Using deep-learning techniques to locate potential human activities in videos

August 30 2018

Credit: AI-generated image (disclaimer)

When a police officer begins to raise a hand in traffic, human drivers realize that the officer is about to signal them to stop. But computers find it harder to work out people's next likely actions based on their current behavior. Now, a team of A*STAR researchers and colleagues has developed a detector that can successfully pick out where human

actions will occur in videos, in almost real-time.

Image analysis technology will need to become better at understanding human intentions if it is to be employed in a wide range of applications, says Hongyuan Zhu, a computer scientist at A*STAR's Institute for Infocomm Research, who led the study. Driverless cars must be able to detect police officers and interpret their actions quickly and accurately, for safe driving, he explains. Autonomous systems could also be trained to identify suspicious activities such as fighting, theft, or dropping dangerous items, and alert security officers.

Computers are already extremely good at detecting objects in static images, thanks to deep learning techniques, which use artificial neural networks to process complex image information. But videos with moving objects are more challenging. "Understanding human actions in videos is a necessary step to build smarter and friendlier machines," says Zhu.

Previous methods for locating potential human actions in videos did not use deep-learning frameworks and were slow and prone to error, says Zhu. To overcome this, the team's YoTube detector combines two types of neural networks in parallel: a static neural network, which has already proven to be accurate at processing still images, and a recurring neural network, typically used for processing changing data, for speech recognition. "Our method is the first to bring detection and tracking together in one deep learning pipeline," says Zhu.

The team tested YoTube on more than 3,000 videos routinely used in computer vision experiments. They report that it outperformed state-of-the-art detectors at correctly picking out potential human actions by approximately 20 per cent for videos showing general everyday activities and around 6 per cent for sports videos. The detector occasionally makes mistakes if the people in the video are small, or if there are many people

in the background. Nonetheless, Zhu says, "We've demonstrated that we can detect most potential human action regions in an almost real-time manner."

More information: Hongyuan Zhu et al. YoTube: Searching Action Proposal Via Recurrent and Static Regression Networks, *IEEE Transactions on Image Processing* (2018). DOI: 10.1109/TIP.2018.2806279

Provided by Agency for Science, Technology and Research (A*STAR), Singapore

Citation: Using deep-learning techniques to locate potential human activities in videos (2018, August 30) retrieved 9 April 2024 from https://phys.org/news/2018-08-deep-learning-techniques-potential-human-videos.html