

The Wild World of Computer Vision Datasets-- “Sports Videos in the Wild”

Executive Summary

Improving algorithmic performance in computer vision research requires constructing ever larger datasets with more challenging conditions, such as low-lighting or poor resolution. Some of these novel computer vision research datasets invite concerns for individuals' privacy and data protection rights, including privacy of children's data. “[Sports Videos in the Wild](#)” is one example of a computer vision research dataset that raises privacy concerns.

What are Computer Vision Research Datasets?

Computer vision researchers strive to build models that automate performance of common human tasks like object detection, human action recognition, and facial recognition. Researchers build datasets to advance research goals, such as improving scores on model performance metrics, like better F1 scores or ROC, for specific tasks, such as human action recognition, under ever more challenging and realistic conditions, such as poor camera resolution. Improving performance in these areas means that these technologies can be introduced into complex systems, such as exercise and physical mobility assisting robots, with greater confidence that they will perform in “real world” settings.

Computer vision research datasets come in two main forms-- controlled and unconstrained. Controlled datasets use professionally captured videos and/or videos taken in carefully controlled environments like physiology laboratories or the Olympic games. Uncontrolled datasets aggregate user-captured videos often taken in unstructured, highly variable, environments like homes, neighborhood fields, and garages. Computer vision algorithms for action recognition perform consistently well on controlled datasets, with high accuracy scores (>90%). But, action recognition accuracy on uncontrolled datasets continues to be poor, with scores in the 60-70% accuracy range. Alas, humans' use of AI to recognize actions in their daily lives will depend on performance in uncontrolled settings.

What is “Sports Videos in the Wild”?

Human action recognition datasets like [UCF Sports](#), [Olympic Sports Dataset](#), and [Hollywood2](#), are built from professionally captured videos of many different actions taken by professional or world-class level performers, from Olympic water polo goals to Humphrey Bogarts' goodbye kisses. Sports Videos in the Wild (SVW) is, at a high level, an unconstrained action recognition dataset available to researchers. It is built for sports action recognition using user-level, uncontrolled, videos of amateur, novice, sports participants that range in age from pre-kindergarten to 60 or even 70+ years. The uncontrolled nature of the videos makes SVW a good dataset against which to test action detection and action recognition algorithms where there is poor video quality (i.e., resolution and stabilization). The amateur nature of the sports subjects makes SVW a singular dataset for sports genre categorization that is robust to the setting in which most of us view, capture, or play sports.

Who Gathered the Sports Videos in the Wild Dataset?

Sports Videos in the Wild (SVW) was gathered from the publicly usable videos posted to the [“Coach’s Eye”](#) mobile app. Publicly usable videos are videos uploaded to the App by account holders who choose to make their content “publicly available”. Importantly, there is no implication that the account holder uploading videos has the consent of the individuals captured to have their likeness and action uploaded. This is assumed to have been done by “coaches” interacting with their “athletes”.

As outlined in the Coach’s Eye [“Terms of Service”](#):

"Content") that you post (“your Content) on or through the Services. By displaying or publishing (“posting”) any of your Content on or through the Services, you hereby grant to TechSmith a non-exclusive, fully paid and royalty- free, irrevocable, worldwide, license (with the right to sublicense third parties and users/account owners of the Service) to use, modify, delete from, add to, publicly perform, publicly display, reproduce, download, bookmark and translate your Content, including without limitation, distributing part or all of your Content in any media formats through any media channels, except your Content that is not shared publicly (“private Content”- Go to settings page to control who sees your Content when this feature is available. Applying the appropriate level of access to your Content is solely your responsibility). As a result of the above, you acknowledge and understand that your Content may be able to be viewed by other users of the Services and through third party services and sites and you should not submit Content that you are not comfortable sharing with others under the Terms.

SVW was created collaboratively between researchers from Michigan State University and TechSmith. The 85,000 publicly usable videos excerpted for the SVW dataset were manually annotated by a research team for 30 sports categories and 44 action categories.

Why do Action Recognition and Sports Genre Categorization Matter?

Action detection and action recognition algorithms are necessary components of robotic or other autonomous/ semi-autonomous systems (e.g., vehicles) that interact with humans. Sports genre classification and sports action recognition will also be included in systems like autonomous vehicles that will need to recognize a human waving at the car or swinging a racket. For the specific case of autonomous vehicles and human interacting robots, learning from controlled data of amateur motion is important so that these systems can identify the action “running” whether this is performed by champion 10k runners racing or teens running for physical education classes. In order for these systems to predict human action and respond accordingly, they must be trained to detect and classify (recognize) actions using training datasets and pre-trained algorithms. Sports genre categorization is a specific form of recognition and classification that can also serve the multi-million dollar sports industry, such as by training auto-captioning services for those who are hard of hearing.

What are the Ethical Concerns?

There are three ethical concerns associated with the SVW dataset: 1) subject awareness and control of this secondary use of their data, 2) inclusion of images of children, possibly without parental consent, 3) inclusion of information that violates’ users’ privacy or causes embarrassment, and 4) unknown methods and representativeness of the data labelling team.

1. The transmission of Coach’s Eye users’ data to third parties is included in the “Terms of Service”. However, the use of videos for this or other secondary, research, purposes is not described in the Terms of Service. Onward transmission of these videos from the

researchers' website to others is also not described as a possibility for uses of the users' posted videos. While Coach's Eye indicates that they do not claim any ownership rights, the researchers' make the dataset available via download to anyone able to answer 3 questions on a [Google Form](#).

2. An examination of the videos included in the SVW dataset reveals that there are images of children performing as well as images of other children watching others play sports. In some cases, these videos are shot with full view of houses, bedrooms, garages, and school facilities in the background. In other cases, children's faces are visible as are their school or team uniforms. These considerations may be particularly important where videos of "cheerleading", "gymnastics", "diving" and "swimming" are involved as these sports often require uniforms that are, even if worn properly, revealing of the participants' body contours. While it could be presumed that parents approved of the coaches' use of the technology, if parents did not extensively review the Terms of Service, they may not be aware of the onward transmission of videos of their children to third parties. The challenges of parental or guardians' awareness of the onward transmission of videos for research purposes to anyone interested, which are described above, create additional complications for the use of children's data here.
3. Examining the videos included showed that the homes, personal property, team uniforms and numbers, and even initials of persons are included in the brief snippets of videos. This includes representations of the persons performing the actions and of bystanders, coaches, and other participants. It also includes business logos and brands and the logos and markings for police, ambulance and other emergency services departments. Their representations here could violate copyright laws but also be used for unrelated purposes in computer vision, such as brand identification models. In addition, some videos represent actions that are "failures" in the techniques, which could constitute embarrassment for persons included.
4. The methods that the researchers collecting the SVW dataset used to select videos or to label them are not well described in the [research paper accompanying the dataset](#). It is not clear who labelled the data: if it was the research team authoring the paper, this is a team of corporate (TechSmith, the parent company for Coach's Eye) and research professionals whose professional qualifications can be evaluated. If the data was labelled through crowdsourced means (e.g., Amazon's Mechanical Turk), it is not clear what information was given to the labelers or what considerations for the privacy of users were imparted to the labelers. The sporting background of the dataset team is not clarified which introduces the possibility of labeling error. For example, under the category "rowing", six videos clearly contain representations of the sport "dragon boating", which is not a rowing discipline.

Concluding Points

Sports Videos in the Wild (SVW) is a unique and valuable dataset. As other action detection and classification datasets are built under controlled conditions or are built using broad use sources, such as YouTube videos, SVW provides action recognition researchers with a new benchmark dataset for performance using data that is uncontrolled on multiple dimensions that mimic the "real world". Such datasets may ultimately improve algorithmic performance for this important area of computer vision research. However, the question whether improvements in algorithmic performance should come at the expense of persons' control of their data, including parents' control of the uses of their children's data, is an important one to ask.