Learning with Noisy Labels by Efficient Transition Matrix Estimation to Combat Label Miscorrection ECCV 2022

Seong Min Kye[1*] Kwanghee Choi[1*] Joonyoung Yi[1*] Buru Chang[1] *Equal contributions.

{harris,kwanghee.choi,joonyoung.yi,buru.chang}@hpcnt.com [1]Hyperconnect, Republic of Korea

HYPERCONNECT

Background

Learning with noisy labels. In the last decade, supervised learning has achieved great success by leveraging an abundant amount of annotated data to solve various tasks. However, we cannot avoid noisy labels of human annotation due to its coarse-grained annotation sources, resulting in performance degradation.

Among various noises, we concentrate on the **class-dependent label noise**. For example, consider classifying cats, dogs, and cars. We can expect that the annotators would be more confused between cats and dogs than cats and cars.

Unlike traditional methods, recent methods often additionally utilize an inexpensively obtained **small clean dataset**. Especially, recent **label correction methods** achieve remarkable performance based on **model-agnostic meta-learning** (MAML). Label correction methods relabel noisy labels to directly reduce the noise level, raising the theoretical upper bound of the predictive performance

Motivation

There are two challenges for these MAML-based label correction methods:

- **The label correction methods blindly trust the already miscorrected labels.** Erroneously corrected labels are often kept throughout the training, which causes the model to learn the miscorrected labels as ground-truth labels.
- **MAML-based methods are inherently slow in training, resulting in excessive computational overhead.** The inefficiency comes from multiple training steps per single iteration of MAML-based methods.

To alleviate these issues, we propose ***FasTEN*** (**Fas**t **T**ransition Matrix **E**stimation for Learning with **N**oisy Labels). Intuitively, our method remains skeptical about its correction by using the transition matrix while training.

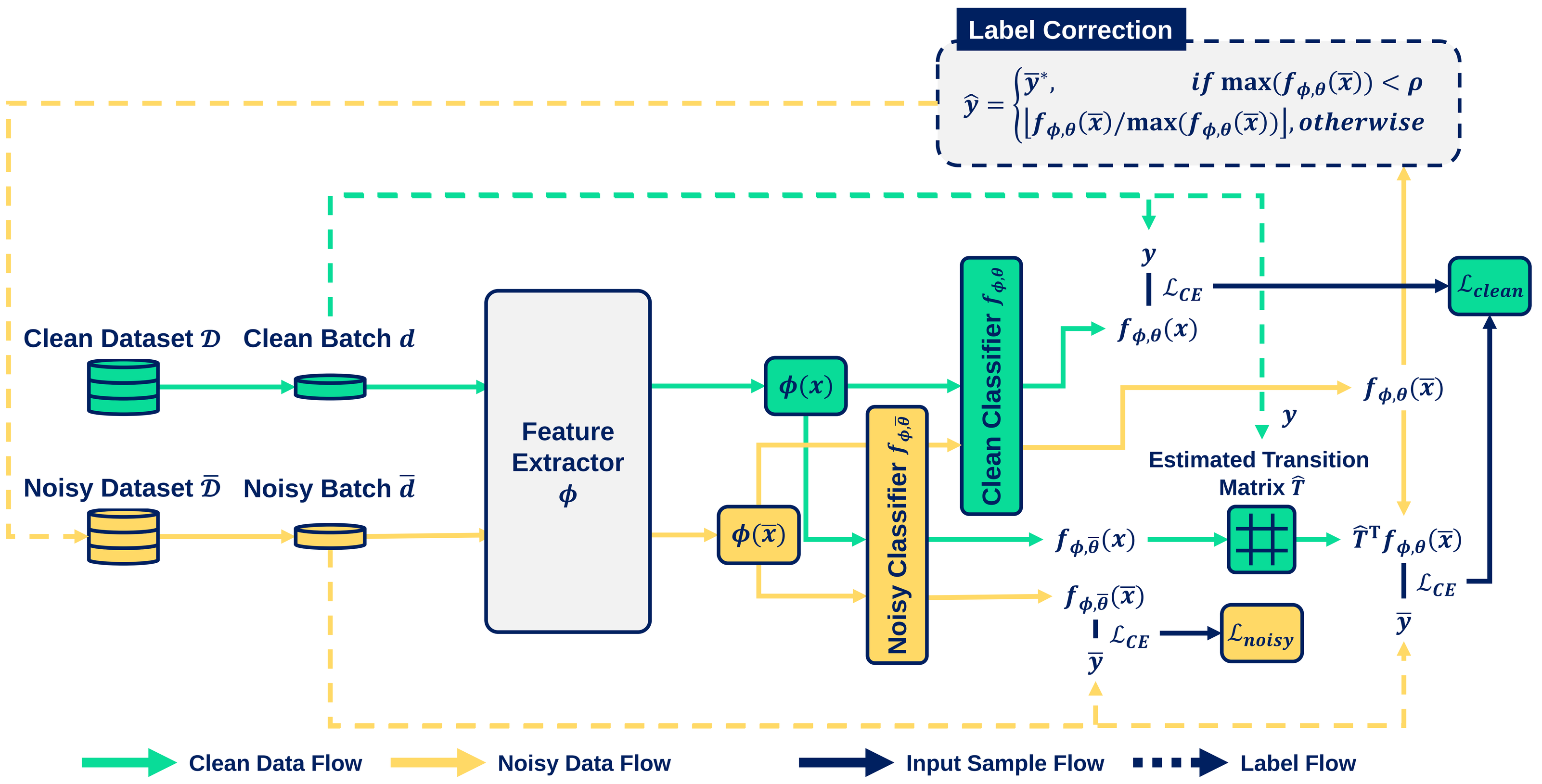
The **label transition matrix** captures the level of confusion between classes. Each element of the matrix is defined as the probability of a clean label i to be corrupted as a noisy label j, i.e. $p(\bar{y} = j | y = i)$. By attaching the matrix to the classifier while training, we can yield a classifier robust to noisy labels.

However, label correction shifts the noise level, changing the optimal transition matrix. Hence, a way to efficiently estimate the matrix is needed.

Our Contributions

- We propose a robust and efficient method that **learns a transition matrix to learn with noisy labels while continuously correcting them on the fly**. To the best of our knowledge, this is the first attempt to improve the label correction with the transition matrix estimation.
- Our proposed method ***FasTEN*** boosts training speed by employing a **two-head architecture** so that the label transition matrix can be efficiently learned with a single back-propagation.
- Extensive experiments validate the efficacy of our proposed method in terms of **both training speed and predictive performance**.

Proposed Method



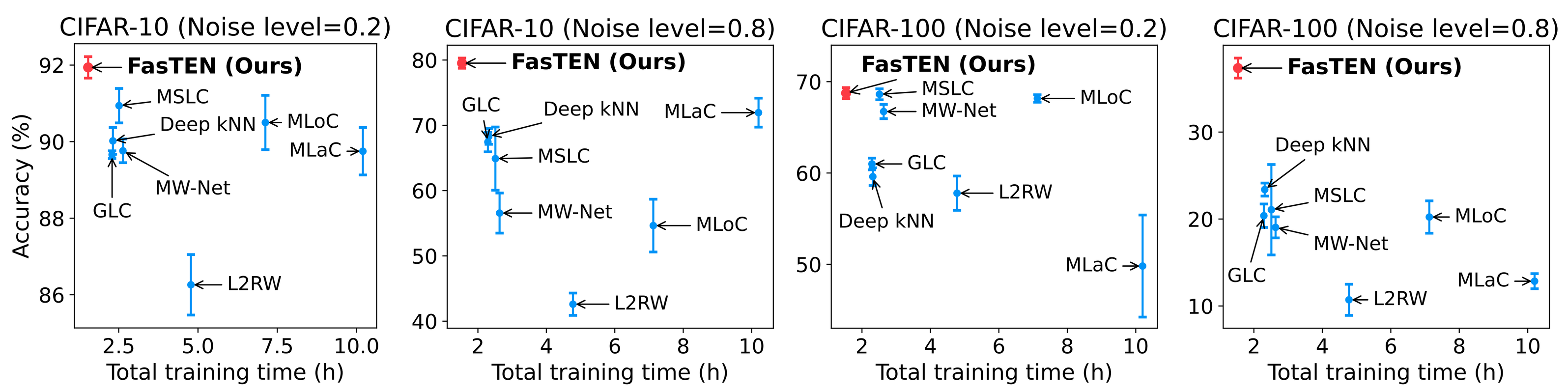
To efficiently estimate the transition matrix, we adopt a **two-head architecture**, a noisy and a clean classifier, with a shared feature extractor. Unlike MAML-based methods, **our FasTEN only requires a single back-propagation every iteration**, accelerating the training speed.

The noisy classifier aims to be fitted to the noisy data so that we can estimate the transition probability $p(\bar{y} = j | y = i)$ by feeding the clean samples to the noisy classifier every iteration. **The clean classifier** is trained to be statistically consistent by leveraging the estimated transition matrix by attaching the matrix to the classifier when applying the cross-entropy loss.

Finally, we employ a naive **label correction** strategy by feeding the noisy samples to the clean classifier to correct the labels. We threshold the samples based on their maximum probability. As it relies only on the most recent prediction, the decision is prone to change, **avoiding blindly trusting the corrections**.

Experimental Results

CIFAR-10/100 with Synthetic Noise. FasTEN is the quickest to train while having comparable or better accuracy on various noise levels.



Clothing1M with Real-world Noise. FasTEN beats the baselines by a large margin on Clothing1M which contains instance-dependent noisy label.

	SotA w/o the clean set	SotA w/ the clean set	FasTEN (Ours)
Accuracy	75.11 (AugDesc)	75.78 (MLaC)	**77.83** $\pm$ 0.17

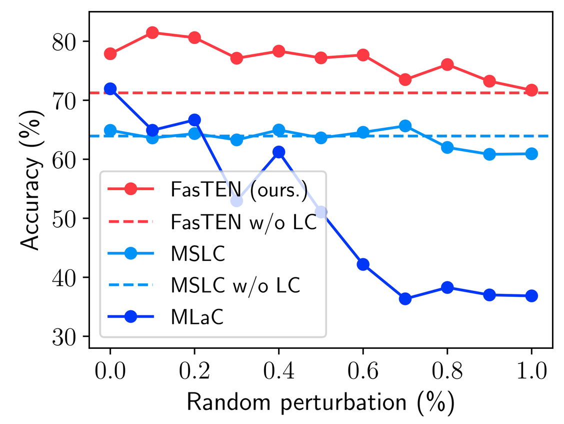
Further Analyses

Theoretical analysis. It is known that the estimation accuracy determines the bounds of the generalization error of the classifier. As the method estimates the transition matrix every iteration, we theoretically analyze the relationship between the batch size and the matrix estimation error bound in Theorem 1.

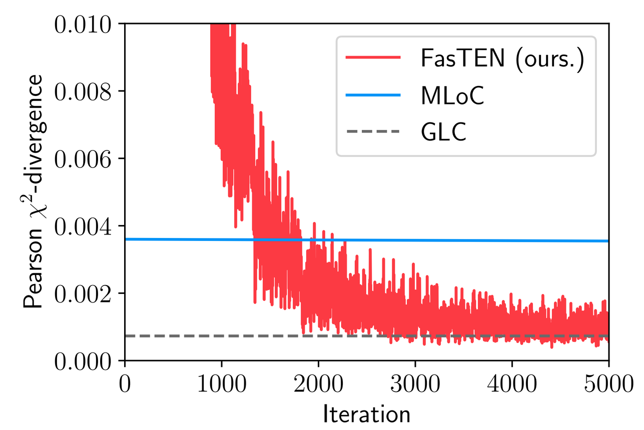
Label correction accuracies. We measure the performance on all training samples (Overall) and incorrectly labeled training samples (Incorrect). The results show that our method can successfully correct the noisy labels, implying that our FasTEN may be helpful in further cleansing the noisy training set. Also, notable under-performance of the meta-model of MAML-based MSLC may indicate the inefficacy of the meta-model.

	Reweighting	Matrix estim.	Meta-model	FasTEN
Overall Acc.	0.6024 (Deep kNN)	0.6900 (GLC)	0.2821 (MSLC)	78.47
Incorrect Acc.	0.6024 (Deep kNN)	0.6903 (GLC)	0.2836 (MSLC)	78.61

Robustness to label miscorrections. We compare the robustness with other methods that blindly trust the miscorrected labels. To simulate the miscorrection, we artificial perturb the corrected labels. FasTEN does not degrade performance even if the correction becomes useless (100% perturbation). However, MSLC shows trivial performance gains, implying the uselessness of label correction. Also, MLaC shows steep performance degradation, revealing the susceptibility of MLaC.



Empirical convergence analysis on estimating the matrix. We compare the estimation error between the true and the estimated transition matrix. GLC (dotted line) estimates only once using all the samples, requiring two-staged training. MLoC error decrease is extremely slow (blue line), implying its ineffectiveness on estimation. Our FasTEN uses only a single mini-batch to estimate every iteration, but shows fast convergence, on par with GLC.



On-the-fly matrix estimation quality. We compare the estimated matrix with the true label transition matrix during training. We verify the overall tendency of the estimated transition matrix (red) to follow the true label matrix (blue).