Supplementary Material for

Paper# IoT-10629-2020.R2 (Multi-View Summarization and Activity Recognition Meet Edge Computing in IoT Environments)

This supplementary material elaborates the results of multi-view summarization (MVS) and activity recognition given in Section 4.2 and 4.3 of the experimental results of the main paper. The codes and necessary upload able models and configuration files are given in respective directories of MVS and activity recognition as provided in the zip file.

1. MVS evaluation (Section IV.B in main paper)

Keyframes and generated summary in video format is given in the zip file for Office dataset [1]. The results for this summary are given in Table 1 of main paper. There is some redundancy in the extracted keyframes of view-2 which reduces the recall value for Office dataset to 0.81. The events happened in the input videos of all the views are completely retrieved by our system, thereby producing precision value of 1 in **Table 1** of main paper. A different behavior in contrast to Office dataset can be observed in Bl-7F [2], where our system misses some events, and produces more redundant keyframes compared to the Office dataset. This results in a lower value of 0.85 in recall measure for this dataset. Sample keyframes extracted by our system from Office dataset are given in **Figure 1**. It is evident from the provided keyframes (summary videos) and **Figure 1** that our system is a suitable option for consideration with lower time complexity and higher accuracy.

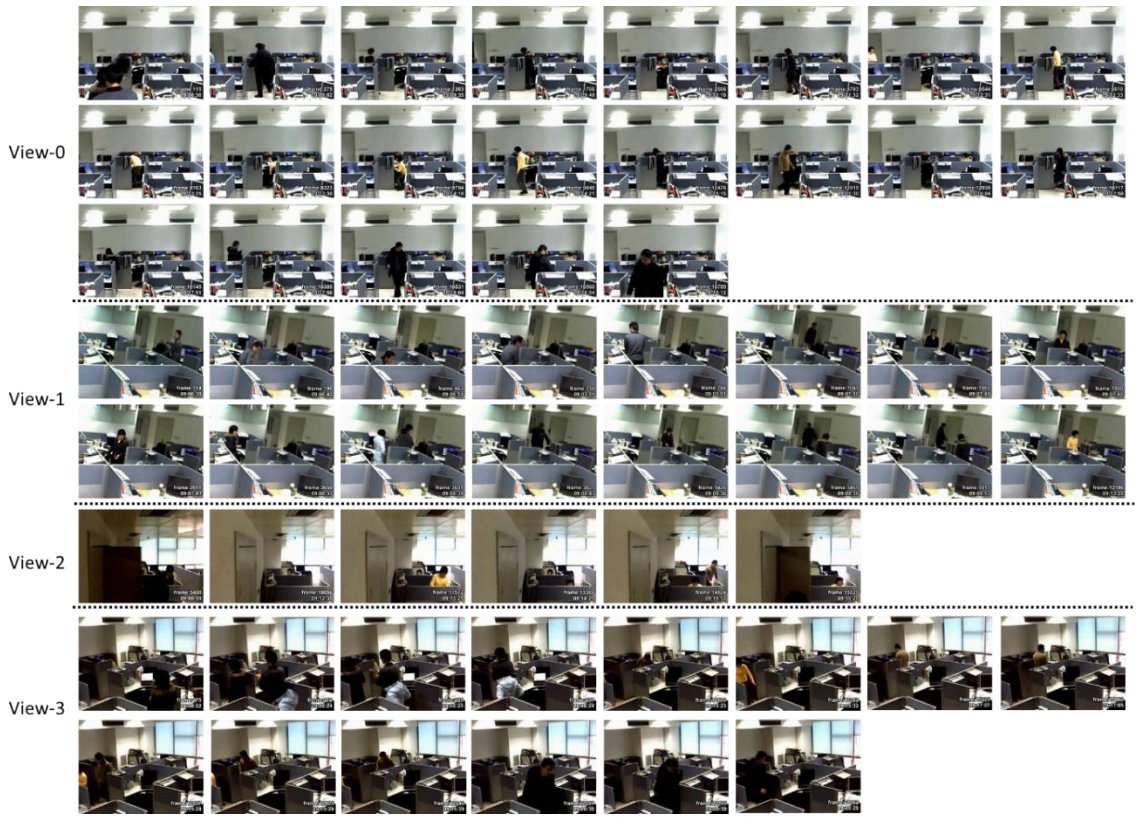

Figure 1: Keyframes extracted by our proposed framework from MVS Office dataset.

2. Activity recognition evaluation (Section IV.C in main paper)

In **Figure 2(a)**, the confusion matrix for YouTube 11 action dataset [3] is given for a model trained over the VGG-19 features without any encoding scheme with a linear SVM. Similarly, the ROC curve of testing data on VGG-19 features is visualized in **Figure 2(b)** for linear SVM. Results of different other SVM variants are also given in **Figure 3, Figure 4, and Figure 5** with respective descriptions. Some sample results in video format are also included in the zip file. The video contains most challenging actions having inter and intra class variations, where it can be seen that most of the actions are predicted accurately and few of them are miss-predicted by our system.

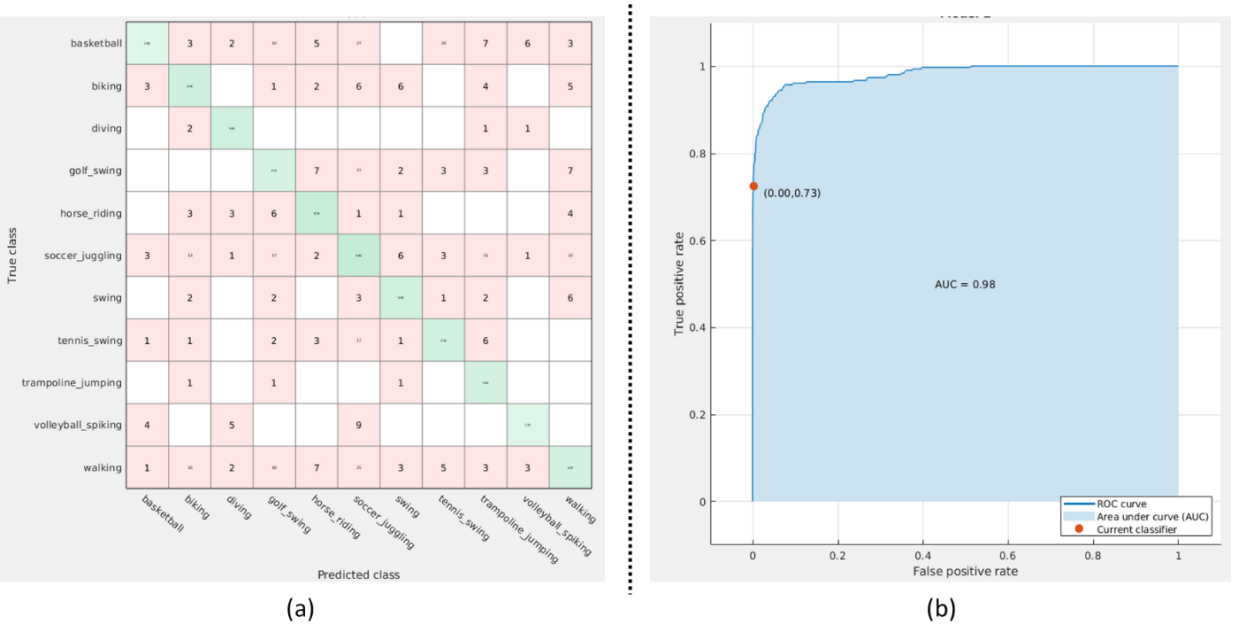

*Figure 2: The test set results on YouTube 11 dataset using **Linear SVM** with **VGG-19 features**. (a) confusion matrix and (b) ROC curve*

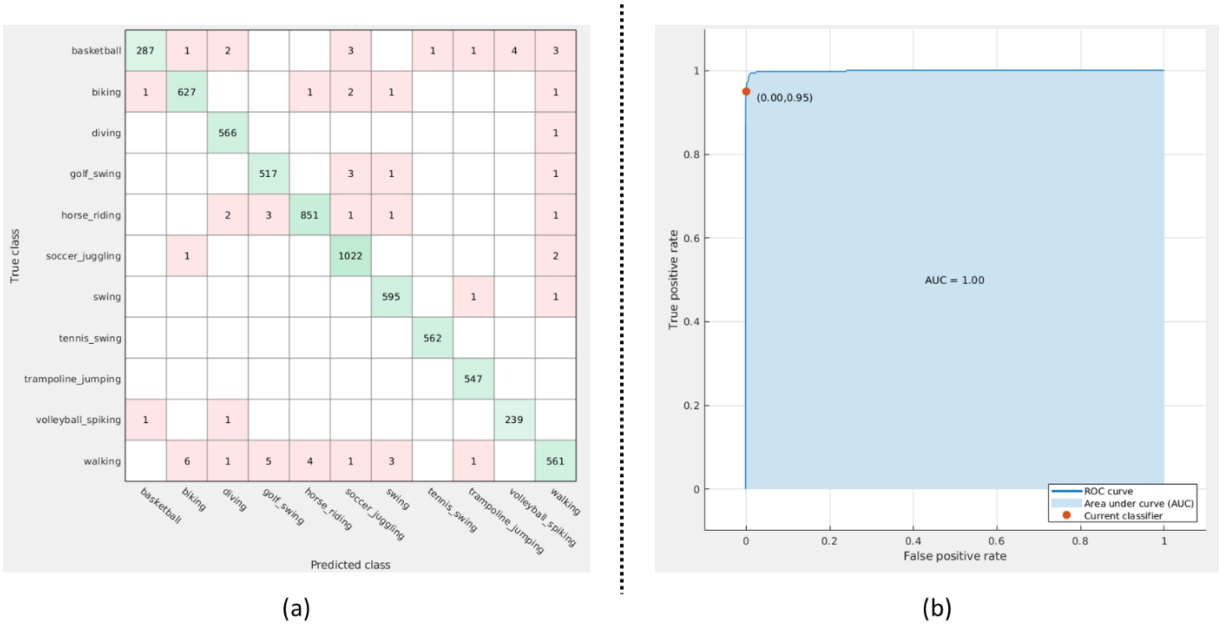

*Figure 3: The test set results on YouTube 11 dataset using **Quadratic SVM** with **VGG-19 features**. . (a) confusion matrix and (b) ROC curve*

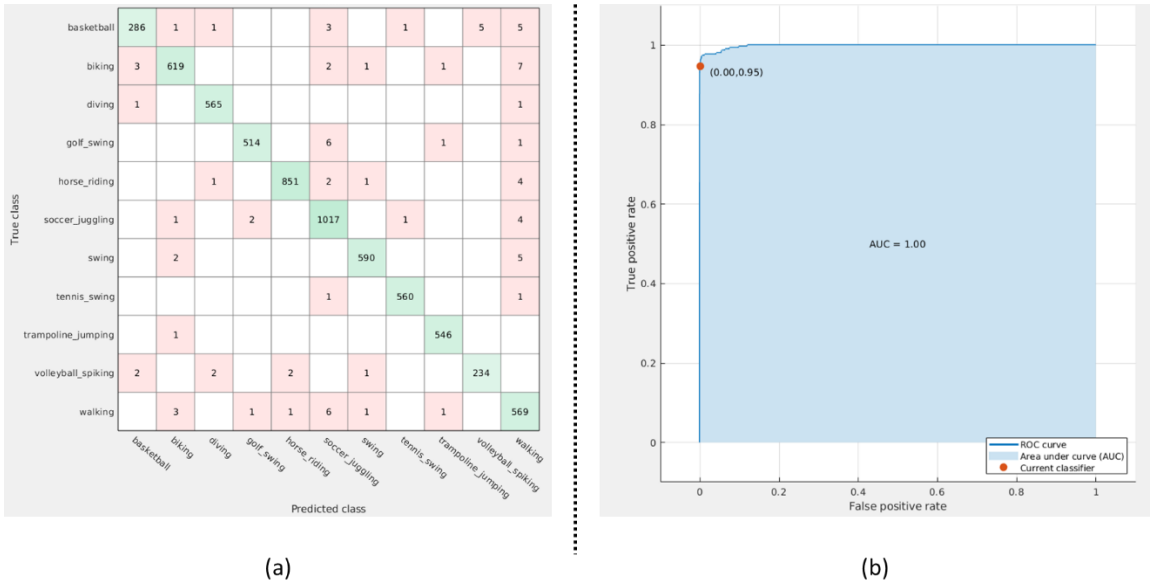

*Figure 4: The test set results on YouTube 11 dataset using **Cubic SVM** with **VGG-19 features**. (a) confusion matrix and (b) ROC curve*

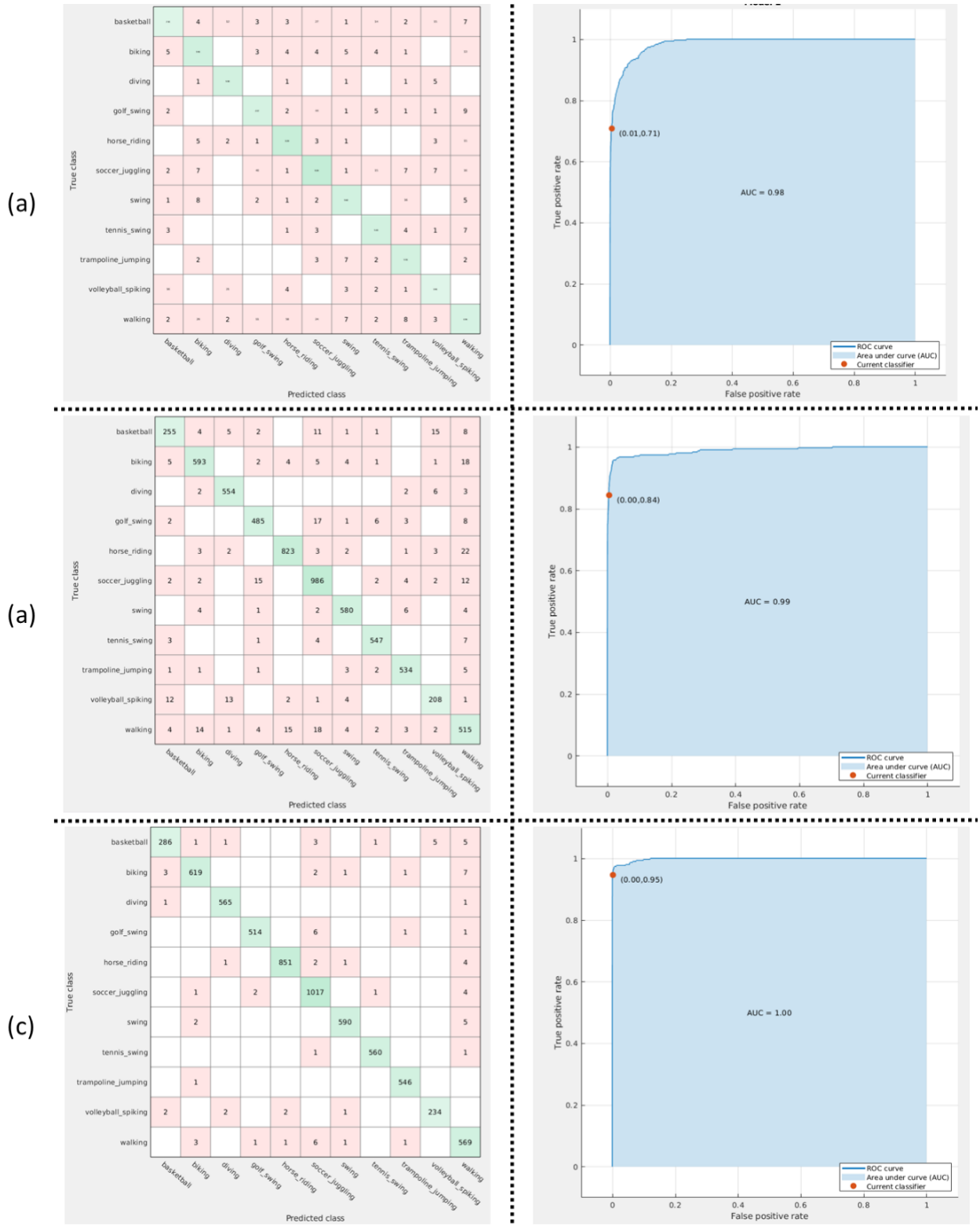

*Figure 5: The test set results on YouTube 11 dataset using SVM variants over **encoded features**. (a) results visualization for **Linear SVM,** (b) graphs for **Quadratic SVM**, and (c) graphs for **Cubic SVM**. These graphs are obtained by encoding features of 15000 dimension to 1000.*

References

[1] Y. Fu, Y. Guo, Y. Zhu, F. Liu, C. Song, and Z.-H. Zhou, "Multi-view video summarization," *IEEE Transactions on Multimedia,* vol. 12, pp. 717-729, 2010.

[2] S.-H. Ou, C.-H. Lee, V. S. Somayazulu, Y.-K. Chen, and S.-Y. Chien, "On-line multi-view video summarization for wireless video sensor network," *IEEE Journal of Selected Topics in Signal Processing,* vol. 9, pp. 165-179, 2015.

[3] J. Liu, J. Luo, and M. Shah, "Recognizing realistic actions from videos “in the wild”," in *Computer vision and pattern recognition, 2009. CVPR 2009. IEEE conference on*, 2009, pp. 1996-2003.