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

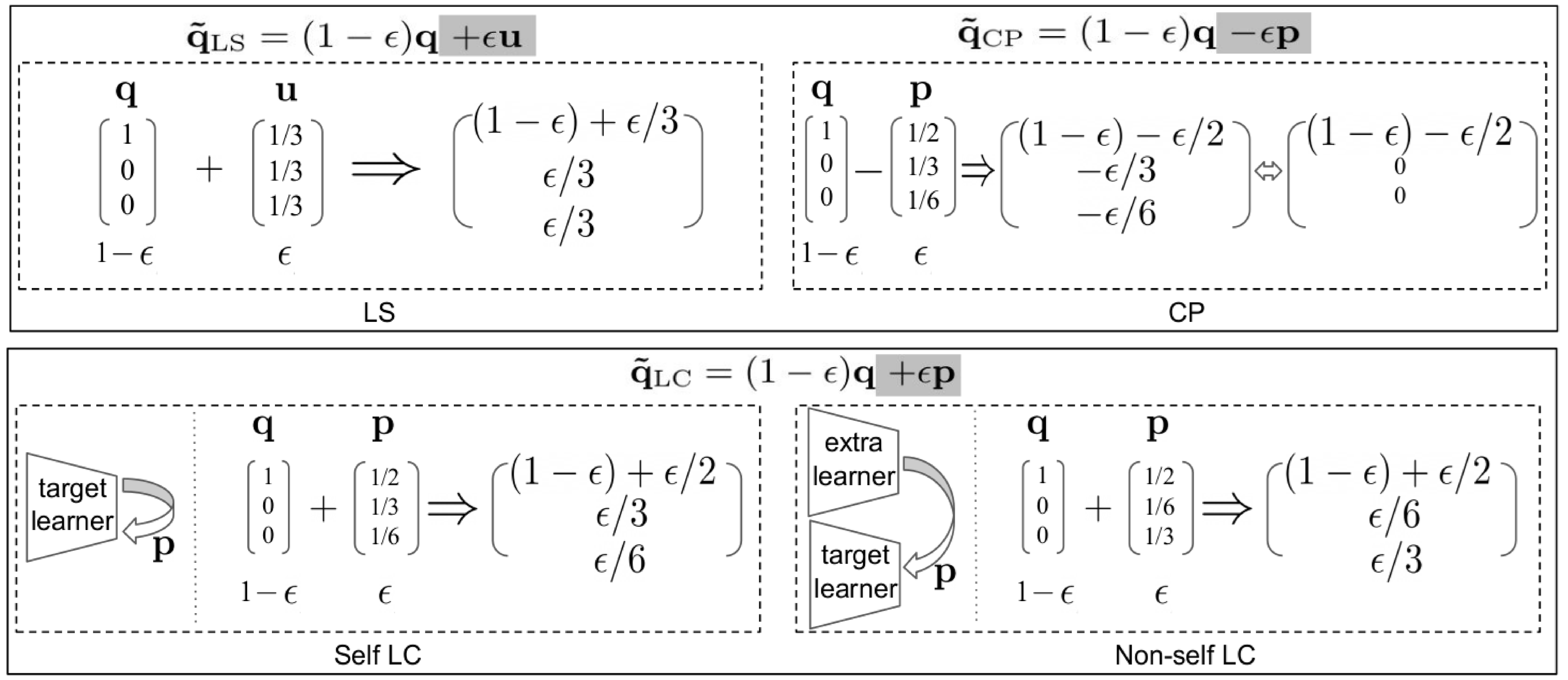
Xinshao Wang[1,2], Yang Hua[3], Elyor Kodirov[1], David A. Clifton[2], Neil M. Robertson[1,3]
[1]Zenith Ai, UK. [2]University of Oxford, UK. [3]Queen's University Belfast, UK
{xinshao, elyor}@zenithai.co.uk, {y.hua, n.robertson}@qub.ac.uk, {xinshao.wang, david.clifton}@eng.ox.ac.uk

A label distribution defines what to learn.

- Semantic class: which class has the largest probability?
- Similarity structure: how each data point is similar to different classes.

An overview of label (target) modification approaches.

- Output regularisation (OR) makes similarity structure smoothed.
- Label correction (LC) corrects semantic class and smoothes similarity structure.



Why Self LC? To exploit a model's self knowledge.

- OR methods naively penalise confident outputs without leveraging easily accessible knowledge from other learners or itself.
- Non-self LC relies on accurate auxiliary models to generate predictions.

The core research questions we study:

- In Self LC, how much should we trust a learner to leverage its knowledge?

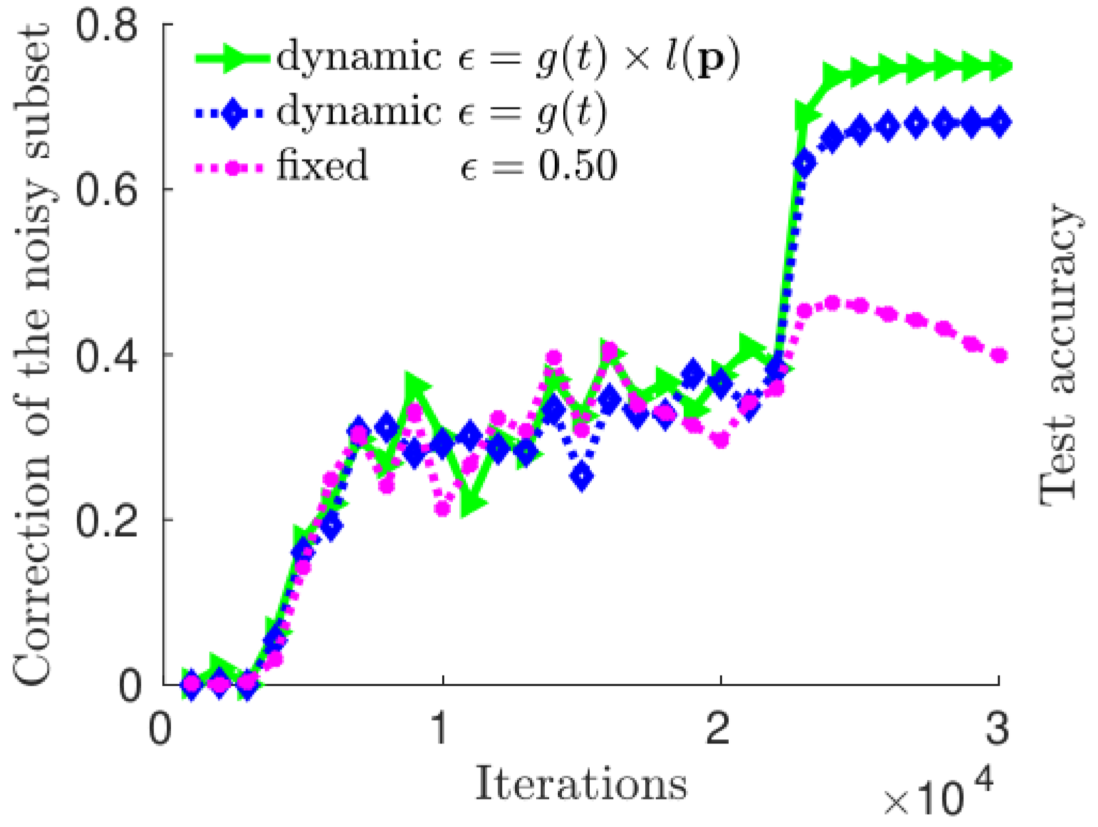
The trust score is fixed or updated stage-by-stage in prior work. ProSelfLC modifies the target **progressively, is end-to-end trainable, and requires negligible extra cost.**

- Should we penalise a low-entropy status or reward it?
  - OR methods penalise low entropy while LC rewards it.
  - ProSelfLC redirects and promotes entropy minimisation, which is in marked contrast to recent practices of confidence penalty [42, 33, 6].

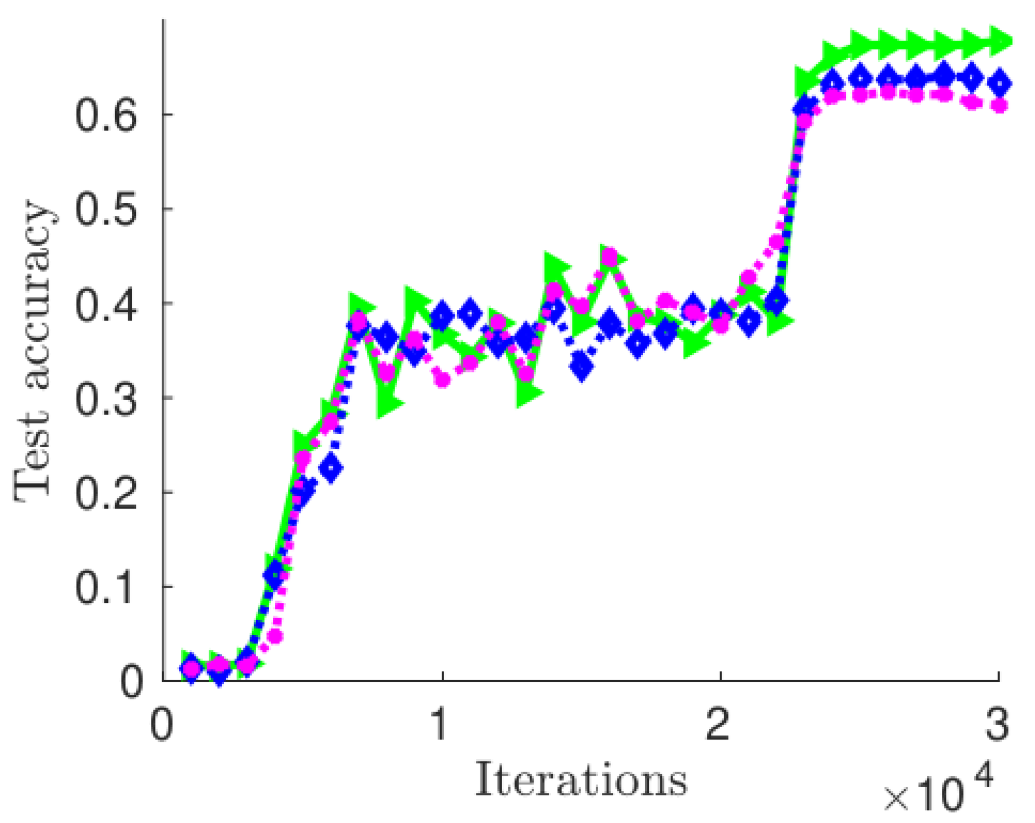
Self knowledge or predefined annotations, how to trust?

- **The design principles of ProSelfLC:**
  - **Deep neural networks learn meaningful patterns before fitting noise**: when a model learns from scratch, human annotations are more reliable than its own predictions in the early phase, during which the model is learning simple meaningful patterns before fitting noise [3].
  - **Minimum entropy regularisation**: as a learner attains confident knowledge as time progresses, we leverage it to revise annotated labels. This is surrounded by the minimum entropy regularisation [9, 10].
- **Progressive and adaptive label correction**:
  - **Global trust score $g(t)$** denotes how much we trust a learner. It is independent of data points, thus being global.
  - **Local trust score $l(\mathbf{p})$** indicates how much we trust an output distribution p, which is data-dependent.

$$\begin{cases} \text{Loss: } L(\tilde{\mathbf{q}}_{\mathrm{ProSelfLC}}, \mathbf{p}; \epsilon_{\mathrm{ProSelfLC}}) = \mathrm{H}(\tilde{\mathbf{q}}_{\mathrm{ProSelfLC}}, \mathbf{p}) \\ \qquad\qquad = \mathrm{E}_{\tilde{\mathbf{q}}_{\mathrm{ProSelfLC}}}(-\log\ \mathbf{p}). \\ \text{Label: } \tilde{\mathbf{q}}_{\mathrm{ProSelfLC}} = (1-\epsilon_{\mathrm{ProSelfLC}})\mathbf{q} + \epsilon_{\mathrm{ProSelfLC}}\mathbf{p}. \\ \epsilon_{\mathrm{ProSelfLC}} = g(t) \times l(\mathbf{p}) \begin{cases} g(t) = h(t/\Gamma - 0.5, B) \in (0,1), \\ l(\mathbf{p}) = 1 - \mathrm{H}(\mathbf{p})/\mathrm{H}(\mathbf{u}) \in (0,1). \end{cases} \end{cases}$$

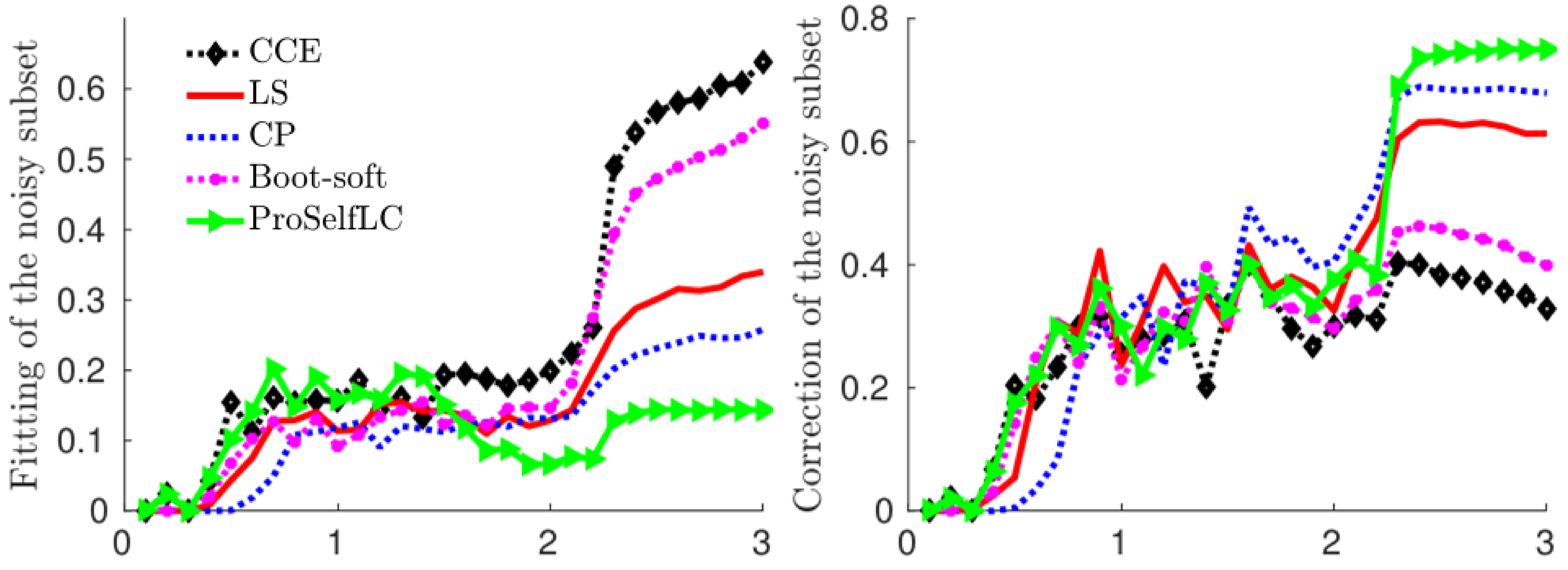


(a)Semantic class correction

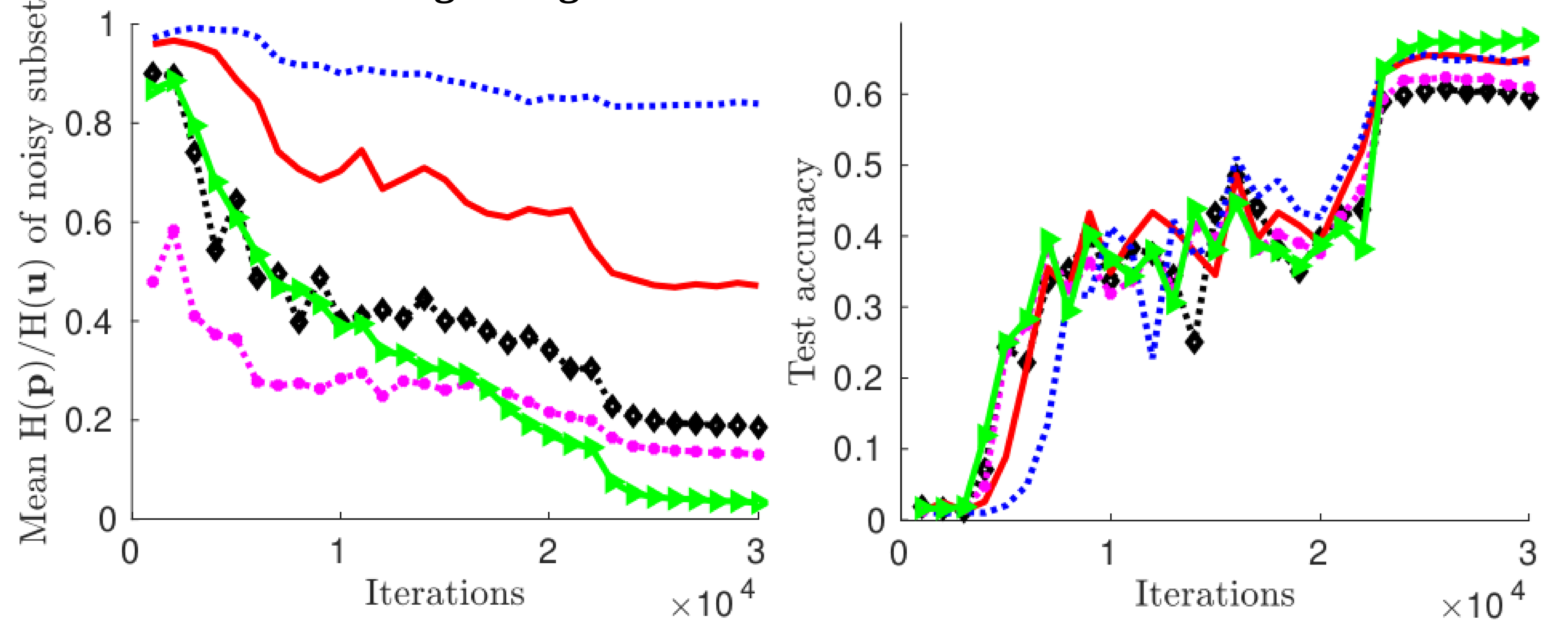


(b) Generalisation

Experimental results



Wrong fitting **Semantic class correction**



Entropy of noisy subset **Generalisation**

Summary/Conclusion

- ProSelfLC:
  - enhance the similarity structure information over training classes;
  - correct the semantic classes of noisy label distributions;
  - is the first method to trust self knowledge progressively and adaptively.
- Our extensive experiments:
  - defend the entropy minimisation principle;
  - demonstrate the effectiveness of ProSelfLC in clean and noisy settings.
- Code: https://github.com/XinshaoAmosWang/ProSelfLC-CVPR2021