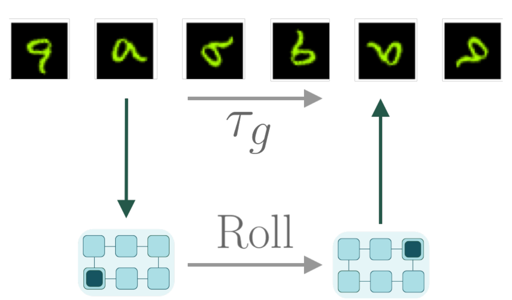
Summary

Topographic generative models are a class a generative models where latent variables are not entirely independent, but instead have a complex correlation structure defined by their position in a predefined topology:

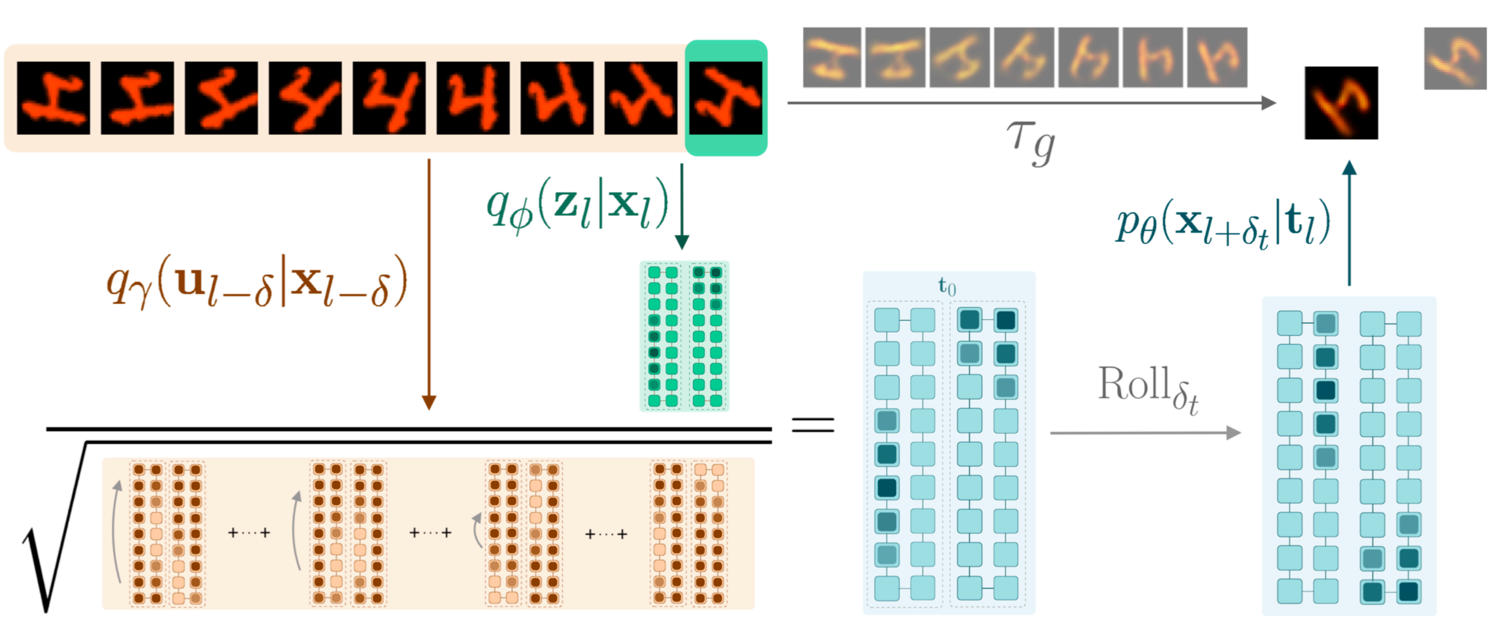
$$p(\boldsymbol{z}) = p(z_{11}, \ldots, z_{18})p(z_{21}, \ldots, z_{28})p(z_{31}, \ldots, z_{34})$$



Recently, the Topographic VAE (TVAE) [1] was introduced as a model capable of training deep generative models with such topographically organized latent variables. Furthermore, it was shown that such topographic organization could be leveraged to learn approximately equivariant capsules directly from transformation sequences. To achieve this, the TVAE learns to encode observed transformations as 'Capsule Rolls' within the capsule dimension:



In this work, we show how the 'Capsule Roll' operation can be further leveraged as a forward prediction operator for transformation sequences, and demonstrate that such a change significantly improves the ability of the model to accurately predict the immediate future.



Overview of the Predictive Coding Topographic VAE. The transformation in input space τ becomes encoded as a Roll within the equivariant capsule dimension. The model is thus able to forward predict the continuation of the sequence by encoding a partial sequence and rolling activations within the capsules.

Contributions

- We Introduce a novel version of the Topographic VAE, called PCTVAE, which depends only on past observations, permitting online training and inference.
- We show that by directly maximizing the likelihood of future inputs through the Capsule Roll, the PCTVAE is able to more accurately learn and predict sequence transformations.

The Generative Model

We assume that the joint distribution over observations $\mathbf{x}$ and latent variables $\mathbf{t}$ factorizes over time steps l, and further factorizes into a product of the generative conditional and the prior:

$$p_{\{\mathbf{X}_{l+1},\mathbf{T}_l\}_l}(\{\mathbf{x}_{l+1},\mathbf{t}_l\}_l) = \prod_l p_{\mathbf{X}_{l+1}|\mathbf{T}_l}(\mathbf{x}_{l+1}|\mathbf{t}_l)p_{\mathbf{T}_l}(\mathbf{t}_l)$$

As with the TVAE, the prior is a Topographic Product of Student's T distribution (TPoT), and is constructed from Gaussian random variables **Z** and **U**:

$$p_{\mathbf{T}_l}(\mathbf{t}_l) \;=\; \mathrm{TPoT}(\mathbf{t}_l;\nu)$$

$$\mathbf{T}_l = \frac{\mathbf{Z}_l - \mu}{\sqrt{\mathbf{W}\left[\mathbf{U}_l^2;\cdots;\mathbf{U}_{l-L}^2\right]}} \qquad \mathbf{Z}_l, \mathbf{U}_l \sim \mathcal{N}(\mathbf{0},\mathbf{I})$$

Importantly, the PCTVAE differs from the TVAE by only including past **U** in the denominator ($\delta > 0$), and by defining the conditional generative distribution to model timestep $l+1$ given the **t** variable from timestep l :

$$\mathbf{W}\left[\mathbf{U}_l^2;\cdots;\mathbf{U}_{l-L}^2\right] = \sum_{\delta=0}^{L}\mathbf{W}_\delta \mathrm{Roll}_\delta(\mathbf{U}_{l-\delta}^2)$$

$$p_{\mathbf{X}_{l+1}|\mathbf{T}_l}(\mathbf{x}_{l+1}|\mathbf{t}_l) = p_\theta(\mathbf{x}_{l+1}|g_\theta(\mathbf{t}_l))$$

The Predictive Coding Topographic VAE

To train the parameters of the generative model, we use the above construction of **T** to parameterize an approximate posterior for **t** in terms of a deterministic transformation of approximate posteriors of simpler Gaussian latent variables **z** and **u**:

$$q_\phi(\mathbf{z}_l|\mathbf{x}_l) = \mathcal{N}\big(\mathbf{z}_l;\mu_\phi(\mathbf{x}_l),\sigma_\phi(\mathbf{x}_l)\mathbf{I}\big) \qquad \mathbf{t}_l = \frac{\mathbf{z}_l-\mu}{\sqrt{\mathbf{W}\left[\mathbf{u}_l^2;\cdots;\mathbf{u}_{l-L}^2\right]}}$$

$$q_\gamma(\mathbf{u}_l|\mathbf{x}_l) = \mathcal{N}\big(\mathbf{u}_l;\mu_\gamma(\mathbf{x}_l),\sigma_\gamma(\mathbf{x}_l)\mathbf{I}\big)$$

We additionally add a capsule roll of one step as the first layer of the decoder to encourage the use of the capsule roll as the forward prediction operator:

$$p_\theta(\mathbf{x}_{l+1}|g_\theta(\mathbf{t}_l)) = p_\theta(\mathbf{x}_{l+1}|\hat{g}_\theta(\mathrm{Roll}_1[\mathbf{t}_l]))$$

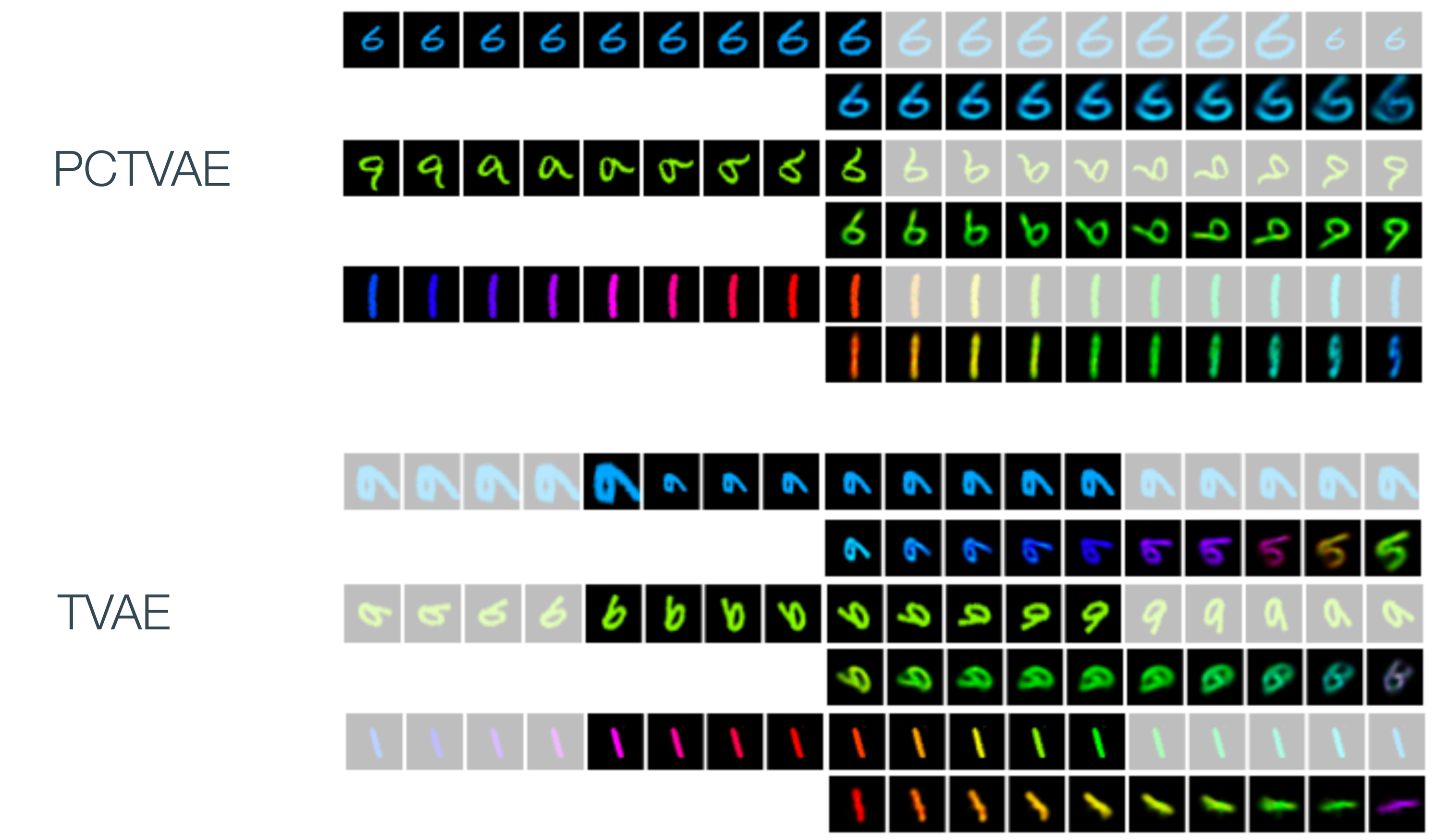
The model is then trained to maximize the likelihood of the data through the ELBO:

$$\sum_{l=1}^{S}\mathbb{E}_{Q_{\phi,\gamma}(\mathbf{z}_l,\mathbf{u}_l|\{\mathbf{x}\})}\Big(\log p_\theta(\mathbf{x}_{l+1}|\hat{g}_\theta(\mathrm{Roll}_1[\mathbf{t}_l])) - D_{KL}[q_\phi(\mathbf{z}_l|\mathbf{x}_l)||p_{\mathbf{Z}}(\mathbf{z}_l)] - D_{KL}[q_\gamma(\mathbf{u}_l|\mathbf{x}_l)||p_{\mathbf{U}}(\mathbf{u}_l)]\Big)$$

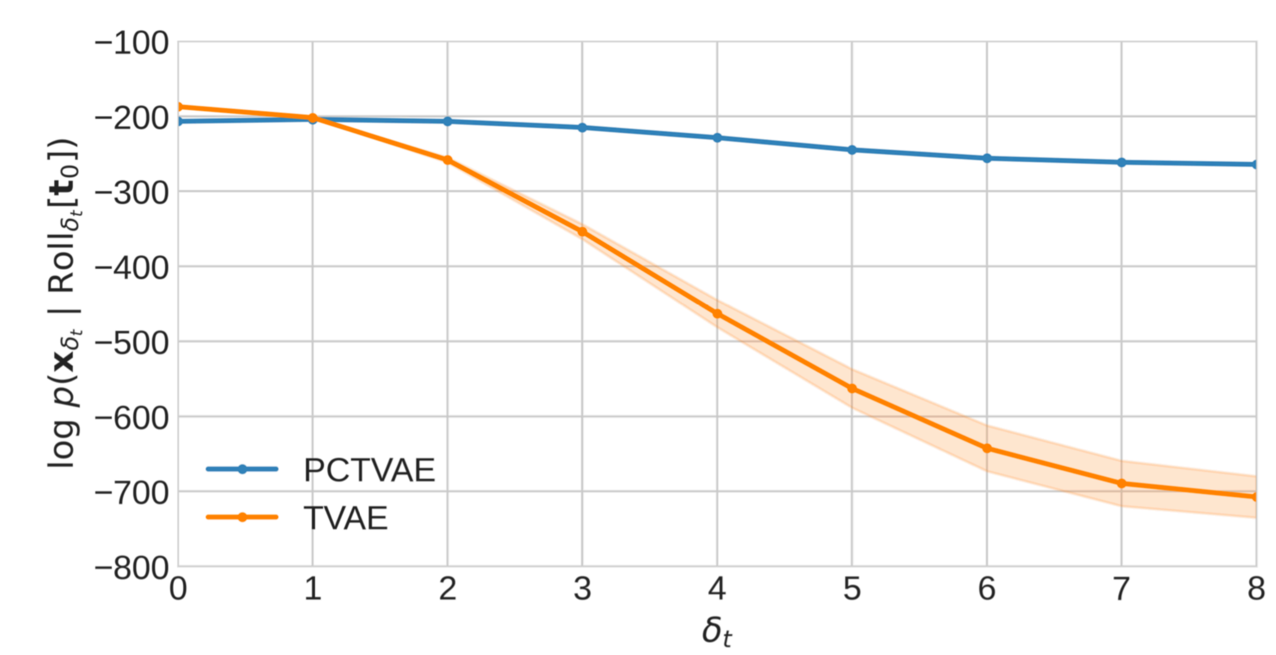
References

[1] T. Anderson Keller and Max Welling. Topographic VAEs learn Equivariant Capsules. NeurIPS 2021.

Experiments



Forward predicted trajectories from the Predictive Coding TVAE (top) and the original TVAE (bottom). The images in the top row show the true input transformation, with greyed out images being held out. The lower row then shows the reconstruction, constructed by starting at t_0, and progressively rolling the capsules forward to decode the remainder of the sequence. We see the PCTVAE is able to predict sequence transformations accurately, while the TVAE forward predictions slowly lose coherence with the input sequence.



Forward prediction log-likelihood vs. future prediction offset. We see that the PCTVAE has consistently high likelihood for sequence elements into the future whereas the likelihood of the TVAE model drops off rapidly.

	NLL @ $\delta_t = 0$	NLL Avg. Seq.	$\mathcal{E}_{eq}$ Avg. Seq.
VAE	190 ± 1	N/A	13274 ± 0
TVAE	$\mathbf{187} \pm 1$	452 ± 16	2122 ± 21
PCTVAE	207 ± 1	$\mathbf{232} \pm 1$	2201 ± 9

Negative log-likelihood (NLL in nats) without forward prediction ($\delta t = 0$), NLL averaged over the forward predicted sequence, and equivariance error E_{eq} for a non-topographic VAE, TVAE, and PCTVAE. The PCTVAE achieves the lowest average NLL over the forward predicted sequence while also maintaining low equivariance error.

Paper: https://openreview.net/pdf?id=WvUOFEESncx

Code: https://github.com/akandykeller/PCTVAE