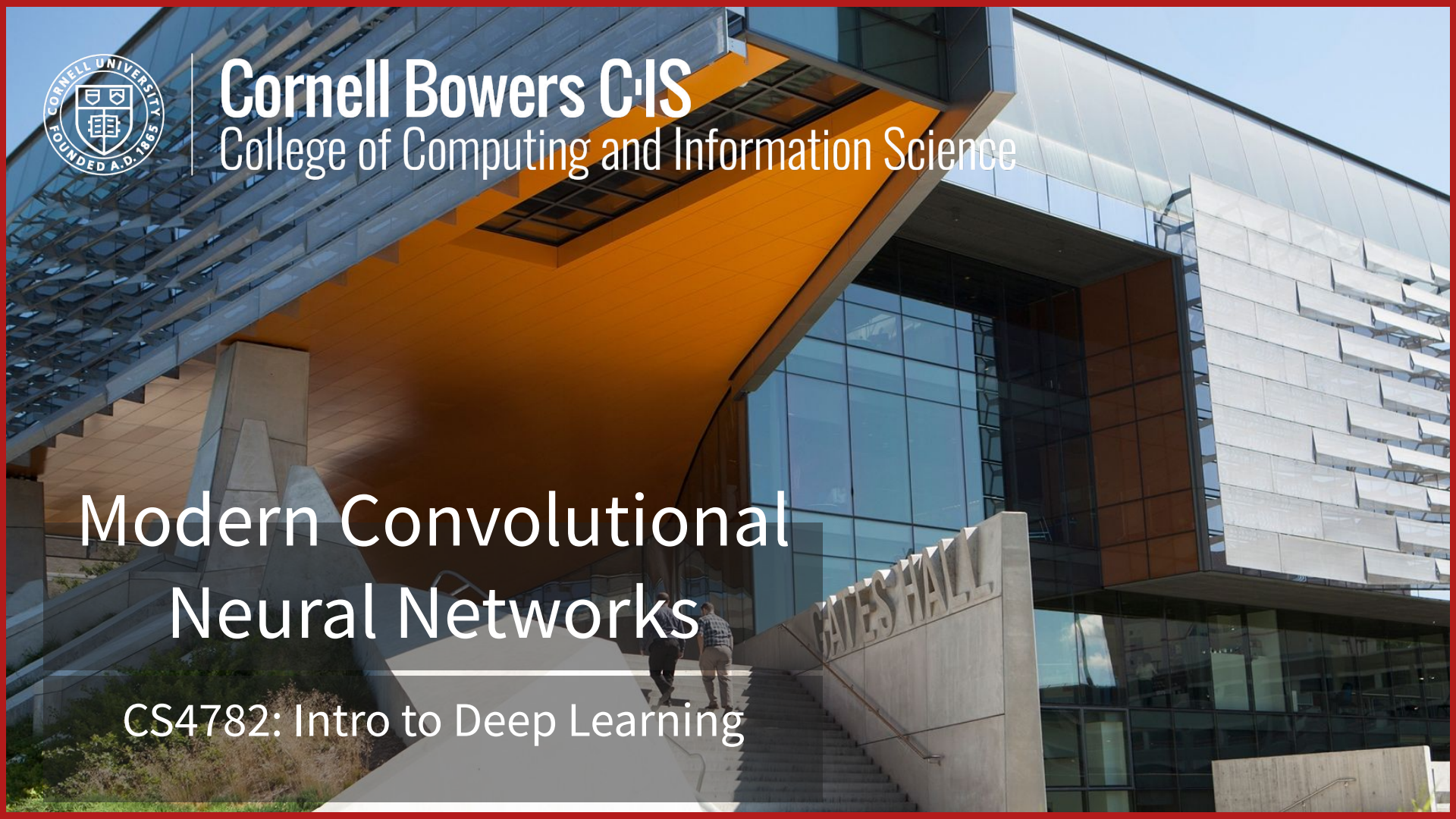


Modern Convolutional Neural Networks

CS4782: Intro to Deep Learning

Thanks to:

Varsha Kishore
Justin Lovelace
Anissa Dallmann
Stephanie Ginting

Logistics

- HW1 is due Thursday (February 13) 11:59 PM
- P1(Coding Assignment 1) is also due Thursday (February 13) 11:59 PM
- Late submissions accepted until Saturday()
- Office hours are listed on the course website
- Homework clarifications are listed as pinned posts under HW1 on Ed
- Post questions on Ed

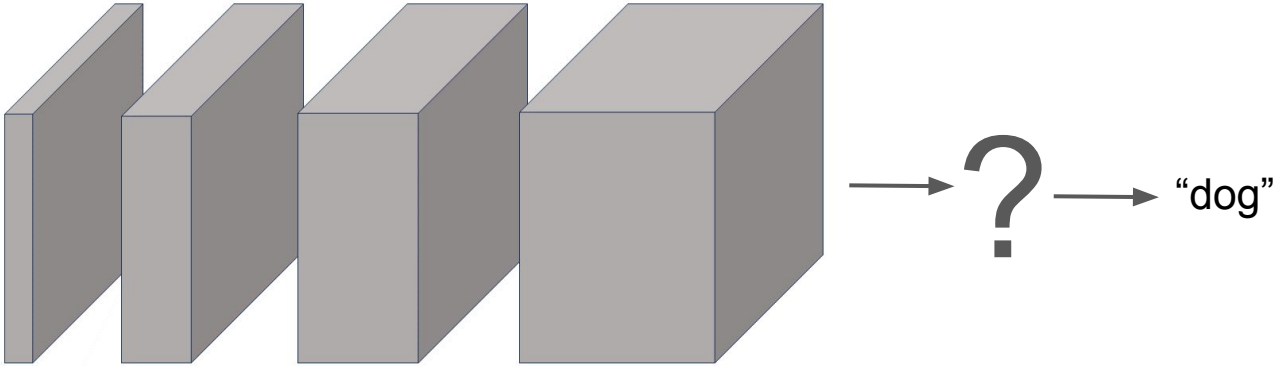
Review: Convolutional Neural Networks (CNNs)

✅ **Convolutions**

Maintain spatial relation between pixels

Reduce number of parameters through weight sharing

input image



Review: Convolutional Neural Networks (CNNs)

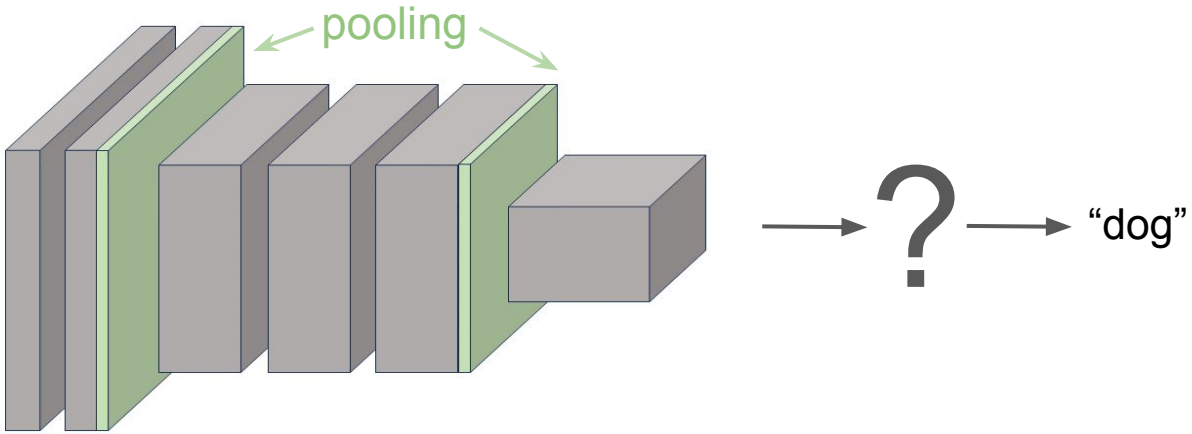
✅ Convolutions

Maintain spatial relation between pixels
Reduce number of parameters through weight sharing

✅ **Pooling**

Captures key information from across different areas of the feature maps
Together with convolutions allows for translational invariance

input image



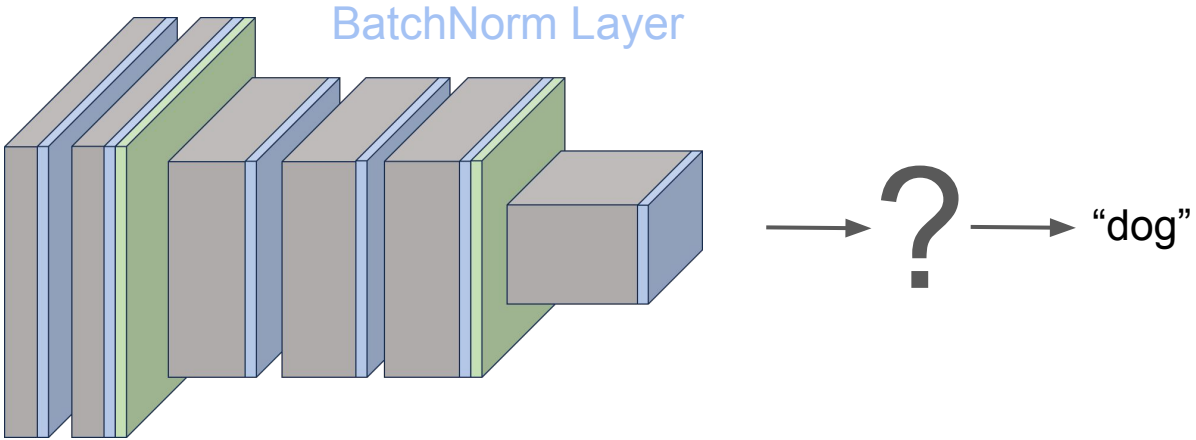
Review: Convolutional Neural Networks (CNNs)

✅ Convolutions — Maintain spatial relation between pixels
Reduce number of parameters through weight sharing

✅ Pooling — Captures key information from across different areas of the feature maps
Together with convolutions allows for translational invariance

✅ **BatchNorm** — Increases speed and stability of training

input image



Review: Image Classification

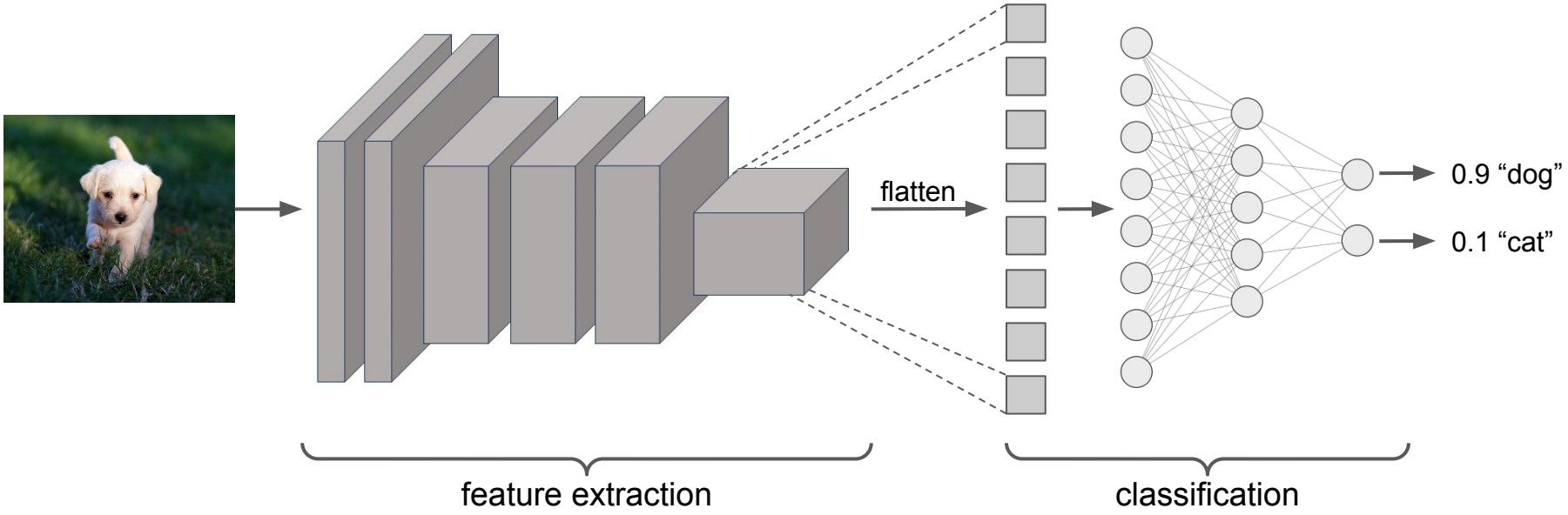
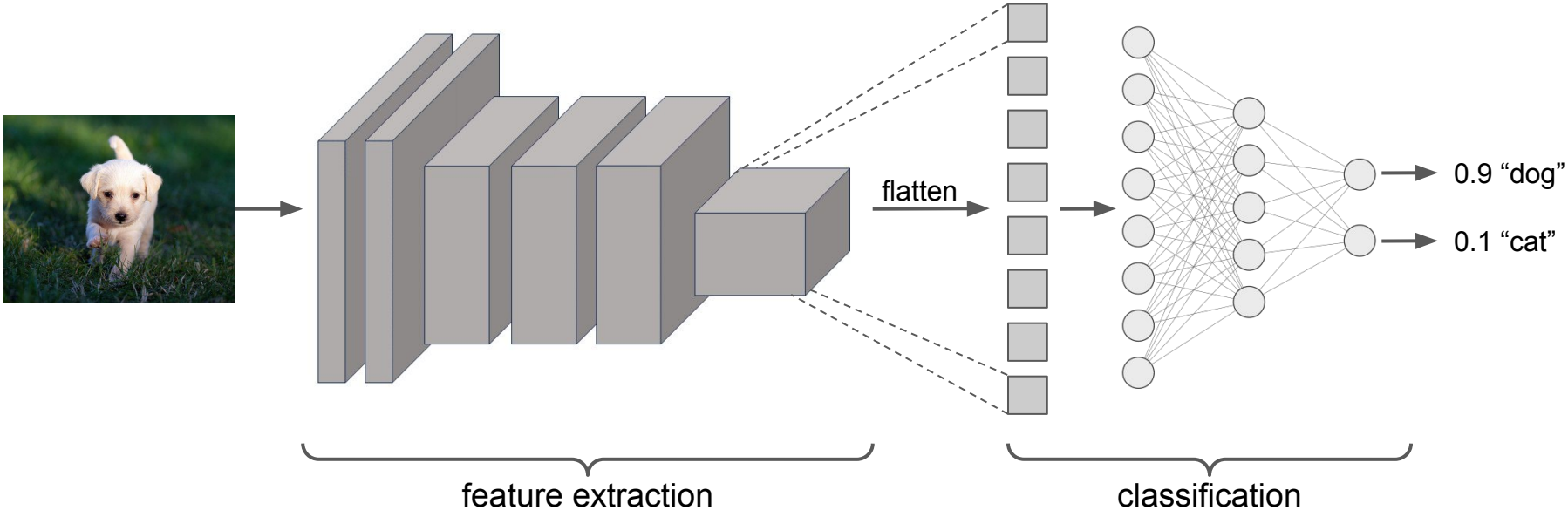


Image Classification

- Important: Everything is differentiable!
- Can calculate gradient of the loss with backpropagation
  - Train with SGD/Adam/etc.
  - Learn convolutional filters and classification head end-to-end!

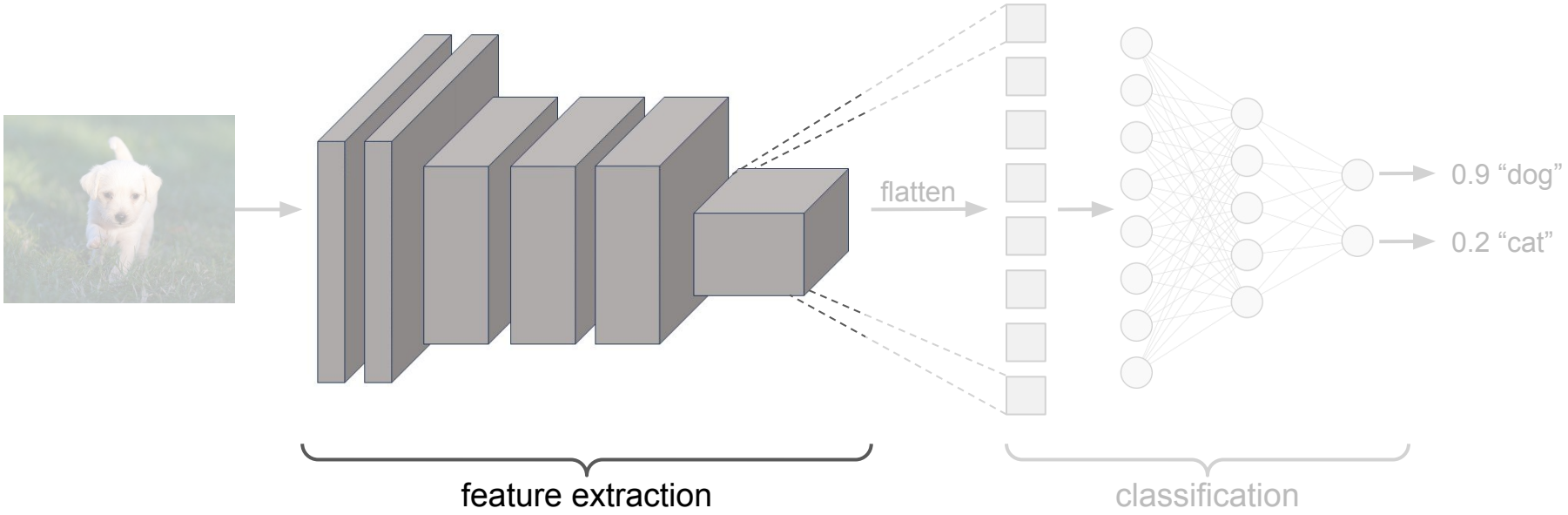


Discuss: Padding

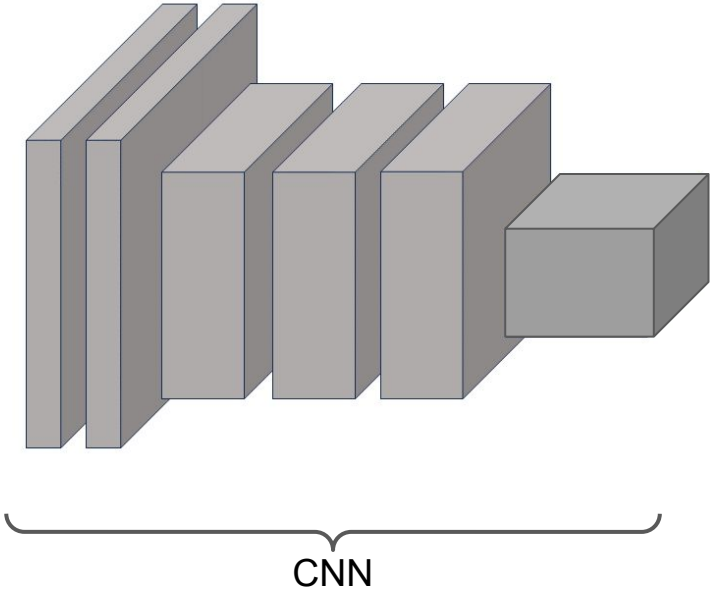
- Given a 5x5 feature map and a 3x3 convolution:
  - How much padding do I need to maintain the spatial size of the feature map (i.e., 5x5)?

- What about when using a 5x5 convolution?

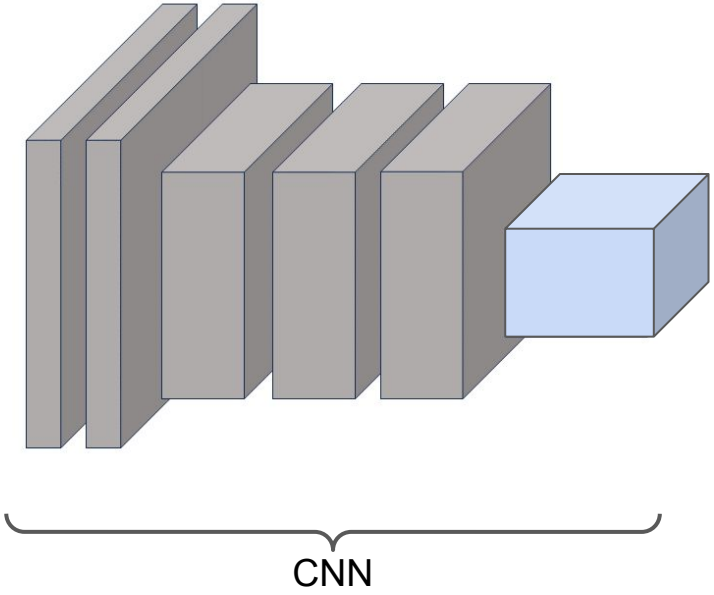
Deeper CNN Architectures

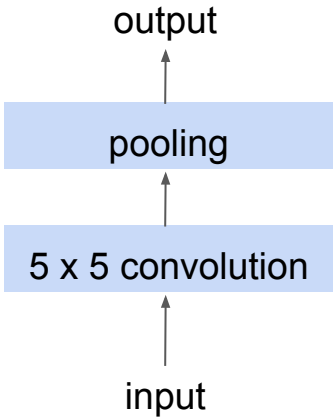


Deeper CNN Architectures

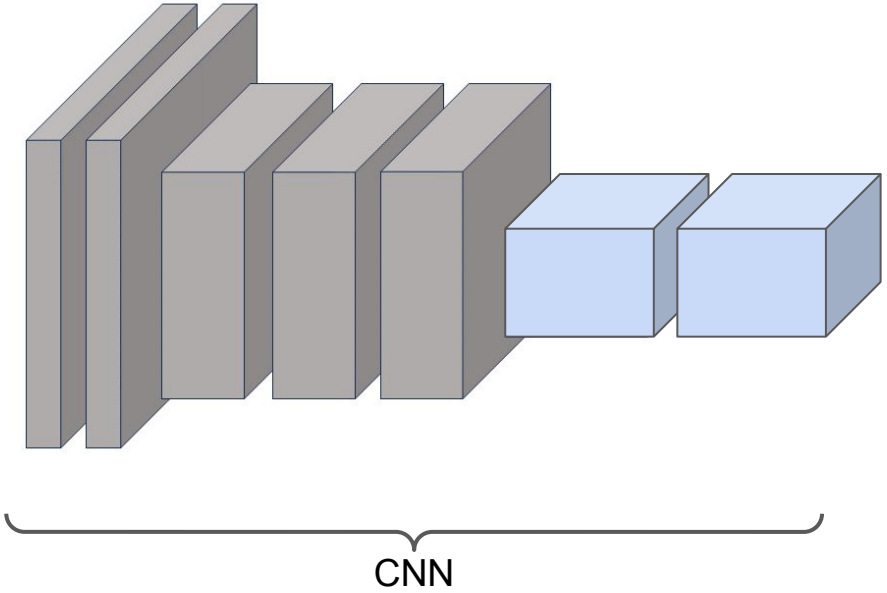


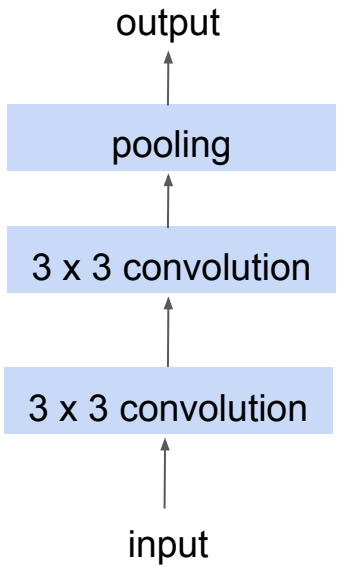
Deeper CNN Architectures



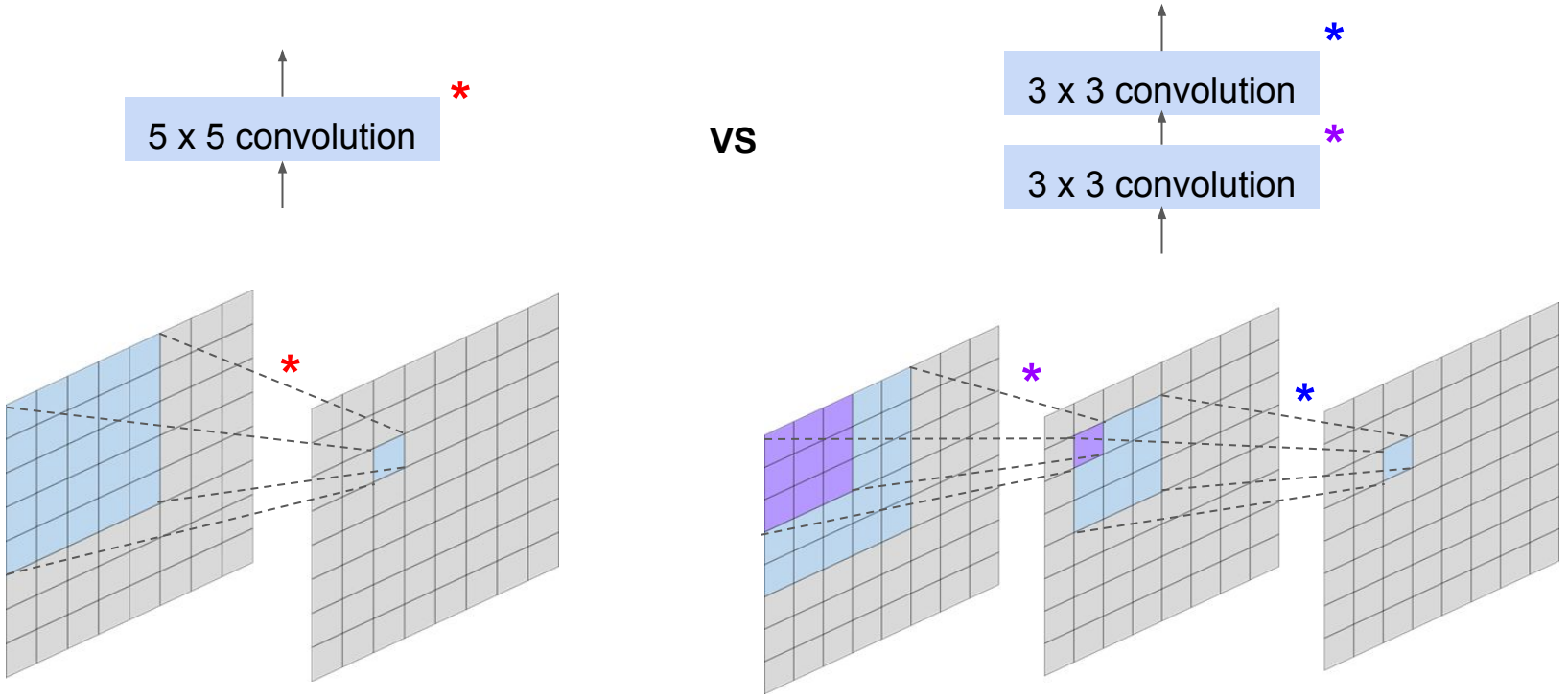


Deeper CNN Architectures

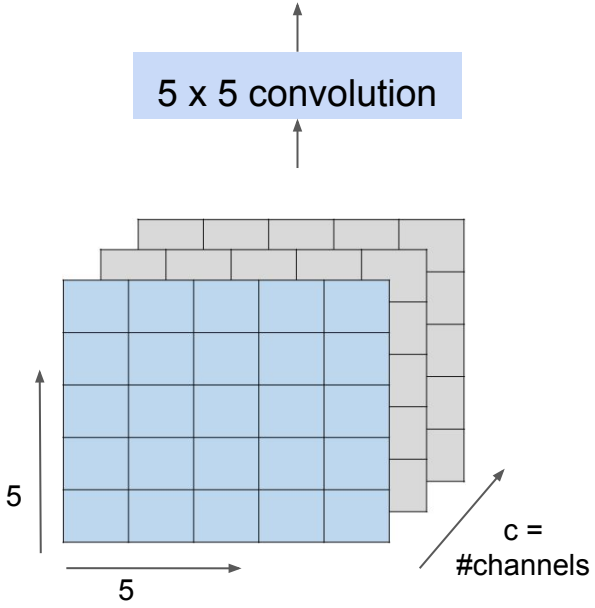




Deeper CNN Architectures

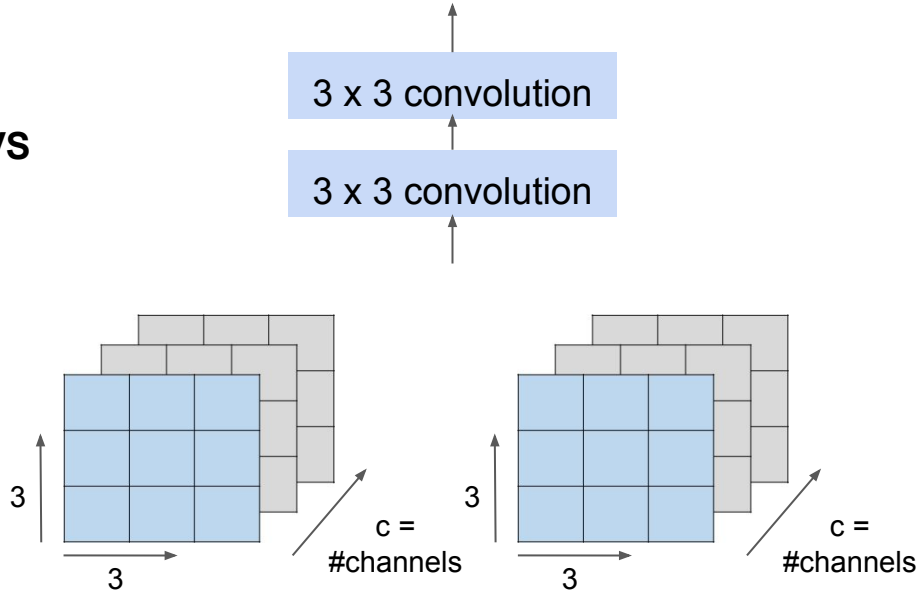


Deeper CNN Architectures



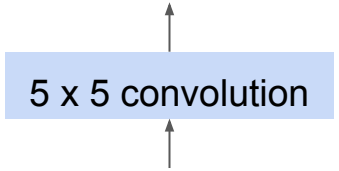
$5 * 5 * c^2$
$= 25c^2$ parameters

VS

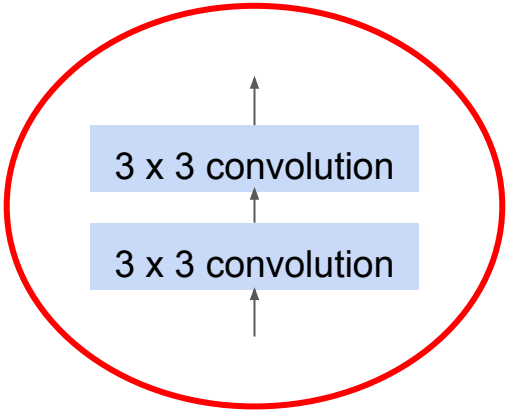


$2 * 3 * 3 * c^2$
$= 18c^2$ parameters

Deeper CNN Architectures

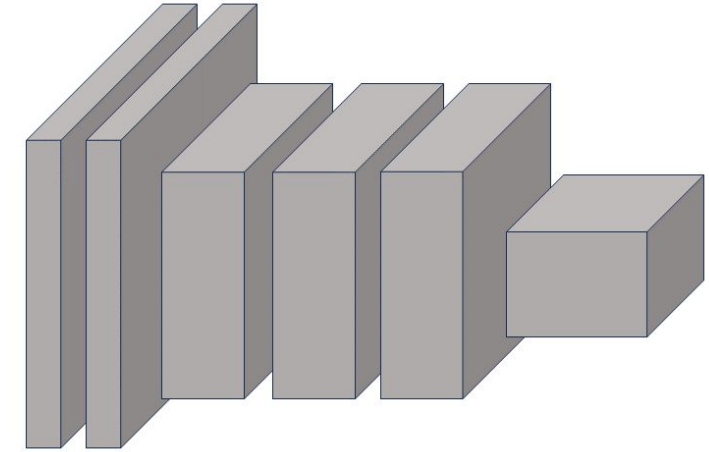


vs

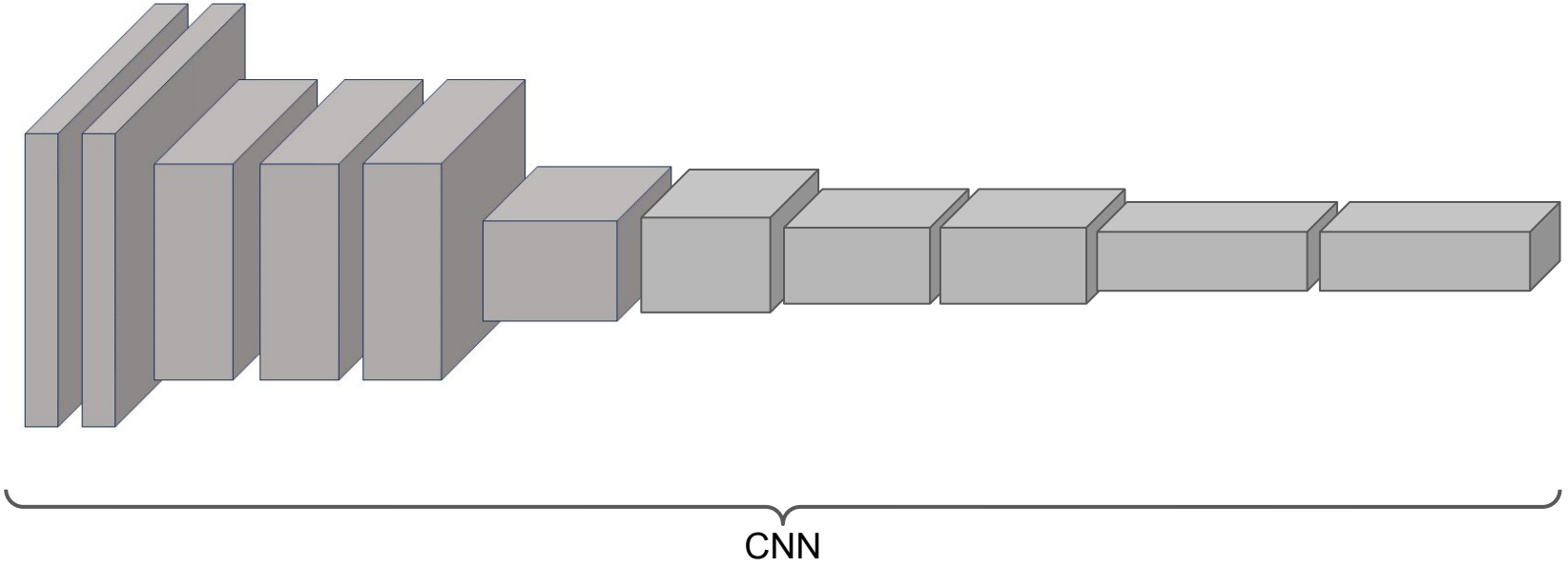


Performed better!

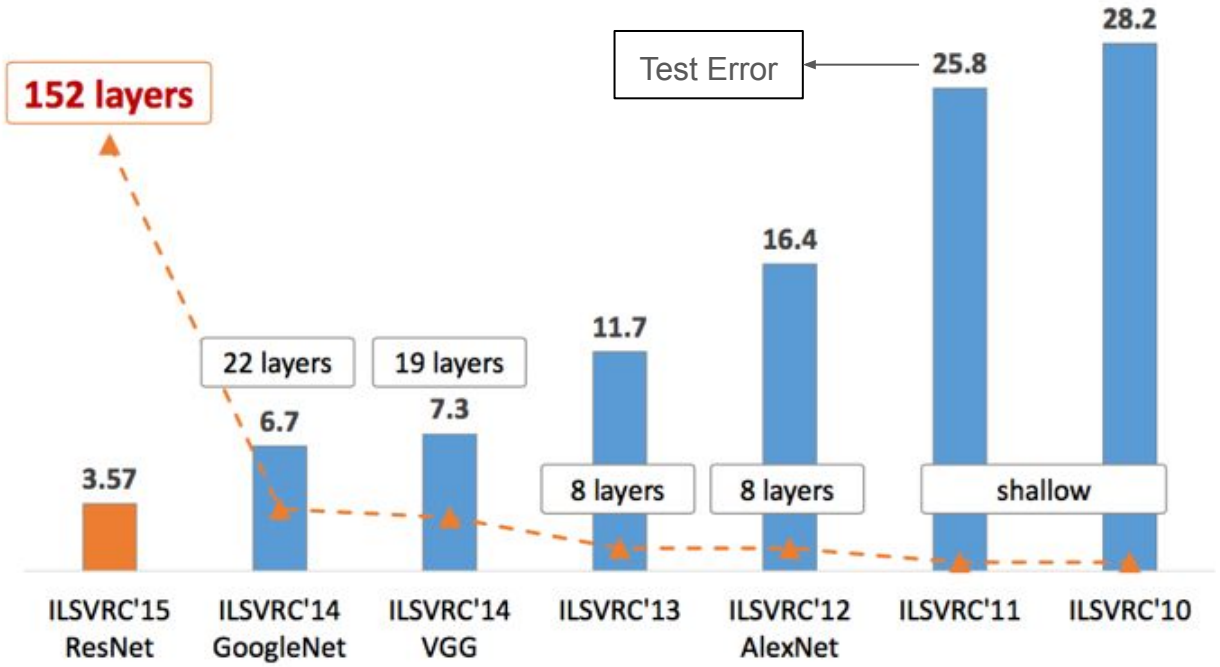
Deeper == better



Deeper == better

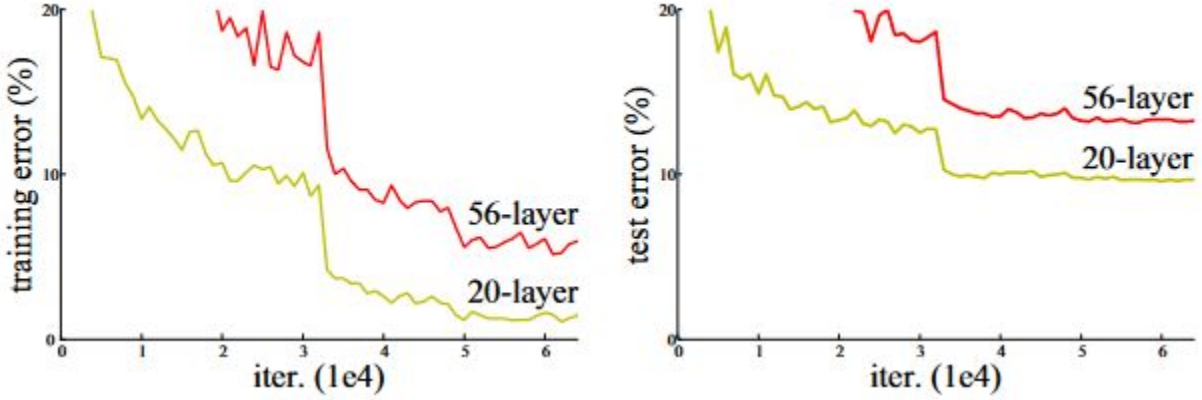


ImageNet Classification Challenge: Deeper == better



[Nguyen, Kien & Fookes, Clinton & Ross, Arun & Sridharan, Sridha. (2017). Iris Recognition with Off-the-Shelf CNN Features: A Deep Learning Perspective. IEEE Access. PP. 1-1. 10.1109/ACCESS.2017.2784352.]

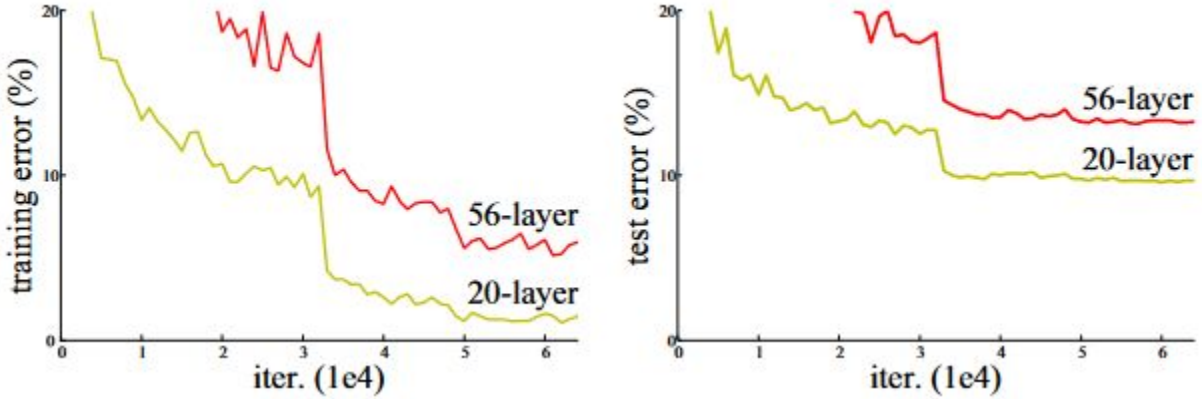
Deeper == better?



56 layer CNN has higher training and test error than 20 layer CNN on CIFAR-10 dataset for image classification

[He, Kaiming, et al. "Deep residual learning for image recognition." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2016.]

Discuss: How can a larger network achieve a higher training error?



56 layer CNN has higher training and test error than 20 layer CNN on CIFAR-10 dataset for image classification

Deeper != better

- Long training times

- Vanishing gradient problem
  - Recall backpropagation to update weights

$$\frac{\partial z}{\partial z_i} = \frac{\partial z}{\partial z_{n-1}} \frac{\partial z_{n-1}}{\partial z_{n-2}} \cdots \frac{\partial z_{i+1}}{\partial z_i}$$

  - If each term <<< 1, gradient “vanishes” as the entire multiplication goes towards 0
  - => Weights not updated properly