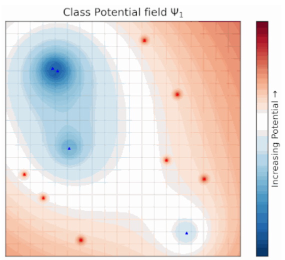
Potential Field Based Deep Metric Learning

Shubhang Bhatnagar, Narendra Ahuja
University of Illinois Urbana-Champaign



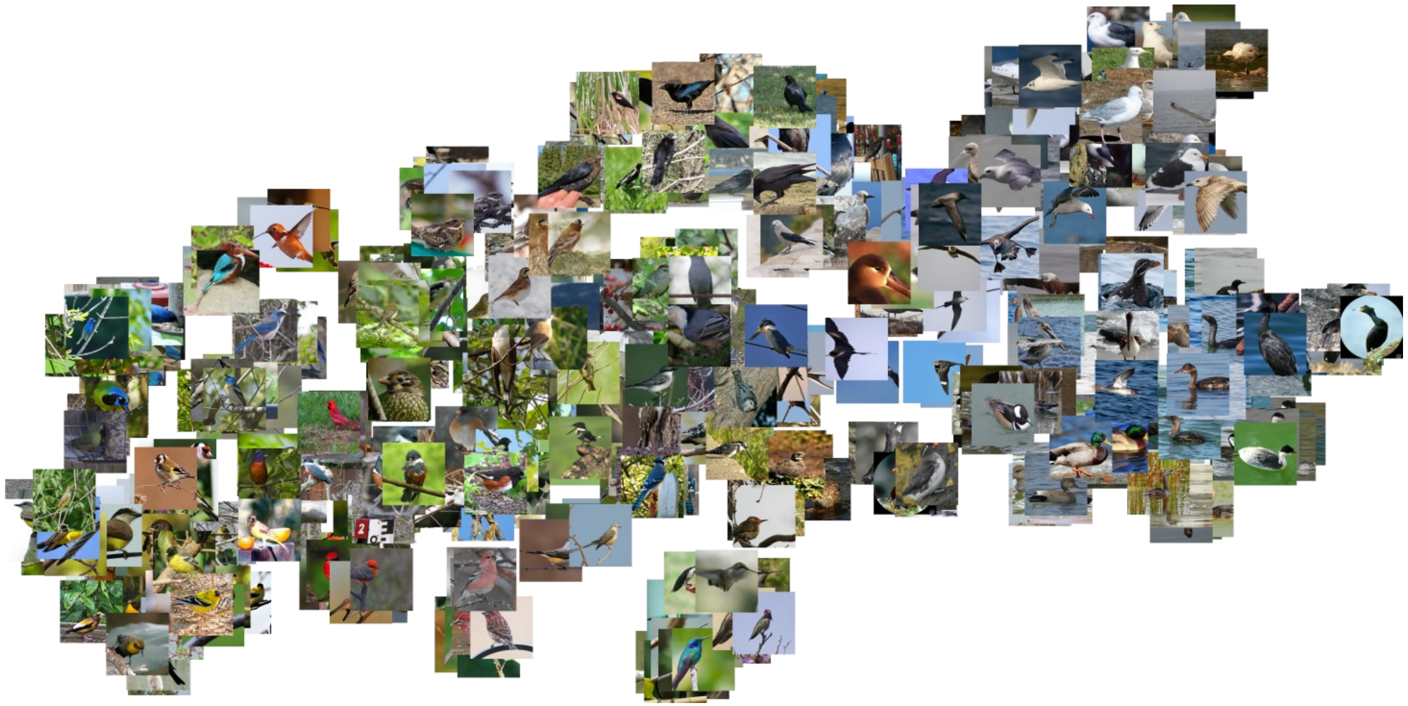
CVPR Nashville JUNE 11-15, 2025

ILLINOIS

Introduction: What is Deep Metric Learning ?

Problem

Learn a representation where semantic similarity encoded as distance
(distance in representation space = semantic dissimilarity)



A t-sne visualization of a semantic representation space learnt using PFML

ILLINOIS

Deep Metric Learning: Past Work & Motivation

Setting: No fine-grained similarity information between images available beyond class labels for training

Current Methods

Tuple Based

Sample tuple of embeddings

(pair of same class samples with rest being dissimilar examples)

Enforce :

Distance between samples of the same class **<** Distance between samples of different class

e.g. triplet, contrastive, n-tuplet loss

Proxy Based

Use class centers (proxies), moving samples of each class towards their respective proxies

e.g. Proxy Anchor, ProxyNCA

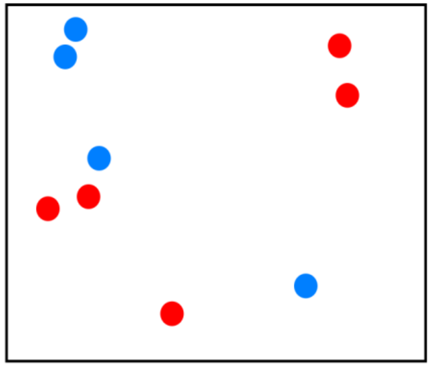
Deep Metric Learning: Past Work & Motivation

Disadvantages:

Tuple Based

Need high complexity sample mining strategies to work
($O(N^3)$ for triplets among N samples)

Disregards overall representation space, modelling only a subset of interactions → Reduced feature quality and more susceptibility to noise



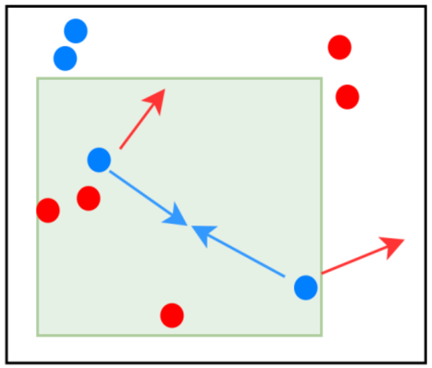
Deep Metric Learning: Past Work & Motivation

Disadvantages:

Tuple Based

Need high complexity sample mining strategies to work
($O(N^3)$ for triplets among N samples)

Disregards overall representation space, modelling only a subset of interactions ⇒ Reduced feature quality and more susceptibility to noise



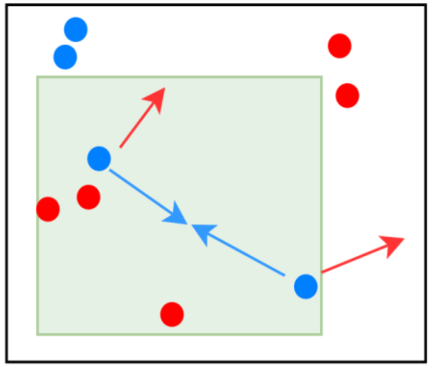
Deep Metric Learning: Past Work & Motivation

Disadvantages:

Tuple Based

Need high complexity sample mining strategies to work ($O(N^3)$ for triplets among N samples)

Disregards overall representation space, modelling only a subset of interactions ⇒ Reduced feature quality and more susceptibility to noise



Proxy Based

Less effective as no sample-sample comparison information is used

Can we better model interactions between the whole embedding space ?

ILLINOIS

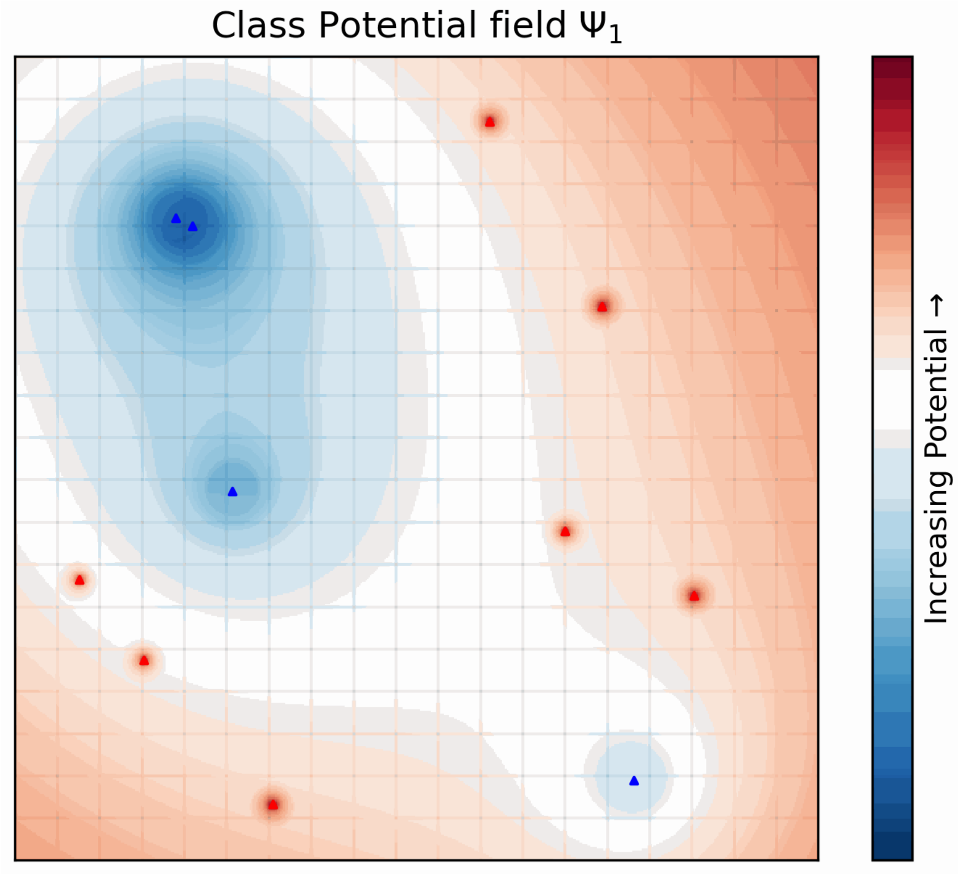
PFML: Potential Field Based Metric Learning

PFML models interactions between samples using a **continuous potential field**

Each point induces both:

- an **attractive potential** toward **same-class** examples
- a **repulsive potential** pushing away from **different-class** examples

This encourages **intra-class clustering** and **inter-class separation** in the learned representation space.



A toy problem with embeddings from 2 classes shown. Similar embeddings move together and dissimilar ones apart under the influence of our potential field

PFML: Potential Field Based Metric Learning

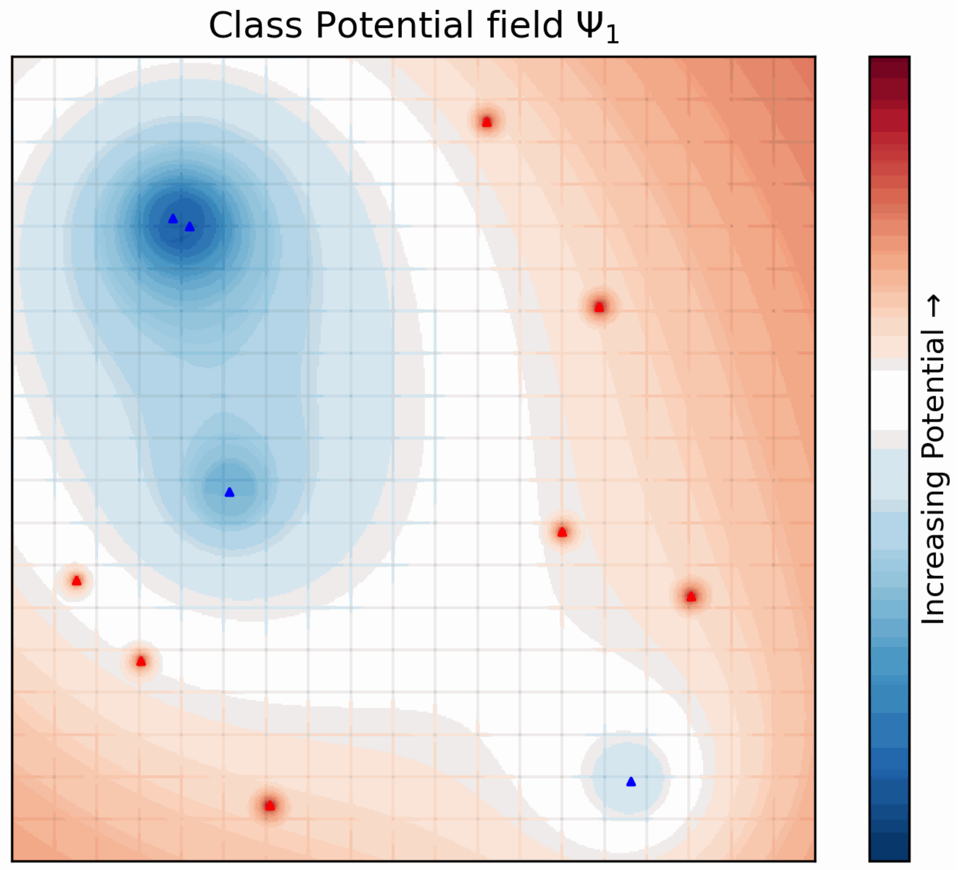
Class Potential Field and Net Force

•A **class potential field** Ψ_1 is the **superposition of fields** from all samples.

Ψ_1 models the **force acting on embeddings of class 1** due to:

- •Repulsion from all other class embeddings
- •Attraction from all same class embeddings

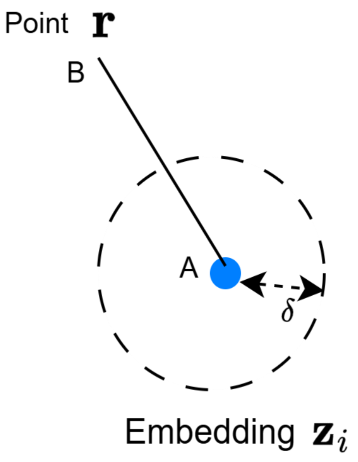
Net force on embedding = Gradient of class potential field Ψ_1 at embedding

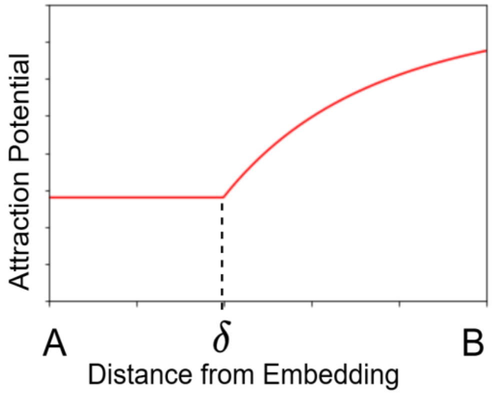


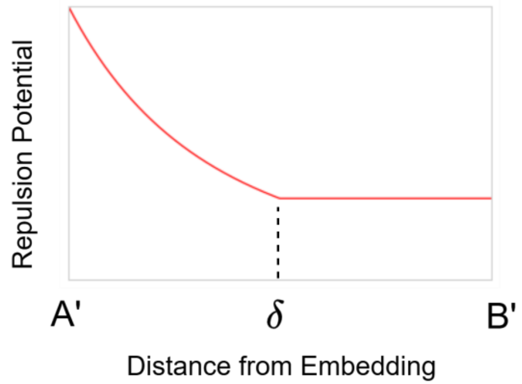
A toy problem with embeddings from 2 classes shown. Similar embeddings move together and dissimilar ones apart under the influence of our potential field

ILLINOIS

PFML: Functional Form of Individual Potential







Design of Individual potential field brings about a major design change

Decaying gradient (force) with distance

In contrast with all current works that have increasing or constant force with distance

a) It helps learning in presence of:
- Annotation/ Label noise
- Large intra-class variations

b) Superposition of potentials still yields local extremum near embeddings (Theoretically Proven in proposition 1 & Corollary 1 of the supplement)

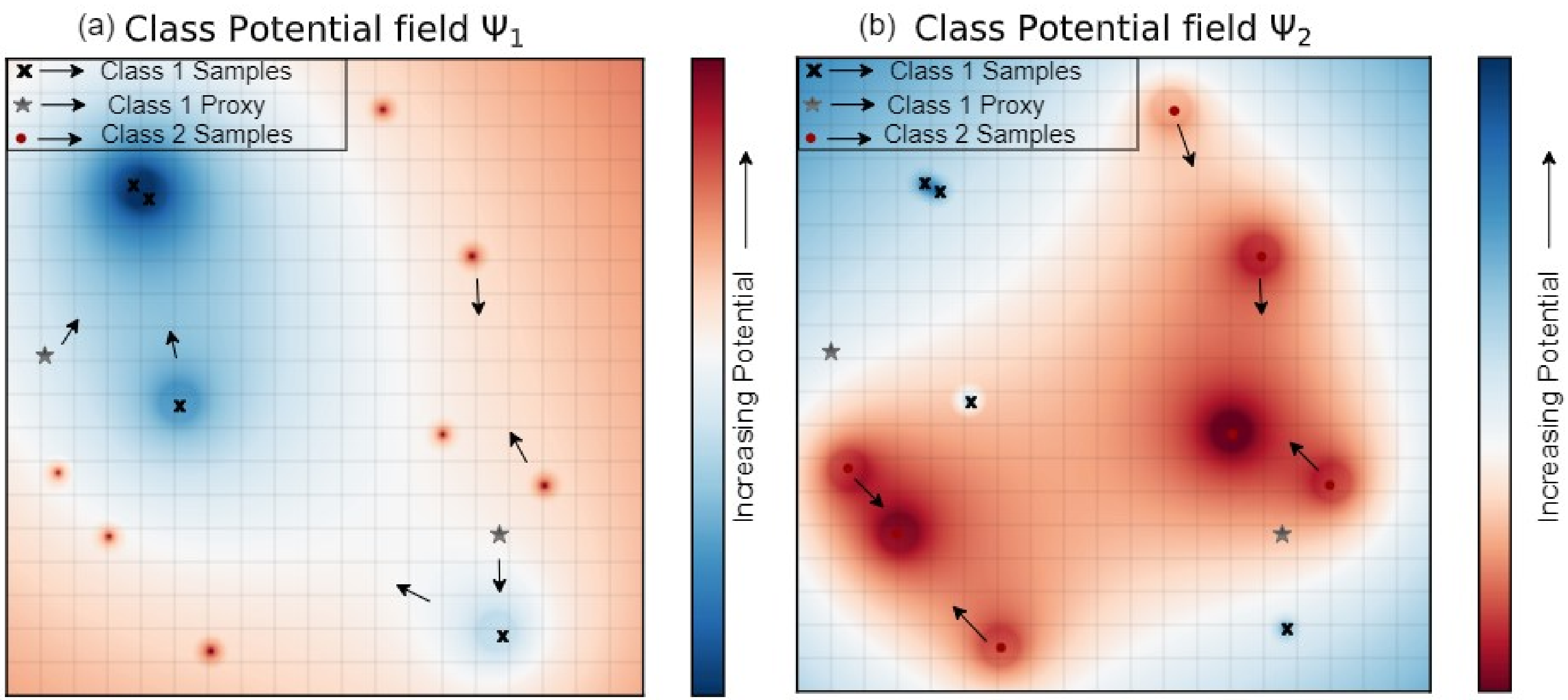
PFML: Approach

Proxies to augment Potential

- Proxies model the potential field due to out of batch embeddings
- They are trainable parameters affected by the field

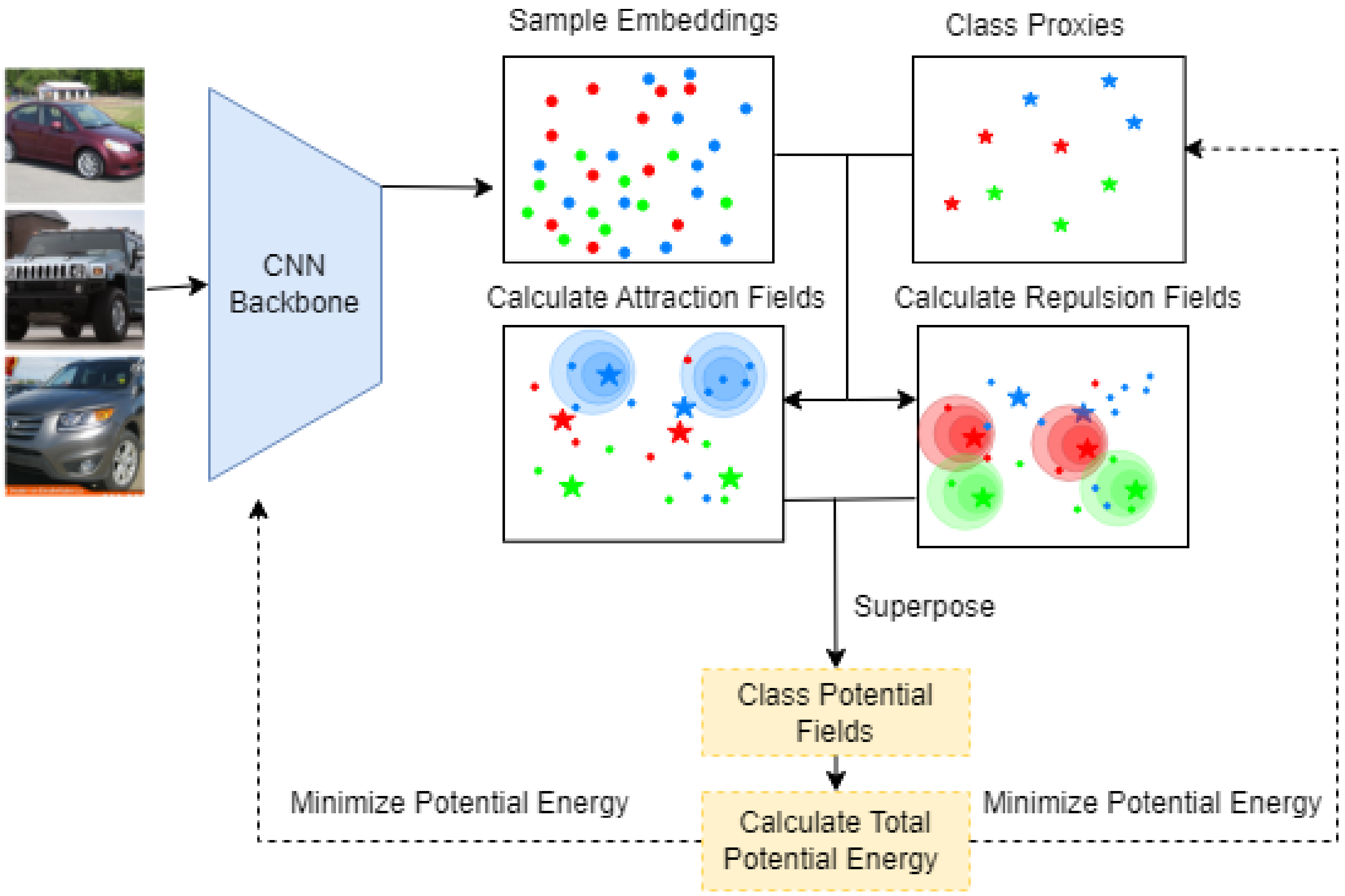
Training by minimizing total potential Energy

$$\mathcal{U} = \sum_{i=1}^{||\mathcal{B}||} \Psi_{y_i}(\mathbf{z}_i) + \sum_{j=1}^{N} \sum_{k=1}^{M} \Psi_j(\mathbf{p}_{j,k})$$



Example fields generated in a 2 class toy example.
Embeddings are drawn to nearest points from the same class

PFML: Summary



ILLINOIS

Results: Benchmarks

DML Benchmark: Zero-Shot Image Retrieval using 3 datasets

(1) Cars-196 dataset

(2) CUB-200-2011 dataset

(3) Stanford Online Products (SOP) dataset

Follows Standard Evaluation Setting : Half of the classes used for training, while remaining half are used for testing

Recall@K used for evaluation

Results: Image Retrieval

State-of-the-art R@K performance

- Gains of 2.1% and 2% R@1 on Cars and CUB over previous SOTA (HIST [1]),
- ≈ double the gains achieved by HIST [1] over SOTA before it
- Consistent gains using all backbones (ResNet, Vision transformers, Inception) on all 3 benchmarks

Benchmarks →	**CUB-200-2011**			**Cars-196**			**SOP**		
Methods ↓ (Chronological)	R@1	R@2	R@4	R@1	R@2	R@4	R@1	R@10	R@100
ResNet50 (512 dim)									
ESPHN [54]	64.9	75.3	83.5	82.7	89.3	93.0	78.3	90.7	96.3
N.Softmax [58]	61.3	73.9	83.5	84.2	90.4	94.4	78.2	90.6	96.2
DiVA [32]	69.2	79.3	-	87.6	92.9	-	79.6	91.2	-
Proxy NCA++ [46]	64.7	-	-	-	85.1	-	- 79.6	-	-
Proxy Anchor [21]	69.7	80.0	87.0	87.7	92.9	95.8	-	-	-
DCML-MDW [59]	68.4	77.9	86.1	85.2	91.8	96.0	79.8	90.8	95.8
MS+DAS [29]	69.2	79.2	87.1	87.8	93.1	96.0	80.6	91.8	96.7
HIST [28]	71.4	81.1	88.1	89.6	93.9	96.4	81.4	92.0	96.7
HIER[22]	70.1	79.4	86.9	88.2	93.0	95.6	80.2	91.5	96.6
HSE-PA [55]	70.6	80.1	87.1	89.6	93.8	96.0	80.0	91.4	96.3
CPML (Sec. 3.4)	68.3	78.7	86.2	85.2	91.5	95.2	79.4	90.7	96.1
Potential Field (Ours)	**73.4 ± 0.3**	**82.4± 0.1**	**88.8± 0.1**	**92.7± 0.3**	**95.5 ± 0.1**	**97.6± 0.1**	**82.9 ± 0.2**	**92.5 ± 0.2**	**96.8 ± 0.1**
BN Inception (512 dim)									
HTL [13]	57.1	68.8	78.7	81.4	88.0	92.7	74.8	88.3	94.8
MultiSimilarity [50]	65.7	77.0	86.3	84.1	90.4	94.0	78.2	90.5	96.0
SoftTriple [35]	65.4	76.4	84.5	84.5	90.7	94.5	78.3	90.3	95.9
CircleLoss [43]	66.7	77.4	86.2	83.4	89.8	94.1	78.3	90.5	96.1
DiVA [32]	66.8	77.7	-	84.1	90.7	-	78.1	90.6	-
ProxyGML [61]	66.6	77.6	86.4	85.5	91.8	95.3	78.0	90.6	96.2
Proxy Anchor [21]	68.4	79.2	86.8	86.1	91.7	95.0	79.1	90.8	96.2
DRML-PA [60]	68.7	78.6	86.3	86.9	92.1	95.2	71.5	85.2	93.0
MS+DAS [29]	67.1	78.11	86.4	85.7	91.6	95.3	78.2	90.3	96.0
HIST [28]	69.7	80.0	87.3	87.4	92.5	95.4	79.6	91.0	96.2
DFML-PA [48]	69.3	-	-	88.4	-	-	-	-	-
HSE-M [55]	67.6	78.0	85.8	82.0	88.9	93.3	-	-	-
PA+niV [23]	69.5	80.0	-	86.4	92.0	-	79.2	90.4	-
Potential Field (Ours)	**71.5± 0.3**	**81.2±0.2**	**88.3±0.2**	**90.1±0.2**	**93.9±0.1**	**96.3±0.1**	**80.6 ± 0.3**	**91.8 ± 0.1**	**96.4 ± 0.1**
DINO (384 dim)									
DINO [5]	70.8	81.1	88.8	42.9	53.9	64.2	63.4	78.1	88.3
Hyp [12]	80.9	87.6	92.4	89.2	94.1	96.7	85.1	94.4	97.8
HIER [22]	81.1	88.2	93.3	91.3	95.2	97.1	85.7	94.6	97.8
Potential Field (Ours)	**83.1 ± 0.3**	**89.3 ± 0.2**	**94.2 ± 0.1**	**94.7 ± 0.1**	**96.5 ± 0.1**	**97.8 ± 0.1**	**86.5 ± 0.3**	**95.1 ± 0.3**	**98.0 ± 0.2**
ViT (384 dim)									
ViT-S [11]	83.1	90.4	94.4	47.8	60.2	72.2	62.1	77.7	89.0
Hyp [12]	85.6	91.4	94.8	86.5	92.1	95.3	85.9	94.9	98.1
HIER [22]	85.7	91.3	94.4	88.3	93.2	96.1	86.1	95.0	98.0
Potential Field (Ours)	**87.8 ± 0.2**	**92.6 ± 0.2**	**95.7 ± 0.1**	**91.5 ± 0.3**	**95.2 ± 0.2**	**97.4 ± 0.1**	**88.2 ± 0.1**	**95.7 ± 0.1**	**98.6 ± 0.1**

[1] Lim, Jongin, Sangdoo Yun, Seulki Park, and Jin Young Choi. "Hypergraph-induced semantic tuplet loss for deep metric learning." In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 212-222. 2022.

Results: Noisy labels

DML Performance in Presence of Label Noise

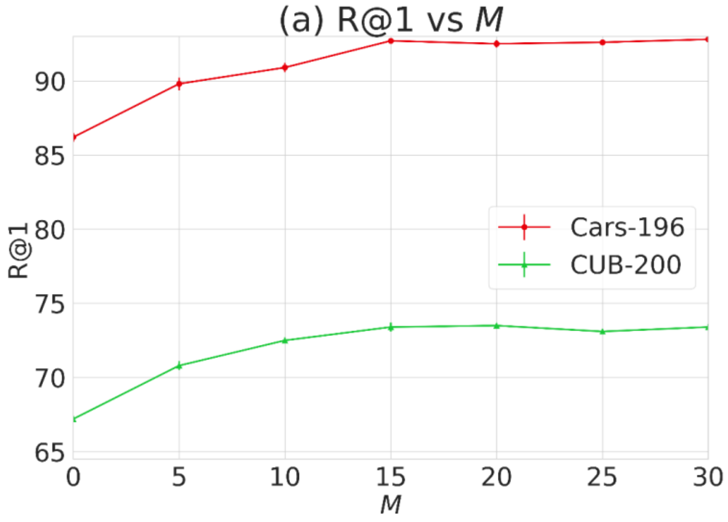
- Real world datasets often contain significant label noise
- Potential Field helps robustness in presence of label noise due to its
  1. Decay
  2. Modelling relations between all samples
- **Outperforms** SOTA by more than 6% and 7.6 % in terms of R@1 on CUB200 and Cars-196 in presence of 20% label noise

	CUB-200-2011		Cars-196	
Methods	R@1	R@2	R@1	R@2
Triplet[51]	55.1	68.7	67.5	77.9
MS [50]	58.9	71.8	70.4	79.8
PNCA[33]	60.1	74.7	74.3	82.4
PA [21]	60.7	75.1	76.9	83.1
HIST[27]	59.7	74.6	72.9	81.8
Ours	**66.7 ± 0.6**	**76.9± 0.3**	**84.5± 0.5**	**88.6± 0.3**

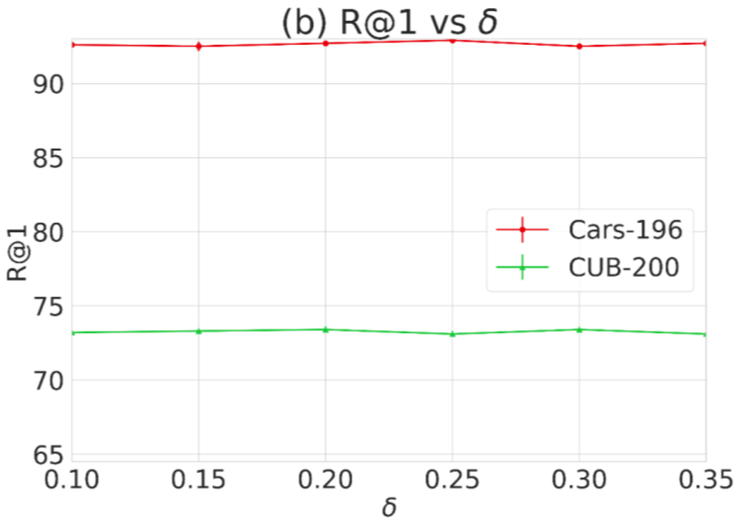
ILLINOIS

Results: Ablation

Trend of PFML Performance on 2 datasets



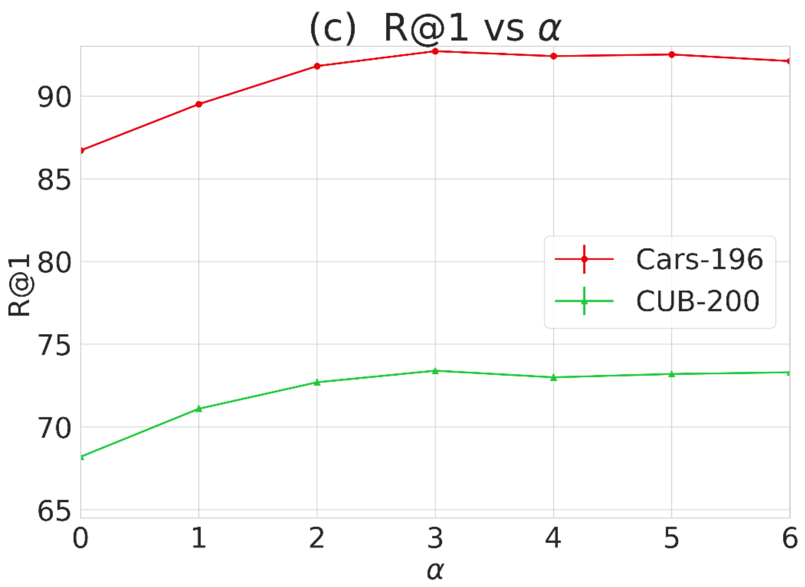
Number of Proxies (M)
Using more proxies boosts performance before saturating



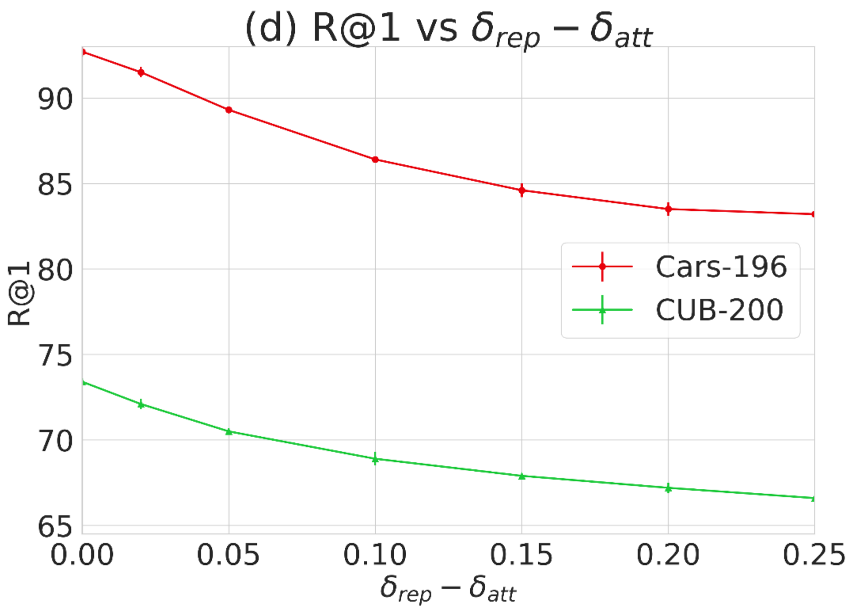
Radius Parameter delta
Doesn’t seem to significantly effect performance

Results: Ablation

Trend of PFML Performance on 2 datasets



Rate of Decay alpha
Stronger decay better than very mild decay



Is extending repulsion good ?
No, as it compressing intra-class features more than needed

ILLINOIS

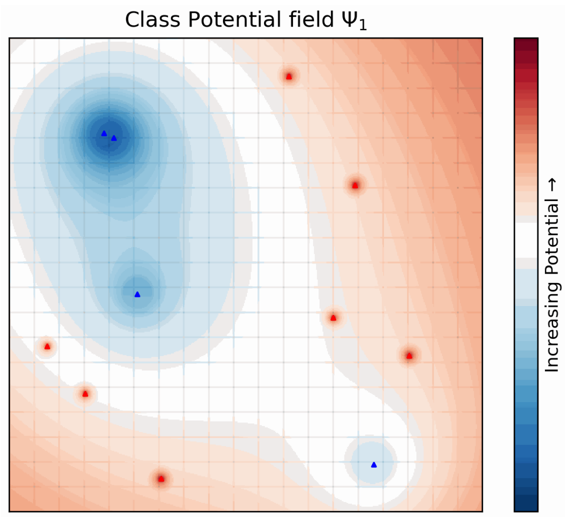
Conclusion

PFML

- Introduces a novel **continuous** potential field model for **capturing all-sample interactions**.

- Its **decaying with distance influence** model
  a) Significantly boosts **robustness to label noise** and
  b) Enhances **proxy alignment with training data**, leading to improved performance.

- Achieves SOTA (State-of-the-Art) performance in zero-shot image retrieval, both in the presence and absence of label noise.



Project Page