Neural Collapse Inspired Contrastive Continual Learning

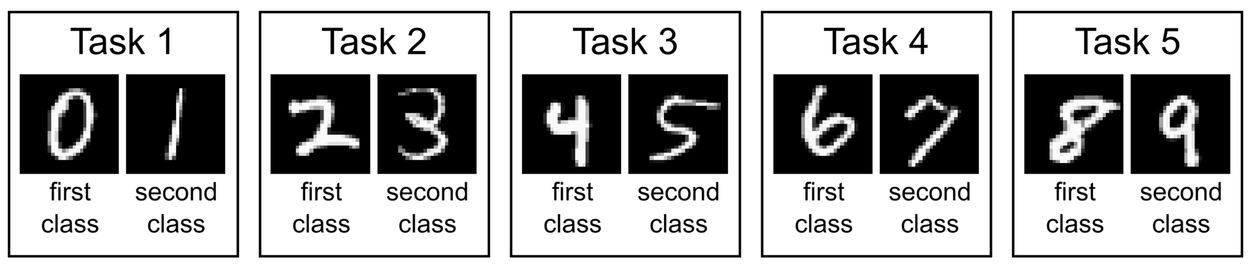
Antoine Montmaur[1], Nicolas Larue[1], Ngoc-Son Vu[1]
[1] ETIS, CYU, ENSEA, CNRS : name.surname@ensea.fr

Contributions

- A novel loss function compelling representations to directly achieve neural collapse by using predefined hard prototypes
- A novel distillation technique based on simplex to enhance consolidation of previously acquired knowledge
- Experiments in class-incremental scenario outperforming replay-based methods

Concepts

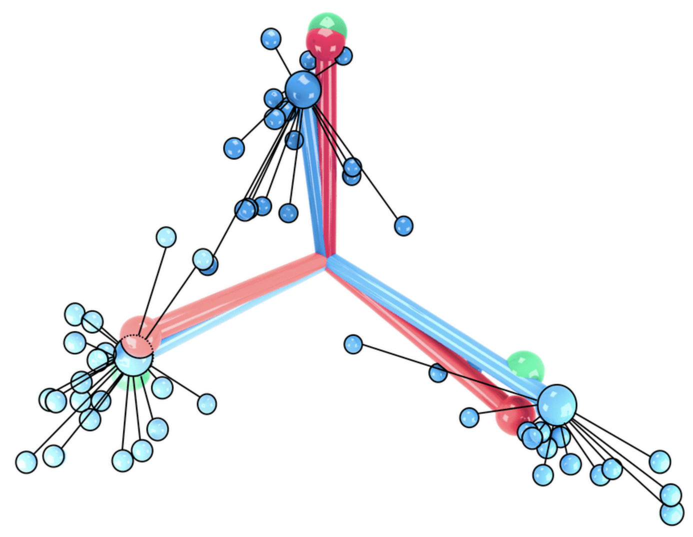
Continual Learning



No access to previous data throughout learning process

Neural Collapse

All representations converge towards a unique prototype for each class



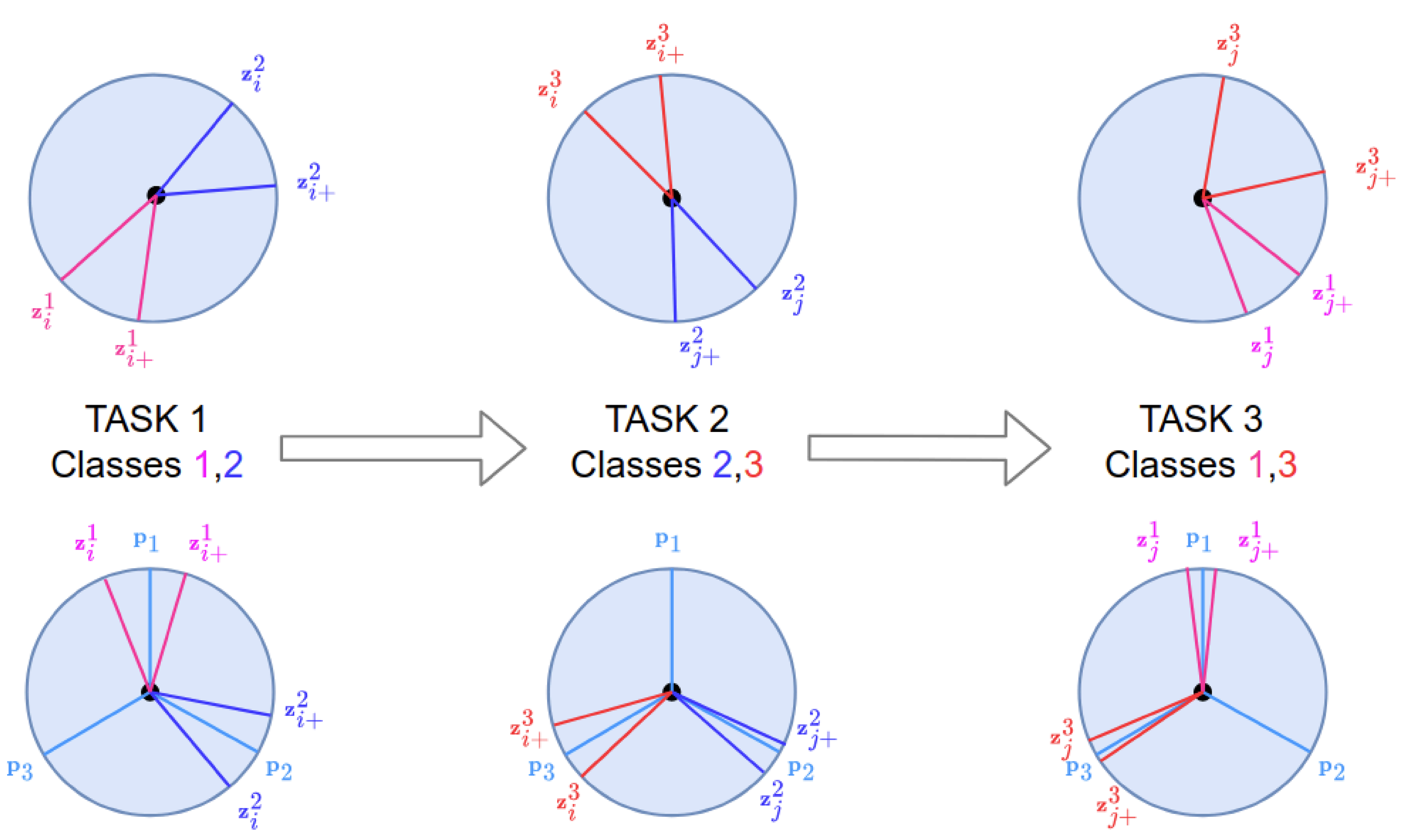
Matrix Representation

$$\mathbf{P} = \sqrt{\frac{L}{L-1}}\mathbf{U}(\mathbf{I}_L - \frac{1}{L}\mathbf{1}_L\mathbf{1}_L^T)$$

Cosine distance

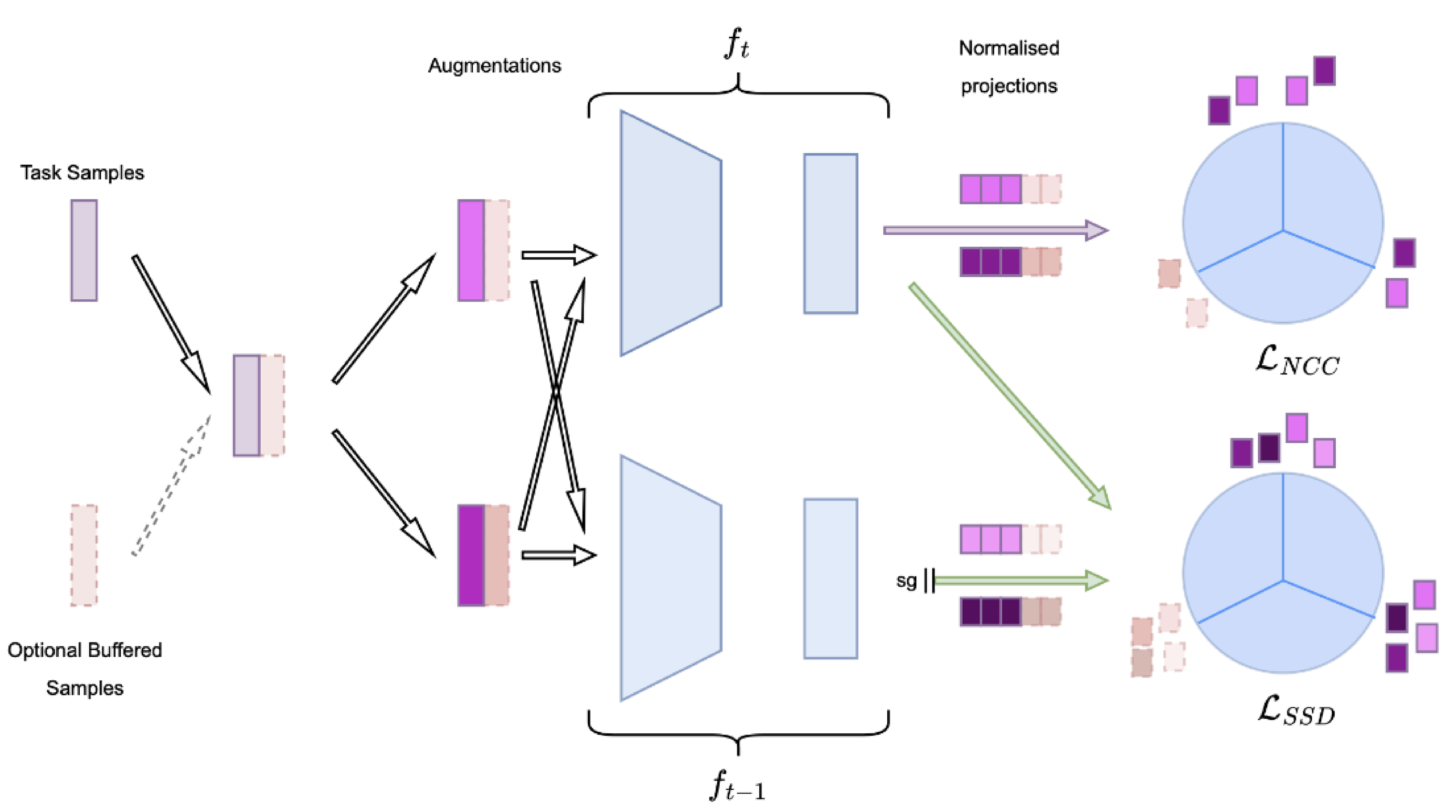
$$\mathbf{p}_i^\top \mathbf{p}_j = \frac{L}{L-1}\delta_{i,j} - \frac{1}{L-1}, \forall i,j \in [1,L]$$

Method



Plasticity loss [3]

$$\mathcal{L}_{NCC} = \sum_{i=1}^{2N} \ell_{NCC}(\mathbf{x}_i) = \sum_{i=1}^{2N} \frac{-1}{|P(i)|}\left[\sum_{p\in P(i)} \log \frac{\exp(\mathbf{z}_i \cdot \mathbf{z}_p/\tau)}{\sum_{\substack{j=1\\ j\neq i}}^{2N} \exp(\mathbf{z}_i \cdot \mathbf{z}_j/\tau)} + \log \frac{\exp(\mathbf{z}_i \cdot \mathbf{p}_{y_i}/\tau)}{\sum_{\substack{j=1\\ j\neq i}}^{L} \exp(\mathbf{z}_i \cdot \mathbf{p}_{y_j}/\tau)}\right]$$



Stability loss

$$\mathcal{L}_{SSD} = \sum_{i=1}^{N} \mathcal{D}_{KL}(P(\mathbf{Z}^{t-1};\mathbf{p}_i)||P(\mathbf{Z}^{t};\mathbf{p}_i))$$

Quantitative Results

Buffer	Method	Seq-CIFAR-10 C-IL	Seq-CIFAR-10 T-IL	Seq-Tiny-ImageNet C-IL	Seq-Tiny-ImageNet T-IL
200	GEM [20]	25.54±0.76	90.44±0.94	-	-
	A-GEM [5]	20.04±0.34	83.88±1.49	8.07±0.08	22.77±0.03
	iCaRL [25]	49.02±3.20	88.99±2.13	7.53±0.79	28.19±1.47
	DER [2]	61.93±1.79	91.40±0.92	11.87±0.78	40.22±0.67
	DER++ [2]	64.88±1.17	91.92±0.60	10.96±1.17	40.87±1.16
	Co^{2}L [3]	65.57±1.37	93.43±0.78	13.88±0.40	42.37±0.74
	Ours	**68.97**±1.28	**95.04**±0.73	**15.32**±0.52	**43.57**±0.41
500	GEM [20]	26.20±1.26	93.81±0.27	-	-
	A-GEM [5]	22.67±0.57	89.48±1.45	8.06±0.04	25.33±0.49
	iCaRL [25]	47.55±3.95	88.22±2.62	9.38±1.53	31.55±3.27
	DER [2]	70.51±1.65	93.40±0.39	17.75±1.14	51.78±0.88
	DER++ [2]	72.70±1.36	93.88±0.50	19.38±1.41	51.91±0.68
	Co^{2}L [3]	74.26±0.77	95.90±0.26	20.12±0.42	53.04±0.69
	Ours	**75.83**±0.79	**97.02**±0.59	**21.51**±0.35	**54.96**±0.43

Evaluation metric:

$$A = \frac{1}{T}\sum_{i=1}^{T} A_{T,i}$$

Learning Scenarii:

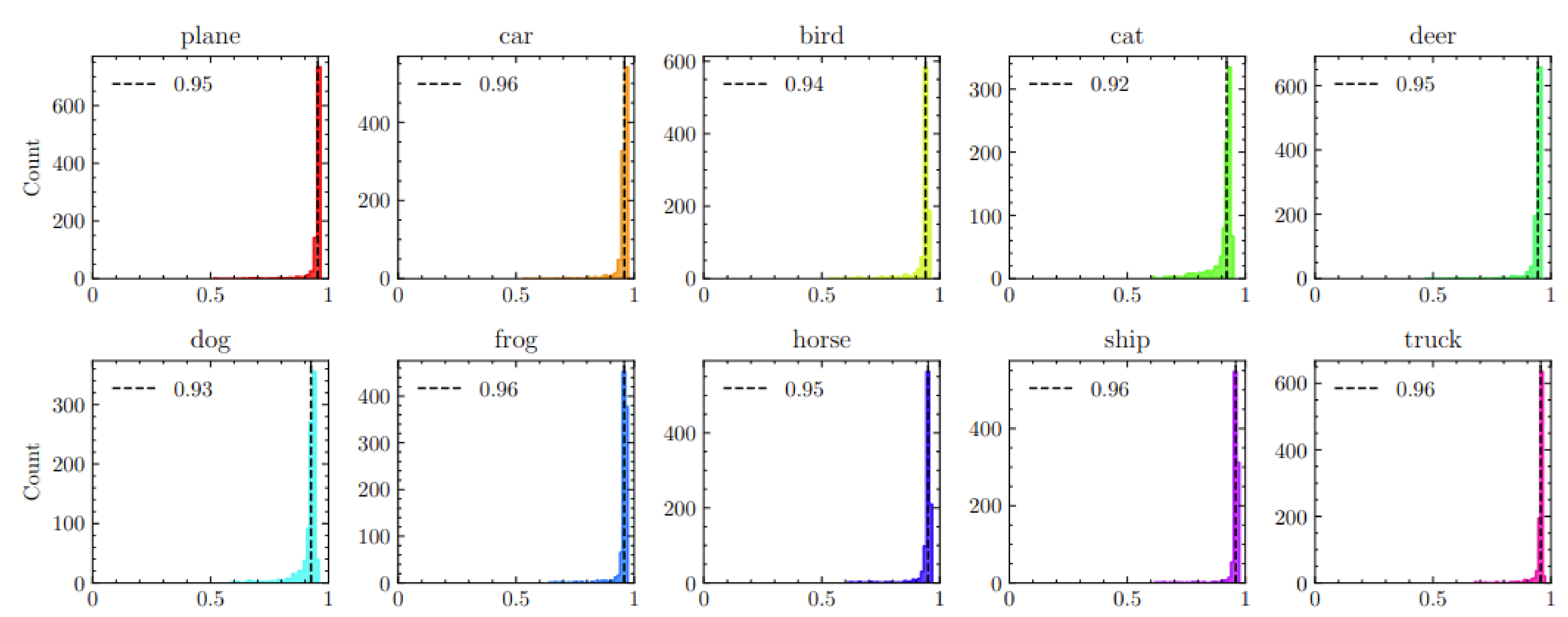
Both class-incremental and task-incremental setups have same structure but indices are given in task-incremental.

Qualitative Results

Method	Buffer size 0	Buffer size 200	Buffer size 500
Co^{2}L with $\mathcal{L}_{SCL}^{asym}$ & $\mathcal{L}_{IRD}$ [3]	53.57±1.03	65.57±1.37	74.26±0.77
$\mathcal{L}_{NCC}$ & $\mathcal{L}_{IRD}$	61.81±1.22	67.72±1.17	75.14±0.67
$\mathcal{L}_{NCC}$ & $\mathcal{L}_{SSD}$	**65.96**±1.17	**68.97**±1.28	**75.83**±0.79

Ablation study shows the importance of our stabilization loss, especially when no replay is allowed

Method	Buffer size 0	Buffer size 200	Buffer size 500
Co^{2}L with $\mathcal{L}_{SCL}^{asym}$ & $\mathcal{L}_{IRD}$ [3]	53.57±1.03	65.57±1.37	74.26±0.77
$\mathcal{L}_{NCC}$ & $\mathcal{L}_{IRD}$	61.81±1.22	67.72±1.17	75.14±0.67
$\mathcal{L}_{NCC}$ & $\mathcal{L}_{SSD}$	**65.96**±1.17	**68.97**±1.28	**75.83**±0.79



Correlation between closeness to prototypes and class accuracy is shown above

[3] Hyuntak Cha, Jaeho Lee, and Jinwoo Shin. Co^{2}l: Contrastive continual learning. In *CVPR*, 2021.

BMVC 2024