All4One: Symbiotic Neighbour Contrastive Learning via Self-Attention and Redundancy Reduction [Supplementary Material]

Imanol G. Estepa
Universitat de Barcelona
Barcelona, Spain
igonzaes42@alumnes.ub.edu

Ignacio Sarasúa
NVIDIA
isarasua@nvidia.com

Bhalaji Nagarajan
Universitat de Barcelona
Barcelona, Spain
bhalaji.nagarajan@ub.edu

Petia Radeva
Universitat de Barcelona
Barcelona, Spain
Computer Vision Center,
Cerdanyola (Barcelona), Spain
petia.ivanova@ub.edu

1. Shift Operation

We define the process of replacing an element of a neighbour sequence as the Shift operation. This operation can be visualized in Figure 1, which illustrates how the final element of a sequence of neighbours is replaced with the output of a predictor, denoted by p_i^{c2}. The resulting sequence is then rearranged so that p_i^{c2} becomes the first element in the sequence. Note that this operation can be applied to sequences of any length.

2. UMAP Visualization

Epochwise UMAP for All4One. In Figure 2 (left), the progress of five CIFAR-100 classes during training is visually depicted. Initially, the feature representations of these classes lack any structure or pattern, appearing highly mixed. However, as the training process proceeds, the model acquires the ability to generate unique features for each class, causing multiple clusters to emerge within the feature space. These clusters effectively group the feature representations according to their respective classes, even in the absence of annotations, effectively separating one class's representations from another's.

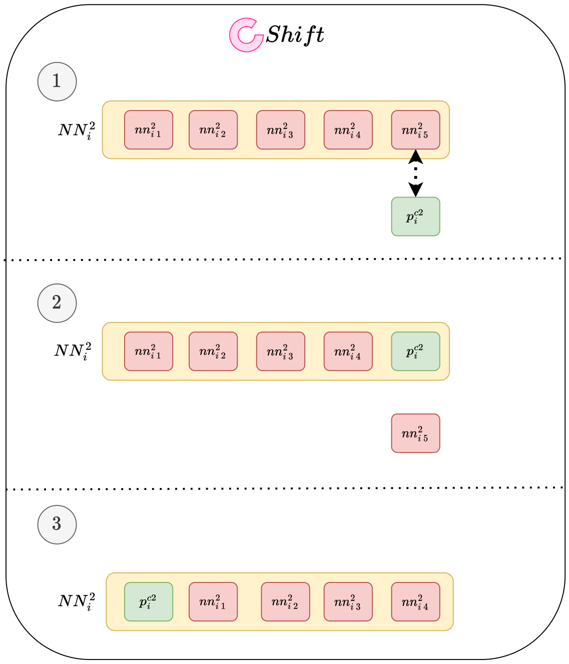

Figure 1: **Overview of the Shift operation.**

Feature representation comparison. All4One outperforms NNCLR in generating feature representations, enabling more effective cluster visualization by UMAP. This superior performance is particularly noteworthy in scenarios featuring classes with high similarity, such as those pertaining to certain animal or tree species. As illustrated in Figure 2 (right), the feature clusters generated from All4One representations for the Maple, Palm, and Willow tree classes demonstrate greater compactness in comparison to those derived from NNCLR representations. Moreover, while the NNCLR approach erroneously groups the Fox and Wolf classes, the All4One method distinguishes between these two classes with precision.

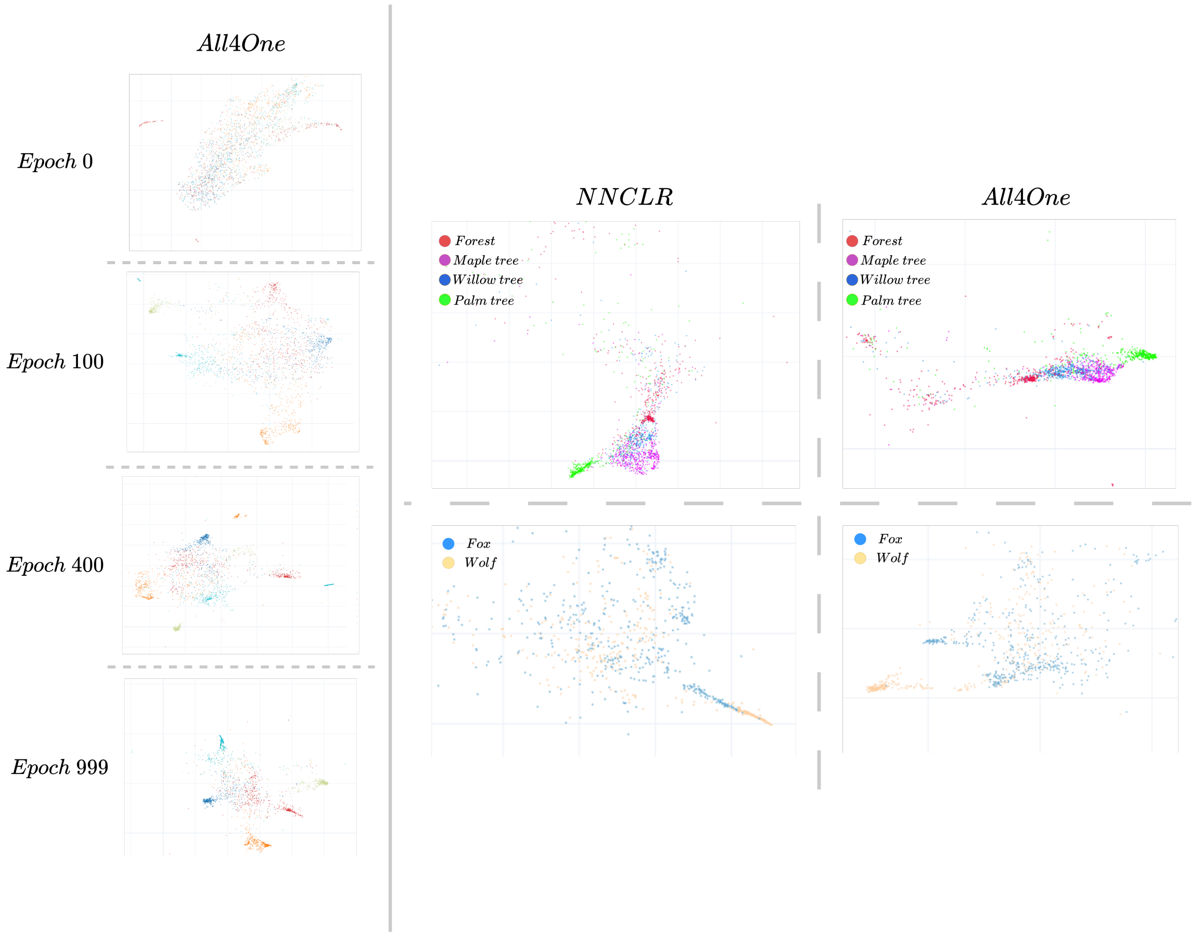

Figure 2: **Left: Epochwise All4One UMAP visualization:** During the training, we select and visualize 5 random classes from the CIFAR-100 dataset. **Right: Per-class UMAP comparison between NNCLR and All4One:** Classes are extracted from CIFAR-100 dataset.

3. Computational complexity

We provide the total training complexity in terms of model parameters in Table 1. Each SSL framework, depending on its architecture, introduces a different number of additional parameters apart from the backbone. With All4One, we introduce 10% more parameters (transformer encoder and a second predictor) than BT when using a ResNet18 backbone. However, it decreases to less than 2% on bigger backbones such as ViT-L. Nevertheless, we have to note that during the inference in all the frameworks, only the backbone is used and the increased complexity does not affect the inference.

	Backbone	Projs.	Preds.	Trans.	Params
BT	11.2M	9.5M	-	-	20.7M
NNCLR	11.2M	5.8M	2.1M	-	19.1M
All4One	11.2M	5.8M	4.2M	1.6M	22.8M

Table 1: Computational complexity of All4One, NNCLR and BT.

4. Training curves

Feature vs Centroid vs Neighbour. In Figure 3a, the training curves for each objective of our model are presented. It is worth noting that the Centroid and Neighbour objectives exhibit very similar starting values, but the Centroid objective is optimized more effectively by our model. We hypothesize that the ease of optimization can be attributed to the contextual information captured by the multiple representations used by the Centroid objective.

NNCLR vs Neighbour objective. Despite the equivalence of the NNCLR objective function and our Neighbour objective function, All4One is able to optimize the latter more effectively, thanks to the symbiotic relationship be-

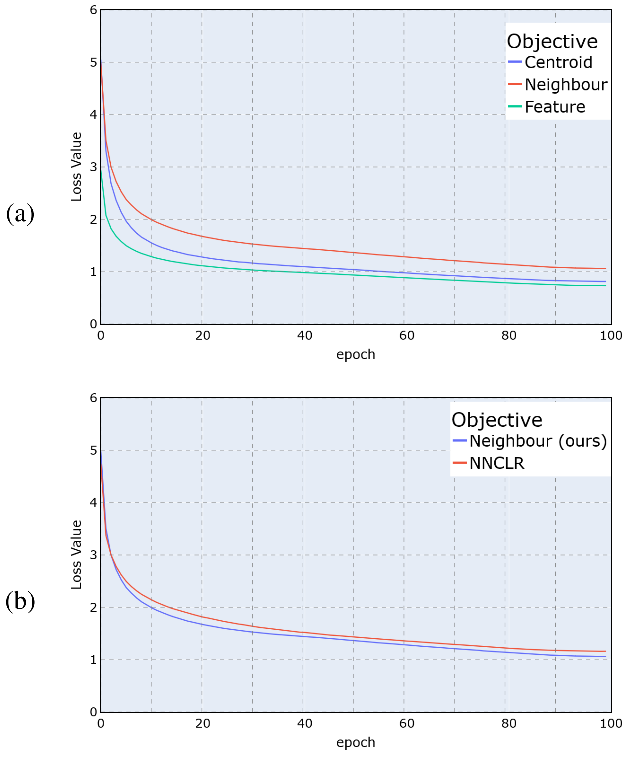

Figure 3: **Training curve comparison (ImageNet). Top:** Centroid vs Neighbour vs Feature. **Bottom:** NNCLR objective vs All4One Neighbour objective.

tween the two components, as demonstrated in Figure 3b, where our Neighbour objective curve is better optimized during training that NNCLR curve.

5. Neighbour Retrieval

Figure 4 showcases some of the image extractions performed by the KNN operator during the training phase. As can be seen, the KNN operator accurately supplies the model with numerous and varied image representations primarily pertaining to the same semantic class.

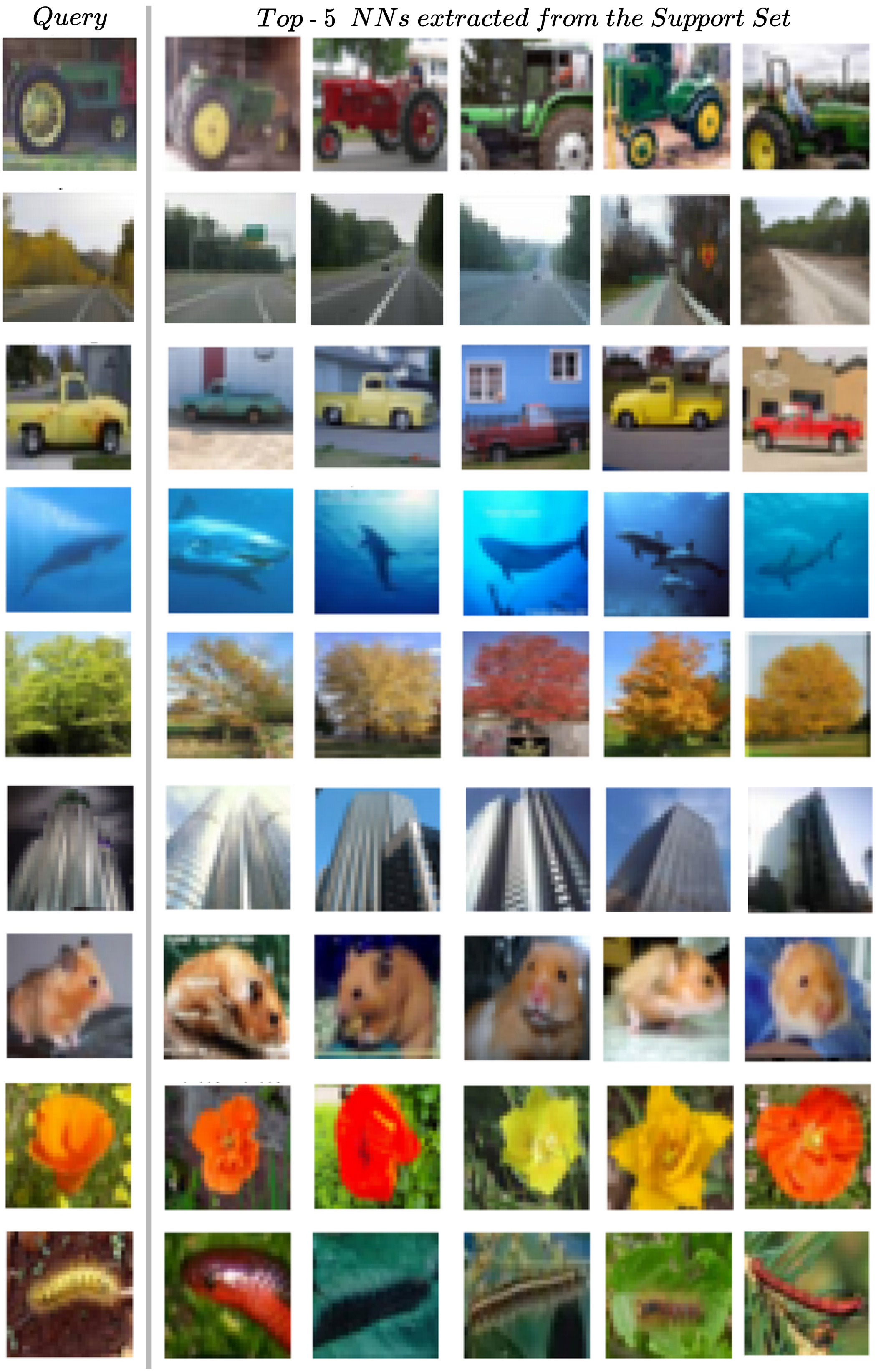

Figure 4: **NN extractions performed by All4One.**