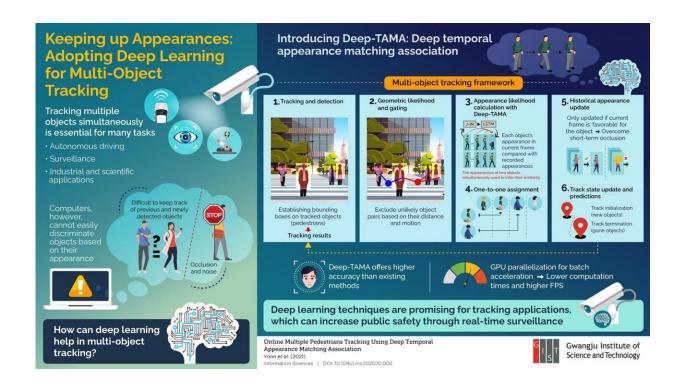# Scientists adopt deep learning for multi-object tracking

July 19 2021

A novel framework achieves state-of-the-art performance without sacrificing efficiency in public surveillance tasks. Credit: Gwangju Institute of Science and Technology

Computer vision has progressed much over the past decade and made its way into all sorts of relevant applications, both in academia and in our daily lives. There are, however, some tasks in this field that are still extremely difficult for computers to perform with acceptable accuracy

and speed. One example is object tracking, which involves recognizing persistent objects in video footage and tracking their movements. While computers can simultaneously track more objects than humans, they usually fail to discriminate the appearance of different objects. This, in turn, can lead to the algorithm to mix up objects in a scene and ultimately produce incorrect tracking results.

At the Gwangju Institute of Science and Technology in Korea, a team of researchers led by Professor Moongu Jeon seeks to solve these issues by incorporating deep learning techniques into a multi-object tracking framework. In a recent study published in *Information Sciences*, they present a new tracking model based on a technique they call 'deep temporal appearance matching association (Deep-TAMA)' which promises innovative solutions to some of the most prevalent problems in multi-object tracking.

Conventional tracking approaches determine object trajectories by associating a bounding box to each detected object and establishing geometric constraints. The inherent difficulty in this approach is in accurately matching previously tracked objects with objects detected in the current frame. Differentiating detected objects based on hand-crafted features like color usually fails because of changes in lighting conditions and occlusions. Thus, the researchers focused on enabling the tracking model with the ability to accurately extract the known features of detected objects and compare them not only with those of other objects in the frame but also with a recorded history of known features. To this end, they combined joint-inference neural networks (JI-Nets) with long-short-term-memory networks (LSTMs).

LSTMs help to associate stored appearances with those in the current frame whereas JI-Nets allow for comparing the appearances of two detected objects simultaneously from scratch—one of the most unique aspects of this new approach. Using historical appearances in this way

allowed the algorithm to overcome short-term occlusions of the tracked objects. "Compared to conventional methods that pre-extract features from each object independently, the proposed joint-inference method exhibited better accuracy in public surveillance tasks, namely pedestrian tracking," highlights Dr. Jeon. Moreover, the researchers also offset a main drawback of deep learning—low speed—by adopting indexing-based GPU parallelization to reduce computing times. Tests on public surveillance datasets confirmed that the proposed tracking framework offers state-of-the-art accuracy and is therefore ready for deployment.

Multi-object tracking unlocks a plethora of applications ranging from autonomous driving to public surveillance, which can help combat crime and reduce the frequency of accidents. "We believe our methods can inspire other researchers to develop novel deep-learning-based approaches to ultimately improve public safety," concludes Dr. Jeon.