Outline

- RNNs and vanishing/exploding gradients
- Solutions

RNNs: Advantages

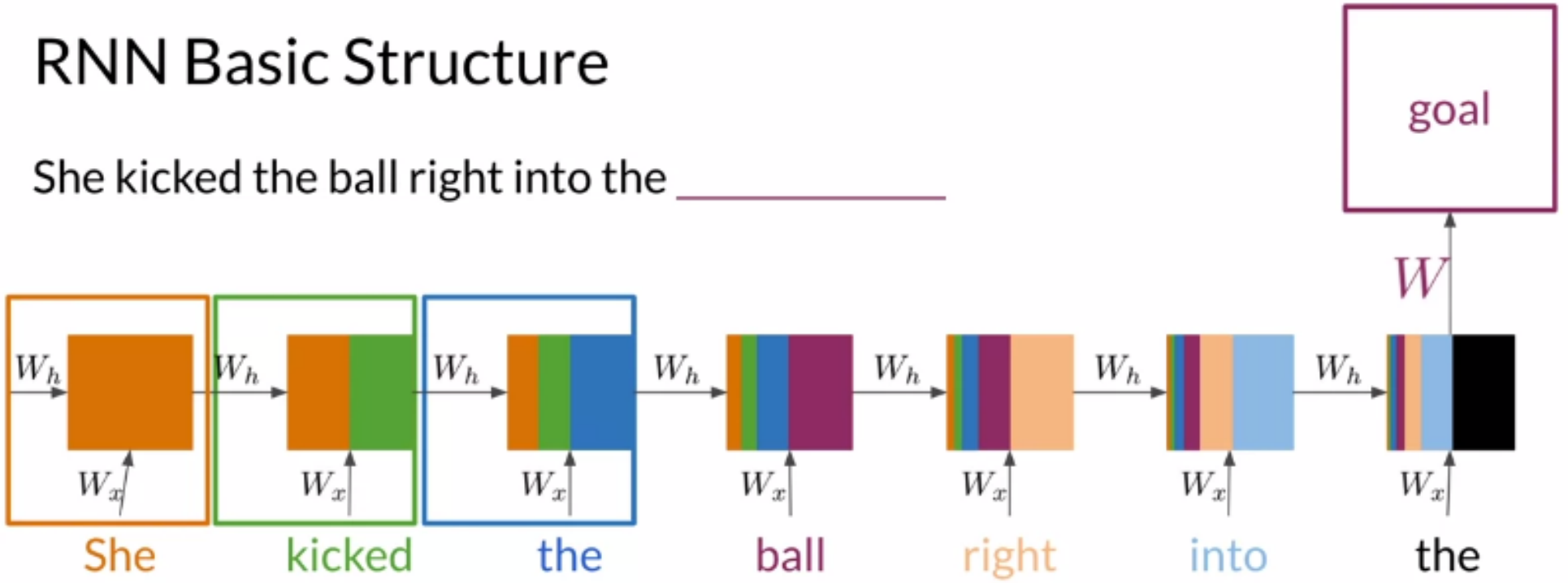
- ✚ Captures dependencies within a short range
- ✚ Takes up less RAM than other n-gram models

RNNs: Disadvantages

- Struggles with longer sequences
- Prone to vanishing or exploding gradients

RNN Basic Structure

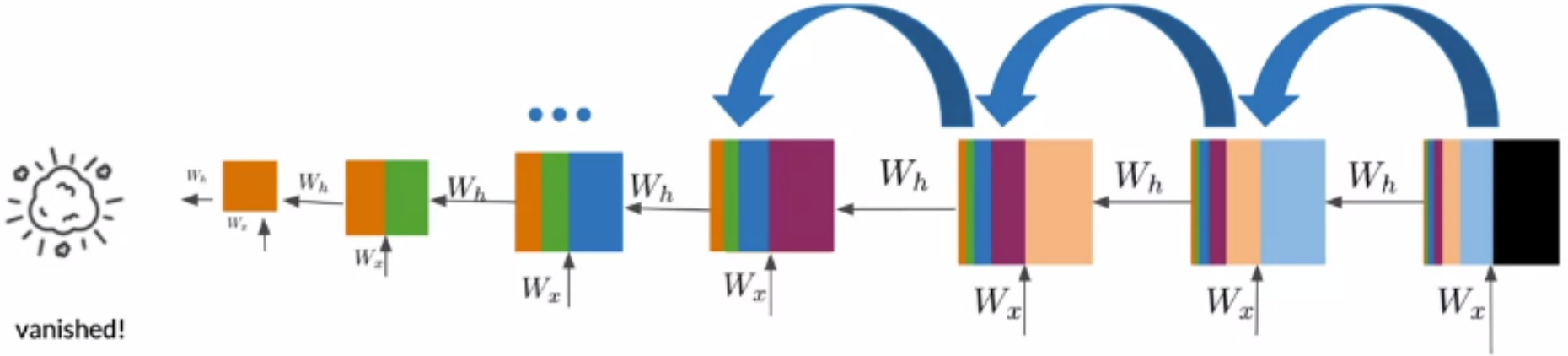
She kicked the ball right into the _____



Learnable parameters

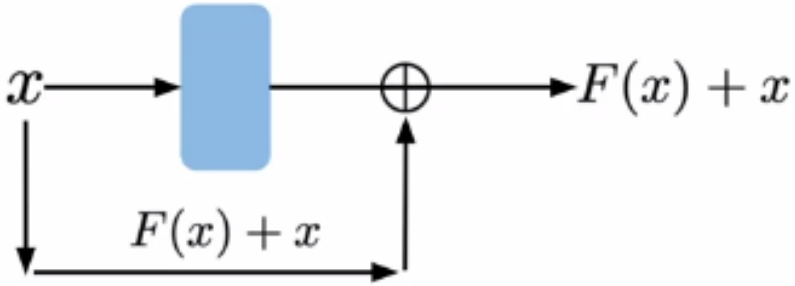
Backpropagation through time

The vanishing gradient problem



Solving for vanishing or exploding gradients

- Identity RNN with ReLU activation $\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}$ -1 ⟶ 0
- Gradient clipping 32 ⟶ 25
- Skip connections



Outline

- Meet the Long short-term memory unit!
- LSTM architecture
- Applications

LSTMs: a memorable solution

- Learns when to remember and when to forget
- Basic anatomy:
  - A cell state
  - A hidden state with three gates
  - Loops back again at the end of each time step
- Gates allow gradients to flow unchanged

LSTMs: Based on previous understanding

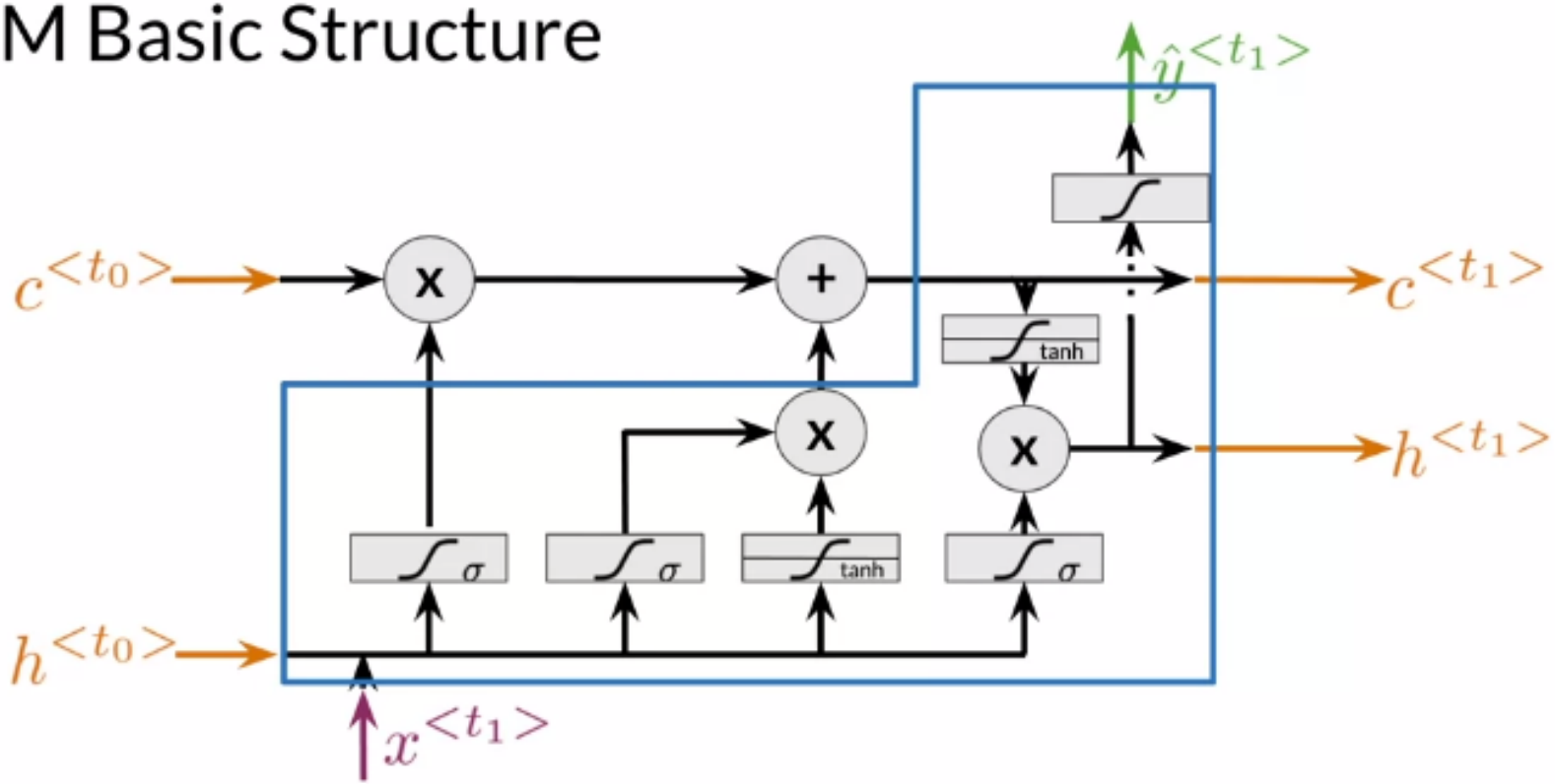
Cell state = before conversation

Forget gate = beginning of conversation

Input gate = thinking of a response
Output gate = responding

Updated cell state = after conversation

LSTM Basic Structure



Applications of LSTMs

Next-character prediction

Chatbots

Music composition

Image captioning

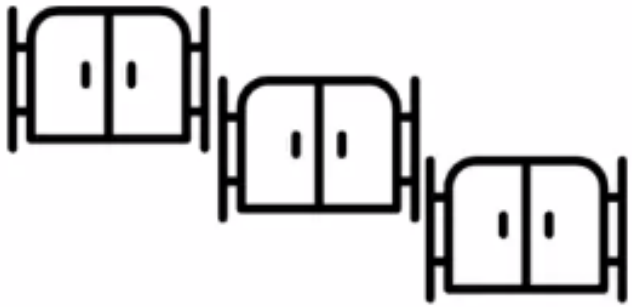
Speech recognition

Summary

- LSTMs offer a solution to vanishing gradients

Summary

- LSTMs offer a solution to vanishing gradients
- Typical LSTMs have a cell and three gates:
  - Forget gate
  - Input gate
  - Output gate



Summary

- LSTMs use a series of gates to decide which information to keep:
  - Forget gate decides what to keep
  - Input gate decides what to add
  - Output gate decides what the next hidden state will be

Summary

- LSTMs use a series of gates to decide which information to keep:
  - Forget gate decides what to keep
  - Input gate decides what to add
  - Output gate decides what the next hidden state will be

- One time step is completed after updating the states