Joint Asymmetric Loss for Learning with Noisy Labels

Jialiang Wang, Xianming Liu, Xiong Zhou

Deming Zhai, Junjun Jiang, Xiangyang Ji

Harbin Institute of Technology

Tsinghua University

Motivation

- Learning with noisy labels is a crucial task for training accurate deep neural networks. Robust loss function is a popular approach to solve this problem.
- Active Passive Loss (APL) jointly optimizes an active and a passive symmetric loss to mutually enhance the overall fitting ability.
- Asymmetric losses, a new class of robust loss functions, possess superior properties compared to symmetric losses.
- However, existing asymmetric losses are not compatible with advanced optimization frameworks such as APL.

Contributions

- We propose a novel asymmetric loss function, ***Asymmetric Mean Square Error (AMSE).*** We rigorously establish the condition for AMSE to satisfy noise-tolerance.
- By incorporating the proposed AMSE into the APL framework, we introduce a novel approach called ***Joint Asymmetric Loss (JAL)***, which ensures robustness and enhances sufficient learning.
- The extensive results highlight the superiority of our method.

Preliminary

For a loss $L(f(\boldsymbol{x}), y) = \sum_{k=1}^{K} \ell(f(\boldsymbol{x})_k, e_k)$:

- **(Active Loss Function)** L_{active} is an active loss function if $\forall(\boldsymbol{x}, y) \in D, \forall k \neq y, \ell(f(\boldsymbol{x})_k, e_k) = 0$
- **(Passive Loss Function)** $L_{passive}$ is a passive loss function if $\forall(\boldsymbol{x}, y) \in D, \exists k \neq y, \ell(f(\boldsymbol{x})_k, e_k) \neq 0$

By combining the two different symmetric loss functions, APL can improve the fitting ability under the premise of ensuring robustness.

Preliminary

- Recently, asymmetric loss functions have been proposed, which are noise-tolerant.
- **(Asymmetric Condition).** On the given weights $w_1, \dots, w_K \geq 0$, where $\exists t \in [K], s.t., w_t > \max_{i \neq t} w_i$, a loss function L is called asymmetric if L satisfies

$$\arg\min_{f(\boldsymbol{x})} \sum_{k=1}^{K} w_k L(f(\boldsymbol{x}), k) = \arg\min_{f(\boldsymbol{x})} L(f(\boldsymbol{x}), t),$$

where we always have $\arg\min_{f(\boldsymbol{x})} L(f(\boldsymbol{x}), t) = \boldsymbol{e_t}$.

Methodology

- In this paper, we extend the asymmetric loss function to a more complex passive loss scenario and propose the Asymmetric Mean Square Error (AMSE).

- **Asymmetric Mean Square Error (AMSE)**

$$L_{\text{AMSE}} = \frac{1}{K} \| a \cdot \boldsymbol{e}_{\boldsymbol{y}} - f(\boldsymbol{x}) \|_2^2$$

- **Theorem (Noise-tolerant for AMSE).** On the given weights $w_1, \ldots, w_K$, where $w_m > w_n$, and $w_n = \max_{i \neq m} w_i$. The loss function $L(f(\boldsymbol{x}), y) = \frac{1}{K} \| a \cdot \boldsymbol{e}_{\boldsymbol{y}} - f(\boldsymbol{x}) \|_q^q = \sum_{k=1}^{K} \frac{1}{K} |a \cdot e_k - f(\boldsymbol{x})_k|^q$, where $q > 0$ and $a \geq 1$ are parameters, is asymmetric if and only if $\frac{w_m}{w_n} \geq \frac{a^{q-1} + \sum_{i \neq m} \frac{w_i}{w_n}}{(a-1)^{q-1}} \cdot \mathbb{I}(q > 1) + \mathbb{I}(q \leq 1)$.

Methodology

We integrate the proposed AMSE into the APL framework to enhance its performance, resulting in a novel approach called **Joint Asymmetric Loss (JAL).**

- By combining Normalized Cross Entropy (NCE), we have JAL-CE:

$$L_{\mathrm{JAL-CE}} = \alpha \cdot L_{\mathrm{NCE}} + \beta \cdot L_{\mathrm{AMSE}}$$

- By combining Normalized Focal Loss (NFL), we have JAL-FL:

$$L_{\mathrm{JAL-FL}} = \alpha \cdot L_{\mathrm{NFL}} + \beta \cdot L_{\mathrm{AMSE}}$$

ICCV HONOLULU HAWAII OCT 19-23, 2025

Experiments

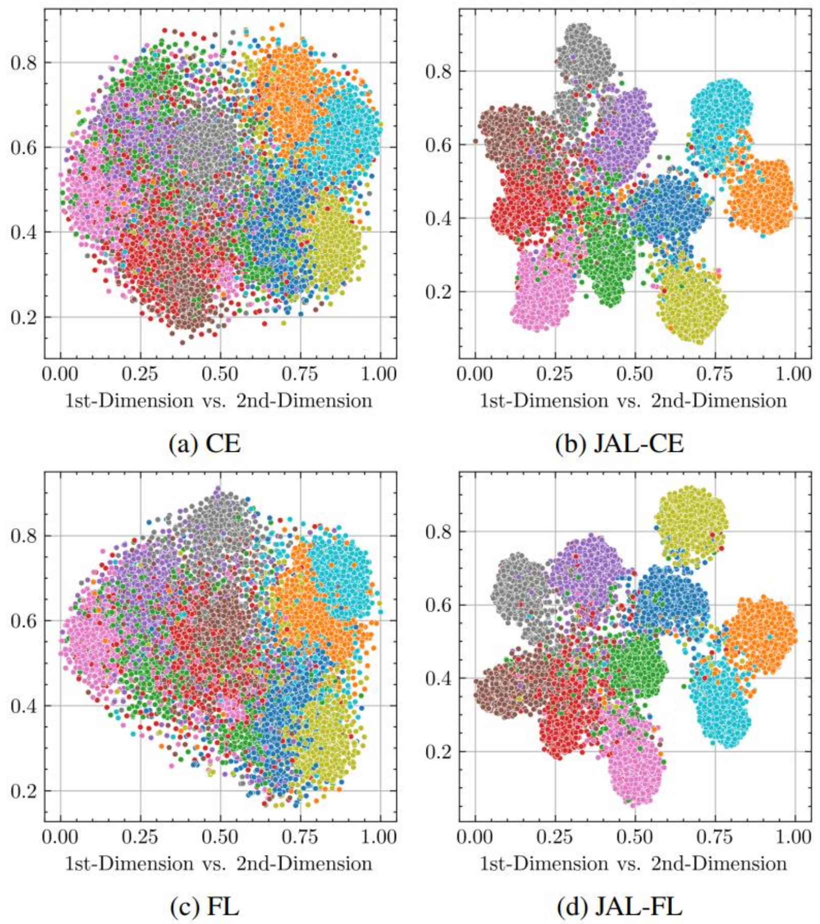


Figure 1. Visualizations of 2D t-SNE [21] embeddings of learned representations on the CIFAR-10 test set, from models trained with 0.4 symmetric noise. The representations learned by the proposed JAL method are with more separated and clearly bound margin.

Table 3. Last epoch test accuracies (%) of different methods on CIFAR-10 and CIFAR-100 with clean, symmetric ($\eta \in [0.2, 0.4, 0.6, 0.8]$), and asymmetric ($\eta \in [0.1, 0.2, 0.3, 0.4]$) label noise. The results (mean±std) are reported over 3 random trials and the top-2 best results are in **bold**.

CIFAR-10	Clean	Symmetric 0.2	Symmetric 0.4	Symmetric 0.6	Symmetric 0.8	Asymmetric 0.1	Asymmetric 0.2	Asymmetric 0.3	Asymmetric 0.4
CE	90.50±0.22	75.21±0.39	58.05±0.53	38.80±0.45	19.74±0.40	86.85±0.15	83.05±0.35	78.37±0.61	73.85±0.07
FL	89.70±0.24	74.50±0.18	58.23±0.40	38.69±0.06	19.47±0.74	86.64±0.12	83.08±0.07	79.34±0.30	74.68±0.31
GCE	89.36±0.19	89.36±0.19	82.19±0.84	68.01±0.40	46.61±0.39	88.41±0.20	85.72±0.22	79.49±0.20	73.36±0.53
SCE	91.51±0.24	87.65±0.36	79.73±0.29	61.79±0.72	28.01±0.92	89.54±0.33	85.94±0.38	80.50±0.09	74.33±0.56
NCE	75.48±0.37	73.22±0.35	69.37±0.22	62.47±0.85	41.20±1.25	74.11±0.24	72.20±0.38	70.14±0.27	65.33±0.40
NCE+RCE	90.80±0.06	88.93±0.04	85.89±0.31	79.89±0.25	54.99±2.13	90.04±0.17	88.62±0.29	85.07±0.27	77.94±0.21
NCE+AUL	91.17±0.18	89.00±0.58	86.05±0.30	79.22±0.22	56.24±0.94	90.06±0.16	88.19±0.07	84.83±0.47	77.60±0.16
NCE+AGCE	91.01±0.20	88.91±0.38	86.16±0.38	79.93±0.33	43.82±1.91	90.29±0.05	88.49±0.28	85.21±0.59	78.47±1.05
CE+LC	90.09±0.13	83.87±0.27	70.36±0.23	46.53±0.29	19.74±1.77	87.74±0.23	83.16±0.33	78.48±0.25	73.32±0.78
ANL-CE	91.74±0.18	89.68±0.29	87.16±0.16	81.28±0.63	62.28±1.10	90.66±0.16	89.09±0.21	85.49±0.49	77.99±0.40
ANL-FL	91.58±0.19	89.93±0.03	86.94±0.03	81.10±0.30	61.89±2.25	**90.72±0.20**	**89.29±0.02**	85.80±0.38	77.89±0.28
LT-APL	-	89.42±0.13	86.82±0.18	80.93±0.30	40.87±1.57	-	89.28±0.24	86.29±0.36	**79.99±0.58**
JAL-CE	91.63±0.21	**89.95±0.22**	**87.53±0.10**	**82.03±0.18**	**65.43±0.99**	90.70±0.21	89.11±0.38	**86.38±0.14**	**79.54±0.34**
JAL-FL	91.56±0.25	**89.99±0.11**	**87.43±0.29**	**82.09±0.08**	**64.84±1.13**	**90.77±0.16**	**89.36±0.27**	**86.18±0.04**	79.51±0.06

CIFAR-100	Clean	Symmetric 0.2	Symmetric 0.4	Symmetric 0.6	Symmetric 0.8	Asymmetric 0.1	Asymmetric 0.2	Asymmetric 0.3	Asymmetric 0.4
CE	70.93±0.77	56.47±1.34	39.68±0.77	22.64±0.53	7.82±0.33	64.14±1.01	58.67±0.45	50.44±1.16	41.51±0.12
FL	70.58±0.34	56.32±1.43	40.83±0.52	22.44±0.54	7.68±0.37	65.00±0.46	58.12±0.44	51.16±1.32	41.46±0.38
GCE	61.73±1.30	60.58±2.51	57.35±0.91	46.15±1.10	20.33±0.31	62.01±1.11	59.19±1.36	53.35±0.65	40.92±0.21
SCE	70.57±0.93	55.50±0.35	40.13±1.48	22.23±1.29	7.84±0.56	64.51±0.45	57.84±0.57	49.66±0.48	41.58±0.87
NCE	29.95±0.56	25.43±0.91	20.26±0.25	14.66±1.04	8.82±0.47	27.16±1.01	26.67±0.73	23.83±0.29	20.83±1.08
NCE+RCE	68.07±0.70	64.57±0.16	58.48±0.51	46.73±1.00	26.94±1.29	66.74±0.30	62.82±0.57	55.86±0.40	41.50±0.39
NCE+AUL	69.95±0.33	65.45±0.49	56.37±0.12	38.68±0.75	12.95±0.37	66.41±0.15	57.39±0.34	48.20±0.19	38.41±0.52
NCE+AGCE	69.05±0.36	65.61±0.27	59.40±0.34	47.66±0.49	26.14±0.01	66.96±0.45	64.08±0.44	57.17±0.33	44.62±1.04
CE+LC	71.80±0.34	56.26±0.09	37.36±0.49	17.46±0.62	6.32±0.16	63.51±0.27	56.19±0.30	48.07±0.38	39.64±0.14
ANL-CE	70.26±0.15	66.93±0.09	61.58±0.33	52.09±0.58	**28.01±1.06**	68.60±0.41	65.96±0.18	60.57±0.07	45.73±0.74
ANL-FL	70.11±0.27	67.03±0.46	61.89±0.25	51.58±0.33	**28.81±0.74**	68.67±0.21	66.12±0.39	60.03±0.48	46.20±0.45
LT-APL	-	63.29±0.49	54.70±1.73	40.52±1.65	22.63±0.78	-	62.59±1.31	56.90±1.29	44.05±1.32
JAL-CE	70.60±0.09	**68.25±0.39**	**64.11±0.55**	**56.73±0.65**	22.80±2.11	**69.29±0.42**	**67.90±0.59**	**64.90±0.27**	**56.17±0.32**
JAL-FL	70.66±0.37	**68.33±0.34**	**64.55±0.61**	**56.44±0.22**	23.11±2.28	**69.25±0.21**	**67.63±0.50**	**65.18±0.26**	**56.26±0.05**

Experiments

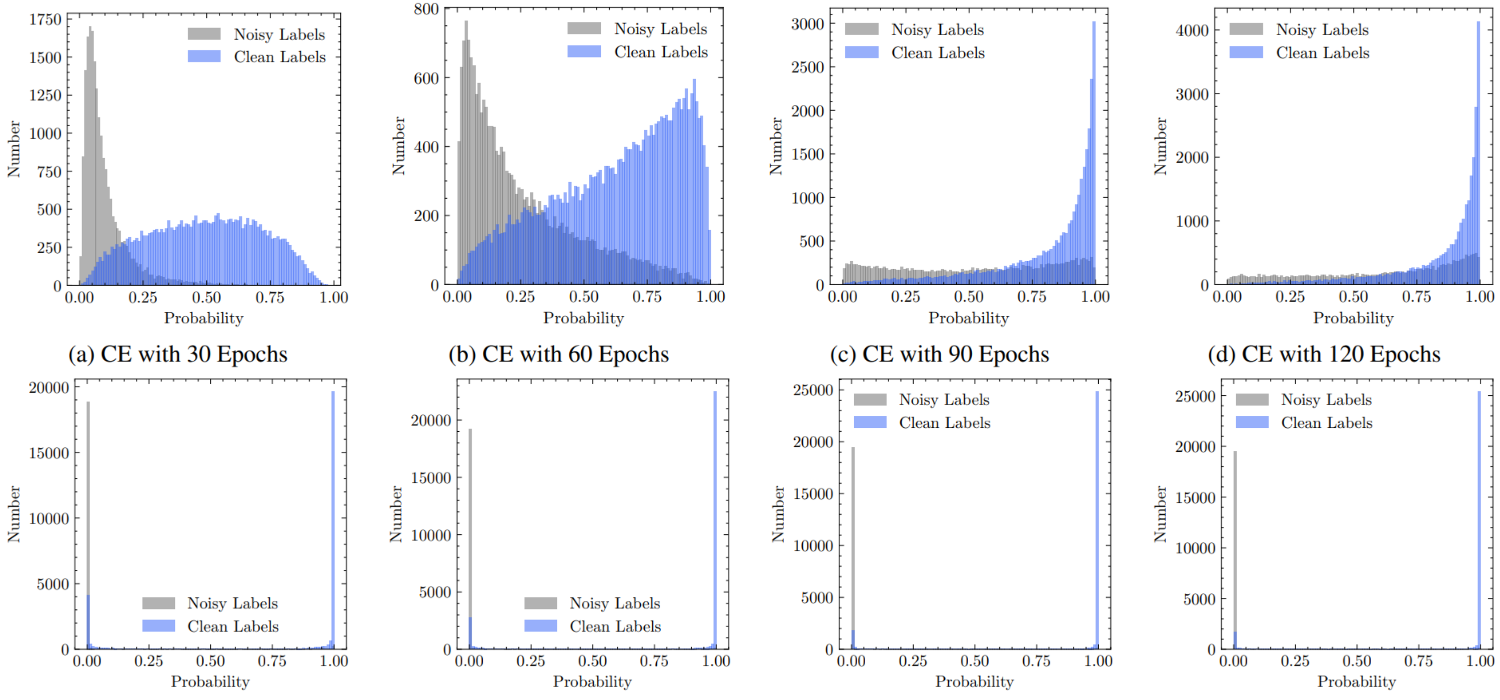
Table 4. Last epoch test accuracies (%) of different methods on CIFAR-10 and CIFAR-100 with instance-dependent noise (IDN) ($\eta \in [0.2, 0.4, 0.6]$). The results "mean±std" are reported over 3 random trials and the top-2 best results are in **bold**.

Loss	CIFAR-10 IDN			CIFAR-100 IDN		
	0.2	0.4	0.6	0.2	0.4	0.6
CE	$75.38_{\pm 0.19}$	$57.63_{\pm 0.27}$	$37.97_{\pm 0.36}$	$57.02_{\pm 0.54}$	$40.91_{\pm 2.05}$	$24.49_{\pm 0.86}$
GCE	$86.66_{\pm 0.14}$	$79.99_{\pm 0.23}$	$51.90_{\pm 0.13}$	$61.43_{\pm 2.24}$	$57.07_{\pm 1.04}$	$42.40_{\pm 0.52}$
SCE	$86.65_{\pm 0.27}$	$74.54_{\pm 0.34}$	$49.83_{\pm 0.40}$	$56.32_{\pm 0.27}$	$39.82_{\pm 1.43}$	$23.19_{\pm 0.87}$
NCE+RCE	$89.06_{\pm 0.26}$	$85.11_{\pm 0.28}$	$71.27_{\pm 0.66}$	$64.33_{\pm 0.46}$	$57.53_{\pm 0.84}$	$40.36_{\pm 0.35}$
NCE+AGCE	$88.95_{\pm 0.07}$	$85.30_{\pm 0.23}$	$71.49_{\pm 0.34}$	$65.18_{\pm 0.17}$	$57.89_{\pm 0.57}$	$43.04_{\pm 0.29}$
CE+LC	$82.77_{\pm 0.09}$	$68.06_{\pm 0.22}$	$43.60_{\pm 0.39}$	$55.93_{\pm 0.39}$	$37.74_{\pm 0.63}$	$18.68_{\pm 0.50}$
ANL-CE	$89.71_{\pm 0.35}$	$85.74_{\pm 0.15}$	$69.83_{\pm 0.38}$	$66.89_{\pm 0.53}$	$60.88_{\pm 0.35}$	$48.12_{\pm 0.48}$
ANL-FL	$89.68_{\pm 0.21}$	$85.97_{\pm 0.16}$	$70.70_{\pm 0.30}$	$67.17_{\pm 0.11}$	$61.07_{\pm 0.38}$	$46.77_{\pm 0.80}$
JAL-CE	$\mathbf{90.01}_{\pm 0.12}$	$\mathbf{86.46}_{\pm 0.15}$	$\mathbf{75.62}_{\pm 0.18}$	$\mathbf{67.51}_{\pm 0.29}$	$\mathbf{63.24}_{\pm 0.16}$	$\mathbf{51.69}_{\pm 0.68}$
JAL-FL	$\mathbf{89.90}_{\pm 0.14}$	$\mathbf{86.78}_{\pm 0.17}$	$\mathbf{75.02}_{\pm 0.48}$	$\mathbf{67.77}_{\pm 0.38}$	$\mathbf{63.56}_{\pm 0.18}$	$\mathbf{51.69}_{\pm 0.59}$

Table 5. Last epoch test accuracies (%) of different methods on CIFAR-10N and CIFAR-100N human-annotated noise [26]. The results "mean±std" are reported over 3 random trials and the top-2 best results are in **bold**.

Loss	CIFAR-10			CIFAR-100
	Aggregate	Random 1	Worst	Noisy
CE	$85.09_{\pm 0.30}$	$79.09_{\pm 0.28}$	$61.43_{\pm 0.52}$	$48.63_{\pm 0.53}$
GCE	$87.39_{\pm 0.09}$	$85.98_{\pm 0.42}$	$77.77_{\pm 0.59}$	$50.97_{\pm 0.60}$
SCE	$88.48_{\pm 0.26}$	$85.65_{\pm 0.30}$	$73.65_{\pm 0.29}$	$48.52_{\pm 0.11}$
NCE+RCE	$89.17_{\pm 0.28}$	$87.62_{\pm 0.34}$	$79.74_{\pm 0.09}$	$54.27_{\pm 0.09}$
NCE+AGCE	$89.27_{\pm 0.28}$	$87.92_{\pm 0.02}$	$79.91_{\pm 0.37}$	$55.96_{\pm 0.20}$
CE+LC	$86.60_{\pm 0.40}$	$83.51_{\pm 0.13}$	$70.11_{\pm 0.10}$	$47.76_{\pm 0.29}$
ANL-CE	$89.66_{\pm 0.12}$	$88.68_{\pm 0.13}$	$80.23_{\pm 0.28}$	$56.37_{\pm 0.42}$
ANL-FL	$89.81_{\pm 0.08}$	$88.57_{\pm 0.18}$	$80.56_{\pm 0.23}$	$57.09_{\pm 0.40}$
JAL-CE	$\mathbf{89.94}_{\pm 0.20}$	$\mathbf{88.85}_{\pm 0.23}$	$\mathbf{81.33}_{\pm 0.34}$	$\mathbf{59.54}_{\pm 0.12}$
JAL-FL	$\mathbf{90.06}_{\pm 0.22}$	$\mathbf{88.71}_{\pm 0.30}$	$\mathbf{81.25}_{\pm 0.10}$	$\mathbf{59.38}_{\pm 0.24}$

Experiments



(a) CE with 30 Epochs (b) CE with 60 Epochs (c) CE with 90 Epochs (d) CE with 120 Epochs

(e) JAL-CE with 30 Epochs (f) JAL-CE with 60 Epochs (g) JAL-CE with 90 Epochs (h) JAL-CE with 120 Epochs

Figure 3. Histograms of the distribution of samples with different prediction probabilities in the training set for CIFAR-10 with 0.4 symmetric noise.

Table 6. Last epoch test accuracies (%) of different methods on ILSVRC12, WebVision, and Clothing1M. The top-2 best results are in **bold**.

Loss	CE	GCE	SCE	NCE+RCE	NCE+AGCE	ANL-CE	ANL-FL	**JAL-CE**	**JAL-FL**
WebVision	66.28	61.84	65.16	66.96	67.16	67.36	67.76	**69.84**	**69.20**
ILSVRC12	60.68	60.32	61.00	63.96	64.36	65.60	64.84	**66.64**	**66.00**
Clothing1M	67.93	68.46	67.71	69.24	67.90	69.75	69.90	**70.31**	**70.11**

Thanks for your attention!

Any question? Please contact us!

Jialiang Wang: cswjl@stu.hit.edu.cn