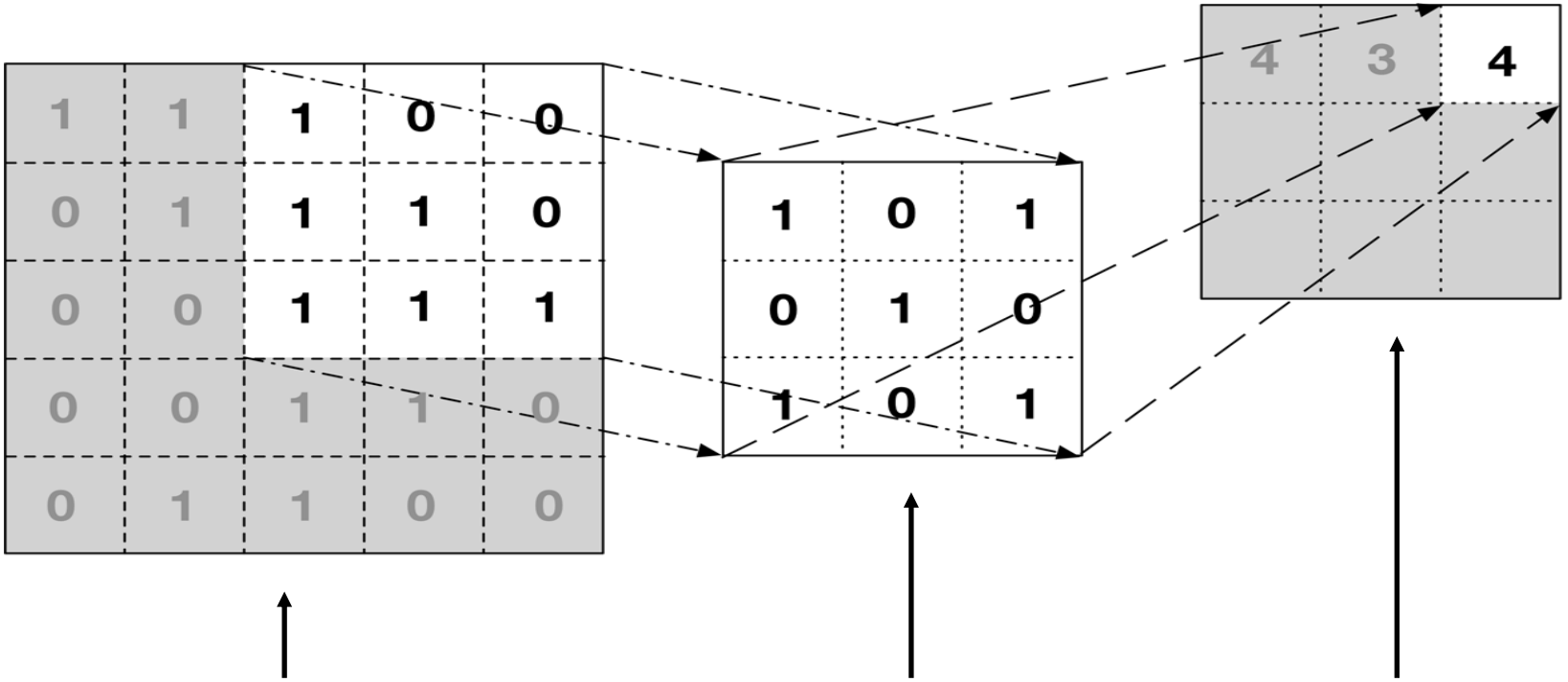
CONVOLUTIONAL NEURAL NETWORKS

Convolutional Neural Networks (CNNs)

- Biggest driver of much recent excitement in machine learning
- Specialized architecture for grid-like data
  - Images
  - Sound
- Key operation: ***discrete convolution***

Discrete Convolution

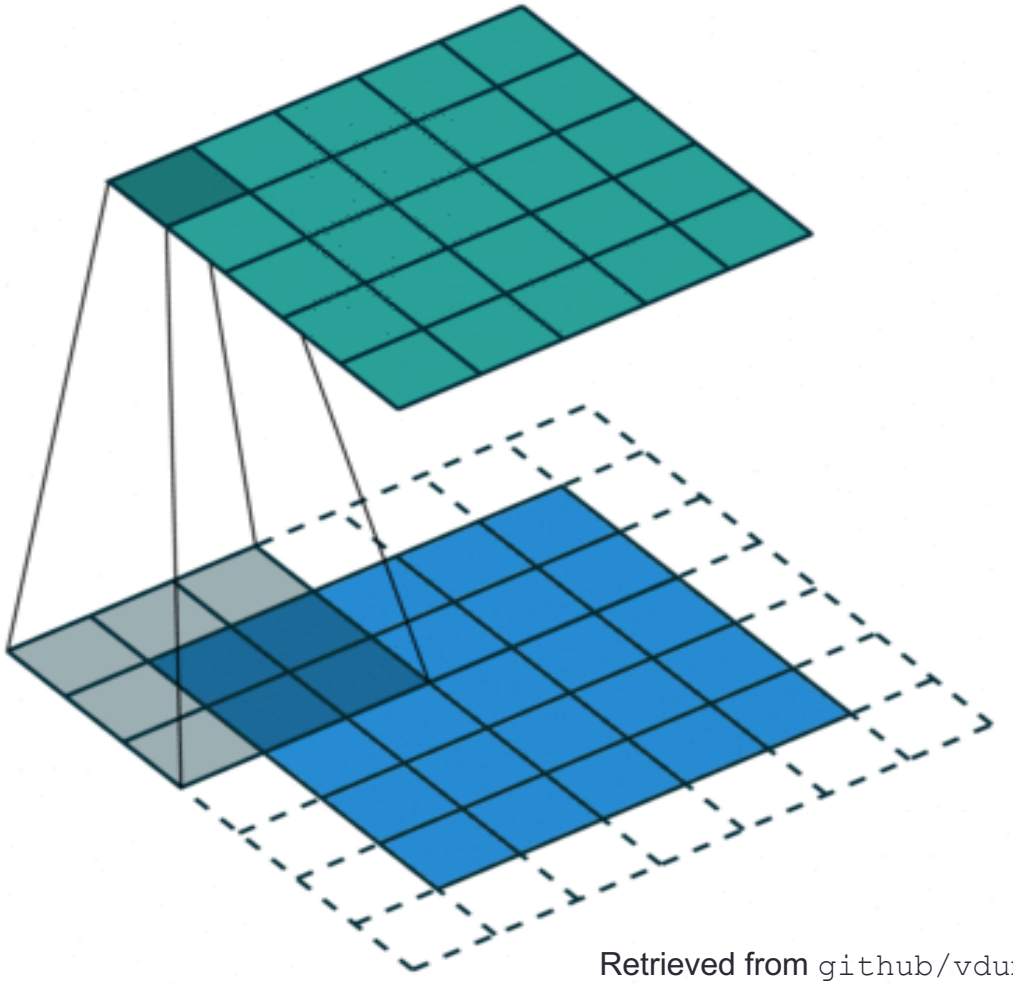


1 * 1 = 1
0 * 0 = 0
0 * 1 = 0
1 * 0 = 0
1 * 1 = 1
0 * 0 = 0
1 * 1 = 1
1 * 0 = 0
\+ 1 * 1 = 1

4

Adapted from *Deep Learning*, Gibson and Patterson.

Discrete Convolution



Retrieved from `github/vdumoulin/conv_arithmetic`

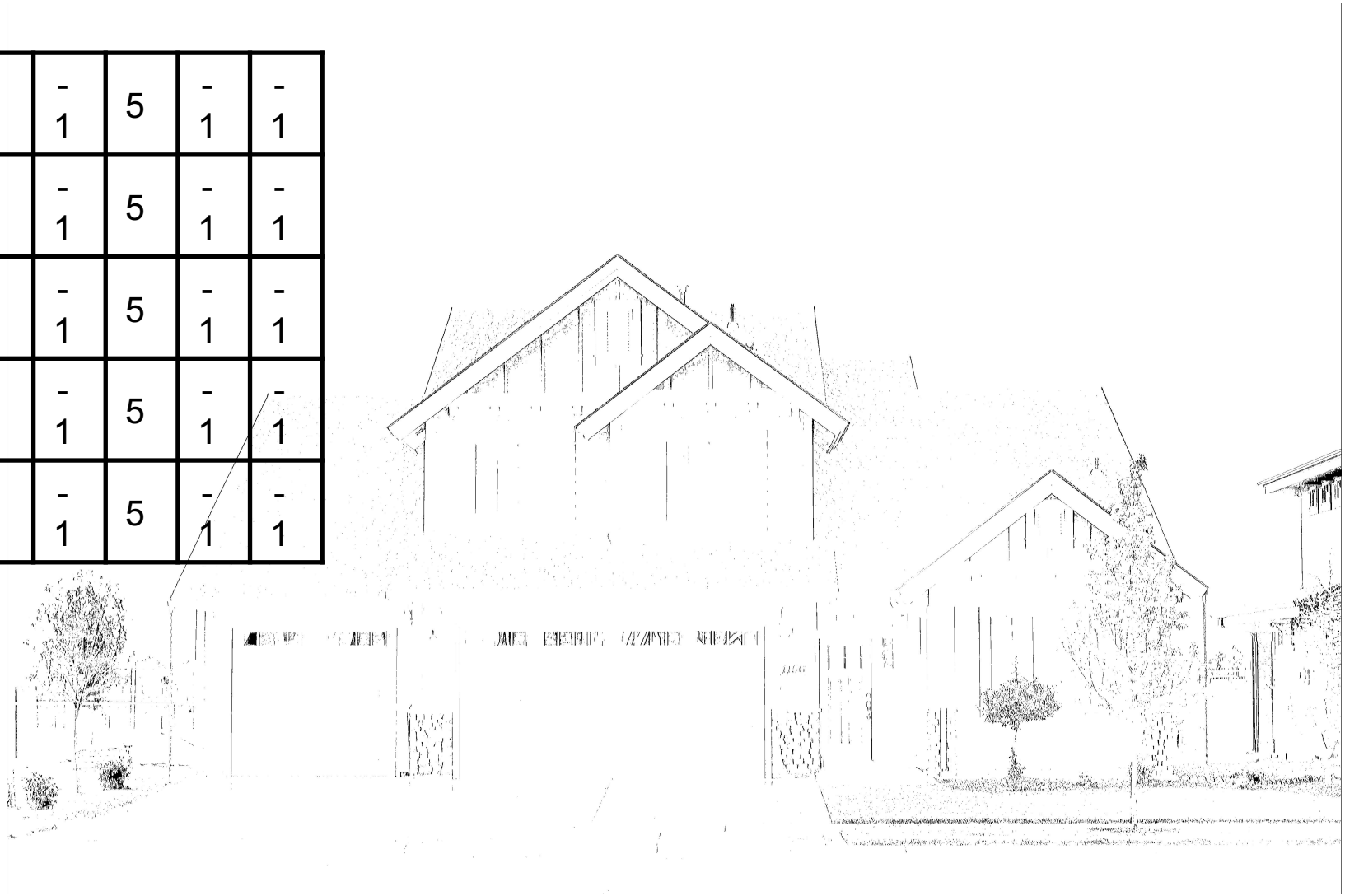
Applications of Convolutions

Applications of Convolutions

-1	-1	-1	-1	-1
-1	-1	-1	-1	-1
5	5	5	5	5
-1	-1	-1	-1	-1
-1	-1	-1	-1	-1

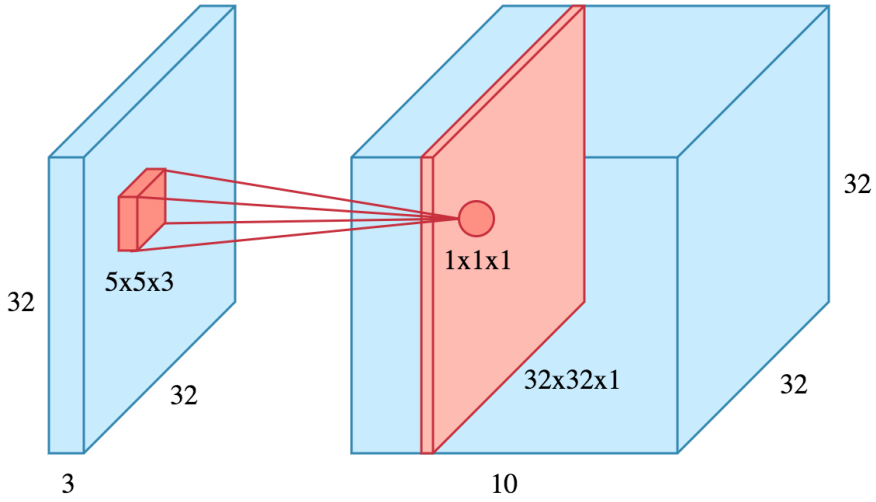
Applications of Convolutions

-1	-1	5	-1	-1
-1	-1	5	-1	-1
-1	-1	5	-1	-1
-1	-1	5	-1	-1
-1	-1	5	-1	-1

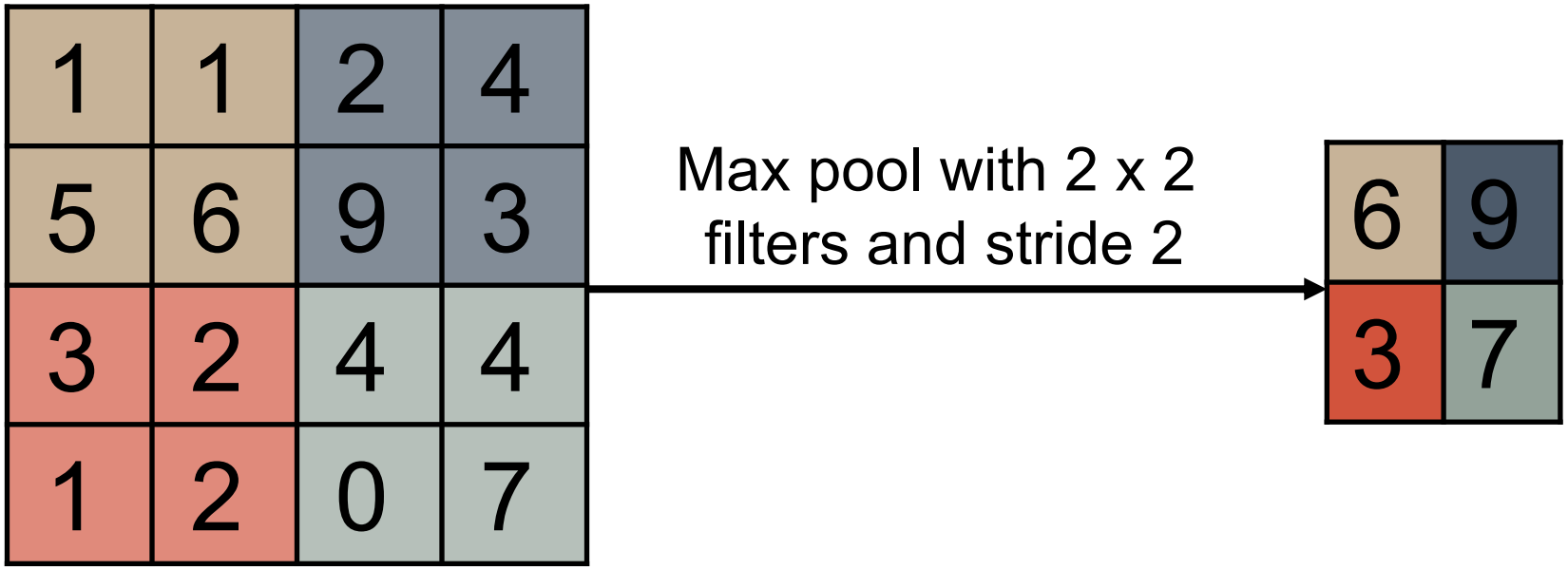


A Convolutional Layer

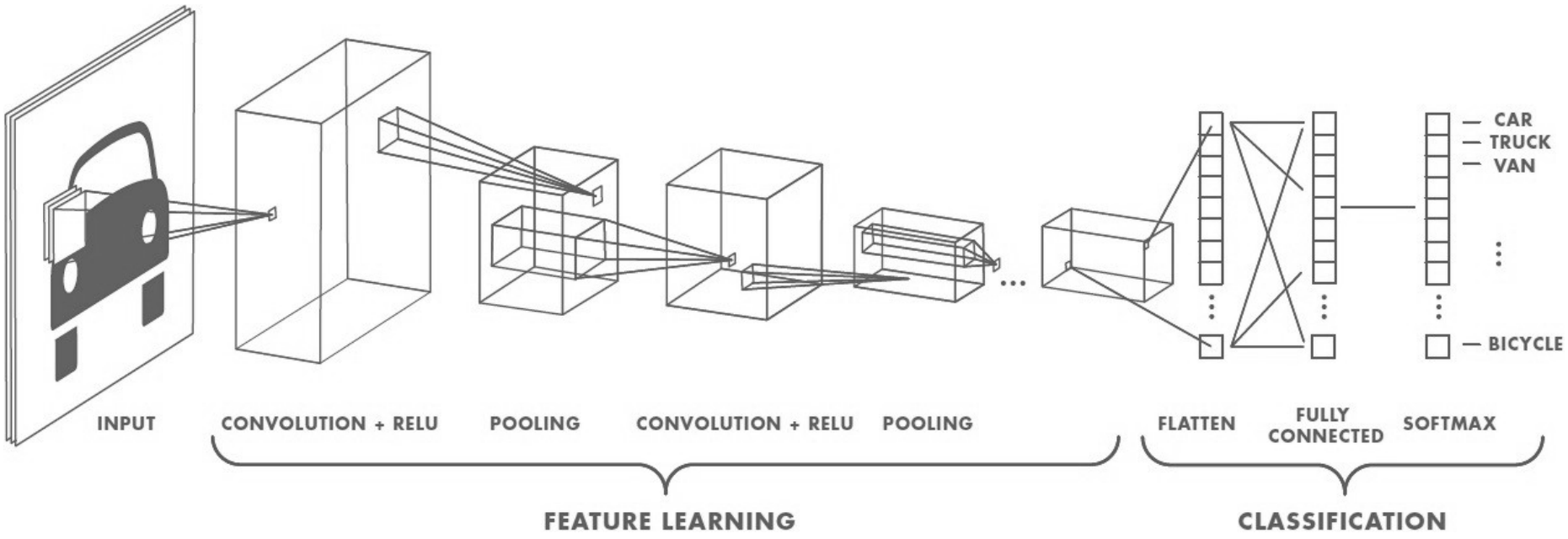
- Maps an input volume to an output volume
- Comprised of many convolutional filters
  - Let's say input volume is ***h x w x d***
  - We have ***s*** filters that are ***m x n***
  - Output volume is ***h x w x s***
- Each filter is highly localized
  - "Receptive field"
  - Weight sharing
- **Kernel weights are learned**



Pooling Layer



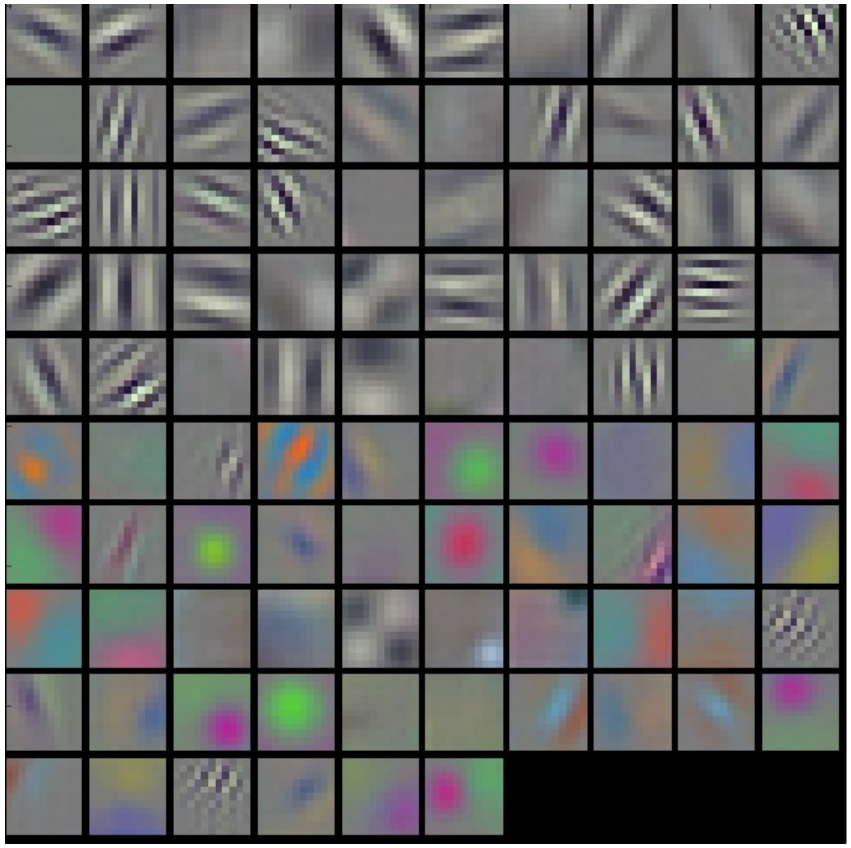
A Typical CNN Architecture



Retrieved from `https://hackernoon.com/what-is-a-capsnet-or-capsule-network-2bfbe48769cc`

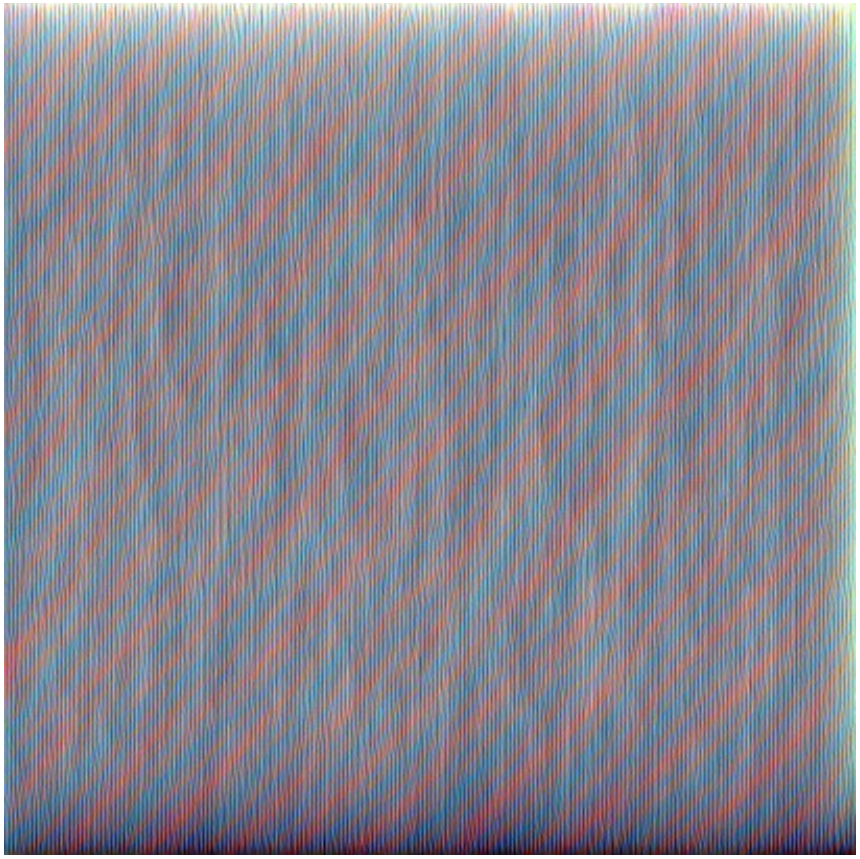
What We Know About Learning in CNNs

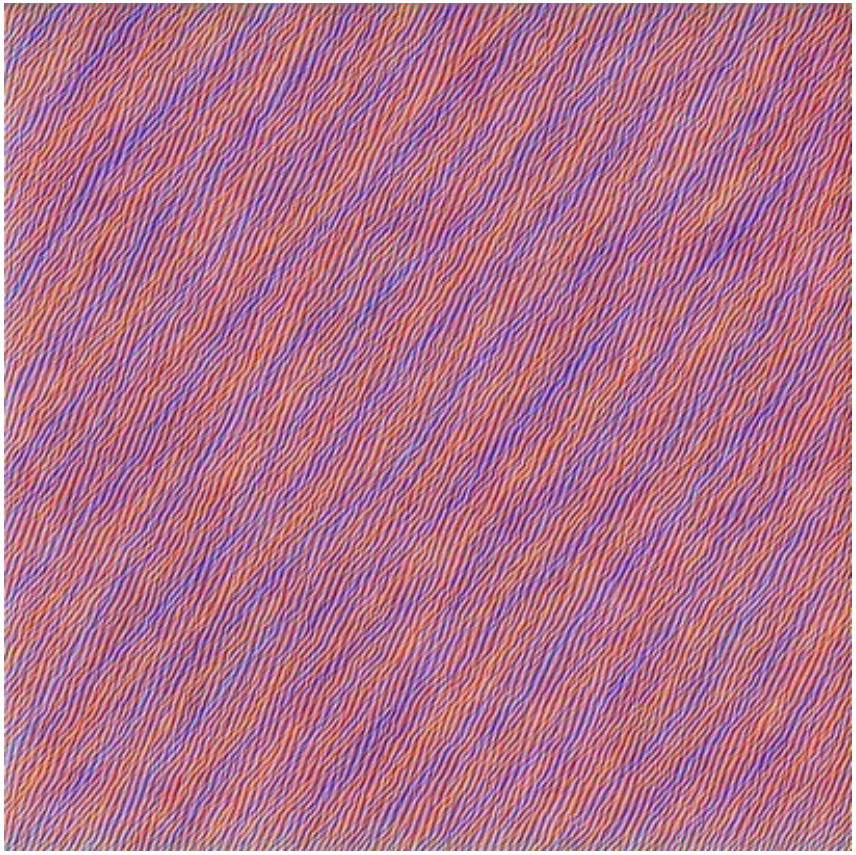
- Early filters are more abstract, later filters less so



What We Know About Learning in CNNs

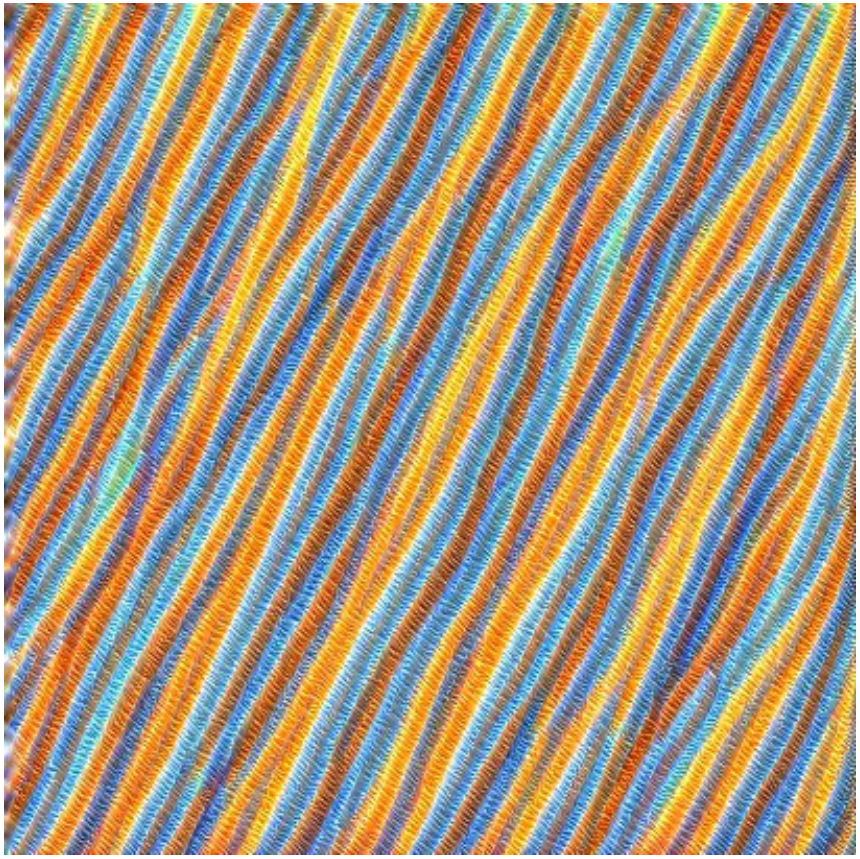
- Early filters are more abstract, later filters less so

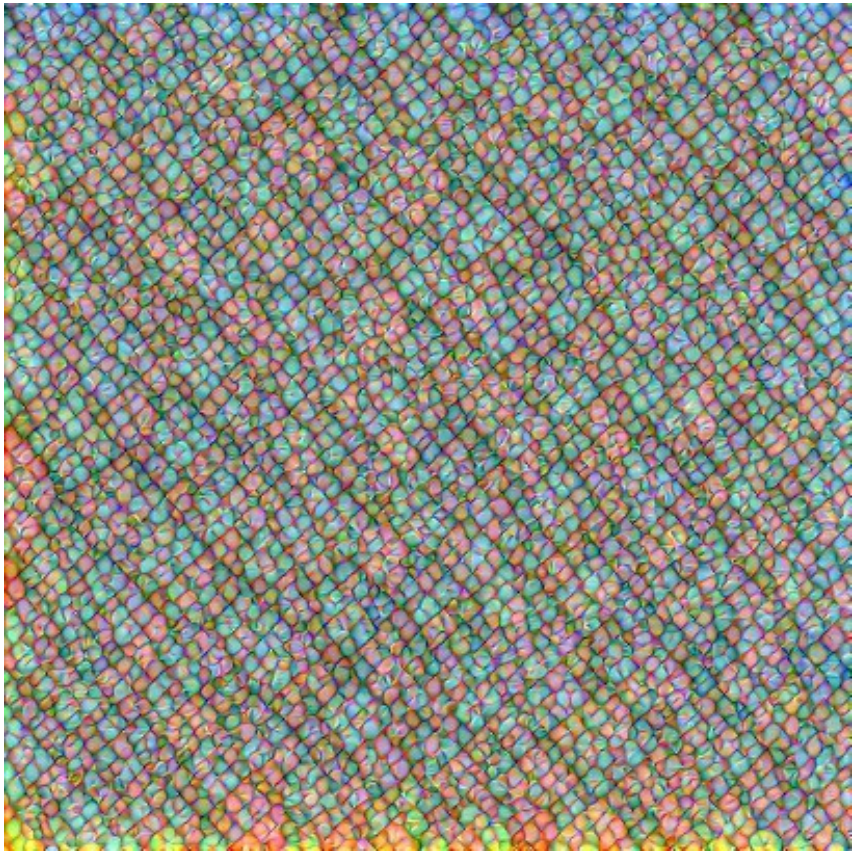




What We Know About Learning in CNNs

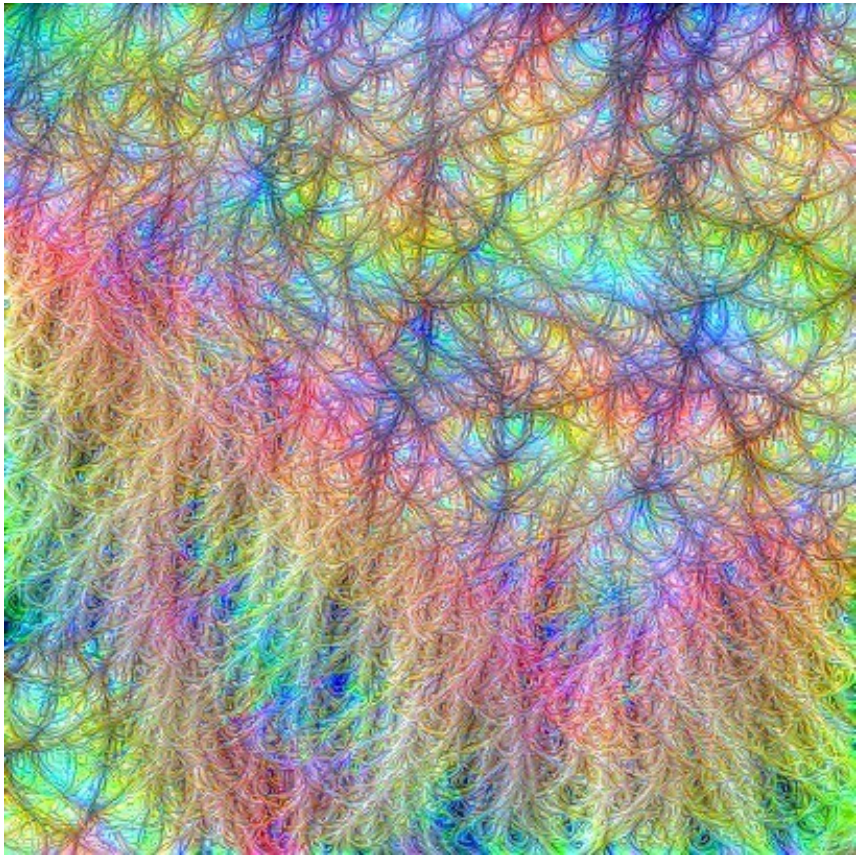
- Early filters are more abstract, later filters less so





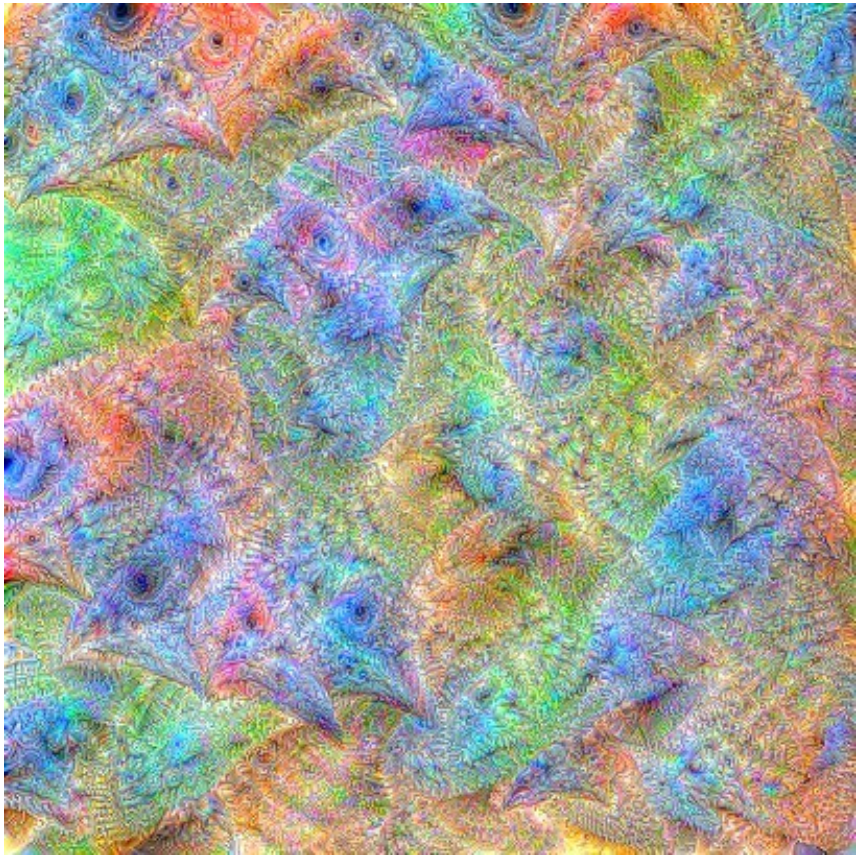
What We Know About Learning in CNNs

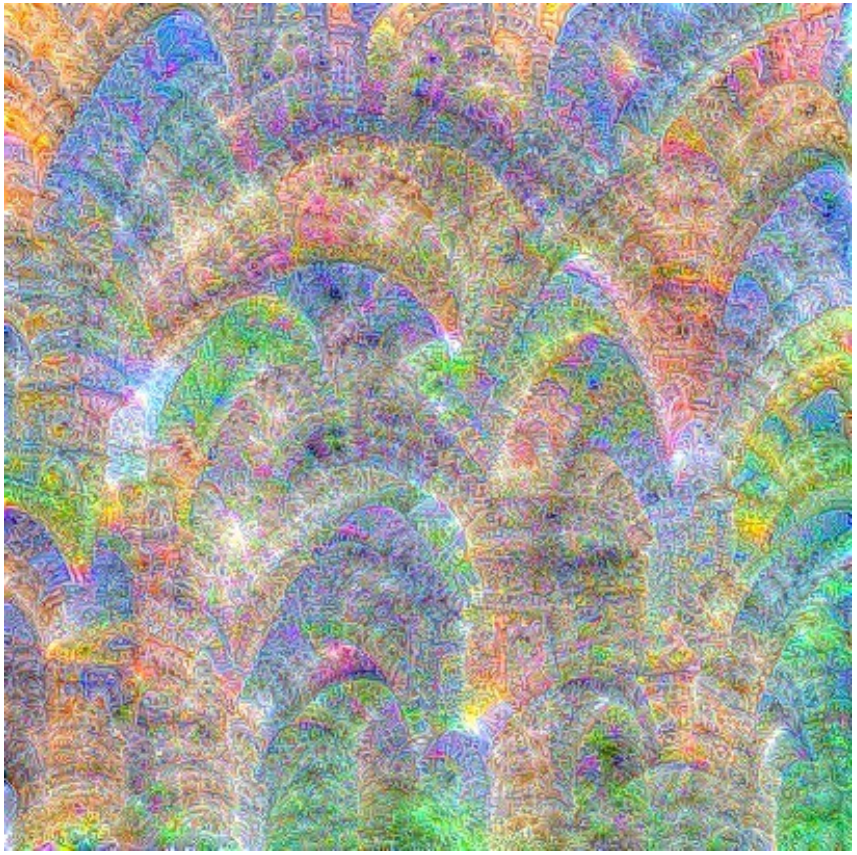
- Early filters are more abstract, later filters less so



What We Know About Learning in CNNs

- Early filters are more abstract, later filters less so





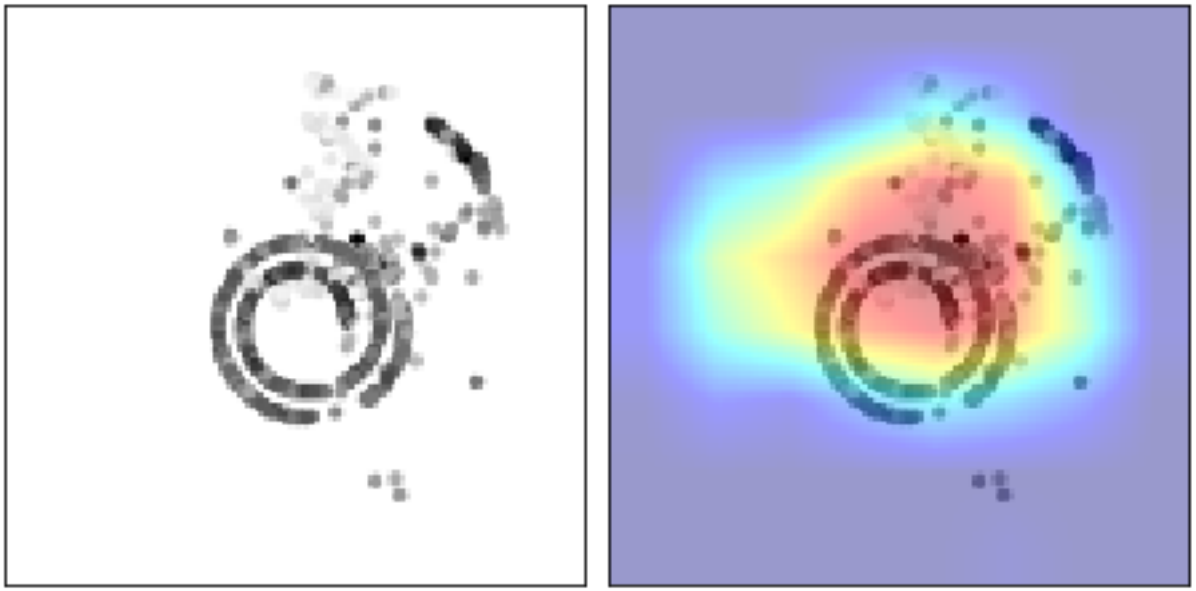
What We Know About Learning in CNNs

- Early filters are more abstract, later filters less so
- Deeper networks are better, but harder to train
  - “Vanishing gradient” problem
- “Skip connections” help
  - Most significant advance since AlexNet (2012)
  - Main enhancement in ResNet, also used by Google’s Inception
- Lessons from earlier:
  - Use ReLU, Adam, careful initialization, batch normalization

Getting Started with CNNs

- Many popular architectures are available in Tensorflow/Keras
- Have a lot of labeled data?
  - If you also have a lot of patience, you can train a CNN from scratch
  - Creative augmentation can help
- Have a lot of cheap, lower-fidelity data?
  - For example, simulations
  - Can train from scratch on simulated data, fine-tune on “real” data
- Not much data?
  - Use pre-trained model
  - Can be a feature extractor
  - Can be starting point for further tuning

Interpreting Predictions



Interpreting Predictions