

What do 15,000 object categories tell us about classifying and localizing actions?

Qilei Zhang

Jul 10 2018

Abstract

This paper helps to automatically classify and localize human behavior in video. Motion is the key component of modern methods. Evaluation has the advantages of objects in video representation. Instead of taking into account a few carefully selected and localized objects, the benefits encoding the positive study of the 15000 object categories use 6 data sets, totaling more than 200 hours of video and covering 180 action classes.

1. Introduction

It helps to automatically classify and locate human behaviors, such as telephone, horse riding and sumo wrestling video content. Unlike the leading technologies in these two challenging problems, research has the benefits of objects in video representation. For action classification, the relationship between objects and actions has been considered earlier, but only for a few selected and localized objects [1].

2. Related work

Advancements in action classification have resulted in a mature repertoire of elegant and reliable techniques yielding good accuracy. Such techniques include sampling at interest points, densely or along dense trajectories. Such samples are represented by powerful local descriptors which are robust to modest appearance and motion changes [2].

3. Conclusion

Experiments show that objects matter for actions, and are often semantically relevant as well, especially when the actions interact with objects. What is more, the object representation is complementary to modern motion

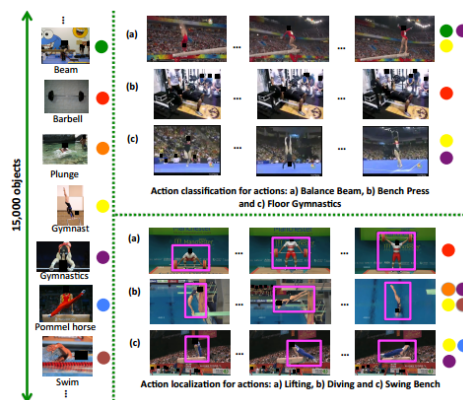


Figure 1. In this paper we ask ourselves the question: What do 15,000 object categories tell us about classifying and localizing actions? and conduct an empirical study on the benefit of encoding object categories for action. We show that: objects matter for actions, actions have object preference, object-action relations are generic, and adding object encodings improves the state-of-the-art in action classification and localization.



Figure 2. Action examples from the UCF101, THUMOS14, Hollywood2, HMDB51, UCF Sports and KTH datasets that we consider in our empirical study to reveal what 15,000 object categories tell us about classifying and localizing actions.

encodings [3].

The determination of action has an object preference. Instead of using all objects, selection is advantageous in terms of compact video representation, as well as in its discriminative action classification ability. The object action relation of learning is universal and transferred from one data set to another. When object representation is combined with modern motion coding and space-time action proposals, it leads to the action classification and localization of the most advanced five data sets in a new country.

References

- [1] L. Hazelhoff. Large scale detection, classification and localization of traffic signs. *Journal of Clinical Microbiology*, 43(2):489–523, 2015. [1](#)
- [2] W. Huang, D. Wang, and P. Ratilal. Fin and humpback whale vocalization classification and localization, northern georges bank. *Journal of the Acoustical Society of America*, 138(3):1753–1753, 2015. [1](#)
- [3] V. Jagadeesh, J. Anderson, B. Jones, R. Marc, S. Fisher, and B. S. Manjunath. Synapse classification and localization in electron micrographs. *Pattern Recognition Letters*, 43(43):17–24, 2014. [1](#)