Decoder ensembling for learned latent geometries

Syrota, Stas; Moreno-Munõz, Pablo; Hauberg, Søren

Published in:
Proceedings of the ICML 2024 Workshop on Geometry-grounded Representation Learning and Generative Modeling

Publication date:
2025

Document Version
Peer reviewed version

Decoder ensembling for learned latent geometries

Stas Syrota [1] **Pablo Moreno-Muñoz** [1] **Søren Hauberg** [1]

Abstract

Latent space geometry provides a rigorous and empirically valuable framework for interacting with the latent variables of deep generative models. This approach reinterprets Euclidean latent spaces as Riemannian through a pull-back metric, allowing for a standard differential geometric analysis of the latent space. Unfortunately, data manifolds are generally compact and easily disconnected or filled with holes, suggesting a topological mismatch to the Euclidean latent space. The most established solution to this mismatch is to let uncertainty be a proxy for topology, but in neural network models, this is often realized through crude heuristics that lack principle and generally do not scale to high-dimensional representations. We propose using ensembles of decoders to capture model uncertainty and show how to easily compute geodesics on the associated expected manifold. Empirically, we find this simple and reliable, thereby coming one step closer to easy-to-use latent geometries.

1. Introduction

Generative models provide state-of-the-art density estimators for high-dimensional data (Lipman et al., 2022; Sohl-Dickstein et al., 2015; Ho et al., 2020; Rombach et al., 2022). In the case of deep latent variable models, such as the *variational autoencoder (*VAE*)* (Kingma & Welling, 2014; Rezende et al., 2014), we assume that data is distributed near a low-dimensional manifold in the spirit of the *manifold hypothesis* (Bengio et al., 2013). Specifically, we assume that data $\mathbf{x} \in \mathcal{X}$ lies near a low-dimensional manifold $\mathcal{M} \subset \mathcal{X}$, which is parametrized through a low-dimensional *latent representation* $\mathbf{z} \in \mathcal{Z}$. Given finite noisy data, we can recover a stochastic estimate of $\mathcal{M}$.

Formally, the VAE is defined through a (usually unit-Gaussian) prior $p(\mathbf{z})$ over the latent variables and a conditional likelihood $p(\mathbf{x}|\mathbf{z})$, which is parametrized by the output of a neural network, $f_\theta : \mathcal{Z} \to \mathcal{X}$, known as the *decoder*. These then define the data density

$$p(\mathbf{x}) = \int p(\mathbf{x}|\mathbf{z})p(\mathbf{z})\mathrm{d}\mathbf{z}. \tag{1}$$

Here the latent space $\mathcal{Z}$ is generally Euclidean $\mathbb{R}^d$ with a significantly lower dimension than the observation space $\mathcal{X}$.

We focus on the latent space $\mathcal{Z}$, which generally lacks physical units even when the data may possess such. Following Arvanitidis et al. (2018), we consider infinitesimal latent distances measured along the data manifold in observation space. If we let $\mathbf{z}$ denote some latent variable and let $\Delta\mathbf{z}_1$ and $\Delta\mathbf{z}_2$ be infinitesimals, then we can compute the squared distance using Taylor's Theorem,

$$\begin{aligned}\|f(\mathbf{z}+\Delta\mathbf{z}_1) - f(\mathbf{z}+\Delta\mathbf{z}_2)\|^2 \\ = (\Delta\mathbf{z}_1 - \Delta\mathbf{z}_2)^\intercal \left(\mathbf{J}_\mathbf{z}^\intercal \mathbf{J}_\mathbf{z}\right)(\Delta\mathbf{z}_1 - \Delta\mathbf{z}_2),\end{aligned} \tag{2}$$

where $\mathbf{J}_\mathbf{z} = \frac{\partial f}{\partial \mathbf{z}}\big|_{\mathbf{z}=\mathbf{z}}$ is the Jacobian of the decoder f. This implies that the natural distance function in $\mathcal{Z}$ changes locally through the Riemannian metric $\mathbf{G}_\mathbf{z} = \mathbf{J}_\mathbf{z}^\intercal \mathbf{J}_\mathbf{z}$, which gives the latent space a rich geometric structure.

The geometry of the manifold has been shown to carry great value when systematically interacting with the latent representations, as it provides meaningful distances that are independent of how the latent space is parametrized (Tosi et al., 2014; Arvanitidis et al., 2018; Hauberg, 2018). For example, this geometry has allowed VAEs to discover latent evolutionary signals in proteins (Detlefsen et al., 2022), provide efficient robot controls (Scannell et al., 2021; Chen et al., 2018; Beik-Mohammadi et al., 2021), improve latent clustering abilities (Yang et al., 2018; Arvanitidis et al., 2018) and more.

The fundamental issue with these geometric approaches is that by assuming the latent space to have an Euclidean topology, we impose the same topology on the manifold $\mathcal{M}$ in observation space. In practice, we have little *a priori* information about the topology of the true manifold and must rely on the observed data to estimate a reasonable

[1]Department of Applied Mathematics and Computer Science, Technical University of Denmark. Correspondence to: Stas Syrota <stasy@dtu.dk>, Pablo Moreno-Muñoz <pabmo@dtu.dk>, Søren Hauberg <sohau@dtu.dk>.

Proceedings of the Geometry-grounded Representation Learning and Generative Modeling Workshop (GRaM) at the 41st International Conference on Machine Learning, Vienna, Austria. PMLR 251, 2024.

topology. As data is finite, we should expect such an estimate to be compact, and empirically it is often observed that manifolds arising from real-world data are disconnected and often have holes. All of which mismatches the Euclidean latent topology.

Hauberg (2018) argues that the uncertainty of the decoder offers a *topological hint*, i.e. when model uncertainty is high we are most likely outside the support of the driving manifold. When the decoder follows a *Gaussian process (GP)*, there is a well-established notion of model uncertainty, and its impact on the latent geometry is reasonably well-understood (Tosi et al., 2014; Pouplin et al., 2023). However, when the decoder is a neural network (the ever-present case), a set of heuristics is commonly applied to mimic the behavior of the GP models (Arvanitidis et al., 2018; 2019; 2021; 2022; Detlefsen et al., 2019; 2022; Beik-Mohammadi et al., 2021). Besides lacking principle, these heuristics also tend to break down when the latent dimension exceeds a handful.

In this paper, we propose to use an ensemble (Lakshminarayanan et al., 2017; Hansen & Salamon, 1990) of decoders in the VAE to capture model uncertainty and provide simple training techniques for the associated model. We then show how to easily incorporate the ensemble into the computation of geodesics on the modeled stochastic manifold. The result is a simple, yet reliable, approach for leveraging uncertainty in learned geometric representations.

2. Background and related work

2.1. Variational autoencoders

We briefly review the variational autoencoder (VAE) as our empirical results are realized with this generative model. Many of our findings, however, extend beyond this model.

The VAE (Kingma & Welling, 2014; Rezende et al., 2014) is a deep latent variable model that generalizes *probabilistic principal component analysis* (Tipping & Bishop, 1999). Commonly the latent variable is assumed *a priori* to follow a unit-Gaussian, $p(\mathbf{z}) = \mathcal{N}(\mathbf{z}|\mathbf{0}, \mathbf{I})$, though more elaborate priors have been studied (Tomczak & Welling, 2018; Kalatzis et al., 2020; Rombach et al., 2022). The conditional likelihood $p(\mathbf{x}|\mathbf{z})$ is then parametrized by a neural network $f_\theta(\mathbf{z})$ known as the *decoder*. For example, for a Gaussian VAE, we let $f_\theta(\mathbf{z}) = (\mu_\theta(\mathbf{z}); \sigma_\theta(\mathbf{z}))$ and

$$p(\mathbf{x}|\mathbf{z}) = \mathcal{N}(\mathbf{x}|\mu(\mathbf{z}), \sigma^2(\mathbf{z})\mathbf{I}), \tag{3}$$

where we omitted the θ subscript for brevity. The data likelihood (1) arise by the marginalization of $\mathbf{z}$, but, alas, the associated integral is generally intractable and we resort to a lower bound, known as the *ELBO*, (Kingma & Welling, 2014; Rezende et al., 2014)

$$\mathcal{L}_{\theta,\psi} = \mathbb{E}_{q_\psi(\mathbf{z}|\mathbf{x})}\left[\log p_\theta(\mathbf{x}|\mathbf{z})\right] - \mathrm{KL}(q_\psi(\mathbf{z}|\mathbf{x})\|p(\mathbf{z})), \tag{4}$$

where $q_\psi(\mathbf{z}|\mathbf{x}) = \mathcal{N}(\mathbf{z}|\mu_\psi(\mathbf{x}), \sigma^2_\psi(\mathbf{x}))$ is a variational approximation to the latent posterior $p(\mathbf{z}|\mathbf{x})$. Details can be found in the original papers.

Despite mentioning priors and posteriors, the VAE is inherently non-Bayesian as it relies on maximum likelihood to arrive at a point estimate of the decoder parameters θ. Daxberger & Hernández-Lobato (2019) gives the model a Bayesian treatment and relies on stochastic gradient Markov chain Monte Carlo for inference.

We focus on the common case where the latent space is assumed to have an Euclidean structure, i.e. $\mathcal{Z} = \mathbb{R}^d$. This is, however, not a strict requirement and other latent structures have been investigated (Davidson et al., 2018; Mathieu et al., 2019).

2.2. Latent representation geometries

The latent variables of the VAE are enticing as they provide low-dimensional 'distillations' of high-dimensional data. This form of representation learning (Bengio et al., 2013) can give us a *glimpse* into the model's inner workings, but also potentially in the mechanisms of the true physical system that generated the data.

Unfortunately, the latent space can be almost arbitrarily deformed without changing the associated model density (Hauberg, 2018). To see this, consider a smooth invertible function $h : \mathcal{Z} \leftarrow \mathcal{Z}$, such that its inverse is also smooth (i.e. a *diffeomorphism*). If the Jacobian of h further has unit determinant, we see that the latent representations $\hat{\mathbf{z}} = h(\mathbf{z})$ yields an unchanged density when combined with the decoder $\hat{f} = f \circ h^{-1}$. This implies that whichever latent representations we may recover from optimizing Eq. 4, are not unique. This lack of uniqueness hinders any form of interpretability of the latent representations, such that the aforementioned 'glimpse' becomes difficult to trust. Hauberg (2023) discuss this issue at greater length, while Detlefsen et al. (2022) show the empirical significance of the problem in a model of proteins.

Fortunately, *differential geometry* provides an elegant solution (Arvanitidis et al., 2018; Shao et al., 2018). The basic idea is to define distances in the latent space by measuring infinitesimally along the spanned manifold in observation space. Specifically, consider a latent curve $\gamma : [0, 1] \to \mathcal{Z}$ and its decoded counterpart $f \circ \gamma : [0, 1] \to \mathcal{X}$. We may then define the length of γ by integrating $f \circ \gamma$, i.e.

$$\mathrm{Length}[\gamma] = \int_0^1 \left\| \frac{\mathrm{d}}{\mathrm{d}t} f(\gamma_t) \right\| \mathrm{d}t, \tag{5}$$

where $\gamma_t = \gamma(t)$. Applying the chain rule quickly reveals that the integrand can be written as

$$\left\| \frac{\mathrm{d}}{\mathrm{d}t} f(\gamma_t) \right\| = \sqrt{\dot{\gamma}_t^\intercal \mathbf{J}_{\gamma_t}^\intercal \mathbf{J}_{\gamma_t} \dot{\gamma}_t}, \tag{6}$$

where $\dot{\gamma}_t = {}^{\mathrm{d}\gamma}/_{\mathrm{d}t}|_{t=t}$ denotes the curve derivative. The matrix $\mathbf{J}_{\gamma_t}^\intercal \mathbf{J}_{\gamma_t}$, thus, defines a local inner product, which is known as a *Riemannian metric*. From this notion of curve length, we can define the associated distance that measures the length of the shortest path, also known as the *geodesic*,

$$\mathrm{dist}(\mathbf{z}_0, \mathbf{z}_1) = \mathrm{Length}[\gamma^*], \qquad \text{where} \tag{7}$$

$$\begin{aligned} \gamma^* &= \operatorname*{argmin}_\gamma \mathrm{Length}[\gamma] \\ \text{s.t.} \quad & \gamma_0 = \mathbf{z}_0 \text{ and } \gamma_1 = \mathbf{z}_1. \end{aligned} \tag{8}$$

This distance measure does not change if we reparametrize the latent space by some diffeomorphism $h : \mathcal{Z} \to \mathcal{Z}$. This construction can be expanded upon to allow for reparametrization invariant measurements of volumes, angles, and more (see Hauberg (2023) for details).

2.3. Topology estimation and the role of uncertainty

Training data is inherently finite, suggesting that we should only expect to be able to learn a compact manifold. Further, it is not unreasonable to expect that the underlying manifold near which the data are distributed can have holes. These considerations lead to a mismatch between the topology of the manifold we seek to estimate and the Euclidean topology of the latent space.

In rare cases, we may have prior topological information about the underlying manifold and we can adapt the latent space accordingly (Davidson et al., 2018; Mathieu et al., 2019). Generally, we, however, must estimate the underlying manifold's topology if we are to reliably estimate its geometry.

One approach to topology estimation is to cover the manifold using multiple charts and learn diffeomorphisms that connect these (Kalatzis et al., 2021; Schonsheck et al., 2019). This, however, notably complicates model estimation, and the approach is rarely followed in practice.

Hauberg (2018) argues that model uncertainty offers a topological hint. The intuition is that the decoder should have high uncertainty in regions of the latent space with little support from training data (i.e. *outside the manifold*). One approach to incorporating model uncertainty into the geometry is to consider the *expected Riemannian metric* (Tosi et al., 2014),

$$\mathbb{E}[\mathbf{G}_\mathbf{z}] = \mathbb{E}[\mathbf{J}_\mathbf{z}^\intercal \mathbf{J}_\mathbf{z}] = \mathbb{E}[\mathbf{J}_\mathbf{z}]^\intercal \mathbb{E}[\mathbf{J}_\mathbf{z}] + \mathrm{cov}(\mathbf{J}_\mathbf{z}) \tag{9}$$

such that distances are larger in regions of high uncertainty (Hauberg, 2018). The property ensures that geodesics stay close to the training data (Fig. 1). The expected metric has been analyzed in great detail when the decoder follows a posterior Gaussian process (Pouplin et al., 2023).

To the best of our knowledge, the expected metric has only been explored for decoders following Gaussian processes and not for neural network decoders. To shape the metric to take large values outside the data support, Arvanitidis et al. (2018) suggests taking the variance of the conditional likelihood $p(\mathbf{x}|\mathbf{z})$ into account. Specifically, for a Gaussian conditional likelihood, Arvanitidis et al. suggests the metric

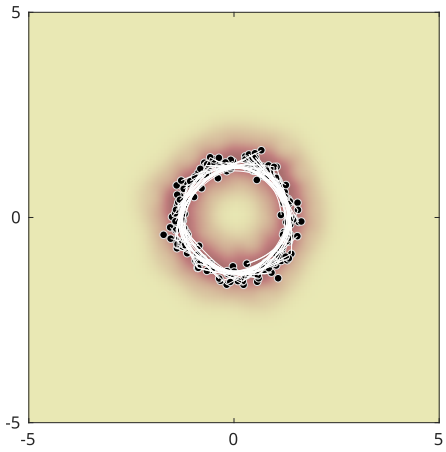

Figure 1. Shortest paths (geodesics) under the expected metric of a decoder following a Gaussian process. The topological hint of uncertainty is, thus, propagated to the metric. *Figure is courtesy of Hauberg (2018).*

$$\mathbf{G} = \mathbf{J}_\mu^\intercal \mathbf{J}_\mu + \mathbf{J}_\sigma^\intercal \mathbf{J}_\sigma. \tag{10}$$

Assuming $\sigma^2(\mathbf{z})$ grows with the distance to training data, then this metric will give rise to geodesics that approach the data. Unfortunately, the neural network $\sigma : \mathcal{Z} \to \mathcal{X}$ does not exhibit such growth on its own, and Arvanitidis et al. (2018) heuristically proposed to model σ^{-2} with a *radial basis function neural network* (Que & Belkin, 2016), which provides such growth. Variants of this heuristic are commonly applied when using learned latent geometries (Arvanitidis et al., 2022; Detlefsen et al., 2022; 2019; Beik-Mohammadi et al., 2021).

2.4. Computing geodesics

There are several ways to compute the geodesic that connects two points. The classic approach amounts to solving the geodesic differential equation as a two-point boundary value problem (Hauberg et al., 2012; Arvanitidis et al., 2019; Miller et al., 2006). This works well for low-curvature manifolds, such as spheres and tori, but is generally unstable on learned manifolds. On low-dimensional manifolds, we can alternatively discretize the manifold into a graph and apply classic algorithms for computing shortest paths on graphs (Beik-Mohammadi et al., 2021). The size of such a graph, however, grows exponentially with the manifold dimension, and the approach is impractical beyond three dimensions.

A more practical approach is to note that minimizers of *curve length* coincides with those of *curve energy* (Carmo,

1992),

$$\mathcal{E}[\gamma] = \int_0^1 \left\| \frac{\mathrm{d}}{\mathrm{d}t} f(\gamma_t) \right\|^2 \mathrm{d}t, \tag{11}$$

which follows from the Cauchy-Schwarz inequality. Minimizing curve energy has the benefit of yielding solution curves with constant speed (Carmo, 1992). This energy can easily be discretized as

$$\mathcal{E}[\gamma] \approx \sum_{t=0}^{T-1} \left\| f(\gamma(t+1/T)) - f(\gamma(t/T)) \right\|^2 \mathrm{d}t. \tag{12}$$

A simple algorithm for computing geodesics then amounts to parametrizing the curve γ and minimizing the discretized energy (12) with respect to the curve parameters. Shao et al. (2018) propose parametrizing the curve as a discrete set of points, Yang et al. (2018) use a second-order polynomial, while Detlefsen et al. (2021) use cubic splines. In our implementation, we opt for the latter.

3. Ensemble of decoders

To capture the model uncertainty of a VAE, we need to access the posterior distribution over the model parameters θ. Since the encoder is not part of the model, but rather an amortization mechanism for the variational inference, we are only interested in the posterior of the decoder weights, $p(\theta|\mathcal{D})$, where $\mathcal{D}$ denotes the training data.

In practice, current Bayesian deep learning techniques often struggle to approximate the posterior over the weights. We, therefore, propose to approximate posterior samples with a *deep ensemble* (Lakshminarayanan et al., 2017), which can, heuristically, be seen as a Bayesian approximation (Gustafsson et al., 2020).

This can trivially be implemented by instantiating S randomly initialized decoders $\{f_{\theta_s}\}_{s=1}^S$. For each mini-batch of data, we randomly sample a decoder f_{θ_s} and take a gradient step to optimize the ELBO $\mathcal{L}_{\theta_s,\psi}$. At convergence, we have access to one encoder and S decoders.

Figure 2 shows an example of the uncertainty of an ensemble of decoders. For ease of visualization, we consider a VAE with a two-dimensional latent space trained on three classes of MNIST. We show the uncertainty in the latent space, which we have calculated as the mean over n pixel standard deviations

$$\text{uncertainty}(\mathbf{z}') = \frac{1}{n} \sum_{i=1}^{n} \sigma_i(\mathbf{z}') \tag{13}$$

with

$$\sigma(\mathbf{z}') = \sqrt{\frac{1}{S} \sum_{j=1}^{S} (f_{\theta_j}(\mathbf{z}') - \mu(\mathbf{z}'))^2}$$

$$\mu(\mathbf{z}') = \frac{1}{S} \sum_{j=1}^{S} f_{\theta_j}(\mathbf{z}')$$

We see that uncertainty generally grows with the distance to the latent representations as one would naturally expect. This gives hope that this uncertainty can be used to shape the Riemannian metric to better respect topology.

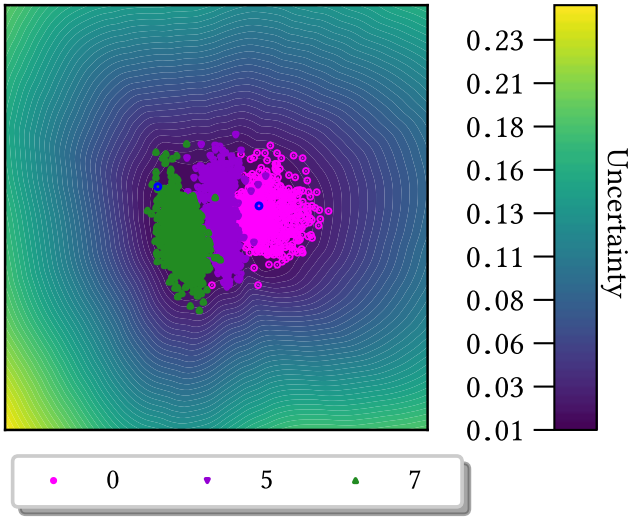

Figure 2. Using an ensemble of decoders ensures that regions of the latent space with limited data support have high uncertainty.

4. Ensemble geodesics

To obtain a practical latent geometry from the decoder ensemble, we think of this as samples from an approximate posterior $f_\theta \sim q(\theta)$. We may, thus, construct an *expected metric* as $\mathbf{G} = \mathbb{E}_{q(\theta)}[\mathbf{J}_{f_\theta}^\intercal \mathbf{J}_{f_\theta}]$. Under this metric, we see that the energy of a curve γ becomes

$$\mathcal{E}[\gamma] = \int_0^1 \mathbb{E}_{q(\theta)} \left[\dot{\gamma}_t^\intercal \mathbf{J}_{f_\theta}^\intercal \mathbf{J}_{f_\theta} \dot{\gamma}_t \right] \mathrm{d}t, \tag{14}$$

and following the discretization of Eq. 12 we get

$$\mathcal{E}[\gamma] \approx \sum_{t=0}^{T-1} \mathbb{E}_{q(\theta)} \left[\left\| f_\theta(\gamma(t+1/T)) - f_\theta(\gamma(t/T)) \right\|^2 \right].$$

Empirically, we have found that geodesics that minimize this discretized energy do not follow the data as closely as one could hope for. We hypothesize that the decoder ensemble *underestimate* model uncertainty since all ensemble

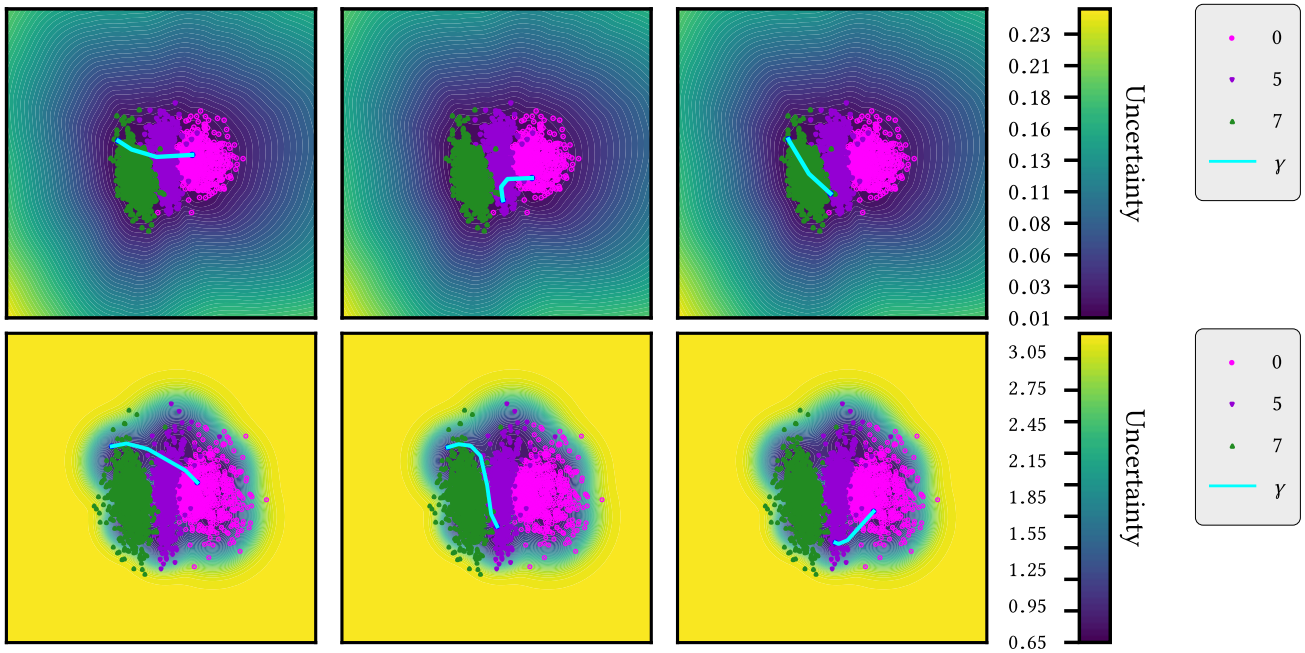

Figure 3. **Upper row:** Three examples of the latent space for ensembles of VAE decoders on a reduced version of MNIST data with *three* classes. Blue curves indicate the geodesic interpolants between two random latent coordinates. **Lower row:** Three examples of the latent space for a VAE with RBF-generated uncertainties on MNIST data with *three* classes.

members are trained always on the same data points. Similar issues have been previously observed with other classical ensemble-based models, i.e. *bootstrap* methods (Efron & Gong, 1983). To counter this, we will modify the energy to amplify the impact of model uncertainty by disregarding correlations.

Particularly, the discretized energy sums expected squared distances $\mathbb{E}[\|f(\mathbf{z}_2) - f(\mathbf{z}_1)\|^2]$. To analyze this, we introduce the short-hand notation $\mathbf{x}_i = f_\theta(\mathbf{z}_i)$ and write the moments of a pair $(\mathbf{x}_1, \mathbf{x}_2)$ as

$$\mathbb{E}_{q(\theta)}\left[\begin{pmatrix}\mathbf{x}_1\\ \mathbf{x}_2\end{pmatrix}\right] = \begin{pmatrix}\mu_1\\ \mu_2\end{pmatrix}, \tag{15}$$

$$\mathrm{cov}\left[\begin{pmatrix}\mathbf{x}_1\\ \mathbf{x}_2\end{pmatrix}\right] = \begin{pmatrix}\Sigma_{11} & \Sigma_{12}\\ \Sigma_{12} & \Sigma_{22}\end{pmatrix}. \tag{16}$$

The difference vector $\Delta = \mathbf{x}_2 - \mathbf{x}_1$ will thus have moments

$$\mathbb{E}[\Delta] = \mathbb{E}[\mathbf{x}_2] - \mathbb{E}[\mathbf{x}_1], \tag{17}$$

$$\mathrm{cov}[\Delta] = \Sigma_{11} + \Sigma_{22} - 2\Sigma_{12}, \tag{18}$$

and the individual summands of the discretized energy are then of the following form

$$\mathbb{E}\left[\|\Delta\|^2\right] = \|\mathbb{E}[\Delta]\|^2 + \mathrm{tr}[\Sigma_{11} + \Sigma_{22}] - 2\mathrm{tr}[\Sigma_{12}]. \tag{19}$$

We explicate these expressions to emphasize that cross-covariances between points along γ decrease the curve energy. Neural network ensembles are known to provide better performance under *de-correlated* predictions, which is *de facto* a way to promote higher degrees of ensemble diversity (Lakshminarayanan et al., 2017; Lee et al., 2016).

Additionally, the correlation terms in posterior cross-covariances in other probabilistic models like *Gaussian processes* (Williams & Rasmussen, 2006), also collapse to zero values as the size of difference vector Δ augments (see Figure 4). This primarily indicates that the discretized energy can be also negatively affected by spurious cross-covariance terms whenever the difference is not sufficiently small given the high flexibility of the ensemble neural networks.

In practice, all of this suggests that cross-covariance terms like Σ_{12} in Eq. 19 are not beneficial for the minimization of the discretized energy with ensembles and we drop them, i.e.

$$\mathbb{E}\left[\|\Delta\|^2\right] \approx \|\mathbb{E}[\Delta]\|^2 + \mathrm{tr}[\Sigma_{11} + \Sigma_{22}]. \tag{20}$$

This can be practically implemented by evaluating the energy directly as

$$\mathcal{E}[\gamma] \approx \sum_{t=0}^{T-1} \mathbb{E}_{\theta,\theta' \sim q(\theta)q(\theta)}\left[\left\|f_\theta(\gamma(t+1/T)) - f_{\theta'}(\gamma(t/T))\right\|^2\right]$$

When minimizing this energy, we use a simple one-sample Monte Carlo estimate.

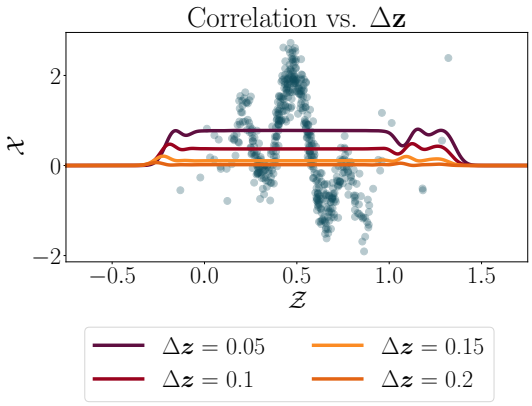

Figure 4. The *correction* term of posterior covariances in a GP tends to be zero as $\Delta\mathbf{z} \gg 0$, even in areas of $\mathcal{Z}$ where is a high-density of training data.

5. Experiments

We compare our method to the current state-of-the-art approach to model uncertainty when learning latent geometries Arvanitidis et al. (2018), which relies on RBF networks to model data uncertainty. In particular, we show that geodesic distances stemming from our ensemble of decoders method are more stable under retraining when compared to distances learned using the RBF neural network. The implementation of our method and code producing the experimental results is available at https://github.com/mustass/ensertainty.

For both approaches, we choose a VAE architecture with dense layers whereas for the RBF neural network part we use a mixture of 10 Gaussians in the latent space. We train the models on the MNIST and FMNIST datasets with two-dimensional latent spaces. We further extend the MNIST analysis to 50-dimensional latents.

We retrain the VAEs using 30 different seeds on both datasets. We subsequently calculate the geodesic distances between the latent representations of 100 pairs of randomly chosen data points from the test set. These points are fixed across all trials. The outcome is 30 measurements per point pair for both methods. This allows us to calculate the coefficient of variation (CV) for each method to compare their variability and, thus, robustness,

$$\text{CV} = \frac{\sigma}{\mu}, \tag{21}$$

where μ and σ are the mean and standard deviations of the distances calculated for the same point pair by 30 different estimations of a model. Note that CV is a unitless measure of relative variability, where a lower value indicates less variability and, thus, allows us to compare the variability of values on different scales.

Table 1 shows results for the one-sided paired Student's t-test with the null hypothesis of ensemble geodesics having a lower coefficient of variation than by using the RBF-based model. The results show that the ensemble of decoders is consistently more reliable than the RBF-based approach, which is currently the most popular approach. Figure 5 visualizes the findings using a histogram of coefficients of variation for different point pairs.

6. Conclusion

Learned latent geometries crucially rely on uncertainty estimation in order to shape the metric according to the underlying manifold's topology (Hauberg, 2018). In Gaussian process models (Tosi et al., 2014; Pouplin et al., 2023) this construction naturally comes in place, but models based on neural networks have required a series of heuristics to behave desirable (Arvanitidis et al., 2018). Unfortunately, these heuristics work poorly beyond a few latent dimensions.

We have proposed to use neural network ensembles to capture model uncertainty. We have shown that this leads to empirical improvements compared to current heuristics. In practice, training ensembles of decoders requires only small code modifications and our proposed approach is generally easy to implement.

Acknowledgments

This work was supported by a research grant (42062) from VILLUM FONDEN. This project received funding from the European Research Council (ERC) under the European Union's Horizon research and innovation programme (grant agreement 101125993). The work was partly funded by the Novo Nordisk Foundation through the Center for Basic Machine Learning Research in Life Science (NNF20OC0062606).

Dataset	Number of classes	d	Null-hypothesis	Alternative	t-statistic	p-value
MNIST-3	3	2	Ensemble geodesics have lower CV	greater	-16.834	1.000
MNIST	10	2	Ensemble geodesics have lower CV	greater	-16.290	1.000
FMNIST	10	2	Ensemble geodesics have lower CV	greater	-15.339	1.000
MNIST	10	50	Ensemble geodesics have lower CV	greater	-6.472	0.999

Table 1. Statistical metrics (p-value and t-statistic) for MNIST and FMNIST with ensembles and RBF uncertainties.

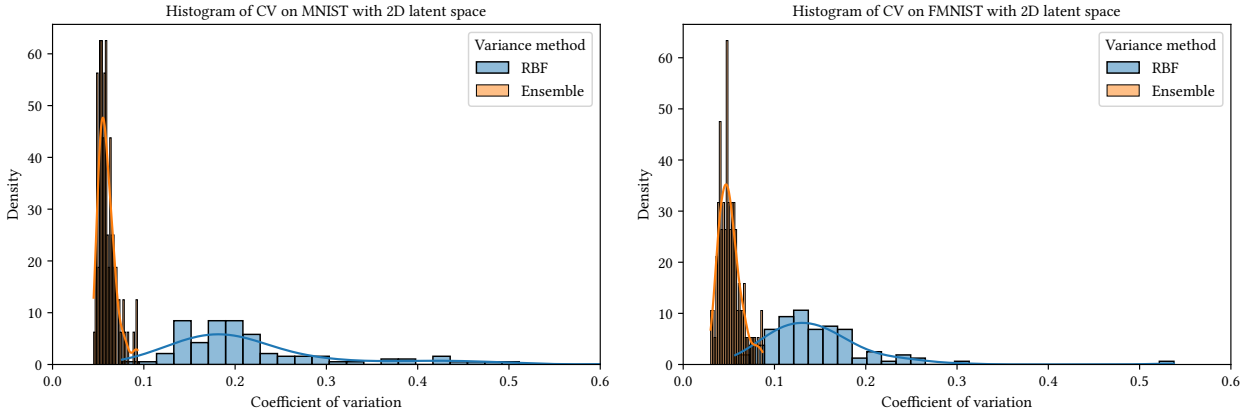

Figure 5. Histogram of coefficients of variation for MNIST and FMNIST data with $d = 2$ in the latent space $\mathcal{Z}$.

References

Arvanitidis, G., Hansen, L. K., and Hauberg, S. Latent space oddity: on the curvature of deep generative models. In *International Conference on Learning Representations (ICLR)*, 2018.

Arvanitidis, G., Hauberg, S., Hennig, P., and Schober, M. Fast and robust shortest paths on manifolds learned from data. In *Artificial Intelligence and Statistics (AISTATS)*, 2019.

Arvanitidis, G., Hauberg, S., and Schölkopf, B. Geometrically Enriched Latent Spaces. In *Artificial Intelligence and Statistics (AISTATS)*, 2021.

Arvanitidis, G., González-Duque, M., Pouplin, A., Kalatzis, D., and Hauberg, S. Pulling back information geometry. In *Artificial Intelligence and Statistics*, 2022.

Beik-Mohammadi, H., Hauberg, S., Arvanitidis, G., Neumann, G., and Rozo, L. Learning riemannian manifolds for geodesic motion skills. In *Robotics: Science and Systems (R:SS)*, 2021.

Bengio, Y., Courville, A., and Vincent, P. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35(8): 1798–1828, 2013.

Carmo, M. P. d. *Riemannian geometry*. Birkhäuser, 1992.

Chen, N., Klushyn, A., Paraschos, A., Benbouzid, D., and van der Smagt, P. Active learning based on data uncertainty and model sensitivity, 2018.

Davidson, T. R., Falorsi, L., De Cao, N., Kipf, T., and Tomczak, J. M. Hyperspherical variational auto-encoders. *arXiv preprint arXiv:1804.00891*, 2018.

Daxberger, E. and Hernández-Lobato, J. M. Bayesian variational autoencoders for unsupervised out-of-distribution detection. *arXiv preprint arXiv:1912.05651*, 2019.

Detlefsen, N. S., Jørgensen, M., and Hauberg, S. Reliable training and estimation of variance networks. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2019.

Detlefsen, N. S., Pouplin, A., Feldager, C. W., Geng, C., Kalatzis, D., Hauschultz, H., González-Duque, M., Warburg, F., Miani, M., and Hauberg, S. Stochman. *GitHub. Note: https://github.com/MachineLearningLifeScience/stochman/*, 2021.

Detlefsen, N. S., Hauberg, S., and Boomsma, W. Learning meaningful representations of protein sequences. *Nature Communications*, 13(1):1–12, 2022.

Efron, B. and Gong, G. A leisurely look at the bootstrap, the jackknife, and cross-validation. *The American Statistician*, 37(1):36–48, 1983.

Gustafsson, F. K., Danelljan, M., and Schon, T. B. Evaluating scalable bayesian deep learning methods for robust computer vision. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pp. 318–319, 2020.

Hansen, L. K. and Salamon, P. Neural network ensembles. *IEEE transactions on pattern analysis and machine intelligence*, 12(10):993–1001, 1990.

Hauberg, S. Only bayes should learn a manifold. 2018.

Hauberg, S. *Differential geometry for generative modeling*. 2023.

Hauberg, S., Freifeld, O., and Black, M. J. A geometric take on metric learning. In *Advances in Neural Information Processing Systems (NeurIPS) 25*, 2012.

Ho, J., Jain, A., and Abbeel, P. Denoising diffusion probabilistic models. *Advances in neural information processing systems*, 33:6840–6851, 2020.

Kalatzis, D., Eklund, D., Arvanitidis, G., and Hauberg, S. Variational autoencoders with riemannian brownian motion priors. In *Proceedings of the 37th International Conference on Machine Learning (ICML)*, volume 119 of *Proceedings of Machine Learning Research (PMLR)*, 2020.

Kalatzis, D., Ye, J. Z., Wohlert, J., and Hauberg, S. Multi-chart flows, 2021.

Kingma, D. P. and Welling, M. Auto-Encoding Variational Bayes. In *International Conference on Learning Representations, ICLR*, 2014.

Lakshminarayanan, B., Pritzel, A., and Blundell, C. Simple and scalable predictive uncertainty estimation using deep ensembles. *Advances in neural information processing systems*, 30, 2017.

Lee, S., Purushwalkam Shiva Prakash, S., Cogswell, M., Ranjan, V., Crandall, D., and Batra, D. Stochastic multiple choice learning for training diverse deep ensembles. *Advances in Neural Information Processing Systems (NIPS)*, 29, 2016.

Lipman, Y., Chen, R. T., Ben-Hamu, H., Nickel, M., and Le, M. Flow matching for generative modeling. *arXiv preprint arXiv:2210.02747*, 2022.

Mathieu, E., Le Lan, C., Maddison, C. J., Tomioka, R., and Teh, Y. W. Continuous hierarchical representations with poincaré variational auto-encoders. *Advances in neural information processing systems*, 32, 2019.

Miller, M. I., Trouvé, A., and Younes, L. Geodesic shooting for computational anatomy. *Journal of mathematical imaging and vision*, 24:209–228, 2006.

Pouplin, A., Eklund, D., Ek, C. H., and Hauberg, S. Identifying latent distances with finslerian geometry. Transactions on Machine Learning Research (TMLR), 2023.

Que, Q. and Belkin, M. Back to the future: Radial basis function networks revisited. In *Artificial Intelligence and Statistics (AISTATS)*, pp. 1375–1383, 2016.

Rezende, D. J., Mohamed, S., and Wierstra, D. Stochastic Backpropagation and Approximate Inference in Deep Generative Models. In *Proceedings of the 31st International Conference on Machine Learning (ICML)*, 2014.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P., and Ommer, B. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10684–10695, 2022.

Scannell, A., Ek, C. H., and Richards, A. Trajectory Optimisation in Learned Multimodal Dynamical Systems Via Latent-ODE Collocation. In *Proceedings of the IEEE International Conference on Robotics and Automation*. IEEE, 2021.

Schonsheck, S., Chen, J., and Lai, R. Chart autoencoders for manifold structured data. *arXiv preprint arXiv:1912.10094*, 2019.

Shao, H., Kumar, A., and Thomas Fletcher, P. The riemannian geometry of deep generative models. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pp. 315–323, 2018.

Sohl-Dickstein, J., Weiss, E., Maheswaranathan, N., and Ganguli, S. Deep unsupervised learning using nonequilibrium thermodynamics. In *International conference on machine learning*, pp. 2256–2265. PMLR, 2015.

Tipping, M. E. and Bishop, C. M. Probabilistic principal component analysis. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 61(3):611–622, 1999.

Tomczak, J. and Welling, M. Vae with a vampprior. In *International conference on artificial intelligence and statistics*, pp. 1214–1223. PMLR, 2018.

Tosi, A., Hauberg, S., Vellido, A., and Lawrence, N. D. Metrics for Probabilistic Geometries. In *Uncertainty in Artificial Intelligence (UAI)*, 2014.

Williams, C. K. and Rasmussen, C. E. *Gaussian processes for machine learning*, volume 2. MIT press Cambridge, MA, 2006.

Yang, T., Arvanitidis, G., Fu, D., Li, X., and Hauberg, S. Geodesic clustering in deep generative models. In *arXiv preprint*, 2018.