Learning from Noisy Labels with Decoupled Meta Label Purifier

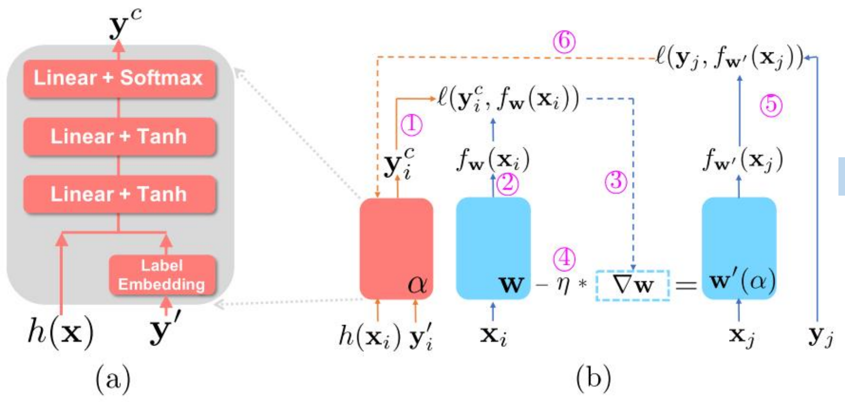
Yuanpeng Tu[1] , Boshen Zhang[2], Yuxi Li[2], Liang Liu[2], Jian Li[2],
Yabiao Wang[2], Chengjie Wang[2,3†], Cai Rong Zhao[1†]

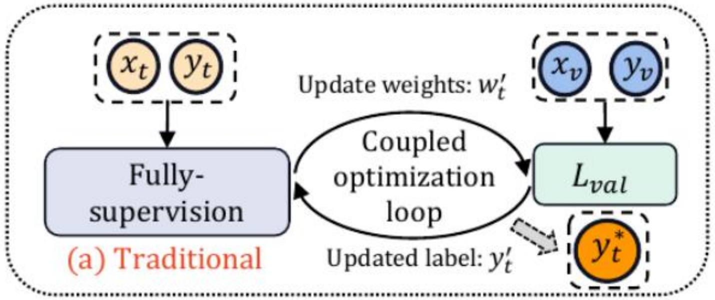
1 Tongji University, 2 Tencent Youtu Lab, 3 Shanghai Jiao Tong University

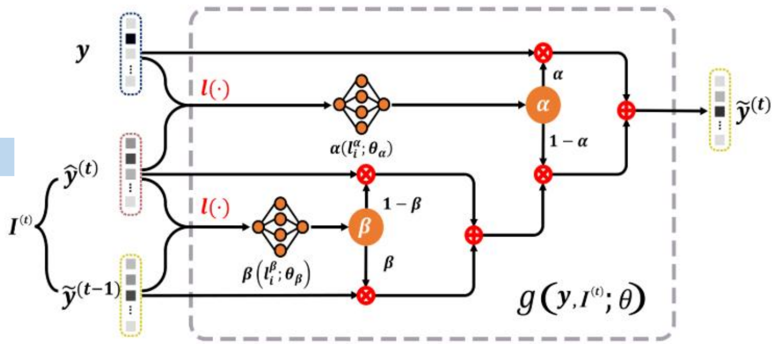
Corresponding authors. Email: zhaocairong@tongji.edu.cn, jasoncjwang@tencent.com

Motivation

Coupled Optimization Methods

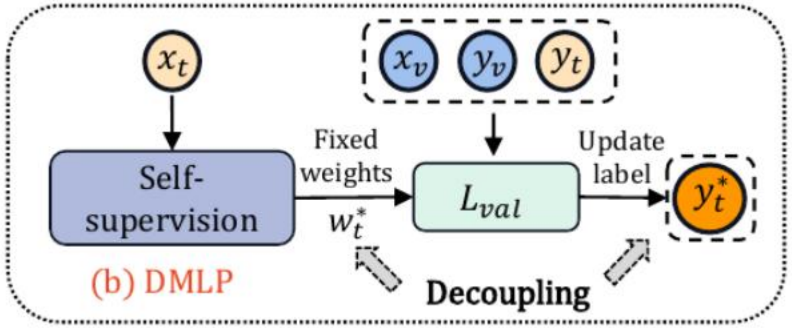






VS

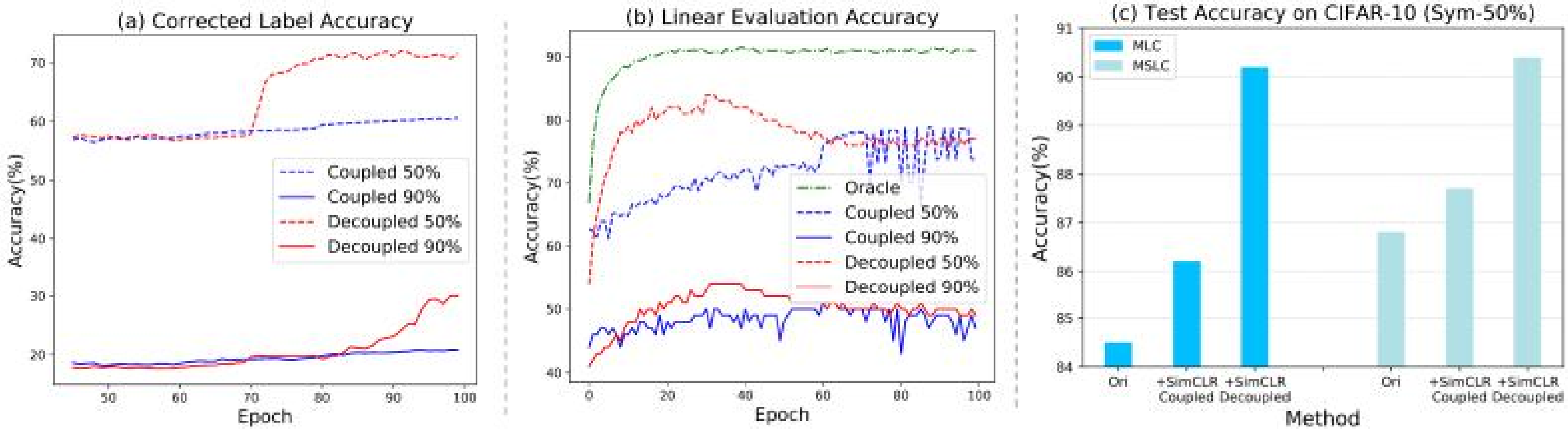
Which one is more suitable for LNL?



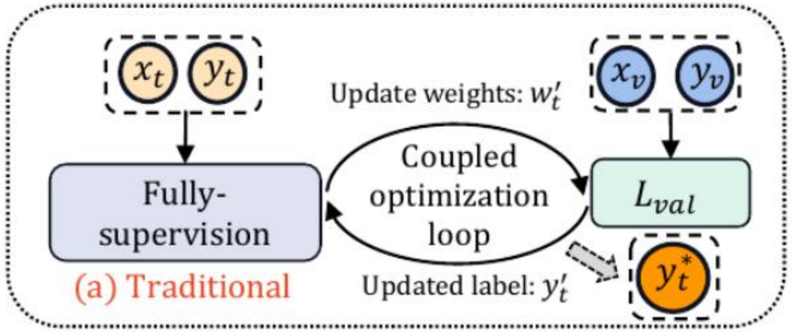
Decoupled Optimization Methods

[1] Learning to Purify Noisy Labels via Meta Soft Label Corrector. AAAI21.
[2] Meta Label Correction for Noisy Label Learning. AAAI21.

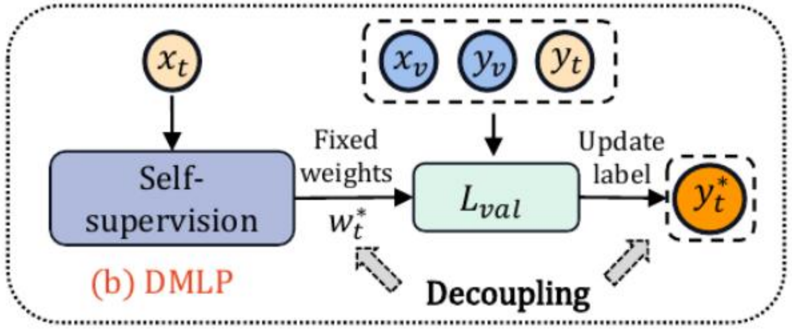
Motivation



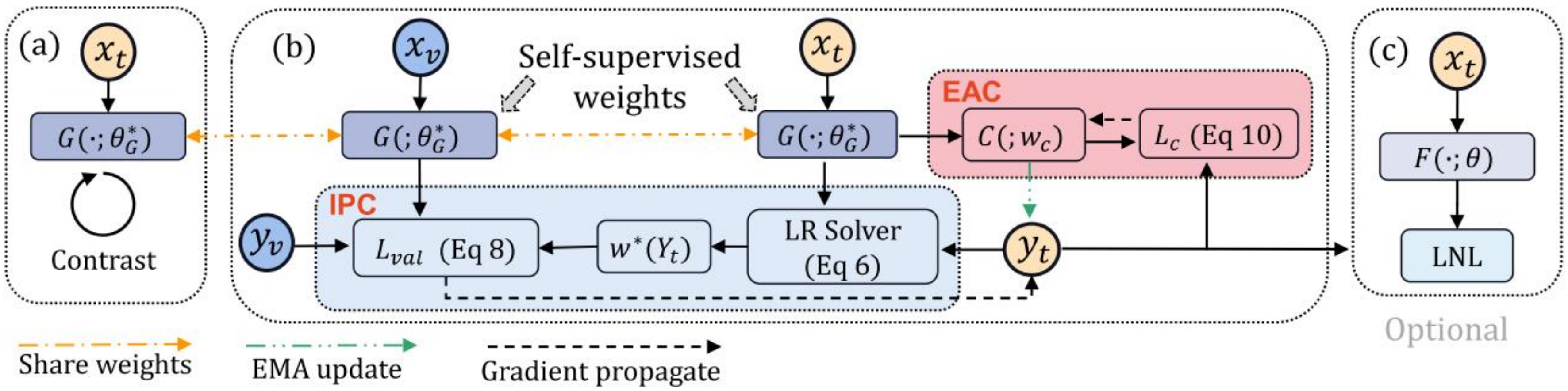
Conclusion: Previous coupled meta label purifier is sub-optimal in terms of both model weights and labels.



<



Framework



Intrinsic Primary Correction

$$\min_{\theta_\alpha} E_{(x_v,y_v)\in D_v}\mathcal{L}_{val}(x_v,y_v;w^*(\theta_\alpha))$$

$$\textbf{s.t.}\quad w^*(\theta_\alpha)=\arg\min_{w} E_{(x_t,y_t)\in D_t}\mathcal{L}_{train}(x_t,y_t;w,\theta_\alpha)$$

Training Data Matrix:

$$F_t=[\mathbf{f}_{t,1},\mathbf{f}_{t,2},\cdots,\mathbf{f}_{t,b}]^T,\quad Y_t=[y_{t,1},y_{t,2},\cdots,y_{t,b}]^T.$$

$$\min_{w}\mathcal{L}_{train}(F_t,Y_t;w)=\|Y_t-F_tw\|^2.$$

Least Square Method

$$w^*(Y_t)=\left({F_t}^TF_t\right)^{-1}{F_t}^TY_t.$$

Optimal Solution of Train set

Classification on Val set

$$y'_{v,i}(Y_t)=w^*(Y_t)^T\mathbf{f}_{v,i}.$$

Minimize Discrepancy

$$\mathcal{L}_{\mathrm{val}}(Y_t)=\frac{1}{N_v}\sum_{i=1}^{N_v}||y'_{v,i}(Y_t)-y_{v,i}||^2+\mathcal{H}(y'_{v,i}(Y_t)),$$

Update Noisy Labels

$$Y_t^{p+1}:=Y_t^p-\eta_I\nabla(\mathcal{L}_{\mathrm{val}}(Y_t^p)).$$

Extrinsic Auxiliary Correction

$$\mathcal{L}_c(w_c)=\mathcal{L}_{ce}(C(\mathbf{f}_t;w_c),y'_t)+\mathcal{H}(C(\mathbf{f}_t;w_c)),$$

Update Noisy Labels

$$Y_t^{p+1}:=(1-\eta_E)Y_t^p+\eta_EC(F_t;w_c)$$

Dataset

Algorithm 1 The workflow of DMLP.

Input: Noisy training set D_t, clean validation set D_v, feature extractor $G(\cdot;\theta_G)$, classifier $C(\cdot;w_c)$, batch size b, max iterations m, period for regular label substitution T.

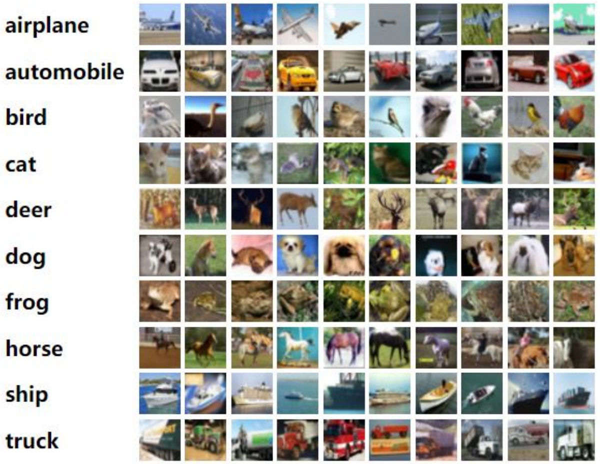
Procedure:

```
Self-supervised training for G(·; θ_G)
Generate features f by Eq. (1)
for i = 1 to m do
    /*IPC starts*/
    {F_t, Y_t} ← SampleMiniBatch(f, D_t, b)
    Calculate closed-form solution w*(Y_t) by Eq. (6)
    Predict validation set labels y'_v by Eq. (7)
    Calculate label purification loss L_val(Y_t) by Eq. (8).
    Update training labels Y_t in backward process.

    /*EAC starts*/
    Calculate loss for the classifier C(·; w_c) by Eq. (10)
    Update classifier parameter w_c in backward process.
    if i = nT then
        Update training labels Y_t by Eq. (11)
    end if
end for
```

Output: The purified labels Y_t^*.

Simulated Noisy Dataset: CIFAR-10/100



Real-world Noisy Dataset: Clothing1M

Symmetric Noise on CIFAR-10/100

Table 1. Comparison with state-of-the-art methods on CIFAR-10/100 datasets with symmetric noise. "CE" is the standard ConvNet trained with Cross-Entropy loss in an end-to-end manner. "Classifier" means adopts the pre-trained SimCLR features to re-train a linear classifier. "Val" denotes using a small clean validation set. DivideMix* denotes training DivideMix with the same validation set as additional data.

Dataset			CIFAR-10				CIFAR-100			
Method	Val	Noise ratio	20%	50%	80%	90%	20%	50%	80%	90%
Cross-Entropy (CE)	✗	Best	86.8	79.4	62.9	42.7	62.0	46.7	19.9	10.1
		Last	82.7	57.9	26.1	16.8	61.8	37.3	8.8	3.5
Co-teaching+ [36]	✗	Best	89.5	85.7	67.4	47.9	65.6	51.8	27.9	13.7
		Last	88.2	84.1	45.5	30.1	64.1	45.3	15.5	8.8
PENCIL [35]	✗	Best	92.4	89.1	77.5	58.9	69.4	57.5	31.1	15.3
		Last	92.0	88.7	76.5	58.2	68.1	56.4	20.7	8.8
REED [38]	✗	Best	95.8	95.6	94.3	93.6	76.7	73.0	66.9	59.6
		Last	95.7	95.4	94.1	93.5	76.5	72.2	66.5	59.4
Sel-CL+ [18]	✗	Best	95.5	93.9	89.2	81.9	76.5	72.4	59.6	48.8
		Last	95.1	93.3	88.7	81.6	76.1	72.0	59.2	48.6
MOIT+ [21]	✗	Best	94.1	91.8	81.1	74.7	75.9	70.6	47.6	41.8
		Last	93.8	91.3	80.6	74.0	75.2	70.1	46.9	41.2
C2D-DivideMix [40]	✗	Best	**96.3**	95.2	94.4	93.5	78.6	76.4	67.7	58.7
		Last	**96.2**	95.1	94.1	93.4	78.3	76.0	67.4	58.4
DivideMix [15]	✗	Best	96.1	94.6	93.2	76.0	77.3	74.6	60.2	31.5
		Last	95.7	94.4	92.9	75.4	76.9	74.2	59.6	31.0
Meta-Learning [16]	✓	Best	92.9	89.3	77.4	58.7	68.5	59.2	42.4	19.5
		Last	92.0	88.8	76.1	58.3	67.7	58.0	40.1	14.3
MLC [41]	✓	Best	92.6	88.1	77.4	67.9	66.8	52.7	21.8	15.0
		Last	91.8	87.5	77.1	67.0	66.5	52.4	18.9	14.2
MSLC [9]	✓	Best	93.4	89.9	69.8	56.1	72.5	65.4	24.3	16.7
		Last	93.3	89.4	68.8	55.2	72.0	64.9	20.5	14.6
DivideMix* [15]	✓	Best	96.1	94.9	93.6	77.3	77.7	74.8	60.7	32.5
		Last	95.9	94.6	93.0	76.5	77.1	74.3	60.5	32.2
DMLP-Naive	✓	Best	94.7	94.2	93.5	92.8	72.7	68.0	63.5	61.3
		Last	94.2	94.0	93.2	92.0	72.3	67.4	63.2	60.9
DMLP-DivideMix	✓	Best	**96.3**	**95.8**	**94.5**	**94.3**	**79.9**	**76.8**	**68.6**	**65.8**
		Last	**96.2**	**95.6**	**94.3**	**94.0**	**79.4**	**76.1**	**68.5**	**65.4**

Asymmetric Noise on CIFAR-10

Table 2. Evaluation results with asymmetric noise of different noisy ratio on CIFAR-10. "Validation" denotes the method exploits a small clean validation set.

Method	Validation	Noisy ratio	
		20%	40%
Joint-Optim [28]	✗	92.8	91.7
PENCIL [35]	✗	92.4	91.2
M-correction [1]	✗	-	86.3
Iterative-CV [4]	✗	-	88.0
DivideMix [15]	✗	93.4	93.4
REED [38]	✗	95.0	92.3
C2D-DivideMix [40]	✗	93.8	93.4
Sel-CL+ [18]	✗	**95.2**	93.4
GCE [11]	✗	87.3	78.1
RRL [17]	✗	-	92.4
Zhang, et al. [39]	✓	92.7	90.2
Meta-Learning [16]	✓	-	88.6
MSLC [9]	✓	94.4	91.6
DMLP-Naive	✓	94.6	93.9
DMLP-DivideMix	✓	**95.2**	**95.0**

Real-world Clothing1M

Table 3. Top-1 testing accuracy on Clothing-1M testset. “Validation” denotes using the validation provided by [31].

Method	Validation	Top-1 Accuracy
PENCIL [35]	✗	73.49
DivideMix [15]	✗	74.76
RRL [17]	✗	74.90
GCE [11]	✗	73.30
C2D-DivideMix [40]	✗	74.30
REED [38]	✗	75.81
Meta-Learning [16]	✔	73.47
Self-Learning [13]	✔	76.44
MLC [41]	✔	75.78
MSLC [9]	✔	74.02
Meta-Cleaner [10]	✔	72.50
Meta-Weight [8]	✔	73.72
FaMUS [34]	✔	74.40
MSLG [7]	✔	76.02
DMLP-Naive	✔	77.77
DMLP-DivideMix	✔	**78.23**

Generality of DMLP

Table 4. Comparison between the LNL methods and their DMLP applications with symmetric noise on CIFAR-10/100. Specifically, the 9-layer CNN is adopted as the backbone network of Co-teaching.

Dataset		CIFAR-10				CIFAR-100			
Method/Noise ratio		20%	50%	80%	90%	20%	50%	80%	90%
Co-teaching [12]	Best	82.6	73.0	24.0	14.6	50.5	38.2	11.8	4.9
	Last	81.9	72.6	23.5	11.7	50.3	38.0	11.3	4.3
DMLP-Co-teaching	Best	**85.8**	**85.8**	**85.4**	**84.6**	**51.2**	**49.8**	**48.1**	**45.3**
	Last	**85.6**	**85.6**	**85.3**	**84.5**	**51.0**	**49.3**	**47.8**	**45.1**
CDR [30]	Best	90.4	85.0	47.2	12.3	63.3	39.5	29.2	8.0
	Last	82.7	49.4	16.6	10.1	62.9	39.5	9.7	4.5
DMLP-CDR	Best	**91.4**	**91.2**	**91.2**	**90.2**	**69.2**	**64.8**	**61.4**	**58.5**
	Last	**91.2**	**90.8**	**90.6**	**89.3**	**68.3**	**64.3**	**61.1**	**57.9**
ELR+ [19]	Best	94.6	93.8	91.1	75.2	77.5	72.4	58.2	30.8
	Last	94.4	93.7	90.5	73.5	76.2	72.2	56.8	30.6
DMLP-ELR+	Best	**94.9**	**94.1**	**93.0**	**92.5**	**77.8**	**73.6**	**63.9**	**60.5**
	Last	**94.6**	**94.0**	**92.7**	**92.1**	**77.1**	**73.4**	**63.6**	**60.5**

Component Analysis

Table 5. Ablation study for the effectiveness of IPC and EAC in DMLP-Naive on CIFAR-10.

Component			CIFAR-10				Clothing
IPC	EAC		20%	50%	80%	90%	1M
✗	✓	Best	93.7	93.3	91.1	67.4	76.5
		Last	93.0	92.9	90.6	66.5	76.1
✓	✗	Best	87.8	85.7	79.9	76.0	76.8
		Last	87.2	85.5	79.4	75.4	76.5
✓	✓	Best	**94.7**	**94.2**	**93.5**	**92.8**	**77.7**
		Last	**94.2**	**94.0**	**93.2**	**92.0**	**77.6**

Comparison against other coupled purifiers with pre-training.

Table 6. Comparison with coupled meta label correction methods MLC [41] and MSLC [9] on CIFAR-10. "*" denotes training with SimCLR pretrained ResNet-18.

Method		Noisy ratio 20%	50%	80%	90%
MLC*	Best	91.8	86.2	77.6	72.9
	Last	91.6	85.9	77.5	72.6
MSLC*	Best	92.0	87.7	78.0	67.8
	Last	92.0	87.5	77.9	67.3
DMLP-Naive*	Best	**94.0**	**93.7**	**93.1**	**92.3**
	Last	**93.9**	**93.4**	**92.9**	**91.9**
MLC*-DivideMix	Best	95.3	94.0	93.0	86.6
	Last	95.0	93.6	92.7	86.5
MSLC*-DivideMix	Best	95.7	94.9	93.8	83.0
	Last	95.5	94.8	93.1	82.8
DMLP*-DivideMix	Best	**96.3**	**95.6**	**94.1**	**93.8**
	Last	**96.0**	**95.2**	**94.0**	**93.6**

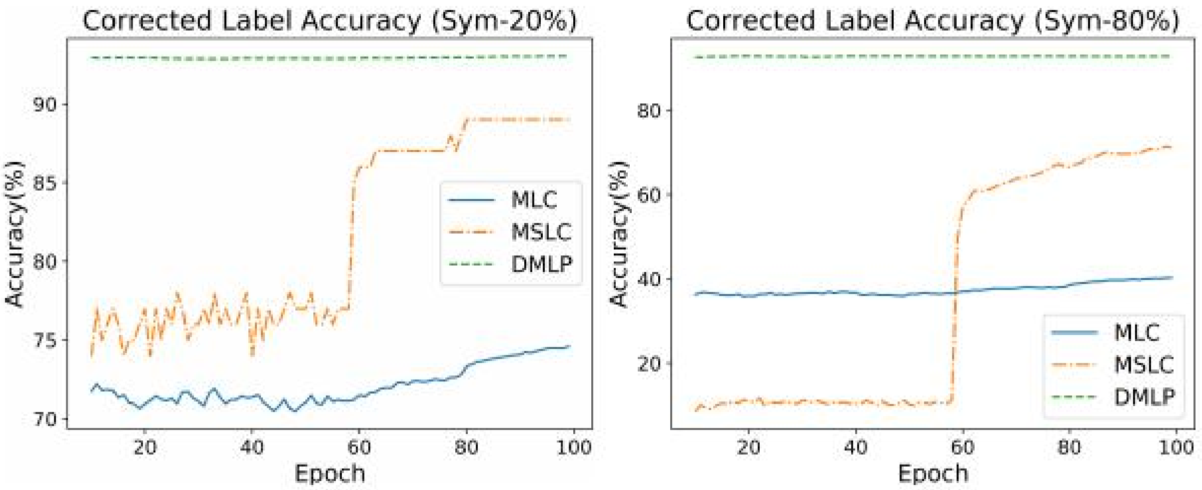


Figure 4. Comparison of corrected label accuracy under symmetric-20% (left), symmetric-80% (right) noise settings on CIFAR-10.

Effect of validation size

Table 7. Investigation of the validation set size τ on Clothing1M.

τ	10%	20%	30%	40%	50%	100%
Accuracy (%)	75.50	76.40	76.61	77.00	77.30	**77.31**

Performance under extremely noisy setting

Table 9. Comparison between recent semi-supervised methods and DMLP-DivideMix on CIFAR-10/100 with 100% noisy ratio.

Method	CIFAR-10	CIFAR-100
MeanTeacher	83.0	31.0
MixMatch	87.9	57.7
FixMatch	88.1	56.3
UDA	88.2	56.1
Ours	**91.7**	**60.1**

Effect of different feature representation for purification

Table 8. Ablation study for adopting different features in DMLP-Naive on CIFAR-10, where "R18/50" denote "ResNet-18/50" and "M/S" represent "MoCo/SimCLR".

Feature Source		Noisy ratio 20%	50%	80%	90%
R18 (M)	Best	93.8	93.3	92.2	90.4
	Last	93.7	92.7	92.1	90.0
R18 (S)	Best	94.0	93.7	93.1	92.3
	Last	93.9	93.4	92.9	91.9
R50 (S)	Best	**94.7**	**94.2**	**93.5**	**92.8**
	Last	**94.2**	**94.0**	**93.2**	**92.0**

同濟大學
TONGJI UNIVERSITY

JUNE 18-22, 2023
CVPR
VANCOUVER, CANADA

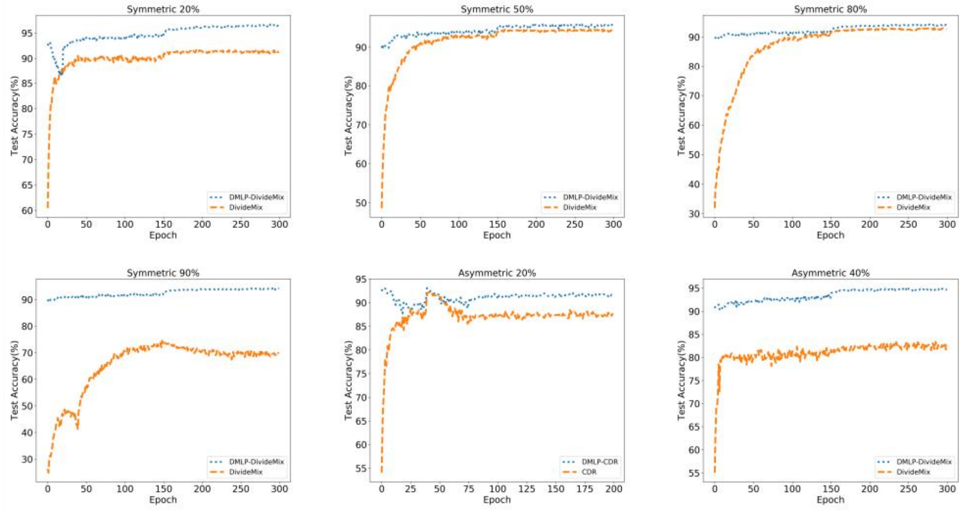


Figure 4. Accuracy curve of DMLP-DivideMix and DivideMix on CIFAR-10 under different noise settings.

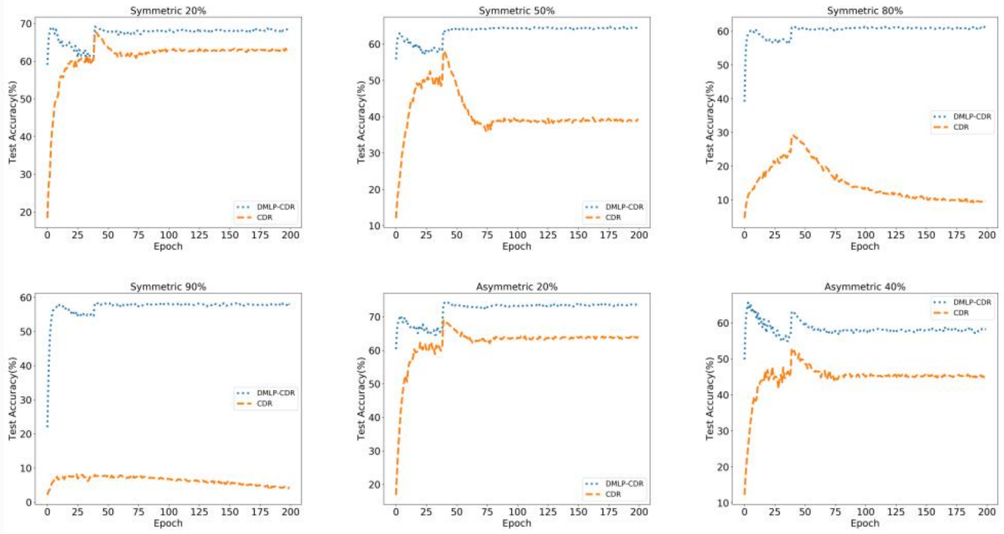


Figure 5. Accuracy curve of DMLP-DivideMix and DivideMix on CIFAR-100 under different noise settings.

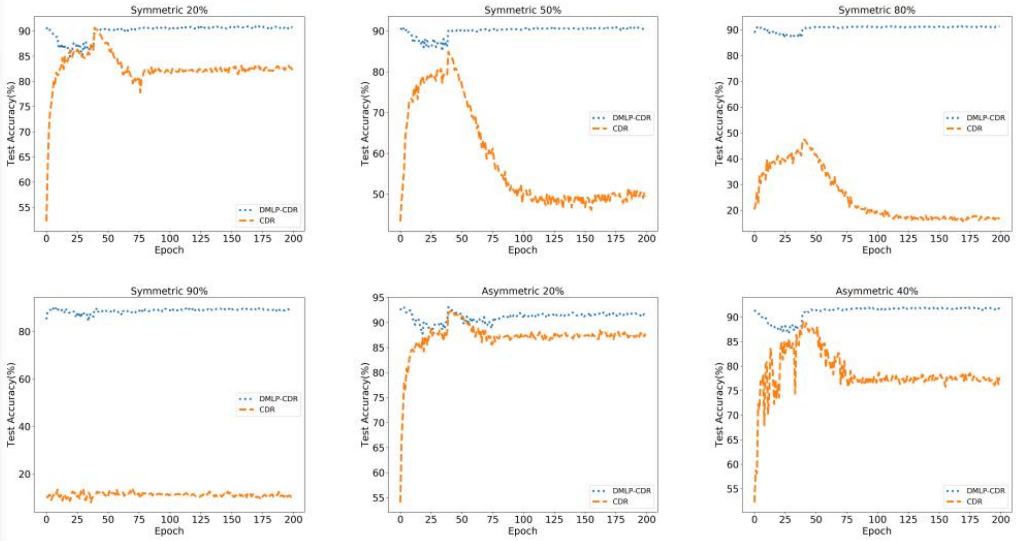


Figure 6. Accuracy curve of DMLP-CDR and CDR on CIFAR-10 under different noise settings.

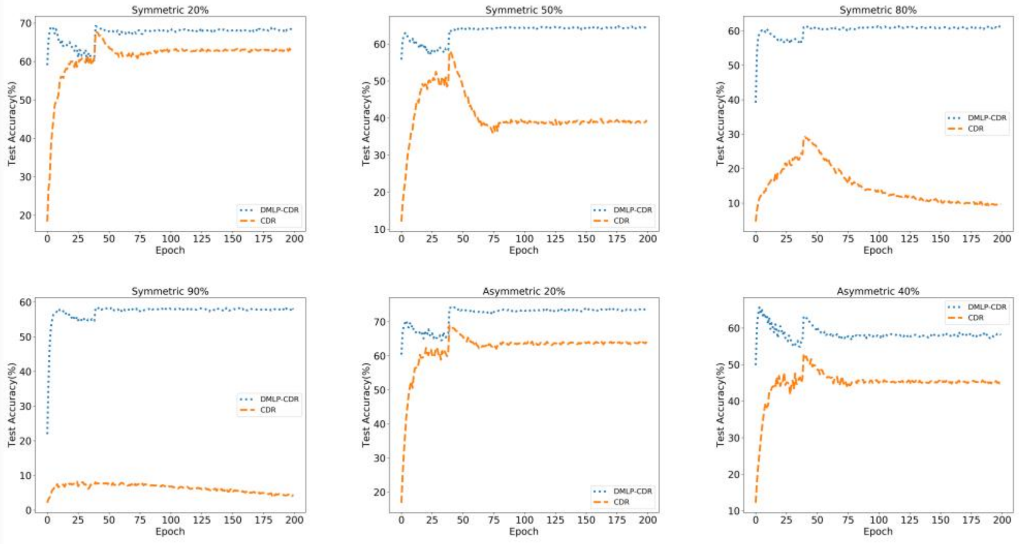


Figure 7. Accuracy curve of of DMLP-CDR and CDR on CIFAR-100 under different noise settings.

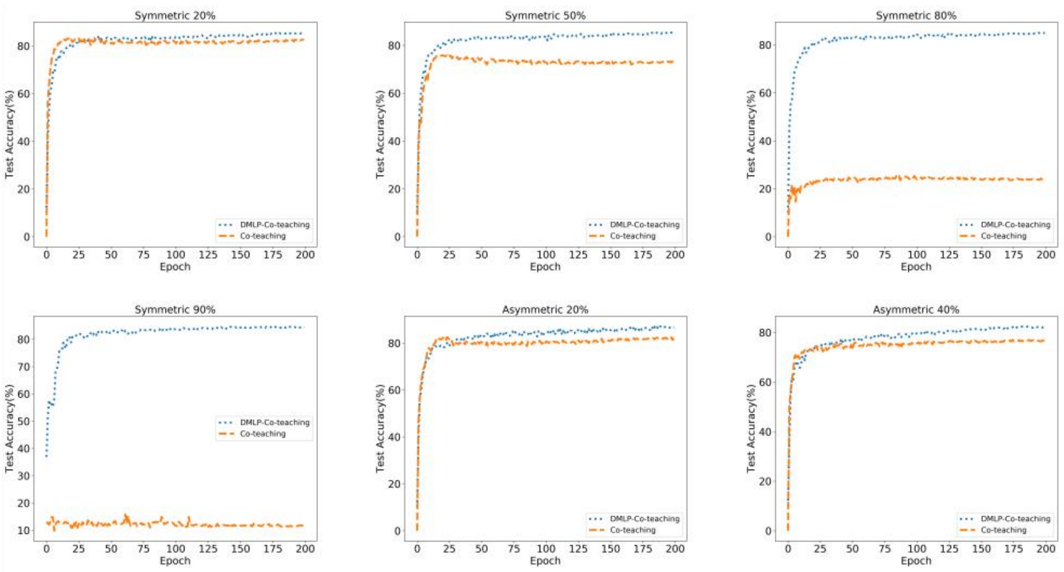


Figure 8. Accuracy curve of DMLP-Co-teaching and Co-teaching on CIFAR-10 under different noise settings.

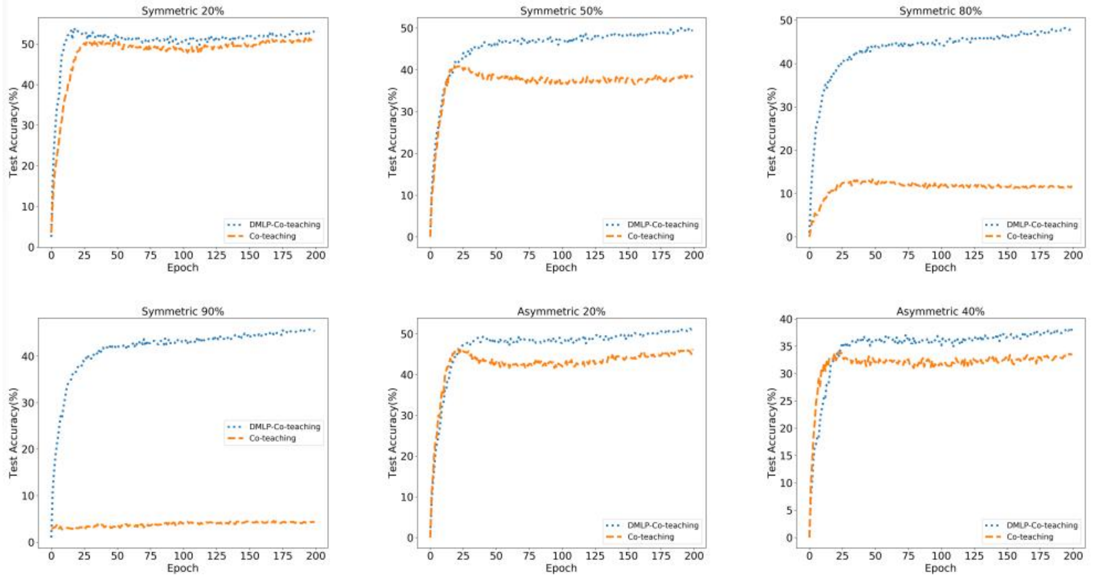


Figure 9. Accuracy curve of of DMLP-Co-teaching and Co-teaching on CIFAR-100 under different noise settings.

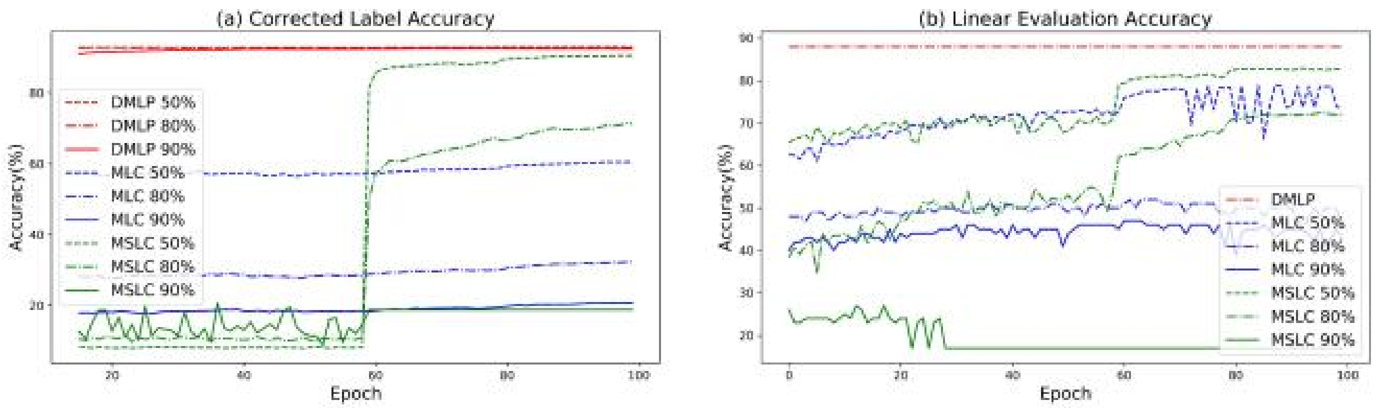


Figure 1. Comparison with two state-of-the-art coupled optimization based meta label correction methods MLC [25] and MSLC [6] on corrected label accuracy and linear evaluation accuracy.

Table 2. Comparison with MLC and MSLC on CIFAR-10/100. "†" denotes training with fixed self-supervised pretrained ResNet-18. "*" denotes training with self-supervised pretrained ResNet-18.

Dataset / Method	Noise ratio	CIFAR-10 20%	CIFAR-10 50%	CIFAR-10 80%	CIFAR-10 90%	CIFAR-100 20%	CIFAR-100 50%	CIFAR-100 80%	CIFAR-100 90%
MLC* [25]	Best	91.8	86.2	77.6	72.9	62.2	53.8	46.5	39.6
	Last	91.6	85.9	77.5	72.6	61.6	53.0	46.2	39.2
MSLC* [6]	Best	92.0	87.7	78.0	67.8	70.8	64.1	36.4	19.8
	Last	92.0	87.5	77.9	67.3	70.2	63.8	34.3	18.7
MLC† [25]	Best	92.0	90.2	89.0	88.9	65.9	59.4	54.4	54.2
	Last	91.6	89.4	88.5	88.1	65.2	59.2	54.1	54.0
MSLC† [6]	Best	92.1	90.4	87.3	84.7	71.7	64.7	53.3	46.8
	Last	92.0	90.0	87.2	84.2	71.6	64.4	53.0	46.4
DMLP-Naive	Best	**94.7**	**94.2**	**93.5**	**92.8**	**72.7**	**68.0**	**63.5**	**61.3**
	Last	**94.2**	**94.0**	**93.2**	**92.0**	**72.3**	**67.4**	**63.2**	**60.9**

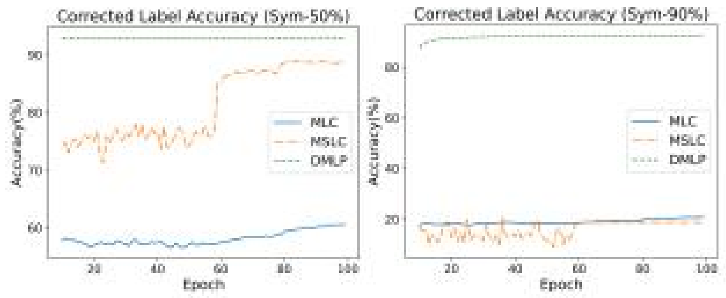


Figure 2. Comparison of corrected label accuracy curve under symmetric-50% (left), symmetric-90% (middle) noise settings on CIFAR-10.

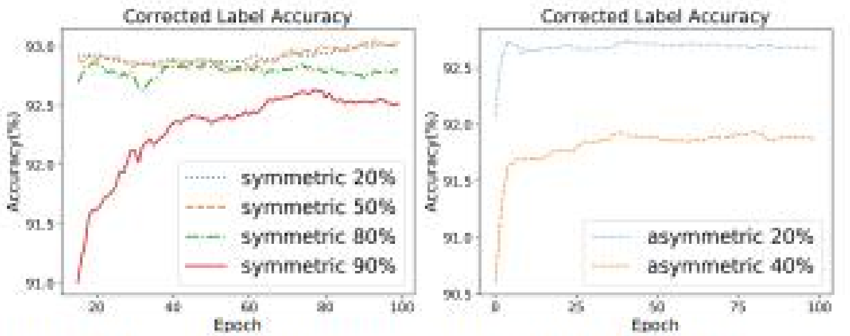


Figure 3. Corrected label accuracy curve of DMLP under symmetric (left), asymmetric (middle) noise settings on CIFAR-10.

Table 4. Comparison on CIFAR-10/100 with symmetric noise.

Dataset / Method	CIFAR-10 20%	CIFAR-10 50%	CIFAR-10 80%	CIFAR-10 90%
DMLP-Naive	94.28±0.10	94.02±0.21	93.31±0.19	92.16±0.20
DMLP-DivideMix	**96.20±0.11**	**95.63±0.13**	**94.22±0.14**	**93.97±0.22**

Dataset / Method	CIFAR-100 20%	CIFAR-100 50%	CIFAR-100 80%	CIFAR-100 90%
DMLP-Naive	72.39±0.08	67.60±0.26	63.17±0.14	61.09±0.20
DMLP-DivideMix	**79.31±0.21**	**76.11±0.10**	**68.42±0.12**	**65.55±0.23**

Learning from Noisy Labels with Decoupled Meta Label Purifier

Thanks for Watching!

https://github.com/yuanpengtu/DMLP

Yuanpeng Tu[1] , Boshen Zhang[2], Yuxi Li[2], Liang Liu[2], Jian Li[2],
Yabiao Wang[2], Chengjie Wang[2,3†], Cai Rong Zhao[1†]
1 Tongji University, 2 Tencent Youtu Lab, 3 Shanghai Jiao Tong University
Corresponding authors. Email: zhaocairong@tongji.edu.cn, jasoncjwang@tencent.com