Supplementary Material: Structured Images for RGB-D Action Recognition

Pichao Wang[1]*, Shuang Wang[2]*, Zhimin Gao[1], Yonghong Hou[2]† and Wanqing Li[1]
[1]Advanced Multimedia Research Lab, University of Wollongong, Australia
[2]School of Electronic Information Engineering, Tianjin University, China
pw212@uowmail.edu.au,wangshuang1993@tju.edu.cn,zg126@uowmail.edu.au,houroy@tju.edu.cn,wanqing@uow.edu.au

1. Confusion Matrices

From the following confusion matrices, we can see that the recognition accuracy improves as the granularity increases. The structured motion images aggregate spatio-temporal information contained in the three hierarchical spatial levels by using rank pooling, and keeps the structure information of human motion explicitly using the proposed structured dynamic images.

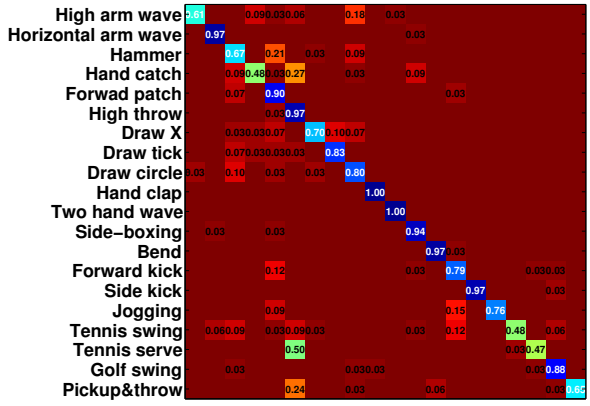

Figure 1: Confusion matrix for structured body DDI on MSRAction3D Dataset.

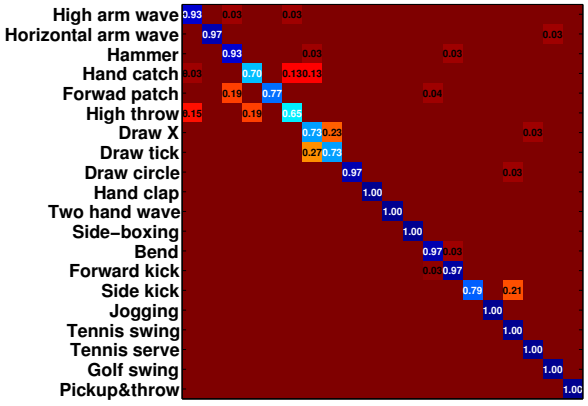

Figure 2: Confusion matrix for structured part DDI on MSRAction3D Dataset.

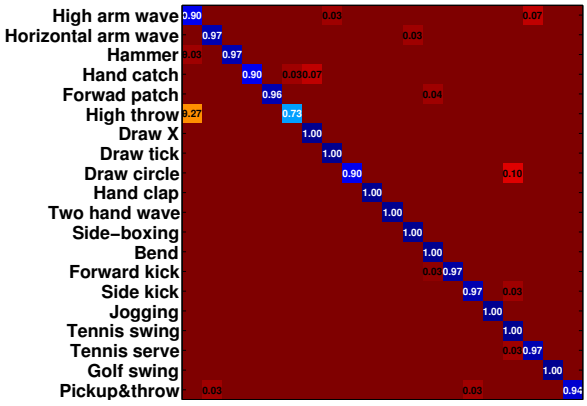

Figure 3: Confusion matrix for structured joint DDI on MSRAction3D Dataset.

*Both authors contributed equally to this work
†Corresponding author

1.1. Confusion Matrix for MSRAction3D Dataset

From the final results we can see that the proposed method can well recognize the simple actions, as it aggregates spatio-temporal information from global to fine-grained by adopting the three hierarchical spatial levels and keeps the structure information of human body by using structured images. Moreover, the multiply-score fusion method works very well on this dataset. For example, action "Forward kick" is confused with "Bend" in part and joint DDIs, but not confused with "Bend" in body DDIs, and after multiply score fusion, "Forward kick" is well recognised from "Bend". It implies that the three structured DDIs are likely to be statistically independent and complementary to each other.

1.2. Confusion Matrix for G3D Dataset

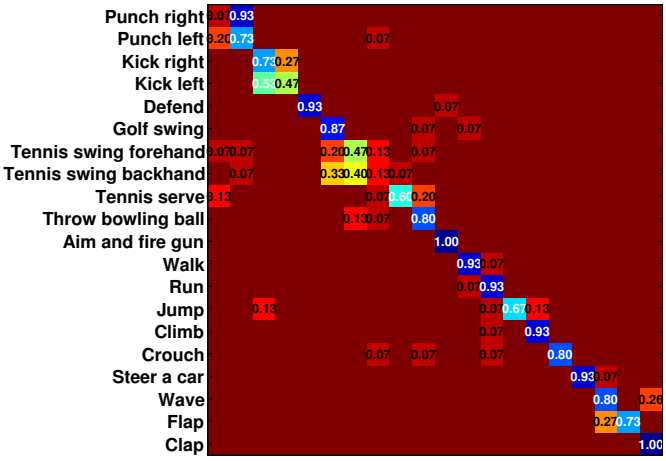

Figure 4: Confusion matrix for structured body DDI on G3D Dataset.

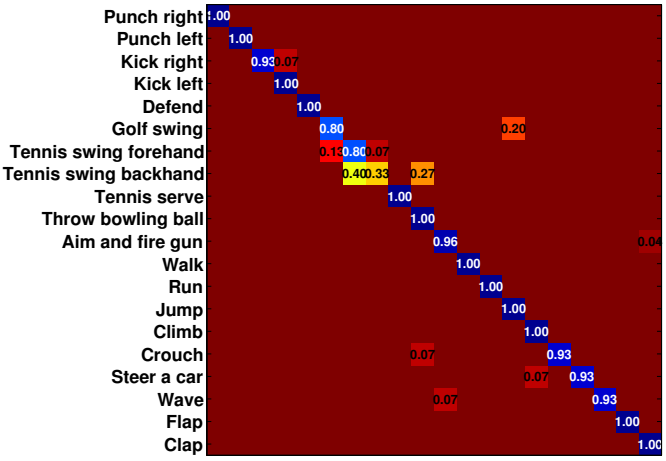

Figure 5: Confusion matrix for structured part DDI on G3D Dataset.

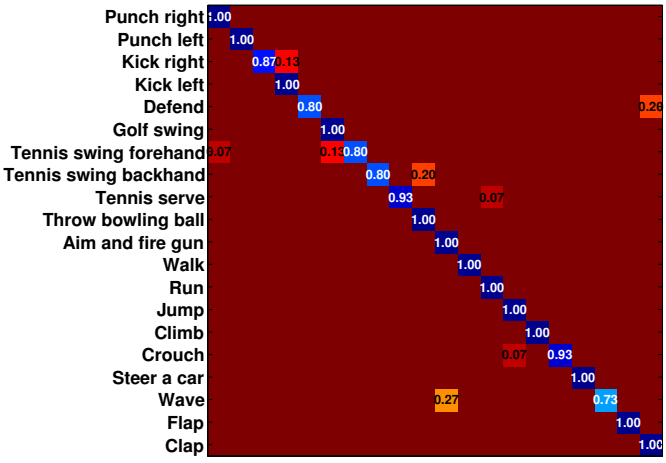

Figure 6: Confusion matrix for structured joint DDI on G3D Dataset.

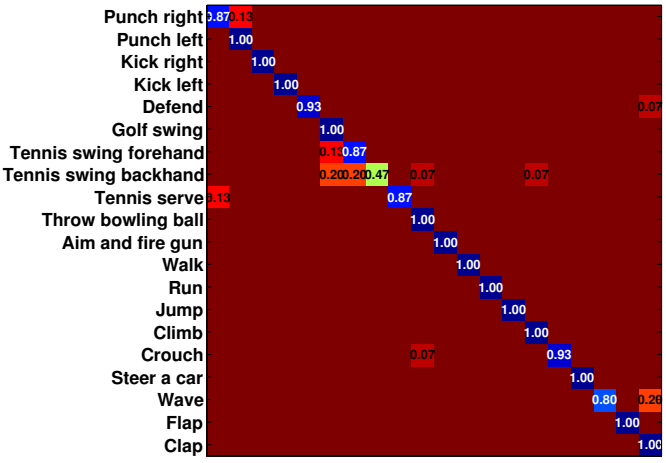

Figure 7: Confusion matrix for S^{2}DDI on G3D Dataset.

From the confusion matrix for S^{2}DDI we can see that the proposed method confuses "Tennis swing backhand" with "Tennis swing forehand" and "Golf swing" after score fusion, even though "Tennis swing backhand" can be recognized much better based on structured joint DDI.

1.3. Confusion Matrix for SYSU 3D HOI Dataset

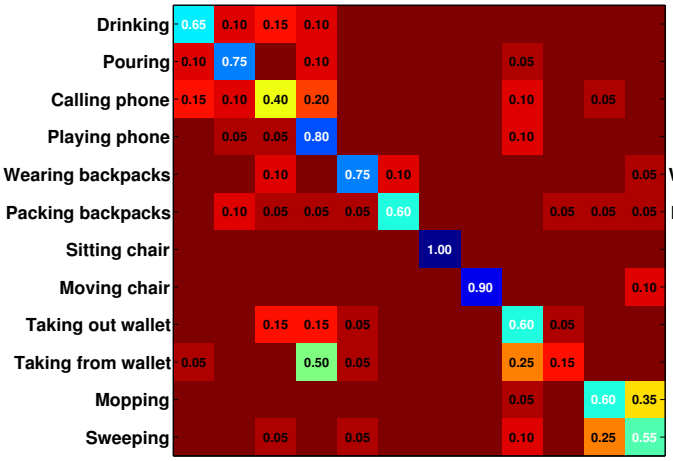

Figure 8: Confusion matrix for structured body DDI on SYSU 3D HOI Dataset.

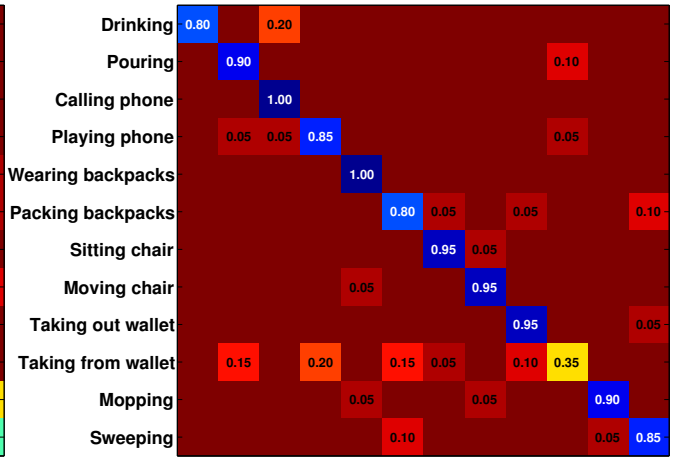

Figure 9: Confusion matrix for structured part DDI on SYSU 3D HOI Dataset.

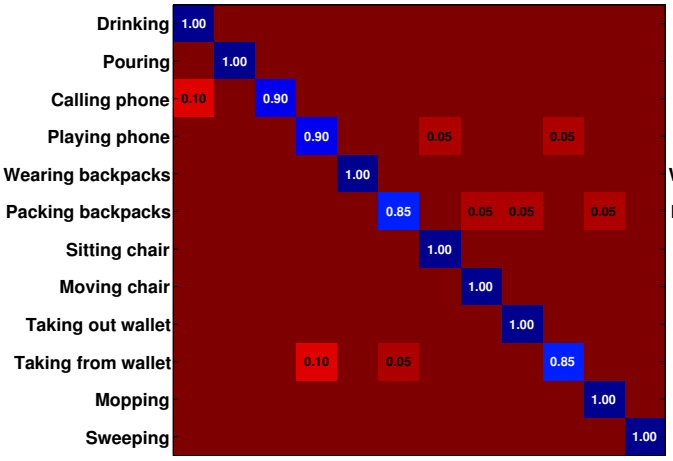

Figure 10: Confusion matrix for structured joint DDI on SYSU 3D HOI Dataset.

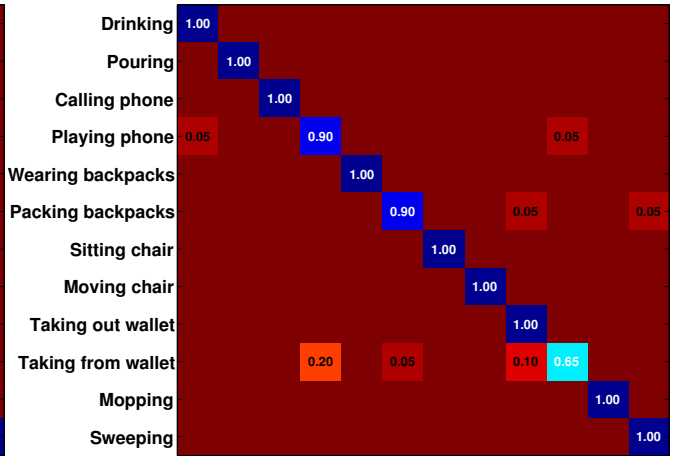

Figure 11: Confusion matrix for S^2DDI on SYSU 3D HOI Dataset.

From the confusion matrices we can see that, the structured joint DDI achieved the best performance. The "Taking from wallet" action is greatly confused in structured body and part DDIs, that affects the final performance of S^2DDI.

1.4. Confusion Matrix for UTD-MHAD Dataset

From the confusion matrices, we can see that as the granularity increases, the proposed method achieves higher accuracy. After multiply score fusion, S^2DDI achieves the best result. It implies that the three structured DDIs are likely to be statistically independent and provide complementary information.

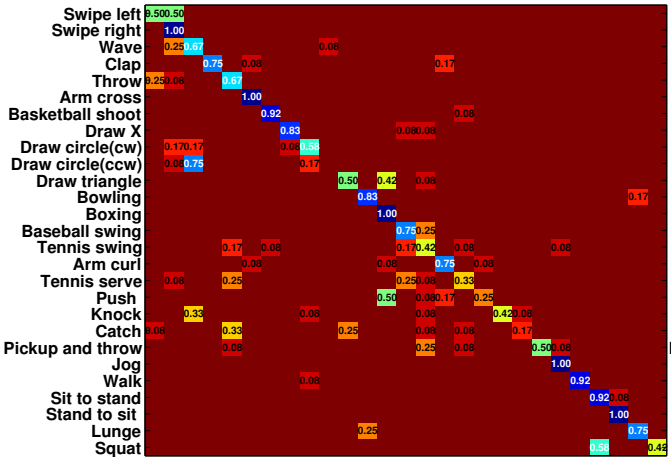

Figure 12: Confusion matrix for structured body DDI on UTD-MHAD Dataset.

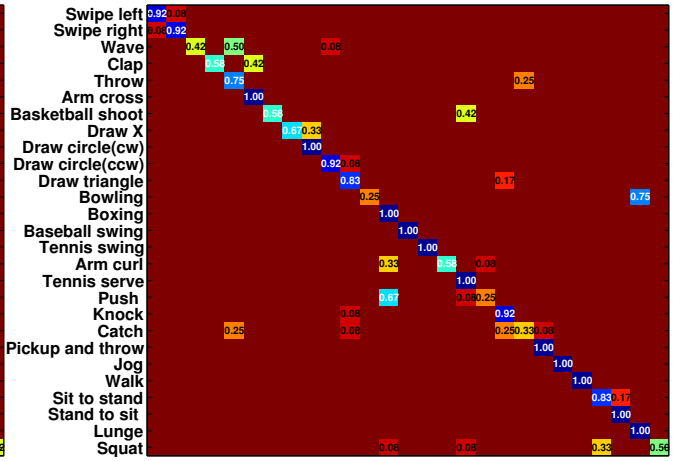

Figure 13: Confusion matrix for structured part DDI on UTD-MHAD Dataset.

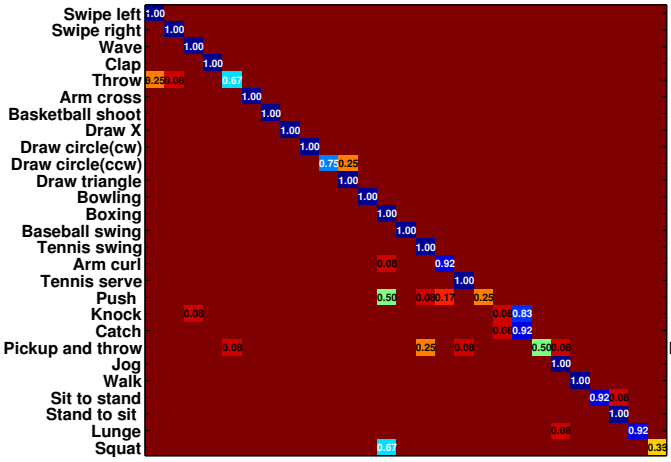

Figure 14: Confusion matrix for structured joint DDI on UTD-MHAD Dataset.

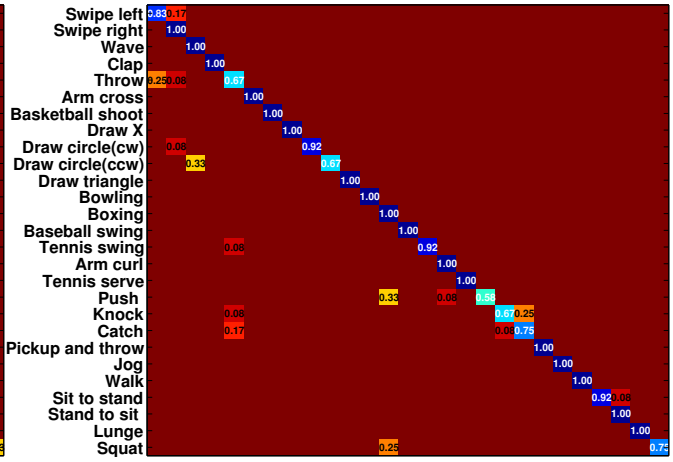

Figure 15: Confusion matrix for S^2DDI on UTD-MHAD Dataset.