1. Continued Figures

We extend Fig.5-6 in the main paper with metrics on \( k = 10 \) and \( k = 50 \). Similarly, we find that the dynamic weighting strategy is a good tradeoff between Video-to-Text and Text-to-Video tasks and manifests stability in different trials.

Figure 1: We compare the model performance with different weighting strategies in the MeVTR loss on Recall@10 on the Video-to-Text task for ActivityNet Captions.

Figure 2: We compare the model performance with different weighting strategies in the MeVTR loss on Recall@50 on the Video-to-Text task for ActivityNet Captions.