Efficient Priors for Scalable Variational Inference in Bayesian Deep Neural Networks

Ranganath Krishnan*, Mahesh Subedar*, Omesh Tickoo

Intel Labs, USA

*equal contribution

ICCV 2019 Seoul, Korea

MOPED: informed weights priors in Bayesian DNN

Specifying meaningful weight priors in Bayesian deep neural network (DNN) is a challenging problem, particularly for scaling variational inference to larger models involving high dimensional weight space. We propose MOPED (MOdel Priors Extracted from Deterministic DNN) method to choose informed weight priors in Bayesian DNN using Empirical Bayes framework.

We illustrate with mean-field variational inference (MFVI) in Bayesian DNNs, where each weight is independently sampled from the Gaussian distribution $\mathcal{N}(\overline{\omega},\ log(1+e^{\rho}))$

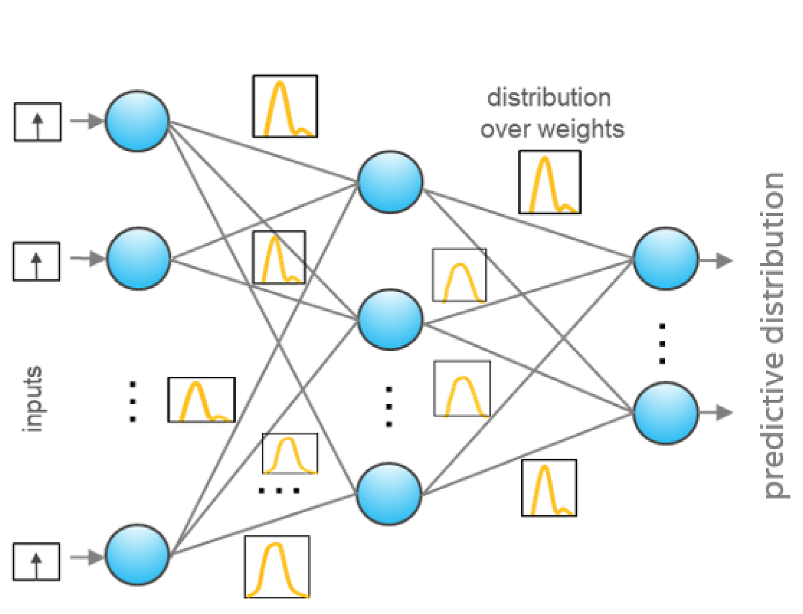
$$\overline{w} = w_d; \quad \rho \sim \mathcal{N}(\overline{\rho}, \Delta\rho)$$
$$w \sim \mathcal{N}(w_d, \log(1+e^{\rho}))$$

$$\overline{w} = w_d; \quad \rho = \log(e^{\delta|w_d|} - 1)$$
$$w \sim \mathcal{N}(w_d,\ \delta \mid w_d \mid)$$

where, ω_d represents weights obtained from standard DNN model of same architecture as BNN. δ is initial perturbation factor in terms of decimal percentage of the mean.

- The proposed method enables scaling of variational inference to larger Bayesian DNN models on complex datasets, and provides reliable uncertainty quantification without compromising on the accuracy provided by the deterministic DNNs
- We empirically evaluated the proposed approach on real-world tasks including image classification, audio classification and video activity recognition with varying complexity of Bayesian DNN architectures

Mean Field Variational Inference in Bayesian DNN



Posterior distribution:

$$p(\omega|x,y) = \frac{p(y|x,\omega)\, p(\omega)}{p(y|x)}$$

Approx. variational posterior:

$$q_\theta(\omega) \approx p(\omega \mid x,y)$$

$$q_\theta(\omega) = \mathcal{N}(\omega \mid \mu, diag(\sigma^2)) \quad ; \ \mu, \sigma \in R^D$$

Evidence lower bound (ELBO) loss:

$$\mathcal{L}_{VI} := -\,\mathbb{E}_{q_\theta(\omega)}[log\ p(y|x,\omega)] + KL[q_\theta(\omega)||p(\omega)]$$

Predictive distribution through Monte Carlo sampling:

$$p(y^*|x^*, D) = \int p(y^*|x^*,\omega)\, q_\theta(\omega) d\omega \approx \frac{1}{T} \sum_{i=1}^{T} p(y^*|x^*, \omega_i) \quad ; \ \omega_i \sim q_\theta(\omega)$$

Results

Dataset	Modality	Architecture	Bayesian DNN Complexity (# parameters)	Validation Accuracy DNN	Bayesian DNN MFVI	Bayesian DNN MOPED-MFVI
UCF-101	Video	ResNet-101 C3D	170,838,181	0.851	**0.029**	**0.867**
UrbanSound8K	Audio	VGGish	144,274,890	0.817	**0.143**	**0.819**
CIFAR-10	Image	ResNet-56	1,714,250	0.926	0.896	**0.927**
		ResNet-20	546,314	0.911	0.878	**0.916**
MNIST	Image	LeNeT	1,090,856	0.994	0.993	**0.995**
Fashion-MNIST	Image	SCNN	442,218	0.921	0.906	**0.923**

Table: Comparison of validation accuracy with varying complexity of Bayesian DNN architectures on large-scale datasets and various input modalities (image, audio, video).

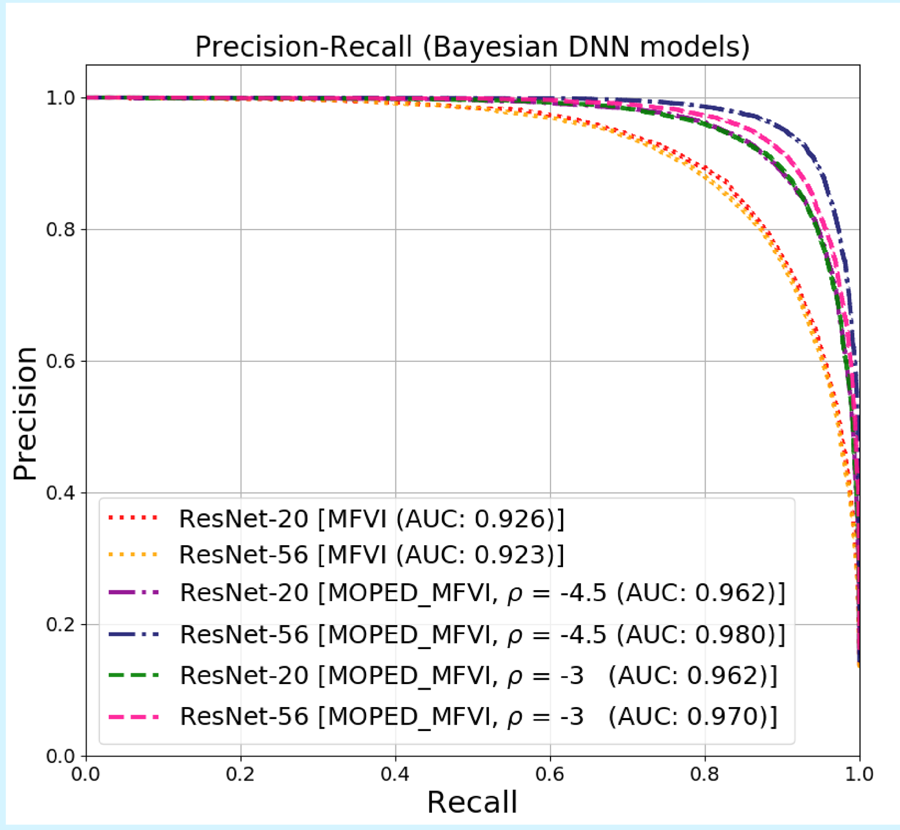


Figure 1: Precision-Recall curve

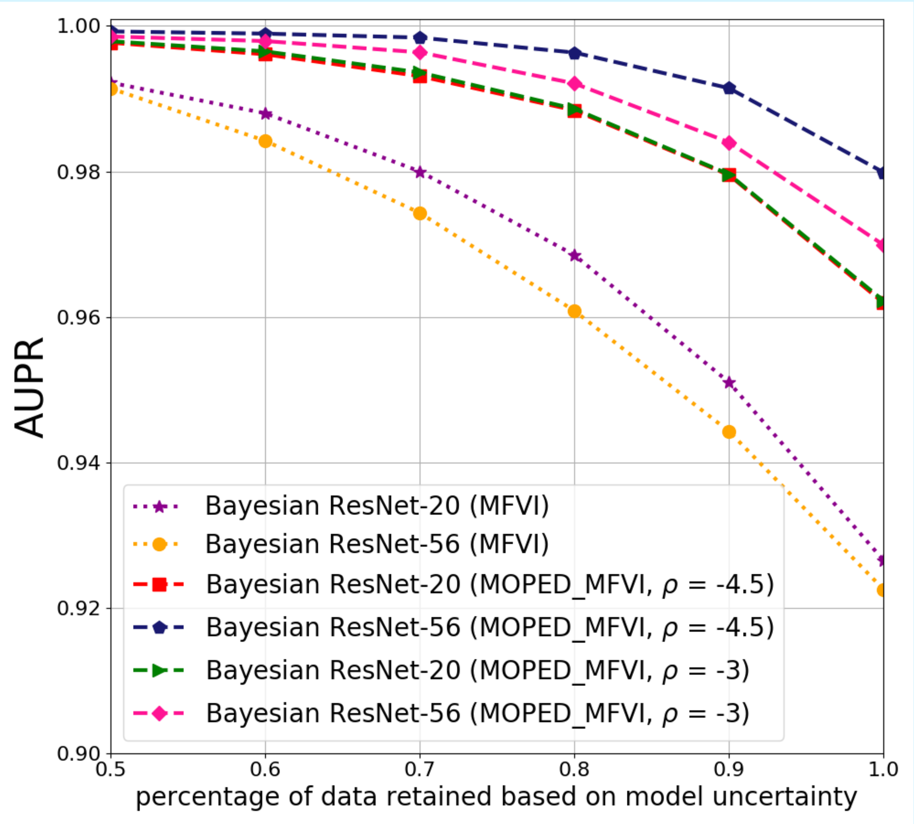
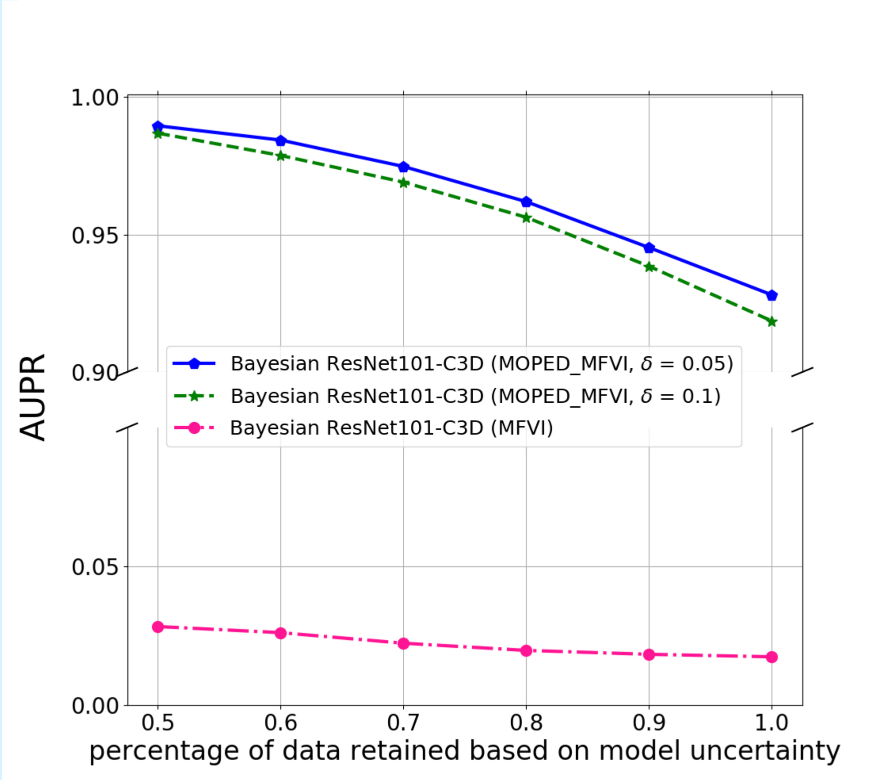


Figure 2: AUPR for different percentage of retained data based on model uncertainty

- Bayesian DNNs with random initialization of weight priors had difficulty in converging to optimal solution for the large-scale models. Whereas, MOPED method with informed priors enabled large-scale models to converge while achieving similar or better predictive accuracies as compared to standard DNN.
- Our proposed method (MOPED_MFVI) provides higher AUPR values than baseline MFVI and also AUPR increases as most uncertain predictions are ignored based on the model uncertainty, indicating our method provides better performance and reliable uncertainty estimates.

Evaluating model performance and uncertainty quantification on large-scale models



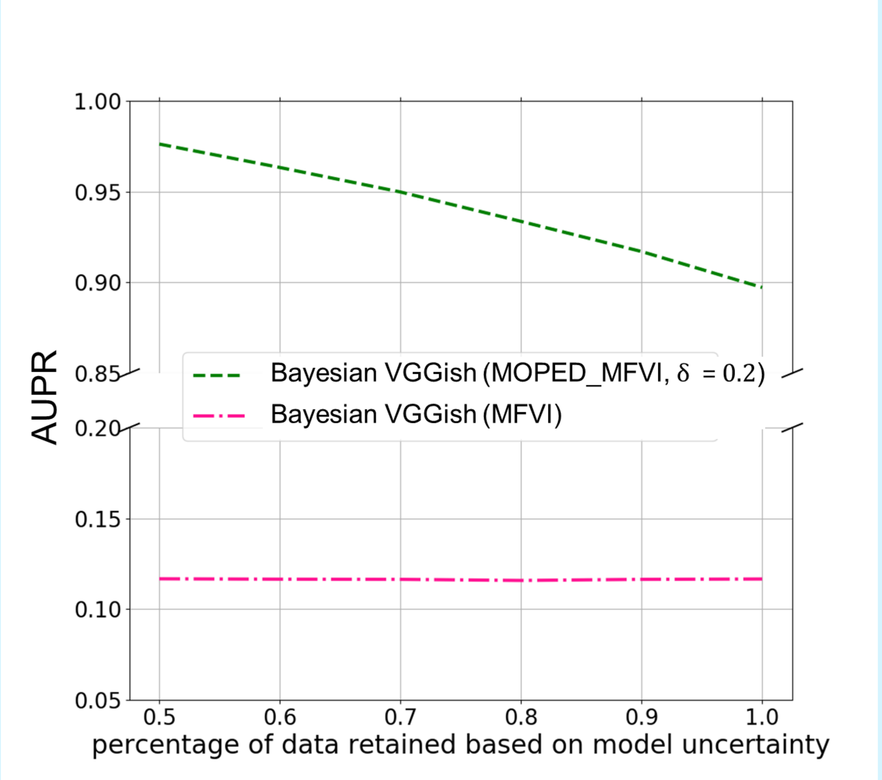


Figure 3 (a) and (b): AUPR for different percentage of retained data based on model uncertainty. (a) ResNet101-C3D on UCF101 and (b) VGGish on UrbanSound8K

Robustness to out-of-distribution (OOD) data

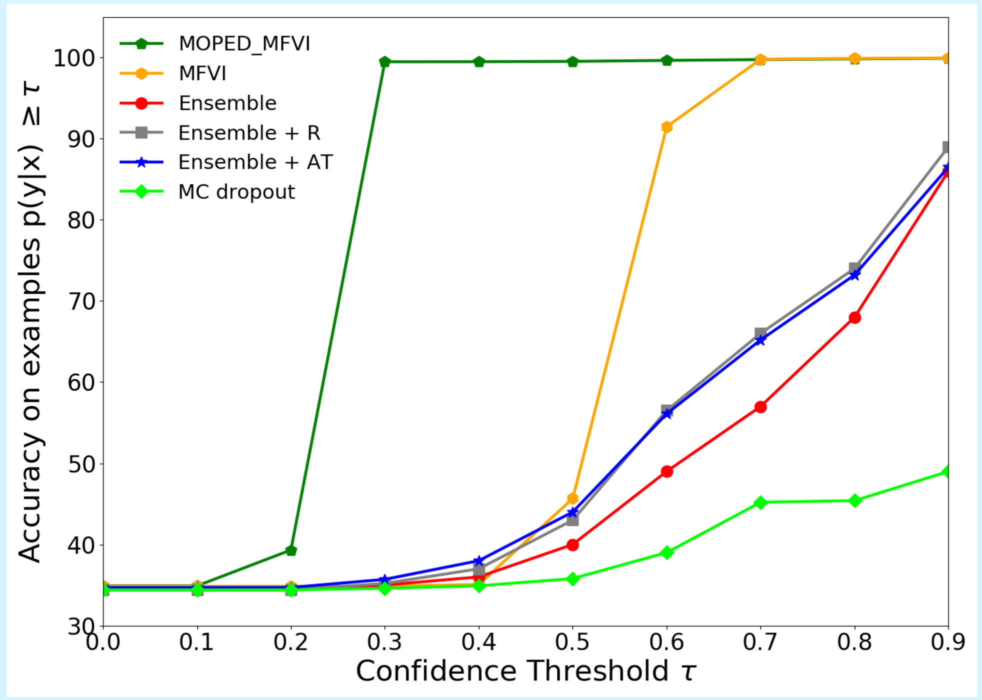


Figure 4: Accuracy vs Confidence curves while testing on both MNIST and not-MNIST (OOD) data

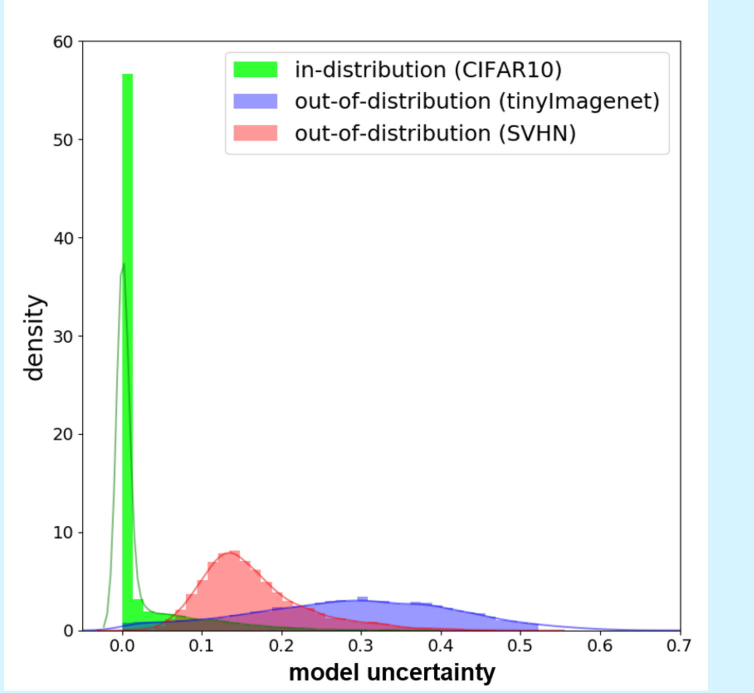


Figure 5: density histogram of model uncertainty for in-distribution and OOD

Conclusion

- Results show that proposed MOPED method based on Empirical Bayes enables scaling of variational inference to larger Bayesian DNN models. Thus offering a new method for statistical deep learning community to apply Bayesian MFVI on large-scale real-world tasks.
- Model confidence/uncertainty measures obtained from the proposed method are reliable to identify out-of-distribution data and out-performs other probabilistic deep learning methods.

Contact: Ranganath.Krishnan@intel.com, Mahesh.Subedar@intel.com