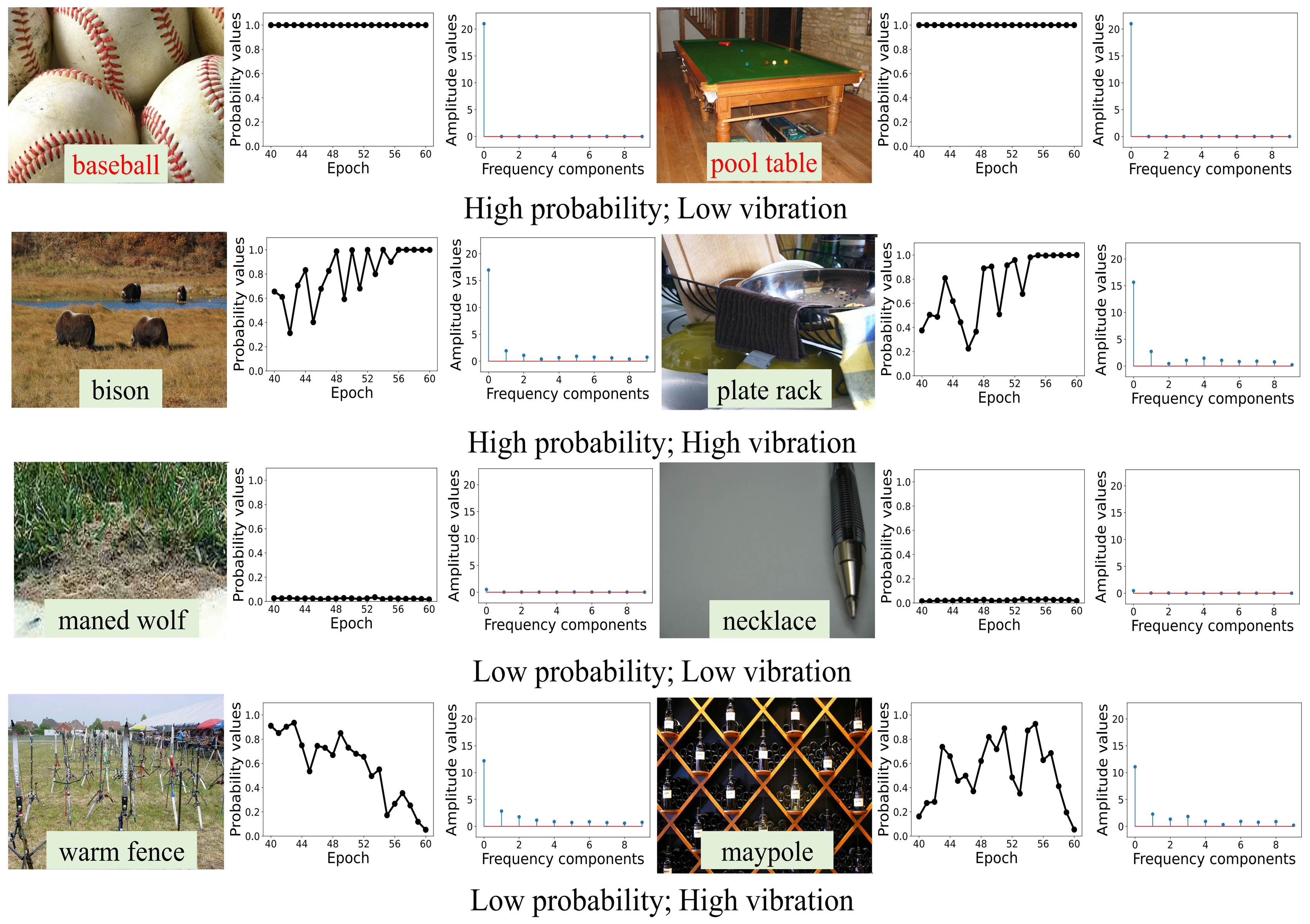
Vibration-based Uncertainty Estimation for Learning from Limited Supervision

Author names: Hengtong Hu, Lingxi Xie, Xinyue Huo, Richang Hong, Qi Tian

ABSTRACT

This paper investigates the problem of estimating uncertainty for training data, so that deep neural networks can make use of the results for learning from limited supervision. However, both prediction probability and entropy estimate uncertainty from the instantaneous information. In this paper, we present a novel approach that measures uncertainty from the vibration of sequential data, e.g., the output probability during the training procedure. The key observation is that, a training sample that suffers heavier vibration often offers richer information when it is manually labeled. Motivated by Bayesian theory, we sample the sequences from the latter part of training. We make use of the Fourier Transformation to measure the extent of vibration, deriving a powerful tool that can be used for semi-supervised, active learning, and one-bit supervision. Experiments on the CIFAR10, CIFAR100, mini-ImageNet and ImageNet datasets validate the effectiveness of our approach.



CONTRIBUTION

In this paper, we propose a novel approach which estimates uncertainty by using the sequential information from training procedure. Different from the methods that use instaneous information, our approach calculates uncertainty by measuring vibration of the obtained sequence. We utilize the Fourier Transformation to calculate the fluctuation baseline and extent to design an accurate measure. We select the appropriate sequence from training process according to Bayesian theory. To further improve this measure, we equip it with the label flipping information. We develop methods to apply this uncertainty measure to the tasks of learning from limited supervision, e.g., semi-supervised learning, active learning and one-bit supervision. In particular, we select reliable pseudo labels for SSL, select highly uncertain samples for AL, and select appropriate samples to conduct mix annotation for one-bit supervision.

ALGORITHM

Learning from limited supervision aims to utilize unlabeled data to reduce model uncertainty. Therefore, we write the objective as,

$$\mathcal{L}(\theta) = \mathbb{E}_{\mathbf{x}\in\mathcal{D}^S} l\left(\mathbf{y}_n^{\star}, \mathbf{f}(\mathbf{x};\boldsymbol{\theta})\right) + \lambda\cdot\mathbb{E}_{\mathbf{x}\in\mathcal{D}^S\cup\mathcal{D}^U}\mathbf{h}\left(\mathrm{q}, \mathbf{f}(\mathbf{x};\boldsymbol{\theta})\right).$$

Since the main idea for this task is the use strategy of unlabeled data, measuring uncertainty to distinguish each of them is very significant. We use FT to capture the vibration of the sequence, and define the uncertainty by its results:

$$v_c = \sum_{i=1}^{(L-M+1)/2} A_i - \mu\cdot A_0$$

By taking the same conduction to the label flipping sequence, we can obtain another measure which denoted as v_l. Then a fused measure is defined as,

$$v_f = (1-\alpha)v_c + \alpha v_l$$

RESULTS

We conduct experiments on CIFAR10, CIFAR100, Mini-ImageNet and ImageNet for semi-supervised learning, active learning and one-bit supervision. For SSL, we compare our approach with Mean Teacher and FixMatch. The results verify that the proposed uncertainty measure can effectively mine reliable pseudo labels to improve SSL baselines. For AL, we compare our approach with some classic AL baselines and other uncertainty estimation methods. The experimental results also show that our approach can select valuable samples to achieve better accuracy. For one-bit supervision, the results also reveal that mix annotation outperforms one-bit supervision by a large margin.

CAP: Part of the results for SSL on CIFAR10 and CIFAR100. We report the mean and standard deviation over 3 runs

	CIFAR10			CIFAR100		
Total Labels	250	1000	4000	2500	4000	10000
MT	16.50±0.18	11.72±0.10	9.48±0.29	49.83±0.10	43.86±0.56	35.60±0.36
Ours+MT	10.37±0.53	7.63±0.64	5.87±0.05	42.12±0.22	36.86±0.46	29.84±0.23
FixMatch	6.16±0.79	5.21±0.08	4.73±0.03	34.28±0.23	31.22±0.16	26.87±0.05
Ours+FM	6.06±0.76	4.84±0.04	4.63±0.12	33.53±0.21	30.40±0.20	26.18±0.12

CAP: Part of the results for AL on CIFAR100. They are all based on the Mean Teacher algorithm

	MCD	AUM	Consistency	Vibration	Vibration_fused
iter1	45.31	45.77	46.18	45.64	43.99
iter2	42.10	42.37	43.24	41.97	41.22
iter3	39.46	39.22	40.30	38.78	38.42
iter4	38.01	38.56	38.36	37.61	37.11
iter5	37.06	37.33	37.90	37.01	36.51