Machine detection of human-object interaction in images and videos

July 1 2019, by Barbara L. Micale

Jia-Bin Huang, assistant professor in the Bradley Department of Electrical and Computer Engineering and a faculty member at the Discovery Analytics Center. Credit: Virginia Tech

Jia-Bin Huang, assistant professor in the Bradley Department of

Electrical and Computer Engineering and a faculty member at the Discovery Analytics Center, has received a Google Faculty Research Award to support his work in detecting human-object interaction in images and videos.

The Google award, which is in the Machine Perception category, will allow Huang to tackle the challenges of detecting two aspects of human-object interaction: modeling the relationship between a person and relevant objects/scene for gathering contextual information and mining hard examples automatically from unlabeled but interaction-rich videos.

According to Huang, while significant progress has been made in classifying, detecting, and segmenting objects, representing images/videos as a collection of isolated object instances has failed to capture the information essential for understanding activity.

"By improving the model and scaling up the training, we aim to move a step further toward building socially intelligent machines," Huang said.

Given an image or a video, the goal is to localize persons and object instances, as well as recognize interaction, if any, between each pair of a person and an object. This provides a structured representation of a visually grounded graph over the humans and the object instances they interact with.

For example: Two men are next to each other on the sidelines of a tennis court, one standing up and holding an umbrella and one sitting on a chair holding a tennis racquet and looking at a bag on the ground beside him. As the video progresses, the two smile at each other, exchange the umbrella and tennis racquet, sit side by side, and drink from water bottles. Eventually, they turn to look at each other, exchange the umbrella and tennis racquet again, and finally, talk to one another.

"Understanding [human activity](#) in images and/or videos is a fundamental step toward building socially aware agents, semantic image/[video](#) retrieval, captioning, and question-answering," Huang said.

He said that detecting human-computer interaction leads to a deeper understanding of human-centric activity.

"Instead of answering 'What is where?' the goal of human-object interaction detection is to answer the question 'What is happening?' The outputs of human-object interaction provide a finer-grained description of the state of the scene and allow us to better predict the future and understand their intent," Huang said.

Ph.D. student Chen Gao will work on the project with Huang. They expect that the research will significantly advance state-of-the-art human-object detection and enable many high-impact applications, such as long-term health monitoring and socially aware robots.

Huang plans to share results of the research via publications at top-tier conferences and journals and will also make the [source code](#), collected datasets, and pre-trained models produced from this project publicly available.

"Our project aligns well with several of Google's on-going efforts to build 'social visual intelligence.' We look forward to engaging with researchers and engineers at Google to exchange and share ideas and foster future collaborative relationships," Huang said.

Provided by Virginia Tech

Citation: Machine detection of human-object interaction in images and videos (2019, July 1) retrieved 27 April 2024 from