

Distribution Consistency Guided Hashing for Cross-Modal Retrieval:

Supplementary Material

Anonymous Authors

1 SUMMARY

This supplementary material provides a comprehensive understanding of our proposed DCGH method. Specifically, we first present the whole algorithm procedure. Then, on the other three datasets, we provide the comparison results, convergence analysis, parameter analysis, and comparison on time cost to further evaluate our method.

2 ALGORITHM PROCEDURE

To better show the overview of the proposed method, we give the details of the proposed DCGH in Algorithm 1.

Algorithm 1: Distribution Consistency Guided Hashing

Require: Training data X^v , semantic label Y , query data K^v , code length r , and anchors k

- 1: Parameter setting: α , β , and λ ;
- 2: Initialize: B^v , H , W^v , U , F^v , and E ;
- 3: **repeat**
- 4: Update B^v with Eq.8;
- 5: Update H with Eq.11;
- 6: Update W^v with Eq.13;
- 7: Update U with Eq.15;
- 8: Update F^v with Eq.17;
- 9: Update E with Eq.19;
- 10: **until** Reach the convergence criteria;
- 11: Learn hash function P^v with Eq.20;
- 12: Obtain hash codes H of query set via Eq.21;

Ensure: Cross-modal retrieval results.

3 EXPERIMENTS

In this section, we conduct more experiments on the other three datasets (*i.e.*, WIKI, MIRFlickr-25K, and NUS-WIDE) to validate the performance of our proposed DCGH. Our experimental environment is MATLAB 2021b for Windows, installed on a host with 64GB of memory.

3.1 Datasets

More descriptions of the four datasets are given as follows.

WIKI: It is a single-label cross-modal dataset. It contains 2866 image-text pairs collected from Wikipedia, divided into 10 categories. In our approach, textual examples are represented as 10-dimensional topic vectors, while image examples are represented as 128-dimensional SIFT vectors.

MIRFlickr-25K: It is a cross-modal dataset comprising 25000 images sourced from the Flickr website, each associated with at

least one of the 24 textual labels. In our experiments, every visual instance is described as a 512-dimensional vector, while every textual instance is described as a 1386-dimensional vector. It is worth mentioning that we only select those instances with more than 20 textual labels.

IAPR-TC12: It comprises 20,000 geographical images, each paired with single or multiple textual descriptions. The textual descriptions are categorized into 255 classes. We employ GIST features to represent image data and BOW features to represent textual data, with dimensions of 512 and 2912, respectively.

NUS-WIDE: It is a large-scale dataset that contains 269684 images along with its corresponding semantic concepts chosen from the total 81 categories. Our experiments only select 186577 instances associated with the top-10 concepts. Each image instance is mapped into a 500-dimensional BOVW vector, and each textual instance is mapped into a 1000-dimensional BOW vector.

3.2 Setting

In our experiments, we perform some comparison experiments on fully paired datasets and partially paired datasets. Fig.1(a) shows the scenario where all data are fully paired. Fig.1(b) portrays the scenario of unpaired data.

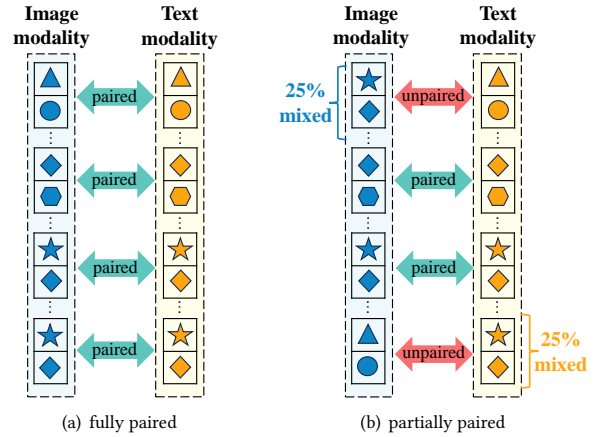


Figure 1: Examples of fully paired and partially paired datasets. (Shapes represent different categories, while colors represent different modalities)

3.3 Experimental Results

We conduct experiments with 11 comparison methods on the other three datasets with fully paired and partially paired. We report the PR and Top-k precision curves in Fig.2 and Fig.3 on fully paired

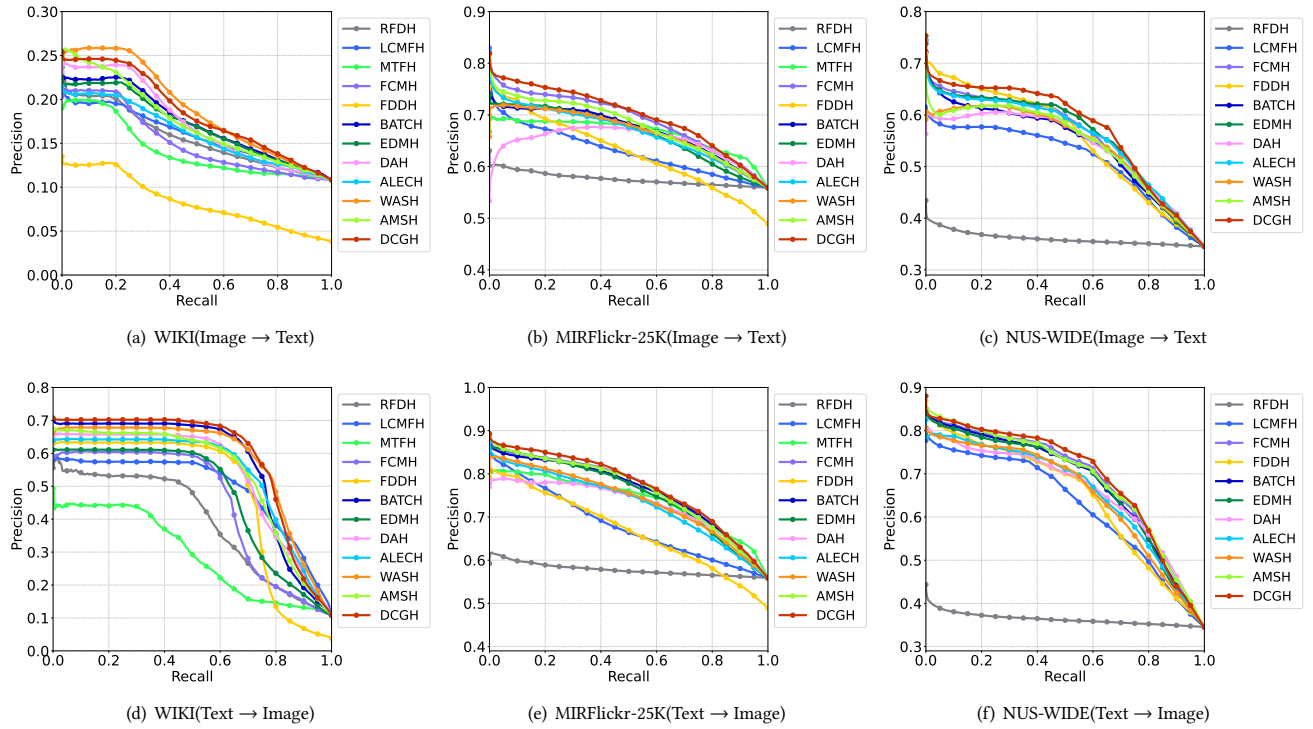


Figure 2: PR curves with 8 bits on three datasets.

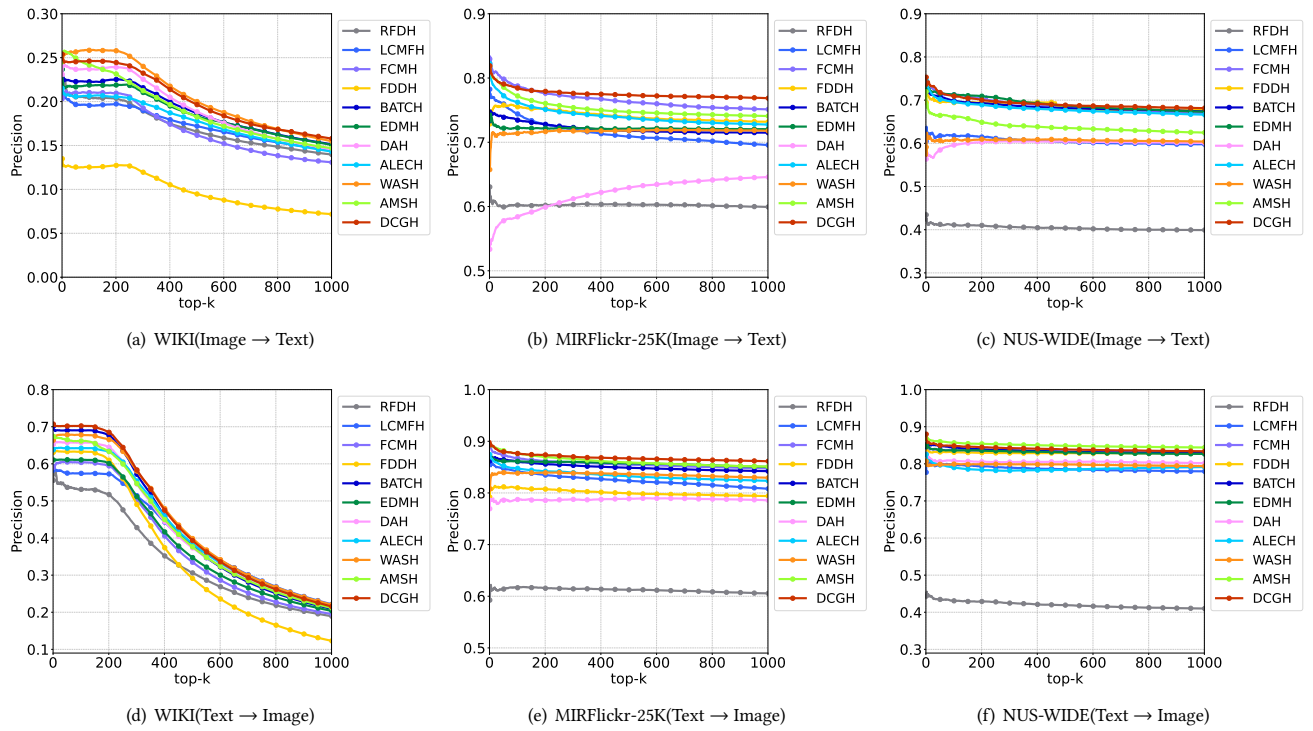


Figure 3: Top-k precision curves with 8 bits on three datasets.

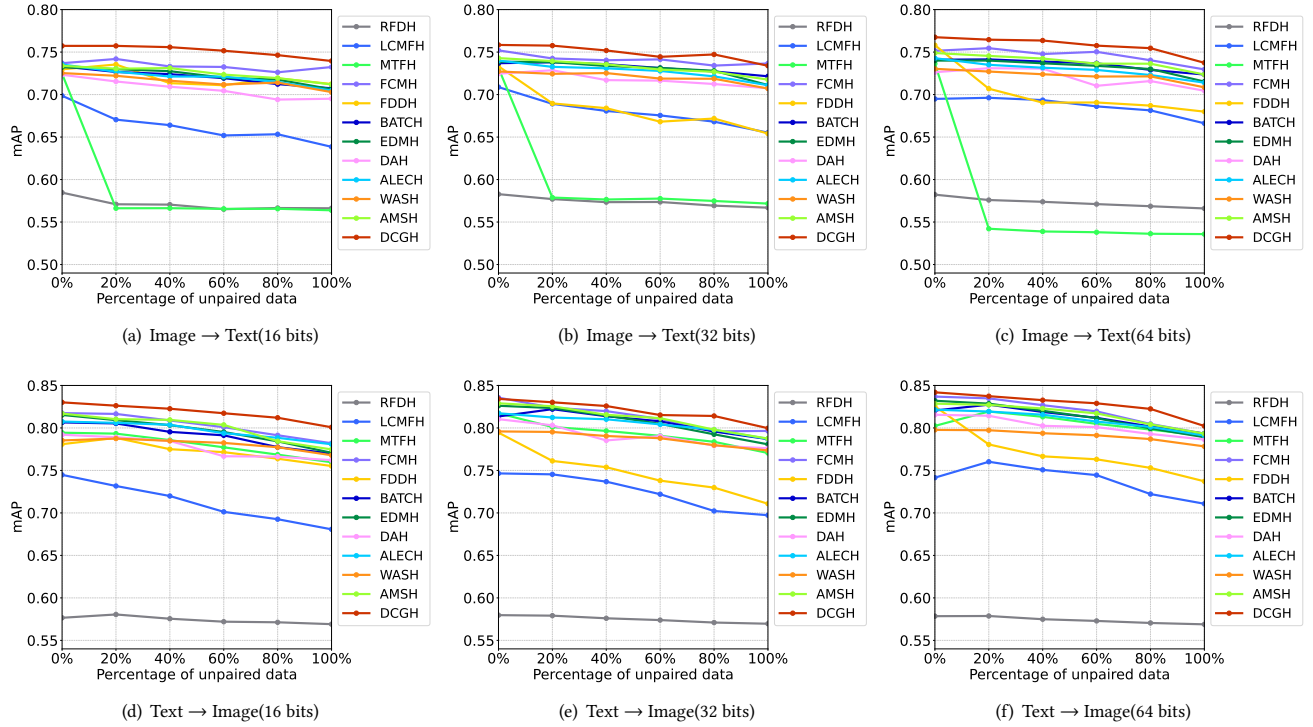


Figure 4: The mAP results with varying proportions of unpaired data with different hash lengths on MIRFlickr-25K dataset.

datasets. Moreover, we draw the corresponding mAP results on MIRFlickr-25K with different bits in Fig.4. These experimental analysis is provided in our manuscript.

3.4 Parameter Sensitivity Analysis

On the other three datasets, we further perform parameter sensitivity analysis. As shown in Fig.5, DCGH consistently achieves superior results across most ranges, which indicates that DCGH is stable to hyperparameters to some extent.

3.5 Convergence Analysis

As shown in Fig.6, we add more convergence analysis on the other datasets. The results indicates the objective value of our algorithm consistently decreases with each iteration, which provides clear evidence of the convergence of our proposed algorithm.

3.6 Comparison on Time Cost

We compare the training time of different methods across four datasets. The outcomes of our experiments are presented in Table 1. It is evident from the table that BATCH achieves the fastest training speed due to its asymmetric strategy and efficient matrix decomposition. Conversely, MTFH is the slowest because its computational complexity in the optimization process reaches $O(n^2)$. It is noticeable that our DCGH achieves highly competitive training efficiency compared with most baselines. Specifically, the training speed of DCGH is slightly slower than that of ALECH, BATCH, and DAH because the process of searching the cross-modal class distribution

center is time-consuming. In general, DCGH outperforms most baselines in terms of training time.

Table 1: Training time (seconds) of different methods with 64-bit hash codes on four datasets.

Method	WIKI	MIRFlickr-25K	IAPR-TC12	NUS-WIDE
RFDH [7]	6.65	76.38	118.16	624.46
LCMFH [6]	0.14	3.42	16.32	27.89
MTFH [3]	79.54	190.24	326.51	/
FCMH [8]	0.79	47.68	230.12	279.28
FDDH [4]	1.56	27.81	28.59	307.64
BATCH [9]	0.12	0.58	0.83	0.41
EDMH [1]	0.65	9.63	22.86	69.11
DAH [11]	0.14	0.59	0.88	5.56
ALECH [2]	0.50	1.15	1.92	9.15
WASH [10]	0.91	2.92	6.65	23.39
AMSH [5]	1.84	6.93	8.13	59.84
Our DCGH	0.24	1.55	3.85	12.68

REFERENCES

- [1] Yong Chen, Hui Zhang, Zhibao Tian, Jun Wang, Dell Zhang, and Xuelong Li. 2022. Enhanced Discrete Multi-Modal Hashing: More Constraints Yet Less Time to Learn. *IEEE Transactions on Knowledge and Data Engineering* 34, 3 (2022), 1177–1190. <https://doi.org/10.1109/TKDE.2020.2995195>
- [2] Huaxiong Li, Chao Zhang, Xiuyi Jia, Yang Gao, and Chunlin Chen. 2023. Adaptive Label Correlation Based Asymmetric Discrete Hashing for Cross-Modal Retrieval. *IEEE Transactions on Knowledge and Data Engineering* 35, 2 (2023), 1185–1199.
- [3] Xin Liu, Zhikai Hu, Haibin Ling, and Yiu-ming Cheung. 2019. MTFH: A matrix tri-factorization hashing framework for efficient cross-modal retrieval. *IEEE transactions on pattern analysis and machine intelligence* 43, 3 (2019), 964–981.

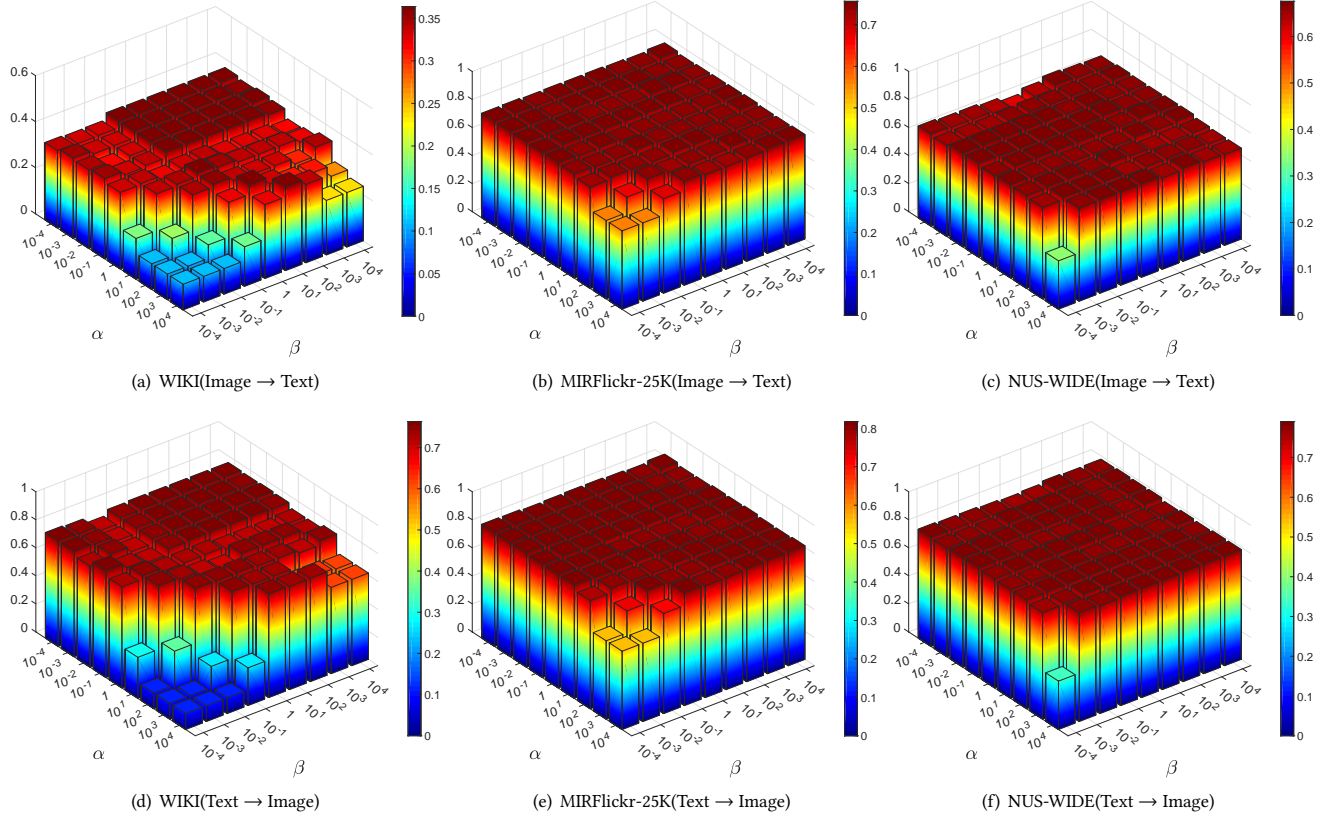


Figure 5: The mAP scores with 8 bits in terms of parameters α and β on three datasets.

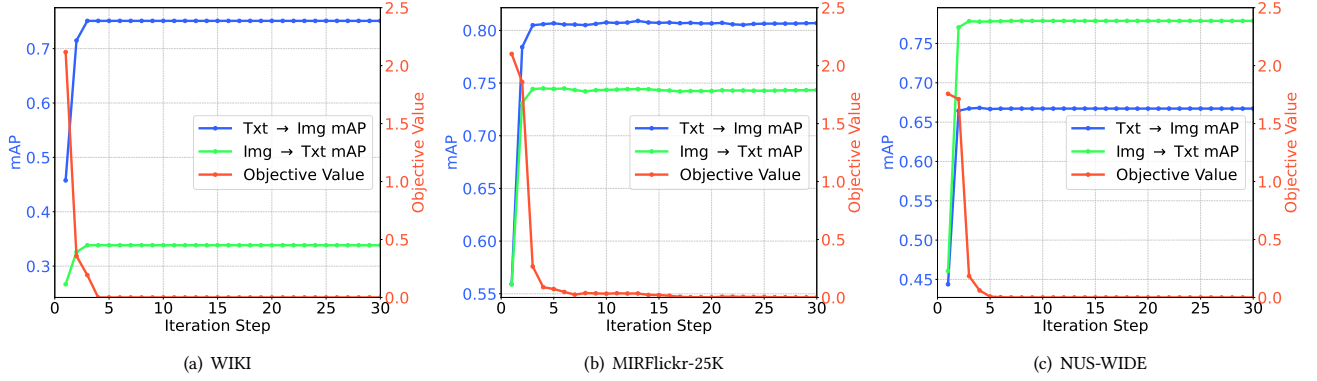


Figure 6: The convergence results with 8 bits on three datasets.

- [4] Xin Liu, Xingzhi Wang, and Yiu-Ming Cheung. 2022. FDDH: Fast Discriminative Discrete Hashing for Large-Scale Cross-Modal Retrieval. *IEEE Transactions on Neural Networks and Learning Systems* 33, 11 (2022), 6306–6320.
- [5] Kaiyi Luo, Chao Zhang, Huaxiong Li, Xiuyi Jia, and Chunlin Chen. 2023. Adaptive marginalized semantic hashing for unpaired cross-modal retrieval. *IEEE Transactions on Multimedia* (2023).
- [6] Di Wang, Xinbo Gao, Xiumei Wang, and Lihuo He. 2018. Label consistent matrix factorization hashing for large-scale cross-modal similarity search. *IEEE transactions on pattern analysis and machine intelligence* 41, 10 (2018), 2466–2479.
- [7] Di Wang, Quan Wang, and Xinbo Gao. 2017. Robust and flexible discrete hashing for cross-modal similarity search. *IEEE transactions on circuits and systems for video technology* 28, 10 (2017), 2703–2715.
- [8] Yongxin Wang, Zhen-Duo Chen, Xin Luo, Rui Li, and Xin-Shun Xu. 2022. Fast Cross-Modal Hashing With Global and Local Similarity Embedding. *IEEE Transactions on Cybernetics* 52, 10 (2022), 10064–10077.
- [9] Yongxin Wang, Xin Luo, Liqiang Nie, Jingkuan Song, Wei Zhang, and Xin-Shun Xu. 2021. BATCH: A Scalable Asymmetric Discrete Cross-Modal Hashing. *IEEE Transactions on Knowledge and Data Engineering* 33, 11 (2021), 3507–3519.
- [10] Chao Zhang, Huaxiong Li, Yang Gao, and Chunlin Chen. 2023. Weakly-Supervised Enhanced Semantic-Aware Hashing for Cross-Modal Retrieval. *IEEE Transactions on Knowledge and Data Engineering* 35, 6 (2023), 6475–6488.
- [11] Donglin Zhang, Xiao-Jun Wu, Tianyang Xu, and He-Feng Yin. 2023. DAH: Discrete Asymmetric Hashing for Efficient Cross-Media Retrieval. *IEEE Transactions on Knowledge and Data Engineering* 35, 2 (2023), 1365–1378.