Exploring the limits of Transfer learning with a Unified Text-to-Text Transformer

2.1. Model

- roughly equivalent to the original Transformer
- removing the Layer Norm bias
- placing the layer normalization outside the residual path
- using different position embedding scheme.

2.4. Input & Output Format

- "text-to-text" format
- provides a consistent training objective both for pre-training & fine-tuning