Information Bound and its Applications in Bayesian Neural Networks

GitHub Repository: AISIGSJTU/IBBNN

Jiaru Zhang, Yang Hua, Tao Song, Hao Wang, Zhengui Xue, Ruhui Ma, Haibing Guan

Shanghai Jiao Tong University, Queen's University Belfast, Louisiana State University

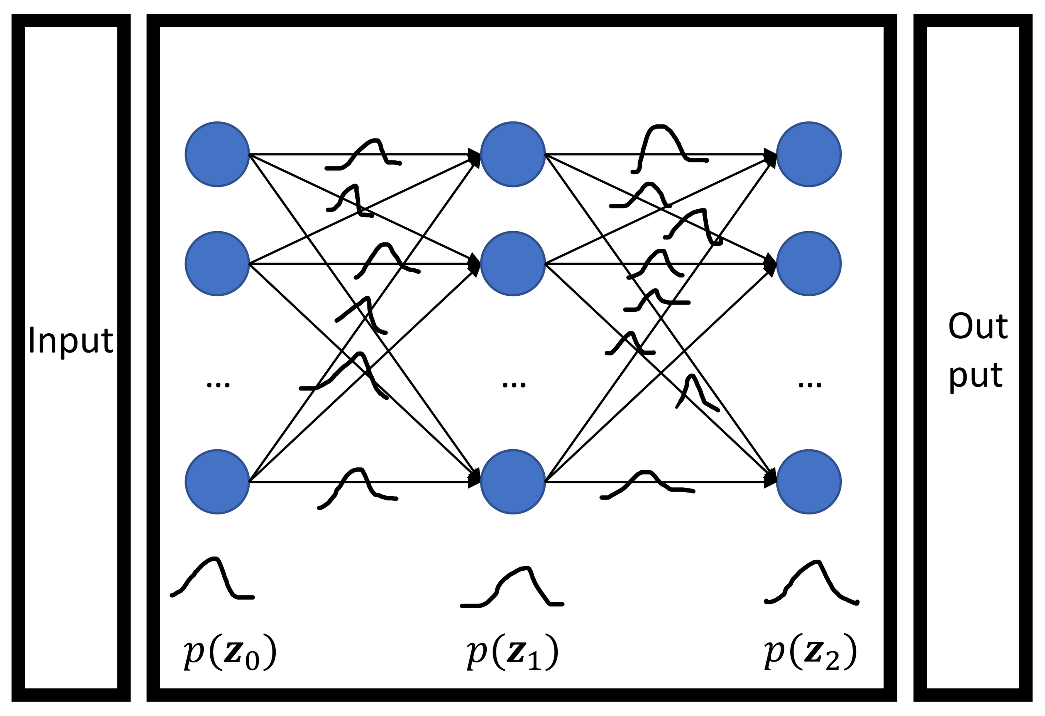
Introduction

Bayesian neural networks (BNNs) provide a natural probabilistic representation of network parameters and model predictions.

However, the relevant research on information and its applications is limited.

- We introduce Information Bound as a metric to measure the quantity of information in Bayesian neural networks. It can be easily estimated without requiring any modifications to the training process or network structure.
- We provide evidence for the existence of a "critical period" in BNNs and show that IB can be used in OOD dataset detection.
- We propose two regularization methods based on the model interpretation for better robustness and generality.

Information Bound in BNNs



Information Bound = KL($p(\mathbf{z}_1)$|| $q(\mathbf{z})$)

- The derivation of the Information Bound:

$$
\begin{aligned}
I(\mathbf{x},\mathbf{z}) &= \iint p(\mathbf{z}\mid\mathbf{x})p(\mathbf{x})\log\frac{p(\mathbf{z}\mid\mathbf{x})}{q(\mathbf{z})}dxdz \\
&\quad + \iint p(\mathbf{z}\mid\mathbf{x})p(\mathbf{x})\log\frac{q(\mathbf{z})}{p(\mathbf{z})}dxdz \\
&= \int p(\mathbf{x})KL(p(\mathbf{z}\mid\mathbf{x})\|q(\mathbf{z}))d\mathbf{x} - KL(p(\mathbf{z})\|q(\mathbf{z})) \\
&< \int p(\mathbf{x})KL(p(\mathbf{z}\mid\mathbf{x})\|q(\mathbf{z})) \\
&= \mathbb{E}_{\mathbf{x}}KL(p(\mathbf{z}\mid\mathbf{x})\|q(\mathbf{z})).
\end{aligned}
$$

- The Information Bound is an upper bound of the mutual information between the input $\mathbf{x}$ and the medium variable $\mathbf{z}$.
- It can be used to estimate the amount of information in Bayesian neural networks

- The distribution of $\mathbf{z}$ can be easily calculated.
  - Linear layer:

$$p(\mathbf{z}_i|\mathbf{x}) \sim \mathcal{N}(\sum_j M_{ij}\mathbf{x}_j, \sum_j A_{ij}^2\mathbf{x}_j^2).$$

  - Convolutional layer:

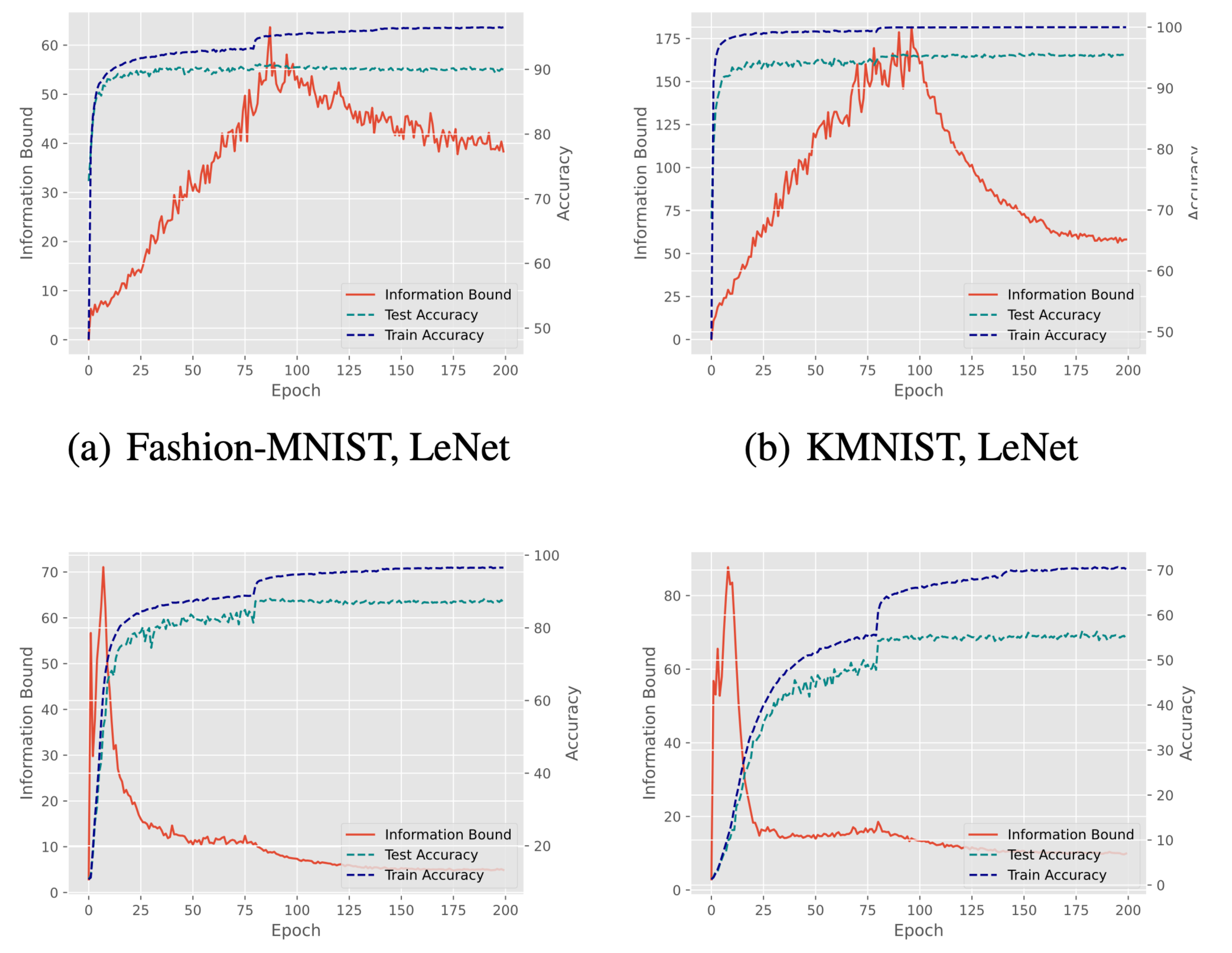
$$p(\mathbf{z}|\mathbf{x}) \sim \mathcal{N}(conv(\mathbf{x}, M), conv(\mathbf{x}^2, A^2)).$$

- Information Bound calculation:

$$
\begin{aligned}
IB(\mathbf{x},\mathbf{z}) &= KL(p(\mathbf{z}|\mathbf{x})\|q(\mathbf{z})) \\
&= -\frac{1}{2}\sum_{i=1}^{m}\Bigg(1 + \log\sum_j A_{ij}^2\mathbf{x}_j^2 - \\
&\quad \sum_j A_{ij}^2\mathbf{x}_j^2 - \sum_j M_{ij}\mathbf{x}_j\Bigg).
\end{aligned}
$$

Model Interpretation

Critical Periods in BNNs

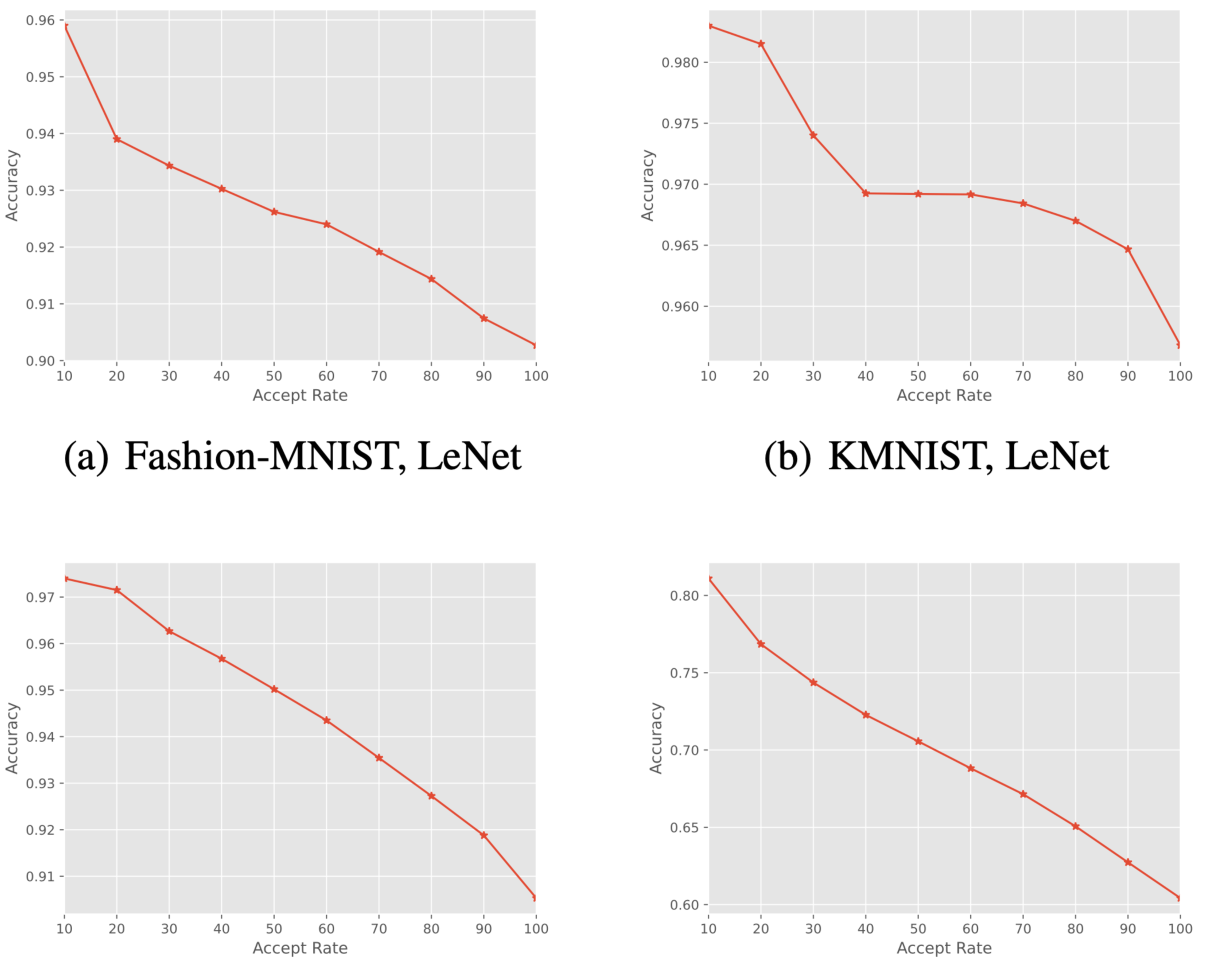


(a) Fashion-MNIST, LeNet (b) KMNIST, LeNet

(c) CIFAR-10, VGG (d) CIFAR-100, VGG

Figure 1. Trends of Information Bound and Accuracy during Bayesian neural networks training.

Model Confidence Evaluation



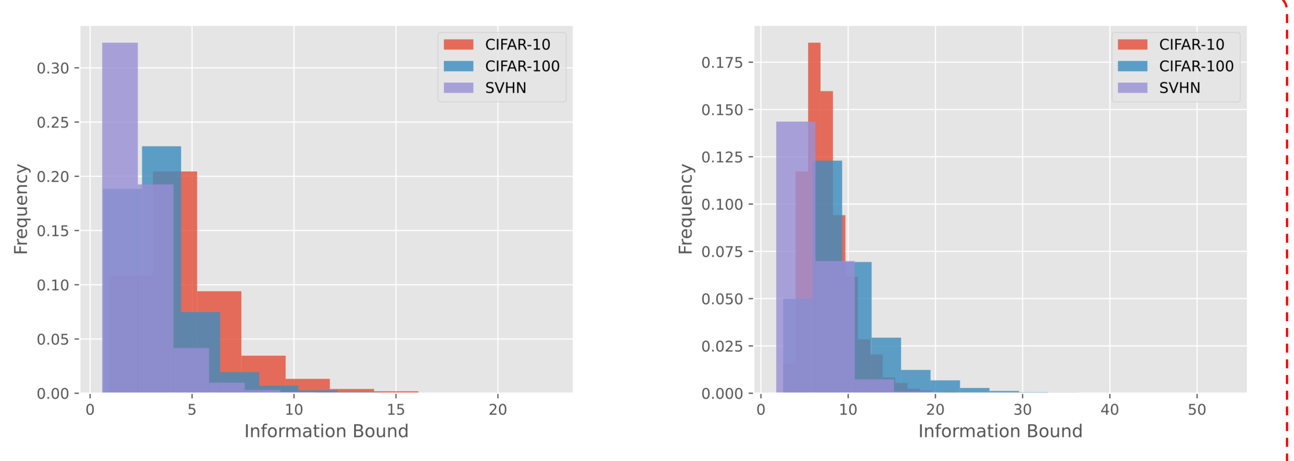
(a) Fashion-MNIST, LeNet (b) KMNIST, LeNet

(c) CIFAR-10, VGG (d) CIFAR-100, VGG

Figure 2. Trends of accuracy with varying accept rates according to Information Bounds. It verifies the effectiveness of Information Bound as a metric of model confidence.

OOD Dataset Detection

- The distribution of Information Bound for in-distribution datasets is higher.



(a) VGG trained on CIFAR-10 (b) VGG trained on CIFAR-100

Figure 3. Information Bounds of VGG models trained on CIFAR-10 and CIFAR-100 with in-distribution data and out-of-distribution dataset.

Regularization Methods

- Information Bound Regularization

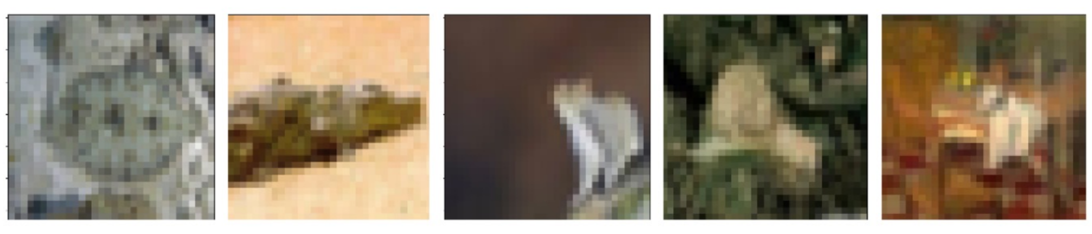
$$\mathcal{L} = \mathcal{L}_p + \mathcal{L}_r + \lambda_1 \cdot \frac{1}{n}\sum_{i=0}^{n} IB(X_i, Z_i)$$

Table 1. Comparison of models trained with Information Bound regularization and without Information Bound regularization.

Model	Dataset	Acc. w/o. IB Reg.	Acc. w. IB Reg.
LeNet	KMNIST	95.49 ± 0.26	$\mathbf{95.73 \pm 0.39}$
LeNet	Fashion-MNIST	90.48 ± 0.37	$\mathbf{90.87 \pm 0.14}$
VGG	CIFAR-10	91.03 ± 0.12	$\mathbf{91.46 \pm 0.20}$
VGG	CIFAR-100	61.06 ± 0.86	$\mathbf{62.13 \pm 0.50}$

- Information Bound Variance Regularization

$$\mathcal{L} = \mathcal{L}_p + \mathcal{L}_r + \lambda_2 \cdot Var_i(IB(X_i, Z_i))$$



Label	flatfish	crocodile	squirrel	snail	table
Old pred, IB	ray, 2.52	lizard, 2.77	possum, 2.72	snail, 2.84	table, 2.86
New pred, IB	flatfish, 7.32	crocodile, 8.12	squirrel, 5.30	porcupine, 8.12	table, 7.03

Figure 5. The five images with the lowest Information Bounds in CIFAR-100, and their labels, predictions, and Information Bounds. The model trained with Information Bound variance regularization keeps more Information and predicts more accurately.