CA2C: A Prior-Knowledge-Free Approach for Robust Label Noise Learning via Asymmetric Co-learning and Co-training

Speaker: Mengmeng Sheng

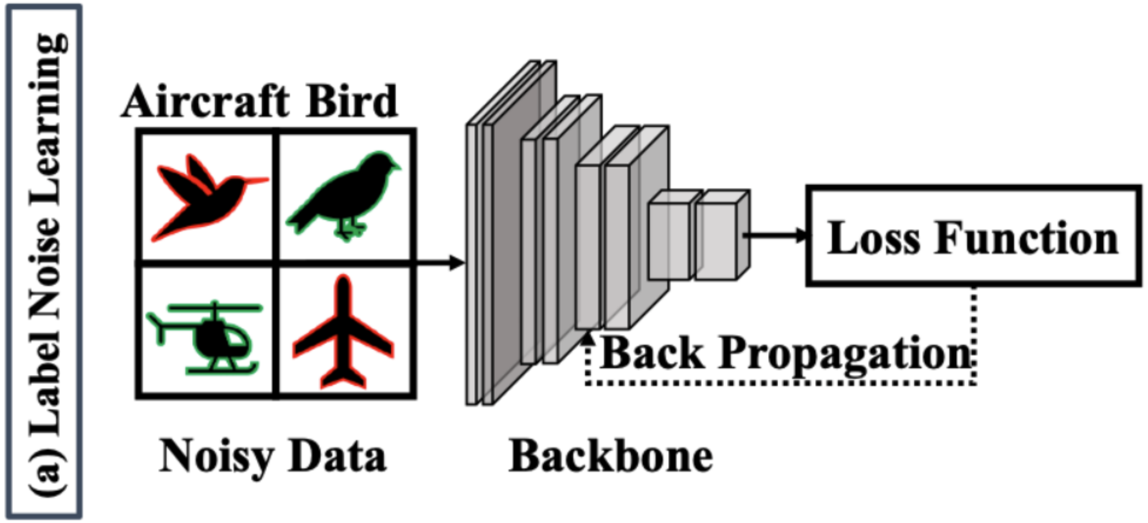
Paper ID: 1201

Project Page : https://github.com/NUST-Machine-Intelligence-Laboratory/CA2C

Authors: Mengmeng Sheng, Zeren Sun, Tianfei Zhou, Xiangbo Shu, Jinshan Pan, Yazhou Yao

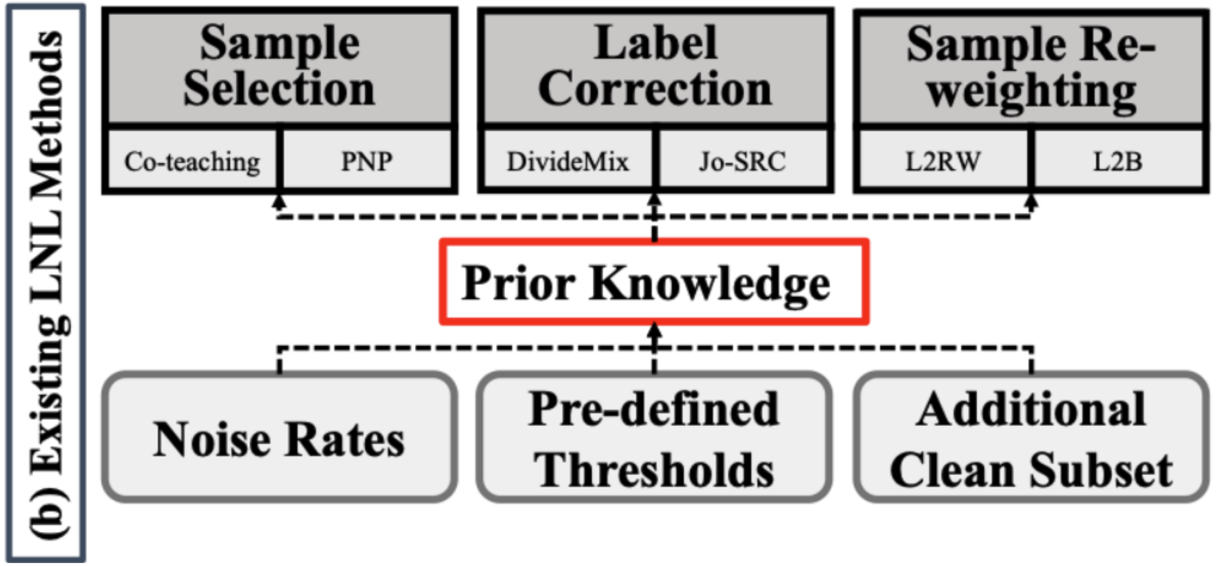
Background

- Label noise learning primarily aims to mitigate the negative effects of noisy labels on model training, a challenge that is inevitable in real-world scenarios.



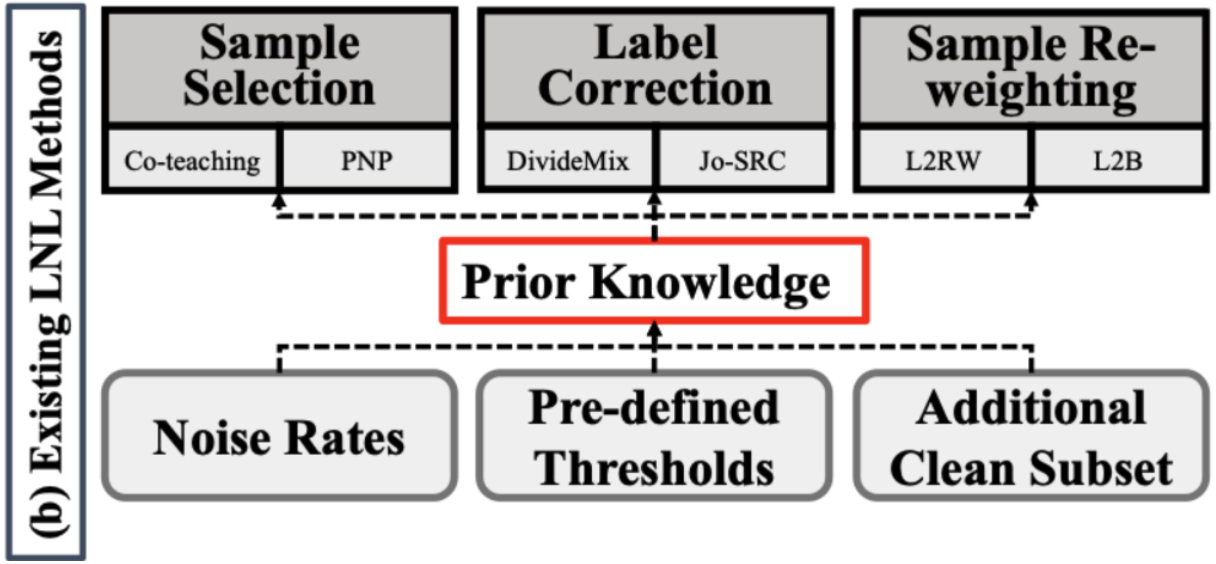
Background

- Existing methods (sample selection, label correction, and sample re-weighting) often face challenges in real-world applications due to their strong dependence on prior knowledge (e.g., noise rates, predefined thresholds, or additional clean subsets) to sustain performance.



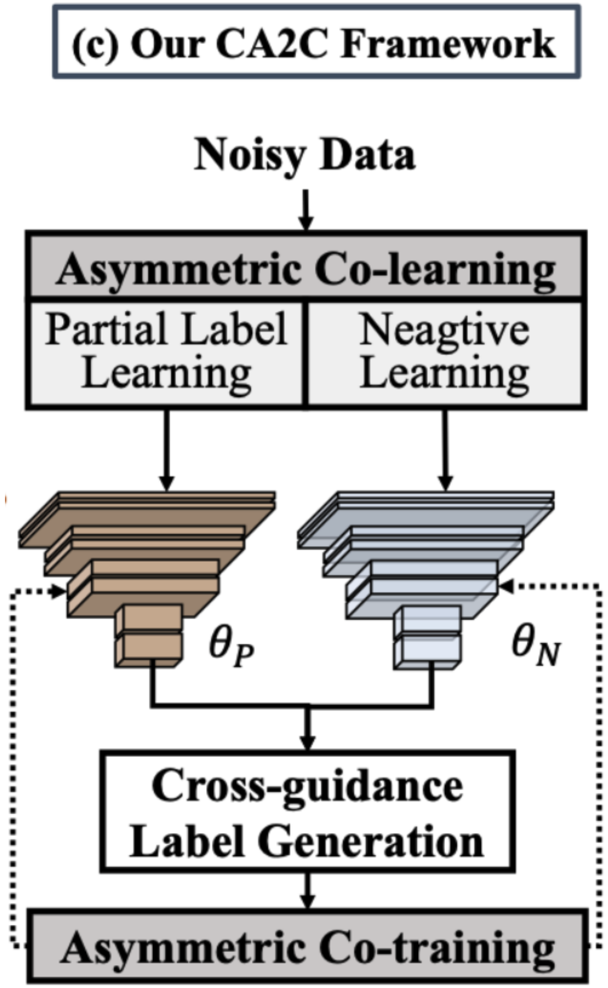
Background

- This dependence limits their adaptability and practicality in real-world scenarios where such priors are usually unavailable.



Idealy,

- We propose a novel approach for learning with noisy labels, termed CA2C.
- This method introduces a combined asymmetric co-learning and co-training framework that eliminates demands for strong prior knowledge.
- CA2C incorporates cross-guidance label generation to promote knowledge exchange between twin models.



Our Proposed CA2C

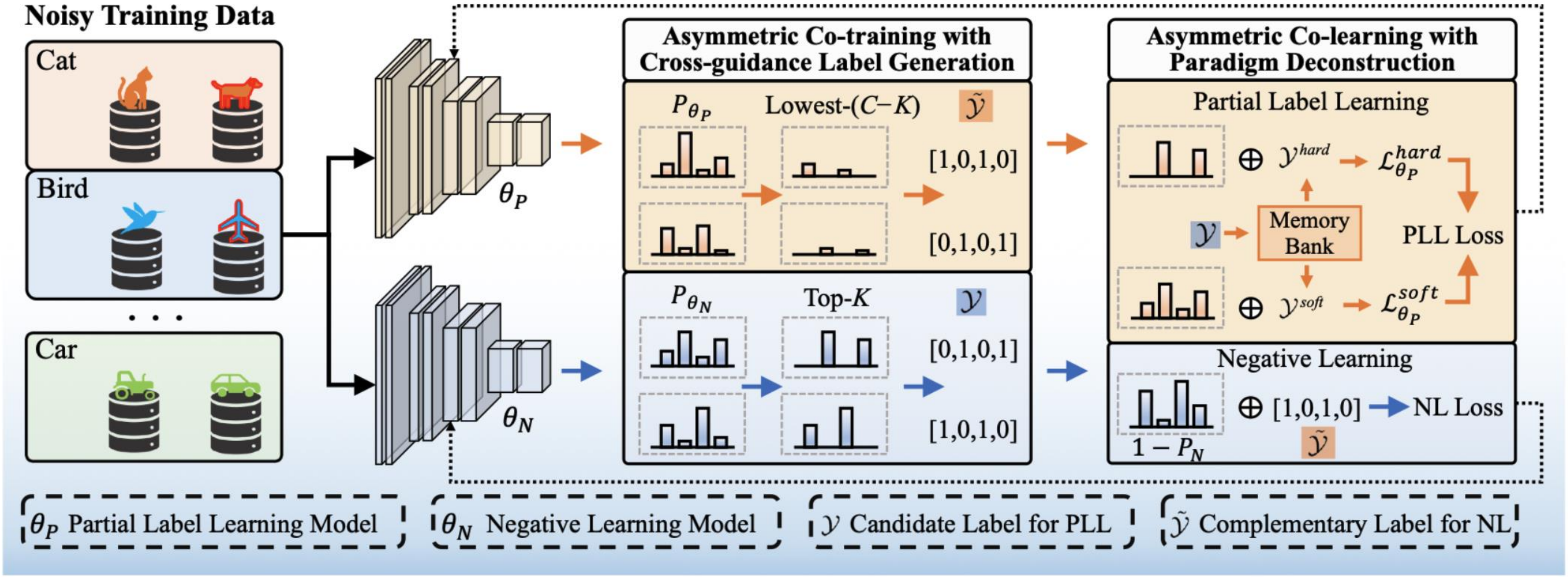


Figure 2. The overall framework of our proposed CA2C. In our CA2C, twin models (*i.e.*, θ_P and θ_N) with identical architectures are trained simultaneously but employ distinct learning paradigms: partial label learning and negative learning. To promote knowledge exchange between θ_P and θ_N, we exploit the paradigm independence inherent in our asymmetric co-learning strategy by using each model's predictions to cross-generate label spaces. For the *P-model*, we implement a memory bank to track the frequency of $\mathcal{Y}$ and design a confidence-based re-weighting strategy for label disambiguation, enhancing θ_P's robustness against disambiguation failures. For the *N-model*, we use the complementary labels $\tilde{\mathcal{Y}}$ generated from the *P-model* for negative learning.

Our Proposed CA2C

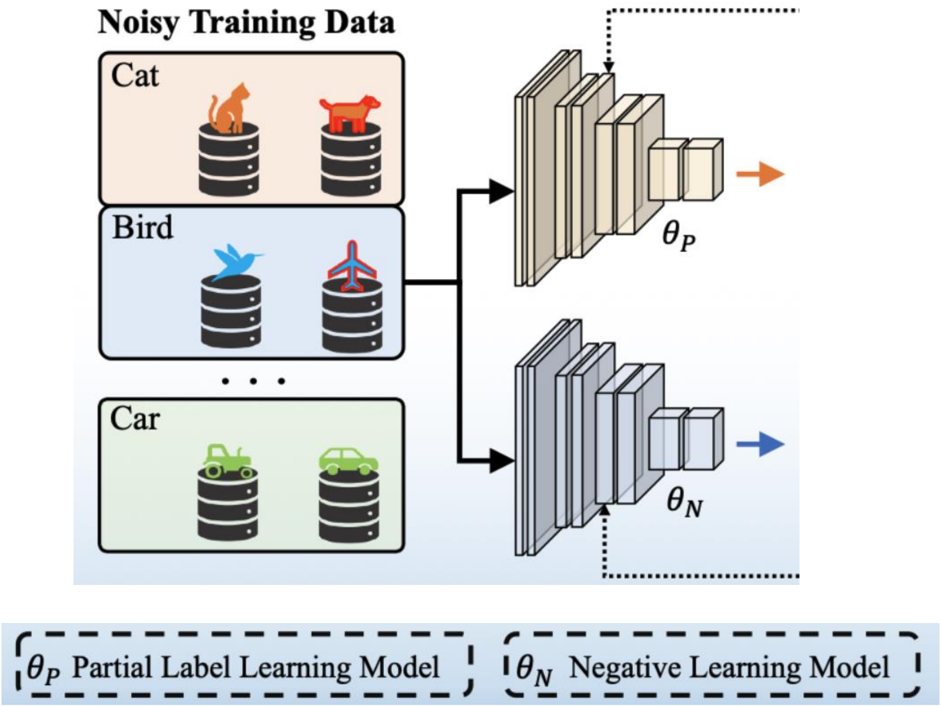
- Asymmetric Co-learning with Paradigm Deconstruction

➢Loss for optimizing partial label learning model:

$$\mathcal{L}_{\theta_P} = -\frac{1}{N}\sum_{n=1}^{N} \mathcal{Y}_n \log(p(x_n, \theta_P))$$

➢Loss for optimizing negative learning model:

$$\mathcal{L}_{\theta_N} = -\frac{1}{N}\sum_{n=1}^{N} \tilde{\mathcal{Y}}_n \log(1 - p(x_n, \theta_N))$$



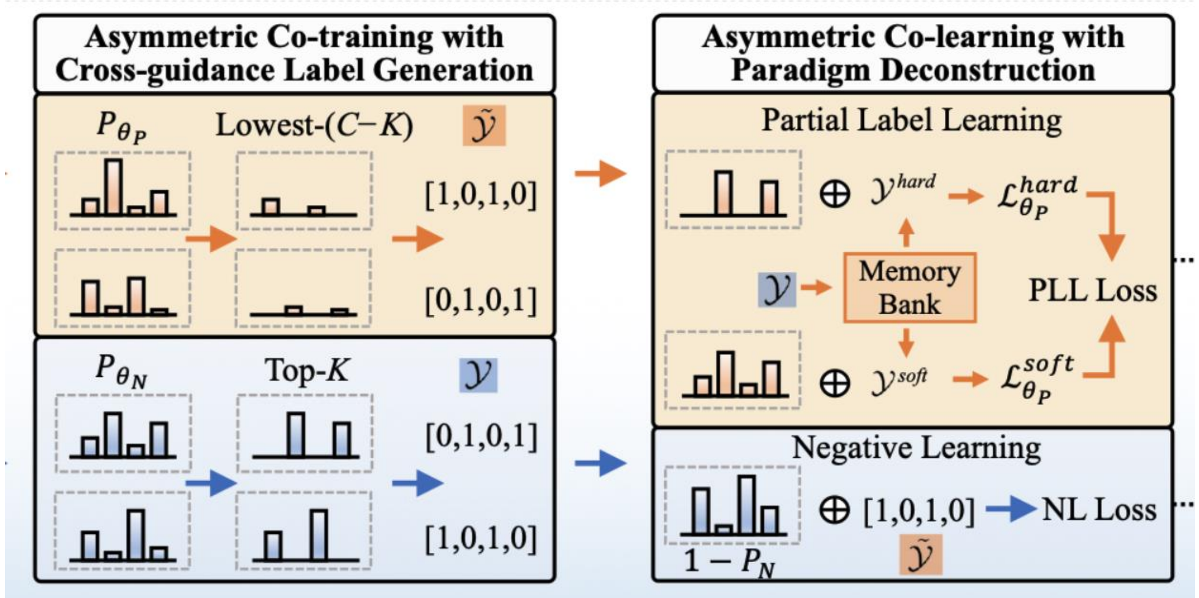
Our Proposed CA2C

- Asymmetric Co-training with Cross-guidance Label Generation
  - ➢Multi-hot candidate labels in PLL:

$$\mathcal{Y}_n^* = \{\hat{y}_n^1, ..., \hat{y}_n^C\},$$
$$\hat{y}_n^c = \mathbb{1}_{c \in \{\text{Top-}K(\{p^1(x_n,\theta_N),...,p^C(x_n,\theta_N)\})\}}$$

  - ➢Multi-hot complementary labels in NL :

$$\tilde{\mathcal{Y}}_n^* = \{\hat{y}_n^1, ..., \hat{y}_n^C\},$$
$$\hat{y}_n^c = \mathbb{1}_{c \notin \{\text{Top-}K(\{p^1(x_n,\theta_P),...,p^C(x_n,\theta_P)\})\}}$$



Our Proposed CA2C

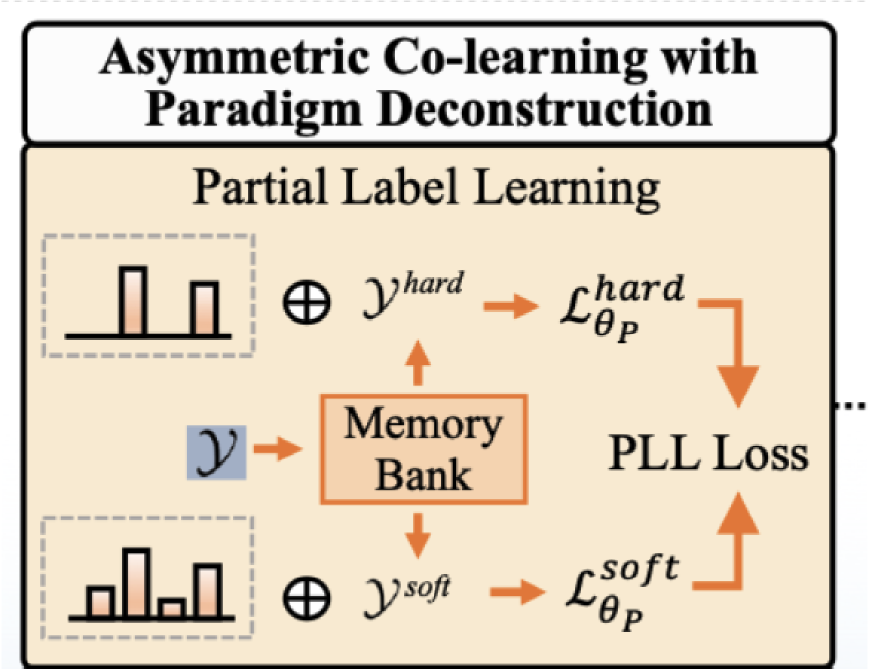
- Confidence-based Re-weighting Strategy for label disambiguation

  - ➢Memory Bank:

$$\mathcal{M}^t(x_n) = \begin{cases} 0, & if\ t = 0 \\ \mathcal{M}^{t-1}(x_n) + \mathcal{Y}_n^*, & if\ t \geq 1 \end{cases}$$

  - ➢Re-weighting:

$$\mathcal{W}(n) = \frac{1}{\max(\mathcal{M}^t(x_n))}(\mathcal{M}^t(x_n))$$



Experiment Results

Synthetic noisy datasets:

Methods	Publication	CIFAR100N			CIFAR80N			Average
		Sym-20%	Sym-80%	Asym-40%	Sym-20%	Sym-80%	Asym-40%	
Standard	-	35.14	4.41	27.29	29.37	4.20	22.25	20.44
Decoupling	NeurIPS 2017	33.10	3.89	26.11	43.49	10.1	33.74	25.07
Co-teaching	NeurIPS 2018	43.73	15.15	28.35	60.38	16.59	42.42	34.44
Co-teaching+	ICML 2019	49.27	13.44	33.62	53.97	12.29	43.01	34.27
JoCoR	CVPR 2020	53.01	15.49	32.70	59.99	12.85	39.37	35.57
DivideMix	ICLR 2020	57.76	28.98	43.75	57.47	21.18	37.47	41.10
Jo-SRC	CVPR 2021	58.15	23.80	38.52	65.83	29.76	53.03	44.85
Co-LDL	TMM 2022	59.73	25.12	52.28	58.81	24.22	50.69	45.14
UNICON	CVPR 2022	55.10	31.49	49.90	54.50	36.75	51.50	46.54
SOP	ICML 2022	58.63	34.23	49.87	60.17	34.05	53.34	48.38
AGCE	TPAMI 2023	59.38	27.41	43.04	60.24	25.39	44.06	43.25
DISC	CVPR 2023	60.28	33.90	50.56	50.33	38.23	47.63	46.82
ANL	NeurIPS 2023	60.20	23.39	44.15	61.35	20.74	47.31	42.86
NPN	AAAI 2024	62.76	31.69	57.11	63.78	25.25	58.50	49.85
ACT	MM 2024	65.51	40.74	63.48	67.09	38.58	64.40	56.63
SED	ECCV 2024	66.50	38.15	58.29	69.10	**42.57**	60.87	55.91
Ours	-	**68.64**	**40.97**	**65.59**	**70.06**	40.47	**65.71**	**58.57**

Table 1. Average test accuracy (%) on CIFAR100N and CIFAR80N over the last ten epochs. Experiments are conducted under various noise conditions ("Sym" and "Asym" denote the symmetric and asymmetric label noise, respectively).

Experiment Results

Real-world noisy datasets:

Methods	Publication	Backbone	Performances(%)			
			Web-Aircraft	Web-Bird	Web-Car	Average
Standard	-	ResNet50	60.80	64.40	60.60	61.93
Decoupling	NeurIPS 2017	ResNet50	75.91	71.61	79.41	75.64
Co-teaching	NeurIPS 2018	ResNet50	79.54	76.68	84.95	80.39
Co-teaching+	ICML 2019	ResNet50	74.80	70.12	76.77	73.90
PENCIL	CVPR 2019	ResNet50	78.82	75.09	81.68	78.53
JoCoR	CVPR 2020	ResNet50	80.11	79.19	85.10	81.47
AFM	ECCV 2020	ResNet50	81.04	76.35	83.48	80.29
DivideMix	ICLR 2020	ResNet50	82.48	74.40	84.27	80.38
Jo-SRC	CVPR 2021	ResNet50	82.73	81.22	88.13	84.03
Co-LDL	TMM 2022	ResNet50	81.97	80.11	86.95	83.01
UNICON	CVPR 2022	ResNet50	85.18	81.20	88.15	84.84
SOP	ICML 2022	ResNet50	84.06	79.40	85.71	83.06
AGCE	TPAMI 2023	ResNet50	84.22	75.60	85.16	81.66
DISC	CVPR 2023	ResNet50	85.27	81.08	88.31	84.89
ANL	NeurIPS 2023	ResNet50	81.78	79.46	86.47	82.57
NPN	AAAI 2024	ResNet50	83.65	79.36	85.46	82.82
ACT	MM 2024	ResNet50	86.56	81.43	88.75	85.58
SED	ECCV 2024	ResNet50	86.62	82.00	88.88	85.83
Ours	-	ResNet50	**87.70**	**82.48**	**89.11**	**86.43**

Table 2. The comparison with SOTA approaches in test accuracy (%) on real-world noisy datasets: Web-Aircraft, Web-Bird, Web-Car.

Methods	Publication	Backbone	Acc (%)
Standard	-	ResNet50	84.50
Decoupling	NeurIPS 2017	ResNet50	85.53
Co-teaching	NeurIPS 2018	ResNet50	61.91
Co-teaching+	ICML 2019	ResNet50	81.61
JoCoR	CVPR 2020	ResNet50	77.94
DivideMix	ICLR 2020	ResNet50	85.88
PLC	ICML 2021	ResNet50	85.28
WarPI	PR 2022	ResNet50	85.91
CoDis	ICCV 2023	ResNet50	86.13
VRI	IJCV 2024	ResNet50	86.24
Ours	-	ResNet50	**86.83**

Table 3. The comparison with SOTA approaches in test accuracy(%) on the large-scale, real-world noisy dataset Food101N.

Experiment Results

Real-world noisy datasets:

ACPD	✗	✓
Sym-20%	35.14	65.22
Sym-80%	4.41	28.53
Asym-40%	27.29	62.02

ACLG	✗	✓
Sym-20%	65.22	67.37
Sym-80%	28.53	35.90
Asym-40%	62.02	65.06

CBRW	✗	✓
Sym-20%	67.37	68.64
Sym-80%	35.90	40.97
Asym-40%	65.06	65.59

ACPD	✗	✓
Sym-20%	29.37	64.81
Sym-80%	4.20	28.01
Asym-40%	22.25	60.39

ACLG	✗	✓
Sym-20%	64.81	68.46
Sym-80%	28.01	36.46
Asym-40%	60.39	64.63

CBRW	✗	✓
Sym-20%	68.46	70.06
Sym-80%	36.46	40.47
Asym-40%	64.63	65.71

Table 4. Effect of key components (*i.e.*, ACPD, ACLG and CBRW) in our CA2C on CIFAR100N (top) and CIFAR80N (bottom). Test accuracy (%) of our CA2C with (✓) and without (✗) the different components is compared under different settings.

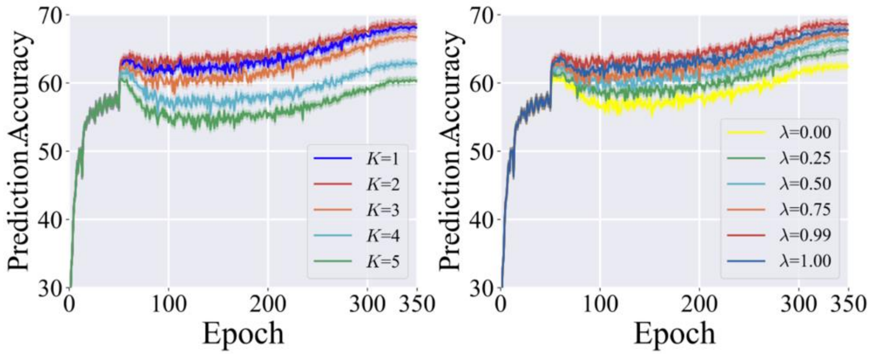


Figure 3. Sensitivity of Hyper-parameters: K (left) and λ (right). Experiments are conducted on CIFAR100N with Sym-20%.

Thanks!