Supplementary Material for Flexible Visual Recognition by Evidential Modeling of Confusion and Ignorance

Lei Fan[1], Bo Liu[2], Haoxiang Li[2], Ying Wu[1] and Gang Hua[2]
[1]Northwestern University [2]Wormpex AI Research
leifan@u.northwestern.edu,yingwu@northwestern.edu,{richardboliu,lhxustcer,ganghua}@gmail.com

Abstract

This document is the supplementary material of "Flexible Visual Recognition by Evidential Modeling of Confusion and Ignorance". We provide results of flexible recognition on hybrid datasets, more synthetic experiments, the result on an imbalanced set, and more qualitative results.

1. Flexible Recognition on Hybrid Datasets

This part intends to provide a more intuitive impression of our method on a hybrid dataset that contains both closed- and open-class samples. To be more specific, the hybrid dataset is composed of closed CIFAR-10 [2] and open ImageNet (crop) [3] datasets, which simulates a potential scenario for real-world recognitions. Both two datasets contain 10000 images.

Recall the requirements of a flexible recognition system. The recognition system is supposed to reject samples that are out of the training distribution and deliver multiple predictions when being unsure. Here, we also show the behavior of the most widely used softmax-based classification model, which exhibits its potential deficiencies in a real-world recognition scenario. The results are demonstrated in Fig. 1.

The proposed method is tested with different thresholds on the accumulation of belief. In other words, the method will give a second prediction when no singleton belief meets the threshold. And the sample will be rejected when the ignorance is too large that even predicting all classes does not meet the bar. As the proposed method holds the additivity between class belief, confusion and ignorance, we could place a single belief threshold for both making multiple predictions and rejections.

	Conf.	Ign.	Cls.	Tr. Samples	Acc.
CIFAR-10 + ResNet-18	1.6	1.9	10	5000	94.6
CIFAR-100 + ResNet-18	2.1	22.1	100	500	75.0
ImageNet + ResNet-18	15.9	20.3	1000	≈1300	57.3
ImageNet + ResNet-50	15.6	11.2	1000	≈1300	61.5

Table 1: The scale of confusion and ignorance on CIFAR-10, CIFAR-100 and ImageNet datasets. The Cls. and Tr. Samples denote the number of classes and the number of training samples for each class, respectively.

2. More Synthetic Experiments

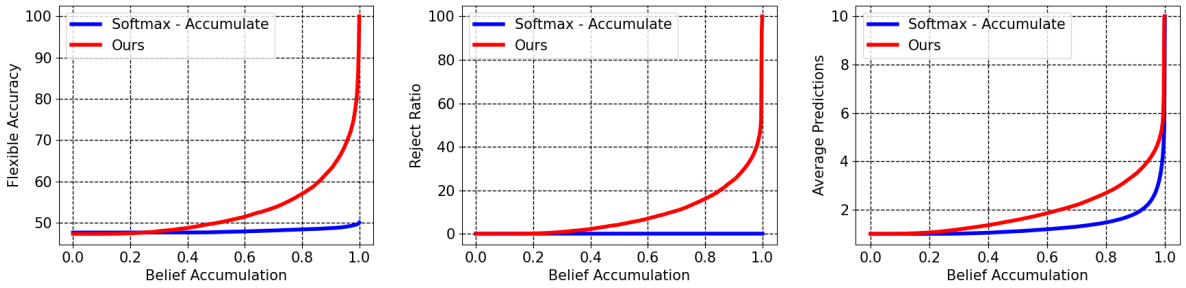
More results on synthetic data are demonstrated in Fig. 2 by changing the scale of standard deviation σ of the training Gaussian distribution. We keep the model architecture and training protocols the same as Sec. 4.1 in our main paper. The confusion estimates arise at the boundary between different classes, while the ignorance is located outside of the training distribution. And both the area of singleton beliefs and confusion develop as the σ increases.

3. Uncertainties on Imbalanced Sets

We visualize the ignorance and confusion of each class of the CIFAR-10-LT test data [1] in Fig. 3. The ignorance and confusion are the averages for each class. CIFAR-10-LT is sampled from the original CIFAR dataset with exponential distributions. The imbalance factor is set to 0.01 in our experiments. The trend of ignorance is consistent with the drop of test accuracy. Therefore, it is easy to infer training data distributions, *i.e.*, head, body, and tail classes, from the estimates of ignorance. And the confusion is not explicitly related to the test accuracy because the misclassification is mostly caused by lacking evidence, *i.e.*, unable to extract meaningful evidence because of insufficient training data.

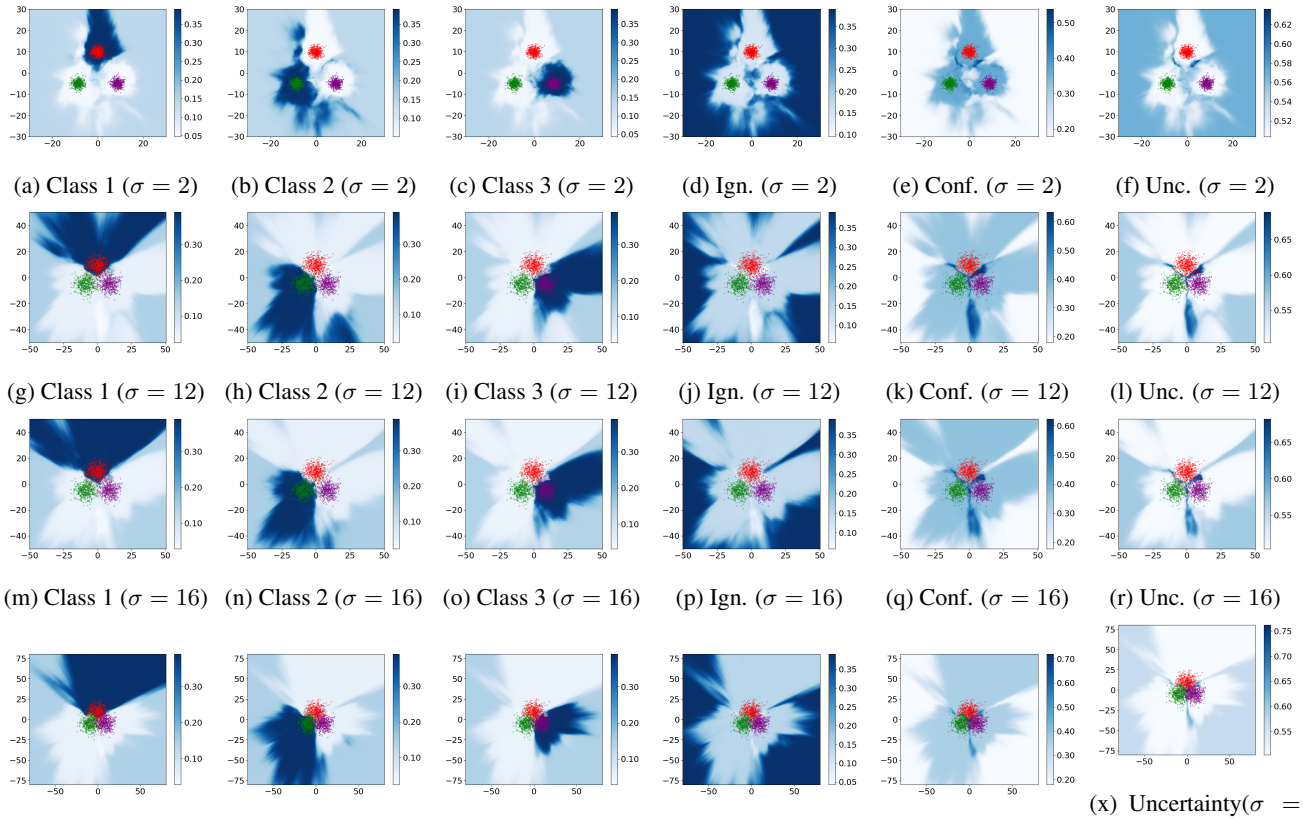
4. Qualitative Results

We provide more qualitative results for the CIFAR-10 dataset [2]. In each sample, we give the ground truth la-



(a) The accuracy on closed samples (b) The reject ratio (c) The average number of predictions

Figure 1: The behavior of our method on flexible recognition with a hybrid dataset.



(a) Class 1 ($\sigma = 2$) (b) Class 2 ($\sigma = 2$) (c) Class 3 ($\sigma = 2$) (d) Ign. ($\sigma = 2$) (e) Conf. ($\sigma = 2$) (f) Unc. ($\sigma = 2$)

(g) Class 1 ($\sigma = 12$) (h) Class 2 ($\sigma = 12$) (i) Class 3 ($\sigma = 12$) (j) Ign. ($\sigma = 12$) (k) Conf. ($\sigma = 12$) (l) Unc. ($\sigma = 12$)

(m) Class 1 ($\sigma = 16$) (n) Class 2 ($\sigma = 16$) (o) Class 3 ($\sigma = 16$) (p) Ign. ($\sigma = 16$) (q) Conf. ($\sigma = 16$) (r) Unc. ($\sigma = 16$)

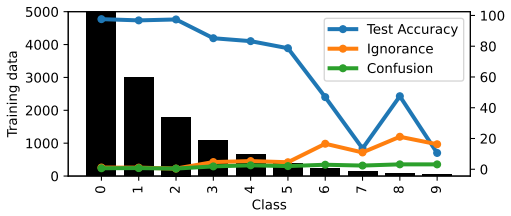
(s) Class 1 ($\sigma = 32$) (t) Class 2 ($\sigma = 32$) (u) Class 3 ($\sigma = 32$) (v) Ign. ($\sigma = 32$) (w) Conf. ($\sigma = 32$) (x) Uncertainty($\sigma = 32$)

Figure 2: More experiments on the 3-class 2D classification benchmark by setting $\sigma = 2, 12, 16, 32$ of Gaussian distributions. For each σ, we demonstrate three beliefs for singletons together with ignorance, confusion, and total uncertainty.

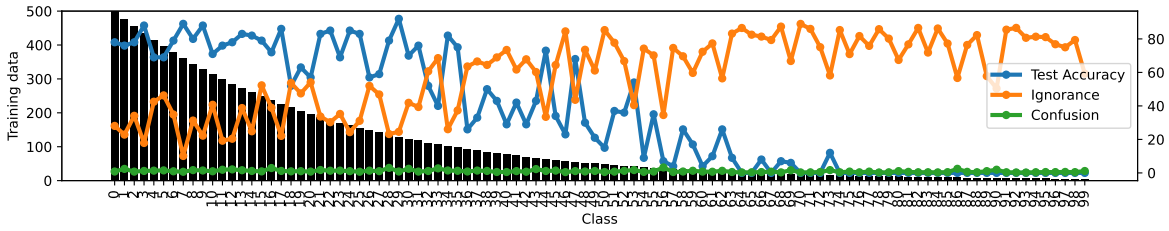
bel, the prediction, the belief, the ignorance, and the highest confusion in the captions. The diagonal of the matrix denotes the singleton belief, while the rest represents the confusion between any two classes. The results are exhibited in Fig. 4.

5. Scale of Confusion and Ignorance

In this part, we further discuss the scale of confusion and ignorance with different datasets or backbones. Please refer to Tab. 1. We find that the ignorance of the same dataset is relevant to the chosen backbone. In other words, a backbone with higher capability would generally achieve better classification accuracy and less ignorance. The scale of confu-



(a) Results on CIFAR-10-LT



(b) Results on CIFAR-100-LT

Figure 3: Data distributions, test accuracies, and average uncertainties on CIFAR-LT datasets.

sion, on the other hand, is more complicated and is affected by the number of classes and the training samples for each class. We think more classes with fewer training samples could lead to a high-confusion model after training. However, this result is only an empirical study. We leave the discussion and research of the scale as our future focus.

References

[1] Yin Cui, Menglin Jia, Tsung-Yi Lin, Yang Song, and Serge Belongie. Class-balanced loss based on effective number of samples. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9268–9277, 2019. 1

[2] Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009. 1

[3] Shiyu Liang, Yixuan Li, and Rayadurgam Srikant. Enhancing the reliability of out-of-distribution image detection in neural networks. *arXiv preprint arXiv:1706.02690*, 2017. 1

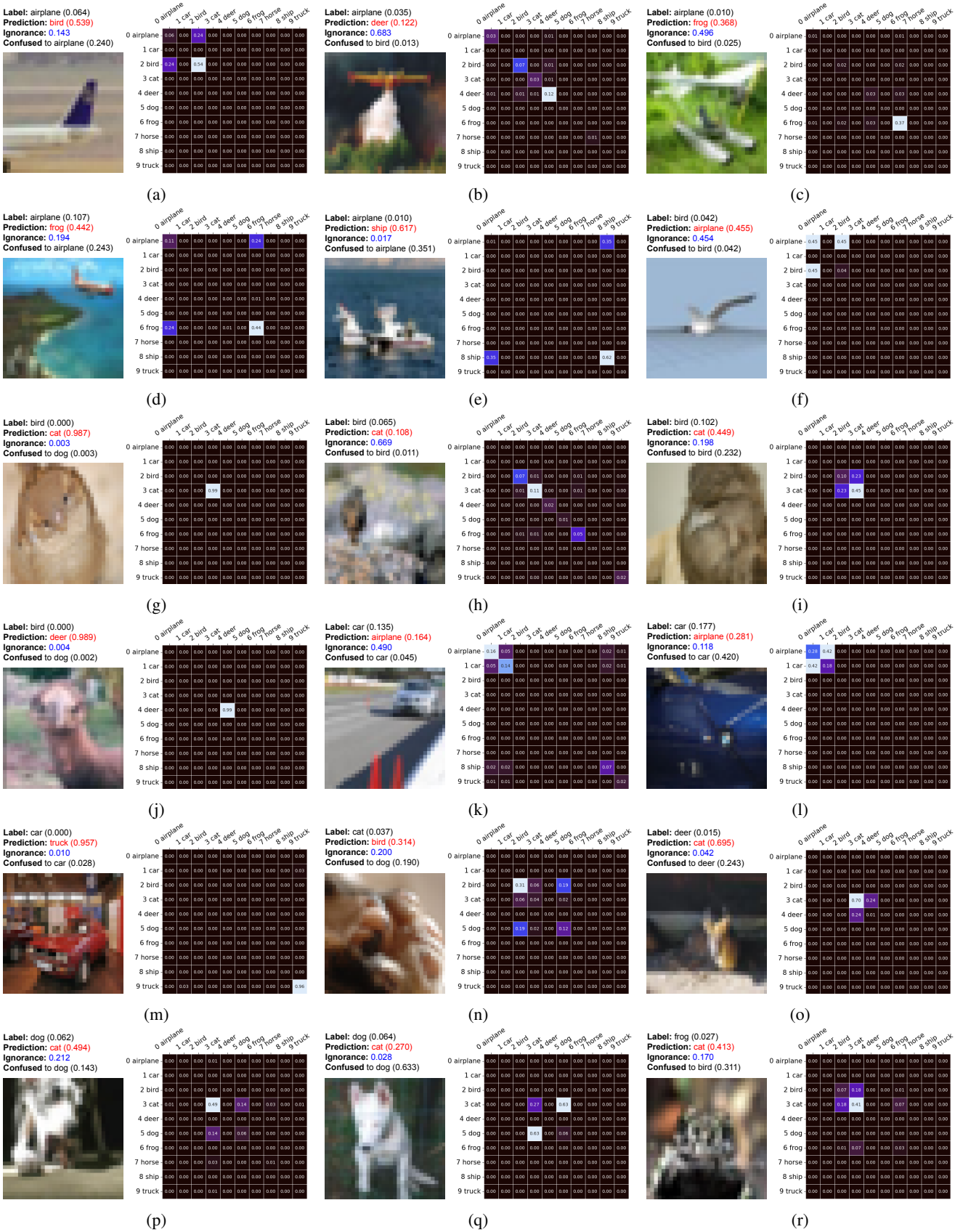


Figure 4: Matrices of confusion of misclassified samples on the CIFAR-10 dataset. The diagonal of each matrix is set to the singleton belief of each class. Notice that each heatmap is normalized individually. The total ignorance is demonstrated in the caption.