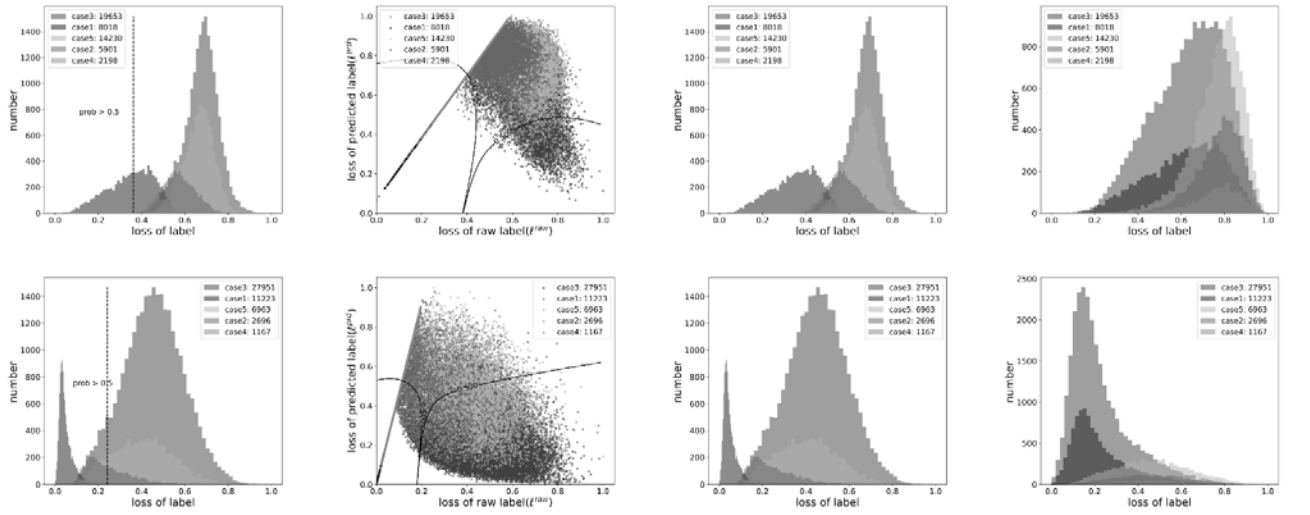
Data fusing and joint training for learning with noisy labels

Yi WEI, Xue MEI, Liu XIN, Pengxiang Xu

Frontiers of Computer Science, DOI: 10.1007/s11704-021-1208-9

Problems & Ideas

- Problems of conventional stereo matching approaches:
  - Labeling samples with the annotator that can annotate large-scale datasets easily, the samples with noisy labels are yielded by these.
  - The Small-loss method used in Learning with Noisy Labels(LNL) has some limitations.
- Ideas: A joint training by fusing sample labels that are generated by data selection method with two kinds of losses.

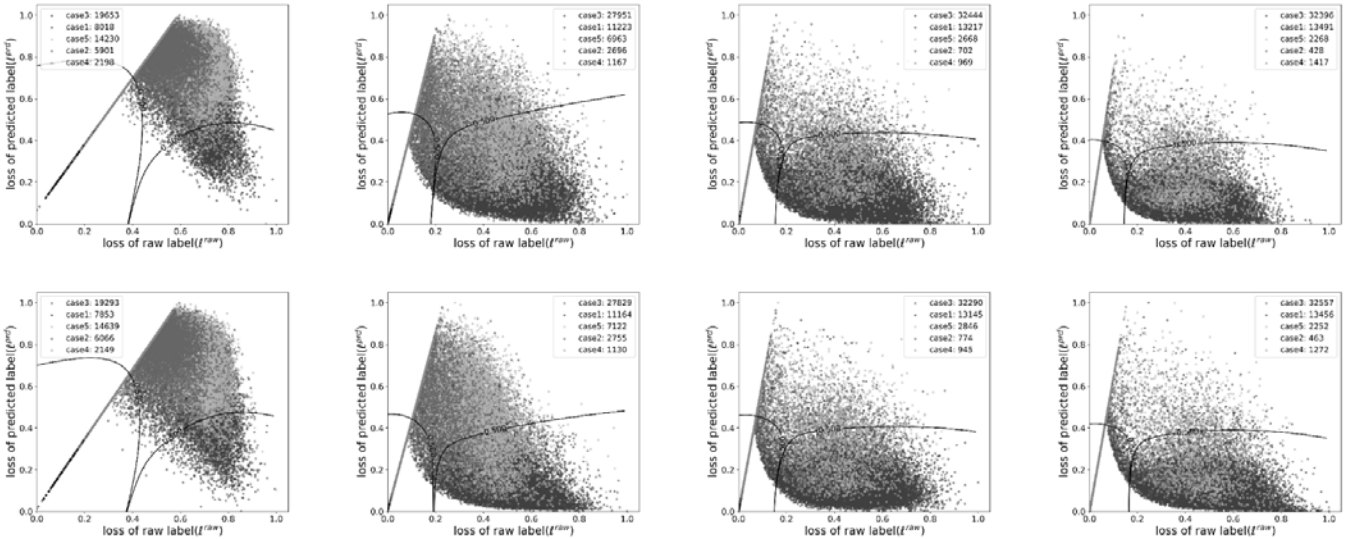


(a) Small-loss method (b) Our method (c) Loss of raw labels (d) Loss of predicted label

Distributions of the normalized loss on CIFAR-10 with 80\% noise ratio. Top: epoch 15; bottom: epoch 50. (a) the distributions of the small-loss method by GMM; (b) the distributions of our method; (c)(d) the loss of raw and predicted labels.

Main Contributions

- Contributions:
  - Two kinds of the sample loss can be used to select the more clean samples than the small-loss methods;
  - A Gaussian Mixture Model (GMM) is dynamically fitted on dataset loss distribution to divide the dataset into a correctly labeled set, a correctly predicted set and a wrong set.
  - The new sample label is generated by fusing the GMM results to train the network. And two networks are used to avoid confirmation bias.



(a) Epoch 15: Cross Entropy (b) Epoch 50: Our method (c) Epoch 150: Our method (d) Epoch 250: Our method

Distributions of the normalized loss on CIFAR-10 with 80% noise ratio. Top: network A; bottom: network B.