CVPR VIRTUAL JUNE 19-25

SLADE: A Self-Training Metric Learning Framework


Jiali Duan[1,2]


Yen-Liang Lin [2]

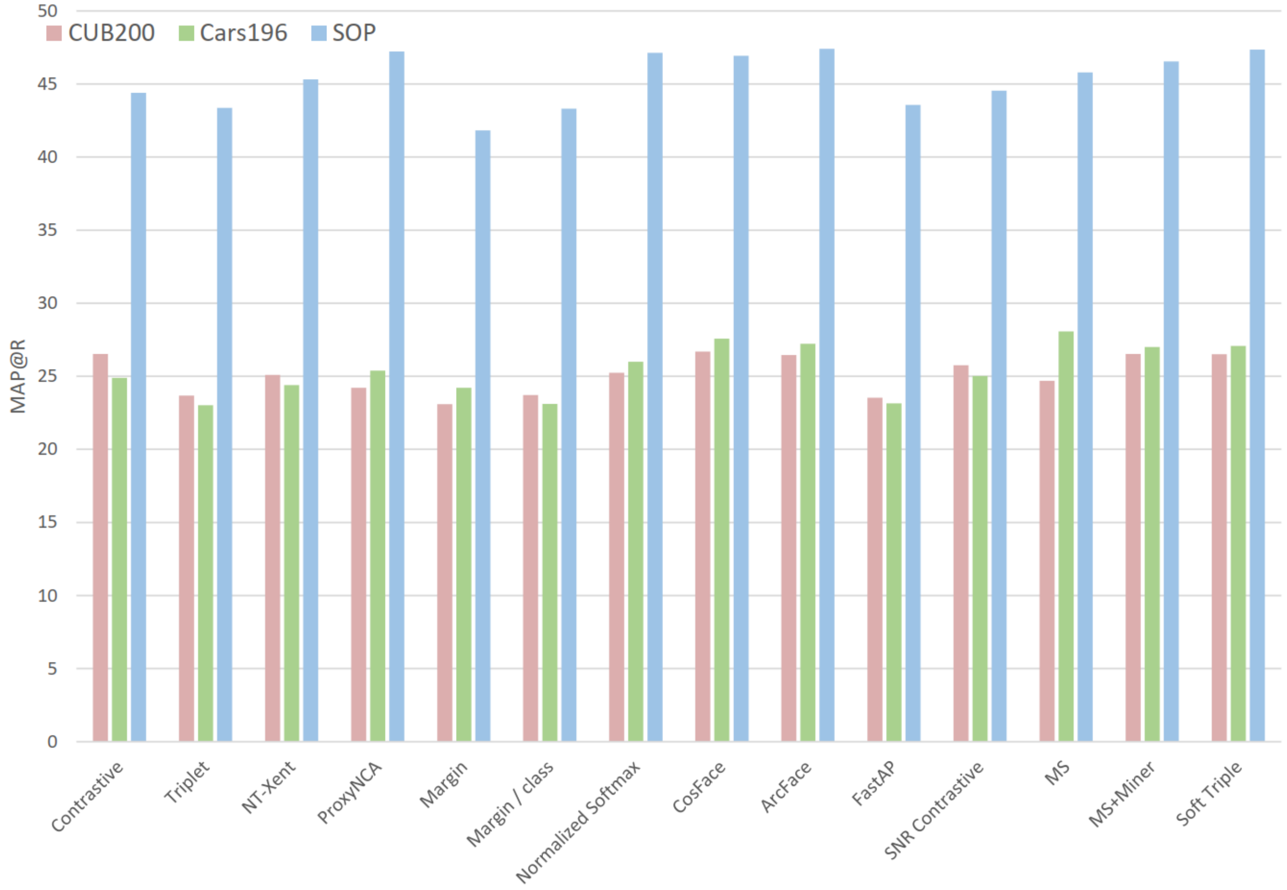

Son Tran [2]


Larry S. Davis [2]


C.-C. Jay Kuo [1]

[1] University of Southern California [2] Amazon

A Review of Deep Metric Learning



Musgrave K, Belongie S, Lim S N. A metric learning reality check. ECCV 2020

Background of Self-Training

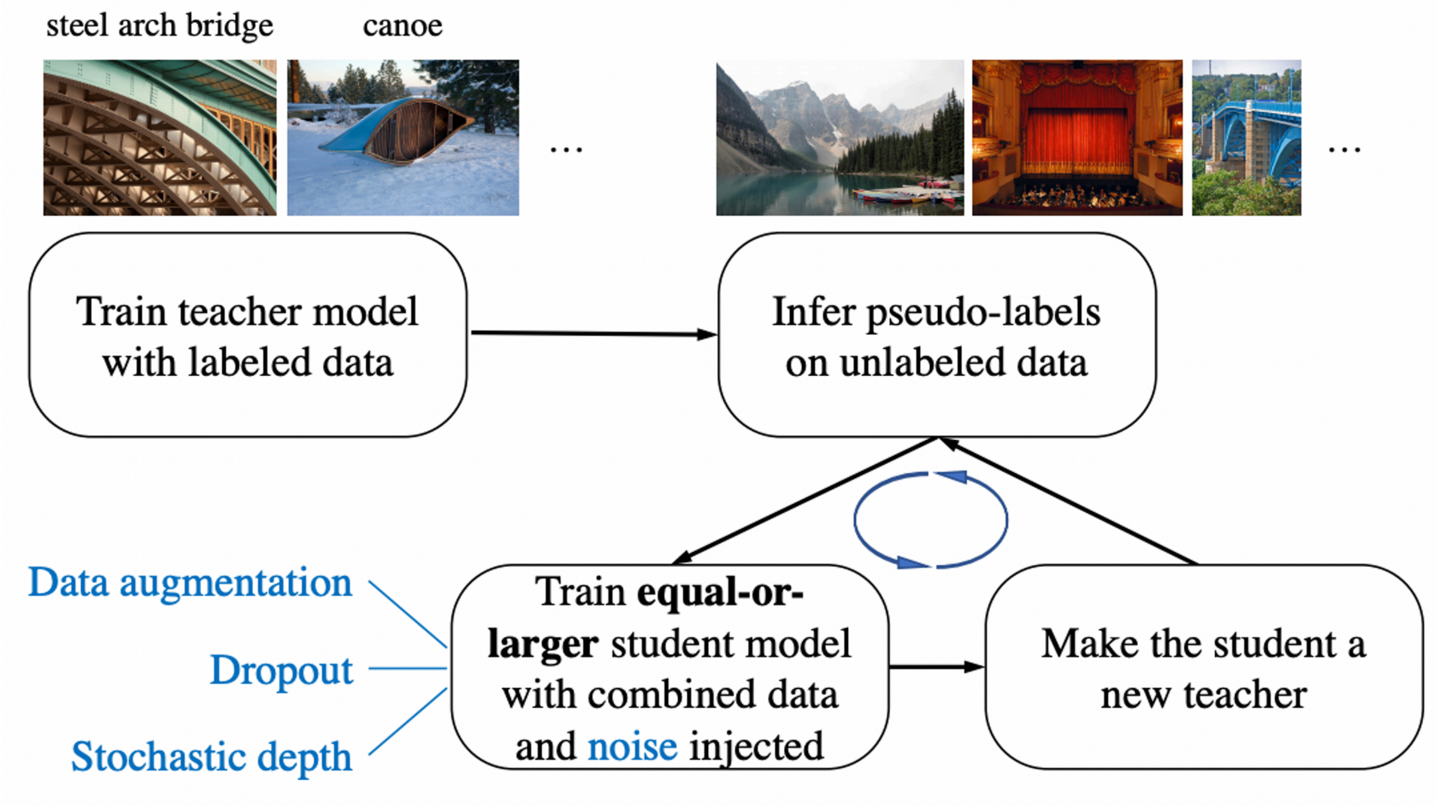
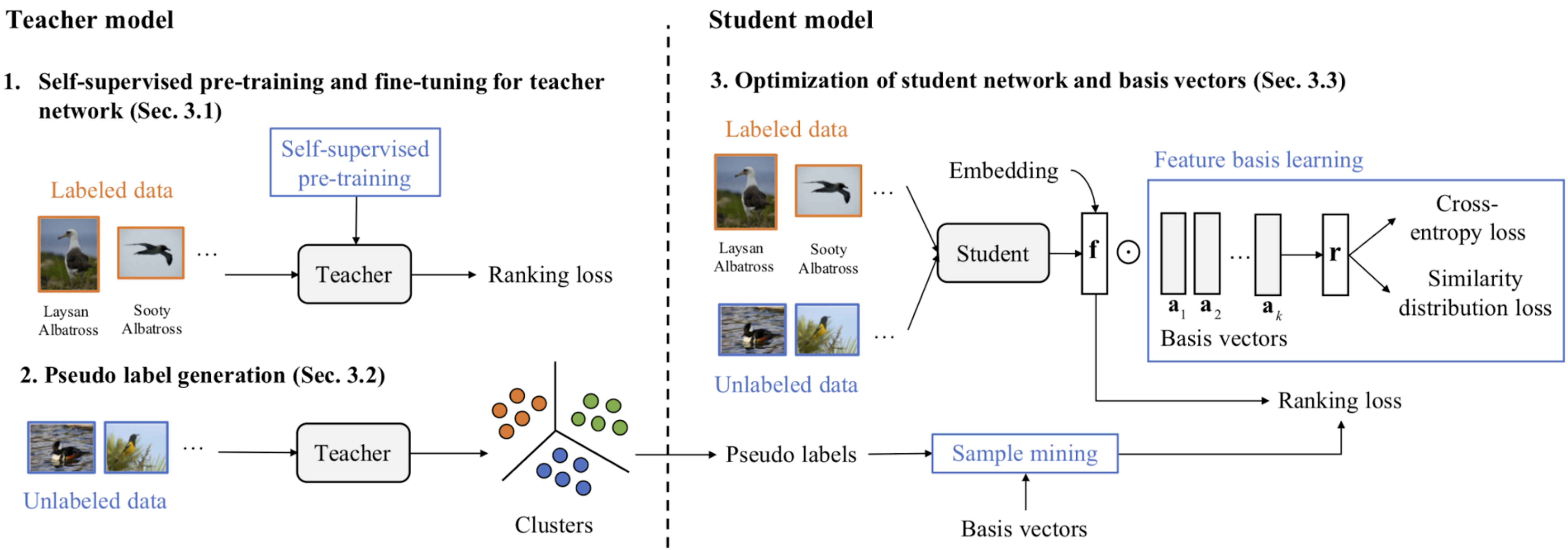
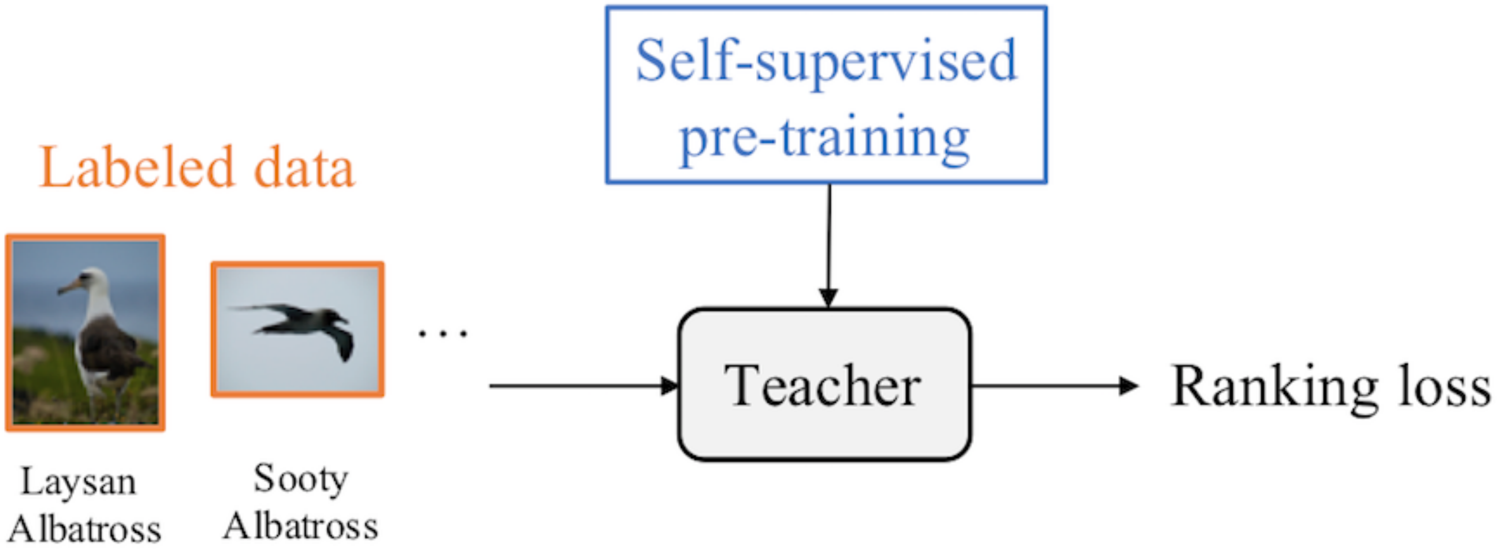


Image credit: Xie Q, Luong M T, Hovy E, et al. Self-training with noisy student improves imagenet classification, CVPR 2020

SLADE: A Self-Training Metric Learning Framework

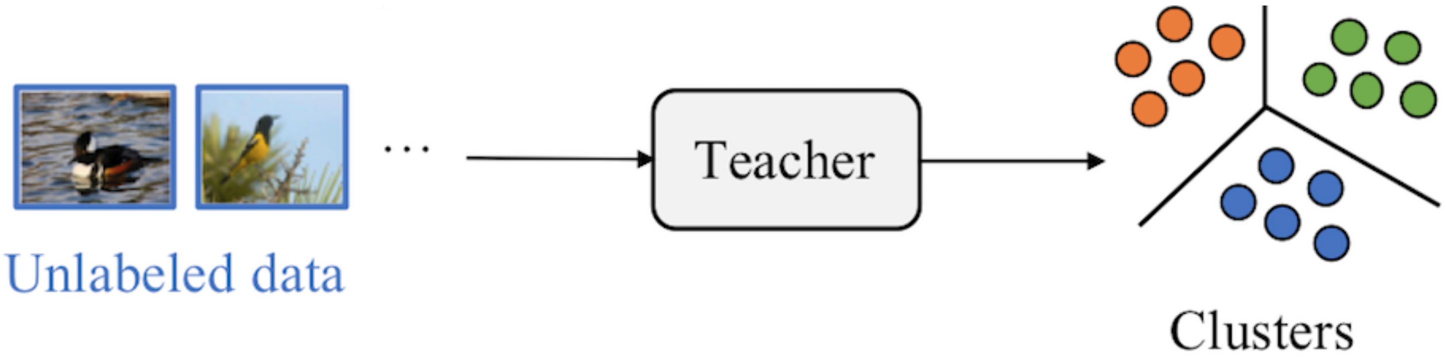


Step 1: Self-supervised Pretraining and Finetuning



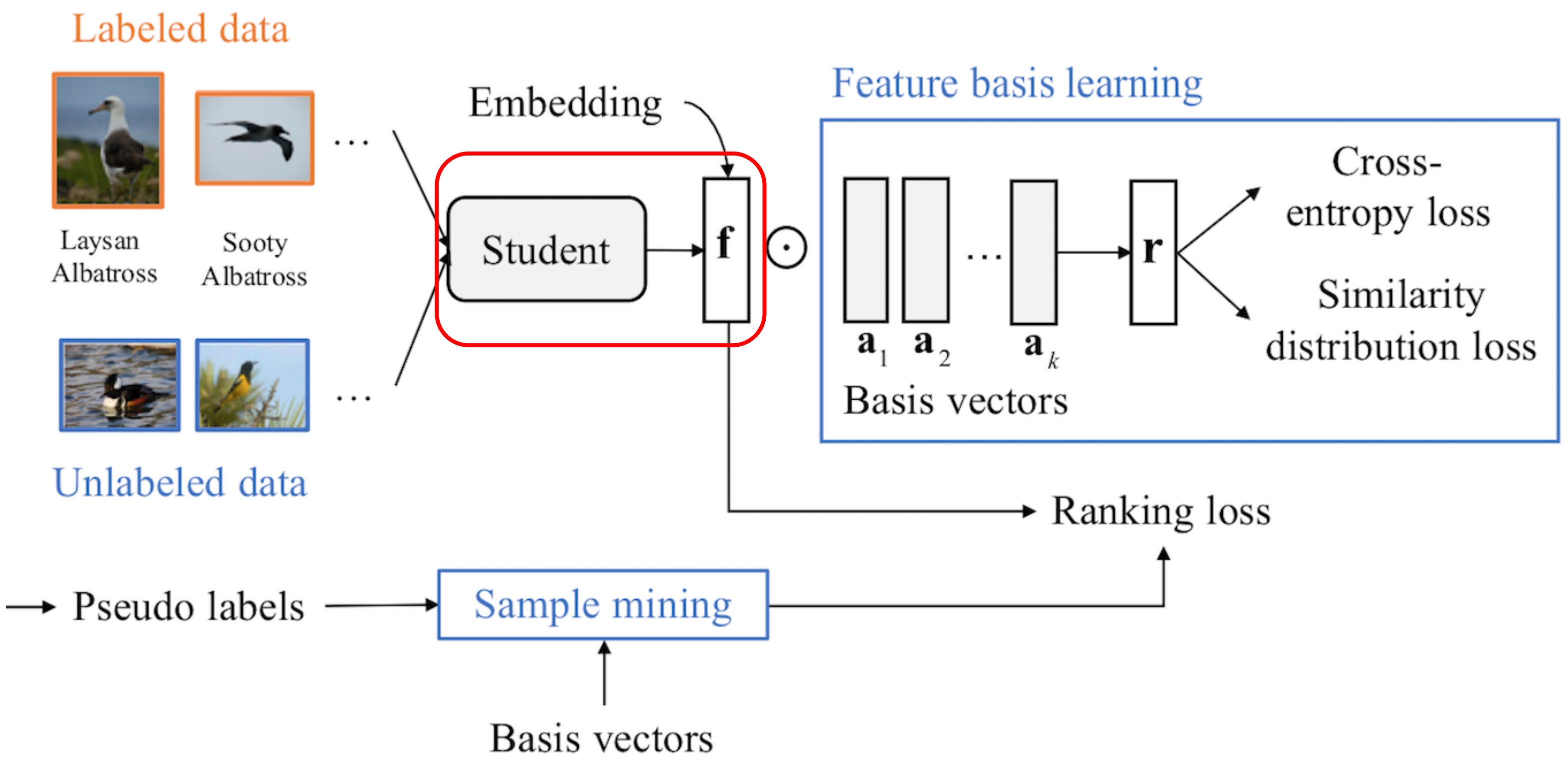
SWAV: Mathilde Caron et al., Unsupervised Learning of Visual Features by Contrasting Cluster Assignments . NeurIPS 2020.
BYOL: Grill J B, Strub F, Altché F, et al. Bootstrap your own latent: A new approach to self-supervised learning.NeurIPS 2020.
MoCo: He K, Fan H, Wu Y, et al. Momentum contrast for unsupervised visual representation learning. CVPR 2020

Step 2: Pseudo Label Generation



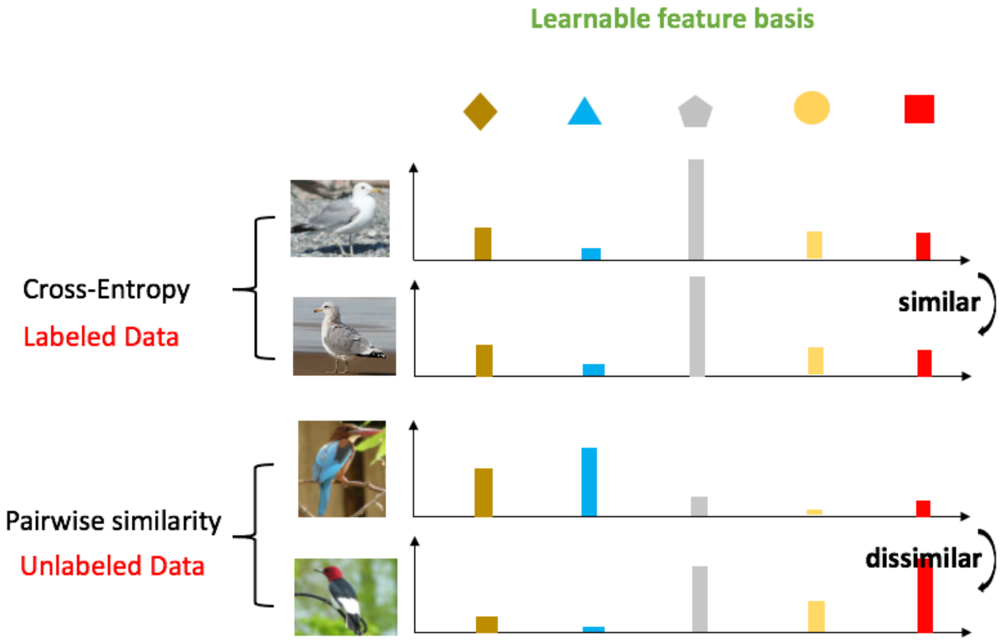
Noisy!

Step 3: Training of Student Network



$$L = L_{rank}(D^l;\theta^s) + L_{rank}(D^u;\theta^s) + L_{basis}(D^l, D^u;\theta^s, W_a)$$

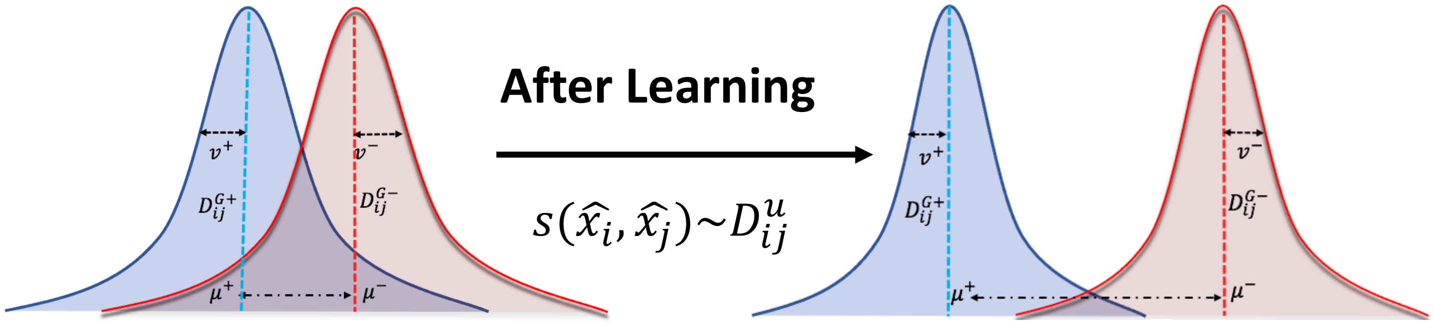
Step 3: Feature basis learning



$$L_{CE} = \sum_{(x_i, y_i) \in D^l} -y_i \log(\sigma(W_a f(x_i, \theta^s)))$$

Goal: reduce overlap between distributions

- Maximize distance between two means
- Reduce variances of two distributions



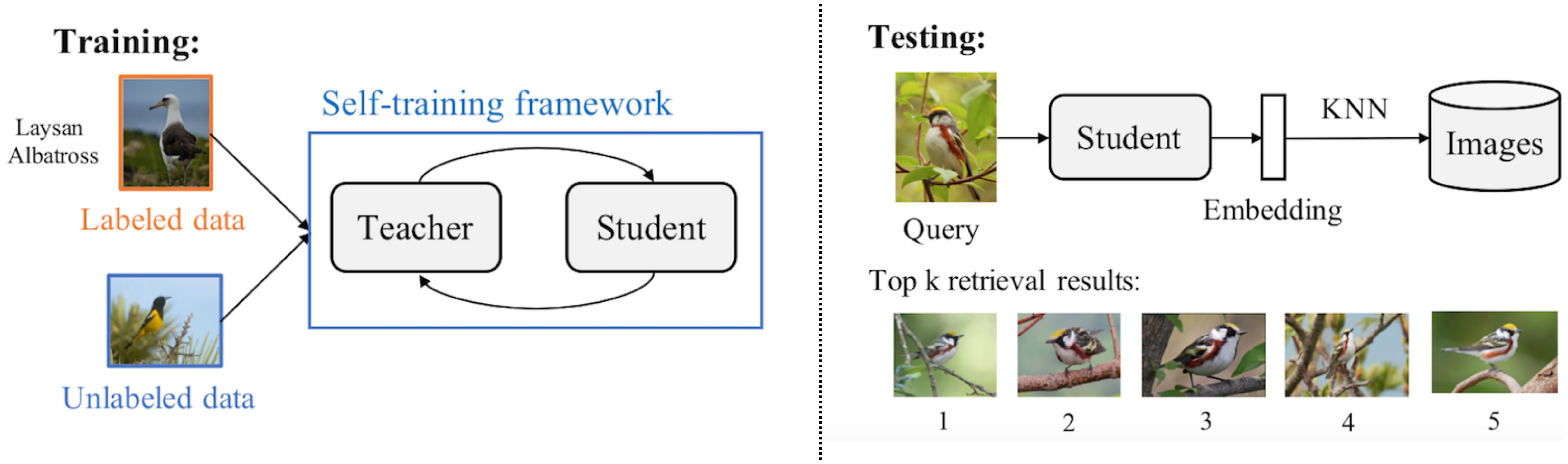
$$L_{SD}(G^+ || G^-) = \max(\mu^- - \mu^+ + m, 0) + \lambda(v^+ + v^-)$$

Momentum Update

$$\mu^+ = (1-\beta) \times \mu_b^+ + \beta \times \mu^+$$

$$v^+ = (1-\beta) \times v_b^+ + \beta \times v^+$$

Putting it Together



Evaluation

CUB-200 (labeled): 200 species/ 12k images
NABIRDS (unlabeled): 400 species/ 48k images

Cars-196 (labeled): 196 brands/ 16k images
CompCars (unlabeled): 145 brands/ 16k images

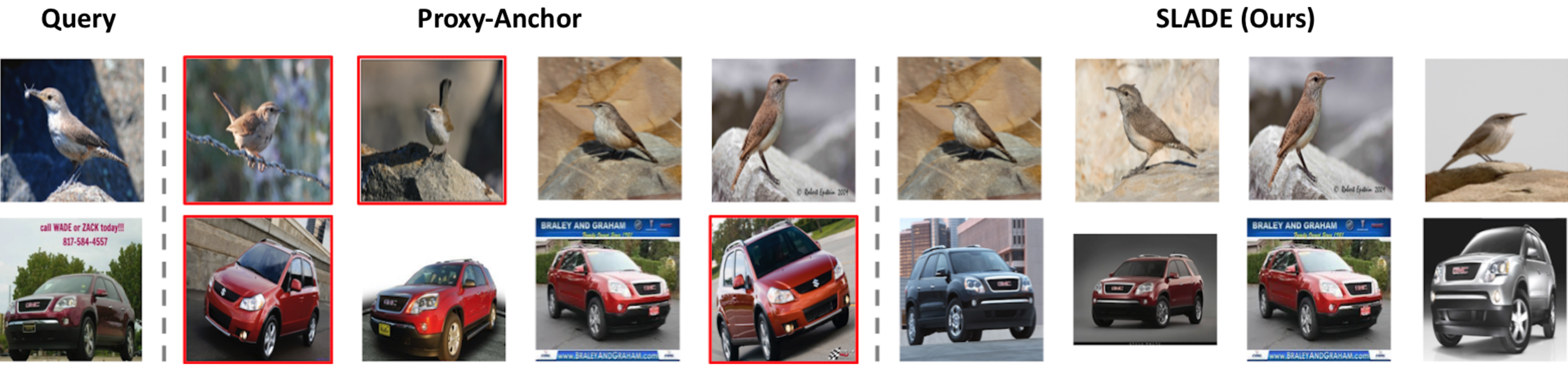
In-shop (labeled): 8k instances/ 52k images
Fashion200k (unlabeled): 1k instances/15k images*

Results: Performance Comparisons

Methods	Frwk	Init	Arc / Dim	CUB-200-2011			Cars-196		
				MAP@R	RP	P@1	MAP@R	RP	P@1
Contrastive [10]	[19]	ImageNet	BN / 512	26.53	37.24	68.13	24.89	35.11	81.78
Triplet [29]	[19]	ImageNet	BN / 512	23.69	34.55	64.24	23.02	33.71	79.13
ProxyNCA [18]	[19]	ImageNet	BN / 512	24.21	35.14	65.69	25.38	35.62	83.56
N. Softmax [35]	[19]	ImageNet	BN / 512	25.25	35.99	65.65	26.00	36.20	83.16
CosFace [25, 26]	[19]	ImageNet	BN / 512	26.70	37.49	67.32	27.57	37.32	85.52
FastAP [3]	[19]	ImageNet	BN / 512	23.53	34.20	63.17	23.14	33.61	78.45
MS+Miner [27]	[19]	ImageNet	BN / 512	26.52	37.37	67.73	27.01	37.08	83.67
Proxy-Anchor[1] [15]	[15]	ImageNet	R50 / 512	-	-	69.9	-	-	87.7
Proxy-Anchor[2] [15]	[19]	ImageNet	R50 / 512	25.56	36.38	66.04	30.70	40.52	86.84
ProxyNCA++ [22]	[22]	ImageNet	R50 / 2048	-	-	72.2	-	-	90.1
Mutual-Info [1]	[1]	ImageNet	R50 / 2048	-	-	69.2	-	-	89.3
Contrastive [10] (T_1)	[19]	ImageNet	R50 / 512	25.02	35.83	65.28	25.97	36.40	81.22
Contrastive [10] (T_2)	[19]	SwAV	R50 / 512	29.29	39.81	71.15	31.73	41.15	88.07
SLADE (Ours) (S_1)	[19]	ImageNet	R50 / 512	29.38	40.16	68.92	31.38	40.96	85.8
SLADE (Ours) (S_2)	[19]	SwAV	R50 / 512	**33.59**	**44.01**	**73.19**	**36.24**	**44.82**	**91.06**
MS [27] (T_3)	[19]	ImageNet	R50 / 512	26.38	37.51	66.31	28.33	38.29	85.16
MS [27] (T_4)	[19]	SwAV	R50 / 512	29.22	40.15	70.81	33.42	42.66	89.33
SLADE (Ours) (S_3)	[19]	ImageNet	R50 / 512	30.90	41.85	69.58	32.05	41.50	87.38
SLADE (Ours) (S_4)	[19]	SwAV	R50 / 512	**33.90**	**44.36**	**74.09**	**37.98**	**46.92**	**91.53**

Qualitative Results

CUB-200 Cars-196



Comparison with Proxy Anchor [1] on CUB200 & Cars-196

Kim S et al. Proxy anchor loss for deep metric learning. CVPR 2020

Conclusions

- We propose a novel self-training framework to further improve the performance of deep metric learning which exploits unlabeled data.
- We propose a feature basis learning approach to deal with the noisy pseudo-labels during training.
- Experimental results demonstrate our approach significantly improves the performance over the state-of-the-art methods with additional unlabeled data.