TAAI 2024
International Track

Adaptive Sample Selection Approach for Learning with Noisy Labels

Jing-Yong Wang and Mei-Chen Yeh

Department of Computer Science and Information Engineering

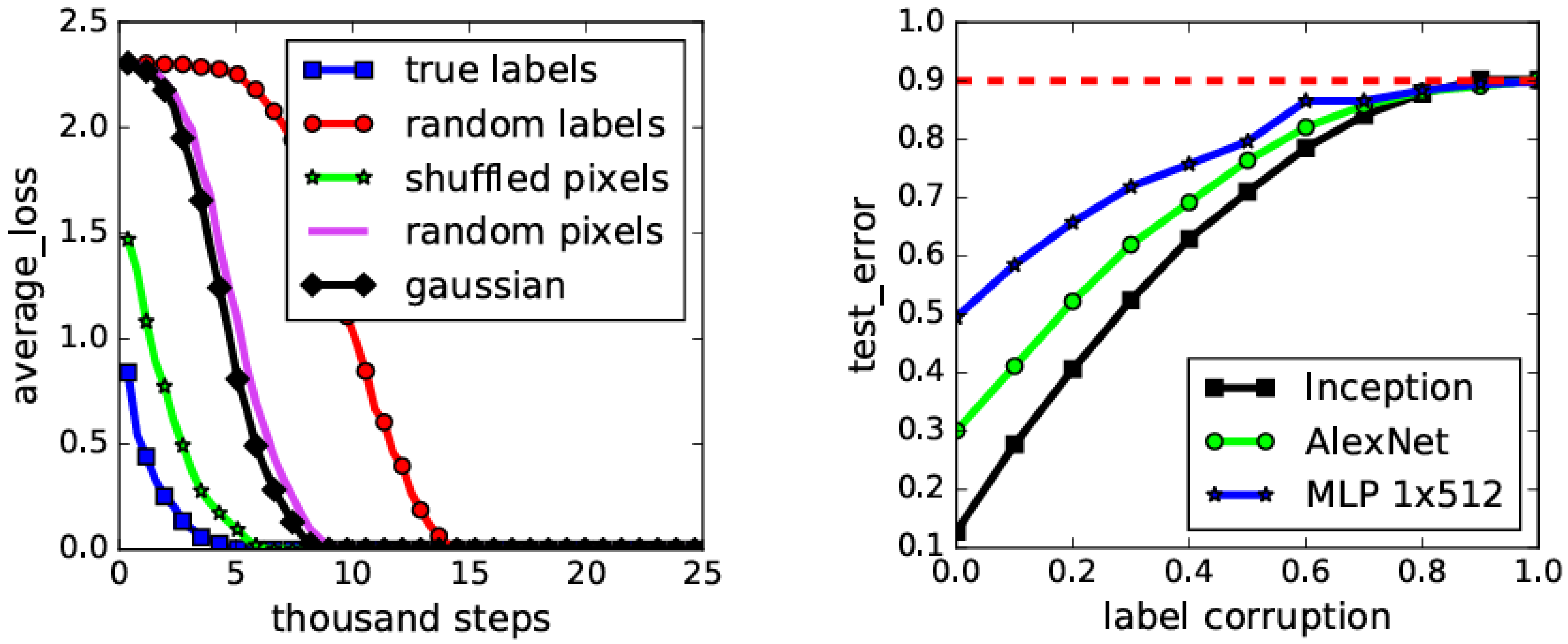
National Taiwan Normal University

Learning with Noisy Labels

- Assumes the label information provided in the training dataset may not be entirely correct
- Why do we have noisy labels?
  - Annotating a large amount of data
  - Dealing with pseudo-labels in semi-supervised learning

Learning with Noisy Labels

- Learning from noisy data decreases model performance.



Chiyuan Zhang, Samy Bengio, Moritz Hardt, Benjamin Recht, Oriol Vinyals, “Understanding deep learning requires rethinking generalization,” *ICLR* 2017.

Learning with Noisy Labels

- Current methods typically resort to segregating the training dataset into distinct *clean* and *noisy* subsets.
- Following this, semi-supervised learning approaches are applied, where clean samples are employed for supervised training, and noisy samples are utilized without corresponding ground-truth labels.

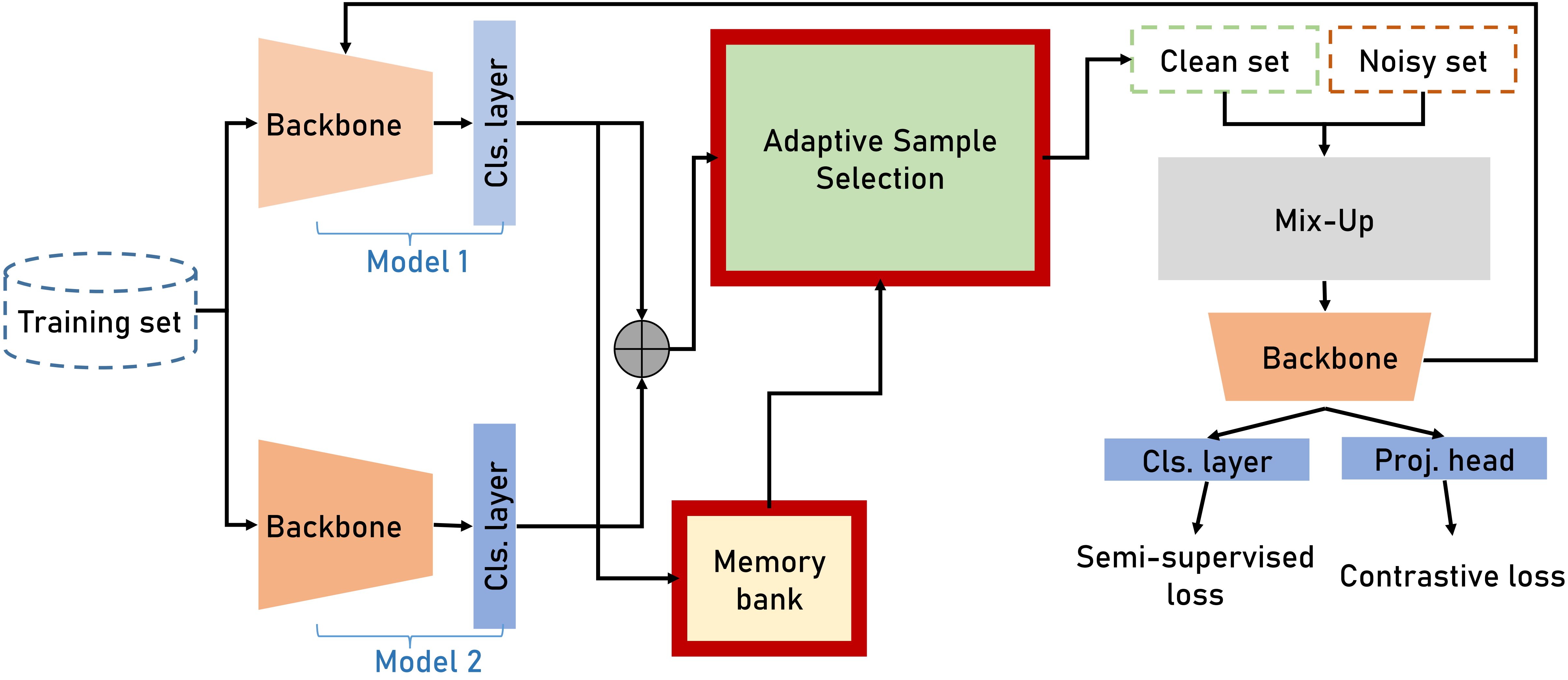
The quality is more important than the quantity!

Distinguishing clean and noisy samples is important!

Contributions

- We propose a *rigorous* and *adaptive* sample selection approach for selecting high-quality clean samples from the training set.
- We progressively extend the clean set by performing *k*NN classification on pre-filtered clean samples, maximizing the value of the clean set filtered in the previous stage.
- This approach can effectively handle data with varying noise levels, without requiring prior knowledge of the noise rate.
- We conduct extensive experiments on several datasets, and show comparable or even better performance than previous methods on these benchmarks.

Method


Training set
Backbone
Cls. layer
Model 1
Backbone
Cls. layer
Model 2
Adaptive Sample Selection
Memory bank
Clean set
Noisy set
Mix-Up
Backbone
Cls. layer
Proj. head
Semi-supervised loss
Contrastive loss

Adaptive Sample Selection

- Assumption: large disagreement between ground-truth labels and predicted probabilities implies a noisy or difficult sample!

$$\begin{aligned} d_i &= \mathrm{JSD}(\mathbf{y}_i, \mathbf{p}_i) \\ &= \frac{1}{2}\mathrm{KLD}(\mathbf{y}_i || \frac{\mathbf{y}_i + \mathbf{p}_i}{2}) + \frac{1}{2}\mathrm{KLD}(\mathbf{p}_i || \frac{\mathbf{y}_i + \mathbf{p}_i}{2}) \end{aligned}$$

- A sample is classified as clean only if it not only exhibits a small disagreement but also consistently maintains this small discrepancy in recent predictions.

Be refined with more training iterations

$$d_i \leq d_{\text{avg}},$$

$$s_i \leq s_{\text{avg}} + s_{\text{avg}} \times \underline{(1 - d_{\text{avg}})} \times \underline{(1 - \frac{t}{T})}$$

Be conservative with a high noise rate

Adaptive Sample Selection

- A *k*-nearest neighbors search retrieves additional clean samples, leveraging the rigorously selected clean set from the previous stage.

$$k = \sqrt{|D_{\text{clean}}|} \times \alpha_k$$

- However, if the *k*NN prediction of a clean sample contradicts the ground truth label, we evaluate further whether this sample should remain in the clean set.

$$s_i \geq \underline{(s_{\text{avg}} + s_{\text{avg}} \times (1 - d_{\text{avg}})) \times (1 - \frac{t}{T})}$$

More rigorous as the threshold is lower!

Algorithm 1 Adaptive Sample Selection

Input:

Training set $D = \{X, Y\}$;

A scaling factor α_k;

Output:

Clean sample subset D_{clean};

Initialize $D_{\text{clean}} \leftarrow \emptyset$;
for $x_i \in D$ **do**
  Calculate d_i and s_i;
  if $d_i \leq d_{\text{avg}}$ **and** $s_i \leq s_{\text{avg}} + s_{\text{avg}} \times (1 - d_{\text{avg}}) \times (1 - \frac{t}{T})$ **then**
    $D_{\text{clean}} \leftarrow x_i$;
  end if
end for
Determine $k = \sqrt{|D_{\text{clean}}|} \times \alpha_k$;
for $x_i \in D$ **do**
  Perform k-NN label classification using D_{clean}
  if k-NN prediction equals ground-truth **then**
    $D_{\text{clean}} \leftarrow D_{\text{clean}} \cup \{x_i\}$
  else
    if $x_i \in D_{\text{clean}}$ **and** Equ. 5 holds **then**
      $D_{\text{clean}} \leftarrow D_{\text{clean}} \setminus \{x_i\}$;
    end if
  end if
end for
Perform Uniform Selection;
return D_{clean};

Experimental settings

- Datasets
  - CIFAR-10 (50K images, 10 classes)
  - CIFAR-100 (50K images, 100 classes)
  - WebVision (66K images, 50 classes)
- Evaluation protocol
  - Symmetric noise rates: 20%, 50%, 80%, 90%
  - Asymmetric noise rates: 10%, 30%, 40%
  - Metric: Top-1 accuracy rate
- Implementation details
  - Backbone: ResNet18 for CIFAR-10 and CIFAR-100, InceptionResNetV2 for WebVision
  - Warm-up: 10 epochs for CIFAR-10, 30 epochs for CIFAR-100, 2 epochs for WebVision
  - Batch size: 64 for CIFAR-10 and CIFAR-100, 32 for WebVision

symmetric noise

Method	CIFAR-10				CIFAR-100			
	20%	50%	80%	90%	20%	50%	80%	90%
DMix [15] (*ICLR'20*)	96.1	94.6	92.9	76.0	77.3	74.6	60.2	31.5
ELR [20] (*NeurIPS'20*)	95.8	94.8	93.3	78.7	77.6	73.6	60.8	33.4
JPL [11] (*CVPR'21*)	93.5	90.2	35.7	23.4	70.9	67.7	17.8	12.8
MOIT [22] (*CVPR'21*)	94.1	91.1	75.8	70.1	75.9	70.1	51.4	24.5
UNICON [10] (*CVPR'22*)	96.0	<u>95.6</u>	<u>93.9</u>	<u>90.8</u>	78.9	<u>77.6</u>	<u>63.9</u>	<u>44.8</u>
Sel-CL+ [17] (*CVPR'22*)	95.5	93.9	89.2	81.9	76.5	72.4	59.6	48.8
DISC [19] (*CVPR'23*)	**96.5**	-	-	-	<u>80.1</u>	-	-	-
DivideMix+ [27] (*ICCV'23*)	<u>96.2</u>	94.7	93.1	-	77.5	74.7	60.4	-
Ours	95.4	**95.9**	**95.1**	**93.7**	**81.2**	**77.8**	**66.9**	**51.7**

asymmetric noise

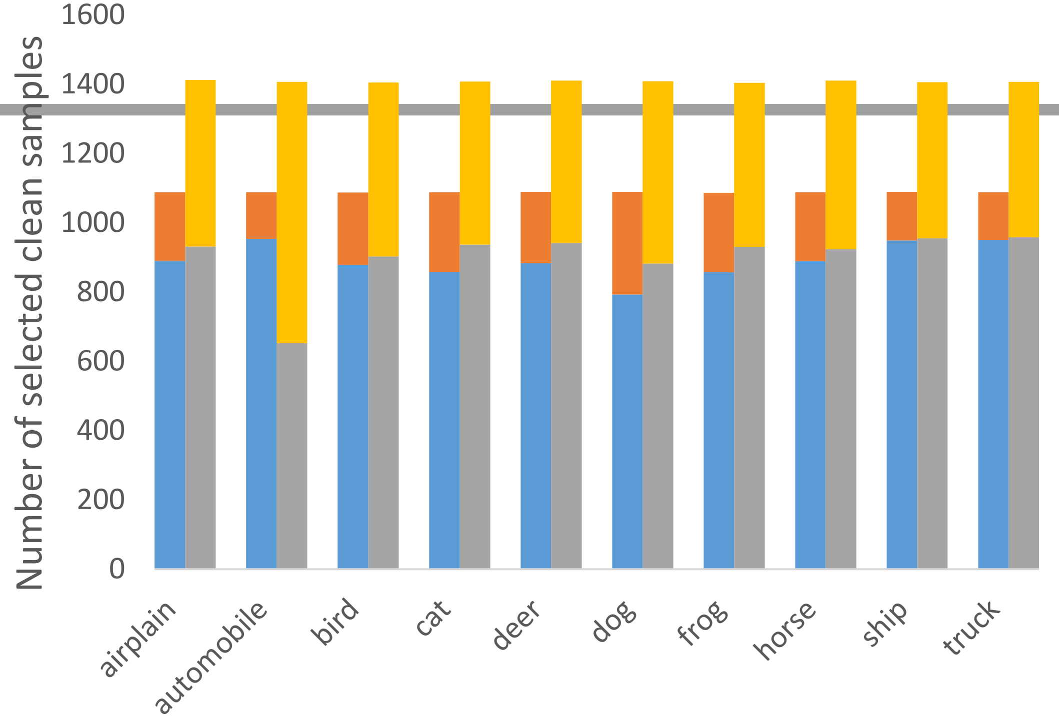
Method	CIFAR-10			CIFAR-100		
	10%	30%	40%	10%	30%	40%
DMix [15] (*ICLR'20*)	93.8	92.5	91.7	71.6	69.5	55.1
ELR [20] (*NeurIPS'20*)	<u>95.4</u>	94.7	93.0	77.3	74.6	73.2
JPL [11] (*CVPR'21*)	94.2	92.5	90.7	72.0	68.1	59.5
MOIT [22] (*CVPR'21*)	94.2	94.1	93.2	77.4	75.1	74.0
UNICON [10] (*CVPR'22*)	95.3	<u>94.8</u>	<u>94.1</u>	78.2	75.6	<u>74.8</u>
Sel-CL+ [17] (*CVPR'22*)	**95.6**	94.5	93.4	<u>78.7</u>	<u>76.4</u>	74.2
Ours	95.1	**95.0**	**94.3**	**81.5**	**80.0**	**78.1**

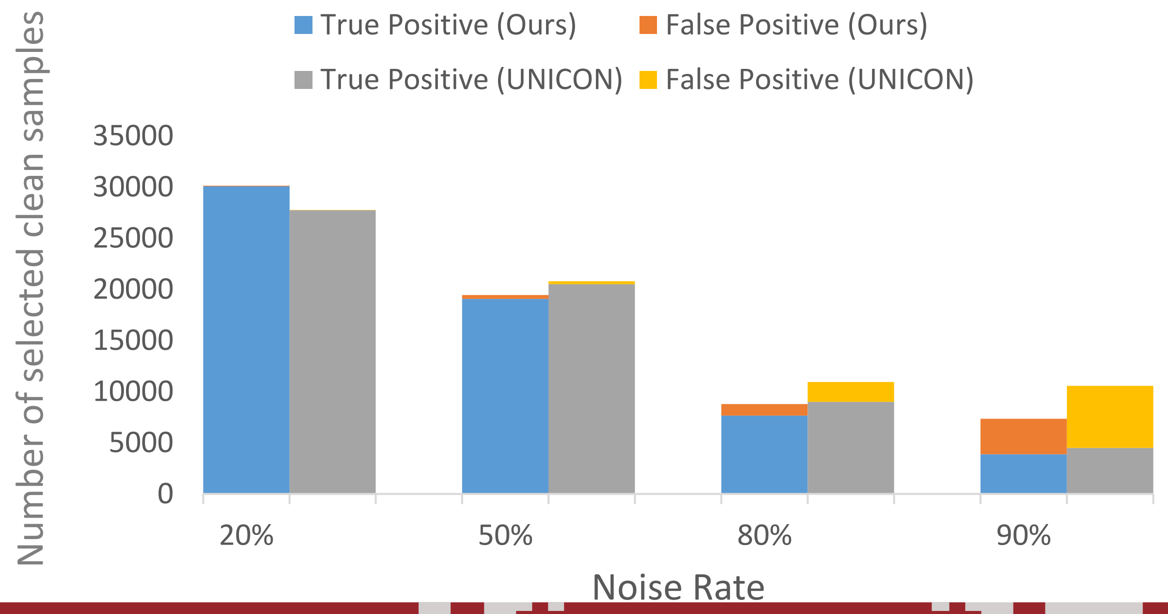
real noise

Method	WebVision		ILSVRC12	
	Top-1	Top-5	Top-1	Top-5
DMix [15] (*ICLR'20*)	77.32	91.64	75.20	90.84
ELR [20] (*NeurIPS'20*)	77.78	91.68	70.29	89.72
MOIT [22] (*CVPR'21*)	78.76	-	-	-
UNICON [10] (*CVPR'22*)	77.60	**93.44**	75.29	**93.72**
DISC [19] (*CVPR'23*)	<u>80.28</u>	92.28	**77.44**	92.28
DivideMix+ [27] (*ICCV'23*)	77.51	91.95	75.51	91.58
Ours	**80.40**	<u>92.72</u>	<u>75.80</u>	<u>92.68</u>

Ablation experiment

Consistency-based selection	kNN	CIFAR-100 (sym) 20%	50%	80%	90%
✗	✗	79.2	77.7	65.8	45.2
✓	✗	77.9	75.7	66.3	45.9
✗	✓	80.7	77.2	66.8	51.3
✓	✓	**81.2**	**77.8**	**66.9**	**51.7**





Conclusion

- We present a sample selection method for learning with noisy labels
  - Considering the disagreement of model predictions and ground-truth, and the stability of model predictions
  - Handling data with varying noise levels
  - Expanding the clean sample set by a traditional machine learning method
- We will integrate the proposed method with advanced semi-supervised or self-supervised learning techniques, which could lead to further improvements in learning with noisy labels.

Questions?

More Information:

https://web.ntnu.edu.tw/~myeh/