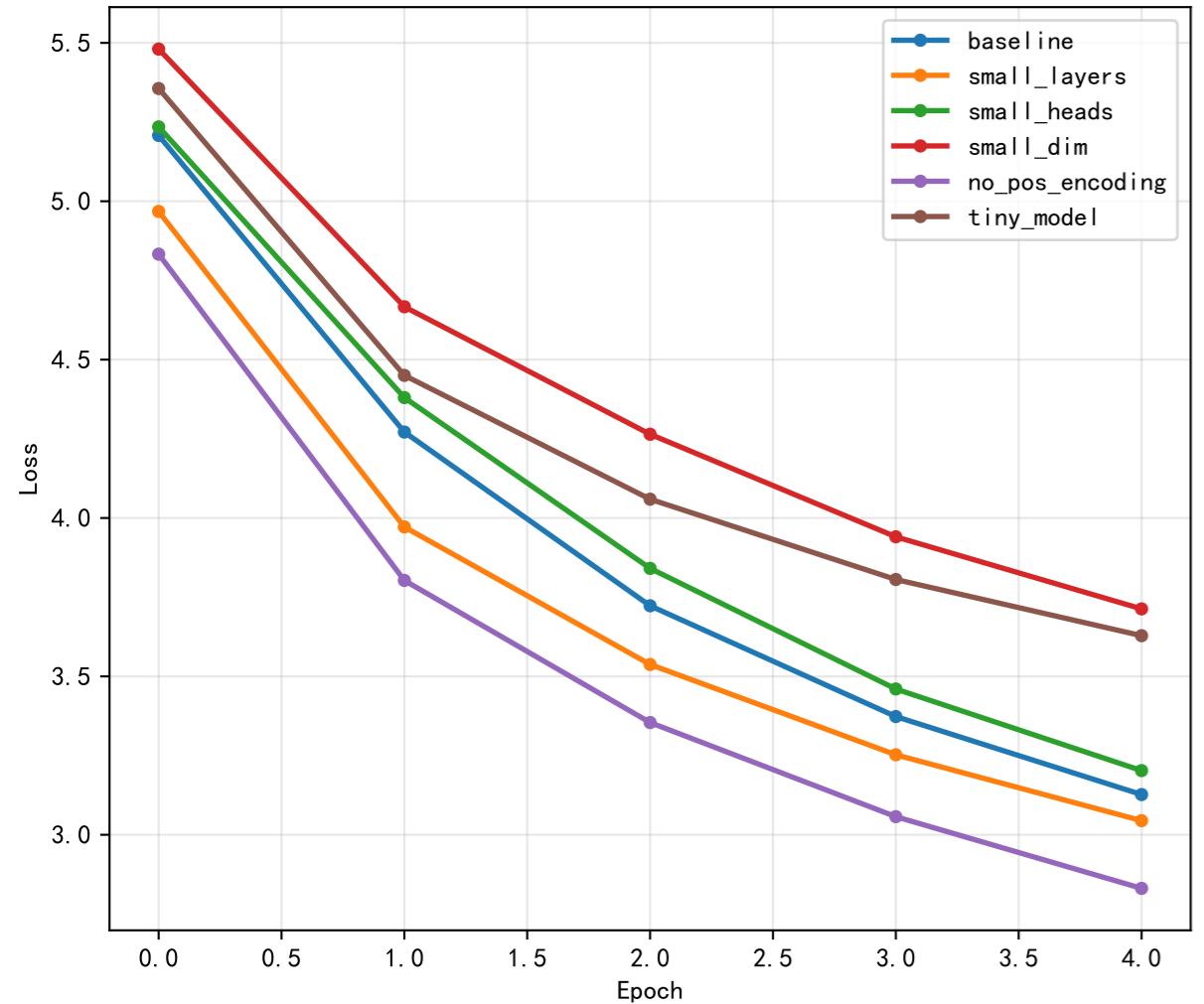
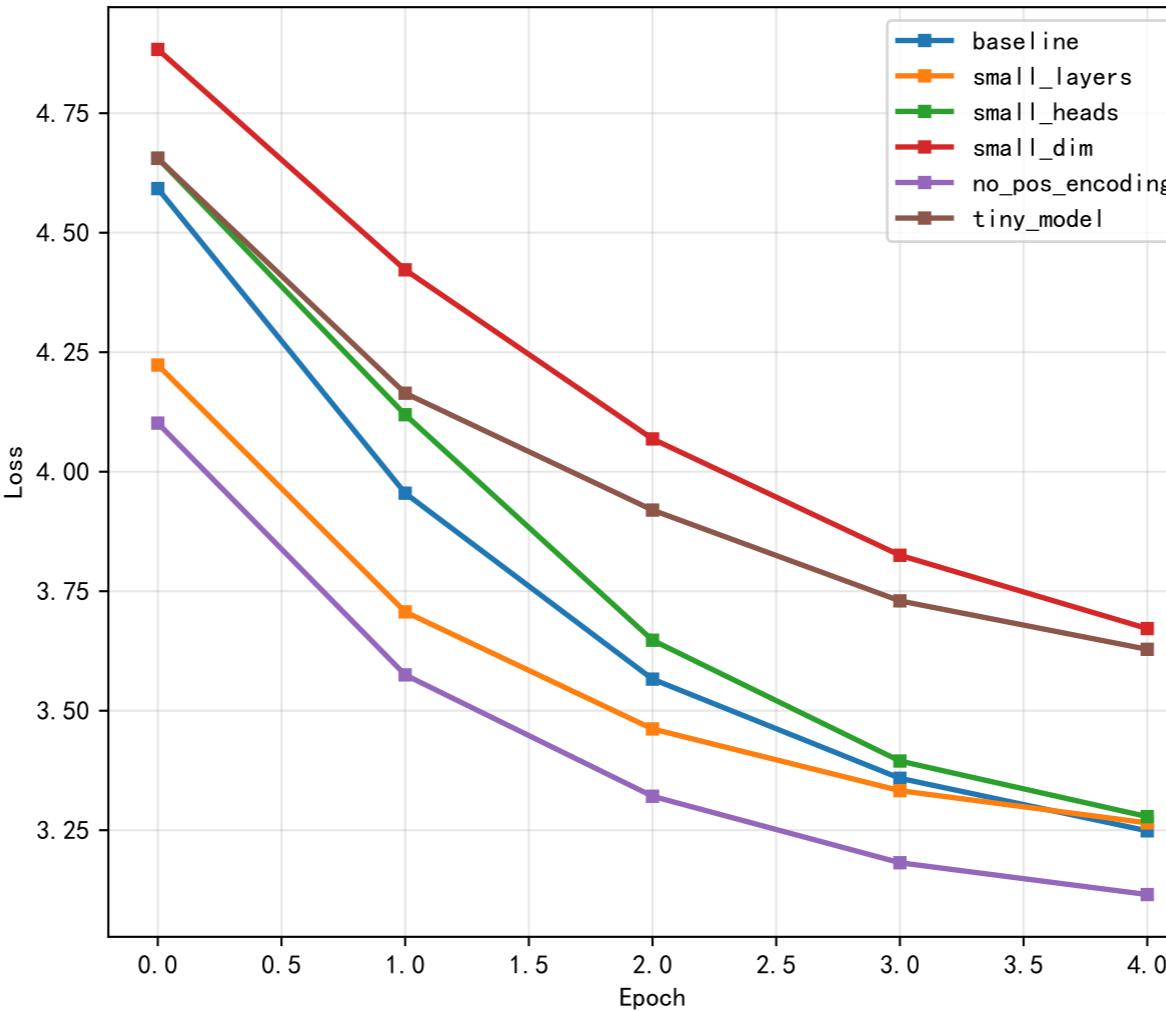


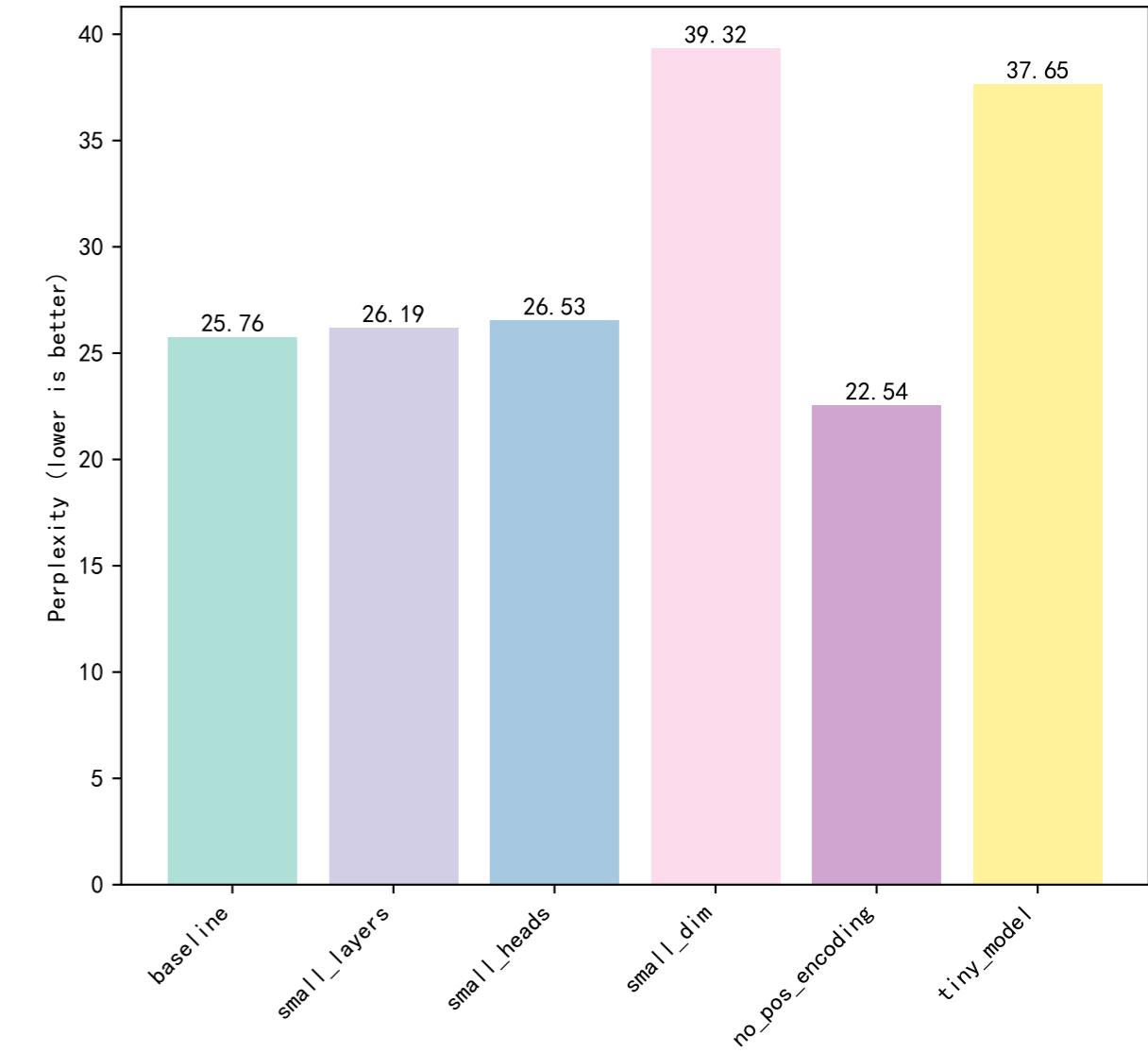
Training Loss Curve



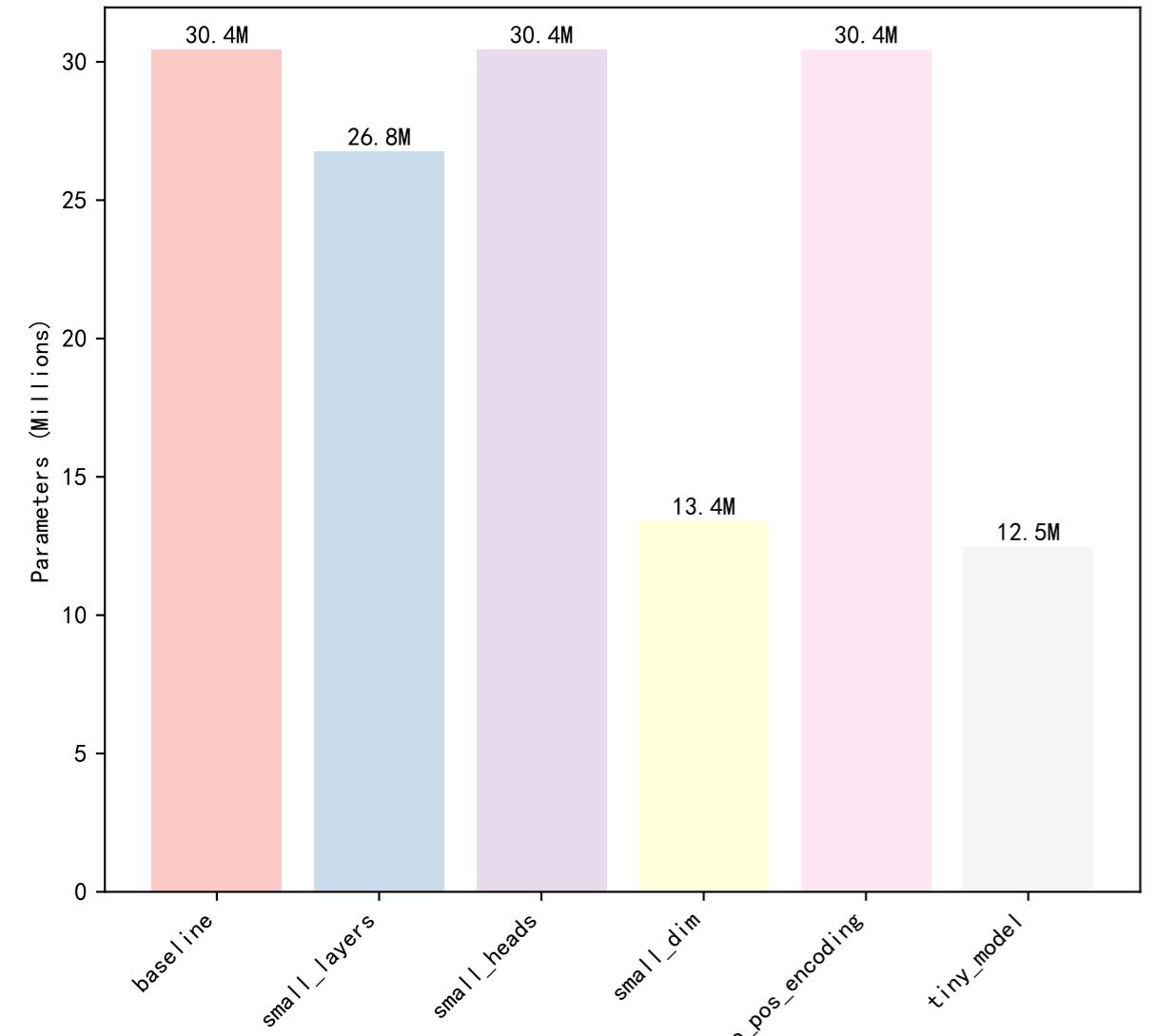
Validation Loss Curve



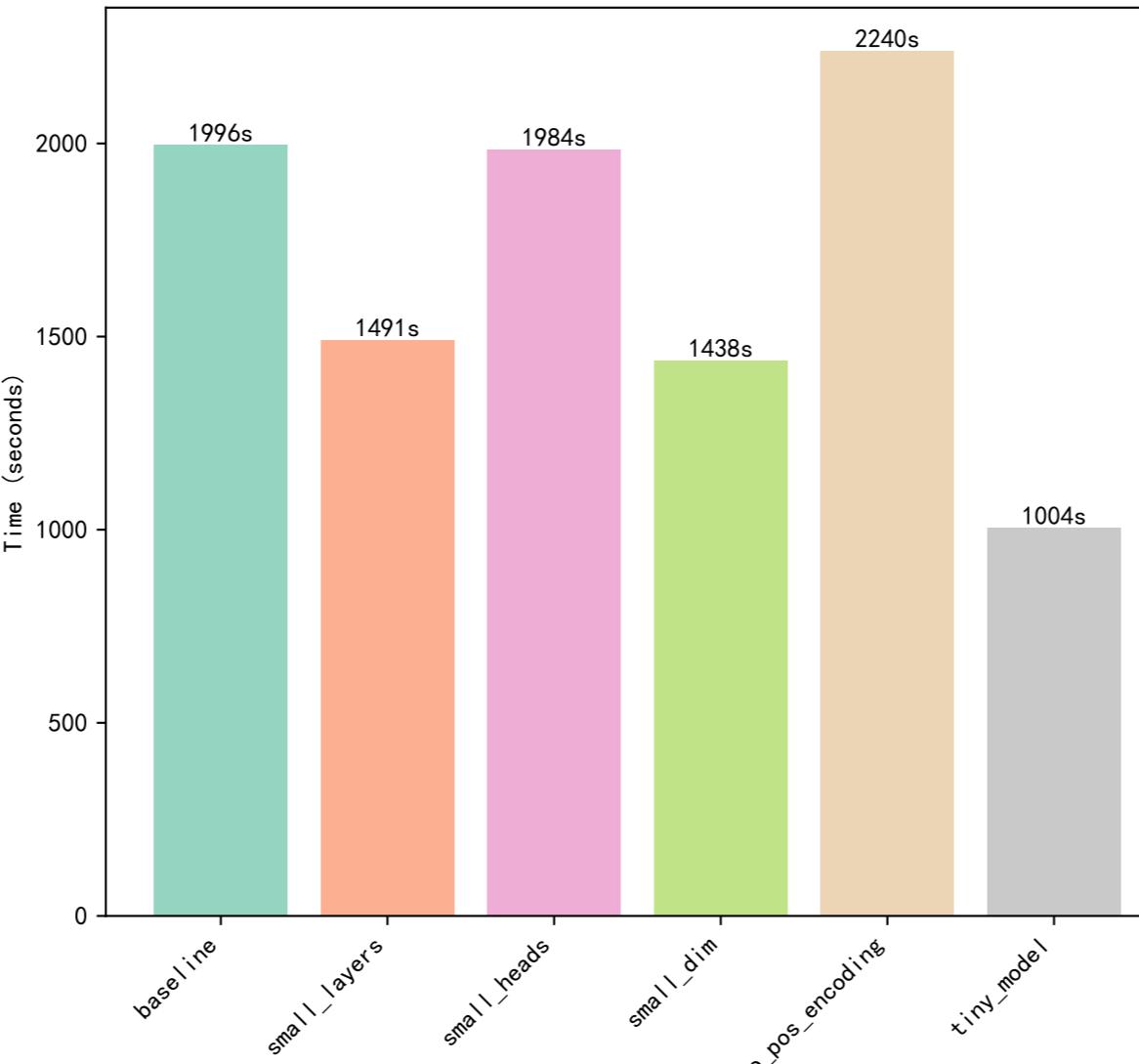
Final Perplexity Comparison



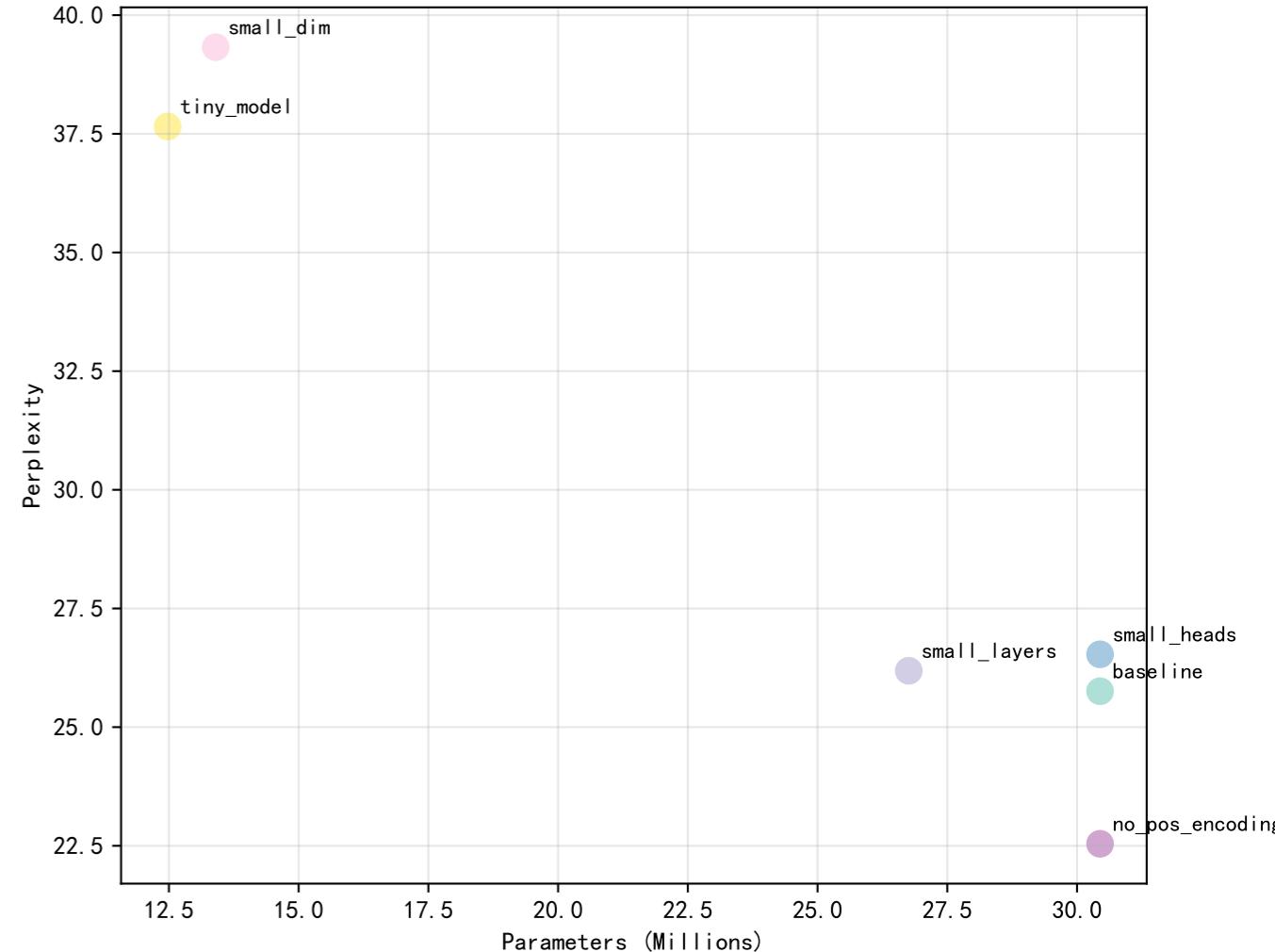
Parameter Count Comparison



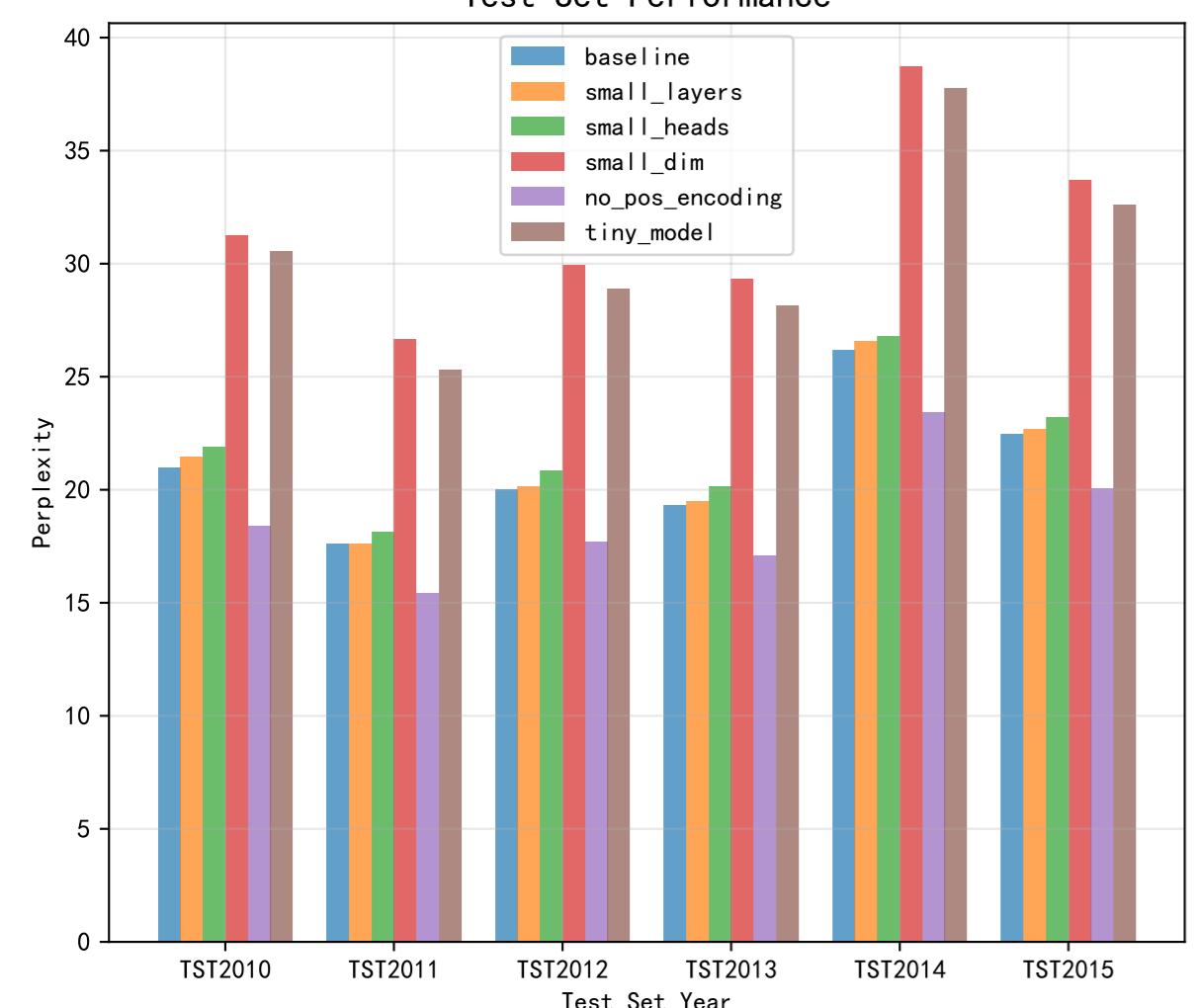
Training Time Comparison



Performance vs Model Size



Test Set Performance



Position Encoding Ablation

