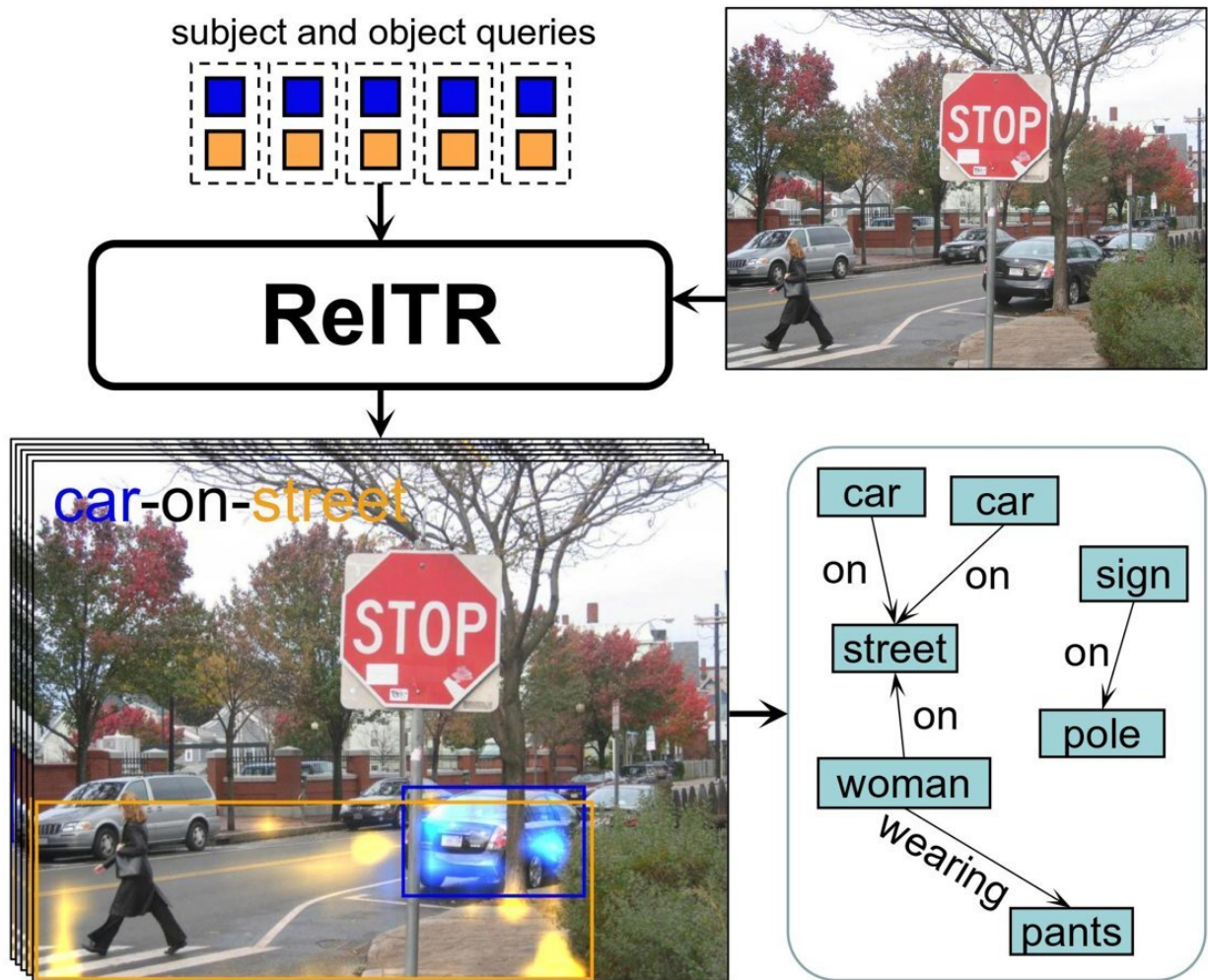# New AI method for graphing scenes from images

June 26 2023, by K. W. Wesselink

Credit: University of Twente

Generative AI programs can generate images from textual prompts. These models work best when they generate images of single objects. Creating complete scenes is still difficult. Michael Ying Yang, a UT-researcher from the faculty of ITC recently developed a novel method that can graph scenes from images that can serve as a blueprint for generating realistic and coherent images. He and his team recently published their findings in the journal *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Humans are excellent at defining relationships between objects. "We can see that a chair is standing on the floor and a dog is walking on the street. AI models find this difficult," explains Yang, assistant professor at the Scene Understanding Group of the Faculty of Geo-Information Science and Earth Observation (ITC). Improving a computer's ability to detect and understand visual relationships is needed for image generation, but could also assist the perception of autonomous vehicles and robots.

From two-stage to single-stage

Currently, methods exist to graph a semantic understanding of an image, but they are slow. These methods use a two-stage approach. At first, they map all objects in a scene. In the second step, some specific neural network goes through all different possible connections and then labels them with the correct relationship.

The number of connections this method has to go through increases exponentially with the number of objects. "Our model takes just a single step. It automatically predicts subjects, objects and their relationships at the same time," says Yang.

Detecting relationships

For this one-stage method, the model looks at the visual features of the objects in the scene and focuses on the most relevant details for determining the relationships. It highlights important areas where objects interact or relate to each other. These techniques and relatively little training data are enough to identify the most important relationships between different objects. The only thing left to do is to generate a description of how they are connected.

"The model detects that in an example picture, the man is very likely to interact with the baseball bat. It's then trained to describe the most likely relationship: 'man-swings-baseball bat,'" says Yang.