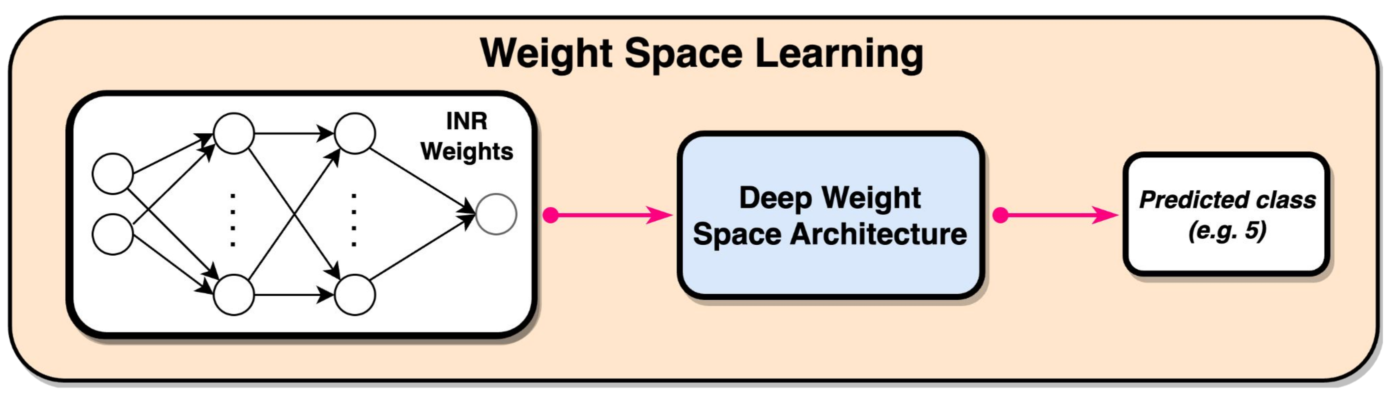
Data Augmentations in Deep Weight Spaces

Aviv Shamsian*[1] David W. Zhang*[3] Aviv Navon[1] Yan Zhang[4] Miltiadis Kofinas[3]
Idan Achituve[1] Riccardo Valperga[3] Gertjan J. Burghouts[5] Efstratios Gavves[3]
Cees G. Snoek[3] Ethan Fetaya[1] Gal Chechik[1,6] Haggai Maron[2,6]

[1]Bar-Ilan University [2]Technion [3]University of Amsterdam [4]Samsung - SAIT AI Lab, Montreal [5]TNO [6]NVIDIA

Overview

- Learning in Deep Weight Spaces (DWS) is a new learning setup in which neural networks take other neural networks as input.
- Recent works address the *weight-symmetry problem* but the performance still lags behind other data modalities. We identify overfitting as the main culprit.
- We demonstrate that data augmentation tailored for learning in DWS is an effective approach for mitigating overfitting.

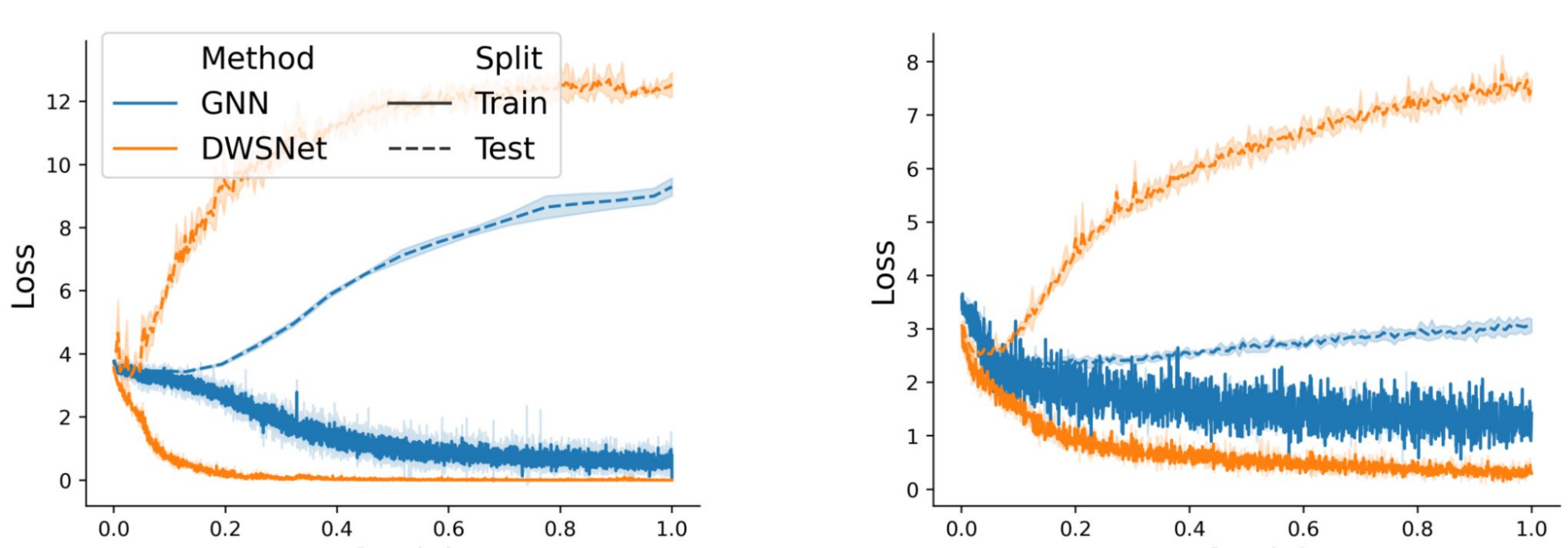


Contributions

- We present a **taxonomy of data augmentation schemes** for DWS and conduct a thorough empirical comparison.
- We propose **Weight Space MixUp** and demonstrate that it is as effective as using up to 10X more data.

Overfitting in Deep Weight Spaces

- Classifiers of Implicit Neural Representations (INRs) still largely underperform their image/point cloud counterparts.
- The main cause of this problem is poor generalization of DWS architectures, as can be seen in the figure below:



Train versus test loss for DWSNet and GNN on the ModelNet40 INR dataset. Overfitting occurs very early regardless the training size (12K and 120K training samples in the left and right panels respectively).

Augmentation Taxonomy

Generic augmentations: Can be applied to any neural network. Examples are *weight dropout, quantile weight dropout,* and *Gaussian noise.*

Activation symmetries: Exploits symmetries of activation functions:

- ReLU scaling - $\frac{1}{c}W_{i+1}\mathrm{ReLU}(cW_ix+cb_i)+b_{i+1} = W_{i+1}\mathrm{ReLU}(W_ix+b_i)+b_{i+1}$
- SIREN negation - $W_{i+1}\mathrm{Sin}(W_ix+b) = -W_{i+1}\mathrm{Sin}(-W_ix-b)$
- SIREN bias - $W_{i+1}\mathrm{Sin}(W_ix+b) = (-1)^k W_{i+1}\mathrm{Sin}(W_ix+b+k\pi)$

Input-space augmentations: Augmentations like random rotations or translations of the underlying object, directly applied to the weights of the INR. For example, rotations by the rotation matrix R replaces W_1 with W_1R.

Weight space MixUp

Due to the weight space symmetry there exists multiple ways of combining neural network weights for MixUp.

MixUp: $W = \lambda W_1 + (1-\lambda)W_2, b = \lambda b_1 + (1-\lambda)b_2$

MixUp + Random permutation: Randomly permute W_2 *and* b_2 before MixUp.

Alignment + MixUp: Align the weights and biases prior to MixUp by the optimal transport map.

Experiments

- We evaluate various weight space augmentations for classifying 2D and 3D INRs using FMNIST and ModelNet40 datasets.
- We apply each augmentation on two types of leading architectures DWS (Navon et al., 2023) and GNN (Zhang et al., 2023).

Augmentation type	Model	ModelNet40		FMNIST	
		1 View	10 View	1 View	10 View
No augmentation	DWS	16.17 ± 0.25	30.25 ± 0.95	68.30 ± 0.62	76.01 ± 1.20
No augmentation	GNN	8.82 ± 1.08	34.51 ± 1.24	68.84 ± 0.41	79.58 ± 3.01
Translate	DWS	18.18 ± 0.97	31.17 ± 0.02	67.90 ± 0.24	77.61 ± 0.36
Rotation	DWS	—	—	68.55 ± 0.28	77.04 ± 0.47
Scale	DWS	16.41 ± 0.57	30.54 ± 0.72	67.99 ± 0.14	75.77 ± 1.09
Gaussian noise	DWS	14.10 ± 0.71	25.31 ± 1.78	68.53 ± 0.09	77.60 ± 0.13
SIREN bias	DWS	4.69 ± 0.10	4.90 ± 0.01	58.20 ± 0.01	62.21 ± 0.55
SIREN negation	DWS	20.14 ± 0.98	32.31 ± 0.70	71.40 ± 0.29	77.71 ± 1.38
Dropout	DWS	11.43 ± 2.44	14.71 ± 1.14	68.48 ± 0.14	75.57 ± 1.91
Quantile dropout	DWS	15.13 ± 2.45	29.88 ± 0.62	68.72 ± 0.27	76.22 ± 0.72
Translate	GNN	8.17 ± 0.81	34.93 ± 1.31	70.17 ± 1.26	$\mathbf{83.83 \pm 0.25}$
Rotation	GNN	—	—	69.35 ± 2.18	83.72 ± 1.14
Scale	GNN	8.58 ± 0.65	34.70 ± 5.19	68.96 ± 1.46	83.67 ± 0.19
Gaussian noise	GNN	9.06 ± 0.27	32.82 ± 1.14	77.55 ± 0.33	81.28 ± 0.50
SIREN bias	GNN	11.63 ± 2.48	34.32 ± 1.57	68.09 ± 0.49	77.20 ± 1.03
SIREN negation	GNN	11.41 ± 3.22	37.93 ± 2.26	72.74 ± 4.29	82.36 ± 3.66
Dropout	GNN	8.10 ± 0.43	18.04 ± 1.24	68.55 ± 1.21	79.72 ± 1.35
Quantile dropout	GNN	8.12 ± 0.85	34.36 ± 1.14	69.96 ± 2.08	83.78 ± 0.76
MixUp	DWS	26.96 ± 0.91	31.92 ± 0.37	74.36 ± 1.17	78.58 ± 0.20
MixUp + random perm.	DWS	26.62 ± 0.18	$\mathbf{33.55 \pm 1.40}$	73.89 ± 0.89	78.04 ± 1.02
Alignment + MixUp	DWS	$\mathbf{27.40 \pm 0.97}$	33.33 ± 0.43	$\mathbf{75.67 \pm 0.36}$	$\mathbf{79.41 \pm 0.56}$
MixUp	GNN	20.45 ± 3.82	42.25 ± 3.83	$\mathbf{80.18 \pm 0.59}$	82.20 ± 0.52
MixUp + random perm.	GNN	24.46 ± 2.92	41.67 ± 4.55	78.45 ± 2.29	82.24 ± 0.68
Alignment + MixUp	GNN	$\mathbf{26.88 \pm 1.75}$	$\mathbf{42.83 \pm 4.18}$	78.80 ± 2.12	82.94 ± 0.31