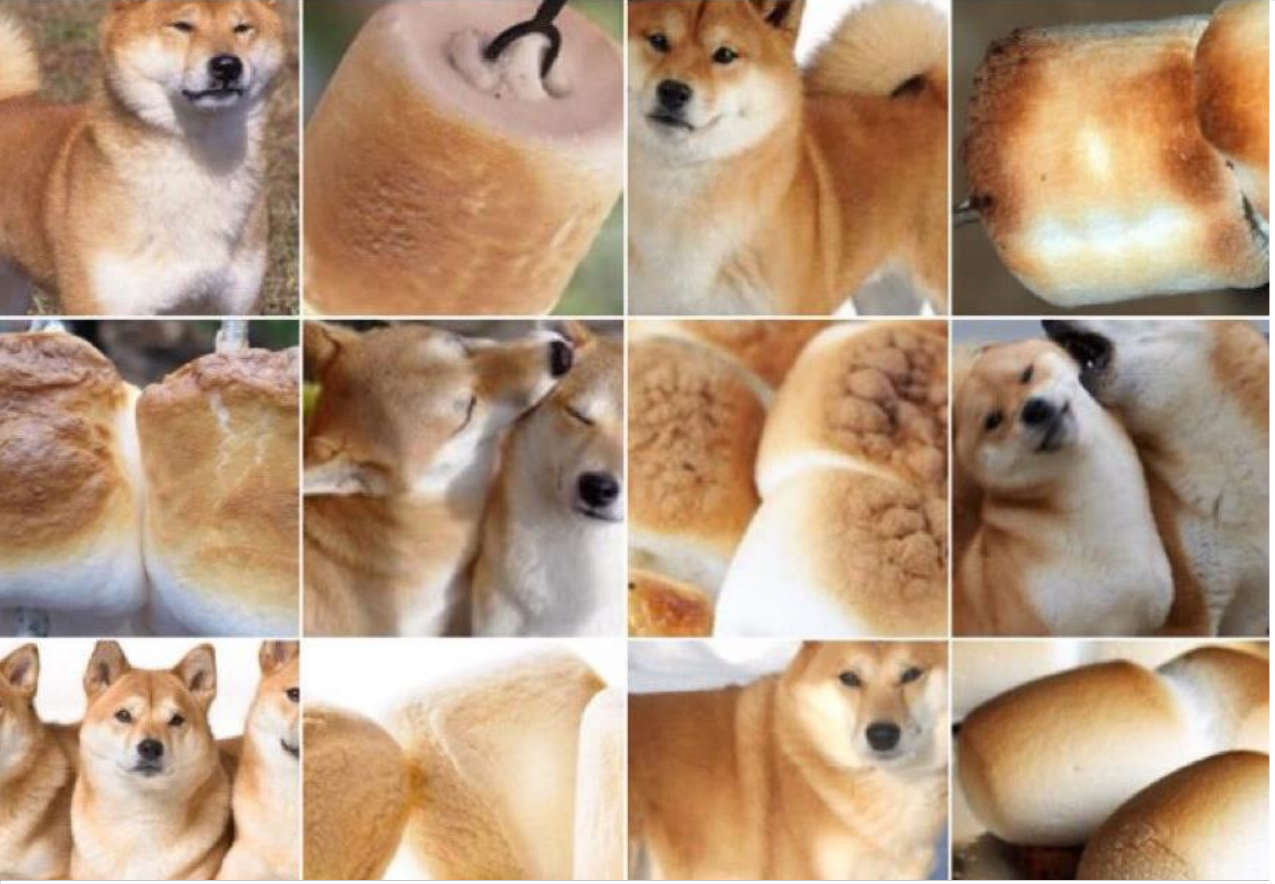
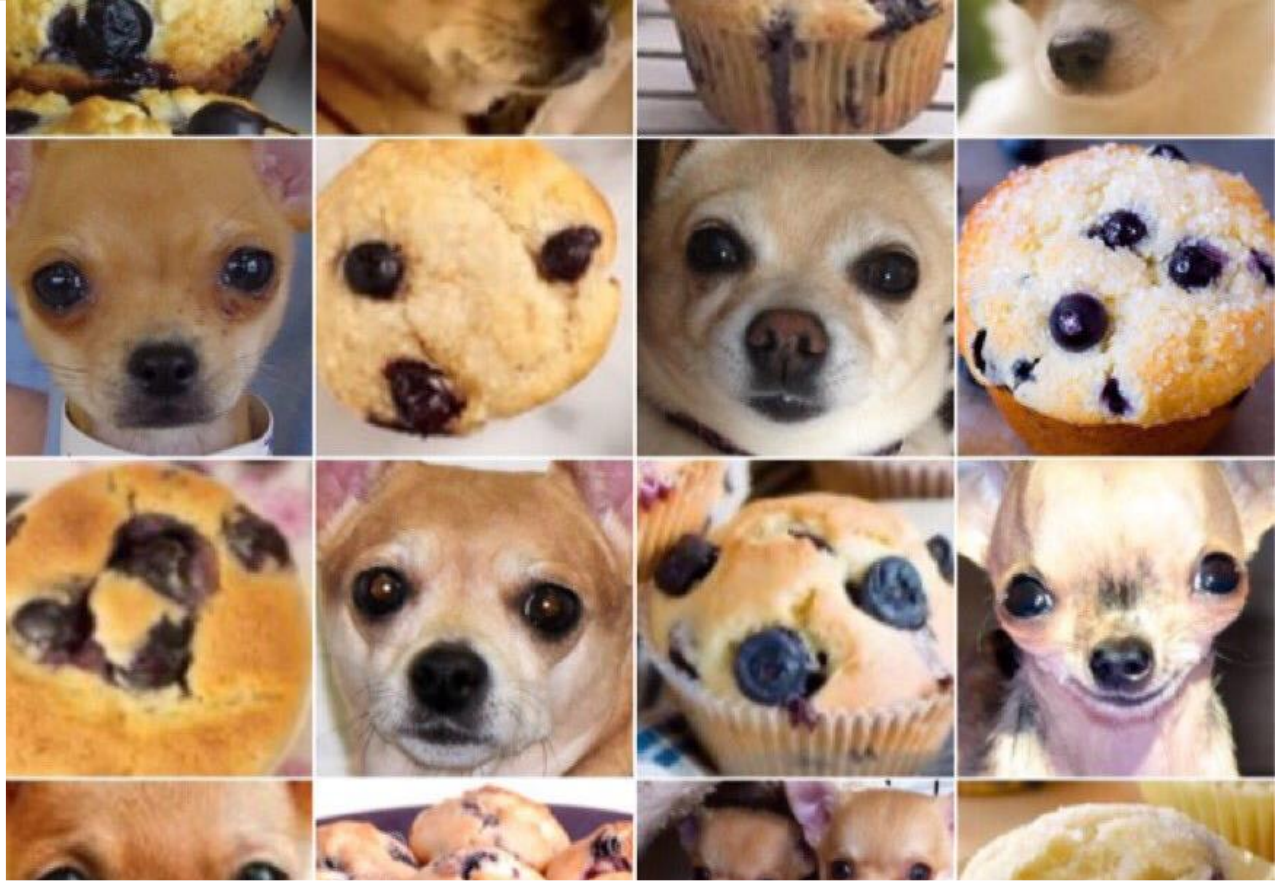




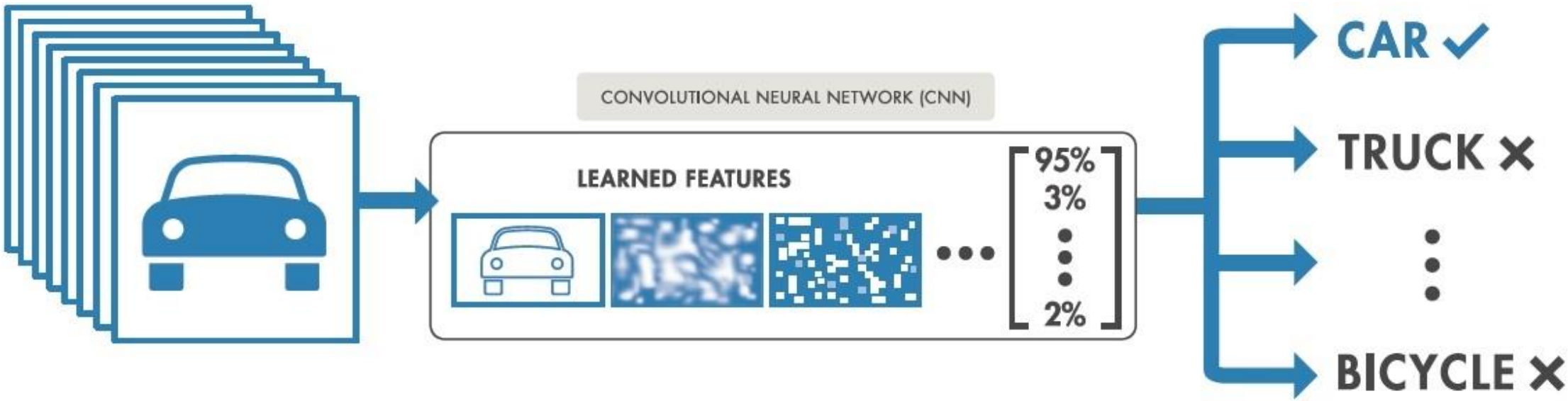
Convolutional Neural Networks

Increasing trust through visualization

Sam Green, samgreen@gmail.com

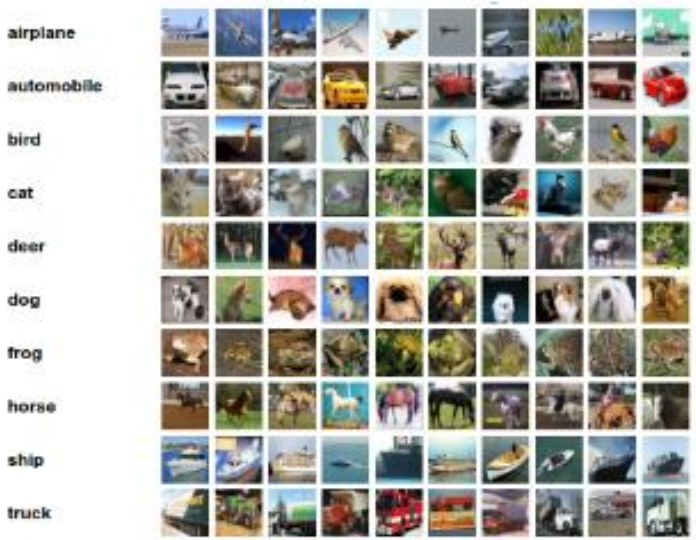
https://twitter.com/teenybiscuit

Convolutional Neural Network (CNN)



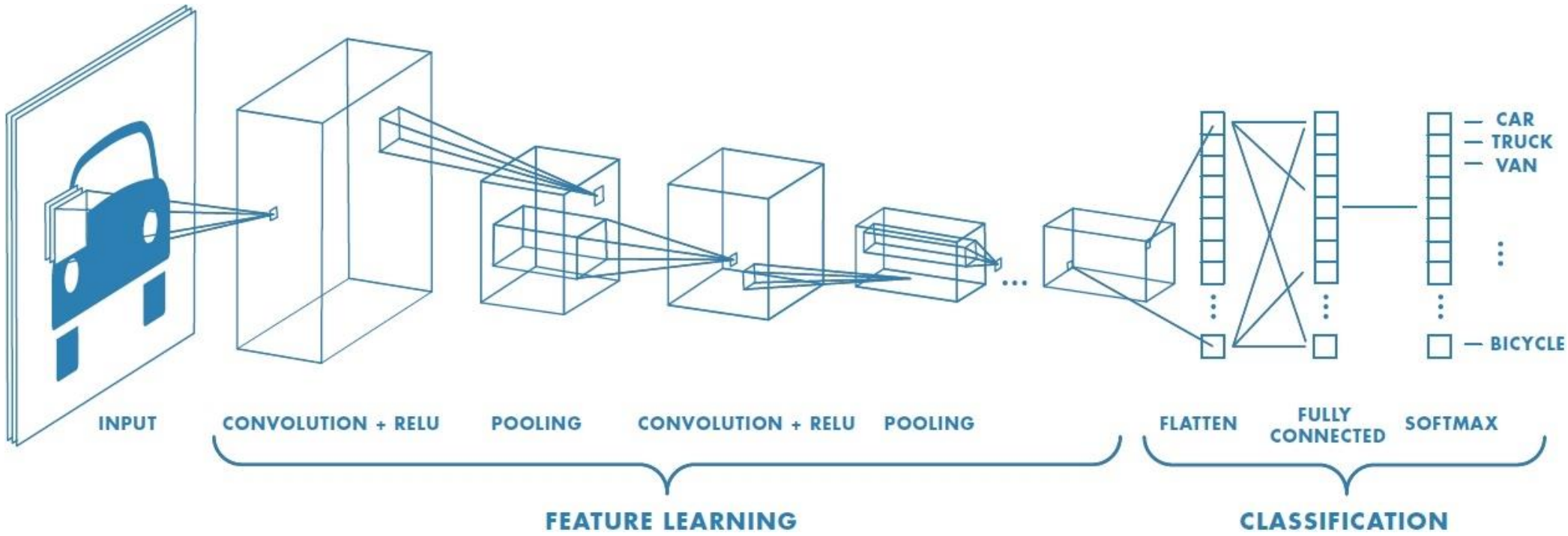
https://www.mathworks.com/discovery/convolutional-neural-network.html

Labeled training images



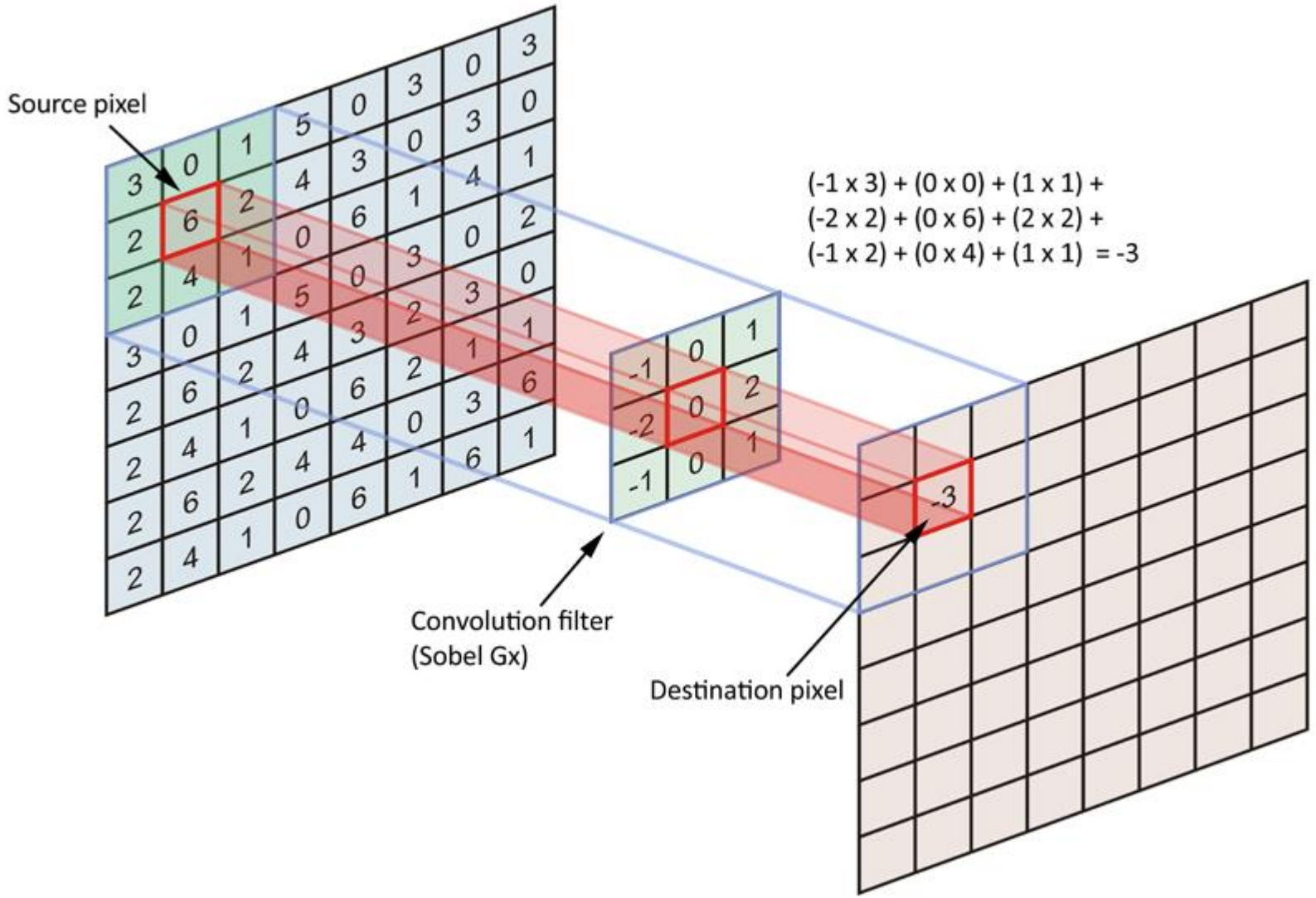
https://blog.rescale.com/wp-content/uploads/2016/05/classification.png

Convolutional Neural Network (CNN)



https://www.mathworks.com/discovery/convolutional-neural-network.html

Convolution



https://github.com/pjreddie/cnn-primer/blob/master/1/README.md

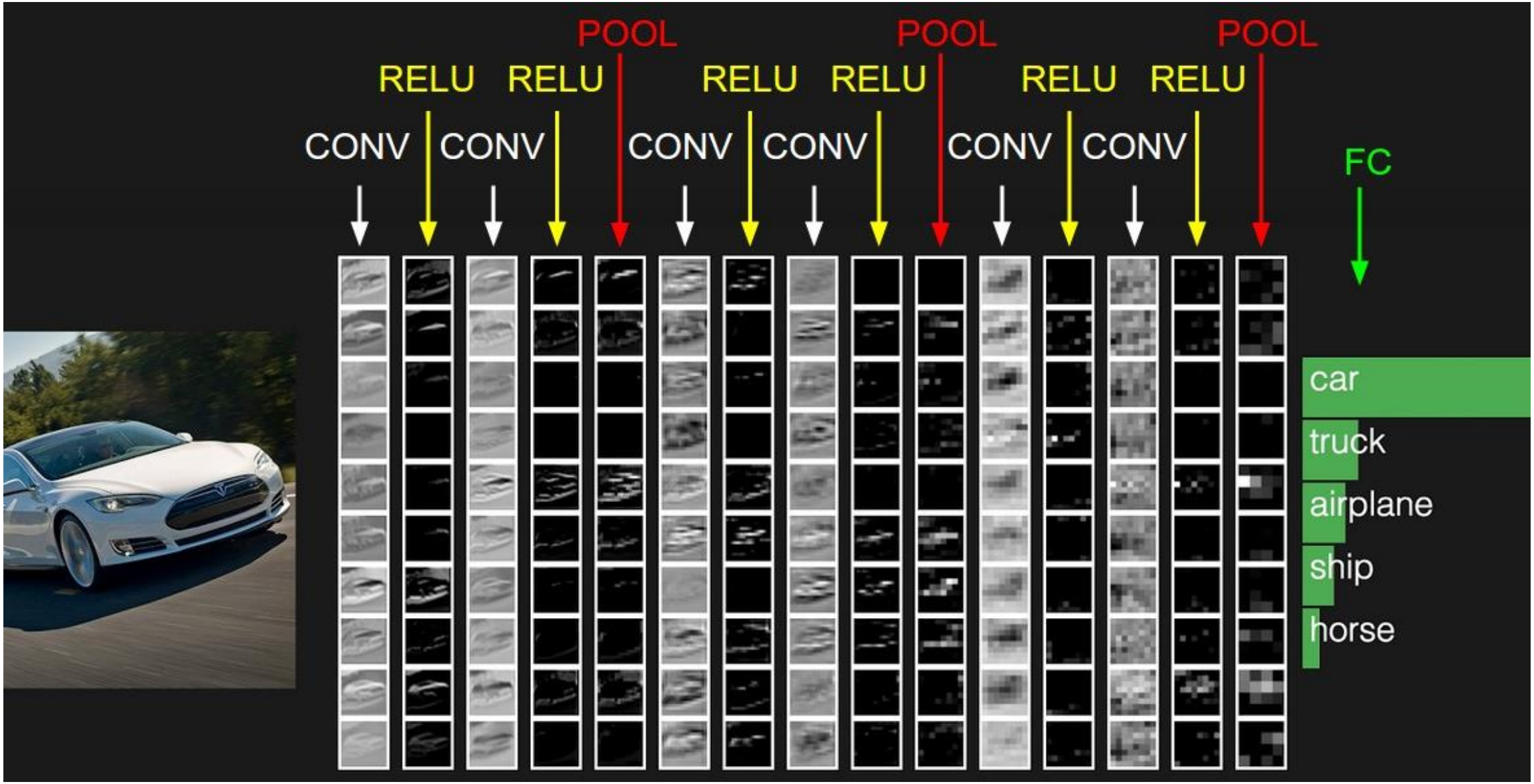
Convolution

$$\frac{1}{16}\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}$$

$$\begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix}$$

$$\begin{bmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{bmatrix}$$

https://en.wikipedia.org/wiki/Kernel_(image_processing)



http://cs231n.github.io/convolutional-networks/

Power of CNNs

Beating Go
(and chess, shogi, checkers, backgammon, Dota 2,…)

Breed recognition

Face recognition

Image caption generation

Prosthetics control

Colorizing black and white images

ATM deposits

Speech recognition

Weirdness

A close up of a lush green field
Tags: grass, field, sheep, standing, rainbow, man

http://aiweirdness.com/post/171451900302/do-neural-nets-dream-of-electric-sheep

Weirdness

A herd of sheep grazing on a lush green hillside
Tags: grazing, sheep, mountain, cattle, horse

http://aiweirdness.com/post/171451900302/do-neural-nets-dream-of-electric-sheep

Weirdness

Left: A man is holding a dog in his hand
Right: A woman is holding a dog in her hand
Image: @SouperSarah

http://aiweirdness.com/post/171451900302/do-neural-nets-dream-of-electric-sheep

Weirdness

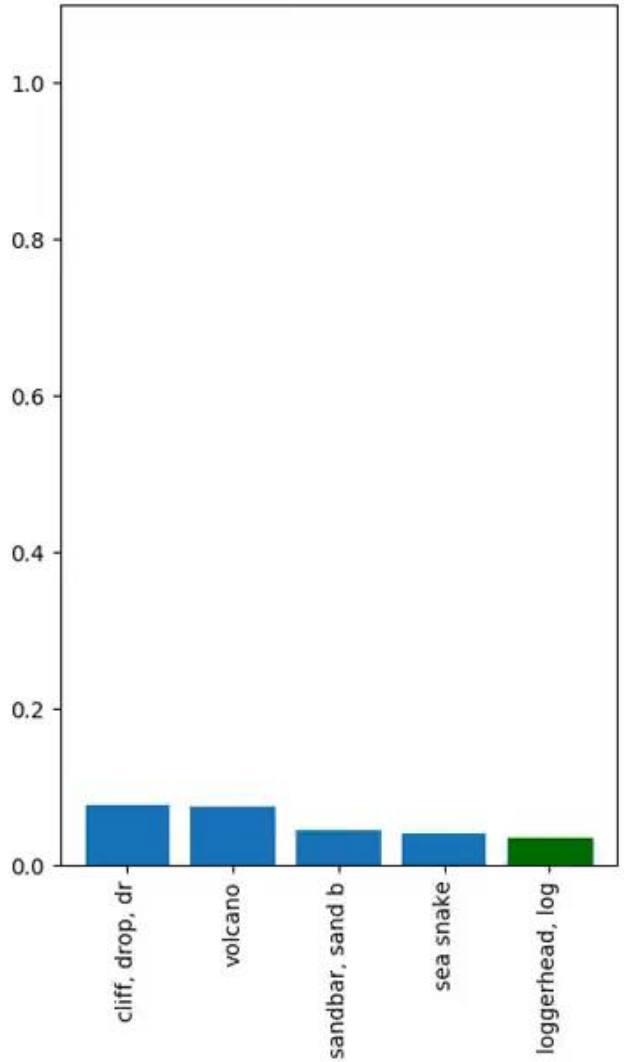
NeuralTalk2: A flock of birds flying in the air

Microsoft Azure: A group of giraffe standing next to a tree

Image: Fred Dunn, https://www.flickr.com/photos/gratapictures - CC-BY-NC

http://aiweirdness.com/post/171451900302/do-neural-nets-dream-of-electric-sheep

Attacks



https://www.youtube.com/watch?v=YXy6oX1iNoA

t-SNE

t-Distributed Stochastic Neighbor Embedding

t-SNE results with AlexNet

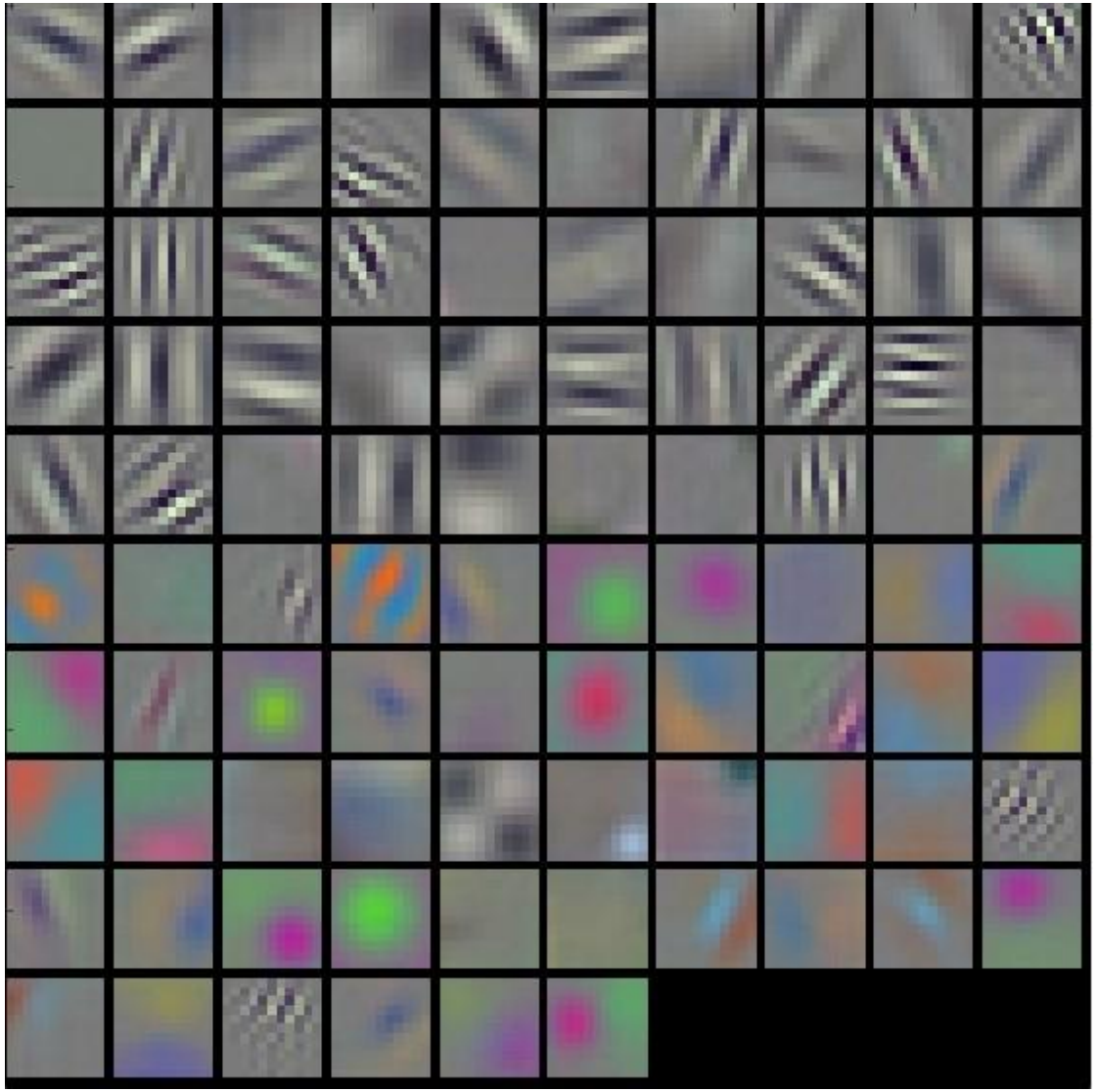
Maps 4096-dimensional CNN output
to 2-dimensions

https://cs.stanford.edu/people/karpathy/cnnembed/cnn_embed_full_1k.jpg

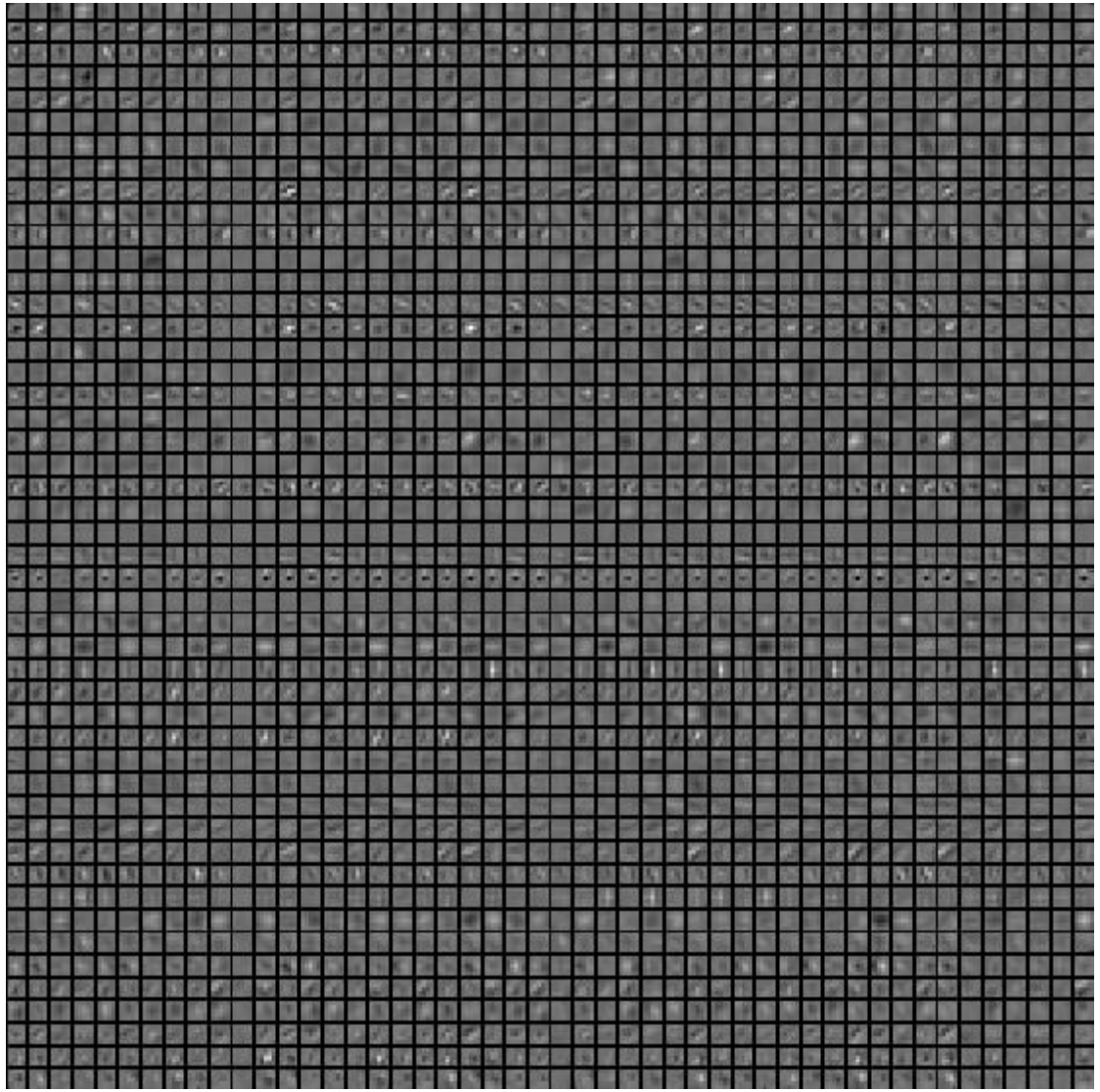
Filter Visualization

Inspecting what the CNN learned

AlexNet | conv1



AlexNet | conv2



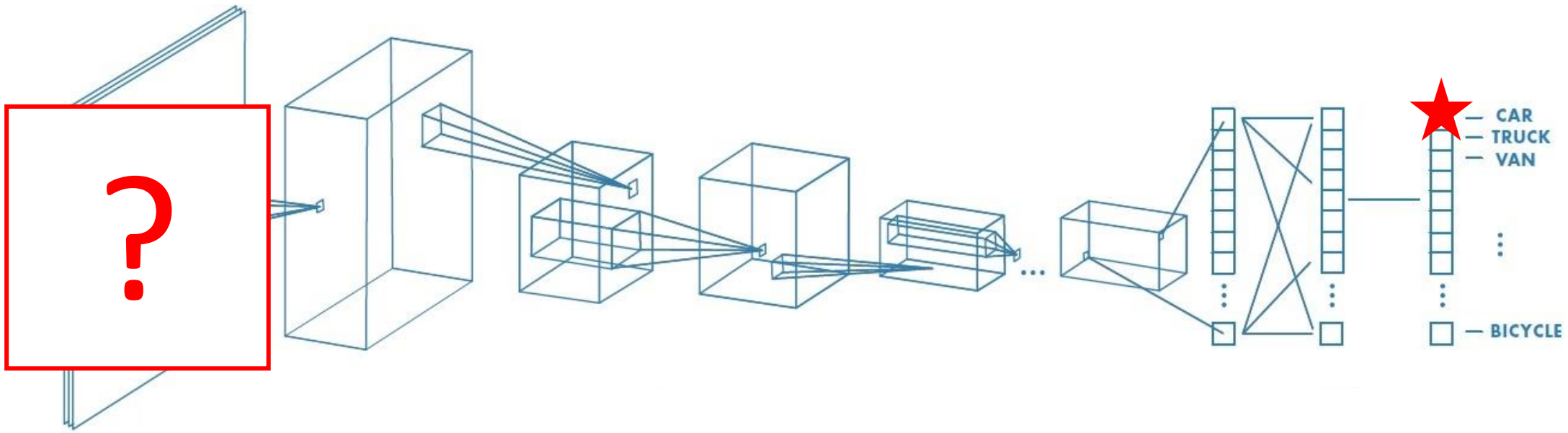
MNIST Filters

https://github.com/jireh-father/tensorflow-cnn-visualization

Class Visualization

Generating inputs to activate classes

What will trigger an output?



Visualizing GoogLeNet Classes

Loggerhead turtle

Saxophone

Screws

https://www.auduno.com/2015/07/29/visualizing-googlenet-classes/

Visualizing GoogLeNet Classes

Shetland sheepdog

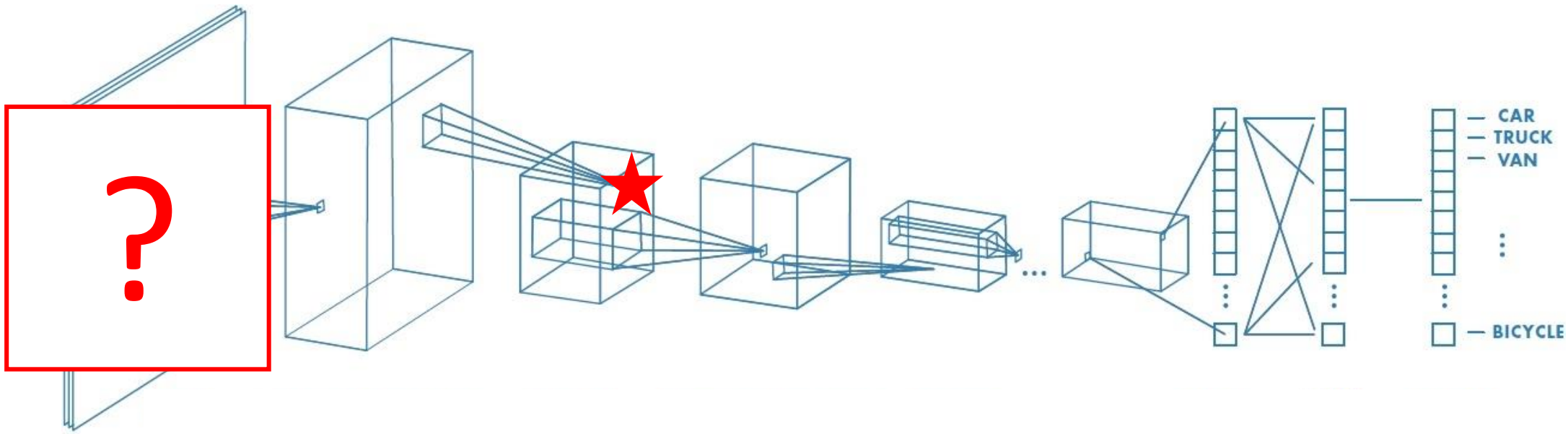
Basset hound

https://www.auduno.com/2015/07/29/visualizing-googlenet-classes/

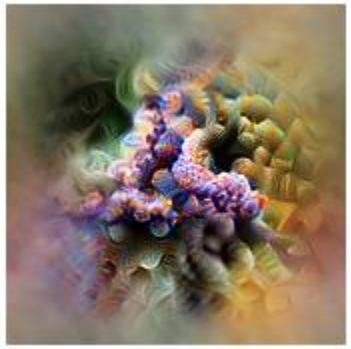
Feature Visualization

Generating inputs to activate neurons

What image triggers a neuron?



Images that activate neurons in GoogLeNet

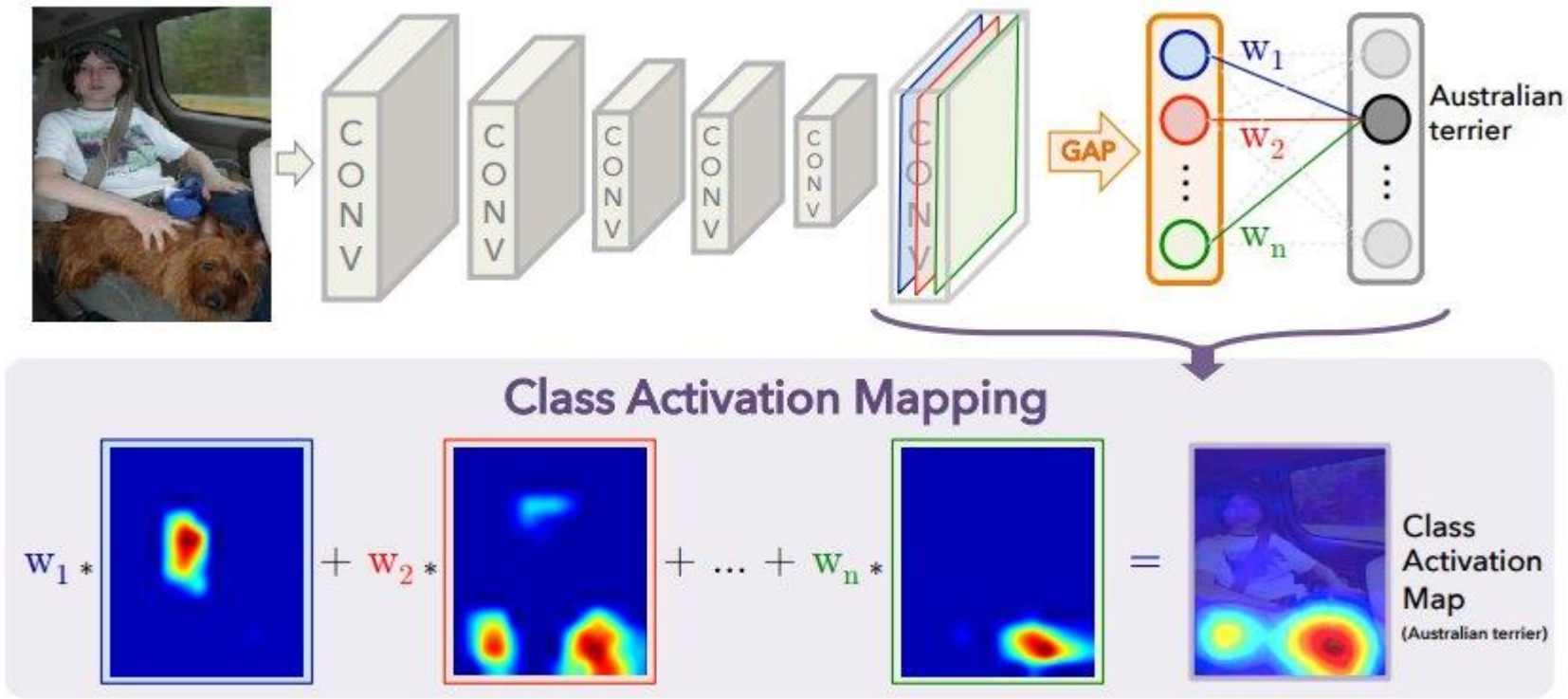
	Mixed4a, neuron 6	Mixed4a, neuron 240	Mixed4a, neuron 492
Triggering images in training set			
Feature visualization			

https://distill.pub/2017/feature-visualization/

Attribution Visualization

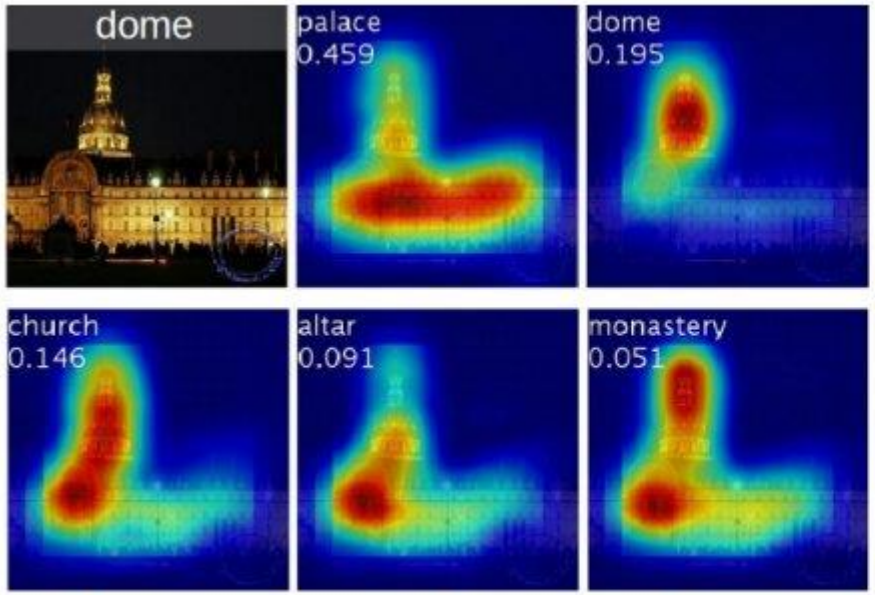
Understanding what triggers a class selection

Class Activation Mapping



https://github.com/metalbubble/CAM

Class Activation Mapping



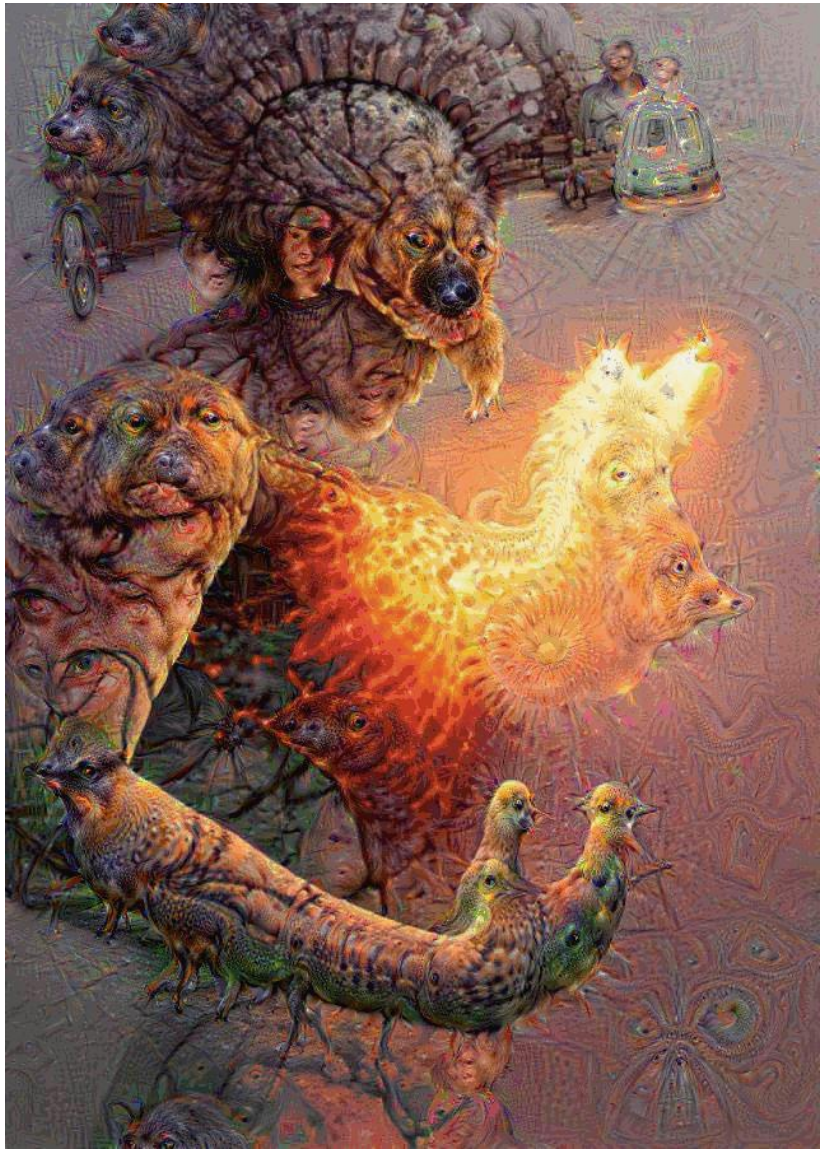
Class activation maps of top 5 predictions

Class activation maps for one object class

https://github.com/metalbubble/CAM

DeepDream

Modifying image to generate class activations



https://www.reddit.com/r/deepdreamgifs/comments/3ezaxe/heart_of_a_warrior_xpost_woahdude/

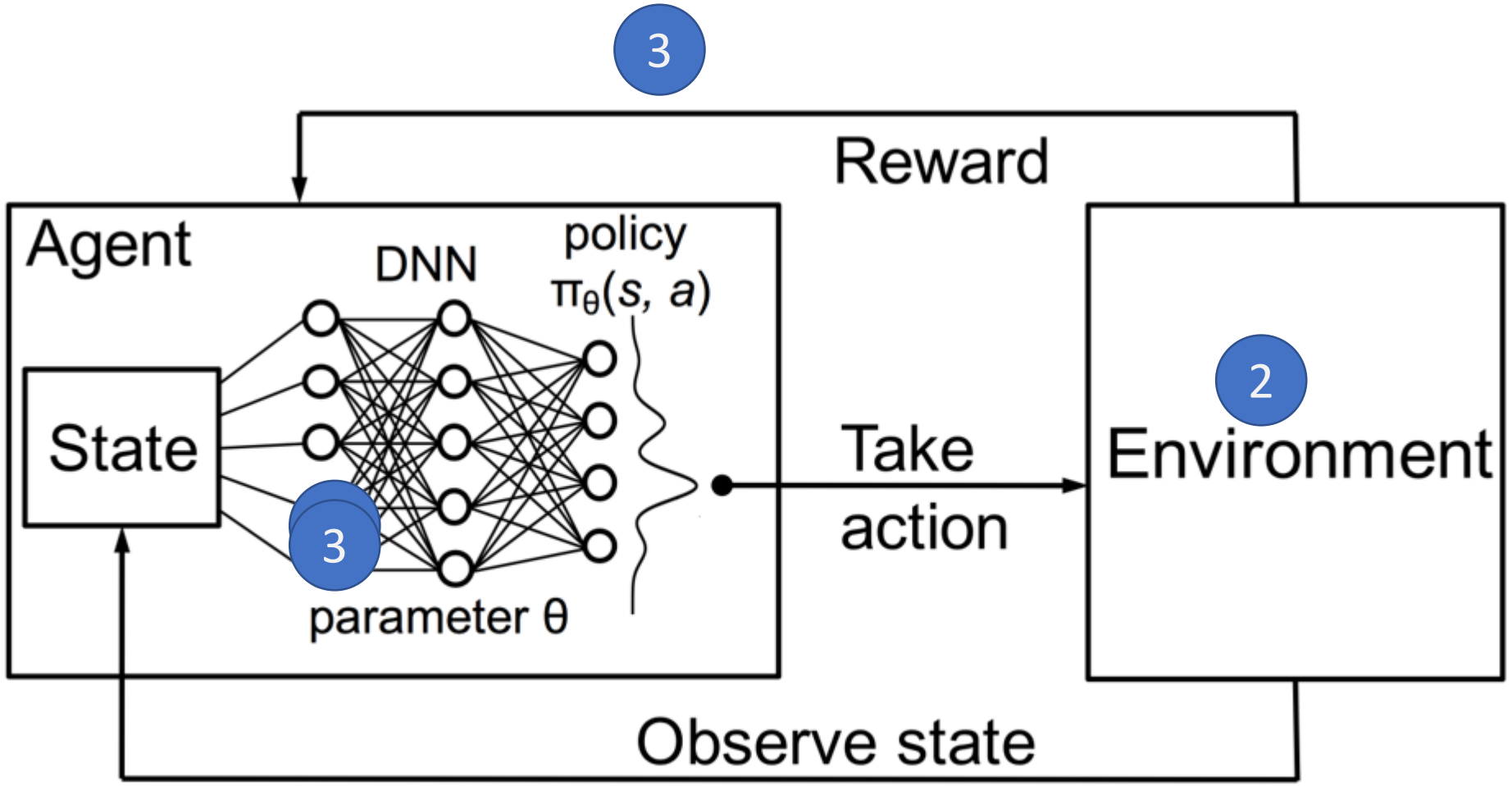
DeepDream

https://research.googleblog.com/2015/06/inceptionism-going-deeper-into-neural.html

Applications to Reinforcement Learning

Visualizing the policy

Reinforcement Learning Paradigm

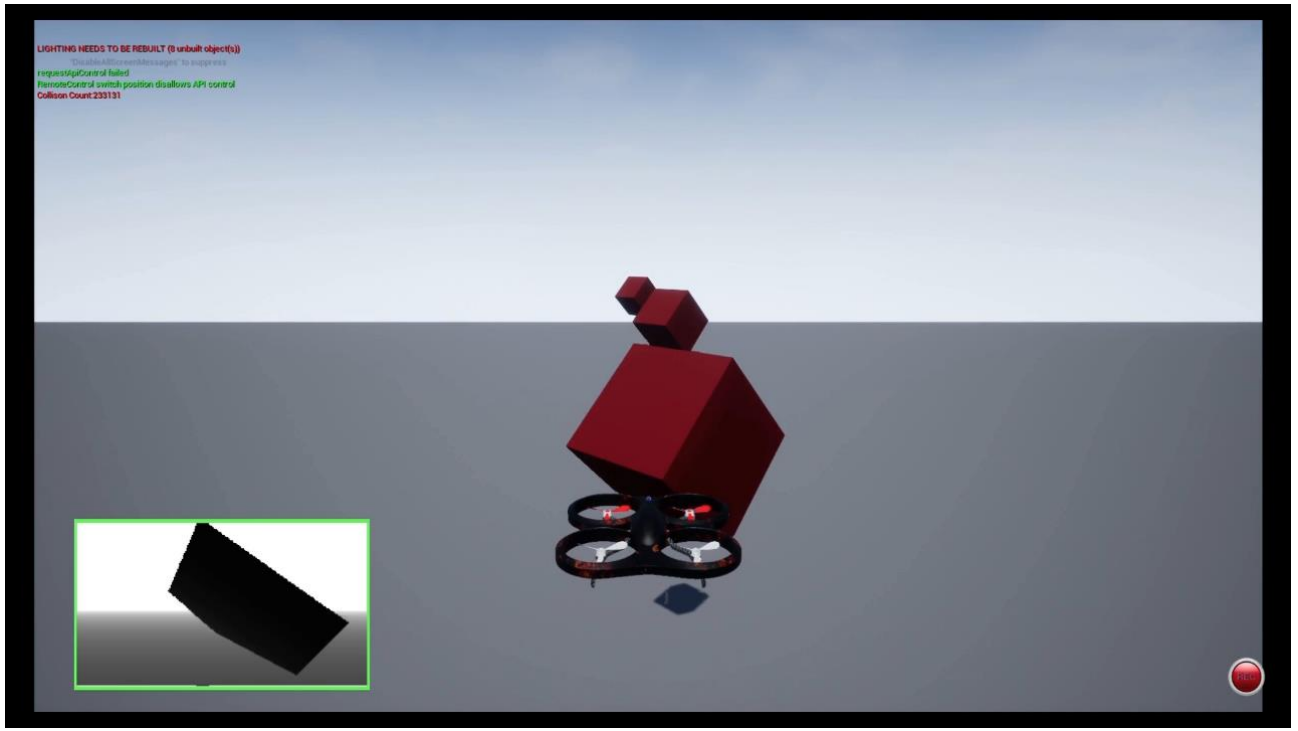


Reinforcement Learning Cube Example

State observation is camera
CNN for policy π_θ
Actions: left, forward, right
Reward +1 for hitting box

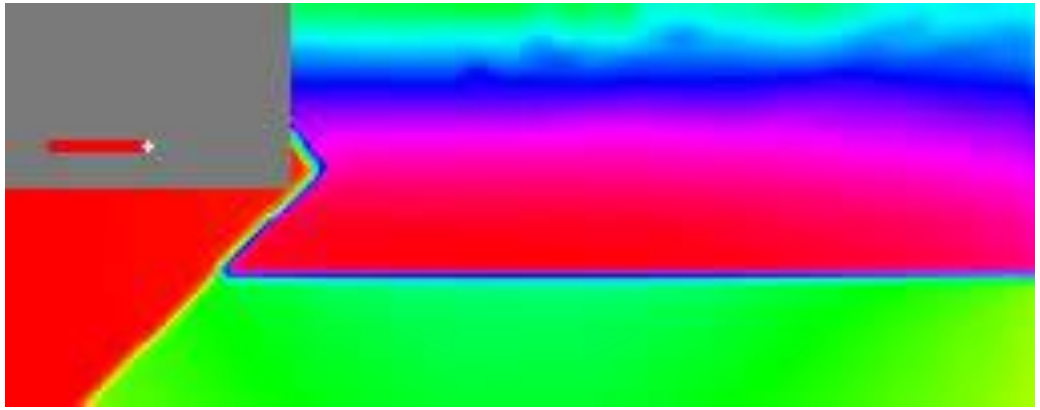
Policy is stochastic:

$$\pi_\theta\left(\quad\right) = \begin{bmatrix} .4 \\ .5 \\ .1 \end{bmatrix}$$



Reinforcement Learning Cube Example

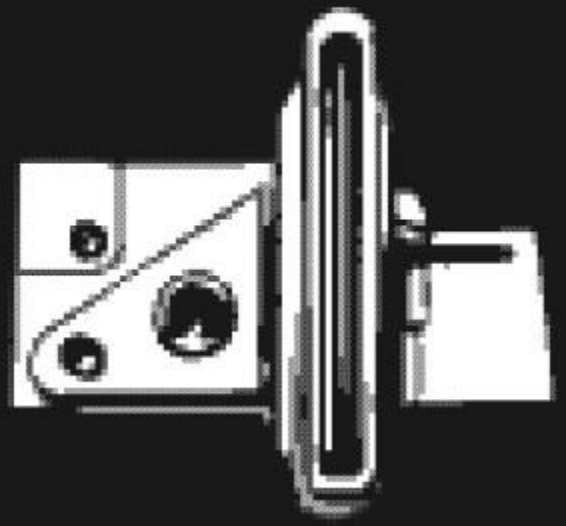
- Using attribution visualization to understand decision making



Credit: Jieliang Luo, Sam Green @ koclab.org

Learning resources

- PyTorch
  http://pytorch.org/
- Linear algebra for deep learning
  https://goo.gl/RJiNQJ
- Calculus for deep learning
  https://goo.gl/zyQp7k



THE INTERROGATOR ASKS YOU:

DO YOU DREAM ABOUT BEING
A UNICORN?

WHAT DO YOU SAY?

> INTERLINKED.
  CELLS.