NAMIBIA UNIVERSITY OF SCIENCE AND TECHNOLOGY

Faculty of Informatics and Computing

Trends in Artificial Intelligence and Machine Learning (TAI911S)

ASSIGNMENT 4

Deadlines	**Released on: 24/05/2025** **Due Date: 06/06/2025**
Total points	100
Student Name and Number	Ms. Theodensia Nakale: **224102907**

RESEARCH PAPER REVIEW

Lightweight Contrastive Distilled Hashing for Online Cross-Modal Retrieval

Authors: Jiaxing Li, Lin Jiang, Zeqi Ma, Kaihang Jiang, Xiaozhao Fang, Jie Wen

Conference: Accepted by AAAI 2025

DOI: https://doi.org/10.48550/arXiv.2502.19751

1. SUMMARY OF THE PAPER

Problem being addressed

The paper addresses challenges in online cross-modal retrieval, concentrating on how to extract coexistent semantic relevance, achieve high performance on real-time data, and transfer offline knowledge to lightweight online models. The focus is on improving the effectiveness and efficiency of retrieval systems that operate across different data modalities (such as images and text) by leveraging both semantic understanding and knowledge transfer techniques.

The main contributions and challenges addressed include:

Extracting coexistent semantic relevance- learning to capture shared semantic meaning across different modalities despite their characteristic heterogeneity.

Real-time performance- developing systems that maintain high retrieval accuracy while managing endless and fast-arriving data streams.

Knowledge transfer- efficiently transferring expertise from a powerful offline 'teacher' model to a lightweight online 'student' model suitable for real-time applications.

Effective feature representation for hashing- ensuring the feature descriptions learned preserve semantic comparison for efficient cross-modal hashing-based retrieval.

Bridging offline and online hashing via distillation- launching a strong tie between offline and online models using a knowledge distillation framework to guide the student model's learning process.

Main contribution of the work

The paper introduces a novel framework called Lightweight Contrastive Distilled Hashing (LCDH), aimed at improving online cross-modal retrieval. The technique tackles the challenge of transferring semantic knowledge from powerful offline models to efficient, lightweight models suitable for real-time use. By combining contrastive learning and knowledge distillation, LCDH ensures semantic alignment among modalities while keeping fast, resource-efficient performance, with the following main contributions and innovations of LCDH:

Bridging offline and online training- LCDH aligns offline and online cross-modal hashing through resemblance matrix estimate within a knowledge distillation framework, allowing the student model to benefit from the richer knowledge of the teacher model.

Semantic relevance extraction and transfer- it extracts coexistent semantic relevance from cross-modal data and extracts this knowledge from the teacher network to the lightweight student network, increasing recovery accuracy in real-time applications.

Use of clip and attention mechanisms- LCDH integrates Contrastive Language-Image Pre-training (CLIP) and attention modules to reserve semantic relationships across different modalities, vital for cross-modal understanding.

Discriminative hash code generation- the framework fuses and improves features to create discriminative hash codes, using both offline and online similarity matrices to supervise and guide the student model's learning process.

Experimental/theoretical results
Experiments on three benchmark datasets (MIRFlickr-25K, IAPR TC-12, and NUS-WIDE) show that LCDH outperforms state-of-the-art methods in mean average precision, demonstrating efficiency and effectiveness in real-time scenarios.

Performance- experimental results show that LCDH consistently produces the best and most stable performance compared to other state-of-the-art baseline methods.

Evaluation Metrics- the performance was evaluated using mean average precision (MAP), precision-recall curves, and top-N precision curves for both text-to-image (T→I) and image-to-text (I→T) retrieval tasks.

Comparison- LCDH notably outperforms knowledge distillation-based techniques like UKD and SKDCH, confirming the efficiency of its knowledge distillation framework and feature extraction modules. Table 1 in the paper presents detailed AP results, showing LCDH's superior performance across different bit lengths and datasets compared to baseline methods.

2. RELATED WORK

The work is built upon prior research in cross-modal hashing, contrastive learning (especially CLIP), and knowledge distillation. It differentiates itself through its integration of these components in a single lightweight retrieval framework.

3. LIMITATIONS OF THE PAPER

The approach depends on CLIP, which may not generalize well in all domains. Moreover, the dataset diversity is limited, and scalability in streaming settings remains underexplored.

Offline vs online performance- the paper states that offline cross-modal hashing baseline approaches outperform online since they learn hash codes and functions in two steps with a full database, while online methods update hash functions with chunks of partial databases. This means that even with LCDH's progress, the important nature of online training with partial data might still present an inherent performance gap compared to methods trained on whole datasets. While LCDH narrows this gap and achieves state-of-the-art results for online hashing, it runs within this constraint.

Discussion of relevant literature

- Chen, Y. et al. (2024). *Denoising High-Order Graph Clustering*. ICDE.
- Chua, T.-S. et al. (2009). *NUS-WIDE: a real-world web image database*. CIVR.
- Cui, C. et al. (2023a). *Effective Multi-View Clustering*. NeurIPS.
- Cui, C. et al. (2023b). *Deep Multi-view Subspace Clustering*. IJCAI.
- Hu, H. et al. (2020). *Unsupervised knowledge distillation for cross-modal hashing*. CVPR.
- Huang, J. et al. (2023). *Two-stage asymmetric similarity preserving hashing*. TKDE.
- Huiskes, M. J., & Lew, M. S. (2008). *The MIR flickr retrieval evaluation*. ACM MIR.
- Jiang, K. et al. (2023a). *Random Online Hashing for Cross-Modal Retrieval*. TNNLS.
- Yao, T. et al. (2019). *Online latent semantic hashing*. Pattern Recognition.
- Zhang, D., & Li, W. (2014). *Supervised multimodal hashing with semantic correlation*. AAAI.
- Zhu, Y. et al. (2024). *Large Language Models for Information Retrieval*. arXiv:2308.07107.

4. EVALUATION

Strengths

LCDH's integration of contrastive learning and distillation is innovative. Its lightweight architecture enables real-time cross-modal retrieval. The modular design enhances adaptability.

Weaknesses

Performance may degrade in domains where CLIP is less effective. Limited dataset diversity impacts generalizability. The paper also lacks extensive ablation studies for component contributions.

Suggestions for improvement

Future work should evaluate broader domains, explore alternatives to CLIP, and provide detailed runtime and ablation analyses.