Learning intra-batch connections for deep metric learning

Jenny Seidenschwarz, Ismail Elezi, Laura Leal-Taixé

Technical University of Munich

Deep Metric Learning (DML)

Goal: Learn distance functions over objects such that:

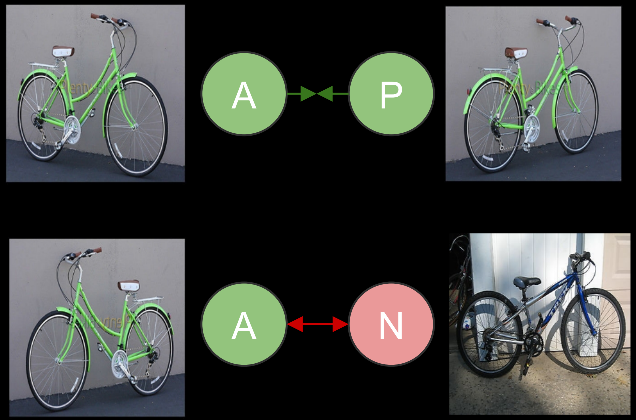
Low similarity score
(large distance)

High similarity score
(small distance)

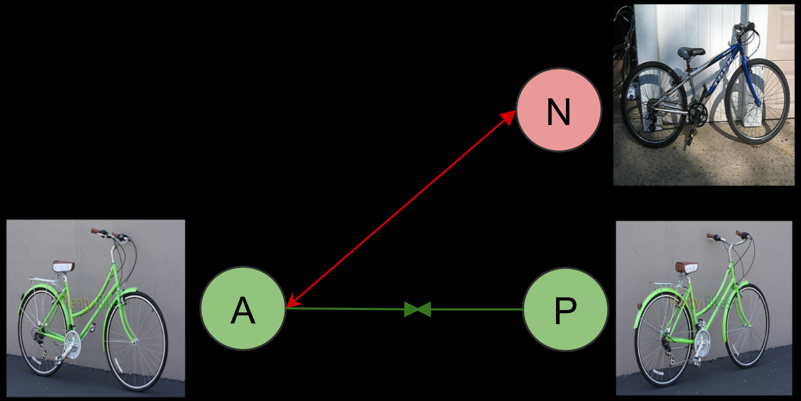
→ Test time: simple threshold or nearest neighbors

Deep Metric Learning (DML)

Contrastive loss



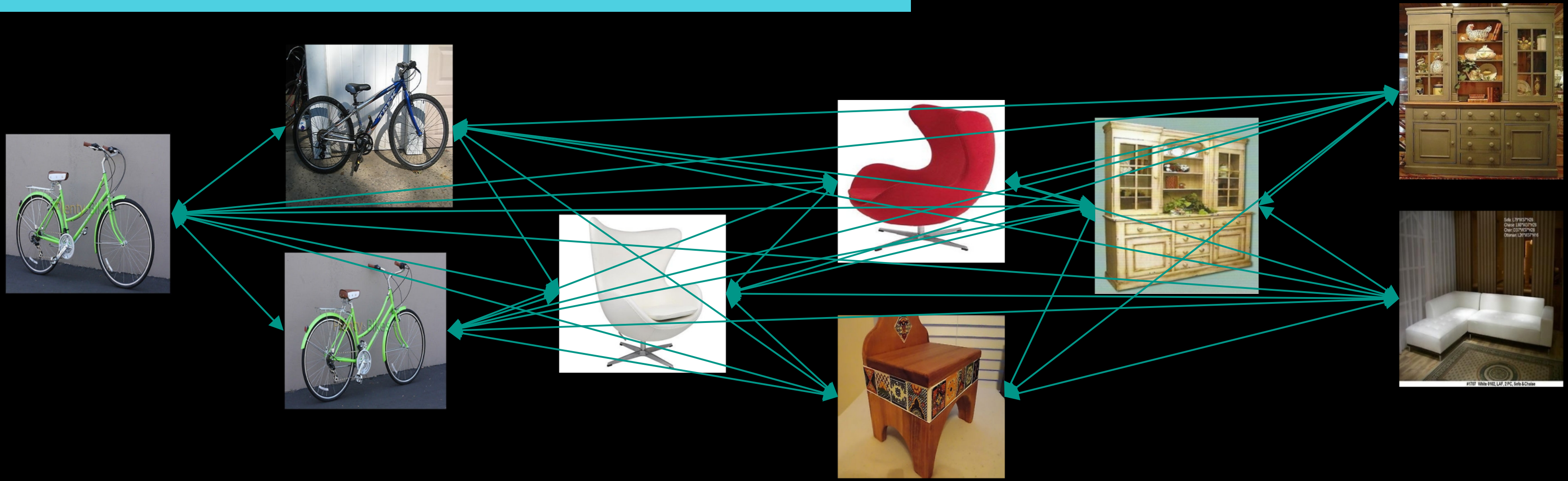
Triplet loss



- Dataset of $\mathbf{N}$ samples contains $\mathbf{O(N^2)}$ pairs and $\mathbf{O(N^3)}$ triplets
- Vast majority is uninformative which leads slow convergence

→ Special sampling techniques required

Training with contrastive and triplet loss



In a batch of size B contrastive and triplet loss only take $\frac{\mathrm{B}}{2}$ and $\frac{2\mathrm{B}}{3}$ relations into account, while there are B^2 relations in a batch

Could we take the global batch structure into account?

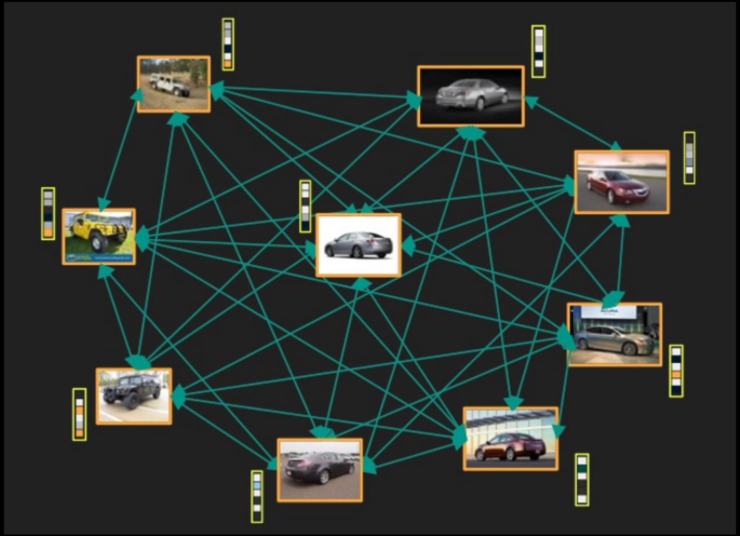
Our approach

Our approach...

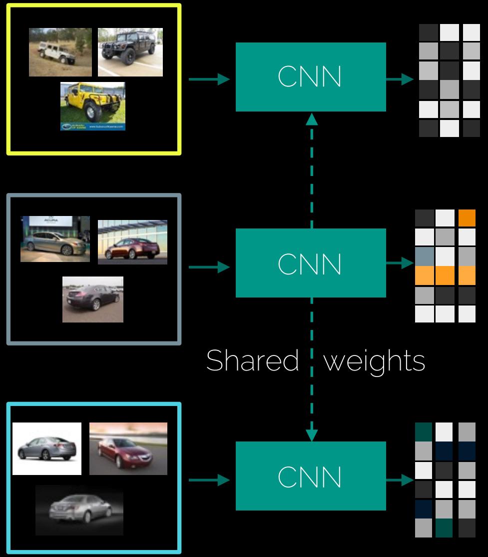
...utilizes a message pasisng network

...exchange contextual information

... take the global structure into account



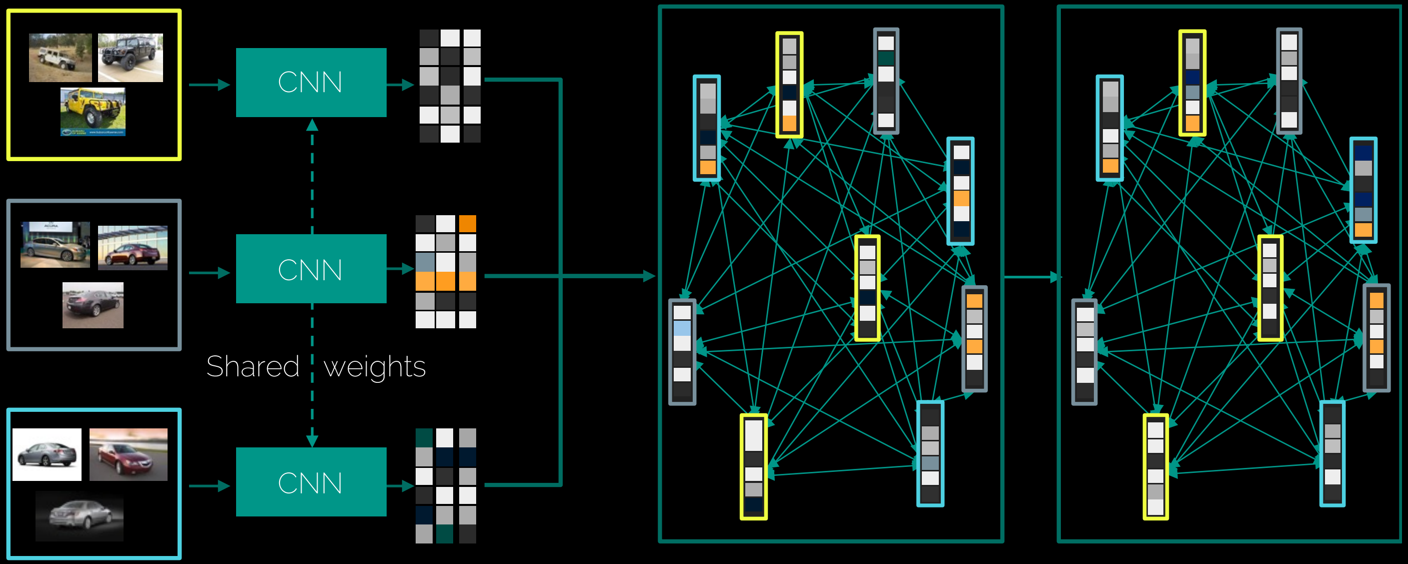
Learning Intra-Batch Connections



1. Feature initialization using a backbone CNN

J. Seidenschwarz, I. Elezi, L. Leal-Taixé. "Learning intra-batch connections for deep metric learning". ICML 2021

Learning Intra-Batch Connections

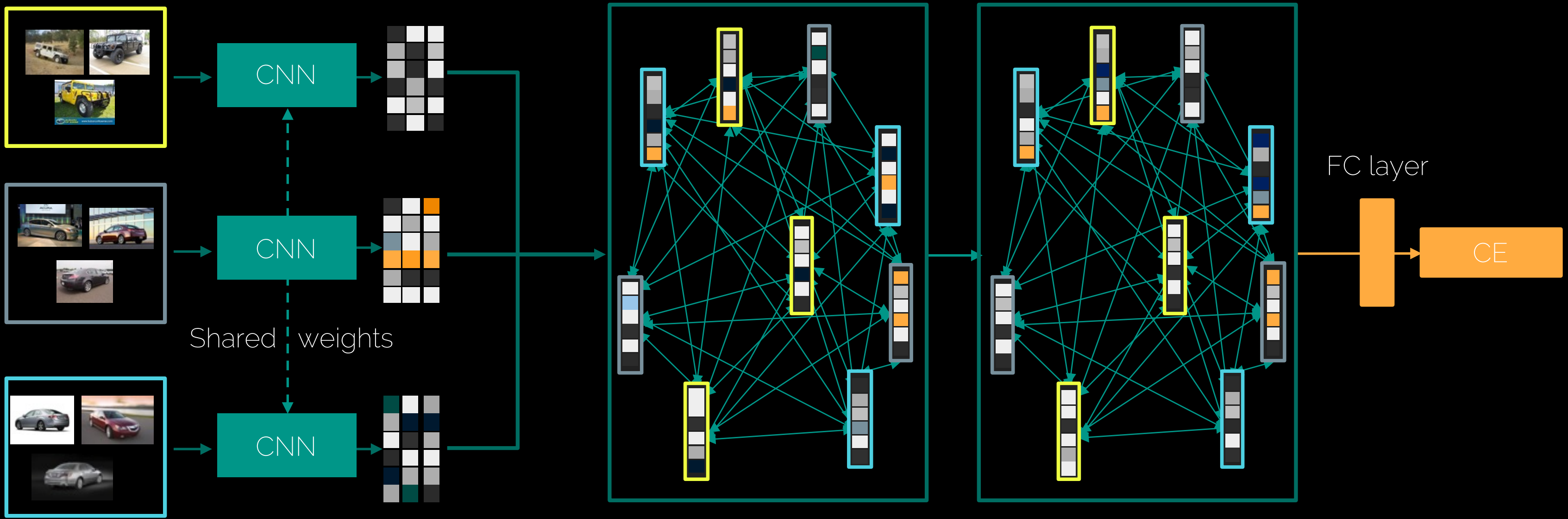


1. Feature initialization using a backbone CNN

2. Create graph and let samples communicate

J. Seidenschwarz, I. Elezi, L. Leal-Taixé. "Learning intra-batch connections for deep metric learning". ICML 2021

Learning Intra-Batch Connections



1. Feature initialization using a backbone CNN

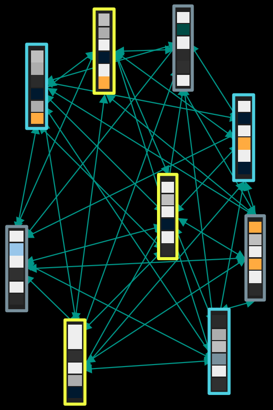
2. Create graph and let samples communicate

3. Optimize CNN and MPN end-to-end

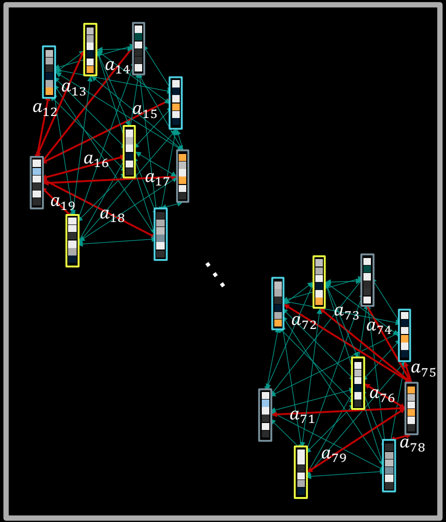
J. Seidenschwarz, I. Elezi, L. Leal-Taixé. "Learning intra-batch connections for deep metric learning". ICML 2021

Message Passing Networks

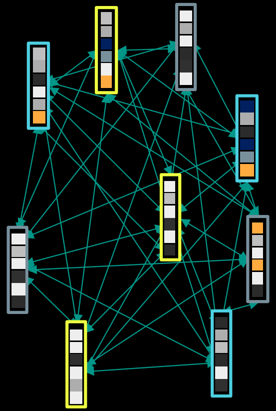
$$h_i^{(l+1)} = \sum_{n=1}^{N} \alpha_{in}^{l} W^{l} h_n^{(l)}$$



Graph with node feature vectors



Information propagation across the graph for several iterations



Graph with updated *context-aware* node feature vector(s)

Optimization and Inference

Optimization:

- We optimize backbone CNN and MPN in an end-to-end fashion
- We can apply cross-entropy loss on updated node feature vectors

Inference:

- As we optimize in end-to-end fashion we can use backbone CNN during inference

We do not add any complexity at test time!

Results

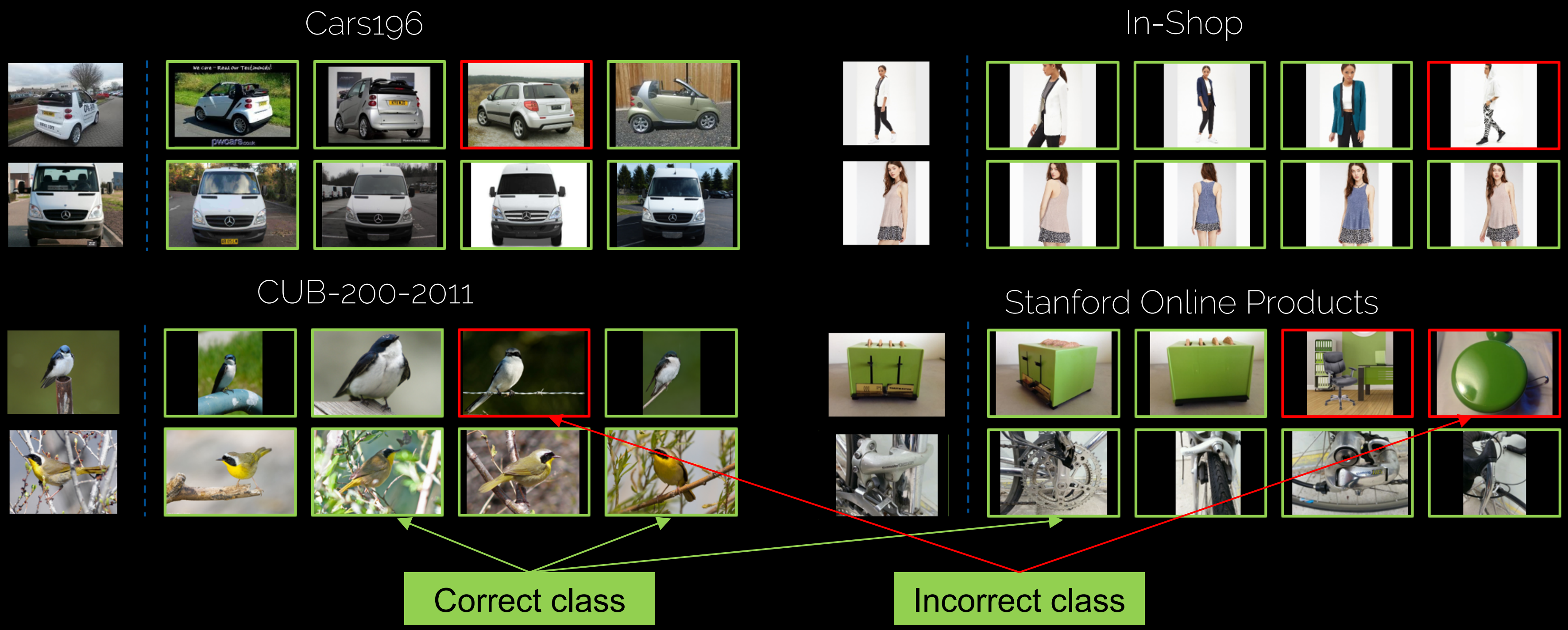
Method	BB	CUB-200-2011 R@1	R@2	R@4	R@8	NMI	CARS196 R@1	R@2	R@4	R@8	NMI	Stanford Online Products R@1	R@10	R@100	NMI
Triplet^{64} (Schroff et al., 2015) *CVPR15*	G	42.5	55	66.4	77.2	55.3	51.5	63.8	73.5	82.4	53.4	66.7	82.4	91.9	89.5
Npairs^{64} (Sohn, 2016) *NeurIPS16*	G	51.9	64.3	74.9	83.2	60.2	68.9	78.9	85.8	90.9	62.7	66.4	82.9	92.1	87.9
Deep Spectral^{512} (Law et al., 2017) *ICML17*	BNI	53.2	66.1	76.7	85.2	59.2	73.1	82.2	89.0	93.0	64.3	67.6	83.7	93.3	89.4
Angular Loss^{512} (Wang et al., 2017) *ICCV17*	G	54.7	66.3	76	83.9	61.1	71.4	81.4	87.5	92.1	63.2	70.9	85.0	93.5	88.6
Proxy-NCA^{64} (Movshovitz-Attias et al., 2017) *ICCV17*	BNI	49.2	61.9	67.9	72.4	59.5	73.2	82.4	86.4	88.7	64.9	73.7	-	-	90.6
Margin Loss^{128} (Manmatha et al., 2017) *ICCV17*	R50	63.6	74.4	83.1	90.0	69.0	79.6	86.5	91.9	95.1	69.1	72.7	86.2	93.8	90.7
Hierarchical triplet^{512} (Ge et al., 2018) *ECCV18*	BNI	57.1	68.8	78.7	86.5	-	81.4	88.0	92.7	95.7	-	74.8	88.3	94.8	-
ABE^{512} (Kim et al., 2018) *ECCV18*	G	60.6	71.5	79.8	87.4	-	85.2	90.5	94.0	96.1	-	76.3	88.4	94.8	-
Normalized Softmax^{512} (Zhai & Wu, 2019) *BMVC19*	R50	61.3	73.9	83.5	90.0	69.7	84.2	90.4	94.4	96.9	**74.0**	78.2	90.6	96.2	91.0
RLL-H^{512} (Wang et al., 2019b) *CVPR19*	BNI	57.4	69.7	79.2	86.9	63.6	74	83.6	90.1	94.1	65.4	76.1	89.1	95.4	89.7
Multi-similarity^{512} (Wang et al., 2019a) *CVPR19*	BNI	65.7	77.0	86.3	91.2	-	84.1	90.4	94.0	96.5	-	78.2	90.5	96.0	-
Relational Knowledge^{512} (Park et al., 2019) *CVPR19*	G	61.4	73.0	81.9	89.0	-	82.3	89.8	94.2	96.6	-	75.1	88.3	95.2	-
Divide and Conquer^{1028} (Sanakoyeu et al., 2019) *CVPR19*	R50	65.9	76.6	84.4	90.6	69.6	84.6	90.7	94.1	96.5	70.3	75.9	88.4	94.9	90.2
SoftTriple Loss^{512} (Qian et al., 2019) *ICCV19*	BNI	65.4	76.4	84.5	90.4	69.3	84.5	90.7	94.5	96.9	70.1	78.3	90.3	95.9	**92.0**
HORDE^{512} (Jacob et al., 2019) *ICCV19*	BNI	66.3	76.7	84.7	90.6	-	83.9	90.3	94.1	96.3	-	**80.1**	**91.3**	96.2	-
MIC^{128} (Brattoli et al., 2019) *ICCV19*	R50	66.1	76.8	85.6	-	69.7	82.6	89.1	93.2	-	68.4	77.2	89.4	95.6	90.0
Easy triplet mining^{512} (Xuan et al., 2020b) *WACV20*	R50	64.9	75.3	83.5	-	-	82.7	89.3	93.0	-	-	78.3	90.7	**96.3**	-
Group Loss^{1024} (Elezi et al., 2020) *ECCV20*	BNI	65.5	77.0	85.0	91.3	69.0	85.6	91.2	94.9	97.0	**72.7**	75.1	87.5	94.2	**90.8**
Proxy NCA++^{512} (Teh et al., 2020) *ECCV20*	R50	66.3	77.8	**87.7**	91.3	**71.3**	84.9	90.6	94.9	97.2	71.5	79.8	**91.4**	**96.4**	-
DiVA^{512} (Milbich et al., 2020) *ECCV20*	R50	**69.2**	**79.3**	-	-	**71.4**	**87.6**	**92.9**	-	-	72.2	79.6	-	-	90.6
PADS^{128} (Roth et al., 2020) *CVPR20*	R50	67.3	78.0	85.9	-	69.9	83.5	89.7	93.8	-	68.8	76.5	89.0	95.4	89.9
Proxy Anchor^{512} (Kim et al., 2020) *CVPR20*	BNI	68.4	79.2	86.8	**91.6**	-	86.1	91.7	95.0	**97.3**	-	79.1	90.8	96.2	-
Proxy Anchor^{512} (Kim et al., 2020) *CVPR20*	R50	**69.7**	**80.0**	**87.0**	**92.4**	-	**87.7**	**92.9**	**95.8**	**97.9**	-	**80.0**	**91.7**	**96.6**	-
Proxy Few^{512} (Zhu et al., 2020) *NeurIPS20*	BNI	66.6	77.6	86.4	-	69.8	85.5	**91.8**	**95.3**	-	72.4	78.0	90.6	96.2	90.2
Ours^{512}	R50	**70.3**	**80.3**	**87.6**	**92.7**	**74.0**	**88.1**	**93.3**	**96.2**	**98.2**	**74.8**	**81.4**	**91.3**	95.9	**92.6**

In-Shop

Method	BB	R@1	R@10	R@20	R@40
FashionNet^{4096} (Liu et al., 2016) *CVPR16*	V	53.0	73.0	76.0	79.0
A-BIER^{512} (Opitz et al., 2020) *PAMI20*	G	83.1	95.1	96.9	97.8
ABE^{512} (Kim et al., 2018) *ECCV18*	G	87.3	96.7	97.9	98.5
Multi-similarity^{512} (Wang et al., 2019a) *CVPR19*	BNI	89.7	**97.9**	98.5	**99.1**
Learning to Rank^{512} (Çakir et al., 2019)	R50	90.9	97.7	98.5	**98.9**
HORDE^{512} (Jacob et al., 2019) *ICCV19*	BNI	90.4	97.8	98.4	**98.9**
MIC^{128} (Brattoli et al., 2019) *ICCV19*	R50	88.2	97.0	98.0	98.8
Proxy NCA++^{512} (Teh et al., 2020) *ECCV20*	R50	90.4	**98.1**	**98.8**	**99.2**
Proxy Anchor^{512} (Kim et al., 2020) *CVPR20*	BNI	**91.5**	**98.1**	**98.8**	**99.1**
Proxy Anchor^{512} (Kim et al., 2020) *CVPR20*	R50	**92.1**	**98.1**	**98.7**	**99.2**
Ours^{512}	R50	**92.8**	**98.5**	**99.1**	**99.2**

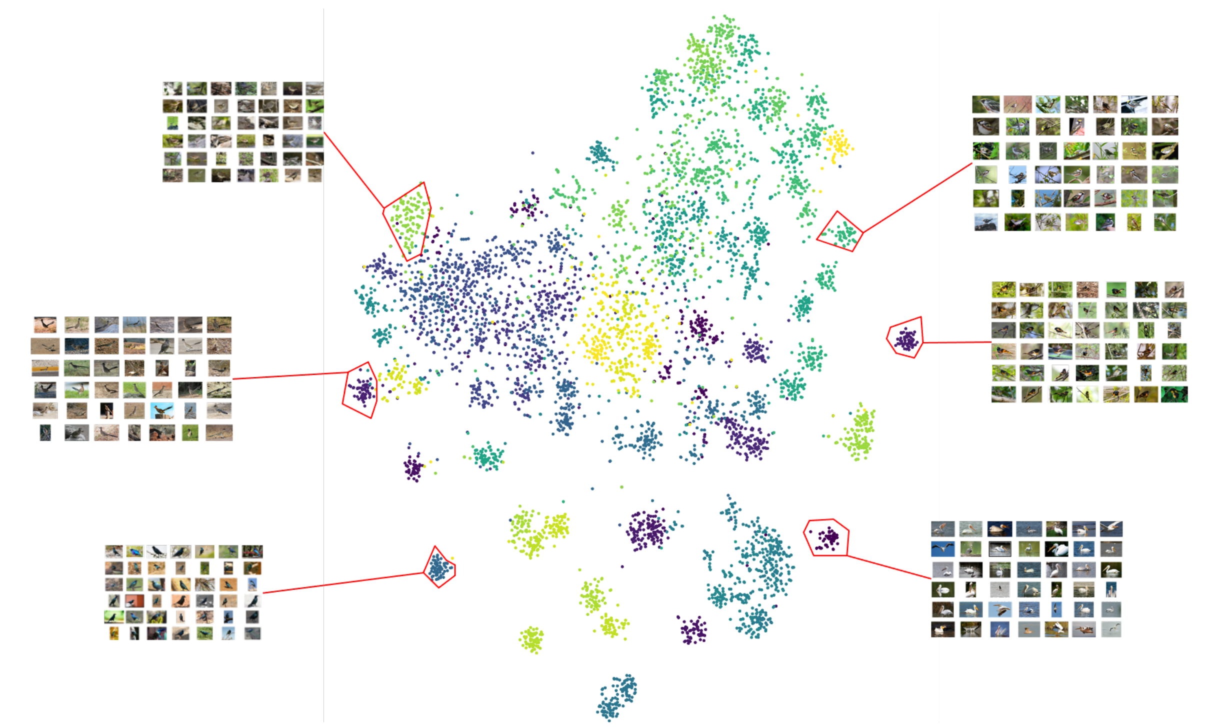
Results

- State-of-the-art results on 4 public datasets



Results

- Clustering performance on CUB-200-2011



Want to know more?

https://dvl.in.tum.de

Jenny
Seidenschwarz

Ismail
Elezi

Laura
Leal-Taixé