

Effect of residual connection on perplexity evolution for 4-layer network

