南京航空航天大學
NANJING UNIVERSITY OF AERONAUTICS AND ASTRONAUTICS

# UNICON Combating Label Noise Through Uniform Selection and Contrastive Learning

Nazmul Karim[†] Mamshad Nayeem Rizve[‡] Nazanin Rahnavard[†] Ajmal Mian[§] Mubarak Shah[‡]

[†]Department of Electrical and Computer Engineering, UCF, USA

[‡]Center for Research in Computer Vision, UCF, USA

[§]Department of Computer Science and Software Engineering, UWA, Australia

{nazmul.karim18, nayeemrizve}@knights.ucf.edu, nazanin.rahnavard@ucf.edu

ajmal.mian@uwa.edu.au, shah@crcv.ucf.edu

CVPR 2022

# Background



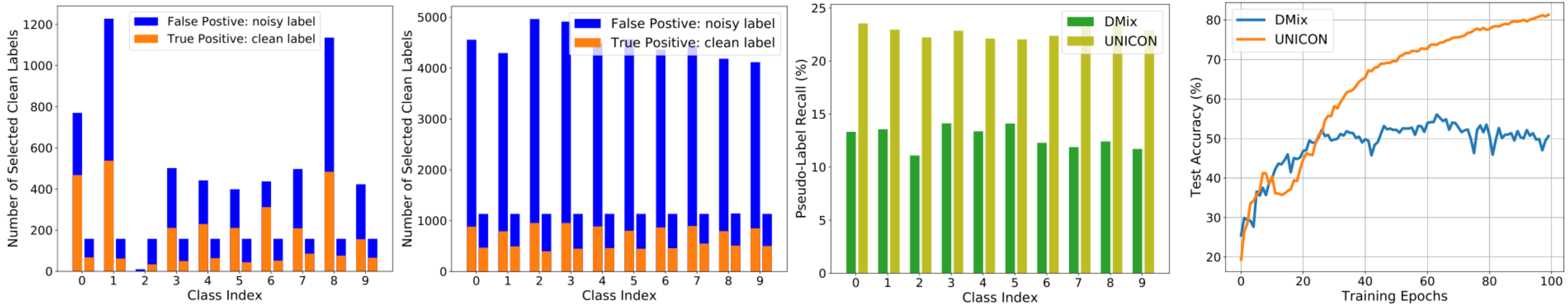
A case of uniform and non-uniform selection for CIFAR10 under 90% noise rate.

(a) Class distribution in $D_{clean}$ after warmup(10 epochs of training). For each class index, the left and right bars indicate non-uniform (DMix) and uniform selection (UNICON), respectively.

(b) Class distribution after 100 epochs. UNICON selects clean samples with higher precision.

(c) Pseudo-label recall (%) after 100 epochs of training. Uniform selection criteria along with contrastive feature learning helps generating higher quality pseudo-labels with better recall.

(d) This in turn boosts the test accuracy significantly.

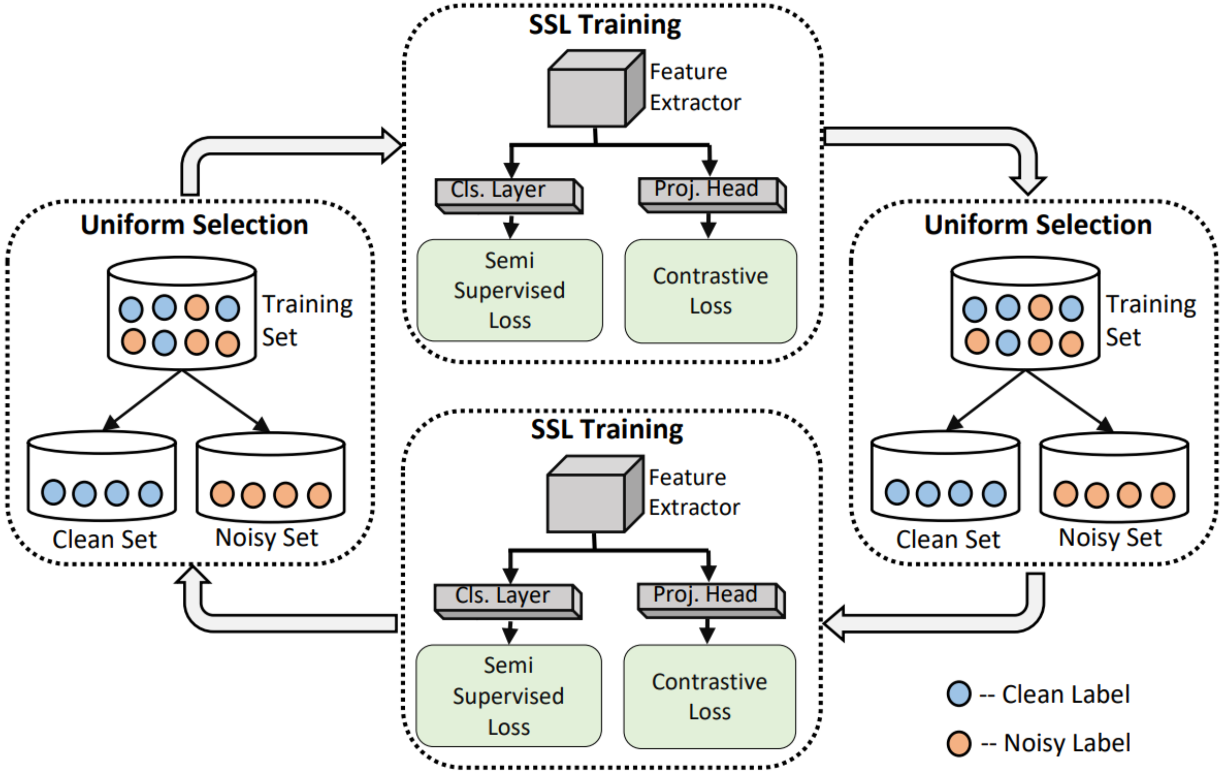



(a) Selected clean set after warmup. (b) Selected clean set after 100 epochs. (c) Pseudo-label recall after 100 epochs. (d) Test accuracy (%).

# Methods

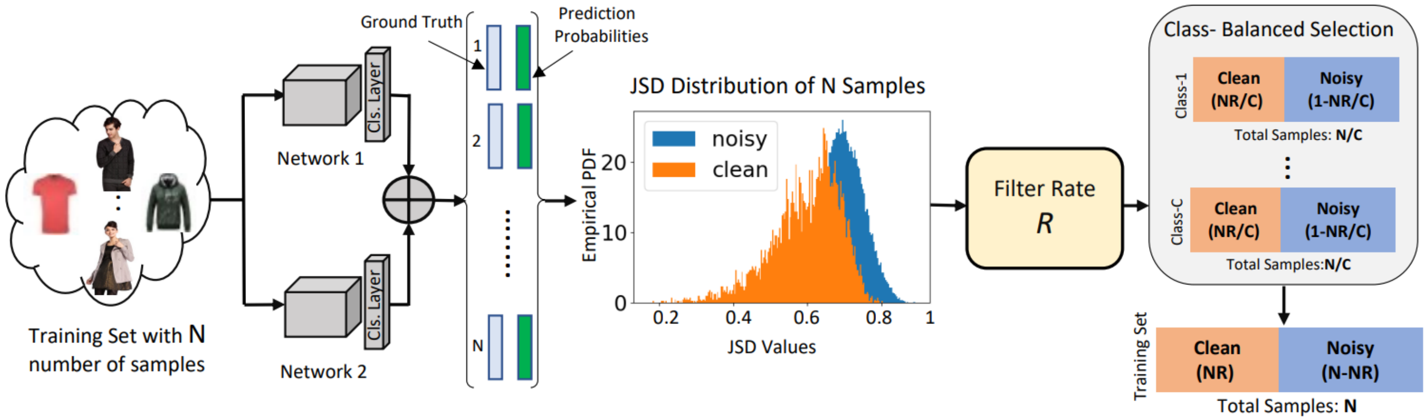


UNICON training overview: At each iteration, we employ a uniform selection technique to partition the training set into clean and noisy sets. Upon separation, we perform SSL-training with an additional contrastive loss function. The uniform selection and subsequent SSL-training is repeated until convergence.




| Noise Rate (%) | 90% | 92% | 95% | 98% |
|---|---|---|---|---|
| DMix [25] | 76.08 | 57.62 | 51.28 | 17.18 |
| UNICON (Ours) | **90.81** | **87.61** | **80.82** | **50.63** |

# Methods




a) Uniform Clean Sample Selection

Kullback-Leibler divergence

$$d_i = \mathrm{JSD}(\mathbf{y}_i, \mathbf{p}_i) = \frac{1}{2}\boxed{\mathrm{KLD}}(\mathbf{y}_i || \frac{\mathbf{y}_i + \mathbf{p}_i}{2}) + \frac{1}{2}\mathrm{KLD}(\mathbf{p}_i || \frac{\mathbf{y_i} + \mathbf{p}_i}{2})$$

filter coefficient　　adjustment threshold

$$d_{cutoff} = \begin{cases} d_{avg} - (d_{avg} - d_{min})\boxed{/\tau,} & \text{if } d_{avg} \geq \boxed{d_\mu} \\ d_{avg}, & \text{otherwise} \end{cases}$$

**Algorithm 1:** Uniform Clean Sample Selection

**Input:** training set $\mathbb{D} = (\mathcal{X}, \mathcal{Y})$, number of samples $N$, number of classes $C$

**for** $i = 1$ **to** $N$ **do**

 $\mathbf{p}_i = (\hat{\mathbf{y}}_i^{(1)} + \hat{\mathbf{y}}_i^{(2)})/2$

 $d_i = JSD(\mathbf{p}_i, \mathbf{y}_i)$ (see Eq. (2))

Determine the cutoff distance, $d_{cutoff}$ using Eq. (3)

$\mathbf{d}_R \leftarrow \{d_i < d_{cutoff} : i \in (1, \ldots, N)\}$

Determine filter rate, $\boxed{R = |\mathbf{d}_R|/N}$

the percentage of samples that have JSDs lower than $d_{cutoff}$

$\mathbb{D}_{clean} = \{\}$

`// Uniform Selection`

**for** $j = 1$ **to** $C$ **do**

 $\mathbf{d}_{filtered}^{(j)} \leftarrow$ Lowest $R$ portion of $\mathbf{d}^{(j)}$

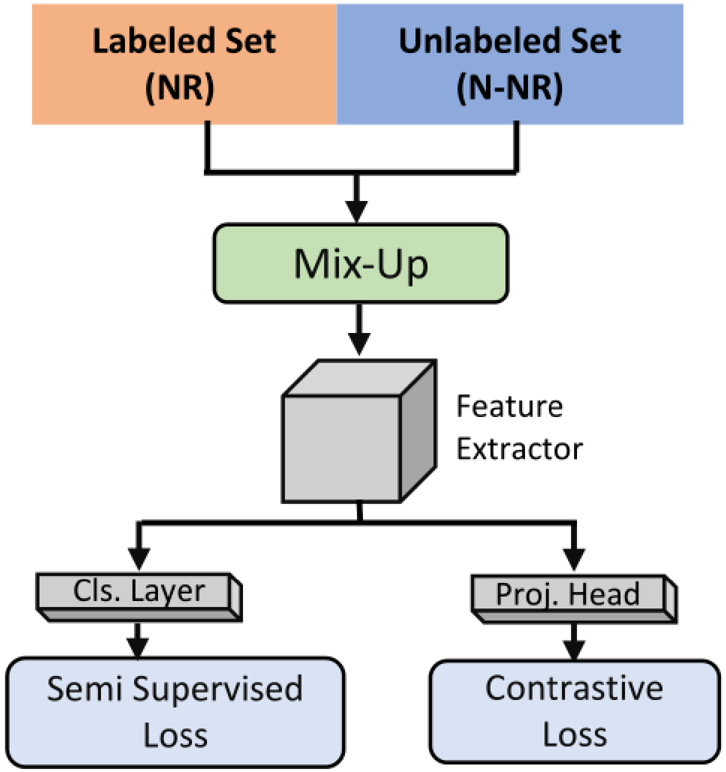
 $\mathbb{D}_{clean}^{(j)} \leftarrow \{(\mathbf{x}_t^{(j)}, \mathbf{y}_t^{(j)}) : \forall \ d_t^{(j)} \in \mathbf{d}_{filtered}^{(j)}\}$

 $\mathbb{D}_{clean} \leftarrow \mathbb{D}_{clean} \cup \mathbb{D}_{clean}^{(j)}$

$\mathbb{D}_{noisy} \leftarrow \mathbb{D} \setminus \mathbb{D}_{clean}$

**Output:** $\mathbb{D}_{noisy}, \mathbb{D}_{clean}$

# Methods




## Contrastive loss:

$$\ell_{i,j} = -\log \frac{\exp(\mathrm{sim}(\mathbf{z}_i, \mathbf{z}_j)/\kappa)}{\sum_{b=1}^{2B} \mathbb{1}_{b \neq i} \exp(\mathrm{sim}(\mathbf{z}_i, \mathbf{z}_b)/\kappa)}$$

$$\mathcal{L}_{\mathcal{C}} = \frac{1}{2B} \sum_{b=1}^{2B} [\ell_{2b-1,2b} + \ell_{2b,2b-1}]$$

## Mix up:

$$\tilde{x} = \lambda x_i + (1-\lambda) x_j, \qquad \text{where } x_i, x_j \text{ are raw input vectors}$$

$$\tilde{y} = \lambda y_i + (1-\lambda) y_j, \qquad \text{where } y_i, y_j \text{ are one-hot label encodings}$$

## Fixmatch:

$$\ell_u = \frac{1}{\mu B} \sum_{b=1}^{\mu B} \mathbb{1}(\max(q_b) \geq \tau)\, \mathrm{H}(\hat{q}_b, p_{\mathrm{m}}(y \mid \mathcal{A}(u_b)))$$

$$\ell_s = \frac{1}{B} \sum_{b=1}^{B} \mathrm{H}(p_b, p_{\mathrm{m}}(y \mid \alpha(x_b)))$$

## Total loss:

$$\mathcal{L}_{tot} = \underbrace{\boxed{\mathcal{L}_{semi}}}_{\text{Fixmatch}} + \lambda_{\mathcal{C}} \mathcal{L}_{\mathcal{C}}$$

# Experiments



Symmetric Noise:

| Method | CIFAR-10 | | | | CIFAR-100 | | | |
|---|---|---|---|---|---|---|---|---|
| | 20% | 50% | 80% | 90% | 20% | 50% | 80% | 90% |
| CE | 86.8 | 79.4 | 62.9 | 42.7 | 62.0 | 46.7 | 19.9 | 10.1 |
| LDMI [59] | 88.3 | 81.2 | 43.7 | 36.9 | 58.8 | 51.8 | 27.9 | 13.7 |
| M-Up [68] | 95.6 | 87.1 | 71.6 | 52.2 | 67.8 | 57.3 | 30.8 | 14.6 |
| PCIL [64] | 92.4 | 89.1 | 77.5 | 58.9 | 69.4 | 57.5 | 31.1 | 15.3 |
| JPL [20] | 93.5 | 90.2 | 35.7 | 23.4 | 70.9 | 67.7 | 17.8 | 12.8 |
| MOIT [39] | 94.1 | 91.1 | 75.8 | 70.1 | 75.9 | 70.1 | 51.4 | 24.5 |
| DMix [25] | **96.1** | 94.6 | 92.9 | 76.0 | 77.3 | 74.6 | 60.2 | 31.5 |
| ELR [30] | 95.8 | 94.8 | 93.3 | 78.7 | 77.6 | 73.6 | 60.8 | 33.4 |
| UNICON | 96.0 | **95.6** | **93.9** | **90.8** | **78.9** | **77.6** | **63.9** | **44.8** |

Symmetric Noise:

| Method | CIFAR-10 | | | CIFAR-100 | | |
|---|---|---|---|---|---|---|
| | 10% | 30% | 40% | 10% | 30% | 40% |
| CE | 88.8 | 81.7 | 76.1 | 68.1 | 53.3 | 44.5 |
| LDMI [59] | 91.1 | 91.2 | 84.0 | 68.1 | 54.1 | 46.2 |
| M-Up [68] | 93.3 | 83.3 | 77.7 | 72.4 | 57.6 | 48.1 |
| JPL [20] | 94.2 | 92.5 | 90.7 | 72.0 | 68.1 | 59.5 |
| PCIL [64] | 93.1 | 92.9 | 91.6 | 76.0 | 59.3 | 48.3 |
| DMix* [25] | 93.8 | 92.5 | 91.7 | 71.6 | 69.5 | 55.1 |
| ELR* [30] | **95.4** | **94.7** | 93.0 | 77.3 | 74.6 | 73.2 |
| MOIT [39] | 94.2 | 94.1 | 93.2 | 77.4 | 75.1 | 74.0 |
| UNICON | 95.3 | 94.6 | **94.1** | **78.2** | **75.6** | **74.8** |

# Experiments



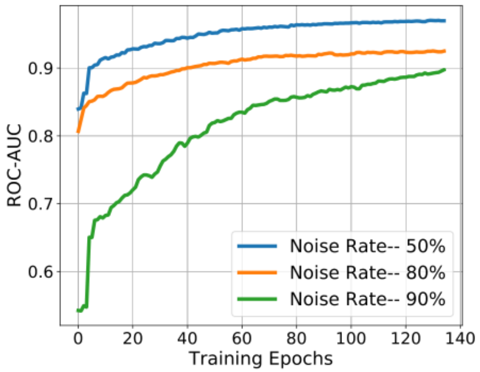
Clothing1M dataset:

| Method | Backbone | Test Accuracy |
|---|---|---|
| Standard CE | ResNet-50 | 69.21 |
| Joint-Optim [51] | ResNet-50 | 72.00 |
| MetaCleaner [69] | ResNet-50 | 72.50 |
| MLNT [26] | ResNet-50 | 73.47 |
| PCIL [64] | ResNet-50 | 73.49 |
| JPL [20] | ResNet-50 | 74.15 |
| DMix [25] | ResNet-50 | 74.76 |
| ELR [30] | ResNet-50 | 74.81 |
| UNICON | ResNet-50 | **74.98** |

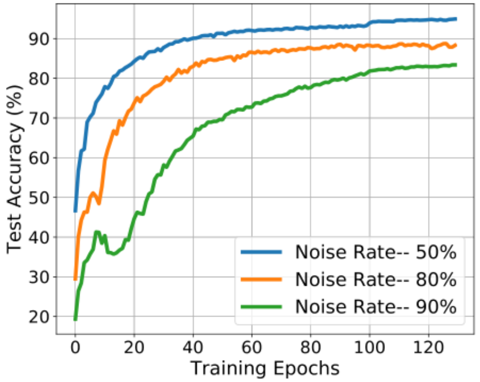
| Dataset | WebVision | | ILSVRC12 | |
|---|---|---|---|---|
| Method | *Top-1* | *Top-5* | *Top-1* | *Top-5* |
| D2L [32] | 62.68 | 84.00 | 57.80 | 81.36 |
| MentrorNet [17] | 63.00 | 81.40 | 57.80 | 79.92 |
| Co-Teaching [12] | 63.58 | 85.20 | 61.48 | 84.70 |
| Iterative-CV [55] | 65.24 | 85.34 | 61.60 | 84.98 |
| DivideMix [25] | 77.32 | 91.64 | 75.20 | 90.84 |
| ELR [30] | 77.78 | 91.68 | 70.29 | 89.76 |
| MOIT [39] | **78.76** | - | - | - |
| UNICON | 77.60 | **93.44** | **75.29** | **93.72** |

# Experiments

ROC-AUC score and test accuracy (%) on CIFAR10 with different noise rates. As the model becomes more precise in selection, the test-time performance improves accordingly.

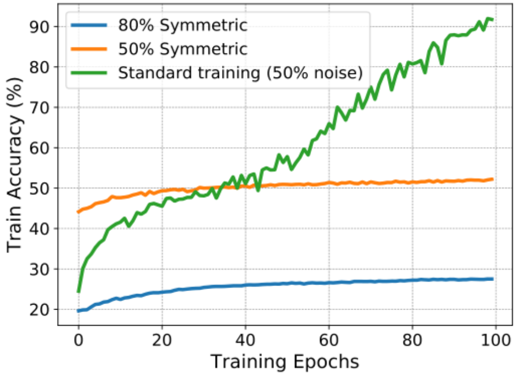



(a) ROC-AUC score vs epochs.




(b) Test Accuracy vs Epochs

Training accuracy at different epochs. Low accuracy indicates that the networks do not memorize the noisy labels even after long training. In contrast to UNICON, standard CE loss based training leads to a high training accuracy (should be~50%), i.e., complete memorization of noisy labels.



# Experiments



Ablation study with different training settings. Both contrastive loss and class-imbalance affects the performance significantly; especially for high noise rates. Ensembling the outputs of both network during separation seems to improve the performance as well. Test results at last epoch are also shown here.

| Dataset | CIFAR10 | | | | | | CIFAR100 | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Noise Rate | 50% | | 80% | | 90% | | 50% | | 80% | | 90% | |
| Method | Best | Last | Best | Last | Best | Last | Best | Last | Best | Last | Best | Last |
| UNICON w/o balancing | 94.28 | 94.06 | 91.41 | 91.16 | 85.49 | 85.28 | 75.26 | 75.01 | 60.51 | 60.16 | 39.87 | 39.02 |
| UNICON w/o CL | 94.92 | 94.24 | 91.67 | 91.21 | 87.28 | 86.34 | 75.75 | 75.09 | 60.54 | 60.17 | 41.83 | 41.11 |
| UNICON w/o ensemble | 95.20 | 94.91 | 92.38 | 92.11 | 88.84 | 88.18 | 76.28 | 76.10 | 62.98 | 62.11 | 42.36 | 41.56 |
| UNICON | **95.61** | **95.24** | **93.97** | **93.97** | **90.81** | **89.95** | **77.63** | **76.91** | **63.98** | **63.13** | **44.82** | **44.51** |

# Thanks