Advanced School in Artificial Intelligence

Introduction to Recurrent Neural Networks

Ing. Zese Riccardo
riccardo.zese@unife.it

Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

RegioneEmilia-Romagna

Università degli Studi di Ferrara

Outline

- Introduction to Python
- Introduction to Neural Networks
- Convolutional NN
- Recurrent NN
- Autoencoders and self supervised learning

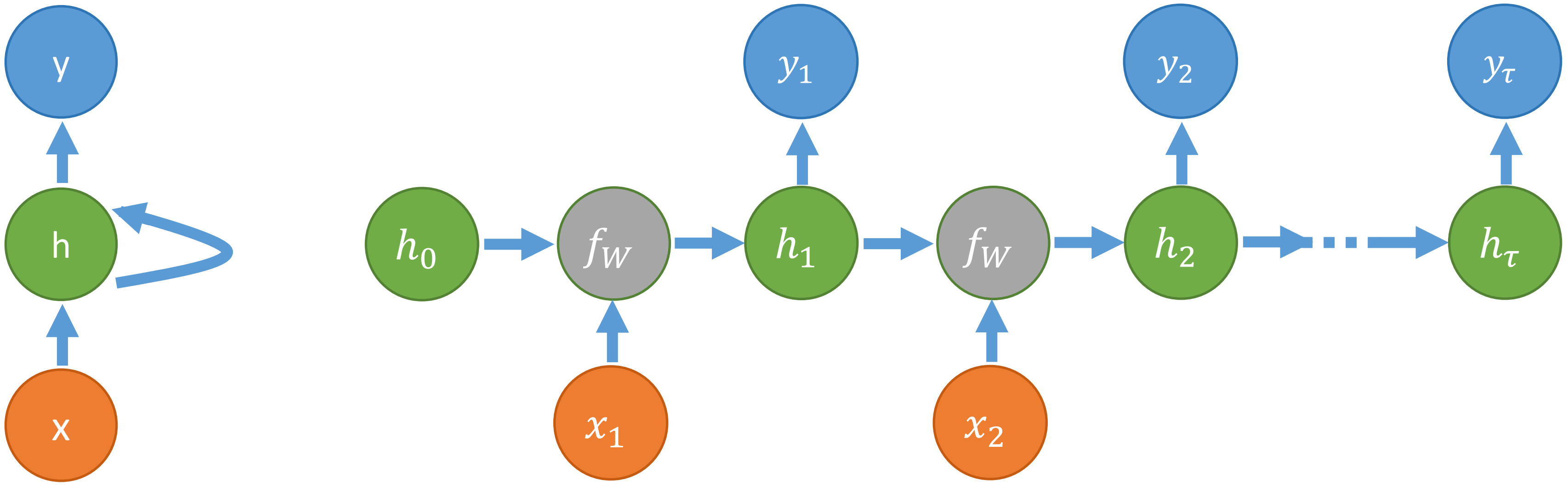
Outline

- Introduction to Python
- Introduction to Neural Networks
- Convolutional NN
- Recurrent NN
- Autoencoders and self supervised learning

Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

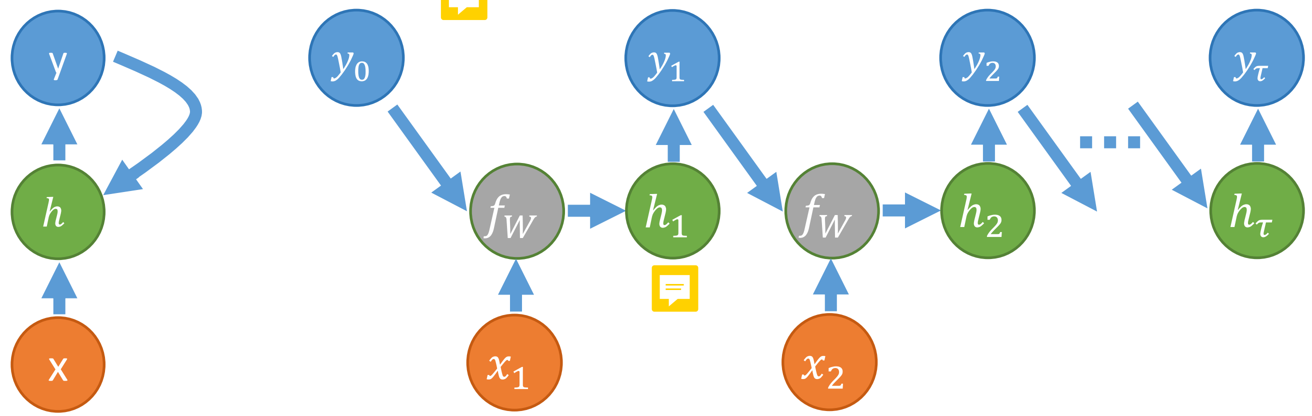
Passing information to the future

We have seen that the state is passed to the next step → the state can contain **any information it wants** about the past



Passing information to the future

However, one can model a net that passes the output y to the next state → It is trained to put a specific output value into y, and y is the only information it is allowed to send to the future.



Unless h is very high-dimensional, it will usually lack important information from the past. This RNN is less powerful, but it may be easier to train because each time step can be trained in isolation from the others, allowing greater parallelization during training

Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

Example: Character-level Language Model

- We have a possibly very long text and we want our model to output the probability distribution of the next character in the sequence given a sequence of previous characters.

- Consider a smaller scenario with the following vocabulary [h,e,l,o]
- Our target is to train an RNN to recognize the word "hello"

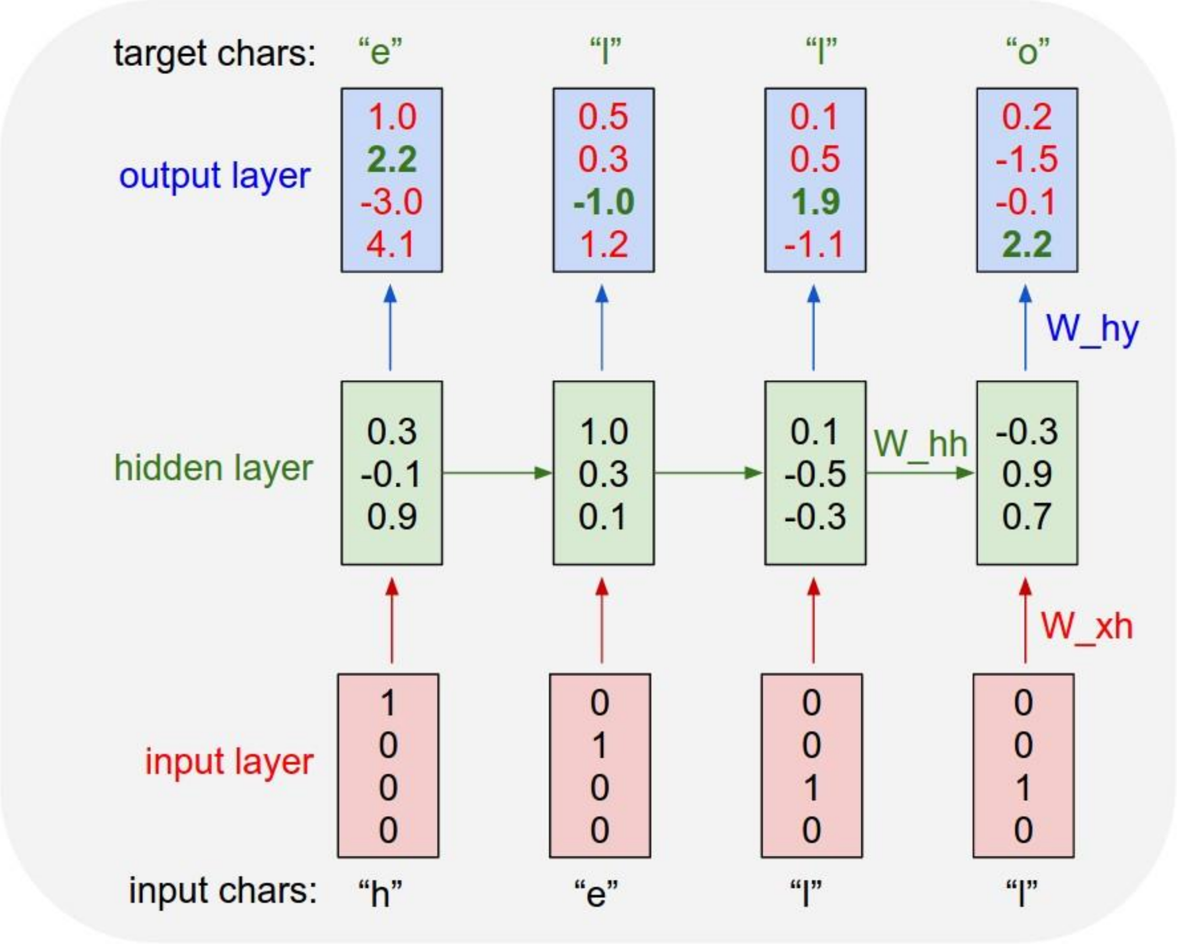
- We have 4 different training examples whit the following characteristics:
  1. After ‘h’, letter ‘e’ must have higher probability
  2. After ‘he’, letter ‘l’ must be the most likely
  3. After ‘hel’, we would like another ‘l’
  4. After ‘hell’, the net should output an ‘o’

From: https://medium.com/explore-artificial-intelligence/an-introduction-to-recurrent-neural-networks-72c97bf0912

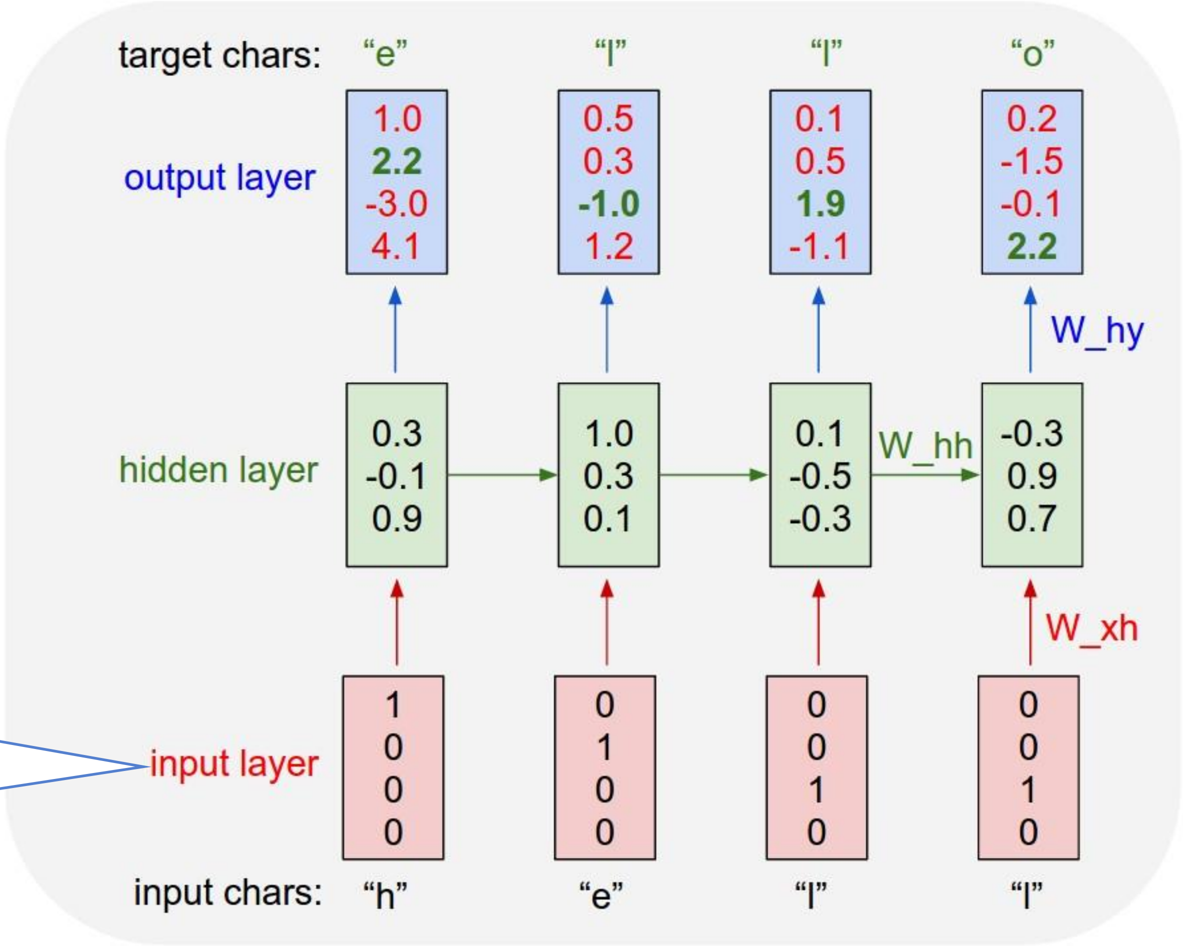
Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

Example: Character-level Language Model

- Graphically



Example: Character-level Language Model

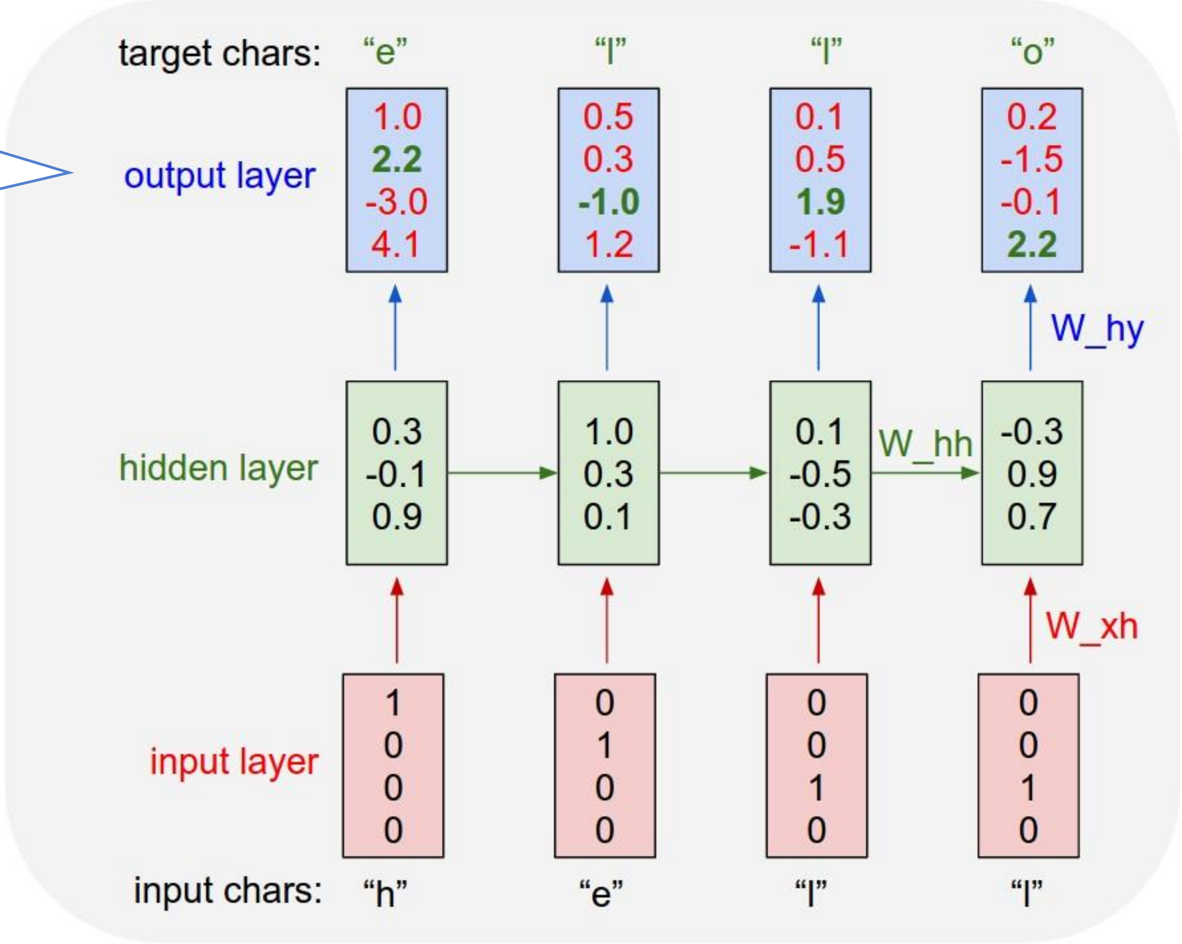


Characters encoded with 4-dimensional one-hot vectors

Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

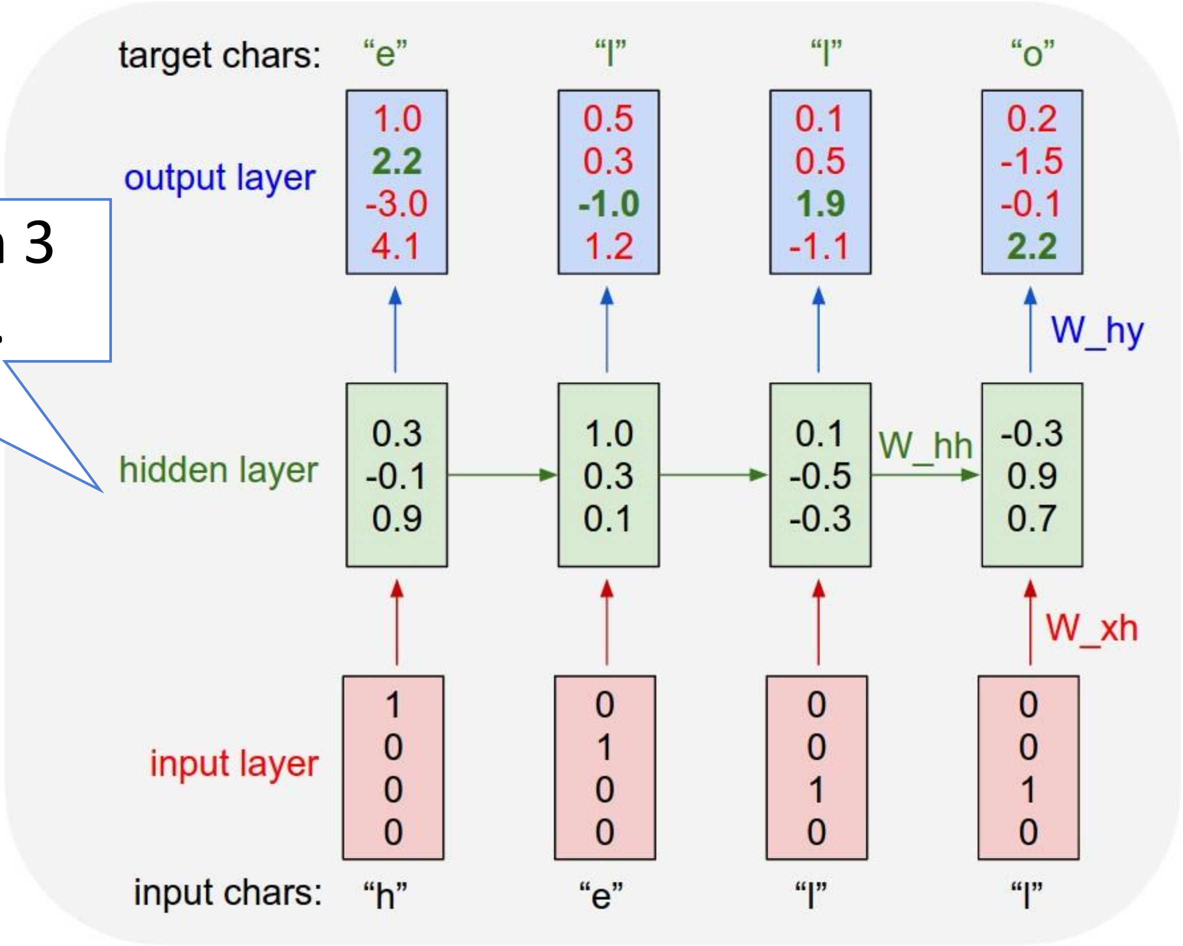
Example: Character-level Language Model

Output is 4-dimensional vector containing the confidence we have on the next character.



Example: Character-level Language Model

Hidden layer with 3 units (neurons).



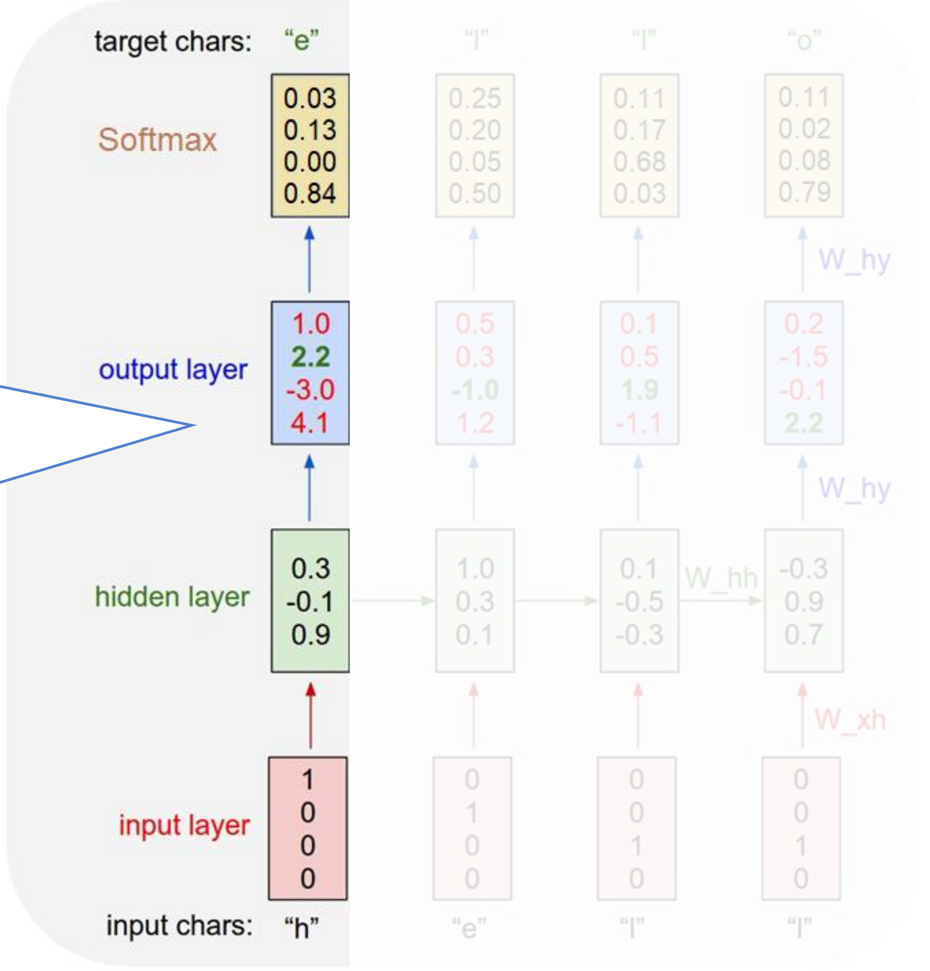
This diagram shows the activations in the forward pass when the RNN is fed the characters "hell" as input.

We want the green numbers in output layer to be high and red numbers to be low.

Example: Character-level Language Model

At test time sample characters one at a time is given to the model → feedforward

First step:
'h' has confidence of 1.0 to be the next letter after 'h', 'e' has 2.2, -3.0 for 'l', and 4.1 for 'o'.
Since the next correct character is 'e', we would like to increase its confidence (green) and decrease the confidence of all other letters (red).

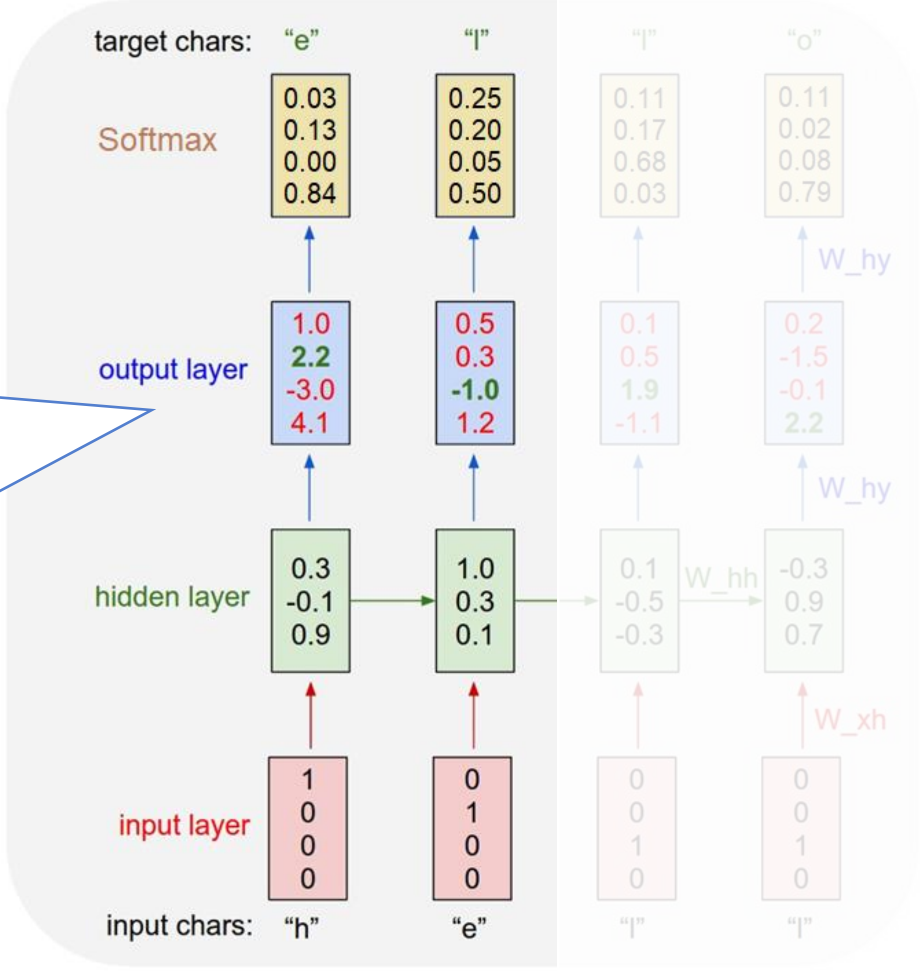


Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

Example: Character-level Language Model

At test time sample characters one at a time is given to the model → feedforward

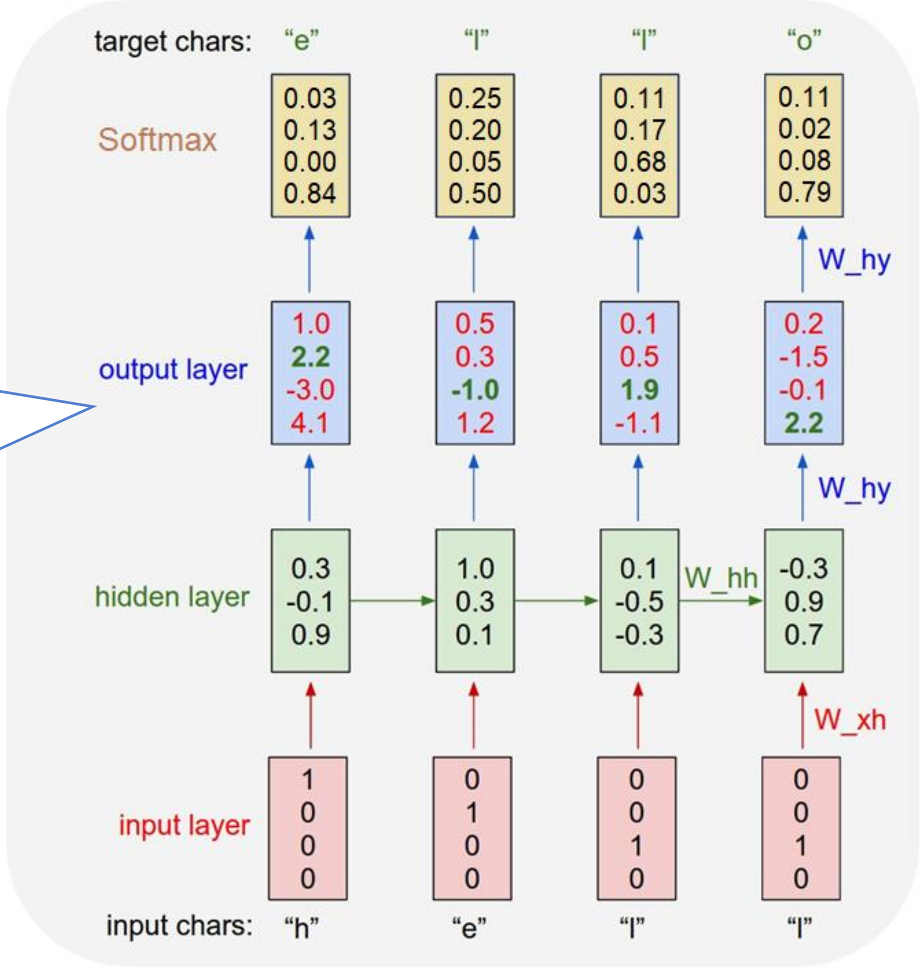
Second step:
The hidden layer is updated using the previous state.



Example: Character-level Language Model

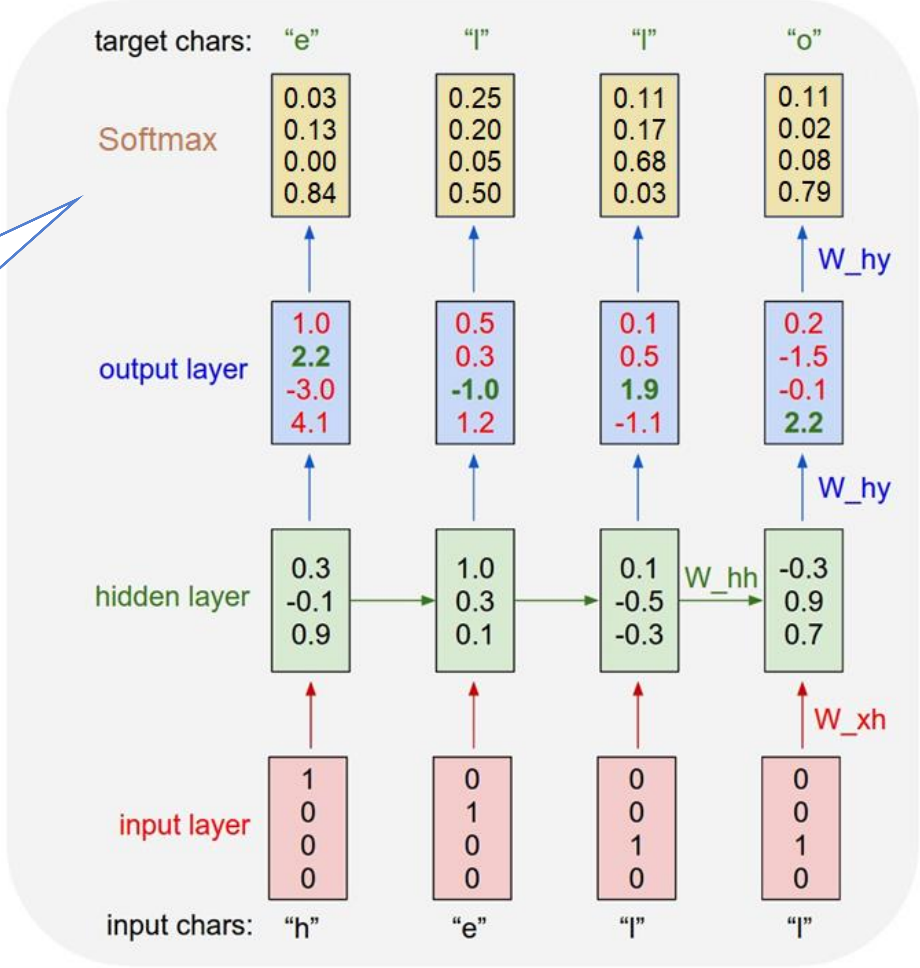
With these weights the resulting output is ‘oolo’ instead of ‘ello’.

Note that all the operations performed are differentiable → use of the back-propagation to tune the weights and obtain the wanted answer.



Example: Character-level Language Model

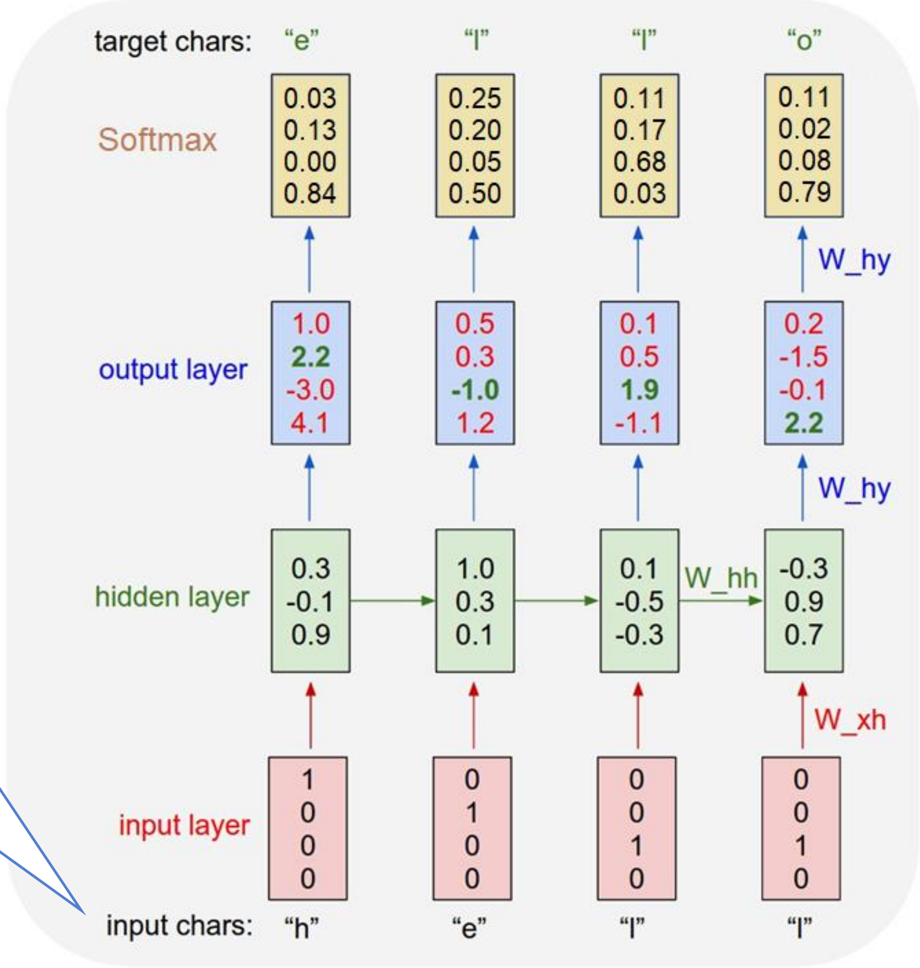
More technically, standard Softmax classifier is used on every output vector simultaneously.



Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

Example: Character-level Language Model

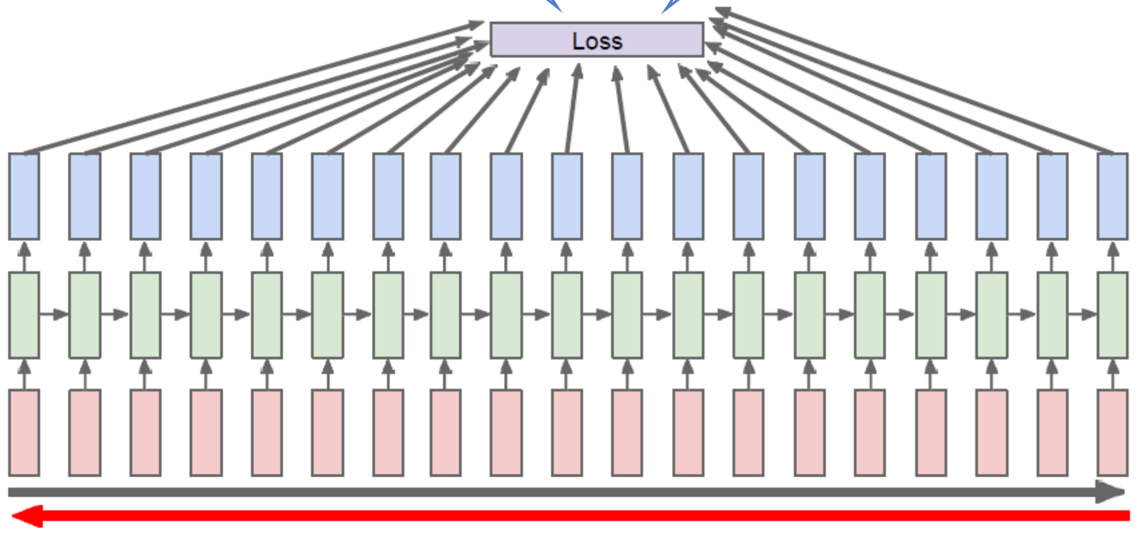
It is clear in this example the need to consider the entire sequence (or at least part of them) instead of single characters, in fact after character 'l' we would like a second 'l' the first time and an 'o' the second time. By considering only the single input character this definition cannot be learned.



Back-propagation Trough Time (BPTT)

Forward through entire sequence to compute loss, then backward through entire sequence to compute gradient

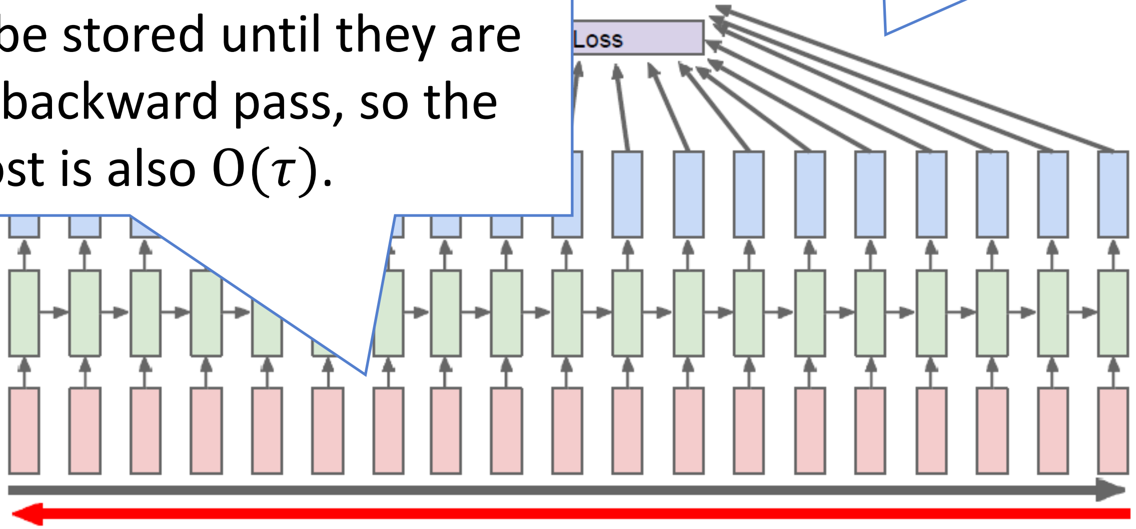
The total loss for a given sequence paired with the sequence of the wanted values is just the sum of the losses over all the time step.



Back-propagation Trough Time (BPTT)

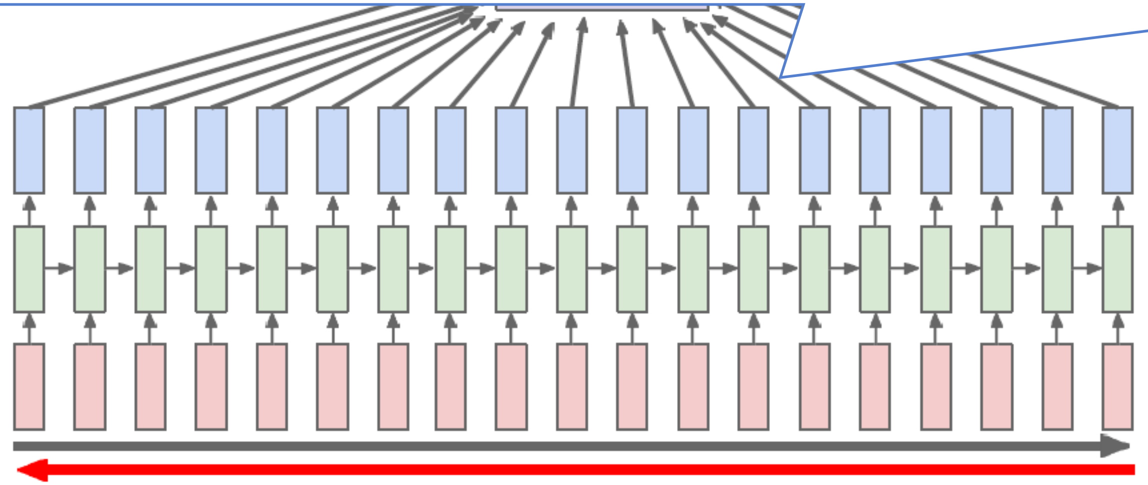
The runtime is $O(\tau)$ and cannot be reduced by parallelization because the forward propagation graph is inherently sequential; each time step may only be computed after the previous one. States computed in the forward pass must be stored until they are reused during the backward pass, so the memory cost is also $O(\tau)$.

The back-propagation algorithm applied to this graph with $O(\tau)$ cost is called **back-propagation through time** or **BPTT**.



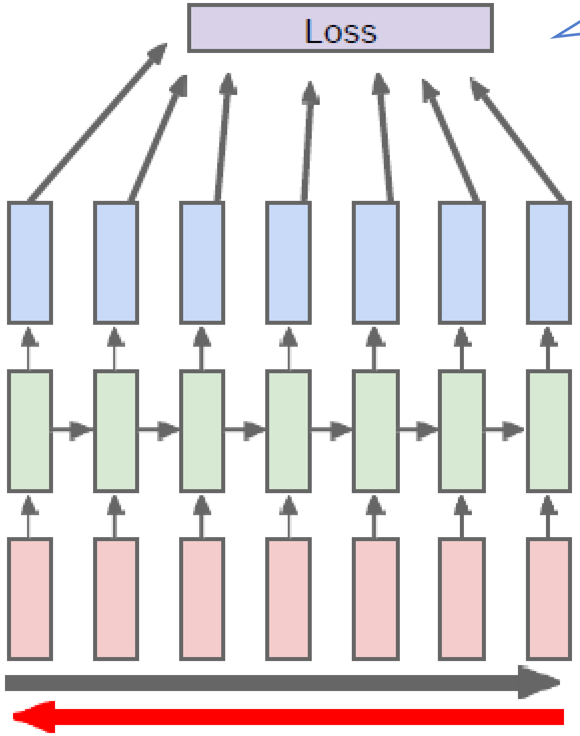
Back-propagation Trough Time (BPTT)

Computing the gradient through a recurrent neural network is straightforward.
One simply applies the generalized back-propagation algorithm to the computational graph. No specialized algorithms are necessary.
Gradients obtained by back-propagation may then be used with any general-purpose gradient-based techniques to train an RNN, for example, train the RNN with mini-batch Stochastic Gradient Descent using RMSProp or Adam.



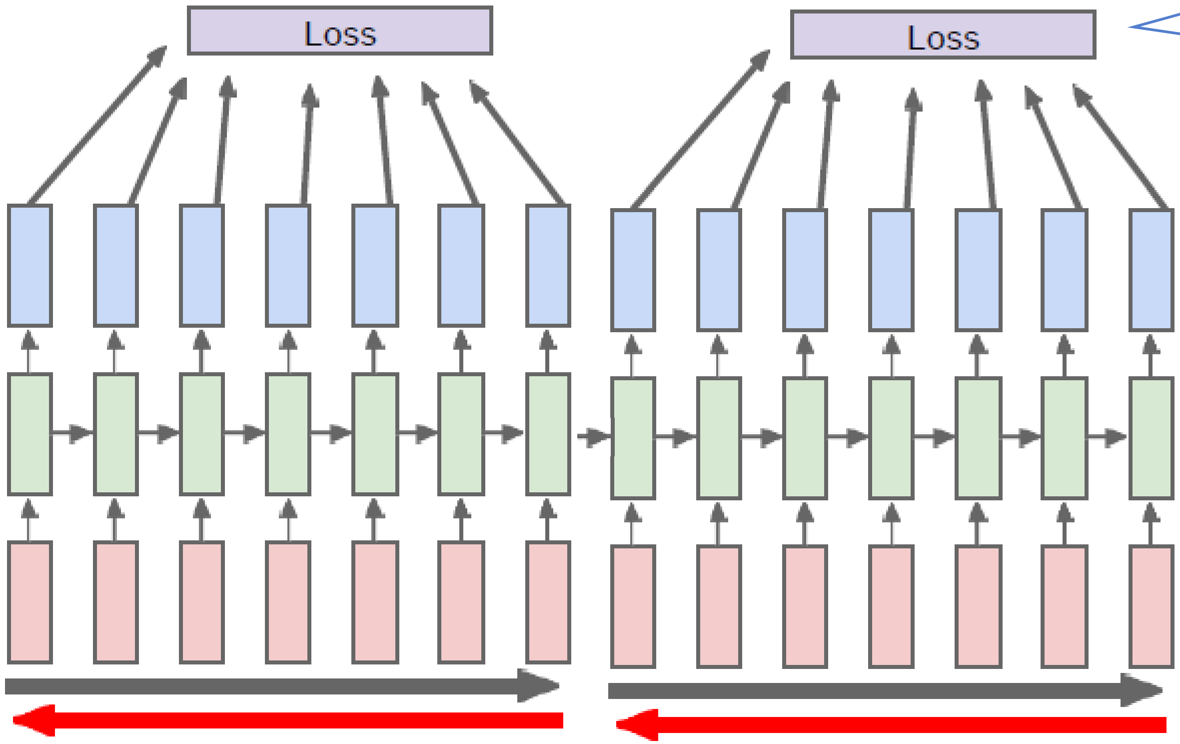
Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

Truncated Back-propagation Through Time



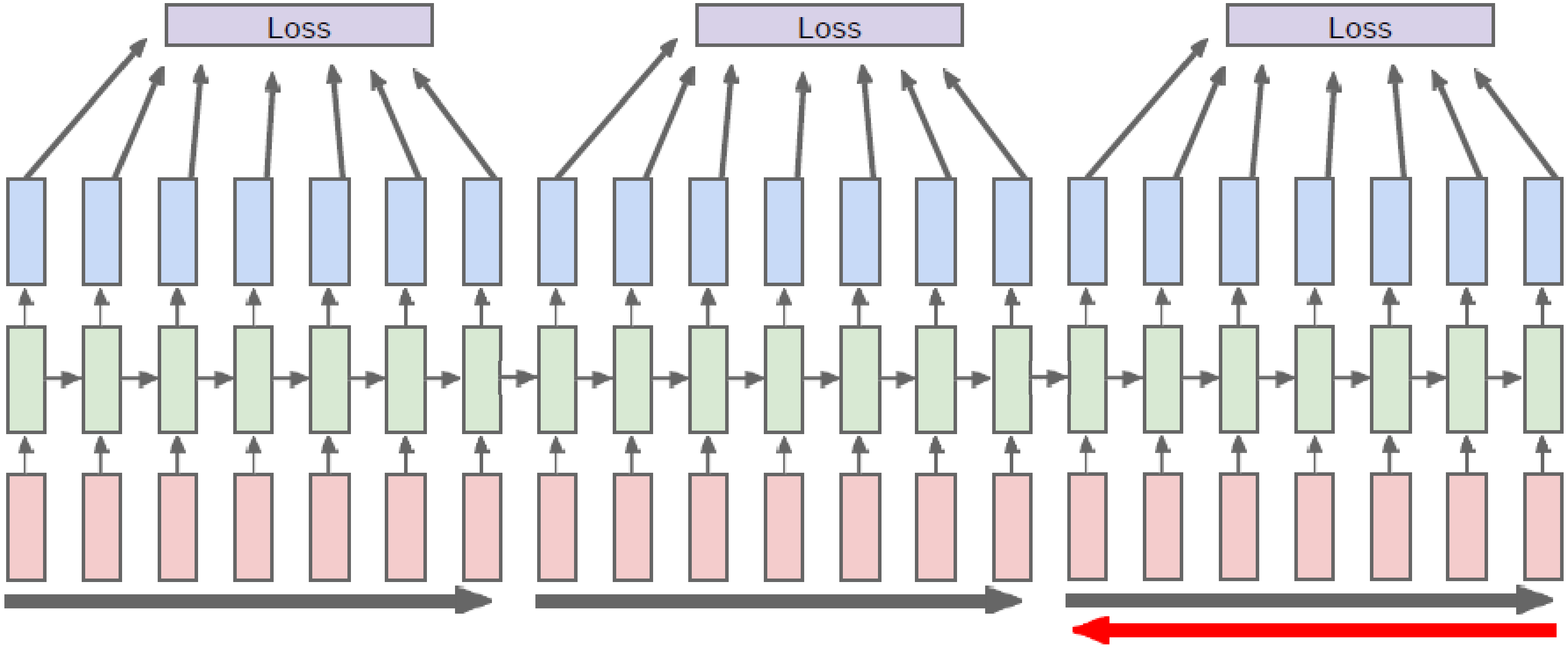
Run forward and Backward through chunks of the sequence instead of whole sequence

Truncated Back-propagation Through Time



Carry hidden states forward in time forever, but only backpropagate for some smaller number of steps

Truncated Back-propagation Through Time



Recurrent Neural Networks

- Once the net is trained, we can use it to create text which follows the learned distribution.

- At test time, we feed a character into the RNN and get a distribution over what characters are likely to come next. We sample from this distribution, and feed it right back in to get the next letter. Repeat this process and you're sampling text!
- To do that, however, the network must have a mechanism to determine the length of the sequence.

Recurrent Neural Networks

- To chose when to stop during the generation of samples we can:
  - Add a special symbol corresponding to the end of a sequence (Schmidhuber , 2012), in the case when the output is a symbol taken from a vocabulary → when that symbol is generated, the sampling process stops. In the training set, we insert this symbol as an extra member of the sequence.
  - Introduce an extra Bernoulli output to the model that represents the decision stop the generation at each time step → more general, because it may be applied to any RNN (generating e.g. real numbers instead of characters).
  - Determine the sequence length τ by adding an extra output to the model that predicts the integer τ itself. The model can sample a value of τ and then sample τ steps → requires adding an extra input to the recurrent update at each time step so that the recurrent update is aware of whether it is near the end of the generated sequence.

Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

Regione Emilia-Romagna

Università degli Studi di Ferrara

Recurrent Neural Networks

- The price recurrent networks pay for their reduced number of parameters is that optimizing the parameters may be difficult.
- The parameter sharing used in recurrent networks relies on the assumption that the same parameters can be used for different time steps.
- Equivalently, the assumption is that the conditional probability distribution over the variables at time t+1 given the variables at time t is **stationary**, meaning that the relationship between the previous time step and the next time step does not depend on t.
- In principle, it would be possible to use t as an extra input at each time step and let the learner discover any time-dependence while sharing as much as it can between different time steps.
- This would already be much better than using a different conditional probability distribution for each t, but the network would then have to extrapolate when faced with new values of t.

Bidirectional RNNs

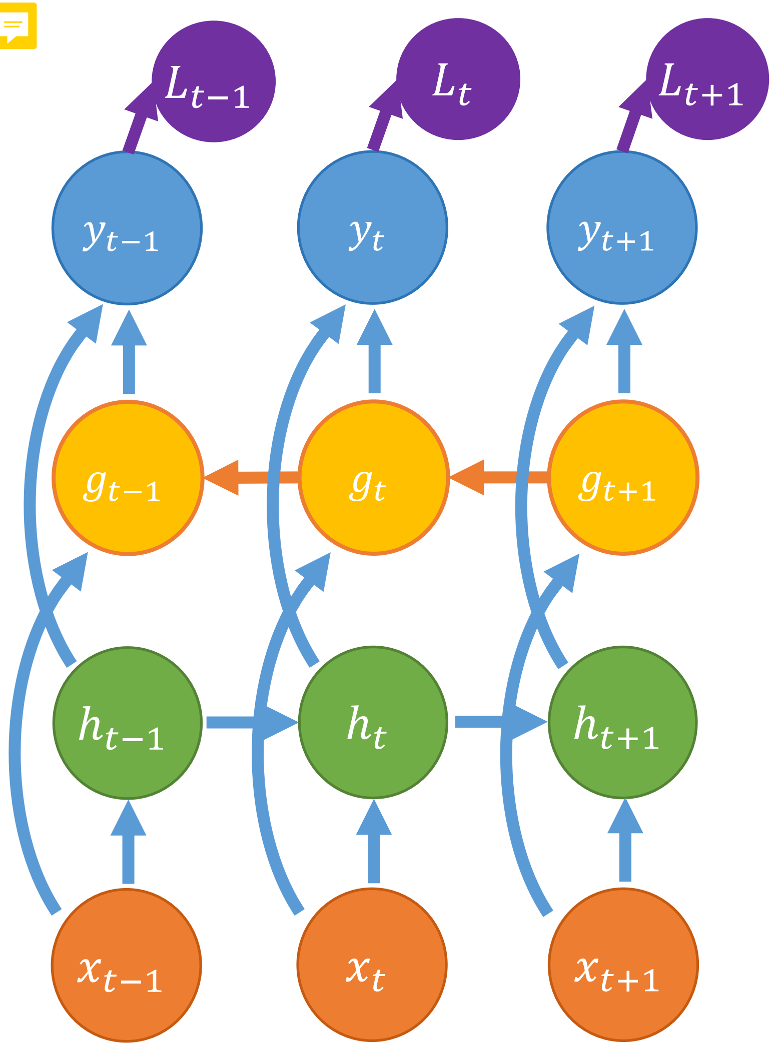
- In many scenarios we may want to predict a value given the entire input sequence, e.g., in speech recognition where knowing the next characters is necessary to correctly spell a word.
- Bidirectional RNNs combine two RNNs: one moves forward and one backward.

Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

Bidirectional RNNs

g propagates information backward in time

h propagates information forward in time



y depends on both past and future but it is more sensitive to the input values around t (we don't know the size of the windows which is not fixed).

Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

Regione Emilia-Romagna

Università degli Studi di Ferrara

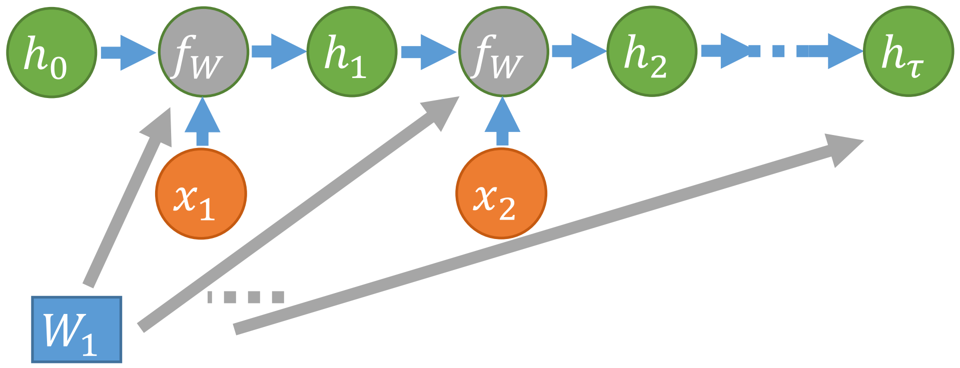
Beyond (Bidirectional) RNNs

- The idea seen before can be extended to 2D input (images) by using 4 directions, i.e., 4 RNNs.
- At each pixel, the output y depends on mostly local information but could also depend on long-range inputs, if the RNN is able to learn to carry that information.
- Compared to a convolutional network, RNNs applied to images are typically more expensive but allow for long-range lateral interactions between features in the same feature map.

Encoder-Decoder
Sequence to Sequence

Many to One + One to Many

Many to one: Encode input sequence in a single vector

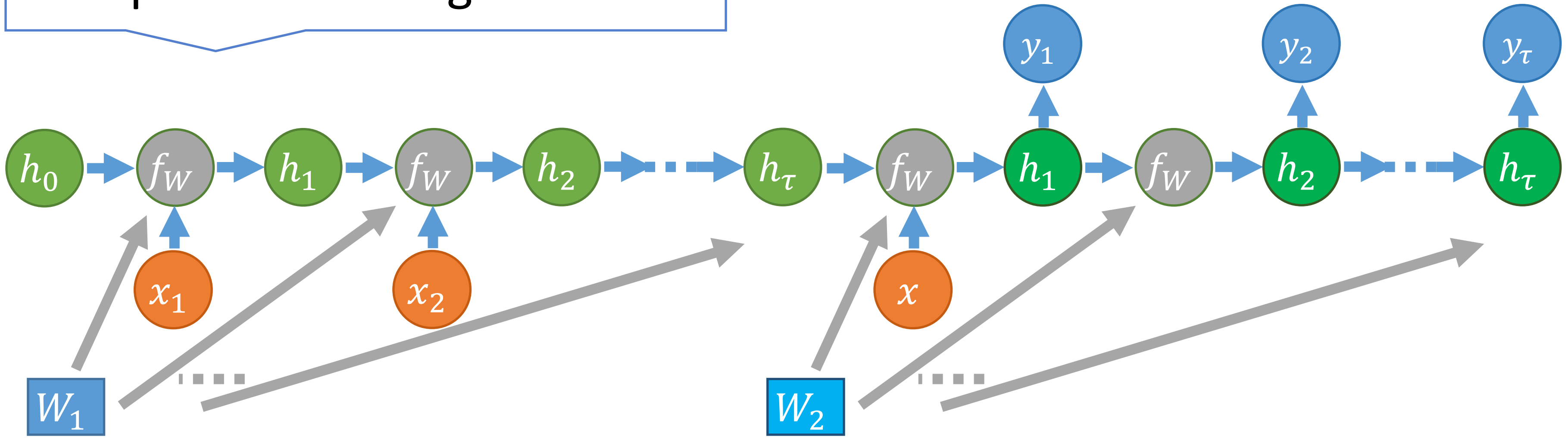


Encoder-Decoder Sequence to Sequence

Many to One + One to Many

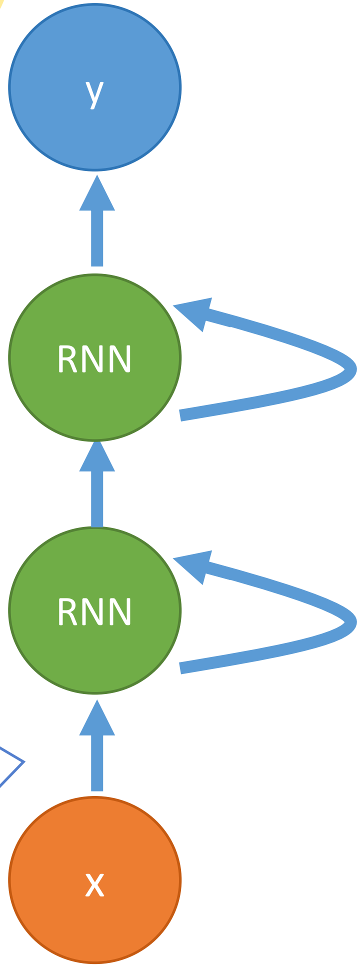
Many to one: Encode input sequence in a single vector

One to many: Produce output sequence from single input vector



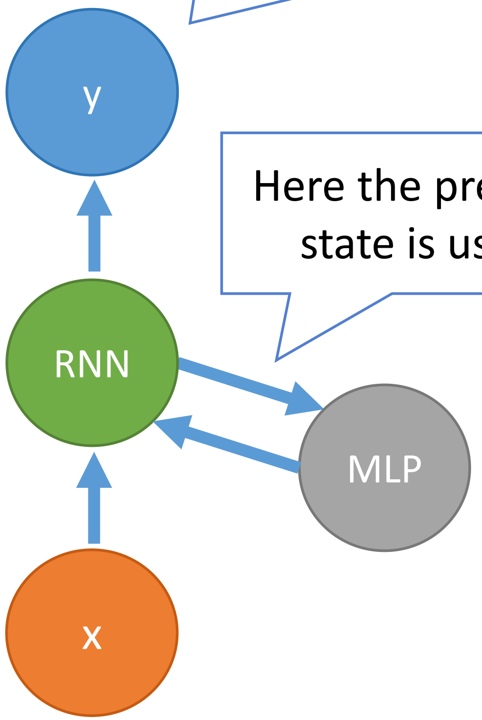
Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

Deep RNNs

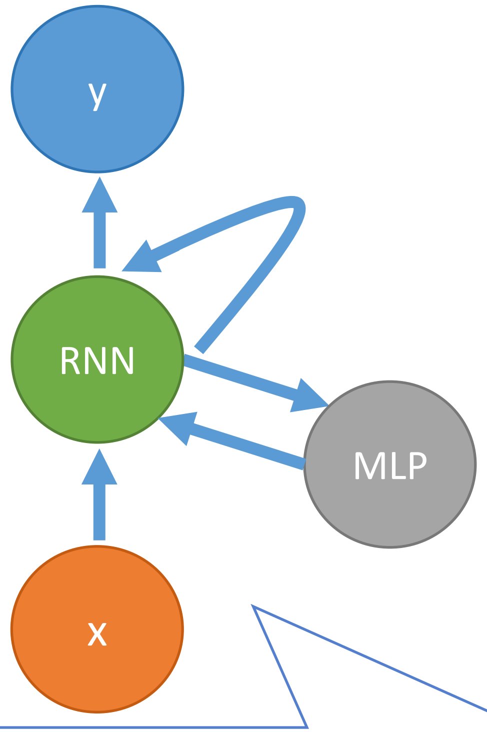


The hidden recurrent state can be broken down into groups organized hierarchically.

Deeper computation (e.g., an MLP) can be introduced in the input-to-hidden, hidden-to-hidden and hidden-to-output parts. This may lengthen the shortest path linking different time steps.



Here the previous state is used.



The path-lengthening effect can be mitigated by introducing skip connections.

Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

Università degli Studi di Ferrara

Possible applications: Image Captioning

- Studied by many researchers:
  - Explain Images with Multimodal Recurrent Neural Networks, Mao et al.
  - Deep Visual Semantic Alignments for Generating Image Descriptions, Karpathy and Fei-Fei
  - Show and Tell: A Neural Image Caption Generator, Vinyals et al.
  - Long term Recurrent Convolutional Networks for Visual Recognition and Description, Donahue et al.
  - Learning a Recurrent Visual Representation for Image Caption Generation, Chen and Zitnick

Image Captioning

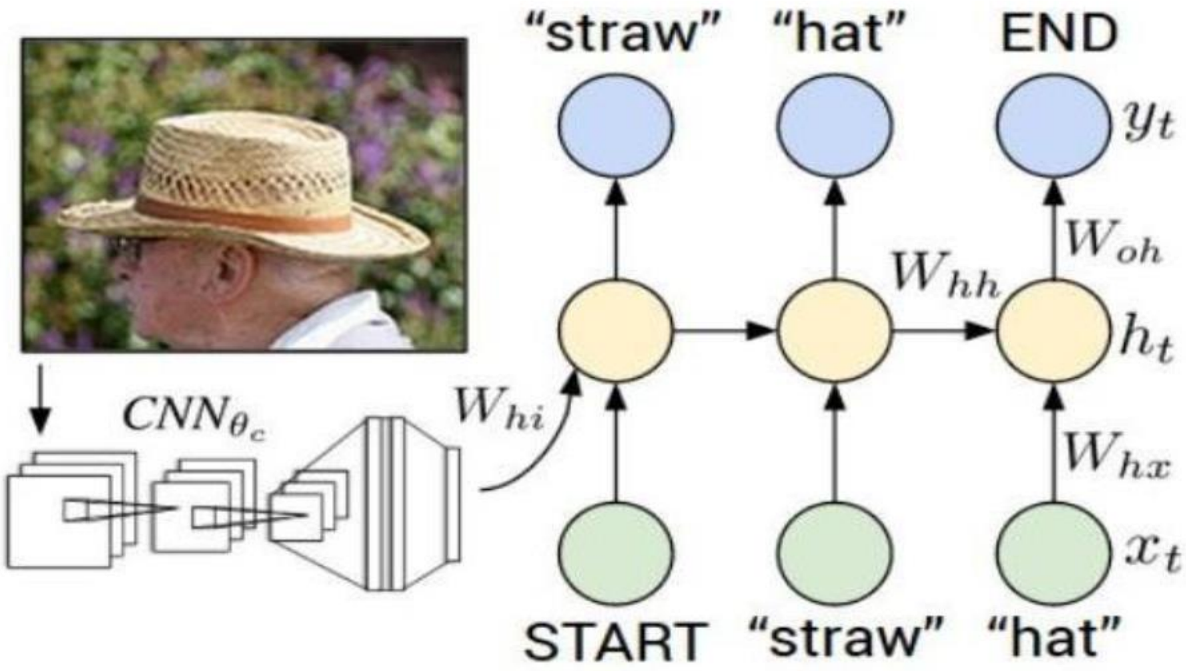


Figure from Karpathy et al, "Deep Visual-Semantic Alignments for Generating Image Descriptions", CVPR 2015; figure copyright IEEE, 2015. Reproduced for educational purposes.

Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

Image Captioning

Recurrent Neural Network

Convolutional Neural Network

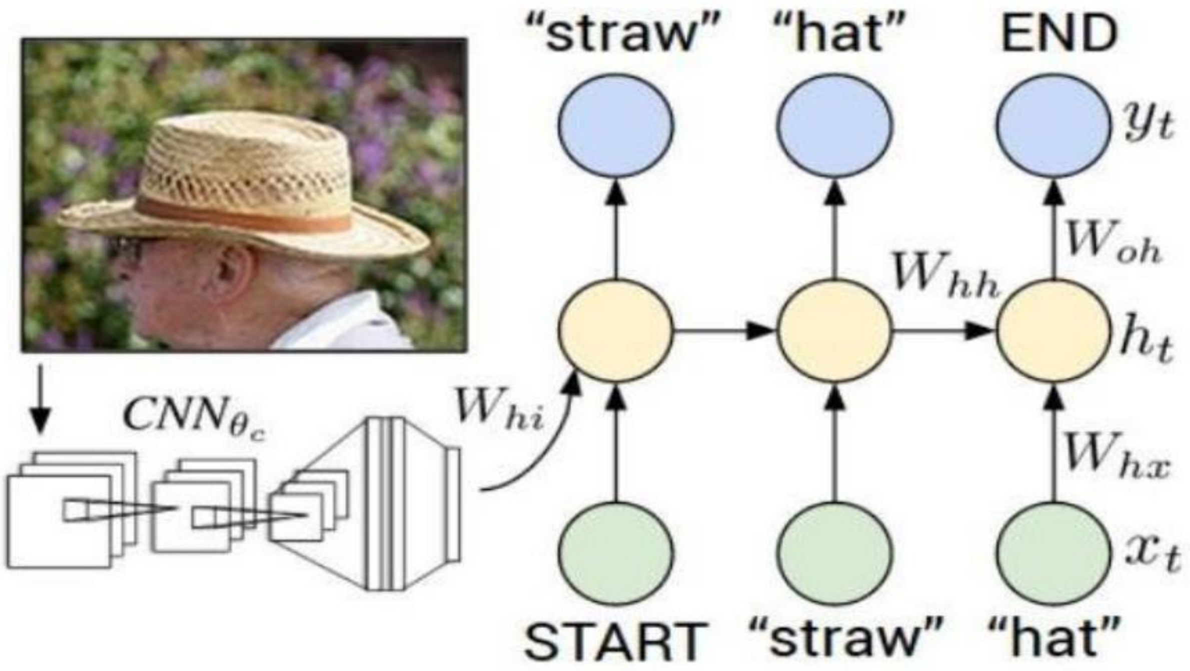
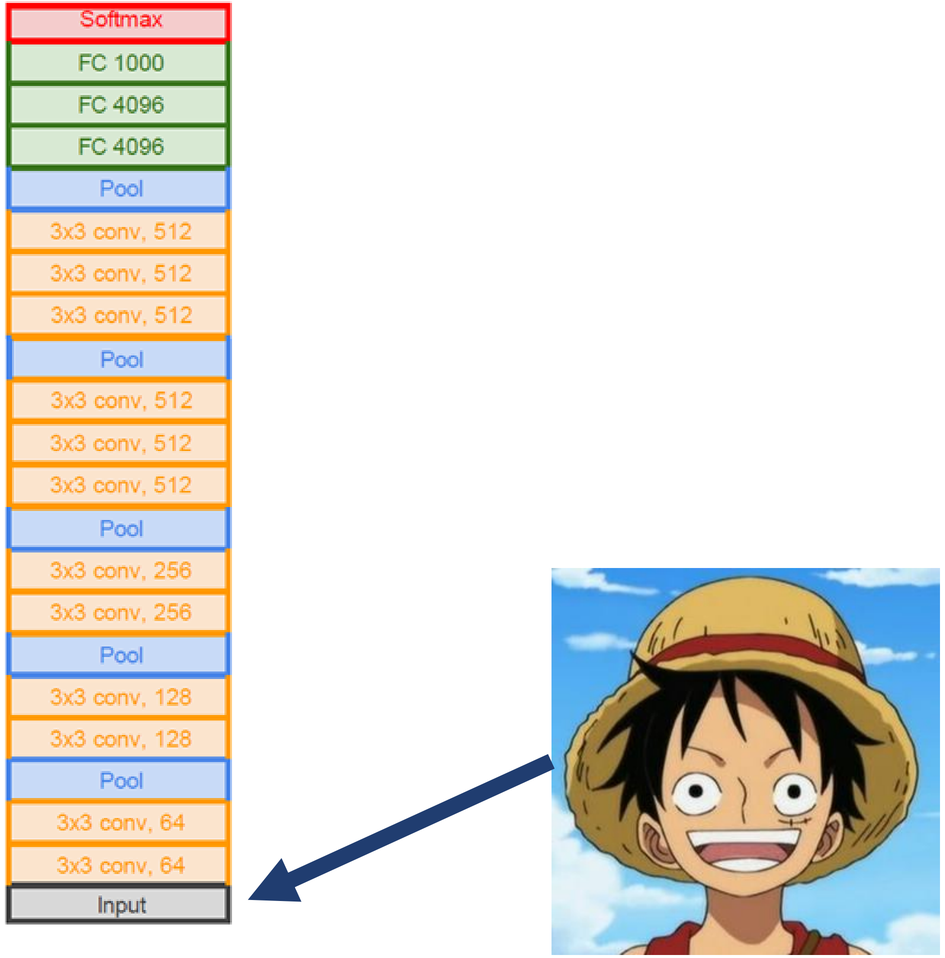


Figure from Karpathy et al, "Deep Visual Semantic Alignments for Generating Image Descriptions", CVPR 2015; figure copyright IEEE, 2015. Reproduced for educational purposes.

Possible Architecture

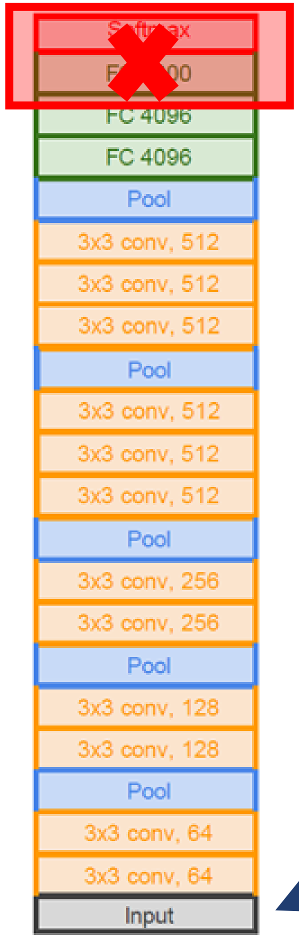


Test Image

Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

Università degli Studi di Ferrara

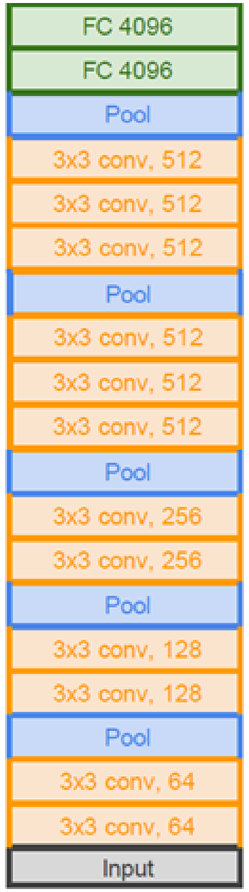
Possible Architecture



Test Image

Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

Possible Architecture



x_0

<START>

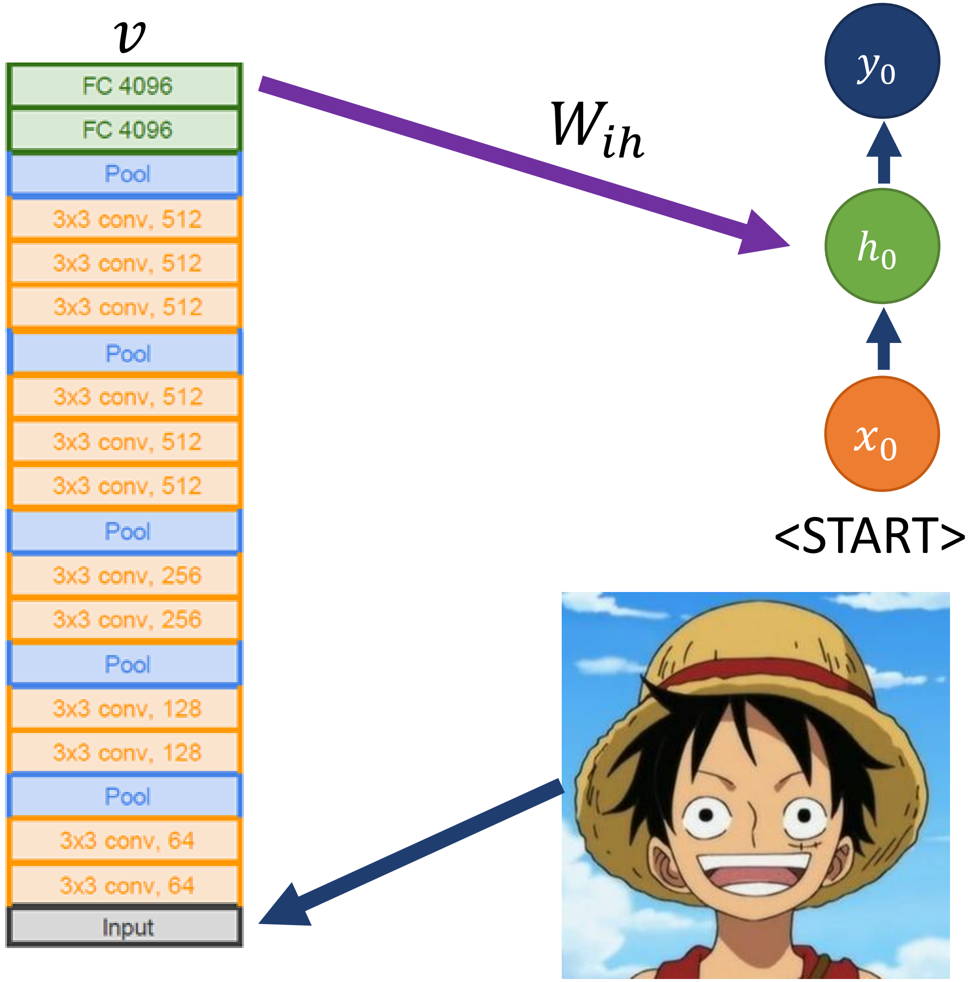
Let's create x_0 containing the token <START>
→ start of the sequence

Test Image

Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

Università degli Studi di Ferrara

Possible Architecture



Test Image

Before:

$$h = \tanh(W_{hh}h_{t-1} + W_{xh}x_t)$$

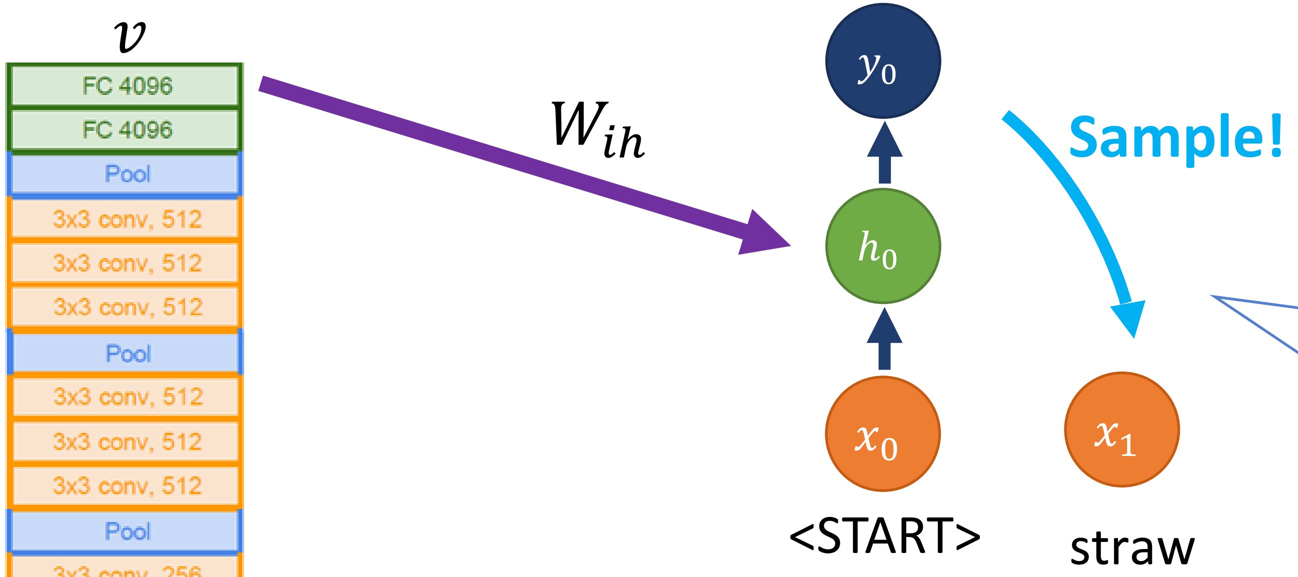
Now:

$$h = \tanh(W_{hh}h_{t-1} + W_{xh}x_t + W_{ih}v)$$

Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

Università degli Studi di Ferrara

Possible Architecture

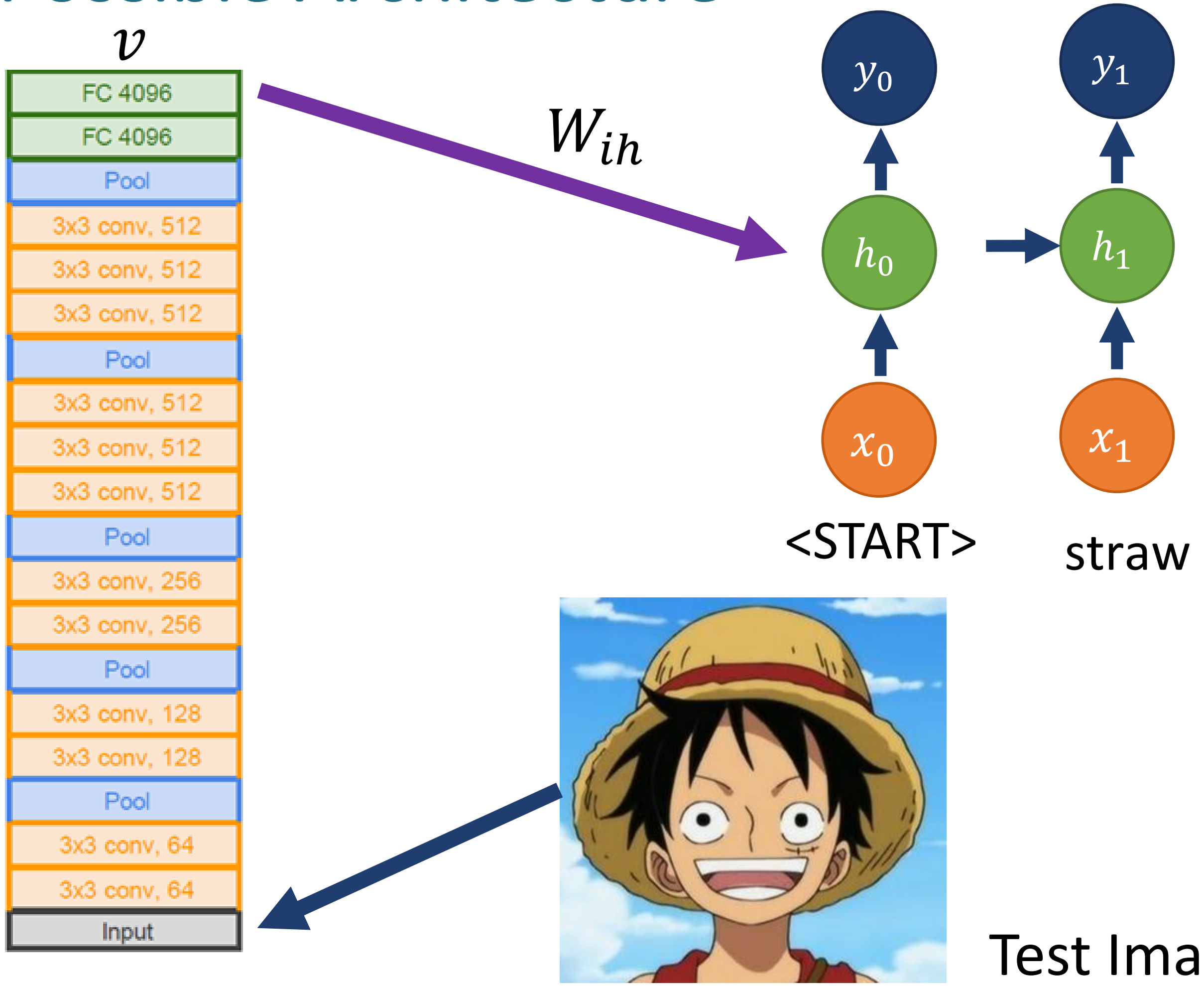


Let's sample the output of the RNN y_0 containing the word "straw" and use it as new input x_1.

Test Image

Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

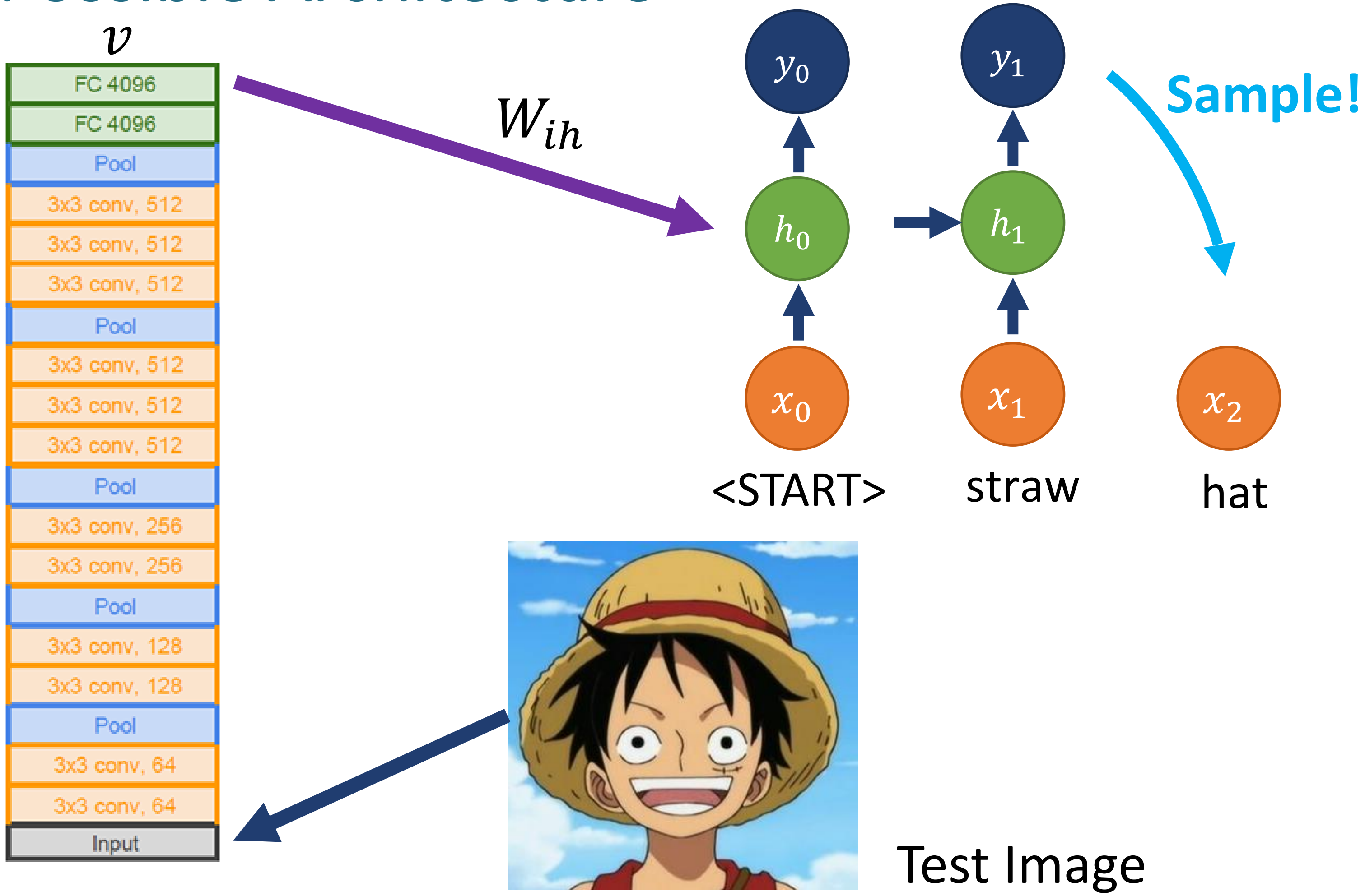
Possible Architecture



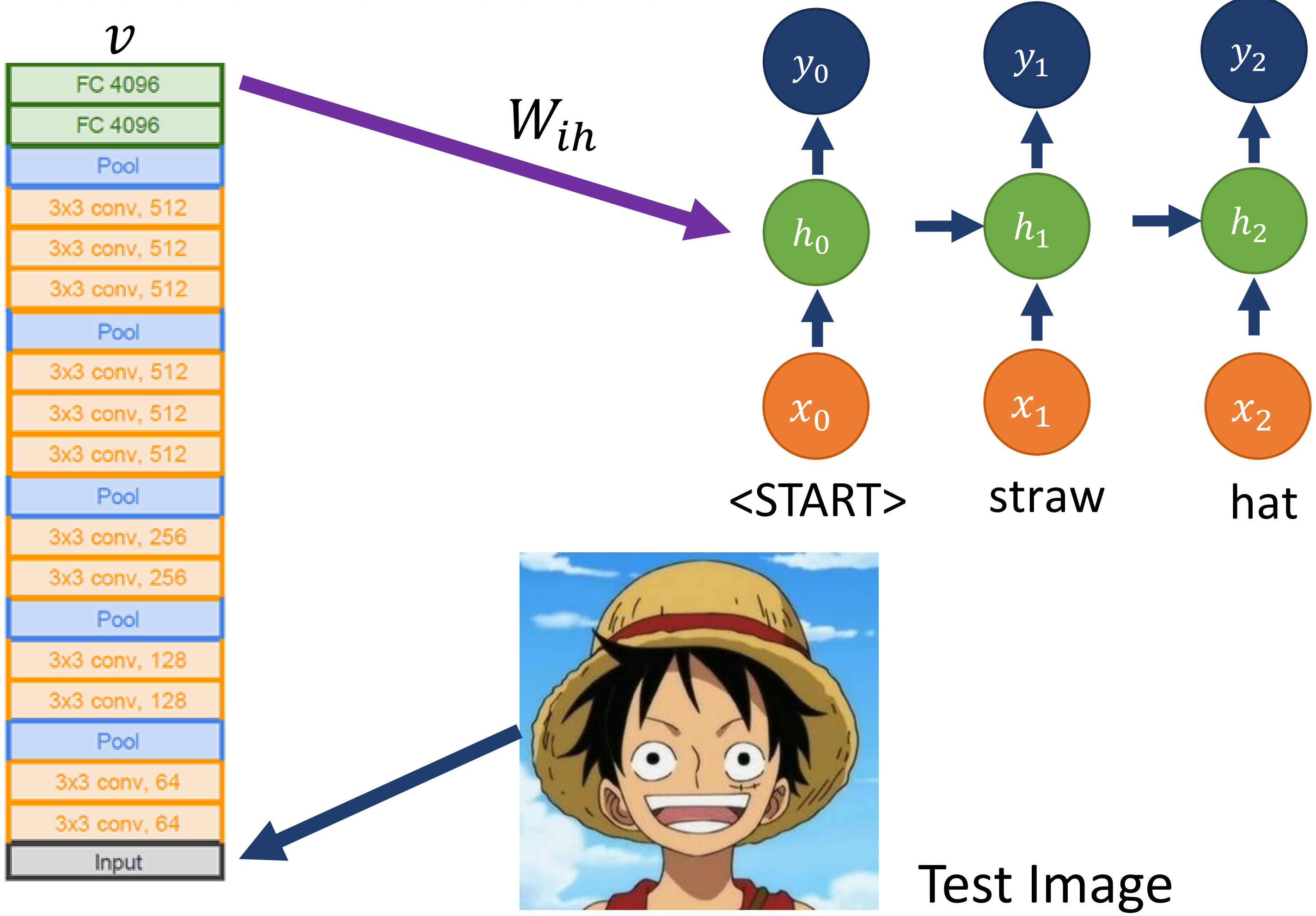
Test Image

Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

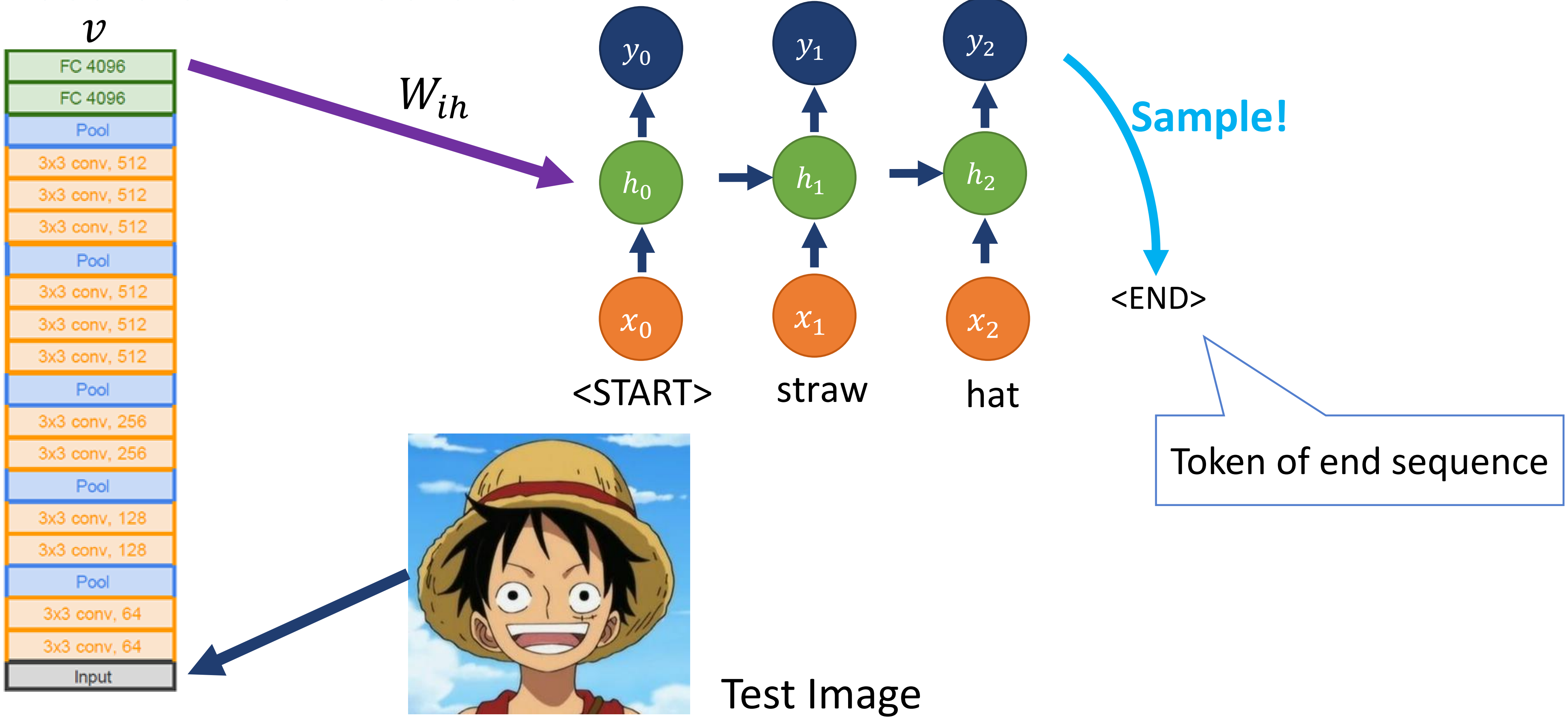
Possible Architecture



Possible Architecture



Possible Architecture



Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

Image Captioning: Example Results

Captions generated using neuraltalk2
All images are CC0 Public domain:
cat suitcase, cat tree , dog, bear,
surfers, tennis, giraffe, motorcycle

A cat sitting on a suitcase on the floor

A cat is sitting on a tree branch

A dog is running in the grass with a frisbee

A white teddy bear sitting in the grass

Two people walking on the beach with surfboards

A tennis player in action on the court

Two giraffes standing in a grassy field

A man riding a dirt bike on a dirt track

Regione Emilia-Romagna

Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021

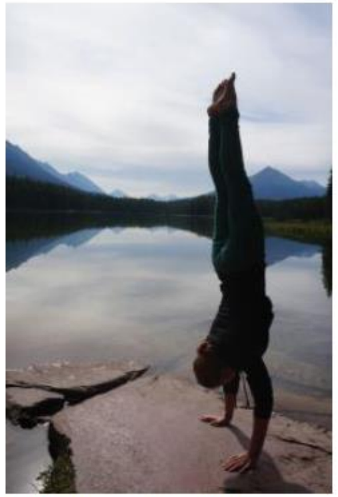
Università degli Studi di Ferrara

Image Captioning: Failure Cases

Captions generated using neuraltalk2
All images are CC0 Public domain:
fur coat, handstand, spider web, baseball

A woman is holding a cat in her hand

A person holding a computer mouse on a desk



A woman standing on a beach holding a surfboard

A bird is perched on a tree branch

A man in a baseball uniform throwing a ball

Progetto di alta formazione in ambito tecnologico economico e culturale per una regione della conoscenza europea e attrattiva approvato e cofinanziato dalla Regione Emilia-Romagna con deliberazione di Giunta regionale n. 1625/2021