Table A. Test accuracy (%) of PLL methods under extreme ambiguity (90% candidate labels) on uniform PLL settings, comparing UMiP with and without renormalization. The best results are highlighted in bold.

		PRODEN	LWS	CAVL	CRDPLL	PiCO	ABLE	DIRK
CIFAR-10	Baseline	44.97±0.23	59.94±0.45	10.06±0.18	62.23±0.56	14.02±0.25	18.14±0.37	10.76±0.12
	w/ renorm.	**51.42±0.19**	**67.92±0.42**	**11.32±0.20**	**62.86±0.49**	**16.69±0.19**	**19.37±0.31**	**12.21±0.13**
	w/o renorm.	11.73±0.47	34.61±0.55	10.00±0.22	11.26±0.58	14.02±0.33	17.31±0.29	10.76±0.17
SVHN	Baseline	70.66±0.60	86.04±0.48	16.01±0.26	95.01±0.71	12.41±0.35	88.90±0.52	34.09±0.19
	w/ renorm.	**75.02±0.39**	**94.89±0.45**	**27.27±0.23**	**96.57±0.68**	**14.83±0.28**	**90.28±0.37**	**34.58±0.14**
	w/o renorm.	20.90±1.03	22.44±0.66	14.83±0.31	19.73±0.94	12.41±0.42	33.54±0.80	32.94±0.36

Table B. Test accuracy (%) on CIFAR-10 under instance-dependent PLL settings with the dynamically adjusted mix ratio λ. The gray rows indicate the PLL method using UMiP, boldface highlights better results, and ↑/↓ denote better or worse performance compared to the original strategy.

Dataset	q_2	PRODEN	LWS	CAVL	CRDPLL	PiCO	ABLE	DIRK
CIFAR-10	0.6	77.41±1.96	81.65±0.23	44.32±2.14	73.40±1.27	77.33±1.54	76.93±1.23	47.80±3.92
		77.53±2.11↓	80.47±0.45↓	**49.15±2.23**↑	**73.97±0.95**↓	**77.94±0.46**↓	75.01±1.03↓	44.98±3.39↓
	0.8	66.48±1.24	73.88±1.11	38.60±1.54	50.08±1.56	66.57±0.27	65.62±2.69	34.72±2.25
		66.37±0.40↓	**74.45±0.88**↓	**39.14±0.11**↓	38.99±2.07↓	**68.51±0.10**↓	60.83±0.18↓	**38.78±0.15**↑
	1.0	55.41±0.54	66.08±0.88	37.50±1.08	43.86±0.46	53.42±1.94	51.13±1.61	30.30±1.33
		47.93±0.17↓	**66.27±0.85**↓	**44.08±0.15**↑	**58.35±0.70**↑	49.46±0.25↓	45.64±0.06↓	**30.46±1.10**↓

Table C. Test accuracy (%) on SVHN under instance-dependent PLL settings with q_2 = 0.6. Results for α = 2 and α = 0.5 are newly included. The best results are highligted in bold.

	PRODEN	LWS	CAVL	CRDPLL	PiCO	ABLE	DIRK
Baseline	91.79±0.16	90.42±1.09	71.76±6.82	79.45±0.10	93.71±0.22	94.92±0.23	54.25±6.28
$\lambda \sim \text{Beta}_{[0.5,1]}(1.0,1.0)$	**95.82±0.16**	**95.76±0.22**	**77.84±3.39**	**79.70±0.31**	**95.34±0.33**	**95.34±0.26**	**56.69±1.61**
$\lambda \sim \text{Beta}_{[0.5,1]}(0.5,0.5)$	95.33±0.23	95.47±0.12	76.15±3.43	79.63±0.21	95.13±0.30	33.22±0.48	49.30±3.74
$\lambda \sim \text{Beta}_{[0.5,1]}(2.0,2.0)$	95.80±0.16	95.74±0.15	74.07±3.33	79.50±0.28	95.31±0.15	34.48±1.29	55.33±0.72

Table D. Test accuracy (%) comparison on ABLE†, DIRK† and PaPi under instance-dependent PLL settings († indicates use of the exponential moving-average strategy). The gray rows indicate the PLL method using UMiP, and boldface indicates better results.

	q_2	ABLE†	DIRK†	PaPi		q_2	ABLE†	DIRK†	PaPi
CIFAR-10	0.6	80.13±0.92	76.98±1.22	81.23±0.76	SVHN	0.6	95.02±0.13	92.64±0.32	89.25±0.54
		82.08±0.42	**79.68±0.81**	**82.81±0.29**			**95.96±0.10**	**94.58±0.22**	**93.89±0.13**
	0.8	67.21±0.89	68.96±1.12	62.86±1.25		0.8	94.35±0.14	92.11±0.44	86.47±0.42
		72.62±0.88	**74.50±0.43**	**69.04±0.98**			**94.36±0.09**	**93.83±0.25**	**89.14±0.15**
	1.0	58.34±0.49	62.48±0.41	52.25±0.35		1.0	92.92±0.13	91.72±0.84	82.36±0.37
		62.98±0.11	**68.23±0.33**	**60.02±0.35**			**93.97±0.14**	**92.72±0.72**	**88.34±0.38**
CIFAR-100	0.01	63.36±0.35	62.66±0.27	64.48±0.57	CUB-200	0.01	58.12±0.92	65.59±0.31	61.23±1.12
		63.57±0.32	**63.86±0.25**	**64.52±0.18**			**60.21±0.34**	**66.45±0.21**	**64.36±0.45**
	0.05	60.82±0.25	59.73±0.04	60.98±0.24		0.02	49.41±1.21	57.35±0.81	56.63±1.23
		60.88±0.27	**61.37±0.05**	**61.96±0.21**			**52.89±0.70**	**59.09±0.66**	**59.44±0.38**
	0.1	55.88±0.36	52.67±0.46	54.78±0.23		0.03	48.25±1.10	53.27±1.24	47.85±1.12
		56.24±0.31	**53.73±0.26**	**54.91±0.13**			**52.43±0.65**	**54.24±0.66**	**52.32±0.41**

Table E. Test accuracy (%) comparison on ABLE†, DIRK†, and PaPi under class-dependent PLL settings († indicates use of the exponential moving-average strategy). The gray rows indicate the PLL method using UMiP, and boldface indicates better results.

	q_1	ABLE†	DIRK†	PaPi		q_1	ABLE†	DIRK†	PaPi
CIFAR-10	0.7	88.04±0.31	83.24±0.46	89.02±0.39	SVHN	0.7	94.68±0.33	93.23±0.43	96.71±0.16
		88.92±0.30	**86.17±0.24**	**89.52±0.22**			**95.44±0.34**	**94.32±0.35**	**96.84±0.14**
	0.8	86.10±0.24	82.48±0.32	86.23±0.42		0.8	94.65±0.14	93.15±0.49	95.87±0.27
		86.42±0.23	**85.12±0.18**	**86.83±0.27**			**95.15±0.12**	**94.08±0.28**	**96.15±0.10**
	0.9	82.32±0.21	80.89±0.28	78.98±0.58		0.9	94.45±0.20	93.45±0.24	95.81±0.12
		83.22±0.17	**81.89±0.24**	**80.03±0.18**			**95.05±0.14**	**93.84±0.14**	**96.05±0.11**

Table F. The magnitude change of **traditional ambiguity degree** before (white columns) and after data augmentation (gray columns). (To be modified)

Ambiguity Level	Class-dependent (0.7 / 0.8 / 0.9)					Instance-dependent (0.6 / 0.8 / 1.0)				
	Baseline	UMiP	RandAugment	Cutout	Mixup	Baseline	UMiP	RandAugment	Cutout	Mixup
Low level	0.715	0.713	0.716	0.718	0.818	0.955	0.951	0.951	0.953	0.974
Medium level	0.810	0.810	0.815	0.814	0.887	0.972	0.975	0.972	0.971	0.986
High level	0.907	0.909	0.907	0.907	0.950	0.986	0.983	0.983	0.984	0.992

Table G. The magnitude change of **local ambiguity degree** (AD↓) and Accuracy (Acc↑) before (white columns) and after UMiP augmentation (gray columns). PRODEN.

Ambiguity Level	Baseline		UMiP		RandAugment		Mixup		Cutout	
	AD	Acc	AD	Acc	AD	Acc	AD	Acc	AD	Acc
$q_1 = 0.7$	0.781	88.12±0.31	0.706	**89.20±0.22**	0.000	80.12±0.42	0.705	84.24±0.24	0.000	83.19±0.18
$q_1 = 0.8$	0.869	86.42±0.55	0.852	**87.91±0.35**	0.000	79.28±0.41	0.854	81.02±0.54	0.000	82.91±0.28
$q_1 = 0.9$	0.978	75.95±0.96	0.925	**80.19±0.39**	0.000	66.99±0.21	0.925	69.29±0.31	0.000	72.23±0.34