Directional Label Diffusion Model for Learning from Noisy Labels

Senyu Hou[1], Gaoxia Jiang[1], Jia Zhang[1], Shangrong Yang[1], Husheng Guo[1],Yaqing Guo[1], Wenjian Wang[2]*

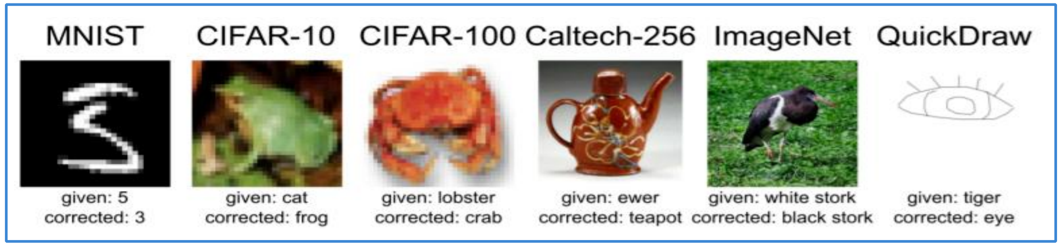
[1]School of Computer and Information Technology, Shanxi University, China
[2]Key Laboratory of Data Intelligence and Cognitive Computing, Shanxi University, China

1. Introduction: Background

The Generality and Impact of Label Noise

- **Generality:** Label noise are **numerous** and **widespread.**

Benchmarks:

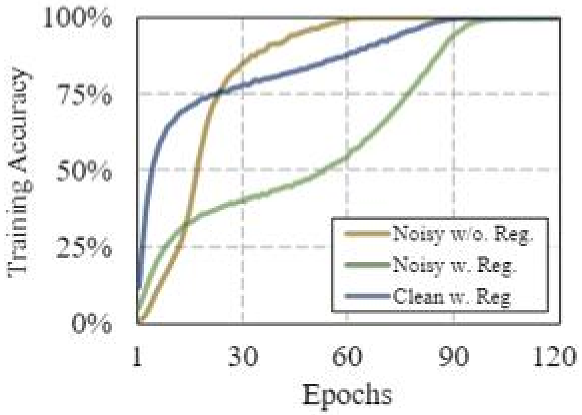


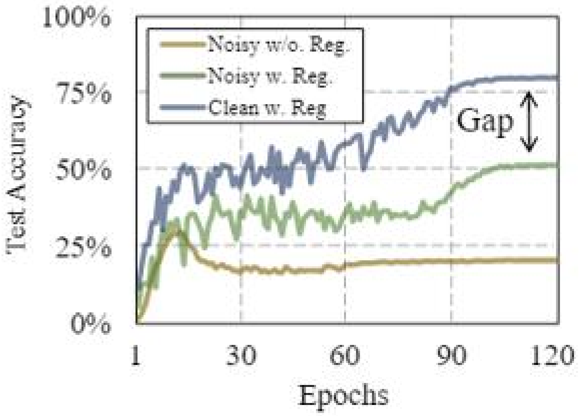
Animal-10N:

WebVision:

Clothing1M:

- **Impact:** Misguides the model training





1. Introduction: Motivation

SHANXI UNIVERSITY

Troditional methods to deal with label noise

- ✓ Methods: Robust Loss Functions, Regularization, Sample Weighting, Robust Architecture, or Ensemble Methods
- ✓ Limitation: Discriminative paradigm

Generative perspective

- ✓ Methods: Take the labels as the diffusion objects and the image features as the conditionl information
- ✓ Limitation: Non-explicit Robustness

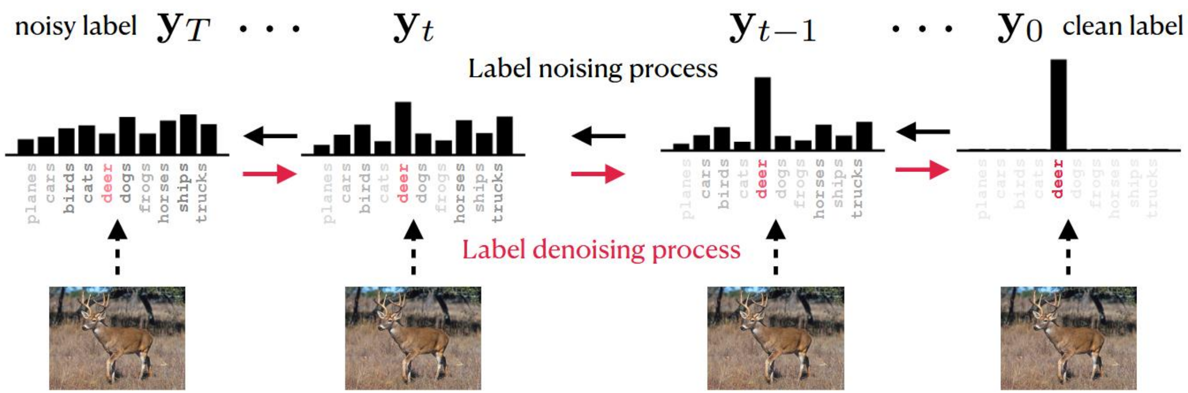
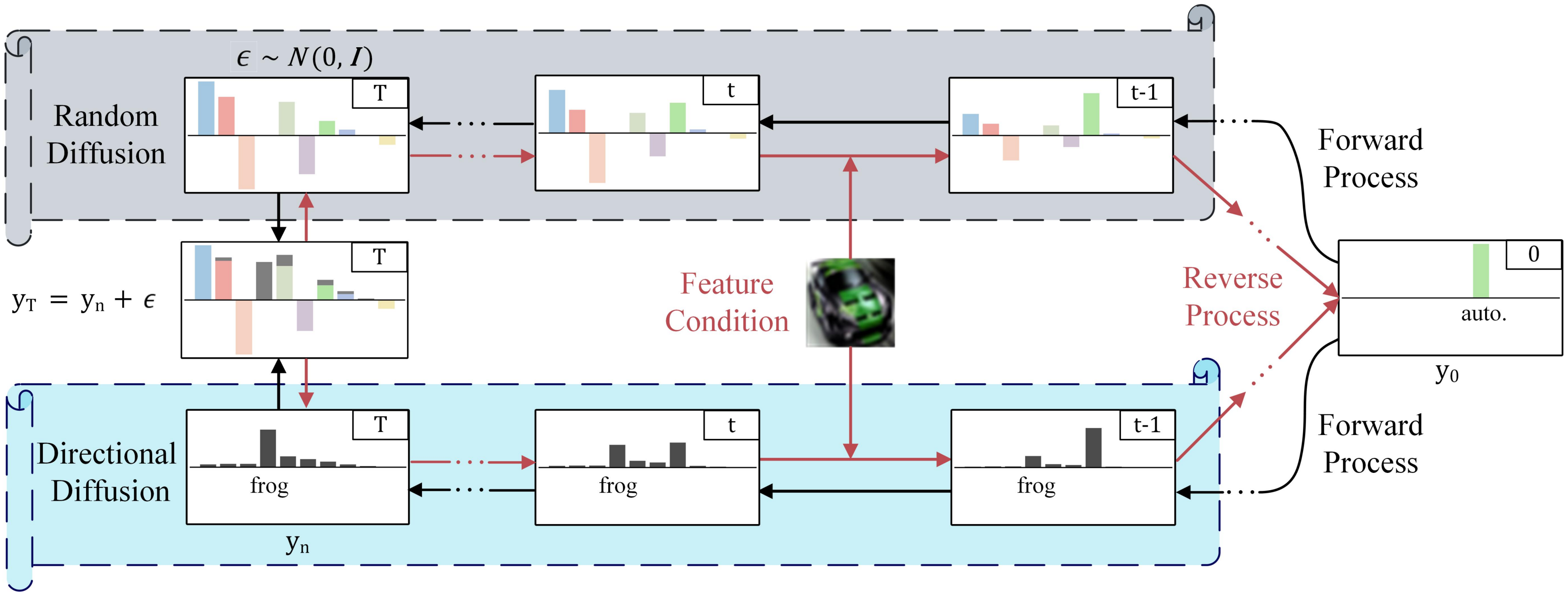


Figure 1: Label denoising as a reverse noising process.

2. Method: Directional Label Diffusion

We proposes a novel directional label diffusion (DLD) paradigm to **expand the diffusion model into a robust classifier** that explicitly accommodates more noise knowledge.



2. Method: Directional Label Diffusion

SHANXI UNIVERSITY

● **Forward Process**

A new forward process containing label noise information

$$
\begin{aligned}
y_t &= y_{t-1} + \boxed{\alpha_t y_d} + \beta_t \epsilon_{t-1}, \\
&= y_{t-2} + \boxed{(\alpha_{t-1} + \alpha_t) y_d} + (\sqrt{\beta_{t-1}^2 + \beta_t^2}) \epsilon_{t-2} \\
&= \ldots \\
&= y_0 + \boxed{\bar{\alpha}_t y_d} + \bar{\beta}_t \epsilon,
\end{aligned}
$$

The new forward distribution

$$
\begin{aligned}
\hat{q}\left(\mathbf{y}_t | \mathbf{y}_{t-1}, \mathbf{y}_d, \mathbf{x}\right) := q\left(\mathbf{y}_t | \mathbf{y}_{t-1}, \mathbf{y}_d\right) \\
= \mathcal{N}\left(\mathbf{y}_t ; \mathbf{y}_{t-1} + \alpha_t \mathbf{y}_d, \beta_t^2 \mathbf{I}\right),
\end{aligned}
$$

$$
\begin{aligned}
\hat{q}\left(\mathbf{y}_{1:T} \mid \mathbf{y}_0, \mathbf{y}_d, \mathbf{x}\right) := \prod_{t=1}^{T} \hat{q}\left(\mathbf{y}_t \mid \mathbf{y}_{t-1}, \mathbf{y}_d, \mathbf{x}\right) \\
= \mathcal{N}\left(\mathbf{y}_t ; \mathbf{y}_0 + \bar{\alpha}_t \mathbf{y}_d, \bar{\beta}_t^2 \mathbf{I}\right),
\end{aligned}
$$

2. Method: Directional Label Diffusion

● Reverse Process

$$\hat{q}\left(\mathbf{y}_{t-1} \mid \mathbf{y}_t, \mathbf{y}_0, \mathbf{y}_d, \mathbf{x}\right) = \mathcal{N}\left(\mathbf{y}_{t-1}; \hat{\mu}_t, \hat{\sigma}_t \mathbf{I}\right), \qquad (10)$$

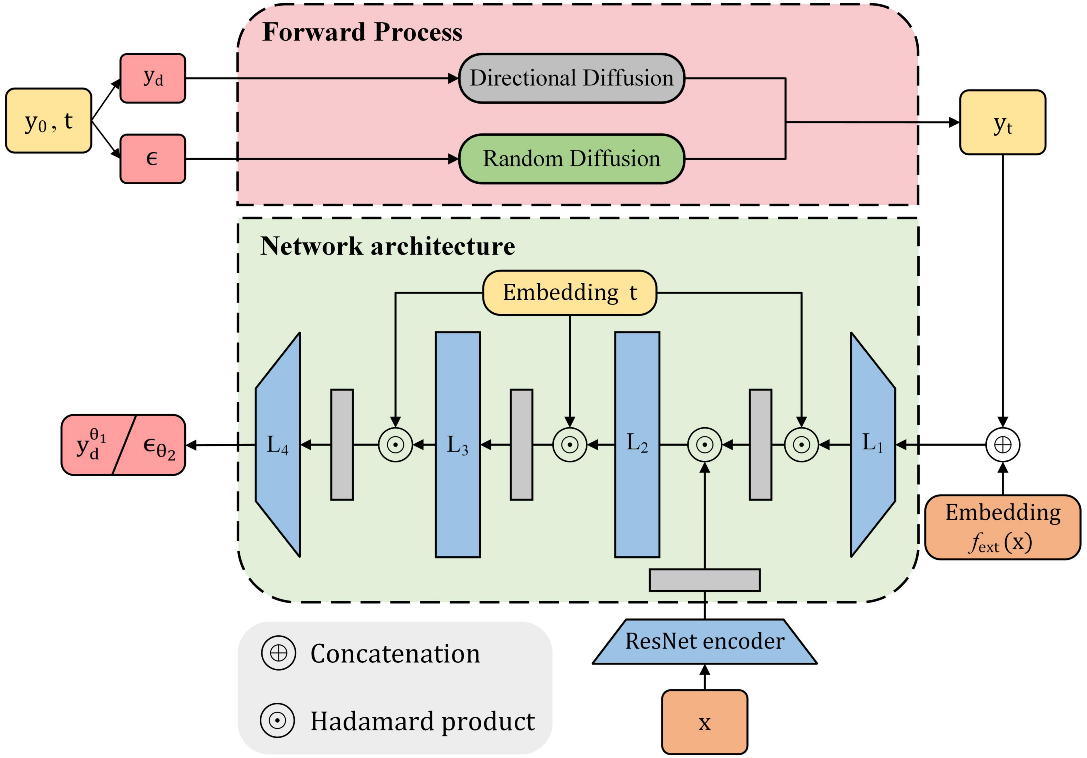
where $\hat{\mu}_t = y_t - \alpha_t y_d - \frac{\beta_t^2}{\bar{\beta}_t}\epsilon + \hat{\sigma}_t \nabla_{y_{t-1}} \log \hat{q}\left(\mathbf{x} | \mathbf{y}_{t-1}, \mathbf{y}_d\right)$ and $\hat{\sigma}_t = \frac{\beta_t^2 \bar{\beta}_{t-1}^2}{\bar{\beta}_t^2}$.

Minimize the KL divergence

$$p_\theta\left(\mathbf{y}_{t-1} \mid \mathbf{y}_t, \mathbf{x}\right) := \mathcal{N}\left(\mathbf{y}_{t-1}; \mu_\theta, \hat{\sigma}_t^2 \mathbf{I}\right). \qquad (11)$$

$$\mathcal{L}_d = \|\, \mathbf{y}_d - \mathbf{y}_d^{\theta_1}\left(\mathbf{y}_t, \mathbf{y}_n, \mathbf{x}, t\right) \|^2, \qquad (12)$$

$$\mathcal{L}_\epsilon = \|\, \epsilon - \epsilon_{\theta_2}\left(\mathbf{y}_t, \mathbf{y}_n, \mathbf{x}, t\right) \|^2, \qquad (13)$$



2. Method: Directional Label Diffusion

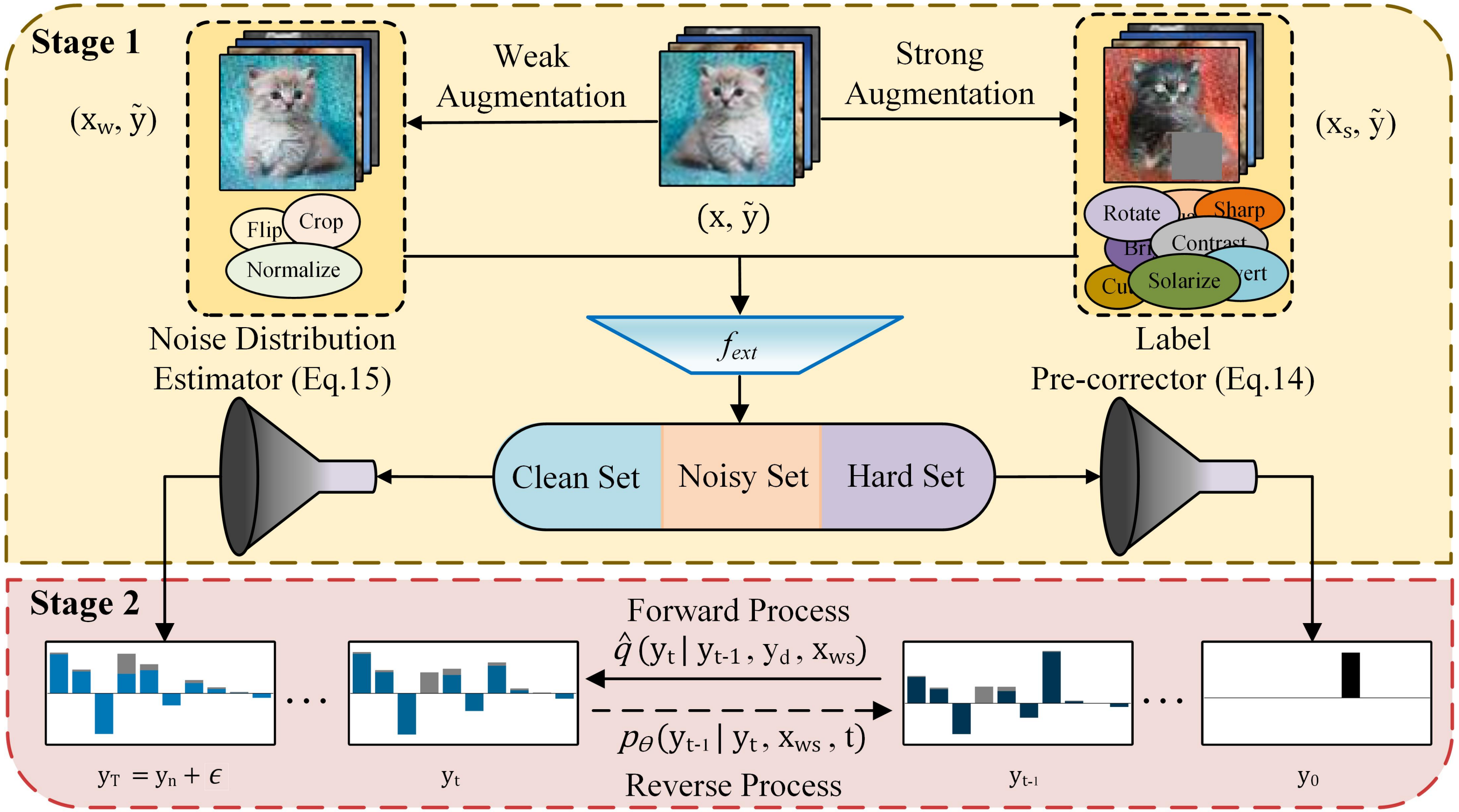
● Pre-correction Method and DLD Model

Label Pre-correction for y_0

$$P(y_0 \mid \tilde{y}, x) = \begin{cases} \tilde{y} & \text{if } x \in \mathcal{D}_C, \\ \operatorname{argmax}(p_{ws}) & \text{if } x \in \mathcal{D}_N, \\ p_{ws} & \text{if } x \in \mathcal{D}_H. \end{cases} \quad (14)$$

Estimate Noise Distribution for y_n

$$P(y_n \mid \tilde{y}, x) = \begin{cases} 0 & \text{if } x \in \mathcal{D}_C, \\ \tilde{y} & \text{if } x \in \mathcal{D}_N, \\ \frac{|p_w - p_s|}{\sum |p_w - p_s|} & \text{if } x \in \mathcal{D}_H. \end{cases} \quad (15)$$



3. Experiments

山西大學 SHANXI UNIVERSITY

Diverse baselines

- ✓ Co-teaching
- ✓ GCE
- ✓ DAC
- ✓ SEAL
- ✓ NCR
- ✓ DISC
- ✓ LRA-diffusion
- ✓ EPL
- ✓ … …

Evaluation metrics

- ✓ Test accuracy
- ✓ Label pre-correction accuracy

Test accuracy

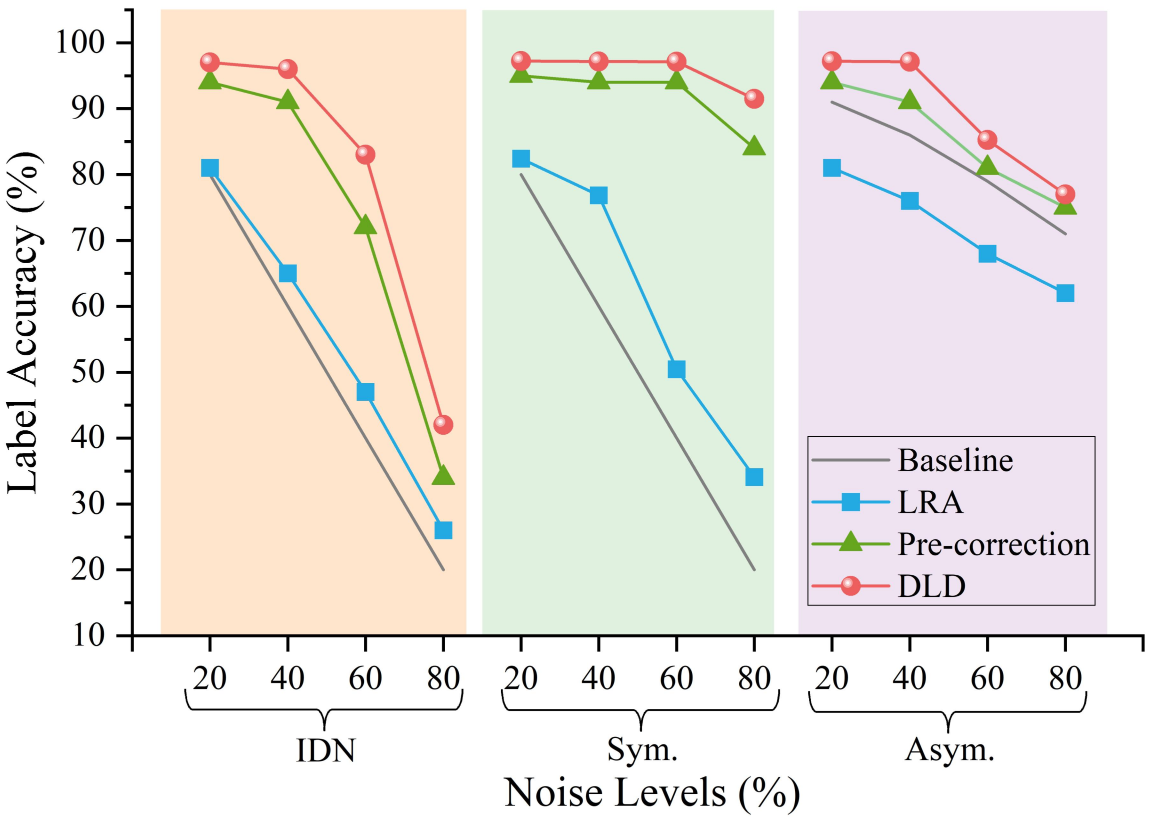
Table 1. Classification accuracies (%) of CIFAR datasets under different levels of IDN.

Method	CIFAR-10			CIFAR-100		
	20%	40%	60%	20%	40%	60%
Co-teaching [17]	87.28±0.20	78.72±0.47	62.51±1.98	69.26±0.39	59.53±0.57	46.80±0.51
GCE [61]	86.44±0.23	76.71±0.39	60.76±3.08	68.24±0.27	55.76±1.18	46.75±2.46
DAC [45]	86.16±0.13	74.80±0.32	59.97±0.90	68.70±0.78	55.71±1.01	46.29±3.35
DMI [54]	86.57±0.16	77.81±0.85	69.94±1.31	68.71±1.21	57.99±0.92	50.96±2.08
SEAL [6]	87.79±0.09	82.98±0.05	70.07±1.54	70.61±0.19	61.18±0.35	51.86±1.32
NCR [22]	87.74±0.15	83.03±0.21	69.44±1.01	70.68±0.37	61.93±0.54	49.13±2.77
CC+DivideMix [62]	88.36±0.11	84.18±0.40	71.52±1.73	70.09±0.19	64.35±0.35	53.20±0.66
DISC [31]	90.68±0.10	85.61±0.32	73.59±1.13	71.49±0.13	64.46±0.29	53.62±0.51
LRA-diffusion+ResNet [4]	89.94±0.15	86.73±0.37	74.66±1.02	72.35±0.14	65.55±0.30	54.43±0.53
DLD+ResNet (Ours)	90.43±0.12	87.17±0.29	76.07±1.10	72.60±0.15	66.59±0.31	55.72±0.24
EPL+ViT [26]	95.95±0.17	93.01±0.33	76.16±1.18	72.24±0.12	67.20±0.20	55.59±0.13
LRA-diffusion+ViT [4]	96.65±0.18	93.68±0.34	78.12±1.17	73.59±0.13	67.67±0.21	57.21±0.12
DLD+ViT (Ours)	**96.98±0.12**	**96.49±0.07**	**83.14±0.81**	**78.62±0.11**	**76.03±0.14**	**69.57±0.17**

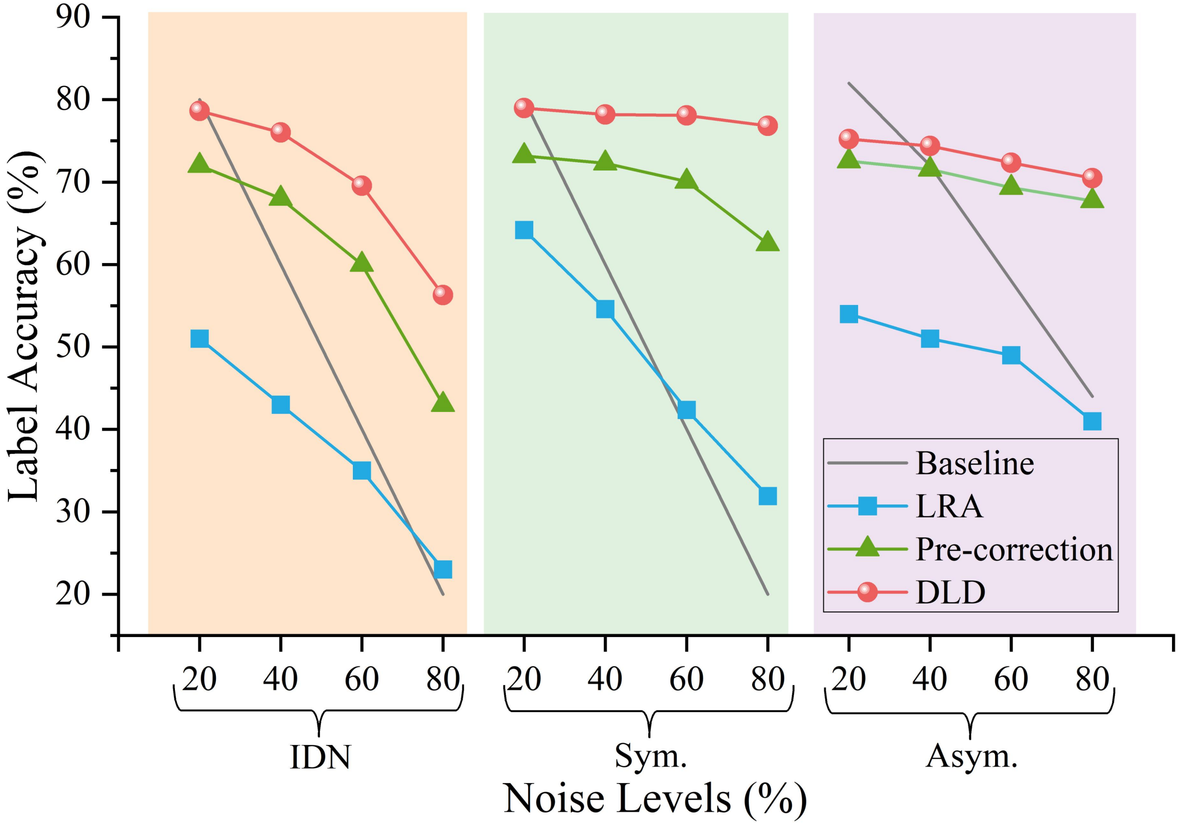
3. Experimental Results

● Label pre-correction accuracy

✓ Results on Noisy CIFAR-10:



✓ Results on Noisy CIFAR-100:



3. Experimental Results

Results on Real-world Noisy Datasets:

Table 2. Classification accur ... acies (%) on Clothing1M.

Method	Accuracy
CE [35]	69.21
GCE [61]	72.46
NCT [8]	74.76
DISC [31]	75.40
SSR [14]	74.79
LRA-diffusion [4]	75.21
J-ViT [9]	74.46
DLD (Ours)	75.63
	75.69

Table 4. Classification accuracies (%) on WebVision and ILSVRC2012 datasets.

Method	WebVision	ILSVRC12
Co-teaching [17]	63.58	61.48
DivideMix [28]	77.32	75.20
ELR [33]	77.78	70.29
EPL [26]	78.77	76.51
C2D [63]	79.42	78.57
CC [62]	79.36	76.08
DISC [31]	80.28	77.44
LRA-diffusion [4]	84.16	82.56
DLD (Ours)	**84.51**	**83.74**

4. Conclusion

SHANXI UNIVERSITY

Main Contibutions

1. Propose dual diffusion architecture for LNL
2. Design fast label pre-correction method
3. Refine feature conditioning

Solved Problems

1. How to address LNL problem from a generative perspective
2. Recovering clean labels from feature with explicit noise information

Thanks for Listening :)

Feel free to contact
Wenjian Wang and
Senyu Hou via:

wjwang@sxu.edu.cn
202312407005@email.sxu.edu.cn

Address:
School of Computer and Information Technology, Shanxi University, Taiyuan 030006, P.R.China

山西大學
SHANXI UNIVERSITY