

Transformer Training (NOVOGRAD): Max/Min Gradient Norm Ratio (Layer Imbalance)

