Cooperative Matching and Correspondence Structure Learning for Person Re-identification

To Reviewer 1

We thank you for your useful comments and suggestions.
Q1. "The key idea of this paper is to learn correspondence structures from image subsets divided by pose views. However, the pose views are basically unknown in practice, and are often difficult to be estimated. However, the authors provided no description about how the pose views are obtained."
A1. For training data, the pose view of each training person image is established manually. That is, the training person images are divided into different pose groups manually. In testing phase, for each testing image we first parse the human pose information for it by using the method [1] [1], and then classify it into its closest pose group obtained in the training phase.
Q2. "The complexity analysis is not clear. Is the running time provided in Table 5 is the training time or the testing time? Does the time for Co-matching include the time for estimating the pose views"
A2. The running time shown in Table 5 only contains training time of our co-matching approach. It does not include the time for estimating pose views. This is the same for the other comparison methods listed in Table 5.

To Reviewer 2

We thank you for your useful comments.
Q1. "It is hard to understand w_{ij}^r in Equation (2)"
A1. The parameter w_{ij} is used to weight/balance the importance of different graphs. Here, we use w_{ij}^r where $r > 1$ instead of w_{ij} to avoid trivial solution. This trick has been commonly used in some other works, such as work [2][2].
Q2. "Multiple repeated statements in this paper. Like the description of related papers in Section 1 and 2."
A2. Thank you for your suggestions. We will revise the related section to avoid repeated statements.

To Reviewer 3

We thank you for your useful comments and suggestions.
Q1. "The key benefits of the proposed method is not well justified. The authors claimed that the proposed method can leverage multiple images to learn correspondence between different pose pairs. Specifically, this is achieved by setting r in Eq (2) to be larger than 1. However, it does not compare the results between $r > 1$ and $r = 1$, and thus it is not clear to what degree "cooperative" matching contributes."

[1][1] Luo et al., Pedestrian parsing via deep decompositional network. ICCV 2013

[2][2] Wang et al., Optimizing multi-graph learning: towards a unified video annotation scheme. ACM MM 2007

A1. When $r = 1$, our co-matching model is degenerated to standard single image-level matching case, which has been similarly done in related work (GCT) [3] [3]. In our experiments, we have compared our co-matching method with related work GCT [2], as shown in Table 1-4 in detail. Experimental results show that, our co-matching performs obviously better than GCT, which demonstrates the benefit of the proposed cooperative matching.
Q2. "Secondly, as shown in Eq (12), the learned Ps works as a regularization term during testing. So its weight, alpha, should be carefully discussed, because it could reveal the learned Ps contributes in the testing phrase. However, alpha is simply set as 0.5 and not discussed in the experimental section."
A2. We add the experiments on different α in the following. Figure 1 shows the performance of the proposed method across different α (in Eq.(12)) values. As shown in Eq.(12), when $\alpha = 1.0$, the method does not use the learned P and when $\alpha = 0$, the method only uses the learned P. Here, we can note that, 1) when $\alpha = 0.1 \sim 0.9$, the method generally performs better than $\alpha = 1.0$, which clearly demonstrates the positive contribution of the learned P in guiding more reliable patch-wise matching in the testing phrase. 2) When $\alpha = 0$, the method also returns feasible results, which indicates the desired correspondence structure information encoded in learned P.

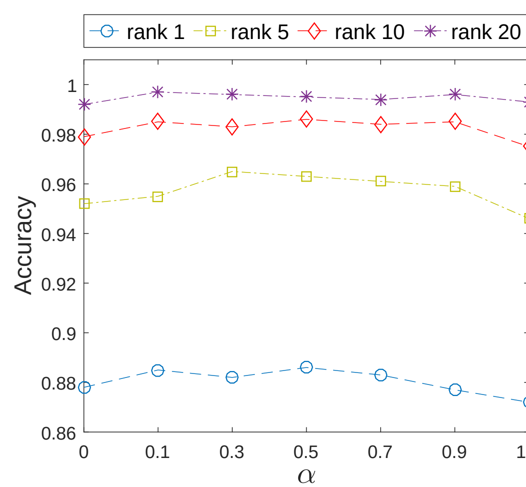

Figure 1. Performance of the proposed co-matching method across different α values on Road dataset.

[3][3] Zhou et al., Graph Correspondence Transfer for Person Re-identification, AAAI 2018