LSD-C: Linearly Separable Deep Clusters

Sylvestre-Alvise Rebuffi*, Sebastien Ehrhardt*, Kai Han*, Andrea Vedaldi, Andrew Zisserman
University of Oxford

General problem

Leverage powerful **self-supervised learning** methods to improve deep clustering.

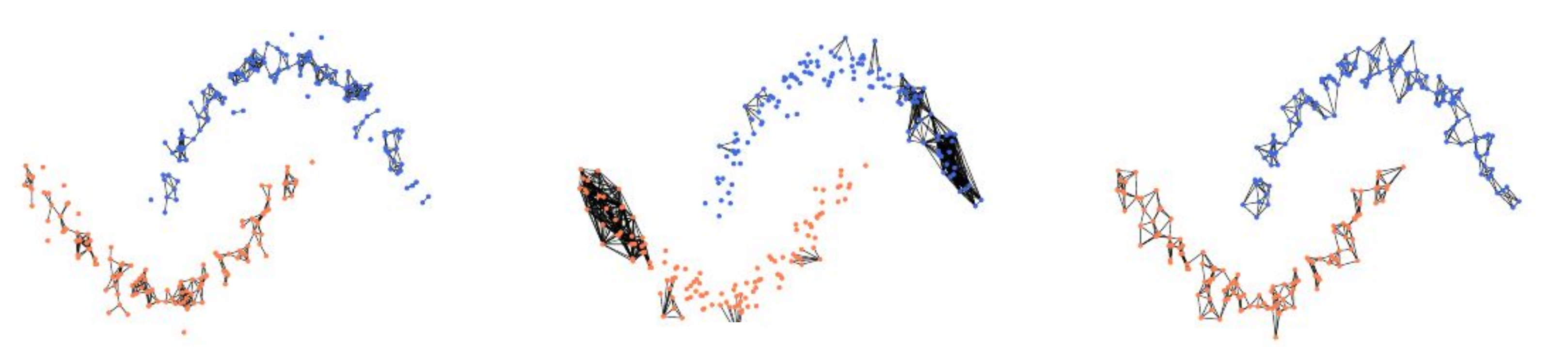
Key points of our method

- Model **initialization** with RotNet.
- Pairwise **labeling** in the feature map.
- A pairwise clustering **loss**.
- Data **augmentation** to avoid local minima.

Pairwise labeling in feature space

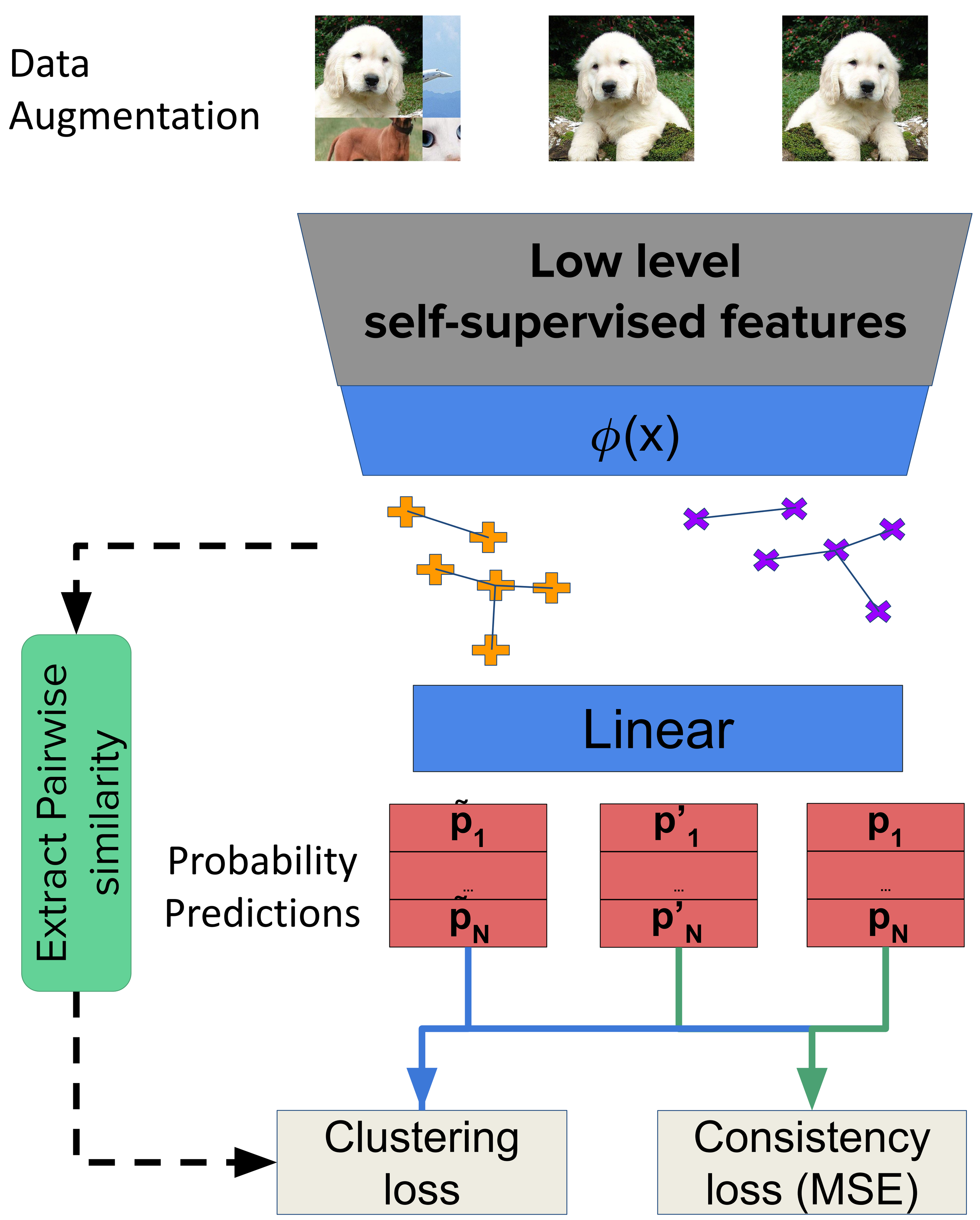
We compute similarity C_{ij} at **the feature level.**
We then assign a similarity matrix $A_{ij} = \mathbb{1}_{C_{ij}}$

L_2 dist.	SNE	kNN
$C_{ij} = \|\mathbf{f}_j - \mathbf{f}_i\|^2 < \tau$	$\frac{\exp(-\|\mathbf{f}_j - \mathbf{f}_i\|^2/T^2)}{H(Z_i, Z_j)} > \tau$	$(j \in \text{kNN}(i)) \vee (i \in \text{kNN}(j))$



L_2 dist. SNE kNN

Pairwise clustering in feature space



Pairwise clustering loss

K clusters linear classifier:

$$P(i=j) = \sum_{k=1}^{K} P(i=k, j=k) = \sum_{k=1}^{K} P(i=k)P(j=k) = \mathbf{p}_i^\top \mathbf{p}_j$$

Loss to match the pairwise labels assumes **independence of samples**:

$$L_{\text{clus}} = -\sum_{i,j} A_{ij} \log P(i=j) + (1 - A_{ij}) \log P(i \neq j)$$

$$L_{\text{clus}} = -\sum_{i,j} A_{ij} \log(\mathbf{p}_i^\top \mathbf{p}'_j) + (1 - A_{ij}) \log(1 - \mathbf{p}_i^\top \mathbf{p}'_j)$$

This loss aims at:

- **maximizing** the number of similarity edges within clusters.
- **minimizing** within clusters the number of edges of the complement of the similarity graph.

Experiments

Methods comparaison

	K-means [40]	JULE [55]	IIC [28]	Ours
CIFAR 10	22.9	27.2	61.7	**81.7** $\pm$ 0.9
CIFAR 100-20	13.0	13.7	25.7	**42.3** $\pm$ 1.0
STL 10	19.2	27.7	59.6	**66.4** $\pm$ 3.2
MNIST	57.2	96.4	**99.2**	98.6 $\pm$ 0.5

Our work is outperforming past method by a constituent margin on standard clustering benchmarks.

Ablation study

	Pairwise labeling				Using the pred. space			Data augmentation		
	L_2	Cosine	kNN	SNE	Cosine	kNN	SNE	RICAP	MixUp	None
CIFAR 10	70.2	81.1	**81.7**	81.5	63.7	64.7	**67.0**	**81.7**	75.3	53.7
CIFAR 100-20	26.1	34.4	**42.3**	40.4	20.4	**32.8**	30.4	**42.3**	37.1	35.4

Summary:

- kNN and SNE are the best labeling strategies.
- Pairwise labeling at the prediction space level hurts the performance.
- Key role of data augmentation (especially for CIFAR-10).

Code and paper link

Scan the QR code to download our publicly available code.

https://arxiv.org/pdf/2006.10039.pdf