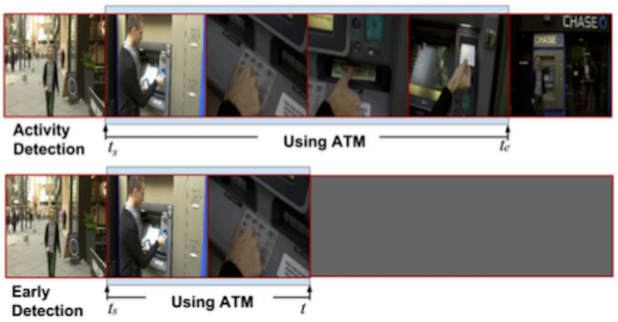
New method detects human activity in videos earlier and more accurately

23 June 2016



Credit: Disney Research

Researchers at Disney Research and Boston University have found that a machine learning program can be trained to detect human activity in a video sooner and more accurately than other methods by rewarding the program for gaining confidence in its prediction the longer it observes the activity.

It seems intuitive that the program would grow more confident that it is detecting, say, a person changing a tire, the longer it observes the person loosening lugnuts, jacking up the car and subsequently removing the wheel, but that's not the way most computer models have been trained to detect activity, said Leonid Sigal, senior research scientist at Disney Research.

"Most training techniques are happy if the computer model gets 60 percent of the video frames correct, even if the errors occur late in the process, when the activity should actually be more apparent," Sigal said. "That doesn't make much sense. If the model predicts a person is making coffee even after it sees the person put pasta into boiling water, it should be penalized more than if it made the same incorrect prediction when the person was still just boiling water."

Shugao Ma, a Ph.D. student in computer science at Boston University and a former intern at Disney Research, found that this change in training methods resulted in more accurate predictions of activities. The computer also was often able to accurately predict the activity early in the process, even after seeing only 20 to 30 percent of the video. Likewise, the program can detect that an activity is finished if its confidence that it is observing that activity begins to drop.

The research team, which included Stan Sclaroff, Boston University professor of computer science, will present their findings June 26 at the Computer Vision and Pattern Recognition conference, CVPR 2016, in Las Vegas.

"Automatic detection of human activities in videos has many potential applications, such as video retrieval and human-computer interaction," said Jessica Hodgins, vice president at Disney Research. "Human-robot interaction applications in particular could benefit from early detection of activities; a caregiving robot, for instance, would want to recognize that an elderly patient was in danger of falling so it could act to steady the patient."

Activity detection remains a challenging technical task because there are so many variables in actors, their appearance, surroundings and viewpoints. Understanding the progression of the activity - "making pasta," for instance, might include setting a pot on a stove, boiling water, boiling noodles, draining, etc. - is thus critical for a computer program to recognize the activity, Ma said.

The researchers used Long Short Term Memory (LSTM), a type of recurrent neural network that is well-suited to learn how to classify, process and predict time series, for this task. They introduced the concept of ranking losses in the learning objectives, which are computed for each time point in the prediction, so that the detection score gets

higher as the action progresses. They found that LSTM models that used the ranking losses had better performance than LSTM models that were trained only based on whether the activity was classified correctly.

More information: "Learning Activity Progression in LSTMs for Activity Detection and Early Detection-Paper"
[PDF, 3.42 MB] s3-us-west-1.amazonaws.com/dis … -Detection-Paper.pdf

Provided by Disney Research

APA citation: New method detects human activity in videos earlier and more accurately (2016, June 23) retrieved 22 October 2019 from https://phys.org/news/2016-06-method-human-videos-earlier-accurately.html

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.