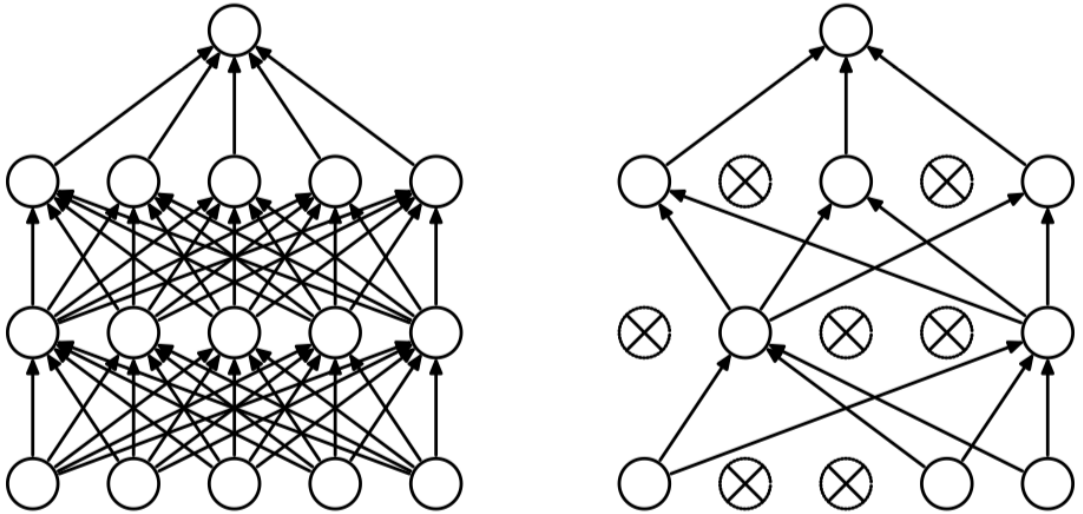
IN3063/INM702 - Programming and Mathematics for Artificial Intelligence

10.A - CNNs

michael.garcia-ortiz@city.ac.uk

Train / eval mode

- A Neural Network is by default in train mode
- When evaluating, the behavior of some components differ
- For example: Dropout



(a) Standard Neural Net

(b) After applying dropout.

Neural Networks should subclass nn.Module

- All weights of the module are registered as parameters
- Automatically sets them as differentiable
- Allows to create the computational graph

CLASS `torch.nn.Module`

Base class for all neural network modules.

Your models should also subclass this class.

Modules can also contain other Modules, allowing to nest them in a tree structure. You can assign the sub
regular attributes:

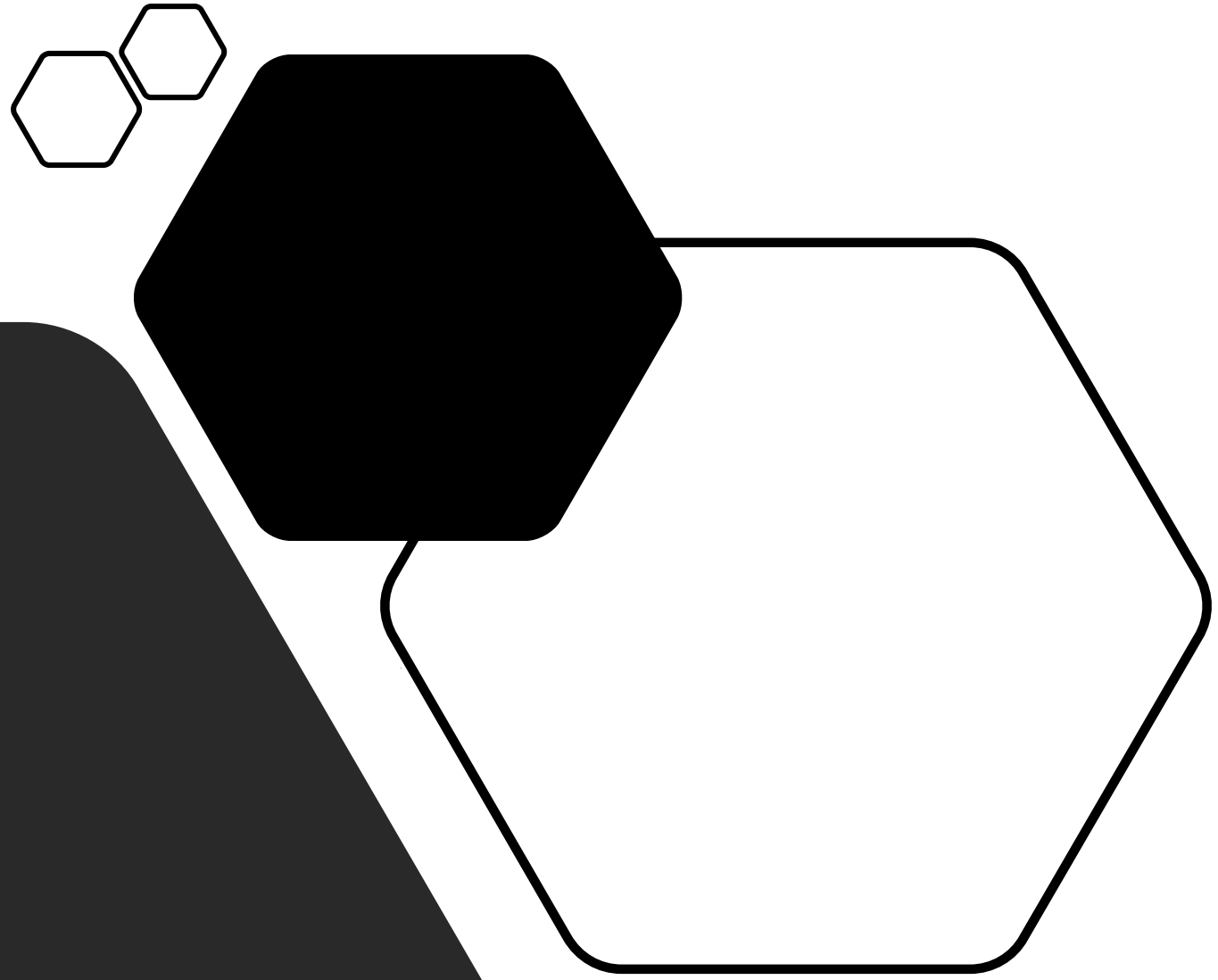
```
import torch.nn as nn
import torch.nn.functional as F

class Model(nn.Module):
    def __init__(self):
        super(Model, self).__init__()
        self.conv1 = nn.Conv2d(1, 20, 5)
        self.conv2 = nn.Conv2d(20, 20, 5)

    def forward(self, x):
        x = F.relu(self.conv1(x))
        return F.relu(self.conv2(x))
```

Loss and Loss with logits

- Some loss functions already implement the activation of the last layer

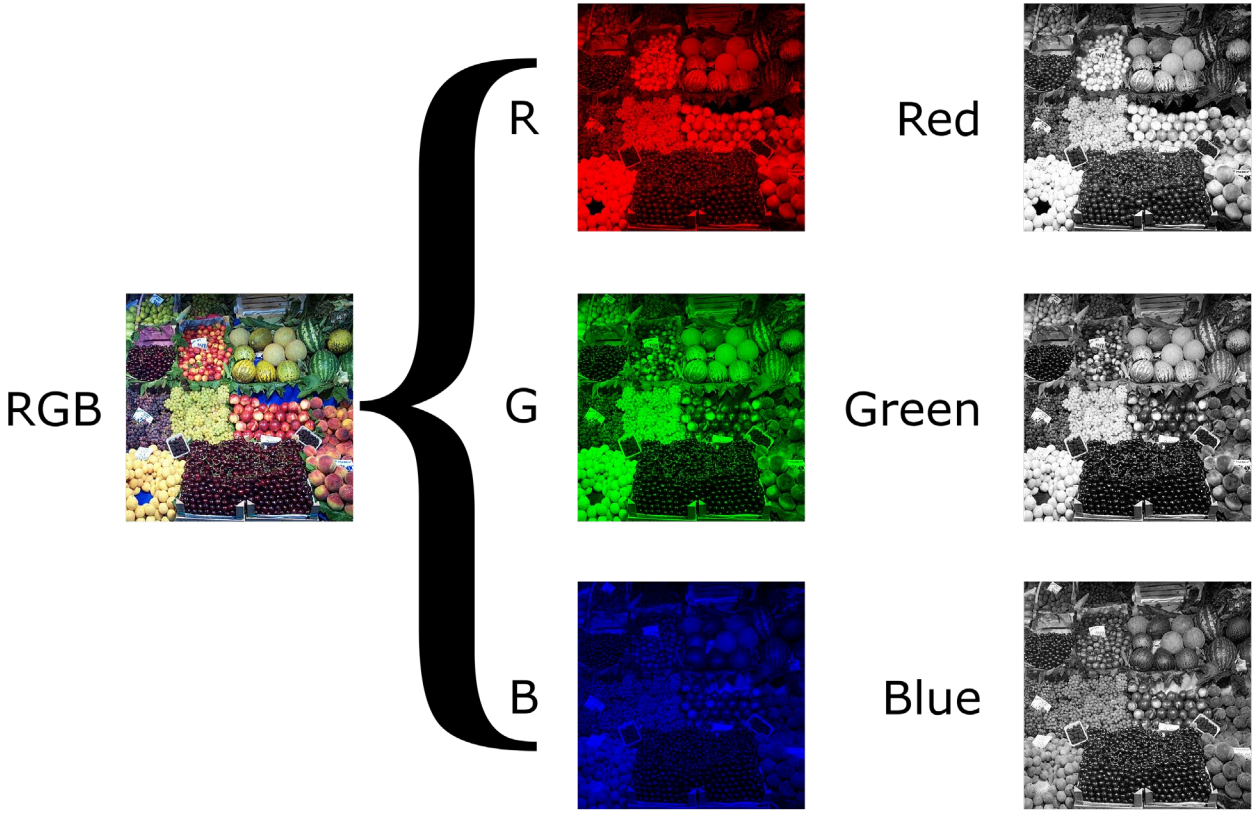
`nn.CrossEntropyLoss`	This criterion combines `nn.LogSoftmax()` and `nn.NLLLoss()` in one single class.
`nn.NLLLoss`	The negative log likelihood loss.

Learning on Images



Properties of Images

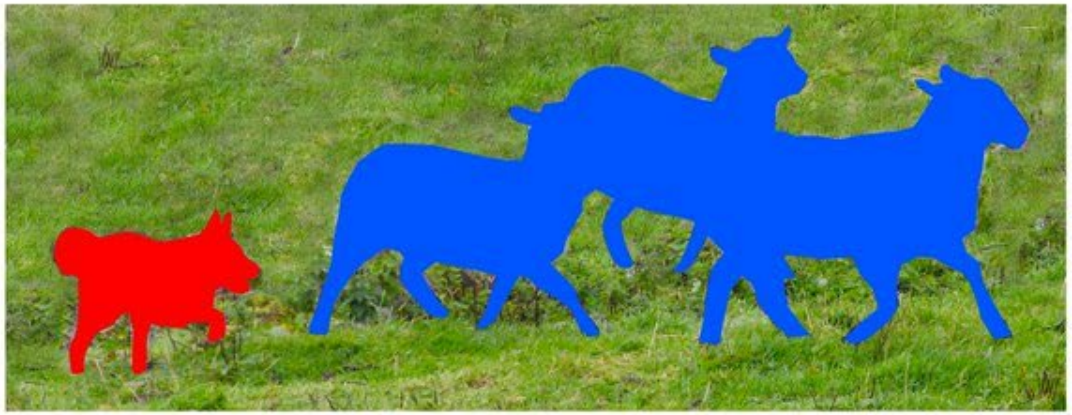
- 2-Dimentional structure (3 if you count the channel)



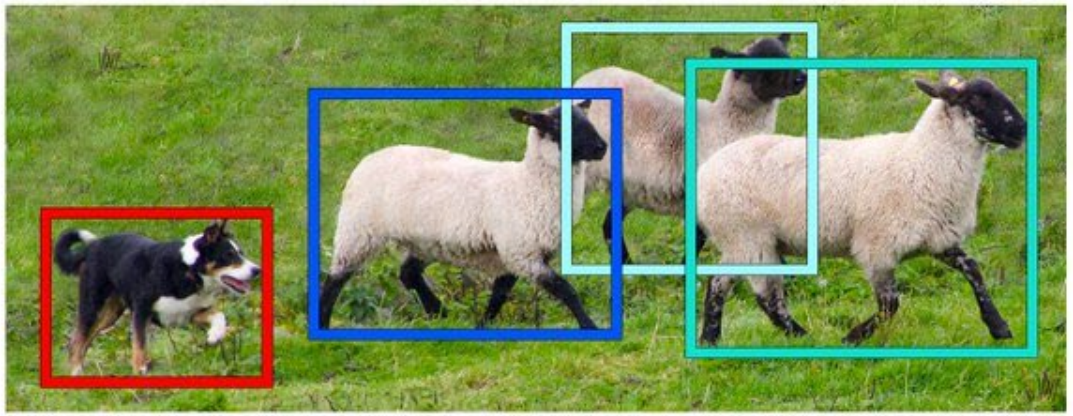
Local features

Different Deep Learning Problems

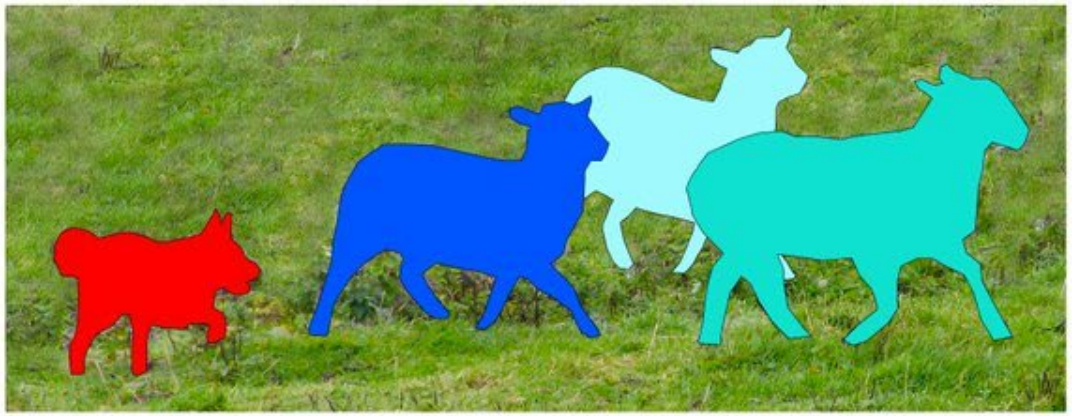
Image Recognition



Semantic Segmentation



Object Detection



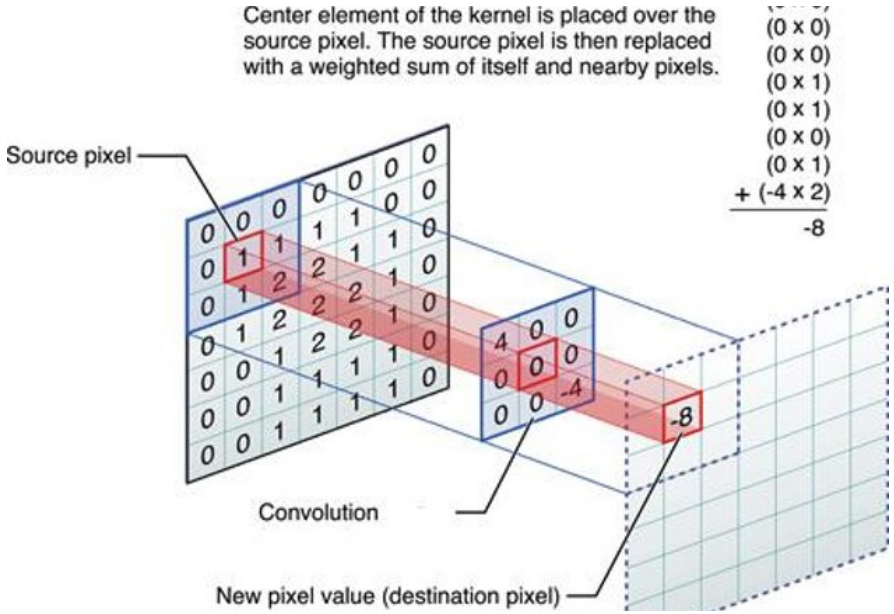
Instance Segmentation

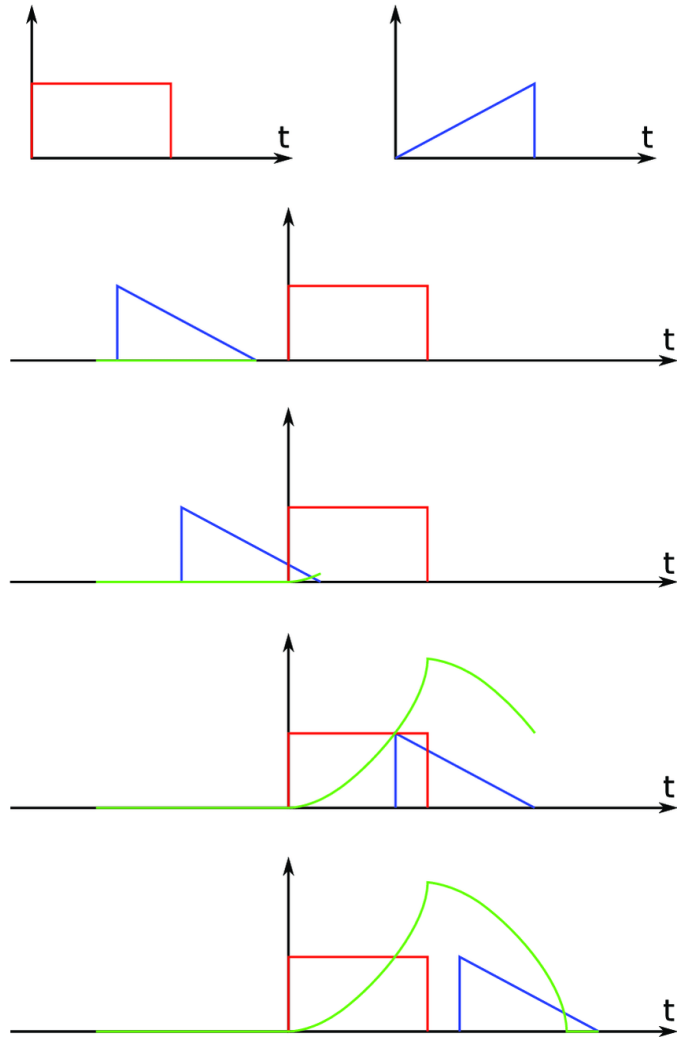
How to solve this broad range of problem?

- Classifying pixel by pixel?
- Classifying the whole image?
- Local features can occur at different locations of the image

Convolution

- Signal processing method
- In the case of images: 3D kernels

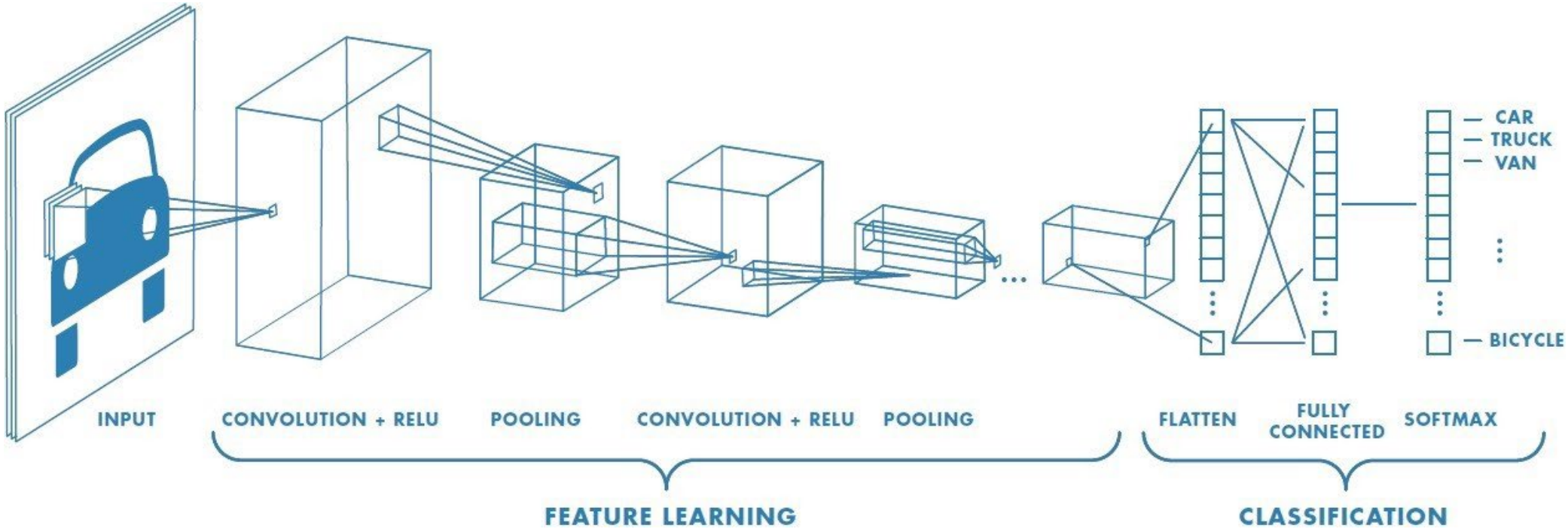




Convolutional Neural Networks

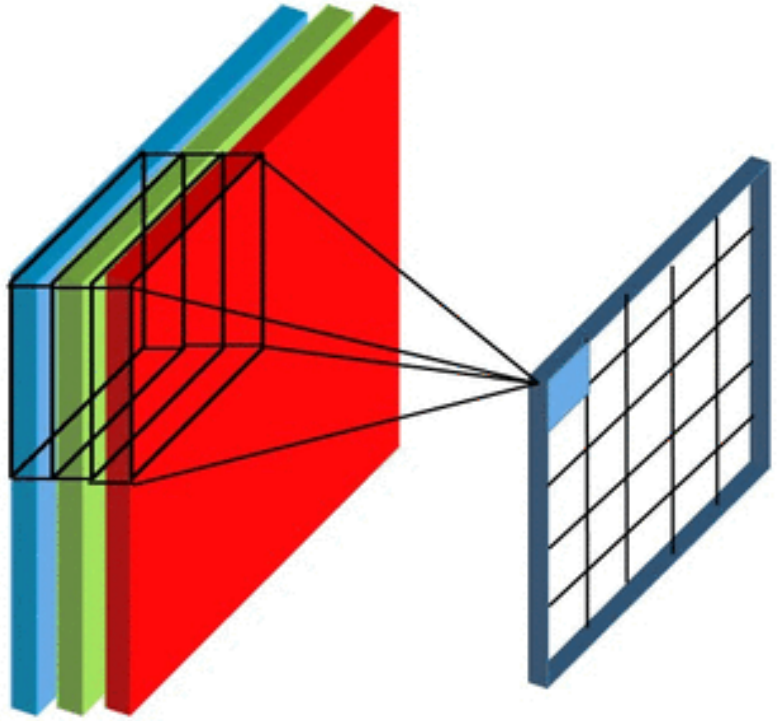
Convolutional Neural Networks

- The kernel is a list of parameter.
- Local feature learning



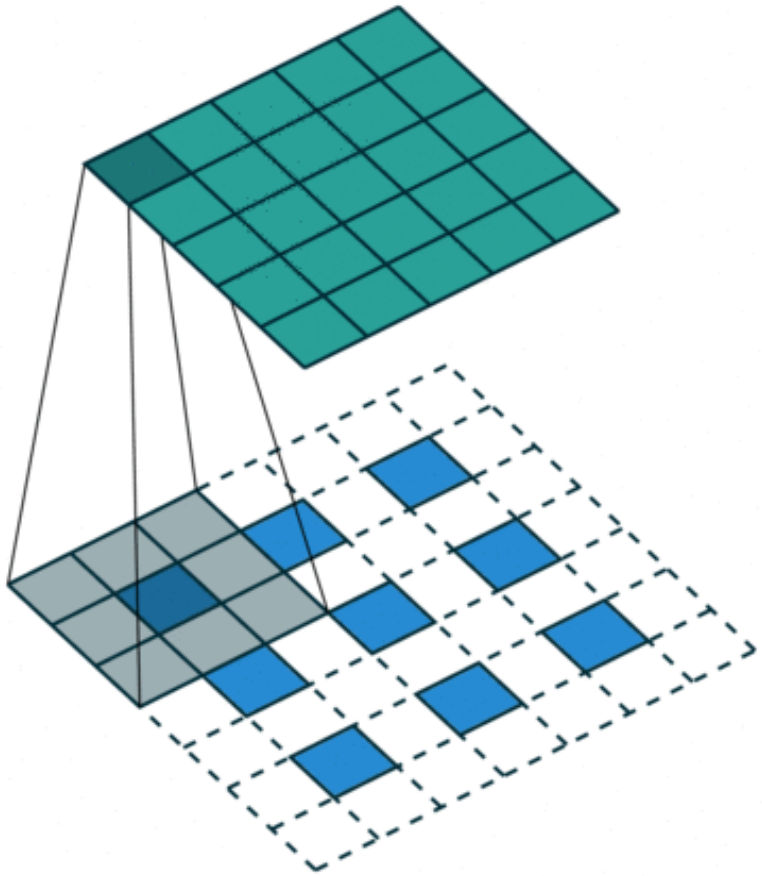
Parameters of a convolutional layer

- Kernel size



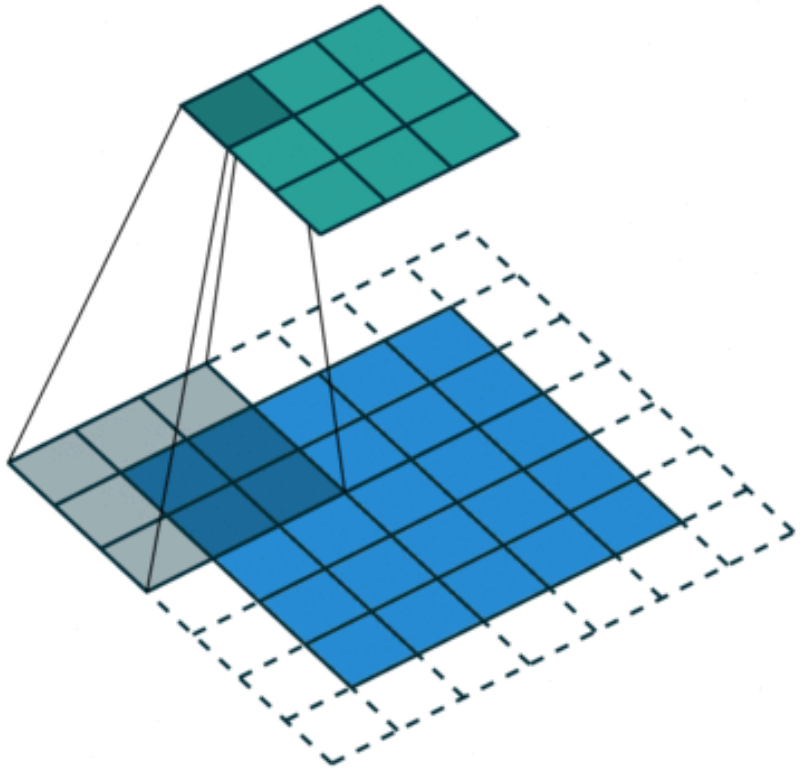
Parameters of a convolutional layer

- Stride



Parameters of a convolutional layer

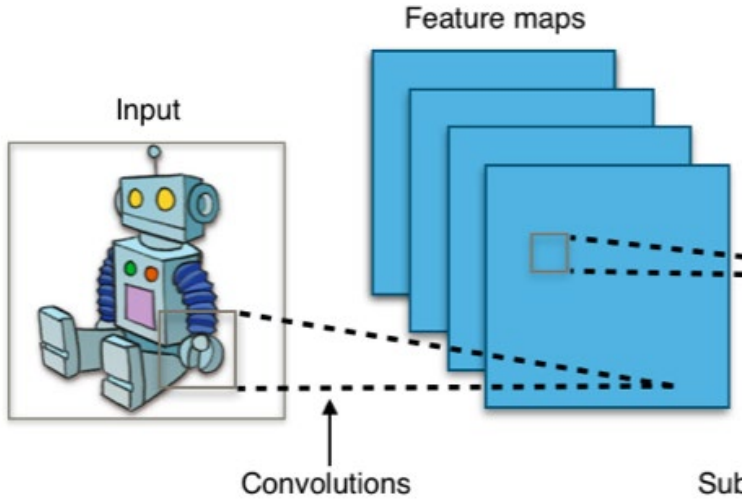
- Padding:
  - Zero
  - Constant
  - Reflection
  - Replication



Parameters of a convolutional layer

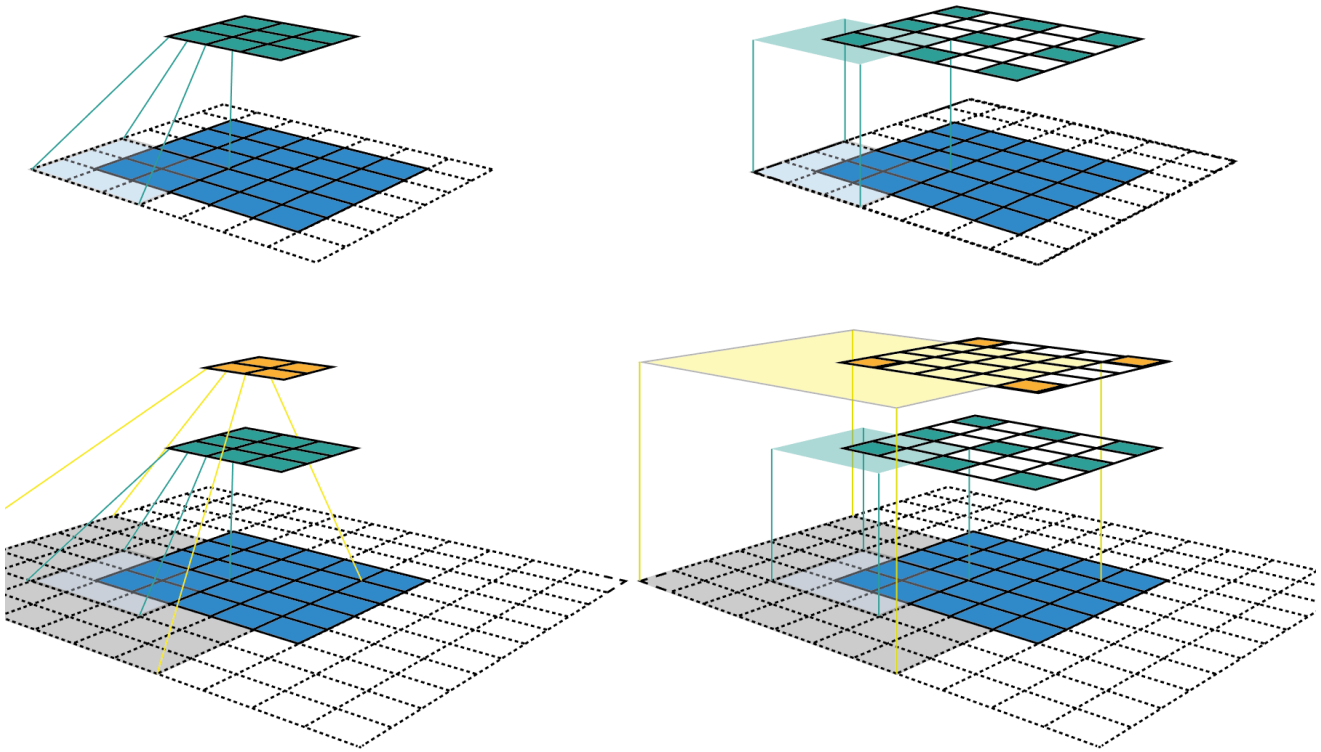
- Number of features

- Stride / kernel size and padding consistent across feature maps of a layer.



Sub-sampling

We want to compress information



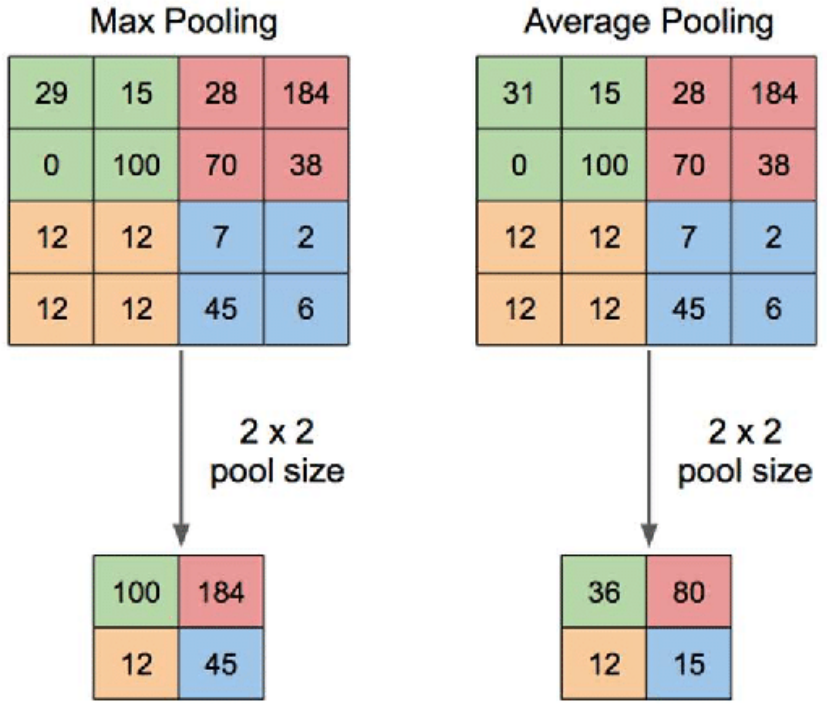
- Receptive Fields become larger and larger
- Size of the feature maps become smaller and smaller
- Features become more and more abstract (folding and non-linearities)

Effect of stride

- A stride greater than one allows for progressive compression of the spatial information

Pooling

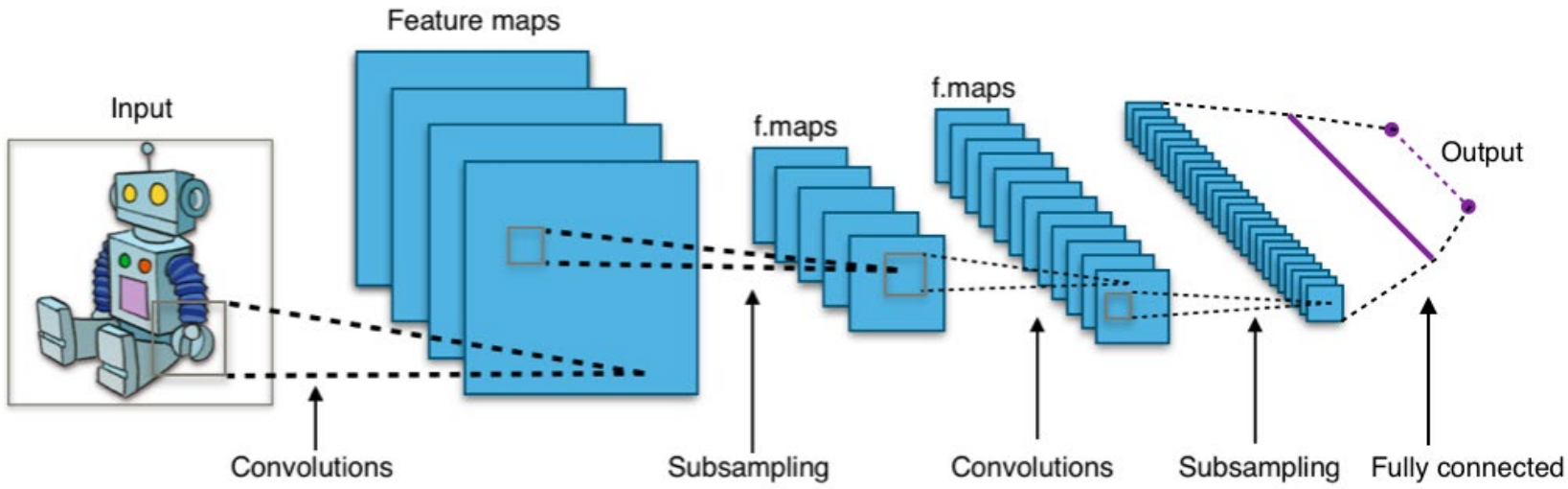
- Mathematical Operation to group/keep only certain features:
  - Max Pooling
  - Average Pooling
- Improve translational invariance



CNN overview

Classical CNN architecture

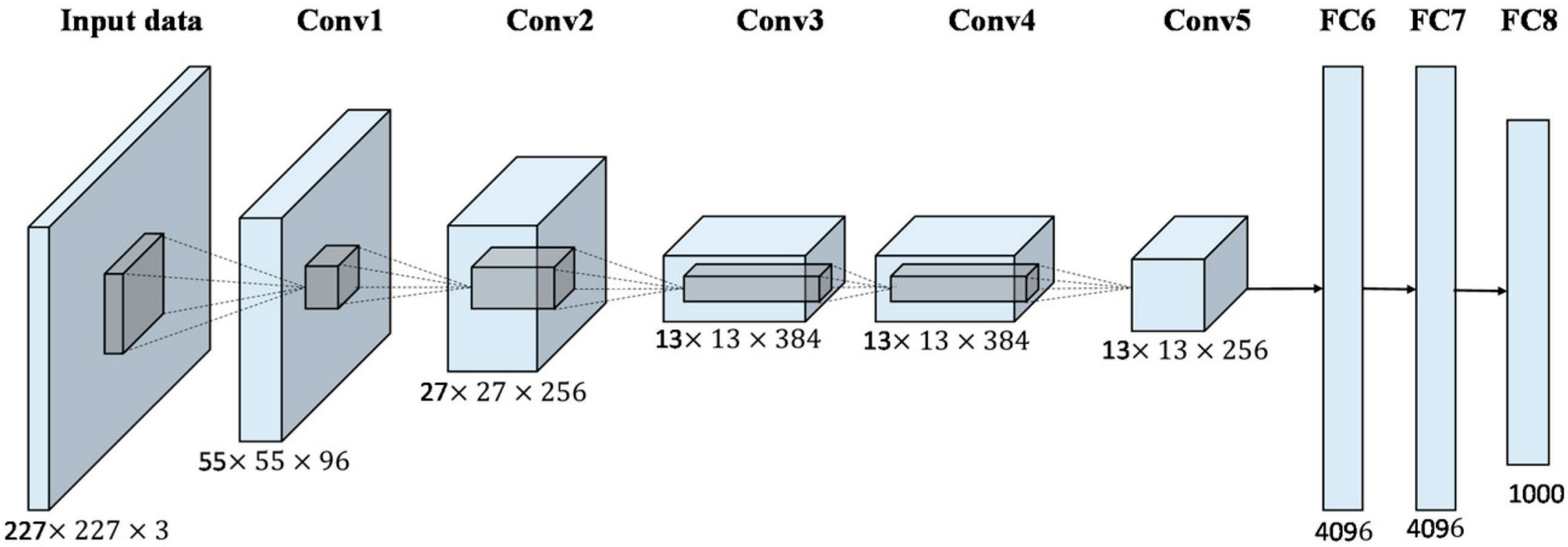
- For classification: last convolutional layer is concatenated and flatten
- Fully connected, then softmax



Classical CNN Architectures

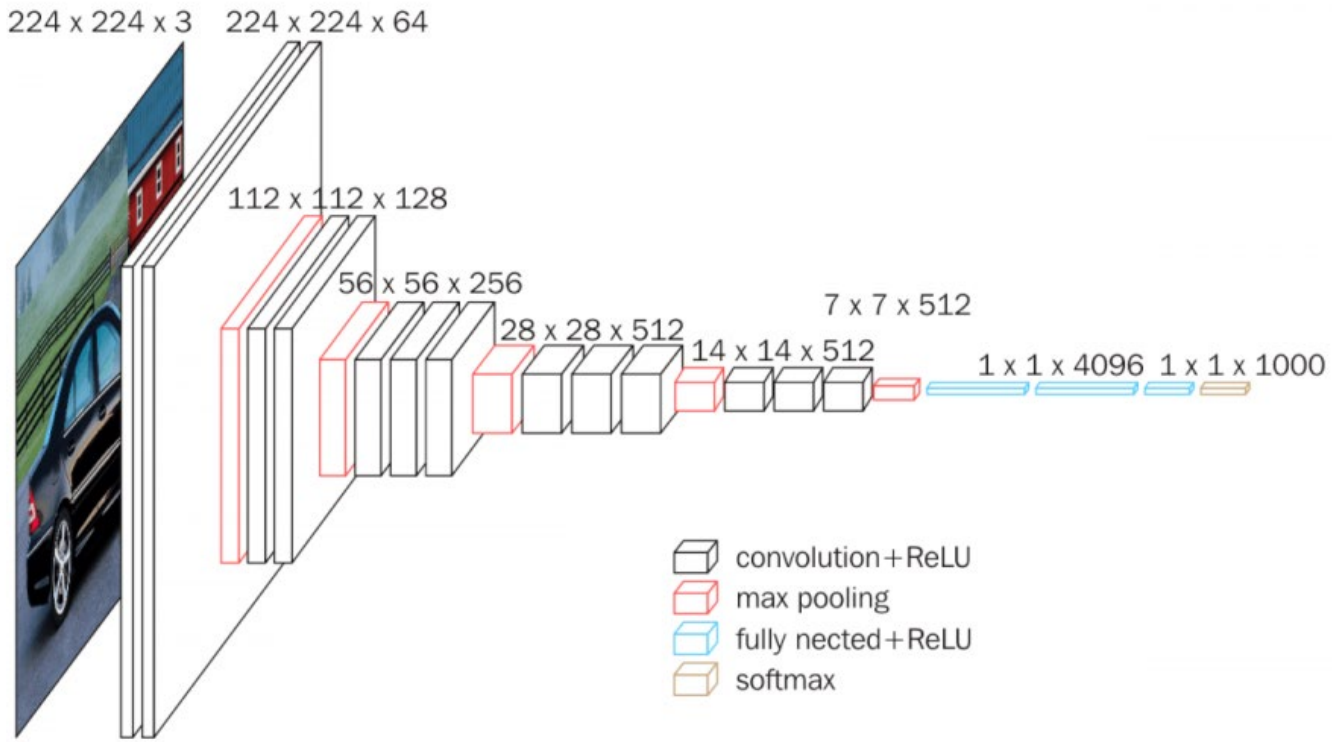
AlexNet (2012)

- Trained on ImageNet (1000 classes)

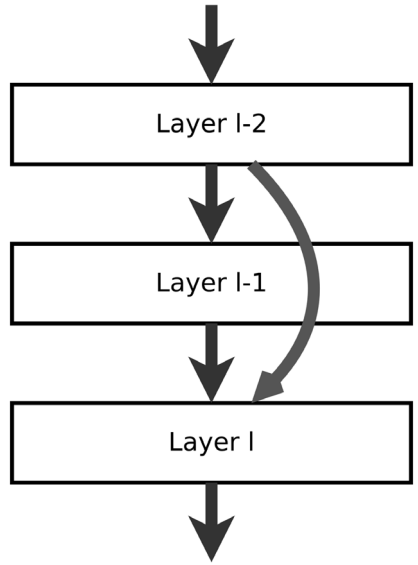


VGG (2014)

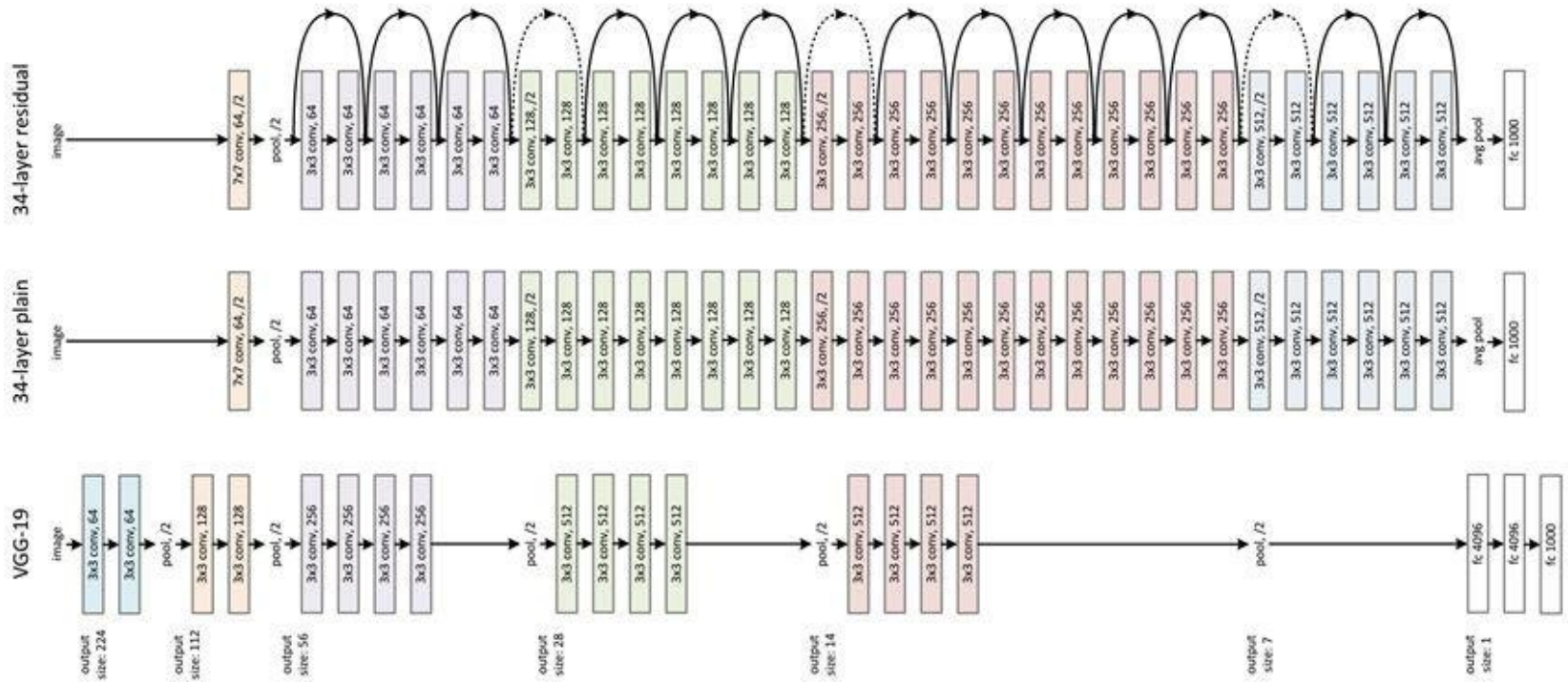
- Trained on ImageNet (1000 classes)



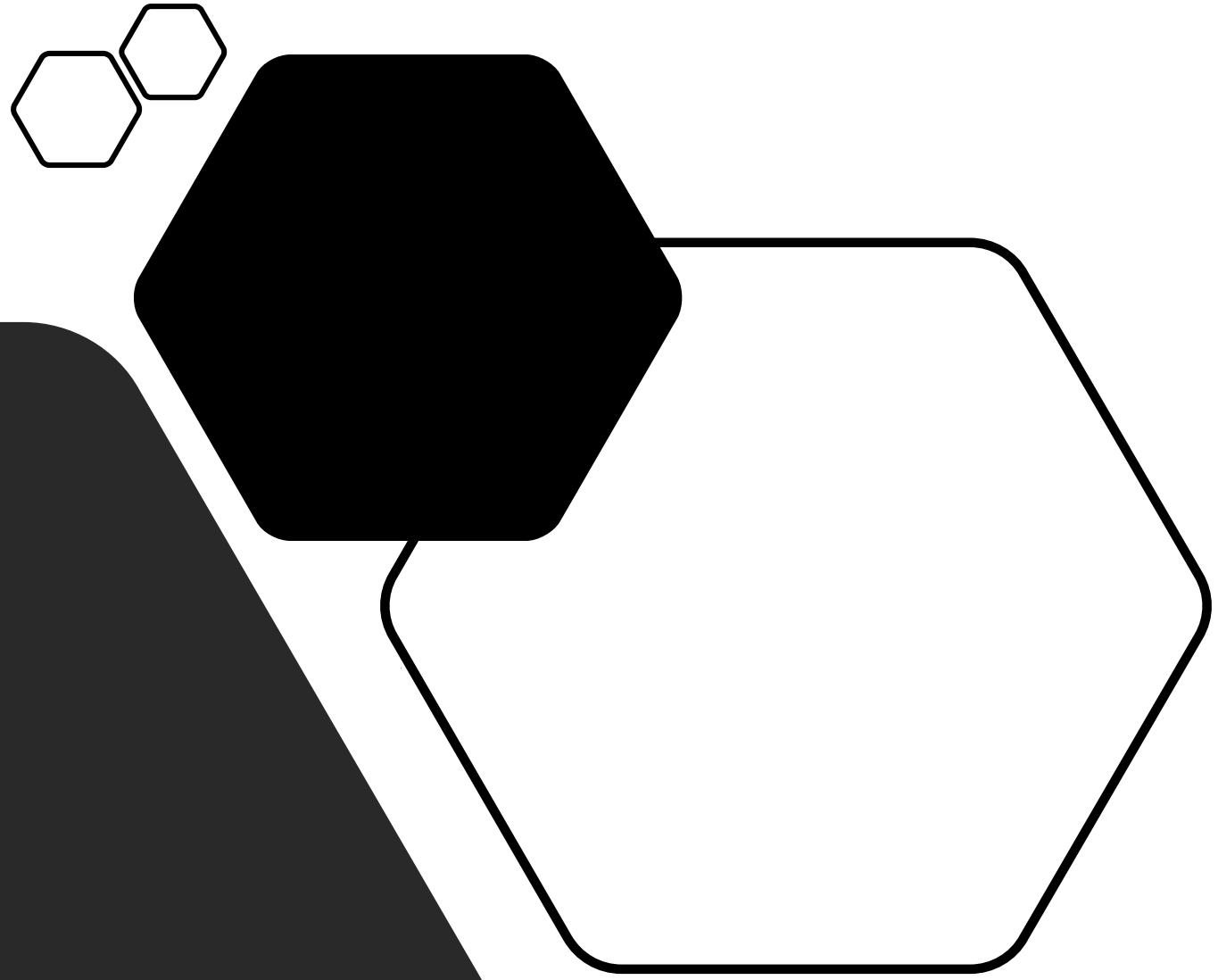
ResNet (2015)



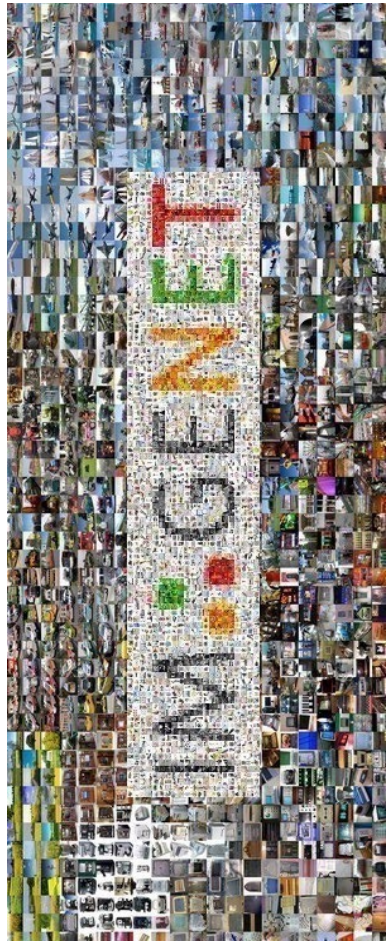
- Skip connections to improve gradient propagation
- Prevents the gradient dying out

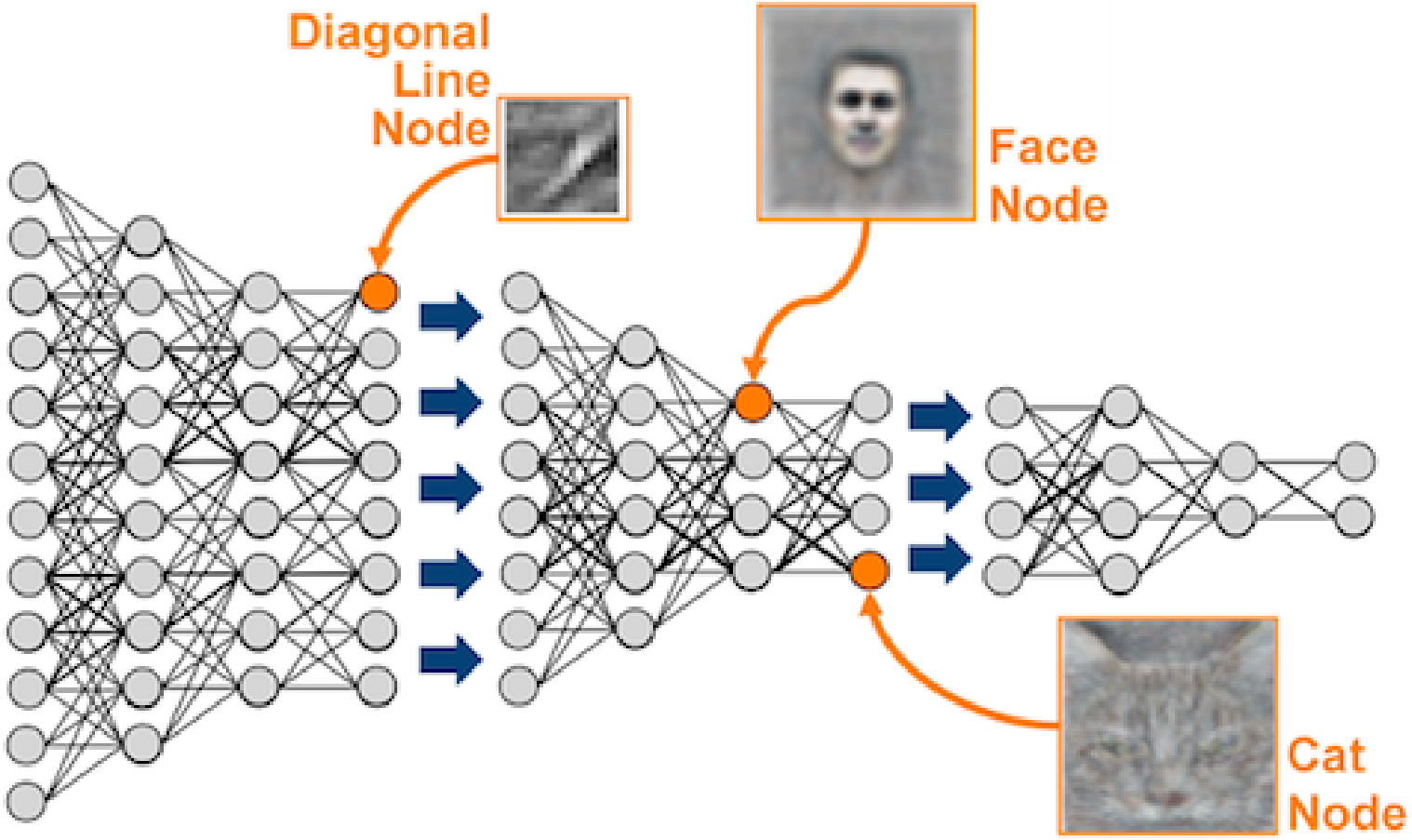


Transfer Learning

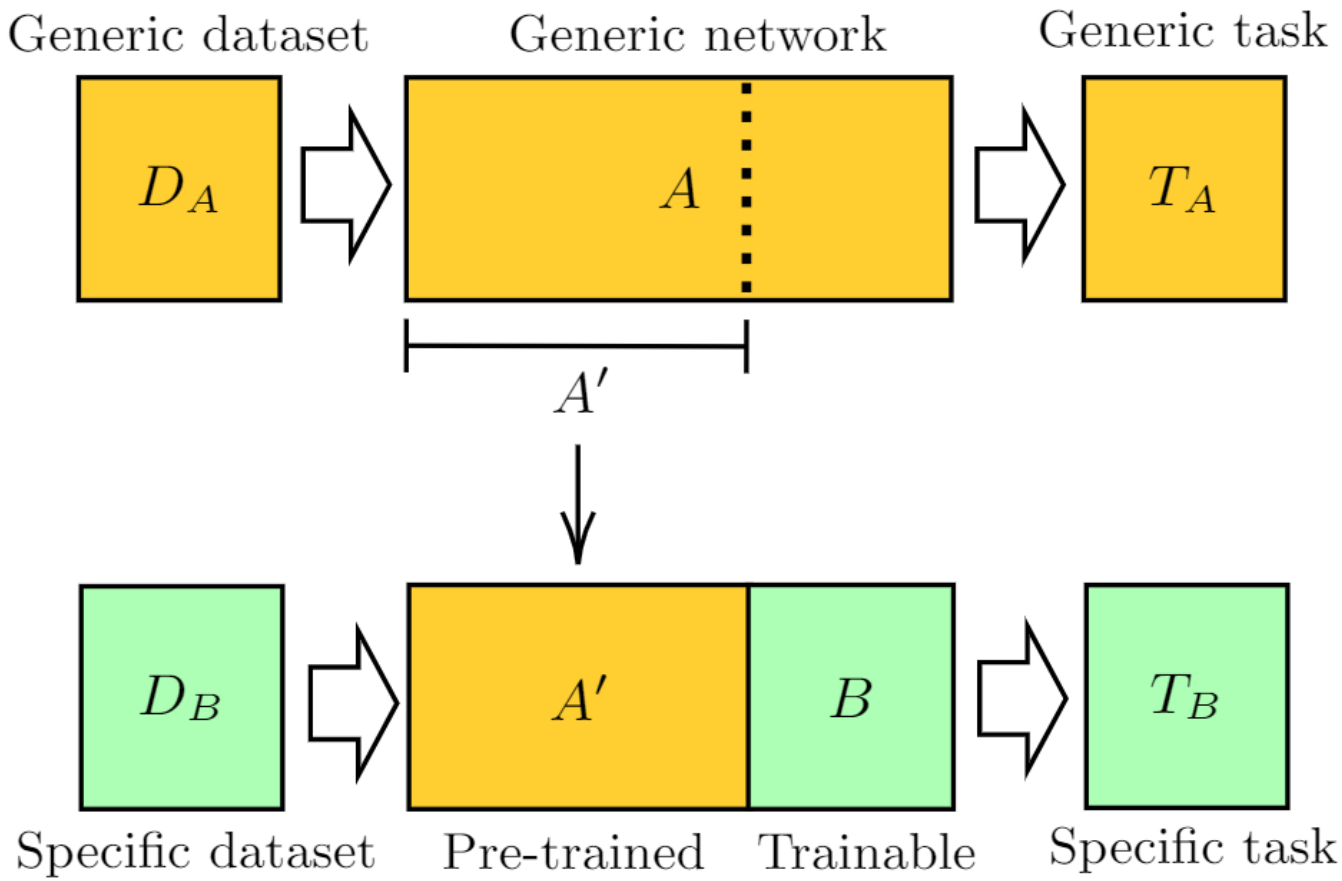


Learning generic features



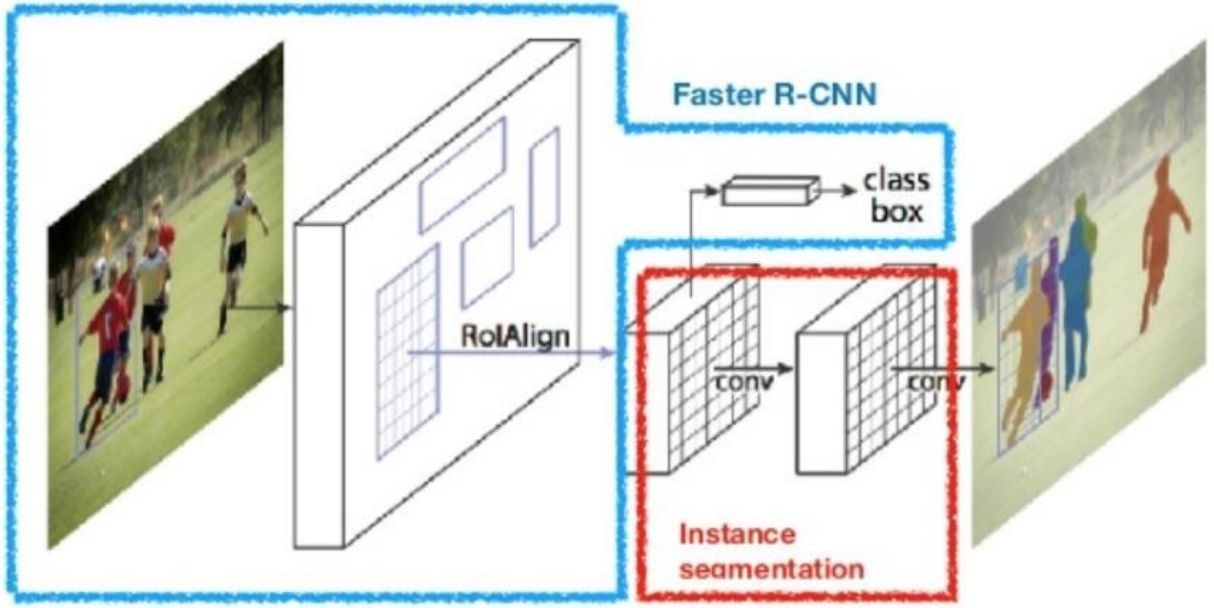


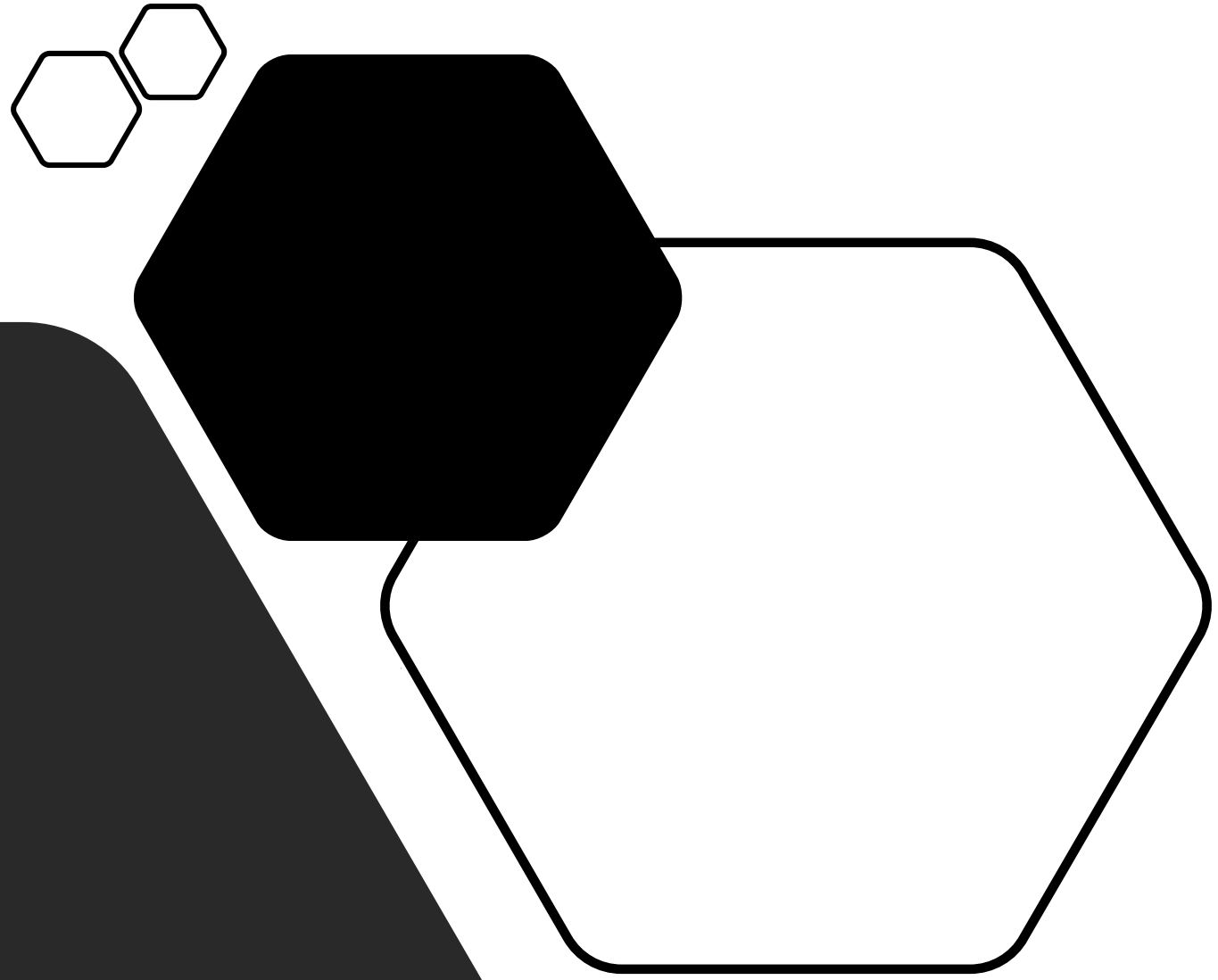
Re-using pre-trained networks



Going forward: Mask-RCNN

- Every patch of image can correspond to an object

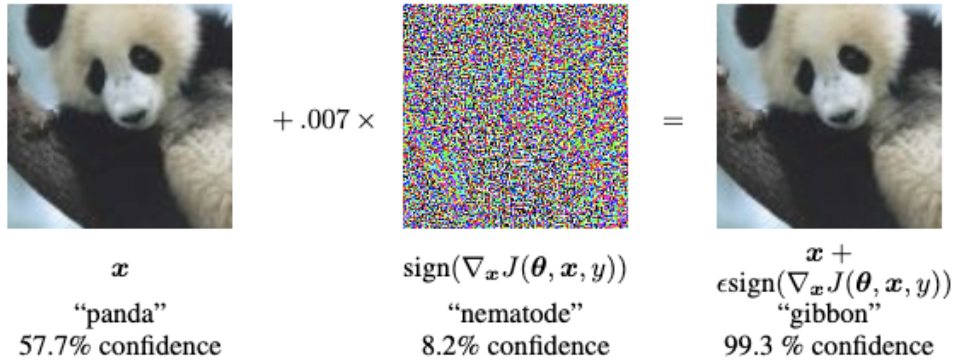




Limitations of CNNs

Dataset vs natural images

- A dataset gives you examples of images + classes
- Not all images and classes are accounted for (data is limited)
- How well does if work in the real world?



"a young boy is holding a baseball bat."

Lab 10 – 1

Lab 10 – part 1

- CIFAR 10:
  - 60000 32x32 images
  - 10 classes
- Use CNNs to solve this
- Use colab if necessary