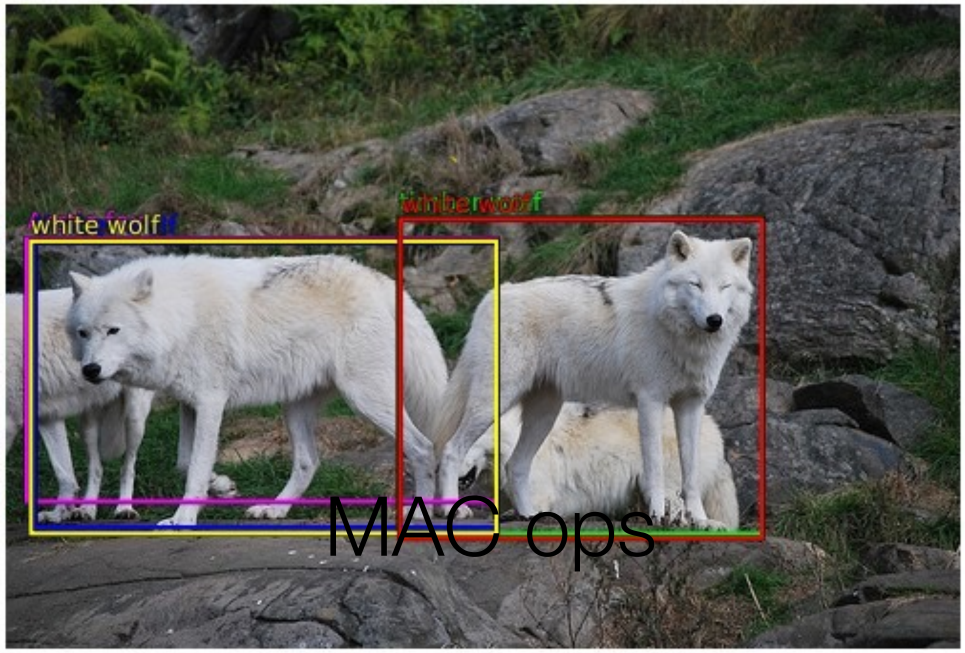
Data Mining & Machine Learning

CS37300
Purdue University

Oct 27, 2023

Example Applications of CNNs

Object Recognition



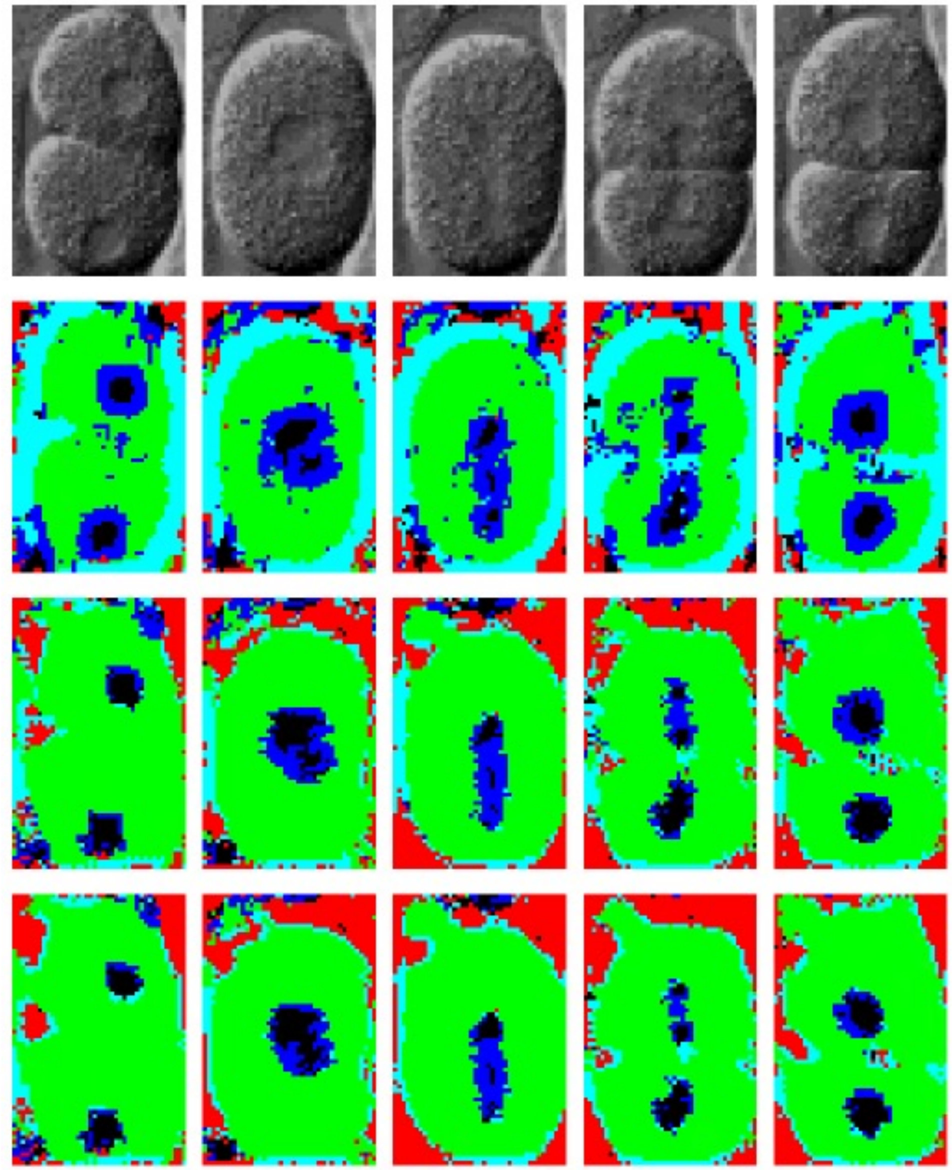
Top 5:
white wolf
white wolf
timber wolf
timber wolf
Arctic fox

ILSVRC2012_val_00000027.JPEG

Credit: Y. LeCun, Facebook

ConvNets for image segmentation

- Biological Image Segmentation
  - [Ning et al. IEEE-TIP 2005]
- Pixel labeling with large context using a convnet
- ConvNet takes a window of pixels and produces a label for the central pixel

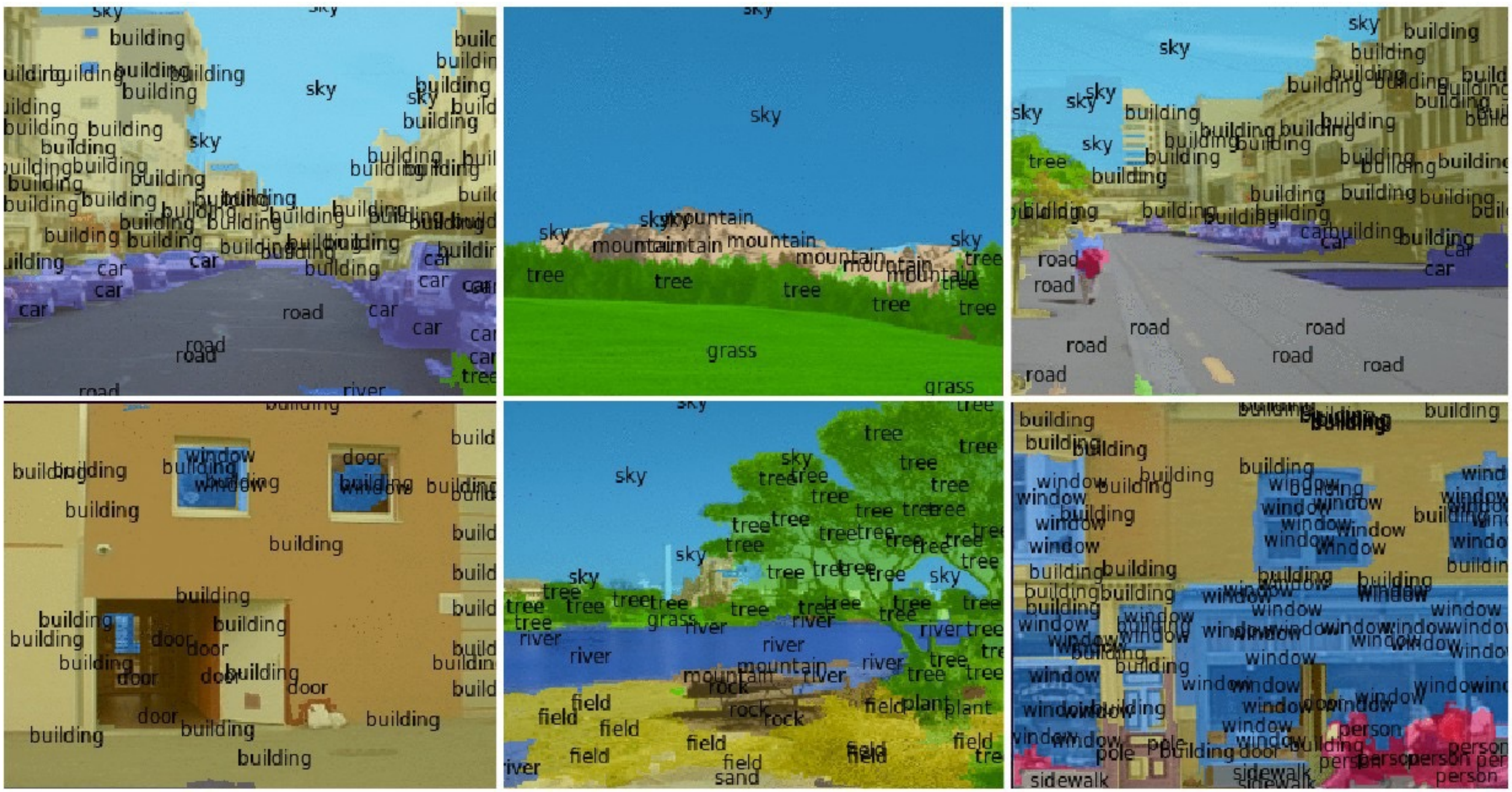


Credit: Y. LeCun, Facebook

Semantic labeling / scene parsing:

Labeling every pixel with the object it belongs to

– Would help identify obstacles, targets, landing sites, dangerous areas
Would help line up depth map with edge maps



[Farabet et al. ICML 2012, PAMI 2013]

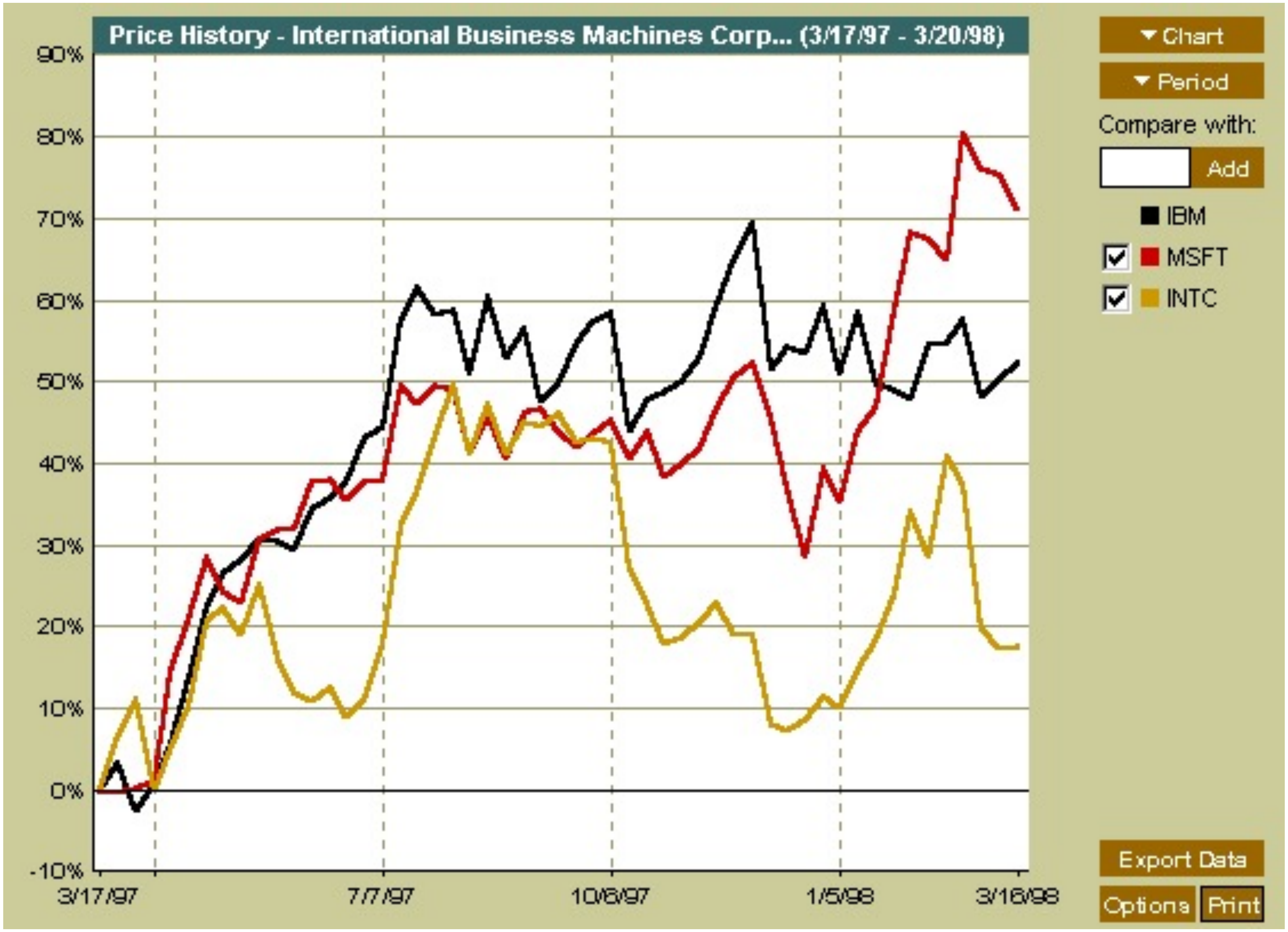
Credit: Y. LeCun, Facebook

Time series

Sequence data

- Time-series database
  - Consists of sequences of values or events changing with time
  - Data is recorded at regular intervals
  - Characteristic time-series components
    - Trend, cycle, seasonal, irregular
- Applications
  - Financial: stock price, inflation
  - Biomedical: blood pressure
  - Meteorological: precipitation

Example: Stocks



Mining sequences

- ”Fit the curve”: Find the curve minimizing the sum of the squares of the deviation of points on the curve from the corresponding data points

- The moving-average statistic to reduce variance

- Eliminate cyclic, seasonal and other patterns

    - e.g. Estimation of seasonal variations through seasonal indices

      Set of numbers showing the relative values of a variable during the months of the year

        - e.g., if the sales during October, November, and December are 80%, 120%, and 140% of the average monthly sales for the whole year, respectively, then 80, 120, and 140 are seasonal index numbers for these months

Searching similar sequences

- Normal database query finds exact literal matches
- Similarity search finds data sequences that differ only slightly from the given query sequence
- Two categories of similarity queries
  - Whole matching: find a sequence that is similar to the query sequence
  - Subsequence matching: find all pairs of similar sequences
- Capture ”invariant” structures e.g. using data transformations
- Typical Applications
  - Financial market
  - Market basket data analysis
  - Scientific databases
  - Medical diagnosis

Sequential Pattern Mining

- Mining of frequently occurring patterns related to time or other sequences
- Sequential pattern mining usually concentrate on symbolic patterns
- Examples
  - Renting “Star Wars”, then “Empire Strikes Back”, then “Return of the Jedi” in that order
  - Collection of ordered events within an interval
- Applications
  - Targeted marketing
  - Customer retention
  - Weather prediction

Mining Sequences (cont.)

Customer-sequence

CustId	Video sequence
1	{(C), (H)}
2	{(AB), (C), (DFG)}
3	{(CEG)}
4	{(C), (DG), (H)}
5	{(H)}

Map Large Itemsets

Large Itemsets	MappedID
(C)	1
(D)	2
(G)	3
(DG)	4
(H)	5

Sequential patterns with support > 0.25

{(C), (H)}

{(C), (DG)}

Recurrent Neural Networks

(learning sequences with identifiable “transition” structures)

Overview

- Deep Learning Architectures
  - Feedforward neural networks
  - Convolutional Neural Networks – Specialized for representing grid data to capture spatial information
  - Recurrent Neural Networks – Specialized for representing sequence data

Importance of parameter sharing

- In CNNs, we used parameter sharing to encode a feature extracted using a filter

- Parameter sharing allows application of a model to generalize to unseen forms of data

    - e.g. detection filter identifies edges from pixels that are further combined to form higher level information features

- In RNNs, these generalization-worthy patterns refers to sequences

    - e.g In 2012, I moved to the US

    - vs I moved to the US in 2012

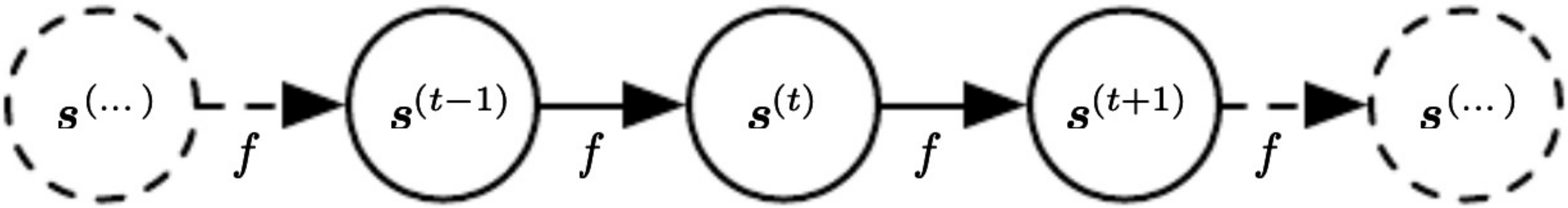
A simple FF NN will require too many features

Motivation – 1D temporal sequence convolution

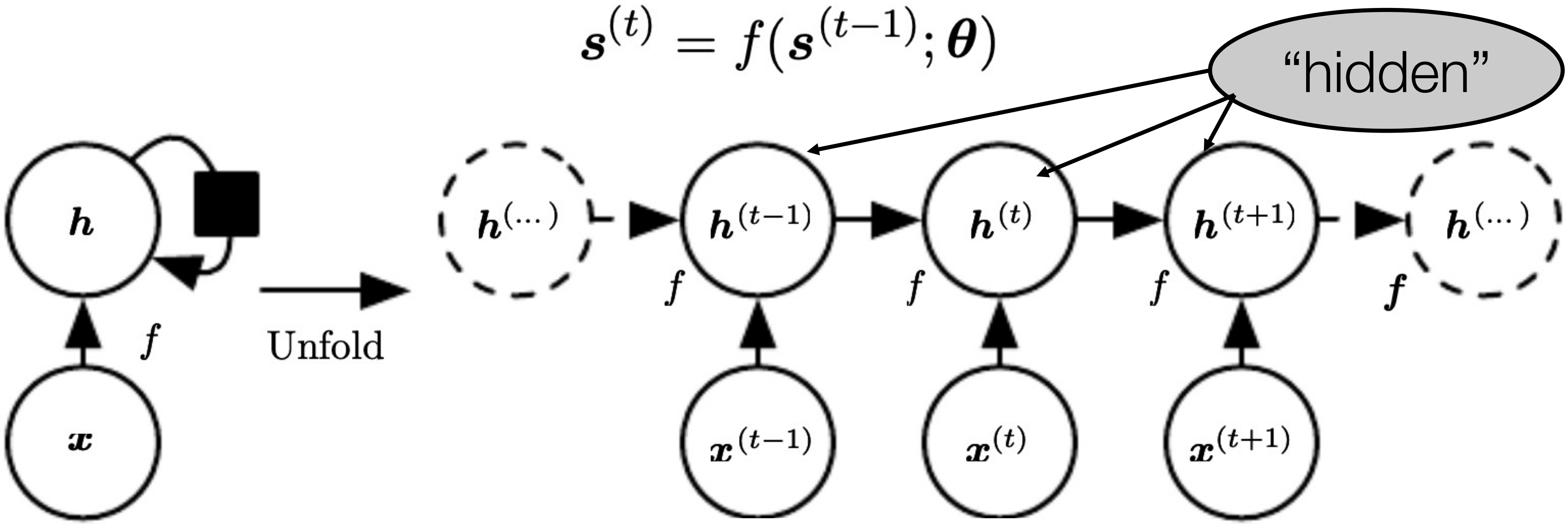
- The technique constructs “Time-Delayed networks”
- Output of convolution is a sequence where each member of output is a function of neighboring elements.
- A shallow network that shares parameters across time
- RNNs on the other hand: Each member of the output is a function of previous members of the output

RNNs encoded as f( )

- Unfolded computation graphs



$$\boldsymbol{s}^{(t)} = f(\boldsymbol{s}^{(t-1)}; \boldsymbol{\theta})$$

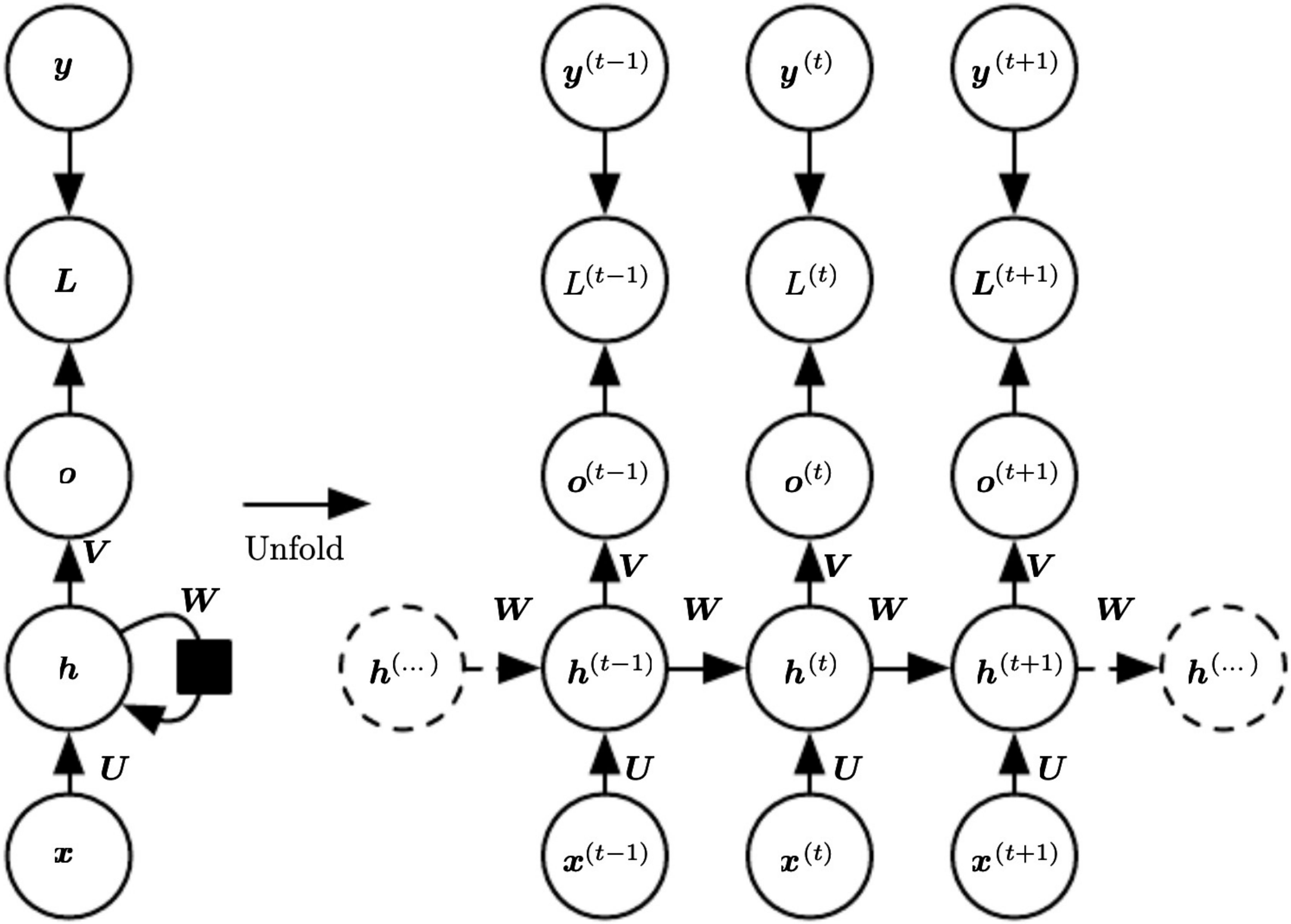


$$\boldsymbol{h}^{(t)} = f(\boldsymbol{h}^{(t-1)}, \boldsymbol{x}^{(t)}; \boldsymbol{\theta})$$

RNNs

- Predict future state based on past state:
    - h( ) encoded as a lossy summary of past sequences relevant for future predictions in the task at hand
    - Inputs x^t, length of input sequence is arbitrary
- Example: Statistical language modeling task to predict next word
- Learned model of a fixed sized irrespective of sequence length, since the model is meant to learn the transition
- Parameters shared for transitions

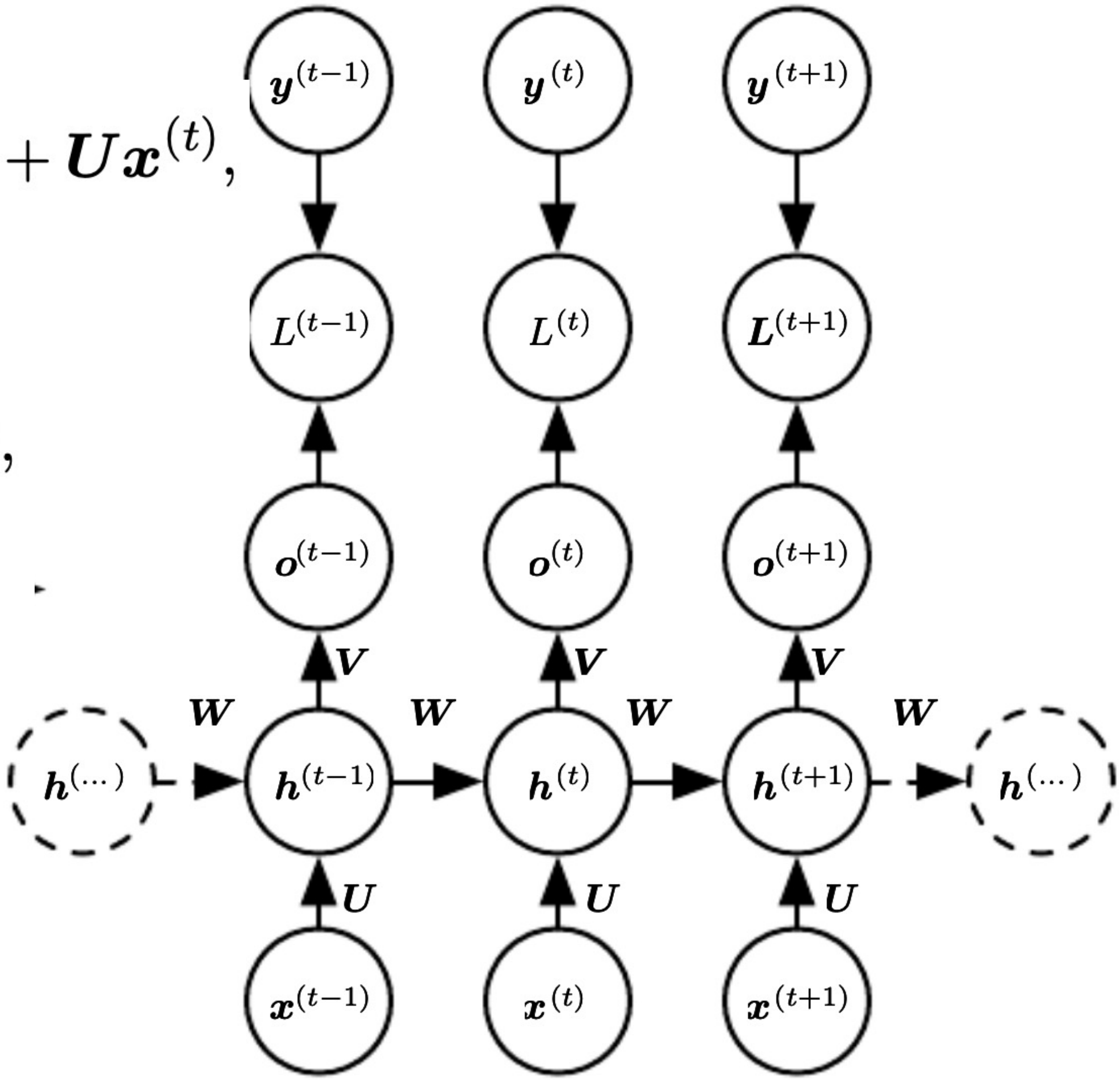
Simple idea with many potential applications



Simple idea with many potential applications

$$
\begin{aligned}
\boldsymbol{a}^{(t)} &= \boldsymbol{b} + \boldsymbol{W}\boldsymbol{h}^{(t-1)} + \boldsymbol{U}\boldsymbol{x}^{(t)}, \\
\boldsymbol{h}^{(t)} &= \tanh(\boldsymbol{a}^{(t)}), \\
\boldsymbol{o}^{(t)} &= \boldsymbol{c} + \boldsymbol{V}\boldsymbol{h}^{(t)}, \\
\hat{\boldsymbol{y}}^{(t)} &= \operatorname{softmax}(\boldsymbol{o}^{(t)}),
\end{aligned}
$$

Any suggestions for L( )?



Simple idea with many potential applications

Would backprop work for gradient calculations?