# Potential Field Based Deep Metric Learning

Shubhang Bhatnagar, Narendra Ahuja

## Problem

**Learn a representation where semantic similarity encoded as distance**

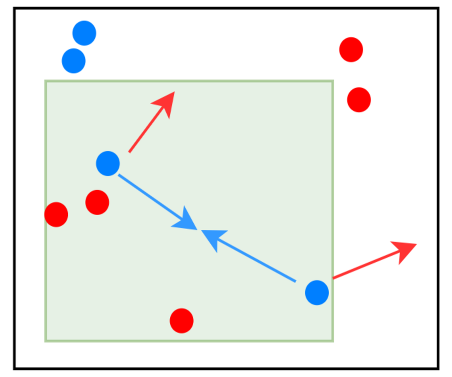
### Disadvantages of Current Loss functions

1) **Pair Based** e.g. triplet, contrastive, n-tuplet loss

- Models only a subset of interactions
- Disregards overall representation space

⇒

- Reduced feature quality
- Susceptible to label noise (useful information discarded)
- Need high complexity sample mining strategies to work (O ($N^3$))



2) **Proxy Based** e.g. Proxy Anchor, ProxyNCA

Less effective as no sample-sample comparison information is used

## Proposed Solution: PFML

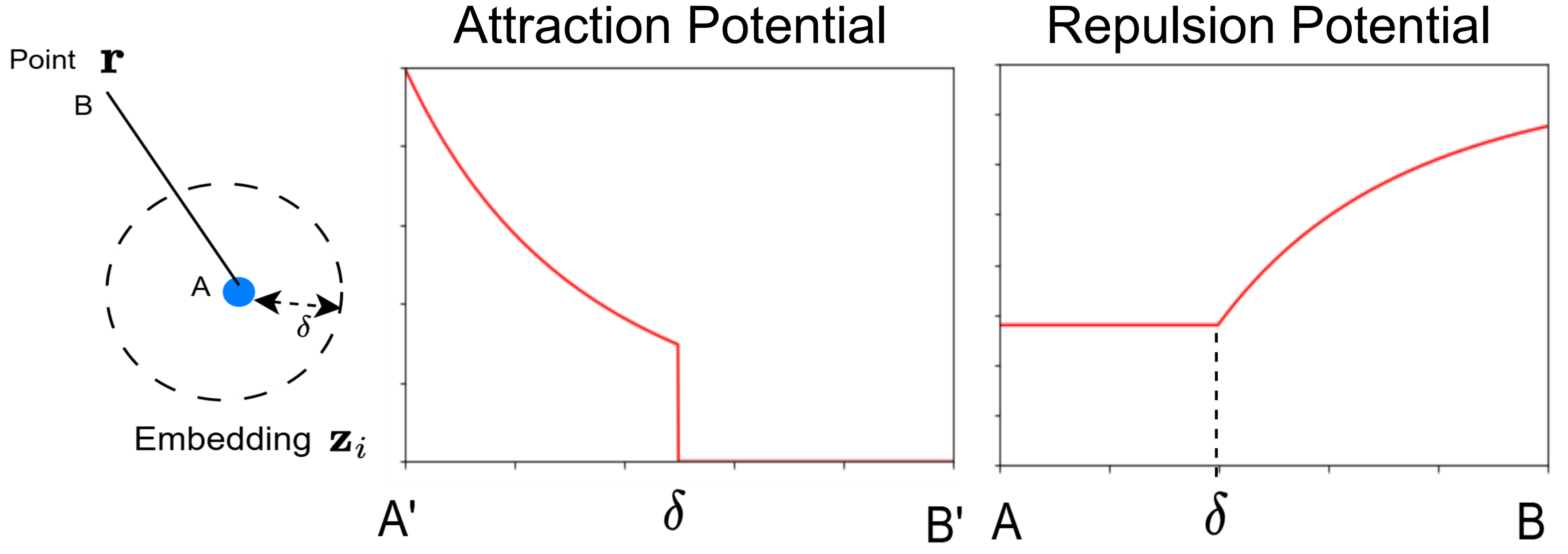
Use a **continuous potential field** to directly model **all** interactions

**Adv #1 No complex mining needed**

**Adv #2 Better features** (all Interactions modeled)

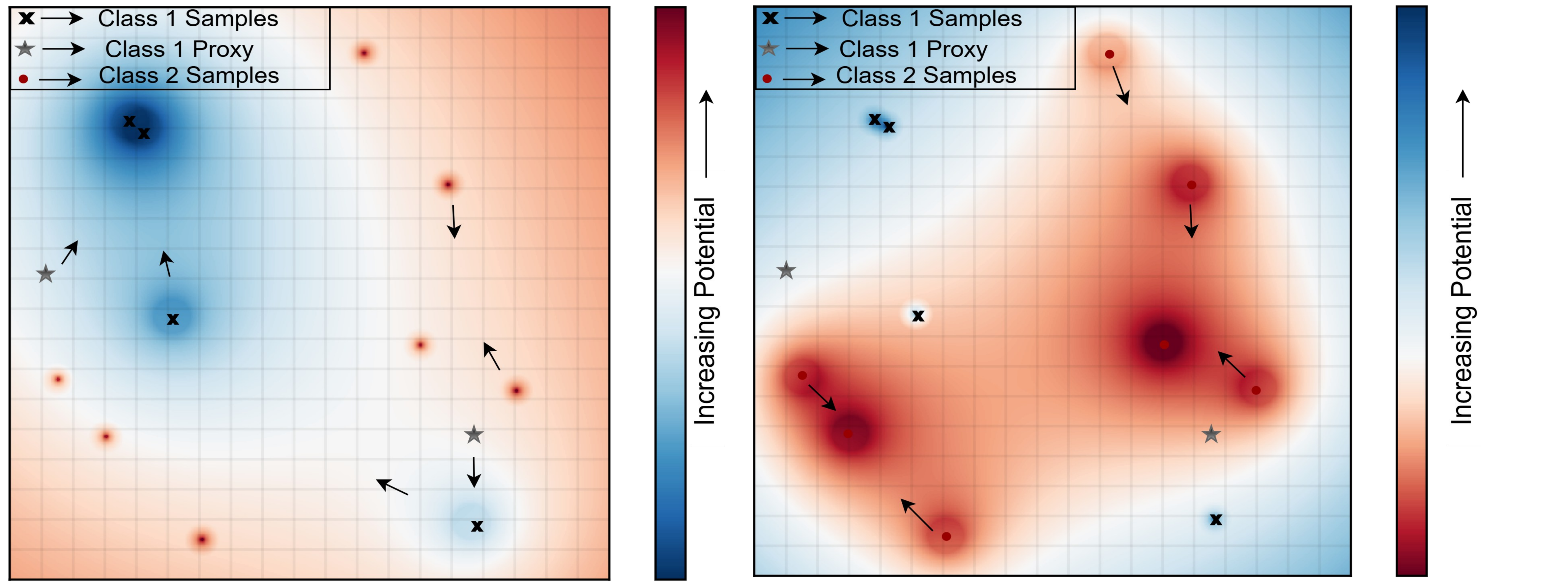
Each point induces **attractive & repulsive potentials**

**Big Change:** Interaction strength (force) decays with distance




## Example Potential fields $\Psi_1$ and $\Psi_2$ by PFML for a 2 class problem

(a) Class Potential field $\Psi_1$ (b) Class Potential field $\Psi_2$

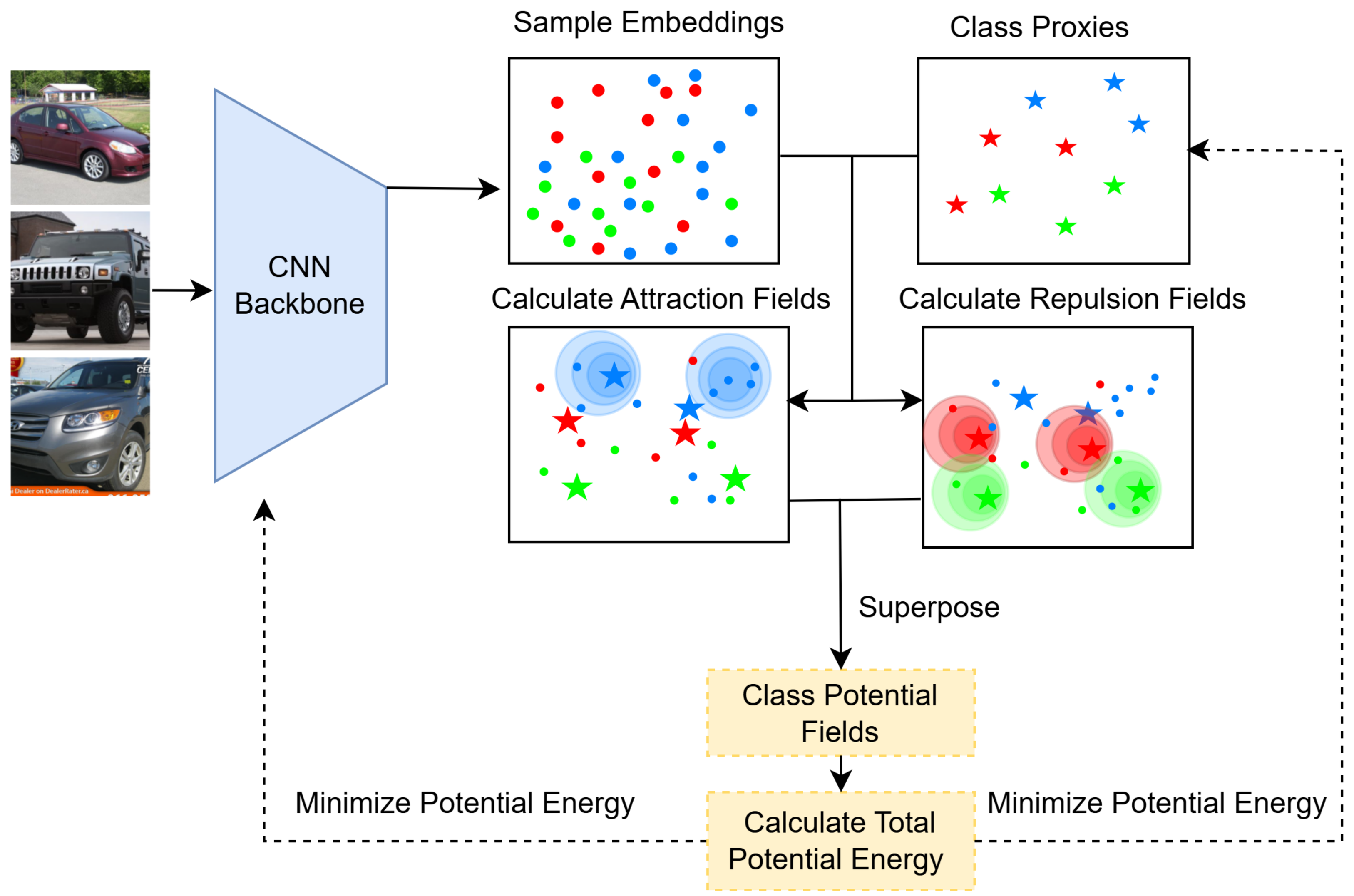



Class potential field ($\Psi$) = Superposition of individual fields, models net force applied **on** embeddings

**Net force on embedding of Class N = Gradient of Class N potential field $\Psi_N$ at its location**

**Adv. #3 Label Noise Resilience & Intra-Class Feature Preservation.** The decaying strength of interaction ensures $\Psi_1$ and $\Psi_2$ draw embeddings towards nearby same class embeddings, unlike past methods that push them toward distant class embeddings, which may be very different variants of the class or a mislabeled point

## Overall Training Pipeline




PFML also uses proxies to augment embedding fields

**Adv #4 Better use of Proxies vs Previous Methods.** As they better represent data distribution due to PFML having decaying interaction strength

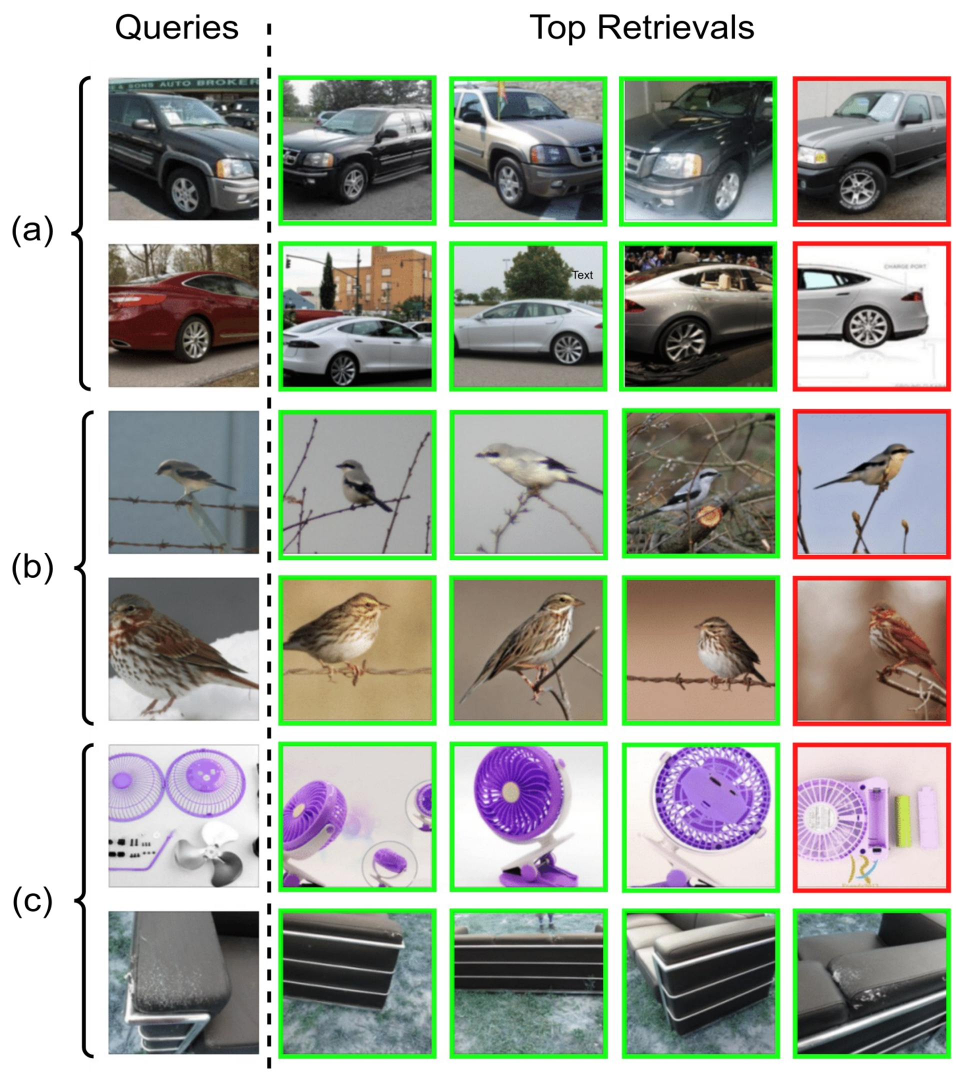
## Evaluation of Metric Space

SOTA for Zero Shot Image Retrieval

| Benchmarks → / Methods ↓ (Chronological) | CUB-200 R@1 | CUB-200 R@2 | Cars-196 R@1 | Cars-196 R@2 | SOP R@1 | SOP R@10 |
|---|---|---|---|---|---|---|
| **ResNet50 (512 dim)** | | | | | | |
| Proxy Anchor | 69.7 | 80.0 | 87.7 | 92.9 | - | - |
| MS+DAS | 69.2 | 79.2 | 87.8 | 93.1 | 80.6 | 91.8 |
| HIST | 71.4 | 81.1 | 89.6 | 93.9 | 81.4 | 92.0 |
| CPML (Sec. 3.4) | 68.3 | 78.7 | 85.2 | 91.5 | 79.4 | 90.7 |
| **Potential Field (Ours)** | **73.4** | **82.4** | **92.7** | **95.5** | **82.9** | **92.5** |
| **DINO (384 dim)** | | | | | | |
| DINO | 70.8 | 81.1 | 42.9 | 53.9 | 63.4 | 78.1 |
| Hyp | 80.9 | 87.6 | 89.2 | 94.1 | 85.1 | 94.4 |
| HIER | 81.1 | 88.2 | 91.3 | 95.2 | 85.7 | 94.6 |
| **Potential Field (Ours)** | **83.1** | **89.3** | **94.7** | **96.5** | **86.5** | **95.1** |
| **ViT (384 dim)** | | | | | | |
| ViT-S | 83.1 | 90.4 | 47.8 | 60.2 | 62.1 | 77.7 |
| Hyp | 85.6 | 91.4 | 86.5 | 92.1 | 85.9 | 94.9 |
| **Potential Field (Ours)** | **87.8** | **92.6** | **91.5** | **95.2** | **88.2** | **95.7** |

Much more robust to Label Noise (20% corruption)

| Benchmarks → / Methods | CUB-200 R@1 | CUB-200 R@2 | Cars-196 R@1 | Cars-196 R@2 |
|---|---|---|---|---|
| Multi Similarity | 58.9 | 71.8 | 70.4 | 79.8 |
| Proxy Anchor | 60.7 | 75.1 | 76.9 | 83.1 |
| HIST | 59.7 | 74.6 | 72.9 | 81.8 |
| **Potential Field** | **66.7** | **76.9** | **84.5** | **88.6** |




Qualitative Zero Shot Retrievals

**Scan To Know More**