

Figure 1. SGT on ZINC: Test MAE v.s. Batch Size (BS). # Training epochs are adjusted per batch-size for the same total update steps: $400 * BS/32$. The first 10% epochs are in the warmup stage.

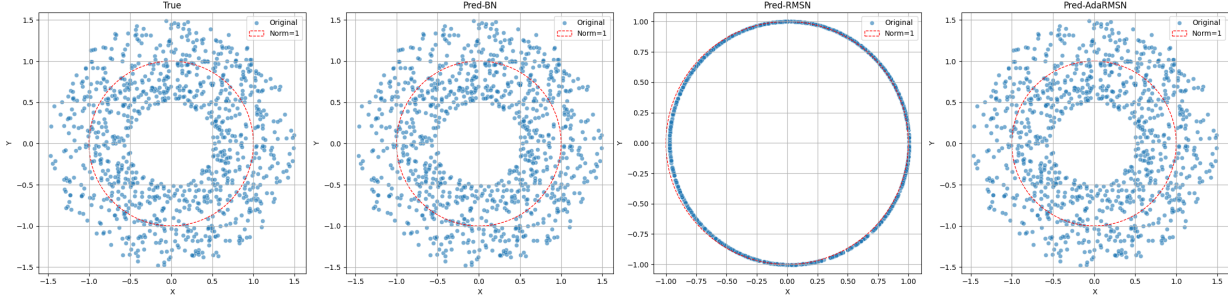


Figure 2. (Case Study of AdaRSMN) Visualization of Input and Pred data points. Overfitting test on Auto-encoders of 2-dim ($Linear \rightarrow BN/RMSN/AdaRMSN \rightarrow Linear$): each model is trained 5000 epochs via AdamW without regularization. (together with Figure. 3)

PCMQM4Mv2	Val MAE (\downarrow)	# Param.
Graphormer	0.0864	48.3M
GPS	0.0858	19.4M
GRIT	0.0859	16.6M
SGT	0.0856	17.6M

Table 1. Performance on PCMQM4Mv2. The eval. pipeline follows Rampásek et al. (2022); no 3D-info included.

ZINC	GPS	GPS+ L_2	SGT
MAE (\downarrow)	0.070 ± 0.004	0.0693 ± 0.0023	0.0566 ± 0.002

Table 2. Performance on ZINC. GPS+s L_2 : integrating s L_2 attention into GPS without changing other parts. (run 3 trials)

Model	GPU Memory (GB)	Training Time (Sec/Epoch)
GRIT	29.16	141.60
SGT	25.07	100.68

Table 3. Comparison of peak GPU memory usage and per-epoch training time for GRIT and SGT.

Dataset: Peptides-Structure (15K graphs);

Model config.: 5 transformer layers, 96 channels, batch size 32.

Hardware: a single Nvidia V100 GPU with 32GB memory, supported by 80 Intel Xeon Gold 6140 CPUs running at 2.30GHz

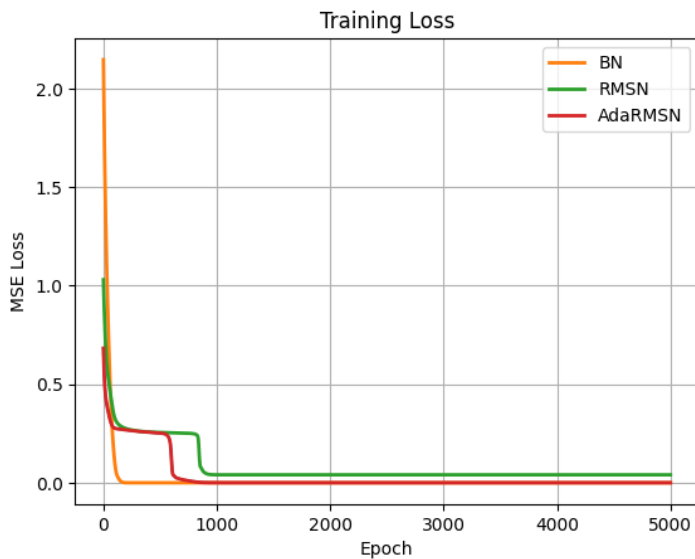


Figure 3. (Case Study of AdaRSMN) Training curves of overfitting test. (together with Figure. 2)

Model	ZINC MAE (\downarrow)	SP-CIFAR Acc. (\uparrow)	SP-MNIST Acc. (\uparrow)	PATTERN W.Acc. (\uparrow)	CLUSTER W.Acc. (\uparrow)	Peptides-Struct MAE (\downarrow)	Peptides-Func AP (\uparrow)
Expformer	-	74.69 \pm 0.125	98.55 \pm 0.037	86.74 \pm 0.015	78.07 \pm 0.037	0.2481 \pm 0.0007	0.6527 \pm 0.0043
GEAET	-	76.634 \pm 0.427	98.513 \pm 0.086	86.993 \pm 0.026	-	0.2445\pm0.0013	-
GEANet	0.193 \pm 0.001	73.857 \pm 0.306	98.315 \pm 0.097	85.607 \pm 0.038	77.013 \pm 0.224	0.2512 \pm 0.0003	0.6722 \pm 0.0065
SGT	0.0566\pm0.002	78.560\pm0.700	98.614\pm0.096	89.752\pm0.030	80.027\pm0.114	0.2450 \pm 0.0017*	0.6961\pm0.0062

Table 4. Performance comparison across different models on various datasets. Best results are highlighted in bold. * indicates the difference to the best is not statistically significant (by two-tail T-test)

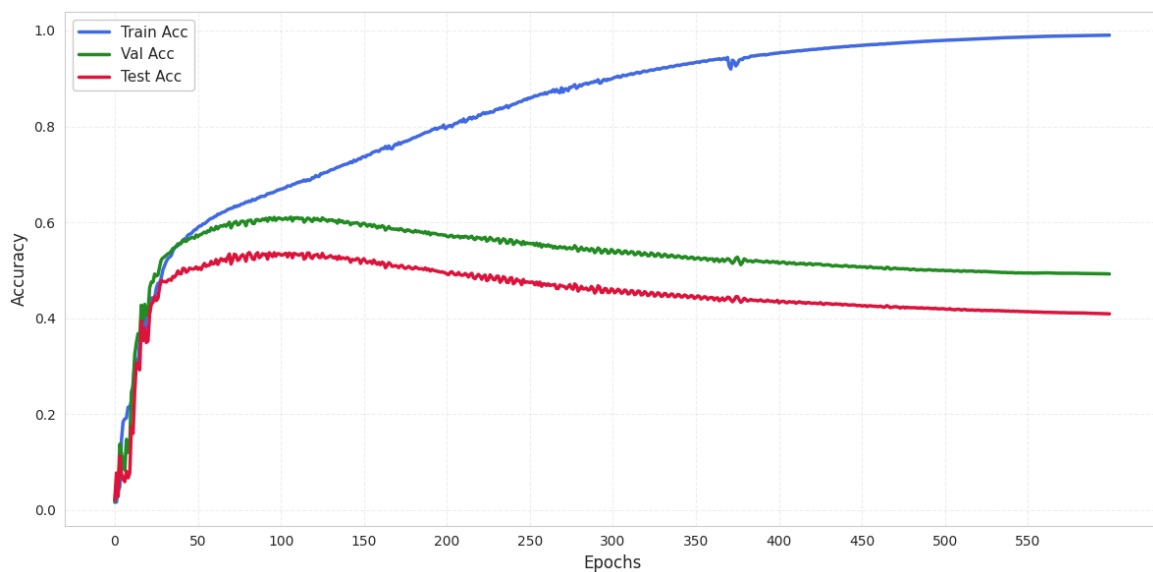


Figure 4. Overfitting Test of Exp-SGT on large-scale graph OGBN-ArXiv (169,343). Using the same configuration as Expformer and remove all regularization to validate the trainability.