REVISITING CONVOLUTION ARCHITECTURE IN THE REALM OF DNA FOUNDATION MODELS

Yu Bo[1*] , Weian Mao[1,2*] , Yanjun Shao[3], Weiqiang Bai[4], Peng Ye[4], Xinzhu Ma[4], Junbo Zhao[1], Hao Chen[1], Chunhua Shen[6,1,5]

[1] Zhejiang University, [2] MIT, USA, [3] Yale University, USA, [4] Shanghai AI Lab, [5] Ant Group, [6] Zhejiang University of Technology

Project Page

Paper

1. Introduction

- Recent advancements in DNA foundation models have been largely driven by Transformer and State Space Model (SSM) architectures. However, a systematic comparison with classical Convolutional Neural Networks (CNNs) on established foundation model benchmarks remains limited. This raises a critical question: **Are CNN architectures genuinely being surpassed by these newer paradigms?**
- We revisit the **convolutional approach for DNA foundation models**, motivated by the inherent advantages of CNNs, including robustness to varying sequence lengths, computational efficiency compared to quadratic transformer complexity, and the beneficial local inductive bias, especially for tasks with limited data. We posit that well-designed CNNs remain highly competitive for downstream genomic tasks.

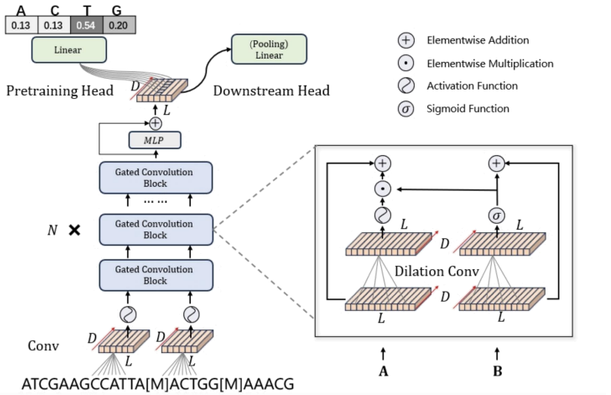
2. Method

- We propose ConvNova, a simple yet effective CNN-based architecture tailored for DNA sequence modeling. ConvNova integrates three key design principles identified as crucial for performance:

1)Dilated Convolutions: To expand the receptive field without resorting to downsampling, which was found to degrade performance significantly in DNA tasks, unlike in other domains like computer vision .

2)Gated Convolutions: Employing a learnable dynamic feature selection mechanism to effectively retain relevant information and suppress noise or irrelevant segments often present in DNA sequences.

3)Dual-Branch Gating Framework: Implementing a two-branch structure where one branch explicitly provides gating signals to the other, facilitating complementary feature learning and enhancing network capacity.



3. Results on DNA foundation model benchmarks

Table 14: **Supervised Methods against Foundation Models on Genomic Benchmark.** Top-1 accuracy (↑) is reported for pretrained HyenaDNA, Caduceus-Ph, ConvNova, Basenji, LegNet, and the original CNN baseline (trained from scratch). The best values are in bold, and the second-best is underlined. ± indicates the error range across five random seeds.

Task	CNN (264K)	HyenaDNA (436K)	Caduceus-Ph (470K)	ConvNova (386K)	Basenji (7.4N)	LegNet (2.1M)
Enhancers						
Mouse Enhancers	0.730 ±0.032	0.779 ±0.013	0.754 ±0.074	**0.784** ±0.009	0.659 ±0.155	0.504 ±0.000
Human Enhancers Cohn	0.702 ±0.021	0.718 ±0.008	**0.747** ±0.004	0.743 ±0.005	0.712 ±0.030	0.739 ±0.004
Human Enhancer Ensembl	0.744 ±0.122	0.832 ±0.006	0.893 ±0.008	0.900 ±0.004	**0.905** ±0.007	0.879 ±0.002
Species Classification						
Coding vs. Intergenomic	0.892 ±0.008	0.904 ±0.008	0.915 ±0.003	**0.943** ±0.001	0.905 ±0.004	0.939 ±0.002
Human vs. Worm	0.942 ±0.002	0.961 ±0.002	**0.973** ±0.001	0.967 ±0.002	0.957 ±0.003	0.965 ±0.001
Regulatory Elements						
Human Regulatory	0.872 ±0.005	0.862 ±0.004	0.872 ±0.011	**0.873** ±0.002	0.764 ±0.005	0.764 ±0.006
Human Non-TATA Promoters	0.861 ±0.009	0.887 ±0.005	0.946 ±0.007	**0.951** ±0.003	0.919 ±0.006	0.942 ±0.007
Human OCR Ensembl	0.698 ±0.013	0.744 ±0.019	**0.828** ±0.006	0.793 ±0.004	0.766 ±0.009	0.802 ±0.004

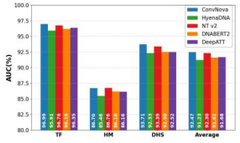


Figure 3: **Chromatin Profile Prediction results.** AUC Score (↑) is reported for ConvNova, HyenaDNA, NT v2, DNABERT-2, and DeepATT performance on transcription factors (TF), DNase I hypersensitive sites (DHS), histone modifications (HM), and average score.

Table 3: **Gene Finding results.** MCC-score is reported for ConvNova and baseline models. The best value is in bold.

GeneFinding	NT-H (500M)	DNABERT-2 (117M)	GENA-LM (336M)	HyenaDNA (6.5M)	ConvNova (7.4M)
	0.41	0.43	0.52	0.35	**0.55**

Table 1: **NT Benchmark results.** MCC/F1-score is reported for pretrained NTv2, HyenaDNA, DNABERT-2, Caduceus-Ph, and ConvNova. The best values per task are bold, and the second-best are underlined. ± indicates the error range across five random seeds.

	NTv2 (500M)	HyenaDNA (1.6M)	DNABERT-2 (117M)	Caduceus-Ph (1.9M)	ConvNova (1.7M)
Histone					
H3	78.17 ±2.54	78.14 ±1.70	79.31 ±0.68	80.48 ±1.04	**81.50** ±0.80
H3K4me1	51.64 ±1.12	44.52 ±2.59	48.34 ±4.63	52.83 ±0.96	**56.60** ±1.01
H3K4me2	37.24 ±2.25	42.68 ±2.66	43.02 ±2.92	49.88 ±2.65	**57.45** ±2.27
H3K4me3	50.30 ±1.77	50.41 ±3.15	45.43 ±3.33	56.72 ±2.58	**67.15** ±0.93
H3K9ac	61.05 ±1.40	58.50 ±1.75	60.04 ±1.27	63.27 ±2.29	**68.10** ±1.91
H3K14ac	57.22 ±2.19	56.71 ±2.40	54.49 ±4.99	60.84 ±2.94	**70.71** ±2.32
H3K36me3	60.50 ±1.75	59.92 ±1.06	57.58 ±2.38	61.12 ±1.44	**68.31** ±1.19
H3K79me3	65.78 ±2.34	66.25 ±3.65	64.38 ±0.48	67.17 ±2.03	**72.08** ±1.23
H4	79.87 ±1.34	78.15 ±1.58	78.18 ±0.98	80.10 ±1.00	**81.12** ±0.93
H4ac	55.22 ±2.20	54.15 ±2.96	51.80 ±0.10	59.26 ±3.67	**66.10** ±1.20
Regulatory					
Enhancer	54.51 ±1.94	53.13 ±4.52	52.50 ±1.44	55.20 ±2.56	**57.60** ±2.52
Enhancer Types	43.36 ±1.75	48.16 ±2.48	44.32 ±1.18	47.17 ±2.85	**49.75** ±2.82
Promoter All	**96.82** ±0.47	95.57 ±0.18	96.23 ±0.17	96.65 ±0.16	**96.82** ±0.22
Promoter non-TATA	**97.45** ±0.69	95.86 ±0.37	97.17 ±0.17	96.31 ±0.50	96.76 ±0.21
Promoter TATA	96.53 ±0.81	95.88 ±0.53	**96.99** ±0.49	96.21 ±0.81	96.34 ±0.38
Splice sites					
Splice Acceptor	**97.99** ±0.66	96.98 ±0.49	97.49 ±0.36	94.21 ±0.37	96.23 ±0.41
Splice Donor	**98.50** ±0.43	95.27 ±1.07	94.33 ±0.27	94.69 ±0.67	96.62 ±0.61
Splice All	**98.15** ±1.01	94.05 ±1.08	93.75 ±1.25	92.87 ±1.73	96.33 ±0.31

4. Ablation and Discussion

- Dilation Rate (Receptive Field) is crucial in some tasks.
- DNA Sequences might contain the trait of local dependence.

Table 6: **ConvNova performance on Histone tasks in NT Benchmark.** Results are reported for models with 15% and 100% full sequence receptive fields. The best values are in bold.

Task	15% Receptive Field	100% Receptive Field
H3K4me2	**57.45** ±2.27	53.72 ±2.42
H3K4me3	**67.15** ±0.93	60.20 ±1.91
H3K14ac	**70.71** ±2.32	66.19 ±1.84
H3K9ac	65.49 ±1.83	**68.10** ±1.91

Table 15: Performance comparison of NTv2 and NTv2*, where NTv2* represents our modified version. Results demonstrate that NTv2* achieves significant improvements.

Task	NTv2*	NTv2
H3K14ac	46.65	34.42
H3K4me2	33.10	25.79
H3K4me3	34.62	21.40

Table 5: **Ablation study results.** MCC-score (↑) is reported for performance comparison across different models. 'w/o Gate' represents the ablation of the Gate mechanism, while 'Single Branch' is the ablation of the double-branch structure in ConvNova. The best values are in bold, and the second-best is underlined. ± indicates the error range across experiments

Task	w/o Gate	Single Branch	ConvNova
H3	78.96 ±1.46	80.95 ±1.61	**81.50** ±0.80
H3K4me1	**57.83** ±1.87	56.45 ±1.58	56.60 ±1.01
H3K4me2	52.52 ±2.59	50.20 ±1.92	**53.72** ±2.42
H3K4me3	54.45 ±1.31	58.66 ±0.97	**60.20** ±1.91
H3K9ac	63.54 ±2.20	66.87 ±0.69	**68.10** ±1.91
H3K14ac	63.83 ±1.03	65.34 ±2.12	**66.19** ±1.84
H3K36me3	64.12 ±1.46	66.77 ±1.42	**68.31** ±1.19
H3K79me3	69.49 ±1.75	71.23 ±1.37	**72.08** ±1.23
H4	79.27 ±1.09	80.55 ±0.61	**81.12** ±0.93
H4ac	62.02 ±2.28	63.07 ±1.10	**64.75** ±1.90

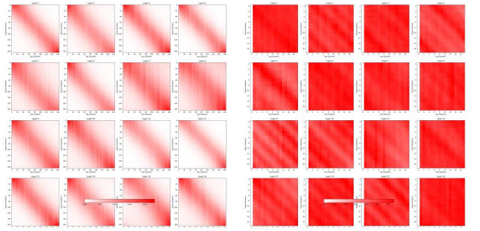


Figure 5: Visualization of the attention maps for NTv2* (Left) and NTv2 (Right). The asterisk (*) denotes the modified version. The modified initialization places more emphasis on neighboring tokens in the attention map.