

In this revision, we clarified the definition of the target distribution in Equations (2)-(3), explaining the rationale behind the choices of distribution center and standard deviation. We also elaborated on the principles and implementation details of kernel density estimation (KDE) and DBSCAN. Furthermore, we provided a clearer explanation of the joint filtering strategy across the two dimensions - text quality and image-text relevance - and added a brief introduction to DBSCAN at its first mention. Previously ambiguous symbols and figure references have been corrected and standardized.

We also substantially revised the experimental analysis and interpretation. In response to concerns about DOSE underperforming random sampling on certain tasks (e.g., GPS@40%, TQA@5%), we analyzed sample diversity and distribution under different selection strategies. We clarified the comparative logic among strategies in Figures 2 and 3, standardized the notation of best and second-best results across all tables, and provided precise definitions for the reported performance metrics.

Additionally, to address concerns about the efficiency-performance trade-off, we expanded our time cost analysis in the appendix. We specifically explained why the Perplexity-based strategy incurs significantly higher time costs at larger sampling ratios. The appendix also details the time breakdown (data selection + fine-tuning) for each method under different sampling ratios, providing a more comprehensive perspective on computational cost.