Angular Gradient Sign Method: Uncovering Vulnerabilities in Hyperbolic Networks

Minsoo Jo[1], **Dongyoon Yang**[2*], **Taesup Kim**[1*]

[1]Graduate School of Data Science, Seoul National University
[2]AI Advanced Technology, SK hynix

Abstract

Adversarial examples in neural networks have been extensively studied in Euclidean geometry, but recent advances in *hyperbolic networks* call for a reevaluation of attack strategies in non-Euclidean geometries. Existing methods such as FGSM and PGD apply perturbations without regard to the underlying hyperbolic structure, potentially leading to inefficient or geometrically inconsistent attacks. In this work, we propose a novel adversarial attack that explicitly leverages the geometric properties of hyperbolic space. Specifically, we compute the gradient of the loss function in the tangent space of hyperbolic space, decompose it into a radial (depth) component and an angular (semantic) component, and apply perturbation derived solely from the angular direction. Our method generates adversarial examples by focusing perturbations in semantically sensitive directions encoded in angular movement within the hyperbolic geometry. Empirical results on image classification, cross-modal retrieval tasks and network architectures demonstrate that our attack achieves higher fooling rates than conventional adversarial attacks, while producing high-impact perturbations with deeper insights into vulnerabilities of hyperbolic embeddings. This work highlights the importance of geometry-aware adversarial strategies in curved representation spaces and provides a principled framework for attacking hierarchical embeddings.

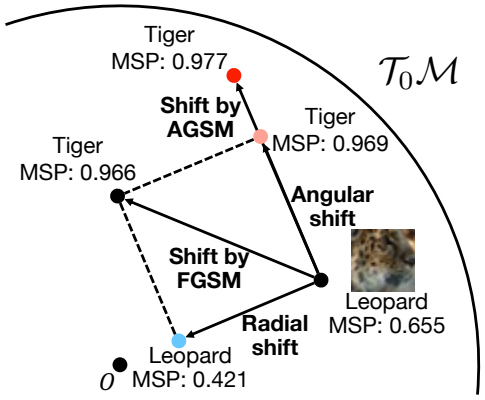

Figure 1: Overview of representation shifts induced by FGSM, AGSM, radial, and angular perturbations. We visualize how FGSM, AGSM, radial, and angular perturbations influence predictions and confidence (MSP; Maximum Softmax Probability). FGSM causes mixed, less semantic shifts, while radial perturbations reduce confidence without changing labels. AGSM amplifies angular deviation, leading to semantically meaningful misclassifications and stronger confidence drops.

1 Introduction

Deep neural networks have achieved remarkable success across a wide range of domains. However, they are also known to be highly sensitive to adversarial examples (Szegedy et al. 2014; Goodfellow, Shlens, and Szegedy 2015). These are specially crafted inputs (i.e., images) that include small, intentional perturbations designed to fool the model into making incorrect predictions. Despite their effect on the model, such perturbations are often imperceptible to human observers, making them appear visually or semantically identical to the original data. This vulnerability has motivated the development of numerous attack methods such as FGSM (Goodfellow, Shlens, and Szegedy 2015) and PGD (Madry et al. 2017), which generate input perturbations by leveraging gradients of the loss function. While these methods apply perturbations to the input, they fundamentally rely on the assumption that the model's representation space is Euclidean. As a result, they compute perturbation directions using gradients defined in a space with zero curvature (i.e., Euclidean geometry), which may not accurately reflect the underlying non-Euclidean structure of more complex models.

However, recent advances in representation learning have demonstrated that Euclidean space is ill-suited for representing structured data, such as trees, taxonomies, or graphs, which exhibit hierarchical relationships. For such data, hyperbolic space provides a more natural geometric setting, offering exponential representational capacity and hierarchy-preserving structure (Nickel and Kiela 2017; Ganea, Becigneul, and Hofmann 2018b). This has led to the development of *hyperbolic networks*, which learn and operate on latent representations embedded in hyperbolic manifolds. These models have demonstrated strong performance not only on structured tasks such as hierarchical classifica-

*Corresponding Authors.

tion (Chen et al. 2020), knowledge graph reasoning (Chami et al. 2020), and graph learning, but also on cross-modal retrieval tasks including text-to-image and image-to-text retrieval (Desai et al. 2023; Ramasinghe et al. 2024; Pal et al. 2025). In these tasks, hierarchical representations are particularly important, as they capture the coarse-to-fine semantic structure that naturally arises between visual and linguistic concepts. However, the study of adversarial robustness in hyperbolic networks remains largely unexplored (van Spengler, Zahálka, and Mettes 2025). Conventional adversarial attacks are geometry-agnostic, which do not account for the curvature or structure of non-Euclidean spaces. As a result, applying them directly to hyperbolic models can lead to ineffective perturbations that do not respect the underlying geometry, resulting in representation shifts that are semantically misaligned with the structure of the manifold.

To address this limitation, we propose a novel *adversarial attack that explicitly respects the geometric structure of hyperbolic space*. Our key insight is that, in hyperbolic geometry, the loss gradient, when computed in the tangent space of a representation point, can be decomposed into a radial (depth) component and an angular (semantic) component. The radial component alters the hierarchical level of the representation, while the angular component modulates it within the same hierarchical level, potentially aligning with semantically relevant directions in the manifold. This phenomenon is evident in Figure 1 and Table 1. Radial shifts of the representation have negligible impact on the final prediction, whereas angular shifts account for a substantial portion of the performance degradation induced by FGSM. Based on this decomposition, we introduce the *Angular Gradient Sign Method (AGSM)*, a novel adversarial attack specifically designed for hyperbolic networks. AGSM operates in the tangent space of hyperbolic representations and isolates only the angular component of the loss gradient, enabling perturbations that align with the semantic geometry of hyperbolic space.

This perturbation yields an adversarial example that exploits semantically sensitive directions in the hyperbolic representation space, thereby enhancing attack effectiveness without explicitly altering the hierarchical structure. We summarize our contributions as follows:

- We argue that conventional adversarial attacks may be suboptimal for hyperbolic networks, as they ignore the geometric properties of curved representation spaces and fail to exploit the structure inherent in hyperbolic embeddings.
- We propose AGSM (Angular Gradient Sign Method), a novel adversarial attack tailored to hyperbolic networks, which leverages radial–angular decomposition of gradients to isolate and perturb semantically sensitive directions in the hyperbolic representation space.
- We empirically demonstrate that AGSM outperforms conventional adversarial attacks on both hyperbolic classification and cross-modal retrieval tasks, including text-to-image and image-to-text retrieval, achieving higher fooling rates with perturbations that are more geometry-aware and effective in manipulating semantic content.

Method	Acc@1 (%)
Clean	53.44
FGSM	19.67
Radial shift	53.44
Angular shift	25.56
AGSM	**13.93**

Table 1: CIFAR-100 top-1 accuracy under five conditions: clean, radial shift, angular shift, FGSM, and AGSM on Poincaré ResNet-32. The radial shift has virtually no impact on accuracy, while the angular shift alone induces a substantial performance drop. FGSM combines both effects to further degrade accuracy, and AGSM elicits the strongest adversarial breakdown by selectively enhancing angular perturbations.

2 Related Works

Hyerpbolic Networks. Deep learning in hyperbolic space has demonstrated outstanding performance in encoding tree-structured hierarchies over the recent few years (Nickel and Kiela 2017; Ganea, Becigneul, and Hofmann 2018b; Peng et al. 2022; Mettes et al. 2024; He et al. 2025). Several studies have shown that language and image data also exhibit hierarchical structures, and have proposed hyperbolic space as an effective solution for representing such datasets (Khrulkov et al. 2020; Ermolov et al. 2022; Mandica et al. 2025; Sinha et al. 2025). To facilitate more explicit learning of hierarchical structure in hyperbolic space, several studies have imposed hierarchical constraints directly on the feature representations. As part of this research, nested geodesically convex cone (Ganea, Becigneul, and Hofmann 2018a) was employed to embed directed acyclic graphs. Wang et al. (2025) enforced hierarchical structure in image embeddings by partitioning each image into constituent parts and the overall scene for training. Similarly, Hi-Mapper (Kwon et al. 2024) preserved semantic relationships by decomposing visual scenes into individual elements and mapping them into hyperbolic space. This trend has likewise persisted in the training of vision–language models. MERU (Desai et al. 2023) jointly embeds visual and textual modalities in hyperbolic space, with the goal of encoding the language-image hierarchical relationships. Building on this, HyCoCLIP (Pal et al. 2025) leverages pre-trained grounding model (Li* et al. 2022; Zhang et al. 2022) to extract box-level image regions and corresponding text from full images and full text, more precisely structuring hierarchical relationships in hyperbolic space. While the majority of hyperbolic networks introduce embeddings only at the penultimate layer, architectures such as Poincaré ResNet (van Spengler, Berkhout, and Mettes 2023), Hyperbolic Neural Networks (HNN) (Ganea, Becigneul, and Hofmann 2018b; Shimizu, Mukuta, and Harada 2021), HyboNet (Chen et al. 2022) and L-CLIP (He, Yang, and Ying 2025) learn their representations entirely within hyperbolic space. While the aforementioned approaches have succeeded in embedding multiple modalities into hyperbolic space both effectively and interpretably, the robustness of

these models to adversarial attacks has not been deeply explored.

Gradient-based Adversarial Attacks. Gradient-based adversarial attacks form the cornerstone of white-box robustness evaluation in deep networks. The Fast Gradient Sign Method (FGSM, Goodfellow, Shlens, and Szegedy (2015)) computes a single-step perturbation by taking the sign of the loss gradient with respect to the input, scaled by a budget ε, to maximize the model's prediction error. Its iterative extension, Projected Gradient Descent (PGD, Madry et al. (2017)), applies multiple FGSM updates of smaller magnitude, projecting the perturbed sample back onto the ℓ_p-ball around the original input at each step, thereby yielding stronger attacks under the same perturbation constraint. Optimization-based approaches such as the Carlini & Wagner (C&W, Carlini and Wagner (2017)) attack further refine this paradigm by framing adversarial example generation as a constrained optimization problem. Moreover, a variety of gradient-based adversarial attack methods have been proposed, including Jacobian-based Saliency Maps (Papernot et al. 2016), the Basic Iterative Method (Kurakin, Goodfellow, and Bengio 2017), Gradient Aligned Adversarial Subspace (Tramèr et al. 2017), Momentum Iterative FGSM (Dong et al. 2018), Meta Gradient Adversarial Attack (Yuan et al. 2021), and Auto-Attack (Croce and Hein 2020, 2021). Although these methodologies have achieved remarkable success in Euclidean space, only a handful of works have attempted to transfer these attacks into hyperbolic space. A notable study (van Spengler, Zahálka, and Mettes 2025) applied FGM and PGD directly to synthetic hyperbolic embeddings, examining perturbation characteristics. However, these initial efforts have largely focused on synthetic hyperbolic embeddings and input space, and have not yet considered the distinct radial and angular components in output space. As a result, there is still an opportunity to explore how these components influence perturbation behavior and to develop attack strategies that more directly incorporate the geometric properties of hyperbolic space.

3 Preliminaries

We assess adversarial robustness using two representative models: Poincaré ResNet (van Spengler, Berkhout, and Mettes 2023) and HyCoCLIP (Pal et al. 2025).

Poincaré Ball Model. The Poincaré ball model provides a Riemannian manifold with constant negative curvature and is widely used to model hierarchical data structures in hyperbolic neural networks. Notably, it serves as the geometric foundation for architectures such as Poincaré ResNet (van Spengler, Berkhout, and Mettes 2023), where feature representations are embedded in hyperbolic space to capture hierarchical relations more effectively. Formally, the n-dimensional Poincaré ball of curvature $K = -c < 0$ is defined as the open ball:

$$\mathbb{B}_c^n \;=\; \big\{\, \mathbf{x} \in \mathbb{R}^n : \|\mathbf{x}\| < 1/c \big\}. \tag{1}$$

The hyperbolic distance between two points $u, v \in \mathbb{B}_c^n$ is given by:

$$d_{\mathbb{B}}(\mathbf{u},\mathbf{v}) = \operatorname{arcosh}\Big(1 + 2c\,\frac{\|\mathbf{u}-\mathbf{v}\|^2}{(1-c\|\mathbf{u}\|^2)\,(1-c\|\mathbf{v}\|^2)}\Big)$$
$$= \frac{2}{\sqrt{c}}\,\tanh^{-1}\big(\sqrt{c}\,\|(-\mathbf{u}) \oplus_c \mathbf{v}\|\big),$$

where the operation $\oplus_c$ denotes Möbius addition in curvature c. Möbius addition generalizes the addition of the Euclidean vector to the hyperbolic space and is defined for any $\mathbf{x}, \mathbf{y} \in \mathbb{B}_c^n$ as:

$$\mathbf{x} \oplus_c \mathbf{y} = \frac{\big(1 + 2c\,\langle \mathbf{x}, \mathbf{y}\rangle + c\|\mathbf{y}\|^2\big)\,\mathbf{x} \;+\; (1 - c\|\mathbf{x}\|^2)\,\mathbf{y}}{1 + 2c\,\langle \mathbf{x}, \mathbf{y}\rangle + c^2\|\mathbf{x}\|^2\,\|\mathbf{y}\|^2},$$

Lorentz Model. The Lorentz (hyperboloid) model offers an alternative realization of hyperbolic geometry, and is particularly useful for its numerical stability and closed-form expressions. This model underpins recent hyperbolic architectures such as HyCoCLIP (Pal et al. 2025), where feature representations are embedded in Lorentzian space to effectively capture hierarchical and semantic structures. The n-dimensional hyperbolic space of constant curvature $K = -c < 0$ is realized as the upper sheet of a two-sheeted hyperboloid in $\mathbb{R}^{n+1}$ equipped with the Lorentzian (Minkowski) inner product:

$$\langle \mathbf{x}, \mathbf{y}\rangle_{\mathbb{L}} = -\,x_0\,y_0 + \sum_{i=1}^{n} x_i\,y_i.$$

The Lorentz manifold is defined as:

$$\mathbb{L}_c^n = \Big\{\, \mathbf{x} = (x_0, x_1, \ldots, x_n) \in \mathbb{R}^{n+1} \;\Big|\; -\,x_0^2 \;+\; \sum_{i=1}^{n} x_i^2 \;=\; -\tfrac{1}{c}, \quad x_0 > 0 \Big\}.$$

The hyperbolic distance between two points $\mathbf{u}, \mathbf{v} \in \mathbb{L}_c^n$ is given by:

$$d_{\mathbb{L}}(\mathbf{u},\mathbf{v}) = \frac{1}{\sqrt{c}}\,\operatorname{arcosh}\big(-c\,\langle \mathbf{u}, \mathbf{v}\rangle_{\mathbb{L}}\big). \tag{2}$$

To perform perturbation in tangent space of Lorentz model, it is often necessary to move between the manifold and its tangent space via the exponential and logarithmic maps. For a point $\mathbf{x} \in \mathbb{L}_c^n$ and a tangent vector $\mathbf{v} \in T_{\mathbf{x}}\mathbb{L}_c^n$ with $\langle \mathbf{x}, \mathbf{v}\rangle_{\mathbb{L}} = 0$, define the Lorentz norm:

$$\|\mathbf{v}\|_{\mathbb{L}} \;=\; \sqrt{\langle \mathbf{v}, \mathbf{v}\rangle_{\mathbb{L}}}\,.$$

Then the exponential map is:

$$\exp_{\mathbf{x}}^c(\mathbf{v}) = \cosh\big(\sqrt{c}\,\|\mathbf{v}\|_{\mathbb{L}}\big)\,\mathbf{x} \;+\; \frac{\sinh\big(\sqrt{c}\,\|\mathbf{v}\|_{\mathbb{L}}\big)}{\sqrt{c}\,\|\mathbf{v}\|_{\mathbb{L}}}\,\mathbf{v}. \tag{3}$$

The corresponding logarithmic map $\log_{\mathbf{x}}^c : \mathbb{L}_c^n \to T_{\mathbf{x}}\mathbb{L}_c^n$ for a tangent point $\mathbf{y} \in \mathbb{L}_c^n$

$$\mathbf{v} = \log_{\mathbf{x}}(\mathbf{y}) = \frac{\cosh^{-1}(-c\,\langle \mathbf{x}, \mathbf{y}\rangle_{\mathbb{L}})}{\sqrt{(c\,\langle \mathbf{x}, \mathbf{y}\rangle_{\mathbb{L}})^2 - 1}}\;\operatorname{proj}_{\mathbf{x}}(\mathbf{y}), \tag{4}$$

where the projection onto the tangent space is defined as $\operatorname{proj}_{\mathbf{x}}(\mathbf{y}) = \mathbf{y} + c\,\mathbf{x}\,\langle \mathbf{x}, \mathbf{y}\rangle_{\mathbb{L}}$. These tools allow for differentiable computations and adversarial manipulations, making it a practical choice for hyperbolic deep learning. For further details, the reader is referred to Ratcliffe (2006).

4 Proposed Method

Existing adversarial attacks are typically developed under the assumption of Euclidean geometry, making them suboptimal for models whose representations lie in curved hyperbolic spaces. In particular, naively applying gradient-based perturbations in hyperbolic networks may result in less effective feature shifts, as they fail to exploit the underlying hierarchical structure encoded in hyperbolic embeddings.

To address this, we propose AGSM (Angular Gradient Sign Method), a novel adversarial attack that leverages the geometric structure of hyperbolic space. Rather than perturbing in arbitrary directions, AGSM isolates the angular component of the representation shift (i.e., the component orthogonal to the radial direction) and uses it to craft perturbations that drive semantically meaningful deviations without altering hierarchical level. This approach is grounded in the observation that, in hyperbolic geometry, radial displacement changes hierarchical depth, whereas angular displacement induces fine-grained semantic variation within the same level. By explicitly targeting angular shifts, AGSM produces more effective and geometry-aware adversarial examples for hyperbolic networks.

We now formalize this approach by describing how to decompose gradients in the tangent space and apply angular perturbations in a principled manner.

Geometric Decomposition of FGSM Perturbations. To investigate how adversarial perturbations affect representations in hyperbolic space, we propose a general framework that applies to both the Poincaré and Lorentz models, two common realizations of hyperbolic geometry in neural networks. We begin by applying the Fast Gradient Sign Method (FGSM) to generate perturbed input samples, and then analyze the resulting representation shift by decomposing it into radial and angular components in the tangent space of the corresponding manifold. This decomposition reflects the hierarchical and semantic structure encoded in hyperbolic embeddings and forms the basis of our geometry-aware attack method. Specifically, in hyperbolic space, the radial direction corresponds to changes in hierarchical depth (e.g., moving from general to specific classes), while the angular direction captures fine-grained semantic variations within the same level of the hierarchy. By isolating the angular component, we are able to generate perturbations that exploit semantically sensitive directions in the representation space, thereby enhancing attack effectiveness without unnecessarily altering the hierarchical structure.

We begin by applying FGSM, which perturbs the input along the sign of the loss gradient:

$$\tilde{\mathbf{x}}_{\text{adv}} = \mathbf{x} + \varepsilon \,\text{sign}\big(\nabla_{\mathbf{x}} \mathcal{L}(\theta, \mathbf{x}, y)\big).$$

Let the original and perturbed representations be defined as:

$$\mathbf{h} = f(\mathbf{x}) \quad \text{and} \quad \tilde{\mathbf{h}}_{\text{adv}} = f(\tilde{\mathbf{x}}_{\text{adv}}),$$

where $f(\cdot)$ denotes a hyperbolic feature encoder such as Poincaré ResNet (van Spengler, Berkhout, and Mettes 2023) or HyCoCLIP (Pal et al. 2025).

We first illustrate the decomposition procedure in the case where the feature space lies in the tangent space $\mathcal{T}_0\mathbb{B}_c^n$ of the Poincaré ball $\mathbb{B}_c^n$ (Equation 1). Since both $\mathbf{h}$ and $\mathbf{h}_{\text{adv}}$ reside in this Euclidean tangent space, the shift in representation can be computed via:

$$\Delta\mathbf{h} = \tilde{\mathbf{h}}_{\text{adv}} - \mathbf{h}.$$

We then decompose $\Delta\mathbf{h}$ into radial and angular components by first computing the unit radial direction:

$$\mathbf{u}_{\mathbf{h}} = \frac{\mathbf{h}}{\|\mathbf{h}\|_2},$$

and project $\Delta\mathbf{h}$ onto it to isolate the radial and angular components as:

$$\mathbf{v}_{\text{rad}} = \langle \Delta\mathbf{h},\, \mathbf{u}_{\mathbf{h}} \rangle\, \mathbf{u}_{\mathbf{h}} \quad \text{and} \quad \mathbf{v}_{\text{ang}} = \Delta\mathbf{h} - \mathbf{v}_{\text{rad}}.$$

This decomposition can be naturally extended to other hyperbolic models such as the Lorentz model. For networks whose output embeddings lie on the Lorentzian hyperboloid $\mathbb{L}_c^n$, we first project the hyperbolic points into the tangent space $\mathcal{T}_0\mathbb{L}^n$ using the logarithmic map (Equation 4), and then perform the same radial–angular decomposition in that tangent space. This allows our method to generalize across different realizations of hyperbolic geometry while maintaining geometric consistency.

Angular-based Adversarial Perturbation. Building on the decomposition above, we now describe how to construct adversarial examples that explicitly maximize angular shifts in hyperbolic representation space. Rather than perturbing the input indiscriminately in the direction of the overall gradient, as done in FGSM, we isolate the angular component $\mathbf{v}_{\text{ang}}$ of the representation shift and backpropagate this direction to the input space. This yields an input-space gradient that selectively promotes semantic variation within the same hierarchical level.

Concretely, we compute the gradient of the inner product between the current feature representation $\mathbf{h}$ and its angular shift component $\mathbf{v}_{\text{ang}}$ using the chain rule:

$$\nabla_{\mathbf{x}} \big\langle \mathbf{h},\, \mathbf{v}_{\text{ang}} \big\rangle \;=\; \left(\frac{\partial \mathbf{h}}{\partial \mathbf{x}}\right)^{\top} \mathbf{v}_{\text{ang}}.$$

This gradient points in a direction that maximally increases the angular displacement of the representation with only negligible impact on its radial depth, thereby concentrating the perturbation on semantically meaningful variations. We then apply a perturbation to the input in this direction, analogously to FGSM, using a normalized step:

$$\mathbf{x}_{\text{adv}} = \mathbf{x} + \varepsilon \,\text{sign}(\nabla_{\mathbf{x}} \langle \mathbf{h},\, \mathbf{v}_{\text{ang}} \rangle),$$

where ε controls the perturbation magnitude. This method, which we term the Angular Gradient Sign Method (AGSM), results in adversarial examples that exploit semantically sensitive directions within the hyperbolic manifold.

Compared to conventional attacks, AGSM yields more effective feature shifts while remaining aligned with the intrinsic geometry of hyperbolic space. To formalize our method, we present the overall procedure in Algorithm 1, which details how angular components are extracted and backpropagated to generate adversarial perturbations.

Algorithm 1: Angular Gradient Sign Method (AGSM)

Input: input $\mathbf{x}$, label y, perturbation budget ε, model f
Output: Adversarial example $\mathbf{x}_{\text{adv}}$

1: Compute Euclidean input gradient.
$\mathbf{g} \leftarrow \nabla_{\mathbf{x}} \mathcal{L}\big(f(\mathbf{x}), y\big)$
2: Generate tentative perturbed input.
$\tilde{\mathbf{x}}_{\text{adv}} \leftarrow \mathbf{x} + \varepsilon\, \text{sign}(\mathbf{g})$
3: Compute feature shift.
$\Delta\mathbf{h} \leftarrow f(\tilde{\mathbf{x}}_{\text{adv}}) - f(\mathbf{x})$
4: Get radial unit vector.
$\mathbf{u} \leftarrow f(\mathbf{x}) / \|f(\mathbf{x})\|_2$
5: Extract angular component (orthogonal to radial).
$\mathbf{v}_{\text{ang}} \leftarrow \Delta\mathbf{h} - \langle \Delta\mathbf{h}, \mathbf{u} \rangle\, \mathbf{u}$
6: Back-propagate angular shift via chain rule.
$\mathbf{d} \leftarrow \big(\partial\mathbf{h}/\partial\mathbf{x}\big)^{\top}\, \mathbf{v}_{\text{ang}}$
7: Apply angular perturbation to input.
$\mathbf{x}_{\text{adv}} \leftarrow \mathbf{x} + \varepsilon\, \text{sign}(\mathbf{d})$
8: **return** $\mathbf{x}_{\text{adv}}$

Extension to Projected Gradient Descent. Our Angular Gradient Sign Method (AGSM) can be naturally extended into a multi-step adversarial attack by adopting the framework of Projected Gradient Descent (PGD; Madry et al. (2017)). Instead of applying a single-step angular perturbation, we iteratively maximize the angular shift at each step, followed by a projection back onto the valid perturbation set. This extension enables stronger adversarial examples that remain aligned with the semantic geometry of hyperbolic space while respecting perturbation constraints such as an ℓ_∞ budget.

At each step, the method recomputes the angular direction in feature space, backpropagates it to the input via the chain rule, and applies a normalized update, followed by a projection back into the allowed perturbation set. This process is repeated over multiple iterations to refine the attack and enhance its effectiveness. The full procedure, Algorithm of Projected Angular Gradient Descent (PAGD), is summarized in supplementary material.

5 Experiments

Experimental Setup

Datasets. To evaluate our method on standard image classification benchmarks, we use CIFAR-10, CIFAR-100 (Krizhevsky and Hinton 2009) and Tiny ImageNet (Le and Yang 2015), covering a range of object categories and difficulty levels. For image-to-text (I2T) and text-to-image (T2I) retrieval experiments, we conduct evaluations on the MS COCO dataset (Lin et al. 2014) and the Flickr30K dataset (Plummer et al. 2015), which provide paired image–caption annotations suitable for cross-modal retrieval tasks.

Models. For the image classification experiments, we employ Poincaré ResNet-20 and Poincaré ResNet-32 architectures, both of which we trained using the exact hyperparameter settings and training protocol specified in the original **Poincaré ResNet** (van Spengler, Berkhout, and Mettes 2023). For cross-modal retrieval (I2T and T2I), we utilize the **HyCoCLIP** framework (Pal et al. 2025) with Vision Transformer backbones (ViT-S and ViT/16), leveraging the pretrained weights officially released by the HyCoCLIP authors.

Model	Dataset	ε	FGSM	AGSM	PGD	PAGD	Clean
PRN 20	C-10	ε_1	56.59	**47.63**	31.83	**22.42**	84.76
		ε_2	54.43	**44.53**	21.40	**14.25**	
		ε_3	49.63	**36.69**	8.86	**8.35**	
	C-100	ε_1	24.67	**20.02**	13.29	**11.05**	49.63
		ε_2	22.62	**17.66**	11.68	**9.28**	
		ε_3	17.78	**12.19**	9.43	**9.43**	
	TIN	ε_1	11.73	**8.90**	7.31	**5.93**	30.48
		ε_2	10.50	**7.78**	6.62	**5.49**	
		ε_3	7.44	**5.43**	5.66	**4.63**	
PRN 32	C-10	ε_1	60.68	**51.09**	28.96	**18.69**	86.21
		ε_2	59.10	**48.05**	18.43	**11.44**	
		ε_3	54.19	**41.56**	8.05	**7.77**	
	C-100	ε_1	26.36	**21.05**	12.71	**10.44**	53.44
		ε_2	24.61	**18.74**	11.18	**9.19**	
		ε_3	19.67	**13.93**	9.24	**7.86**	
	TIN	ε_1	11.90	**9.56**	7.03	**5.74**	30.46
		ε_2	10.71	**8.23**	6.61	**5.49**	
		ε_3	8.02	**5.57**	5.69	**5.00**	

Table 2: Robust accuracy (%) of Poincaré ResNet-20 and ResNet-32 on CIFAR-10, CIFAR-100, and Tiny ImageNet under ℓ_∞ attacks with $\varepsilon \in \{2.4/255, 3.2/255, 8.0/255\}$. For each attack type (FGSM and PGD), the lower accuracy (indicating a stronger attack) is highlighted in **bold**.

Results on Classification and Retrieval Tasks

Poincaré ResNet Robustness (Table 2). Across both ResNet-20 and ResNet-32 on CIFAR-10, Angle-only FGSM (AGSM) consistently inflicts an extra 9–11% drop in robust accuracy over standard FGSM, while PAGD compounds PGD's effect by roughly the same amount (around 9-10%). For instance, at $\varepsilon = 8.0/255$ on ResNet-32, AGSM lowers clean accuracy **12.63%** more than FGSM, and PAGD further lowers clean accuracy **10.27%** more than PGD at $\varepsilon = 2.4/255$. ResNet-20 shows a similar pattern, with AGSM undercutting FGSM by about **13%** and PAGD lowering PGD by over **9%** in its strongest case.

On CIFAR-100 and Tiny ImageNet, AGSM still outperforms its one-step counterpart by about 5–6%, and PAGD delivers an additional 1–2% degradation beyond PGD. These results confirm that angular-maximizing perturbations more effectively exploit hyperbolic geometry than conventional gradient-based methods.

HyCoCLIP Retrieval Robustness (Table 3, 4). Across both COCO and Flickr30K, and for both ViT-S/16 and ViT-B/16 backbones, our Angle-only FGSM (AGSM) consistently deepens the drop in recall by roughly 2–5% compared

Model	Dataset	ε	T2I R@5					T2I R@10				
			FGSM	AGSM	PGD	PAGD	Clean	FGSM	AGSM	PGD	PAGD	Clean
ViT-S/16	COCO	3.2/255	11.20	**8.20**	3.00	**2.60**	55.10	16.50	**12.70**	5.10	**4.40**	66.60
		8.0/255	7.60	**4.80**	1.50	**1.00**		11.60	**7.60**	2.70	**1.90**	
	Flickr30K	3.2/255	19.70	**15.40**	8.40	**7.30**	81.50	27.30	**22.50**	13.50	**12.10**	88.20
		8.0/255	12.90	**9.10**	4.40	**4.00**		19.10	**14.30**	7.50	**7.00**	
ViT-B/16	COCO	3.2/255	15.90	**12.60**	4.50	**4.00**	58.40	22.70	**18.80**	7.30	**6.40**	69.30
		8.0/255	10.80	**7.60**	2.20	**1.80**		16.20	**11.90**	3.80	**3.10**	
	Flickr30K	3.2/255	27.20	**24.90**	10.90	**9.80**	84.90	36.30	**33.20**	17.70	**16.00**	90.30
		8.0/255	18.60	**14.10**	5.30	**5.00**		25.10	**21.00**	9.50	**8.40**	

Table 3: Performance of the Text-to-Image (T2I) task at Recall@5 and Recall@10 under adversarial attacks (FGSM, AGSM, PGD, PAGD) on COCO and Flickr30K using ViT-S/16 and ViT-B/16. For each attack type (FGSM and PGD), the lower accuracy (indicating a stronger attack) is highlighted in **bold**.

Model	Dataset	ε	I2T R@5					I2T R@10				
			FGSM	AGSM	PGD	PAGD	Clean	FGSM	AGSM	PGD	PAGD	Clean
ViT-S/16	COCO	3.2/255	13.20	**9.00**	4.10	**3.10**	69.50	18.50	**13.60**	6.10	**4.70**	79.50
		8.0/255	7.30	**4.20**	1.90	**1.40**		11.10	**6.80**	3.10	**2.30**	
	Flickr30K	3.2/255	20.10	**15.30**	8.10	**6.90**	89.10	27.10	**21.70**	13.30	**9.90**	93.90
		8.0/255	11.60	**8.50**	4.20	**3.90**		16.70	**12.20**	7.70	**7.20**	
ViT-B/16	COCO	3.2/255	20.10	**15.00**	5.40	**4.60**	72.00	26.90	**21.60**	8.30	**7.50**	82.00
		8.0/255	11.10	**7.40**	2.40	**1.90**		15.90	**10.90**	3.80	**3.20**	
	Flickr30K	3.2/255	29.50	**26.60**	11.70	**10.70**	92.60	38.40	**35.00**	16.90	**16.20**	95.40
		8.0/255	18.10	**13.70**	6.10	**4.40**		24.60	**19.70**	9.00	**7.30**	

Table 4: Performance of the Image-to-Text (I2T) task at Recall@5 and Recall@10 under adversarial attacks (FGSM, AGSM, PGD, PAGD) on COCO and Flickr30K using ViT-S/16 and ViT-B/16.

to standard FGSM, while its multi-step counterpart PAGD yields an additional 0.5–1% reduction over PGD.

For example, on COCO with ViT-S/16 at $\varepsilon = 3.2/255$, AGSM adds a further 3.0% of degradation in T2I R@5 beyond FGSM's already severe drop, and PAGD compounds PGD's effect by another 0.8%. With the larger ViT-B/16, AGSM's advantage reaches up to about a 5% extra drop in I2T R@5, and PAGD still provides nearly a 1% degradation over PGD. These largest observed gains underline that angular-maximising perturbations more effectively disrupt cross-modal retrieval in hyperbolic embedding spaces.

Summary. Across both Poincaré ResNet and HyCoCLIP backbones, standard one-step (FGSM) and multi-step (PGD) attacks already degrade performance, but their angular maximizing counterparts, AGSM and PAGD, consistently inflict an additional drop in accuracy or recall. This highlights the critical role of angular movement in breaking hierarchical representations.

Analysis on Perturbed Sample and Representation

Distance Between Hyperbolic Embeddings. Table 5 reports the average hyperbolic distance in the Lorentz model (Equation 2) between original and perturbed feature vectors, after mapping them to the hyperbolic manifold via the exponential map (Equation 3). Features were produced by HyCoCLIP under FGSM and AGSM at $\varepsilon \in \{3.2/255, 8.0/255\}$. On both dataset, AGSM increases the mean geodesic at both ε values, indicating that angular-maximizing updates push representations farther along hyperbolic geodesics than standard gradient-sign perturbations. Figure 2 qualitatively compares retrieval outputs under different perturbations. The radial shift preserves the correct caption, FGSM and the standard angular shift yield semantically incorrect sentences, and AGSM produces the most semantically misaligned caption.

Confidence Drop. Table 6 shows that AGSM consistently produces larger confidence reductions than FGSM across both CIFAR-10 and CIFAR-100, and at both moderate and high perturbation levels. In particular, the gap between FGSM and AGSM widens as ε increases, indicating that emphasizing the angular shift becomes even more destructive under stronger attacks. Qualitatively, this demonstrates that angular-focused perturbations more effectively undermine the model's predictive certainty than conventional gradient-sign methods.

Dataset	ε	Clean vs FGSM	Clean vs AGSM
COCO	3.2/255	0.3058	**0.3639**
	8.0/255	0.3883	**0.4457**
Flickr30K	3.2/255	0.3119	**0.3675**
	8.0/255	0.3875	**0.4400**

Table 5: Hyperbolic feature distances (in $\mathbb{L}_c^n$) between the original and perturbed samples for HyCoCLIP (ViT/b/16) under FGSM and AGSM.

Dataset	ε	FSGM	AGSM
CIFAR-10	3.2/255	0.3870	**0.4860**
	8.0/255	0.4364	**0.5597**
CIFAR-100	3.2/255	0.4242	**0.4628**
	8.0/255	0.4566	**0.4935**

Table 6: Under FGSM and AGSM, drop in MSP of the model on initially predicted label.

Dataset	ε	FGSM	AGSM	PGD	PAGD	Clean
C-10	3.2/255	58.20	**49.38**	32.79	**25.47**	84.76
	8.0/255	51.74	**40.16**	10.06	**8.62**	
C-100	3.2/255	24.97	**20.96**	13.34	**10.84**	49.63
	8.0/255	19.43	**14.31**	9.46	**7.74**	

Table 7: Top-1 accuracy (%) of Poincaré ResNet-20 under ℓ_2-constrained Fast Gradient Sign Method (FGSM) and Projected Gradient Descent (PGD), together with their angular-only variants, on CIFAR-10 and CIFAR-100. "Original" denotes clean accuracy.

Dataset	Attack	Training Data		
		Clean	FGSM-Aug	AGSM-Aug
CIFAR-10	Clean	**84.76**	81.65	82.31
	FGSM	8.67	**56.58**	55.08
	AGSM	8.30	**52.46**	51.07
CIFAR-100	Clean	**49.63**	47.81	46.96
	FGSM	9.61	26.45	**27.99**
	AGSM	7.91	23.61	**25.66**

Table 8: Adversarial Training Results. Top-1 accuracy (%) of Poincaré ResNet-20 trained on clean, FGSM-augmented, and AGSM-augmented datasets, evaluated under clean, FGSM, and AGSM attacks on CIFAR-10 and CIFAR-100.

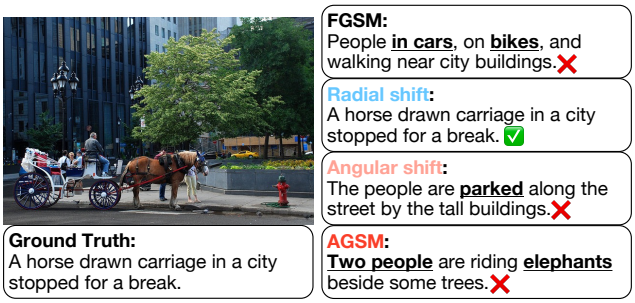

Figure 2: Qualitative comparison of Image-to-Text retrieval under FGSM, radial shift, angular shift, and AGSM. While the radial shift preserves the correct caption, FGSM and the standard angular shift generate semantically incorrect outputs, and AGSM yields the most misaligned caption.

Summary. Table 5 and Table 6 together underscore the central contribution of our approach: by isolating and maximizing the angular component of the gradient, AGSM not only drives feature vectors to traverse significantly farther along hyperbolic geodesics but also precipitates a more severe collapse of the model's predictive confidence compared to conventional gradient-sign attacks. This dual effect confirms that angular-maximizing perturbations provide a principled mechanism to undermine both representational integrity and output certainty in hierarchical models.

Ablation Study

Under an ℓ_2-constraint (Table 7), results demonstrate that isolating and maximizing the angular component delivers powerful attacks that are largely agnostic to the choice of norm. Whether measured in ℓ_∞ or ℓ_2, AGSM consistently exploits angular vulnerabilities in hyperbolic embeddings more effectively than FSGM.

6 Conclusion and Limitation

In this work, we introduced the Angular Gradient Sign Method (AGSM) and its multi-step extension PAGD to craft adversarial perturbations that explicitly maximize angular shifts in hyperbolic embedding spaces. Through extensive experiments on Poincaré ResNet for image classification and HyCoCLIP for cross-modal retrieval, we demonstrated that angular-focused attacks consistently outperform standard FGSM and PGD baselines. Our ablation studies further revealed that the angular component alone drives the majority of the adversarial effect, and that these attacks remain effective under both ℓ_∞ and ℓ_2 norm constraints.

However, training with AGSM-perturbed examples yields only modest gains in robustness and incurs a trade-off in clean accuracy relative to FGSM augmentation. On CIFAR-100, AGSM augmentation improves robustness specifically against angular perturbations but at the cost of a larger drop in standard accuracy. These results suggest that naively incorporating adversarial examples perturbed by AGSM does not uniformly strengthen hyperbolic models and may incur dataset-dependent trade-offs. Taken together, our findings underscore the pivotal role of angular misalignment in hyperbolic vulnerability and point to the need for geometry-aware defense strategies that explicitly accommodate the curved, hierarchical structure of hyperbolic embeddings.

Acknowledgements

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (No. RS-2024-00345809, "Research on AI Robustness Against Distribution Shift in Real-World Scenarios"), the Institute of Information & Communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT) (No. RS-2025-02263754, "Human-Centric Embodied AI Agents with Autonomous Decision-Making"),and the Korea Health Technology R&D Project through the Korea Health Industry Development Institute (KHIDI), funded by the Ministry of Health & Welfare, Republic of Korea (No. RS-2025-02307233).

References

Carlini, N.; and Wagner, D. 2017. Towards Evaluating the Robustness of Neural Networks . In *2017 IEEE Symposium on Security and Privacy (SP)*, 39–57. Los Alamitos, CA, USA: IEEE Computer Society.

Chami, I.; Wolf, A.; Juan, D.-C.; Sala, F.; Ravi, S.; and Ré, C. 2020. Low-Dimensional Hyperbolic Knowledge Graph Embeddings. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, 6901–6914.

Chen, B.; Huang, X.; Cai, Z.; and Jing, L. 2020. Hyperbolic Interaction Model for Hierarchical Multi-Label Classification. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34: 7496–7503.

Chen, W.; Han, X.; Lin, Y.; Zhao, H.; Liu, Z.; Li, P.; Sun, M.; and Zhou, J. 2022. Fully Hyperbolic Neural Networks. In Muresan, S.; Nakov, P.; and Villavicencio, A., eds., *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 5672–5686. Dublin, Ireland: Association for Computational Linguistics.

Croce, F.; and Hein, M. 2020. Reliable evaluation of adversarial robustness with an ensemble of diverse parameter-free attacks. In *ICML*.

Croce, F.; and Hein, M. 2021. Mind the box: l_1-APGD for sparse adversarial attacks on image classifiers. In *ICML*.

Desai, K.; Nickel, M.; Rajpurohit, T.; Johnson, J.; and Vedantam, R. 2023. Hyperbolic Image-Text Representations. In *Proceedings of the International Conference on Machine Learning*.

Dong, Y.; Liao, F.; Pang, T.; Su, H.; Zhu, J.; Hu, X.; and Li, J. 2018. Boosting Adversarial Attacks with Momentum . In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 9185–9193. Los Alamitos, CA, USA: IEEE Computer Society.

Ermolov, A.; Mirvakhabova, L.; Khrulkov, V.; Sebe, N.; and Oseledets, I. 2022. Hyperbolic Vision Transformers: Combining Improvements in Metric Learning. arXiv:2203.10833.

Ganea, O.; Becigneul, G.; and Hofmann, T. 2018a. Hyperbolic Entailment Cones for Learning Hierarchical Embeddings. In Dy, J.; and Krause, A., eds., *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, 1646–1655. PMLR.

Ganea, O.; Becigneul, G.; and Hofmann, T. 2018b. Hyperbolic Neural Networks. In Bengio, S.; Wallach, H.; Larochelle, H.; Grauman, K.; Cesa-Bianchi, N.; and Garnett, R., eds., *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc.

Goodfellow, I. J.; Shlens, J.; and Szegedy, C. 2015. Explaining and Harnessing Adversarial Examples. In Bengio, Y.; and LeCun, Y., eds., *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*.

He, N.; Madhu, H.; Bui, N.; Yang, M.; and Ying, R. 2025. Hyperbolic Deep Learning for Foundation Models: A Survey. In *Proceedings of the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. ACM.

He, N.; Yang, M.; and Ying, R. 2025. HyperCore: The Core Framework for Building Hyperbolic Foundation Models with Comprehensive Modules. arXiv:2504.08912.

Khrulkov, V.; Mirvakhabova, L.; Ustinova, E.; Oseledets, I.; and Lempitsky, V. 2020. Hyperbolic Image Embeddings. In *The IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*.

Krizhevsky, A.; and Hinton, G. 2009. Learning multiple layers of features from tiny images. Technical Report 0, University of Toronto, Toronto, Ontario.

Kurakin, A.; Goodfellow, I. J.; and Bengio, S. 2017. Adversarial examples in the physical world. In *5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Workshop Track Proceedings*. OpenReview.net.

Kwon, H.; Jang, J.; Kim, J.; Kim, K.; and Sohn, K. 2024. Improving Visual Recognition with Hyperbolical Visual Hierarchy Mapping. *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 17364–17374. Publisher Copyright: © 2024 IEEE.; 2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2024 ; Conference date: 16-06-2024 Through 22-06-2024.

Le, Y.; and Yang, X. S. 2015. Tiny ImageNet Visual Recognition Challenge.

Li*, L. H.; Zhang*, P.; Zhang*, H.; Yang, J.; Li, C.; Zhong, Y.; Wang, L.; Yuan, L.; Zhang, L.; Hwang, J.-N.; Chang, K.-W.; and Gao, J. 2022. Grounded Language-Image Pre-training. In *CVPR*.

Lin, T.-Y.; Maire, M.; Belongie, S. J.; Bourdev, L. D.; Girshick, R. B.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; and Zitnick, C. L. 2014. Microsoft COCO: Common Objects in Context. *CoRR*, abs/1405.0312.

Madry, A.; Makelov, A.; Schmidt, L.; Tsipras, D.; and Vladu, A. 2017. Towards Deep Learning Models Resistant to Adversarial Attacks. *ArXiv*, abs/1706.06083.

Mandica, P.; Franco, L.; Kallidromitis, K.; Petryk, S.; and Galasso, F. 2025. Hyperbolic Learning with Multimodal Large Language Models. In *Computer Vision – ECCV 2024 Workshops: Milan, Italy, September 29–October 4,*

2024, Proceedings, Part XVII, 382–398. Berlin, Heidelberg: Springer-Verlag. ISBN 978-3-031-91584-0.

Mettes, P.; Ghadimi Atigh, M.; Keller-Ressel, M.; Gu, J.; and Yeung, S. 2024. Hyperbolic Deep Learning in Computer Vision: A Survey. *Int. J. Comput. Vision*, 132(9): 3484–3508.

Nickel, M.; and Kiela, D. 2017. Poincaré Embeddings for Learning Hierarchical Representations. In Guyon, I.; Luxburg, U. V.; Bengio, S.; Wallach, H.; Fergus, R.; Vishwanathan, S.; and Garnett, R., eds., *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc.

Pal, A.; van Spengler, M.; di Melendugno, G. M. D.; Flaborea, A.; Galasso, F.; and Mettes, P. 2025. Compositional Entailment Learning for Hyperbolic Vision-Language Models. In *The Thirteenth International Conference on Learning Representations, ICLR 2025, Singapore, April 24-28, 2025*.

Papernot, N.; McDaniel, P. D.; Jha, S.; Fredrikson, M.; Celik, Z. B.; and Swami, A. 2016. The Limitations of Deep Learning in Adversarial Settings. In *IEEE European Symposium on Security and Privacy, EuroS&P 2016, Saarbrücken, Germany, March 21-24, 2016*, 372–387. IEEE.

Peng, W.; Varanka, T.; Mostafa, A.; Shi, H.; and Zhao, G. 2022. Hyperbolic Deep Neural Networks: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(12): 10023–10044.

Plummer, B. A.; Wang, L.; Cervantes, C. M.; Caicedo, J. C.; Hockenmaier, J.; and Lazebnik, S. 2015. Flickr30k Entities: Collecting Region-to-Phrase Correspondences for Richer Image-to-Sentence Models. In *2015 IEEE International Conference on Computer Vision (ICCV)*, 2641–2649.

Ramasinghe, S.; Shevchenko, V.; Avraham, G.; and Thalaiyasingam, A. 2024. Accept the modality gap: An exploration in the hyperbolic space. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 27263–27272.

Ratcliffe, J. G. 2006. *Foundations of hyperbolic manifolds*. New York: Springer. ISBN 978-0-387-47322-2.

Shimizu, R.; Mukuta, Y.; and Harada, T. 2021. Hyperbolic Neural Networks++. arXiv:2006.08210.

Sinha, A.; Zeng, S.; Yamada, M.; and Zhao, H. 2025. Learning structured representations with hyperbolic embeddings. In *Proceedings of the 38th International Conference on Neural Information Processing Systems*, NIPS '24. Red Hook, NY, USA: Curran Associates Inc. ISBN 9798331314385.

Szegedy, C.; Zaremba, W.; Sutskever, I.; Bruna, J.; Erhan, D.; Goodfellow, I. J.; and Fergus, R. 2014. Intriguing properties of neural networks. In Bengio, Y.; and LeCun, Y., eds., *2nd International Conference on Learning Representations, ICLR 2014, Banff, AB, Canada, April 14-16, 2014, Conference Track Proceedings*.

Tramèr, F.; Papernot, N.; Goodfellow, I. J.; Boneh, D.; and Mcdaniel, P. 2017. The Space of Transferable Adversarial Examples. *ArXiv*, abs/1704.03453.

van Spengler, M.; Berkhout, E.; and Mettes, P. 2023. Poincare ResNet. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 5419–5428.

van Spengler, M.; Zahálka, J.; and Mettes, P. 2025. Adversarial Attacks on Hyperbolic Networks. In *Computer Vision – ECCV 2024 Workshops: Milan, Italy, September 29–October 4, 2024, Proceedings, Part XVII*, 363–381. Berlin, Heidelberg: Springer-Verlag. ISBN 978-3-031-91584-0.

Wang, Z.; Ramasinghe, S.; Xu, C.; Monteil, J.; Bazzani, L.; and Ajanthan, T. 2025. Learning Visual Hierarchies in Hyperbolic Space for Image Retrieval. arXiv:2411.17490.

Yuan, Z.; Zhang, J.; Jia, Y.; Tan, C.; Xue, T.; and Shan, S. 2021. Meta Gradient Adversarial Attack. *arXiv preprint arXiv:2108.04204*.

Zhang, H.; Zhang, P.; Hu, X.; Chen, Y.-C.; Li, L. H.; Dai, X.; Wang, L.; Yuan, L.; Hwang, J.-N.; and Gao, J. 2022. GLIPv2: Unifying Localization and Vision-Language Understanding. *arXiv preprint arXiv:2206.05836*.