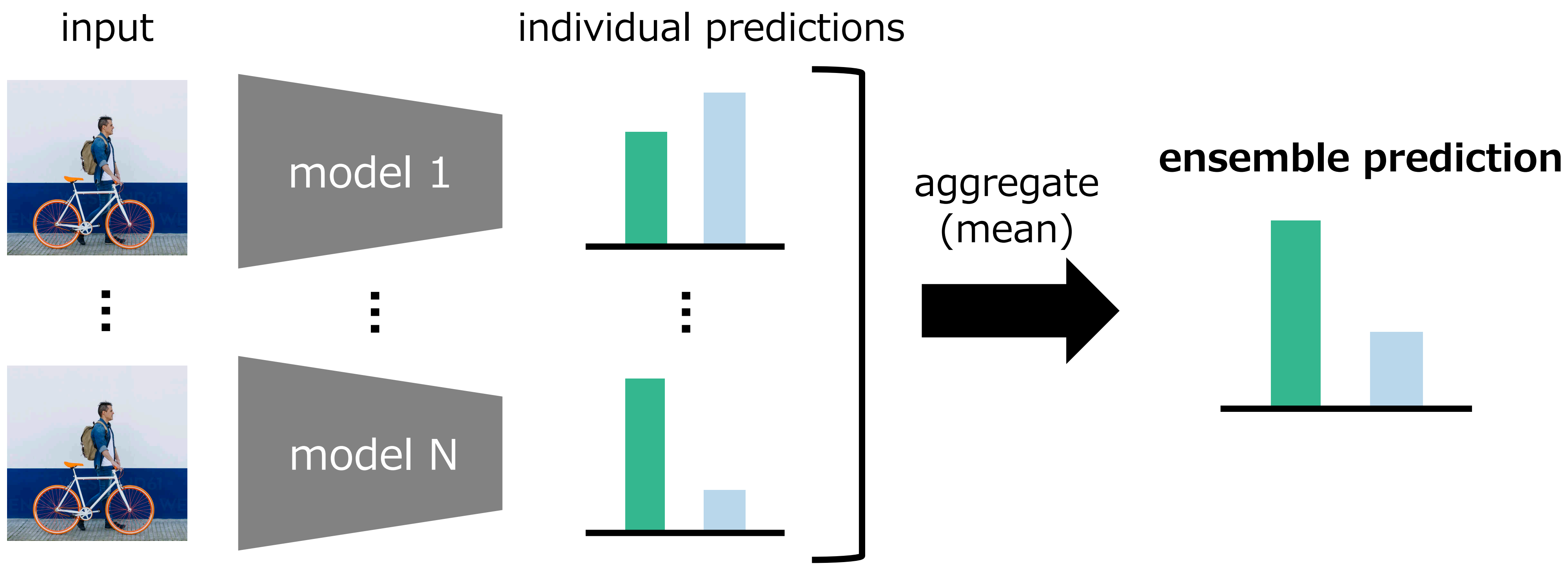
ICML 2022

Feature Space Particle Inference for Neural Network Ensembles

Shingo Yashima[1], Teppei Suzuki[1],
Kohta Ishikawa[1], Ikuro Sato[12], Rei Kawakami[12]

[1]Denso IT Laboratory, Inc., Japan
[2]Tokyo Institute of Technology, Japan

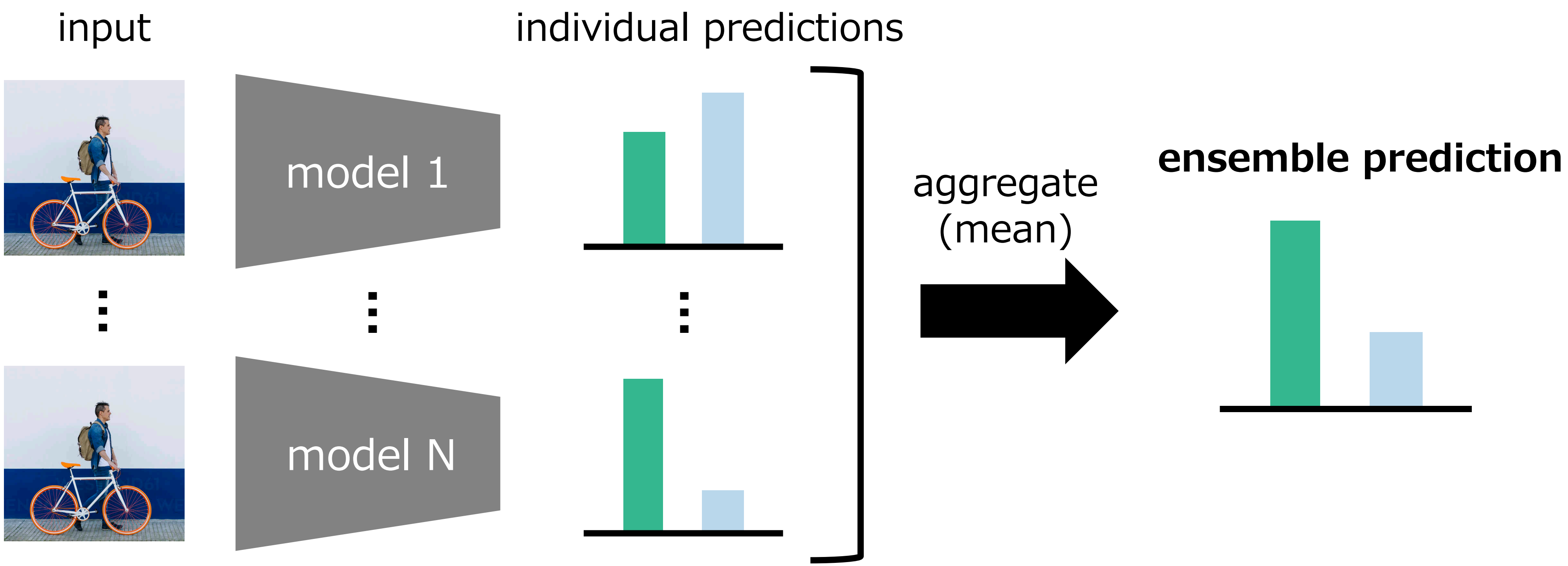
Backgound: Model Ensemble



Popular approach to improve
- Generalization performance
- Uncertainty quantification
- Robustness to perturbation

© DENSO IT LABORATORY, INC. All Rights Reserved.

Backgound: Model Ensemble

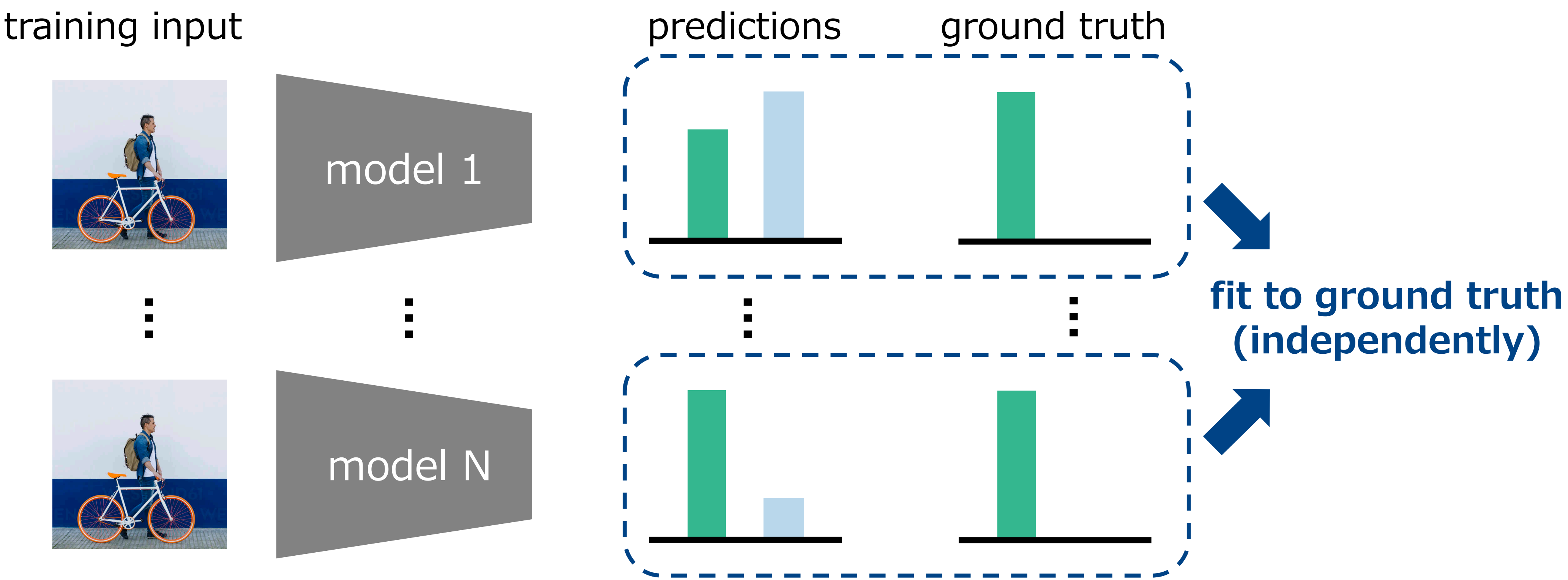


- For robust predictions, individual predictions should be diverse
- There is a **trade-off** between **individual performance** and **ensemble diversity**

What diversity is effective for model ensembles?

© DENSO IT LABORATORY, INC. All Rights Reserved.

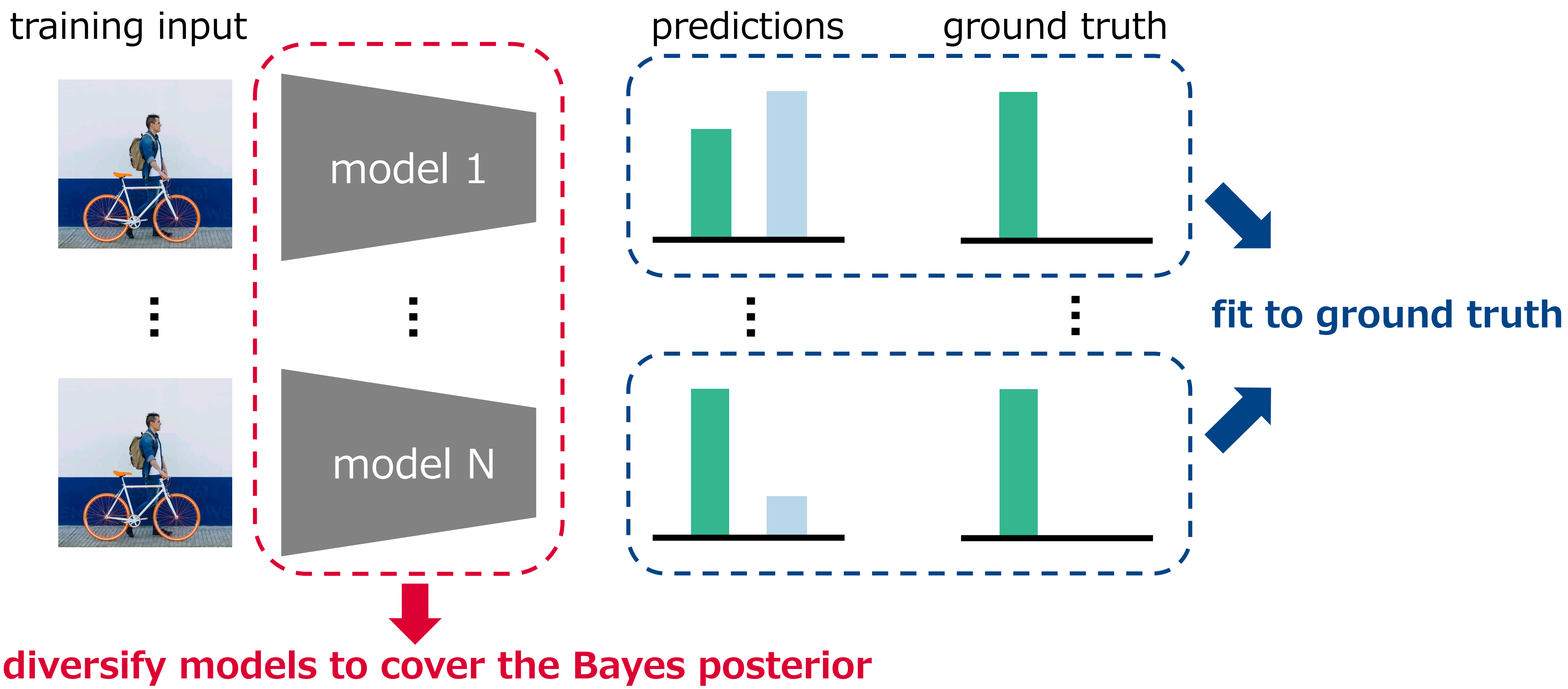
Deep Ensembles (Lakshminarayanan et al., 2017)



- Train individual models independently
  (**implicit diversity** from randomness in initialization)
- Decent (SOTA) empirical performance

© DENSO IT LABORATORY, INC. All Rights Reserved.

Particle-based Variational Inference

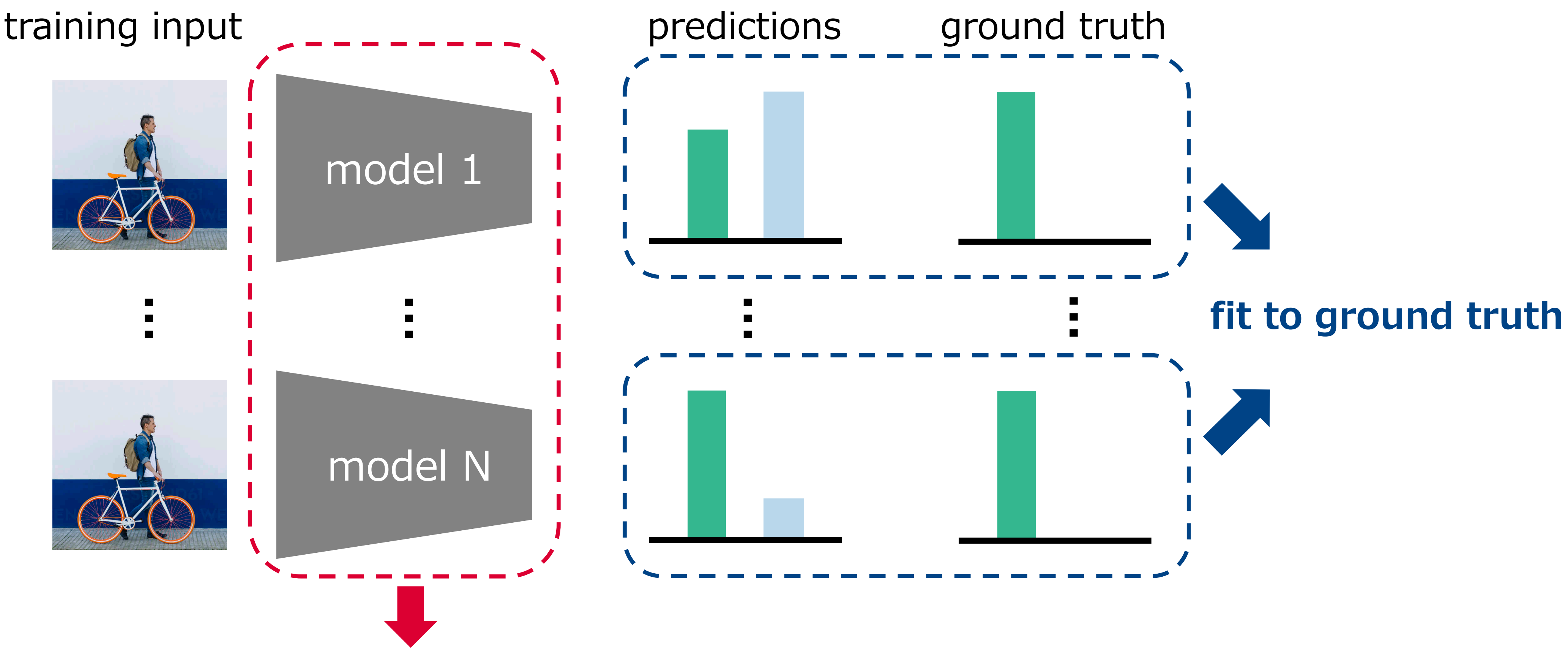


- Nonparametric method to obtaining samples from the Bayes posterior
- When applying to DNNs, they can be seen as Deep Ensembles with repulsive forces

(D'Angelo and Fortuin, 2021)

© DENSO IT LABORATORY, INC. All Rights Reserved.

Particle-based Variational Inference



diversify models to cover the Bayes posterior

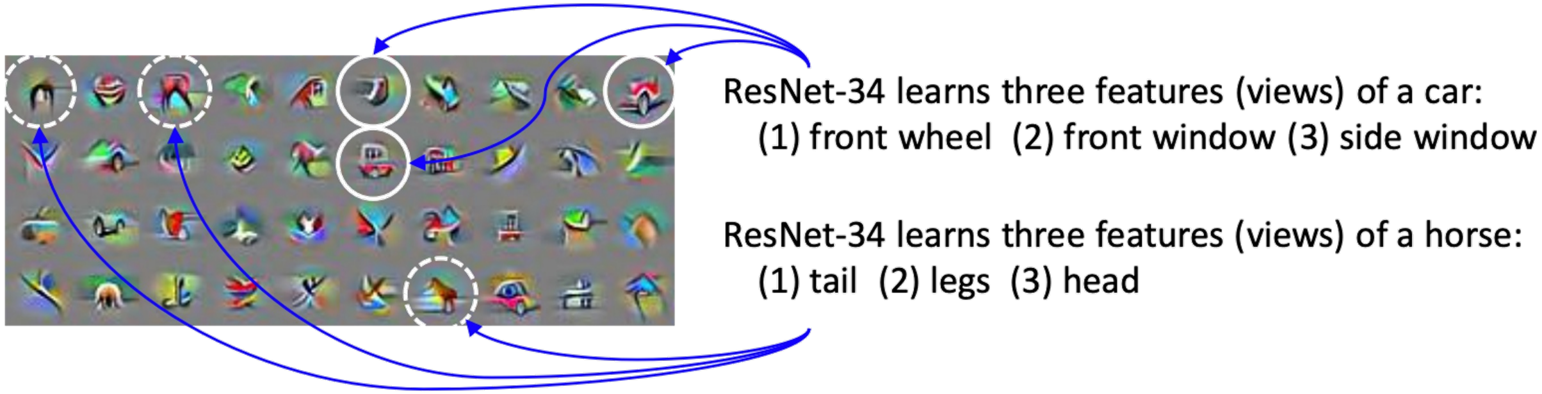
However, the best way to apply these methods to DNNs is unclear:

- Sample from weight-space posterior $p(w|\mathcal{D})$ suffers from **overparameterization**
- Sample from function-space posterior $p(f|\mathcal{D})$ often shows severe **underfitting**

© DENSO IT LABORATORY, INC. All Rights Reserved.

Multi-View Structure (Allen-Zhu et al., 2020)

Figure 4: Illustration of images with multiple views (features) in the ImageNet dataset.



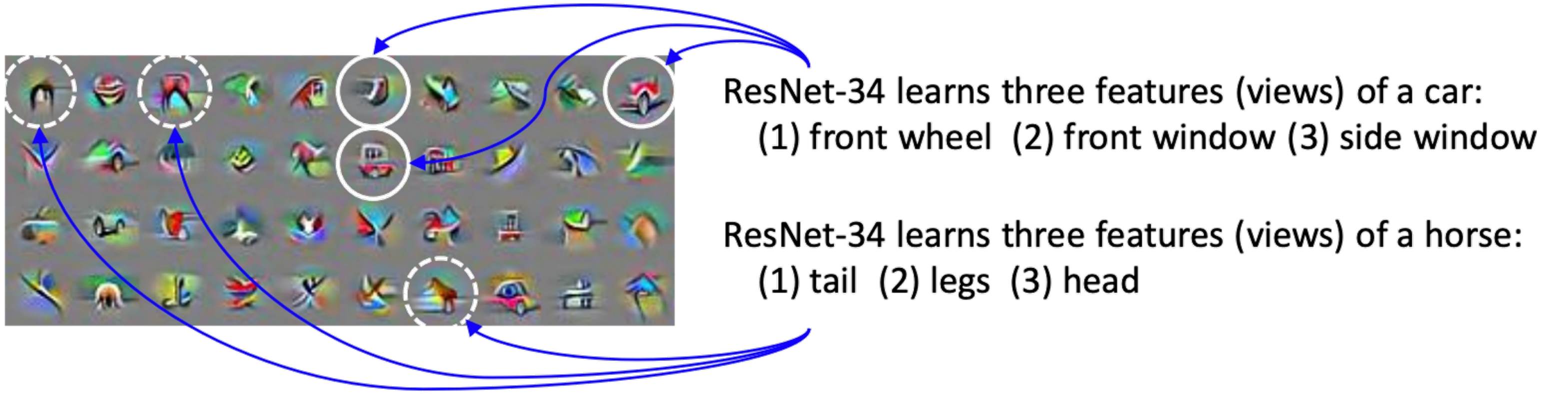
An ensemble of DNNs improve its performance when

- Input images have **multiple features which explain label** (multi-view structure)
- Each model captures **different features from each other**

© DENSO IT LABORATORY, INC. All Rights Reserved.

Multi-View Structure (Allen-Zhu et al., 2020)

Figure 4: Illustration of images with multiple views (features) in the ImageNet dataset.



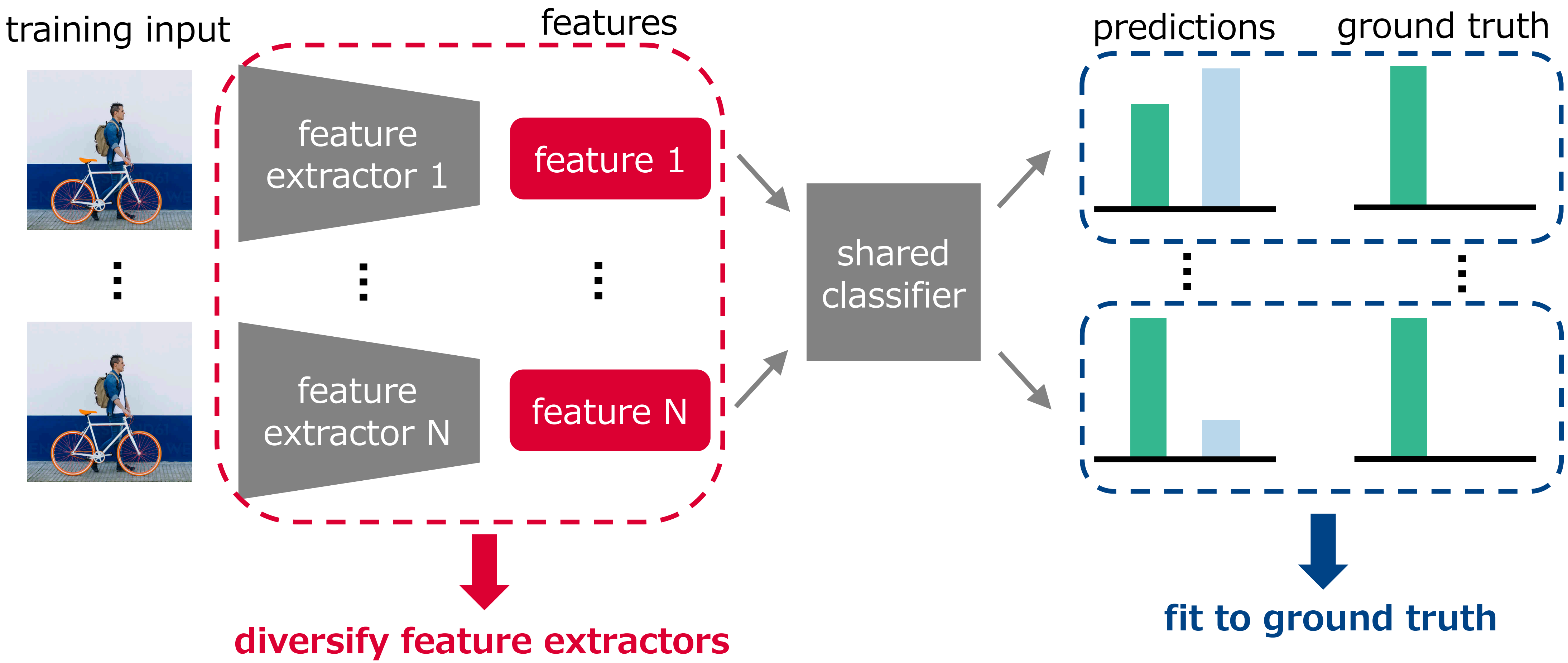
An ensemble of DNNs improve its performance when

- Input images have **multiple features which explain label** (multi-view structure)
- Each model captures **different features from each other**

Hypothesis: Explicitly promoting feature diversity improves ensemble performance

© DENSO IT LABORATORY, INC. All Rights Reserved.

Method Overview



- Divide a model into a feature extractor and a classifier (typically a final dense layer)
- Promote each model to capture distinct feature while classifying data correctly

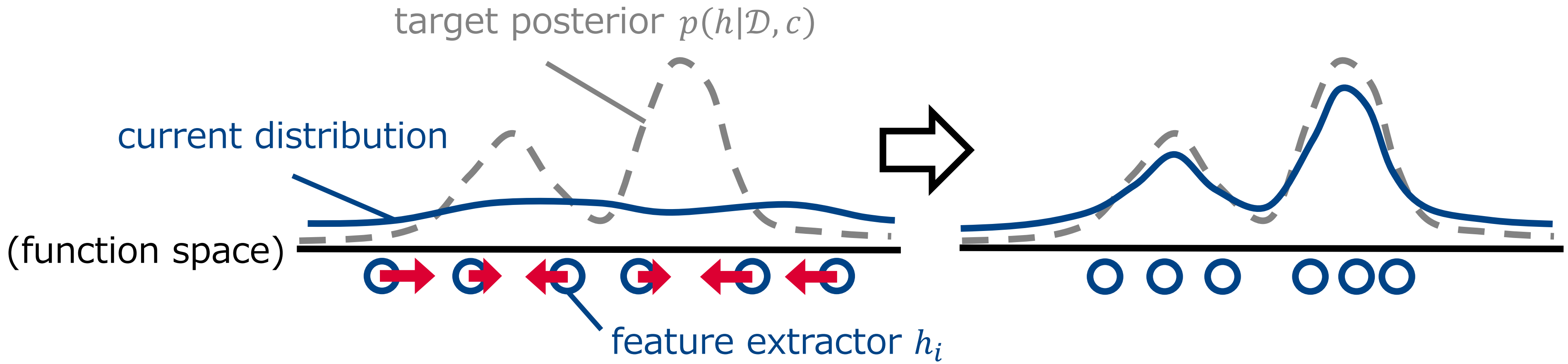
© DENSO IT LABORATORY, INC. All Rights Reserved.

Formulation

Consider the **Bayes posterior** of a feature extractor h given training data and classifier c :

$$\underbrace{p(h|\mathcal{D},c)}_{\text{posterior}} \propto \underbrace{p(h)}_{\text{prior}} \underbrace{\prod_{x,y\in\mathcal{D}} p(y|c(h(x)))}_{\text{data likelihood}}$$

Optimize feature extractors $\{h_i\}_{i=1}^N$ so that they approximate the above posterior using particle-based variational inference



© DENSO IT LABORATORY, INC. All Rights Reserved.

Algorithm

1. Calculate gradient of feature: $\phi(\boldsymbol{h}_i) = \nabla_{\boldsymbol{h}_i} \log p(\boldsymbol{y}|\boldsymbol{h}_i) + \nabla_{\boldsymbol{h}_i} \log p(\boldsymbol{h}_i) + \frac{\sum_j \nabla_{\boldsymbol{h}_i} k(\boldsymbol{h}_i, \boldsymbol{h}_j)}{\sum_j k(\boldsymbol{h}_i, \boldsymbol{h}_j)}$

data fitting term

diversify term
(k : p.d. kernel)

2. Update weights by backprop: $w_i \leftarrow w_i + \alpha \frac{\partial \boldsymbol{h}_i}{\partial w_i} \phi(\boldsymbol{h}_i)$

$\boldsymbol{h}_i \coloneqq h_i(\boldsymbol{X})$

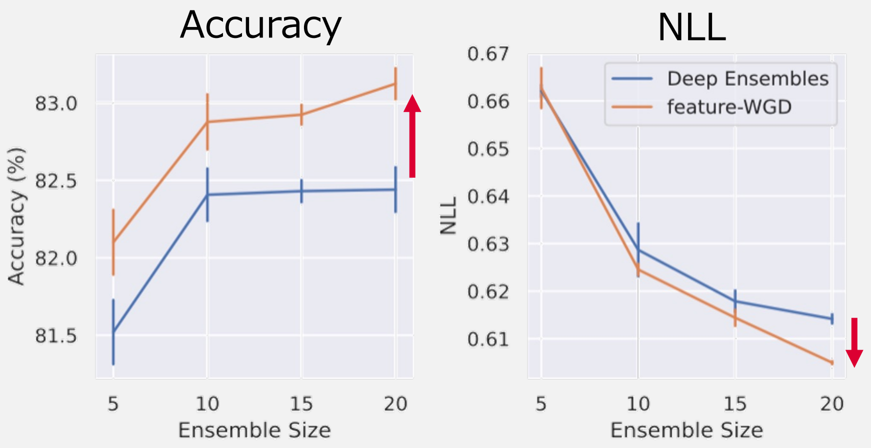
- Performing an inference on feature extractors projected on training data points
  $\boldsymbol{h}_i \coloneqq h_i(\boldsymbol{X})$ (Wang et al., 2019)
- To update weights, backpropagate the gradient of feature values

© DENSO IT LABORATORY, INC. All Rights Reserved.

Evaluation: Classification Performance

Classification results on CIFAR-100 with an ensemble size of 10

	METHOD	ACCURACY(↑)	NLL(↓)	BRIER(↓)	ECE(↓)	CA / CNLL / CBRIER / CECE
			uncertainty estimation			corruption robustness
	SINGLE	77.4 ± 0.3	0.835 ± 0.007	0.316 ± 0.003	0.030 ± 0.003	46.7 / 2.279 / 0.658 / 0.035
independent	DEEP ENSEMBLES	82.3 ± 0.2	0.632 ± 0.004	0.249 ± 0.001	0.020 ± 0.001	52.9 / 1.971 / 0.590 / 0.032
	WEIGHT-WGD	82.3 ± 0.1	0.633 ± 0.002	0.250 ± 0.001	0.021 ± 0.001	52.8 / 1.967 / 0.589 / 0.031
	FUNCTION-WGD	79.0 ± 0.1	0.715 ± 0.003	0.286 ± 0.001	0.018 ± 0.002	49.5 / 2.133 / 0.623 / 0.034
diversifying features	FEATURE-WGD	**82.9 ± 0.2**	**0.624 ± 0.002**	**0.243 ± 0.001**	**0.017 ± 0.001**	**53.5 / 1.955 / 0.584 / 0.029**

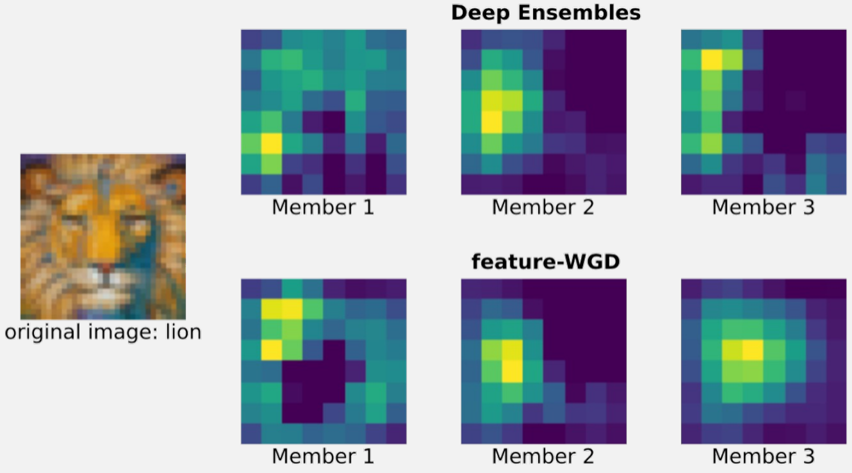


Increasing an ensemble size

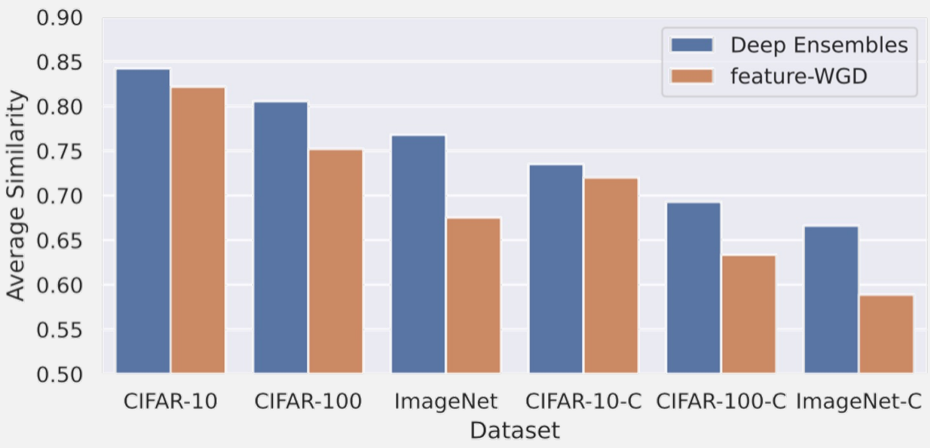
- Ours consistently improves Deep Ensembles and weight/function space inferences
- Achieve comparable performance with fewer number of models

© DENSO IT LABORATORY, INC. All Rights Reserved.

Evaluation: Feature Diversity



Class activation map (CAM) on a lion image



Similarity of CAM between ensemble members

- Visualize attention maps using Grad-CAM
- Ours capture more diverse features (e.g., face, mane)

© DENSO IT LABORATORY, INC. All Rights Reserved.

Summary

- Proposed an ensemble method of DNN that explicitly promote feature diversity using Bayesian particle-based variational inference

- Confirmed that proposed method improves Deep Ensembles and weight/function space inference in terms of accuracy, calibration, and robustness

- Code is available at: https://github.com/DensoITLab/featurePI

© DENSO IT LABORATORY, INC. All Rights Reserved.