

# Improving Noisy Fine-Grained Datasets using Active Label Cleaning Framework

## Supplementary Material

### 7. Stanford Dogs - breed hierarchy

As discussed with a small subsection of the hierarchy in Section 4.1, Figure 5 shows the complete hierarchy with the semantic category relations between different dog breeds using WordNet. The sibling labels are chosen to be those categories that share the same parent node and do not have any child nodes.

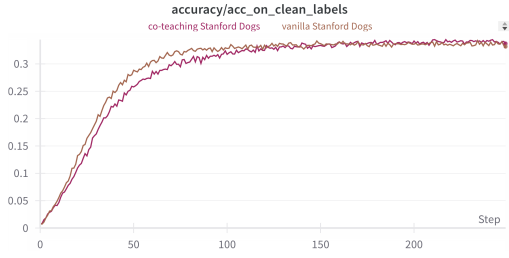
### 8. Implementation details

We train the vanilla models on both Stanford Dogs and CUB-200 datasets using the same hyperparameters with the only change in the number of classes in the final layer of the classifier. We used a ResNet-50 architecture with a bottleneck unit containing convolution layers with batch-normalization. We trained the models for 250 epochs with Adam optimizer keeping  $\beta = (0.9, 0.999)$  and base learning rate of  $10^{-3}$ . A batch size of 64 was used for gradient updates with cross-entropy loss. The images were initially normalized with ImageNet statistics and further augmented with random crop, random horizontal flip, and random affine.

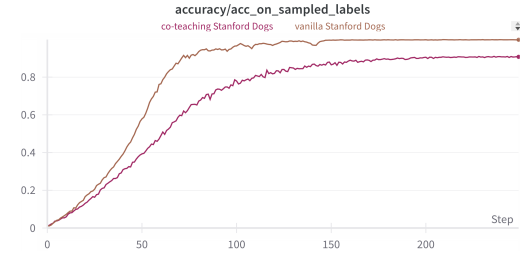
We also train two ResNet-50 models using the co-teaching scheme using the same parameters as mentioned for vanilla. Additionally, the drop rate for excluding noisy samples from the batch is kept as the expected noise rate in the datasets. All samples of the batch including those with large loss are used in weight updates in the initial 10 warm-up epochs.

The training was performed on one NVIDIA-A100 GPU which took approximately 2 hours to train the vanilla model and 2.5 hours to train using the co-teaching scheme. The accuracy curves of the training and validation sets of both Stanford Dogs and CUB-200 datasets are plotted in Figure 6.

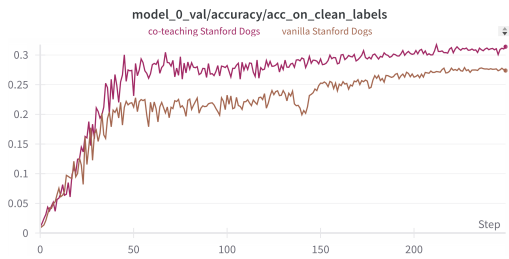




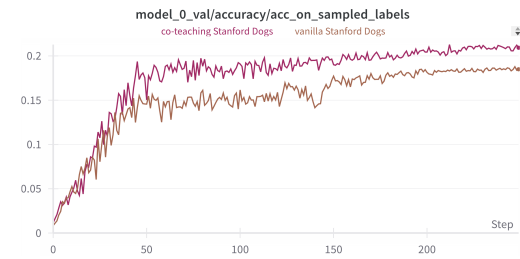
(a) Stanford Dogs - accuracy on clean training set.



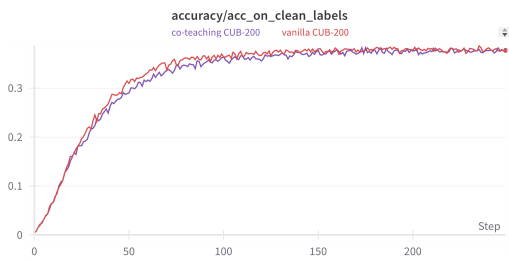
(b) Stanford Dogs - accuracy on noisy training set.



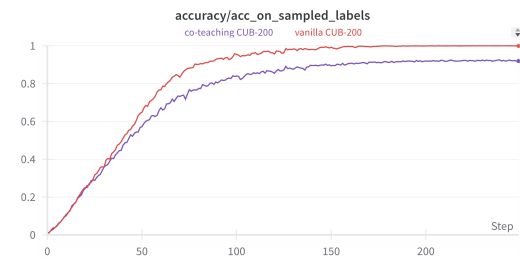
(c) Stanford Dogs - accuracy on clean validation set.



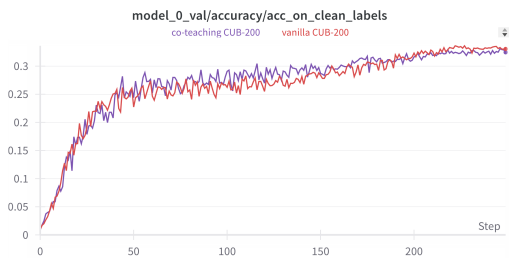
(d) Stanford Dogs - accuracy on noisy validation set.



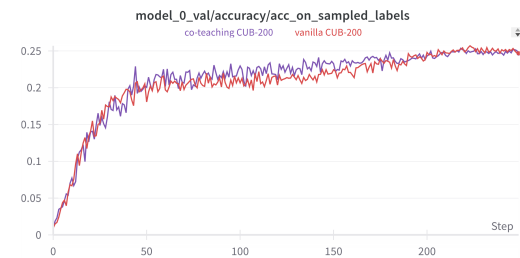
(e) CUB-200 - accuracy on clean training set.



(f) CUB-200 - accuracy on noisy training set.



(g) CUB-200 - accuracy on clean validation set.



(h) CUB-200 - accuracy on noisy validation set.

Figure 6. Classification accuracy curves of training vanilla and co-teaching models on clean and noisy versions of both training and validation sets of Stanford Dogs (a)-(d) and CUB-200 (e)-(h) datasets.