Responses to Reviewers

Manuscript ID: TCSVT-07299-2021

Manuscript Title: Adversarial Decoupling and Modality-invariant Representation Learning for Visible-Infrared Person Re-identification

---

Editor-in-Chief (comments and responses):

We understand that the reason why you select the IEEE TCSVT for your manuscript is that your manuscript has a good match with this journal----Many related papers should have already been published in this journal. Therefore, before your new submission, you have to answer two questions clearly in your revised manuscript and responses:

1. What are the 3-5 papers published in the IEEE Transactions on Circuits and Systems for Video Technology, which are most closely related to your manuscript.

Response: There are 6 papers published in the IEEE Transactions on Circuits and Systems for Video Technology, which have been cited and are most closely related to our manuscript. Please refer to citations [3] [13] [15] [24] [28] [46] in pp. 13-14.

2. What is distinctive/new about your current manuscript related to these previously published papers.

Response: We spare no efforts to reveal our technical contributions and have analyzed the differences between our work and these previously published papers in pp.2 and pp.4, where the corresponding words are marked with red and blue color respectively. To be specific, the most distinctive novelty of the proposed DMiR is that the model naturally integrates Domain-related Representation Disentanglement (DrRD), Modality-invariant Discriminative Representation (MiDR) and Representation Orthogonal Decorrelation (ROD) into an end-to-end framework. The DrRD is introduced to decouple modal information and purify identity-specific information. Subsequently, MiDR and ROD are designed to enhance and further purify identity-discriminative features, respectively. Totally speaking, our DMiR not only reduces the modality gaps between visible and infrared images, but also elaborates on how to enhance identity-related features and make the two decomposed features unrelated to each other. Yet, DRAH [24] and ReIDCaps [28]

aim to learn discriminative pedestrian representation mapping, CBN [3] is designed to learn camera-invariant distribution, and SDL [13], OMDRA [15] and DFAL [46] aim at reducing domain-related variations by optimizing disentanglement loss. Different from the above methods [3] [13] [15] [24] [28] [46], the DMiR network considers representation decorrelation to avoid introducing spectrum information to identity representations. Moreover, the DMiR disentangles domain-specific features via a min-max adversarial disentanglement process, and enhances intra-class compactness and reduces domain variations by exploring positive and negative pair variations, semantic-wise differences, and pair-wise semantic variations.

Associate Editor (comments and responses)

The revision has resolved most of the concerns. But it will be good to address the remaining ones. Hence AC decides to rank as a minor revision.

Response: We are very appreciated with your kind words. In the revised paper, we have addressed all the reviewers' concerns. Please refer to our point-to-point responses to the reviewer's comments below for more details.

Reviewer #1 (comments and responses)

1. The author has updated the missing references I mentioned. The overall presentation is good and the contributions are clear. I suggest thus accepted once the authors have considered the advice from the other reviewers.

Response: Thank you for your kind words.

Reviewer #2 (comments and responses)

Some of my remarks have been addressed, but some points in the paper are still not clear enough. In particular, some additional details or explanations in the authors' reply to my remarks have not been included in the paper, which remains vague and sometimes misleading in the corresponding points, especially formulations and definitions.

Response: We are very appreciated with the precious suggestion by the reviewer. In the revised paper, we have provided more details and explanations to make the points of the paper clearer. Moreover, to improve the readability of the paper, we have revised the formulations and definitions

based on the reviewers' suggestions. Please refer to our point-to-point responses below for more details.

1. What is intra-representation compactness? What is inter-spectrum variations? More official definitions (equations are better) are needed.

Response: We appreciate the precious suggestion by the reviewer. In the revised paper, we have provided more official definitions to improve the readability of the paper. Specifically, we have used the official definitions 'intra-class compactness' and 'domain variations' to replace 'intra-representation compactness' and 'inter-spectrum variations' respectively.

2. There are many wrong equation formulations. For instance

- In Eq.4 and Eq.2, y_i is for one sample, while L^D and L^ID the training loss for samples, which is not suitable.

- Similar error also occurs in Eq.7. LWRT is defined for the single input, which is directly used in the Eq.11.

Response: We are very appreciated with the precious suggestion by the reviewer. In the revised paper, we have corrected the corresponding equation formulations. Please refer to Eq.2, Eq.4, Eq.7 and Eq.11.

3. Wrong annotations.

- \sum_i^{RFB,IR} should be \sum_i=0^{1}, 0 is RGB, 1 is IR. i is number, can not be the string.

Response: We follow the suggestion.

4. The definition of D is ambiguous. D denotes for domain in Eq.3 while another D denotes for Eq.7. And I am confused for the D in Eq.12.

Response: We appreciate the precious suggestion by the reviewer. In the revised paper, to improve the readability of the paper, we have replaced 'D' with 'Eud' in Eq.7. In particular, y_i^I and y_i^D denote identity-specific features and domain-specific features in Eq.12, and these two features are defined in Eq.1 and Eq.3, respectively.