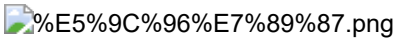
Today's Topic: What is Convolution?

Let's explore more about the theoretical aspect of deep learning

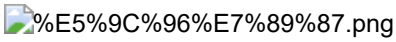
Part I

- In probability/signal processing/differential equation,we also see this term.
  1. Example.The probability density function for sum of two independent random variables, $z = x + y$
  2. Example. Convolution theorem for FT

PDF(probability density function)



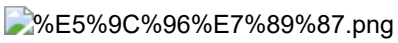
Covolution theorem



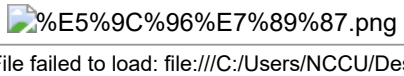
- Brief review about CNN
  - Processing grid-format data
    - 1D:
      1. sound wave (amptitude x time)
      2. skeletion animation(angles x time)
    - 2D:
      1. audio data (after FT $\rightarrow$ row:freq x col:time)
      2. colored image (RGB channel x height x width)
    - 3D: 1.volume data(CT)
      1. colored video data (time x height x width)

- Mathematical foundation
  - Mathematical operation
  - A kind of moving average
  - Input(x),Kernel(w) function and feature map(ouput,s)
  - Communative property due to flip operation

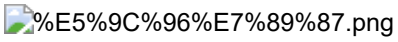
Continous form:



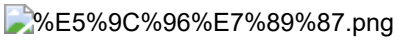
Notation:

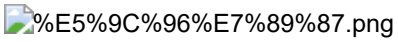


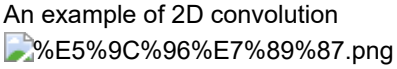
Discrete form:



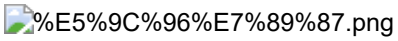
Communativeness





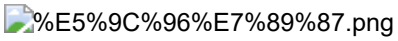
- (continue)
  - An example of 2D convolution
    

Cross-correlation:

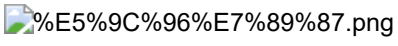


- Convolution in DL
  - Implementation in the DL packages: Cross Correlation (without flip)
  - Sparse interactions(sparse connectivity)
    - $O(m*n)$ to $O(k*n)$, m : input, n : ouput, k : number of connection, let $k << m$
    - Reduce computation complexity

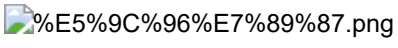
Forward aspect



Backward aspect

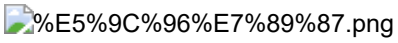


Feature extraction through multiple layers

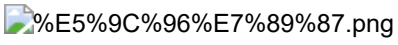


- Convolution in DL (continue)
  - Parameter sharing
    - Tied weight
    - Margin detection
      - Example.Right neighbor of original pixel subtract it self: 280∗320 pixels $\rightarrow$ 280∗319 pixels
      - To represent this transform, the number of operations
        1. matrix: $320 * 280 * 319 * 280 = 8$ billion
        2. covolution: $319 * 280 * 3 = 0.3$ million

Parameter sharing



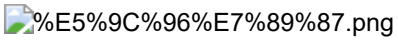
Margin detection



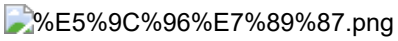
- Convolution in DL (continue)
  - Convolution as a infinitely strong prior (probability distribution)
    - Low (information) entropy / Highly centralized probability density
    - We can think convolution layer as a fully connective layer with a prior: the weight of element in this hidden layer must equal to its neighbors',but can move in the space.
  - The variation of basic convolution
    - Stride $\neq 1$
    - Valid convolution (fill zeros)
    - Unshared convolution (skip)
    - Tiled convolution (skip)

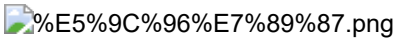
- Cognitive science foundation of CNN
  - Primary visual cortex(V1)
    - Function: space mapping (2D structure)
    - Simple cell $\rightarrow$ relu
    - Complex cell $\rightarrow$ pooling
      - Supplement: pooling
        - local shift invariance
        - we care the apperance of features rather than location
    - "Grandma cell"(Halle Berry Neuron)
  - IT(顳下皮質)
  - Fovea(中央凹)
    - Saccade
    - Attention mechanism $\rightarrow$ NLP
  - Garbor function can describe V1 cells
    - detector

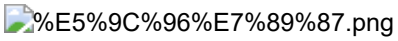
Pooling



Garbor function





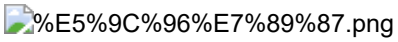


s : response, I : image, w : weight, τ : direction, (x_0, y_0) : origin, α : scaling factor, β : decay rate, f : frequency, ϕ : phrase

Gaussion term : threshold

Cosine term : reponse to the changes along x/y axis

- C(I) describe the behavior of complex cell



s_0,s_1 are same except ϕ (special case: ϕ = 1/4 period) $\rightarrow$ quadrature pair, whcih causes invariance

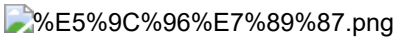
Part II

- An toy model of CNN (without any DL computation framework)
  - Run the code
  - Description for the detail about convolution function in the code

Function for convolution layer (./common/layer.py)

A simple CNN (simple_convnet.py)

Training process (train_convnet.py)



Reference:

- https://github.com/oreilly-japan/deep-learning-from-scratch (https://github.com/oreilly-japan/deep-learning-from-scratch)
- Deep Learning, by Ian Goodfellow, Yoshua Bengio and Aaron Courville