ICME Summer Workshops 2020

Introduction to Deep Learning

Session 4: 3:30—4:45 PM
Instructor: Sherrie Wang

icme-workshops.github.io/deep-learning

Eykholt *et al.* “Robust Physical-World Attacks on Deep Learning Visual Classification” (2018)

Workshop Schedule

Session 1 (9:00—10:30 AM)
- Introduction
- Current state-of-the-art in deep learning
- Math review
- Fully connected neural networks

Session 2 (10:45—12:00 PM)
- Loss functions
- Gradient descent
- Backpropagation
- Overfitting and underfitting

Lunch (12:00—2:00 PM)

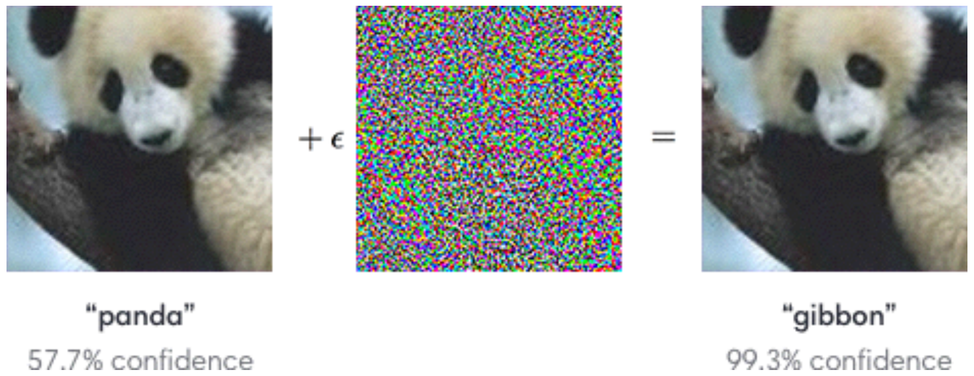
Session 3 (2:00—3:15 PM)
- Convolutional neural networks
- Recurrent neural networks
- Other architectures
- Deep learning libraries
- Hands-on coding session in Tensorflow

Session 4 (3:30—4:45 PM)
- Hands-on coding session in Keras
- Hands-on coding session on transfer learning
- Failures of deep learning

Failures and Limits

Adversarial attacks

- Noise that is imperceptible to humans has been shown to dramatically change neural network output



Real-world adversarial attacks

- Many self-driving cars rely on convolutional neural networks…

Table 1: Sample of physical adversarial examples against LISA-CNN and GTSRB-CNN.

Distance/Angle	Subtle Poster	Subtle Poster Right Turn	Camouflage Graffiti	Camouflage Art (LISA-CNN)	Camouflage Art (GTSRB-CNN)
5′ 0°					
5′ 15°					
10′ 0°					
10′ 30°					
40′ 0°					
Targeted-Attack Success	100%	73.33%	66.67%	100%	80%

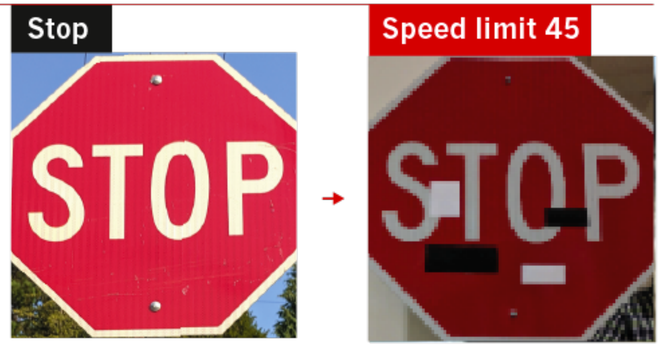
Eykholt *et al.* “Robust Physical-World Attacks on Deep Learning Visual Classification” (2018)

Adversarial attacks

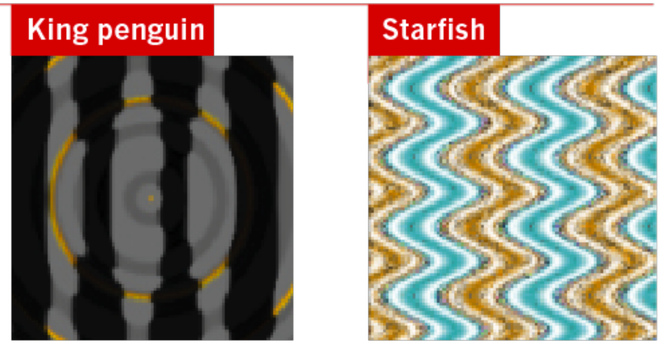
FOOLING THE AI

Deep neural networks (DNNs) are brilliant at image recognition — but they can be easily hacked.

These stickers made an artificial-intelligence system read this stop sign as ‘speed limit 45’.

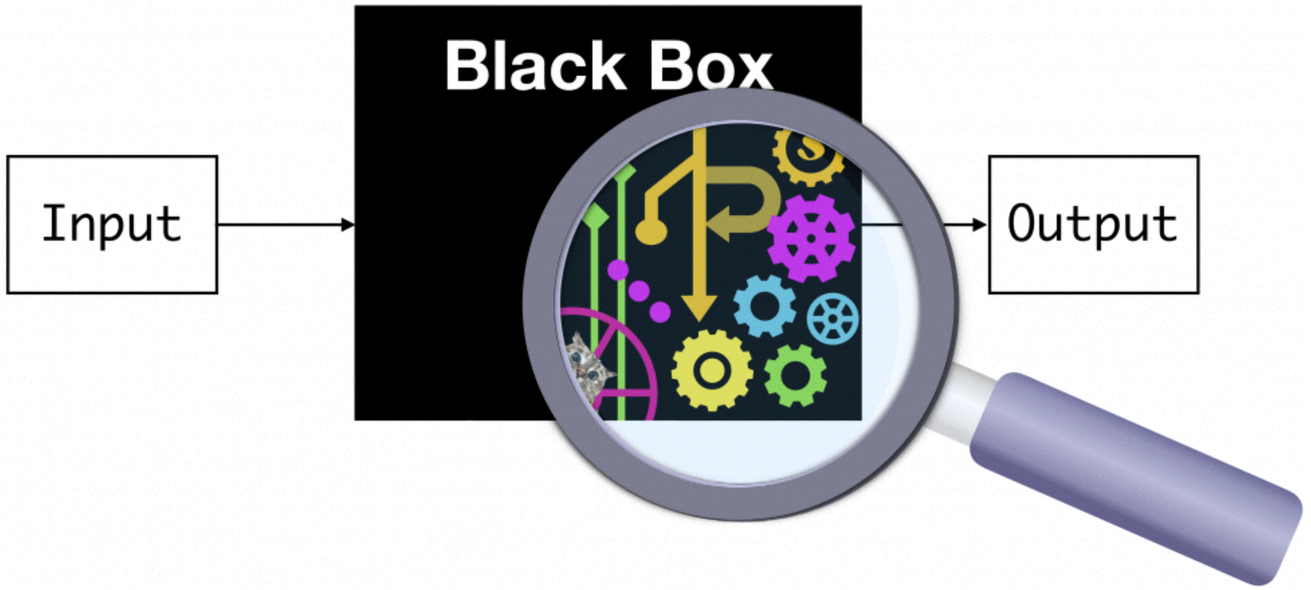


Scientists have evolved images that look like abstract patterns — but which DNNs see as familiar objects.



“Brittle, greedy, opaque, and shallow”

“People can rationalize what’s going on in their thought processes. Deep learning can’t: These systems have no idea how they’re thinking or how they’re categorizing themselves.” —Rodney Brooks



https://www.flagshippioneering.com/stories/brittle-greedy-opaque-and-shallow-the-promise-and-limits-of-todays-artificial-intelligence

“Brittle, greedy, opaque, and shallow”

“True intelligence is being able to approach a new problem you haven’t had a lot of direct experience with. A human being can play a game that they’ve never played before and in a matter of minutes figure out something about what’s going on. Machines still can’t do that.”

—Gary Marcus



??

Deep learning can be notoriously brittle

https://www.flagshippioneering.com/stories/brittle-greedy-opaque-and-shallow-the-promise-and-limits-of-todays-artificial-intelligence

Thank you!