Supplementary Materials: Multi-view Hashing Classification

Yuhang Lan
College of Computer Science, Sichuan University
Chengdu, China
yhanglan@163.com

Shilin Xu
College of Computer Science, Sichuan University
Chengdu, China
xushilin990@gmail.com

Chao Su
College of Computer Science, Sichuan University
Chengdu, China
suchao.ml@gmail.com

Run Ye
School of Automation Engineering, University of Electronic Science and Technology of China
Chengdu, China
rye@uestc.edu.cn

Dezhong Peng
College of Computer Science, Sichuan University
Tianfu Jincheng Laboratory
Chengdu, China
pengdz@scu.edu.cn

Yuan Sun*
National Key Laboratory of Fundamental Algorithms and Models for Engineering Numerical Simulation, Sichuan University
Chengdu, China
sunyuan_work@163.com

1 Introduction

This supplementary material provides additional details about MHC to facilitate a deeper understanding of the proposed framework. Section 2 provides a detailed description of the datasets used in our experiments. Section 3 introduces the prompt templates used in the Class-prompt Contrastive Learning (CCL) module. Section 4 describes the training process of MHC. And Section 5 presents an analysis of additional visual results and comparisons.

2 Dataset Details

The multi-view data used in this paper include:

- **MSRC**[2]: The dataset contains 210 images across 7 classes, with each instance described by five feature types: CM, HOG, GIST, CENTRIST, and LBP.
- **Leaves**[1]: The dataset comprises 1600 leaf samples from 100 plant species, with each sample represented using three views: shape descriptors, fine-scale edges, and texture histograms.
- **UCI**[2]: The dataset includes 2000 handwritten digit samples categorized into 10 classes. Each sample is described using three distinct views: average pixel values over 240 windows, 47 Zernike moments, and 6 morphological features.
- **HW**[3]: The dataset consists of 2000 instances of handwritten numerals from 0 to 9, with 200 patterns per class, represented using six different feature sets.
- **LandUse**[3]: The dataset contains 2100 satellite images categorized into 21 classes, with each image represented using three views extracted by GIST, PHOG, and LBP descriptors.
- **Scene**[4]: The dataset comprises 4485 images from 15 indoor and outdoor scene categories. Each image is represented using features extracted with GIST, PHOG, and LBP.
- **ALOI-100**[5]: The dataset is a four-view dataset consisting of HSB, RGB, COLORSIM, and HARALICK features. It includes 1000 small objects and a total of 10,800 images.
- **Animal**[6]: The dataset includes 11,673 images across four views and 20 categories.
- **YoutubeFace**[7]: The dataset contains 101,499 instances belonging to 31 classes. Each instance is represented using five views: audio volume stream, vision cuboids histogram, vision motion estimate histogram, vision HOG features, and miscellaneous vision features.
- **ALOI-1000**[1]: The dataset consists of 110,250 samples distributed across 1000 classes, with each sample represented using four different views.

Table 1: The detailed statistics of all multi-view datasets.

Dataset	Size	Categories	View	Dimensionality
MSRC	210	7	6	1,302;48;512;100;256;210
Leaves	1,600	100	3	64;64;64
UCI	2,000	10	3	6;240;47
HW	2,000	10	6	216;76;64;6;240;47
LandUse	2,100	21	3	20;59;40
Scene	4,485	15	3	20;59;40
ALOI-100	10,800	100	4	77;13;64;125
Animal	11,673	20	4	2,689;2,000;2,001;2,000
YoutubeFace	101,499	31	5	64;512;64;647;838
ALOI-1000	110,250	1,000	4	125;77;13;64

3 Prompt template

In this paper, the ten predefined prompt templates used in the Class-prompt Contrastive Learning module are as follows:

- A clear and detailed image of a <class>.
- A well-composed photograph of a <class>.
- A high-quality picture of a <class>.
- A visually appealing shot of a <class>.
- A crisp and vibrant photo of a <class>.
- A beautifully captured image of a <class>.
- A stunning and sharp picture of a <class>.

*Corresponding author

[1] https://archive.ics.uci.edu/dataset/241/one+hundred+plant+ species+leaves+data+set
[2] http://archive.ics.uci.edu/dataset/72/multiple+features
[3] https://archive.ics.uci.edu/ml/datasets/Multiple+Features
[4] https://doi.org/10.6084/m9.figshare.7007177.v1
[5] http://elki.dbs.if.lmu.de/wiki/DataSets/MultiView.
[6] https://cvml.ista.ac.at/AwA/
[7] https://www.cs.tau.ac.il/ wolf/ytfaces/

- A professional-grade photo of a <class>.
- A striking and memorable image of a <class>.
- A captivating and clear photograph of a <class>.

4 Training Procedure

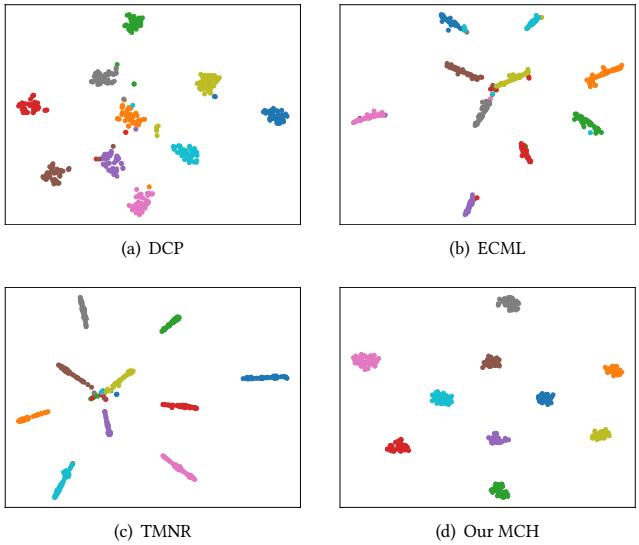
To clearly illustrate the training process of our MHC framework, we give the pseudocode in Algorithm 1.

Algorithm 1 The pseudo code of the proposed MHC

Input: Multi-view dataset: $O = \{\{x_i^v\}_{v=1}^V, y_i\}_{i=1}^N$;
Expand label y using predefined prompt templates and obtain its embedding representation $\mathcal{F}^T$;
for *epoch* = 1 to $T_{\max}$ **do**
  for i = 1 to N **do**
    for v=1 to V **do**
      Obtain multi-view hash codes b_i^v, by Eq.1;
    end for
  end for
  for i = 1 to N **do**
    Obtain multi-view fusion hash codes e_i by Eq.11;
  end for
  for i = 1 to N **do**
    Obtain anchor hash codes m_i by Eq.4;
    Compute the expectation of anchor hash codes μ_i by Eq.13;
    Compute the bit-level calibration vector δ_i by Eq.14;
  end for
  Compute $\mathcal{L}_{CCL}$ by Eq.6;
  Compute $\mathcal{L}_{SCC}$ by Eq.10;
  Compute $\mathcal{L}_{BIH}$ by Eq.15;
  Calculate $\mathcal{L}_{MHC}$ by Eq.16;
  Optimize MHC parameters through backpropagation;
end for

5 Visualization Analysis

To reveal the intrinsic structure of the data and explore the relationships between different classes, we visualize our proposed MHC and several advanced methods(i.e., DCP, ECML, and TMNR) on the HW dataset using the t-SNE method. As shown in the Fig.1, the DCP method exhibits a relatively disordered inter-class distribution, with blurry class boundaries and significant overlap. Similarly, the ECML and TMNR methods show a concentration of multiple classes in the central region, increasing the risk of overlap. In contrast, our MHC benefits from class-prompt contrastive learning, which enhances intra-class compactness, and a boundary-aware independent hashing that promotes inter-class separability. As a result, clear boundaries between different categories are observed, highlighting the superior feature representation and discriminative capability of our MHC.



(a) DCP (b) ECML (c) TMNR (d) Our MCH

Figure 1: t-SNE visualization on the HW dataset.

References

[1] Jan-Mark Geusebroek, Gertjan J Burghouts, and Arnold WM Smeulders. 2005. The Amsterdam library of object images. *International Journal of Computer Vision* 61 (2005), 103–112.
[2] John Winn and Nebojsa Jojic. 2005. Locus: Learning object classes with unsupervised segmentation. In *Tenth IEEE International Conference on Computer Vision (ICCV'05) Volume 1*, Vol. 1. IEEE, 756–763.
[3] Yi Yang and Shawn Newsam. 2010. Bag-of-visual-words and spatial extensions for land-use classification. In *Proceedings of the 18th SIGSPATIAL international conference on advances in geographic information systems*. 270–279.