PHOTREK

Photrek Robust Data Generation Using the CVAE

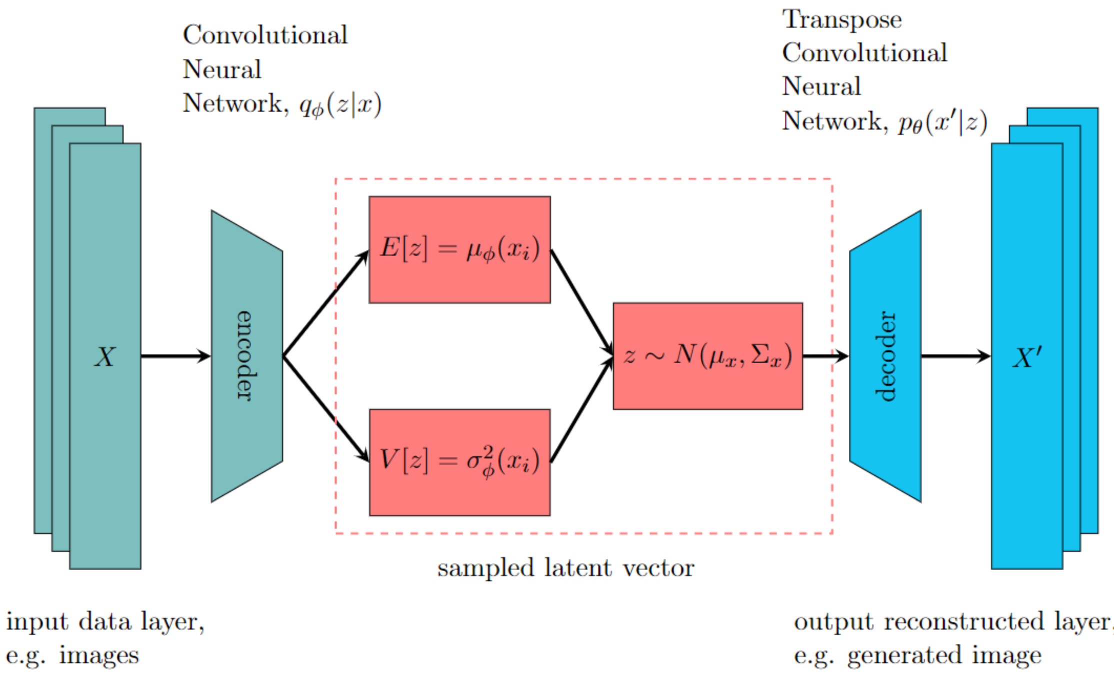
William Thistleton, Kenric Nelson, Blake Anderton, and Hong Yue
March 2023

A Variational Auto-Encoder (VAE) (*Kingma and Welling, 2019)* generates an object of interest by learning how similar objects are composed. The applications and use cases are continually growing, but new image generation from image libraries such as *MNIST (Deng, 2012)* and *CIFAR10 (Krizhevsky, 2009)*, the generation of novel faces (*Nakagawa et al., 2020)*, the generation of synthetic time series via sequential approaches (*Kaewkiriya and Woraratpanya, 2022)*, and intrusion detection (*Nguyen et al., 2022)*, and (*Singh and Ogunfunmi, 2021)* are just a few examples. Since the foundational papers *Kingma and Welling (2014)* and *Rezende et al. (2014)*, VAEs have been used to generate independent data such as images, but also sequential data such as audio, video, and text (*Girin et al., 2021)*.

Given a high dimensional data set, such as a library of images, a VAE will produce new samples that share important properties of the original data and mimic these originals, but are new creations. In some respects, a VAE is a generalization of an autoencoder, which compresses or encodes a higher dimensional input vector into a lower dimensional object in the so-called latent space. The lower dimensional object is then reconstructed or decoded back to a good approximation of the original vector.

In an autoencoder, almost all the images created by arbitrarily selecting an object in the latent space for decoding will be noise. High dimensional data sets descriptive of phenomena of interest typically live in a very narrow band within the space of potential objects. The overwhelming majority of the original space will be occupied by noise. For example, a grayscale image is typically represented by a set of pixels with values between 0 and 255. An image of the same dimensions as the commonly studied MNIST images will be represented on a grid of $28 \times 28 = 784$ pixels. Thus, the potential number of images is $256^{784} \approx 10^{1888}$. Compared with 10^{82} atoms in the known, observable universe, MNIST images (real or simulated) occupy a dramatically smaller subset. With an autoencoder, typically only vectors in the latent space arrived at through the encoding process will contain meaningful information for the task at hand.

The VAE extends this brittle and deterministic design to incorporate a stochastic representation in the latent space, allowing for the generation of novel vectors with many of the interesting features of the originals. An object of interest is encoded as a *distribution* in the latent space rather than as an *individual point*.



We briefly review the formalism of the VAE and show the improvement obtained with our Coupled Variational autoencoder, or CVAE. The foundational paper for VAEs is (*Kingma and Welling, 2019)*. A very useful tutorial is (*Doersch, 2016)*. Our project builds on recent work (*Cao et al., 2022)* which extends the traditional VAE framework to a CVAE. This allows us to generate images in a way which is robust to corruptions or noise in the original image data. Our approach replaces the traditional loss function used in training with one derived from a generalized coupled entropy, a modification of Tsallis entropy (*Tsallis, 1988)*. This coupled entropy improves upon the usual normalization of nonextensive statistical mechanics based on the escort distribution. We use Convolutional Neural Networks (*Lecun et al., 1998)* for the encoder and decoder rather than the fully connected networks previously studied to allow for representations of more complex images than MNIST. Our goal is to develop a product that can also handle video.

We consider the use case of novel image generation. Ideally, one would represent images by a model distribution $P(X)$ defined over a space of possible images X and then sample from $P(X)$ to generate new, unseen images. Taking inspiration from MNIST, given a 28×28 grid and a multivariate normal model for pixel values, we would need to estimate $28 \times 28 = 784$ means and $\binom{784}{2} = 306936$ variances and covariances. Using a dataset of several thousand images, this is a dramatically under-determined problem, so we will need to use approximating distributions.

The latent space Z must be sufficiently detailed so as to produce outputs consistent with the original data set. It should also be of much lower dimension than the original space. We sample from the latent space, then ”decode” using a neural network. VAEs create this stochastic representation in the latent space by learning distributions rather than fixed vectors in the latent space. There have been many extensions and adaptations since the original formulation, such as the β-VAE (*Higgins et al., 2017)* which increases the importance of the regularization term in the loss function in order to produce a more interpretable representation of the independent data at the expense of reconstruction accuracy. Other models include the q-VAE (*Kobayashi, 2020)* which creates an improved disentangled latent space with a new lower bound and which also allows a relaxing of the independence condition on the data vectors.

We augment the original VAE framework with a single hyperparameter, the coupling κ, that modulates the relative weight of small and large probability events in the learning process. The subsequent Coupled Evidence Lower Bound (CELBO) enables us to learn models of complex datasets with reduced variation, thereby improving robustness against outliers in the dataset. The CVAE CELBO has two loss components, the image loss, which measures the coupled log-likelihood of the reconstructed image given its input sample and the latent loss, which measures the coupled divergence between a the latent layer prior (a multivariate standard normal distribution) and the posterior distribution. By tuning the coupling parameter we can adjust the degree to which the model learns a canonical or broad representation of the dataset. CVAE with $\kappa = 0$ corresponds to the original VAE framework (*Kingma Welling, 2014; Rezende et al., 2014*). With $\kappa > 0$ the model is pushed to learn a more robust model of the data, which reduces the variation of the output samples.

This simple modification allows CVAE to significantly improve the degree of disentanglement in learnt latent representations compared to the unmodified VAE framework (*Kingma Welling, 2014; Rezende et al., 2014*). Furthermore, we show that CVAE achieves state of the art disentangling performance against both the best unsupervised (*Chen et al., 2016)* and semi-supervised (*Kulkarni et al., 2015)* approaches for disentangled factor learning on a number of benchmark datasets, such as CelebA (*Liu et al., 2015)*, chairs (*Aubry et al., 2014)* and faces (*Paysan et al., 2009)* using qualitative evaluation.

Results

We generate images free of corruptions and allow the level of coupling to vary from $\kappa = 0$ (no coupling) to $\kappa = 1/2$ (heavy tailed with variance $= \frac{1}{1-2\kappa} = \infty$). Some representative images are shown below. Reconstruction improves as coupling increases.

Figure 1 Original Corrupted Images

No Coupling

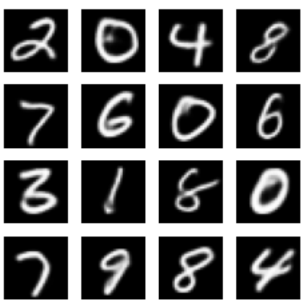


Figure 2 Generated Images, Coupling=0

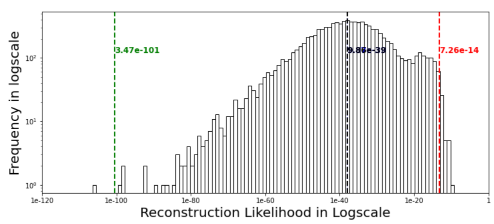


Figure 3 Reconstruction Performance Measure, Coupling=0

Light Coupling, kappa=0.05

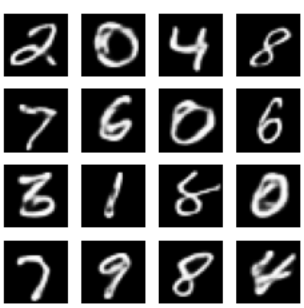


Figure 4 Generated Images, Coupling=0.05

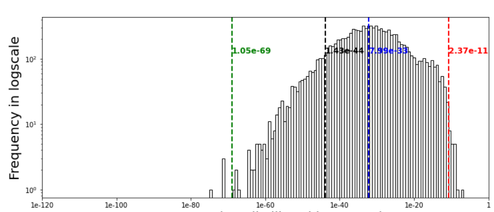


Figure 5 Reconstruction Performance Measure, Coupling=0.05

More Coupling, kappa=0.3

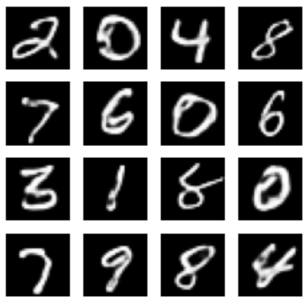


Figure 6 Generated Images, Coupling=0.3

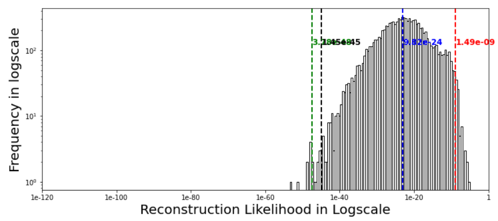


Figure 7 Reconstruction Performance Measure, Coupling=0.3

References

Mathieu Aubry, Daniel Maturana, Alexei A. Efros, Bryan C. Russell, Josef Sivic; Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2014, pp. 3762-3769

Cao, S., Li, J., Nelson, K. P., & Kon, M. A. (2022). Coupled VAE: Improved Accuracy and Robustness of a Variational Autoencoder. Entropy, 24(3). https://doi.org/10.3390/e24030423

Chen, Xi and Duan, Yan and Houthooft, Rein and Schulman, John and Sutskever, Ilya and Abbeel, Pieter (2016). InfoGAN: Interpretable Representation Learning by Information Maximizing Generative Adversarial Nets. Advances in Neural Information Processing Systems, Curran Associates, Inc.

Deng, Li (2012). The mnist database of handwritten digit images for machine learning research. IEEE Signal Processing Magazine, volume 29, number 6, pages 141--142

Doersch, Carl (2016). Tutorial on Variational Autoencoders http://arxiv.org/abs/1606.05908

Irina Higgins and Loic Matthey and Arka Pal and Christopher Burgess and Xavier Glorot and Matthew Botvinick and Shakir Mohamed and Alexander Lerchner (2017). Learning Basic Visual Concepts with a Constrained Variational Framework, International Conference on Learning Representations

Kaewkiriya, Thongchai and Woraratpanya, Kuntpong (2022) 3DVAE-LSTM for Extremely Rare Anomaly Signal Generation, The 14th International Conference on Information Technology and

Electrical Engineering (ICITEE 2022)

Kingma, Diederik P. and Welling, Max (2014). Auto-Encoding Variational Bayes. 2nd International Conference on Learning Representations, {ICLR} 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings.

Kingma, D. P., & Welling, M. (2019). An introduction to variational autoencoders. Foundations and Trends in Machine Learning, 12(4), 307–392. https://doi.org/10.1561/2200000056

Kobayashi, Taisuke, (2020). Q-VAE for Disentangled Representation Learning and Latent Dynamical Systems, IEEE Robotics and Automation Letters, number = 4, pages 5669--5676.

Alex Krizhevsky, G Hinton (2009). Learning multiple layers of features from tiny images. Technical report, University of Toronto.

Tejas D Kulkarni, William F Whitney, Pushmeet Kohli, and Josh Tenenbaum (2015). Deep convolutional inverse graphics network. In Advances in Neural Information Processing Systems, pages 2530–2538.

Lecun, Y. and Bottou, L. and Bengio, Y. and Haffner, P. (1998). Gradient-based learning applied to document recognition in Proceedings of the IEEE volume=86, number=11, pages=2278-2324.

Liu et al. (2015). Analyzing effects of climate change on streamflow in a glacier mountain catchment using an ARMA model (2015). Quaternary International, 137-145

Nakagawa, A., Kato, K., & Suzuki, T. (2020). Quantitative Understanding of VAE as a Non-linearly Scaled Isometric Embedding. International Conference on Machine Learning.

Nelson, K. P., & Umarov, S. (2010). Nonlinear statistical coupling. Physica A: Statistical Mechanics and Its Applications, 389(11), 2157–2163. https://doi.org/10.1016/j.physa.2010.01.044

Nguyen, V., Nguyen, V., Hoang, T., Shone, N (2022). A Novel Deep Clustering Variational Auto-Encoder for Anomaly-based Network Intrusion Detection 2022 14th International Conference on Knowledge and Systems Engineering (KSE). Pages 1-7.

P. Paysan, R. Knothe, B. Amberg, S. Romdhani, and T. Vetter (2019). A 3d face model for pose and illumination invariant face recognition. AVSS.

PHOTREK

Razavi, A., van den Oord, A., & Vinyals, O. (2022). *Generating Diverse High-Resolution Images with VQ-VAE*. https://github.com/deepmind/sonnet/blob/master/sonnet/python/modules/nets/vqvae.py

Razavi, A., van den Oord, A., & Vinyals, O. (2019). Generating diverse high-fidelity images with VQ-VAE-2. *Advances in Neural Information Processing Systems*, *32*(NeurIPS 2019), 1–11.

Danilo Jimenez Rezende and Shakir Mohamed and Daan Wierstra (2014). Stochastic Backpropagation and Approximate Inference in Deep Generative Models}, arXiv

Singh, A., & Ogunfunmi, T. (2022). An Overview of Variational Autoencoders for Source Separation, Finance, and Bio-Signal Applications. In Entropy (Vol. 24, Issue 1). MDPI. https://doi.org/10.3390/e24010055

Soelch, M., Bayer, J., Ludersdorfer, M., & van der Smagt, P. (2016). Variational Inference for On-line Anomaly Detection in High-Dimensional Time Series. http://arxiv.org/abs/1602.07109

Singh A, Ogunfunmi T. An Overview of Variational Autoencoders for Source Separation, Finance, and Bio-Signal Applications. Entropy. 2022; 24(1):55. https://doi.org/10.3390/e24010055

Tsallis J Stat Phys 52(1–2):479–487, 1988)

Zhu, Y., Renqiang Min, M., Kadav, A., & Peter Graf, H. (2020). S3VAE: Self-Supervised Sequential VAE for Representation Disentanglement and Data Generation Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition

PHOTREK

Appendix: Nonlinear Statistical Coupling

The Coupled Loss Function

The innovation of the CVAE is in the loss function. VAEs uses the while the CVAE uses the coupled-ELBO. The coupled-ELBO is defined as:

$$L\kappa(q, \Theta) = log\kappa p(x|\Theta)) - CD\kappa(q(z)||p(z)|x, \Theta)$$

where κ is the coupling value of the loss function. Since we are minimizing this function, the loss function is the negative coupled-ELBO.

To compute the loss function, a definition for the coupled logarithm is needed. The coupled logarithm is defined as *Nelson and Umarov 2009*:

$$\log_\kappa(x) = \frac{1}{\kappa}(x^{\frac{\kappa}{1+d\kappa}} - 1)$$

With that defined, the computation of the coupled log-likelihood and divergence can be shown. The coupled log-likelihood is calculated as:

$$log_\kappa p(x|\Theta)) = \Sigma_{i=1}^{n} p_i log_\kappa(q_i) + (1-p_i) log_\kappa(1-q_i)$$

where p_i is the true probability, q_i is the probability from the reconstruction, and *n* is the number of images in the batch.

The CVAE is an application of Nonlinear Statistical Coupling (NSC). Here we briefly describe some of the basic properties of NSC. The modeling of complex systems and their heavy-tailed distributions has several established approaches,

1) focus on the scale-invariant power-law tail decay, often called Pareto distributions,
2) the generalization of the Pareto distributions,
3) the class of Levy stable distributions,
4) the q-distributions of Tsallis statistics, and
5) the Pearson classification of distributions, which is the most comprehensive.

Given the variety of existing methods for describing heavy-tail distributions and their relationship with the modeling of complex systems, one might ask why interject yet another schema. In the immortal words of Occam, “the simplest solution is almost always the best”. Each of the existing classifications of heavy-tailed distributions has origins that pre-date a thorough understanding of complex systems. While that understanding remains incomplete, we do know today that nonlinear dynamics is central to the properties of complex systems; and our characterization of these systems, including their underlying distributions, has the potential to be simplified if one isolates the degree of nonlinearity.

Tsallis’ own investigations started from the observation that the construct p^q provides the foundation for a generalization of thermodynamic analysis. Unfortunately, to the authors knowledge, the extensive literature on Nonextensive Statistical Mechanics does not include an explanation that given a probability raised to a power, the power q must represent the number of independent samples having the probability p or how this property of independence relates to the nonlinearity of the complex system. Theoretical and statistical evidence that q is the number of independent samples was provided in [4]. A further difficulty is that q is dependent on the dimensions d and the power of the random variable . The random variable power refers to whether the heavy-tail distribution is a deviation from the exponential $\alpha = 1$, Gaussian $\alpha = 2$ distribution, or a non-integer power defined below. To facilitate modeling and analysis of the statistical properties of complex systems, the methods of Nonlinear Statistical Coupling were defined such that the degree of nonlinear deformation κ from the linear domain is clearly isolated. Linear analysis naturally involves the logarithmic and exponential functions. Examples include the solution of linear differential equations being exponential, the additive properties of information lead to use of the logarithmic function for a measure of information, and the maximum entropy solution constrained by the mean, standard deviation, or higher moments lead to the exponential family of distributions. In contrast nonlinear systems are described by power functions, which the Nonextensive Statistical Mechanics community showed can be expressed as generalized functions of the exponential and logarithm. Here we will specify this generalization such that the degree of deformation from the exponential/logarithm.

Definition 1: The Coupled Exponential and Coupled Logarithm The coupled exponential is a generalization of the exponential function defining a family of power-functions parameterized by the nonlinear coupling . To facilitate the isolation the deformation by the nonlinear statistical coupling, the parameterization also includes the dimension d and power of the variable (1.5) The coupled logarithm is the inverse of the coupled exponential, (1.6) This leads to definitions which specify a family of heavy-tail (or compact-support) distributions which maximize a coupled-entropy function and Lagrangian constraints on generalized moments, particularly the location and scale of the random variable. The coupled cross-entropy serves as a foundational function from which the coupled-entropy (densities are equal) and the coupled-divergence (generalization of Kullback-Leibler divergence) are derived.

Definition 2: The Coupled Cross-Entropy, Coupled-Entropy, and Coupled-Divergence The Coupled Cross-Entropy defines a generalization of the Boltzmann-Gibbs-Shannon entropy based on the degree of nonlinear deformation from the exponential/logarithmic and two supporting parameters the dimension d and power of the random variable X. (1.7) The Coupled Entropy is equal to the Coupled Cross-Entropy with (1.8) The Coupled Divergence (a generalization of the Kullback-Leibler Divergence) is the Coupled Cross-Entropy minus the Coupled Entropy (1.9) Comment 1: Relationship of Coupled-Entropy with other generalized entropy definitions The Coupled Entropy derives from research in complex systems in which nonlinear dynamics have been shown to result in growth of statistic systems which are nonextensive, that is to say information is not combined additively. The nonextensive statistical mechanics (NSC) community has parameterized generalizations to entropy functions using an index q but has not been clear about what the physical property q represents. In recent work by one of the authors (Nelson), it was shown that given the origin of these generalizations using escort probabilities formulated from the expression , q is simply the number of independent samples of the ith state. The connection to nonlinear systems with long-range correlation arises in that evaluating the statistical properties of q equally-valued random variables rather than one random variable dampens the heavy-tail characteristics of these systems, enabling estimation of otherwise unmeasurable moments.