

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 attention weights, ranging from 0.10 (dark purple) to 0.14 (light orange). The heatmap shows a strong diagonal pattern, indicating that each layer primarily attends to its corresponding input tokens. A notable dark purple square is visible at Layer 5, Head 10, suggesting a specific attention mechanism or a potential artifact.

