

Computer Vision Workshop

Kevin Forest

DATA
LAB

Tasks in computer vision

- Image classification
- Object localization
- Object segmentation
- Image generation

Image classification

- Classify a whole image in a category
 - Binary classification (Hot Dog vs Not Hot Dog)
 - Multi-class classification (ImageNet with 1 000 classes)



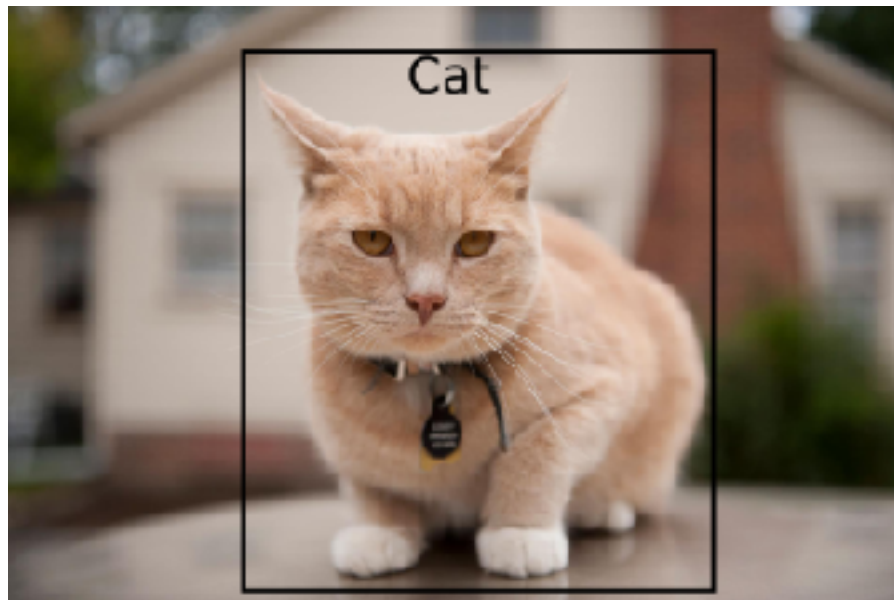
Image classification

- Classes are defined prior to training
 - Let's say we trained a multi-class classifier with three classes
 - Cat
 - Dog
 - Rabbit
 - We present an image of a car to this model
 - The model will classify the car either as a cat, a dog or a rabbit
- Popular neural network architecture include
 - VGG
 - Inception
 - ResNet
 - ResNet Inception
 - Xception



Object localization

- Localize an object in an image
- Will only localize specific classes of object
- It's basically like putting bounding boxes around objects
- Bounding boxes can overlap
- Popular neural network architecture include
 - YOLO (You Only Look Once)
 - R-CNN
 - SSD
 - RetinaNet



Object segmentation

- Similar to image localization
- Localize an object in an image
- Instead of putting bounding boxes, does a pixel level segmentation
- Popular network architecture include
 - Masknet
 - Mask RCNN



Image generation

- Given an input, generate a new image or enhance an image
- Image generation from scene graphs - sg2sim

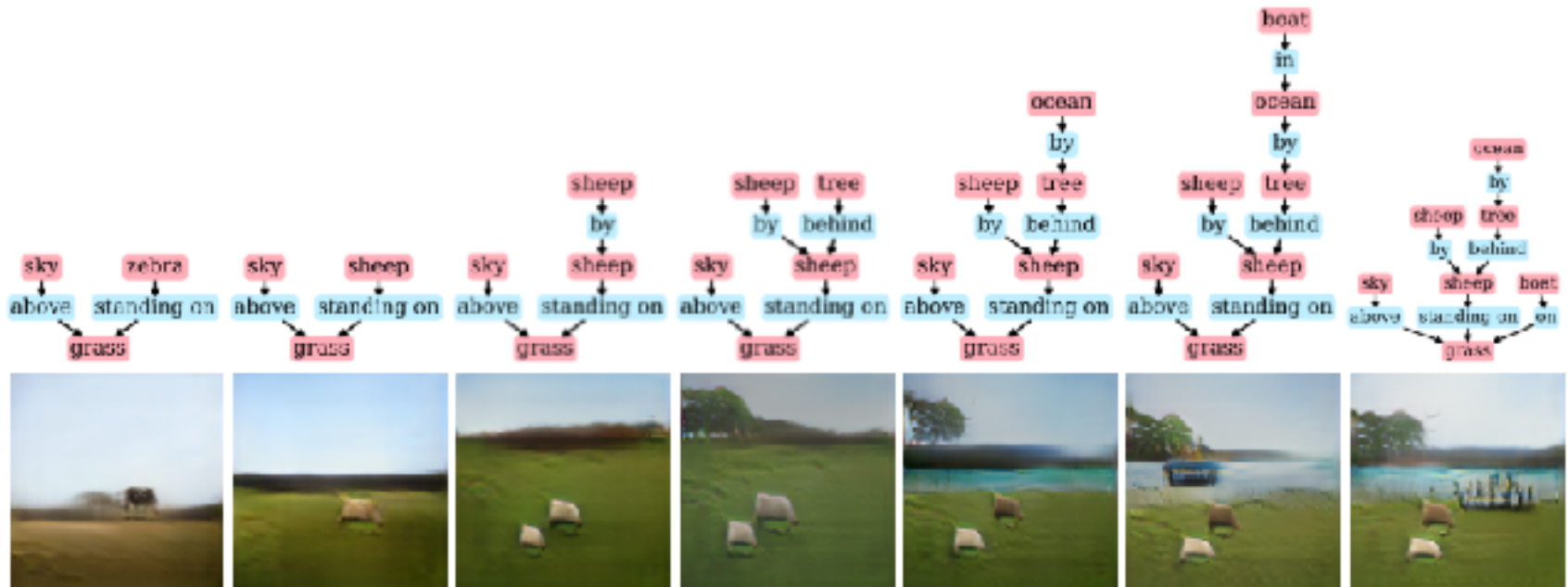


Image generation

- Style transfert
- Take two image as input and create a new image
- Both style transfert and scene generation leverage Generative Adversarial Network (GAN)



Image classification

- How does human recognize a car?
 - Rectangular-box shape
 - 4 wheels
 - Pair of headlights
 - Etc...
- How do human differentiate between car make/model?
 - Square-shaped vs round-shaped
 - Number of doors
 - **It's hard to precisely identify rules**

Convolutional Neural Network (CNN)

- What is a kernel conceptually?
 - A small matrix
 - For each pixel of the image, we apply a convolution between the image and the kernel
- A convolution is similar to a matrix multiplication
- Can be use for multiple purpose
 - Edge detection
 - Blur

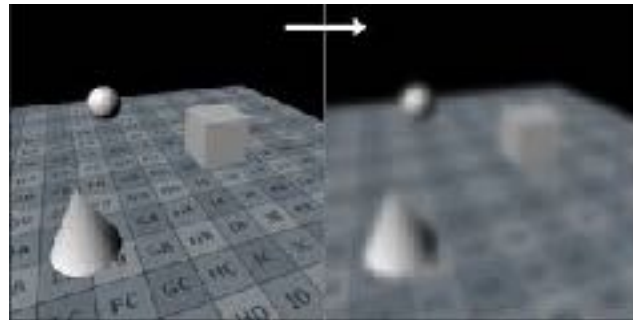
Edge detection

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



Blur

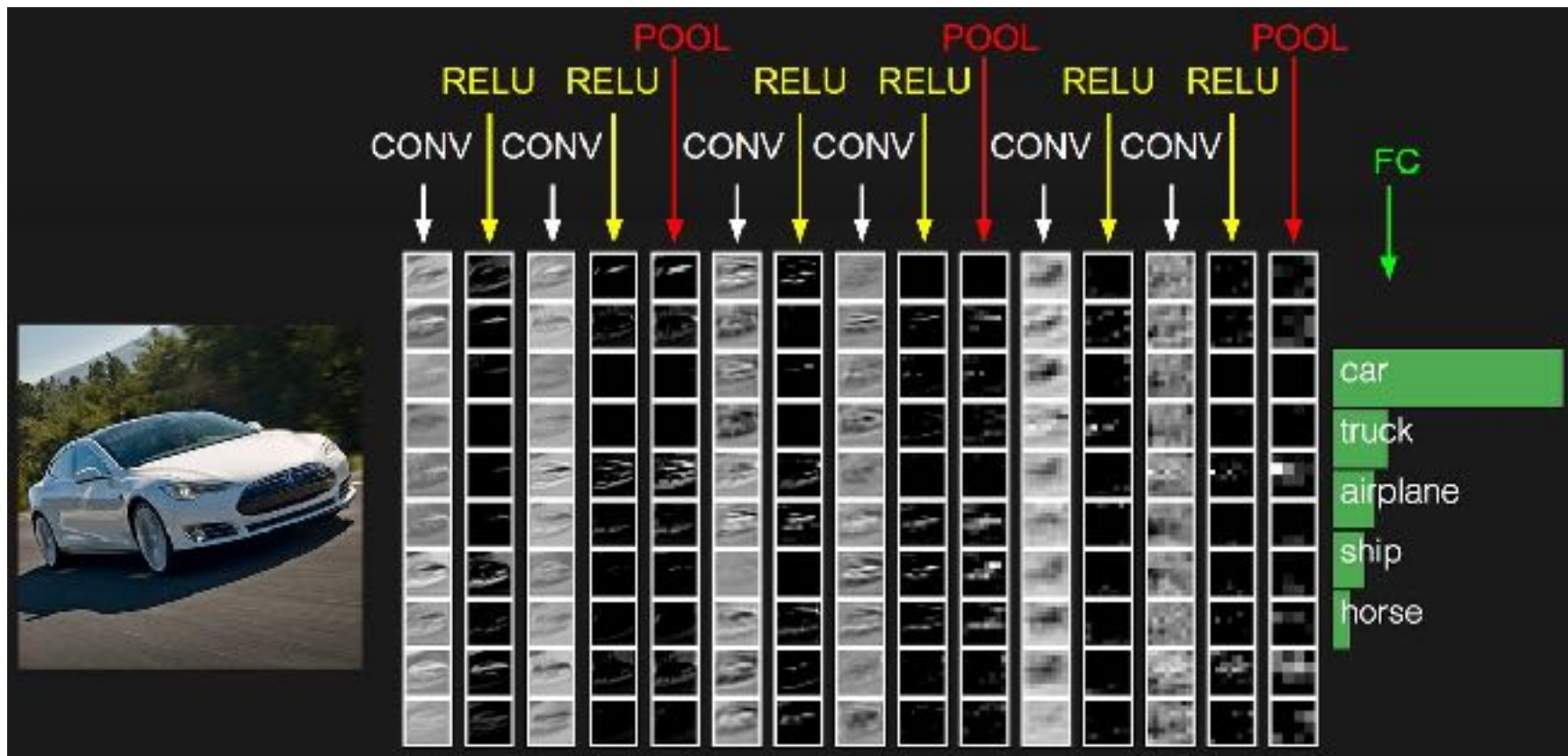
| | | |
|-------|-------|-------|
| $1/9$ | $1/9$ | $1/9$ |
| $1/9$ | $1/9$ | $1/9$ |
| $1/9$ | $1/9$ | $1/9$ |



Learning the kernel

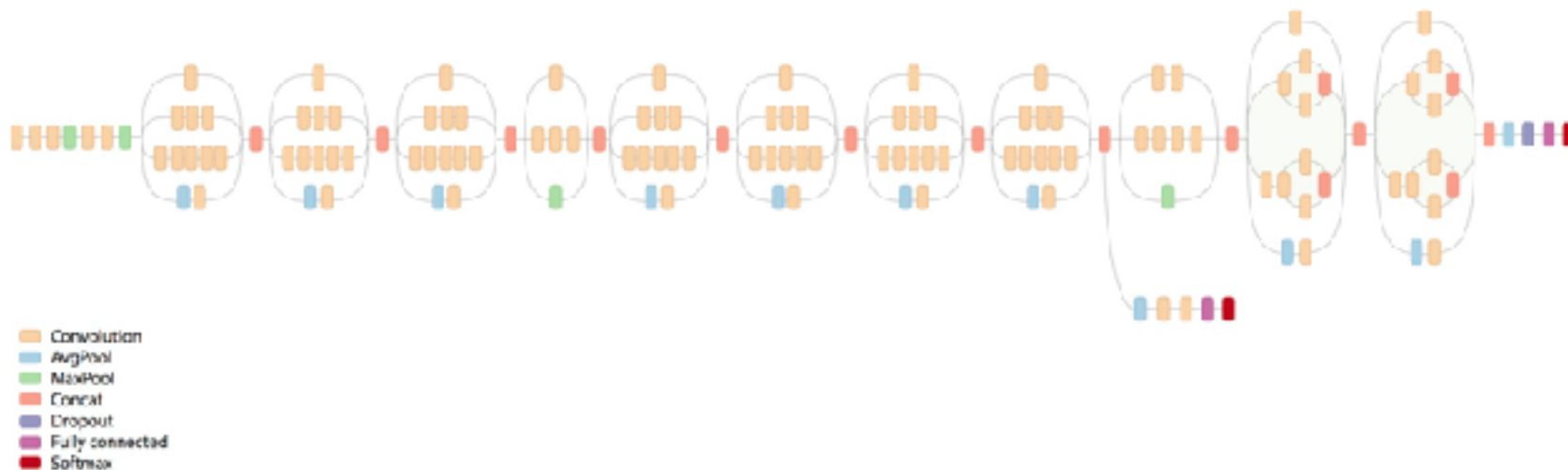
- Convolutional Neural Network allows learning the kernel
- The idea behind CNN is to stack convolution in a deep network
- The intuition is that by stacking convolution, the network will learn feature relevant to the task
 - Edges
 - Shapes
 - Shape patterns
 - Color patterns
- The deeper, the more complex and specific to the task the features are

CNN



Inception v3

- Proposed by Google in 2015
- When introduced, was the state of the art
- Trained on ImageNet2012
 - A big dataset of images with labels
- Original paper - <https://arxiv.org/abs/1512.00567>





Questions