

Supplementary Material: Multimodal LLM Enhanced Cross-lingual Cross-modal Retrieval

Anonymous Author(s)

ACM Reference Format:

Anonymous Author(s). 2024. *Supplementary Material: Multimodal LLM Enhanced Cross-lingual Cross-modal Retrieval*. In *MM '24: Proceedings of the 32th ACM International Conference on Multimedia*, October 10–14, 2024, Lisbon, Portugal. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/nnnnnnn.nnnnnnn>

We report more experimental results and technical details which are not included in the paper due to space limit:

- Analysis of the parameter λ (Section 1).
- Analysis of the parameter β (Section 2).

1 PARAMETER λ

As shown in Figure 1, we conduct an experiment to investigate the impact of the parameter λ , which balances the weight between English guidance loss and vision-non-English contrastive loss. When $\lambda = 0.0$, it indicates that we only use the English guidance loss. We can observe that the performance is best when we set the λ to

0.5. This result demonstrates the complementarity between English guidance loss and contrastive loss, enabling the model to learn comprehensive and robust inter-modal correspondence by incorporating English guidance.

2 PARAMETER β

The parameter β is used to control the weights of global similarities S_g and local similarities S_l during inference. Among them, global similarities S_g are calculated by global visual features and query features, and local similarities S_l are calculated by local semantic contexts and query features. To further investigate the influence of the parameter β during inference, we conduct an ablation study. As shown in Figure 2, we observe that incorporating the similarity score S_l leads to additional performance improvement. This can be attributed to the fact that the similarity score S_l can provide the local correspondence between modalities, which is complementary to the global similarity scores S_g . The performance is best when $\beta = 0.8$, and we set it to 0.8 in our experiments.

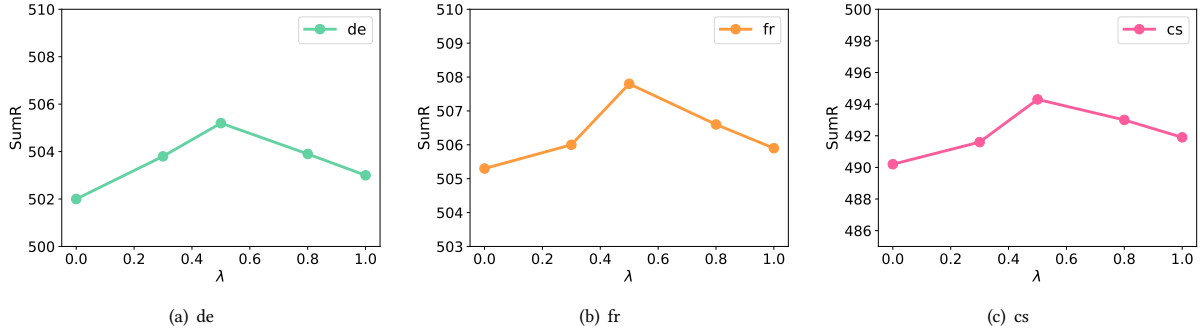


Figure 1: The performance of different values of λ .

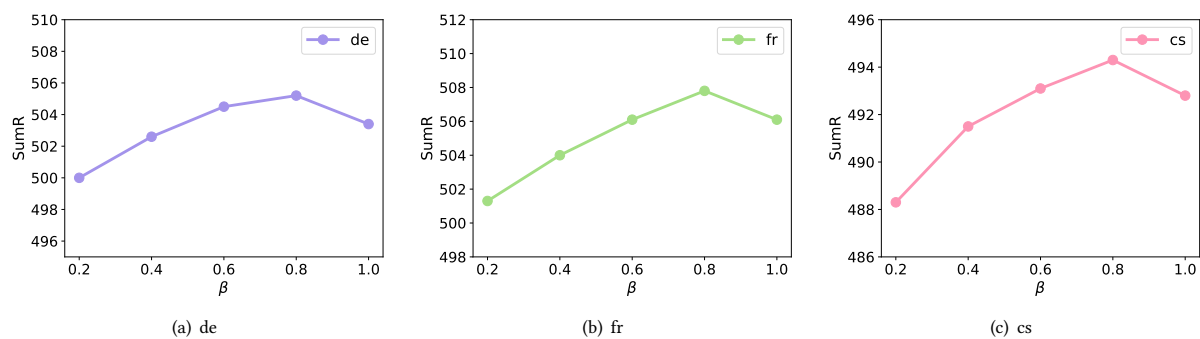


Figure 2: The performance of different values of β .