BEYOND FLAT SPACES: HYPERBOLIC MANIFOLDS FOR UNEXPLODED ORDNANCE IDENTIFICATION

Anonymous authors
Paper under double-blind review

ABSTRACT

Unexploded Ordnance (UXO) identification is a safety-critical visual recognition problem characterized by extreme data scarcity and substantial visual similarity across ordnance categories. We analyse UXO embeddings using complementary geometric diagnostics and find evidence of low effective dimensionality and stronger tree-like structure relative to a calibrated null model. Motivated by this evidence, we introduce LORENTZTIM, a transductive information-maximization method formulated directly on the Lorentz hyperboloid. LORENTZTIM represents samples and class prototypes in hyperbolic space and performs transductive prototype adaptation using geodesic distances and Riemannian optimization. On the CTX-UXO benchmark, LORENTZTIM consistently improves F1 over the evaluated Euclidean and hyperbolic baselines, with consistent gains across all shots regime. The improvements persist across different feature extractors and extend to CUB-200-2011, FGVC-Aircraft, and tieredImageNet, supporting the broader applicability of the proposed approach beyond UXO identification.

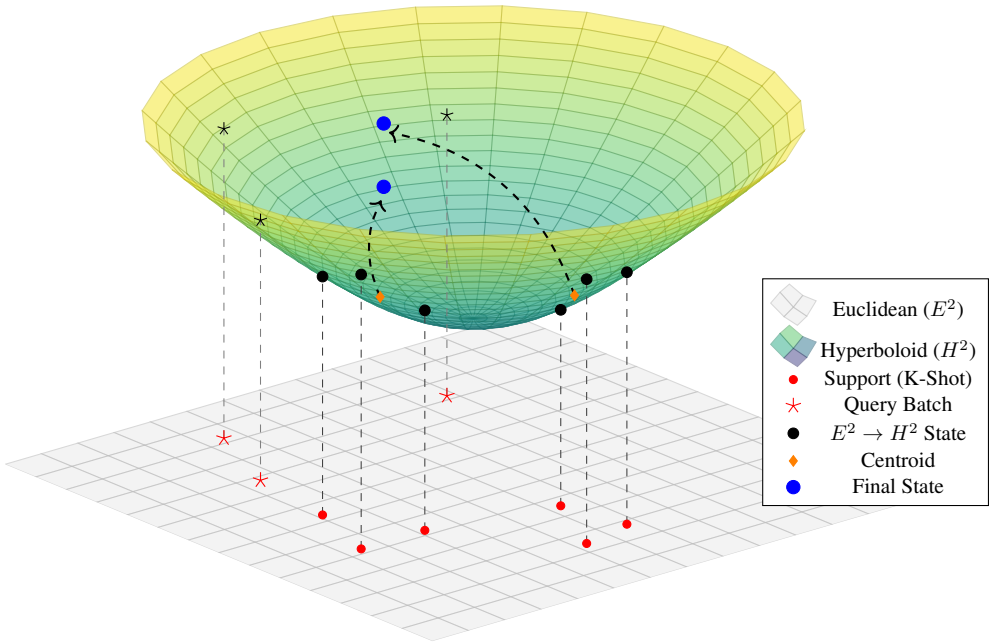


Figure 1: **LORENTZTIM: geometric workflow.** Given S labelled support samples per class and an unlabelled query batch Q, features are mapped from Euclidean space to the Lorentz hyperboloid (H^2). Class prototypes are initialised as Lorentzian centroids $\mathbf{w}_a^{(0)}$ of the support samples and refined via Riemannian optimisation of a transductive objective combining support cross-entropy with mutual-information maximisation over the query batch. The resulting prototypes $\mathbf{w}_a^{(T)}$ adapt to the query distribution, accounting for domain variations such as corrosion and lighting. Final predictions are obtained by geodesic nearest-prototype assignment.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

1 INTRODUCTION

According to United Nations (UNMAS, 2026), UneXploded Ordnance (UXO) or UneXploded Bombs (UXB) represents ammunitions, such as bombs and projectiles, that did not detonate and present major hazards. Statistics from the International Committee of the Red Cross (ICRC, 2025) show a rise in UXO discoveries in recent years, raising concern among populations, as millions of people continue to live or work in regions heavily affected by unexploded ordnance. Fine-grained visual UXO identification remains challenging under limited labeled data, posing serious risks to affected communities. Accurate identification of these energetic material cases is essential for applying appropriate Explosive Ordnance Disposal (EOD) procedures and safeguarding human lives.

Recent advances in deep representation learning have greatly improved general object recognition (Qiao et al., 2025), but their application to UXO identification exposes key limitations. Real-world UXO data, such as CTX-UXO (Craioveanu & Stamatescu, 2024; Craioveanu et al., 2026), are scarce and strongly long-tailed. Simultaneously, ordnance shows extreme visual variability from corrosion, occlusion, and partial burial, yet shares fine-grained morphological traits dictated by ballistic physics (e.g., standardised fuzes and fins). This combination of high intra-class variance and strong inter-class similarity is difficult (Khrulkov et al., 2020; Qiao et al., 2025) to model with Euclidean feature extractors, which distort hierarchical structure and face capacity limits in low-data regimes.

We introduce LORENTZTIM, a novel algorithm that uses the Lorentz model of hyperbolic space. With a custom test-time Transductive Information Maximisation (TIM) strategy, the method exploits the topology of the unlabelled query set at inference. The main contributions of this work are:

- **Empirical geometric motivation:** Rather than assuming a priori hierarchies, we motivate the use of hyperbolic geometry by directly studying the tree-like structure and intrinsic dimensionality of UXO representations;
- **Transductive adaptation of the Lorentz hyperboloid model:** For this humanitarian task, we introduce a transductive algorithm grounded in the Lorentz hyperboloid model;
- **Extensive evaluation:** We benchmark LORENTZTIM on a recognized UXO benchmark against both Euclidean and hyperbolic baselines. We conducted additional experiments on different datasets/feature extractors (see Appendix G).

The paper is organised as follows. Section 2 reviews UXO recognition and few-shot learning. Section 3 introduces notation, while Section 4 describes the hyperbolic properties and the algorithm. Section 5 reports quantitative and qualitative results. Section 6 concludes the paper. The Appendix provides pseudocode, generalisability analysis, code, hyperparameters, human-in-the-loop recommendations, additional related works, memory requirements and future directions.

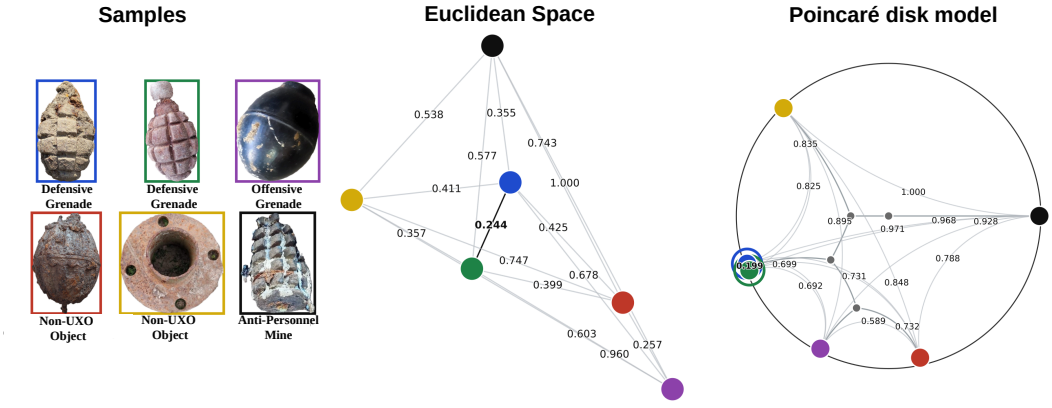


Figure 2: **Resolving UXO class ambiguity through hyperbolic embeddings. Simplified view.** Connections show pairwise distances, scaled by the maximum distance in each space to lie within $[0, 1]$. In Euclidean space, common embedding methods can blur category boundaries, placing visually similar but categorically distinct objects close together. We use the Poincaré disk for this preliminary view as it naturally represents the tree-like structure of the data.

2 Related Work

Table 1: **A comparative study of embedding-based methods for unexploded ordnance recognition**. We investigated how these methods address the fundamental challenges of UXO. Traditional Euclidean approaches struggle to capture the fine-grained differences and hierarchical visual structures of real-world UXOs. In our results in Section 5, we report and compare methods that are applicable to UXO. For completeness and empirical validation, we extend our evaluation to baselines that do not fully satisfy all requirements for UXO.

Few-Shot Method	Intra-Heterogeneity	Cross-Domain	Small Support	Meta Training-Free
SimpleShot (CL2N) (Wang et al., 2019)	✗	✗	✓	✓
Proto Networks (Snell et al., 2017)	✗	✗	✓	✗
Matching Networks (Vinyals et al., 2016)	✗	✗	✓	✗
Relation Networks (Sung et al., 2018)	✗	✗	✓	✗
CAN (Hou et al., 2019)	✓	✗	✓	✗
DSN (Simon et al., 2020)	✓	✗	✓	✗
LaplacianShot (Ziko et al., 2020)	✓	✓	✓	✓
PT-MAP (Hu et al., 2021)	✓	✓	✓	✓
TIM (Boudiaf et al., 2020)	✓	✓	✓	✓
Hyp-ProtoNet (Khrulkov et al., 2020)	✓	✗	✓	✗
Busemann Learning (Ghadimi Atigh et al., 2021)	✓	✗	✓	✓
Taylor-Series (Choudhary & Reddy, 2022)	✓	✗	✓	✗

The identification of UXO based on machine learning has been difficult and has focused mainly on ground-penetrating radar (GPR) (Núñez-Nieto et al., 2014; Machado Brito-da Costa et al., 2022) or magnetometry (Bray & Link, 2014; Wen et al., 2024). These approaches are limited to initial detection (UXO vs. non-UXO) and do not address the assignment of a specific UXO type. The identification problem arises once an item has been excavated or localised: a metallic object with UXO-like shape and material is found, but its type is unknown. Fine-grained analysis in the RGB domain was long infeasible due to a lack of annotated UXO datasets. The CTX-UXO dataset (Craioveanu et al., 2026) partially fills this gap by providing imagery that reflects real operational environments, with high class imbalance and substantial visual ambiguity between UXO classes. These UXO characteristics are challenging because conventional deep learning requires large, balanced, labelled datasets, which are impractical to obtain quickly for unexploded ordnance. Few-shot learning offers a viable alternative by enabling models to generalise from only a small number of labelled examples per class. In image-related tasks, few-shot approaches are commonly divided into two main categories: *embedding-based* and *optimization-centric* (Ali et al., 2021; Ren et al., 2024). Embedding-based methods learn discriminative feature representations and perform classification by applying similarity measures within the embedding space. In contrast, optimisation-centric methods prioritise identifying suitable initial parameters and quickly adapting them to the target domain. Traditionally, these approaches have relied on spherical or Euclidean embedding spaces. Few-shot learning methods can also be categorised by their inference paradigm: inductive or transductive (Ali et al., 2021; Ren et al., 2024). In the inductive setting, the model learns a general function from the labelled support set ($\mathcal{S}$) and classifies each query independently, without using the test data distribution. In contrast, transductive few-shot learning uses the entire unlabelled query set ($\mathcal{Q}_{\text{unlabeled}}$) for inference, exploiting the statistical and latent structure of the current test batch. Predictions are iteratively refined through domain adaptation. This often improves task accuracy by adapting the decision boundaries to the query distribution. Euclidean space, with zero curvature and only polynomial volume growth, poorly captures rich hierarchies without heavy distortion. Constant negative curvature spaces may be better suited for real-world datasets (Khrulkov et al., 2020). Early methods (Nickel & Kiela, 2017; Khrulkov et al., 2020) used the Poincaré ball, whose exponentially growing volume and boundary metric blow-up naturally match tree-like and hierarchical graphs. However, optimisation there suffers from numerical instabilities and floating-point errors near the boundary due to the ill-conditioned Riemannian metric. To address this, later work (Nickel & Kiela, 2018; Bdeir et al., 2024; Hasan et al., 2026) adopted the Lorentz (hyperboloid) model, realising the hyperbolic space as the upper sheet of a two-sheet hyperboloid in Minkowski space. This unbounded representation avoids boundary singularities, supports stable Riemannian gradient optimisation, and provides closed-form geodesic distances using the Minkowski inner product while preserving the hierarchical advantages of hyperbolic geometry.

In Table 1, we compare the main few-shot techniques from the literature which are suitable for UXO identification. More methodological information and more extensive comparisons can also be found in the Appendix B. For future directions on taxonomy training, see the Appendix H.

3 PRELIMINARIES

We formalise UXO identification as a *C-way S-shot* few-shot classification problem. Let $\mathcal{X} \subset \mathbb{R}^{H\times W\times 3}$ denote the input space and $\mathcal{Y} = \{c_1, \ldots, c_C\}$ the set of labels comprising the UXO subtypes. A feature extractor $f_\theta : \mathcal{X} \to \mathbb{R}^d$ maps each image to a vector of embedding dimensional d. We use DINO v3 (Siméoni et al., 2025) as feature extractor, d=768. Also, CLIP (Radford et al., 2021) with d=768 is considered in the Appendix. The feature extractors are described in the Appendix B. Each episode samples a labelled support set $\mathcal{S} = \{(x_i, y_i)\}_{i=1}^{C\cdot S}$, containing S examples per class, and an unlabelled query set $\mathcal{Q}$ with Q samples per class. To avoid notational confusion, we define $\mathcal{U} = \mathcal{Q}$ (query) and perform evaluation on fixed-seed episodes M=1,000 with $S \in \{1, 3, 5, 10\}$. We use the Lorentz model of the n-dimensional hyperbolic space with curvature $K < 0$ (Nickel & Kiela, 2018), given by:

$$\mathbb{L}_K^n = \{\mathbf{x} \in \mathbb{R}^{n+1} : \langle \mathbf{x}, \mathbf{x}\rangle_L = 1/K,\ x_0 > 0\}, \tag{1}$$

The following denotes the Minkowski inner product:

$$\langle \mathbf{x}, \mathbf{y}\rangle_L = -x_0 y_0 + \sum_{j=1}^{n} x_j y_j \tag{2}$$

The tangent space at $\mathbf{x}$ is:

$$T_{\mathbf{x}}\mathbb{L}_K^n = \{\mathbf{u} \in \mathbb{R}^{n+1} : \langle \mathbf{x}, \mathbf{u}\rangle_L = 0\}, \tag{3}$$

We use the induced norm $\|\mathbf{u}\|_L = \sqrt{\langle \mathbf{u}, \mathbf{u}\rangle_L}$ and set the manifold origin to $\mathbf{o} = (1/\sqrt{-K}, 0, \ldots, 0)$. The manifold dimension is $n = d_{\text{eff}}$, the effective intrinsic dimension (Section 4.1). To move vectors between tangent spaces, we use parallel transport $\mathrm{PT}_{\mathbf{x}\to\mathbf{y}}^K$ along the geodesic from $\mathbf{x}$ to $\mathbf{y}$ (see equation 25).

For few-shot identification, class prototypes $\mathbf{W} = \{\mathbf{w}_a\}_{a=1}^C \subset \mathbb{L}_K^{d_{\text{eff}}}$ are learned with Riemannian adaptive optimization (Bécigneul & Ganea, 2018). Let $\mathbf{g}_a \in T_{\mathbf{w}_a}\mathbb{L}_K^{d_{\text{eff}}}$ be the Riemannian gradient w.r.t. $\mathbf{w}_a$. We maintain exponential moving averages: $\boldsymbol{\chi}_a \in T_{\mathbf{w}_a}\mathbb{L}_K^{d_{\text{eff}}}$ (first moment) and $\varsigma_a \in \mathbb{R}_+$ (second moment), with decay $(\rho_1, \rho_2) \in [0, 1)^2$. Because $T_{\mathbf{w}_a}\mathbb{L}_K^{d_{\text{eff}}}$ changes as $\mathbf{w}_a$ updates, we parallel-transport the previous momentum into the current tangent space using PT^K before updating the moving averages.

The objective uses a shot-specific hyperparameter tuple $\psi_S^* = (T, \beta, w_{\text{ce}}, w_{\hat{h}}, w_h, \omega, \tau)$, selected separately for each shot count S. The weights $w_{\text{ce}}, w_{\hat{h}}, w_h$ scale the cross-entropy, marginal-entropy, and conditional-entropy terms (see equation 17); T, β, ω, and τ control the remaining terms. We denote the loss by $\mathcal{L}$. Each component is described in the following section.

4 METHODS

We introduce LORENTZTIM, a method based on a custom transductive information maximisation algorithm formulated in the Lorentz hyperbolic space model. Our Euclidean baseline follows the TIM framework of Boudiaf et al. (2020). Section 4.1 analyses the hyperbolic properties of the UXO representations, while Section 4.2 describes the transductive algorithm.

4.1 HYPERBOLIC PROPERTIES

Khrulkov et al. (2020) have shown that visual data can exhibit an underlying tree-structured organisation, whose exponential growth in the number of nodes conflicts with the growth of polynomial volume of Euclidean space. Moreover, given the real-world characteristics (not pure noise) of UXO-related tasks, we draw on the manifold hypothesis (Bengio et al., 2013) that the high-dimensional visual data associated with UXO naturally reside on a lower-dimensional curved surface. As a result, because hyperbolic space expands exponentially and can represent tree-like hierarchies with

minimal distortion, mapping these representations onto a curved manifold (such as the Lorentz hyperboloid) helps mitigate the metric crowding that limits standard Euclidean embeddings. We observe that in the UXO context, this hierarchy can be explicit: the delivery mechanism (e.g. dropped, projected) constrains the ballistic form factor (shape, size, internal components), which then determines the ordnance family and, in turn, the subtype, implying that $\mathcal{Y}$ represents the leaves of the rooted taxonomy $\mathcal{T}$. Since the geometry induced by f_θ cannot be presumed to reflect $\mathcal{T}$ a priori, instead we evaluate it directly on $\mathcal{D}$ (CTX-UXO) using three diagnostics: (i) Gromov δ-hyperbolicity, which characterises the global tree-like structure; (ii) intrinsic dimension, highlighting the contrast between effective and ambient dimensionality; and (iii) cophenetic agreement, which tests whether the embedding distances at the sample-level are well approximated by a tree structure. Diagnostic (i) is evaluated with respect to a spectrum-matched null model, whose effective dimensionality is determined by (ii). This allows the study to separate genuine hierarchical organisation from artefacts arising purely from dimensionality.

For global δ-hyperbolicity, the Gromov product (Gromov, 1987) of $p, q \in \mathcal{Z}$ with respect to a base point $w \in \mathcal{Z}$ is:

$$(p \mid q)_w \;=\; \tfrac{1}{2}\big(d_{\mathbb{E}}(w,p) + d_{\mathbb{E}}(w,q) - d_{\mathbb{E}}(p,q)\big) \tag{4}$$

In equation 4, the space $\mathcal{Z}$ is said to be δ-hyperbolic if, for every choice of points $p, q, r, w \in \mathcal{Z}$, the following condition holds:

$$(p \mid r)_w \;\geq\; \min\big((p \mid q)_w,\, (q \mid r)_w\big) - \delta. \tag{5}$$

Denoting $A_w(p,q) \equiv (p \mid q)_w$, the associated hyperbolicity constant is given by:

$$\delta(\mathcal{Z}) \;=\; \max_{w \in \mathcal{Z}} \max_{p,q,r \in \mathcal{Z}} \Big[\min\big(A_w(p,q),\, A_w(q,r)\big) - A_w(p,r)\Big], \tag{6}$$

In accordance with equation 6, and using the fact that $\mathrm{diam}(\mathcal{Z}) = \max_{p,q \in \mathcal{Z}} d_{\mathbb{E}}(p,q)$, this may be written in the following scale-invariant form:

$$\delta_{\mathrm{rel}}(\mathcal{Z}) \;=\; \frac{2\,\delta(\mathcal{Z})}{\mathrm{diam}(\mathcal{Z})} \;\in\; [0,1]. \tag{7}$$

In equation 7, $\hat{\delta}_{\mathrm{rel}}$ is averaged over B batches of size m, using percentile-bootstrap confidence intervals (Raschka, 2018). We estimate intrinsic dimensionality using Levina–Bickel MLE and the participation ratio. We define the effective dimension d_{eff} using the participation ratio to parameterize the dimensionality-matched control. Let $r_1(z_i) \leq \cdots \leq r_k(z_i)$ be the ordered k-nearest-neighbour distances of z_i. We compute Levina–Bickel MLE (Levina & Bickel, 2004) as:

$$\hat{d}_k(z_i) \;=\; (k-2)\left[\sum_{j=1}^{k-1} \log \frac{r_k(z_i)}{r_j(z_i)}\right]^{-1}, d_{\mathrm{MLE}} \quad = \; \frac{1}{N}\sum_{i=1}^{N} \hat{d}_k(z_i), \tag{8}$$

With $\lambda_1 \geq \cdots \geq \lambda_d$ the eigenvalues of the empirical covariance of $\mathcal{Z}$, the participation ratio is evaluated as:

$$d_{\mathrm{PR}} \;=\; \frac{\big(\sum_{i=1}^{d} \lambda_i\big)^2}{\sum_{i=1}^{d} \lambda_i^2}. \tag{9}$$

We use d_{MLE} as a secondary consistency check and we set $d_{\mathrm{eff}} = \mathrm{round}(d_{\mathrm{PR}})$. To quantify δ-hyperbolicity, which measures how tree-like a metric space is, we evaluate how well pairwise distances are represented in a hierarchical tree using cophenetic correlation. Cophenetic agreement (Sokal & Rohlf, 1962) is the Pearson and Spearman correlation, over all pairs of sampled points, between their original ambient distance and their cophenetic distance from an average-linkage dendrogram fitted to the same distance matrix.

$$\rho_{\mathrm{coph}} \;=\; \mathrm{corr}\Big(\{d_{\mathrm{coph}}(z_i,z_j)\}_{i<j},\; \{d_{\mathbb{E}}(z_i,z_j)\}_{i<j}\Big) \;\in\; [-1,1], \tag{10}$$

We test whether the observed tree structure is stronger than what would be expected from a control that matches Z's empirical covariance spectrum. Let $\lambda_1 \geq \cdots \geq \lambda_{d_{\mathrm{eff}}}$ be the leading d_{eff} eigenvalues of the empirical covariance of Z, with eigenvectors $\mathbf{v}_1, \ldots, \mathbf{v}_{d_{\mathrm{eff}}} \in \mathbb{R}^d$, and set $\boldsymbol{V} = [\mathbf{v}_1, \ldots, \mathbf{v}_{d_{\mathrm{eff}}}]$. For $i = 1, \ldots, N$, sample $u_{i,k} \sim \mathcal{N}(0, \lambda_k)$ independently for $k = 1, \ldots, d_{\mathrm{eff}}$. We define:

$$\tilde{z}_i \;=\; \frac{\boldsymbol{V}\mathbf{u}_i}{\|\boldsymbol{V}\mathbf{u}_i\|_2} \;\in\; \mathbb{S}^{d-1}, \qquad \tilde{\mathcal{Z}} \;=\; \{\tilde{z}_i\}_{i=1}^{N}. \tag{11}$$

By construction, $\tilde{\mathcal{Z}}$ matches Z's dominant anisotropy but contains no latent tree signal (independent Gaussian coordinates), so any residual $\hat{\delta}_{\text{rel}}$/cophenetic gap cannot be associated with anisotropy or elongated class clusters alone. In equation 7, the results are reported accordingly as a gap relative to $\tilde{\mathcal{Z}}$. All diagnostics are reported in Section 5, Table 2.

4.2 ALGORITHM

Motivated by the geometric evidence in Sec. 4.1, transductive adaptation operates directly on $\mathbb{L}_K^{d_{\text{eff}}}$ with fixed (d_{eff}, K), $K = -1$, optimizing only the class centroids W. We adapt the TIM framework (Boudiaf et al., 2020) and CL2N-normalization (Wang et al., 2019) combining Riemannian Adam (Bécigneul & Ganea, 2018), per-shot hyperparameter tuning ψ_S^* (Algorithm 2), and Lorentzian centroids (Eq. 16). Transductive leakage is prevented at the centering step: features are centered using exclusively the base-split mean $\bar{f}_{\text{base}}$ before ℓ_2-normalization:

$$z_i = \frac{f_\theta(x_i) - \bar{f}_{\text{base}}}{\|f_\theta(x_i) - \bar{f}_{\text{base}}\|_2} \in \mathbb{S}^{d-1}, \qquad \bar{f}_{\text{base}} = \frac{1}{|\mathcal{D}_{\text{base}}|} \sum_{x \in \mathcal{D}_{\text{base}}} f_\theta(x). \tag{12}$$

The PCA-reduced features are treated as tangent vectors at the origin, $(0, U_{\text{PR}}^\top z_i) \in T_{\mathbf{o}}\mathbb{L}_K^{d_{\text{eff}}}$, where $U_{\text{PR}} \in \mathbb{R}^{d \times d_{\text{eff}}}$ is obtained per episode using SVD. These are lifted to the manifold using the exponential map (Nickel & Kiela, 2018; Khrulkov et al., 2020), defined for $\mathbf{u} \in T_{\mathbf{x}}\mathbb{L}_K^{d_{\text{eff}}}$ as:

$$\exp_{\mathbf{x}}^K(\mathbf{u}) = \cosh\big(\sqrt{-K}\,\|\mathbf{u}\|_L\big)\mathbf{x} + \frac{\sinh\big(\sqrt{-K}\,\|\mathbf{u}\|_L\big)}{\sqrt{-K}\,\|\mathbf{u}\|_L}\,\mathbf{u}, \qquad \mathbf{z}_i = \exp_{\mathbf{o}}^K\big((0, U_{\text{PR}}^\top z_i)\big), \tag{13}$$

The geodesic distance is obtained by directly applying it to the lifted points (from the Euclidean space to the Lorentz hyperboloid), as follows:

$$d_L(\mathbf{x}, \mathbf{y}) = \tfrac{1}{\sqrt{-K}} \operatorname{arcosh}\big(K\langle \mathbf{x}, \mathbf{y}\rangle_L\big) \tag{14}$$

With the geodesic classifier, class posteriors follow a distance-based softmax, and they are computed as:

$$p_{ia} = \frac{\exp\big(-\tau\, d_L^2(\mathbf{z}_i, \mathbf{w}_a)\big)}{\sum_{b=1}^{C} \exp\big(-\tau\, d_L^2(\mathbf{z}_i, \mathbf{w}_b)\big)}, \tag{15}$$

The prototypes are initialised by the Lorentzian centroid (Law et al., 2019) in closed-form:

$$\mathbf{w}_a^{(0)} = \frac{\bar{\mathbf{z}}_a}{\sqrt{K\langle \bar{\mathbf{z}}_a, \bar{\mathbf{z}}_a\rangle_L}}, \qquad \bar{\mathbf{z}}_a = \frac{1}{|\mathcal{S}_a|} \sum_{i \in \mathcal{S}_a} \mathbf{z}_i, \tag{16}$$

When $T = 0$, the method becomes the closed-form Hyp-SimpleShot baseline, performing nearest-centroid classification from $\mathbf{w}_a^{(0)}$ with no transductive optimisation. The objective preserves the information-theoretic structure of TIM Boudiaf et al. (2020), with geometry entering through the likelihood of equation 15:

$$\mathcal{L}(\mathbf{W}) = w_{\text{ce}}\, \mathcal{L}_{\text{CE}}(\mathcal{S}; \mathbf{W}) + w_h\, \hat{\mathcal{H}}(Y \mid X) - w_{\hat{h}}\, \hat{\mathcal{H}}(Y), \tag{17}$$

For Eq. 17, the cross-entropy term is defined as:

$$\mathcal{L}_{\text{CE}} = -\frac{1}{|\mathcal{S}|} \sum_{i \in \mathcal{S}} \sum_{a=1}^{C} \mathbf{1}_{[y_i = c_a]} \log p_{ia}, \tag{18}$$

Also, empirical mutual-information regularisation is:

$$\hat{\mathcal{H}}(Y \mid X) = -\frac{1}{|\mathcal{U}|} \sum_{i \in \mathcal{U}} \sum_{a=1}^{C} p_{ia} \log p_{ia}, \qquad \hat{\mathcal{H}}(Y) = -\sum_{a=1}^{C} \hat{p}_a \log \hat{p}_a \tag{19}$$

In equation 19, $\hat{p}_a$ is defined as:

$$\hat{p}_a = \frac{1}{|\mathcal{U}|} \sum_{i \in \mathcal{U}} p_{ia}. \tag{20}$$

Minimising equation 19 tightens the per-sample posteriors through the conditional term, while the marginal term prevents collapse onto a single class. The updates minimise the regularised objective:

$$\mathcal{L}_{\text{reg}} = \mathcal{L} + \frac{\omega}{2} \sum_{a=1}^{C} d_L^2(\mathbf{w}_a, \mathbf{o}), \tag{21}$$

Since $\mathbf{W}$ lies on a product manifold, Riemannian Adam maintains, per prototype, a first-moment estimate in the tangent space and a scalar second-moment estimate. It computes the Riemannian gradient by applying the inverse metric and projecting onto the tangent space.

$$\mathbf{g}_a = \text{proj}_{\mathbf{w}_a}\big(J \nabla_{\mathbf{w}_a} \mathcal{L}_{\text{reg}}\big), \qquad \text{proj}_{\mathbf{x}}(\mathbf{u}) = \mathbf{u} - K \langle \mathbf{x}, \mathbf{u} \rangle_L \mathbf{x}. \tag{22}$$

In equation 22, $J = \eta = \text{diag}(-1, 1, \ldots, 1)$ is the Minkowski metric, used to map the Euclidean gradient to the ambient Riemannian gradient before projecting onto the tangent space. With $\boldsymbol{\chi}_a^{(0)} = \mathbf{0}$ and $\varsigma_a^{(0)} = 0$, iteration t performs:

$$\boldsymbol{\chi}_a \leftarrow \rho_1 \boldsymbol{\chi}_a + (1 - \rho_1)\, \mathbf{g}_a, \qquad \varsigma_a \leftarrow \rho_2 \varsigma_a + (1 - \rho_2)\, \|\mathbf{g}_a\|_L^2, \tag{23}$$

$$\mathbf{w}_a^+ = \exp_{\mathbf{w}_a}^K\left(-\beta \, \frac{\boldsymbol{\chi}_a \, (1 - \rho_1^t)^{-1}}{\sqrt{\varsigma_a \, (1 - \rho_2^t)^{-1}} + 10^{-8}}\right), \qquad \boldsymbol{\chi}_a \leftarrow \text{PT}_{\mathbf{w}_a \to \mathbf{w}_a^+}^K(\boldsymbol{\chi}_a) \tag{24}$$

In equation 24, PT^K denotes parallel transport of momentum along the connecting geodesic, being defined as:

$$\text{PT}_{\mathbf{x} \to \mathbf{y}}^K(\mathbf{u}) = \mathbf{u} - \frac{K \langle \mathbf{y}, \mathbf{u} \rangle_L}{1 + K \langle \mathbf{x}, \mathbf{y} \rangle_L} (\mathbf{x} + \mathbf{y}), \tag{25}$$

Predictions $\hat{y}_q = c_{a^\star}$, where $a^\star = \arg\max_a p_{qa}$, are equivalent to choosing the prototype that is the geodesically closest. Each trial starts with equation 16, then runs T iterations of equation 24 to minimise equation 21. Full pseudocode appears in Algorithms 1 and 2 (Appendix C).

5 RESULTS

Table 2: **Geometric characterization of the DINOv3 embedding space $\mathcal{Z}$ for UXO.** Brackets denote 95% bootstrap confidence intervals. The participation ratio sets the effective dimension at 29 for PCA projection prior to hyperbolic lifting, while MLE serves as a complementary diagnostic.

Diagnostic	Z **(UXOs)**	$\tilde{Z}$ **(control)**	**Obs.**
$\hat{\delta}_{\text{rel}}$ (tree-likeness)	.2453 [.213,.274]	.4215 [.394,.462]	$\Delta = -.1762$ (tree-like)
d_{MLE}	16.43	18.05	–
d_{PR}	29.04	29.04 / 12.88	$d_{eff} = 29$
ρ_{coph} (Pearson/Spearman)	.768 / .613	.520 / .454	tree-metric consistent

Table 2 reports diagnostics from Sec. 4.1. UXO embeddings show a significantly stronger hierarchical structure than a spectrum-matched null control, demonstrating stronger tree-likeness than a covariance-matched Euclidean control. Figure 3 illustrates LORENTZTIM on a 5-way, 5-shot episode, serving as a representative example of how the Lorentzian centroids are optimized through our proposed algorithm. We evaluate on the CTX-UXO benchmark using the predefined standard splits (without augmentation) and report macro F1 scores. To ensure a fair comparison, we report baselines in their native protocol alongside enhanced versions. LorentzTIM *with fixed params* applies TIM's fixed hyperparameters (T=1000, lr=10^{-4}, τ=7.5) within $\mathbb{L}_K^{d_{\text{eff}}}$, thus decoupling and isolating the effect of the geometry. LORENTZTIM further performs the per-shot search described in Algorithm 2.

We evaluate on standard C-way S-shot episodes with $S \in \{1, 3, 5, 10\}$ and 15 queries per class, over M=1000 fixed-seed episodes shared by all methods for paired comparison. The results can be found in Table 3. LORENTZTIM consistently achieves a significant improvement in F1-score across all values of S. The ablation study in Table 4 shows that combining hyperbolic geometry with transductive adaptation delivers the largest gains.

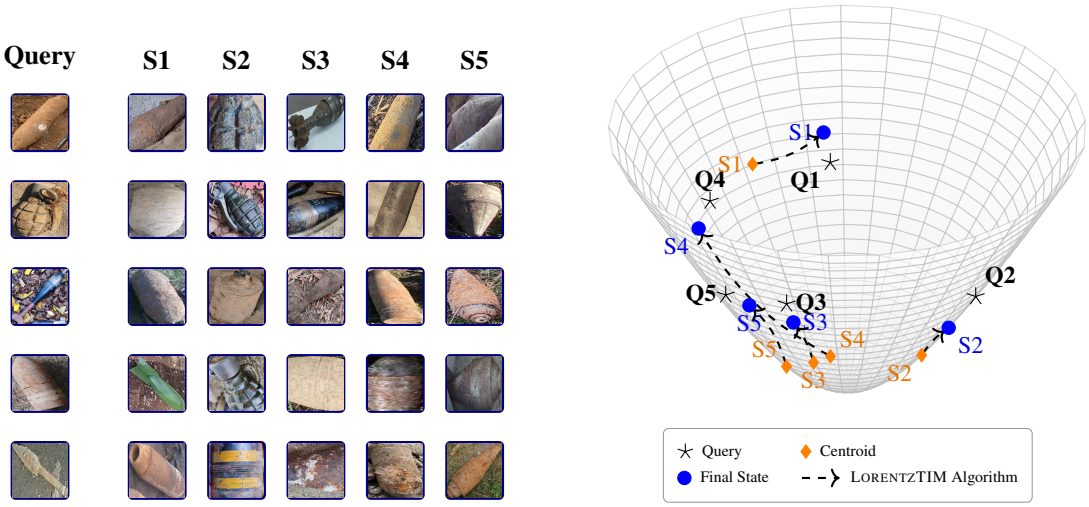
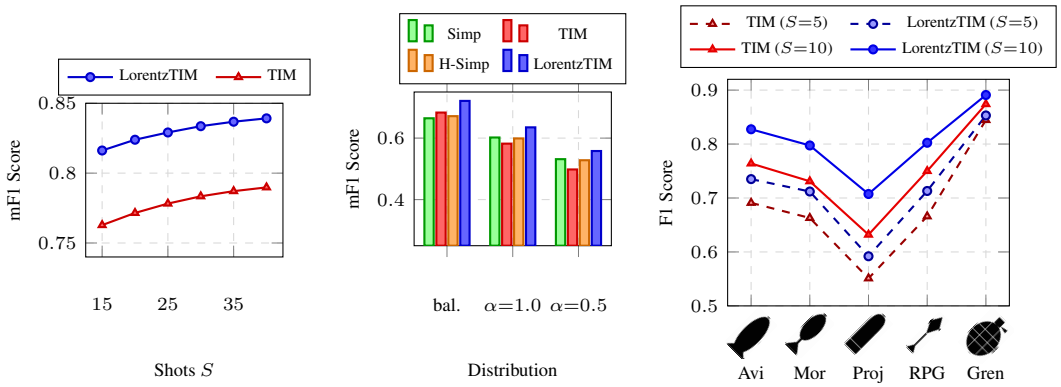


Figure 3: **Qualitative visualization of LorentzTIM.** A representative 5-way, 5-shot episode with a query batch of 5 samples. S denotes the support samples used to initialise the class centroids. The figure illustrates the transductive refinement of the initial Lorentzian centroids toward their final states. For visual clarity, the z-axis is rescaled.

Table 3: **Primary results on CTX-UXO.** F1 scores are reported as mean ± half-width of the 95% confidence interval over M=1000 paired episodes. An asterisk (*) denotes a statistically significant difference from TIM (original protocol), assessed using paired Wilcoxon tests with Holm correction ($p < 0.05$). All methods are evaluated on the same PCA-projected representation to ensure a controlled comparison. Random-search variants use the same trial budget.

Method	S=1	S=3	S=5	S=10
Inductive, Euclidean				
SimpleShot	.4690*±.0066	.6050*±.0054	.6645*±.0047	.7299*±.0037
DSN	.4692*±.0066	.6097*±.0053	.6638*±.0049	.7160*±.0039
Inductive, hyperbolic				
Hyp-SimpleShot	.4759*±.0066	.6107*±.0054	.6714*±.0047	.7340*±.0036
Hyp-ProtoNet	.4672*±.0066	.5952*±.0054	.6581*±.0048	.7249*±.0038
Taylor-HNN	.4685*±.0065	.5984*±.0053	.6615*±.0046	.7276*±.0036
Hyp-Busemann	.4718*±.0066	.6064*±.0054	.6670*±.0046	.7320*±.0036
Transductive, Euclidean				
TIM (original protocol)	.4841±.0071	.6432±.0059	.6831±.0044	.7502±.0036
TIM (with random search)	.4954*±.0072	.6518*±.0055	.6935*±.0042	.7614*±.0037
PT-MAP	.4794*±.0065	.6086*±.0057	.6628*±.0050	.7255*±.0040
LaplacianShot	.4809*±.0073	.6191*±.0057	.6755*±.0050	.7356*±.0037
Transductive, hyperbolic				
LorentzTIM (ours, fixed params)	.5128*±.0087	.6644*±.0055	.7167*±.0047	.7957*±.0037
LorentzTIM (ours, random search)	**.5383***±.0094	**.6710***±.0058	**.7210***±.0050	**.8051***±.0039

Moreover, the observed advantage over Euclidean TIM (ΔmF1) aligns with our geometric indicators: when the data are more intrinsically tree-like, hyperbolic methods tend to outperform flat Euclidean baselines. We additionally sweep $S \in \{15, 20, \ldots, 40\}$ to study the large-shot regime (see Figure 4a) and consider a realistic protocol with varying query-class proportions drawn from a symmetric Dirichlet prior (see Figure 4b). As illustrated in 4c, a key result is that the observed performance improvements are driven mainly by visual similarity between classes, rather than by the frequency with which each class occurs. This difference matters because it indicates that hyperbolic geometry offers practical, real-world advantages for dealing with UXO categories that look very similar and are consequently hard to differentiate.



(a) Extended-shot. (b) Imbalanced queries. (c) Per-class F1 performance.

Figure 4: **Robustness and per-class performance analysis.** LORENTZTIM consistently outperforms baselines across extended-shot regimes (4a) and under severe query imbalance (4b). Per-class results (4c) show the largest gains for visually similar UXO categories. Abbreviations: Avi (Aviation Bomb), Mor (Mortar Bomb), Proj (Projectile), RPG (Rocket-Propelled Grenade), Gren (Grenade).

Table 4: **Ablation study.** We report macro F1 change (ΔmF1) relative to the full LORENTZTIM model as mean $\pm$ half-width of the 95% bootstrap confidence interval. Each ablation modifies a single component while holding all others fixed.

Configuration	S=1	S=3	S=5	S=10
Full model LORENTZTIM	**.5383**$\pm$**.0094**	**.6710**$\pm$**.0058**	**.7210**$\pm$**.0050**	**.8051**$\pm$**.0039**
Impact of Geometry & Transduction				
(a) Curvature $c \to 0$	$-$.0497$\pm$.0039	$-$.0239$\pm$.0026	$-$.0328$\pm$.0022	$-$.0493$\pm$.0019
(b) T=0 (w/o transductive steps)[1]	$-$.0624$\pm$.0049	$-$.0603$\pm$.0029	$-$.0496$\pm$.0028	$-$.0711$\pm$.0030
Impact of Loss Terms				
(c) w/o marginal entropy $\hat{\mathcal{H}}(Y)$	$-$.1650$\pm$.0068	$-$.0520$\pm$.0034	$-$.0185$\pm$.0017	$-$.0080$\pm$.0013
(d) w/o Mutual Information	$-$.0412$\pm$.0048	$-$.0354$\pm$.0028	$-$.0289$\pm$.0018	$-$.0152$\pm$.0020
Impact of Normalization & Optimization				
(e) CL2N $\to$ plain ℓ_2 normalization	$-$.0154$\pm$.0040	$-$.0085$\pm$.0019	$-$.0042$\pm$.0016	$-$.0028$\pm$.0012
(f) R-Adam $\to$ R-SGD optimizer	$-$.0082$\pm$.0030	$-$.0125$\pm$.0019	$-$.0094$\pm$.0014	$-$.0051$\pm$.0008

[1] Setting T=0 mathematically reduces the model to Hyp-SimpleShot (see Table 3).

Additional Experiments. To demonstrate robustness beyond the UXO domain, we extend our evaluation to an alternative backbone (CLIP) and across three standard fine-grained and hierarchical benchmarks (CUB-200-2011, FGVC-Aircraft, and tieredImageNet). In all configurations, LORENTZTIM achieves the strongest performance, outperforming the baselines by up to 6%. A summary with the per-shot evolution of LORENTZTIM for each hierarchical benchmark is shown in Figure 5. Full results are provided in Appendix G.

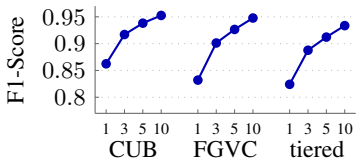


Figure 5: **LORENTZTIM performance across benchmarks.**

6 CONCLUSIONS

This paper addresses the humanitarian problem of identifying unexploded ordnance (UXO). A measure-first geometric analysis shows UXO embeddings exhibit low intrinsic dimensionality and significant tree-like structure. Based on this evidence, we propose LORENTZTIM, a transductive information maximization method on the Lorentz hyperboloid. LORENTZTIM achieves statistically significant improvements, and ablations suggest these gains arise primarily from combining transduction with hyperbolic geometry. Future directions are outlined in Appendix H.

REPRODUCIBILITY STATEMENT

The notation and initial problem setup are presented in Section 3. Hyperparameter choices are reported in Appendix E. The objectives for transductive adaptation, along with the optimization procedure, are outlined in Section 4.2, with pseudocode given in Algorithm 1. The evaluation protocol, which follows standard episodic few-shot classification practices, is described in Section 5. Additional related work, including research on feature extractors, Dirichlet-based imbalance techniques, and the paired Wilcoxon method, is discussed in Appendix B.

For code and datasets, see the Appendix A.

AI USE STATEMENT

In this work, we used Google Gemini Pro for sentence-level grammatical refinement, code readability polish, and resolving complex plot-layout/rendering issues in multi-panel figures (see layout of Figure 3 and Figure 4c). We have not used generative AI tools for research conceptualisation, theoretical developments, mathematical derivations, experimental design, or data analysis. All LLM edits were manually verified, tested and reviewed by the authors. We take full responsibility for the final content of this work, including all text, claims, mathematical formulations, and artefacts.

ETHICS STATEMENT

We confirm that we have read and understood the ICLR Code of Ethics and comply with it. The proposed work is developed with the objective of supporting life-saving operations, and, as demonstrated in the Appendix, it can be effectively generalised to other domains. The datasets used in this work are publicly available and widely used for research, and we are not aware of any metadata in these datasets that would introduce additional risks of misuse in the context of this study. We encourage end-users to contact us with any questions and to see our human-in-the-loop recommendations in the Appendix D.

REFERENCES

Rashid Ali, Ran Liu, Yongping He, Anand Nayyar, and Basit Qureshi. Systematic review of dynamic multi-object identification and localization: Techniques and technologies. *IEEE Access*, 9: 122924–122950, 2021.

Ahmad Bdeir, Kristian Schwethelm, and Niels Landwehr. Fully hyperbolic convolutional neural networks for computer vision. In *International Conference on Learning Representations*, volume 2024, pp. 47687–47711, 2024.

Gary Bécigneul and Octavian-Eugen Ganea. Riemannian adaptive optimization methods. *arXiv preprint arXiv:1810.00760*, 2018.

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

Kevin S. Beyer, Jonathan Goldstein, Raghu Ramakrishnan, and Uri Shaft. When is "nearest neighbor" meaningful? In Catriel Beeri and Peter Buneman (eds.), *Database Theory - ICDT '99, 7th International Conference, Jerusalem, Israel, January 10-12, 1999, Proceedings*, volume 1540 of *Lecture Notes in Computer Science*, pp. 217–235. Springer, 1999. ISBN 3-540-65452-6.

Malik Boudiaf, Imtiaz Masud Ziko, Jérôme Rony, José Dolz, Pablo Piantanida, and Ismail Ben Ayed. Transductive information maximization for few-shot learning. *CoRR*, abs/2008.11297, 2020. URL https://arxiv.org/abs/2008.11297.

Matthew P Bray and Curtis A Link. Learning machine identification of ferromagnetic uxo using magnetometry. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 8(2):835–844, 2014.

Nurendra Choudhary and Chandan K. Reddy. Towards scalable hyperbolic neural networks using taylor series approximations, 2022. URL https://arxiv.org/abs/2206.03610.

Gheorghe Marian Craioveanu and Grigore Stamatescu. Ctx-uxo: A comprehensive dataset for detection and identification of unexploded ordnances. *IEEE Dataport*, 2024. doi: 10.21227/cwnm-de53. URL https://dx.doi.org/10.21227/cwnm-de53.

Gheorghe-Marian Craioveanu, Vasile Calofir, and Grigore Stamatescu. CTX-UXO: Contextual visual dataset for localisation, classification and identification of unexploded ordnance. *Nature Scientific Data*, 13(1):1340, 2026. ISSN 2052-4463. doi: 10.1038/s41597-026-07818-8. URL https://doi.org/10.1038/s41597-026-07818-8.

Janez Demšar. Statistical comparisons of classifiers over multiple data sets. *Journal of Machine learning research*, 7(Jan):1–30, 2006.

Guneet S. Dhillon, Pratik Chaudhari, Avinash Ravichandran, and Stefano Soatto. A baseline for few-shot image classification. *CoRR*, abs/1909.02729, 2019. URL http://arxiv.org/abs/1909.02729.

Zhi Gao, Yuwei Wu, Yunde Jia, and Mehrtash Harandi. Curvature generation in curved spaces for few-shot learning. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 8671–8680. IEEE, 2021.

Mina Ghadimi Atigh, Martin Keller-Ressel, and Pascal Mettes. Hyperbolic busemann learning with ideal prototypes. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan (eds.), *Advances in Neural Information Processing Systems*, volume 34, pp. 103–115. Curran Associates, Inc., 2021. URL https://proceedings.neurips.cc/paper_files/paper/2021/file/01259a0cb2431834302abe2df60a1327-Paper.pdf.

M. Gromov. *Hyperbolic Groups*, pp. 75–263. Springer New York, New York, NY, 1987. ISBN 978-1-4613-9586-7. doi: 10.1007/978-1-4613-9586-7_3. URL https://doi.org/10.1007/978-1-4613-9586-7_3.

Zahid Hasan, Masud Ahmed, and Nirmalya Roy. Lorentz entailment cone for semantic segmentation. In *2026 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, pp. 5216–5225. IEEE, 2026.

Ruibing Hou, Hong Chang, Bingpeng Ma, Shiguang Shan, and Xilin Chen. Cross attention network for few-shot classification. *Advances in neural information processing systems*, 32, 2019.

Yuqing Hu, Vincent Gripon, and Stéphane Pateux. Leveraging the feature distribution in transfer-based few-shot learning. In *International conference on artificial neural networks*, pp. 487–499. Springer, 2021.

ICRC. Casualties surge from landmines and unexploded ordnance. https://www.icrcnewsroom.org/, April 2025. Accessed: 2026-08-22.

Valentin Khrulkov, Leyla Mirvakhabova, Evgeniya Ustinova, Ivan Oseledets, and Victor Lempitsky. Hyperbolic image embeddings. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 6418–6428, 2020.

Marc Law, Renjie Liao, Jake Snell, and Richard Zemel. Lorentzian distance learning for hyperbolic representations. In *International conference on machine learning*, pp. 3672–3681. PMLR, 2019.

Elizaveta Levina and Peter Bickel. Maximum likelihood estimation of intrinsic dimension. *Advances in neural information processing systems*, 17, 2004.

Shuai Ma, Ying Lei, Xinru Wang, Chengbo Zheng, Chuhan Shi, Ming Yin, and Xiaojuan Ma. Who should i trust: Ai or myself? leveraging human and ai correctness likelihood to promote appropriate trust in ai-assisted decision-making. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pp. 1–19, 2023.

Andreia Machado Brito-da Costa, Diana Martins, Diogo Rodrigues, Luís Fernandes, Rui Moura, and Áurea Madureira-Carvalho. Ground penetrating radar for buried explosive devices detection: A case studies review. *Australian journal of forensic sciences*, 54(4):559–578, 2022.

Subhransu Maji, Esa Rahtu, Juho Kannala, Matthew B. Blaschko, and Andrea Vedaldi. Fine-grained visual classification of aircraft. *CoRR*, abs/1306.5151, 2013. URL http://arxiv.org/abs/1306.5151.

Maximillian Nickel and Douwe Kiela. Poincaré embeddings for learning hierarchical representations. *Advances in neural information processing systems*, 30, 2017.

Maximillian Nickel and Douwe Kiela. Learning continuous hierarchies in the lorentz model of hyperbolic geometry. In *International conference on machine learning*, pp. 3779–3788. PMLR, 2018.

Xavier Núñez-Nieto, Mercedes Solla, Paula Gómez-Pérez, and Henrique Lorenzo. Gpr signal characterization for automated landmine and uxo detection based on machine learning techniques. *Remote sensing*, 6(10):9729–9748, 2014.

Qi Qiao, Shengguo Ge, Ke Wang, and Huiying Hu. A systematic literature review on few-shot image classification. *IEEE Access*, 2025.

Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pp. 8748–8763. PmLR, 2021.

Sebastian Raschka. Model evaluation, model selection, and algorithm selection in machine learning. *arXiv preprint arXiv:1811.12808*, 2018.

Jie Ren, Changmiao Li, Yaohui An, Weichuan Zhang, and Changming Sun. Few-shot fine-grained image classification: A comprehensive review. *AI*, 5(1):405–425, 2024.

Mengye Ren, Eleni Triantafillou, Sachin Ravi, Jake Snell, Kevin Swersky, Joshua B. Tenenbaum, Hugo Larochelle, and Richard S. Zemel. Meta-learning for semi-supervised few-shot classification. *CoRR*, abs/1803.00676, 2018. URL http://arxiv.org/abs/1803.00676.

Oriane Siméoni, Huy V Vo, Maximilian Seitzer, Federico Baldassarre, Maxime Oquab, Cijo Jose, Vasil Khalidov, Marc Szafraniec, Seungeun Yi, Michaël Ramamonjisoa, et al. Dinov3. *arXiv preprint arXiv:2508.10104*, 2025.

Christian Simon, Piotr Koniusz, Richard Nock, and Mehrtash Harandi. Adaptive subspaces for few-shot learning. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4135–4144. IEEE, 2020.

Jake Snell, Kevin Swersky, and Richard S. Zemel. Prototypical networks for few-shot learning. *CoRR*, abs/1703.05175, 2017. URL http://arxiv.org/abs/1703.05175.

Robert R Sokal and F James Rohlf. The comparison of dendrograms by objective methods. *Taxon*, 11(2):33–40, 1962.

Flood Sung, Yongxin Yang, Li Zhang, Tao Xiang, Philip HS Torr, and Timothy M Hospedales. Learning to compare: Relation network for few-shot learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1199–1208, 2018.

UNMAS. *Glossary of mine action terms, definitions and abbreviations*. New York, USA, amendment 13, second edition, 2026.

Olivier Veilleux, Malik Boudiaf, Pablo Piantanida, and Ismail Ben Ayed. Realistic evaluation of transductive few-shot learning. *Advances in Neural Information Processing Systems*, 34:9290–9302, 2021.

Oriol Vinyals, Charles Blundell, Timothy P. Lillicrap, Koray Kavukcuoglu, and Daan Wierstra. Matching networks for one shot learning. *CoRR*, abs/1606.04080, 2016. URL http://arxiv.org/abs/1606.04080.

C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie. Caltech-ucsd birds-200-2011. Technical Report CNS-TR-2011-001, California Institute of Technology, 2011.

Yan Wang, Wei-Lun Chao, Kilian Q. Weinberger, and Laurens van der Maaten. Simpleshot: Revisiting nearest-neighbor classification for few-shot learning. *CoRR*, abs/1911.04623, 2019. URL http://arxiv.org/abs/1911.04623.

Zhu Wen, Songtong Han, Chengwei Gao, Yuze Chen, Limei Guo, and Ya Zhang. A deep learning method for recognizing types of unexploded ordnance based on magnetic detection. *IEEE Transactions on Geoscience and Remote Sensing*, 62:1–13, 2024.

Han-Jia Ye, Hexiang Hu, De-Chuan Zhan, and Fei Sha. Few-shot learning via embedding adaptation with set-to-set functions. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 8808–8817, 2020.

Imtiaz Ziko, Jose Dolz, Eric Granger, and Ismail Ben Ayed. Laplacian regularized few-shot learning. In *International conference on machine learning*, pp. 11660–11670. PMLR, 2020.