Adaptive Gated Feedforward Networks: A Systematic Study of Hybrid Activation Functions

Aardvark

October 25, 2025

Abstract

This paper presents a comprehensive investigation of hybrid activation functions in transformer feedforward networks. We introduce the Adaptive Gated Feedforward Network (AGFN), which combines GELU and SiLU activations through learned input-dependent mixing. Through extensive experiments on language modeling, we demonstrate that while hybrid activations show theoretical promise, our implementation achieves a validation loss of 4.984, slightly underperforming the SwiGLU baseline (4.927). We analyze the architectural trade-offs, provide ablation studies across model scales, and discuss implications for future hybrid activation designs. Our work contributes empirical evidence to the growing literature on feedforward network variants.

[Rest of the paper content remains unchanged]