

1 We thank all reviewers for providing us valuable and insightful comments. Below, we answer all of the questions.

2 **R1 Q1)** The main assumption in this method for part-dependent label noise is not realistic.

3 A1) There are lots of psychological and physiological evidences showing that we human perceive objects starting  
4 from their parts. Thus, we believe the assumption makes sense in reality. In this paper, we first learn the parts of  
5 instances using all training examples rather than only using anchor points. Then, we learn the parts-dependent transition  
6 matrices by exploiting anchor points. Lastly, the instance-dependent transition matrix can be well approximated by a  
7 combination of the parts-dependent transition matrices. We think the reason for your concern may be resulted by our  
8 description or presentation, and we will make them clearer in our final version.

9 **R1 Q2)** When the deep model is trained using the noisy labels, the features maybe not the accurate or reliable features.

10 A2) We agree with this point. The memorization effect of deep neural networks shows that they would first memorize  
11 training data of clean labels and then those of noisy labels. Therefore, we can perform early stopping with a noisy/clean  
12 validation set to prevent deep neural networks from overfitting noisy labels. Then, we can obtain relatively reliable  
13 features. This strategy is widely used in existing works such as Forward [44] and T-Revision [58].

14 **R1 Q3)** Please be more specific in references and elaborate on the conjecture of the parameters for combination.

15 A3) Thanks. We are the first to do so, and we will be more specific in discussions and elaborate the conjecture clearly.

16 **R1 Q4)** It would be nice to visualize parts for some classes of datasets used in the experiments, like Figure 1.

17 A4) Due to the page limit, we will visualize parts for some classes of datasets in our final supplementary material.

18 **R2 Q5)** The presentation of the subsection "Learning the parts-dependent transition matrices" could be further improved.

19 A5) We will take your advice and improve the presentation to make the key point clearer to readers.

20 **R2 Q6)** For the experimental results, it would be better to show whether the performance gap is significant or not.

21 A6) We have taken your advice and conducted significance tests. Almost all results were statistically significant. Due to  
22 the page limit, we will provide detailed experimental results in our final supplementary material.

23 **R3 Q7)** It will be better to provide more related work about NMF or other parts-based learning techniques.

24 A7) We will add more detailed related work about NMF or other parts-based learning techniques in this paper.

25 **R3 Q8)** How can the proposed method utilize a small trusted dataset in the experiments?

26 A8) If a small trusted dataset is available, it is helpful. It can be used to (1) better learn the parts/instance-dependent  
27 transition matrix (2) better validate the selected deep learning model, and to (3) fine-tune the deep neural networks.

28 **R3 Q9)** Please give some explanation or emphasis for data augmentation techniques.

29 A9) We don't use any data augmentation technique. We will emphasize this in the experiment settings.

30 **R3 Q10)** Whether it is possible to introduce different slack variables for instance-dependent transition matrices?

31 A10) It is possible, but we believe the current way of introducing a single slack variable to revise instance-dependent  
32 transition matrices is better. If we introduce different slack variables for different instance-dependent transition matrices,  
33 this will make optimization difficult as there are too many parameters in slack variables to learn.

34 **R3 Q11)** Please verify the effectiveness of the proposed methods through experiments on text dataset.

35 A11) We verify our approach on the text dataset *NEWS* (<http://qwone.com/~jason/20Newsgroups/>). A network  
36 architecture with three convolutional layers and one fully connected layer is employed. Other experiment settings are  
37 the same as those in this paper. We follow Co-teaching+ and borrow the pre-trained word embeddings from GloVe  
38 (<https://nlp.stanford.edu/projects/glove/>). The results are presented in Table 1. We can see that the proposed method  
39 consistently outperforms the baselines.

Table 1: Means and standard deviations of classification accuracy on *NEWS*.

	IDN-10%	IDN-20%	IDN-30%	IDN-40%	IDN-50%
CE	69.58±0.42	66.80±0.36	63.11±0.74	58.37±0.88	54.75±1.62
Decoupling	69.35±0.41	65.32±0.43	58.75±0.84	51.63±0.77	43.05±1.52
MentorNet	69.03±0.35	66.92±0.54	62.87±1.31	54.35±1.21	48.35±1.45
Co-teaching	69.37±0.29	67.99±0.76	64.15±0.89	56.36±0.71	52.32±1.03
Co-teaching+	69.35±0.73	64.03±0.91	56.37±0.61	41.88±1.74	10.78±5.87
Joint	69.73±0.51	67.45±0.49	64.54±0.74	60.67±0.83	56.72±2.10
DMI	70.35±0.62	68.01±0.45	64.28±0.61	60.73±0.62	56.33±1.35
Forward	69.24±0.45	66.01±0.55	62.07±0.58	56.33±0.71	53.25±1.43
Reweight	70.25±0.30	68.42±0.77	65.05±0.93	59.37±1.32	57.31±3.51
T-Revision	70.72±0.32	69.91±0.49	67.28±0.81	61.78±0.99	59.29±2.07
PTD-F	70.01±0.47	66.78±0.68	62.16±0.77	59.54±0.63	53.63±1.31
PTD-R	71.03±0.45	70.02±0.53	68.32±0.72	62.37±0.45	62.01±1.21
PTD-F-V	70.27±0.28	66.81±0.48	62.80±0.76	59.71±0.46	54.23±1.17
PTD-R-V	<b>71.92±0.34</b>	<b>71.33±0.42</b>	<b>69.01±0.85</b>	<b>63.17±0.58</b>	<b>62.77±0.98</b>

40 **R4 Q12)** Descriptions on the state-of-the-art  
41 approaches are not sufficient.  
42 A12) We will add descriptions on the state-  
43 of-the-art approaches in this paper.  
44 **R4 Q13)** What is the rationale of achieving  
45 the outperformance theoretically and method-  
46 ologically?  
47 A13) Existing methods focus on class-  
48 dependent noise, but in many real-world ap-  
49 plications, noise can be instance-dependent.  
50 Then the existing methods give a biased solu-  
51 tion. On the other hand, completely instance-  
52 dependent noise is too flexible and is actually not identifiable in general. In this paper, we consider an intermediate and  
53 practical situation of parts-dependent noise, which produces a less biased but still reliable solution. To be specific, there  
54 are many psychological and physiological evidences showing that we human perceive instances by decomposing them  
55 into parts. Motivated by these evidences, we learn the parts of instances and use the combination of parts-dependent  
56 transition matrices to well approximate instance-dependent transition matrix. Empirical results demonstrate that our  
57 proposed method consistently outperforms existing methods.