Hyperbolic Image Segmentation

Mina GhadimiAtigh[1,*], Julian Schoep[1,*], Erman Acar[2], Nanne van Noord[1], Pascal Mettes[1]
[1]University of Amsterdam, [2]Leiden University, [2]Vrije Universiteit Amsterdam

CVPR 2022

Abstract

Abstract

For image segmentation, the current standard is to perform pixel-level optimization and inference in Euclidean output embedding spaces through linear hyperplanes. In this work, we show that hyperbolic manifolds provide a valuable alternative for image segmentation and propose a tractable formulation of hierarchical pixel-level classification in hyperbolic space. Hyperbolic Image Segmentation opens up new possibilities and practical benefits for segmentation, such as uncertainty estimation and boundary information for free, zero-label generalization, and increased performance in low-dimensional output embeddings.

- 三句式

1. 目前的图像分割的做法
2. 本论文中的做法
3. 指出本方法的三个优点

+简洁

+适用于某一方法在该领域的首次尝试

-没有突出已有方法的不足

Introduction

1. Introduction

A ubiquitous goal in visual representation learning is to obtain discriminative and generalizable embeddings. Such visual embeddings are learned in a deep and highly non-linear fashion. On top, a linear layer separates categories through Euclidean hyperplanes. [The choice for a zero curvature Euclidean embedding space, although a *de facto* standard, requires careful re-consideration as it has direct consequences for how well a task can be optimized given the latent structure that is inherently present in both the data and the category space [19, 22, 29].]

- Intro第一段
- 介绍视觉表示学习任务的一般范式，并简要说明其中的关键问题
- 通过说明embedding space选择的重要性，引出下一段中新的embedding space

Introduction

This work takes inspiration from recent literature advocating hyperbolic manifolds as embedding spaces for machine learning and computer vision tasks. [Foundational work showed that hyperbolic manifolds are able to embed hierarchies and tree-like structures with minimal distortion [29]. Follow up work has demonstrated the benefits of hyperboles for various tasks with latent hierarchical structures, from text embedding [42, 55] to graph inference [8, 12, 22]. Notably, Khrulkov *et al*. [19] showed that hyperbolic embeddings also have profound connections to visual data, due to latent hierarchical structures present in vision datasets. This connection has brought along early hyperbolic success in computer vision for few-shot and zero-shot learning [15, 19, 23], unsupervised learning [32, 46], and video recognition [25, 40].]

- Intro第二段
- 介绍双曲流形在ML和CV的进展

（一句总述+一句非视觉领域+两句视觉领域）

Introduction

Common amongst current hyperbolic computer vision works is that the task at hand is global, *i.e.* an entire image or video is represented by a single vector in the hyperbolic embedding space [3, 19, 23, 25]. Here, our goal is to take hyperbolic deep learning to the pixel level. This generalization is however not trivial. The change of manifold brings different formulations for basic operations such as addition and multiplication, each with different spatial complexity. [Specifically, the additional spatial complexity that comes with the Möbius addition as part of the hyperbolic multinomial logistic regression makes it intractable to simultaneously optimize or infer all pixels of even a single image.] [Here, we propose an equivalent re-formulation of multinomial logistic regression in the Poincaré ball that bypasses the explicit computation of the Möbius addition, allowing for simultaneous segmentation optimization on batches of images in hyperbolic space.] [We furthermore outline how to incorporate hierarchical knowledge amongst labels in the hyperbolic embedding space, as previously advocated in image and video recognition [23, 25].] The proposed approach is general and can be plugged on top of any segmentation architecture. The code is available at https://github.com/MinaGhadimiAtigh/HyperbolicImageSegmentation.

- Intro第三段
- 叙述和已有双曲视觉工作的区别---在像素级别进行双曲空间的操作
- 所遇到的问题---空间复杂性
- 解决方案---提出了可并行的Mobinus加法操作
- 其他贡献

缺少：为什么要使用双曲空间的表示/双曲空间能给图像分割带来什么好处

Introduction

We perform a number of analyses to showcase the effect and new possibilities that come with Hyperbolic Image Segmentation. We present the following: *(i)* Hyperbolic embeddings provide natural measures for uncertainty estimation and for semantic boundary estimation in image segmentation, see Figure 1. Different from Bayesian uncertainty estimation, our approach requires no additional parameters or multiple forward passes, *i.e*. this information comes for free. *(ii):* Hyperbolic embeddings with hierarchical knowledge provide better zero-label generalization than Euclidean counterparts, *i.e*. hyperboles improve reasoning over unseen categories. *(iii):* Hyperbolic embeddings are preferred for fewer embedding dimensions. Low-dimensional effectiveness is a cornerstone in hyperbolic deep learning [29]. We find that these benefits extend to image segmentation, with potential for explainability and on-device segmentation [3]. We believe these findings bring new insights and opportunities to image segmentation.

- Intro第四段
- 从三个方面介绍本文的贡献

Hyperbolic Geometry

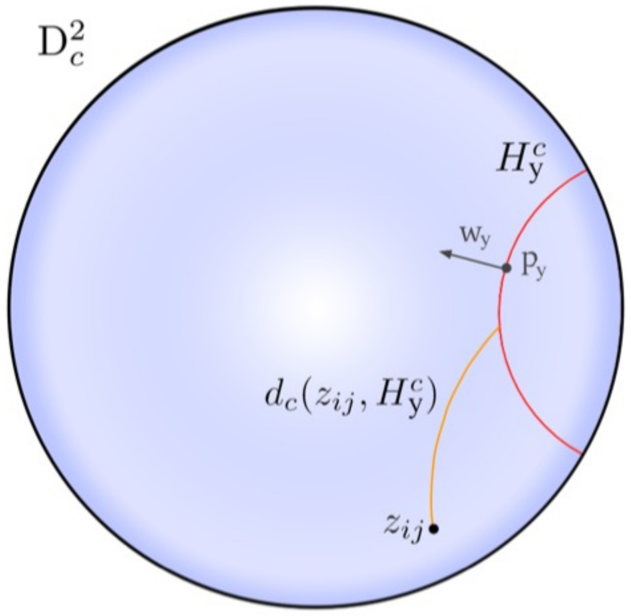
- The Poincare ball model

$$\mathbb{D}_c^n = \{x \in \mathbb{R}^n : c||x|| < 1\}$$

- c : a hyperparameter governing the curvature and radius of the ball

- different operation in $\mathbb{D}_c^n$:
  - Euclidean metric: $g^E = \mathbb{I}^n$
  - Riemannian metric: $g_x^{\mathbb{D}_c} = (\lambda_x^c)^2 g^E = \frac{2}{1 - c||x||^2}\mathbb{I}^n$

  - $\exp_v^c(x) = v \oplus_c \left(\tanh\left(\sqrt{c}\frac{\lambda_v^c||x||}{2}\right)\frac{x}{\sqrt{c}||x||}\right)$
  - $v \oplus_c w = \frac{(1 + 2c\langle v, w\rangle + c||w||^2)v + (1 - c||v||^2)w}{1 + 2c\langle v, w\rangle + c^2||v||^2||w||^2}.$



Hyperbolic Geometry

- A hyperplane in the Poincare ball:

$$H^c = \{z_{ij} \in \mathbb{D}_c^n, \langle -p \oplus_c z_{ij}, w \rangle = 0\}$$

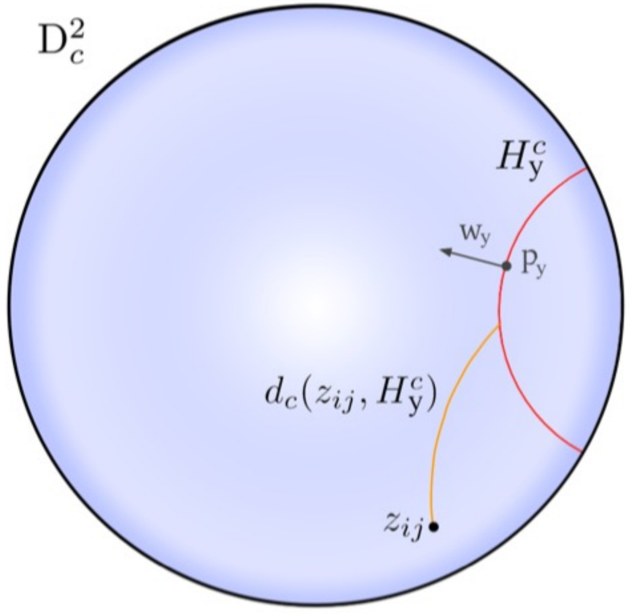
  - offset: p
  - orientation: w
  - exponential map of the network output at pixel location (i, j):

$$z_{ij} = \exp_0(f(X)_{ij})$$

$$d_c(z_{ij}, H_y^c) = \frac{1}{\sqrt{c}} \sinh^{-1}\left(\frac{2\sqrt{c}\langle -p_y \oplus_c z_{ij}, w_y \rangle}{(1 - c|| - p_y \oplus_c z_{ij}||^2)||w_y||}\right)$$

$$\zeta_y(z_{ij}) = \frac{\lambda_{p_y}^c ||w_y||}{\sqrt{c}} \sinh^{-1}\left(\frac{2\sqrt{c}\langle -p_y \oplus_c z_{ij}, w_y \rangle}{(1 - c|| - p_y \oplus_c z_{ij}||^2)||w_y||}\right)$$

$$p(\hat{y} = y | z_{ij}) \propto \exp(\zeta_y(z_{ij}))$$



Issue of spatial complexity

- Euclidean space: 0.5 GB
- Hyperbolic space: 132 GB
- Alternative operations in Hyperbolic space： 1.1 GB

$$\zeta_y(z_{ij}) = \frac{\lambda_{p_y}^c ||w_y||}{\sqrt{c}} \sinh^{-1}\left(\frac{2\sqrt{c}\langle -p_y \oplus_c z_{ij}, w_y\rangle}{(1-c||-p_y \oplus_c z_{ij}||^2)||w_y||}\right)$$

-> $\langle -p_y \oplus_c z_{ij}, w_y\rangle$ $||-p_y \oplus_c z_{ij}||^2$

$$\begin{aligned}\langle \hat{p}_y \oplus_c z_{ij}, w_y\rangle &= \langle \alpha\hat{p}_y + \beta z_{ij}, w_y\rangle, \\ &= \alpha\langle \hat{p}_y, w\rangle + \beta\langle z_{ij}, w\rangle\end{aligned}$$

$$||\hat{p}_y \oplus_c z_{ij}||^2 = \sum_{m=1}^{n} (\alpha\hat{p}_y^m + \beta z_{ij}^m)^2$$

Hierarchical Hyperbolic Class Embedding

- Preset a tree structure for all classes
- y : a class
- $\mathcal{A}_y$: the ancestors of y
- $\mathcal{H}_y = \{y\} \cup \mathcal{A}_y$
- S_h : the siblings of h

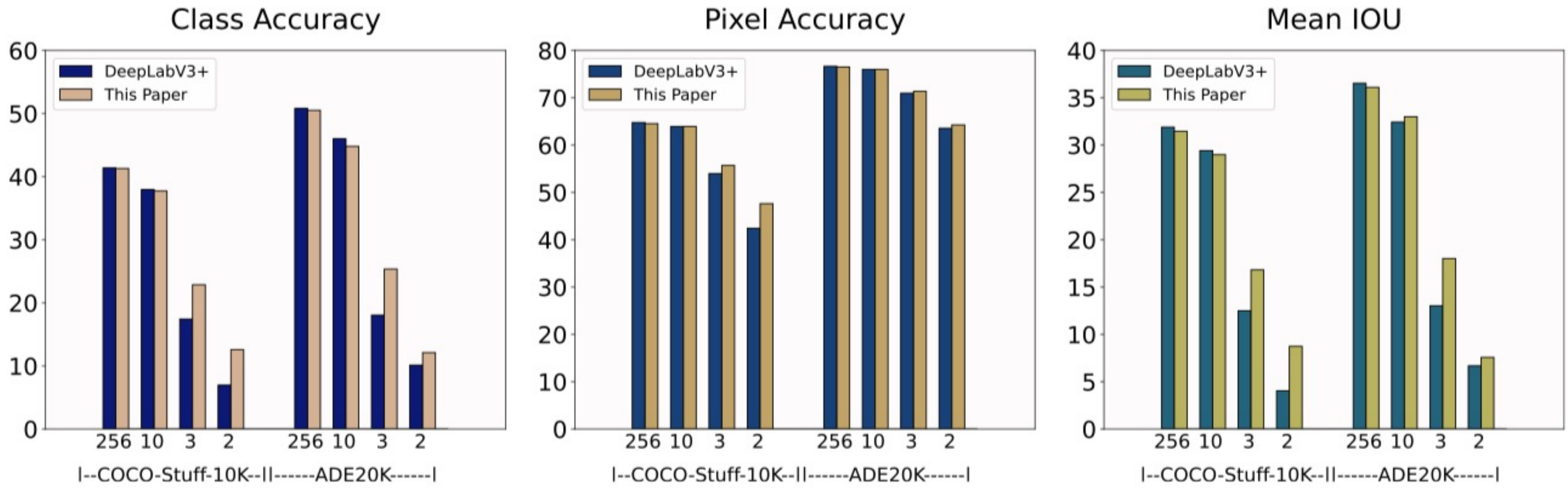
$$p(\hat{y} = y | z_{ij}) = \prod_{h \in \mathcal{H}_y} p(h | \mathcal{A}_h, z_{ij})$$
$$= \prod_{h \in \mathcal{H}_y} \frac{\exp(\zeta_h(z_{ij}))}{\sum_{s \in S_h} \exp(\zeta_s(z_{ij}))}$$

Experiment

- Dataset:
  - COCO-Stuff-10K(10k images, 171 classes)
  - Pascal VOC(12k images, 21 classes)
  - ADE20K(22k images, 150 classes)
- Backbone:
  - DeeplabV3+ with a ResNet101 backbone

Experiment

- Low-dimensional embedding effectiveness



Experiment

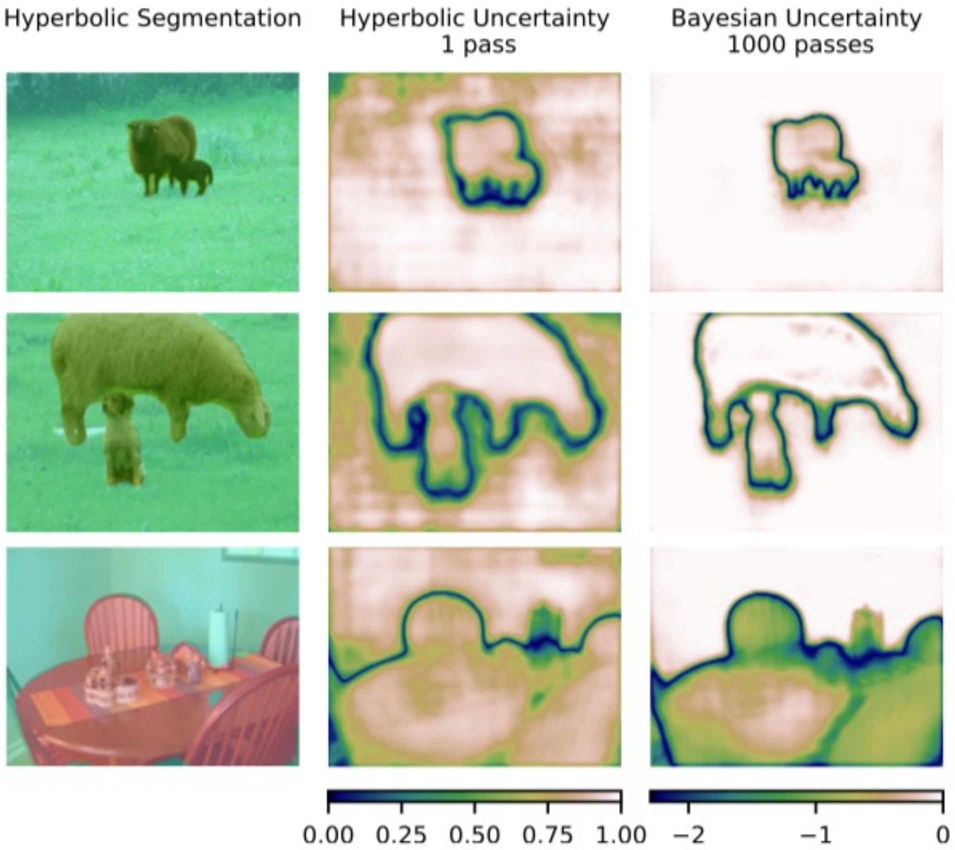
- Zero-label generalization
- unseen classes in training phase: replace them with an ignore label

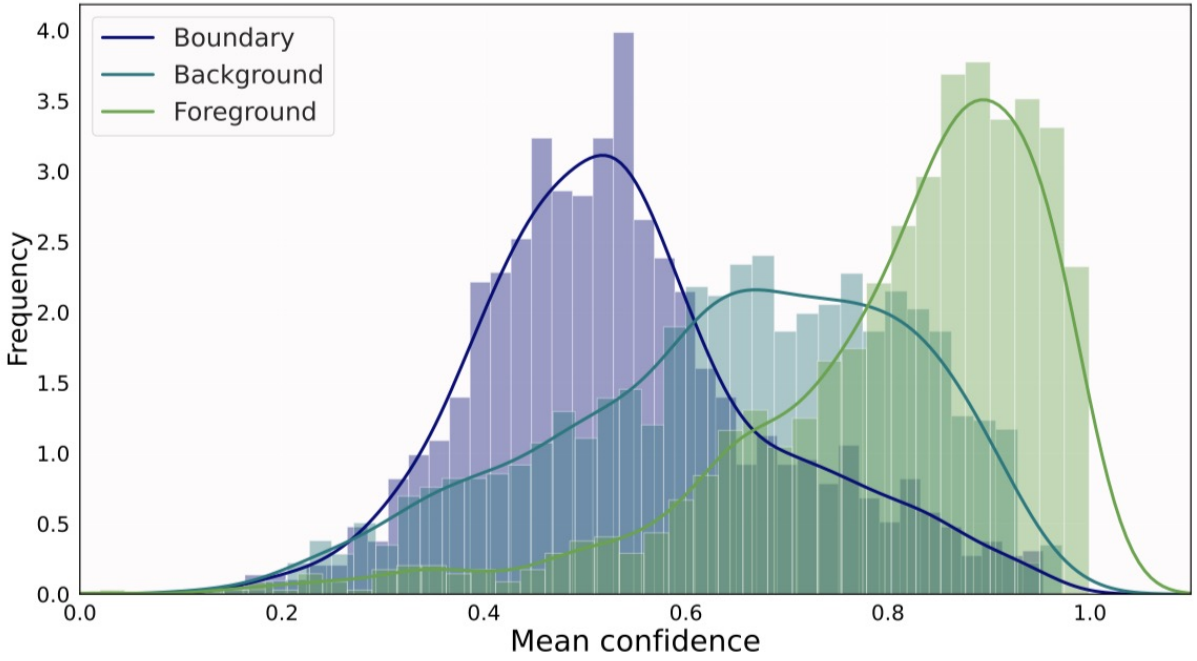
COCO-Stuff-10k				
Manifold	Hierarchical	Class Acc	Pixel Acc	mIOU
$\mathbb{R}$		0.44	0.33	0.23
$\mathbb{R}$	✓	3.29	48.65	18.53
$\mathbb{D}$	✓	**3.46**	**51.70**	**21.15**

Pascal VOC				
Manifold	Hierarchical	Class Acc	Pixel Acc	mIOU
$\mathbb{R}$		4.88	10.84	2.59
$\mathbb{R}$	✓	7.80	31.04	16.15
$\mathbb{D}$	✓	**12.15**	**47.92**	**34.87**

Experiment

- Uncertainty and boundary information for free
- Hyperbolic uncertainty: the distance to the origin of each pixel in the hyperbolic embedding space





Summary

Hyperbolic Class Embedding space:

+ A new representation space
+ Some geometric properties different from Euclidean space

- Additional computational cost
- An effective format for the visual task is needed to explore