TONGJI UNIVERSITY

JUNE 18-22, 2023 CVPR VANCOUVER, CANADA

Learning with Noisy labels via Self-supervised Adversarial Noisy Masking

Yuanpeng Tu[1] , Boshen Zhang[2], Yuxi Li[2], Liang Liu[2], Jian Li[2],
Jiangning Zhang[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, caseywang@tencent.com

Motivation

(a) Clean-trained Mis-predicted (b) Noisy-trained Mis-predicted (c) Noisy-trained Correctly-predicted

Figure 1. Activation maps of mis-predicted (a-b) and correctly-predicted (c) samples when training PreAct ResNet-18 with clean (i.e., clean-trained) and noisy (i.e., noisy-trained) data on CIFAR-10 (1st and 2nd row) and Clothing1M [38] (3rd row).

Difference in Activation Maps

Noise-unaware Mask VS Noise-aware Mask

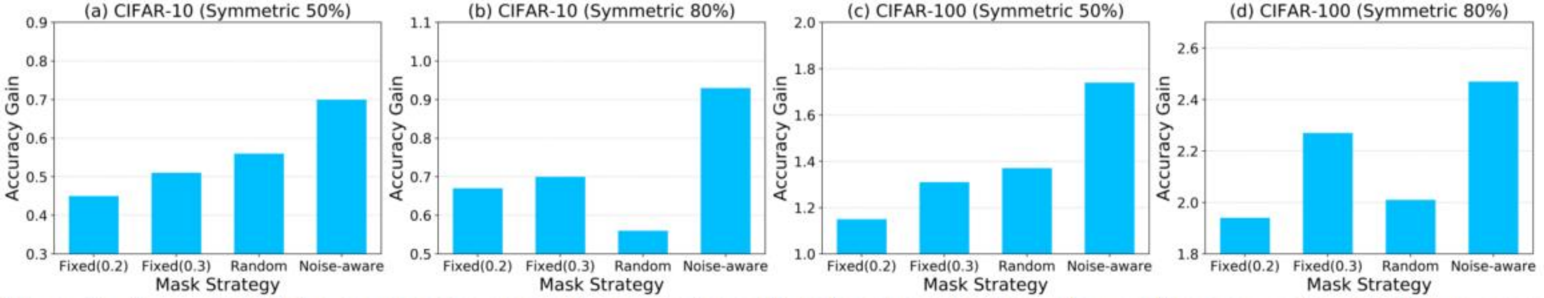
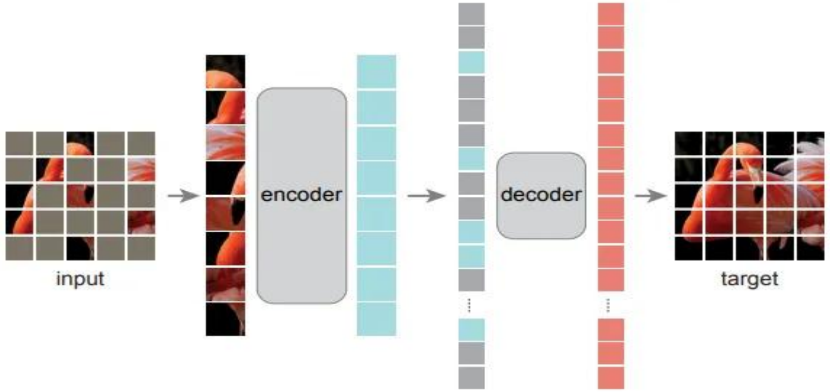


Figure 2. A experiment for masking the max-activated region with different mask ratios. The performance gains of different mask strategies under 50% and 80% symmetric noise of CIFAR-10/100 [14] are reported, where DivideMix [15] is adopted as the baseline. "Fixed(0.2/0.3)" denotes masking all the images with the same mask ratio of 0.2/0.3. "Random" represents masking images with a random mask ratio between 0.2 and 0.4. "Noise-aware" is masking noisy samples with a mask ratio of 0.3 while the ratio for clean ones is 0.2.

Noise-aware Mask induces Better Results!

Contribution

- We propose a novel self-supervised adversarial noisy masking method named SANM to explicitly impose regularization for LNL problem, preventing the model from overfitting to less informative regions from noisy data;
- A label quality guided masking strategy is proposed to differently adjust the process for clean and noisy samples according to the label quality estimation. This strategy modulates the image label and the ratio of image masking simultaneously;
- A self-supervised mask reconstruction auxiliary task is designed to reconstruct the original images based on the features of masked ones, which aims at enhancing generalization by providing noise-free supervision signals.



[1] Masked Autoencoders Are Scalable Vision Learners. CVPR2022

Framework

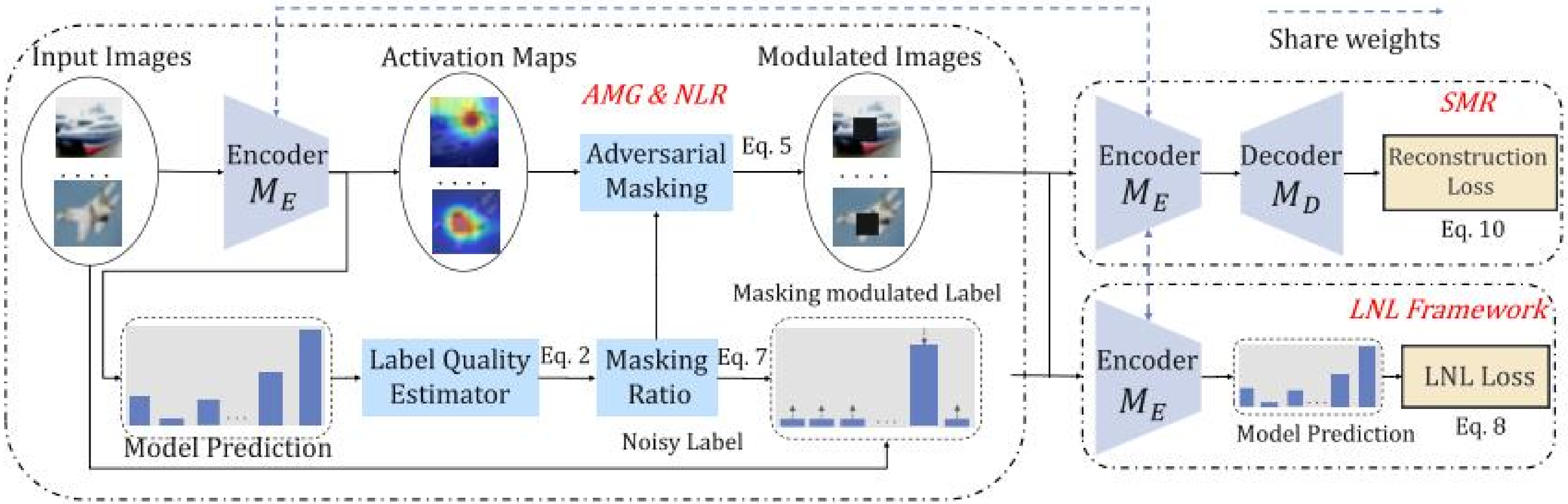


Figure 3. The technical workflow of the proposed SANM. Three components are included in SANM: AMG (Adversarial Masking Generation), NLR (Noisy Label Regularization), SMR (Self-supervised Masking Reconstruction). In AMG and NLR, firstly feed the images to an encoder and generate activation maps. And a label quality guided adversarial masking strategy is proposed to modulate the images and noisy labels simultaneously. Further, an auxiliary decode branch is designed in SMR to reconstruct input images from the features of masked images. Finally, the generated modulated images and labels of SANM together with the reconstruction loss can be directly adopted for the training of existing LNL framework.

AMR: Adversarial Mask Generation ***NLR: Noisy Label Regularization*** ***SMR: Self-supervised Masking Reconstruction***

Adversarial Noisy Masking

$$A_i = \mathrm{CAM}(F_i,\ \mathrm{argmax}(\tilde{y_i}))$$

$$\begin{cases} h_i^{up} = \max\left(h_i^{\max} - \sqrt{\frac{(H_x \times W_x) \times r_i \times \delta_i}{4}}, 0\right) \\ h_i^{\mathrm{dn}} = \min\left(h_i^{\max} + \sqrt{\frac{(H_x \times W_x) \times r_i \times \delta_i}{4}}, H_x\right) \\ w_i^{\mathrm{lt}} = \max\left(w_i^{\max} - \sqrt{\frac{(H_x \times W_x) \times r_i}{4\delta_i}}, 0\right) \\ w_i^{\mathrm{rt}} = \min\left(w_i^{\max} + \sqrt{\frac{(H_x \times W_x) \times r_i}{4\delta_i}}, W_x\right) \end{cases}$$

Masked Ratio

$$r_i = \mu \times (1 - G_i)$$

Aspect Ratio

$$\delta_i \sim \mathrm{Uniform}(\delta,\ \frac{1}{\delta})$$

$$x_i^m(m,n) = \begin{cases} U(0,1), & \text{if } m \in [h_i^{up}, h_i^{dn}], n \in [w_i^{lt}, w_i^{rt}] \\ x_i(m,n), & \text{otherwise.} \end{cases} \quad (5)$$

Noisy Label Regularization

$$\mathbf{f_i}^m = M_E(x_i^m; \theta_E)$$

$$y_i^r(j) = \begin{cases} y_i(j) - r_i + r_i/c, & \text{if } j = \mathrm{argmax}\,(y_i(j)) \\ y_i(j) + r_i/c, & \text{otherwise} \end{cases},$$

Label Regularization

$$\mathcal{L}_{\mathrm{c}} = \mathcal{L}_{\mathrm{ce}}(\widetilde{y_i^m}, y_i^r)$$

Self-supervised Masking Reconstruction

$$x_i{}^r = M_D(\mathbf{f_i}^m; \theta_D)$$

$$\mathcal{L}_{\mathrm{r}} = \|x_i^r - x_i\|^2$$

Overall Objective

$$\mathcal{L}_{\mathrm{train}} = \mathcal{L}_{\mathrm{c}} + \beta \mathcal{L}_{\mathrm{r}}$$

Algorithm

Algorithm 1 The proposed SANM framework

Input: Noisy training set D, encoder model $M_E(\cdot;\theta_E)$, decoder model $M_D(\cdot;\theta_D)$, batch size b, max iterations m, basic mask ratio μ.

Procedure:

1: **for** $i = 1$ to m **do**

2: $\{x_i, y_i\}_{i=1}^{b} \leftarrow$ SampleMiniBatch(D, b).

3: Feed $\{x_i\}_{i=1}^{b}$ into M_E and generate feature maps $\{F_i\}_{i=1}^{b}$ and predictions $\{\widetilde{y_i}\}_{i=1}^{b}$.

4: Generate activation maps $\{A_i\}_{i=1}^{b}$ by Eq. (1).

5: Calculate mask ratios $\{r_i\}_{i=1}^{b}$ and adversarial masked images $\{x_i^m\}_{i=1}^{b}$ by Eq. (2-5).

6: Feed $\{x_i^m\}_{i=1}^{b}$ into M_E and generate predictions $\{\widetilde{y_i^m}\}_{i=1}^{b}$ and features $\{\mathbf{f}_i^m\}_{i=1}^{b}$ by Eq. (6).

7: Calculate the regularized labels $\{y_i^r\}_{i=1}^{b}$ by Eq. (7).

8: Calculate cross-entropy loss $\mathcal{L}_c$ by Eq. (8).

9: Feed $\{\mathbf{f}_i^m\}_{i=1}^{b}$ into M_D and generate reconstructed images $\{x_i^r\}_{i=1}^{b}$ by Eq. (9).

10: Calculate self-supervised reconstruction loss $\mathcal{L}_r$ by Eq. (10).

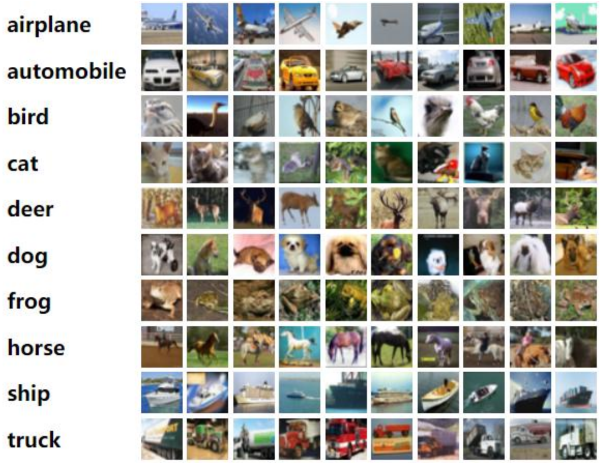
11: Update parameter θ_E, θ_D in backward process.

12: **end for**

Output: The final encoder $M_E(\cdot;\theta_E)$.

Dataset

Simulated Noisy Dataset: CIFAR-10/100



Real-world Noisy Dataset: Clothing1M

Experiment

Results on Simulated & Real-world Noisy Datasets

Table 1. Comparison with state-of-the-art methods on CIFAR-10/100 datasets with symmetric noise.

Dataset	CIFAR-10				CIFAR-100			
Method/Noise ratio	20%	50%	80%	90%	20%	50%	80%	90%
Cross-Entropy (CE)	86.8	79.4	62.9	42.7	62.0	46.7	19.9	10.1
Co-teaching+ [41]	89.5	85.7	67.4	47.9	65.6	51.8	27.9	13.7
Mixup [43]	95.6	87.1	71.6	52.2	67.8	57.3	30.8	14.6
PENCIL [40]	92.4	89.1	77.5	58.9	69.4	57.5	31.1	15.3
Meta-Learning [16]	92.9	89.3	77.4	58.7	68.5	59.2	42.4	19.5
M-correction [1]	94.0	92.0	86.8	69.1	73.9	66.1	48.2	24.3
DivideMix [15]	96.1	94.6	93.2	76.0	77.3	74.6	60.2	31.5
C2D [47]	96.3	95.2	94.4	93.5	78.6	76.4	67.7	58.7
AugDesc [22]	96.3	95.4	93.8	91.9	79.5	77.2	66.4	41.2
GCE [5]	90.0	89.3	73.9	36.5	68.1	53.3	22.1	8.9
Sel-CL+ [19]	95.5	93.9	89.2	81.9	76.5	72.4	59.6	48.8
MOIT+ [23]	94.1	91.8	81.1	74.7	75.9	70.6	47.6	41.8
SANM(DivideMix)	96.4	95.8	94.6	92.3	81.2	78.2	68.7	43.5
SANM(C2D)	**96.6**	**96.4**	**95.7**	**95.1**	**81.9**	**79.3**	**71.6**	**61.9**

Table 2. Asymmetric noise on CIFAR-10.

Method	Noisy ratio	
	20%	40%
Joint-Optim [32]	92.8	91.7
PENCIL [40]	92.4	91.2
F-correction [24]	89.9	-
Distilling [46]	92.7	90.2
Meta-Learning [16]	-	88.6
M-correction [1]	-	86.3
Iterative-CV [3]	-	88.0
DivideMix [15]	93.4	93.4
REED [44]	95.0	92.3
C2D [47]	93.8	93.4
Sel-CL+ [19]	95.2	93.4
GCE [5]	87.3	78.1
RRL [17]	-	92.4
SANM(DivideMix)	**95.4**	**94.8**

Table 3. Testing accuracy on Clothing-1M.

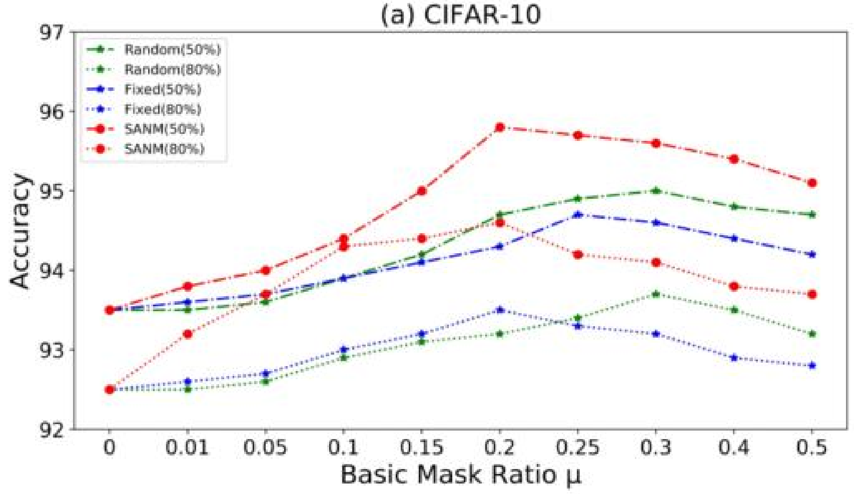
Method	Acc
CrossEntropy	69.21
F-correction [24]	69.84
M-correction [1]	71.00
Joint-Optim [32]	72.16
Meta-Cleaner [45]	72.50
Meta-Learning [16]	73.47
PENCIL [40]	73.49
Self-Learning [10]	74.45
DivideMix [15]	74.76
Nested [4]	74.90
AugDesc [22]	75.11
RRL [17]	74.90
GCE [5]	73.30
C2D [47]	74.30
SANM(DivideMix)	**75.63**

Experiment

Component Analysis

Table 5. Ablation study for the effectiveness of each key component. AMG: adversarial noisy masking generation, NLR: noisy label regularization, SMR: self-supervised masking reconstruction.

AMG	NLR	SMR		CIFAR-10 20%	CIFAR-10 50%	CIFAR-10 80%	CIFAR-10 90%	CIFAR-100 20%	CIFAR-100 50%	CIFAR-100 80%	CIFAR-100 90%
✗	✗	✗	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
✓	✗	✗	Best	96.3	95.3	94.0	91.4	80.2	77.3	68.0	42.7
			Last	96.2	95.1	93.6	90.8	79.7	77.0	67.5	42.5
✓	✓	✗	Best	96.4	95.5	94.2	91.6	80.5	77.5	68.3	43.0
			Last	96.3	95.4	94.0	91.5	80.2	77.1	68.2	42.7
✓	✓	✓	Best	**96.4**	**95.8**	**94.6**	**92.3**	**81.2**	**78.2**	**68.7**	**43.5**
			Last	**96.3**	**95.6**	**94.3**	**92.1**	**80.4**	**78.0**	**68.3**	**43.0**

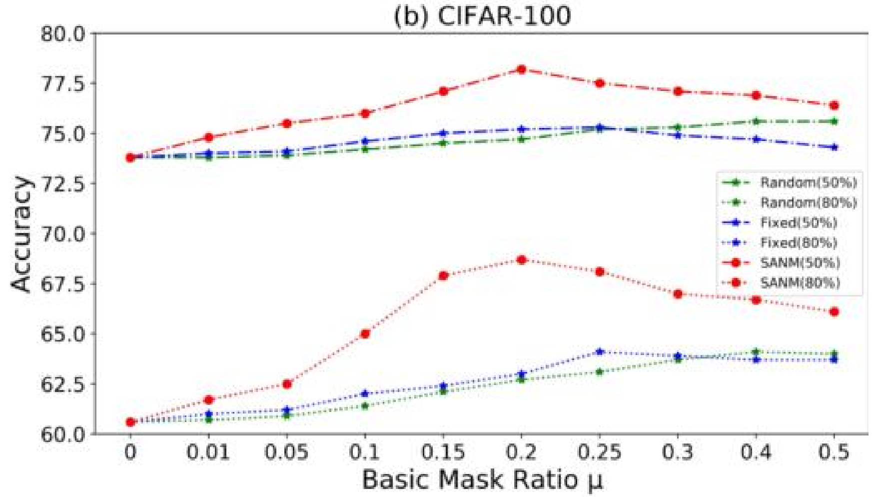


Comparison with Masks from Pre-trained Backbones

Table 4. Comparison with masks generated from pre-trained backbones on CIFAR-10/100. M: Method. P: Pretrained Backbone. S: SANM. D: DivideMix. C: C2D.

M	P	S	CIFAR-10 20%	50%	80%	90%
D	✓	✗	96.0	95.1	93.7	81.5
	✗	✓	**96.4**	**95.8**	**94.6**	**92.3**
C	✓	✗	96.3	95.4	95.0	93.9
	✗	✓	**96.6**	**96.4**	**95.7**	**95.1**

M	P	S	CIFAR-100 20%	50%	80%	90%
D	✓	✗	78.8	76.6	64.9	36.4
	✗	✓	**81.2**	**78.2**	**68.7**	**43.5**
C	✓	✗	80.0	77.1	69.2	58.4
	✗	✓	**81.9**	**79.3**	**71.6**	**61.9**



Experiment

Table 6. Comparison between the LNL methods and their SANM 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%
CE	Best	86.8	79.4	62.9	42.7	62.0	46.7	19.9	10.1
SANM(CE)	Best	**92.4**	**89.7**	**72.1**	**51.5**	**70.9**	**53.1**	**34.8**	**18.6**
Co-teaching [9]	Best	82.6	73.0	24.0	14.6	50.5	38.2	11.8	4.9
SANM(Co-teaching)	Best	**89.2**	**78.2**	**36.4**	**20.7**	**58.2**	**51.3**	**19.4**	**13.4**
CDR [37]	Best	90.4	85.0	47.2	12.3	63.3	39.5	29.2	8.0
SANM(CDR)	Best	**92.6**	**91.6**	**55.3**	**16.7**	**72.7**	**56.4**	**36.6**	**20.8**
ELR+ [21]	Best	94.6	93.8	91.1	75.2	77.5	72.4	58.2	30.8
SANM(ELR+)	Best	**96.3**	**95.7**	**94.1**	**82.9**	**79.8**	**77.3**	**65.0**	**38.7**

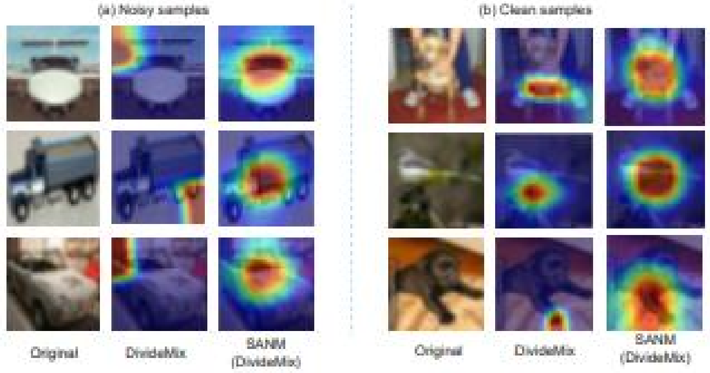


Figure 1. Activation maps for samples with noisy and clean labels between DivideMix and SANM (DivideMix).

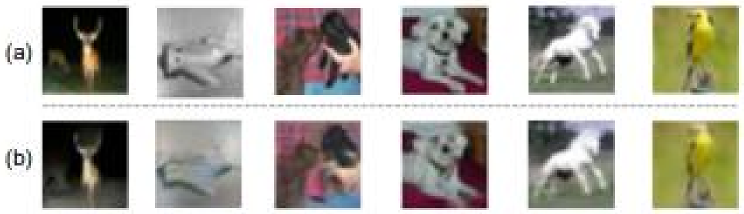


Figure 2. The reconstruction result of SANM on CIFAR-10 of SANM (DivideMix). (a) Original images. (b) The corresponding reconstruction results.

Experiment

Table 3. Comparison with state-of-the-art methods in test accuracy on Animal-10N.

Method	Test Accuracy (%)
Cross-Entropy	79.4
ActiveBias [1]	80.5
PLC [9]	83.4
Co-teaching [3]	80.2
SELFIE [6]	81.8
CREMA [8]	84.2
SSR [2]	88.5
SANM(SSR)	**89.3**

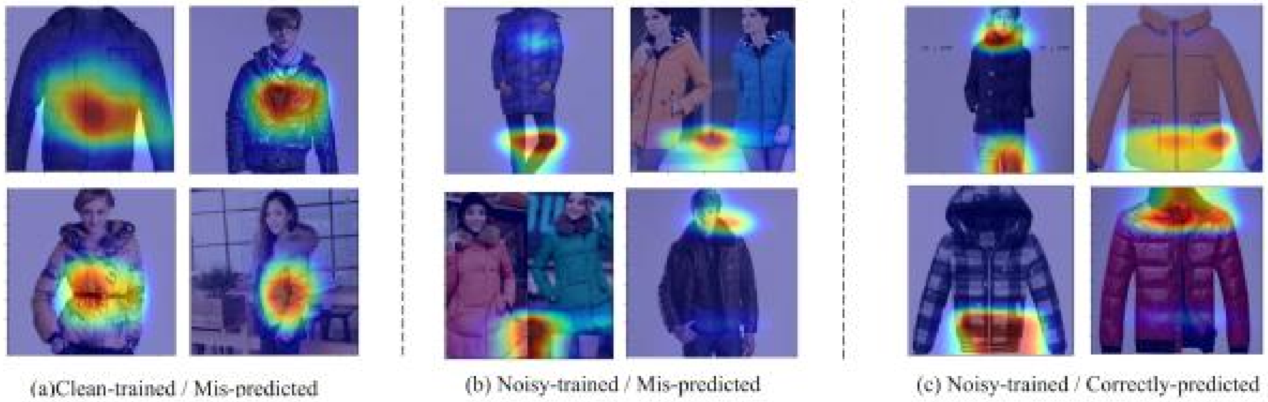


Figure 4. The activation maps of the trained base model on Clothing1M dataset.

Table 5. Comparison on CIFAR-10/100 with symmetric noise.

Dataset / Method	CIFAR-10 20%	CIFAR-10 50%	CIFAR-10 80%	CIFAR-10 90%
SANM(DivideMix)	94.41±0.11	95.79±0.09	94.62±0.16	92.28±0.13
Dataset / Method	**CIFAR-100 20%**	**CIFAR-100 50%**	**CIFAR-100 80%**	**CIFAR-100 90%**
SANM(DivideMix)	81.21±0.10	78.22±0.14	68.71±0.11	43.49±0.18

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

Dataset Method/Noise ratio		CIFAR-10 20%	50%	80%	90%	CIFAR-100 20%	50%	80%	90%
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
SANM(CE)	Best	**92.4**	**89.7**	**72.1**	**51.5**	**70.9**	**53.1**	**34.8**	**18.6**
	Last	**92.1**	**89.0**	**69.6**	**47.3**	**70.5**	**50.9**	**32.0**	**18.1**
Co-teaching [4]	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
SANM(Co-teaching)	Best	**89.2**	**78.2**	**36.4**	**20.7**	**58.2**	**51.3**	**19.4**	**13.4**
	Last	**88.6**	**76.7**	**35.2**	**18.4**	**56.9**	**50.1**	**17.9**	**12.7**
CDR [7]	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
SANM(CDR)	Best	**92.6**	**91.6**	**55.3**	**16.7**	**72.7**	**56.4**	**36.6**	**20.8**
	Last	**91.8**	**90.8**	**48.6**	**15.5**	**71.2**	**53.2**	**30.0**	**19.7**
ELR+ [5]	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
SANM(ELR+)	Best	**96.3**	**95.7**	**94.1**	**82.9**	**79.8**	**77.3**	**65.0**	**38.7**
	Last	**96.2**	**95.4**	**94.0**	**81.7**	**79.2**	**77.1**	**64.1**	**37.9**

Learning with Noisy labels via Self-supervised Adversarial Noisy Masking

Thanks for Watching!

https://github.com/yuanpengtu/SANM

Yuanpeng Tu[1] , Boshen Zhang[2], Yuxi Li[2], Liang Liu[2], Jian Li[2], Jiangning Zhang[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, caseywang@tencent.com