Attraction Diminishing and Distributing for Few-Shot Class-Incremental Learning

Supplementary Material

In Sec. 6, we provide the supplementary proof for the analysis in Sec. 3.2. In Sec. 7, we present more extensive comparison experimental results.

6. Proof of Output Distribution

In this section, we present the detailed proof of the output distribution of the feature extractor after base training, as described in the main manuscript. To facilitate the subsequent analysis, we rewrite the optimization objective for the base training phase (Eq. (1)) as follows,

$$\mathcal{L}_{base} = \frac{1}{|\mathcal{D}^0_{train}|}\mathcal{L}_{\text{CE}}, \tag{19}$$

$$\begin{aligned}\mathcal{L}_{\text{CE}} &= \sum_{c=1}^{|\mathcal{C}^0|}\sum_{i=1}^{Num_c} -\log\frac{\exp((\mathbf{z}_{c,i})_c)}{\exp((\mathbf{z}_{c,i})_c)+\sum_{j\neq c}\exp((\mathbf{z}_{c,i})_j)} \\ &= \sum_{c=1}^{|\mathcal{C}^0|}\sum_{i=1}^{Num_c} \log\left(1+\frac{\sum_{j\neq c}\exp((\mathbf{z}_{c,i})_j)}{\exp((\mathbf{z}_{c,i})_c)}\right),\end{aligned} \tag{20}$$

where $(\mathbf{z}_{c,i})_j = \mathbf{w}'^{\top}_j\mathbf{f}_{c,i} \in \mathbb{R}$, $\mathbf{f}_{c,i} = f(\mathbf{x}_{c,i})$ represents the normalized feature vector of $\mathbf{x}_{c,i}$, $\mathbf{x}_{c,i}$ is the i-th sample in class c, and $\mathbf{w}'_j$ denotes the normalized classifier weight of class c.

Proof. Using Jensen's inequality, we obtain:

$$\sum_{j\neq c}\exp((\mathbf{z}_{c,i})_j) \geq (|\mathcal{C}^0|-1)\exp\left(\frac{\sum_{j\neq c}(\mathbf{z}_{c,i})_j}{|\mathcal{C}^0|-1}\right),$$

with equality *iff* $(\mathbf{z}_{c,i})_j$ are equal for all $j \neq c$. Substituting this bound in the loss function:

$$\begin{aligned}\mathcal{L}_{\text{CE}} \geq \sum_{c=1}^{|\mathcal{C}^0|}\sum_{i=1}^{\text{Num}_c}\log\Bigg(1+(|\mathcal{C}^0|-1) \\ \times\exp\left(\frac{\sum_{j\neq c}(\mathbf{z}_{c,i})_j}{|\mathcal{C}^0|-1}-(\mathbf{z}_{c,i})_c\right)\Bigg).\end{aligned} \tag{21}$$

Let $\bar{\mathbf{w}} = \frac{1}{|\mathcal{C}^0|}\sum_{c=1}^{|\mathcal{C}^0|}\mathbf{w}'_c$, the following inequality holds:

$$(\bar{\mathbf{w}}-\mathbf{w}'_c)^{\top}\mathbf{f}_{c,i} \geq -\|\bar{\mathbf{w}}-\mathbf{w}'_c\|_2\|\mathbf{f}\|_2. \tag{22}$$

Equality is achieved if $\mathbf{f}_{c,i} = -\|\mathbf{f}\|_2\frac{\bar{\mathbf{w}}-\mathbf{w}'_c}{\|\bar{\mathbf{w}}-\mathbf{w}'_c\|_2}$. Substituting these results into the loss,

$$\begin{aligned}\mathcal{L}_{\text{CE}} &\geq \sum_{c=1}^{|\mathcal{C}^0|}\sum_{i=1}^{\text{Num}_c}\log\Bigg(1+(|\mathcal{C}^0|-1) \\ &\quad\times\exp\left(\frac{\sum_{j\neq c}\mathbf{w}'^{\top}_j\mathbf{f}_{c,i}}{|\mathcal{C}^0|-1}-\mathbf{w}'^{\top}_c\mathbf{f}_{c,i}\right)\Bigg) \qquad (23)\\ &= \sum_{c=1}^{|\mathcal{C}^0|}\sum_{i=1}^{\text{Num}_c}\log\Bigg(1+(|\mathcal{C}^0|-1) \\ &\quad\times\exp\left(\frac{|\mathcal{C}^0|(\bar{\mathbf{w}}-\mathbf{w}'_c)^{\top}\mathbf{f}_{c,i}}{|\mathcal{C}^0|-1}\right)\Bigg) \qquad (24)\\ &\geq \text{Num}\sum_{c=1}^{|\mathcal{C}^0|}\log\Bigg(1+(|\mathcal{C}^0|-1) \\ &\quad\times\exp\left(-\frac{|\mathcal{C}^0|\|\bar{\mathbf{w}}-\mathbf{w}'_c\|_2\|\mathbf{f}\|_2}{|\mathcal{C}^0|-1}\right)\Bigg). \qquad (25)\end{aligned}$$

Since the number of training samples for each class is the same or approximately the same, we replace Num_c with a unified Num.

Applying Jensen's inequality for $\sqrt{x}$:

$$\begin{aligned}\mathcal{L}_{\text{CE}} &\geq \text{Num}\log\Bigg(1+(|\mathcal{C}^0|-1) \\ &\quad\times\exp\left(-\frac{|\mathcal{C}^0|}{|\mathcal{C}^0|-1}\sum_{c=1}^{|\mathcal{C}^0|}\|\bar{\mathbf{w}}-\mathbf{w}'_c\|_2\|\mathbf{f}\|_2\right)\Bigg) \qquad (26)\\ &\geq \text{Num}\log\Bigg(1+(|\mathcal{C}^0|-1) \\ &\quad\times\exp\left(-\frac{|\mathcal{C}^0|}{|\mathcal{C}^0|-1}\sqrt{|\mathcal{C}^0|\sum_{c=1}^{|\mathcal{C}^0|}\|\bar{\mathbf{w}}-\mathbf{w}'_c\|_2^2}\|\mathbf{f}\|_2\right)\Bigg) \qquad (27)\\ &\geq \text{Num}\log\Bigg(1+(|\mathcal{C}^0|-1) \\ &\quad\times\exp\left(-\frac{|\mathcal{C}^0|}{|\mathcal{C}^0|-1}\sqrt{|\mathcal{C}^0|}\|\mathbf{W}\|_F\|\mathbf{f}\|_2\right)\Bigg), \qquad (28)\end{aligned}$$

with equality *iff* $\forall c$, $\|\bar{\mathbf{w}}-\mathbf{w}'_c\|_2$ is equal, and $\bar{\mathbf{w}} = 0$.

Therefore, the cross-entropy loss achieves its lower bound, *iff*:

1. For every feature vector $\mathbf{f}_{c,i}$ belonging to class c,

$$\mathbf{w'}_m^{\top}\mathbf{f}_{c,i} = \mathbf{w'}_n^{\top}\mathbf{f}_{c,i}, \quad \forall m, n \neq c,$$

2. Each feature vector $\mathbf{f}_{c,i}$ satisfy:

$$\mathbf{f}_{c,i} = -\|\mathbf{f}\|_2 \frac{\bar{\mathbf{w}} - \mathbf{w'}_c}{\|\bar{\mathbf{w}} - \mathbf{w'}_c\|_2},$$

3. The weight vectors $\mathbf{w'}_c$ satisfy:

$$\forall c, \|\bar{\mathbf{w}} - \mathbf{w'}_c\|_2 = \text{const}\,, \bar{\mathbf{w}} = 0.$$

When the above conditions are met, all features of the same class converge to a single point (i.e., $\mathbf{f}_{c,i} = \mathbf{w'}_c, \forall i$), we have

$$\mu_c = \frac{1}{\text{Num}_c}\sum_{i=1}^{\text{Num}_c}\mathbf{f}_{c,i} = \mathbf{w'}_c,$$

$$\mathbf{\Sigma}_c = \frac{1}{\text{Num}_c}\sum_{i=1}^{\text{Num}_c}(\mathbf{f}_{c,i} - \mu_c)(\mathbf{f}_{c,i} - \mu_c)^{\top} = 0,$$

and

$$\langle \mathbf{w'}_c, \mathbf{w'}_j \rangle = -\frac{1}{|\mathcal{C}^0| - 1}, \quad \forall j \neq c.$$

That is, when the loss function reaches its lower bound, the output distribution collapses to $|\mathcal{C}^0|$ points defined by the classifiers, and these points are uniformly distributed across the feature space. At this point, the output of the feature extractor $f(\mathbf{x}) \in \mathbb{R}^d$ for i.i.d. samples can be represented as a mixture distribution:

$$\mathcal{P}(\mathbf{a}) = \sum_{c=1}^{|\mathcal{C}^0|+\delta} \pi_c \mathbf{N}(\mathbf{a}; \mu_c, \mathbf{\Sigma}_c),$$

where $\delta = 0$. Each Gaussian component in the mixture distribution degenerates into a Dirac delta distribution, i.e.,

$$\mathbf{N}(\mathbf{a}; \mu_c, \mathbf{\Sigma}_c) \rightarrow \delta(\mathbf{a} - \mu_c),$$

where μ_c is the mean of the c-th component, and $\delta(\cdot)$ is the Dirac delta function. In this case, any given point $\mathbf{a}$ lies within one of the components of the mixture.

7. More Results

In this section, we conduct a more extensive and detailed performance comparison with state-of-the-art FSCIL methods on three datasets using different backbones, as shown in Tabs. 4 to 9. Additionally, we evaluate our method on larger and more powerful models, as shown in Tabs. 10 and 11.

Our method consistently achieves performance improvements across all backbones, with particularly significant gains observed in the later sessions. The performance of SAVC varies noticeably across different backbones, indicating the potential for further enhancement through model modifications and optimized training strategies. Consequently, we introduce targeted adjustments to SAVC to achieve additional performance gains. In the tables, † represents model adjustments, and * denotes training strategy adjustments. Notably, our method achieves remarkable improvements even when starting from higher baseline accuracy levels, demonstrating its robustness and effectiveness.

Backbone	Method	Overall accuracy in each session (%)								
		0	1	2	3	4	5	6	7	8
ResNet18	Decoupled-Cosine [29]	70.05	65.15	61.09	57.87	55.05	52.19	49.56	47.58	45.92
	CEC [37]	72.00	66.83	62.97	59.43	56.70	53.73	51.19	49.24	47.63
	FACT [41]	75.32	70.34	65.84	62.05	58.68	55.35	52.42	50.42	48.51
	CLOM [45]	72.08	67.28	63.30	59.85	56.82	53.82	51.08	49.15	47.95
	TEEN [32]	73.53	70.55	66.37	63.23	60.53	57.95	55.24	53.44	52.08
	Bidist [40]	74.65	70.43	66.29	62.77	60.75	57.24	54.79	53.65	52.22
	OSHHG [7]	60.65	59.00	56.59	54.78	53.02	50.73	48.46	47.34	46.75
	EHS [8]	71.25	66.65	62.84	59.65	56.90	54.14	51.63	50.05	49.06
	DyCR [20]	73.18	70.16	66.87	63.43	61.18	58.79	55.00	52.87	51.08
	M2SD [17]	82.11	79.92	75.44	71.31	68.29	64.32	61.13	58.64	56.51
	CLOSER [19]	76.02	71.61	67.99	64.69	61.70	58.94	56.23	54.52	53.33
	SAVC [28]	80.93	75.97	71.83	68.21	65.54	61.91	58.93	57.41	55.97
	SAVC+D2A (Ours)	81.20	75.06	71.44	68.09	66.16	63.05	60.02	58.48	57.72
	SAVC†* [28]	83.93	78.82	74.63	70.96	67.74	63.93	60.59	58.55	57.42
	*SAVC+D2A (Ours)†**	85.25	79.43	75.34	71.84	69.24	65.64	62.54	60.62	59.60
	SAVC† [28]	85.92	81.19	77.13	73.16	70.63	66.79	63.89	61.96	60.57
	SAVC+D2A (Ours)†	**86.90**	**81.97**	**78.09**	**74.05**	**71.73**	**67.98**	**64.72**	**63.19**	**61.99**
ResNet12	C-FSCIL [12]	76.40	71.14	66.46	63.29	60.42	57.46	54.78	53.11	51.41
	NC-FSCIL [34]	84.02	76.80	72.00	67.83	66.35	64.04	61.46	59.54	58.31
	SAVC* [28]	86.80	82.32	78.61	75.23	72.21	68.56	65.14	63.83	62.98
	*SAVC+D2A (Ours)**	**87.23**	**82.49**	**78.94**	**75.45**	**73.14**	**69.93**	**66.74**	**65.56**	**64.29**

Table 4. Comparison with SOTA methods on the *mini*ImageNet dataset.

Backbone	Method	Incremental class accuracy in each session (%)							
		1	2	3	4	5	6	7	8
ResNet18	Decoupled-Cosine [29]	11.20	12.70	13.87	14.10	13.36	12.43	13.09	13.40
	CEC [37]	15.40	17.10	16.67	15.90	14.20	13.63	14.20	14.78
	FACT [41]	15.80	14.40	15.40	14.55	13.64	12.20	12.66	13.20
	CLOM [45]	13.00	14.20	14.07	14.00	12.88	12.20	13.26	15.05
	TEEN [32]	40.20	35.60	32.47	32.70	29.96	28.33	28.89	29.35
	Bidist [40]	27.00	30.30	29.60	27.85	28.44	26.80	26.23	25.62
	OSHHG [7]	39.26	32.26	31.31	30.12	26.95	24.08	24.51	25.89
	DyCR [20]	14.40	13.80	12.93	13.95	12.12	11.20	12.31	13.90
	CLOSER [19]	30.20	28.70	27.93	26.90	25.16	23.67	24.26	25.78
	SAVC [28]	38.00	33.90	30.80	30.60	26.68	24.80	26.23	27.13
	SAVC+D2A (Ours)	46.00	40.20	38.40	**40.05**	36.12	**33.83**	34.17	**36.33**
	SAVC†* [28]	27.20	27.00	27.40	25.95	22.68	20.50	21.54	23.88
	*SAVC+D2A (Ours)†**	30.20	34.00	32.70	32.70	28.68	27.07	28.06	30.18
	SAVC† [28]	**49.20**	**43.10**	37.93	38.15	34.16	32.43	32.69	33.60
	SAVC+D2A (Ours)†	45.80	42.90	**38.73**	39.75	**36.32**	33.50	**34.89**	36.10
ResNet12	C-FSCIL [12]	5.20	12.40	17.27	19.90	23.40	22.60	25.91	25.95
	NC-FSCIL [34]	**54.20**	**48.10**	**42.73**	**40.55**	34.88	32.00	32.09	31.33
	SAVC* [28]	38.40	38.90	39.13	39.35	35.64	32.97	34.94	37.05
	*SAVC+D2A (Ours)**	34.00	36.20	36.20	39.05	**37.20**	**35.30**	**38.06**	**40.13**

Table 5. Comparison with SOTA methods on the *mini*ImageNet dataset.

Backbone	Method	Overall accuracy in each session (%)								
		0	1	2	3	4	5	6	7	8
ResNet20	Decoupled-Cosine [29]	71.92	67.60	63.39	59.45	56.20	53.53	51.14	48.80	46.70
	CEC [37]	73.07	68.88	65.26	61.19	58.09	55.57	53.22	51.34	49.14
	FACT [41]	78.80	72.40	68.33	64.31	61.07	58.11	56.23	54.07	52.13
	CLOM [45]	74.07	70.06	65.90	61.92	58.83	55.75	53.62	51.55	49.56
	TEEN [32]	**78.92**	72.32	68.16	64.43	61.19	58.48	56.11	54.03	51.87
	EHS [8]	73.98	70.11	66.66	62.75	60.11	57.33	55.59	53.75	51.59
	DyCR [20]	75.73	**73.29**	68.71	64.80	62.11	59.25	56.70	54.56	52.24
	CLOSER [19]	75.72	71.83	68.32	64.62	61.91	59.25	57.53	55.43	53.32
	SAVC [28]	78.47	72.31	67.49	62.41	59.10	55.95	53.81	51.54	49.16
	SAVC+D2A (Ours)	77.13	71.91	**69.16**	**65.23**	**62.25**	**59.60**	**57.89**	**56.23**	**53.96**
ResNet18	Bidist [40]	69.33	65.52	61.83	58.04	55.09	52.14	49.91	48.36	46.05
	OSHHG [7]	63.55	62.88	61.05	58.13	55.68	54.59	52.93	50.39	49.48
	SAVC†* [28]	81.60	75.95	71.36	67.17	63.75	60.59	57.77	55.51	53.20
	*SAVC+D2A (Ours)†**	**83.12**	**78.25**	**72.44**	**67.53**	63.51	**60.61**	**58.33**	56.02	53.67
	SAVC† [28]	81.62	76.08	71.66	67.48	63.69	60.34	57.60	55.25	53.13
	SAVC+D2A (Ours)†	82.58	77.88	72.41	67.20	**63.78**	60.55	58.32	**56.11**	**53.73**
ResNet12	C-FSCIL [12]	77.47	72.40	67.47	63.25	59.84	56.95	54.42	52.47	50.47
	NC-FSCIL [34]	82.52	76.82	73.34	69.68	66.19	62.85	60.96	59.02	56.11
	RDI [42]	82.52	78.00	73.69	69.73	66.27	63.48	61.46	59.66	57.48
	SAVC [28]	85.72	80.28	76.21	71.37	68.16	65.66	63.56	61.54	59.65
	SAVC+D2A (Ours)	**86.02**	**82.18**	**78.26**	**73.97**	**70.90**	**68.06**	**66.41**	**64.36**	**61.71**

Table 6. Comparison with SOTA methods on the CIFAR100 dataset.

Backbone	Method	Incremental class accuracy in each session (%)							
		1	2	3	4	5	6	7	8
ResNet20	Decoupled-Cosine [29]	22.20	19.30	17.33	16.05	16.80	16.70	15.71	15.45
	CEC [37]	27.20	23.80	20.07	19.50	20.56	20.57	20.11	19.50
	FACT [41]	31.20	28.40	24.33	23.00	22.44	23.43	22.29	21.55
	CLOM [45]	24.00	19.90	17.40	16.85	16.20	17.07	17.00	17.10
	TEEN [32]	33.00	30.10	27.33	26.00	25.56	25.00	24.20	23.60
	DyCR [20]	31.60	27.00	22.87	21.85	21.36	21.30	20.43	20.22
	CLOSER [19]	**35.80**	**33.30**	**29.33**	**30.05**	**28.28**	29.20	28.34	27.00
	SAVC [28]	30.60	23.50	20.13	20.15	21.36	22.43	21.77	21.15
	SAVC+D2A (Ours)	34.00	29.70	26.73	27.20	28.04	**29.47**	**29.17**	**27.70**
ResNet18	Bidist [40]	36.80	31.20	28.80	25.75	24.36	22.87	22.43	20.35
	OSHHG [7]	54.79	46.03	36.43	32.05	33.07	31.70	27.84	28.38
	SAVC†* [28]	33.60	24.80	21.47	19.95	19.12	18.30	17.97	18.40
	SAVC+D2A (Ours)†*	53.40	51.30	43.80	40.80	40.76	41.00	38.77	38.93
	SAVC† [28]	30.40	25.60	22.20	20.05	19.04	18.50	18.00	18.88
	SAVC+D2A (Ours)†	**57.80**	**53.70**	**44.33**	**42.15**	**41.80**	**41.20**	**39.20**	**39.13**
ResNet12	C-FSCIL [12]	18.00	13.60	13.00	12.65	15.36	16.30	15.74	16.73
	NC-FSCIL [34]	44.00	41.60	36.47	31.95	31.32	33.97	31.31	29.30
	SAVC [28]	**54.60**	**45.60**	38.73	37.25	38.20	37.83	37.00	36.70
	SAVC+D2A (Ours)	45.40	44.30	**39.53**	**38.70**	**38.28**	**39.87**	**38.86**	**37.10**

Table 7. Comparison with SOTA methods on the CIFAR100 dataset.

Method	Overall accuracy in each session (%)										
	0	1	2	3	4	5	6	7	8	9	10
Decoupled-Cosine [29]	75.35	71.30	67.41	63.09	60.30	56.80	54.94	53.13	51.04	50.11	48.81
CEC [37]	75.85	71.94	68.50	63.50	62.43	58.27	57.73	55.81	54.83	53.52	52.28
FACT [41]	75.90	73.23	70.84	66.13	65.56	62.15	61.74	59.83	58.41	57.89	56.94
CLOM [45]	79.57	76.07	72.94	69.82	67.80	65.56	63.94	62.59	60.62	60.34	59.58
TEEN [32]	77.26	76.13	72.81	68.16	67.77	64.40	63.25	62.29	61.19	60.32	59.31
Bidist [40]	75.91	72.32	70.12	66.04	64.37	62.18	60.71	59.62	57.41	56.68	55.94
NC-FSCIL [34]	80.45	75.98	72.30	70.28	68.17	65.16	64.43	63.25	60.66	60.01	59.44
OSHHG [7]	63.20	62.61	59.83	56.82	55.07	53.06	51.56	50.05	47.50	46.82	45.87
MgSvF [39]	72.29	70.53	67.00	64.92	62.67	61.89	59.63	59.15	57.73	55.92	54.33
DyCR [20]	77.50	74.73	71.69	67.01	66.59	63.43	62.66	61.69	60.57	59.69	58.46
M2SD [17]	81.49	76.67	73.58	68.77	68.73	65.78	64.73	64.03	62.70	62.09	60.96
CLOSER [19]	79.40	75.92	73.50	70.47	69.24	67.22	66.73	65.69	64.00	64.02	63.58
SAVC [28]	81.46	**78.55**	75.60	71.51	69.89	66.94	65.86	64.65	62.87	62.54	61.56
SAVC+D2A (Ours)	81.46	78.43	75.28	71.19	69.71	66.67	65.37	64.28	62.49	62.40	61.46
SAVC† [28]	**81.60**	78.27	**75.66**	**72.45**	71.00	68.66	67.75	66.18	64.58	64.43	63.55
SAVC+D2A (Ours)†	**81.60**	78.17	**75.66**	**72.45**	**71.05**	**68.72**	**67.97**	**66.59**	**65.04**	**64.94**	**64.19**

Table 8. Comparison with SOTA methods using ResNet18 on the CUB200 dataset.

Method	Incremental class accuracy in each session (%)									
	1	2	3	4	5	6	7	8	9	10
Decoupled-Cosine [29]	40.50	34.31	27.72	29.09	26.26	26.74	26.89	25.77	27.01	26.99
CEC [37]	40.14	37.45	30.90	33.82	31.32	32.75	32.91	31.46	33.22	33.02
FACT [41]	52.33	47.02	37.61	39.70	37.59	39.25	39.79	38.25	40.14	39.74
CLOM [45]	37.99	37.44	32.94	36.49	35.70	37.97	39.08	38.69	40.60	40.67
TEEN [32]	57.71	52.56	45.55	47.41	44.97	46.10	45.62	43.54	44.94	44.84
Bidist [40]	55.20	46.11	38.19	41.67	37.62	38.11	39.28	36.81	39.73	39.83
NC-FSCIL [34]	**66.67**	45.41	42.59	45.88	41.72	44.11	44.99	41.16	42.66	43.07
OSHHG [7]	56.81	43.17	35.16	34.99	33.03	32.41	31.51	28.13	28.86	28.73
DyCR [20]	46.24	42.76	37.64	41.64	40.41	42.01	42.08	39.51	41.04	41.90
CLOSER [19]	52.69	51.59	45.49	**48.20**	46.31	47.77	47.97	46.93	48.75	49.32
SAVC [28]	60.57	53.39	44.18	46.28	43.23	45.11	45.67	44.16	45.82	45.79
SAVC+D2A (Ours)	61.65	**55.15**	**45.59**	48.07	44.87	46.48	47.13	45.52	47.45	47.43
SAVC† [28]	46.24	47.49	43.65	46.55	44.95	47.12	46.72	45.71	47.54	47.61
SAVC+D2A (Ours)†	49.46	50.14	45.42	48.04	**46.33**	**48.85**	**48.83**	**47.68**	**49.46**	**49.63**

Table 9. Comparison with SOTA methods using ResNet18 on the CUB200 dataset.

Backbone	Method	Overall accuracy in each session (%)										
		0	1	2	3	4	5	6	7	8	9	10
ResNet50	Decoupled-Cosine	79.89	75.56	71.63	66.71	64.30	61.44	59.69	57.99	55.65	55.08	53.54
	Decoupled-Cosine+D2A (Ours)	**81.25**	**77.06**	**73.06**	**68.75**	**66.81**	**63.29**	**62.07**	**60.82**	**58.55**	**57.99**	**56.52**
ViT-B/16	Decoupled-Cosine	**88.09**	85.65	84.61	82.19	81.70	79.66	79.09	79.03	77.81	77.74	78.01
	Decoupled-Cosine+D2A (Ours)	87.88	**85.91**	**85.01**	**83.07**	**82.85**	**80.61**	**80.34**	**80.25**	**79.47**	**79.71**	**79.75**

Table 10. Comparison using larger models on the CUB200 dataset.

Backbone	Method	Incremental class accuracy in each session (%)									
		1	2	3	4	5	6	7	8	9	10
ResNet50	Decoupled-Cosine	45.88	38.70	29.24	32.39	31.06	32.13	32.38	30.62	32.39	31.76
	Decoupled-Cosine+D2A (Ours)	**55.56**	44.59	35.77	**38.79**	**35.89**	**37.55**	**38.72**	**36.54**	**38.16**	**37.43**
ViT-B/16	Decoupled-Cosine	67.03	70.09	64.79	67.83	64.65	65.86	67.64	66.48	67.59	69.12
	Decoupled-Cosine+D2A (Ours)	**69.89**	**72.72**	**68.56**	**71.65**	**67.71**	**69.21**	**70.57**	**70.14**	**71.76**	**72.68**

Table 11. Comparison using larger models on the CUB200 dataset.