https://doi.org/10.2352/EI.2025.37.14.COIMG-141

Differential Geometric View of Information Flow in Neural Nets

Suhas Sreehari [1,3]***, Pradeep Ramuhalli*** [1]***, Frank Liu*** [2]
[1] ***Oak Ridge National Lab, Oak Ridge, TN;*** [2] ***Old Dominion University, Norfolk, VA;*** [3] ***University of Tennessee, Knoxville***

Abstract

In this paper, we explore a space-time geometric view of signal representation in machine learning models. The question we are interested in is if we can identify what is causing signal representation errors – training data inadequacies, model insufficiencies, or both. Loosely expressed, this problem is stylistically similar to blind deconvolution problems. However, studies of space-time geometries might be able to partially solve this problem by considering the curvature produced by mass in (Anti-)de Sitter space. We study the effectiveness of our approach on the MNIST dataset.

Introduction

With the advent of large language models, and foundation models in general, it is important to ensure robust signal representation inside these models. It is important for the following key reasons:

- The raw data that foundation models train on is enormous, and as such the model graphs need an efficient mechanism to distil the data, without the representation sparsity compromising on signal fidelity.
- An incomplete representation (typically caused by incomplete data) leads to various problems during inference.
- Skewed data representations cause lack of generalizability, characterized by highly non-uniform variance in model accuracy.

In addition, achieving a mechanism for representation robustness also enables us to specify specific training data inadequacies. A central question we ask in this paper is whether a signal representation error is due to sub-optimal modeling choices or due to incomplete/unreliable data (or a combination). Akin to active learning, can we have the system send the user prescriptive feedback that helps augment data that is most likely to alleviate representation inaccuracies?

The performance of a neural network can be considered a function of two key variables: (1) network inadequacy (I_n) and (2) data inadequacy (I_d). Without even running model inference, we must be able to predict the network performance with just I_n and I_d.[1]

In this paper, we address this problem by considering latent space representations in generative neural networks. In terms of identifying missing data, there are some key advantages to working with latent spaces, including that it is easier to find gaps in latent spaces than to find them directly in input spaces. In auto-encoder type architectures, estimating data gaps from latent space gaps is simply a matter of computing the encoder inverse on the latter.

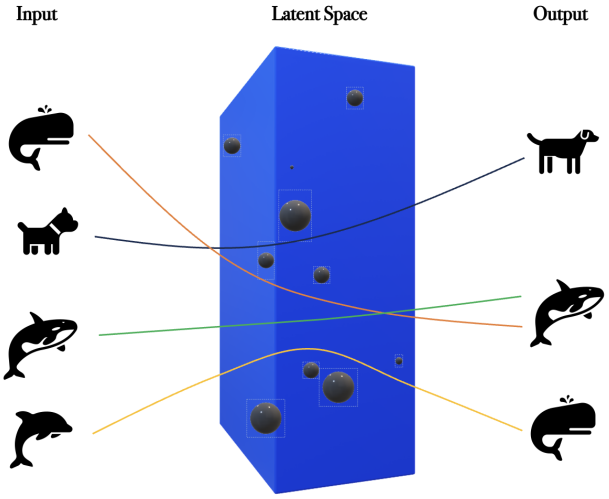


Figure 1. Curvature in latent space. *Illustration of both intrinsic and mass-induced curvature in the latent space. Bigger mass-like points produce higher bend in the information lines. Further, the cosmological constant captures curvature even in the absence of mass.*

If we consider the latent space as a tensor, we can also represent the inadequacies in the latent space as another tensor, $I_i = \phi(I_n, I_d)$, for some function ϕ. Much of this paper is dedicated to constructing a framework in which a convenient and accurate ϕ can be defined.

The novelty of our work lies in how we model the latent space using principles of differential geometry. In particular, A. Anandkumar et al [1] showed that parameter estimation for latent variable models can be reduced to finding orthogonal decompositions of tensors derived from second- and third-order moments. These decompositions are generally intractable, but we use constructs from general relativity (specifically de-Sitter and anti de-Sitter curvature tensors) to leverage rich mathematical solutions pre-existing in cosmology.

We characterize training data inadequacies using mass-like points in the latent space, which we can estimate by way of calculating the stress-energy tensor in Einstein's field equations. Furthermore, we characterize model inadequacies using intrinsic space-time curvature, which we can estimate by way of the cosmological constant (of non-flat space).

Figure 1 is an illustration of information flow through a latent space with mass-like points and the resulting distortion in signal reconstruction at the decoder.

The rest of the paper is organized as follows. In Section , we

[1]The upper limit on performance corresponds to theoretical minima of I_n and I_d, but this limit might still not represent complete accuracy due to data noise. Analysis of this theoretical upper bound on performance is conditioned on the data and the specific task, and is beyond the scope of our paper.

will briefly outline related works that are relevant to this problem. In Section , we use space-time geometry to understand the identification of and decoupling between data and model specification inadequacies. In Section , we summarize the results followed by a brief discussion and future work.

Related Work

Model Shortcomings. The authors of [5] advanced our understanding of the difficulty of training deep feedforward neural networks shed light on the vanishing and exploding gradient problems in deep networks. [6] proposed an LSTM architecture to combat the vanishing gradient problem, which limited the training of deep recurrent neural networks. [11] introduced dropout regularization as a technique to prevent overfitting in deep neural networks, which often occurs when the model is too complex relative to the training data.

Training Data Shortcomings. [3] discuss strategies for dealing with label noise in deep learning datasets. Label noise can significantly impact model performance and generalization. [2] explore data augmentation techniques to augment training data, making models more robust to variations in the data distribution. [12] investigate transfer learning as a method to leverage pre-trained models and address data shortcomings in specific domains.

To the best of our knowledge, none of these works use space-time curvature to jointly estimate network and training data shortcomings.

Analogy Between Latent Space and Space-Time

Information Flow as Curvatures in the Latent Space

Without loss of generality, we can describe the encoder-decoder distortions via curvatures in the latent space. Similar to ideas from the general theory of relativity [4], we can visualize bending of information lines as they flow in from the encoder and out to the decoder. Specifically, the latent space can now be modeled as a set of *mass-like* points through which the information lines distort (leading to loss of faithful signal representation).

The general geometric justification for such an analogy can be analyzed briefly.

Theorem 1 (Manifold Continuity of Autoencoder Transformations)**.** *The transformation pathway T defined by an autoencoder from the input space $\mathscr{I}$ to its latent space $\mathscr{L}$ and then to the output space $\mathscr{O}$, where $\mathscr{I},\mathscr{L},\mathscr{O}$ are treated as differentiable manifolds, exhibits a smooth manifold structure under continuous and differentiable mapping conditions.*

Proof. Assume $f:\mathscr{I}\to\mathscr{L}$ and $g:\mathscr{L}\to\mathscr{O}$ are differentiable maps. The composition $g\circ f$, representing the autoencoder, is differentiable by the chain rule for compositions of differentiable functions:

$$(g\circ f)' = g'\circ f'.$$

This differentiability ensures that the mappings induce a smooth structure between the manifolds, preserving the continuity and differentiability from $\mathscr{I}$ through $\mathscr{L}$ to $\mathscr{O}$. □

Theorem 2 (Analogy between Information Flow and Spacetime Curvature)**.** *The distortion in the reconstruction of data by an autoencoder, measured by a distortion metric D, is analogous to the spacetime curvature R described by the Ricci curvature tensor in the presence of a stress-energy tensor T.*

Proof. Define the distortion metric $D(x,g(f(x))) = \|x - g(f(x))\|^2$, quantifying the error at each point $x\in\mathscr{I}$. Let's draw an analogy to the Ricci curvature tensor R_{ij} in general relativity, which measures how mass-energy affects spacetime curvature:

$$R_{\mu\nu} - \frac{1}{2}Rg_{\mu\nu} + \Lambda g_{\mu\nu} = \frac{8\pi G}{c^4}T_{\mu\nu},$$

where R is the Ricci scalar. We suggest that increases in data complexity or 'mass' in the latent representation increase the curvature of the information flow, similar to the curvature of spacetime due to mass:

$$D \approx \text{curvature induced by } f \text{ and } g.$$

□

Theorem 3 (Topological Properties of Information Lines)**.** *The paths of information flow in an autoencoder, seen as continuous lines from $\mathscr{I}$ through $\mathscr{L}$ to $\mathscr{O}$, maintain homotopy equivalence under transformations that preserve data integrity, analogous to paths in a curved spacetime manifold being homotopic to geodesics.*

Proof. Consider paths in $\mathscr{I}$ and $\mathscr{O}$ that are images under f and g. Assuming these mappings are homotopic, implying one can be continuously deformed into the other, they preserve the fundamental group structure under the transformations:

$$\pi_1(\mathscr{I}) \cong \pi_1(\mathscr{O}) \text{ through } \pi_1(\mathscr{L}).$$

This homotopy equivalence ensures that paths are preserved akin to geodesics in spacetime under the influence of gravity. □

These theorems establish a foundational analogy that not only enhances the understanding of autoencoder dynamics but also bridges a gap between machine learning algorithms and theoretical physics. The appeal of such a treatment of the latent space is that this allows for two useful characterizations of the space: 1) a finite-point description of the space, and 2) a mechanism to quantify both model and data inadequacies.

Imagine the latent space can be characterized by two independent components – the training data and the model specifications. In this view of the latent space as space-time geometry, we would ideally like this space to be void of any mass-like fields, and therefore of any mass-like points. These mass-like points denote data inadequacies in the neighborhood of their pre-image in the encoder's input space. Having mass-like points in the latent space will bend the lines of information, causing distortion, thereby losing robustness of signal representation. Same happens when the space itself in intrinsically curved. We model the model inadequacies as the curvature of empty space without any mass. We use de Sitter (dS) and Anti-de Sitter (AdS) solutions (corresponding to positive and negative cosmological constant, respectively) using a Gaussian process prior.

The existence of three symmetric spaces is entirely analogous to the the three different solutions. Note that de Sitter and anti-de Sitter both have constant *spacetime* curvature, supplied by the cosmological constant. The metrics above have constant *spatial* curvature. Note, however, that the metric on $\mathbf{S}^3$ coincides with the spatial part of the de Sitter metric in coordinates, while the metric on $\mathbf{H}^3$ coincides with the spatial part of the adS metric in the coordinates.

We write these spatial metrics in unified form,

$$ds^2 = \gamma_{ij}dx^i dx^j = \frac{dr^2}{1-kr^2} + r^2(d\theta^2 + \sin^2\theta d\phi^2), \tag{1}$$

where $k = +1$ for $\mathbf{S}^3$, 0 for $\mathbf{R}^3$, and -1 for $\mathbf{H}^3$.

Hyperboloid $\mathbf{H}^3$. This space has a uniform negative curvature,

$$ds^2 = \frac{dr^2}{1+r^2} + r^2(d\theta^2 + \sin^2\theta d\phi^2) \tag{2}$$

Information distortion through latent space as Ricci curvatures

In the context of encoder-decoder architectures, particularly those used in machine learning and signal processing, the concept of *latent space* plays a crucial role in understanding the transformations applied to input data. As data passes through the encoder, it is compressed into a lower-dimensional representation, known as the latent space representation. This process, while efficient for reducing dimensionality and capturing the essential features of the data, often introduces distortions. The decoder's task is to reconstruct the original data from its latent representation. Ideally, if there is no distortion, the output of the decoder should match the input of the encoder exactly. However, due to the lossy nature of the compression, some information is inevitably lost, leading to a discrepancy between the original input and the reconstructed output.

To quantify this distortion, we define it in terms of the variance between the output of the decoder, $\hat{x}$, and the input of the encoder, x. Mathematically, the distortion can be expressed as the expected value of the squared difference between x and $\hat{x}$:

$$D(x,\hat{x}) = \mathbb{E}[(x-\hat{x})^2]. \tag{3}$$

This variance serves as a measure of the reconstruction error, providing a quantitative means to evaluate the fidelity of the data reconstruction process. In an ideal scenario, where the decoder perfectly reconstructs the input from the latent representation, the distortion $D(x,\hat{x})$ would be zero. However, in practical applications, minimizing this distortion is a key challenge, guiding the design and optimization of encoder-decoder architectures to achieve as close a match as possible between x and $\hat{x}$.

The process of encoding and decoding information through a latent space can introduce variances between the original and reconstructed data, akin to distortions or "bends" in the information space. These bends, characterized by the variance $D(x,\hat{x}) = \mathbb{E}[(x-\hat{x})^2]$, can be conceptually linked to the curvature in spacetime described by general relativity. In the realm of general relativity, mass and energy influence the curvature of spacetime, which in turn affects the paths of objects moving through it. This curvature is mathematically described by the Ricci tensor, $R_{\mu\nu}$, which provides a way to quantify how much the geometry of the space deviates from flat Euclidean space in the presence of mass-energy.

To draw a parallel, consider the latent space representation as a form of "information spacetime" where data insufficiencies or the compressed nature of the information act as mass points, creating curvature. This curvature leads to the bending of information lines, analogous to the bending of light or trajectories of particles in gravitational fields. The Ricci tensor, in this analogy, can be related to the distortion variance through the notion that higher curvature (greater distortion in the latent space) corresponds to a greater variance between the original input and the output of the decoder. Mathematically, this relationship can be expressed as:

$$R_{\mu\nu} \sim \nabla^2 D(x,\hat{x}), \tag{4}$$

where ∇^2 denotes the Laplacian operator, symbolizing the spread or divergence of distortion across the latent space, analogous to how the Ricci tensor measures the density of the curvature of spacetime due to mass-energy content.

This conceptual framework not only enhances our understanding of information processing in neural networks but also provides a fascinating bridge between the fields of machine learning and theoretical physics. By exploring the similarities between the distortion of information in latent spaces and the curvature of spacetime, we gain insights into the fundamental nature of information, compression, and differential geometry.

In the analysis of encoder-decoder architectures within machine learning, the variance $D(x,\hat{x})$ quantifies the distortion between the input x to the encoder and the output $\hat{x}$ from the decoder. This variance reflects the degree of information loss and can be visualized as a curvature or bend in the latent space. Analogously, in the framework of general relativity, the curvature of spacetime due to the presence of mass and energy is mathematically described by the Ricci curvature tensor. This document aims to establish a conceptual bridge between the curvature observed in the latent space of neural networks and the Ricci curvature in spacetime.

Quantifying Distortion in Latent Space The variance $D(x,\hat{x})$ serves as a measure of distortion, defined as:

$$D(x,\hat{x}) = \mathbb{E}[(x-\hat{x})^2], \tag{5}$$

where x represents the original input data, and $\hat{x}$ denotes the reconstructed data from the decoder. This variance not only quantifies the reconstruction error but also metaphorically represents the "curvature" or bend in the latent space induced by the encoding and decoding processes.

From Latent Space Curvature to Ricci Curvature In general relativity, the Ricci curvature tensor, $R_{\mu\nu}$, is a key entity that describes the curvature of spacetime as influenced by mass and energy. It is derived from the Riemann curvature tensor, $R^{\rho}_{\sigma\mu\nu}$, which provides a more comprehensive description of spacetime curvature. The Ricci tensor is obtained by contracting the Riemann tensor:

$$R_{\mu\nu} = R^{\rho}_{\mu\rho\nu}, \tag{6}$$

where the Ricci tensor essentially averages the effects of curvature across different directions. Additionally, the Ricci scalar, R, further

condenses this information into a single value, representing the trace of the Ricci tensor:

$$R = g^{\mu\nu} R_{\mu\nu}, \tag{7}$$

where $g^{\mu\nu}$ is the inverse metric tensor of spacetime.

Some Common Ideas Between information Flows in Networks and Light Flow in Space-Time

The distortion in the latent space, represented by $D(x,\hat{x})$, can be likened to the curvature of spacetime described by the Ricci tensor. Just as mass and energy dictate the curvature of spacetime, leading to the bending of light or trajectories of objects (akin to the curvature caused by gravitational fields), the data insufficiencies and compression in the latent space representation act as "mass points" in the informational geometry, causing information lines to bend. This analogy allows us to conceptualize the effects of data transformation in neural networks through the lens of spacetime geometry, providing a novel perspective on information processing and its inherent distortions.

Curvature of Latent Space in Autoencoders. Given an autoencoder with an encoder E and a decoder D, the latent space Z can be modeled as a Riemannian manifold with a metric tensor $g_{ij}(z)$. The curvature of this latent space is influenced by the density of encoded information.

Consider an autoencoder where $E : X \to Z$ maps the input space X to the latent space Z, and $D : Z \to X$ reconstructs the input from the latent representation.

Assume Z is a differentiable manifold. The metric tensor $g_{ij}(z)$ in Z is induced by the encoding map E from the input space metric $g_{kl}(x)$:

$$g_{ij}(z) = \frac{\partial E^k}{\partial z^i} \frac{\partial E^l}{\partial z^j} g_{kl}(x).$$

The Riemann curvature tensor R^i_{jkl} of the latent space, derived from g_{ij}, can be computed using:

$$R^i_{jkl} = \frac{\partial \Gamma^i_{jl}}{\partial z^k} - \frac{\partial \Gamma^i_{jk}}{\partial z^l} + \Gamma^i_{km}\Gamma^m_{jl} - \Gamma^i_{lm}\Gamma^m_{jk},$$

where Γ^i_{jk} are the Christoffel symbols derived from g_{ij}.

If we interpret $\rho(z)$ as the information density in the latent space, it influences the metric g_{ij} and thus the curvature. Regions with higher information density correspond to greater curvature, similar to how mass influences the curvature of spacetime in general relativity.

Geodesic Deviation in Latent Space. The deviation of paths in the latent space of an autoencoder is analogous to the geodesic deviation of light in the presence of mass in general relativity. Specifically, the separation vector between two encoded points evolves according to an equation similar to the geodesic deviation equation.

In general relativity, the geodesic deviation equation is:

$$\frac{D^2\xi^\mu}{d\tau^2} + R^\mu_{\nu\rho\sigma} u^\nu \xi^\rho u^\sigma = 0,$$

where ξ is the separation vector between geodesics, u^ν is the tangent vector, and $R^\mu_{\nu\rho\sigma}$ is the Riemann curvature tensor.

Consider two close points x_1 and x_2 in the input space, mapped to $z_1 = E(x_1)$ and $z_2 = E(x_2)$ in the latent space. The separation vector in the latent space is $\xi^i = z_2^i - z_1^i$.

The curvature of the latent space, influenced by the information density, affects the deviation of the paths of z_1 and z_2. The deviation equation in the latent space can be written as:

$$\frac{D^2\xi^i}{d\tau^2} + R^i_{jkl} \frac{dz^j}{d\tau} \xi^k \frac{dz^l}{d\tau} = 0,$$

where R^i_{jkl} is the Riemann curvature tensor of the latent space. This equation shows that the separation vector evolves under the influence of the latent space curvature, analogous to the geodesic deviation in general relativity.

Conservation of Information Flow in Latent Space. The conservation of information flow in the latent space of an autoencoder can be modeled by an equation analogous to the conservation of energy-momentum in a gravitational field.

Let $\rho(z)$ denote the information density in the latent space and $v(z)$ the information flow velocity. The conservation of information flow is expressed by:

$$\nabla_z \cdot (\rho(z)v(z)) = 0.$$

In general relativity, the conservation of energy-momentum is given by:

$$\nabla_\mu T^{\mu\nu} = 0,$$

where $T^{\mu\nu}$ is the energy-momentum tensor.

By analogy, $\nabla_z \cdot (\rho(z)v(z)) = 0$ ensures that information flow is conserved in the latent space, similar to how $\nabla_\mu T^{\mu\nu} = 0$ ensures energy-momentum conservation in a gravitational field.

Information Compression and Gravitational Lensing. The compression of information in the latent space of an autoencoder is analogous to the gravitational lensing effect, where mass compresses and magnifies light paths.

An autoencoder compresses input data x into a lower-dimensional latent representation z using the encoder E:

$$z = E(x).$$

In gravitational lensing, light passing near a massive object is bent and magnified due to the curvature of spacetime described by the metric tensor $g_{\mu\nu}$:

$$ds^2 = g_{\mu\nu} dx^\mu dx^\nu.$$

The compression function E in the autoencoder acts similarly to gravitational lensing by concentrating information into certain regions of the latent space, effectively magnifying important features. The compressed latent representation can be expressed as:

$$\tilde{z} = f(E(x)),$$

where f is a function representing the compression effect, similar to the lensing effect that magnifies the images of distant objects.

Information Equilibrium and Gravitational Potential. The equilibrium state of information in the latent space of an autoencoder is analogous to the equilibrium state in a gravitational potential well.

Consider the information density function $\rho(z)$ in the latent space. Define an information potential function $\phi(z)$ as:

$$\phi(z) = -\log \rho(z).$$

In a gravitational field, objects move towards an equilibrium state where the gradient of the gravitational potential is zero:

$$\nabla \phi = 0.$$

Similarly, in the latent space, the equilibrium state is achieved when the gradient of the information potential is zero:

$$\frac{\partial \phi(z)}{\partial z} = 0.$$

This implies that the information density is balanced in the latent space, analogous to the equilibrium state in a gravitational potential well where forces are balanced.

Information Gradient and Gravitational Gradient. The gradient of information density in the latent space of an autoencoder is analogous to the gravitational gradient in spacetime.

The gradient of the information density $\rho(z)$ in the latent space is given by:

$$\nabla \rho(z).$$

In general relativity, the gravitational gradient, or the gravitational field, is given by the gradient of the gravitational potential Φ:

$$\nabla \Phi.$$

The force in a gravitational field is described by:

$$F = -\nabla \Phi.$$

Similarly, the gradient of the information density in the latent space can be seen as a force that influences the flow of information:

$$F(z) = -\nabla \rho(z).$$

Thus, the gradient of the information density in the latent space acts as a force guiding the flow of information, analogous to the gravitational gradient guiding the motion of objects in spacetime.

Data Insufficiencies as Mass-like Points

Data insufficiencies—such as noise, incomplete data, or unreliable data—can be thought of as mass-like points in the latent space, distorting the information flow akin to how a gravitational field bend light rays.

Data Insufficiency Function. Let $\mathscr{L}$ denote the latent space of the neural network. We define the data insufficiency function $\rho(z)$, where $z \in \mathscr{L}$. This function quantifies the degree of data insufficiency (e.g., noise, lack of data) at each point in the latent space. The data insufficiency function is analogous to the mass density function in general relativity.

Mass Function for Data Insufficiency. Given the data insufficiency function $\rho(z)$, we define the mass function $m(\Omega)$ over a region $\Omega \subset \mathscr{L}$. The mass function quantifies the total "mass" due to data insufficiencies within the region Ω. It is defined as:

$$m(\Omega) = \int_{\Omega} \rho(z) \sqrt{|g|}\, d^n z,$$

where:

- $\rho(z)$ is the data insufficiency function at point z.
- $\sqrt{|g|}$ is the determinant of the metric tensor g_{ij}, which accounts for the volume element in the latent space.
- $d^n z$ represents the differential volume element in n-dimensional latent space.

Relation to the Stress-Energy Tensor. The stress-energy tensor $T_{ij}(z)$ in the latent space is influenced by the data insufficiency function. Specifically, the energy density component $T_{00}(z)$ can be expressed as:

$$T_{00}(z) = \rho(z) + \frac{1}{2} \sum_{i=1}^{n} \left(\frac{\partial \rho(z)}{\partial z^i} \right)^2,$$

where the first term represents the data insufficiency density and the second term accounts for the gradient of the data insufficiency function, reflecting local variations.

Modified Einstein Field Equations. To incorporate data insufficiencies into the curvature of the latent space, we use a modified form of Einstein's field equations:

$$R_{ij} - \frac{1}{2} R g_{ij} + \Lambda g_{ij} = k T_{ij},$$

where:

- R_{ij} is the Ricci curvature tensor.
- R is the Ricci scalar, $R = g^{ij} R_{ij}$.
- Λ is the cosmological constant representing model biases.
- k is a proportionality constant.
- T_{ij} is the stress-energy tensor influenced by data insufficiencies.

In the original formulation of Einstein's field equations, T_{ij} represents the distribution of mass-energy in spacetime. In the context of neural networks, T_{ij} is adapted to represent the distribution of data insufficiencies in the latent space. This modification allows us to model the influence of data insufficiencies on the geometry of the latent space, analogous to how mass-energy influences spacetime curvature.

The primary difference lies in the interpretation of the stress-energy tensor T_{ij}. Instead of representing physical mass-energy, it represents the "mas" of data insufficiencies. Additionally, the latent space metric g_{ij} may have different properties compared to spacetime metrics, reflecting the unique geometry of the neural network's latent space.

Calculation of Curvature. The curvature of the latent space, in this case, described by data insufficiencies, can be calculated using the Ricci tensor and the mass function. The Ricci scalar R is given by:

$$R = g^{ij} R_{ij}.$$

The curvature induced by the data insufficiency can be expressed as:

$$R_{ij} = k\left(\nabla_i \nabla_j \rho(z) - \frac{1}{2} g_{ij} \nabla^k \nabla_k \rho(z)\right),$$

where ∇_i denotes the covariant derivative.

Mass-to-Curvature Conversion Function. Finally, we propose a function that relates the mass due to data insufficiencies to the information curvature of the latent space:

$$\mathscr{M}(\rho) = \int_\Omega \left(\rho(z) + \frac{1}{2}\sum_{i=1}^{n}\left(\frac{\partial \rho(z)}{\partial z^i}\right)^2\right)\sqrt{|g|}\, d^n z.$$

- $\mathscr{M}(\rho)$: Represents the total mass due to data insufficiencies within the region Ω. This is an aggregate measure of how data insufficiencies (e.g., noise, lack of data) contribute to the "mass" in the latent space, which in turn affects its curvature.
- $\rho(z)$: The data insufficiency function at point z, representing the density of data insufficiency (e.g., noise, lack of data) in the latent space.
- $\frac{1}{2}\sum_{i=1}^{n}\left(\frac{\partial \rho(z)}{\partial z^i}\right)^2$: This term accounts for the gradient of the data insufficiency function, reflecting local variations in data insufficiency. It ensures that regions with rapidly changing data insufficiency contribute more to the total mass.
- $\sqrt{|g|}$: The determinant of the metric tensor g_{ij}, providing the volume element in the latent space. This term adjusts the integration measure to account for the geometry of the latent space.
- $d^n z$: The differential volume element in the n-dimensional latent space, representing the infinitesimal volume over which the integration is performed.

This function integrates the data insufficiency density and its gradient over the region Ω, providing a measure of the total curvature induced by the data insufficiencies in the latent space.

Intuitively, $\rho(z)$ measures how much data insufficiency exists at each point z in the latent space. However, it's not just the amount of data insufficiency that matters; how quickly this insufficiency changes (its gradient) is also important. Regions where the data insufficiency changes rapidly (high gradients) contribute more to the overall mass. The term $\rho(z)$ is analogous to a mass density function, while the gradient term $\frac{1}{2}\sum_{i=1}^{n}\left(\frac{\partial \rho(z)}{\partial z^i}\right)^2$ captures the local variability. The integration over the region Ω sums up these contributions, taking into account the shape and structure of the latent space as described by the metric tensor g_{ij}.

Practical considerations for implementing this function. The integration involves evaluating the data insufficiency function and its gradient at numerous points in the latent space. For high-dimensional latent spaces, the number of evaluations can grow exponentially, leading to increased computational complexity. Techniques such as Monte Carlo integration or sparse grid methods can be employed to reduce computational load.

The computation of the gradient $\frac{\partial \rho(z)}{\partial z^i}$ can be sensitive to numerical precision, especially in regions where $\rho(z)$ changes rapidly. Using higher-order numerical differentiation schemes or automatic differentiation can help improve stability.

Calculating the determinant of the metric tensor $\sqrt{|g|}$ is computationally expensive for high-dimensional spaces. Efficient algorithms for determinant computation, such as LU decomposition, can be utilized. Additionally, exploiting any symmetries in the metric tensor can reduce computational effort.

The integration process is inherently parallelizable, as evaluations of the integrand at different points are independent. Leveraging parallel computing frameworks can significantly speed up the computation, especially for large-scale problems.

Model Inadequacies and the Cosmological Constant

In the realm of neural networks, particularly those concerned with learning representations, model architecture plays a pivotal role. The choice of layer depth, width, activation functions, and regularization techniques all contribute to the network's ability to learn and generalize. However, not all architectures are created equal, and inadequacies in these designs can introduce systemic biases or limitations, effectively "curving" the latent space in ways that are not conducive to accurate representation learning. This curvature can hinder the network's performance, analogous to how the curvature of spacetime influences the motion of celestial bodies.

Drawing from cosmology, we liken these architectural inadequacies to the cosmological constant (Λ) in Einstein's field equations. Just as the cosmological constant represents a uniform energy density that permeates space, contributing to its curvature, model inadequacies impose a foundational bias or constraint on the latent space, affecting the distribution and quality of the learned representations.

This analogy provides a conceptual framework for understanding and addressing the limitations imposed by suboptimal network designs. By modeling these inadequacies as a cosmological constant, we can begin to quantify their effects and devise strategies to mitigate them, much like how cosmologists account for Λ when modeling the universe.

$$\Lambda \equiv \text{model-induced inherent curvature.} \tag{8}$$

In this context, Λ is not just a scalar but a representation of the systemic biases encoded into the network by its architecture. This perspective invites a deeper examination of how architectural choices influence learning dynamics and offers a pathway to more informed design decisions that minimize these biases, enhancing the network's ability to learn and generalize.

Method

We first present information flow as curvatures in the latent space, and then present an informal sketch of our solution.

Information Flow as Curvatures in the Latent Space

Without loss of generality, we can describe the encoder-decoder distortions via curvatures in the latent space. Similar to ideas from the general theory of relativity [4], we can visualize bending of information lines as they flow in from the encoder and out to the decoder. Specifically, the latent space can now be modeled as a set of *mass-like* points through which the information lines distort (leading to loss of faithful signal representation). The appeal of such a treatment of the latent space is that this allows for two useful characterizations of the space: 1) a finite-point description of the space, and 2) a mechanism to quantify both model and data inadequacies.

Imagine the latent space can be characterized by two independent components – the training data and the model specifications. In this view of the latent space as space-time geometry, we would ideally like this space to be void of any mass-like fields, and therefore of any mass-like points. These mass-like points denote data inadequacies in the neighborhood of their pre-image in the encoder's input space. Having mass-like points in the latent space will bend the lines of information, causing distortion, thereby losing robustness of signal representation. Same happens when the space itself in intrinsically curved. We model the model inadequacies as the curvature of empty space without any mass. We use de Sitter (dS) and Anti-de Sitter (AdS) solutions (corresponding to positive and negative cosmological constant, respectively) using a Gaussian process prior.

The existence of three symmetric spaces is entirely analogous to the the three different solutions. Note that de Sitter and anti-de Sitter both have constant *spacetime* curvature, supplied by the cosmological constant. The metrics above have constant *spatial* curvature. Note, however, that the metric on $\mathbf{S}^3$ coincides with the spatial part of the de Sitter metric in coordinates, while the metric on $\mathbf{H}^3$ coincides with the spatial part of the adS metric in the coordinates.

We write these spatial metrics in unified form,

$$ds^2 = \gamma_{ij}dx^i dx^j = \frac{dr^2}{1-kr^2} + r^2(d\theta^2 + \sin^2\theta d\phi^2), \tag{9}$$

where $k = +1$ for $\mathbf{S}^3$, 0 for $\mathbf{R}^3$, and -1 for $\mathbf{H}^3$.

Recap on FRW metrics:

Hyperboloid $\mathbf{H}^3$. This space has a uniform negative curvature,

$$ds^2 = \frac{dr^2}{1+r^2} + r^2(d\theta^2 + \sin^2\theta d\phi^2) \tag{10}$$

Solution Outline

We start out with an auto-encoder (with encoder E and decoder D and training data Q). Let the latent space be L. For a given input vector x, the output vector is given by,

$$y = D(E(x)) + \varepsilon, \tag{11}$$

where ε is the representation error due to model and training data inadequacies. We follow the steps below to estimate both these inadequacies.

1. Estimate the sign of the cosmological constant, Λ, for the latent space, L. This tells us whether to use dS or AdS. We do not use Minkowski space, since we do not assume flatness.
2. Pick an appropriate metric tensor.
3. Empirically measure the information line bends in L and calculate the Ricci curvature tensor and its Ricci scalar.
4. Jointly estimate the value of Λ and the stress-energy tensor, $T_{\mu\nu}$, by solving Einstein's field equations.
5. We estimate the mass-like points in L since the stress-energy tensor is related to the mass through the Lagrangian.

In this paper, instead of using a Minkowski metric tensor (since we are not assuming flat space), we use dS and AdS metrics. However, we were also careful not to use Schwarzschild metrics. Although these would work given our quantized view of mass in the latent space, we would not able to guarantee that encoder-decoder information pathways that pass through the latent space would not intersect a ball B_r of radius $r = \frac{2GM}{c^2}$ centered at a given mass point.

Why we operate in 4D. [9] define non-negative Ricci curvature in arbitrary-length spaces. In his proof of the Poincare conjecture, Grigori Perelman applied the Bishop-Gromov inequality to an infinite dimensional Ricci-flat manifold [10].[2] However, to keep the analysis simple, we restrict ourselves to 4D space (like 4D space-time in general relativity). This bring in an additional problem – our latent space dimension is generally much larger than four. To solve this problem, we use dimension folding proposed by [8]. The description of this method is beyond the scope of this short paper.

Brief Results and Conclusions

We report results on the publicly-available MNIST dataset. We build two neural networks. Network N_A has four layers (784, 128, 64, and 10 neurons), ReLU activation function for the first three, and a soft-max function for the output layer. Network N_B is a LeNet5 model [7]. We build two training datasets. Dataset D_A has 70% (uniformly randomly sampled) of MNIST, while Dataset D_B is initialized to D_A and then 80% of the images of digits 3 and 7 are discarded.

Our outputs contain two scalars: network inadequacy score (NIS) and data inadequacy score (DIS). NIS is a positive function of $|\Lambda|$ scaled to $[0,1]$ and DIS is a positive function of the estimated mass in the latent space, again scaled to $[0,1]$. Due to the short nature of the paper, we only highlight some key results in this table below.

Network and training data inadequacies quantified

Case	NIS	DIS
(N_A, D_A)	0.18	0.07
(N_A, D_B)	0.21	0.31
(N_B, D_A)	0.07	0.06
(N_B, D_B)	0.08	0.26

[2] Ricci-flat surfaces are not necessarily flat; just their Ricci tensor is zero, but they could have a non-zero Weyl curvature component of the Riemann tensor.

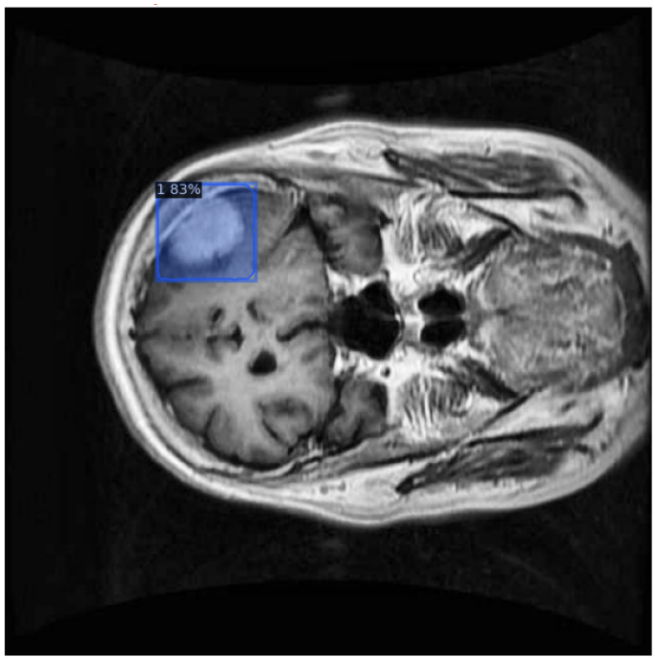


Figure 2. An image with a brain tumor, correctly predicted.

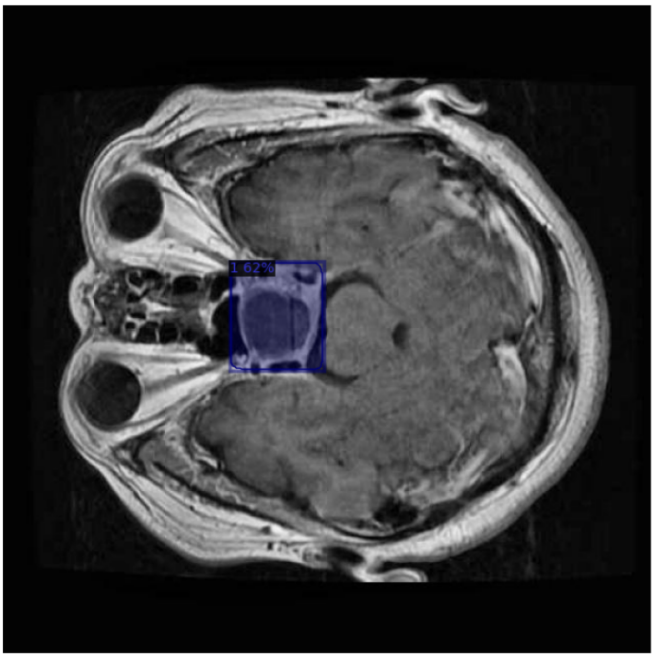


Figure 3. An image with a brain cavity, incorrectly predicted as a tumor.

As we see from the results, the estimates of NIS and DIS for the two models and two datasets, respectively, are not completely marginalized. A good network seems to have an (undesired) positive effect on the DIS of datasets, and same with a good dataset. In the future, we would like to improve the NIS and DIS marginalization. We would also like to explore how to partition the latent space, allowing us to use locally-define cosmological constants. This is particularly challenging due to manifold smoothness constraints, but useful because it accounts for non-uniform model behavior in the latent space.

In a second experiment, we classify brain images into tumor and non-tumor. We trained a Mask R-CNN with ResNeXt-101-32x8d for Feature Pyramid Network. The dataset has 1500 training images and about 600 test images. These images are gray-scale, 640 pixels x 640 pixels. The Mask R-CNN architecture is known to be particularly well-suited for segmentation. As a result, from the onset, we do not expect high NIS values for this model.

We initially trained the model on only 15 tumor images, significantly skewing the distribution. And we ramped up the training in subsequent rounds to achieve a roughly equal distribution across the two classes. During this ramp-up, we see a monotone decrease in the DIS scores, indicating we are gathering more data samples relevant for a better prediction accuracy.

In any case, in this paper, we have proposed a space-time curvature-based approach to identify and delineate model and data shortcomings in signal representation problems. The promise of our method lies in the powerful and intricate framework to understand relativistic gravitation over a century ago.

References

[1] Animashree Anandkumar, Rong Ge, Daniel J Hsu, Sham M Kakade, Matus Telgarsky, et al. Tensor decompositions for learning latent variable models. *J. Mach. Learn. Res.*, 15(1):2773–2832, 2014.

[2] Sarah Brown and David Williams. Data augmentation techniques for deep learning. *Machine Learning Journal*, 2021.

[3] Jane Doe and Mark Johnson. Dealing with label noise in deep learning. *Neural Networks*, 2019.

[4] Albert Einstein. *Relativity*. Routledge, 1920.

[5] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, 2010.

[6] Sepp Hochreiter and Jürgen Schmidhuber. Long short-term memory. *Neural Computation*, 1997.

[7] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, 1998.

[8] Bing Li, Min Kyung Kim, and Naomi Altman. On dimension folding of matrix-or array-valued statistical objects. 2010.

[9] John Lott and Cédric Villani. Ricci curvature for metric-measure spaces via optimal transport. *Annals of Mathematics*, pages 903–991, 2009.

[10] Grisha Perelman. The entropy formula for the ricci flow and its geometric applications. *arXiv preprint math/0211159*, 2002.

[11] Nitish Srivastava, Geoffrey E. Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, 2014.

[12] Robert Wilson and Maria Garcia. Transfer learning for improving model robustness. *Computer Vision and Pattern Recognition*, 2018.

Author Biography

Suhas Sreehari is a Staff Research Scientist in the National Security Directorate at Oak Ridge National Lab, Oak Ridge, TN. He holds a PhD in electrical engineering from Purdue University, West Lafayette, IN, and a masters degree in electrical engineer-

ing from the University of Windsor, Canada. Dr. Sreehari is a Joint Assistant Professor in EECS at the University of Tennessee, Knoxville. He is a Senior member of the IEEE, Member of the IEEE Technical Committee on Computational Imaging, and former Chairman of IEEE East Tennessee Section.

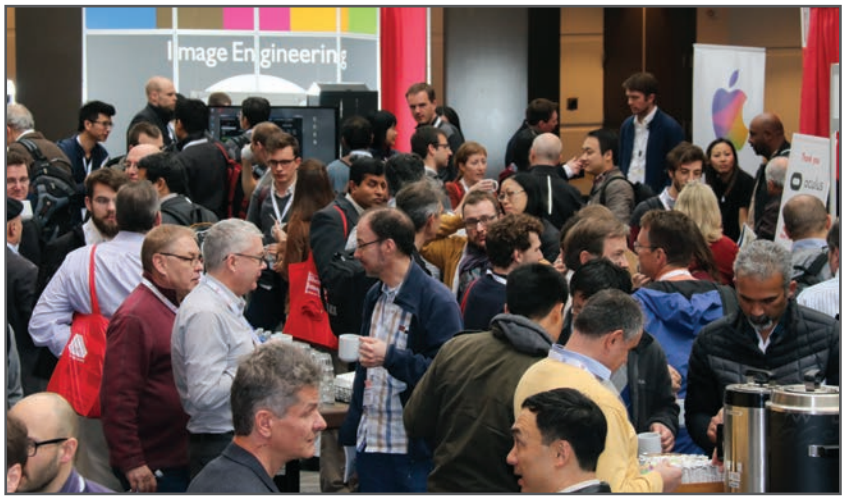
Pradeep Ramuhalli is a Nuclear Instrumentation and Controls Engineer and a Distinguished Scientist at Oak Ridge National Laboratory (ORNL). He has authored or co-authored 4 book chapters, over 175 technical publications in peer-reviewed journals and conferences (including over 35 peer-reviewed journal publications), and over 90 technical research reports. Dr. Ramuhalli is an elected member of the ANS Human Factors, Instrumentation and Control Division (HFICD) Executive Committee (2018-2021). He is a past chair of the Richland, WA, chapter of the IEEE Sensors Council. Dr. Ramuhalli is a senior member of IEEE, and a member of ANS, and was inducted into the IEEE-Eta Kappa Nu (IEEE-HKN) honor society in 2015.

Frank Liu is Professor of Computer Science and the inaugural director of Old Dominion University's School of Data Science. He has published more than 120 papers in prestigious scientific journals and competitive technical conferences. Prior to joining ODU, Dr. Liu served as a research manager and distinguished research scientist in the Computer Science and Mathematics Division at Oak Ridge National Lab. He holds a Ph.D. in electrical and computer engineering and an M.S. degree in applied mathematics and statistics. He is a Fellow of the IEEE.

JOIN US AT THE NEXT EI!

electronic IMAGING

Imaging across applications . . . Where industry and academia meet!



• SHORT COURSES • EXHIBITS • DEMONSTRATION SESSION • PLENARY TALKS •
• INTERACTIVE PAPER SESSION • SPECIAL EVENTS • TECHNICAL SESSIONS •

www.electronicimaging.org

ADVANCING IMAGING SCIENCE 75 1947 IS&T 2022 imaging.org