Supplementary Material for Multi-Memory Matching for Unsupervised Visible-Infrared Person Re-Identification

I Overview

In this document, we first introduce the ARI as a metric for evaluating the reliability of cross-modality correspondences and pseudo-labels. Secondly, we present experimental results for four widely-used cluster evaluation metrics, providing convincing and solid evidence for the reliability of our pseudo-labels. Finally, we supplement the selection of the sub-clustering approach.

II ARI as a Metric for Evaluating Unsupervised Cross-Modal Re-Identification

To evaluate the reliability of cross-modality correspondences and pseudo-labels in unsupervised cross-modality re-identification, this work introduces the Adjusted Rand Index (ARI) metric, ARI is a measure of the similarity between two data clusterings, the ARI is calculated using the formula:

$$\text{ARI} = \frac{\sum_{ij}\binom{n_{ij}}{2} - \left[\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2}\right] / \binom{N}{2}}{\frac{1}{2}\left[\sum_i \binom{a_i}{2} + \sum_j \binom{b_j}{2}\right] - \left[\sum_i \binom{a_i}{2} \sum_j \binom{b_j}{2}\right] / \binom{N}{2}} \tag{I}$$

where n_{ij} is the number of samples in the common cluster between pseudo-labels and ground-truth labels, a_i is the number of samples in the i-th cluster of pseudo-labels, b_j is the number of samples in the j-th cluster of ground-truth labels, and N is the total number of samples in the dataset. The binomial coefficient $\binom{n}{k}$, representing the number of ways to choose (k) elements from n distinct elements, is calculated as:

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \tag{II}$$

where $n!$ denotes n factorial, the product of all positive integers up to n. The ARI value ranges from -1 to 1, with 1 indicating perfect agreement, 0 indicating no better agreement than chance, and negative values indicating less agreement than chance.

To demonstrate the computation of the Adjusted Rand Index (ARI) in the context of unsupervised cross-modal person re-identification, given two elements T (ground-truth labels) and P (pseudo-labels), the ground-truth labels are not used during the training process.

Clusterings Defined:

- Clustering T: {'abc', 'de', 'fgh'}
- Clustering P: {'ab', 'cde', 'fgh'}

Table I. Confusion matrix between pseudo-labels and ground-truth labels.

	CLuster 'abc'	CLuster 'de'	CLuster 'fgh'	sum
CLuster 'ab'	2	0	0	2
CLuster 'cde'	1	2	0	3
CLuster 'fgh'	0	0	3	3
sum	3	2	3	8

where a,b,d,f are visible samples and c,e,g,h are infrared samples.

Step 1: Calculating Pairwise Combinations Within Each Clustering

Based on Tab. I, the number of pairwise combinations within each cluster in both clusterings T and P is calculated as follows:

- In Clustering T:
  - Cluster 'abc': $\binom{3}{2} = 3$ pairs
  - Cluster 'de': $\binom{2}{2} = 1$ pair
  - Cluster 'fgh': $\binom{3}{2} = 3$ pairs
- In Clustering P:
  - Cluster 'ab': $\binom{2}{2} = 1$ pair
  - Cluster 'cde': $\binom{3}{2} = 3$ pairs
  - Cluster 'fgh': $\binom{3}{2} = 3$ pairs

Step 2: Identifying Shared Pairs Between Clusterings

The number of shared pairs between Clustering T and Clustering P is identified:

- Shared pairs: 'ab', 'de', 'fg', 'fh', 'gh'
- Total shared pairs: 5 pairs

Step 3: Computing the ARI

The ARI is computed using the following formula:

- $\sum_{ij} \binom{n_{ij}}{2} = 5$
- $\sum_{i} \binom{a_i}{2} = 3 + 1 + 3 = 7$
- $\sum_{j} \binom{b_j}{2} = 1 + 3 + 3 = 7$
- $\binom{N}{2} = \binom{8}{2} = 28$

Substituting these values into the ARI formula:

$$\text{ARI} = \frac{5 - \frac{7\times7}{28}}{\frac{7+7}{2} - \frac{7\times7}{28}} = \frac{5 - 1.75}{7 - 1.75} = \frac{3.25}{5.25} \approx 0.619$$

The ARI value of approximately 0.619 indicates a moderate level of agreement between Clustering T and Clustering P, suggesting some consistency in the clustering results, yet not perfectly aligned. This example demonstrates the utility of ARI in providing an objective assessment of the reliability of cross-modality correspondences.

III Comparison of General Cluster Evaluation Metrics

To provide more compelling evidence of the reliability of our pseudo-labeling, we present experimental results for four widely-used cluster evaluation metrics. Our results consistently outperform state-of-the-art methods in various cluster evaluation metrics, as shown in Table II.

Table II. The different evaluation metrics of clustering on SYSU-MM01 compared with several state-of-the-art methods.

Metrics	ADCA			PGM			GUR			Ours		
	ALL	RGB	IR	ALL	RGB	IR	ALL	RGB	IR	ALL	RGB	IR
Adjusted Mutal Information [4]	0.729	0.822	0.899	0.746	0.837	0.920	0.779	0.874	0.945	0.857	0.858	0.944
Completeness Score [3]	0.777	0.841	0.933	0.789	0.851	0.949	0.836	0.921	0.977	0.904	0.923	0.974
Fowlkes-Mallows Index [1]	0.344	0.550	0.703	0.355	0.549	0.771	0.417	0.615	0.819	0.616	0.613	0.830
Normalized Mutal Information [2]	0.816	0.889	0.944	0.825	0.897	0.955	0.835	0.910	0.968	0.890	0.894	0.967

IV The Selection of Sub-cluster

The key idea of MMM is to learn multi-memory through sub-clusters, rather than focusing on the specific methods of sub-clustering. In our experiments, the simple yet effective K-means algorithm surpassed more complex dynamic clustering methods like DBSCAN, as shown in Table III. Therefore, we selected K-means as the sub-clustering algorithm due to its simplicity, efficiency, and competitive performance, making it an effective choice for sub-clustering within the MMM framework.

Table III. Ablation study of dynamic clustering on SYSU-MM01.

Methods	R-1	R-10	R-20	mAP
DBSCAN	58.4	90.3	95.7	55.8
K-means (Our)	61.5	93.3	98.0	57.9

References

1. Fowlkes, E.B., Mallows, C.L.: A method for comparing two hierarchical clusterings. Journal of the American statistical association **78**(383), 553–569 (1983)
2. McDaid, A.F., Greene, D., Hurley, N.: Normalized mutual information to evaluate overlapping community finding algorithms. arXiv preprint arXiv:1110.2515 (2011)
3. Rosenberg, A., Hirschberg, J.: V-measure: A conditional entropy-based external cluster evaluation measure. In: EMNLP. pp. 410–420 (2007)
4. Vinh, N.X., Epps, J., Bailey, J.: Information theoretic measures for clusterings comparison: is a correction for chance necessary? In: ICML. pp. 1073–1080 (2009)