

From Manual to Real-World: Deep Unfolded Neural Network for Cross-Domain UXO Retrieval

Gheorghe Marian Craioveanu¹ and Grigore Stamatescu¹ (✉)

University of Science and Technology Politehnica Bucharest, Romania
{marian.craioveanu, grigore.stamatescu}@upb.ro

Abstract. Unexploded Ordnance (UXO), conventional bombs deployed in past conflicts that failed to detonate, represents a significant challenge for robust deep learning identification in realistic field conditions. This is primarily due to the severe domain shift between noisy field imagery and clean technical manual references. At the same time, UXO mitigation represents a difficult task even for human operators due to stress, fatigue, and the associated risks. Despite major advances in deep learning for binary classification, localisation, and segmentation, practical field identification has remained largely unsolved. This work helps to bridge the gap in smart UXO identification, an application with direct impact on society, by unfolding a parametric shape-fitting procedure into a neural network, thereby embedding domain-specific geometric constraints directly into the learnable architecture. Unlike black-box few-shot encoders, this structured approach preserves the interpretability of classical optimisation while gaining data-driven robustness to occlusions and noise common in operational demining/clearance operations. The learnt unfolded representation ensures the correspondence of munitions in latent space, regardless of whether they originate from field conditions or controlled environments, enabling direct nearest-neighbour retrieval without relying on RGB appearance or class supervision. Experiments on a reference dataset demonstrate that the proposed architecture improves retrieval robustness compared to the baseline methods. In our experimental setup, Recall@1 improves from 53.4% to 62.1%, and Recall@5 improves from 65.4% to 82.1%. These findings show that deep unfolded shape refinement based on UXO characteristics offers a practical, interpretable, and physically grounded approach to identifying unexploded ordnances in real-world conditions.

Keywords: Unexploded Ordnance · Deep Unfolded Network · Cross-Domain Retrieval

1 Introduction

Unexploded Ordnance (UXO), also known as Unexploded Bombs (UXB), poses a serious hazard and an ongoing challenge in contemporary society. According to the North Atlantic Treaty Organisation (NATO) [22], UXO represents an explosive object that has been primed, fuzed, armed, or otherwise prepared for use,



Fig. 1: UXO in the field can be oxidised, partially covered by vegetation/soil, and sometimes even component parts are missing, being difficult to use common retrieval approaches. The illustration is produced by running the Segment Anything Model [24] on the annotated bounding box of a UXO sample from the CTX-UXO dataset [8].

may have been fired, dropped, launched, or projected, but remains unexploded due to malfunction, design, or any other reason. Based on statistics provided by the United Nations Mine Action Service [17,31], hundreds of millions of explosive remnants of war are present in at least 70 countries around the world. Across Europe, statistical data suggest that this contamination affects all European nations to varying degrees. In Germany, up to 1.6 million tonnes of ammunition are estimated [19] to be located along its coasts or in its waters. In Norway, up to 200,000 tonnes of munitions are believed to remain undiscovered [19]. In Italy, the authorities have estimated [2] that on average up to 60,000 munitions items (including parts of UXOs) are found annually. UXO are also present in other European and non-European countries, but the projected ranges for the total number of explosive items are very broad as they are difficult to quantify. Beyond the risks posed by direct contact, munitions also affect civilian society through soil and water contamination, releasing toxic substances (mainly TNT) that can cause diseases [13,19]. The phenomenon is global in nature and difficult to delimit [31]. It is important to emphasize that unexploded ordnance (UXO) represents a long-standing and continuing hazard, constituting a persistent societal challenge and a major public safety concern that has endured for more than eight decades. We provide a qualitative example with a real-world projectile in Figure 1.

In this paper, we propose a dual approach, integrating domain-specific Explosive Ordnance Disposal (EOD) expertise with a deep learning architecture, to advance applied research in safety-critical applications represented by UXO. For context, in operational settings UXO identification typically relies on establishing visual correspondence between items observed in the field and reference

sources, such as technical documentation and specialised databases. Automating this workflow is non-trivial due to the substantial domain shift between real-world field imagery and the idealised visualisations found in reference materials. The targets may be partially occluded, truncated by image borders, or obscured by environmental clutter (e.g., soil, vegetation) [9]. In contrast, technical manuals provide clear, complete and canonical views of the ordnance. It is important to consult manuals as references in order to replicate the field methodology, since these references typically include technical details (such as calibre, explosive content, origin, and associated risks) that are crucial for the stages that follow identification. This discrepancy results in a domain gap, meaning the statistical characteristics of the field data differ substantially from those of the reference data. As a result, appearance-driven methods, often exhibit sub-optimal performance, as they implicitly rely on texture, illumination, and visual context, features which fail to transfer reliably between domains. The main contributions of this work are:

- We introduce a deep unfolded shape refinement network that reformulates an iterative parametric shape fitting process as a learnable neural architecture, yielding a compact and interpretable shape representation that is robust to noise, partial occlusion, and missing shape information;
- We develop a training strategy based solely on field segmentation masks and demonstrate that the learnt unfolded representation enables robust cross-domain retrieval;
- We improve class-level identification under severe domain shift, a necessary step toward fine-grained identification when subtype labels become available. The task is formulated as class-level retrieval inside a common latent space, allowing the method to generalise to UXO types that are previously unseen or that differ substantially in visual appearance;
- We introduce a benchmark that contrasts few-shot methods based on RGB images, traditional shape-matching techniques, and our shape-only refinement neural network. In addition, we conduct an ablation study in Section 4 to examine how changes in the number of unfolding steps or in the chosen physically grounded parameters influence performance.

2 Related Work

With the increasing availability of open-source optical images, such as CTX-UXO [8] or CAT-UXO [7], recent works explored the localisation and/or classification of UXO directly from RGB images using convolutional neural networks or transformer architecture configured for UXO image classification and localisation. In this setting, most approaches focus on coarse categorisation [34,25]. Notably, in their dual-domain study [4], the researchers Baur et al. emphasised that deep learning detection models must be easily deployable in portable UAV/UGV systems to facilitate their practical use in the field. Using a drone and a two-stage detection architecture with intermediate region proposals, they achieved results of

71.5% mAP. However, it should be noted that the analysis was conducted in a single class (landmine, binary classification) with only one subtype (butterfly mine, PFM-1). Under favourable conditions, such methods have demonstrated promising performance, mostly when objects are fully visible and imaging conditions are controlled. In a recent preliminary contribution [18], the authors suggested that UXO mitigation should be addressed using an uncertainty-aware framework that relies on Bayesian Networks combined with Monte Carlo Dropout. This work indicates that implementations focusing on architecture-level tailored to the UXO detection problem, while keeping safety measures and explainability, can yield substantial performance gains. However, in the real world, the optical appearance varies significantly between environments, and performance often degrades in realistic field scenarios involving occlusion, clutter, and background variability. This was addressed by introducing contextual representative datasets with real field UXO, such as CTX-UXO [8]. Even if the dataset is representative, it does not include intrinsic metadata, such as calibre or other relevant attributes, meaning that operators are still required to rely on reference materials (databases) [7] augmented with all the metadata necessary for UXO mitigation. Appearance-based methods struggle to infer physically meaningful characteristics, and classification decisions are often driven by texture or contextual cues rather than object geometry. UXO identification has traditionally been addressed using geophysical sensing modalities such as ground penetration radar (GPR) [3] and magnetometry [26]. In these domains, identification is rarely performed through direct classification. Instead, the dominant paradigm relies on interpreting the shape of the measured signal response, such as spatial signatures, temporal waveforms, or polarisation curves, and comparing them against libraries of known responses.

Deep unfolding has emerged as a powerful paradigm for combining the interpretability of iterative algorithms with neural networks [28]. With good results for sparse coding [1] and compressed sensing [36], unfolded networks replace fixed update rules with learnable functions while preserving the algorithmic structure. More recently, shape analysis tasks have been addressed through deep unfolding [35]. Xinyi Wei et al. [35] propose a model-based deep learning framework that incorporates explicit generative priors within a deep unfolding architecture for inverse imaging. They embed a probabilistic flow-based prior directly into the iterative reconstruction, enabling task-adaptive optimisation. Evaluated on denoising, inpainting, and deblurring, and compared with strong ResNet- and U-Net-based proximal baselines, the method delivers competitive or superior results. In deblurring, it reaches 40.55 dB PSNR, outperforming U-Net proximal (38.66 dB) and ResNet-based approaches (35.55 dB). It attains 35.93 dB PSNR for inpainting and 28.24 dB PSNR for denoising, demonstrating robustness across tasks. Our work extends the deep unfolding neural network paradigm to cross-domain UXO shape retrieval, where the unfolding structure enables robust estimation of parametric shape descriptors from noisy observations. We acknowledge that classical shape matching methods, including Iterative Closest Point (ICP) [37], Procrustes analysis [12], and shape context descriptors [16],

provide geometrically grounded representations, but these methods are sensitive to noise and partial observations. ICP iteratively minimises the distance between corresponding point sets through rigid transformations. Procrustes analysis aligns shapes by removing translation, rotation, and scale, allowing direct shape comparison. Statistical shape analysis using Procrustes distances is a core technique in image analysis, as detailed by Dryden et al. [11], and with wide-ranging applications in biology, chemistry, bioinformatics, geography, archaeology, and related disciplines. Principal Component Analysis (PCA) further enhances these methods [11] by reducing noise due to object orientation and rotation and will therefore be applied in the preprocessing stage in this study. For fair benchmarking, Procrustes distance will also serve as a reference measure in the current manuscript. Alternative approaches based on point clouds, such as PointNet [23], were originally designed for 3D point cloud data. Adapting it to the context of our present study would require architectural constraints and modifications, this would constitute a separate line of research in its own right and would considerably broaden the scope of the current manuscript. We further acknowledge a range of recent methods [29,14,27,33], which may prove to be effective. However, these approaches are predicated upon fundamentally divergent paradigms, such that a direct and methodologically transparent comparison cannot, in good faith, be undertaken at present. For this reason, we leave their examination to future work, where they will be explored in a more focused and thorough way.

The current State-Of-The-Art (SoA) enables the localisation of UXO in the field and supports coarse categorisation. However, the precise determination of the exact UXO type remains mostly unresolved for real-world applications, particularly in relation to EOD procedures that rely on cross-referencing technical manuals. It can be observed that challenges highlight the gap between traditional geophysical approaches, which emphasise shape-based signal interpretation, and modern optical methods, which primarily rely on appearance. This observation motivates a return to shape-centric representations, adapted to the optical domain and capable of operating under severe domain mismatch between noisy field imagery and idealised technical references that have the metadata required for post-identification operations.

3 Methods

3.1 Preliminary

We formulate UXO identification as class-level cross-domain retrieval from binary segmentation masks, without relying on RGB appearance, texture, or scale calibration. Let $M \in \{0, 1\}^{H \times W}$ denote an instance mask. We extract the foreground coordinates, compute their centroid and covariance matrix, and apply PCA to align the object with its principal axes. If the eigenvalue ratio falls below an isotropy threshold κ , no alignment is performed. The direction of the first principal axis is resolved by a slope-asymmetry criterion, enforcing consistent front-back orientation. In the resulting canonical frame, each foreground

point has coordinates (z_i, y_i) , with z along the principal axis and y orthogonal. The axial coordinate is normalised to $[0, 1]$ and discretised into N bins, the last bin including the end point to preserve the tip of the object. For each bin, we estimate a radial half-width profile as half the robust width of the orthogonal coordinate distribution, computed by trimming the extreme values to suppress outliers from segmentation noise. Bins with fewer than m points are set to zero and marked as invalid. The maximum calibre $D = 2 \max_k r_{\text{raw}}(z_k)$ is used for normalisation to $[0, 1]$. A bin-wise validity vector $\mathbf{v} \in \{0, 1\}^N$ is defined as $v_k = 1$ if the occupancy $n_k \geq m$ and $r_{\text{raw}}(z_k) \geq \alpha \cdot \text{median}\{r_{\text{raw}}(z_j) \mid n_j \geq m\}$, and $v_k = 0$ otherwise. This vector masks unreliable regions during both training and inference.

Additional Notations. Throughout this section: $r_{\text{obs}} \in \mathbb{R}^N$ denotes the raw observed radial profile, $p^{(t)}$ the profile estimate at unfolding iteration t , $\hat{p} := p^{(T)}$ the final unfolded output, and $r_s = \text{SG}(\hat{p})$ the smoothed profile from which derivative-based descriptors are extracted.

3.2 Parametric Shape Representation

We model UXO shapes using a compact parametric vector $\theta \in \mathbb{R}^d$, where d denotes the total number of shape parameters, that captures key geometric features. Given a profile p (either the raw measurement r_{obs} or the unfolded estimate $p^{(T)}$), we generate a smoothed profile $r_s = \text{SG}(p)$ via a Savitzky–Golay filter with window size ω_{SG} and polynomial order d_{SG} , chosen to retain inflection points while attenuating high-frequency quantisation noise. All derivative-based descriptors described below are derived from r_s . Let $\mathcal{I}_f = \{k : z_k \leq \beta\}$ and $\mathcal{I}_b = \{k : z_k \geq 1 - \beta\}$ denote the sets of bins whose centres fall in the frontal and rear portions of the normalised axis, respectively, where $\beta \in (0, 0.5)$ is the consensus coverage threshold. We then introduce end-coverage scores c_f and c_b with the following equations:

$$c_f = \frac{1}{|\mathcal{I}_f|} \sum_{k: z_k \in \mathcal{I}_f} v_k \quad (1)$$

$$c_b = \frac{1}{|\mathcal{I}_b|} \sum_{k: z_k \in \mathcal{I}_b} v_k \quad (2)$$

In Equation 1 and Equation 2, $|\mathcal{I}_f|, |\mathcal{I}_b|$ denote the number of bins with centres in the respective intervals. An end region is considered reliably observed when its coverage score exceeds β ; these scores are used to gate the taper-ratio computation (Eq. 3), preventing unreliable slope estimates from corrupting the shape descriptor. Based on domain expertise in ordnance identification, validated by public references [7], and international standards [31], we define five complementary shape parameters. The **length-to-Diameter ratio (L/D)** represents the aspect ratio computed from the *unnormalised* aligned coordinates as L/D ,

where $L = z_{\max} - z_{\min}$ and $D = 2 \max_k r_{\text{raw}}(z_k)$, discriminating elongated projectiles, *teardrop* shaped mortar bombs and grenades. The **profile taper ratio** quantifies the asymmetry between nose and tail using the mean absolute slope of the smoothed profile near each end (central finite differences), defined as:

$$\tau = \frac{\frac{1}{|\mathcal{I}_f|} \sum_{z \in \mathcal{I}_f} |r'_s(z)| + \varepsilon}{\frac{1}{|\mathcal{I}_b|} \sum_{z \in \mathcal{I}_b} |r'_s(z)| + \varepsilon}, \quad (3)$$

The **encoded axial profile** corresponds to the normalised radial profile $p(z)$ sampled at equispaced points N along the principal axis (using $p = r_{\text{obs}}$ before unfolding and $p = p^{(T)}$ after unfolding). **Inflection Point Count** denotes the number of sign changes in the second derivative of $r_s(z)$ (central differences), counted only where $|r''_s(z)| > \gamma$, with γ an adaptive noise floor estimated from the variance of the second derivative on smooth reference shapes. Note that n_{infl} is a discrete, non-differentiable variable that is not involved in gradient-based training and is used only as a manually designed retrieval feature, computed after the fact from $p^{(T)}$. **Fourier Descriptors** are defined as the magnitudes of the first K_F non-zero DFT coefficients of r_s , normalised by the DC component, i.e., $|c_k| / \max(|c_0|, \varepsilon)$ for $k = 1, \dots, K_F$.

We define the deterministic parameter extraction function $h : \mathbb{R}^N \times \mathbb{R}_+ \rightarrow \mathbb{R}^d$ that maps a profile p and the raw scale to the complete parameter vector:

$$\theta = \left[\frac{L}{D}, \tau, p(z_1), \dots, p(z_N), n_{\text{infl}}, \frac{|c_1|}{\max(|c_0|, \varepsilon)}, \dots, \frac{|c_{K_F}|}{\max(|c_0|, \varepsilon)} \right]^\top \quad (4)$$

In Equation 4, $d = 2 + N + 1 + K_F = 75$ parameters for $N = 64$ and $K_F = 8$. Before unfolding, θ is $h(r_{\text{obs}}, L/D)$. After unfolding, the vector is determined by $\hat{\theta} = h(p^{(T)}, L/D)$.

3.3 Datasets and Data Preparation

We used for training and evaluation the Contextual Vision for Unexploded Ordnances (CTX-UXO) dataset [8], which provides labelled UXO images captured under realistic field conditions. In our study, we employ the following classes: Projectile, Mortar Bomb, and Grenade, with a total of 13,789 field instances. We extract the corresponding segmentation mask from the annotations. The masks exhibit realistic degradations that include partial occlusion, boundary truncation, and segmentation noise. The reference database is constructed from CAT-UXO [7], a collection of idealised UXO images sourced from a community EOD Knowledge Hub. Each mask undergoes a sequence of operations: connected component analysis is applied while retaining only the largest foreground region, morphological closing with a disc-shaped structuring element of radius 3 is used to fill small gaps, the orientation is aligned to the principal axis by PCA using an isotropy threshold κ , profiles are obtained through percentile-based radial estimation (p_5, p_{95}) with N bins, and a validation mask is derived from a minimum occupancy threshold m combined with a rejection factor α .

We mention that we used the original train/validation/test split of the datasets, as defined by their authors.

3.4 Unfolded Architecture

At the core of our method lies a deep unfolded network that refines an axial profile from noisy observations. Given an observed profile r_{obs} and the current estimate $p^{(t)}$, each iteration applies a learnable direct mapping:

$$p^{(t+1)} = \Pi_{[0,1]} \left(g_{\phi_t} \left(p^{(t)}, r_{\text{obs}}, v \odot (r_{\text{obs}} - p^{(t)}) \right) \right), \quad (5)$$

In Equation 5, $\Pi_{[0,1]}$ denotes element-wise clipping to $[0, 1]$, and $g_{\phi_t} : \mathbb{R}^{3N} \rightarrow \mathbb{R}^N$ is an MLP whose input is the concatenation of $p^{(t)}$, r_{obs} , and $v \odot (r_{\text{obs}} - p^{(t)})$, each of dimension N (concretely, $\mathbb{R}^{192} \rightarrow \mathbb{R}^{64}$), with architecture $192 \rightarrow 128 \rightarrow 64 \rightarrow 32 \rightarrow 64$ (ReLU hidden layers, linear output). Each iteration uses its own parameters ϕ_t (unshared weights). In implementation, each linear layer is spectral-normalised and scaled by a spectral bound $s = 0.95$, constraining the Lipschitz constant of each block and thereby limiting the sensitivity of unfolded iterates to input perturbations. The mapping from the latent noise-free profile p^* to the observed profile can be conceptually written as $r_{\text{obs}} = \mathcal{A}(p^*) + \eta$, where $\mathcal{A}$ encapsulates the effects of segmentation and occlusion that lack a tractable closed form. Each step g_{ϕ_t} implicitly approximates the corresponding inverse refinement operator, replacing a fixed gradient or proximal step with a data-driven mapping while retaining the iterative structure as an inductive bias. The initial profile estimate is set to $p^{(0)} = \bar{r}_{\text{train}}$, the mean normalised profile computed over the CTX-UXO training set; the same $p^{(0)}$ is used for both field queries and reference shapes. After the final iteration, we compute the full parameter vector deterministically as $\hat{\theta} = h(p^{(T)}, L/D)$, where h is the parameter extraction function defined in Eq. 4 and L/D is computed directly from the aligned mask. This enforces consistency between scalar descriptors and the reconstructed profile by construction. The validity weighting ensures that updates are driven only by regions where reliable observations exist, preventing error propagation from truncated or occluded portions of the object.

3.5 Training Objectives

The network is trained using only field segmentation masks, without class labels or paired cross-domain examples. The primary objective encourages the predicted profile $\hat{p} = p^{(T)}$ to match observations where valid:

$$\mathcal{L}_{\text{reconstruction}} = \frac{1}{\max(1, \sum_i v_i)} \sum_{i=1}^N v_i \cdot \rho_{\delta} (r_{\text{obs}}(z_i) - \hat{p}(z_i)), \quad (6)$$

In Equation 6, ρ_{δ} denotes the Huber loss, whose parameter δ is chosen as the empirical standard deviation of the residuals $|r_{\text{obs}} - p^{(T)}|$ computed on clean reference shapes. Note that the reconstruction target in training stage is the observed profile itself, rather than a noise-free reference. Denoising arises [6] from three complementary factors: (i) the constrained capacity of the MLP bottleneck (32 dimensions), which compels the network to represent profiles on a

low-dimensional manifold of plausible solutions instead of memorising noise; (ii) the validity mask, which excludes unreliable bins from the loss computation; and (iii) the consistency loss (described below), which promotes robustness to perturbations and mitigates overfitting to noise specific to individual observations. To improve robustness to augmentation, we enforce consistency between estimates from augmented views of the same mask:

$$\mathcal{L}_{\text{cons}} = \left\| \hat{\theta}_{\text{prof}}(M) - \hat{\theta}_{\text{prof}}(\mathcal{T}(M)) \right\|_2^2, \quad (7)$$

In Equation 7, $\hat{\theta}_{\text{prof}}$ represents θ with both the L/D term and the inflection count n_{infl} removed, the L/D term because truncation is not length-invariant, and n_{infl} because it is non-differentiable and thus cannot support gradient backpropagation. Before PCA alignment and profile extraction, $\mathcal{T}$ applies random local perturbations: truncation up to f_{trunc} , erosion/dilation with kernel size $k \in \mathcal{K}$, and Gaussian boundary noise of scale σ_n .

The final training objective is:

$$\mathcal{L} = \mathcal{L}_{\text{reconstruction}} + \mathcal{L}_{\text{cons}} \quad (8)$$

Weights can be assigned to each value of the losses involved depending on the objective of the researchers. At every unfolding step, the update relies on the current profile, the observation, and a residual that reflects its validity, while the output is clipped to the physical interval $[0, 1]$.

3.6 Retrieval Protocol

At inference time, for a query field mask M_q , we first obtain its profile estimate $p^{(T)}(M_q)$, then compute its parameter vector $\hat{\theta}_q = h(p^{(T)}(M_q), (L/D)_q)$ using the trained unfolded network. CAT-UXO reference shapes are processed identically (connected-component filtering, morphological closing, PCA alignment, profile extraction), then encoded by feeding their observed profiles through the same T -step unfolded network with shared trained weights and the same initialisation $p^{(0)} = \bar{r}_{\text{train}}$, followed by the same deterministic mapping h . Consequently, query and reference embeddings lie in a common latent space.

Retrieval is performed via nearest-neighbour search in a standardised parameter space. Let μ_θ and σ_θ denote the mean and standard deviation of θ calculated in component terms on the CTX-UXO training set. Let $\mathcal{R}$ denote the reference database and define $\bar{\theta}_\ell = (\theta_\ell - \mu_{\theta,\ell})/\sigma_{\theta,\ell}$ for each component ℓ . Retrieval is then carried out according to:

$$\text{match}(M_q) = \arg \min_{j \in \mathcal{R}} \sum_{\ell=1}^d (\bar{\theta}_{q,\ell} - \bar{\theta}_{j,\ell}^{\text{ref}})^2 \quad (9)$$

4 Experiments

4.1 Evaluation Methodology

We evaluate all methods under a unified retrieval protocol, where embeddings are compared via nearest-neighbour search in the reference database. We used the CTX-UXO [8] test subset for evaluation and CAT-UXO (UXORD-10K) [5,7] as a reference database. For a complete analysis, we compare against several alternative methods (baseline). First, to obtain a non-learnt geometric baseline, we apply Direct Shape Matching [32], which uses classical Procrustes alignment [12] followed by computing point-set distances on contours extracted with marching squares and uniformly resampled at 256 points. Second, we employ Prototypical Networks [30], a popular few-shot learning method, which constructs class prototypes from 5 reference examples per class and classifies CTX-UXO test queries based on the Euclidean distance to these prototypes. The feature extractor is trained on the CTX-UXO training split. Third, we consider a Siamese network [21] using a shared ResNet-50 backbone trained with contrastive loss on CTX-UXO training pairs. Finally, we employ a CNN with cosine-similarity-based retrieval, using a pretrained ResNet-50 with fine-tuning on the CTX-UXO training split. For methods requiring RGB input (Prototypical Networks, Siamese, CNN+Cosine), we use masked RGB crops. In this way, we construct a representative benchmark comprising the principal methods that can be used for retrieval in the context of UXO mitigation. We acknowledge that other methods and models may also be considered in the future, such as those based on Foundation Model architectures (Transformer or Mamba). However, we emphasise that a life-saving solution should not be constrained by computational costs. Moreover, to ensure a fair comparison, we employ different approaches from current State-Of-The-Art (RGB and shape-based) that are feasible for retrieval, specifically in order to highlight the advantages of shape analysis. In the Future Works section, we detail the directions we intend to pursue in our studies.

4.2 Metrics

We report multiple retrieval metrics following standard information retrieval evaluation practices [10,15]. We quantify performance using Recall@K (R@K), which represents the fraction of queries in which at least one correct-class reference appears in the top results K . We report R@1, R@3, and R@5 to capture both precision-critical and recall-orientated scenarios. To assess ranking efficiency, we employ the mean reciprocal rank (MRR), where a higher value indicates that correct matches appear earlier in the ranking, defined as

$$\text{MRR} = \frac{1}{|Q|} \sum_{q \in Q} \frac{1}{\text{rank}(q)}, \quad (10)$$

In Equation 10, rank is the position of the first correct-class retrieval. Finally, we employ Normalised Discounted Cumulative Gain (NDCG@10), a rank-sensitive

metric that assigns higher importance to correct retrievals at top positions, formulated as

$$\text{NDCG@}K = \frac{\text{DCG@}K}{\text{IDCG@}K} \quad (11)$$

For Equation 11, Discounted Cumulative Gain (DCG) is defined as:

$$\text{DCG@}K = \sum_{i=1}^K \frac{\text{rel}_i}{\log_2(i+1)}, \quad (12)$$

In Eq. 12, $\text{rel}_i = 1$ if the i -th retrieved item is of the correct class, and IDCG represents the ideal DCG obtained when all correct items are ranked first. We also report the Domain Shift Score (DSS), defined as the mean Euclidean distance in the standardised parameter space between each query and its nearest correct-class reference. For this metric (DSS), lower values indicate better cross-domain alignment. All values are expressed as mean $\pm$ standard deviation, calculated over 10 independent experimental runs using the same hyperparameters. All hyperparameters are available in the public repository, as described in the Reproducibility section at the end of the paper.

4.3 Results

Figure 2 shows a successful retrieval case and known limitations. Tables 1 and 2 report the cross-domain retrieval performance on the CTX-UXO test set. Our unfolded method attains a Recall@1 of 62.1%, surpassing the strongest Recall@1 baseline (Direct Shape Matching, $53.4 \pm 0.4\%$) by 8.7 percentage points and maintaining the highest scores across all ranking metrics. Notably, even the Direct Shape Matching baseline outperforms traditional RGB-based techniques, indicating that munition shape is a more robust cue than appearance for heterogeneous datasets.

Table 1: Cross-domain retrieval recall (%).

Method	R@1	R@3	R@5
<i>Classical Shape Method</i>			
Direct Shape Matching	53.4 $\pm$ 0.4	61.2 $\pm$ 0.5	65.4 $\pm$ 0.4
<i>Few-Shot & Metric Learning</i>			
Prototypical Networks [†]	20.5 $\pm$ 1.4	-	-
Siamese Network	46.9 $\pm$ 1.8	72.8 $\pm$ 1.5	80.2 $\pm$ 1.2
CNN + Cosine	31.1 $\pm$ 1.5	42.5 $\pm$ 1.4	46.3 $\pm$ 1.3
<i>Ours</i>			
Unfolded Shape Network	62.1$\pm$1.2	77.3$\pm$1.6	82.1$\pm$1.3

[†] Prototypical Networks assigns each query to the nearest class prototype. All top- K retrievals belong to the same predicted class. Prototype-based classification is not well suited for fine-grained cross-domain retrieval, motivating the need for instance-level embedding methods such as the proposed approach.



(a) Qualitative example of when the proposed architecture worked in a difficult case (an oxidised object, partially covered by vegetation). Type of UXO: Mortar Bomb

(b) Limitations: Objects that share closely resembling shapes, typically ones belonging to the same class, are prone to being mistaken for one another. Type of UXO: Grenade

Fig. 2: For each case, on the left is the test instance (CTX-UXO) and on the right is the reference matched instance (CAT-UXO). As a known limitation, the system may partially fail in granular identification when the shapes of UXO subtypes are very similar. For example, an offensive grenade (with grooves) can easily be confused with a defensive grenade (smooth on the surface), as they share a similar ovoid shape.

Table 2: Cross-domain ranking and alignment quality

Method	MRR	NDCG@10	DSS↓
Direct Shape Matching	0.590±0.004	0.529±0.005	1.318±0.008
Prototypical Networks	0.206±0.014	0.205±0.013	1.159±0.015
Siamese Network	0.619±0.016	0.438±0.014	1.162±0.018
CNN + Cosine	0.429±0.013	0.301±0.012	1.463±0.019
Unfolded Shape Network (Ours)	0.712±0.012	0.592±0.010	1.112±0.014

4.4 Ablation Study

We conduct ablation experiments to analyse the contribution of each implemented component, namely architecture depth and shape-parameter groups.

Figure 3 shows the effect of varying the number of steps (depth). All three metrics increase up to $T = 5$, after which performance saturates and then gradually declines, suggesting that excessive depth introduces over-refinement. Table 3 evaluates the contribution of each parameter group. The largest drop appears when the L/D ratio is removed, followed by Fourier descriptors. In contrast, removing the axial profile and inflexion count causes only minor changes, suggesting that these features provide complementary information.

Figure 4 shows UMAP visualisations [20] of the Grenade embeddings for different unfolding depths. Increased depth leads to tighter clustering of field queries around their reference counterparts.

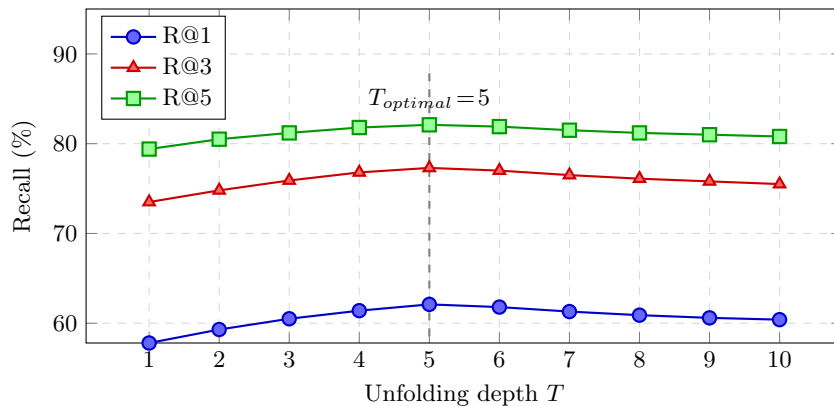
Table 3: Ablation on shape parameters

Configuration	R@1	R@3	R@5	MRR
Full model ($d=75$)	62.1$\pm$1.2	77.3$\pm$1.6	82.1$\pm$1.3	0.712$\pm$0.012
w/o L/D ratio	55.2 $\pm$ 1.4	73.6 $\pm$ 1.8	79.8 $\pm$ 1.5	0.672 $\pm$ 0.014
w/o Taper ratio	59.4 $\pm$ 1.3	75.8 $\pm$ 1.7	81.0 $\pm$ 1.4	0.697 $\pm$ 0.013
w/o Fourier desc.	58.1 $\pm$ 1.3	74.9 $\pm$ 1.7	80.5 $\pm$ 1.4	0.689 $\pm$ 0.013
w/o Axial profile	61.4 $\pm$ 1.2	76.8 $\pm$ 1.6	81.7 $\pm$ 1.3	0.707 $\pm$ 0.012
w/o Inflection count	61.8 $\pm$ 1.2	77.0 $\pm$ 1.6	81.9 $\pm$ 1.3	0.710 $\pm$ 0.012

5 Conclusion

We introduced a deep unfolded neural framework for cross-domain UXO retrieval that addresses the fundamental real-world challenge of matching degraded field observations against idealised technical references, with exemplification on UnEXPloded Ordnance (UXO). Key insights from applying the architecture to UXO are: (1) shape-based representations yield more reliable cross-domain transfer than fully appearance-based features; (2) the unfolding structure allows for effective processing of incomplete observations; and (3) a compact parametric model is sufficient to distinguish the main ordnance categories. From an operational standpoint, the proposed approach helps humanitarian UXO clearance efforts by offering automated decision support for UXO recognition. Rather than committing to a single label, operators can examine the top- K retrieved references to guide their judgement.

Human in the Loop Recommendation. This system should be deployed as a decision support system (DSS) rather than an autonomous identification system. The retrieval paradigm serves as a vital safeguard when human cognitive

Fig. 3: Impact of unfolding depth T on retrieval recall.

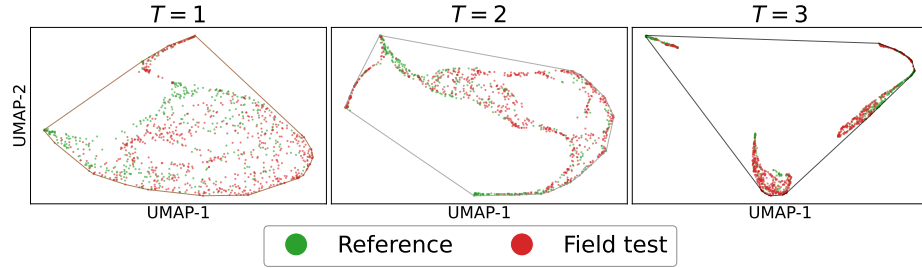


Fig. 4: UMAP visualisation of Grenade embeddings, for different depths (T).

performance is compromised by fatigue, high stress environments, or uncertainty in decision making.

Limitations. The dependence on hyperparameters requires careful tuning. However, a sensitivity analysis on the most influential parameters (unfolding depth T) is provided in Section 4.4. We acknowledge that only three classes (Projectile, Mortar Bomb, Grenade) currently have sufficient representation in both CTX-UXO and CAT-UXO for cross-domain retrieval. Extending the evaluation to additional types of ordnance requires expanding and consolidating the available reference databases. Additionally, the current pipeline assumes access to reasonably accurate segmentation masks. Limitations in shape-only identification are exemplified in Figure 2 from Section 4.

Future Work. We plan to integrate foundation models in the form of Vision Language Models to derive visual primitives (attributes), such as generic shape descriptors, and to apply neurosymbolic approaches. The unfolding framework can be used on its own, with the main immediate benefit being its reduced computational usage (a checkpoint of only 2.3 MB). We will address all limitations, as well as extend the study to additional architectures, in future work.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

Reproducibility

Reproducibility Statement. The notation and initial resources are detailed in Section 3.1. The parameter settings are given in Section 3.2. The preprocessing procedure is described in Section 3.3. Training objectives are described in Section 3.5. We elaborate on the retrieval protocol in Section 3.6. Evaluation follows standard retrieval methodologies, as discussed in Section 4.2. As a life-saving solution should not be restricted, we release the full implementation code, default hyperparameter configurations, model weights, and evaluation scripts in a public (anonymous) repository: https://anonymous.4open.science/r/UXO_DeepUnfolded-8881/. Both datasets are publicly available for research purposes under their respective licences.

Datasets are available at:

- **CTX-UXO**: <https://zenodo.org/records/17052675>
- **CAT-UXO**: <https://m4d.iti.gr/uxord-10k-dataset/> or official website <https://cat-uxo.com/>

Acknowledgments. We gratefully acknowledge the financial publication support provided by the University of Science and Technology POLITEHNICA of Bucharest through the PubArt program.

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Ablin, P., Moreau, T., Massias, M., Gramfort, A.: Learning step sizes for unfolded sparse coding. *Advances in Neural Information Processing Systems* **32** (2019)
2. Barone, P.M.: Bombed archaeology: Towards a precise identification and a safe management of wwii’s dangerous unexploded bombs. *Heritage* **2**(4), 2704–2711 (2019). <https://doi.org/10.3390/heritage2040167>, <https://www.mdpi.com/2571-9408/2/4/167>
3. Baur, J.: Towards Operational UAS-Based Landmine Detection: Vegetation, Validation, and Geophysical Application. Ph.D. thesis, Columbia University (2025)
4. Baur, J., Steinberg, G., Nikulin Ph D, A., Chiu Ph D, K., de Smet Ph D, T.: How to implement drones and machine learning to reduce time, costs, and dangers associated with landmine detection. *The Journal of Conventional Weapons Destruction* **25**(1), 29 (2021)
5. Begkas, G., Giannakeris, P., Ioannidis, K., Kalpakis, G., Tsikrika, T., Vrochidis, S., Kompatsiaris, I.: Automatic visual recognition of unexploded ordnances using supervised deep learning. In: *Association for Computing Machinery*. p. 286–294. ICMR ’22, New York, NY, USA (2022). <https://doi.org/10.1145/3512527.3531383>, <https://doi.org/10.1145/3512527.3531383>
6. Burger, H.C., Schuler, C.J., Harmeling, S.: Image denoising with multi-layer perceptrons, part 1: comparison with existing algorithms and with bounds. *CoRR* **abs/1211.1544** (2012), <http://arxiv.org/abs/1211.1544>
7. CAT-UXO: Cat-uxo — collective awareness to exploded ordnance. <https://cat-uxo.com/> (2026), <https://cat-uxo.com/>, accessed: February 23, 2026
8. Craioveanu, G.M., Stamatescu, G.: Ctx-uxo: A comprehensive dataset for detection and identification of unexploded ordnances (2024). <https://doi.org/10.21227/cwnm-de53>, <https://dx.doi.org/10.21227/cwnm-de53>
9. Craioveanu, M.G., Stamatescu, G.: Detection and identification of unexploded ordnance using a two-step deep learning methodology. In: *2024 32nd Mediterranean Conference on Control and Automation (MED)*. pp. 257–262 (2024). <https://doi.org/10.1109/MED61351.2024.10566207>
10. Deselaers, T., Keysers, D., Ney, H.: Features for image retrieval: an experimental comparison. *Information retrieval* **11**(2), 77–107 (2008)
11. Dryden, I.L., Mardia, K.V.: *Statistical shape analysis: with applications in R*. John Wiley & Sons (2016)

12. Goodall, C.: Procrustes methods in the statistical analysis of shape. *Journal of the Royal Statistical Society: Series B (Methodological)* **53**(2), 285–321 (1991)
13. Heiderscheidt, D.: The impact of world war one on the forests and soils of europe. *Ursidae: The Undergraduate Research Journal at the University of Northern Colorado* **7**(3), 3 (2018)
14. Iijima, L., Giakoumoglou, N., Stathaki, T.: A multimodal approach for cross-domain image retrieval. *arXiv preprint arXiv:2403.15152* (2024)
15. Jung, D., Kang, D., Kwak, S., Cho, M.: Few-shot metric learning: Online adaptation of embedding for retrieval. In: *Proceedings of the Asian Conference on Computer Vision*. pp. 1875–1891 (2022)
16. Kokkinos, I., Bronstein, M.M., Litman, R., Bronstein, A.M.: Intrinsic shape context descriptors for deformable shapes. In: *2012 IEEE Conference on Computer Vision and Pattern Recognition*. pp. 159–166. IEEE (2012)
17. Lacroix, P., Herzog, J., Eriksson, D., Weibel, R.: Methods for visualizing the explosive remnants of war. *Applied Geography* **41**, 179–194 (2013). <https://doi.org/https://doi.org/10.1016/j.apgeog.2013.04.007>, <https://www.sciencedirect.com/science/article/pii/S0143622813001021>
18. Lekhak, S., Ientilucci, E.J., Dera, D., Ghosh, S.: Uncertainty quantification in surface landmines and uxo classification using mc dropout. In: *arXiv* (2025), <https://arxiv.org/abs/2510.06238>, arXiv:2510.06238v1
19. Maser, E., Strehse, J.S.: “don’t blast”: blast-in-place (bip) operations of dumped world war munitions in the oceans significantly increase hazards to the environment and the human seafood consumer. *Archives of toxicology* **94**(6), 1941–1953 (2020)
20. McInnes, L., Healy, J., Melville, J.: Umap: Uniform manifold approximation and projection for dimension reduction (2020), <https://arxiv.org/abs/1802.03426>
21. Melekhov, I., Kannala, J., Rahtu, E.: Siamese network features for image matching. In: *2016 23rd international conference on pattern recognition (ICPR)*. pp. 378–383. IEEE (2016)
22. North Atlantic Treaty Organization: Glossary of Terms and Definitions. NATO Standardization Agency, modified version 02 edn. (aug 2000), updated 07.08.2000
23. Qi, C.R., Su, H., Mo, K., Guibas, L.J.: Pointnet: Deep learning on point sets for 3d classification and segmentation. *CoRR* **abs/1612.00593** (2016), <http://arxiv.org/abs/1612.00593>
24. Ravi, N., Gabeur, V., Hu, Y.T., Hu, R., Ryali, C., Ma, T., Khedr, H., Rädle, R., Rolland, C., Gustafson, L., Mintun, E., Pan, J., Alwala, K.V., Carion, N., Wu, C.Y., Girshick, R., Dollár, P., Feichtenhofer, C.: Sam 2: Segment anything in images and videos. *arXiv preprint arXiv:2408.00714* (2024), <https://arxiv.org/abs/2408.00714>
25. Samadzadegan, F., Qaderi, F., Haghighi, T.L., Toosi, A., Mousavi, S.M.: Unmanned aerial remote sensing for unexploded ordnance detection: Concepts, methods, current status and trends. *Sensing and Imaging* **26**(1), 137 (2025)
26. Ściegienka, P., Blachnik, M.: On the development of a neural network architecture for magnetometer-based uxo classification. *Applied Sciences* **15**(15), 8274 (2025)
27. Shakibajahromi, B., Kim, E., Breen, D.E.: Rimeshgnn: A rotation-invariant graph neural network for mesh classification. In: *2024 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. pp. 3138–3148 (2024)
28. Shlezinger, N., Segarra, S., Zhang, Y., Avrahami, D., Davidov, Z., Routtenberg, T., Eldar, Y.C.: Deep unfolding: Recent developments, theory, and design guidelines. *arXiv preprint arXiv:2512.03768* (2025)

29. Siméoni, O., Vo, H.V., Seitzer, M., Baldassarre, F., Oquab, M., Jose, C., Khalidov, V., Szafraniec, M., Yi, S., Ramamonjisoa, M., Massa, F.: Dinov3. arXiv preprint arXiv:2508.10104 (Aug 2025)
30. Snell, J., Swersky, K., Zemel, R.: Prototypical networks for few-shot learning. *Advances in neural information processing systems* **30** (2017)
31. United Nations Mine Action Service (UNMAS): United nations mine action service. <https://www.unmas.org/> (2026), <https://www.unmas.org/>, accessed: 23 feb. 2026
32. Veltkamp, R.C., Hagedoorn, M.: State of the art in shape matching. *Principles of visual information retrieval* pp. 87–119 (2001)
33. Wang, X., Peng, D., Yan, M., Hu, P.: Correspondence-free domain alignment for unsupervised cross-domain image retrieval. *Proceedings of the AAAI Conference on Artificial Intelligence* **37**(8), 10200–10208 (2023)
34. Ward, C.M., Harguess, J., Hilton, C., Mediavilla, C., Sullivan, K., Watkins, R.: Deep learning for automatic ordnance recognition. In: *Geospatial Informatics IX*. vol. 10992, pp. 103–112. SPIE (2019)
35. Wei, X., Van Gorp, H., Gonzalez-Carabarin, L., Freedman, D., Eldar, Y.C., van Sloun, R.J.: Deep unfolding with normalizing flow priors for inverse problems. *IEEE Transactions on Signal Processing* **70**, 2962–2971 (2022)
36. Zhang, J., Chen, B., Xiong, R., Zhang, Y.: Physics-inspired compressive sensing: Beyond deep unrolling. *IEEE Signal Processing Magazine* **40**(1), 58–72 (2023)
37. Zhang, J., Yao, Y., Deng, B.: Fast and robust iterative closest point. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **44**(7), 3450–3466 (2021)