2026 / SEOUL, SOUTH KOREA / 43RD INT'L CONFERENCE ON MACHINE LEARNING

PMLR vol. 306 · arXiv:2601.22068 · Poster #2810

Quantifying the Uncertainty of Foundation Models with Singular Value Ensembles

Mehmet Özgür Türkoğlu[1] · **Dominik J. Mühlematter**[2*] · **Alexander Becker**[2] · **Konrad Schindler**[2] · **Helge Aasen**[1]

1 Agroscope — Earth Observation of Agroecosystems

2 ETH Zürich — Photogrammetry & Remote Sensing

* presenting author

Deep Ensemble quality, under 1% parameter overhead

- Foundation models are now the default starting point for vision and language tasks, but their predictions are often overconfident. In high-stakes settings, accuracy alone is not enough, the model must also know when it does not know.
- Deep ensembles remain one of the strongest tools for epistemic uncertainty, but they require multiple independently trained models. For large pretrained backbones, this makes memory, training, and deployment costs grow with the number of ensemble members.

EFFICIENCY

<1% — added parameters, vs 700% for deep ensemble

700× — smaller memory than a deep ensemble

22M–7B — params validated, DINO · BERT · LLaMA-2

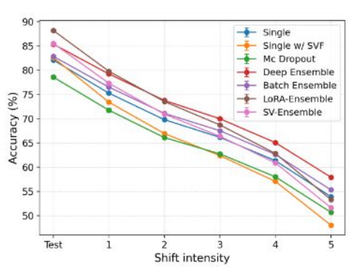
Method	Parameter Overhead(%)	Memory Overhead(%)	Infer. Flops(G)	Infer. Time(ms)
Deep Ensemble	700	700	8×21.9	8×3.1
LoRA-Ensemble	10.3	10.3	175.2	**21.9**
Bayes-LoRA	4.0	4.0	**68.2**	313
SV-Ensemble	**1.0**	**1.0**	175.2	**21.9**

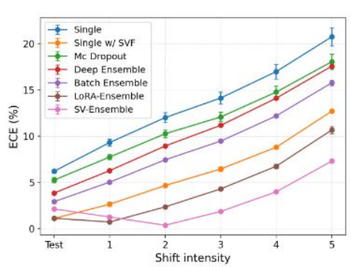
M	Acc(%)↑	ECE(%)↓	NLL↓
1	83.2	12.1	0.77
2	84.5	8.5	0.78
4	84.9	5.8	0.47
8	85.7	5.2	0.50
16	**85.8**	**3.8**	**0.43**

EPISTEMIC UNCERTAINTY

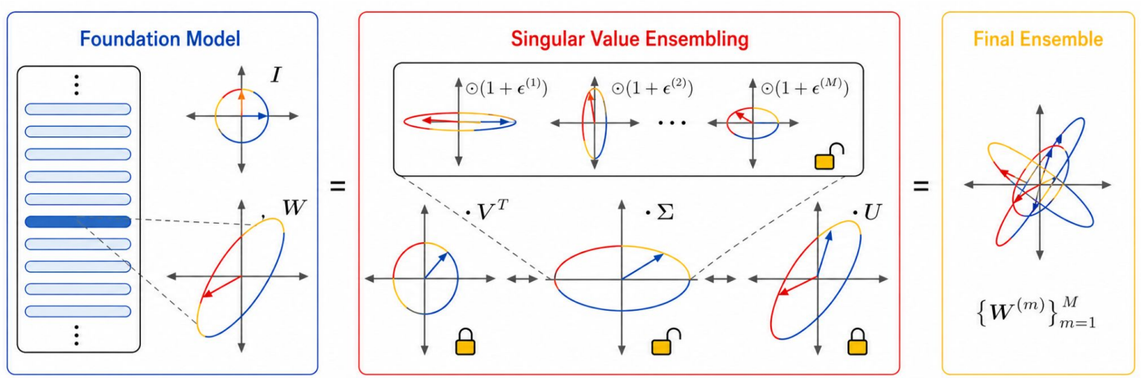
Score	Method	AUROC(↑)	AUPRC(↑)	FPR 95% TPR(↓)	ID mean	OOD mean	OOD/ID
Entropy	Deep Ens.	81.9	79.1	56.7	0.451	1.247	2.8×
Entropy	SVE	**83.9**	**83.3**	**56.5**	0.578	1.728	**3.0**×
MI	Deep Ens.	82.5	79.3	**57.0**	0.110	0.354	3.2×
MI	SVE	**82.7**	**80.6**	58.6	0.007	0.026	**3.4**×

DATASET SHIFT ROBUSTNESS





ONE BASIS, MANY MODELS



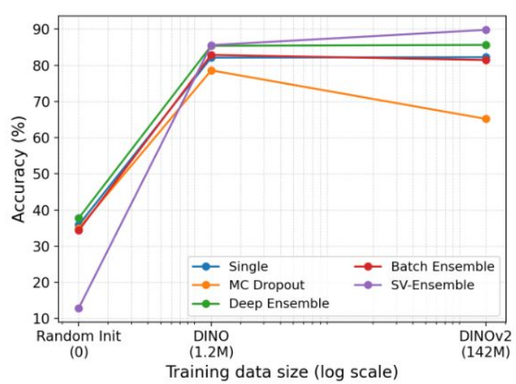
KEY IDEA

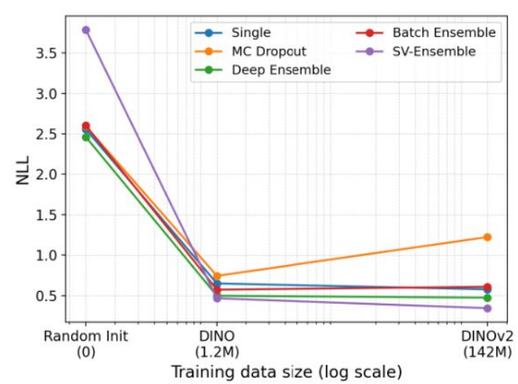
A pretrained model already contains useful knowledge directions in its singular vectors. SVE turns one foundation model into an ensemble by keeping these directions fixed and learning only how strongly each member uses them.

HOW IT WORKS

1. **Freeze the basis.** Decompose each pretrained weight matrix as $W = U\Sigma V^T$. The singular vectors U and V^T define directions learned during pretraining. SVE keeps this basis fixed and shared across all members.
2. **Free the values.** Each ensemble member learns its own singular values Σ^m. These values reweight the shared directions, producing different predictors without learning new directions from scratch.
3. **Diversity for free.** Small initialization noise and stochastic mini-batch training push members toward different rescalings of the same basis.

BACKBONE PRETRAINING SCALE





VISION RESULT

Method	Acc (%)↑	ECE (%)↓	NLL↓	Brier↓
Flowers102 – DINO ViT-S/16 – Fine-tuning for 10 epochs				
Single	86.3±0.8	3.9±0.8	0.56±0.05	0.20±0.02
EDL	80.4	71.9	2.85	0.85
EDL+Temp. Scaling	80.4	4.8	0.89	0.26
Single w/ SVF	91.8±0.4	1.8±0.4	0.33±0.01	0.12±0.01
MC Dropout	76.6±2.5	4.4±0.3	0.92±0.01	0.32±0.03
Deep Ensemble	91.5±0.3	**0.9**±0.2	0.33±0.01	0.12±0.00
Batch Ensemble	91.7±0.4	2.2±0.2	0.32±0.02	0.12±0.01
LoRA-Ensemble	94.6±0.1	1.1±0.1	0.21±0.01	0.08±0.00
SV-Ensemble	**95.4**±0.2	1.0±0.1	**0.18**±0.01	**0.07**±0.00
CIFAR100 – DINO ViT-S/16 – Fine-tuning for 10 epochs				
Single	82.1±0.1	6.2±0.3	0.65±0.01	0.26±0.00
Single w/ SVF	82.6±0.1	**1.1**±0.1	0.57±0.01	0.25±0.00
MC Dropout	78.6±0.7	5.3±0.5	0.74±0.02	0.30±0.01
Deep Ensemble	85.4±0.1	3.9±0.1	0.50±0.00	0.21±0.00
Batch Ensemble	82.9±0.5	2.9±0.3	0.57±0.01	0.24±0.01
LoRA-Ensemble	**88.2**±0.1	**1.1**±0.2	**0.37**±0.00	**0.17**±0.00
SV-Ensemble	85.6±0.2	2.1±0.2	0.47±0.00	0.21±0.00
DTD Texture – DINOv2 ViT-S/14 – Fine-tuning for 10 epochs				
Single	60.2±0.9	11.6±1.4	1.56±0.05	0.55±0.02
Single w/ SVF	78.4±0.9	8.2±0.6	0.82±0.04	0.32±0.01
MC Dropout	53.1±2.8	9.2±2.1	1.76±0.07	0.62±0.02
Deep Ensemble	65.4±0.6	7.6±0.6	1.32±0.01	0.48±0.00
Batch Ensemble	67.2±0.5	8.1±0.4	1.20±0.01	0.47±0.01
LoRA-Ensemble	78.4±0.7	6.6±0.9	0.80±0.01	0.32±0.01
SV-Ensemble	**79.5**±0.1	**6.3**±0.2	**0.74**±0.01	**0.30**±0.01
Oxford Pets – DINOv2 ViT-S/14 – Fine-tuning for ≈ 200 iter.				
Single	81.4±1.1	8.6±0.8	0.61±0.03	0.27±0.01
Single w/ SVF	87.2±1.4	**1.9**±0.8	0.40±0.06	0.19±0.02
MC Dropout	25.9±15.7	6.1±2.4	2.53±0.63	0.85±0.12
Deep Ensemble	89.2±0.4	13.3±0.5	0.43±0.02	0.17±0.04
Batch Ensemble	70.5±2.4	48.7±1.7	1.75±0.05	0.69±0.02
LoRA-Ensemble	86.1±1.0	9.0±4.2	0.49±0.06	0.22±0.02
SV-Ensemble	**90.1**±1.3	2.2±0.9	**0.30**±0.03	**0.15**±0.02

NLP RESULT

Method	Acc (%)↑	ECE (%)↓	NLL↓
Single	84.7	13.4	1.26
Single w/ SVF	83.2	12.1	0.77
MC Dropout (N=10) (Yang et al., 2024)	85.0	12.4	1.11
Deep Ensemble (M=3) (Yang et al., 2024)	85.8	9.9	0.83
Last-Layer Ensemble (M=5) (Wang et al., 2023)	72.0	6.0	0.79
Ckpt Ensemble (M=3) (Yang et al., 2024)	85.8	9.8	0.80
LoRA-Ensemble (M=5) (Wang et al., 2023)	**86.0**	9.0	0.92
C-LoRA (Rahmati et al., 2026)	84.4	4.3	0.48
Bayes-LoRA (LLLA) (Yang et al., 2024)	84.7	11.6	0.87
Bayes-LoRA (LA) (Yang et al., 2024)	85.1	5.4	0.49
Blob (N=5) (Wang et al., 2024)	84.7	6.2	0.46
Blob (N=10) (Wang et al., 2024)	85.5	**3.6**	**0.40**
SV-Ensemble (M=16)	85.8	3.8	0.43

CONCLUSION

SVE turns one pretrained foundation model into a parameter-efficient ensemble by learning only member-specific singular values.

It delivers Deep Ensemble quality uncertainty with under **1%** extra parameters, making calibrated foundation models practical at scale.

code & paper: github.com/moturkoglu/Singular-Value-Ensemble · arXiv:2601.22068

correspondence: moturkoglu@gmail.com