Safeguarded Dynamic Label Regression for Noisy Supervision

Jiangchao Yao[1,2] Hao Wu[1] Ya Zhang[1]
Ivor W. Tsang[2] Jun Sun[1]

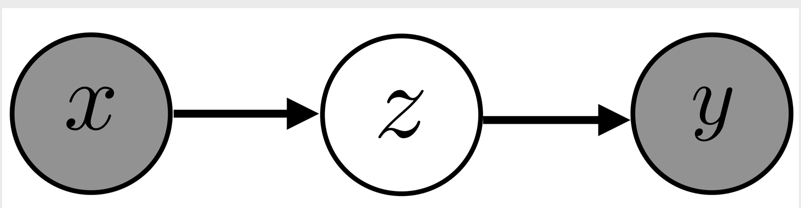
[1]CMIC, Shanghai Jiao Tong University [2]CAI, University of Technology Sydney

Overview

TL;DR: Safeguarded dynamic label regression is a method to dynamically learn the noise transition and train the classifier in a **stochastic** and **stable** way, which reduces computational burden in the EM variant of [Patrini, CVPR-17] and safeguard the transition update in [Goldberger, ICLR-18].

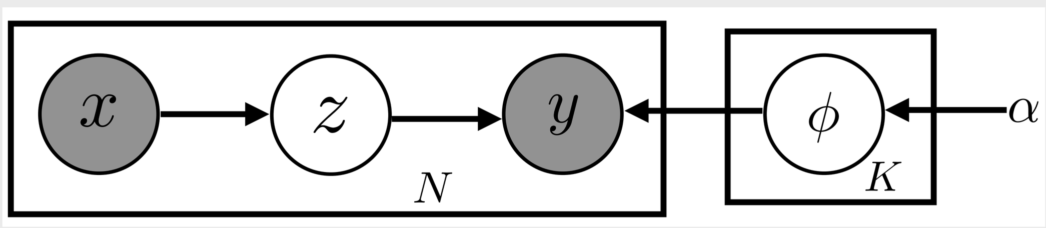
- **Noisy labels** are corrupted from ground-truth labels, which degenerates the robustness of learning models.
- **Deep neural networks** have the high capacity to fit any noisy labels. The solutions are as follows.
  - ⋄ **Noise transition**. E.g., F-correction.
  - ⋄ **Sample re-weighting**. E.g., Bootstrapping.
  - ⋄ **Model regularization**. E.g., Dropout.
- We present a stochastic and stable way for deep learning with noisy supervision.
  - ⋄ Give a Latent Class-Conditional Noise model (**LCCN**).
  - ⋄ Optimize LCCN via Dynamic Label Regression (**DLR**).
- Empirical results on **CIFAR10, CIFAR100** with different noise rates and **Clothing1M, WebVision** demonstrate our approach improves the robustness of classifiers.

Deficiency of Benchmarks



- **Inaccurate** transition estimation [Patrini, CVPR-17].
- **Unstable** transition adaptation [Goldberger, ICLR-18].

Latent Class-Conditional Noise model



- Explicitly model the noise transition on a **simplex**.
- Build **global data dependency** for the noise transition.

Dynamic Label Regression

The optimization of LCCN refers to the following iteration.

- Infer the latent label via Gibbs sampling

$$P(z_n|Z^{\neg n}, X, Y; \alpha) \propto \underbrace{P(z_n|x_n)}_{\text{Classifier}} \underbrace{\frac{\alpha_{y_n} + N^{\neg n}_{z_n y_n}}{\sum_{k'}^{K}(\alpha_{k'} + N^{\neg n}_{z_n k'})}}_{\text{Conditional transition}}$$

- Loss minimization for parameter learning

$$\min\ \ell_1(z_n, P(z_n|x_n)) \text{ and } \min\ \ell_2(y_n, P(y_n|z_n))$$

Although the optimization iterates above two steps, it is still computational effective in a **stochastic** way, thus scalable.

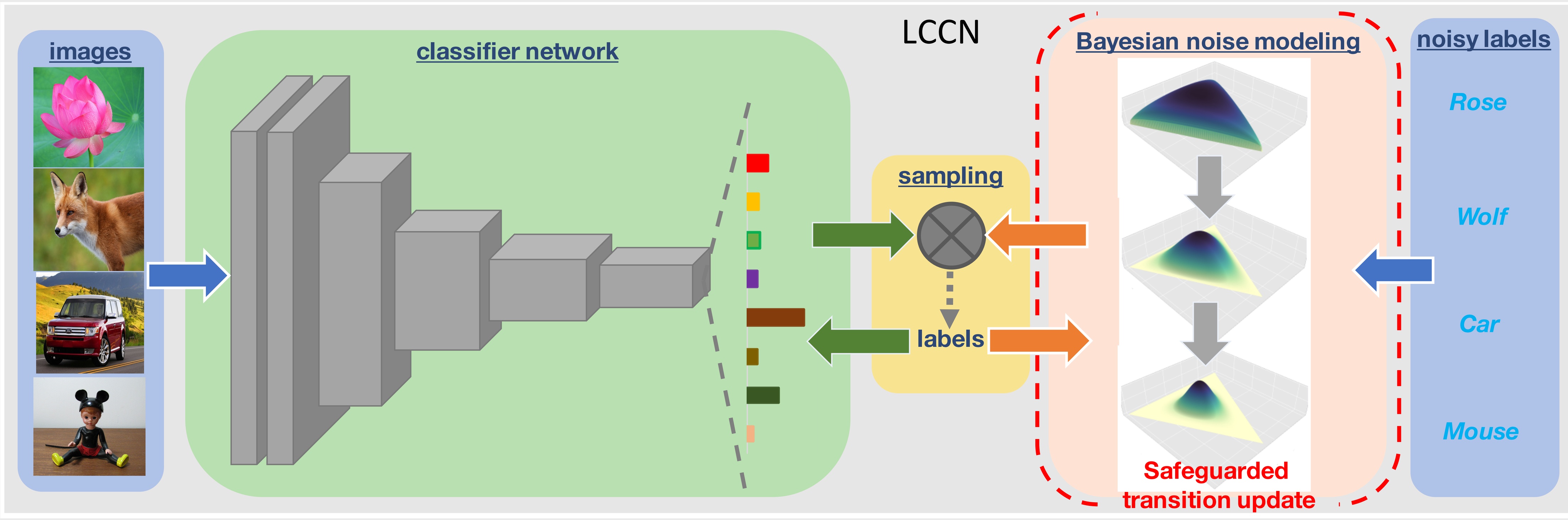
Safeguarded transition update:

$$|\phi_i^{new} - \phi_i^{old}| \leq 2 * \frac{M}{N_i},$$

where the RHS of above inequality correlated to 2 times of the ratio between the batch size M and the sample number for each category N_i, is usually in a small scale if $M \ll N_i$.

QR Code for Implementation

Illustration of Safeguarded Dynamic Label Regression



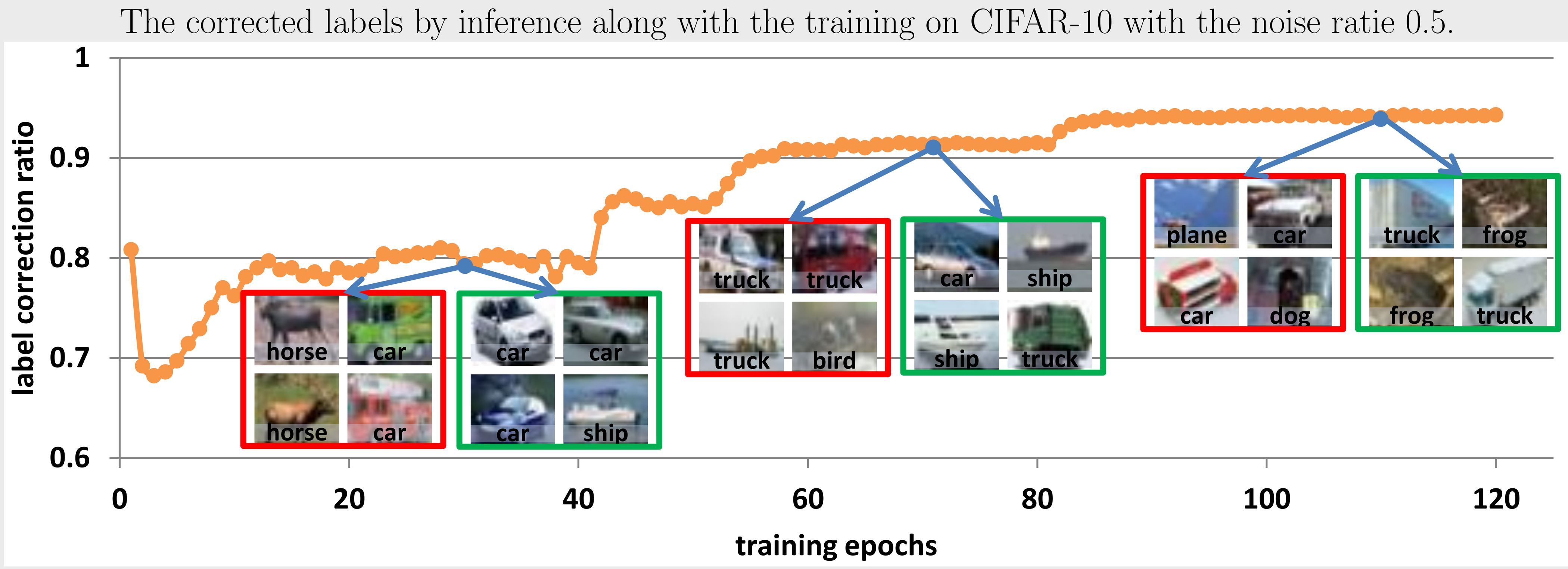
CIFAR-10 and CIFAR-100

The average accuracy (%) over 5 trials on CIFAR-10 and CIFAR-100 with different noise levels.

	Dataset	CIFAR-10					CIFAR-100				
#	Method \ Noise Ratio	0.1	0.3	0.5	0.7	0.9	0.1	0.2	0.3	0.4	0.5
1	CE	90.10	88.12	76.93	59.01	56.85	66.15	64.31	60.11	51.68	33.37
2	Bootstrapping	90.73	88.12	76.29	57.04	56.79	66.48	64.61	63.01	55.27	**34.52**
3	Forward	90.86	89.03	82.47	67.11	57.29	65.43	62.72	61.28	52.64	33.82
4	S-adaptation	91.02	88.83	86.79	72.74	60.92	65.52	64.11	62.39	52.74	30.07
5	LCCN	**91.35**	**89.33**	**88.41**	**79.48**	**64.82**	**67.83**	**67.63**	**66.86**	**65.52**	33.71
6	CE with the clean data	91.63					69.41				

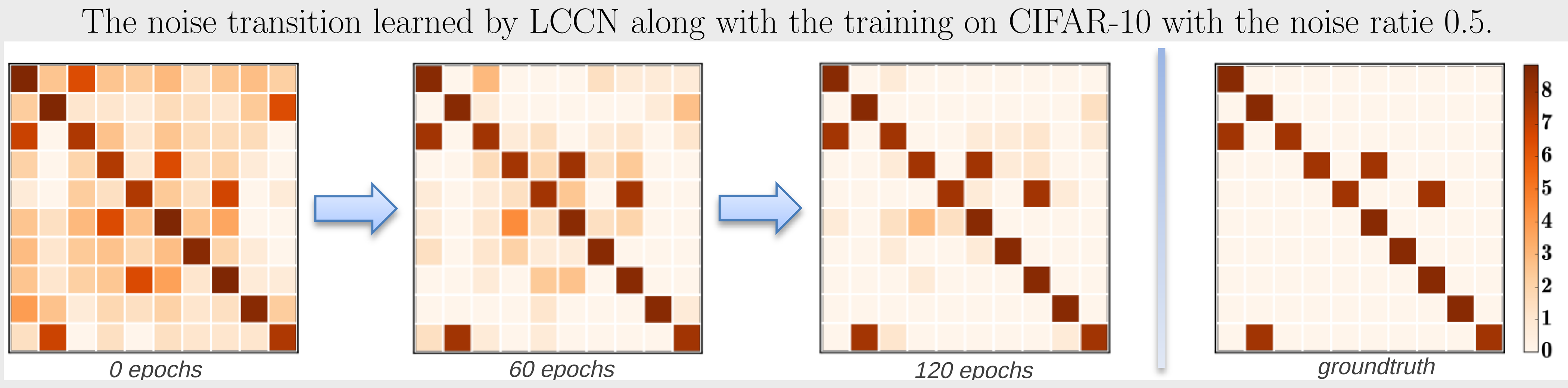
DLR: Latent Label Inference

The corrected labels by inference along with the training on CIFAR-10 with the noise ratie 0.5.



DLR: Noise Transition Estimation

The noise transition learned by LCCN along with the training on CIFAR-10 with the noise ratie 0.5.



Clothing1M (Left) and WebVision (Right)

#	Method	Accuracy
1	CE	68.94
2	Bootstrapping	69.12
3	Forward	69.84
4	S-adaptation	70.36
5	Joint Optimization	72.23
6	LCCN	**73.07**
7	CE with the clean data	75.28

#	Method	Accuracy@1	Accuracy@5
1	CE	63.11	83.69
2	Bootstrapping	63.20	83.81
3	Forward	63.10	83.78
4	S-adaptation	62.54	81.73
5	LCCN	**63.52**	**84.27**