Densely-Anchored Sampling for Deep Metric Learning

Lizhao Liu[1,2], Shangxin Huang[1], Zhuangwei Zhuang[1], Ran Yang[1], Mingkui Tan[1,3], and Yaowei Wang[2]

[1]South China University of Technology [2]PengCheng Laboratory

[3]Key Laboratory of Big Data and Intelligent Robot, Ministry of Education

Problem Definition and Contribution

Goal: To train a deep model that projects semantically similar data into nearby embedding space, Deep Metric Learning (DML) methods often highly depends on sampling effective data from the embedding space.

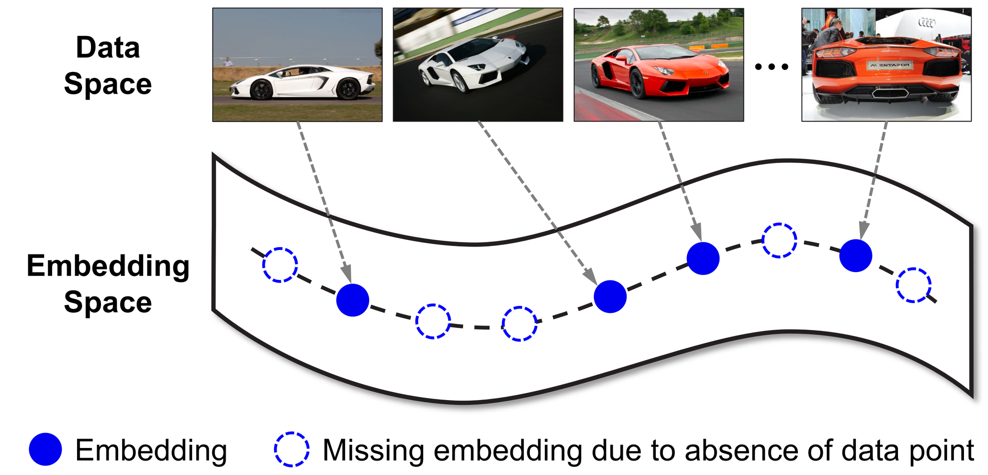
Key Contributions: A PnP Densely-Anchored Sampling (DAS) scheme for DML that

- exploits embeddings' nearby embedding space and densely produces embeddings.
- includes two models, namely Discriminative Feature Scaling (DFS) and Memorized Transformation Shifting (MTS).
- consistently improves existing DML baselines without bells and whistles.

Problem Formulation

Motivations:

- The basic hypothesis of metric learning: Embeddings close to each other in the embedding space have similar semantics.
- The missing embedding issue: The embedding space often has a barren area due to the absence of data points.



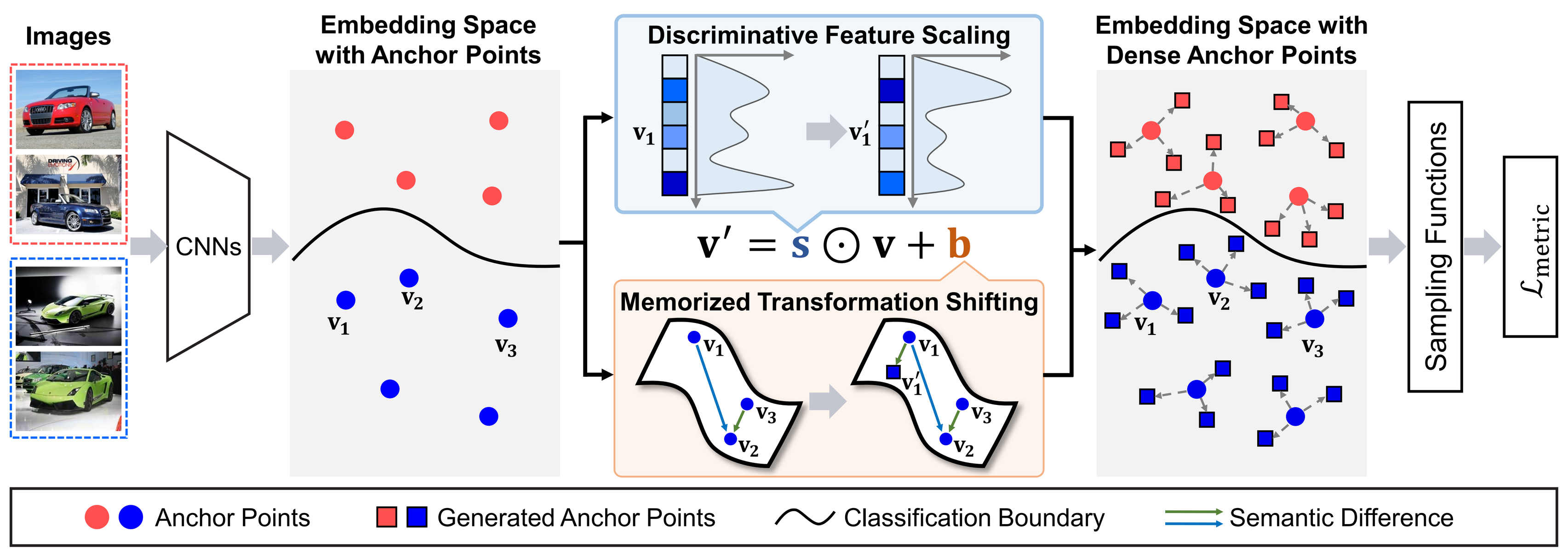
Main Idea: Considering the embeddings $\mathbf{v}$ with data points as anchor points, we alter anchor points semantics by semantic scaling and shifting to produce embedding $\mathbf{v}'$:

$$\mathbf{v}' = \mathrm{DAS}(\mathbf{v};\, \mathbf{s}, \mathbf{b}) = \underbrace{\mathbf{s} \odot}_{scaling} \mathbf{v} \underbrace{+\ \mathbf{b}}_{shifting}, \quad (1)$$

where $\mathbf{s}$ and $\mathbf{b}$ denote the semantic scaling and shifting factors, respectively.

Method

DAS comprises two parts, namely a discriminative feature scaling for producing semantic scaling factors, and a memorized transformation for producing semantic shifting factors.



Discriminative Feature Scaling:

$$\mathbf{s} = \boldsymbol{\gamma} \odot \mathbf{M}[y_{\mathbf{v}}] \ +\ \mathbf{1}^d \odot (1 - \mathbf{M}[y_{\mathbf{v}}]), \quad (2)$$

- $\mathbf{s}$ is the semantic scaling factor.
- $\mathbf{M} \in \mathbb{R}^{c \times d}$, a binary mask, records the frequently activated neurons for each class.
- $\boldsymbol{\gamma} \sim \mathrm{Uniform}[1 - r_s, 1 + r_s]^d$ and $r_s \in (0, 1)$ is a hyper-parameter.

Memorized Transformation Shifting:

$$\mathbf{b} = r_b \mathbf{t}, \quad \mathbf{t} \sim \{\mathbf{B}[y_{\mathbf{v}}, z] \mid z{=}1, 2, \ldots, Z\}. \quad (3)$$

- $\mathbf{b}$ is the semantic shifting factor.
- $\mathbf{B}$ is the memory bank that records the intra-class transformations according to the FIFO principle.
- $r_b \in (0, +\infty)$ is a hyper-parameter.

Experiments & Results

Dataset:

We use three popular benchmarks:

- CUB2011-200 (CUB), a fine-grained bird dataset, #Train / #Test classes: 100 / 100.
- CARS196 (CARS), a fine-grained vehicle dataset, #Train / #Test classes: 98 / 98.
- Stanford Online Products (SOP), a large-scale online products dataset, #Train / #Test classes: 11,318 / 11,316.

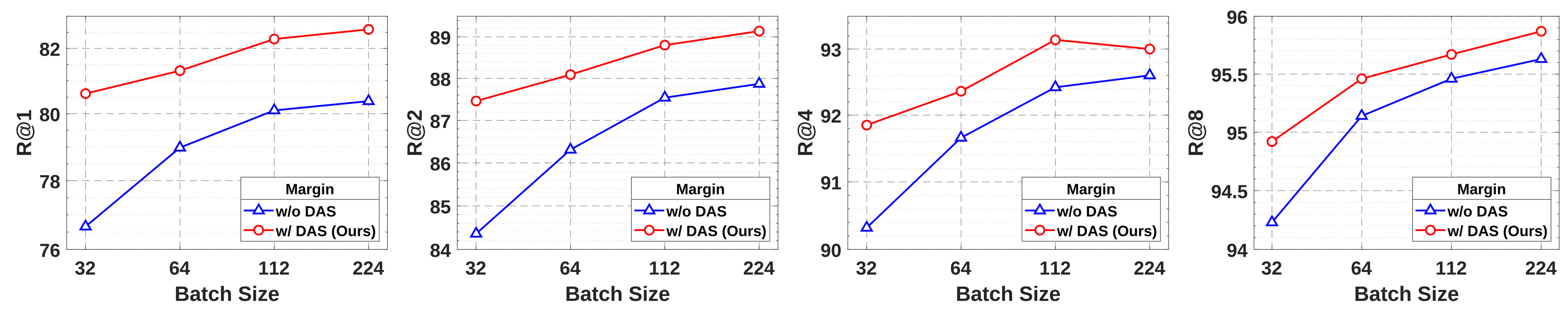
Improvements over Pair-based Methods:

Method	CUB			CARS			SOP		
	R@1	F1	NMI	R@1	F1	NMI	R@1	F1	NMI
Triplet [S]	60.25	32.82	64.64	74.64	31.98	63.22	73.51	33.47	89.33
Triplet [S] + DAS	**60.82**	**33.86**	**65.67**	**77.21**	**33.88**	**64.84**	**73.99**	**33.91**	**89.42**
Triplet [D]	62.68	36.39	67.03	78.86	35.80	65.85	77.54	37.10	90.05
Triplet [D] + DAS	**64.28**	**38.16**	**68.06**	**82.63**	**39.14**	**68.12**	**77.95**	**37.64**	**90.18**
Contrastive [D]	61.65	35.23	66.58	76.03	32.77	64.09	73.13	35.60	89.78
Contrastive [D] + DAS	**63.67**	**36.25**	**67.15**	**80.74**	**36.07**	**65.93**	**74.80**	**36.21**	**89.89**
Margin	62.61	37.33	67.58	80.10	37.85	67.15	78.69	39.20	90.50
Margin + DAS	**64.50**	**37.86**	**68.04**	**82.29**	**38.22**	**67.94**	**79.14**	**39.52**	**90.56**
GenLifted	58.81	34.64	65.50	72.45	32.43	64.00	76.18	37.26	90.13
GenLifted + DAS	**59.94**	**35.09**	**66.07**	**73.55**	**32.85**	**64.11**	**76.92**	**37.64**	**90.21**
N-Pair	60.55	36.94	67.19	77.35	36.26	66.74	77.71	37.13	90.15
N-Pair + DAS	**62.81**	**38.37**	**68.43**	**79.93**	**38.06**	**68.20**	**77.98**	**37.82**	**90.28**
MS	62.63	38.88	68.19	82.04	40.85	69.45	78.89	37.53	90.12
MS + DAS	**64.13**	**39.18**	**69.08**	**83.31**	**42.78**	**70.77**	**79.44**	**38.77**	**90.40**

Comparison with State-of-The-Arts:

Method	Backbone	CUB				CARS				SOP			
		R@1	R@2	R@4	R@8	R@1	R@2	R@4	R@8	R@1	R@10	R@100	R@1000
Margin	R^{128}	63.60	74.40	83.10	90.00	79.60	86.50	91.90	95.10	72.70	86.20	93.80	98.00
HDC	G^{384}	53.60	65.70	77.00	85.60	73.70	83.20	89.50	93.80	69.50	84.40	92.80	97.70
A-BIER	G^{384}	57.50	68.70	78.30	86.20	82.00	89.00	93.20	96.10	74.20	86.90	94.00	97.80
ABE	G^{512}	60.60	71.50	79.80	87.40	85.20	90.50	94.00	96.10	76.30	88.40	94.80	98.20
HTL	IBN^{512}	57.10	68.80	78.70	86.50	81.40	88.00	92.70	95.70	74.80	88.30	94.80	98.40
RLL-H	IBN^{512}	57.40	69.70	79.20	86.90	74.00	83.60	90.10	94.10	76.10	89.10	95.40	N/A
SoftTriple	IBN^{512}	65.40	76.40	84.50	90.40	84.50	90.70	94.50	96.90	78.30	90.30	95.90	N/A
MS	IBN^{512}	65.70	77.00	86.30	91.20	84.10	90.40	94.00	96.50	78.20	90.50	96.00	98.70
ProxyGML	IBN^{512}	66.60	77.60	86.40	N/A	85.50	91.80	95.30	N/A	78.00	90.60	96.20	N/A
ProxyAnchor	IBN^{512}	68.40	79.20	86.80	91.60	86.10	91.70	95.00	97.30	79.10	90.80	96.20	98.70
Contrastive + XBM	IBN^{512}	65.80	75.90	84.00	89.90	82.00	88.70	93.10	96.10	79.50	90.80	96.10	98.70
MS*	IBN^{512}	64.50	76.20	84.60	90.50	82.10	88.80	93.20	96.10	76.30	89.70	96.00	98.80
MS + EE*	IBN^{512}	65.10	76.80	86.10	91.00	82.70	89.20	93.80	96.40	77.00	89.50	96.00	98.80
ProxyAnchor + MemVir	IBN^{512}	69.00	79.20	86.80	91.60	86.70	92.00	95.20	97.40	79.70	91.00	96.30	98.60
MS†	IBN^{512}	65.72	77.19	85.74	91.56	83.86	90.41	94.64	96.99	76.89	89.58	95.59	98.60
MS + DAS (Ours)	IBN^{512}	67.07	78.11	86.43	91.88	85.66	91.60	95.27	97.37	78.16	90.26	95.99	98.76
MS†	R^{512}	66.46	77.28	85.85	91.69	83.99	90.39	94.51	96.80	79.53	91.06	96.30	98.83
MS + DAS (Ours)	R^{512}	**69.19**	**79.25**	**87.09**	**92.62**	**87.84**	**93.15**	**95.99**	**97.85**	**80.59**	**91.80**	**96.68**	**98.95**

Experiments on Different Batch Size:



Evolution of Training Process:

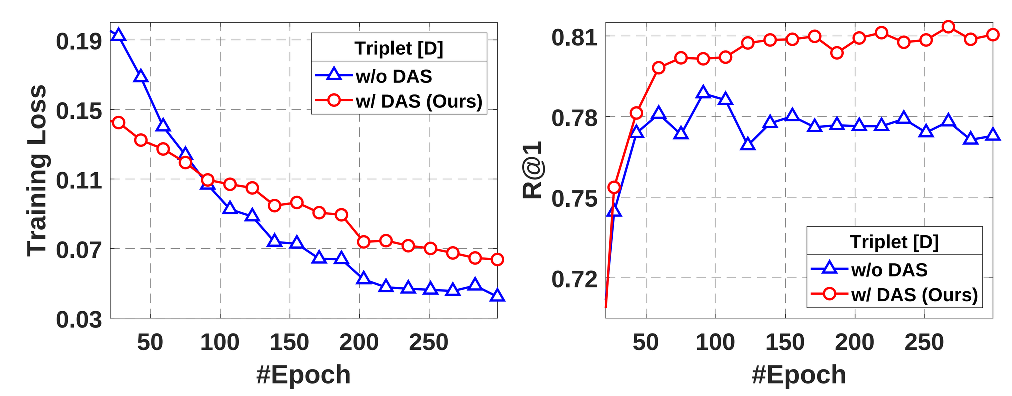


Image Retrieval Results:

Comparisons with Proxy-based Approaches:

Method	CUB			CARS			SOP		
	R@1	F1	NMI	R@1	F1	NMI	R@1	F1	NMI
Softmax	61.58	36.12	66.73	79.07	37.11	67.01	77.92	37.20	90.05
Softmax + DAS	**62.02**	**36.24**	**67.42**	**81.23**	**39.95**	**68.91**	**79.36**	**38.72**	**90.40**
ArcFace	61.56	35.73	66.83	79.50	37.75	67.82	78.08	37.79	90.18
ArcFace + DAS	**62.80**	**37.63**	**67.80**	**82.22**	**40.82**	**69.82**	**78.12**	**38.08**	**90.26**

More details: https://github.com/lizhaoliu-Lec/DAS

Contact: {selizhaoliu,sevtars}@mail.scut.edu.cn