

Figure 10 is a line graph illustrating the performance of various attention heads (Head-0 to Head-11) across different relative distances (0 to 500). The x-axis represents the Relative Distance, and the y-axis represents the performance metric. The graph shows that performance generally decreases as the relative distance increases for most heads. Head-0 consistently shows the highest performance, while Head-11 shows the lowest performance, especially at larger distances. The performance of most heads stabilizes or slightly declines after a distance of 400.

