On The Expressive Power of Permutation-Equivariant Weight-Space Networks

Adir Dayan[1,*] Yam Eitan[1,*] Haggai Maron[1,2]

[1] TECHNION Israel Institute of Technology

[2] NVIDIA

ICML 2026 – Spotlight

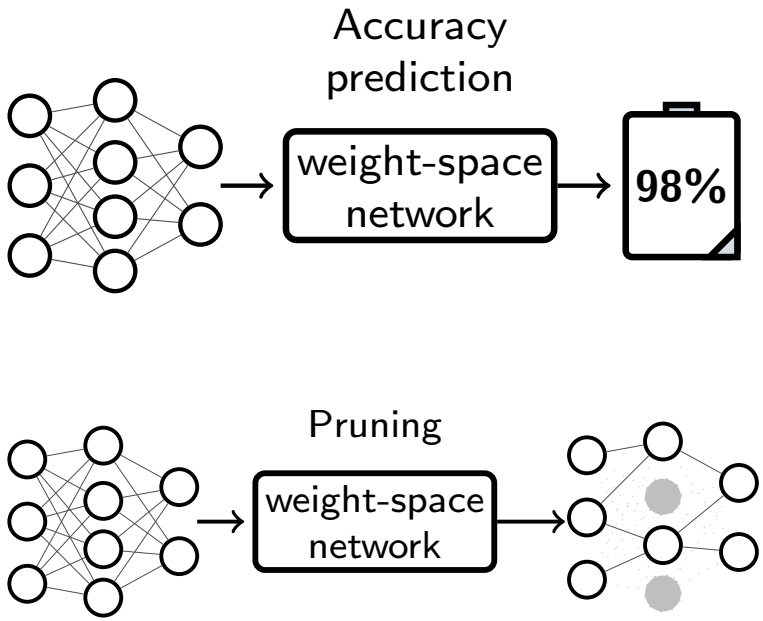
*Equal contribution

How powerful are neural networks that learn from other neural networks?

- Pretrained models are everywhere.
- Models are larger and more black-box.
- Direct evaluation is costly.

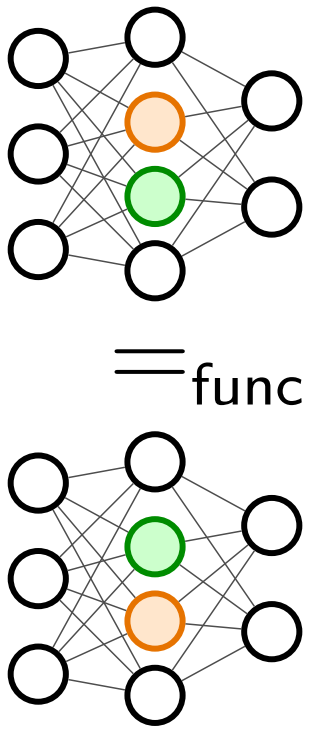
Weight-space learning:

- A new class of networks treats **model weights as the input data**.
- Weight-space tasks:
  - model accuracy prediction
  - model editing
  - pruning
  - domain adaptation



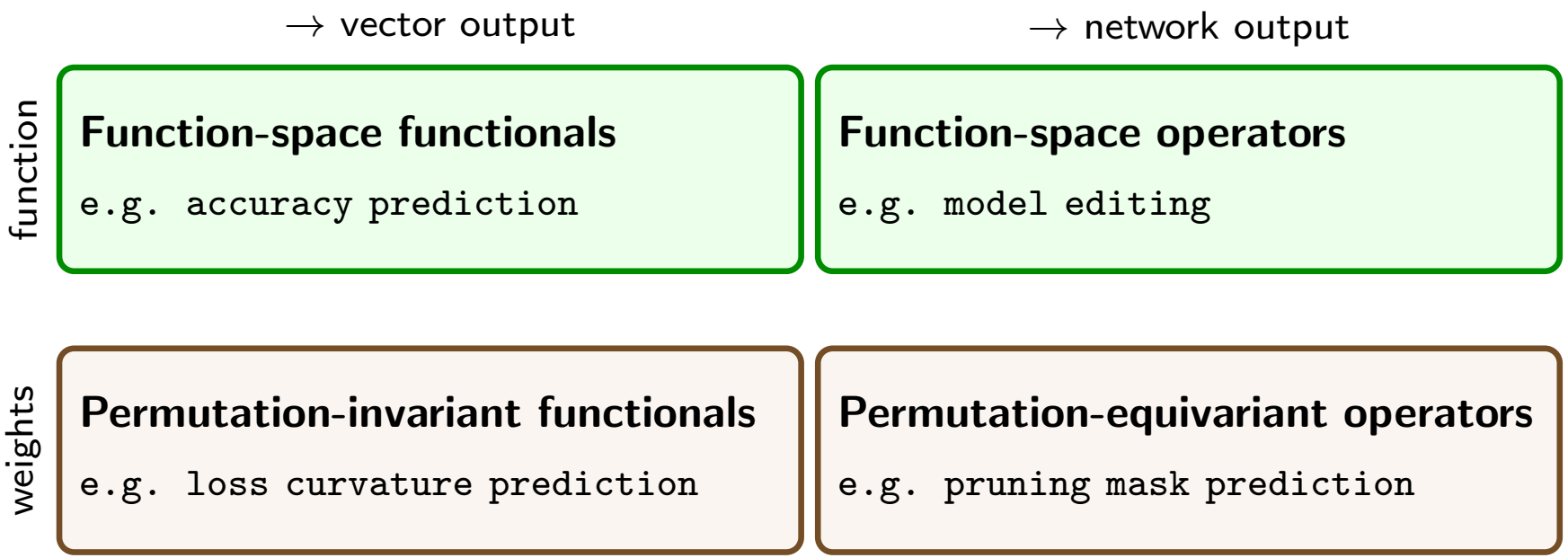
Symmetry VS Expressivity

- Permuting the neurons of a network changes its weights **but not the function it computes**.
- Therefore, SOTA models are built to be permutation-equivariant.
- But this raises a key question: does enforcing symmetry reduce expressive power?



Approximation Framework

- We identify four approximation settings for weight-space tasks:
  - The task objective might depend on the function, or the raw weights.
  - The prediction can be a vector, or different network weights.
- We analyze all four settings.



Our theoretical findings

- Prominent permutation-equivariant architectures (DWS, NFN, GMN, NG-GNN) have the **same expressive power**.
- They are **universal approximators** in both weight-space and function-space settings, assuming input weights are in *General Position (GP)*.

* Weights are in *GP* when all neuron biases distinct within each layer (satisified almost surely).

Universality results

Setting	**Status**
Function-space functionals	Always universal
Permutation-invariant functionals	Universal on weights in GP
Permutation-equivariant operators	Universal on weights in GP
Function-space operators	Requires large-enough output network

Empirical Gains from Increased Output Capacity

- Guided by our theory, we propose a **simple modification** to existing weight-space models that achieve up to a **34%** improvement over prior SOTA on a standard model editing benchmark (**MNIST INR dilation**).
- The modification is to *expand the output capacity* by predicting k different networks and ensemble them - **no extra parameters**, any architecture.

Method	**MSE** $\times 10^{-2} \downarrow$
NG-T-64	1.75
ScaleGMN + GradMetaNet++ (prev. SOTA)	1.60
DWS (k=1, base)	2.29
DWS + OCE (k=8)	1.36
GMN + OCE (k=8)	**1.06**

k = networks in the output ensemble.

Conclusions

- Symmetry does **not weaken** permutation-equivariant models.
- We retain **full expressive power** while gaining structure, efficiency, and better generalization.
- Our work provides a **unified theoretical foundation** for weight-space learning.
- And gives **principled guidance** for designing future architectures that operate on other neural networks.

Thank you!

Adir Dayan
adir.dayan@campus.technion.ac.il

Yam Eitan
yameitan1997@gmail.com

Haggai Maron
haggaimaron@ef.technion.ac.il