Divide-and-Conquer for Enhancing Unlabeled Learning, Stability, and Plasticity in Semi-supervised Continual Learning

Yue Duan[1], Taicai Chen[1], Lei Qi[2], and Yinghuan Shi[1]

[1] Nanjing University, China [2] Southeast University, China

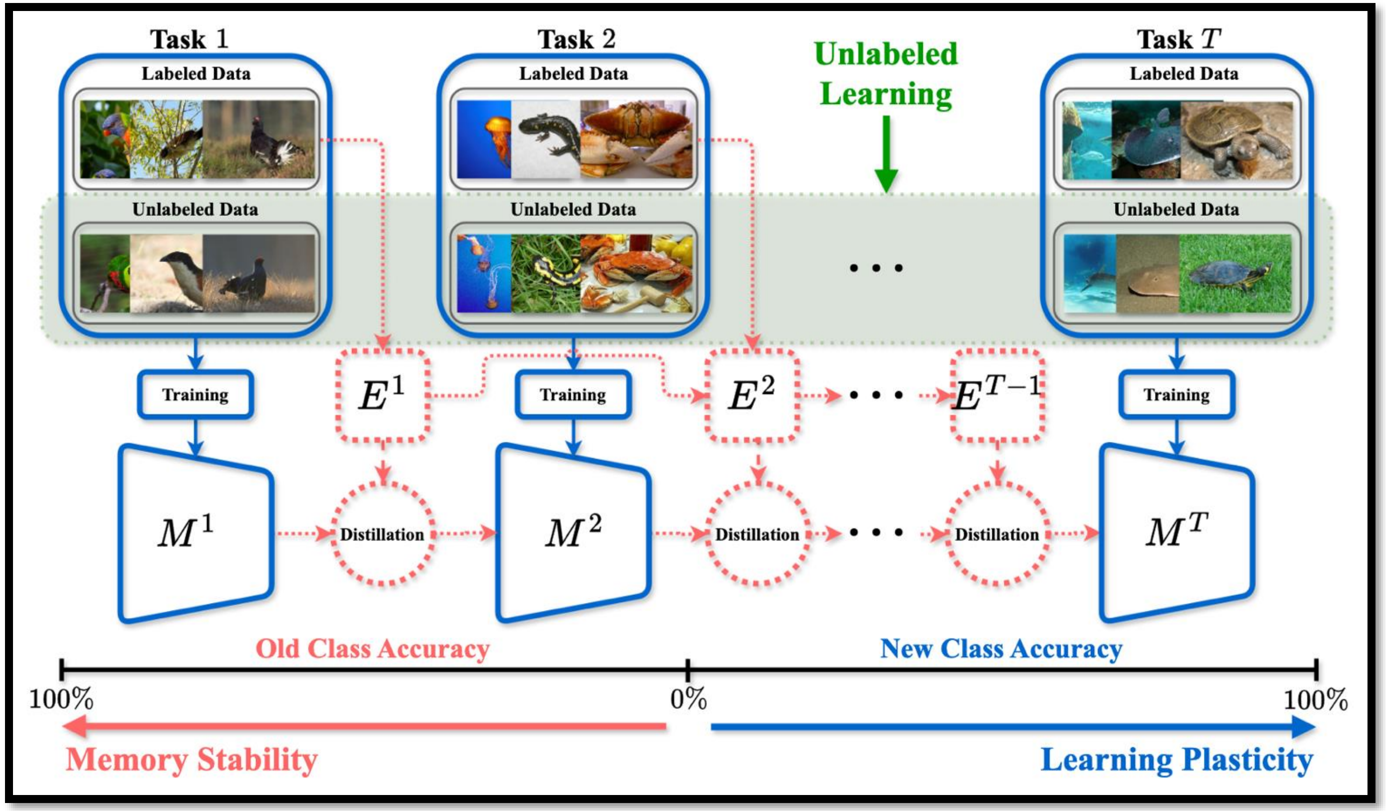
arXiv：https://arxiv.org/abs/2508.05316

GitHub：https://github.com/NJUyued/USP4SSCL

Zhihu（知乎）: https://zhuanlan.zhihu.com/p/1940131476688577441

Semi-supervised Continual Learning

- **Semi-supervised continual learning (SSCL)** faces a complex trade-off between
  - **Unlabeled Learning (UL)**
  - **Memory Stability (MS)**
  - **Learning Plasticity (LP)**
- This paper presents **USP**, a "divide-and-conquer" framework that systematically addresses these challenges with three synergistic modules.



Introduction

SSCL requires simultaneously learning from both labeled and unlabeled data across sequential tasks. Key challenges in SSCL include:

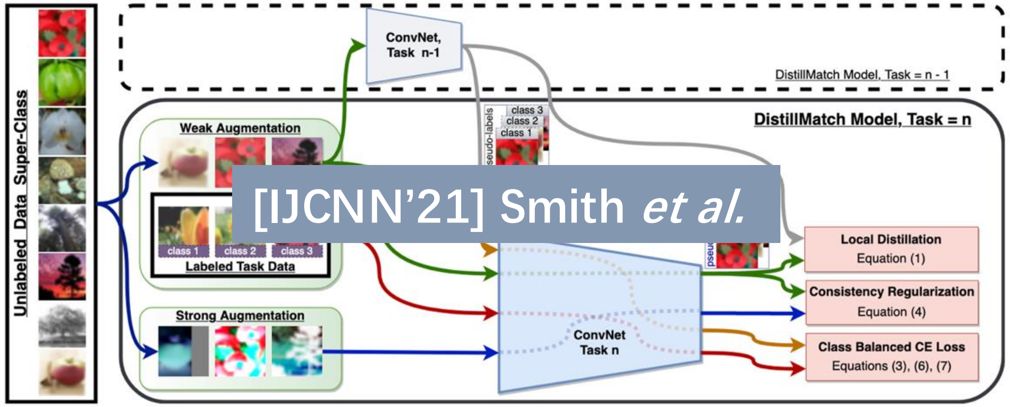
- **UL**: Anti-forgetting processes disrupt effective unlabeled learning, and common CL tech niques (e.g., replay) often underutilize unlabeled data.
- **MS**: Catastrophic forgetting of past tasks.
- **LP**: Overfitting to limited labeled samples.

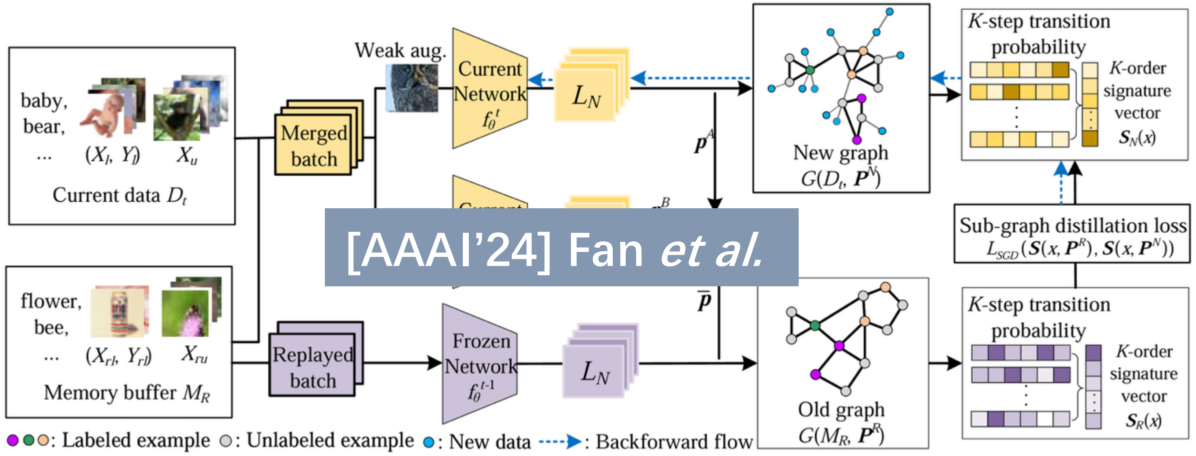
Prior SSCL methods typically address only one or two of these challenges.
We propose USP to holistically enhance UL, MS, and LP for SSCL.

Motivation

Intuition: **Unlabeled learning (UL)**, **memory stability (MS)**, and **learning plasticity (LP)**: Can they not all be achieved simultaneously?

- Previous approaches primarily focus on just one or two of the three core challenges:
  - DistillMatch employs pseudo-labeling technique to utilize unlabeled data for training **(UL)**
  - DSGD leverages semantic and structural information from unlabeled data **(MS)**





Baseline SSCL Learner

We can review SSCL as a optimization task:

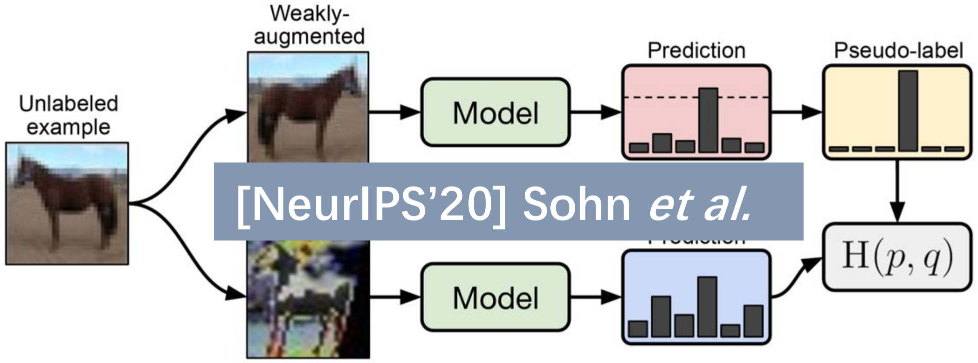
$$\min_{\theta} \sum_{t=1}^{T} \boxed{\mathcal{L}_{\mathtt{ssl}}(D^t)} + \boxed{\mathcal{L}_{\mathtt{cl}}(E^t)}$$

$$\mathcal{L}_{\mathtt{ssl}}(D^t) = \boxed{\mathcal{L}_{\mathtt{sup}}(D_l^t)} + \lambda_{\mathtt{uns}} \boxed{\mathcal{L}_{\mathtt{uns}}(D_u^t)}$$

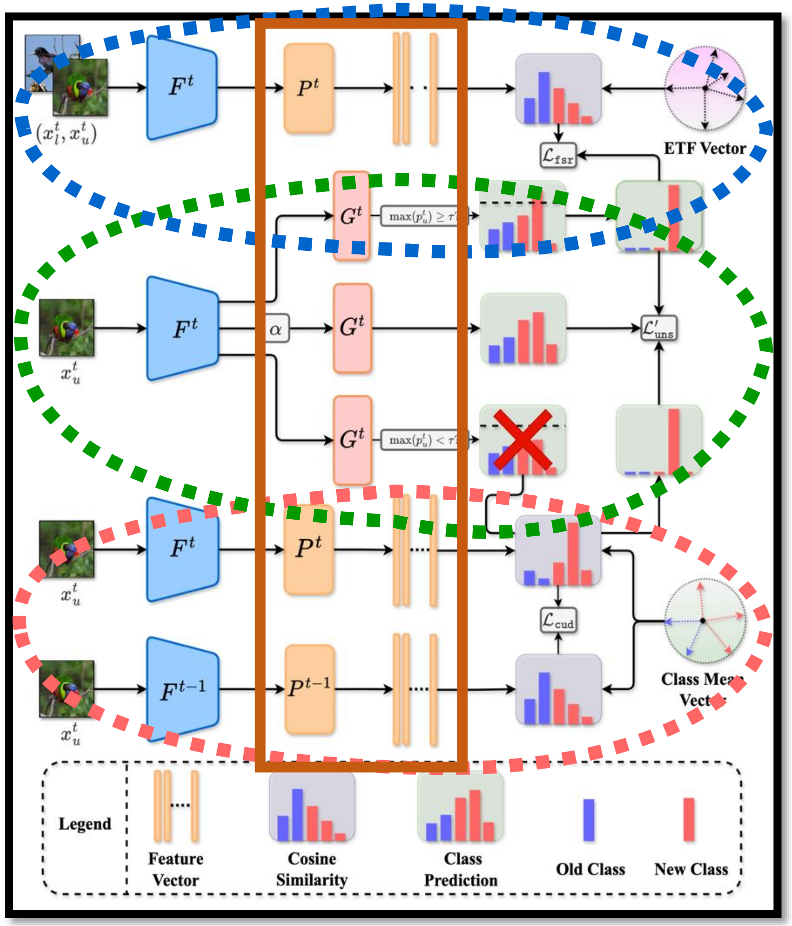
$$\mathcal{L}_{\mathtt{cl}}(E^t) = \mathbb{E}_{x_e^t \sim E^t}\left[\mathtt{KL}\left(\frac{p_e^t}{\beta} \,\|\, \frac{p_e^{t-1}}{\beta}\right)\right]$$

$$\mathcal{L}_{\mathtt{sup}}(D_l^t) = \mathbb{E}_{(x_l^t, y_l^t) \sim D_l^t}\left[H\left(p_{x_l^t}^t, y_l^t\right)\right]$$

$$\mathcal{L}_{\mathtt{uns}}(D_u^t) = \mathbb{E}_{x_u^t \sim D_u^t}\left[\mathbb{1}\left(\tilde{p}_{x_u^t}^t \geq \tau\right) H\left(p_{\alpha(x_u^t)}^t, \hat{p}_{x_u^t}^t\right)\right]$$

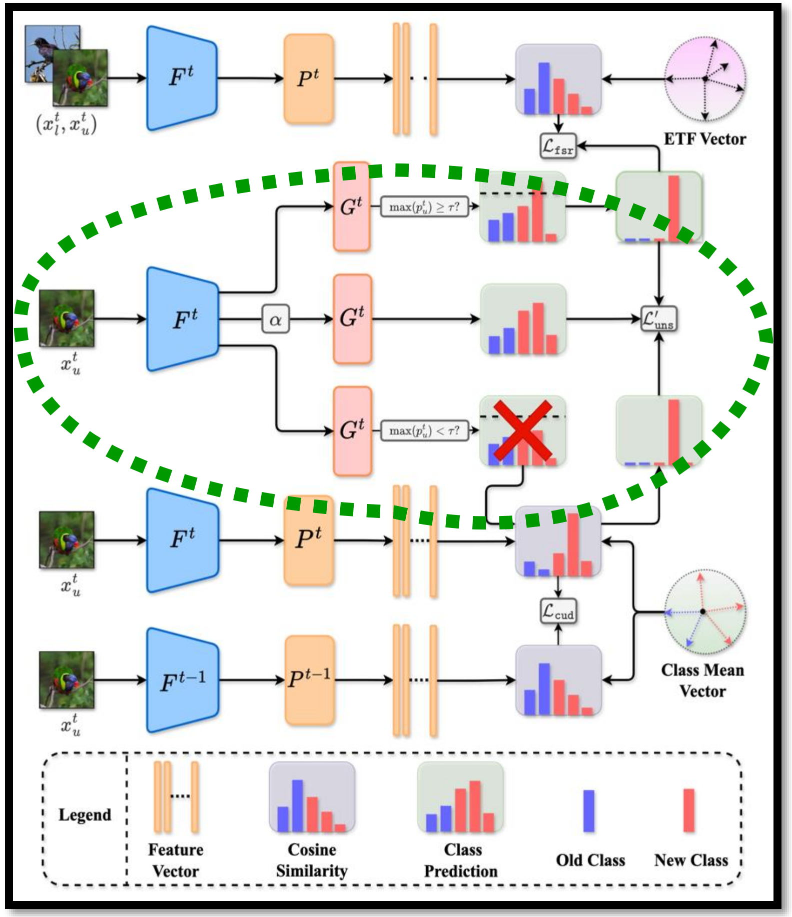


Method

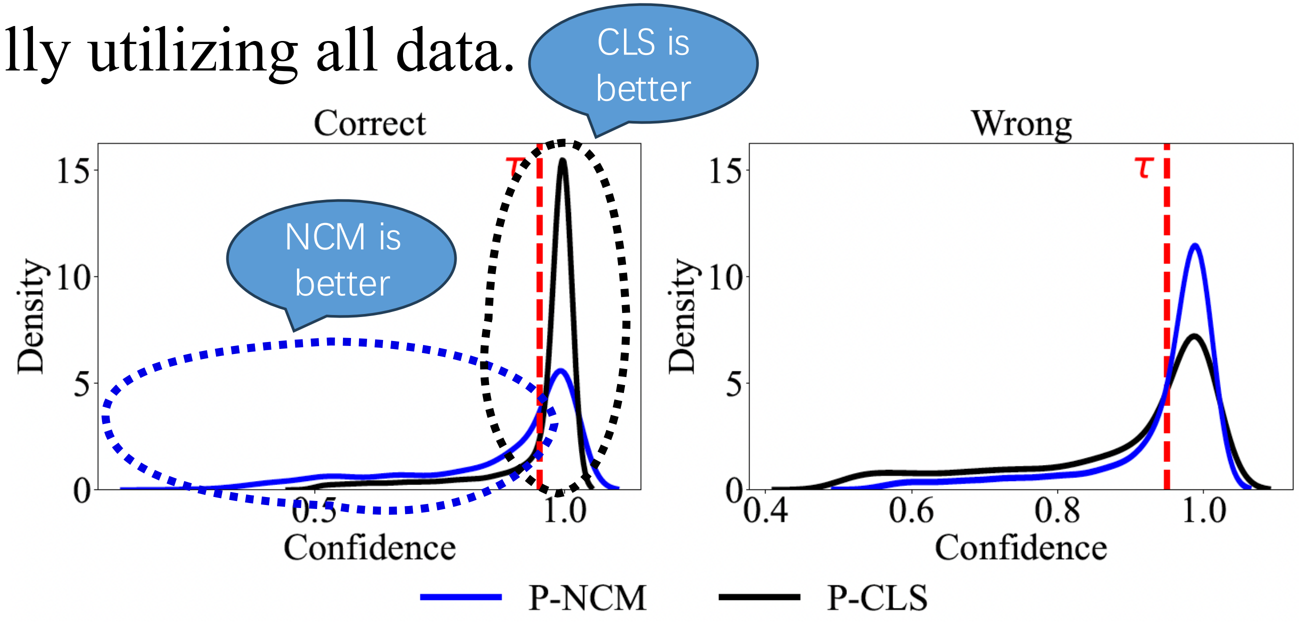


- **Feature Space Reservation (FSR) [For Plasticity]**
- **Divide-and-Conquer Pseudo-labeling (DCP) [For Unlabeled Learning]**
- **Class-mean-anchored Unlabeled Distillation (CUD) [For Stability]**
- All components of USP consistently leverage the output features from $\boxed{P(\cdot)}$ aiming to strengthen the coupling of all components for mutual reinforcement.

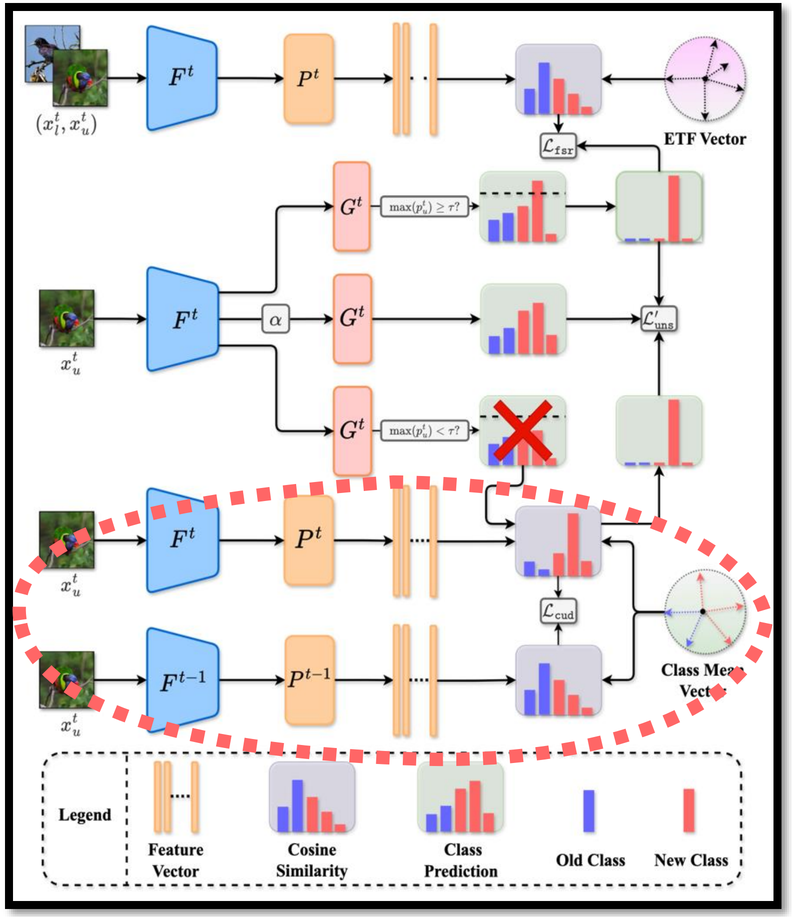
Divide-and-Conquer Pseudo-labeling (DCP)



- A dual-track strategy for unlabeled data: uses *classifier predictions* for *high-confidence* samples and a more robust *NCM method* for *low-confidence* ones, fully utilizing all data.



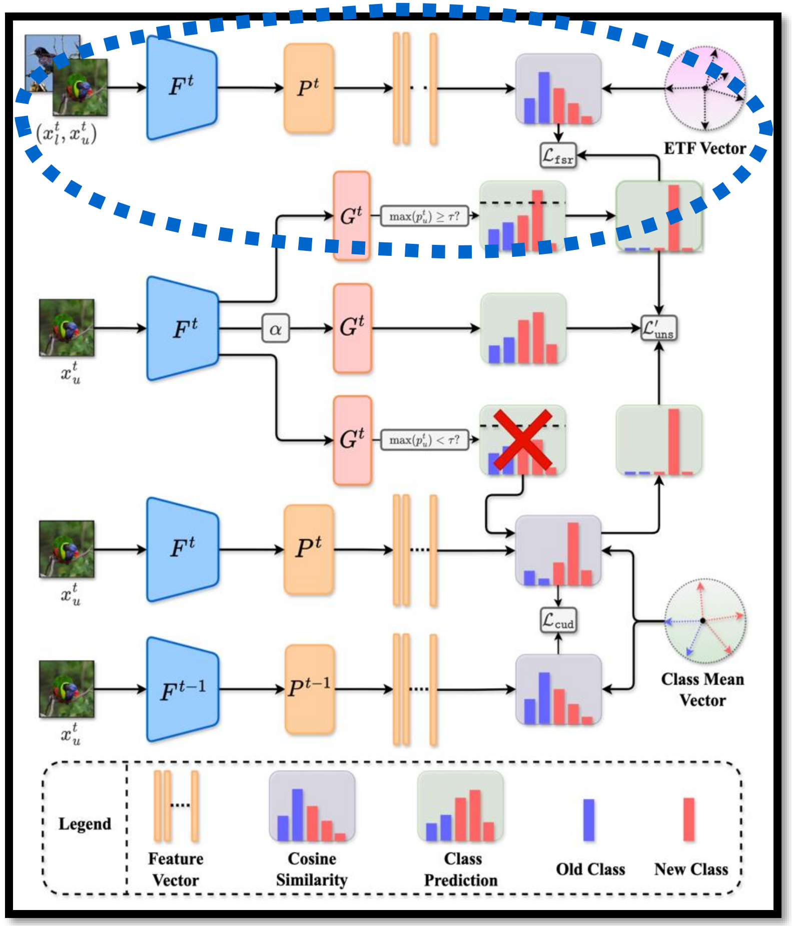
Class-mean-anchored Unlabeled Distillation (CUD)



- *Reuses class means from DCP* as stable anchors.
- CUD distills the combined *relationships between labeled and unlabeled data* by anchoring unlabeled samples to the stable class mean features derived from labeled data.
- CUD v.s. logit distillation and feature distillation

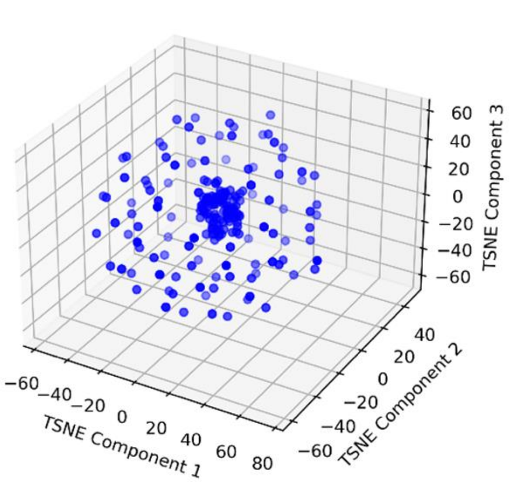
Method	Avg	Last
logit	53.91	37.97
feature	48.16	33.56
CUD	**54.36**	**38.25**

Feature Space Reservation (FSR)



- Reserves feature space for future classes using an *Equiangular Tight Frame (ETF)*. A contrastive loss aligns sample features to these pre-defined, geometrically optimal positions, preventing feature conflicts.

3D TSNE Visualization of ETF Vectors



$$\mathcal{L}_{\mathtt{fsr}}(D^t) = \mathbb{E}_{(x_l^t,y_l^t)\sim D_l^t}\left[-\log \frac{\exp\left(\frac{S(\mathcal{E}_{:,y_l^t}, f_{x_l^t}^t)}{\gamma}\right)}{\sum_{i=1}^k \exp\left(\frac{S(\mathcal{E}_{:,i}, f_{x_l^t}^t)}{\gamma}\right)}\right]$$

$$+ \mathbb{E}_{x_u^t\sim D_u^t}\left[\mathbb{1}\left(\tilde{p}_{x_u^t}^t \geq \tau\right) - \log \frac{\exp\left(\frac{S(\mathcal{E}_{:,\hat{p}_{x_u^t}^t}, f_{x_u^t}^t)}{\gamma}\right)}{\sum_{i=1}^k \exp\left(\frac{S(\mathcal{E}_{:,i}, f_{x_u^t}^t)}{\gamma}\right)}\right]$$

Experiment

Method	CIFAR10-30 Avg	CIFAR10-30 Last	CIFAR10-150 Avg	CIFAR10-150 Last	CIFAR100-20 Avg	CIFAR100-20 Last	CIFAR100-25 Avg	CIFAR100-25 Last	CIFAR100-80 Avg	CIFAR100-80 Last	CIFAR100-125 Avg	CIFAR100-125 Last
iCaRL [40]	34.16	21.84	60.86	53.65	26.43	13.92	28.14	15.29	36.32	19.10	44.14	30.73
DER [54]	40.41	31.48	64.77	61.06	31.01	23.53	32.82	26.53	53.32	41.55	57.21	48.86
CCIC [4]	-	55.20	-	74.30	-	29.50	-	29.50	-	-	-	44.30
ORDisCo [48]	-	-	74.77	65.91	-	-	-	-	-	-	-	-
DistillMatch [41]	-	-	-	-	-	-	-	-	-	37.00	-	-
NNCSL [24]	-	-	-	-	55.19	43.53	57.45	46.00	67.27	55.35	67.58	56.40
iCaRL&Fix [17]	45.98	30.71	78.36	69.08	45.75	23.40	49.83	31.25	53.46	32.21	56.87	41.38
+ DSGD [17]	77.33	**76.41**	84.14	**79.69**	52.80	35.47	53.42	35.95	57.92	37.81	58.08	43.14
+ USP (Ours)	79.66	70.43	**84.78**	78.21	53.20	41.30	54.36	38.25	58.59	44.20	59.96	43.80
DER&Fix [17]	66.71	61.41	81.1	77.00	51.76	40.86	52.03	44.47	64.03	50.25	66.69	53.57
+ DSGD [17]	75.04	72.59	83.08	79.39	55.63	44.63	57.94	46.68	65.48	55.4	69.14	58.5
+ USP (Ours)	**81.43**	73.65	84.43	77.74	**58.79**	**45.22**	**59.87**	**47.44**	**68.67**	**60.45**	**71.60**	**63.08**

- Dataset-X: number of labelled samples per class.
- CIFAR-10: 5 tasks; *Avg: **81.43** v.s. 75.04; Last: **73.65** v.s. 72.59*
- CIFAR-100: 10 tasks; *Avg: **68.67** v.s. 65.48; Last: **60.45** v.s. 55.40*

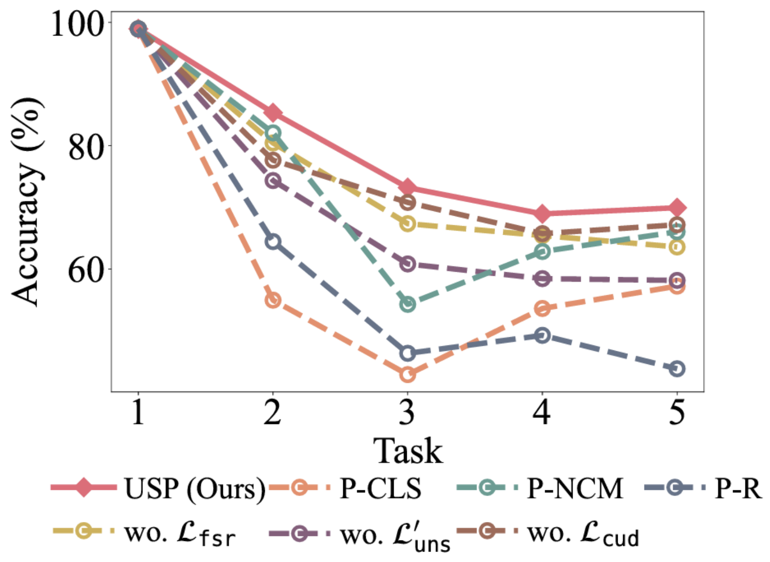
Experiment

Method	ImageNet100-13 Avg	ImageNet100-13 Last	ImageNet100-100 Avg	ImageNet100-100 Last
iCaRL [40]	19.89	12.88	30.78	16.68
NNCSL* [24]	42.19	33.64	56.78	53.84
iCaRL&Fix [17]	26.37	15.58	37.49	21.02
+ DSGD [17]	28.35	19.14	50.53	32.10
+ USP (Ours)	**43.91**	**35.40**	**56.84**	**50.36**
DER&Fix [17]	35.40	29.22	61.96	52.91
+ DSGD [17]	35.73	31.53	62.27	52.82
+ USP (Ours)	**46.09**	**39.58**	**62.29**	**55.01**

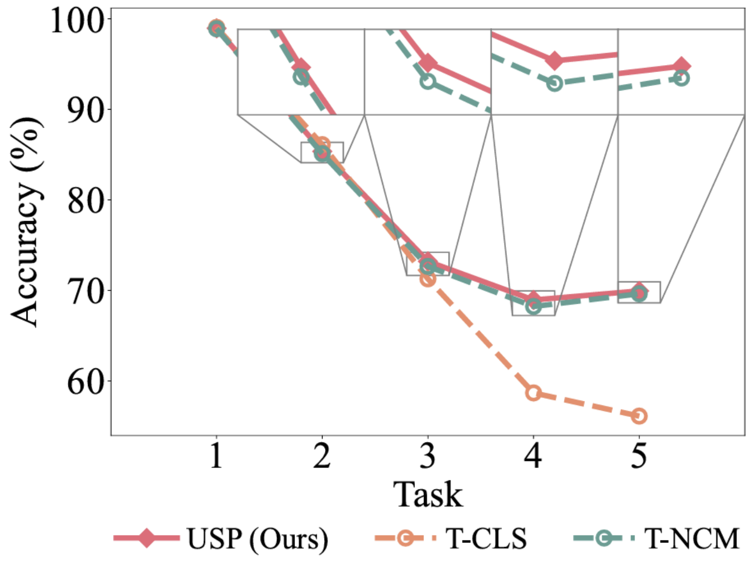
Method	Classes	Task ID 1	2	3	4	5	6	7	8	9	10	11	Avg
SS-iCaRL[10]	Base	69.89	62.32	60.62	58.99	58.59	57.77	59.88	56.21	54.46	50.54	46.11	57.76
	Novel	-	53.22	32.38	24.07	22.76	23.34	17.58	16.40	16.39	16.13	16.32	23.86
SS-NCM-CNN [10]	Base	69.89	65.80	64.97	63.79	63.81	61.08	65.24	63.73	58.77	55.74	51.88	62.24
	Novel	-	56.37	34.70	26.03	24.04	24.68	19.14	18.60	17.70	17.79	18.36	25.74
UaD-CIE* [12]	Base	75.87	74.58	74.09	**73.46**	72.24	**71.68**	**71.33**	**70.50**	**70.15**	**69.27**	**69.13**	**72.03**
	Novel	-	57.35	46.29	39.58	45.02	42.54	45.37	42.75	40.82	41.98	42.49	44.42
+ USP (Ours)*	Base	**78.21**	**75.00**	**74.13**	73.36	**72.42**	70.74	68.99	68.44	67.25	66.62	66.66	71.07
	Novel	-	**69.18**	**60.07**	**50.93**	**55.67**	**53.08**	**52.86**	**53.64**	**51.07**	**53.00**	**54.57**	**55.41**

- ImageNet-100: 10 tasks; *Avg: **43.91** v.s. 28.35; Last: **50.36** v.s. 32.10*
- CUB:
  - Few-show SSCL, 5 labeled images per class
  - 11 tasks; *Avg: **66.43** v.s. 64.3*

Ablation Analysis



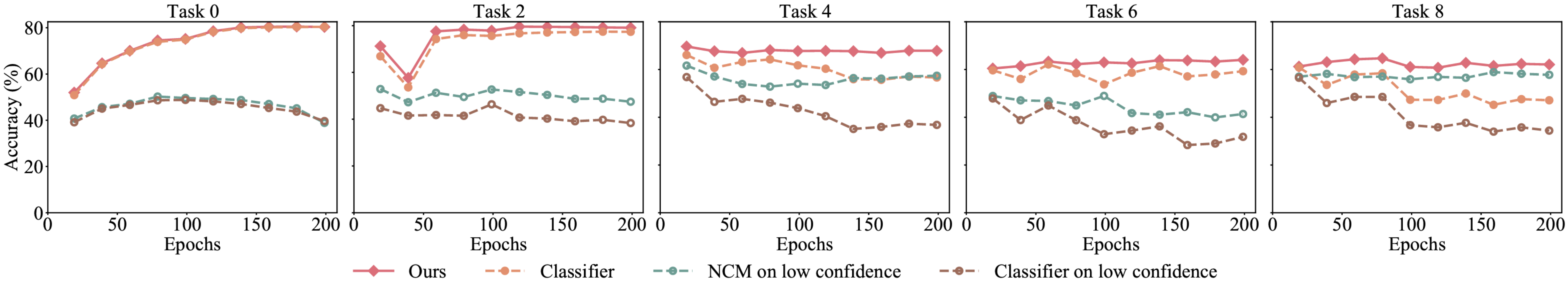
(a) Training phase



(b) Testing phase

- Ablation studies on the *main components* of USP
- The experiments are conducted on CIFAR-10 with 30 labels per
- *USP* v.s. classifier-only inference ("*T-CLS*") and NCM-only inference ("*T-NCM*")

Ablation Analysis



- Ours: *Divide-and-Conquer* Pseudo-Labeling
- Classifier: Use MLP classifier for *all samples*
- NCM: *NCM-based* classification results on low-confidence samples
- Classifier on low confidence: *MLP-classifier-based* classification results on low-confide nce samples

Thanks!

Homepage: https://njuyued.github.io/

arXiv：https://arxiv.org/pdf/2508.05316

GitHub：https://github.com/NJUyued/USP4SSCL

Zhihu（知乎）: https://zhuanlan.zhihu.com/p/1940131476688577441