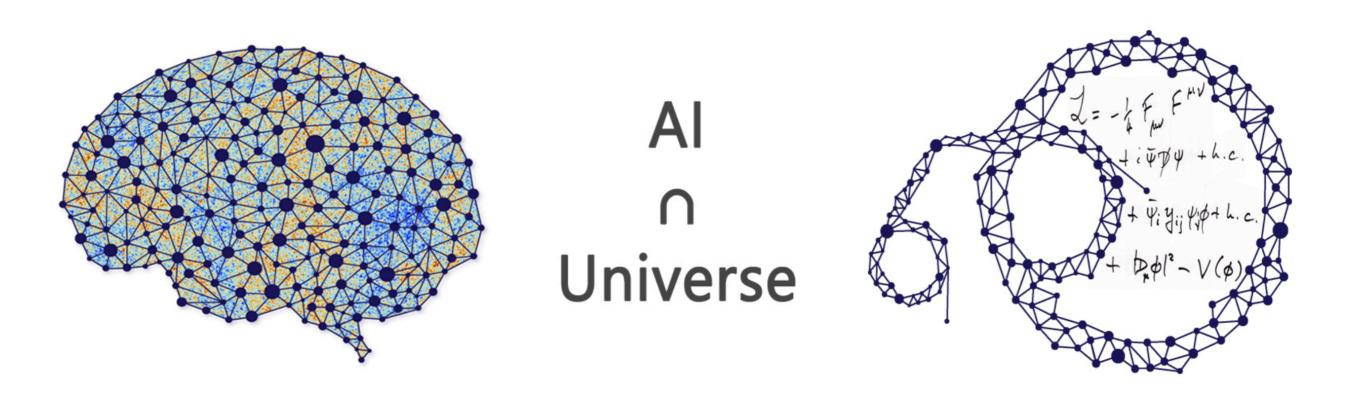
PHY 835: Machine Learning in Physics

Lecture 10: Convolutional Neural Network

February 22, 2024



Outline for today

- Convolutional neural networks (CNNs)
- Convolutional Layer and Pooling layer
- Workflow for Deep Learning

References: Deep Learning Book, 1803.08823

Stanford CS23 (Andrej Karpathy & Fei-Fei Li): https://cs231n.github.io

training works best when imputs are centered around gers with respect to bies Observation: My for activations like tent, signoid y neurons not saturated & gradients not vanishing good training such points

Batch Normalization: additional layers which stand by the mean and variance of m

layer litte d'unions : yt (21, --, 21) Example:

$$\overline{z}_{i} \rightarrow \hat{z}_{i} = \frac{\overline{z}_{i} - \overline{w}_{i}}{\sqrt{Var(\overline{z}_{i})}}$$

might charge representational power of NN Problem: pieces are pided out.

> 4) Sol: 2: + 2: + B; this shifts normalised values back to so.

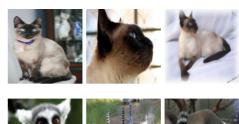














Synset: tiger cat

Definition: a cat having a striped coat. Popularity percentile:: 78% Depth in WordNet: 8

Synset: lesser panda, red panda, panda, bear cat, cat bea

Definition: reddish-brown Old World raccoon-like carnivor giant pandas.

Popularity percentile:: 68% Depth in WordNet: 12

Synset: Egyptian cat

Definition: a domestic cat of Egypt. Popularity percentile:: 67%

Depth in WordNet: 8

Synset: Persian cat

Definition: a long-haired breed of cat. Popularity percentile:: 59% Depth in WordNet: 8

Synset: tabby, tabby cat

Definition: a cat with a grey or tawny coat mottled with bla

Popularity percentile:: 58% Depth in WordNet: 8

Synset: Siamese cat, Siamese

Definition: a slender short-haired blue-eyed breed of cat h

Popularity percentile:: 57% Depth in WordNet: 8

Synset: Madagascar cat, ring-tailed lemur, Lemur catta

Definition: small lemur having its tail barred with black.

Popularity percentile:: 45% Depth in WordNet: 12

Advantage: improves learning speed, acts as regularizer

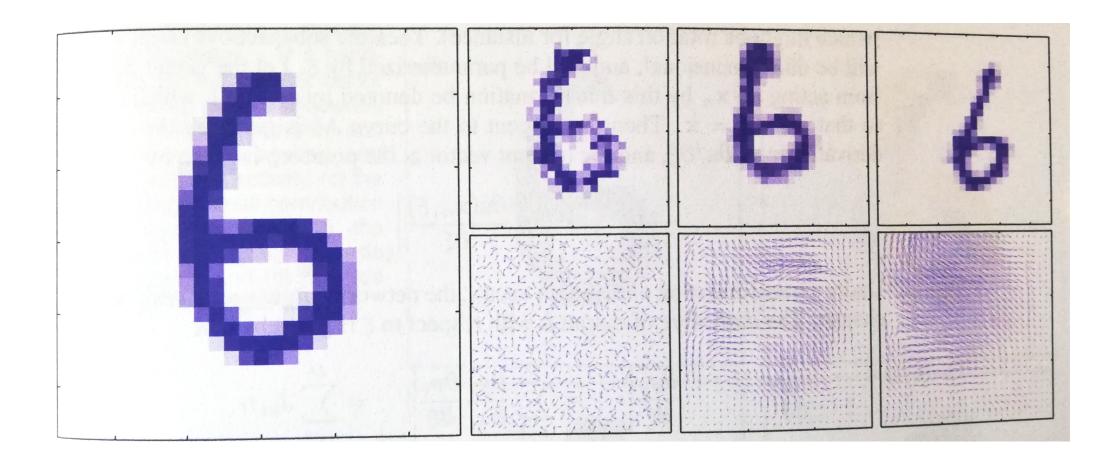
Learning with Symmetries



- Locality: features that define a "cat" are local in the picture: whiskers, tail, paws, ...
- Translational invariance: Cats can be anywhere in the image.
- Rotational invariance: Relative position of features must be respected (e.g. whiskers and tail should appear on opposite sides)
- Our classifier should exhibit all these high-level structures.

Learning with Symmetries

Consider classification of digits:



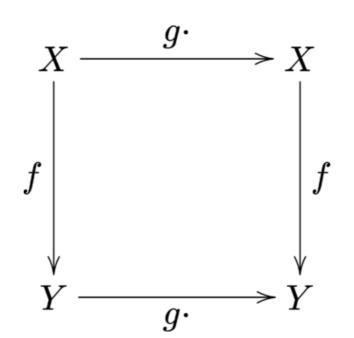
What symmetries should be built-in in ML classifiers?

Translation, scaling, small rotations, smearing, elastic deformations.

 $\mathbf{0} = \nabla \mathcal{C}(\mathbf{w}) = \sum_{i=1}^{n} \left[\sigma(\mathbf{x}_{i}^{T}\mathbf{w}) - y_{i} \right] \mathbf{x}_{i}, \text{ usually supplement the Symmetries of binary rlussification Symmetries of discussion of these in the second states of the second states of the second states are supplementation.$ Lewherewshowade use sefut hetiogistics function identity is rection train for the marketien. Of subjection has been been as the company of th //phyniamerical-methods/such asothose lintroduced hin Sections by local interactions. Think applit Off this cost further sections. 7. relativity indestatistical physics: D Ising model
The goal of this example is to show now one can employ logistic the gray resident. according to their phase of matter. The Hartilus no vos bow lasowatos ingentogistic n^za tegistic regressor to classify bin nkaj m/Minotebool merical methods suc Mehta, M. Bukov, C.-H. Wang et $+ | \mathcal{D}_{\varphi} |^2 - V(\phi)$ te indices i, j run over all near thermodynamic final fibils af x or all leg ferfors agree 10 TVisience the itempeters in the artificial terms of the confusion of the c Matter of the state of the stat 20 of the Ising model. If successful, this can be used 30 where an exact analytical sostion has sostate rem other words, given an ising state, we would like to classify writing it being

Locality and Symmetries

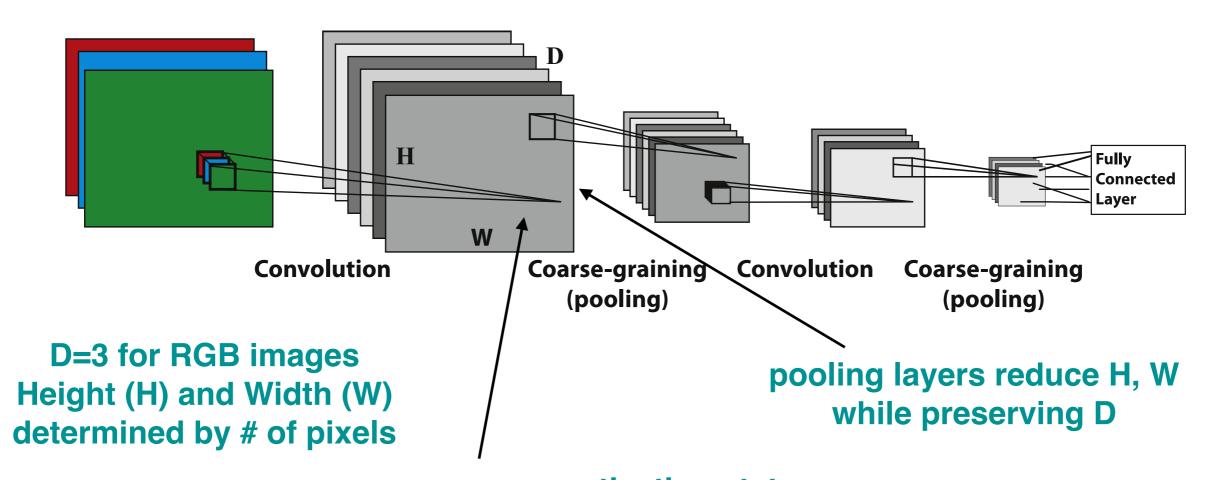
- Symmetries are at the heart of physics. For example, translation invariance allows to work in momentum space → less parameters
- In relativity and quantum field theory, Poincare-symmetry (translations, rotations, boosts) is essential.
- Gauge symmetries are ubiquitous in QFT and gravity. Equivariant CNNs (Cohen, Welling 2016). We will come back to this...
- f(x) is equivariant if we change the input in a particular way as $x' = g \cdot x$, the output changes in the same way: $f(g \cdot x) = g \cdot f(x)$:



- The simplest approach would be to input the images to a fully connected NN which given enough training data (and time) would learn the symmetries by example.
- However, a crucial property is ignored: nearby pixels are strongly correlated we should aim instead first to identify local features that depend on small subregions.
- For example, treating the spin configuration of the 2d Ising model as a $L \times L$ dimensional vector (L = number of sites in each linear direction) throws away spatial information (e.g., domain wall)
- Convolutional Neural Networks (CNNs) are architectures that take advantage of this additional high-level structures that all-to-all coupled networks fail to exploit.

A CNN is a translationally invariant neural network that respects locality of the input data.

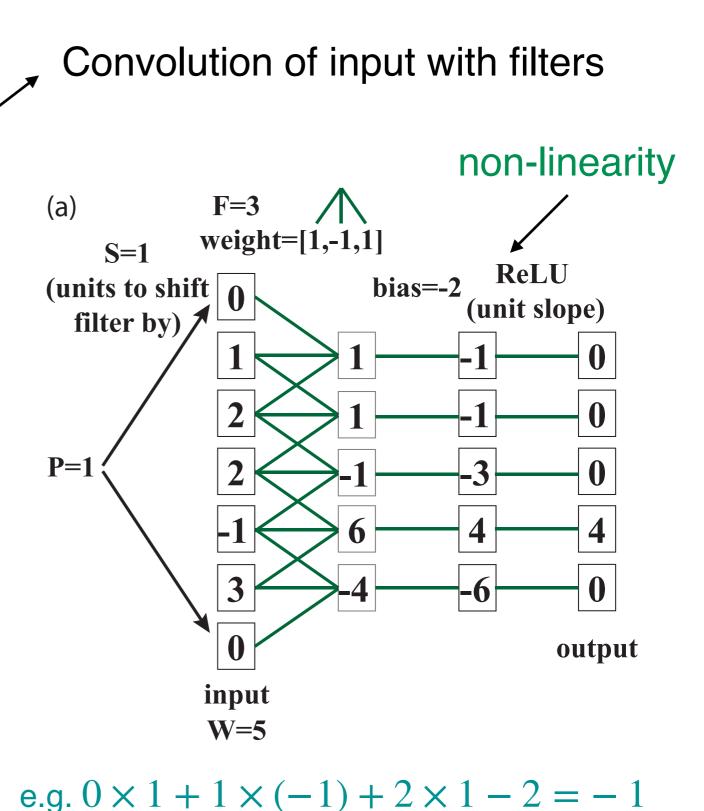
Depth: number of input channels (not depth of neural network)



neuron activation state: convolution with local spatial filter (e.g., 3 x 3 pixel grid)

CNNs are composed by **two** kinds of layers

example of convolutional layer



F=receptive field size of the Conv Layer neurons

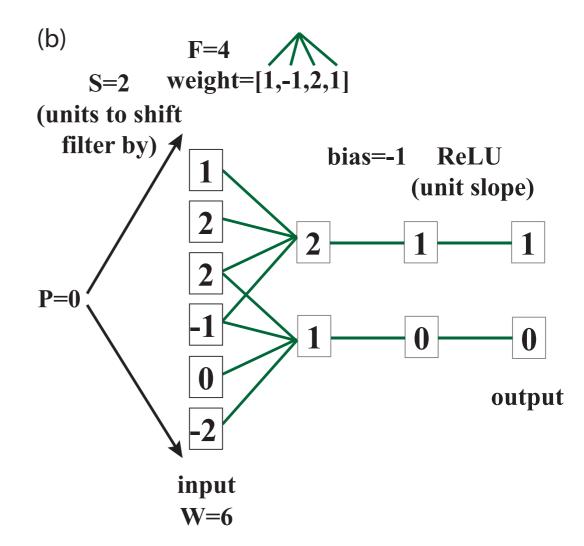
another example (with no padding)

S=stride

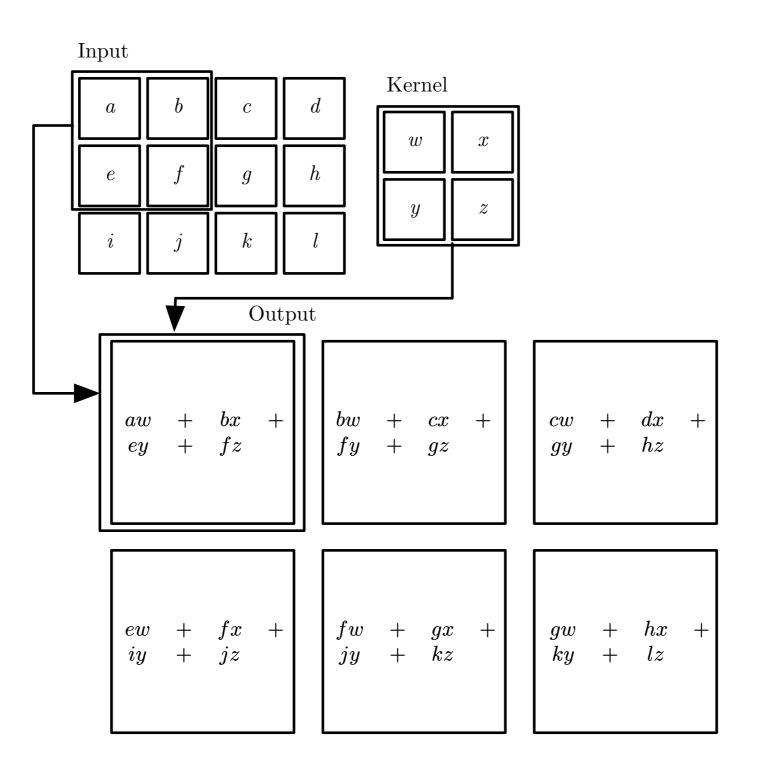
P=amount of zero padding on the border

Number of neurons (outputs) in the layer:

(W-F+2P)/S+1



2d Convolution



nassinate of the spaceship is position, we would nike to average the several nassinate the several property of the position of the spaceship.

We example such that the sense of the position of the spaceship.

CHAPTER 9. CONVOLUTIONAL SET WORKS
$$(t) = \int_{s(t)}^{s(t)} \int_{$$

This operation is sentially as the convolution of t In oumputemplieneuwile tees discretized valid probability wile provide unation, regular in our example, worders to be puvalide the bability density function, or the output list notes. We green see it ansign whereas to be the provide a segretive arguments, output is not a weighted une halo. Also, w needs to be 0 for all negative arguments, or it will igoes into the confinese. initations are barticular to the exhibiting presumably beyond our capabilities. These limitations for which the above integral is defined, and may be used for other ∞ d may be used for other our one functions for wh purposes besides taking $s(t) = (x * w)(t) = \sum_{a = -\infty} x(a)w(t - a)$ in this example, the function convolutional network resources received to as our entropt (in this example, the function of the learning lupide as on the formed at the second the second

of data and the kernel is usually a multidimensional array of parameters that are

adapted by the learning algorithm. We will refer to these multidimensional arrays

Convolution of input with filters

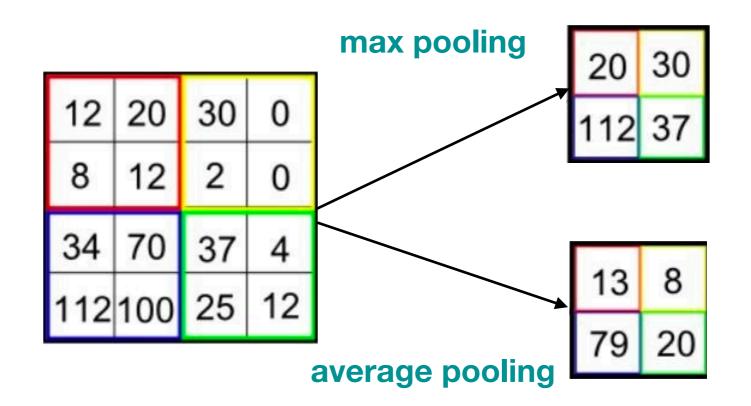
CNNs are composed by **two** kinds of layers

Pooling layer that coarse-grains the input while maintaining locality and spatial structure

~ decimation in RG

reduces the dim. of outputs (depth is kept fixed)

In this example, by pooling over 2x2 blocks, H and W are reduced by half.

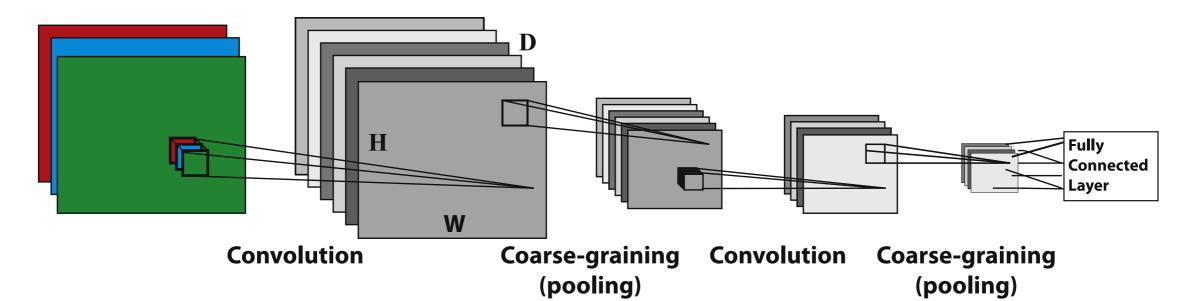


Convolution of input with filters

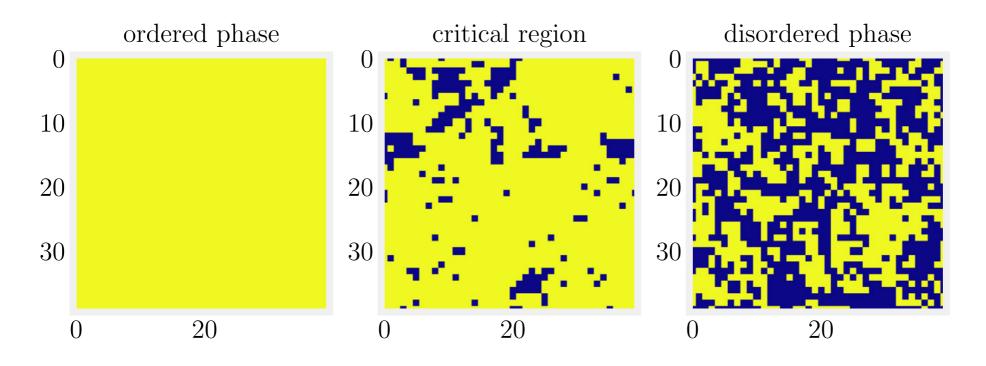
CNNs are composed by two kinds of layers

Pooling layer that coarse-grains the input while maintaining locality and spatial structure

These layers are followed by an **all-to-all connected layer and a high-level classifier**, so that one can train CNNs using the standard backpropagation algorithm:

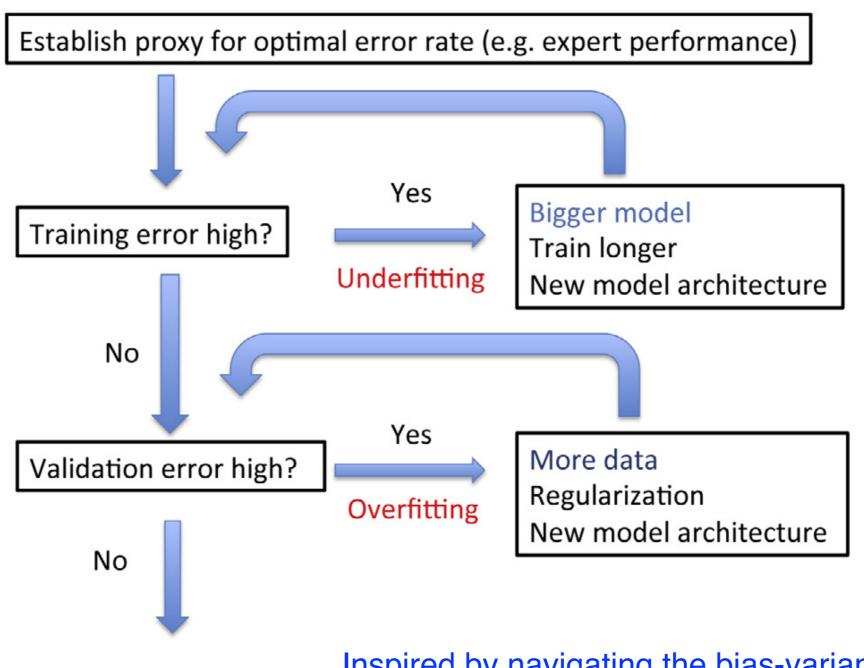


- Significantly reduce the number of parameters: determined by the number and size of the filters. Further reduced by pooling.
- Only problems characterized by spatial locality are amenable to CNNs, e.g., Ising model and MNIST but not SUSY datasets.



- See Notebook 14: Pytorch CNN (Ising); MNIST example (later).
 http://physics.bu.edu/~pankajm/ML-Notebooks/HTML/NB14_CX-CNN-ising-Pytorch.html
- Can you think of the types of datasets in particle physics and cosmology that are amendable to CNNs?

Workflow for Deep Learning



DONE!

Inspired by navigating the bias-variance tradeoff.

See Andrew Ng's talk at the 2016 Deep Learning School: https://www.youtube.com/watch?v=F1ka6a13S9I