This is an unofficial letter from me to let peers know better about our work. Here, I declaim:
1. In this letter, ‘we’ means ‘I’, all replies only represent my personal viewpoint, not my co-authors.
2. If there is something improper, please kindly let me know. It is greatly appreciated.

-Xinshao Wang, 15/08/2020

#2786: ProSelfLC: Progressive Self Label Correction for Training Robust Deep Neural Networks

We thank all reviewers #1, #2, #3, #4 for their comments. Some of them are detailed, constructive and insightful.

Unfortunately, we have to remark:
1. Most comments from the reviewer#1 make no sense. Please see Lines 7-15 of this letter.
2. Most comments from the reviewer#3 are harsh. Please see Lines 38-46 of this letter.

R#1 (S3-F3): "The presented result is a simple extension of previous label correcting methods...” Sorry, we absolutely disagree. Although the mathematical modelling and format are simple, our study presents some significant analysis: (1) relying on a learner’s knowledge versus human annotations; (2) rewarding or penalising a low entropy status? Please kindly read Lines 52-61, 63-69 and 232-253 in more detail, and clarify the reasons if you still disagree.

R#2 (S5-F4): "...It would be better to have a theoretical analysis that demonstrates how ProSelfLC and CCE react differently to the label noise.” Our purpose is to provide a practical way for training DNNs robustly. It is quite natural to believe that noise exists in large-scale datasets. We empirically compare their learning dynamics in Figure 3. Theoretical comparison generally requires many assumptions, which we try to avoid in this work. Please refer to Lines 60-61 and Table 1 for comparing the underlying ideas and mathematical expression analysis.

R#3 (S4-F3): "The presented result is a simple extension of previous label correcting methods...” Sorry, we absolutely disagree. Although the mathematical modelling and format are simple, our study presents some significant analysis: (1) relying on a learner’s knowledge versus human annotations; (2) rewarding or penalising a low entropy status? Please kindly read Lines 52-61, 63-69 and 232-253 in more detail, and clarify the reasons if you still disagree.

R#4 (S7-F3): "The related work could be more thorough and balanced...knowledge distillation...” We discussed knowledge distillation in related work, please see Lines 88-93. We will make it more thorough as suggested.

"CCE is undefined. I assume it's class cross entropy?" Yes, CCE = categorical cross entropy.

"Figure 3: "We remark at training, a learner is given whether an example is clean or not." How is that information used in your method?” Really sorry, it’s a typo/mistake. In fact, we would like to remark: "a learner is NOT GIVEN whether an example is clean or not.”

"(1) A simple way to reduce the proposed loss is to always produce confident predictions ...simply not change its predictions ... (2) Have you considered adding a loss that penalized large epsilons, to avoid such a degenerate solution?" (1) You are partially right here. Rewarding confident predictions is a selling point in this work, as it challenges a recently popular practice–confidence penalty. However, rewarding confidence and without changing confident predictions happens only when a model becomes confident at the later training phase; (2) Large epsilons will not lead to a degenerated solution. Epsilon becomes large when a learner becomes confident at the later training phase, which is what we desire to properly correct labels by exploiting a model’s confident knowledge.