

Heatmap visualization of attention weights for the Transformer encoder. The y-axis represents the Layer (0 to 11), and the x-axis represents the Head (0 to 11). The color scale on the right indicates the magnitude of the attention weight, ranging from 0.00 (dark purple) to 0.05 (light yellow). The heatmap shows a strong diagonal pattern, indicating that each layer primarily attends to its own input. Notable high-attention regions (yellow/orange) are observed in Layer 5, Head 8, and Layer 6, Head 0.

