

Supplementary Material: Recognizing Human-Object Interaction via Exemplar based Modelling

Jian-Fang Hu¹, Wei-Shi Zheng², Jianhuang Lai², Shaogang Gong³, and Tao Xiang³

¹School of Mathematics and Computational Science, Sun Yat-sen University, China

²School of Information Science and Technology, Sun Yat-sen University, China

³School of Electronic Engineering and Computer Science, Queen Mary University of London, UK

hujianfang21@126.com, wszheng@ieee.org, stsljh@mail.sysu.edu.cn, {sgg,txiang}@eecs.qmul.ac.uk

Abstract

As mentioned in the paper, experimental results that are omitted in the main paper due to space constrain would be presented in a supplementary document. Rather than providing a detailed discussion on our method, we just present some results on object/person detection here. Readers are recommended to read our main paper for technique details.

1. Experiments

1.1. Detection

We would detect the location of manipulated objects as well as the actor. Same to other related works, we train a set of detectors using DPM (deformable part based model) with two mixture components. Here, we report the detection results in Tab 1. Then the bounding box with maximum response are selected as our final prediction. A prediction is considered correct if it overlaps more than 50% with a ground-truth bounding box. The performance is evaluated in the term of accuracy. Rather than exploring an approach to improve our detection precious for HOI recognition, we address the issue of modeling the mutual structure between human and object in a probabilistic way in a novel probabilistic framework. Experimental results show that our approach is robust to faulty detection.

1.2. HOI recognition

Here, we present the confuse table of our method on PPMI set. As illustrated in Fig. 1, our method obtain an accuracy of (49.34, 47.56), which is comparable to the state-of-the-art results on this set. It can also be observed that our approach achieves better performances on playing-related human-object interactions than with-related interactions. This is because that for actions of playing instruments, it is much easier to mine a set of representative hu-

Data Set	Performance	
	object	Detection(Localization)
Sports	cricket bat	35% (35%)
	bowling ball	10% (40%)
	croquet mallet	50% (65%)
	tennis racket	33% (55%)
	volleyball	75% (95%)
PPMI	bassoon	69%
	cello	76.5%
	clarine	66%
	erhu	69.5%
	flute	50.5%
	frenchhorn	71.5%
	guitar	71.5%
	harp	69%
	recorder	46%
	saxophone	81%
	trumpet	55%
	violin	69.5%

Table 1. Detection results on sports and PPMI data sets

man poses as people often plays music instrument under some certain poses. This largely limits the performance of our method. To further verify this claim. We also present detailed comparison with state-of-the-art (Yao’s Pami2012) in Fig. 2. We can observe that our method outperforms the state-of-the-art with quite large margin in most playing interactions.

Percentage Accuracy: 49.3366%

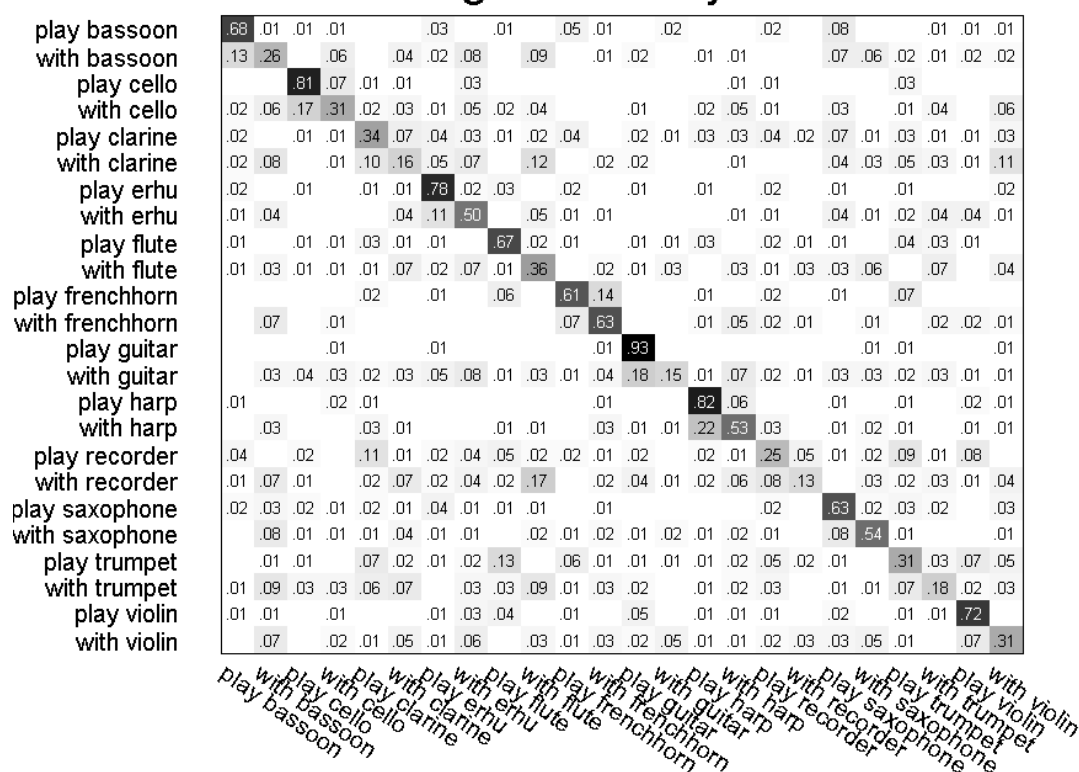


Figure 1. Confusion table of our proposal on PPMI

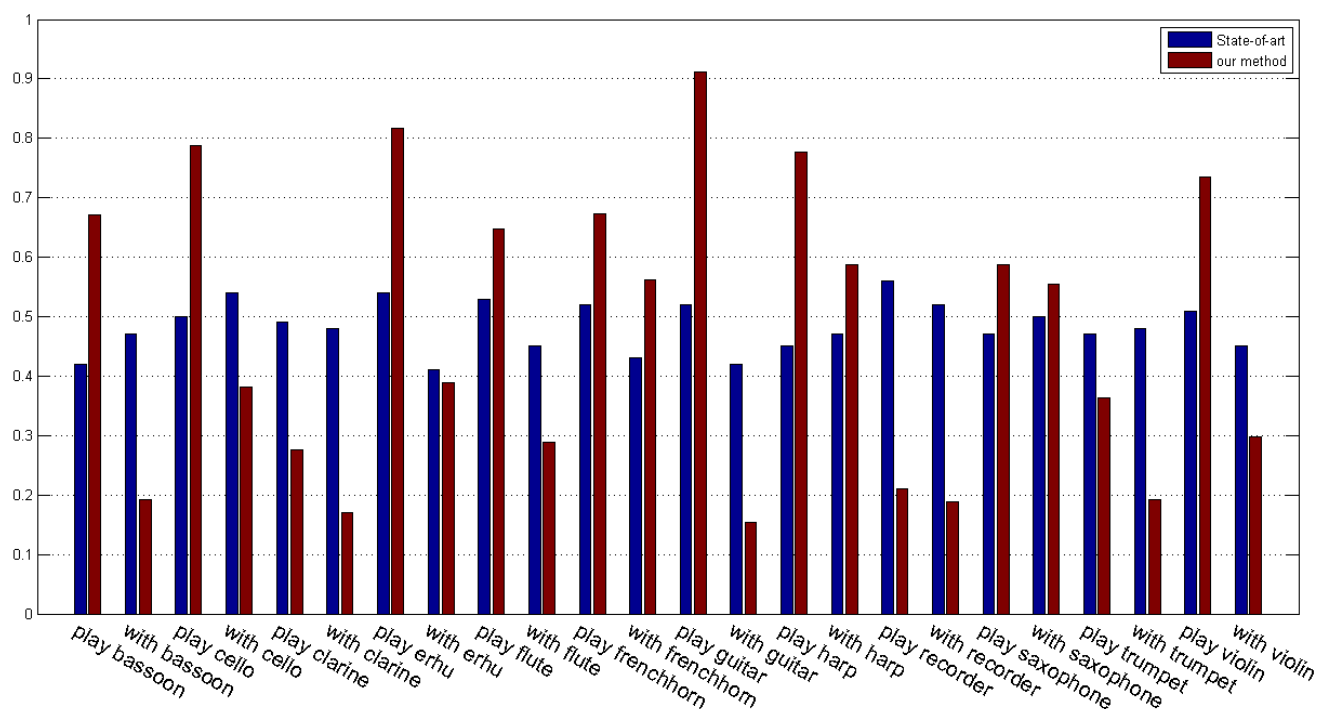


Figure 2. Comparison results on PPMI.