The Visual Chatbot
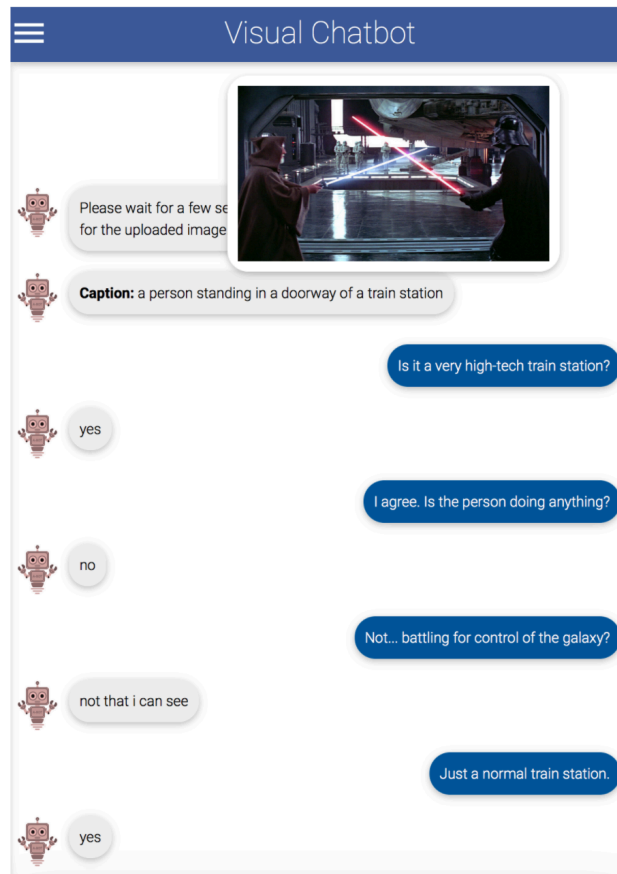
By Janelle Shane On June 21, 2018 - 3 min read

There is a delightful algorithm called Visual Chatbot that will answer questions about any image it sees. It's a demo by a team of academic researchers that goes along with a recent machine learning research paper (and a challenge for anyone who'd like to improve on it), and its performance is pretty state-of-the-art, meant to demonstrate image recognition, language comprehension, and spatial awareness.

However, there are a couple of interesting things to note about this algorithm.

1. It was trained on a large but very specific set of images.
2. It is not prepared for images that aren't like the images it saw in training.
3. When confused, it tends not to admit it.

Now, Visual Chatbot was indeed trained on a huge variety of images. It can answer fairly involved questions about a lot of different things, and that's impressive. The problem is that humans are very weird, and there are still many things it's never seen. (This turns out to be a major challenge for self-driving cars.) And given Visual Chatbot's tendency to react to confusion by digging itself a deeper hole, this can lead to some pretty surreal interactions.