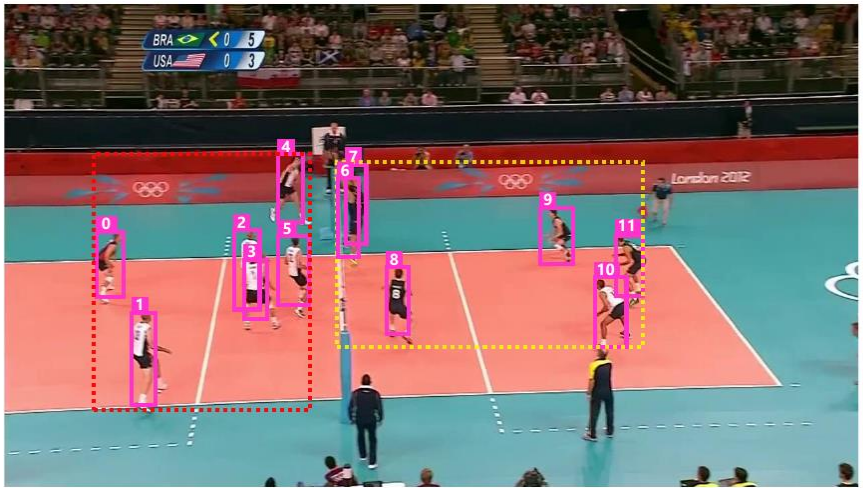
Learning Group Interaction for Sports Video Understanding From a Perspective of Athlete

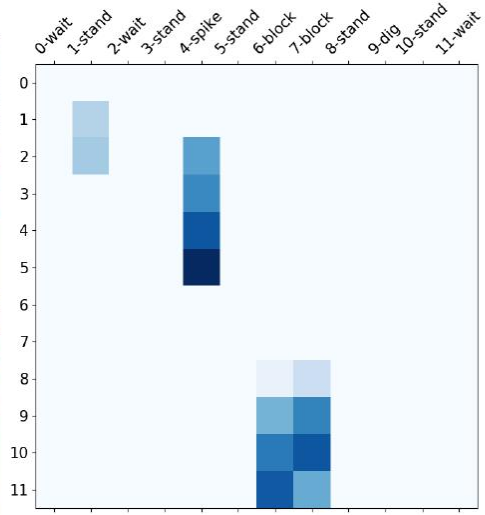
Rui HE, Zehua FU, Qingjie LIU,
Yunhong WANG, Xunxun CHEN

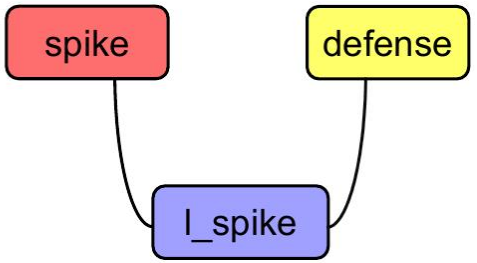
Frontiers of Computer Science, DOI: 10.1007/s11704-023-2525-y

Problems & Ideas

- Problems of sports video understanding approaches:
  - All players in a sports video are usually regarded as a whole by Group Activity Recognition methods.
  - Scene Graph Generation can be used to represent detailed visual relationships in a single image but few in sports videos.
- Ideas: Employing Scene Graph Generation to model group interaction between teams' activities for understanding sports





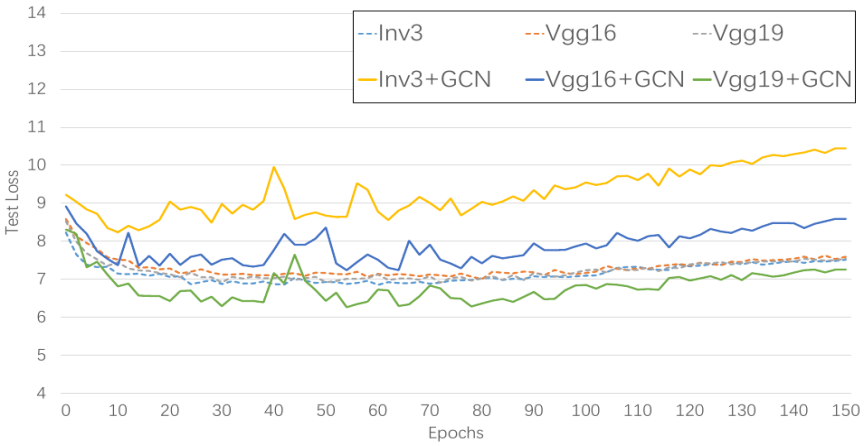


The left figure shows the detected personal actions and the virtual team field. In the middle, recognizing *left_spike*, our method can clearly concentrate on a key actor for both team, such as one spiking man in the left team, and two blocking men in the right team. The color of the spiking player is darker than the two blockers means through Graph Convolutional Network more relations are focusing on the spiker. In the right, The generated Group Scene Graph triplet <*spike – l_spike – defense*> is correct.

Main Contributions

- Contributions:
  - We explore a novel Group Scene Graph Generation method to understand team sports videos from the perspective of an athlete;
  - We construct a novel Hierarchical Relation Network to establish intra-team and inter-team relationship features in a video clip by Graph Convolutional Networks;
  - We build a new *Volleyball+* dataset which is based on the *Volleyball* dataset with additional 9660 team activity labels.

Method	Indi.	Team	Pred.	GSGG
B1-Inv3	53.8	42.9	85.5	1.2
B2-VGG16	53.5	43.0	83.3	1.4
B3-VGG19	58.8	43.1	85.0	1.5
Ours-Inv3+GCN	61.2	43.4	83.5	14.4
Ours-VGG16+GCN	63.3	72.8	82.7	46.5
Ours-VGG19+GCN	63.9	74.5	84.7	49.8



Experimental results of our method. Left: Comparison of baseline and our method on the *Volleyball+* dataset. Indi. means Individual Action Prediction, Team for Team Activity Classification, Pred. for Predicate Recognition and GSGG for Group Scene Graph Generation. The relative gap in performance with respect to the baseline increases significantly: from 1.5 percent to 49.8 percent noticeably with the VGG19 backbone.; Right: the test loss curve. The best GSGG performance of VGG19 with GCN in the left table is corresponding to the lowest curve in the right.