Supplementary Material for Wavy Transformer

Anonymous Author(s)
Affiliation
Address
email

Note to reviewers. The following section is an *extended version* of material already included in the main submission. It is provided here for ease of review only; in the camera-ready version, it will be moved to Appendix F of the main manuscript and supplied as part of the official appendix. Please note that this relocation will update the reference and table numbers accordingly.

Throughput–Accuracy Trade-off Comparison

Table 1 compares the inference throughput (images/s) of various transformer variants, measured on the ImageNet validation set using V100 GPUs and a batch size of 256. Introducing wave dynamics (Diffuse+Wave or Wave alone) reduces throughput by roughly 50% from about 2 600 to 1 200–1 300 images per second which may pose a limitation in real-time applications. However, this slowdown must be balanced against the gains in representational power and accuracy that wave-enhanced models provide. A simple mitigation is to insert wavy blocks in only a subset of layers to recover part of the lost throughput. Concretely, replacing only the last six blocks with wavy blocks (Wave (6)) increases throughput by about 33 % relative to the full Diffuse+Wave with FeatScale variant (1 649 img/s vs. 1 241 img/s) while sacrificing 0.18 pt in Top-1 accuracy (72.44 % vs. 72.62 %). In this experiment, the velocity tensor was initialized to zero at the first wavy block, and all other conditions were identical to those of the full Diffuse + Wave with FeatScale variant.

Table 1: Inference throughput and accuracy on ImageNet. Wave (6) indicates that wavy blocks are inserted only in the final six layers.

Model	Residual Connections	Throughput	Top-1 (%)
DeiT-Ti [1]	Diffuse	**2632.1**	72.17
DeiT-Ti + FeatScale [2]	Diffuse	2483.2	72.35
DeiT-Ti	Diffuse+Wave	1266.0	72.33
DeiT-Ti + FeatScale	Diffuse+Wave	1241.4	**72.62**
DeiT-Ti	Wave	1312.2	–
DeiT-Ti + FeatScale	Diffuse+Wave (6)	1649.3	72.44

References

[1] Hugo Touvron, Matthieu Cord, Matthijs Douze, Francisco Massa, Alexandre Sablayrolles, and Herve Jegou. Training data-efficient image transformers & distillation through attention. In *International Conference on Machine Learning*, pages 10347–10357, 2021.

[2] Peihao Wang, Wenqing Zheng, Tianlong Chen, and Zhangyang Wang. Anti-oversmoothing in deep vision transformers via the fourier domain analysis: From theory to practice. In *International Conference on Learning Representations*, 2022.

Submitted to 39th Conference on Neural Information Processing Systems (NeurIPS 2025). Do not distribute.