Regularizing Vector Embedding in Bottom-Up Human Pose Estimation

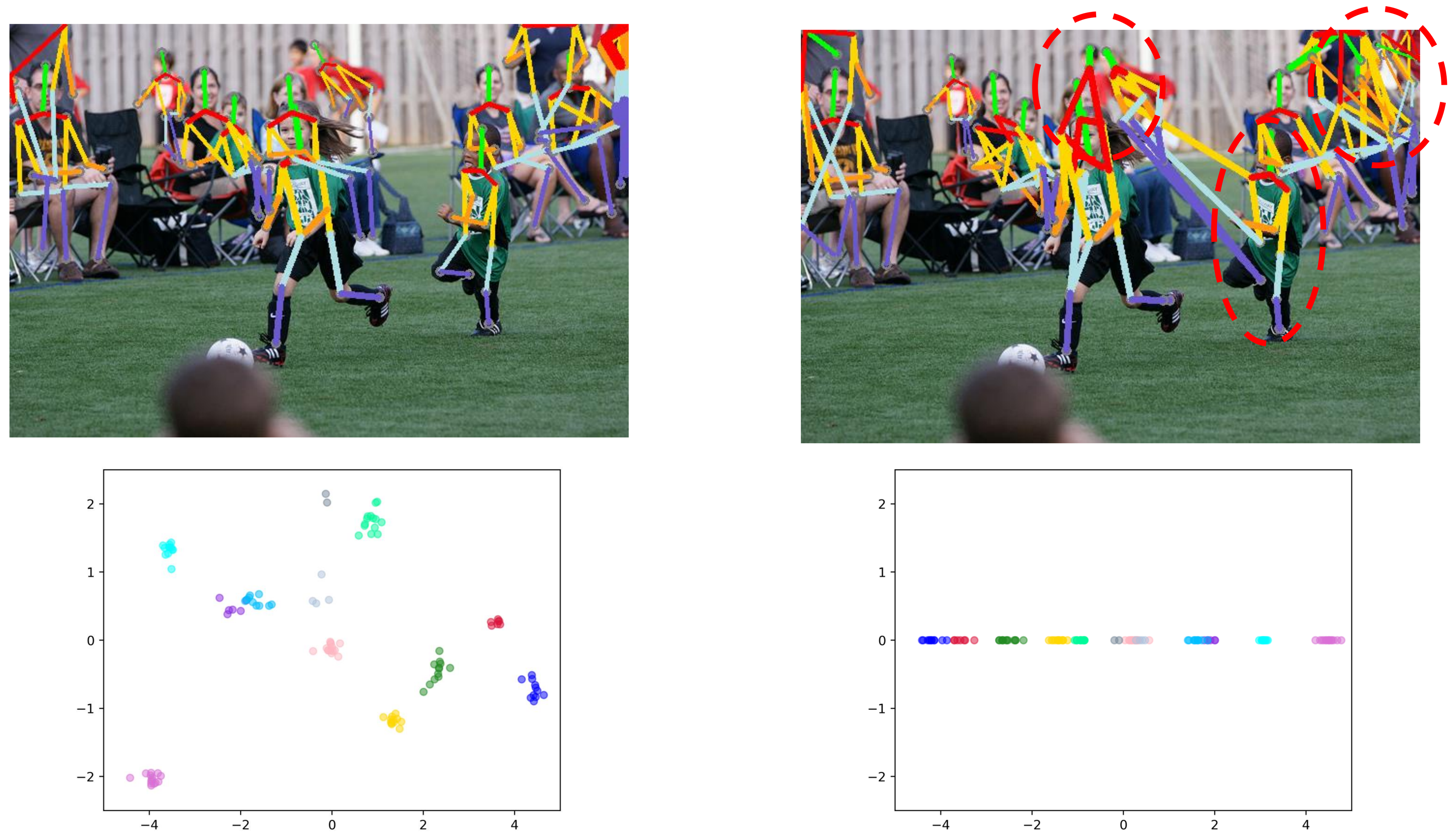
Haixin Wang[1,2], Lu Zhou[2], Yingying Chen[2], Ming Tang[1,2], Jinqiao Wang[1,2]
[1]School of Artificial Intelligence, University of Chinese Academy of Sciences, China
[2]Institute of Automation, Chinese Academy of Sciences, China
{haixin.wang, lu.zhou, yingying.chen, tangm, jqwang}@nlpr.ia.ac.cn

TEL AVIV 2022

Contributions

- We find that different dimensions of embeddings are highly linearly correlated in Associative Embedding, which may lead to incorrect identity assignments in keypoint grouping.
- To address the above-mentioned issue, our method imposes an additional constraint on embeddings to learn sparse multidimensional embeddings.
- Extensive experiments demonstrate that our method has significant advantages in crowded scenes compared with other bottom-up methods.

Ours vs. Previous



The identity embeddings (reducing dimension via PCA) in previous methods (**right**, based on Associative Embedding, incorrect keypoint grouping occurs in red dashed circles) are almost distributed on a line but the embeddings in our methods (**left**) are scattered.

Proposed Method

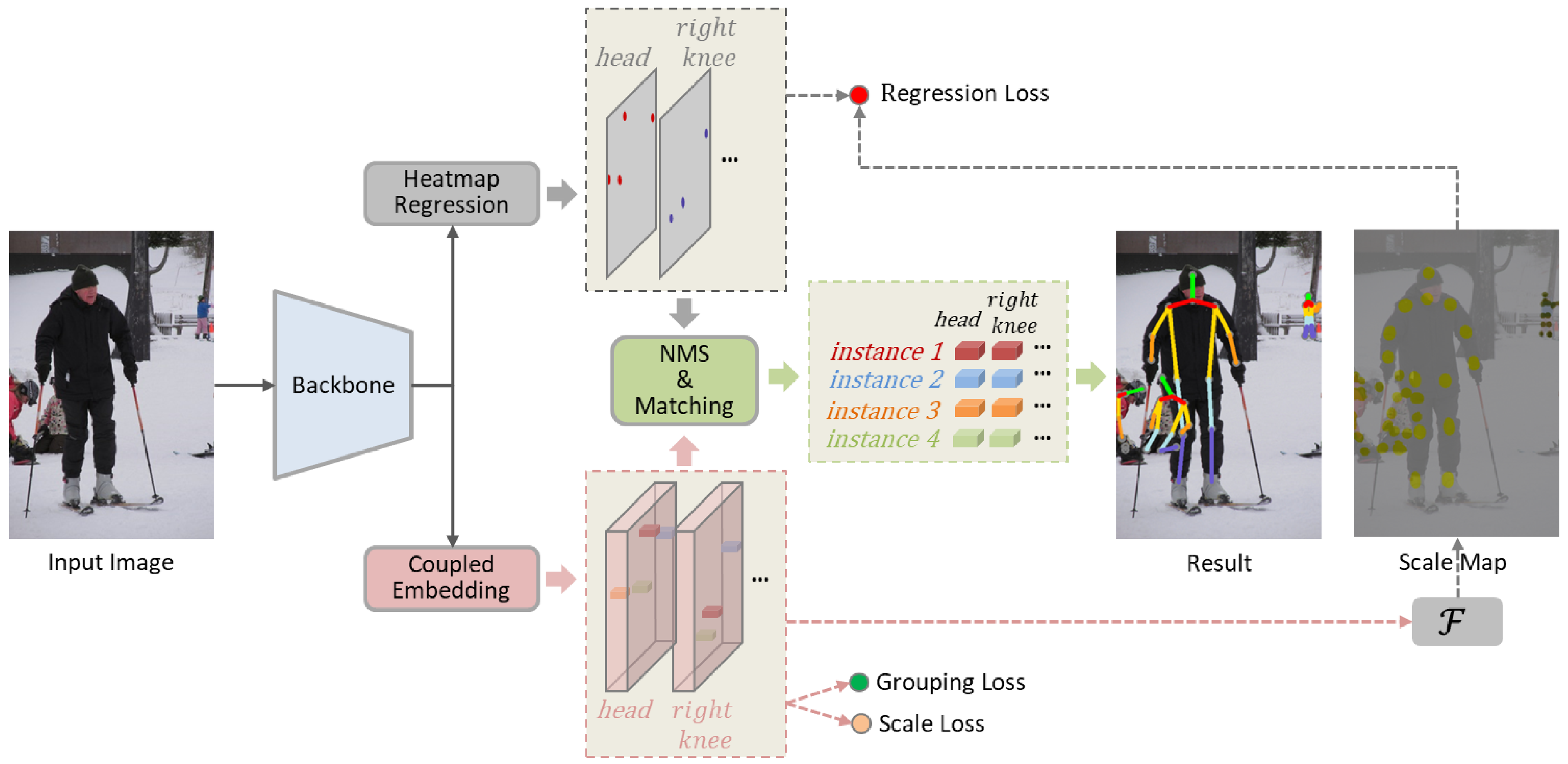
For keypoint embedding, we simultaneously impose scale loss and grouping loss on keypoint embeddings. We denote regularization loss and grouping loss as $\mathcal{L}_s$ and $\mathcal{L}_g$, respectively, where $\boldsymbol{t}$ means the predicted embedding and $\boldsymbol{s}$ means scale ground truth.

$$\mathcal{L}_s = \frac{1}{NK}\sum_n\sum_k\left(1 - \langle \hat{\boldsymbol{t}}_{n,k}, \boldsymbol{s}_{n,k}\rangle\right)$$

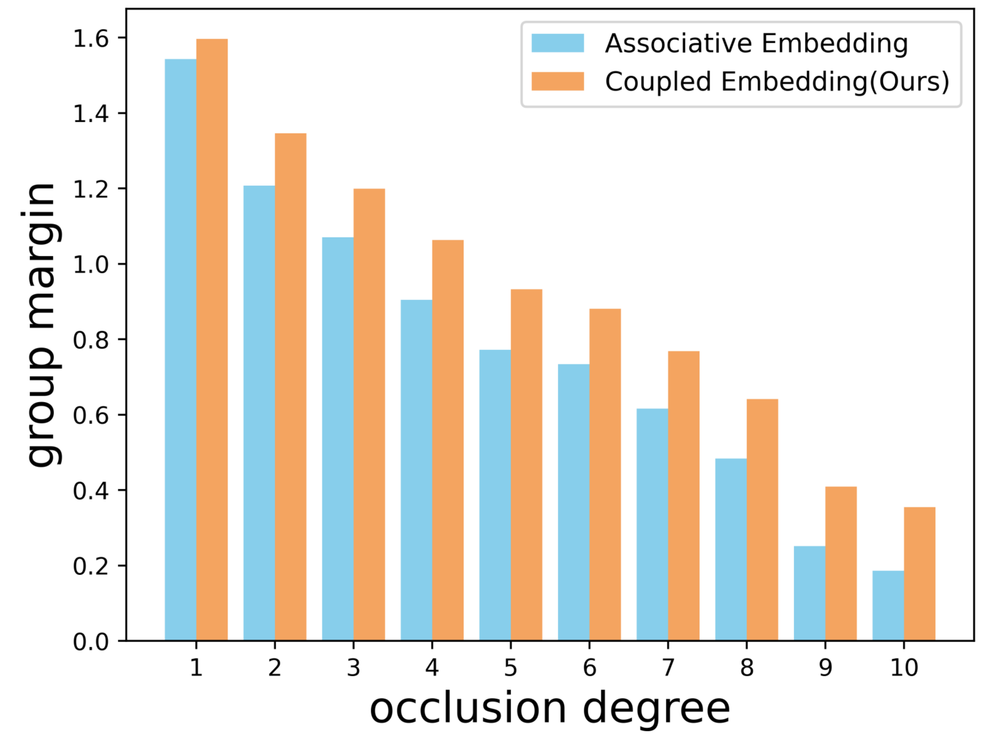
$$\mathcal{L}_g = \frac{1}{NK}\sum_n\sum_k \|\boldsymbol{t}_{n,k} - \bar{\boldsymbol{t}}_n\|_2^2 + \frac{2}{N(N-1)}\sum_n\sum_m e^{-\|\bar{\boldsymbol{t}}_n - \bar{\boldsymbol{t}}_m\|_2^2/2}$$

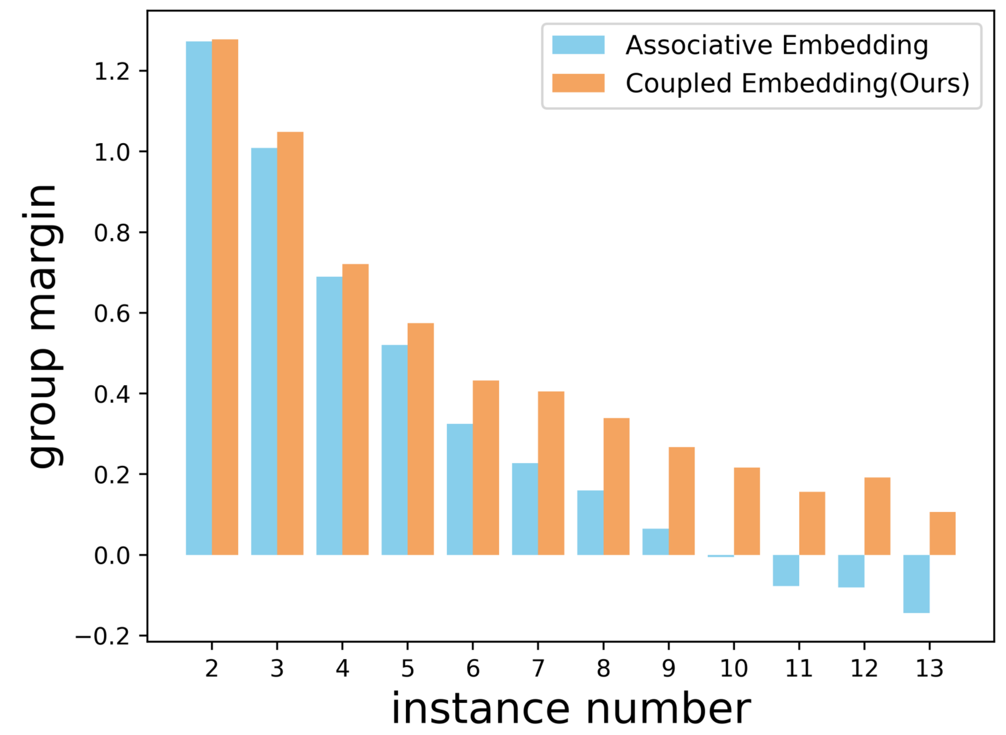
For heatmap regression, we utilize the embedding to generate the scale factor Γ for adaptively adjusting the standard deviation of gaussian kernel in heatmap ground truth. The regression loss can be written as:

$$\mathcal{L}_r = \boldsymbol{W}\cdot\|\boldsymbol{H}_p - \boldsymbol{H}_g^{\sigma\cdot\Gamma}\|_2^2$$



Group Margin





In order to measure the grouping competence of keypoint embedding, we introduce an index named group margin which is defined as the minimum embedding distance minus grouping threshold. Compared with Associative Embedding, our method achieves larger group margin. It indicates our method has more powerful grouping capacity in crowded scenes.