Abstract

In a Riemannian manifold, the Ricci flow is a partial differential equation for evolving the metric to become more regular. We hope that topological structures from such metrics may be used to assist in the tasks of machine learning. However, this part of the work is still missing. In this paper, we propose Ricci flow assisted Eucl2Hyp2Eucl neural networks that bridge this gap between the Ricci flow and deep neural networks by mapping neural manifolds from the Euclidean space to the dynamically stable Poincaré ball and then back to the Euclidean space. As a result, we prove that, if initial metrics have an $L^2$-norm perturbation which deviates from the Hyperbolic metric on the Poincaré ball, the scaled Ricci-DeTurck flow of such metrics smoothly and exponentially converges to the Hyperbolic metric. Specifically, the role of the Ricci flow is to serve as naturally evolving to the stable Poincaré ball. For such dynamically stable neural manifolds under the Ricci flow, the convergence of neural networks embedded with such manifolds is not susceptible to perturbations. And we show that Ricci flow assisted Eucl2Hyp2Eucl neural networks outperform with their all Euclidean counterparts on image classification tasks.

1. Introduction

In the field of machine learning, Euclidean embeddings for representation learning are the universal and successful method, which benefits from simply convenience and closed-form formulas in the Euclidean space $\mathbb{R}^n$ endowed with the Euclidean metric $g^E$. Moreover, in the broad application of neural networks, Euclidean embeddings are also showing off, including image classification (Krizhevsky et al., 2012; Simonyan & Zisserman, 2014), semantic segmentation (Long et al., 2015; Chen et al., 2014), object detection (Girshick, 2015), etc.

Recently, some studies have shown that the latent non-Euclidean structure in many data will affect the representation of Euclidean embeddings (Bronstein et al., 2017). In such cases, the advantages of Hyperbolic embeddings are highlighted. Firstly, the Hyperbolic space provides more powerful or meaningful geometrical representations than the Euclidean space. Secondly, for the weight initialization of a neural network, the distribution is generally Gaussian (Glorot & Bengio, 2010; He et al., 2015), which is a manifold of constant negative curvature (Amari, 2016), i.e., the Hyperbolic space (Amari et al., 1987). Obviously, the adoption of Hyperbolic embeddings in neural networks and deep learning has become very attractive.

For hierarchical, taxonomic or entailment data, Hyperbolic embeddings outperformed Euclidean embeddings in machine learning (Sala et al., 2018; Ganea et al., 2018). For tree structure, the Euclidean space with infinite dimensions cannot be embedded with arbitrary distortion, but the Hyperbolic space with only 2 dimensions can preserve their metric (Sala et al., 2018). As for basic operations (e.g. matrix addition, matrix-vector multiplication, etc.) in the Hyperbolic space, Ganea et al. gave appropriate deep learning tools (Ganea et al., 2019).
Despite the successful application of Hyperbolic embeddings in deep neural networks, there are two other research areas that have not yet been involved, i.e., how to avoid the perturbation of the parameter update on the stability of the Hyperbolic space and how to combine the common advantages of Euclidean space and Hyperbolic space. In this paper, we propose Ricci flow assisted Eucl2Hyp2Eucl neural networks by mapping neural manifolds from the Euclidean space to the dynamically stable Poincaré ball and then back to the Euclidean space. Specifically, the Ricci flow can naturally make the Riemannian manifold converge by evolving metrics and provide nonlinearity for alternating Euclidean and Hyperbolic spaces (the lack of Ricci flow causes Eucl2Hyp2Eucl neural networks to be indistinguishable from their all Euclidean counterparts).

For the Ricci flow, Ye has considered stability of negatively curved manifolds on compact spaces (Ye, 1993). Suneeta has investigated linearised stability of the Hyperbolic space under the Ricci flow (Suneeta, 2009). Li et al. have proven stability of the Hyperbolic space in dimensions \( n \geq 6 \) when the deviation of the curvature of the initial metric from Hyperbolic space exponentially decays (Li & Yin, 2010). Based on the above work, Schnüter et al. yielded stability of the Hyperbolic space in dimensions \( n \geq 4 \) when the initial metric is close to the Hyperbolic metric (Schnüter et al., 2010). A series of works have shown that, in certain cases, the Ricci flow can be used to eliminate the influence of a perturbation in the Hyperbolic space.

There are two main contributions in this paper. On the one hand, we prove that, when initial metrics \( \bar{g} \) is close to the Hyperbolic metric \( g^H \) on the Poincaré ball (Definition 3.2), the scaled Ricci-DeTurck flow is exponential convergence for an \( L^2 \)-norm perturbation (Lemma 3.7). Apparently, such results are very meaningful, which shows that the application of the Ricci flow can help Poincaré embedded neural network manifolds to eliminate an \( L^2 \)-norm perturbation for the metric. The Ricci flow guarantees the dynamical stability of neural manifolds during the training of neural networks with respect to the input data with transforms.

On the other hand, furthermore, we propose Ricci flow assisted Eucl2Hyp2Eucl neural networks that have been in an alternate state between the Euclidean space \( \mathbb{R}^n \) and the dynamically stable Poincaré ball \( \mathbb{D}^n_r \). The illustration is shown in Figure 1. Specifically, we map the neural network (for the output before the softmax) from the Euclidean space \( (\mathbb{R}^n, g^E) \) to the Poincaré ball \( (\mathbb{D}^n_r, \bar{g}) \) with an \( L^2 \)-norm perturbation. Then, the Ricci flow is used to exponentially converge to the Poincaré ball \( (\mathbb{D}^n_r, g^H) \) without an \( L^2 \)-norm perturbation. Finally, we map the neural network from the Poincaré ball \( (\mathbb{D}^n_r, g^H) \) to the Euclidean space \( (\mathbb{R}^n, g^E) \).

In general, Eucl2Hyp2Eucl neural networks take into account the common advantages of Euclidean and Poincaré embeddings, i.e., simply convenience and meaningful geometrical representations. The algorithm is shown in Algorithm 1.

The rest of this paper is organized as follows. Section 2 summarizes basic works on Ricci flow. The proofs of the convergence of the Poincaré ball under Ricci flow are presented in Section 3. In Section 4, we yield the illustration of Ricci flow assisted Eucl2Hyp2Eucl neural networks. For the performance on classification tasks, we compare Ricci flow assisted neural networks with their all Euclidean counterparts in Section 5. The conclusion is given in Section 6.

2. Ricci Flow

For a Riemannian manifold \( \mathcal{M} \) with metric \( g_0 \), the Ricci flow was introduced by Hamilton to prove Thurston’s Geometrization Conjecture and consequently the Poincaré Conjecture (Hamilton et al., 1982). The Ricci flow is a partial differential equation that evolves the metric:

\[
\frac{\partial}{\partial t} g(t) = -2 \text{Ric}(g(t))
\]

where \( g(t) \) is a time-dependent Riemannian metric and \( \text{Ric}() \) denotes the Ricci curvature tensor whose definition can be found in Appendix A.

The idea is to try to evolve the Riemannian metric in some way that make the manifold become more regular, which can also be understood as rounding from the topological structure point of view. This continuous process is known as manifold “surgery”.

2.1. Short Time Existence

If the Ricci flow is strongly parabolic, there exists a unique solution for a short time.

**Theorem 2.1.** When \( u : \mathcal{M} \times [0, T) \to \mathcal{E} \) is a time-dependent section of the vector bundle \( \mathcal{E} \) where \( \mathcal{M} \) is some Riemannian manifold, if the system of the Ricci flow is strongly parabolic at \( u_0 \) then there exists a solution on some time interval \([0, T)\), and the solution is unique for as long as it exists.

**Proof.** The proofs can be found in (Ladyzhenskaia et al., 1988). \( \square \)

**Definition 2.2.** The Ricci flow is strongly parabolic if there exists \( \delta > 0 \) such that for all covectors \( \varphi \neq 0 \) and all symmetric \( h_{ij} = \frac{\partial g_{ij}(t)}{\partial t} \neq 0 \), the principal symbol of \(-2 \text{Ric}\) satisfies

\[
-2 \text{Ric}((\varphi)(h))_{ij} h^{ij} \\
= g^{pq} (\varphi_p \varphi_q h_{ij} + \varphi_i \varphi_j h_{pq} - \varphi_q \varphi_i h_{jp} - \varphi_q \varphi_j h_{ip}) h^{ij} \\
> \delta \varphi_k \varphi^k h_{rs} h^{rs}.
\]
Since the above inequality cannot always be satisfied, the Ricci flow is not strongly parabolic. Empirically, one cannot use Theorem 2.1 to prove the existence of the solution directly.

To understand which parts have an impact on its non-parabolic, one linearizes the Ricci curvature tensor.

**Lemma 2.3.** The linearization of $-2$ $\text{Ric}$ can be rewritten as

$$D[-2\text{Ric}](h)_{ij} = g^{pq} \nabla_p \nabla_q h_{ij} + \nabla_i V_j + \nabla_j V_i + O(h_{ij})$$

where $V_i = g^{pq}\left(\frac{1}{2} \nabla_i h_{pq} - \nabla_q h_{pi}\right).$

**Proof.** The proofs can be found in Appendix B.1.

In particular, the term $O(h_{ij})$ will have no contribution to the principal symbol of $-2$ $\text{Ric}$. For convenience of our discussion, we just ignore this term. By carefully observing the above equation, one finds that the impact on the non-parabolic of the Ricci flow comes from the terms in $V$, not the term $g^{pq} \nabla_p \nabla_q h_{ij}$. The solution is followed by the DeTurck Trick (DeTurck, 1983) that has a time-dependent reparameterization of the manifold:

$$\frac{\partial}{\partial t} \bar{g}(t) = -2\text{Ric}(\bar{g}(t)) - \mathcal{L}_{\partial \phi(t)} \bar{g}(t)$$

$$\bar{g}(0) = \bar{g}_0,$$

**Proof.** The proofs can be found in Appendix B.2 for details. By choosing $\frac{\partial \phi(t)}{\partial t}$ to cancel the effort of the terms in $V$, the reparameterized Ricci flow is strongly parabolic. Thus, one gets that the Ricci-DeTurck flow has a unique solution for a short time.

### 2.2. Curvature Explosion at Singularity

Subsequently, we will present the behavior of the Ricci flow in finite time and show that the evolution of the curvature tends to develop singularities. Before giving the core demonstration, Theorem 2.7, some foreshadowing proofs need to be prepared.

**Theorem 2.4.** Given a smooth Riemannian metric $g_0$ on a closed manifold $\mathcal{M}$, there exists a maximal time interval $[0, T)$ such that a solution $g(t)$ of the Ricci flow, with $g(0) = g_0$, exists and is smooth on $[0, T)$, and this solution is unique.

**Proof.** The proofs can be found in (Sheridan & Rubinstein, 2006).

**Theorem 2.5.** Let $\mathcal{M}$ be a closed manifold and $g(t)$ a smooth time-dependent metric on $\mathcal{M}$, defined for $t \in [0, T)$. If there exists a constant $C < \infty$ for all $x \in \mathcal{M}$ such that

$$\int_0^T \left| \frac{\partial}{\partial t} g_x(t) \right|_{g(t)} dt \leq C,$$

then the metrics $g(t)$ converge uniformly as $t$ approaches $T$ to a continuous metric $g(T)$ that is uniformly equivalent to $g(0)$ and satisfies

$$e^{-C} g_x(0) \leq g_x(T) \leq e^C g_x(0).$$

**Proof.** The proofs can be found in Appendix B.3.

**Corollary 2.6.** Let $(\mathcal{M}, g(t))$ be a solution of the Ricci flow on a closed manifold. If $|\text{Rm}|_{g(t)}$ is bounded on a finite time $[0, T)$, then $g(t)$ converges uniformly as $t$ approaches $T$ to a continuous metric $g(T)$ which is uniformly equivalent to $g(0)$.

**Proof.** The bound on $|\text{Rm}|_{g(t)}$ implies one on $|\text{Ric}|_{g(t)}$. Based on Eq. 1, we can extend the bound on $|\frac{\partial}{\partial t} g(t)|_{g(t)}$. Therefore, we obtain an integral of a bounded quantity over a finite interval is also bounded, by Theorem 2.5.

**Theorem 2.7.** If $g_0$ is a smooth metric on a compact manifold $\mathcal{M}$, the Ricci flow with $g(0) = g_0$ has a unique solution $g(t)$ on a maximal time interval $t \in [0, T)$. If $T < \infty$, then

$$\lim_{t \to T} \left( \sup_{x \in \mathcal{M}} |\text{Rm}_x(t)| \right) = \infty.$$
3. The Poincaré Ball under Ricci Flow

3.1. Basics of Hyperbolic Space and The Poincaré Ball

The Hyperbolic space has several isometric models (Anderson, 2006). In this paper, similarly as in (Nickel & Kiela, 2017) and (Ganea et al., 2018), we choose an $n$-dimensional Poincaré ball $\mathbb{D}^n_r$ with radius $1/\sqrt{r}$.

Empirically, the Poincaré ball can be defined by the background manifold $\mathbb{D}^n_r := \{x \in \mathbb{R}^n \mid r \|x\|^2 < 1\}$ endowed with the Hyperbolic metric:

$$g^H_x = \lambda_x^2 g^E, \quad \text{where } \lambda_x := \frac{2}{1 - r \|x\|^2}. \quad (6)$$

Note that the Euclidean metric $g^E$ is equal to the identity matrix $I$. For $r > 0$, $\mathbb{D}^n_r$ denotes the open ball of radius $1/\sqrt{r}$. In particular, if $r = 0$, then one recovers the Euclidean space $\mathbb{D}^n_0 = \mathbb{R}^n$.

By Corollary 3.1, the Riemannian gradient endowed with $g^H$ for any point $x \in \mathbb{D}^n$ is known to be given by \footnote{Note that the Riemannian gradient is similar to the natural gradient (Martens & Grosse, 2015; Martens, 2020) in Riemannian manifold defined by the KL divergence.}

$$\partial^H = \frac{1}{\lambda_x} \partial^E, \quad \text{where } \lambda_x := \frac{2}{1 - r \|x\|^2}. \quad (7)$$

**Corollary 3.1.** In view of information geometry (Amari, 2016), the steepest descent direction in a Riemannian manifold endowed with $g$ satisfies

$$\partial^g = g^{-1} \partial^E, \quad (8)$$

with respect to the steepest descent direction $\partial^E$ in Euclidean space.

**Proof.** The proofs can be found in Appendix D. \quad ☐

3.2. The Hyperbolic Metric

As the Hyperbolic space evolves under Ricci flow, it is convenient to consider the rescaled Ricci-DeTurck flow (Schnürer et al., 2010) by Eq. 3

$$\frac{\partial}{\partial t} \bar{g}(t) = -2 \text{Ric} (\bar{g}(t)) + \nabla_i W_j + \nabla_j W_i - 2(n - 1) \bar{g}(t), \quad \bar{g}(0) = \bar{g}_0, \quad \text{where } W_i = g^{pq} g_{ij} \left( \bar{g} r_{pq}^j - g^H r_{pq}^j \right). \quad (9)$$

The Hyperbolic metric $g^H$ on $\mathbb{D}^n_r$ of sectional curvature $-r$ is a stationary point to Eq. 9.

Subsequently, we will discuss that the Poincaré ball $(\mathbb{D}^n_r, \bar{g}(t))$ uniformly converges to the Poincaré ball $(\mathbb{D}^n_r, g^H)$ under the rescaled Ricci-DeTurck flow when the given perturbation satisfies Definition 3.2. Obviously, the introduction of the Ricci flow can ensure dynamically stable Poincaré ball.

**Definition 3.2.** Let $\bar{g}$ be a metric on $\mathbb{D}^n_r$. There exists a $\epsilon > 0$ such that

$$(1 + \epsilon)^{-1} g^H \leq \bar{g} \leq (1 + \epsilon) g^H,$$

which can be said that $\bar{g}$ is $\epsilon$-close to $g^H$.

3.3. Finite Time Existence

We denote the norm of a tensor as $|\cdot|$, then

**Lemma 3.3.** Given a Riemannian metric $g(t)$ on $\mathbb{D}^n_r$, there exists a maximal time interval $(0,T)$ such that a solution to Eq. 9 exists and is smooth \footnote{Smooth is equivalent to $C^\infty$, i.e., any derivative is continuous.} on $(0,T)$. Specifically, $g(t)$ is $\epsilon$-close to $g^H$ of sectional curvature $-r$. If a small enough $\epsilon > 0$, then there exists

$$\frac{\partial}{\partial t} |g - g^H|^2 \leq \Delta |g - g^H|^2 - 2 |\nabla (g - g^H)|^2 + 4 |g - g^H|^2 \quad (10)$$

where $\Delta$ is the Laplacian defined by $\bar{g} \nabla_i \nabla_j$.

**Proof.** The proofs can be found in Appendix C. \quad ☐

**Corollary 3.4.** Given the Poincaré ball $\mathbb{D}^n_r$, where $\partial \mathbb{D}^n_r$ denotes the boundary, there exists a maximal interval $(0,T)$ such that a solution $\bar{g}(t)$ on $\mathbb{D}^n_r$ to Eq. 9 exists and is smooth on $(0,T)$. Specifically, $g|_{\partial \mathbb{D}^n_r} = g^H|_{\partial \mathbb{D}^n_r}$. There exists a constant $C > 0$ such that

$$\sup_{\mathbb{D}^n_r} |g - g^H| \leq C. \quad (11)$$

**Proof.** As long as Definition 3.2 is satisfied, Lemma 3.3 gives the proofs. \quad ☐

We denote the $L^\infty$-norm with respect to the Hyperbolic metric $g^H$ as $\|| \cdot ||_{L^\infty}$, then

**Theorem 3.5.** For a solution $\bar{g}(t)$ on $\mathbb{D}^n_r$ to Eq. 9 that exists and is smooth on a maximal time interval $(0,T)$, if $\bar{g}(0)$ is a metric on $\mathbb{D}^n_r$ satisfying $\|\bar{g}(0) - g^H\|_{L^\infty} \leq \epsilon$ where $\epsilon > 0$, then there exists a constant $C$ such that

$$\|\bar{g}(t) - g^H\|_{L^\infty} \leq C \|\bar{g}(0) - g^H\|_{L^\infty} \leq \epsilon \cdot C. \quad (12)$$

**Proof.** The proof follows the similar statement (Simon, 2002; Bamler, 2010). \quad ☐

Empirically, Theorem 3.5 yields that the solutions to the rescaled Ricci-DeTurck flow exists in finite time. Otherwise, Corollary 3.4 gives an upper bound on $|\bar{g} - g^H|$, which allows us to integrate it.
3.4. Exponential Convergence

Theorem 3.6. Based on Corollary 3.4, we further have
\[ \int_{\mathbb{D}^n_r} |\bar{g}(t) - g^H|^2 \, d\Omega \leq e^{-(A(n,r)-4)t} \int_{\mathbb{D}^n_r} |\bar{g}(0) - g^H|^2 \, d\Omega \]
where \( \Omega \) is the volume element with respect to \( \mathbb{D}^n_r \).

Proof. Using Lemma 3.3, we yield
\[
\frac{\partial}{\partial t} \int_{\mathbb{D}^n_r} |\bar{g} - g^H|^2 \, d\Omega \leq 4 \int_{\mathbb{D}^n_r} |\bar{g} - g^H|^2 \, d\Omega \\
+ \int_{\mathbb{D}^n_r} |\bar{g}^i j \nabla_i \nabla_j (g^H)|^2 \, d\Omega = 4 \int_{\mathbb{D}^n_r} |\bar{g} - g^H|^2 \, d\Omega - 2 \int_{\partial \mathbb{D}^n_r} |(\bar{g}^i j \nabla_i \nabla_j (g^H)|^2 \, d\Omega \\
+ \int_{\mathbb{D}^n_r} \nabla_i (\bar{g}^i j \nabla_j) |\bar{g} - g^H|^2 = \nabla_i (\bar{g}^i j \nabla_j) |\bar{g} - g^H|^2 = \nabla_i (\bar{g}^i j \nabla_j) |\bar{g} - g^H|^2.
\]

In the second step, we use \( \nabla_i (\bar{g}^i j \nabla_j) |\bar{g} - g^H|^2 = \nabla_i (\bar{g}^i j \nabla_j) |\bar{g} - g^H|^2 + \nabla_i \bar{g}^i j \nabla_j |\bar{g} - g^H|^2 \).

As \( |\bar{g} - g^H|_{\mathbb{D}^n_r} = 0 \), we compute, using Stokes theorem (Wald, 2010),
\[
\int_{\mathbb{D}^n_r} \nabla_i (\bar{g}^i j \nabla_j) |\bar{g} - g^H|^2 \, d\Omega = \int_{\partial \mathbb{D}^n_r} n_i \bar{g}^i j \nabla_j |\bar{g} - g^H|^2 \, dS = 0
\]
where \( dS \) is the area element with respect to \( \partial \mathbb{D}^n_r \) and \( n_i \) is the outer normal vector with respect to \( \partial \mathbb{D}^n_r \). We define
\[
A(n,r) = \inf \int_{\mathbb{D}^n_r} \frac{2 |\nabla (\bar{g} - g^H)|^2 + \nabla_i \bar{g}^i j \nabla_j |\bar{g} - g^H|^2 \, d\Omega}{\int_{\mathbb{D}^n_r} |\bar{g} - g^H|^2 \, d\Omega}
\]
then
\[
\frac{\partial}{\partial t} \int_{\mathbb{D}^n_r} |\bar{g} - g^H|^2 \, d\Omega \leq (4 - A(n,r)) \int_{\mathbb{D}^n_r} |\bar{g} - g^H|^2 \, d\Omega.
\]

In view of differential equation, the above inequality extends, using \( F(t) = \int_{\mathbb{D}^n_r} |\bar{g}(t) - g^H|^2 \, d\Omega \) and \( F(0) = \int_{\mathbb{D}^n_r} |\bar{g}(0) - g^H|^2 \, d\Omega \),
\[
\int \frac{\partial F(t)}{F(t)} \, dt \leq (4 - A(n,r)) \int \partial t \\
\rightarrow \log F(t) \leq (4 - A(n,r)) t + \log F(0) \\
\rightarrow F(t) \leq e^{-(A(n,r)-4)t} F(0).
\]

Based on Theorem 3.5, we have \( A(n,r) \geq 4 \) because \( |\bar{g} - g^H|^2 \) decays.

Lemma 3.7. Based on Theorem 3.6, we yield the estimate
\[
\|\bar{g}(t) - g^H\|_{L^2(\mathbb{D}^n_r)}^2 \leq e^{-(A(n,r)-4)t} \|\bar{g}(0) - g^H\|_{L^2(\mathbb{D}^n_r)}^2
\]
where \( \| \cdot \|_{L^2} \) is the \( L^2 \)-norm with respect to the Hyperbolic metric \( g^H \).

Proof. The proofs follow directly from Theorem 3.5 and Corollary 3.4.

Consequently, we see that the scaled Ricci-DeTurck flow is exponential convergence for an \( L^2 \)-norm perturbation.

4. Ricci Flow Assisted Neural Networks

4.1. Ricci Curvature in Neural Networks

On the one hand, a neural network is trained on the given dataset and its geometrical structure will gradually become regular. On the other hand, the Ricci flow is a process of “surgery” on a manifold, which will make the manifold also become regular. There seems to be a common goal of these two evolution methods. In this way, we can embed a Riemannian manifold into the neural network, and utilize the Ricci flow to assist in the training of neural networks on dynamically stable manifolds.

Based on the previous section, we have embedded the Poincaré ball into a neural network. And initial metrics \( \bar{g} \) for an \( L^2 \)-norm perturbation that deviates from the Hyperbolic metric \( g^H \) will converge to \( g^H \) under the Ricci flow. That is excellent because we embed a dynamically stable Poincaré ball for neural networks, which will not affect the convergence of neural networks.

Empirically, for each input, we can embed an \( n \)-dimensional Poincaré ball into the output of neural networks before the \textit{softmax}. Apparently, time-dependent metrics \( \bar{g}(t) \) corresponding to the output can be well-defined.

Now, let us see the Ricci curvature of neural networks. According to Eq. 1, the tensor \(-2 \text{Ric}(\bar{g}(t))\) approaches zero, as \( \frac{\partial}{\partial t} \bar{g}(t) \) approaches zero. We can yield, referring to Appendix A,
\[
-2 \text{Ric}(\bar{g}) = -2 R_{ij}^k = \bar{g}^{ip} \left( \partial_i \partial_j \bar{g}_{pk} - \partial_i \partial_k \bar{g}_{pj} + \partial_p \partial_k \bar{g}_{ij} + \partial_p \partial_j \bar{g}_{ik} \right).
\]

Inspired by (Kaul & Lell, 2019), we treat the term \( i \) and \( p \) as the translation and rotation by considering translation invariance instead of rotation invariance. As for rotations, the standard data augmentation does not include such transformations. For the fairness of ablation studies, we just exclude rotations, i.e., \( \partial_i \partial_j \bar{g}_{pk} + \partial_p \partial_k \bar{g}_{ij} \) is zero. Therefore, \( \partial_k \bar{g} \) and \( \partial_j \bar{g} \) can be the row and column transformation respectively.
We consider alternating neural manifolds between Euclidean embeddings and Poincaré embeddings in back-propagation, which greatly retains the common advantages of the Euclidean space and Hyperbolic space.

Firstly, we map the neural manifold from the Euclidean space \( \mathbb{R}^n \) to the Poincaré ball \( \mathbb{D}^n_r \), where \( \bar{g} \) is a solution to the Ricci flow. By adding the regularization to the neural network, the tensor \(-2 \text{Ric}(\bar{g})\) approaches zero to make \( \bar{g} \) satisfy the Definition 3.2. Secondly, we perform the Ricci flow for evolving the metric \( \bar{g} \) to the Hyperbolic metric \( g^H \). Thirdly, we map the neural manifold from the Poincaré ball \( \mathbb{D}^n_r \) to the Euclidean space \( \mathbb{R}^n \).

Fourthly, we complete the backpropagation of the gradient for the neural network.

Since the Poincaré ball is conformal to the Euclidean space, we give the exponential and logarithmic maps.

**Lemma 4.1.** With respect to the origin \( (x = 0) \), the exponential map \( \exp_r : T_x \mathbb{D}^n_r \rightarrow \mathbb{D}^n_r \) and the logarithmic map \( \log_r : \mathbb{D}^n_r \rightarrow T_x \mathbb{D}^n_r \) are given for \( \mu \in T_x \mathbb{D}^n_r \setminus \{0\} \) and \( \nu \in \mathbb{D}^n_r \setminus \{0\} \) by:

\[
\exp_r(\mu) = \tanh(|\mu|) \frac{\mu}{\sqrt{\mu}||\mu||},
\]

\[
\log_r(\nu) = \tanh^{-1}(\frac{|\nu|}{|\sqrt{\nu}||\nu||}) \frac{\nu}{\sqrt{\nu}||\nu||}.
\]

**Proof.** The proofs are followed with (Ganea et al., 2018). The algebraic check concludes the identity \( \log_r(\exp_r(\mu)) = \mu \).

The overall process is illustrated in Figure 1 where the map \( \psi \) and \( \psi^{-1} \) are the exponential map and logarithm map respectively.

4.3. Ricci Flow Assisted Eucl2Hyp2Eucl Neural Networks

The above discussion leaves the foreshadowing for designing neural networks. The precise computation of neural networks with \( l \) layers is performed as follows:

\[
x = f(a; \theta, b),
\]

\[
f(a; \theta, b) = \sigma_1[\cdots \sigma_2(\sigma_1(a\theta_1 + b_1)\theta_2 + b_2) + \cdots + b_l],
\]

\[
y = \text{softmax}(x),
\]

where \( a \) is the input, \( \sigma \) is a nonlinear “activation” function, \( \theta \) is the weight and \( b \) is the bias.

In addition to requiring the neural networks to converge, we also require that \( \bar{g} \) is \( \epsilon \)-close to \( g^H \) based on Definition 3.2, which is the necessary condition for the evolution of the Ricci flow. Obviously, we may achieve the goal by adding a regularization into loss function. Followed by Eq. 19, we yield the regularization

\[
\bar{g}^N = \left\| \frac{\bar{g}|_{k1} - \bar{g}|_{k2}}{k1 - k2} - \frac{\bar{g}|_{l1} - \bar{g}|_{l2}}{l1 - l2} \right\|_{L^2}^2.
\]

Combined with Definition 3.2, the upper bound of Eq. 25 is estimated by

\[
\bar{g}^N \leq \left\| \frac{(1 + \epsilon)^2 g^H|_{k1} - g^H|_{k2}}{(1 + \epsilon)(k1 - k2)} - \frac{g^H|_{l1} - (1 + \epsilon)^2 g^H|_{l2}}{(1 + \epsilon)(l1 - l2)} \right\|_{L^2}^2
\]

\[
= \frac{g^E}{{1 + \epsilon}} \left\| \frac{(1 + \epsilon)^2 \lambda^2_{2, k1} - \lambda^2_{2, k2}}{k1 - k2} - \frac{\lambda^2_{2, l1} - (1 + \epsilon)^2 \lambda^2_{2, l2}}{l1 - l2} \right\|_{L^2}^2.
\]

As the evolution of the Ricci flow approaches to converge, the estimate of Eq. 25 tends to be stable:

\[
\bar{g}^N \text{ Ricci flow } N = \left\| \frac{\lambda^2_{2, k1} - \lambda^2_{2, k2}}{k1 - k2} - \frac{\lambda^2_{2, l1} - \lambda^2_{2, l2}}{l1 - l2} \right\|_{L^2}^2.
\]
Ricci flow will complete the final convergence. Each training of the neural network includes these two stages, therefore, Ricci flow assisted Eucl2Hyp2Eucl neural networks are trained on dynamically stable Poincaré embeddings as shown in Algorithm 1.

**Algorithm 1** For a gradient update of Ricci flow assisted Eucl2Hyp2Eucl neural networks, we choose 4 different translations: k1, k2, j1, and j2. We define a target z and a balancing parameter α(ε). For x embedded in the Euclidean space, one uses ∇E as the gradients. Otherwise, for x embedded in the Poincaré ball with gH and ˜g, one uses ∇E and ˜g−1∇E as the gradients based on Corollary 3.1.

1: {Inference}
2: for m in k1, k2, j1, j2 do
3: \[ x_m = f(a_m; \theta, b) \] based on Eq. 24
4: \[ ˜g x_m = \exp_{x_m}(x_m) \] based on Eq. 22
5: end for
6: {Regularization}
7: \[ \bar{N} = \parallel \frac{\tilde{g}_{k1} - \tilde{g}_{k2}}{k1 - k2} - \frac{\tilde{g}_{j1} - \tilde{g}_{j2}}{j1 - j2} \parallel^2 \] based on Eq. 25
8: \[ \bar{g} x_{k1} \rightarrow g^H x_{k1} \] by computing the Ricci flow (\( \tilde{g} \rightarrow g^H \)) based on Eq. 1
9: \[ N = \parallel \frac{\lambda^2_{\tilde{g}_{k1} - \lambda^2_{k1 - k2}}}{k1 - k2} - \frac{\lambda^2_{\tilde{g}_{j1} - \lambda^2_{j1 - j2}}}{j1 - j2} \parallel^2 \] based on Eq. 28
10: {Loss}
11: \[ x_{k1} = \log_{x_{k1}}(g^H x_{k1}) \] based on Eq. 23
12: \[ y = \text{softmax}(x_{k1}) \]
13: loss = \[ \parallel y - z \parallel^2_L + \alpha(\varepsilon) \cdot N \]

5. Experiment

**CIFAR datasets.** The two CIFAR datasets (Krizhevsky et al., 2009) consist of natural color images with 32×32 pixels, respectively 50,000 training and 10,000 test images, and we hold out 5,000 training images as a validation set from the training set. CIFAR10 consists of images organized into 10 classes and CIFAR100 into 100 classes. We adopt a standard data augmentation scheme (random corner cropping and random flipping) that is widely used for these two datasets. We normalize the images using the channel means and standard deviations in preprocessing.

**Settings.** We set total training epochs as 200 while the learning rate of each parameter group is set as a cosine annealing schedule. The learning strategy is a weight decay of 0.0005, a batch size of 128, SGD optimization. On CIFAR10 and CIFAR100 datasets, we apply ResNet18 (He et al., 2016), ResNet50 (He et al., 2016), VGG11 (Simonyan & Zisserman, 2014) and MobileNetV2 (Sandler et al., 2018) to test the classification accuracy. All experiments are conducted for 5 times, and the statistics of the last 10/5 epochs’ test accuracy are reported as a fair comparison.

**Details.** For Ricci flow assisted Eucl2Hyp2Eucl neural networks and all Euclidean neural networks, we both use the same training strategy and network structure. Note that we both train neural networks from scratch with the initialization Xavier (Glorot & Bengio, 2010).

5.1. Classification Tasks

In this experiment, we compare the classification accuracy of Ricci flow assisted Eucl2Hyp2Eucl neural networks and all Euclidean neural networks on CIFAR datasets. As Table 1 shows, our proposed Ricci flow assisted Eucl2Hyp2Eucl neural networks have better performance than all Euclidean neural networks. Meanwhile, compared to CIFAR10 dataset, the improvement on CIFAR100 dataset seems to be more remarkable. We conjecture that more complex classification tasks bring more meaningful geometric structures to the neural network, and Ricci flow assisted Eucl2Hyp2Eucl neural networks can just mine these geometric representations as much as possible.

Note that experiment on ImageNet can be found in Appendix E.

5.2. Metrics Evolution Analysis

For the training of Ricci flow assisted Eucl2Hyp2Eucl neural networks, we hope to observe the evolution of neural manifolds by the change of metrics. Meanwhile, as far as we define the metric \( \tilde{g}(t) \), we can use the length \( |ds^2| = \sqrt{\sum_{i,j} \tilde{g}_{ij}(t) d\xi_i d\xi_j} \) to intuitively reflect the change of metrics. Specifically, we define a ball whose radius is equal to \( |ds^2| \):

\[
B_r(t) := \left\{ r = \left( \sum_{i,j} \tilde{g}_{ij}(t) d\xi_i d\xi_j \right)^{\frac{1}{2}} \right\}. \quad (29)
\]

By observing the change of the ball in Figure 2, we can know the change of the metric. Through simple observation, metrics \( \tilde{g}(t) \) have a rapid convergence at the beginning of training, and then become relatively flat. It seems that the convergence behavior of metrics is affected by the network structure rather than the depth (There has the similar evolution behavior on ResNet18 and ResNet50). In the training, experiments show that all metrics for Ricci flow assisted Eucl2Hyp2Eucl neural networks have a stable convergence. It is consistent with the evolution of scaled Ricci-DeTurck flow in Section 3.

---

Note: Ricci flow assisted Eucl2Hyp2Eucl neural networks are only used in the training, and we also use all Euclidean neural networks in offline inference.
Table 1. The classification accuracy results on CIFAR datasets with ResNet18, ResNet50, VGG11 and MobileNetV2.

| Network     | Method                                               | CIFAR10          | CIFAR100         |
|-------------|------------------------------------------------------|------------------|------------------|
| ResNet18    | All Euclidean Neural Network                         | 95.25±0.13%      | 77.25±0.15%      |
|             | Ricci Flow Assisted Eucl2Hyp2Eucl Neural Network     | 95.73±0.09%      | 77.87±0.12%      |
| ResNet50    | All Euclidean Neural Network                         | 95.58±0.17%      | 78.47±0.33%      |
|             | Ricci Flow Assisted Eucl2Hyp2Eucl Neural Network     | 96.01±0.06%      | 79.79±0.28%      |
| VGG11       | All Euclidean Neural Network                         | 92.28±0.16%      | 71.66±0.21%      |
|             | Ricci Flow Assisted Eucl2Hyp2Eucl Neural Network     | 93.02±0.12%      | 73.28±0.27%      |
| MobileNetV2 | All Euclidean Neural Network                         | 92.28±0.25%      | 72.33±0.31%      |
|             | Ricci Flow Assisted Eucl2Hyp2Eucl Neural Network     | 93.75±0.22%      | 73.42±0.21%      |

5.3. Training Time Analysis

Our hardware environment is conducted with an Intel(R) Xeon(R) E5-2650 v4 CPU(2.20GHz), GeForce GTX 1080Ti GPU. We test the training time per iteration as for Ricci flow assisted Eucl2Hyp2Eucl neural networks and all Euclidean neural networks with ResNet18 in CIFAR10. For finishing one iteration of training, Ricci flow assisted Eucl2Hyp2Eucl neural network costs 81.06s and all Euclidean neural network costs 39.76s.

6. Conclusion

Ricci flow assisted Eucl2Hyp2Eucl neural networks not only provide the simply convenience and closed-form formula in the Euclidean space (for offline inference), but also take into account the geometric representation of the Hyperbolic space. This provides a new idea for neural networks to obtain meaningful geometric representations. Empirically, we found that Ricci flow assisted neural networks outperform with their all Euclidean counterparts on classification tasks. The Ricci flow plays a vital role in Eucl2Hyp2Eucl neural networks, not only eliminating an $L^2$-norm perturbation of the Hyperbolic metric, but also acting as a smooth evolution from the Euclidean space to the Poincaré ball. Actually, Eucl2Hyp2Eucl neural networks without Ricci flow will become the same as all Euclidean neural networks. We hope that this paper will open an exciting future direction which will use the Ricci flow to assist neural network training in a dynamically stable manifold.

References

Amari, S., Barndorff-Nielsen, O. E., Kass, R. E., Lauritzen, S. L., and Rao, C. Differential geometry in statistical inference. IMS, 1987.

Amari, S.-i. *Information geometry and its applications*, volume 194. Springer, 2016.
Anderson, J. W. *Hyperbolic geometry*. Springer Science & Business Media, 2006.

Bamler, R. H. Stability of hyperbolic manifolds with cusps under ricci flow. *arXiv preprint arXiv:1004.2058*, 2010.

Bronstein, M. M., Bruna, J., LeCun, Y., Szlam, A., and Vandergheynst, P. Geometric deep learning: going beyond euclidean data. *IEEE Signal Processing Magazine*, 34(4): 18–42, 2017.

Chen, J., Huang, T., Chen, W., and Liu, Y. Thoughts on the consistency between ricci flow and neural network behavior. *arXiv preprint arXiv:2111.08410*, 2021.

Chen, L.-C., Papandreou, G., Kokkinos, I., Murphy, K., and Yuille, A. L. Semantic image segmentation with deep convolutional nets and fully connected crfs. *arXiv preprint arXiv:1412.7062*, 2014.

DeTurck, D. M. Deforming metrics in the direction of their ricci tensors. *Journal of Differential Geometry*, 18(1): 157–162, 1983.

Ganea, O., Bécigneul, G., and Hofmann, T. Hyperbolic entailment cones for learning hierarchical embeddings. In *International Conference on Machine Learning*, pp. 1646–1655. PMLR, 2018.

Ganea, O.-E., Bécigneul, G., and Hofmann, T. Hyperbolic neural networks. *Advances in Neural Information Processing Systems 31*, pp. 5345–5355, 2019.

Girshick, R. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pp. 1440–1448, 2015.

Glorot, X., and Bengio, Y. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pp. 249–256. JMLR Workshop and Conference Proceedings, 2010.

Hamilton, R. S. et al. Three-manifolds with positive ricci curvature. *J. Differential geom*, 17(2):255–306, 1982.

He, K., Zhang, X., Ren, S., and Sun, J. Delving deep into rectifiers: Surpassing human-level performance on imagenet classification. In *Proceedings of the IEEE international conference on computer vision*, pp. 1026–1034, 2015.

He, K., Zhang, X., Ren, S., and Sun, J. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.

Kaul, P. and Lall, B. Riemannian curvature of deep neural networks. *IEEE transactions on neural networks and learning systems*, 31(4):1410–1416, 2019.

Krizhevsky, A., Hinton, G., et al. Learning multiple layers of features from tiny images. 2009.

Krizhevsky, A., Sutskever, I., and Hinton, G. E. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25: 1097–1105, 2012.

Ladyzhenskaia, O. A., Solonnikov, V. A., and Ural’tseva, N. N. *Linear and quasi-linear equations of parabolic type*, volume 23. American Mathematical Soc., 1988.

Li, H. and Yin, H. On stability of the hyperbolic space form under the normalized ricci flow. *International Mathematics Research Notices*, 2010(15):2903–2924, 2010.

Long, J., Shelhamer, E., and Darrell, T. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3431–3440, 2015.

Martens, J. New insights and perspectives on the natural gradient method. *Journal of Machine Learning Research*, 21:1–76, 2020.

Nickel, M. and Kiela, D. Poincaré embeddings for learning hierarchical representations. *Advances in neural information processing systems*, 30:6338–6347, 2017.

Sala, F., De Sa, C., Gu, A., and Ré, C. Representation tradeoffs for hyperbolic embeddings. In *International conference on machine learning*, pp. 4460–4469. PMLR, 2018.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., and Chen, L.-C. Mobilenetv2: Inverted residuals and linear bottlenecks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4510–4520, 2018.

Schnürer, O. C., Schulze, F., and Simon, M. Stability of hyperbolic space under ricci flow. *arXiv preprint arXiv:1003.2107*, 2010.
Sheridan, N. and Rubinstein, H. Hamilton’s ricci flow. Honour thesis, 2006.

Shi, W.-X. Deforming the metric on complete riemannian manifolds. Journal of Differential Geometry, 30(1):223–301, 1989.

Simon, M. Deformation of $c^0$ riemannian metrics in the direction of their ricci curvature. Communications in Analysis and Geometry, 10(5):1033–1074, 2002.

Simonyan, K. and Zisserman, A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.

Suneeta, V. Investigating the off-shell stability of anti-de sitter space in string theory. Classical and Quantum Gravity, 26(3):035023, 2009.

Wald, R. M. General relativity. University of Chicago press, 2010.

Ye, R. Ricci flow, einstein metrics and space forms. Transactions of the american mathematical society, 338(2): 871–896, 1993.
A. Differential Geometry

1. Riemann curvature tensor (Rm) is a (1,3)-tensor defined for a 1-form $\omega$:

$$R_{ijkl} \omega_l = \nabla_i \nabla_j \omega_k - \nabla_j \nabla_i \omega_k$$

where the covariant derivative of $F$ satisfies

$$\nabla_p F_{ijkl} = \partial_p F_{ijkl} + \sum_{s=1}^l F_{ijkl;q} \Gamma^q_{ps} - \sum_{s=1}^k F_{ijkl;p} \Gamma^p_{qs}.$$

In particular, coordinate form of the Riemann curvature tensor is:

$$R_{ijkl} = \partial_i \Gamma_{jk}^l - \partial_j \Gamma_{ik}^l + \Gamma_{pj}^k \Gamma_{ik}^l - \Gamma_{pi}^k \Gamma_{kj}^l.$$

2. Christoffel symbol in terms of an ordinary derivative operator is:

$$\Gamma^k_{ij} = \frac{1}{2} g^{pq} (\partial_i g_{jq} + \partial_j g_{qi} - \partial_q g_{ij}).$$

3. Ricci curvature tensor (Ric) is a (0,2)-tensor:

$$R_{ij} = R^p_{pij}.$$

4. Scalar curvature is the trace of the Ricci curvature tensor:

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

5. Lie derivative of $F$ in the direction $\frac{d\varphi(t)}{dt}$:

$$\mathcal{L}_{\frac{d\varphi(t)}{dt}} F = \left( \frac{d}{dt} \varphi^*(t) F \right)_{t=0}$$

where $\varphi(t) : M \to M$ for $t \in (-\epsilon, \epsilon)$ is a time-dependent diffeomorphism of $\mathcal{M}$ to $\mathcal{M}$.

B. Proof for the Ricci Flow

B.1. Proof for Lemma 2.3

**Definition B.1.** The linearization of the Ricci curvature tensor is given by

$$D[Ric](h)_{ij} = -\frac{1}{2} g^{pq} (\nabla_p \nabla_q h_{ij} + \nabla_i \nabla_j h_{pq} - \nabla_q \nabla_i h_{jp} - \nabla_q \nabla_j h_{ip}).$$

**Proof.** Based on Appendix A, we have

$$\nabla_q \nabla_i h_{jp} = \nabla_i \nabla_q h_{jp} - R^r_{qij} h_{rp} - R^r_{qip} h_{jm}.$$

Combining with Definition B.1, we can obtain the deformation equation because of $\nabla_k g_{ij} = 0$,

$$D[-2Ric](h)_{ij} = g^{pq} \nabla_p \nabla_q h_{ij} + \nabla_i \left( \frac{1}{2} \nabla_j h_{pq} - \nabla q h_{jp} \right) + \nabla_j \left( \frac{1}{2} \nabla_i h_{pq} - \nabla q h_{ip} \right) + O(h_{ij}).$$

$$= g^{pq} \nabla_p \nabla_q h_{ij} + \nabla_i V_j + \nabla_j V_i + O(h_{ij}).$$
B.2. Description of the DeTurck Trick

Using a time-dependent diffeomorphism \( \varphi(t) \), we express the pullback metrics \( \bar{g}(t) \):

\[
g(t) = \varphi^*(t) \bar{g}(t)
\]

is a solution of the Ricci flow. Based on the chain rule for the Lie derivative in Appendix A, we can calculate

\[
\frac{\partial}{\partial t} g(t) = \frac{\partial}{\partial t} \left( \varphi^*(t) \bar{g}(t) \right) = \left( \frac{\partial}{\partial \tau} \left( \varphi^*(t+\tau) \bar{g}(t+\tau) \right) \right)_{\tau=0} + \left( \frac{\partial}{\partial \tau} \left( \varphi^*(\tau) \bar{g}(t) \right) \right)_{\tau=0}
\]

\[
= \varphi^*(t) \frac{\partial}{\partial \tau} \bar{g}(t) + \varphi^*(t) \mathcal{L}_{\frac{\partial}{\partial \tau} \bar{g}(t)} \bar{g}(t).
\]

With the help of Equation (1), for the reparameterized metric, we have

\[
\frac{\partial}{\partial t} \bar{g}(t) = \varphi^*(t) \frac{\partial}{\partial \tau} \bar{g}(t) + \varphi^*(t) \mathcal{L}_{\frac{\partial}{\partial \tau} \bar{g}(t)} \bar{g}(t) = -2 \text{Ric}(\varphi^*(t) \bar{g}(t)) = -2 \varphi^*(t) \text{Ric}(\bar{g}(t)).
\]

The diffeomorphism invariance of the Ricci curvature tensor is used in the last step. The above equation is equivalent to

\[
\frac{\partial}{\partial t} \bar{g}(t) = -2 \text{Ric}(\bar{g}(t)) - \mathcal{L}_{\frac{\partial}{\partial \tau} \bar{g}(t)} \bar{g}(t).
\]

B.3. Proof for Theorem 2.5

Considering any \( x \in M, t_0 \in [0, T), V \in T_xM \), we have

\[
\left| \log \left( \frac{g_x(t_0)(V,V)}{g_x(0)(V,V)} \right) \right| = \left| \int_{t_0}^t \frac{\partial}{\partial t} \log g_x(t)(V,V) \, dt \right|
\]

\[
= \left| \int_{t_0}^t \frac{\partial}{\partial t} g_x(t)(V,V) \, dt \right|
\]

\[
\leq \int_{t_0}^t \left| \frac{\partial}{\partial t} g_x(t) \left( \frac{V}{|V|_{g_x(t)}}, \frac{V}{|V|_{g_x(t)}} \right) \right| dt
\]

\[
\leq \int_{t_0}^t \left| \frac{\partial}{\partial t} g_x(t) \right|_{g_x(t)} dt
\]

\[
\leq C.
\]

By exponentiating both sides of the above inequality, we have

\[
e^{-C} g_x(0)(V,V) \leq g_x(t_0)(V,V) \leq e^C g_x(0)(V,V).
\]

This inequality can be rewritten as

\[
e^{-C} g_x(0) \leq g_x(t_0)(V,V) \leq e^C g_x(0)(V,V)
\]

because it holds for any \( V \). Thus, the metrics \( g(t) \) are uniformly equivalent to \( g(0) \).

Now, we have the well-defined integral:

\[
g_x(T) - g_x(0) = \int_0^T \frac{\partial}{\partial t} g_x(t) \, dt.
\]
We say that this integral is well-defined because of two reasons. Firstly, as long as the metrics are smooth, the integral exists. Secondly, the integral is absolutely integrable. Based on the norm inequality induced by $g(0)$, one has

$$|g_x(T) - g_x(t)|_{g(0)} \leq \int_t^T \left| \frac{\partial}{\partial t} g_x(t) \right|_{g(0)} dt.$$  

For each $x \in \mathcal{M}$, the above integral will approach to zero as $t$ approaches $T$. Because $\mathcal{M}$ is compact, the metrics $g(t)$ converge uniformly to a continuous metric $g(T)$ which is uniformly equivalent to $g(0)$ on $\mathcal{M}$. Moreover, we can show that

$$e^{-C} g_x(0) \leq g_x(T) \leq e^C g_x(0).$$

**C. Proof for Lemma 3.3**

Empirically, we have

$$\frac{\partial}{\partial t} \bar{g} - g^H = \sum \frac{\partial}{\partial t} (\bar{g}_{ij}^2 - 2\bar{g}_{ij} g^H_{ij}).$$

At this time, $\bar{g}_{ij}$ or $g^H_{ij}$ is the given coordinate representation. Based on (Shi, 1989), we can write $\frac{\partial}{\partial t} \bar{g}_{ij}$:

$$\frac{\partial}{\partial t} \bar{g}_{ij} = \bar{g}^{ab} \nabla_a \nabla_b \bar{g}_{ij} + 2\bar{g}_{ij} (\bar{g}_{kl} (\bar{g}_{kl} - \bar{g}_{kl})) + 2 (\bar{g}_{ij} - g^H_{ij}) + \frac{1}{2} \bar{g}^{ab} \bar{g}^{pq} (\nabla_i \bar{g}_{pa} \nabla_j \bar{g}_{pb} + 2 \nabla_a \bar{g}_{jp} \nabla_q \bar{g}_{ib} - 2 \nabla_a \bar{g}_{jp} \nabla_b \bar{g}_{iq} - 2 \nabla_i \bar{g}_{pa} \nabla_b \bar{g}_{jq}).$$

In the third line of the above formula, we denote contractions as in (Hamilton et al., 1982):

$$(\nabla \bar{g} \ast \nabla g)_{ij} = \bar{g}^{ab} \bar{g}^{pq} (\nabla_i \bar{g}_{pa} \nabla_j \bar{g}_{pb} + 2 \nabla_a \bar{g}_{jp} \nabla_q \bar{g}_{ib} - 2 \nabla_a \bar{g}_{jp} \nabla_b \bar{g}_{iq} - 2 \nabla_i \bar{g}_{pa} \nabla_b \bar{g}_{jq}).$$

Followed by Definition 3.2, we have the estimate

$$\frac{\partial}{\partial t} \left| \bar{g} - g^H \right|^2 = 2 \sum (\bar{g}_{ij} - g^H_{ij}) \left( \frac{\partial}{\partial t} \bar{g}_{ij} \right)$$

$$= \bar{g}^{ij} \nabla_i \nabla_j \left| \bar{g} - g^H \right|^2 - 2 |\nabla \bar{g}|^2 + 2 \sum (\bar{g}_{ij} - g^H_{ij}) \left[ 2 (\bar{g}_{ij} - g^H_{ij}) - 2 \bar{g}_{ij} \sum (g^H_{kl} (\bar{g}_{kl} - g^H_{kl})) \right]$$

$$+ \sum (\bar{g}_{ij} - g^H_{ij}) (\nabla \bar{g} \ast \nabla \bar{g})_{ij}$$

$$= \bar{g}^{ij} \nabla_i \nabla_j \left| \bar{g} - g^H \right|^2 - 2 |\nabla \bar{g} - g^H|^2$$

$$+ 4 \left| \bar{g} - g^H \right|^2 - 4 \sum \left[ (\bar{g}_{ij} - g^H_{ij}) \bar{g}_{ij} \sum (g^H_{kl} (\bar{g}_{kl} - g^H_{kl})) \right]$$

$$\leq \bar{g}^{ij} \nabla_i \nabla_j \left| \bar{g} - g^H \right|^2 - 2 |\nabla \bar{g} - g^H|^2 + 4 \left| \bar{g} - g^H \right|^2$$

$$+ \frac{4\epsilon}{1 + \epsilon} \sum \left[ g^H_{ij} \bar{g}_{ij} \sum (g^H_{kl} (\bar{g}_{kl} - g^H_{kl})) \right].$$

In the second line, we have $\nabla_i \nabla_j (\bar{g}^2) = \nabla_i (2\bar{g} \cdot \nabla_j \bar{g}) = 2 \nabla_i \bar{g} \cdot \nabla_j \bar{g} + 2 \bar{g} \nabla_i \nabla_j \bar{g}$ and $\nabla_i \nabla_j (g^H)^2 = 0$ because $\nabla$ is compatible with $g^H$. 
D. Proof for Corollary 3.1

**Definition D.1.** $D[P : Q]$ is called a divergence when it satisfies the following criteria:

1) $D[P : Q] \geq 0$.
2) $D[P : Q] = 0$ when and only when $P = Q$.
3) When $P$ and $Q$ are sufficiently close, by denoting their coordinates by $\xi_P$ and $\xi_Q = \xi_P + d\xi$, the Taylor expansion of $D$ is written as

$$D[\xi_P : \xi_P + d\xi] = \frac{1}{2} \sum_{i,j} g_{ij}(\xi_P)d\xi_id\xi_j + O(|d\xi|^3),$$

and Riemannian metric $g_{ij}$ is positive-definite, depending on $\xi_P$.

With constraint the divergence in a constant $C$, we do the minimization of the loss function $\text{loss}(\theta)$ in Lagrangian form:

$$d\theta^* = \arg \min_{d\theta} \text{loss}(\theta + d\theta)$$

$$s.t. \: D[\theta : \theta + d\theta] = C$$

$$= \arg \min_{d\theta} \text{loss}(\theta + d\theta) + \lambda (D[\theta : \theta + d\theta] - C)$$

$$\approx \arg \min_{d\theta} \text{loss}(\theta) + \partial_\theta \text{loss}(\theta)^\top d\theta + \frac{\lambda}{2} d\theta^\top g_{ij} d\theta - C\lambda.$$

To solve the above minimization, we set its derivative with respect to $d\theta$ to zero:

$$0 = \partial_\theta L(\theta) + \frac{\lambda}{2} g_{ij} d\theta$$

$$- \frac{\lambda}{2} g_{ij} d\theta = \partial_\theta \text{loss}(\theta)$$

$$d\theta = \frac{2}{\lambda} g_{ij}^{-1} \partial_\theta \text{loss}(\theta).$$

Where a constant factor $2/\lambda$ can be absorbed into learning rate. The opposite direction $-d\theta$ is the steepest descent direction in a Riemannian manifold endowed with $g_{ij}$.

E. ImageNet

**ImageNet dataset.** ILSVRC2012 (Deng et al., 2009) image classification dataset consists of 1.2 million high-resolution natural images where the validation set contains 50k images. These images are organized into 1000 categories of the object for training, which are resized to $224 \times 224$ pixels before fed into the network. In the next experiment, we report our single-crop evaluation results using top-1 and top-5 accuracy. In preprocessing, images are resized randomly to $256 \times 256$ pixels, and then a random crop of $224 \times 224$ is selected for training.

**Settings.** We set total training epochs as 100 where the learning rate of each parameter group is set as a cosine annealing schedule. The learning strategy is a weight decay of 0.0001, a batch size of 128, SGD optimization. On Table 2, we apply ResNet18 and ResNet50 to test the classification accuracy.

| Network | Method | Top-1 Acc | Top-5 Acc |
|---------|--------|-----------|-----------|
| ResNet18 | All Euclidean Neural Network | 69.76% | 89.08% |
|         | Ricci Flow Assisted Eucl2Hyp2Eucl Neural Network | 70.55% | 89.67% |
| ResNet50 | All Euclidean Neural Network | 76.13% | 92.86% |
|         | Ricci Flow Assisted Eucl2Hyp2Eucl Neural Network | 77.41% | 93.54% |