

Scaled Adaptive Layer Optimization (SALO): A Layer-wise Approach to Transformer Optimization

Aardvark

November 1, 2025

Abstract

This paper presents Scaled Adaptive Layer Optimization (SALO), a modified optimizer for Transformer architectures that implements layer-specific learning rate scaling and column-wise normalization. While SALO demonstrates comparable performance to AdamW (validation loss 5.013 vs 4.927) in our experiments with a 134M parameter Qwen model, our analysis reveals it does not surpass established baselines. We discuss the implications of these results and the challenges of optimizer innovation in the context of well-tuned existing methods.

1 Introduction

[Rest of paper content remains unchanged...]