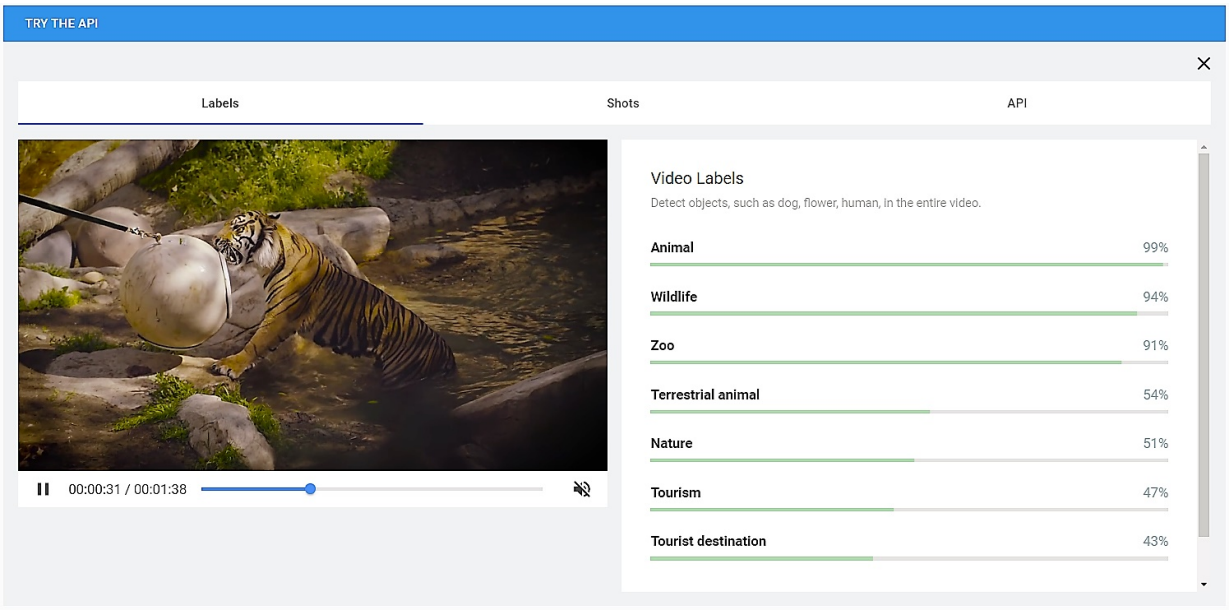
# Security researchers show that Google's AI tool for video searching can be easily deceived

April 3 2017, by Jennifer Langston




This screenshot shows Google Video Intelligence API's output for a sample video named 'animals.mp4,' which is provided by the API website. Google's tool accurately returns video labels that relate to animals and wildlife. Credit: Google Video Intelligence API

University of Washington researchers have shown that Google's new tool that uses machine learning to automatically analyze and label video content can be deceived by inserting a photograph periodically and at a

very low rate into videos. After they inserted an image of a car into a video about animals, for instance, the system returned results suggesting the video was about an Audi.

Google recently released its Cloud Video Intelligence API to help developers build applications that can automatically recognize objects and search for content within videos. Automated video annotation would be a breakthrough technology, helping law enforcement efficiently search surveillance videos, sports fans instantly find the moment a goal was scored or video hosting sites weed out inappropriate content.

Google launched a demonstration website that allows anyone to select a video for annotation. The API quickly identifies the key objects within the video, detects scene changes and provides shot labels of the video events over time. The API website says the system can be used to "separate signal from noise, by retrieving relevant information at the video, shot or per frame" level.

In a new research paper, the UW electrical engineers and security researchers, including doctoral students Hossein Hosseini and Baicen Xiao and professor Radha Poovendran, demonstrated that the API can be deceived by slightly manipulating the videos. They showed one can subtly modify the video by inserting an image into it, so that the system returns only the labels related to the inserted image.

An image of a car that the UW research team inserted into the sample video once every 50 frames. Credit: University of Washington

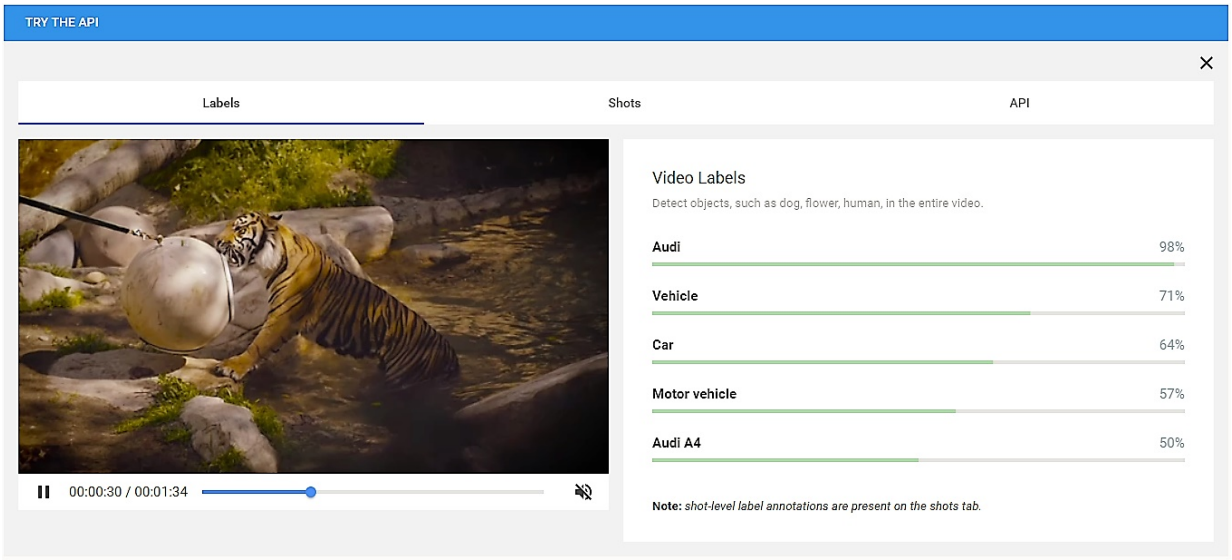
The same research team recently showed that Google's machine-learning-based platform designed to identify and weed out comments from internet trolls can be easily deceived by typos, misspelling offensive words or adding unnecessary punctuation.

"Machine learning systems are generally designed to yield the best performance in benign settings. But in real-world applications, these systems are susceptible to intelligent subversion or attacks," said senior author Radha Poovendran, chair of the UW electrical engineering department and director of the Network Security Lab. "Designing systems that are robust and resilient to adversaries is critical as we move forward in adopting the AI products in everyday applications."

As an example, a screenshot of the API's output in Figure 1 for a sample video named "animals.mp4," which is provided by the API website.

Google's tool does indeed accurately identify the video labels.

The researchers then inserted an image of an Audi car (shown in Figure 2) into the video once every two seconds. The modification is hardly visible, since the image is added once every 50 video frames, for a frame rate of 25.




This screenshot shows the API's output for the wildlife video that was subtly manipulated by UW researchers, who periodically inserted an image of an Audi into it. Now, the Google tool believes with high confidence that the manipulated video is all about the car. Credit: Google Video Intelligence API

Figure 3 shows a screenshot of the API's output for the manipulated video. As seen below, the Google tool believes with high confidence that the manipulated video is all about the car.

"Such vulnerability of the video annotation system seriously undermines its usability in real-world applications," said lead author and UW

electrical engineering doctoral student Hossein Hosseini. "It's important to design the system such that it works equally well in adversarial scenarios."

"Our Network Security Lab research typically works on the foundations and science of cybersecurity," said Poovendran, the lead principal investigator of a recently awarded MURI grant, where adversarial machine learning is a significant component. "But our focus also includes developing robust and resilient systems for machine learning and reasoning systems that need to operate in adversarial environments for a wide range of applications."

Provided by University of Washington