Bootcamp 2023

Convolutional Neural Networks

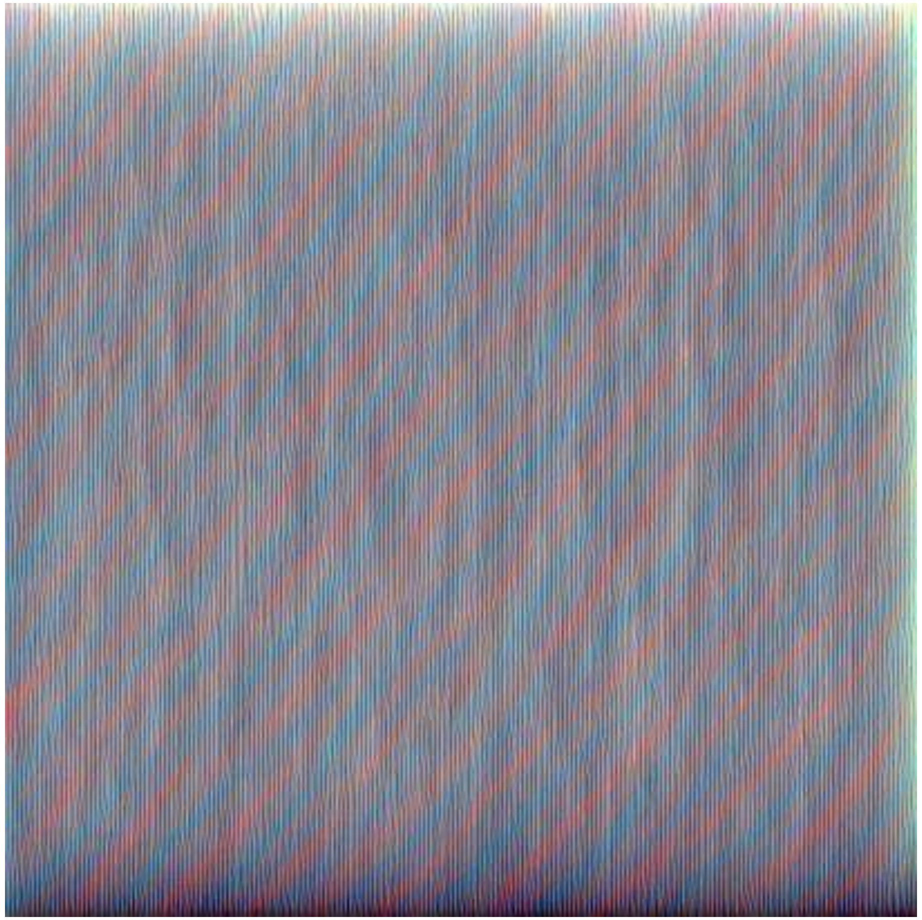
UNIVERSITY OF TORONTO
FACULTY OF APPLIED SCIENCE & ENGINEERING
Centre for Analytics and Artificial Intelligence Engineering

Going past the fully connected network

- In many image tasks, we want to be able to recognize something regardless of where it is in the image
- For fully-connected networks, the order of the inputs is fixed
- No “shift invariance”

Going past the fully connected network

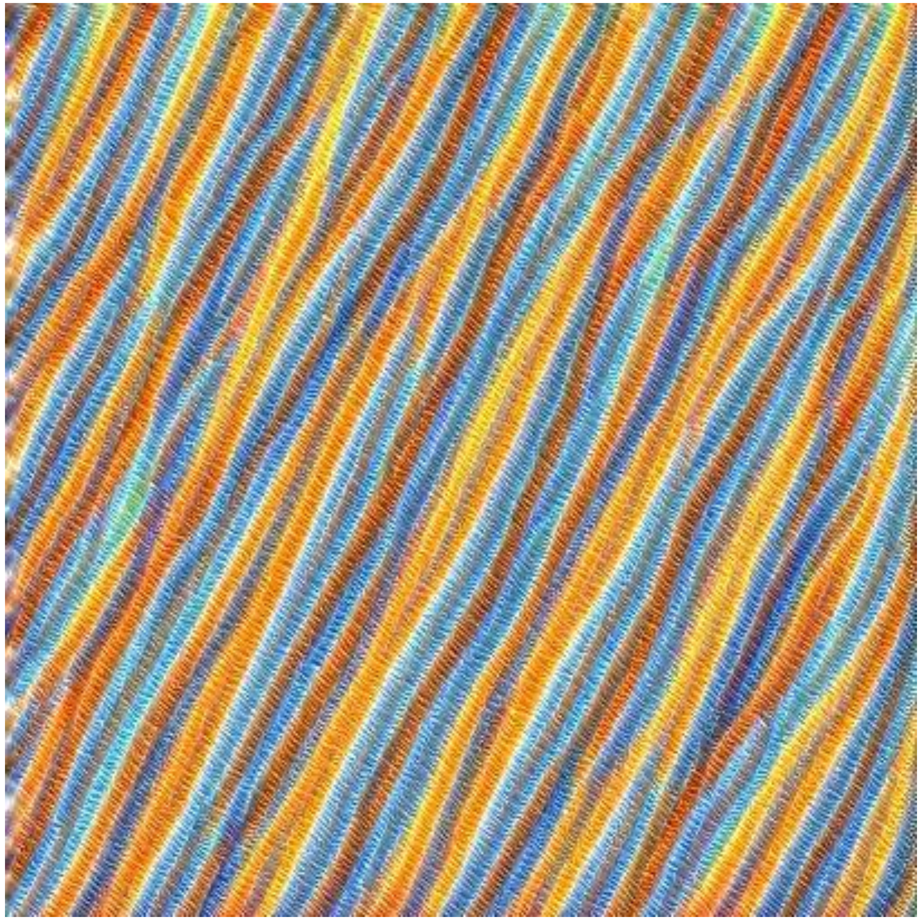
- In the 1950s and 60s, researchers showed that the brain contains neurons which respond to specific patterns, regardless of where they appear
- Combinations of very basic patterns can then be recognized as a more complicated one!



VGG-16, neuron in layer 7

Going past the fully connected network

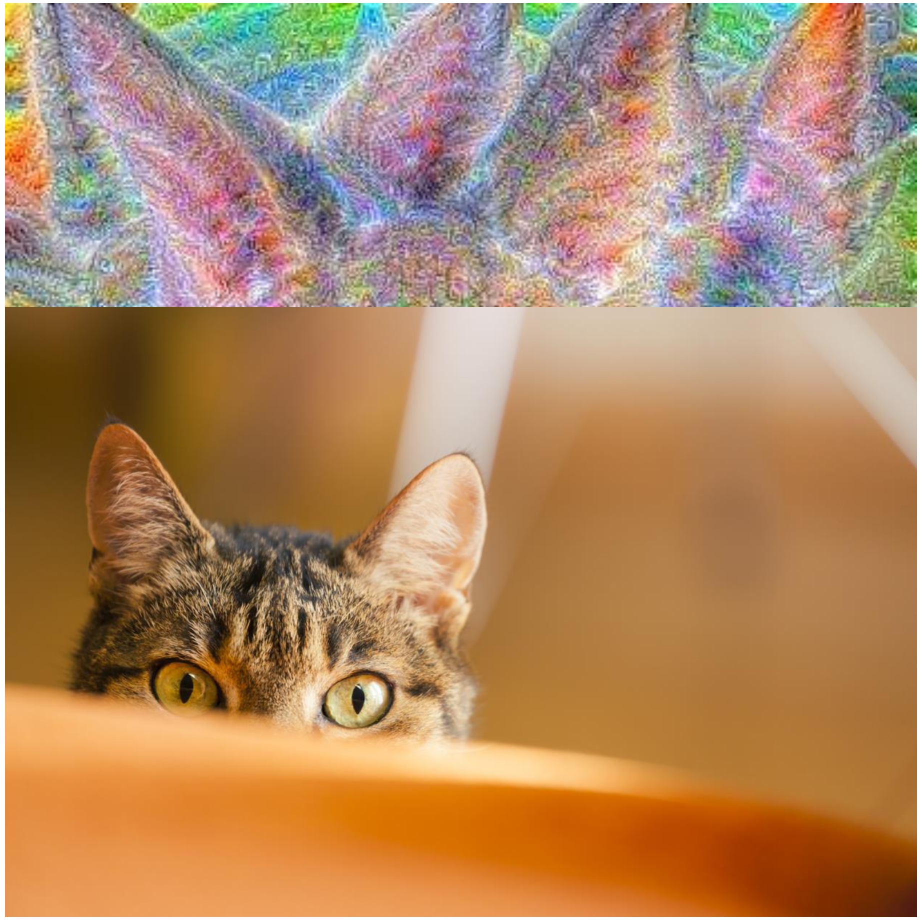
- In the 1950s and 60s, researchers showed that the brain contains neurons which respond to specific patterns, regardless of where they appear
- Combinations of very basic patterns can then be recognized as a more complicated one!



VGG-16, neuron in layer 14

Going past the fully connected network

- In the 1950s and 60s, researchers showed that the brain contains neurons which respond to specific patterns, regardless of where they appear
- Combinations of very basic patterns can then be recognized as a more complicated one!



VGG-16, neuron in layer 40

Interactive CNNs

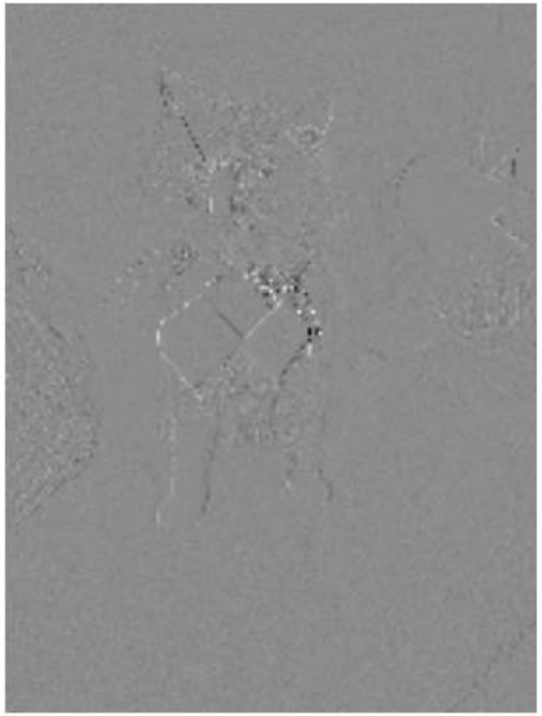
https://adamharley.com/nn_vis/cnn/2d.html

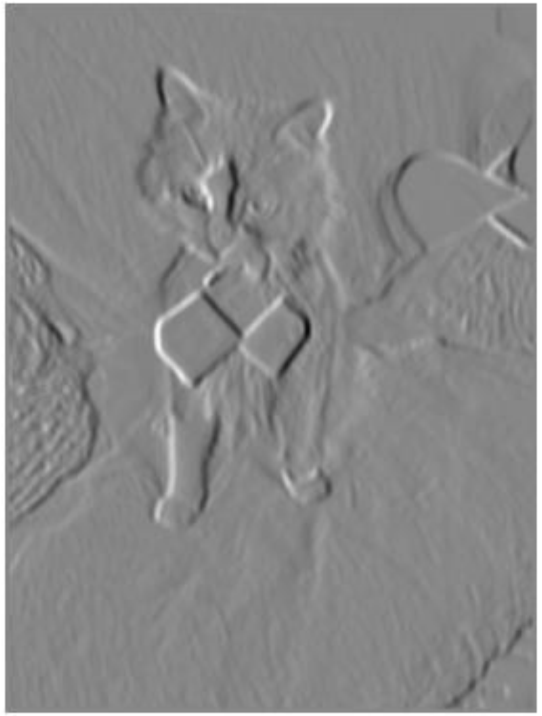
CNNs in depth

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

Stride 10

25x25 Filter





CNNs in depth: Kernel/Filter

- Values are learned during training
- Tunable hyperparameters:
  - Filter size: capture larger areas or finer details
  - Stride: reduce number of computations by moving the filter by more pixels at a time
  - Number of filters in each layer: more variety in feature maps at a cost to compute

CNNs in depth: Activation Layers

- Typically, ReLU is used in CNNs: $f(x) = \max(0, x)$
- Introduces non-linearity without significantly altering feature maps
- Faster training and convergence due to a simple derivative

ReLU

Tanh

CNNs in depth: Pooling

- Pooling layers reduce size of intermediate feature maps, helping with regularization and reducing complexity
- Max pooling most common, but other options exist

Max pooling (2x2)

Max pooling (8x8)

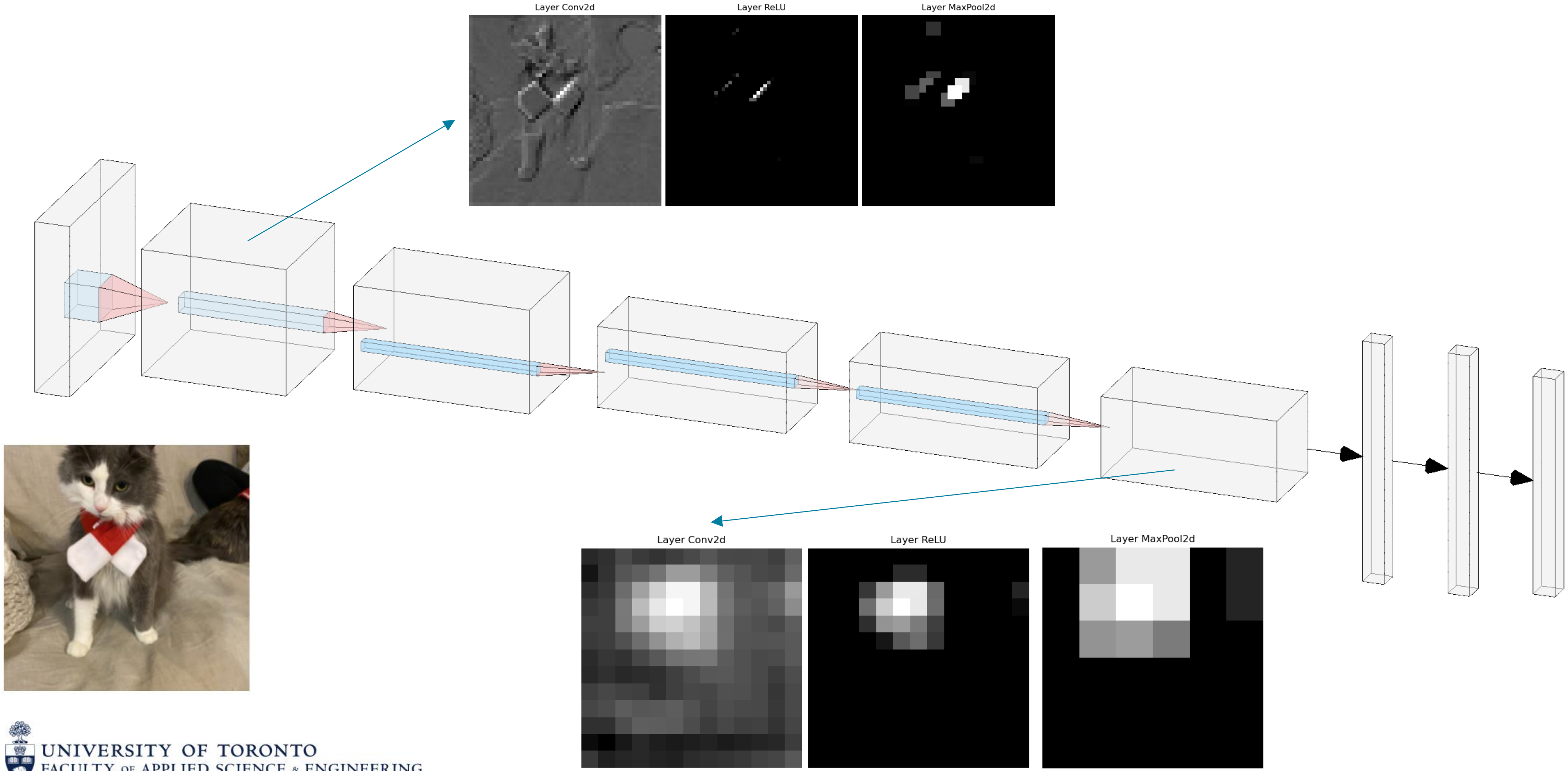
CNNs in depth: Fully Connected Layers

- Once final feature maps are created, output is fed to a standard neural network
- Information from feature maps is used to predict the content of the image

Many Hyperparameters!

- Number of Layers
- Number of Filters
- Filter Size
- Stride Size
- Pooling Size
- Pooling Type
- Activation Function

AlexNet (2012)

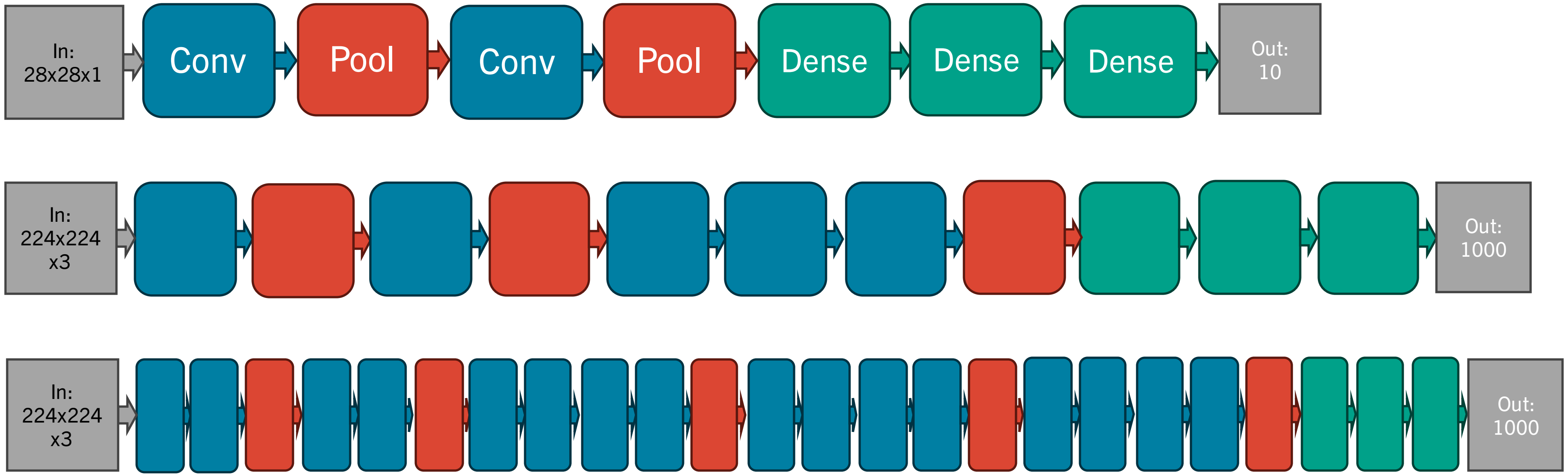


Going Deeper

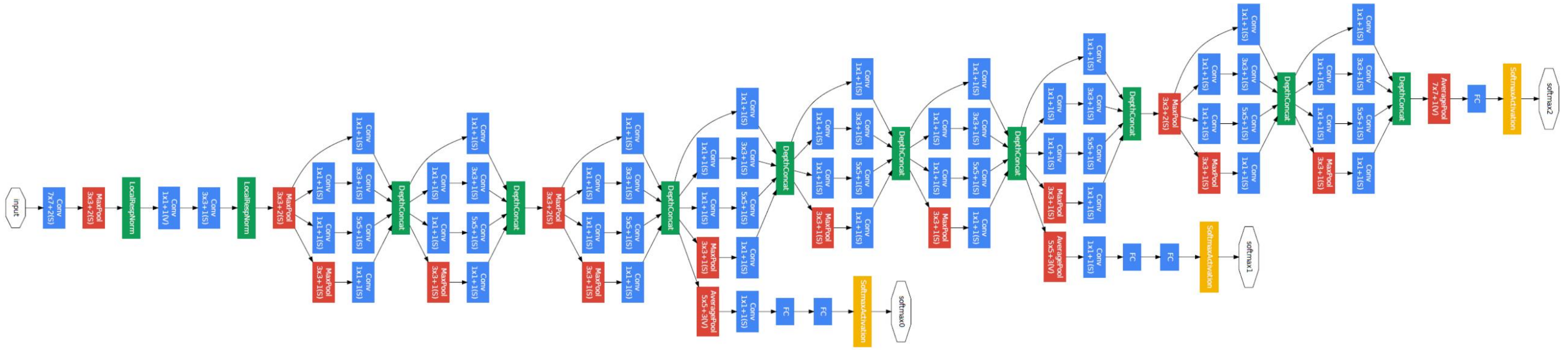
LeNet (1990s)

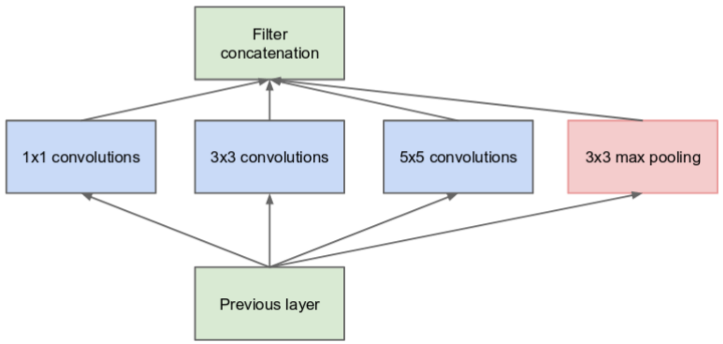
AlexNet (2012)

VGG (2014)

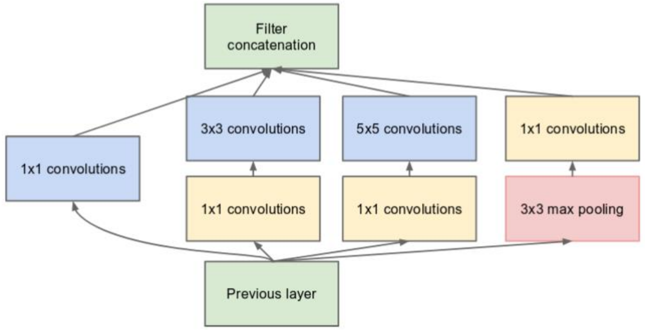


GoogLeNet (Inception)



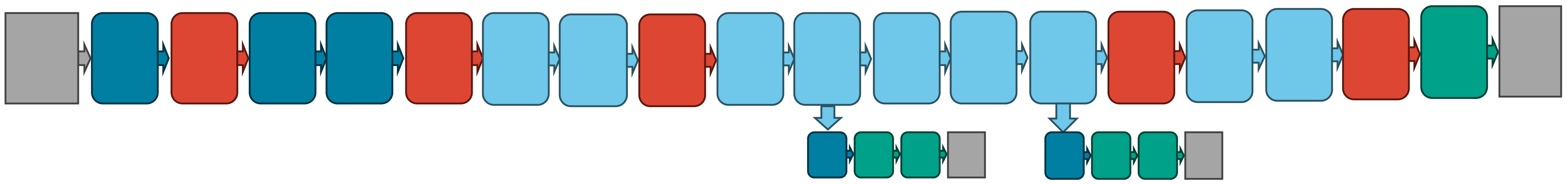


(a) Inception module, naïve version

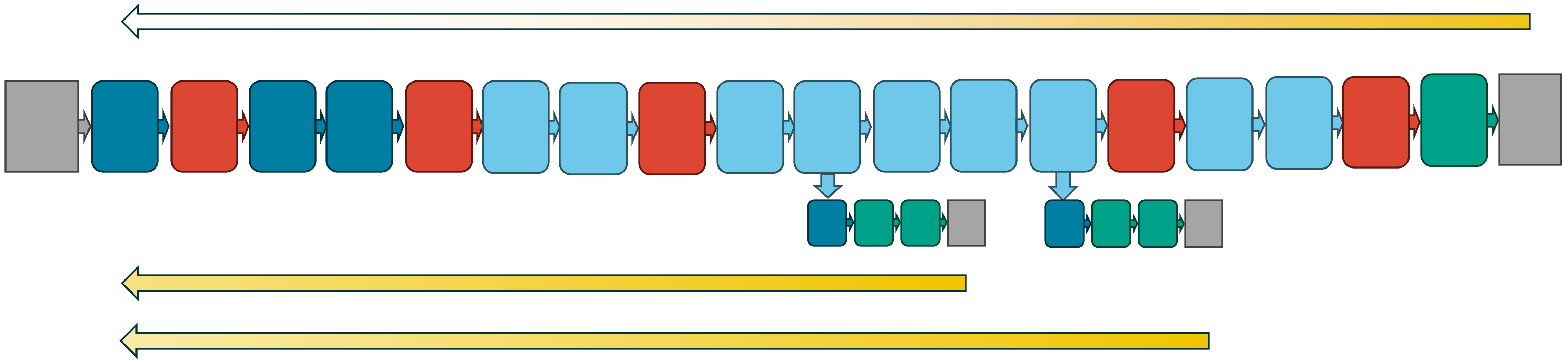


(b) Inception module with dimension reductions

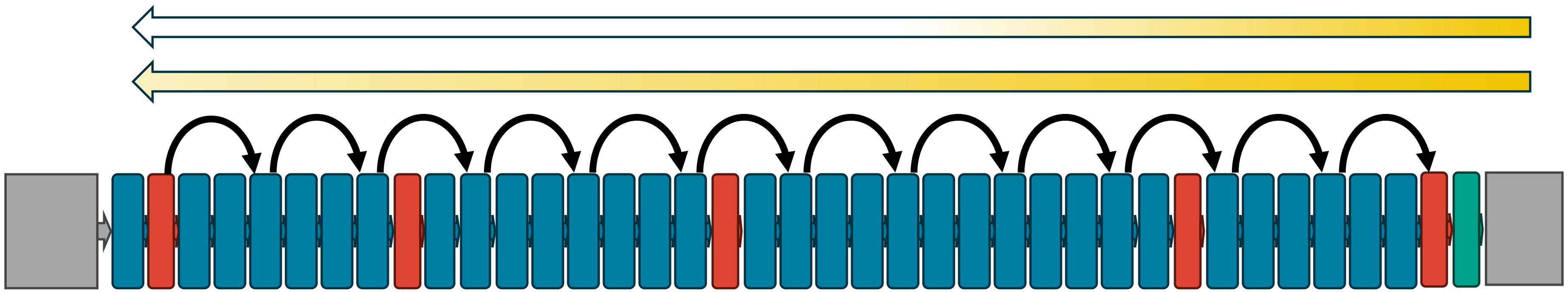
GoogLeNet (Inception)



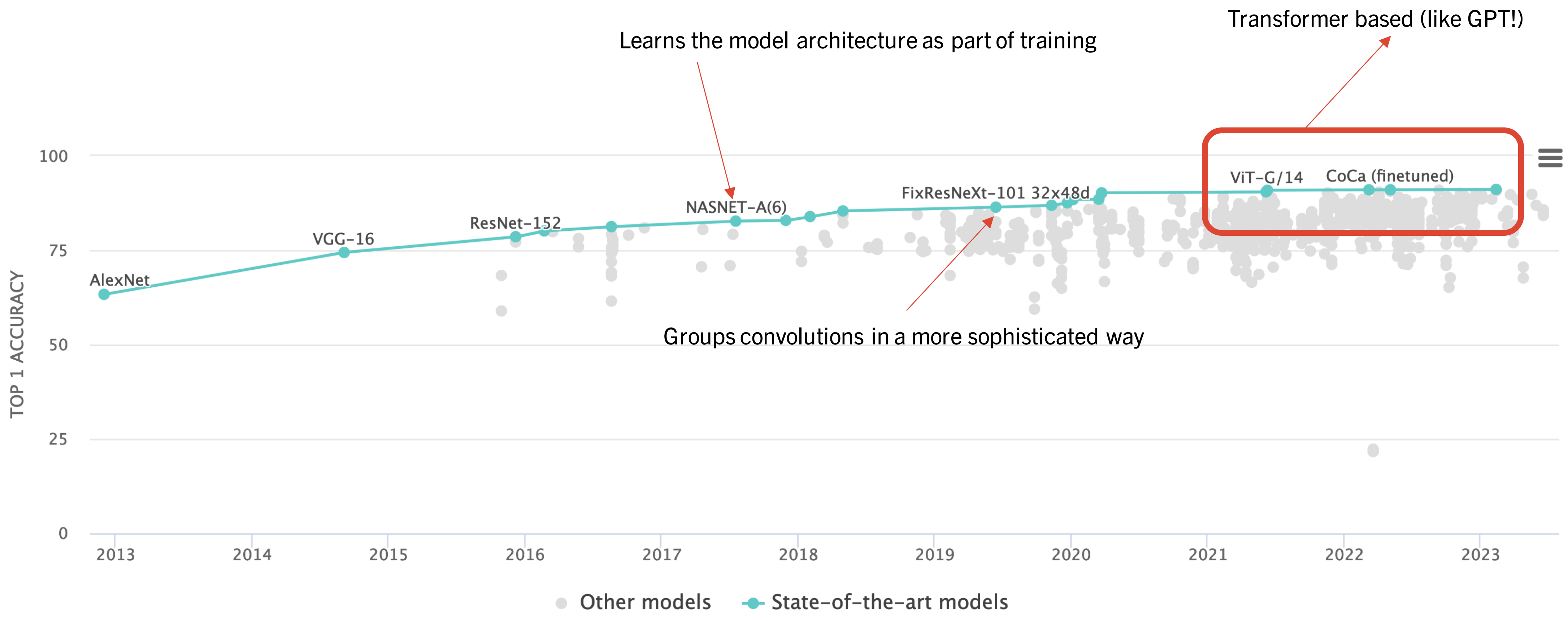
GoogLeNet (Inception)



ResNet



What about the state of the art?



CNNs Are Everywhere!

- **Image and Video Recognition:** Modern smartphones run CNNs directly to power features like facial recognition and categorizing photos
- **Medical Imaging:**
  CNNs are utilized to detect anomalies in X-rays, MRIs…
- **Time-Series Analysis:**
  CNNs are used for signal processing, such as predicting data from ECGs

Interactive CNN Websites

- https://poloclub.github.io/cnn-explainer/
- Deep interactive dive into CNN operations

- https://cs.stanford.edu/people/karpathy/convnetjs/demo/cifar10.html
- Train a CNN in your browser (very computationally expensive!)