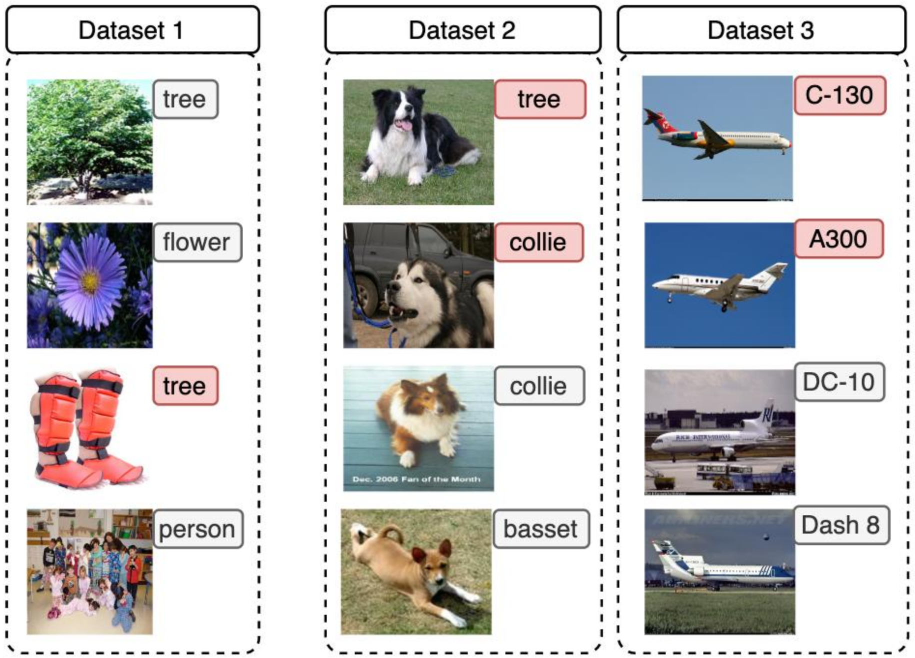
Fine-grained Classification with Noisy Labels

Qi Wei, Lei Feng, Haoliang Sun, Ren Wang, Chenhui Guo, Yilong Yin

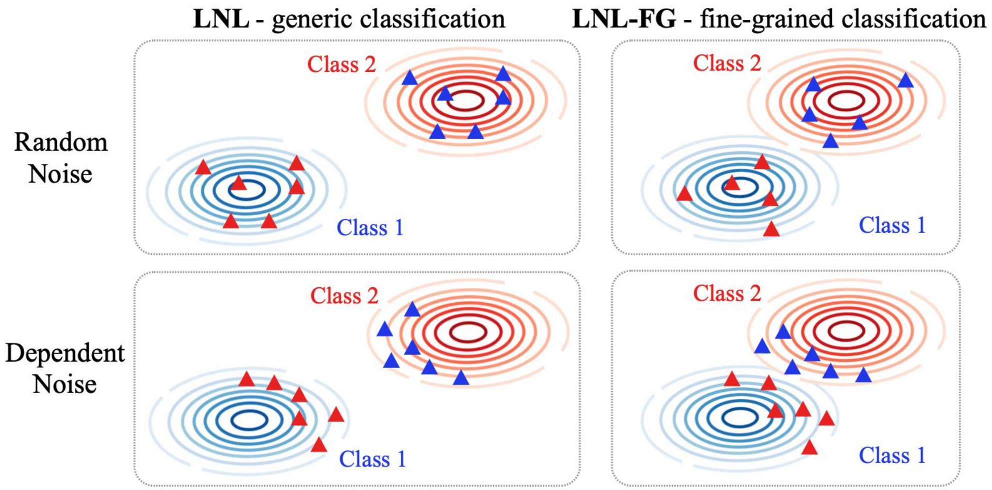
A Novel task in Learning with Noisy Labels (LNL)

Task comparison

Typical **LNL** LNL on fine-grained dataset (**LNL-FG**)



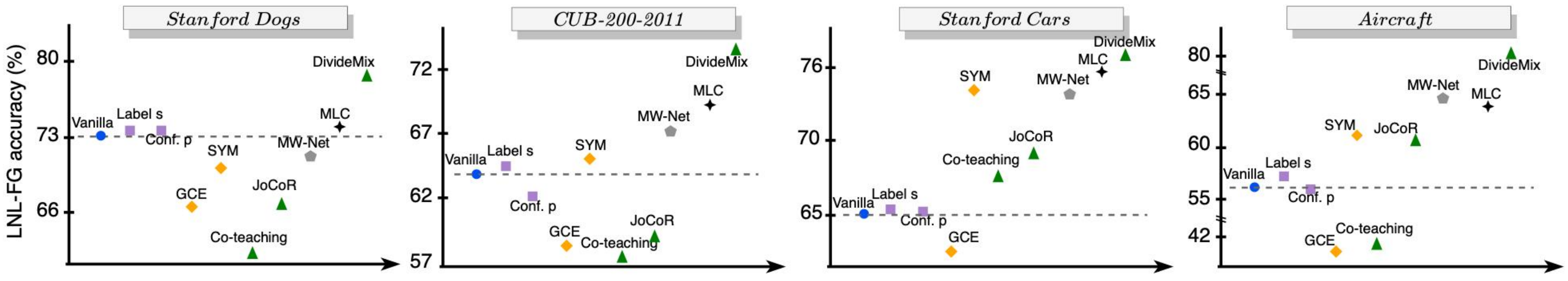
Analysis from feature space



LNL-FG is a more realistic scenario and poses greater challenge !

A prior study on LNL-FG

Results of ten methods on four noisy fine-grained datasets with sym. 20% label noise



Not all robust methods outperform the performance of vinalla cross-entropy on LNL-FG task

Preliminaries

- Problem definition
  - Input space: $\mathcal{X}$
  - Label space: $\mathcal{Y} = \{1, 2, \ldots, C\}$
  - A training set drawn *i.i.d.* from a distribution: $P_{\mathcal{X}\mathcal{Y}}$

- Motivation
  - Challenge: Large inter-class ambiguity among classes in LNL-FG leads to severe overfitting of deep models to noisy labels.
  - Our solution: Encouraging discrimitive feature not only confronts overfitting to label noise but also facilitates the learning of fine-grained task.

Contrastive learning meets noisy labels

We leverage the framework of contrastive learning to enhance discrimitive feature

- Supervised contrastive learning

$$\mathcal{L}_{\mathrm{SCL}} = -\log \frac{\sum\limits_{\mathrm{k_P} \in \mathrm{Pos}} \exp\left(\mathrm{q} \cdot \mathrm{k_P} / \tau\right)}{\sum\limits_{\mathrm{k_P} \in \mathrm{Pos}} \exp\left(\mathrm{q} \cdot \mathrm{k_P} / \tau\right) + \sum\limits_{\mathrm{k_N} \in \mathrm{Neg}} \exp\left(\mathrm{q} \cdot \mathrm{k_N} / \tau\right)},$$

$$\mathbf{q}\text{: a anchor point, }\mathbf{Pos}\text{: the positive list, }\mathbf{Neg}\text{: the negative list}$$

Noisy labels degrades the quality of the anchor, the positive and negative list.

Method：Stochastic Noise-tolerated contrastive learning

➢ Noise-tolerated supervised contrastive learning

- A weight-aware mechanism

Give a sample x_i, the weight can be written as

$$\omega_i = \begin{cases} 1 & \text{if} \quad \gamma_i > t \\ \gamma_i & \text{otherwise} \end{cases}, \qquad \gamma_i = \mathrm{GMM}(l_i \,|\, \{l_i\}_{i=0}^{n})$$

- Two weighted strategies

weighted correction strategy for noisy anchor

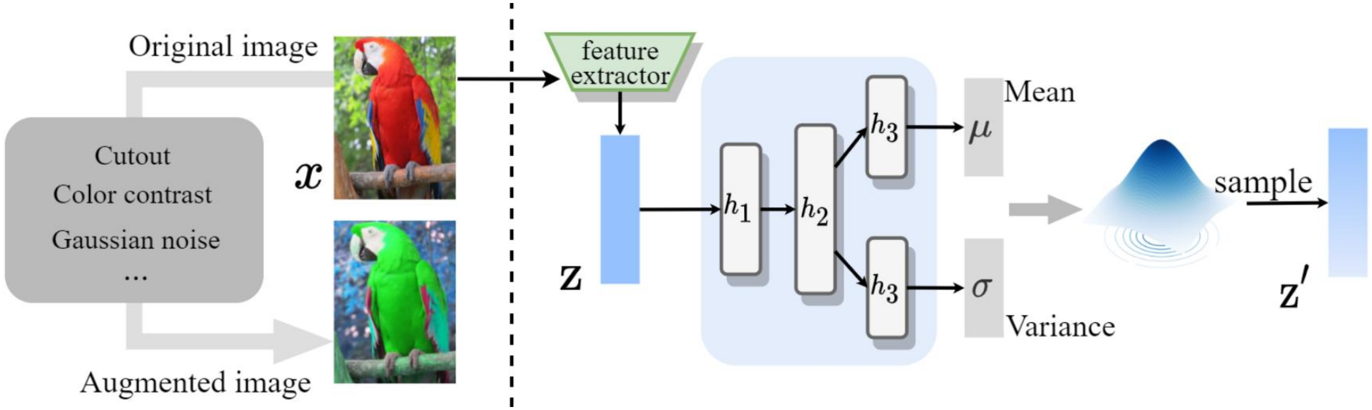
$$\hat{\mathrm{y}} = (1-\omega)\mathrm{y}^{\mathrm{c}} + \omega \mathrm{y}$$
$$\hat{\mathrm{y}}^e = \alpha \hat{\mathrm{y}}^{(e-1)} + (1-\alpha)\hat{\mathrm{y}}^e$$

weighted update strategy for noisy Pos/Neg

Update Pos/Neg with the probability of ω_i

Method：Stochastic Noise-tolerated contrastive learning

➢ Stochastic feature embedding

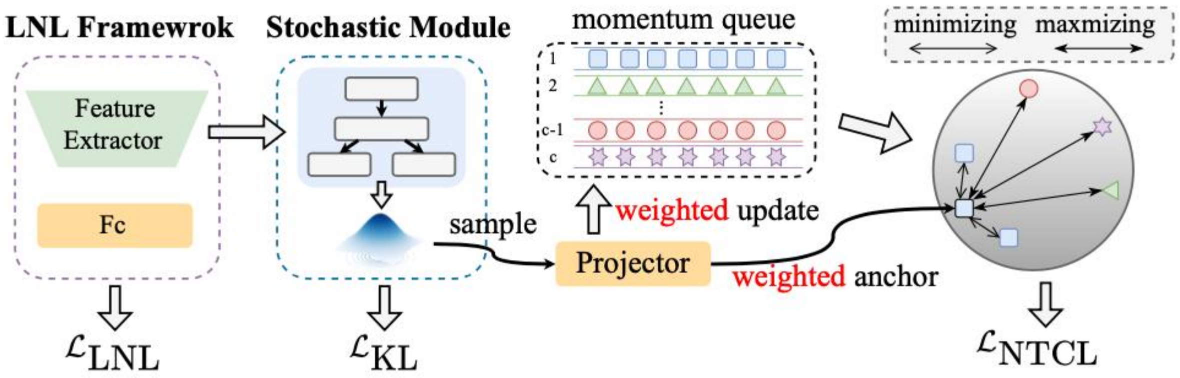


the stochastic feature can be sampled from a generated gaussian distribution

$$p(Q|\mathbf{z}) \sim \mathcal{N}(\mu, \sigma^2)$$

$$\mathbf{z}' = \mu + \epsilon \cdot \sigma \quad \text{with} \quad \epsilon \sim \mathcal{N}(0, \mathbf{I})$$

Our framework



Total training objective

$$\mathcal{L} = \mathcal{L}_{\text{LNL}} + \lambda_1 \mathcal{L}_{\text{NTCL}} + \lambda_2 \mathcal{L}_{\text{KL}}$$

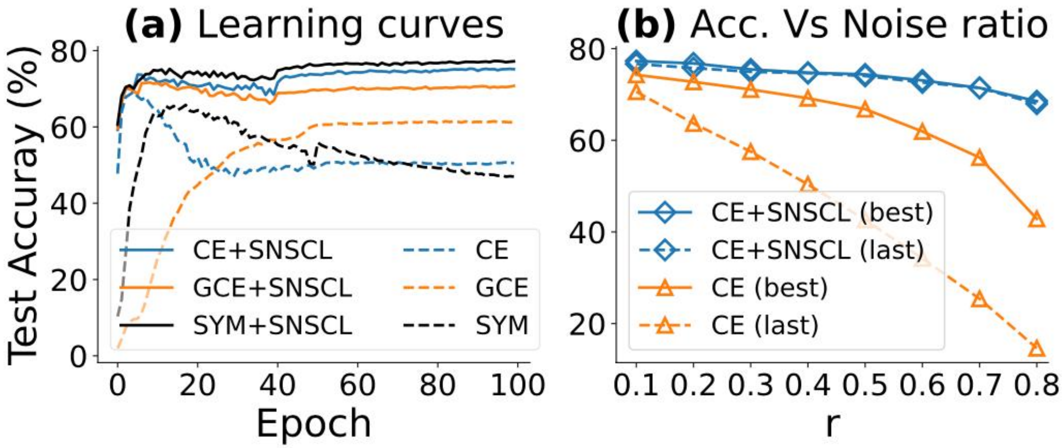
Our proposed SNSCL is generally applicable to prevailing LNL methods and significantly improves their performance on LNL-FG

Experimental results

SNSCL significantly improves current methods on four fine-grained datasets with **sym.** noise

	Stanford Dogs		Standford Cars		Aircraft		CUB-200-2011	
	20%	40%	20%	40%	20%	40%	20%	40%
Cross-Entropy	73.01 (63.82)	69.20 (50.45)	65.74 (64.08)	51.42 (45.62)	56.51 (54.67)	45.67 (38.89)	64.01 (60.77)	54.14 (45.85)
+ SNSCL	76.33 (75.83)	75.27 (75.00)	83.24 (82.99)	76.72 (76.36)	76.45 (76.45)	70.48 (69.64)	73.32 (72.99)	68.83 (68.67)
Label Smooth [24]	73.51 (64.42)	70.22 (50.97)	65.45 (64.24)	51.57 (45.19)	58.21 (54.73)	45.24 (38.01)	64.76 (60.60)	54.39 (45.28)
+ SNSCL	76.85 (76.12)	74.64 (74.60)	83.21 (83.01)	76.07 (75.90)	76.24 (75.70)	70.36 (70.06)	73.46 (73.09)	69.14 (68.64)
Conf. Penalty [33]	73.22 (66.89)	68.69 (52.98)	64.74 (64.46)	48.15 (43.71)	56.32 (55.51)	43.64 (39.54)	62.75 (61.10)	52.04 (45.13)
+ SNSCL	76.14 (75.73)	74.72 (74.49)	83.07 (83.00)	75.67 (75.38)	75.04 (74.23)	67.99 (66.85)	73.90 (73.51)	68.42 (67.86)
GCE [60]	66.96 (66.93)	61.47 (60.32)	62.77 (61.23)	47.44 (46.13)	39.54 (39.24)	32.34 (32.28)	58.74 (57.20)	49.71 (48.11)
+ SNSCL	75.99 (74.56)	71.68 (70.62)	73.78 (73.55)	58.11 (57.41)	72.67 (71.53)	60.19 (59.83)	70.83 (70.56)	61.67 (61.46)
SYM [52]	69.20 (62.13)	65.76 (46.99)	74.65 (73.21)	52.83 (51.61)	62.29 (60.51)	54.36 (45.39)	65.34 (63.60)	50.19 (50.15)
+ SNSCL	77.55 (77.24)	76.28 (76.25)	84.59 (83.54)	79.07 (78.87)	79.64 (79.09)	74.02 (73.63)	76.67 (76.06)	72.71 (72.58)
Co-teaching [9]	63.71 (58.43)	49.15 (48.92)	68.60 (67.95)	56.92 (55.95)	42.55 (40.62)	35.21 (32.16)	57.84 (55.98)	46.57 (46.22)
+ SNSCL	74.18 (73.09)	60.71 (58.84)	78.94 (78.13)	75.98 (75.06)	74.61 (74.19)	65.47 (63.81)	69.77 (69.34)	60.59 (58.94)
JoCoR [53]	66.94 (60.81)	49.62 (48.62)	69.99 (68.25)	57.95 (56.71)	61.37 (59.16)	52.11 (49.93)	58.79 (57.74)	52.64 (49.35)
+ SNSCL	75.79 (74.99)	63.42 (62.84)	79.67 (78.77)	76.80 (76.21)	75.88 (75.16)	71.65 (70.67)	71.86 (70.90)	64.43 (63.81)
MW-Net [38]	71.99 (69.20)	68.14 (65.17)	74.01 (73.88)	58.30 (55.81)	64.97 (61.84)	57.61 (55.90)	67.44 (65.20)	58.49 (54.81)
+ SNSCL	77.49 (77.08)	74.92 (74.38)	85.96 (85.37)	77.76 (77.13)	80.08 (78.94)	73.55 (73.18)	76.94 (76.24)	69.51 (68.83)
MLC [61]	74.08 (70.51)	69.44 (66.28)	76.02 (71.24)	59.44 (55.76)	63.81 (60.33)	58.11 (54.86)	69.44 (68.19)	60.27 (58.49)
+ SNSCL	78.92 (78.56)	76.49 (78.96)	85.92 (84.91)	78.49 (77.80)	79.19 (78.40)	75.21 (74.67)	77.58 (76.68)	71.54 (70.86)
DivideMix [18]	79.22 (77.86)	77.93 (76.28)	78.35 (77.99)	62.54 (62.50)	80.62 (80.50)	66.76 (66.13)	75.11 (74.54)	67.35 (66.96)
+ SNSCL	81.40 (81.16)	79.12 (78.91)	86.29 (85.94)	80.09 (79.51)	82.31 (82.03)	76.22 (75.67)	78.36 (78.04)	73.66 (73.28)
Avg. ↑	**5.88 (9.34)**	**7.76 (15.83)**	**12.44 (13.29)**	**20.82 (23.06)**	**18.60 (19.86)**	**21.41 (24.49)**	**9.87 (11.25)**	**12.22 (16.46)**

Analysis



(a) SNSCL imporves the performance of three loss functions
(b) SNSCL mitigates overfitting under extreme noise ratios

	Stanford Dogs	CUB-200-2011
CE	69.20 (50.45)	54.14 (45.85)
CE + SCL	68.49 (54.77)	53.30 (45.92)
CE + SNSCL	**75.27 (75.00)**	**68.83 (68.67)**
w/o Weight corr.	70.91±0.6	62.71±0.5
w/o Weight update	73.45±0.3	65.29±0.4
w/o Stoc. module	74.11±0.3	67.44±0.3
DivideMix	77.93 (76.28)	67.35 (66.96)
DivideMix + SCL	78.20 (77.89)	70.28 (70.02)
DivideMix + SNSCL	**79.12 (78.91)**	**73.66 (73.28)**
w/o Weight corr.	78.30±0.2	70.41±0.3
w/o Weight update	78.52±0.1	72.59±0.2
w/o Stoc. module	78.85±0.1	73.06±0.1

Effectiveness of each component has been verified

Summary

- We consider a hardly studied LNL task, dubbed LNL-FG and conduct empirical investigation to show that some existing methods in LNL cannot achieve satisfying performance for LNL-FG.
- We design a novel framework dubbed stochastic noise-tolerated supervised contrastive learning (SNSCL), which alters the noisy labels for anchor samples and selectively updates the momentum queue, avoiding the effects of noisy labels on SCL.
- We design a stochastic module to avoid manually-defined augmentation, improving the performance of SNSCL on representation learning.
- Our proposed SNSCL is generally applicable to prevailing LNL methods and significantly improves their performance on LNL-FG.

Thank you！