

# Supplementary Materials: Enhancing Unsupervised Visible-Infrared Person Re-Identification with Bidirectional-Consistency Gradual Matching

Anonymous Authors

In this supplementary material, we provide more details of our experiments. We also provide the details about the training process of our method for convenience.

---

## Algorithm 1: Training process of the first stage

---

**Require:** Unlabeled training data  $X_v$  and  $X_i$  collected from visible and infrared modalities;

**Require:** Initialize the visible encoder  $f_\theta^v$  and infrared encoder  $f_\theta^i$  with ImageNet-pretrained parameters;

**for**  $n$  in  $[1, num\_epochs]$  **do**

    Use encoders  $f_\theta^v$  and  $f_\theta^i$  to extract feature vector sets  $U_v$  and  $U_i$  from  $X_v$  and  $X_i$ , respectively;

    Cluster  $U_v$  and  $U_i$  into cluster sets  $\mathcal{H}_v$  and  $\mathcal{H}_i$  based on DBSCAN;

    Initialize modality-specific prototypes  $\Phi_v$  and  $\Phi_i$  by averaging the corresponding feature vectors in each cluster ;

**for**  $i$  in  $[1, num\_iterations]$  **do**

        Sample  $P \times K/2$  visible images and  $P \times K/2$  augmented visible images from  $X_v$ ;

        Sample  $P \times K$  infrared images from  $X_i$ ;

        Minimize the objective function according to Eq. (2) in the main manuscript ;

        Update prototypes  $\Phi_v$  and  $\Phi_i$  according to Eq. (3) in the main manuscript;

**end**

**end**

---

## A MORE DETAILS ABOUT EXPERIMENTS

In Sec. 4.2 in the main manuscript, we conduct experiments on SYSU-MM01 and RegDB datasets to evaluate the effectiveness of different components in our method. In this section, we describe the implementation details of different combinations in Tab. 2 of the main manuscript. Index 1 (DCL + BGM): we use the bipartite graph matching algorithm to generate cross-modality correspondences. Index 2 (DCL + BCR): we use our designed bidirectional-consistency criteria to generate cross-modality correspondences. For a pair of clusters from different modalities, if the score calculated by this criteria is positive, then they will be regarded as correspondences. Index 3 (DCL + BGM + GM): the bipartite graph matching algorithm and gradual matching strategy are combined to select cross-modality correspondences. Specifically, for those cross-modality correspondences selected by bipartite graph matching algorithm, we further sort them according to the direct similarity scores and apply our gradual matching strategy to filter some

of them. Index 4 (DCL + BCR + GM): our designed bidirectional-consistency criteria and gradual matching strategy are combined to generate cross-modality correspondences. We first adopt the bidirectional-consistency criteria to calculate the reliability scores among them, then the gradual matching strategy is used to select the correspondences. Index 5 (DCL + BCR + GM + CMCP): based on the combination of Index 4, we further introduce the Cross-modality correlation preserving module to maintain the coherence of correlations across modalities.

---

## Algorithm 2: Training process of the second stage

---

**Require:** Unlabeled training data  $X_v$  and  $X_i$  collected from visible and infrared modalities;

**Require:** Visible encoder  $f_\theta^v$  and infrared encoder  $f_\theta^i$  trained after the first stage;

**Require:** Hyper-parameter  $\lambda_1$  and  $\lambda_2$  for Eq. (15) in the main manuscript;

**for**  $n$  in  $[1, num\_epochs]$  **do**

    Use encoders  $f_\theta^v$  and  $f_\theta^i$  to extract feature vector sets  $U_v$  and  $U_i$  from  $X_v$  and  $X_i$ , respectively;

    Cluster  $U_v$  and  $U_i$  into cluster sets  $\mathcal{H}_v$  and  $\mathcal{H}_i$  based on DBSCAN;

    Initialize modality-specific prototypes  $\Phi_v$  and  $\Phi_i$  by averaging the corresponding feature vectors in each cluster ;

**for**  $i$  in  $[1, num\_iterations]$  **do**

        Sample  $P \times K/2$  visible images and  $P \times K/2$  augmented visible images from  $X_v$ ;

        Sample  $P \times K$  infrared images from  $X_i$ ;

        Minimize the objective function according to Eq. (15) in the main manuscript ;

        Update prototypes  $\Phi_v$  and  $\Phi_i$  according to Eq. (3) in the main manuscript;

**end**

**end**

---

## B DETAILS OF THE TRAINING PROCESS

As shown in the main manuscript, we proposed includes two stages during training. For convenience, we also provide the details of the training process of our method, which is shown in Alg. 1 and Alg. 2 in this supplementary material.

In the first stage, as the initialized model is limited in dealing with the discrepancy between different modalities, we employ Eq. (2) to facilitate learning within each modality, supplemented by

random channel augmentation to alleviate modality gaps. Subsequently, in the second stage, with the alleviation of modality discrepancies from the preceding phase, we integrate our proposed

bidirectional-consistency gradual matching and cross-modality correlation preserving modules to further guide the model towards acquiring modality-shareable feature representations, as outlined in Eq. (15) in the main manuscript.

117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174

175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232