MACHINE LEARNING IN PHYSICS
CONVOLUTIONAL NEURAL NETWORKS

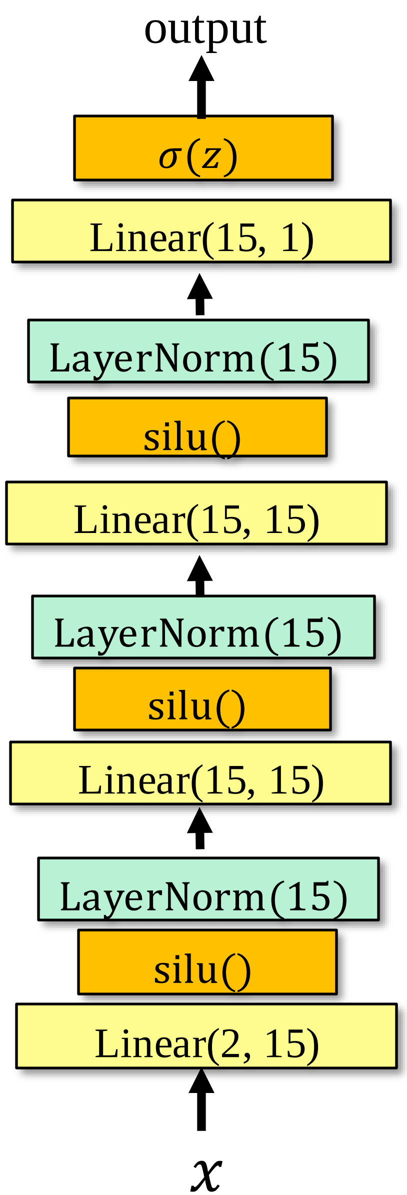
HARRISON B. PROSPER

PHY6938

Recap

We used the simplest machine learning architecture, namely a multi-layer perceptron (MLP) or feed-forward neural network (FFNN), such as the one shown here, to review a few key concepts in ML, including

1. the loss function,
2. the empirical risk (or average loss),
3. the risk functional and the equation $\int dy \, \frac{\partial L}{\partial f} p(y \mid x) = 0$,
4. and empirical risk minimization via stochastic gradient descent.



Introduction

The plan for the next several lectures is to introduce the following machine learning models:

1. Convolutional neural networks (CNN)
2. Autoencoder (AE)
3. Physics-informed neural networks (PINN)
4. Flow and diffusion models
5. Graph neural networks (GNN)
6. Transformer neural networks (TNN)

We shall try to understand each class of models using a simple example.

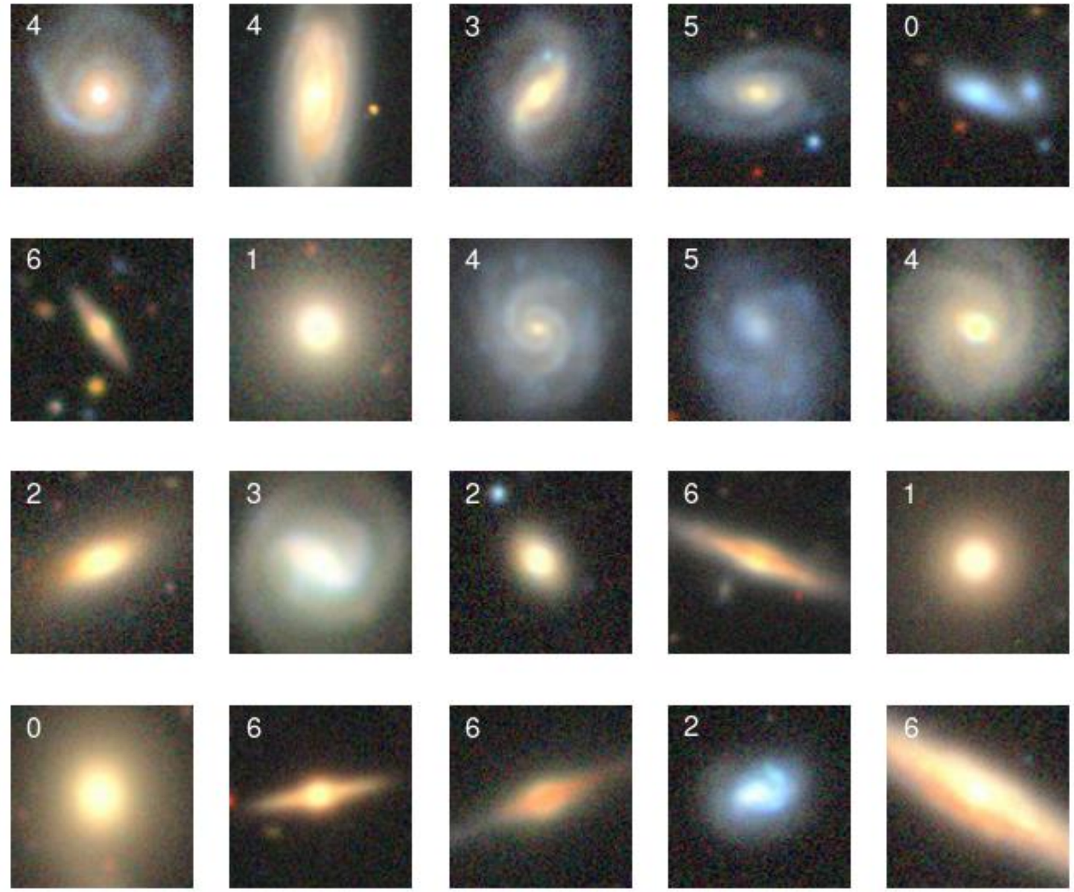
Introduction

- A great deal of our behavior is guided by our ability to interpret visual data in real time.
- This ability is now available in everyday machines. Indeed, there exist image recognition systems that are superhuman in their abilities.
- The breakthrough that has allowed such advances is the convolutional neural network (CNN) and its many, many, variations.
- Our focus this week is CNNs for image classification.

Introduction

Here are images of 20 galaxies from the Galaxy 10 DECals Dataset at the astroNN website.

Goal: classify galaxies using images like these.

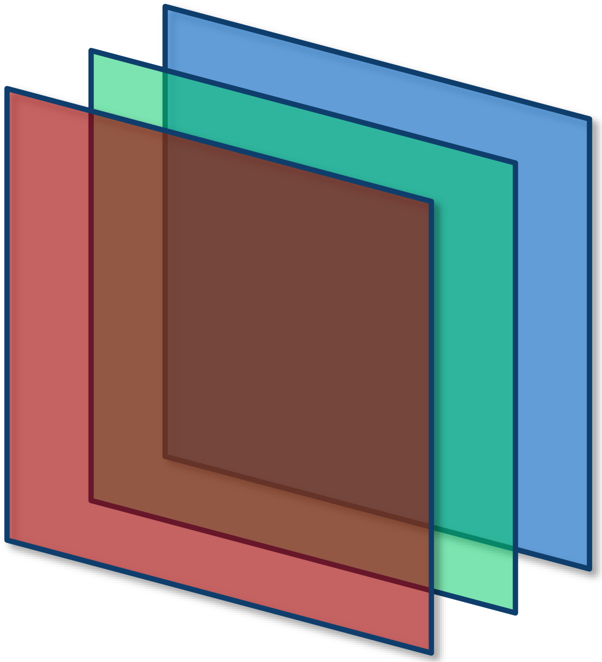


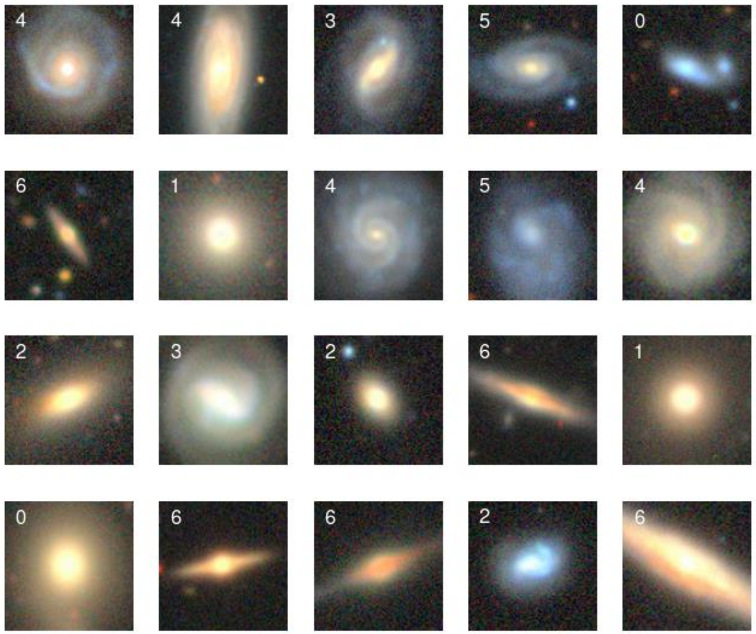
https://astronn.readthedocs.io/en/stable/galaxy10.html

Introduction

Dataset

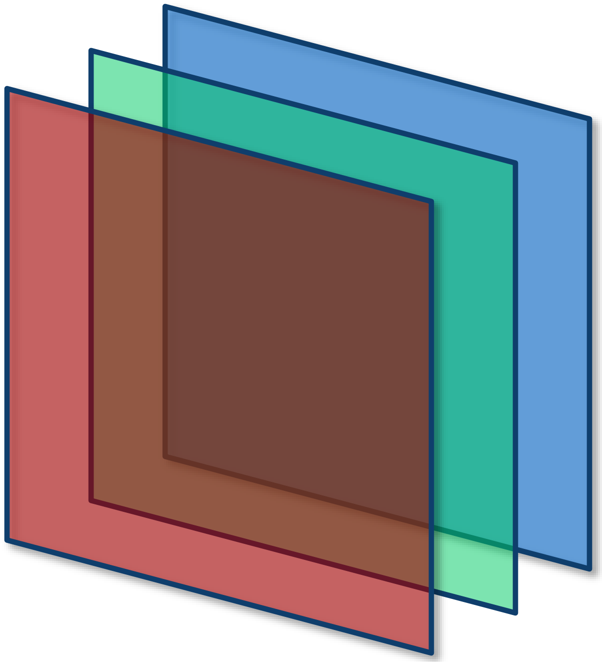
1. We'll use a subset comprising 12,600 images divided into 7 galaxy morphology classes with 1800 galaxies per class.
2. Each 3-channel image is cropped to 96 x 96 pixels.

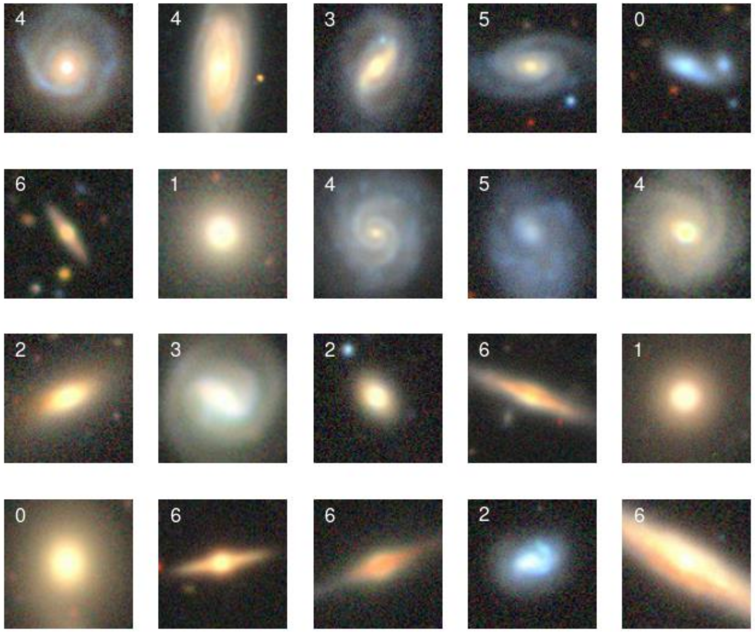




Introduction

Since this is a multi-class problem, we cannot use the binary cross entropy loss. Instead, we use its generalization, called the cross-entropy loss.





CROSS ENTROPY

What is Cross Entropy?

We start with a very important quantity from mathematical statistics, namely, the Kullback-Leibler (KL) divergence

$$D(p,f) = \sum_x p(x) \log\left(\frac{p(x)}{f(x)}\right)$$

between two probability distributions $p(x)$ and $q(x)$.

Key Properties:

- $D(p,f) \geq 0$
- $D(p,f) = 0$ if and only if $f(x) = p(x)$.

What is Cross Entropy?

The KL divergence

$$D(p, f) = \sum_x p(x) \log\left(\frac{p(x)}{f(x)}\right)$$

can be rewritten as

$$D(p, f) = -\sum_x p(x) \log f(x) - \left[-\sum_x p(x) \log p(x)\right]$$

We recognize the term in brackets as the entropy $H(p)$ of the distribution $p(x)$.

The first term is called the cross entropy $H(p, f)$. Notice that $H(p, p) = H(p)$.

Cross Entropy Loss

Our goal is to classify galaxies into one of $K = 7$ classes according to their morphology.

To do so, we need a function $f_k(x)$ that approximates the class probabilities

$$f_k(x) \approx p(k|x) = \frac{p(x|k)\,\pi(k)}{\sum_{j=0}^{K-1} p(x|j)\,\pi(j)}$$

where $p(k|x) \equiv p(y = k|x)$, x is a galaxy image, $p(x|k)$ is the probability distribution of galaxy images for morphology class k, and $\pi(k) = \pi(y = k)$ are the class priors.

Cross Entropy Loss

Assignment 2

Every image x is associated with a class label y. Show that the average of the loss function

$$L(y, f\,) = -\log f_y(x),$$

over an infinite about of data, where $y \in [0, \dots, K-1]$ yields

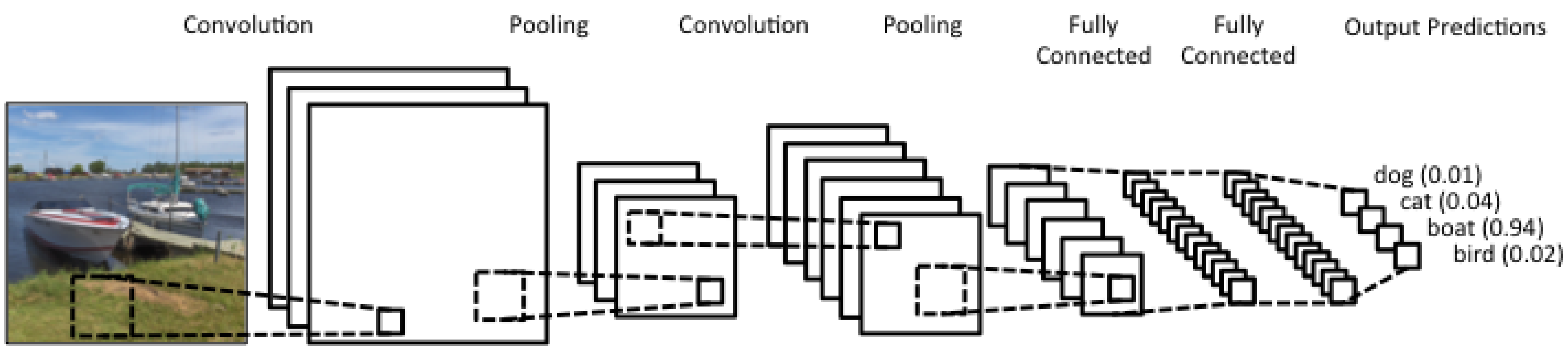
$$f_k(x) \approx p(k|x) = \frac{p(x|k)\,\pi(k)}{\sum_{j=0}^{K-1} p(x|j)\,\pi(j)}$$

CONVOLUTIONAL NEURAL NETWORKS (CNN)

Convolutional Neural Network

- Convolutional neural networks (CNN) are *functions* that code 2D (or 3D) data and classify objects using their coded representations via a fully-connected neural network.

- The key insight that underpins CNNs is the approximate translational invariance of natural images.



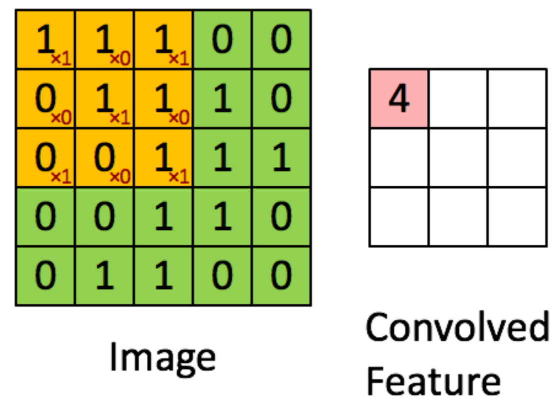
Source: https://www.clarifai.com/technology

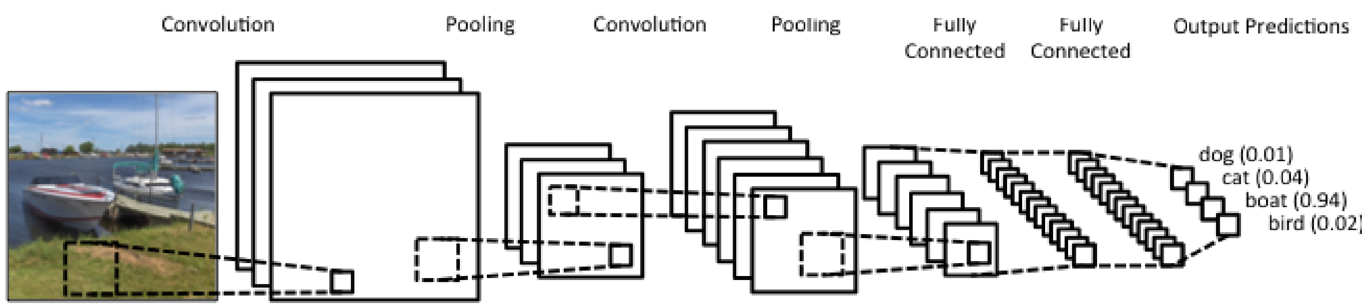
CNN

A standard CNN comprises three types of processing layers:
1. convolution, 2. pooling, and 3. classification.

1. Convolution layers

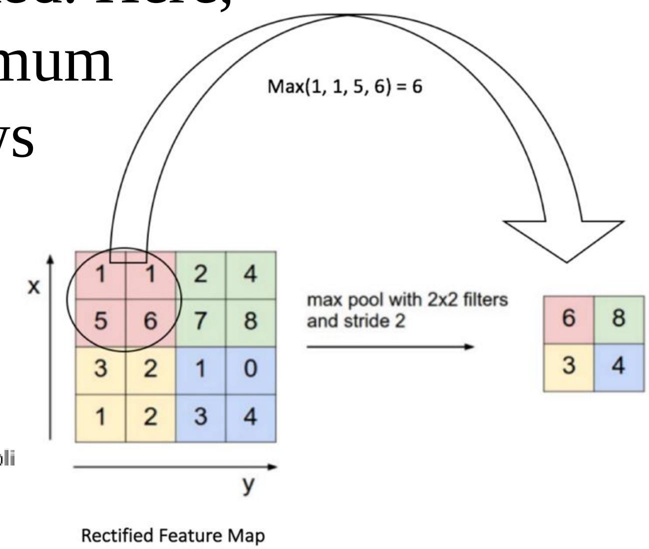
The input to a layer is *convolved* with one or more matrices using element-wise products that are then summed. In this example, the convolution compresses the image from 5 x 5 to a to a 3 x 3 matrix.

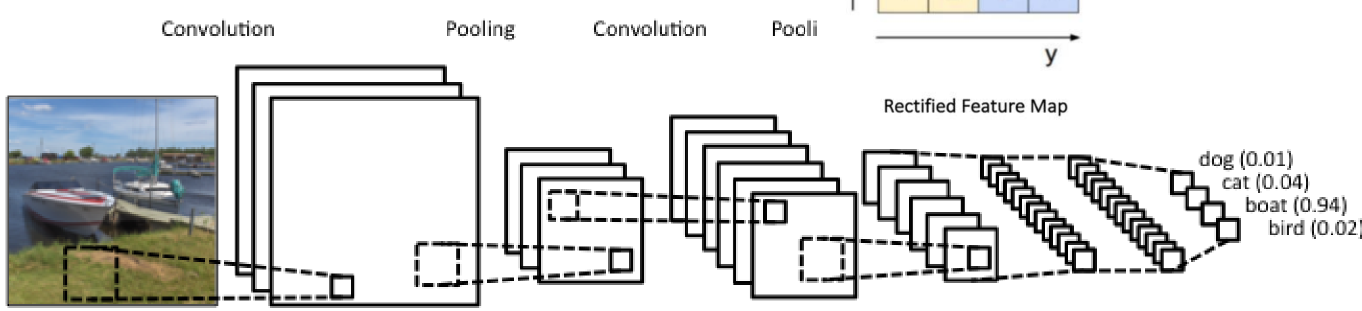




CNN

2. **Pooling Layers**
   After convolution, and a pixel-by-pixel non-linear map (using, e.g., the function $y = \max(0, x) = \mathrm{ReLU}(x)$), a coarse-graining of the layer is performed. Here, called max pooling in which the maximum values within a series of small windows are selected and become the output of a pooling layer.



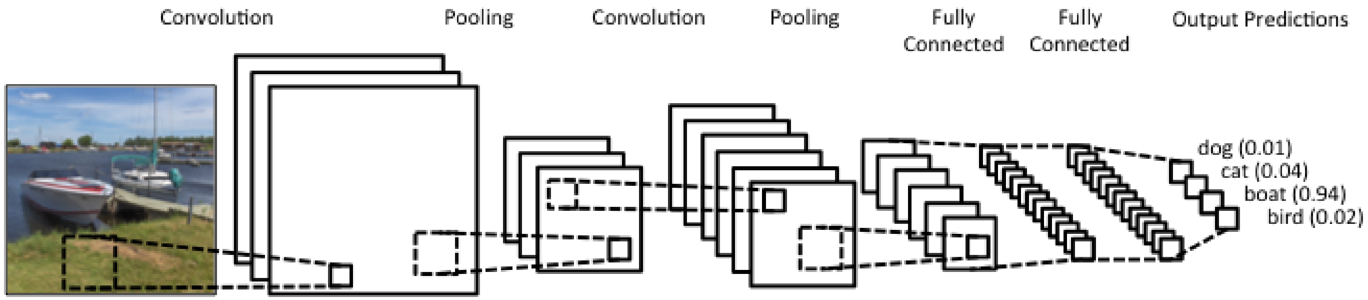


CNN

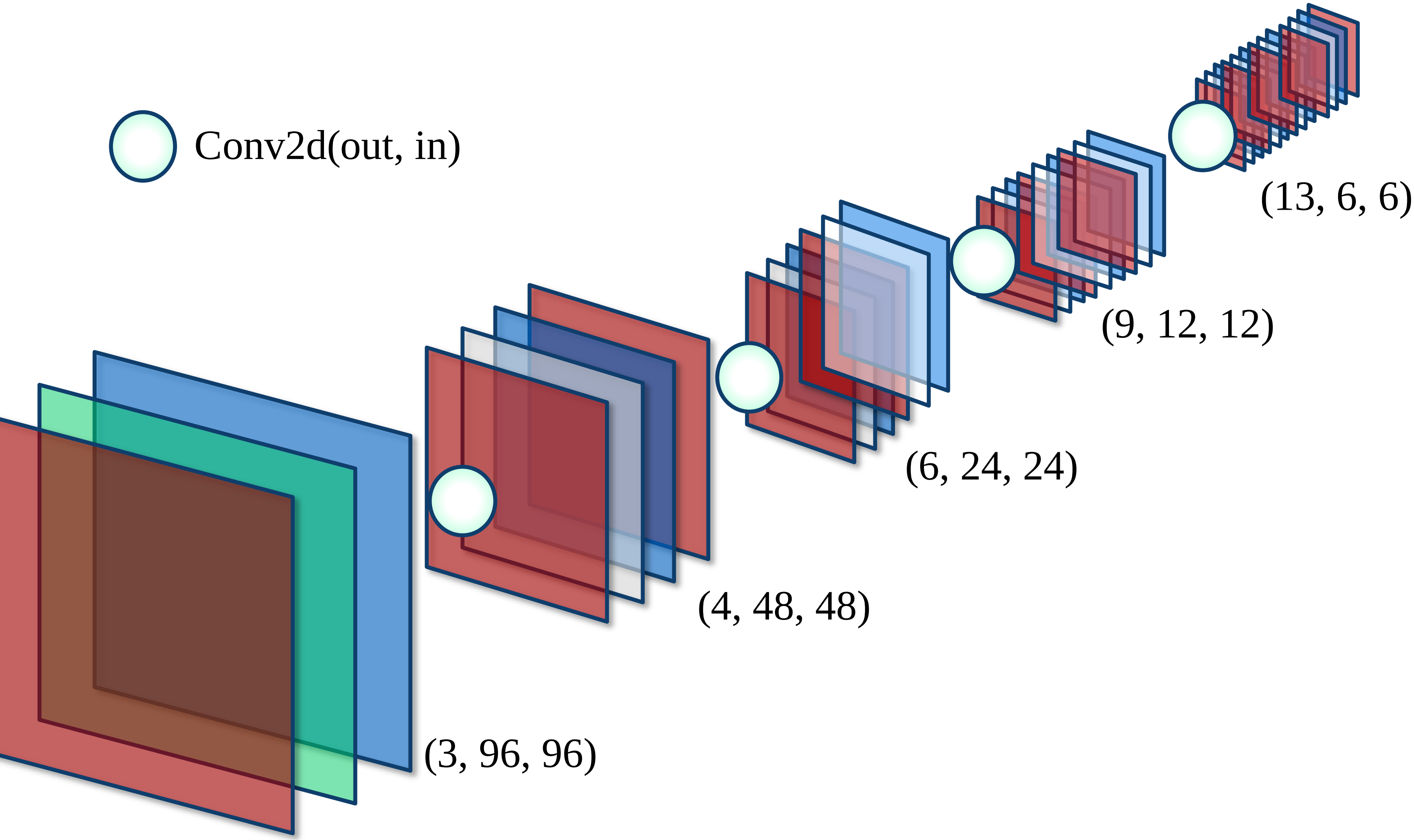
3. Classification Layers

After an alternating sequence of convolution and pooling layers, the outputs go to a fully-connected neural network whose outputs approximate the class probabilities:

$$p(k|x) = p(x|k)\pi(\mathrm{k}) / \sum_{j=0}^{K-1} p(x|j)\pi(\mathrm{j})$$



Tutorial 4: Our Model



Summary

- Convolutional neural networks have revolutionized image interpretation and are now used in many applications.

- The model exploits the fact that natural images are often locally approximately translation invariant.

- The convolutions automatically extract features from images that can later be used to interpret them.