

Figure 1 is a line graph illustrating the performance of different models (raw, avg, thrs) across various tasks. The x-axis represents different tasks, and the y-axis represents performance. The legend indicates that 'raw' lines are highly volatile, 'avg' lines are smoother, and 'thrs' lines represent a threshold performance level. The graph shows that the 'avg' lines generally perform better than the 'raw' lines, and the 'thrs' lines are consistently high, indicating a high threshold performance level.

