# 1 ADDITIONAL MATERIAL

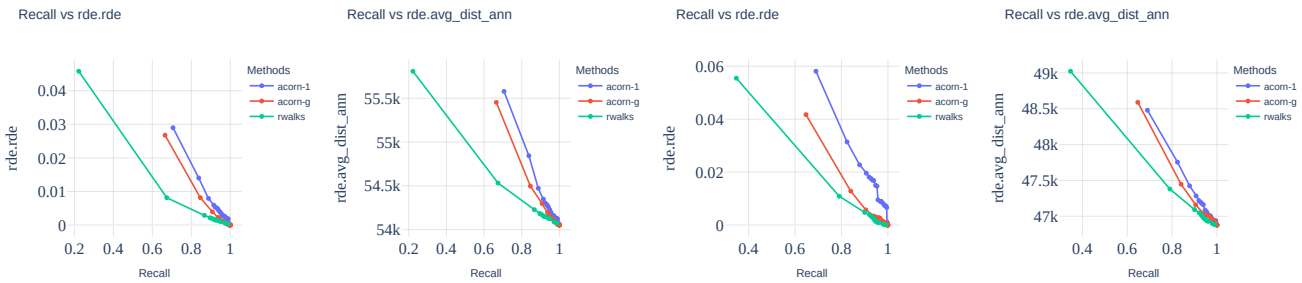
## 1.1 Query performance (Relative distance error)

In evaluating approximate nearest neighbors (ANN) algorithms, recall versus queries per second (QPS) is a common metric to assess the trade-off between accuracy and efficiency. However, recall alone does not fully capture the quality of retrieved neighbors, as it only measures set overlap without considering the relative distances to the query. To address this limitation, we incorporate relative distance error (RDE) as an additional metric. We define RDE as follows:

$$\text{RDE} = \frac{1}{N}\sum_{q=1}^{N}\left(\frac{\text{ANN\_avg}(q)}{\text{GT\_avg}(q)} - 1\right),$$

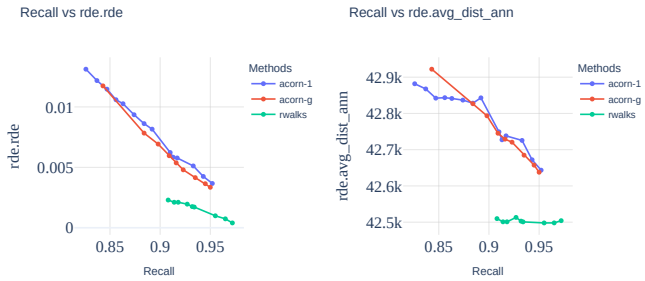
where $\text{ANN\_avg}(q)$ and $\text{GT\_avg}(q)$ are the average distances of intersecting ANN and ground truth neighbors for query $q$, and $N$ is the number of queries. RDE measures how closely ANN distances approximate the true distances.

We evaluate the performance of our method, **RWalks**, by tracking both recall and distance metrics. These results are compared against **Acorn**, the most competitive baseline.



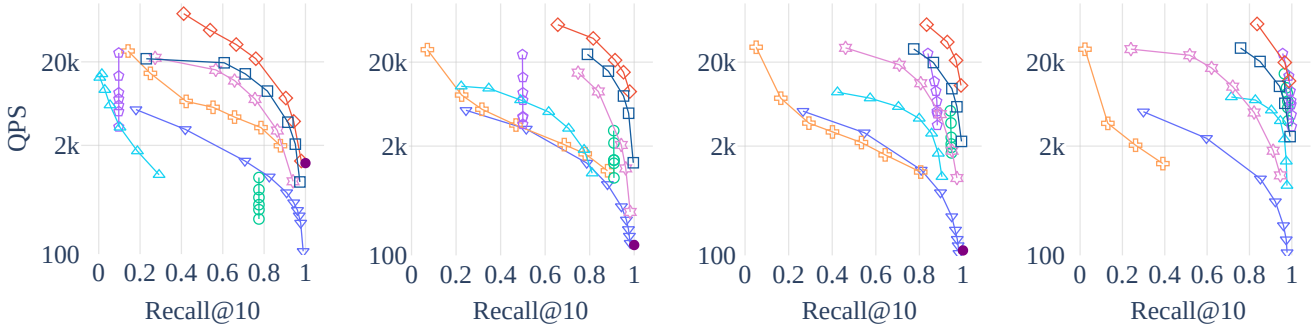

**(a) RDE - 5%** **(b) AVG distance - 5%** **(c) RDE - 10%** **(d) AVG distance - 10%**

**Figure 1: Relative distance error with varied specificity (Sift1M)**



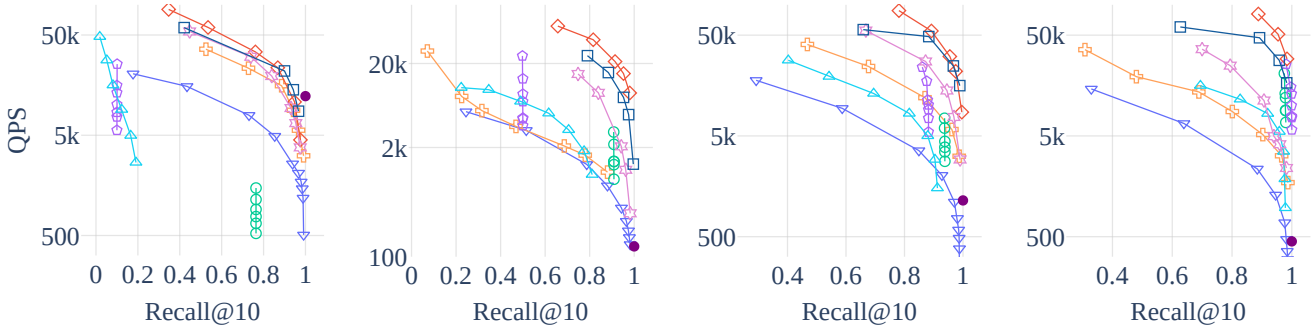

**(a) RDE - 10%** **(b) AVG distance - 10%**

**Figure 2: Relative distance error with varied specificity (Yfcc10M)**



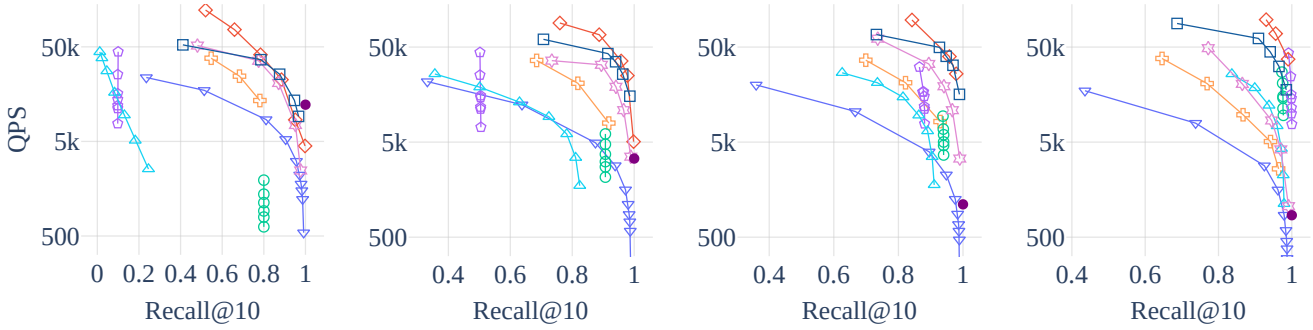

(a) 1% (b) 5% (c) 10% (d) 30%

**Figure 3: Query performance with varied specificity (Sift10M)**



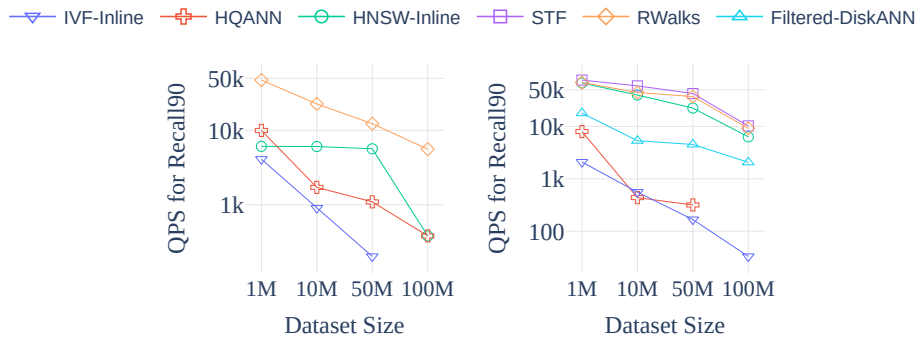

(a) 1% (b) 5% (c) 10% (d) 30%

**Figure 4: Query performance with varied specificity (Sift1M)**



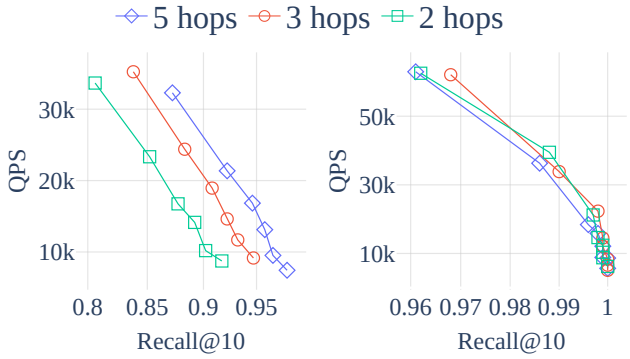

(a) 1% (b) 5% (c) 10% (d) 30%

**Figure 5: Query performance with varied specificity (Deep1M)**




(a) Low specificity (5%) (b) High specificity (30%)

**Figure 6: Query answering scalability with dataset size (Deep, Recall@10 = 0.9).**




(a) Low specificity (5%) (b) High specificity (30%)

**Figure 7: Effect of Random walks depth on QPS-Recall trade-off.**


1.2 Query performance on Sift10M
1.3 Query performance on Sift1M
1.4 Query performance on Deep1M
1.5 Impact of Dataset Size on Query Efficiency at High Recall (Deep dataset)
1.6 Impact of Random Walks Depth on the Accuracy/Efficiency Trade-offs
1.7 Impact of Random Walks number on the Accuracy/Efficiency Trade-offs
1.8 Impact of Search hyper-parameters on the Accuracy/Efficiency Trade-offs

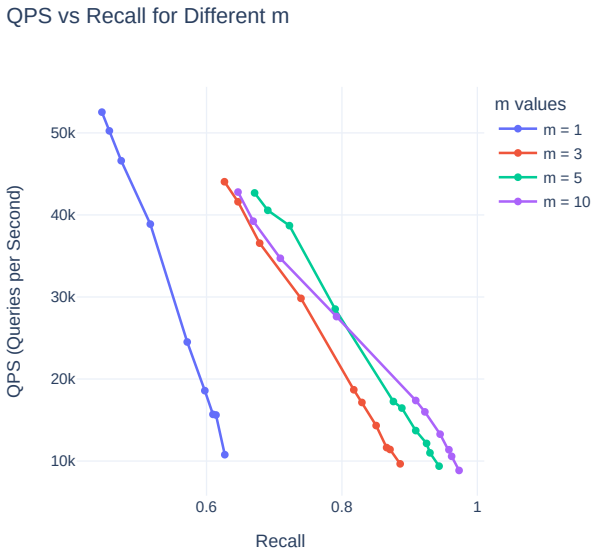
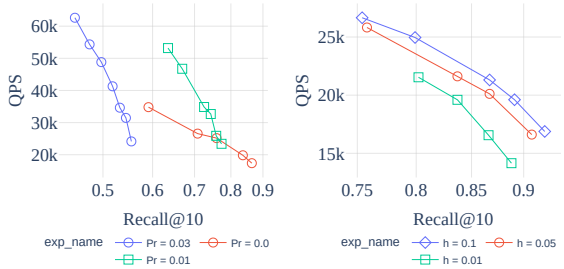



**Figure 8: Effect of Random walks number (m) on QPS-Recall trade-off.**




**(a) Pruning factor ($Pr$)** **(b) Hybrid factor ($h$)**

**Figure 9: Effect of search hyper-parameters on QPS-recall trade-off.**

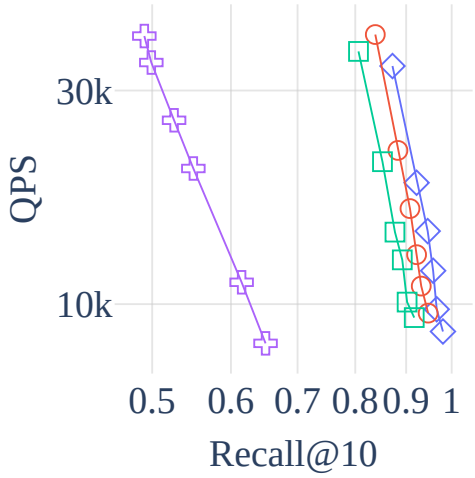



**Figure 10: Increasing walks depth (D) up to 10 decrease performances.**