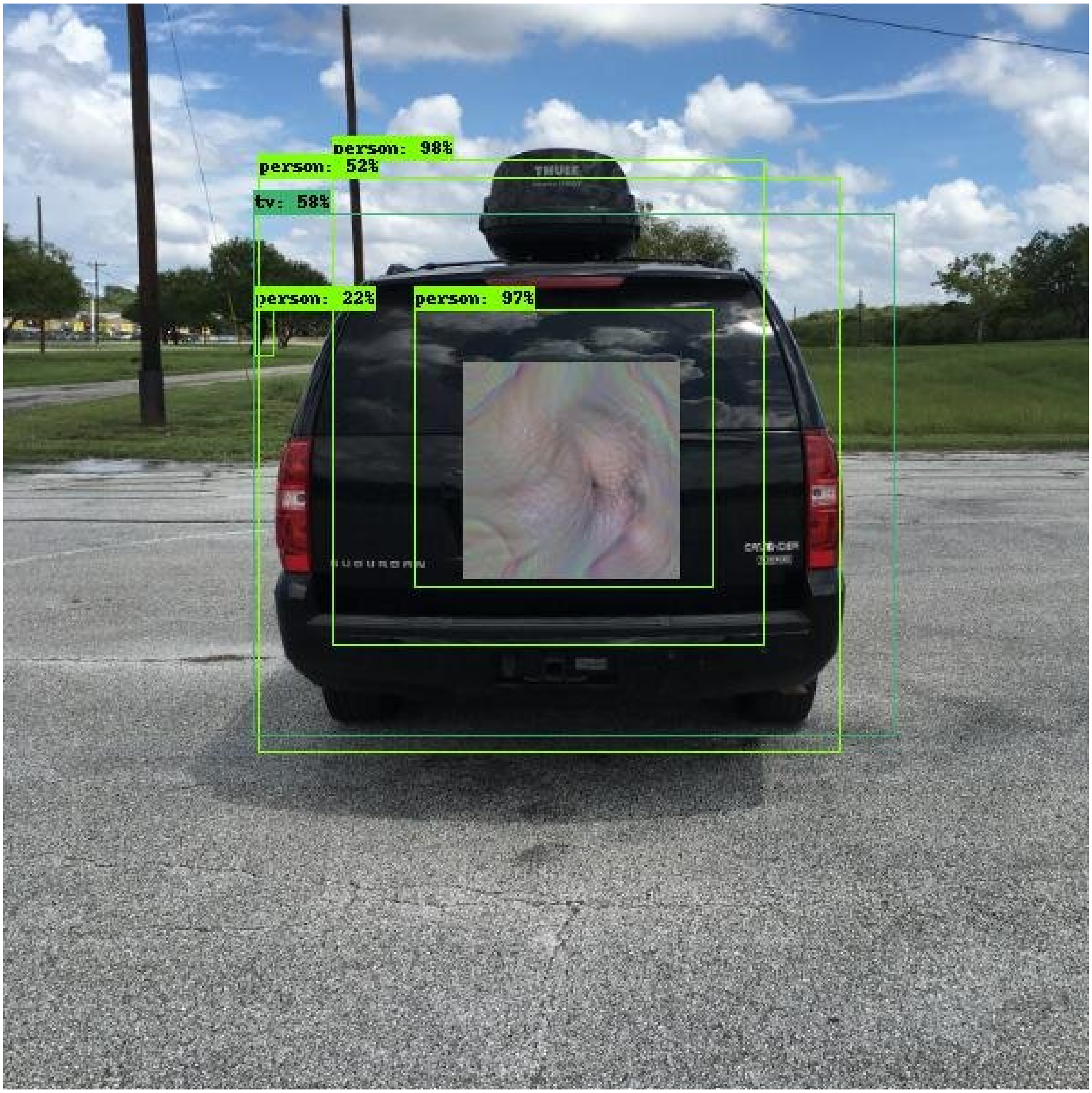
TechXplore

Engineers develop novel techniques to trick object detection systems

April 4 2019



Many of today's vehicles use object detection systems to help avoid collisions. SwRI engineers developed unique patterns that can trick these systems into seeing something else, seeing the objects in another location or not seeing the objects at all. In this photo, the object detection system sees a person rather than a vehicle. This research will allow engineers to thoroughly test object detection systems and improve the security of the deep-learning algorithms they use. Credit: Southwest Research Institute

New adversarial techniques developed by engineers at Southwest Research Institute can make objects "invisible" to image detection systems that use deep-learning algorithms. These techniques can also trick systems into thinking they see another object or can change the location of objects. The technique mitigates the risk for compromise in automated image processing systems.

"Deep-learning neural networks are highly effective at many tasks," says Research Engineer Abe Garza of the SwRI Intelligent Systems Division. "However, deep learning was adopted so quickly that the security implications of these algorithms weren't fully considered."

Deep-learning algorithms excel at using shapes and color to recognize the differences between humans and animals or cars and trucks, for example. These systems reliably detect objects under an array of conditions and, as such, are used in myriad applications and industries, often for safety-critical uses. The automotive industry uses deep-learning object detection systems on roadways for lane-assist, lane-departure and collision-avoidance technologies. These vehicles rely on cameras to detect potentially hazardous objects around them. While the image processing systems are vital for protecting lives and property, the algorithms can be deceived by parties intent on causing harm.

Security researchers working in "adversarial learning" are finding and documenting vulnerabilities in deep- and other machine-learning algorithms. Using SwRI internal research funds, Garza and Senior Research Engineer David Chambers developed what look like futuristic, Bohemian-style patterns. When worn by a person or mounted on a vehicle, the patterns trick object detection cameras into thinking the objects aren't there, that they're something else or that they're in another location. Malicious parties could place these patterns near roadways, potentially creating chaos for vehicles equipped with object detectors.

What looks like a colorful pattern to the human eye looks like a bicycle to an object detection system. While deep-learning algorithms used in these systems are reliable, they can be deceived with special imagery. SwRI researchers are developing techniques to mitigate the risk of compromise in these systems. Credit: Southwest Research Institute

"These patterns cause the algorithms in the camera to either misclassify or mislocate objects, creating a vulnerability," said Garza. "We call these patterns 'perception invariant' adversarial examples because they don't need to cover the entire object or be parallel to the camera to trick the algorithm. The algorithms can misclassify the object as long as they sense some part of the pattern."

While they might look like unique and colorful displays of art to the human eye, these patterns are designed in such a way that object-detection camera systems see them very specifically. A pattern disguised as an advertisement on the back of a stopped bus could make a collision-avoidance system think it sees a harmless shopping bag instead of the bus. If the vehicle's camera fails to detect the true object, it could continue moving forward and hit the bus, causing a potentially serious collision.

"The first step to resolving these exploits is to test the deep-learning algorithms," said Garza. The team has created a framework capable of repeatedly testing these attacks against a variety of deep-learning detection programs, which will be extremely useful for testing solutions.

SwRI researchers continue to evaluate how much, or how little, of the pattern is needed to misclassify or mislocate an object. Working with clients, this research will allow the team to test object detection systems and ultimately improve the security of deep-learning algorithms.

Provided by Southwest Research Institute

Citation: Engineers develop novel techniques to trick object detection systems (2019, April 4) retrieved 14 August 2024 from https://techxplore.com/news/2019-04-techniques.html