LSVC2017: Large-Scale Video Classification Challenge

Zuxuan Wu[1], Yu-Gang Jiang[2], Larry S. Davis[1], Shih-Fu Chang[3]
[1] University of Maryland, [2] Fudan University, [3] Columbia University
{zxwu,lsd}@umiacs.umd.edu,ygj@fudan.edu.cn,sfchang@cs.columbia.edu

ABSTRACT

Recognizing visual contents in unconstrained videos has become a very important problem for many applications, such as Web video search and recommendation, smart advertising, robotics, etc. This workshop and challenge aims at exploring new challenges and approaches for large-scale video classification with large number of classes from open source videos in a realistic setting, based upon an extension of Fudan-Columbia Video Dataset (FCVID). This newly collected dataset contains over 8000 hours of video data from YouTube and Flicker, annotated into 500 categories. We hope this dataset can stimulate innovative research on this challenging and important problem.

KEYWORDS

Video classification; Challenge; Video dataset

ACM Reference format:
Zuxuan Wu[1], Yu-Gang Jiang[2], Larry S. Davis[1], Shih-Fu Chang[3]. 2017. LSVC2017: Large-Scale Video Classification Challenge. In *Proceedings of MM '17, Mountain View, CA, USA, October 23–27, 2017,* 2 pages.
DOI: https://doi.org/10.1145/3123266.3138874

1 INTRODUCTION

Today's digital contents are inherently multimedia: text, audio, image, video and etc. Video, in particular, becomes a new way of communication between Internet users with the proliferation of sensor-rich mobile devices. Accelerated by the tremendous increase in Internet bandwidth and storage space, video data has been generated, published and spread at an astounding speed every hour and every day, becoming an indispensable part of today's big data. This has encouraged the development of advanced techniques for a broad range of video understanding applications. A fundamental issue that underlies the success of these technological advances is the understanding of video contents.

The objectives of the challenge are twofold: a) to serve as a benchmark and enable a comparison of different approaches on the task of video classification in large-scale realistic video settings; b) to advance the state-of-the-art. The great success of deep learning in the image domain largely relies on the large-scale dataset: ImageNet. However, for videos, existing datasets for video content recognition are either small or do not have reliable manual labels. Therefore, we hope this challenge could help stimulate future research on video classification by expanding the FCVID dataset to be large scale in terms of video categories as well as number of videos.

MM '17, October 23-27, 2017, *Mountain View, CA, USA*
 978-1-4503-4906-2/17/10…$15.00
DOI: https://doi.org/10.1145/3123266.3138874

2 DATASET

This newly collected dataset contains over 8000 hours of video data from YouTube and Flicker, annotated into 500 categories. The categories cover a wide range of popular topics like social events (e.g., "tailgate party"), procedural events (e.g., "making cake"), objects (e.g., "panda"), scenes (e.g., "beac), etc. Compared with FCVID, new categories are added to enrich the original hierarchy. For example, 76 new categories are added to "cooking" totaling 93 classes, and 75 new classes are added to "sports". During annotation, multiple labels have been considered as much as possible for each video. When labeling a particular category, categories that are not likely to co-occur are filtered out manually with the remaining labels considered for annotation.

The following components will be publicly available under this challenge:

- Training Set: over 62,000 temporally untrimmed videos from 500 classes. We also provide pre-extracted features and frames (1 fps). Validation Set: around 15,000 videos with annotations of classes.
- Test Set: over 78,000 temporally untrimmed videos with withheld ground truth.

We evaluate the success of the proposed methods based on mean Average Precision (mAP) across all categories.

3 ORGANIZERS

Zuxuan Wu is currently pursuing the Ph.D. at the University of Maryland, College Park. His current research interests include computer vision, machine learning, and multimedia. He received the M.S. degree from Fudan University, Shanghai, China.

Yu-Gang Jiang is a Professor in School of Computer Science and Vice Director of Shanghai Engineering Research Center for Video Technology and System at Fudan University, China. His Lab for Big Video Data Analytics conducts research on all aspects of extracting high-level information from big video data, such as video event recognition, object/scene recognition and large-scale visual search. He is the lead architect of a few best-performing video analytic systems in worldwide competitions such as the annual U.S. NIST TRECVID evaluation. His visual concept detector library (VIREO-374) and video datasets (e.g., CCV, FCVID and THUMOS) are widely used resources in the research community. His work has led to many awards, including "emerging leader in multimedia" award from IBM T.J. Watson Research in 2009, early career faculty award from Intel and China Computer Federation, the inaugural ACM China Rising Star Award, the 2015 ACM SIGMM Rising Star Award, and the research award for outstanding young researchers from NSF China. He holds a PhD in Computer Science from City University of Hong

Kong and spent three years working at Columbia University before joining Fudan in 2011.

Larry S. Davis received the BA degree from Colgate University in 1970 and the MS and PhD degrees in computer science from the University of Maryland in 1974 and 1976, respectively. From 1977 to 1981, he was an assistant professor in the Department of Computer Science at the University of Texas, Austin. He returned to the University of Maryland as an associate professor in 1981. From 1985 to 1994, he was the director of the University of Maryland Institute for Advanced Computer Studies. He is currently a professor in the institute and in the Computer Science Department.

Shih-Fu Chang received the Ph.D. degree in electrical engineering and computer sciences from the University of California at Berkeley, Berkeley, CA, USA, in 1993. In his current capacity as Senior Executive Vice Dean of Columbia Engineering School, Columbia University, New York, NY, USA, he plays a key role in strategic planning, research initiatives, and faculty development. He is a leading researcher in multimedia information retrieval, computer vision, signal processing, and machine learning. His work set trends in areas such as content-based image search, video recognition, image authentication, hashing for large image database, and novel application of visual search in brain-machine interface and mobile systems. Impact of his work can be seen in more than 300 peer-reviewed publications, best paper awards, 25 issued patents, and technologies licensed to many companies. Dr. Chang has been recognized with the IEEE Signal Processing Society Technical Achievement Award, the ACM Multimedia SIG Technical Achievement Award, the IEEE Kiyo Tomiyasu Award, the ONY YIA award, the IBM Faculty Award, and the Great Teacher Award from the Society of Columbia Graduates. He served as the Editor-in-Chief of the IEEE SIGNAL PROCESSING MAGAZINE during 2006âĂŞ2008. He is a Fellow of the American Association for the Advancement of Science.