# SURE: SUrvey REcipes for building reliable and robust deep networks

Yuting Li, Yingyi Chen, Xuanlong Yu, Dexiong Chen†, Xi Shen†

## Motivation and contribution

**Motivation:**

Model robustness in handling complex real-world data challenges, such as long-tailed classification, learning with noisy labels and data corruptions.

**Contribution:**

1. Simple and effective approach **SURE** for building reliable and robutst deep networks.

2. SOTA performance in **failure prediction** across various datasets and model architectures.

3. Competitive results to **SOTA** specialized methods in real-world scenarios : long-tailed distribution, label noise and data corruption.

## Overview of recipes

**Total Loss:**

$$\mathcal{L}_{total} = \mathcal{L}_{ce} + \lambda_{mix}\mathcal{L}_{mix} + \lambda_{crl}\mathcal{L}_{crl}$$

**RegMixup regularization:**

$$\tilde{x}_i = mx_i + (1-m)x_i, \tilde{y}_i = my_i + (1-m)y_i$$

$$m \sim Beta(\beta, \beta), \beta \in (0, \infty)$$

$$\mathcal{L}_{mix}(\tilde{x}_i, \tilde{y}_i) = \mathcal{L}_{ce}(\tilde{x}_i, \tilde{y}_i)$$

**Correctness ranking loss:**

$$\mathcal{L}_{crl}(x_i, x_j) = \max(0, |c_i - c_j| - \text{sign}(c_i - c_j)(s_i - s_j))$$

**Cosine Similarity Classifier:**

$$s_i^k = \tau \cdot \cos(f_\theta(x_i), w^k) = \tau \cdot \frac{f_\theta(x_i)}{\|f_\theta(x_i)\|_2} \cdot \frac{w^k}{\|w^k\|_2}$$

**Sharpness-Aware Minimization:**

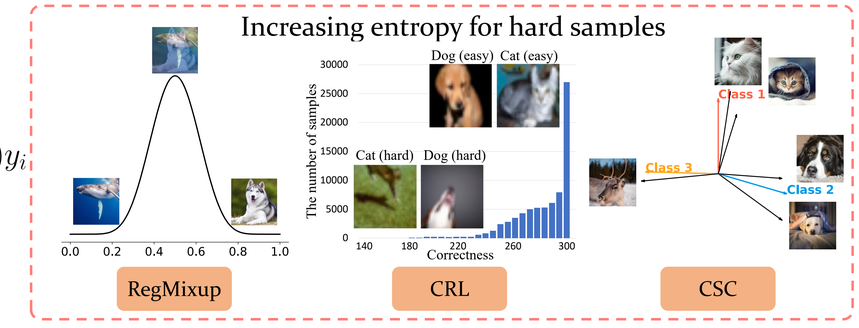
$$\min_{\theta} \max_{\|\epsilon\|_2 \le \rho} \mathcal{L}_{total}(\theta + \epsilon)$$

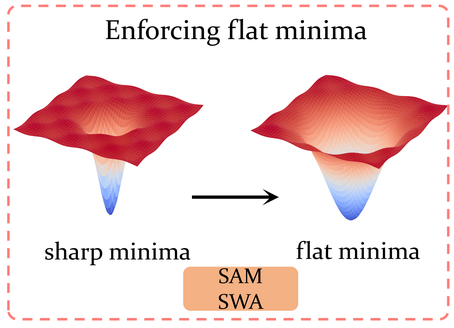
**Stochastic Weight Averaging:**

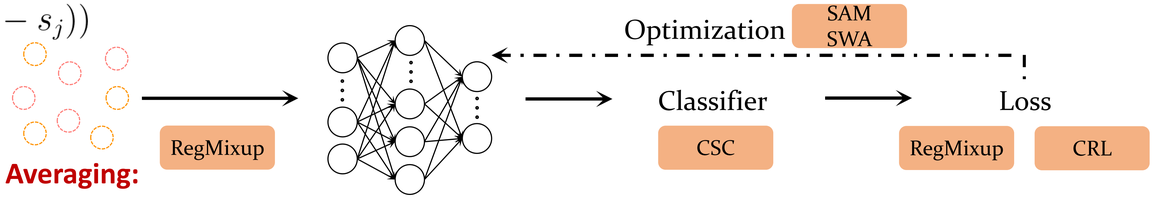
$$\theta_{\text{SWA}} = \frac{1}{T}\sum_{t=1}^{T} \theta_t$$

**SURE** contains two aspects:

**· Increasing entropy for hard samples · Enforcing flat minima during optimization.**

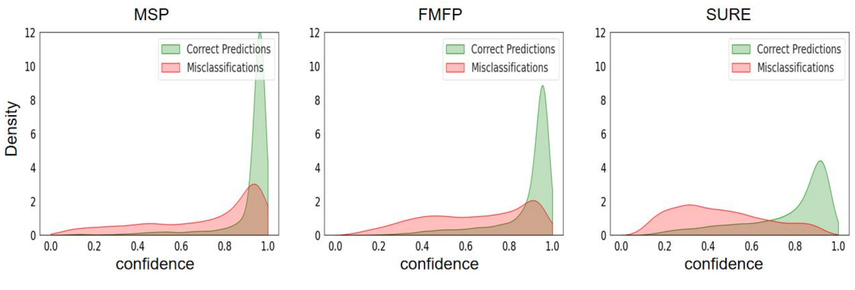


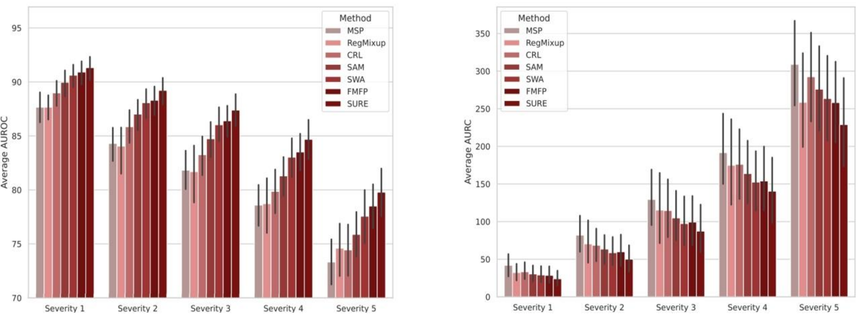

## Experiments

### Visual results of an example from CIFAR100-LT IF=10




*SURE leads to clearly better confidence separation than MSP and FMFP.*

### Failure prediction under distribution shift (CIFAR10-C)







*SURE **enhances** the failure prediction performance across a spectrum of **corruptions***

## Learning with noisy labels: Animal-10N dataset (left) and Food-101N dataset (right)

| Methods | CE [78] | SELFIE [65] | PLC [78] | NCT [6] | Dynamic Loss [37] | SSR+ [17] | Jigsaw-ViT * [7] | **SURE** |
|---|---|---|---|---|---|---|---|---|
| Acc. (%) | 79.4 | 81.8 | 83.4 | 84.1 | 86.5 | 88.5 | 89.0 | **89.0** |

| Methods | CE [78] | CleanNet [42] | MWNet [63] | SMP [27] | NRank [62] | PLC [78] | WarPI [67] | Jigsaw-ViT * [7] | **SURE** |
|---|---|---|---|---|---|---|---|---|---|
| Acc. (%) | 81.7 | 83.5 | 84.7 | 85.1 | 85.2 | 85.3 | 85.9 | 86.7 | **88.0** |

*SURE achieves **SOTA** performance on **learning with noisy label task** without any task-specific adjustments*

## Failure prediction

| Backbones | Methods | CIFAR-10 [40] Acc. ↑ | AURC ↓ | AUROC ↑ | FPR95 ↓ | CIFAR-100 [40] Acc. ↑ | AURC ↓ | AUROC ↑ | FPR95 ↓ | Tiny-ImageNet [41] Acc. ↑ | AURC ↓ | AUROC ↑ | FPR95 ↓ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **ResNet-18** [28] | MSP [31] | 94.89±0.20 | 6.78±0.33 | 92.20±0.55 | 38.73±2.89 | 75.87±0.31 | 69.44±2.11 | 87.00±0.21 | 60.73±1.16 | 63.39±0.59 | 136.50±1.08 | 85.62±0.35 | 63.99±0.64 |
| | RegMixup [59] | 95.69±0.13 | 4.74±0.27 | 92.96±0.29 | 34.26±1.98 | 77.90±0.37 | 59.23±1.65 | 87.61±0.13 | 58.65±0.43 | 66.36±0.43 | 115.08±1.98 | 86.53±0.27 | 62.54±0.43 |
| | CRL [54] | 94.85±0.10 | 5.09±0.28 | 93.64±0.48 | 35.33±1.73 | 76.42±0.21 | 62.78±0.21 | 88.07±0.17 | 59.02±0.39 | 65.50±0.03 | 117.46±0.56 | 87.01±0.13 | 61.15±0.07 |
| | SAM [19] | 95.30±0.25 | 3.97±0.33 | 94.53±0.31 | 31.13±3.62 | 76.60±0.21 | 62.97±1.02 | 87.72±0.10 | 59.35±0.87 | 64.95±0.21 | 120.04±2.11 | 87.19±0.57 | **59.98±0.55** |
| | SWA [35] | 95.38±0.09 | 4.00±0.21 | 94.40±0.50 | 35.70±1.44 | 77.65±0.19 | 55.87±0.32 | 88.55±0.25 | 60.43±1.90 | 68.09±0.19 | 102.11±0.51 | 87.27±0.15 | 60.63±1.38 |
| | FMFP [81] | 95.60±0.09 | 3.56±0.06 | 94.74±0.10 | 33.49±0.33 | 77.82±0.08 | 55.03±0.52 | 88.59±0.07 | 59.79±0.31 | 68.18±0.42 | 100.93±2.12 | 87.45±0.05 | 60.18±1.26 |
| | **SURE** | **96.14±0.16** | **2.97±0.13** | **95.08±0.04** | **28.64±0.66** | **80.49±0.18** | **45.81±0.15** | **88.73±0.24** | **58.91±0.58** | **69.55±0.10** | **93.46±0.82** | **87.67±0.12** | 60.13±0.32 |
| **VGG** [64] | MSP [31] | 93.30±0.21 | 10.41±0.33 | 90.71±0.04 | 44.66±1.81 | 72.43±0.42 | 91.40±1.95 | 85.69±0.90 | 64.41±1.66 | 59.52±0.62 | 156.45±2.51 | 86.33±0.63 | 63.79±0.95 |
| | RegMixup [59] | 94.11±0.28 | 9.89±0.81 | 89.90±0.26 | 39.93±1.58 | 73.51±0.18 | 85.98±1.05 | 86.35±0.32 | 61.70±1.83 | 63.04±0.57 | 146.72±2.59 | 85.60±0.39 | **59.00±1.27** |
| | CRL [54] | 93.42±0.09 | 7.61±0.44 | 92.88±0.56 | 39.66±2.83 | 72.63±0.27 | 80.94±0.47 | 87.37±0.28 | 61.96±0.77 | 60.20±0.36 | 146.76±1.42 | **87.42±0.28** | 59.26±1.44 |
| | SAM [19] | 94.11±0.06 | 5.97±0.08 | 93.68±0.13 | 37.21±2.92 | 73.33±0.36 | 77.44±0.75 | 87.42±0.33 | 63.19±0.58 | 61.24±0.07 | 142.54±1.04 | 86.82±0.25 | 62.93±1.12 |
| | SWA [35] | 93.76±0.25 | 6.64±0.24 | 93.43±0.16 | 40.44±1.27 | 73.98±0.16 | 74.23±0.58 | 87.30±0.14 | 62.89±1.80 | 62.48±0.19 | 137.01±0.71 | 86.29±0.16 | 62.15±1.64 |
| | FMFP [81] | 94.26±0.23 | 5.89±0.16 | 93.46±0.26 | 40.67±3.14 | 74.77±0.31 | 70.07±1.26 | 87.58±0.19 | 60.98±1.16 | 62.95±0.16 | 134.04±1.42 | 86.36±0.12 | 61.71±1.08 |
| | **SURE** | **95.00±0.11** | **4.98±0.24** | **93.79±0.62** | **35.92±2.95** | **76.51±0.07** | **65.25±0.17** | **87.59±0.07** | **60.27±0.60** | **63.75±0.11** | **131.40±0.28** | 86.12±0.19 | 63.04±1.05 |
| **DenseNet** [34] | MSP [31] | 94.72±0.23 | 5.94±0.23 | 93.00±0.45 | 37.00±0.31 | 75.14±0.07 | 74.68±0.32 | 86.22±0.22 | 62.79±0.80 | 57.90±0.25 | 180.08±2.52 | 83.65±0.29 | 68.61±0.37 |
| | RegMixup [59] | 95.13±0.22 | 6.03±0.50 | 92.20±0.80 | 38.63±1.63 | 77.29±0.16 | 63.96±1.15 | 86.57±0.07 | 63.76±1.10 | 61.96±0.09 | 147.22±1.57 | **84.91±0.17** | 65.92±0.40 |
| | CRL [54] | 94.79±0.02 | 5.58±0.42 | 93.22±0.61 | 37.34±2.73 | 76.09±0.06 | 65.96±0.62 | 87.41±0.11 | 60.67±0.72 | 58.80±0.56 | 169.44±3.74 | 84.49±0.04 | 66.05±0.60 |
| | SAM [19] | 95.31±0.10 | 4.25±0.17 | 94.15±0.46 | 33.33±1.27 | 78.17±0.26 | 57.20±0.73 | 86.99±0.23 | 61.42±0.74 | 60.49±0.31 | 158.94±3.86 | 84.39±0.57 | 66.51±1.85 |
| | SWA [35] | 94.86±0.09 | 4.65±0.18 | 94.27±0.27 | 35.78±4.61 | 78.17±0.26 | 57.20±0.73 | 87.23±0.22 | 63.33±0.63 | 60.74±0.46 | 159.68±3.12 | 83.83±0.07 | 68.03±0.75 |
| | FMFP [81] | 95.07±0.15 | 4.11±0.19 | 94.74±0.06 | 34.67±0.48 | 78.33±0.40 | 54.88±1.62 | 87.92±0.46 | 60.52±1.12 | 61.18±0.72 | 154.98±3.72 | 84.29±0.26 | 66.66±1.21 |
| | OpenMix [82]§ | 95.51±0.23 | 4.68±0.72 | 93.57±0.81 | 33.57±3.70 | 78.97±0.31 | 53.83±0.93 | 87.45±0.18 | 62.22±1.15 | - | - | - | - |
| | **SURE** | **95.57±0.06** | **3.51±0.09** | **94.91±0.25** | **29.52±0.56** | **80.02±0.13** | **46.69±0.59** | **88.78±0.26** | **58.37±0.39** | **62.61±0.18** | **142.59±2.16** | 84.31±0.42 | **65.39±2.12** |
| **WRNet** [76] | MSP [31] | 95.71±0.17 | 5.90±0.89 | 92.19±0.82 | 35.95±3.75 | 79.15±0.19 | 53.02±0.89 | 88.21±0.06 | 59.46±1.23 | 67.52±0.18 | 107.97±0.80 | 86.78±0.20 | 61.68±0.99 |
| | RegMixup [59] | 97.03±0.04 | 3.47±0.26 | 93.10±0.56 | 26.16±1.17 | 82.14±0.47 | 47.01±2.12 | 87.70±0.17 | 55.24±1.19 | 69.63±0.09 | 95.96±0.21 | 87.38±0.21 | 59.09±0.75 |
| | CRL [54] | 95.87±0.08 | 3.85±0.20 | 94.10±0.06 | 32.73±1.22 | 80.10±0.28 | 47.99±1.08 | 88.43±0.34 | 59.44±1.45 | 69.00±0.22 | 97.46±0.90 | 87.42±0.23 | 61.02±1.71 |
| | SAM [19] | 96.47±0.11 | 2.91±0.38 | 94.79±0.29 | 28.05±1.56 | 80.67±0.31 | 44.93±0.87 | 89.01±0.31 | 56.60±1.30 | 69.86±0.37 | 93.66±2.03 | 87.49±0.30 | 60.44±1.19 |
| | SWA [35] | 94.86±0.09 | 4.65±0.18 | 94.27±0.27 | 35.78±4.61 | 81.31±0.33 | 41.15±0.89 | 89.39±0.16 | 57.57±1.97 | 71.27±0.16 | 84.97±0.12 | 87.71±0.26 | 60.00±2.42 |
| | FMFP [81] | 96.47±0.12 | 2.33±0.08 | 95.73±0.01 | 26.68±2.62 | 81.66±0.12 | 39.60±0.15 | 89.51±0.10 | 56.41±1.44 | 71.62±0.04 | 83.04±0.16 | 87.78±0.03 | 60.09±0.83 |
| | OpenMix [82]§ | **97.16±0.10** | 2.32±0.15 | 94.81±0.34 | 22.08±1.86 | 82.63±0.06 | 39.61±0.54 | 89.06±0.11 | 55.00±1.29 | - | - | - | - |
| | **SURE** | 97.02±0.20 | **1.79±0.16** | **96.18±0.01** | **19.53±1.23** | **83.71±0.10** | **32.10±0.28** | **90.33±0.18** | **54.34±0.29** | **73.34±0.36** | **74.11±0.97** | **88.23±0.31** | **58.17±1.50** |
| **DeiT-B** * [70] | MSP [31] | 98.28±0.08 | 0.97±0.02 | 95.76±0.28 | 20.47±5.38 | 89.71±0.03 | 17.66±0.56 | 90.40±0.25 | 50.99±0.61 | - | - | - | - |
| | RegMixup [59] | 98.90±0.04 | 0.89±0.05 | 94.30±0.25 | 24.98±3.87 | 90.79±0.11 | 15.38±0.51 | 90.34±0.33 | 52.01±1.76 | - | - | - | - |
| | CRL [54] | 98.27±0.04 | 0.99±0.11 | 95.85±0.44 | 19.65±2.51 | 89.74±0.16 | 17.61±0.71 | 90.30±0.18 | 51.58±0.23 | - | - | - | - |
| | SAM [19] | 98.62±0.10 | 0.58±0.09 | 96.89±0.34 | **15.74±1.71** | 90.43±0.17 | 15.29±0.19 | 90.75±0.15 | 50.02±1.52 | - | - | - | - |
| | SWA [35] | 98.44±0.07 | 0.82±0.03 | 96.11±0.20 | 17.78±3.23 | 90.17±0.34 | 15.37±0.44 | 90.86±0.38 | 50.64±3.37 | - | - | - | - |
| | FMFP [81] | 98.76±0.02 | **0.46±0.02** | **97.15±0.16** | 16.17±0.55 | 90.53±0.13 | 14.30±0.18 | **91.15±0.32** | 51.90±1.50 | - | - | - | - |
| | **SURE** | **98.92±0.07** | 0.86±0.08 | 94.37±0.69 | 27.52±3.11 | **91.18±0.01** | **13.79±0.29** | 90.85±0.05 | **48.81±0.39** | - | - | - | - |

§ reports the results given by models training on extra outliers and all the training data on CIFAR10 [40] CIFAR100 [40]

* reports the results given by finetuning ImageNet [14] pre-trained DeiT-B [70] for 50 epochs

*SURE **consistently** outperforms other methods across **various backbones** and **all evaluated metrics***

## Long-tailed classification

| Methods | CIFAR10-LT [12] IF=100 | IF=50 | IF=10 | CIFAR100-LT [12] IF=100 | IF=50 | IF=10 |
|---|---|---|---|---|---|---|
| CE | 70.40 | 74.80 | 86.40 | 38.30 | 43.90 | 55.70 |
| Mixup [77] | 73.06 | 77.82 | 87.1 | 39.54 | 54.99 | 58.02 |
| CB-Focal [12] | 74.57 | 79.27 | 87.10 | 39.60 | 45.17 | 57.99 |
| LDAM-DRW [4] | 77.03 | 81.03 | 88.16 | 42.04 | 46.62 | 58.71 |
| SSP [73] | 77.83 | 82.13 | 88.53 | 43.43 | 47.11 | 58.91 |
| BBN [80] | 79.82 | 81.18 | 88.32 | 42.56 | 47.02 | 59.12 |
| Casual model [69] | 80.60 | 83.60 | 88.50 | 44.10 | 50.30 | 59.60 |
| MetaSAug-LDAM [45] | 80.66 | 84.34 | 89.68 | 48.01 | 52.27 | 61.28 |
| Hybrid-SC [71] | 81.40 | 85.36 | 91.12 | 46.72 | 51.87 | 63.05 |
| ResLT [11] | 82.40 | 85.17 | 89.70 | 48.21 | 52.71 | 62.01 |
| Dynamic Loss [37] | 82.95 | 88.30 | 91.24 | 50.14 | 54.51 | 63.99 |
| BCL [83] | 84.32 | 87.24 | 91.12 | 51.93 | 56.59 | 64.87 |
| GLMC [16] | **87.75** | 90.18 | 94.04 | 55.88 | 61.08 | 70.74 |
| **SURE** | 83.28 | 87.72 | 93.73 | 51.60 | 58.57 | 71.13 |
| GLMC + MaxNorm [1] | 87.57 | **90.22** | 94.03 | 57.11 | 62.32 | 72.33 |
| **SURE + re-weighting** | 86.93 | **90.22** | **94.96** | **57.34** | **63.13** | **73.24** |

*SURE achieves **SOTA** performance under **long-tailed distribution***

**Paper:**

**Code:**

Let's Start!