

Lecture 7/23 : chapter 6 Advanced NNets (3 weeks)

cca  
1995 - [Convolution NN], image applications

2009 - [Recurrent NN] : sequence data/architecture

?

HW6

- translation
- RNN
- LSTM
- TRN

2014 - [Attention] for NNets

2017 - [Transformer] Architecture

TRN

2022 Large Scale GPT

not in HW  
CNN for images: Convolution layers for pixels - rectangles (part of image) + NN architecture  
- feed forward  $\Rightarrow$  errors  
- backprop  $\Rightarrow$  weights

output

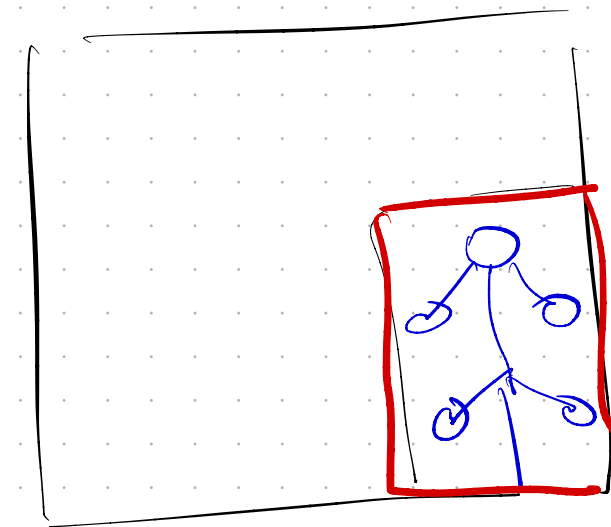
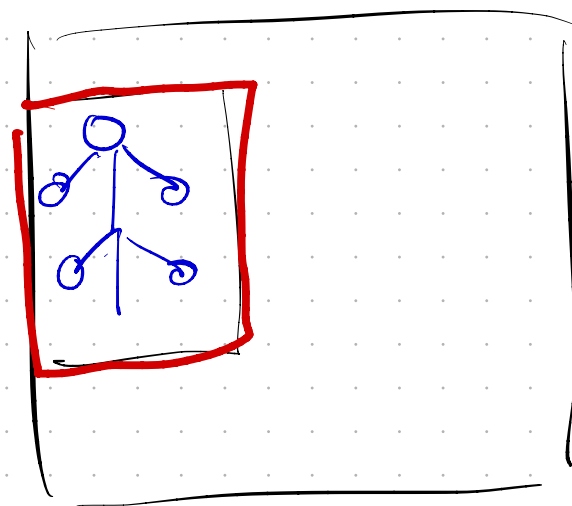
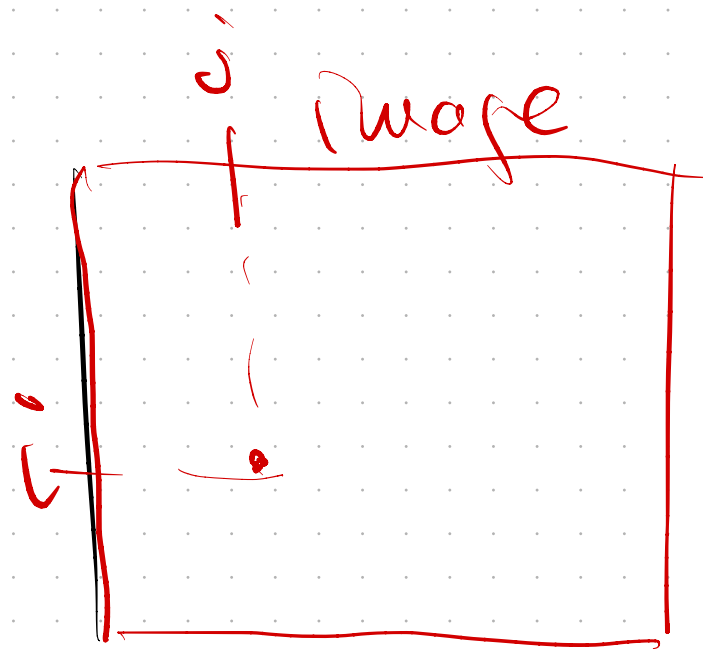
hidden

input

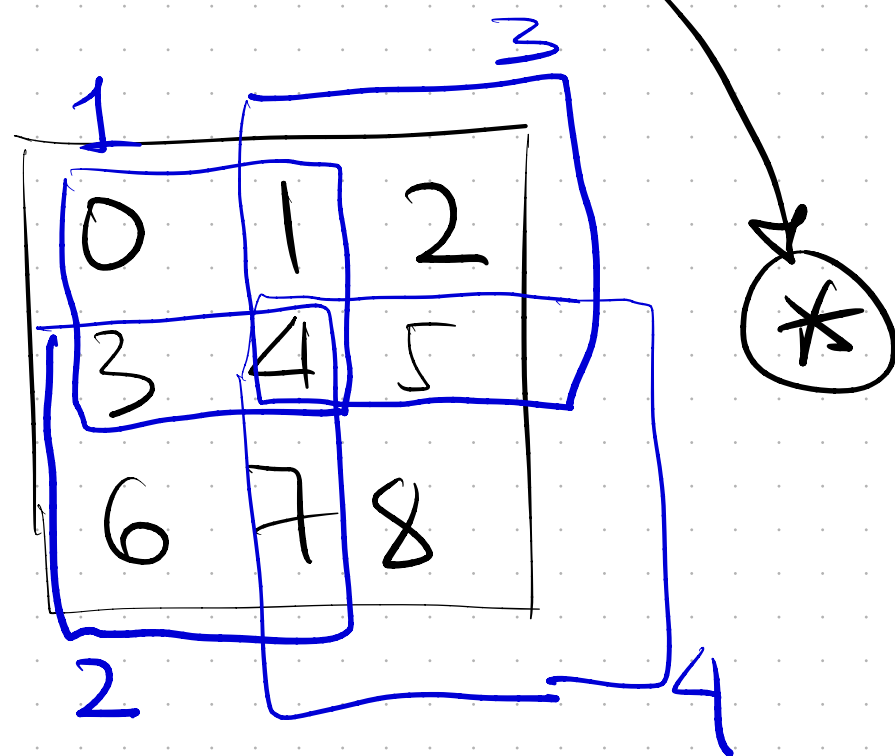
$$H_{ij} = \underbrace{\mu}_{\text{bias}} + \sum_{a,b}^{\text{range}(a,b)} \underbrace{W_{ab}}_{\text{coef}} \underbrace{X_{i+a,j+b}}_{\text{input}}$$

$a$ -row shift  
 $b$ -col shift  
shifts in image

useful if a particular region - rectangle has to be walked with a diff region



# Cross Correlation



convolution  
window

$$\begin{bmatrix} 0 & 1 \\ 2 & 3 \end{bmatrix}$$

=

1

$$\begin{aligned} &0 \times 0 + 1 \times 1 + \\ &3 \times 2 + 4 \times 3 \\ &= 19 \end{aligned}$$

3

$$\begin{aligned} &1 \times 0 + 2 \times 1 + \\ &+ 4 \times 2 + 5 \times 3 = \\ &= 25 \end{aligned}$$

2

$$\begin{aligned} &3 \times 0 + 1 \times 4 \\ &+ 6 \times 2 + 7 \times 3 \\ &= 37 \end{aligned}$$

4

$$\begin{aligned} &4 \times 0 + 5 \times 1 \\ &+ 7 \times 2 + 8 \times 3 \\ &= 43 \end{aligned}$$

Padding: adding rows + cols of "0" on edges.

Stride: vert, horizontal amount to jump-around  
with conv. window

Pooling - averaging/max with neighboring pixels

Deep CNN : multiple layers

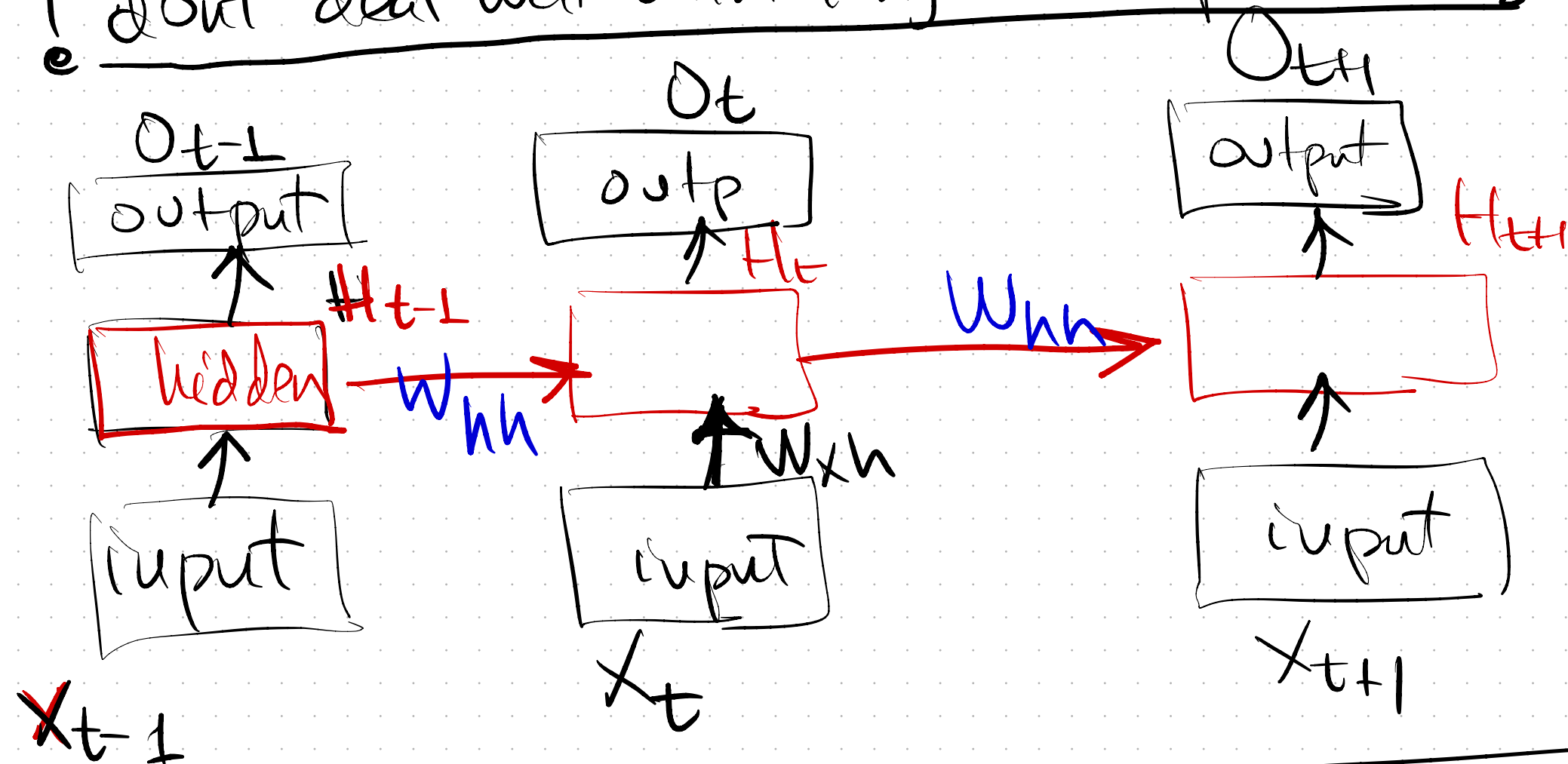
VGG : Visual Geometry Blocks

Computation Issues .

HW6 Recurrent NN : time/episode  $t$

! don't deal well with long-term dependencies!

hidden layer at time  $t$  = hidden state



$$H_t = \phi \left( x_t \cdot \overset{\text{regular } w}{W_{xh}} + H_{t-1} \cdot \overset{\text{rec weights}}{W_{hh}} + \text{bias}_h \right)$$

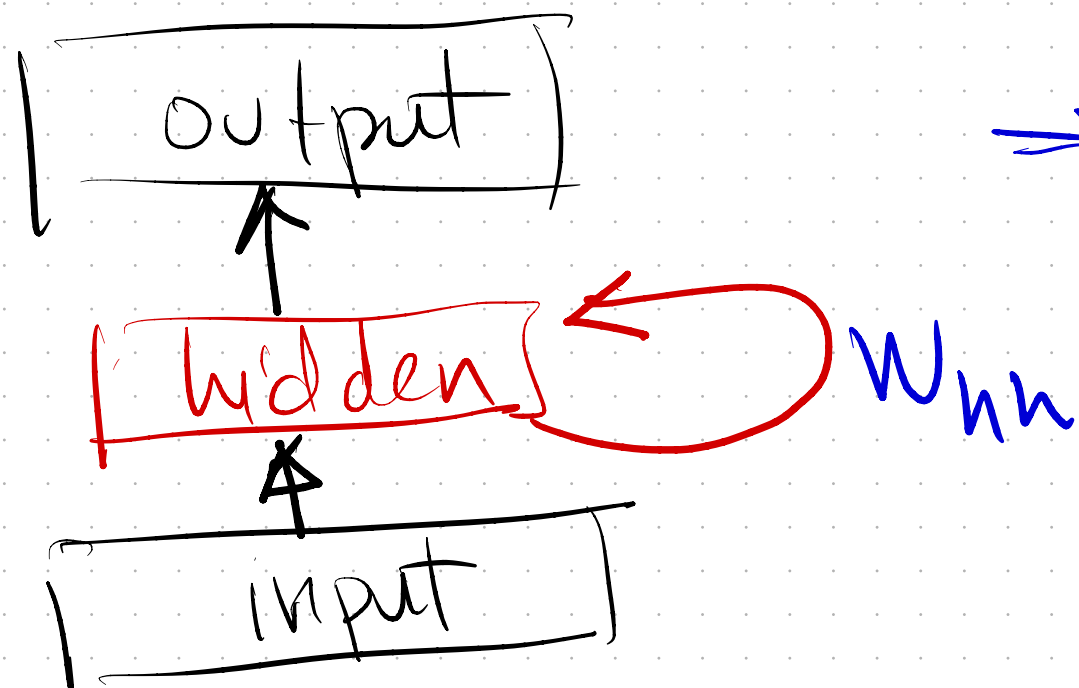
activation

Backpropagation through time (BPTT)

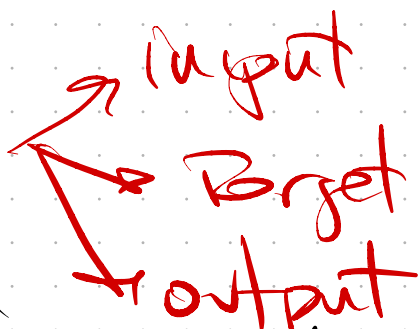
⇒ update both sets of weights

⇒  $w_{xh}$  reg network weights

⇒  $w_{hh}$  recurrent weights



# Adv RNN (libraries)

- Gated memory (cells)   
"sequence of history of events/computation"

⇒ L S T M  
long Short-term memory ⇒ HW6

- Gated Recurrent Units ("simplification of LSTM")

- Deep RNN: multiple hidden layers/states

• Beam Search

- Bidirectional RNN

- Encoder-Decoder architecture for RNN

input sequence  $\xrightarrow{\text{ENCODER}}$  hidden state  $\xrightarrow{\text{DECODER}}$  figures/predict next item

Transformers NN  $\Rightarrow$  Simple TRN in HW6

- Attention mechanism (queries, keys, values)
- Attention Pooling / Similarity
- Attention Scoring — formula for attention
- Bahdanau Attention
- Multihead Attention
- Self Attention / Transformer