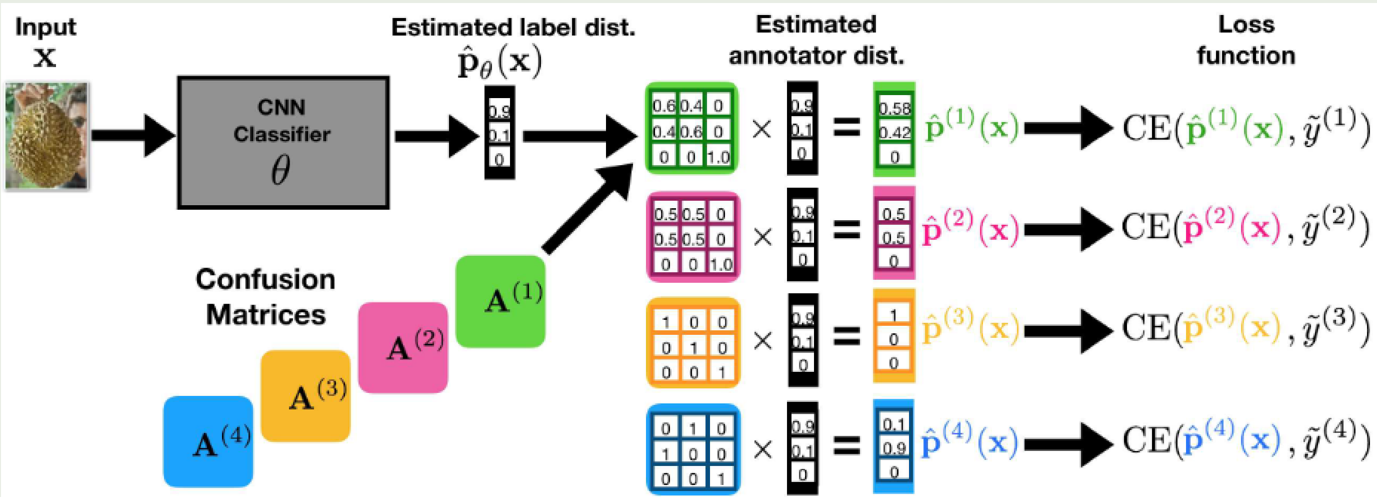
要点

- 该方法属于混淆矩阵(CM)和base classifier同时优化的噪声建模方法
- 加入基于每个annotator的$trace(CM)$的正则化
- 这种方法在1 label per image的情况下表现良好

方法



假设base CNN classifier 输出的$\hat{y}$和y的CM为P；每个标注者的CM为$\hat{A}^r$. 那么联合训练可以得到$P\hat{A}^r \to A^r$

本文证明了：若$trace(\hat{A}^r)$达到最大，那么就可以收敛到A^r. 前提是A^r和$\hat{A}^r$必须为diagoncal dominant

A初始化为identity matrix

实验

噪声种类汇总

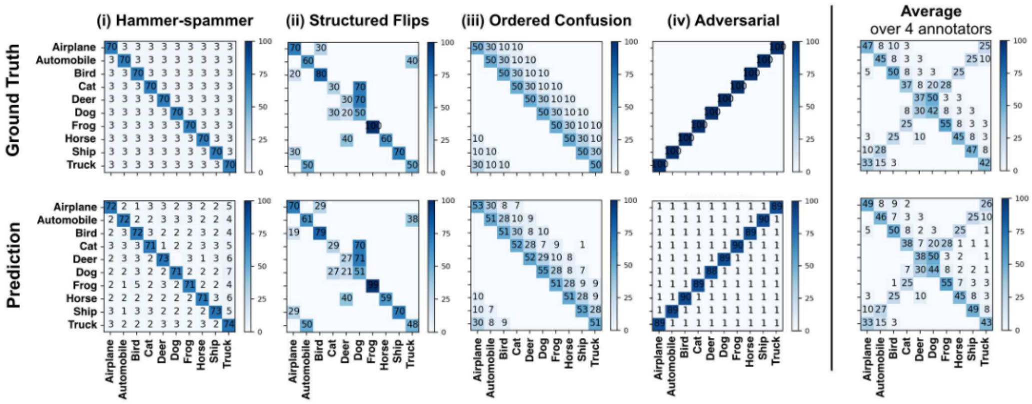
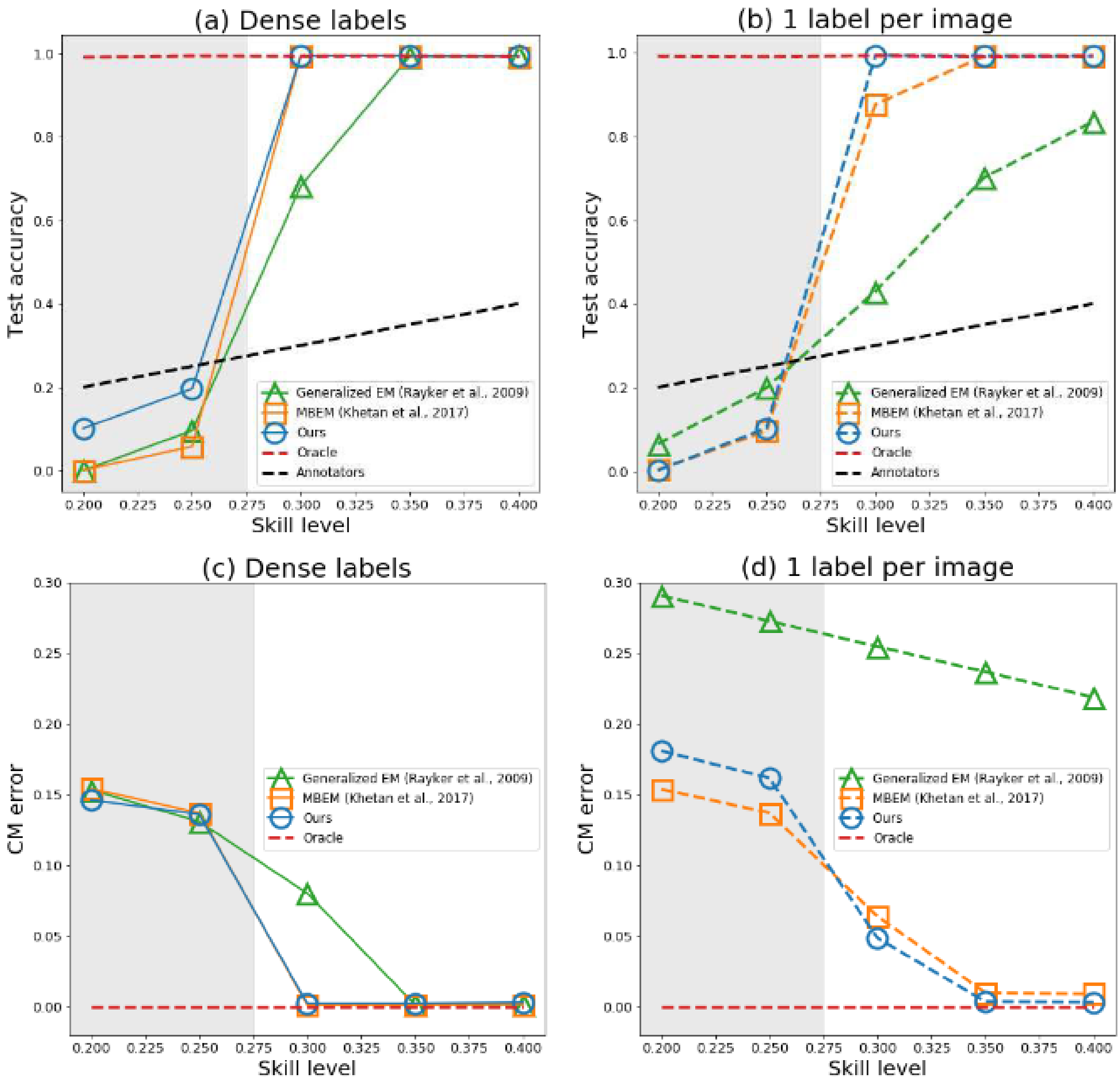
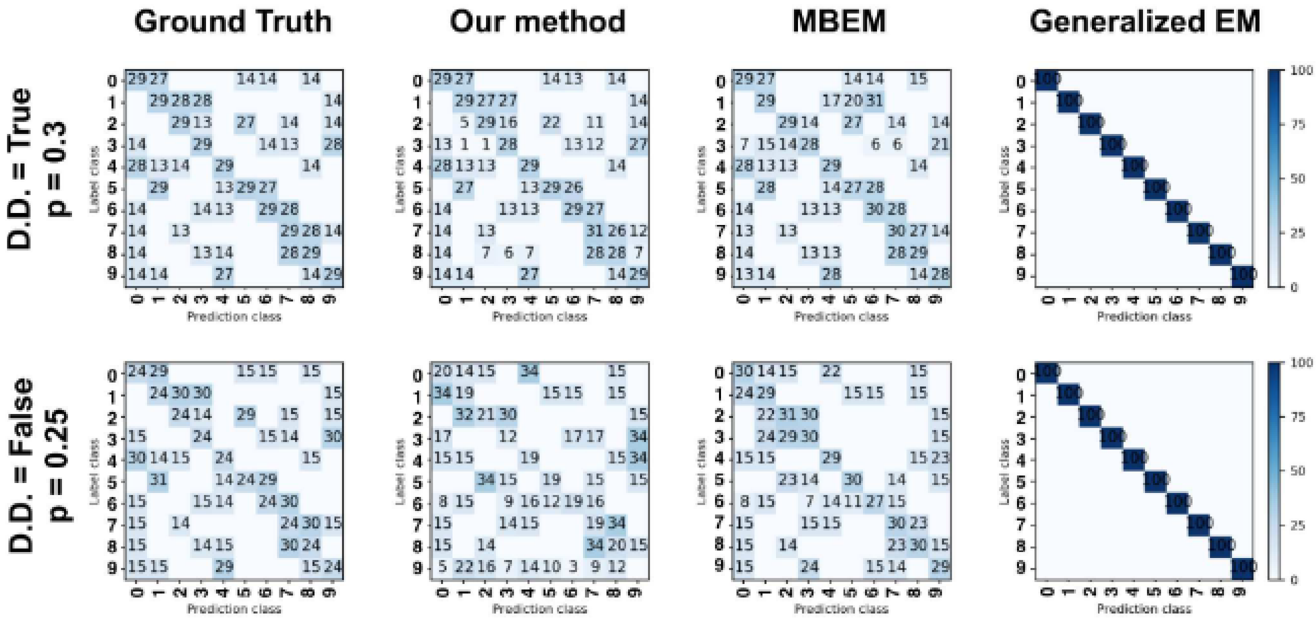


Figure 2: A diverse set of 4 simulated annotators on CIFAR-10. The top row shows the ground truths while the bottom row are the estimation from our method, trained with only one label per image.

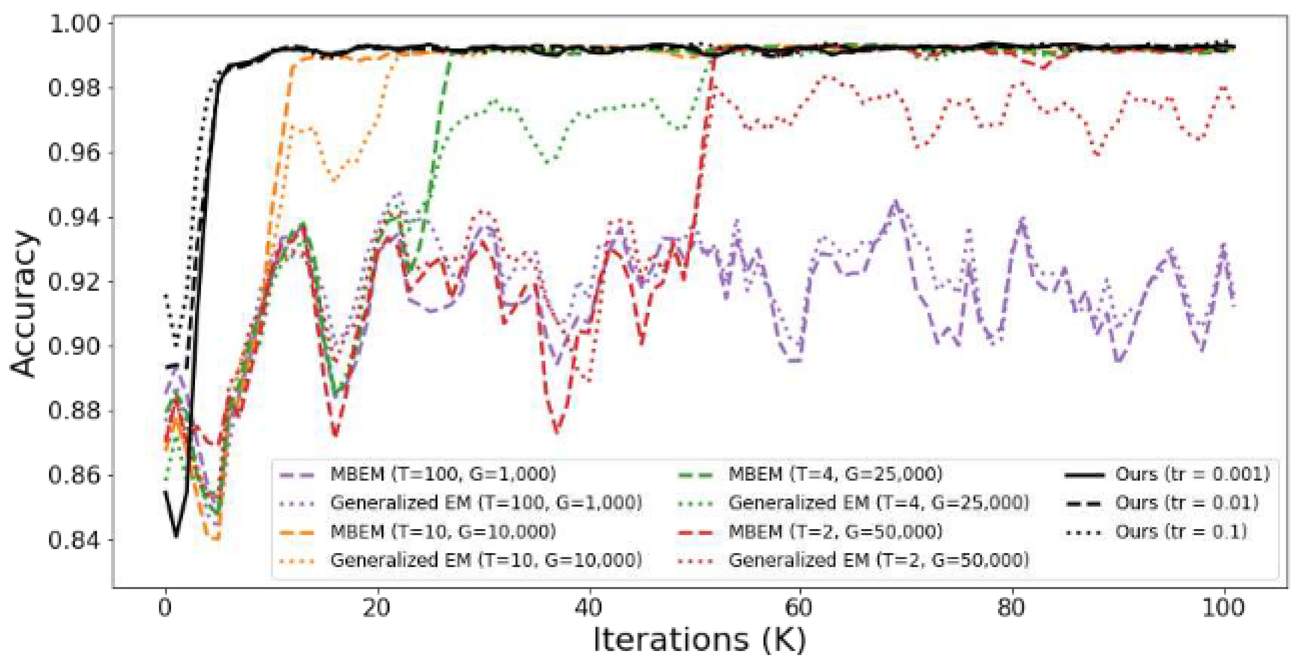
实验1：比较1 label per image的效果



CM error是指：真CM和预测CM的Frobenius norm



实验2：对超参数的敏感性



T指总的EM次数，G指每次M步梯度下降的次数

实验3：抗噪

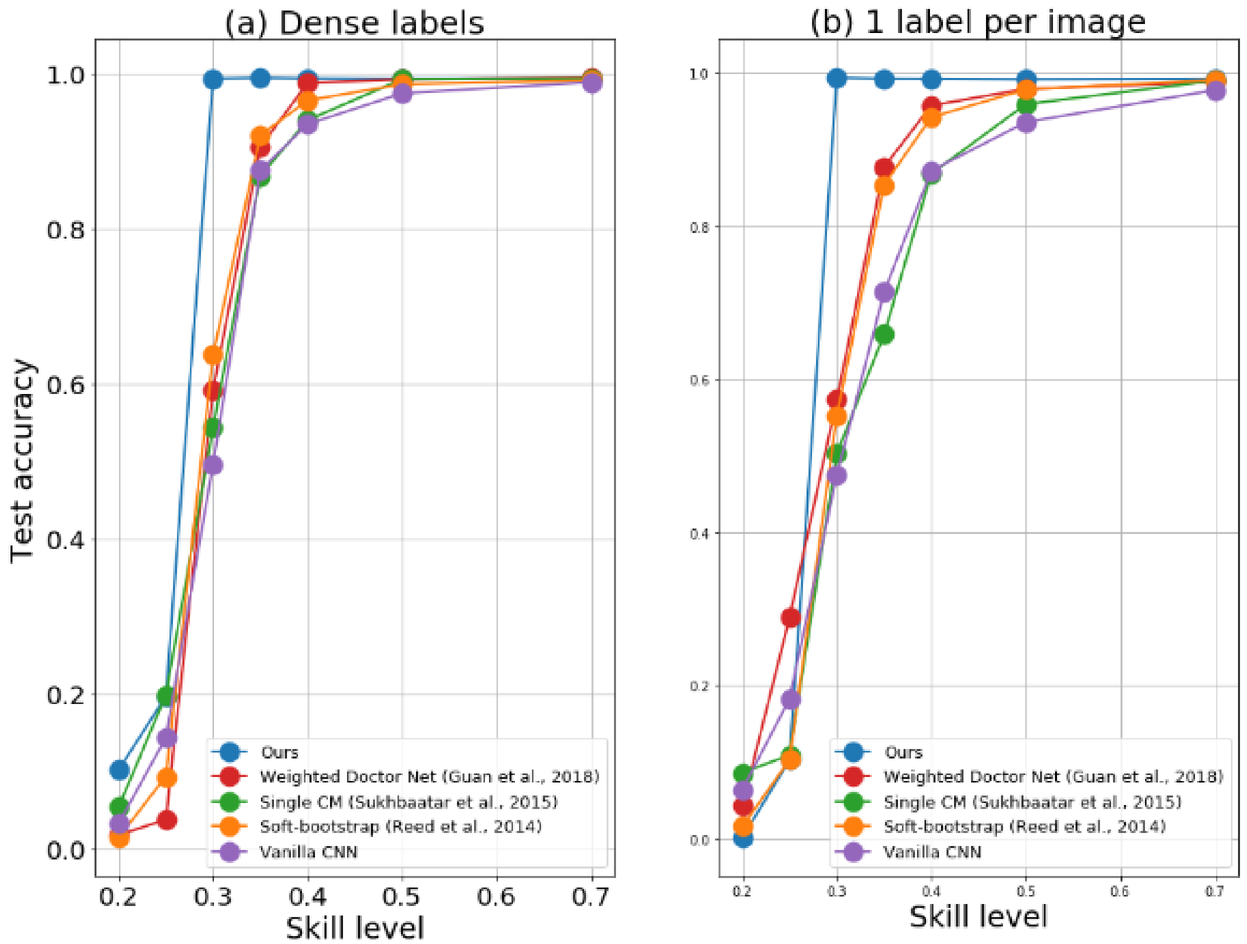
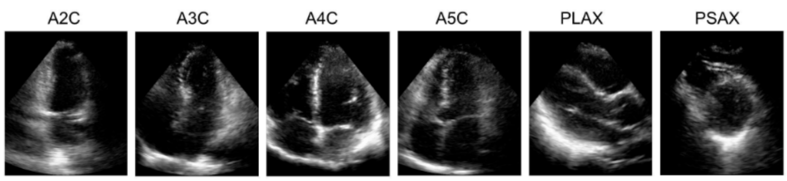
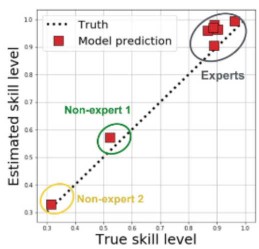


Figure 6: Classification accuracy on MNIST of different noise-robust models as a function of the mean annotator skill level p in two cases. Here, for each mean skill-level p, a group of 5 “pairwise flippers” is formed and used to generate labels. (a). each example receives labels from all the annotators. (b). each example is labelled by only 1 randomly selected annotator.

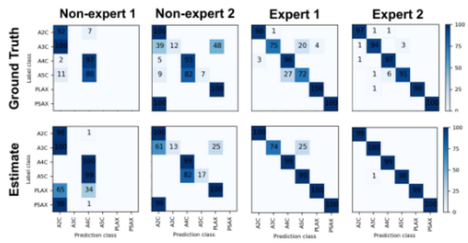
实验4：真实实验（关于心脏的超声波图找了4位标注者）



(a) Different classes of cardiac views



(b) Skill estimation



(c) Learned CMs

Method	Accuracy	CM error
Ours	$\mathbf{75.57 \pm 0.16}$	$\mathbf{11.48 \pm 0.48}$
w/o trace norm	70.99 ± 3.31	15.22 ± 0.94
MBEM [19]	73.91 ± 0.11	12.18 ± 0.29
WDN [24]	59.15 ± 1.60	-
Single-CM [22]	74.38 ± 0.29	-
Soft-bootstrap [21]	72.99 ± 0.17	-
Vanilla CNN	70.95 ± 0.44	-

(d) Performance comparison

Figure 7: Results on the cardiac view classification dataset: (a) illustrates examples of different cardiac view images. (b) plots the estimated skill level of each annotator (average of the diagonal elements of its estimated CM) against the ground truth (c) compares the estimated CMs of the two least skilled and two most skilled annotators according the GT labels (d) summarizes the classification accuracy and error of CM estimation for different methods.