Generative Particle Variational Inference via Estimation of Functional Gradients

Neale Ratzlaff[†]
Oregon State University

Qinxun Bai[†]
Horizon Robotics

Li Fuxin
Oregon State University

Wei Xu
Horizon Robotics

Oregon State University

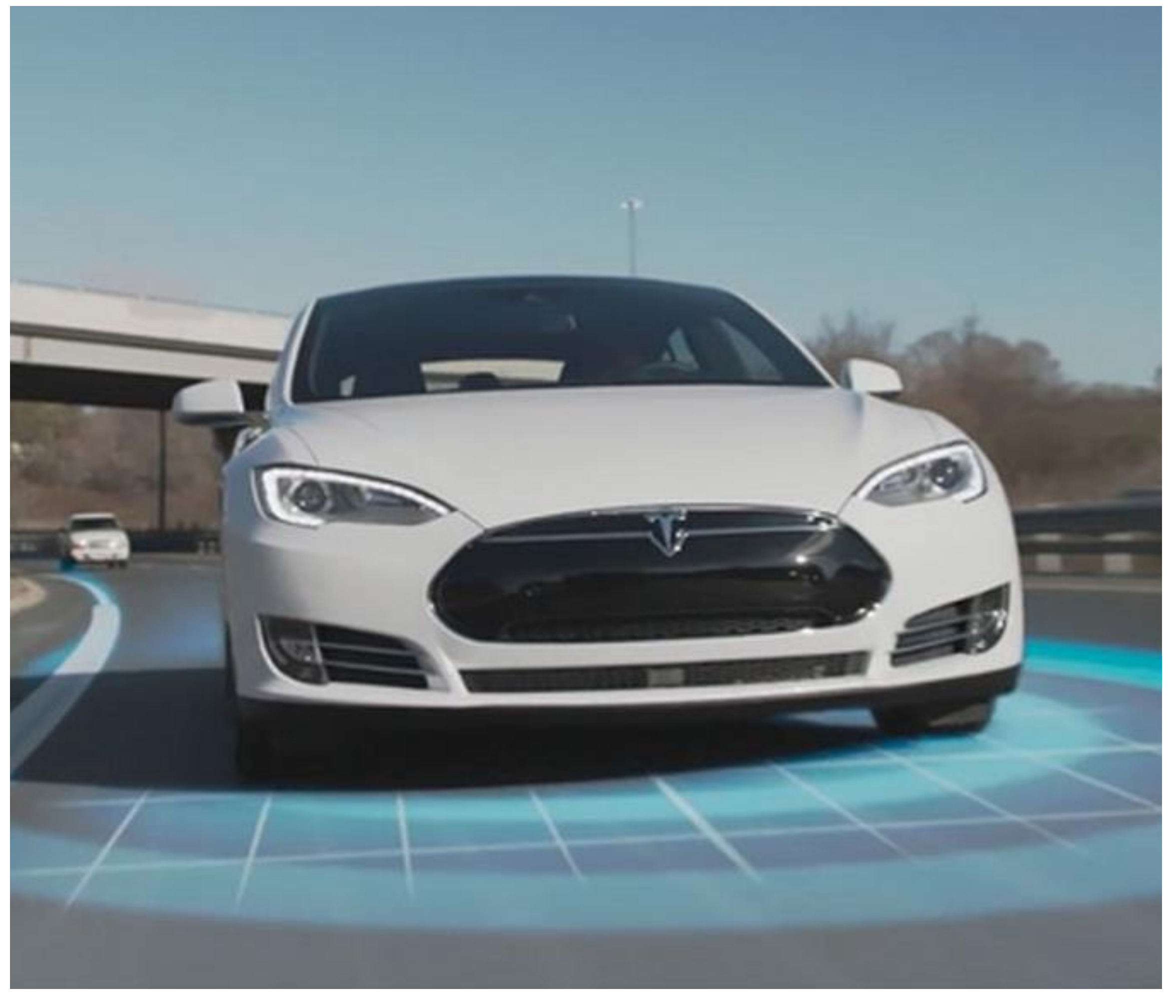
Horizon Robotics

† Equal Contribution

Epistemic Uncertainty in Deep Learning

We study the problem of variational inference (VI) for quantifying uncertainty in deep learning

By reasoning under uncertainty, we can apply deep learning to safety-critical domains



Bayesian Neural Networks

Variational Inference has proven difficult to apply to neural networks.

Prior work assumes posterior over neural network parameters from a simple family [1] [2] [3]

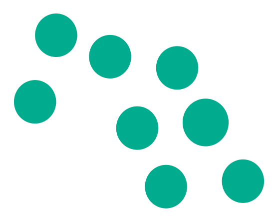
- Analytically known distributions not flexible enough to model large neural networks
- Known to underestimate epistemic uncertainty

Particle-based Variational Inference

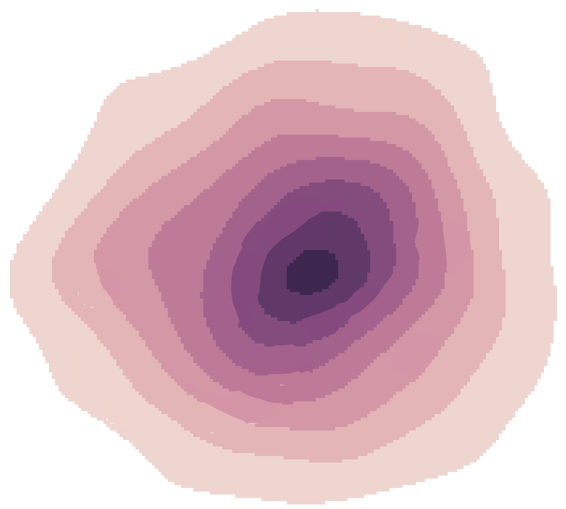
Particle-based variational inference (ParVI) is a recent nonparametric method for Bayesian inference

ParVI approximates the posterior with an empirical distribution of samples [4] [5]

- Quantify epistemic uncertainty by measuring entropy in posterior predictive distribution
- No way to draw additional samples



initial particles ~ q_0



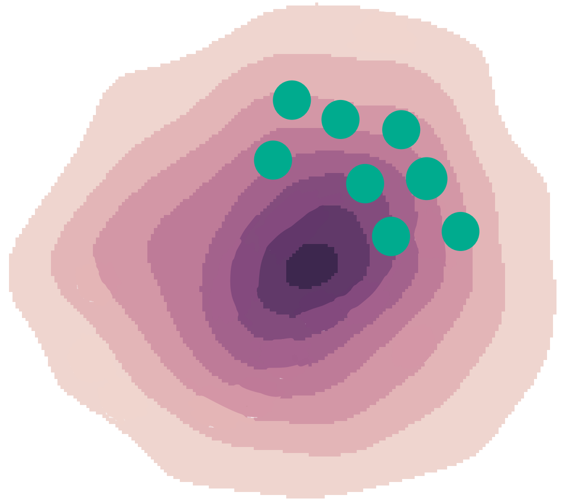
target p

Particle-based Variational Inference

Particle-based variational inference (ParVI) is a recent nonparametric method for Bayesian inference

ParVI approximates the posterior with an empirical distribution of samples [4] [5]

- Quantify epistemic uncertainty by measuring entropy in posterior predictive distribution
- No way to draw additional samples



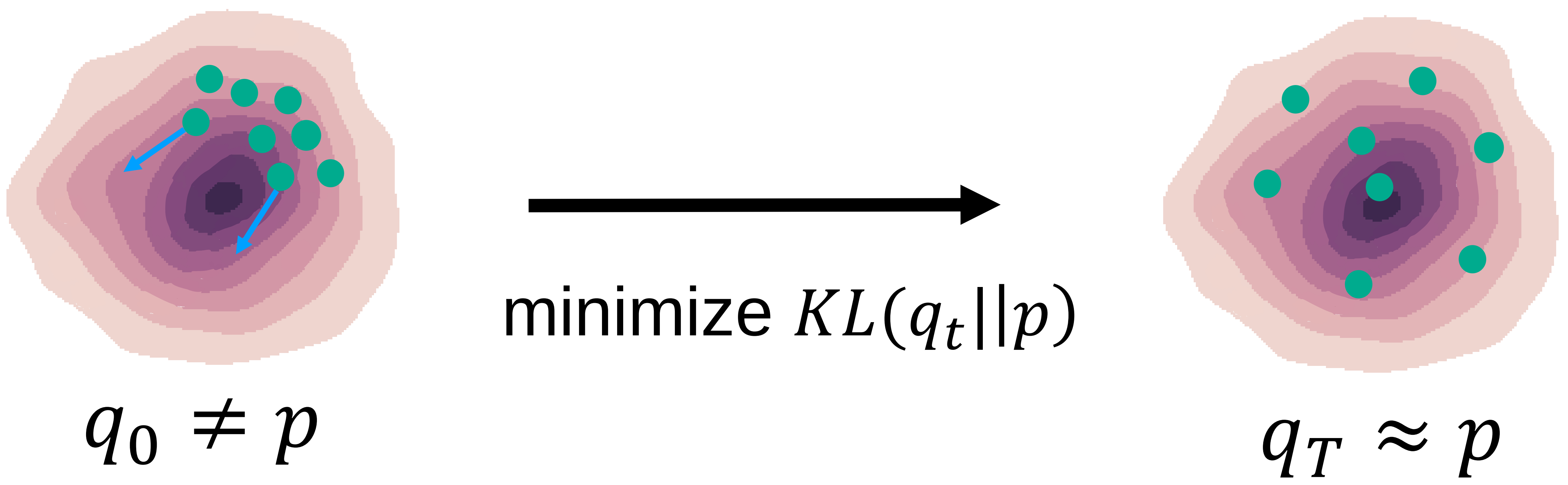
$q_0 \neq p$

Particle-based Variational Inference

Particle-based variational inference (ParVI) is a recent nonparametric method for Bayesian inference

ParVI approximates the posterior with an empirical distribution of samples [4] [5]

- Quantify epistemic uncertainty by measuring entropy in posterior predictive distribution
- No way to draw additional samples

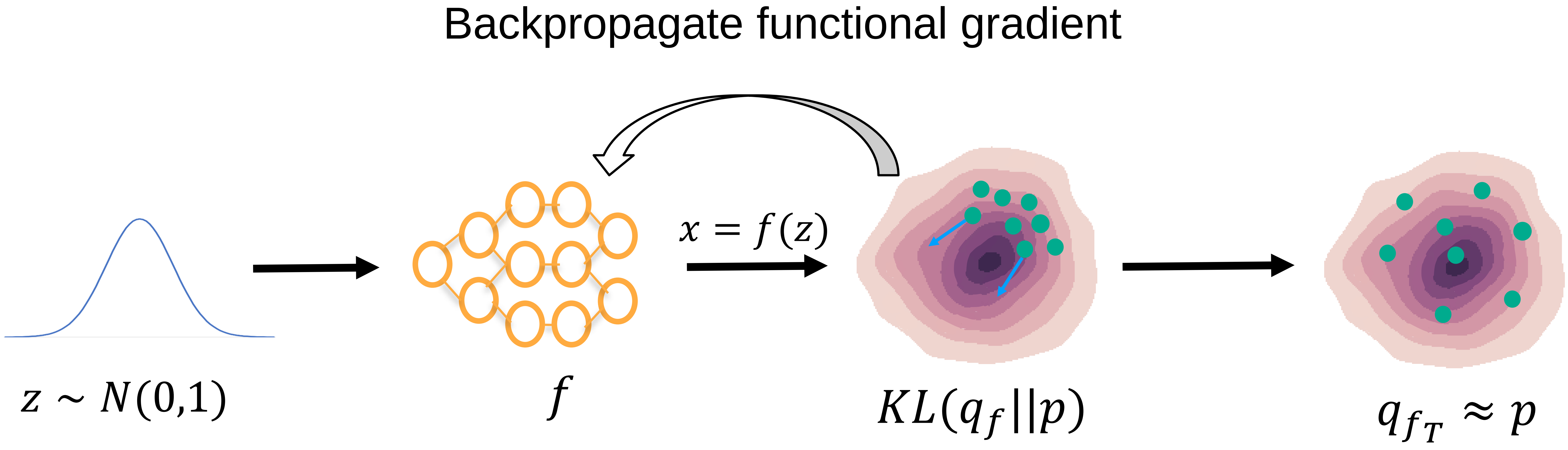


Generative Particle Variational Inference

We propose GPVI: a generative counterpart to particle based VI

The generator f minimizes $KL[q_f(x)||p(x)]$, where $q_f(x)$ is the generated distribution

To apply GPVI on BNNs, the generator outputs weight vectors for NNs



Functional Gradient of $\mathcal{J}(f) = KL[q_f || p]$

The functional gradient $\nabla_f \mathcal{J}(f)$ tells us how we should change f to fit $p(x)$

We can express this in closed form when f is from an RKHS with kernel k

$$\nabla_f \mathcal{J}(f)(z) = \mathbf{E}_{z'}\left[\underbrace{-\nabla_x \log p(x)\Big|_{x=f(z')}}_{\text{Log-likelihood}} k(z', z) - \underbrace{\left(\frac{\partial f}{\partial z'}\right)^{-1} \nabla_{z'} k(z', z)}_{\text{Repulsive Term}} \right]$$

To update the parameters θ of f , we backpropagate the functional gradient to θ

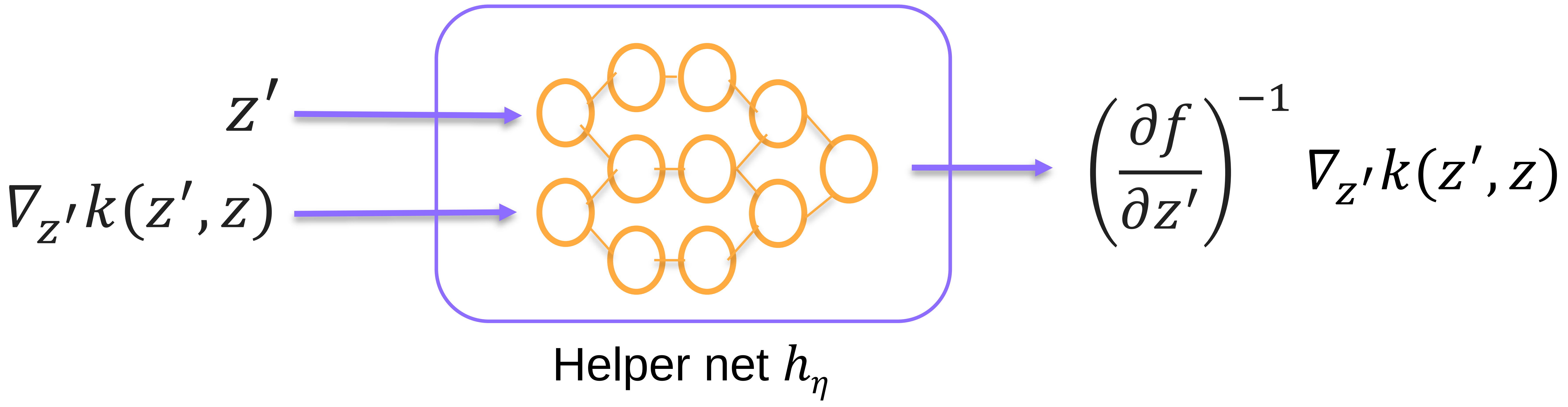
$$\nabla_\theta \mathcal{J} = \mathbf{E}_z \left[\frac{\partial f(z)}{\partial \theta} \nabla_f \mathcal{J}(f)(z) \right]$$

Estimating the Repulsive Term

The repulsive term $\left(\frac{\partial f}{\partial z'}\right)^{-1} \nabla_{z'} k(z', z)$ is difficult to compute due to Jacobian inverse

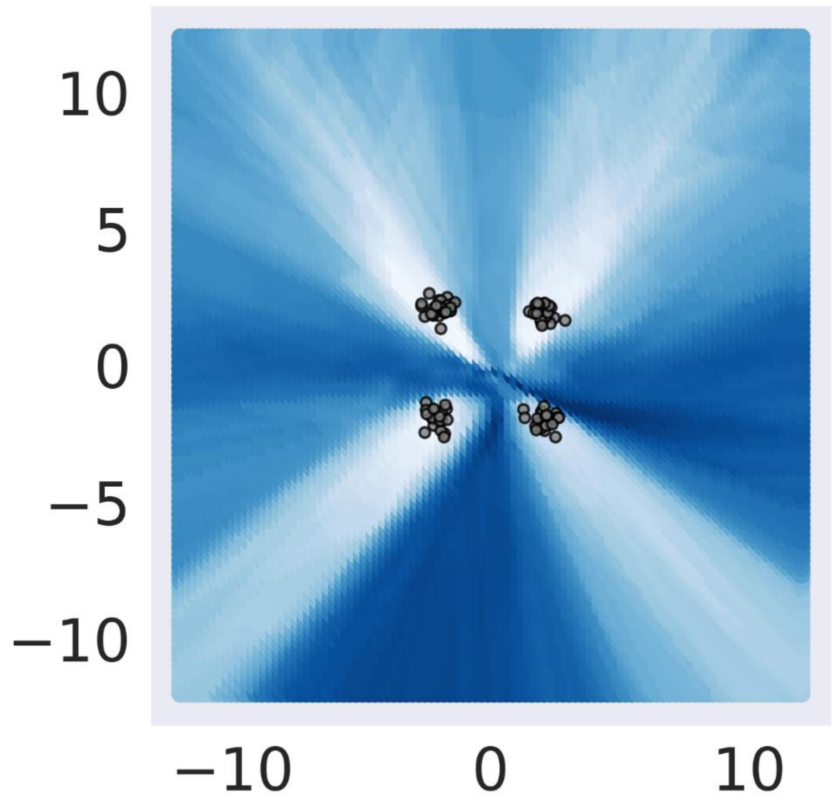
We use a helper network h_η to predict $\left(\frac{\partial f}{\partial z'}\right)^{-1} \nabla_{z'} k(z', z)$

Train with 1 step of gradient descent, per training step of f

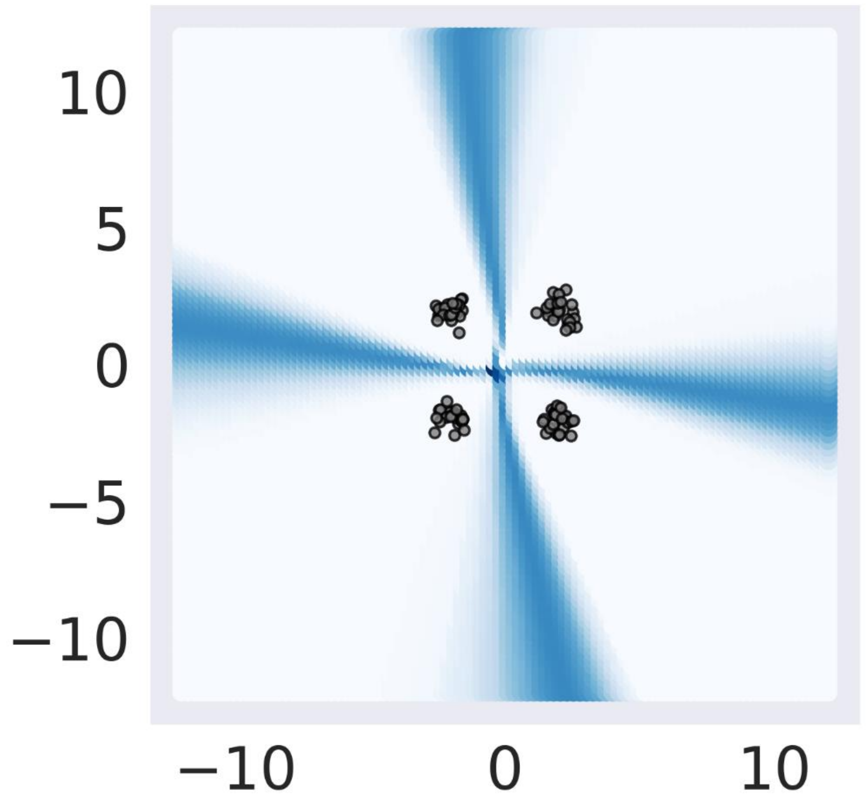


Bayesian Neural Networks: Classification

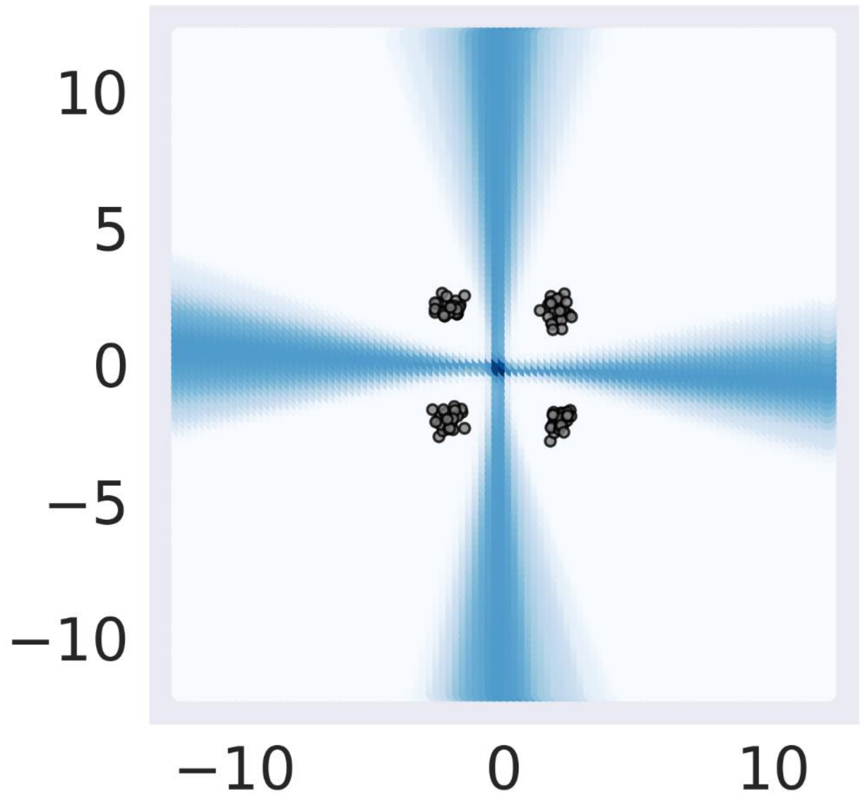
- GPVI: sampled classification functions have intuitive predictive uncertainty



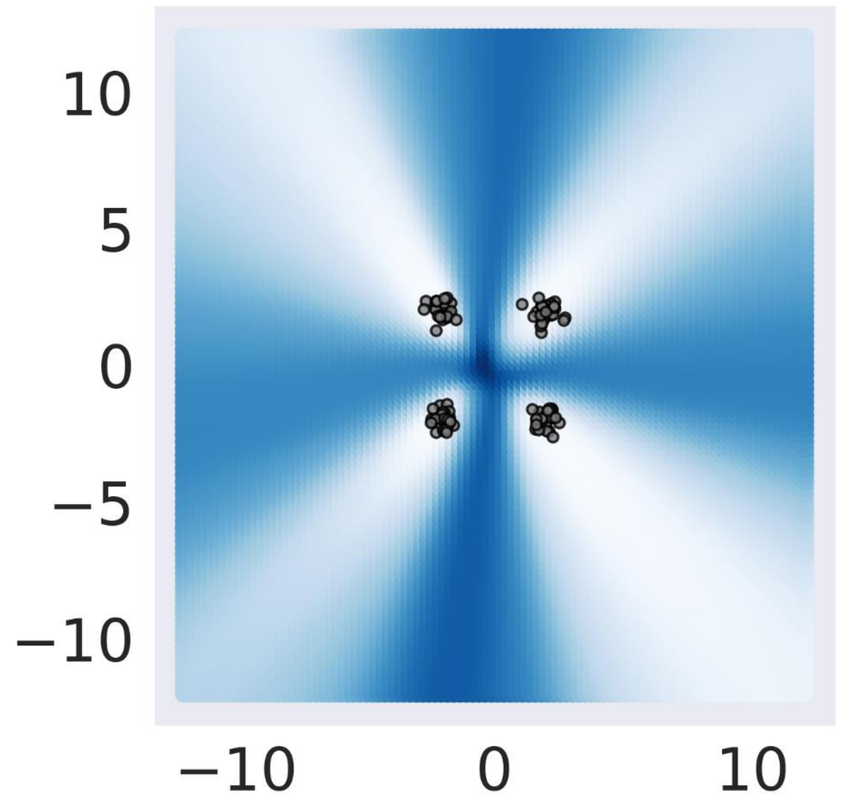
GPVI (Ours)



Amortized SVGD [6]



Deep Ensembles [7]



HMC [8]

More Uncertain  **Less Uncertain**

Bayesian Neural Networks: Open Category Prediction

Open Category Prediction: detect new classes unseen during training

- MNIST & CIFAR-10.
- 6 training classes, 4 evaluation classes

MNIST

Method	Clean	AUC↑	ECE ↓
SVGD	99.3	**.989** $\pm$ **.001**	**.001** $\pm$ **.0002**
GFSF	99.2	**.988** $\pm$ **.003**	.002 $\pm$ .0003
KSD	97.7	.964 $\pm$ .005	.014 $\pm$ .0007
Amortized SVGD	99.1	.958 $\pm$ .015	**.002** $\pm$ **.0007**
Amortized GFSF	99.2	.978 $\pm$ .005	.004 $\pm$ .0013
Amortized KSD	97.7	.951 $\pm$ .008	.017 $\pm$ .0010
MF-VI	98.6	.951 $\pm$ .008	.014 $\pm$ .0027
Deep Ensemble	99.3	.972 $\pm$ .002	.008 $\pm$ .0060
GPVI	99.3	**.988** $\pm$ **.001**	**.001** $\pm$ **.0005**

CIFAR-10

Method	Clean	AUC ↑	ECE ↓
SVGD	80.3	**.683** $\pm$ **.008**	.055 $\pm$ .004
GFSF	80.6	**.681** $\pm$ **.004**	.068 $\pm$ .012
Amortized SVGD	71.12	.636 $\pm$ .018	.073 $\pm$ .029
Amortized GFSF	71.09	.583 $\pm$ .007	.042 $\pm$ .029
MF-VI	70.0	.649 $\pm$ .006	**.016** $\pm$ **.002**
Deep Ensemble	73.54	.652 $\pm$ .018	.033 $\pm$ .011
GPVI	76.2	.677 $\pm$ .008	**.018** $\pm$ **.015**

Summary of Contributions

- GPVI is a new method for approximate Bayesian inference
- We retain the asymptotic accuracy of particle-based VI, also allow sampling
- Helper network allows for generators with flexible architectures
- Competitive uncertainty estimation for Bayesian neural networks
  - Classification, Regression, Open-category prediction

Thank you!

Come view our poster in GatherTown!

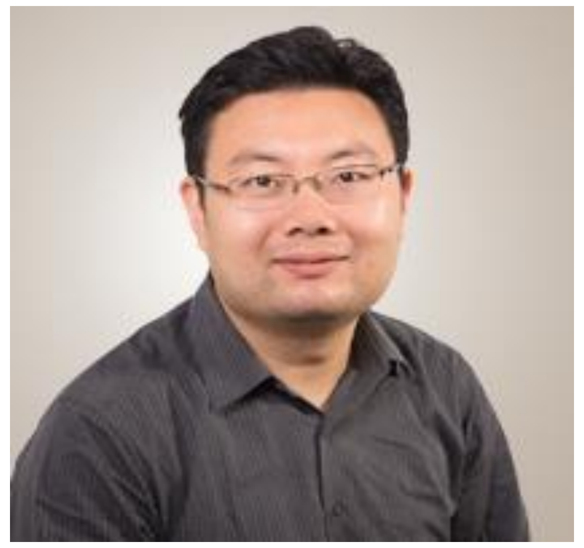
Horizon Robotics

Oregon State University

Qinxun Bai

Wei Xu

Neale Ratzlaff



Li Fuxin

Code Link

References

[1] Blundell, Charles, et al. "Weight uncertainty in neural network." *International Conference on Machine Learning*. PMLR, 2015.

[2] Hoffman, Matthew D., et al. "Stochastic variational inference." *Journal of Machine Learning Research* 14.5 (2013).

[3j Graves, Alex. "Practical variational inference for neural networks." *Advances in neural information processing systems*. 2011.

[4] Liu, Qiang, and Dilin Wang. "Stein variational gradient descent: A general purpose bayesian inference algorithm." *Advances in neural information processing systems.* 2016.

[5] Liu, Chang, et al. "Understanding and accelerating particle-based variational inference." *International Conference on Machine Learning*. PMLR, 2019.

[6] Wang, Dilin, and Qiang Liu. "Learning to draw samples: With application to amortized mle for generative adversarial learning." *arXiv preprint arXiv:1611.01722* (2016).

[7] Lakshminarayanan, Balaji, Alexander Pritzel, and Charles Blundell. "Simple and scalable predictive uncertainty estimation using deep ensembles." *Advances in neural information processing systems*. 2017.

[8] Neal, Radford M. "MCMC using Hamiltonian dynamics." *Handbook of markov chain monte carlo* 2.11 (2011): 2.