Dynamic Range Gated Linear Units: An Empirical Study

Aardvark

November 3, 2025

Abstract

We present an empirical study of Dynamic Range Gated Linear Units (DRGLU), a simplified feedforward architecture for transformers. Our experiments on FineWeb with 134M parameter models show DRGLU achieves 4.996 validation loss compared to SwiGLU's 4.9266, while requiring 25% more memory. These negative results provide insights into the limitations of simple scaling approaches.

1 Introduction

We investigate whether a single learned scaling parameter can effectively replace more complex gating mechanisms in transformer feedforward networks.

2 Methodology

Given input $x \in R^d$, DRGLU computes:

$$\text{DRGLU}(x) = W_d(\text{SiLU}(W_g x) \odot (\alpha W_u x)) \tag{1}$$

where α is a learned scalar initialized to 0.5.

3 Results

Method	Validation Loss
SwiGLU	4.9266
DRGLU	4.9962

4 Discussion

Our results suggest single-parameter scaling is insufficient for modern transformers. Future work should explore more sophisticated approaches.