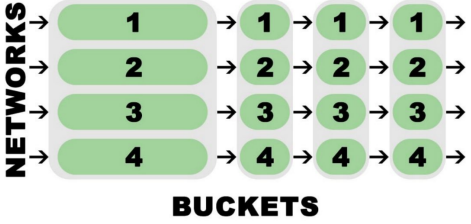
Cross-Pollinated Deep Ensembles

SAMSUNG AI Center - Moscow
p(BA)yesgroup.ru

Alexander Lyzhov*[13], Daria Voronkova*[2], Dmitry Vetrov[21]
* Equal contribution 1 Samsung AI Center Moscow 2 Samsung-HSE Laboratory, HSE university 3 Center on Long-Term Risk

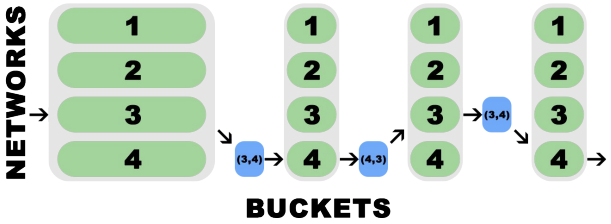
CENTER ON LONG-TERM RISK

How to estimate uncertainty more efficiently? Form an ensemble by combining blocks of multiple networks



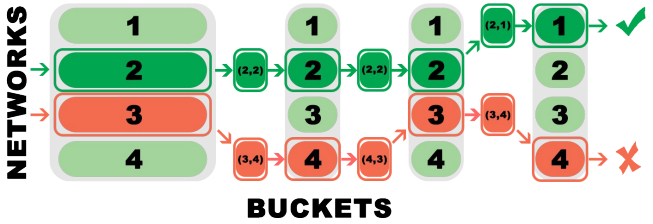
I. Independent pre-training

Train a Deep Ensemble [1].
Split networks into blocks.



II. Joint finetuning

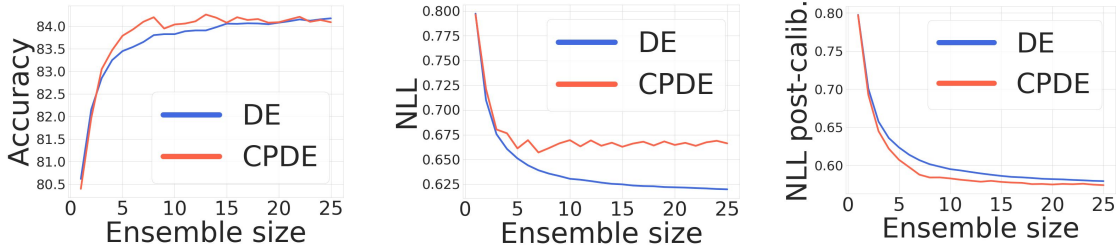
On each step, finetune a random combination of blocks (path).
Learn 3x3 conv **interfaces between each consecutive pair of blocks**.
Also use **separate batchnorms** for each consecutive pair of blocks.
Use low learning rates to **preserve the diversity from stage I**.



III. Ensemble pruning

Start with an empty ensemble.
Greedily add paths by minimizing NLL of the ensemble on validation data after calibration with temperature scaling [2].

CPDE (Cross-Pollinated Deep Ensembles) vs. DE (Deep Ensembles) on image classification (CIFAR-100 / WideResNet28x10):



	Training time	Test time	Memory
CPDE	8	N/3 . . . N	8
DE	N	N	N

Requirements in budgets for an ensemble of size N
(1 budget = resources required by 1 network)

	Accuracy	NLL post-calibration
CPDE	**84.1**	**0.571**
w/o separate batchnorms	83.6	0.616
w/o interfaces	83.8	0.626
w/o separate batchnorms, w/o interfaces	83.7	0.623
with 2 (instead of 4) buckets	83.9	0.611
with random (instead of greedy) path selection	80.8	0.717
with joint (instead of independent) pre-training	70.3	1.094

CPDE ablations (10 ensemble members trained on 45K objects)

CPDE enables better generalization and uncertainty estimation than DE with equal test-time compute, and is much cheaper to train and store given large compute budgets.
In the future we plan to compare and combine CPDE with MIMO [3]

References

[1] Lakshminarayanan, B., Pritzel, A., & Blundell, C. (2017). Simple and scalable predictive uncertainty estimation using deep ensembles. NeurIPS'17.
[2] Guo, C., Pleiss, G., Sun, Y., & Weinberger, K. Q. (2017). On calibration of modern neural networks. arXiv preprint arXiv:1706.04599.
[3] Havasi, M., Jenatton, R., Fort, S., Liu, J. Z., Snoek, J., Lakshminarayanan, B., ... & Tran, D. (2020). Training independent subnetworks for robust prediction. arXiv preprint arXiv:2010.06610.

Acknowledgements

Dmitry Vetrov was supported by the Russian Science Foundation grant no. 19-71-30020.