Deep Metric Learning Solutions

Thermal UAV Waterfowl Analysis

Introduction

This document presents a comprehensive theoretical analysis of deep metric learning applied to thermal UAV imagery of waterfowl. The analysis explores:

1. **Training dynamics of the triplet loss function** and the impact of margin selection
2. **Out-of-distribution detection methods** using the learned embedding space

The solutions demonstrate how metric learning can effectively distinguish between waterfowl species and identify non-bird objects (vegetation, artifacts, other wildlife) in thermal imagery for wildlife conservation applications.

Task 1: Training Dynamics of the Triplet Loss

1.1 Impact of Margin on Inter-Class Distance

Triplet Loss Function

The triplet loss is defined as:

$$\mathcal{L} = \max(0, d(a,p) - d(a,n) + m)$$

Where:

- **a**: anchor sample (reference thermal image)
- **p**: positive sample (same waterfowl species as anchor)
- **n**: negative sample (different waterfowl species)
- **d(x,y)**: Euclidean distance between embeddings of x and y
- **m**: margin parameter ($m > 0$)

Analytical Derivation

The triplet loss becomes zero when the constraint is satisfied:

$$d(a,p) - d(a,n) + m \leq 0$$

Rearranging this inequality:

$$d(a,n) \geq d(a,p) + m$$

Interpretation: The negative sample must be at least margin *m* farther from the anchor than the positive sample.

Condition for Zero Expected Loss

For the expected loss $\mathbb{E}[\mathscr{L}]$ to become zero, we require:

$$\mathbb{E}[d(a,n)] \geq \mathbb{E}[d(a,p)] + m$$

This can be rewritten as:

$$\mathbb{E}[d(a,n)] - \mathbb{E}[d(a,p)] \geq m$$

Effect on Inter-Class Distance

Let us define:

- **D_intra =** $\mathbb{E}$**[d(a,p)]**: average intra-class distance (within same species)
- **D_inter =** $\mathbb{E}$**[d(a,n)]**: average inter-class distance (between different species)

For convergence ($\mathbb{E}[\mathscr{L}] \to 0$), we require:

$$D\_inter \geq D\_intra + m$$

Key Insights

1. **Direct Control:** The margin *m* directly controls the minimum separation between classes in the embedding space.
2. **Proportional Effect:** Larger margins force the model to push different classes farther apart, requiring the average inter-class distance to exceed the average intra-class distance by at least *m*.
3. **Training Objective:** The triplet loss drives the embedding space to:
   - Pull similar samples (same species) together: minimize d(a,p)
   - Push dissimilar samples (different species) apart: maximize d(a,n)
   - Enforce separation: d(a,n) - d(a,p) ≥ m

Conclusion: The margin directly increases the minimum required inter-class distance by exactly *m*, making it a crucial hyperparameter for controlling class separation.

1.2 Extreme Margin Values and Training Dynamics

Case 1: Margin Too Large

Problem: When the margin exceeds the embedding space's capacity, training stagnates.

Definitions:

- **$d_{max} = \max\{d(a,n)\}$**: maximum possible inter-class distance in embedding space
- **$d_{min} = \min\{d(a,p)\}$**: minimum possible intra-class distance

Critical Condition:

If the margin is set such that:

$$m > d_{max} - d_{min}$$

Then even for the "hardest" triplet (closest negative, farthest positive):

$$d(a,p) - d(a,n) + m \geq d_{min} - d_{max} + m > 0$$

Mathematical Justification:

Since $d(a,p) \geq d_{min}$ and $d(a,n) \leq d_{max}$:

$$d(a,p) - d(a,n) + m \geq d_{min} - d_{max} + m$$

If $m > d_{max} - d_{min}$, then:

$$d_{min} - d_{max} + m > d_{min} - d_{max} + (d_{max} - d_{min}) = 0$$

Therefore, $\mathscr{L} > 0$ for all possible triplets.

Consequences:

1. **Infeasible Constraint:** The condition $d(a,n) \geq d(a,p) + m$ becomes geometrically impossible to satisfy
2. **No Convergence:** Loss remains positive indefinitely ($\mathscr{L} > 0$ always)
3. **Training Stalls:** Gradients exist but optimization leads nowhere
4. **Practical Issue:** The embedding space cannot accommodate the required separation

Example: If embeddings are normalized to unit sphere ($\|e\| = 1$), then $d_{max} = 2$. If we set $m = 3$, the constraint can never be satisfied since $d(a,n) \leq 2 < 2 + 3$.

Case 2: Margin Too Small

Problem: When the margin is too small, the constraint is trivial to satisfy, preventing meaningful learning.

Trivial Satisfaction Scenario:

If:

$$m < \min\{d(a,n)\} - \max\{d(a,p)\}$$

Then for all triplets:

$$d(a,p) - d(a,n) + m < \max\{d(a,p)\} - \min\{d(a,n)\} + m < 0$$

This means $\mathscr{L} = 0$ for all triplets immediately, providing no learning signal.

Analysis:

If $m \approx 0$ (very small margin), the loss becomes:

$$\mathscr{L} = \max(0, d(a,p) - d(a,n))$$

This only requires $d(a,n) > d(a,p)$, with **no enforced separation**.

Why this prevents good convergence:

1. **Weak constraint**: Model only needs negatives slightly farther than positives
2. **Overlapping classes**: Classes can still overlap significantly
3. **No robustness**: Small perturbations can cause misclassification
4. **Numerical issues**: Optimization may oscillate near $d(a,n) \approx d(a,p)$

Formal statement:

For small $m < \varepsilon$ where ε is the numerical precision threshold:

Even if we achieve $d(a,n) = d(a,p) + \varepsilon$, this provides no meaningful separation. The loss will remain positive for many triplets because:

- Random noise: $\sigma_{noise} > m$ causes violations
- Optimization steps may overshoot the tiny margin
- Hard negatives near decision boundary always violate

The model oscillates without achieving stable, well-separated clusters.

Task 2: Using the Trained Embedding Space for OOD Detection

2.1 OOD Detection Method

Context: In waterfowl conservation, we need to distinguish between:

- **In-Distribution (ID):** Known waterfowl species (trained classes)
- **Out-of-Distribution (OOD):** Background vegetation, water reflections, debris, other animals

Step 1: Use Class Prototypes (Class Centers)

For each waterfowl species class *c*, compute the **class prototype**:

$$\mu_c = (1/N_c) \Sigma f(x_i)$$

This is simply the **average embedding** of that class.

Why prototypes? They represent the "central point" of each species cluster.

Step 2: Measure Distances to Prototypes

For any test sample *x*, we compute:

$$\text{Distance}(x) = \min_c \|f(x) - \mu_c\|_2$$

This tells us **how close the sample is to the nearest known species**.

Interpretation:

- **Small distance** → looks similar to some waterfowl species → in-distribution (ID)
- **Large distance** → far from all species → likely OOD

Step 3: Compute Class Radii

For each class:

$$r_c = \max_{x_i \in c} \|f(x_i) - \mu_c\|$$

This gives the **maximum distance of real ID samples from their prototype**.

We can think of each class as a "ball" with radius r_c.

Step 4: Define OOD Score

Simple OOD score:

$$\text{OOD\_score}(x) = \min_c \|f(x) - \mu_c\|$$

Optional more robust version:

$$\text{OOD\_score}(x) = \min_c \left(\|f(x) - \mu_c\| - r_c \right)$$

This accounts for the "size" of each class cluster.

Step 5: Decision Rule (Thresholding)

Choose a threshold τ, then:

$$\text{If OOD\_score}(x) > \tau, \text{ classify as OOD}$$

How to choose τ?

- **Option A:** $\tau = \max(r_c) + \varepsilon$
- **Option B:** Use validation set to choose threshold that gives good accuracy

Summary of the Method

1. Compute the center (prototype) of each species
2. Measure how far a test sample is from the nearest prototype
3. If it is **too far**, it is considered **not a waterfowl** → OOD
4. If it is **close enough**, we treat it as a known species → ID

This method is simple, fast, and works because metric learning already separated species in the embedding space.

2.2 Why OOD Samples Are Farther Away (Proof)

We now justify **why this method works** when the embedding space is well-learned.

Assumptions (Geometric Picture)

We assume:

1. Each species forms a compact cluster

For class *c*:

$$C_c = \{z : \|z - \mu_c\| \leq r_c\}$$

So each class is like a **solid ball** around its prototype.

2. Clusters are well separated

For any two classes $c \neq c'$:

$$\|\mu_c - \mu_{c'}\| > r_c + r_{c'} + \text{margin}$$

This means class-balls **do not touch or overlap**.

3. ID samples fall inside their cluster

If x is waterfowl of class c:

$$f(x) \in C_c \implies \|f(x) - \mu_c\| \leq r_c$$

4. OOD samples lie outside all class clusters

If x_ood is not a waterfowl:

$$f(x_{ood}) \notin \bigcup_c C_c$$

Thus, for the nearest class c*:

$$\|f(x_{ood}) - \mu_{c^*}\| > r_{c^*}$$

Proof

Goal: Show OOD distance is always **bigger** than ID distance.

Let:

- x_id = any ID sample from class c
- x_ood = any OOD sample
- c* = class whose prototype is closest to OOD sample

For ID sample:

$$\min_c \|f(x_{id}) - \mu_c\| = \|f(x_{id}) - \mu_c\| \leq r_c$$

For OOD sample:

Because it lies outside all clusters:

$$\min_c \|f(x_{ood}) - \mu_c\| = \|f(x_{ood}) - \mu_{c^*}\| > r_{c^*}$$

Combine the two inequalities:

$$\|f(x_{ood}) - \mu_{c^*}\| > r_{c^*} \geq \min_c \|f(x_{id}) - \mu_c\|$$

Thus:

$$\min_c \|f(x_{ood}) - \mu_c\| > \min_c \|f(x_{id}) - \mu_c\|$$

Interpretation

- In-distribution samples lie **inside** their species cluster → **small distance**
- OOD samples lie **outside all clusters → large distance**

Therefore:

OOD samples must always be farther from every prototype than ID samples.

This is why a simple distance-based threshold is enough to detect OOD.

Conclusion

This analysis demonstrates that:

1. The margin parameter in triplet loss directly controls class separation in the embedding space
2. Extreme margin values (too large or too small) prevent effective learning
3. Well-trained metric learning models naturally enable OOD detection through distance-based methods
4. The geometric structure of the embedding space provides theoretical guarantees for OOD detection performance