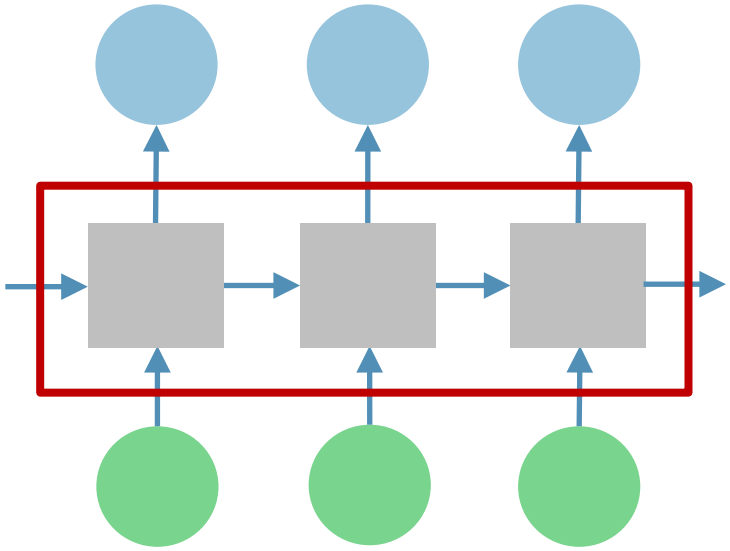
Practical use cases

We're going to dive into the considerations of RNN architecture to pick given our problem.

Previously on this week: Recurrent architecture

Here, we won't care about the implementation of the layers. We just care about the possibilities of the RNN architecture.

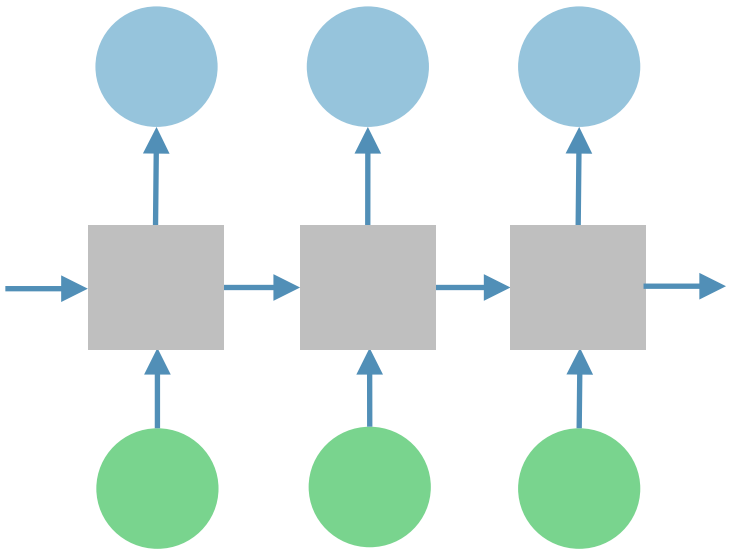


One or more layers:

- Simple RNN
- LSTM
- GRU
- …

Elements-wise classification

For each element of the input sequence, we want to get the label.
Hence element-wise classification.



Input sequence
Output sequence

Input and output are synchronized

Tasks

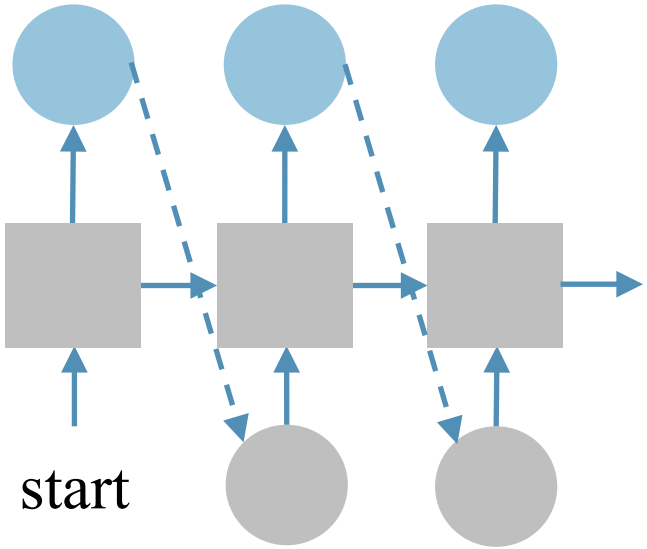
- POS tagging
- Video frames classification

Sequence generation

We use the element-wise output at timestep t as input in timestep t+1.

Input — no input. ---

Output — sequence



'start' is the initial state. Could be some token, OR even the encoding of some picture chopped from a CNN autoencoder.

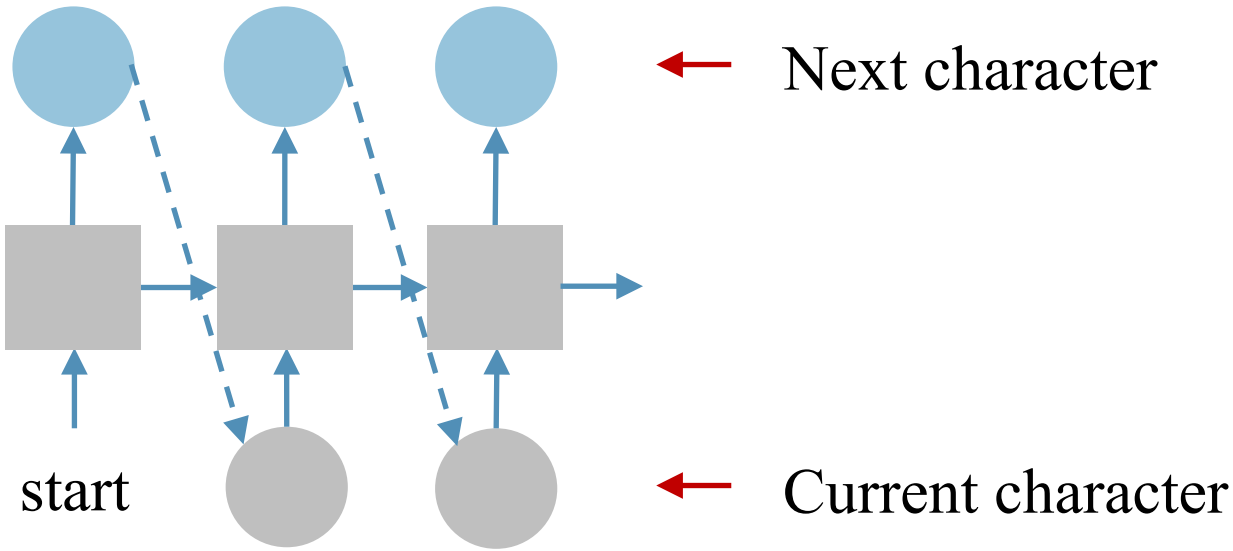
Tasks

- Character-based language model
- Word-based language model
- Music generation
- Speech generation
- Handwriting generation
- …

  Name generation, etc.

Char-based language model: Shakespeare

Model	3 layer LSTM of 512 units
Training data	all the works of Shakespeare



Char-based language model: Shakespeare

Second Senator:
They are away this miseries, produced upon my soul,
Breaking and strongly should be buried, when I perish
The earth and thoughts of many states.

DUKE VINCENTIO:
Well, your wit is in the care of side and that.

Second Lord:
They would be ruled after this chamber, and
my fair nues begun out of the fact, to be conveyed,
Whose noble souls I'll have the heart of the wars.

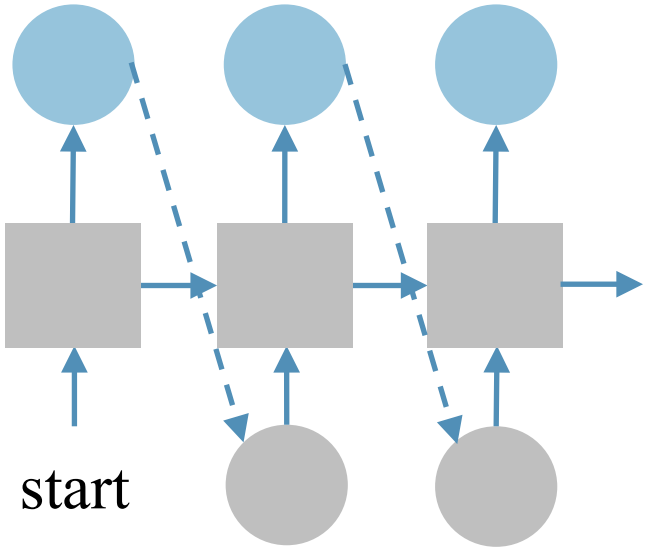
Clown:
Come, sir, I will make did behold your worship.

This is interesting.

Andrej Karpathy, http://karpathy.github.io/2015/05/21/rnn-effectiveness/

Handwriting generation

We predict handwriting point by point



ooo ok.
smart
smart

Next pen position:

$x,\ y$ – mixture of Gaussians

z – Bernoulli distribution

Current pen position:

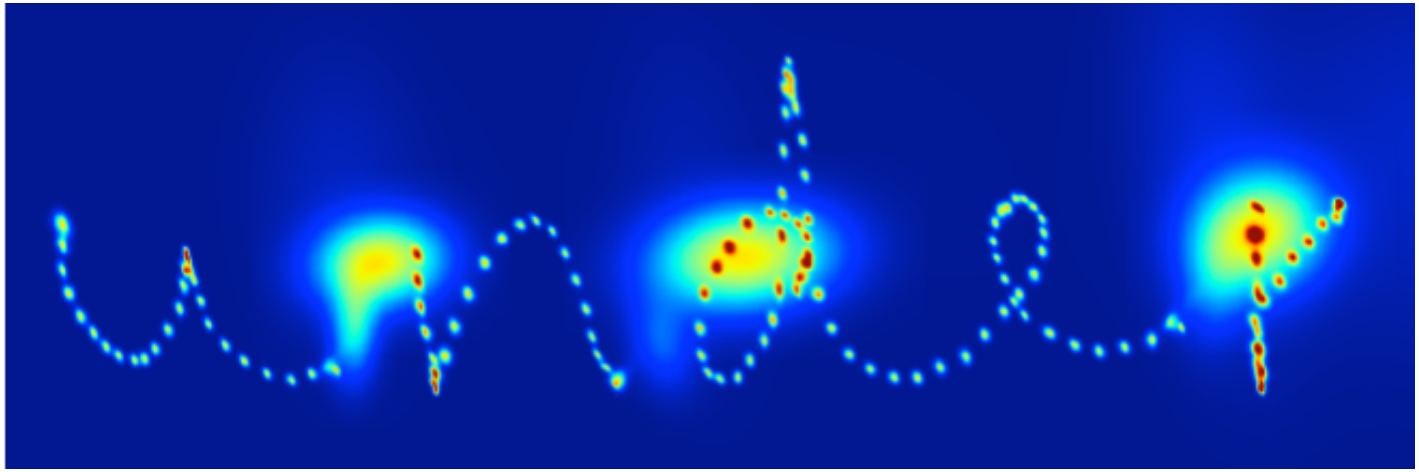
$x,\ y$ – pen offset

z – do we end a stroke here?

z - 0 or 1.

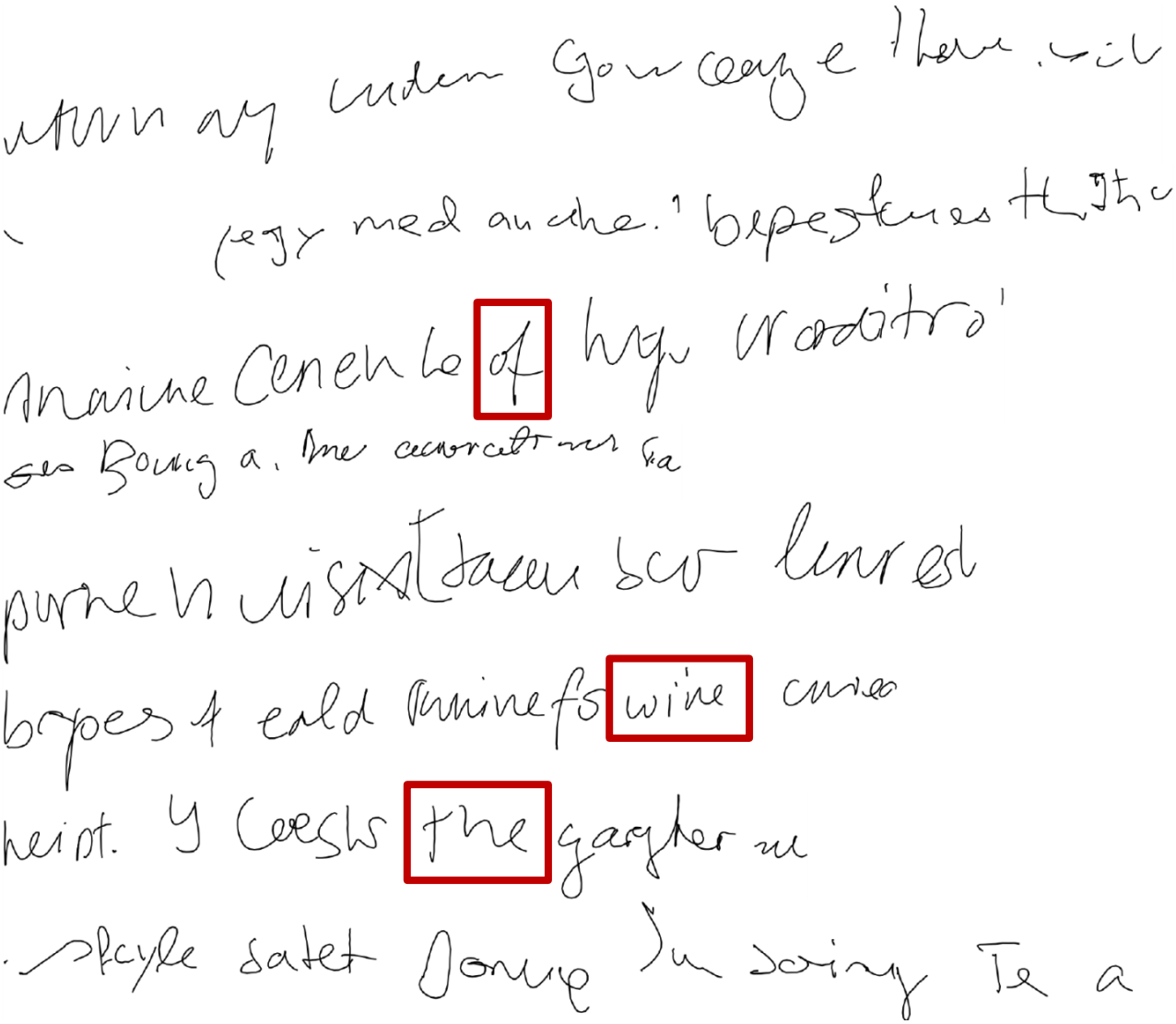
Handwriting generation

We predict handwriting point by point



Alex Graves, https://arxiv.org/pdf/1308.0850.pdf

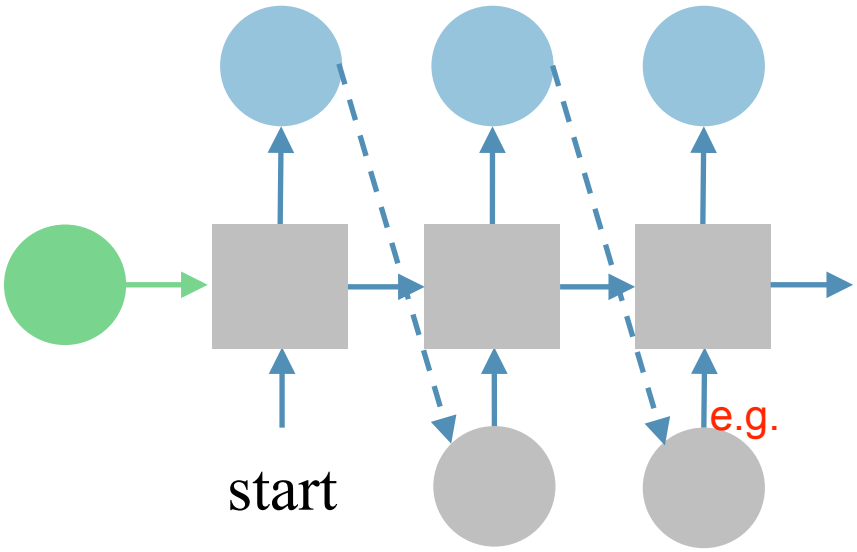
Handwriting generation



Alex Graves, https://arxiv.org/pdf/1308.0850.pdf

Conditional sequence generation

Essentially, we can 'start' sequence generation with some condition (not just by the 'start' token alone like last time but also by some input object)



Input some object

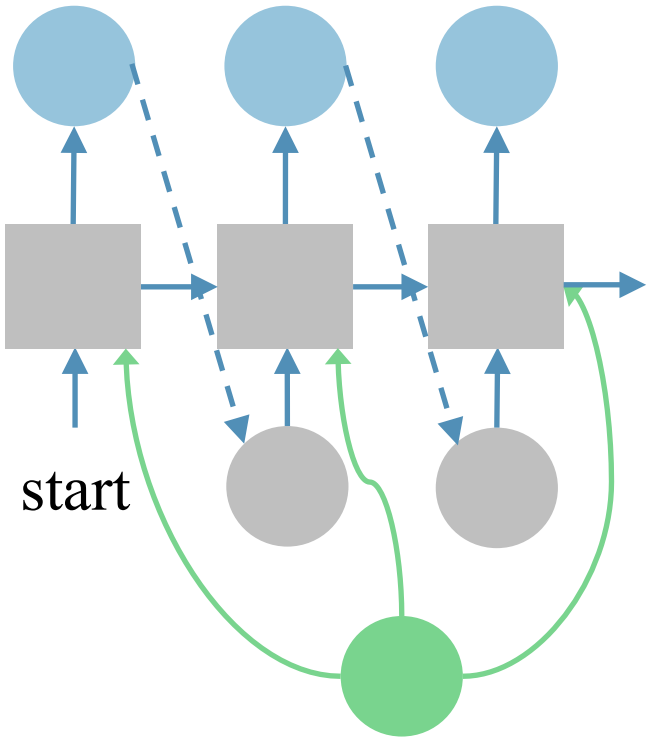
Output sequence

Tasks

- Speech generation
- Handwriting generation
- Image captioning
- …

Conditional sequence generation

Now we can condition the sequence generator on ALL timesteps, not just the start.



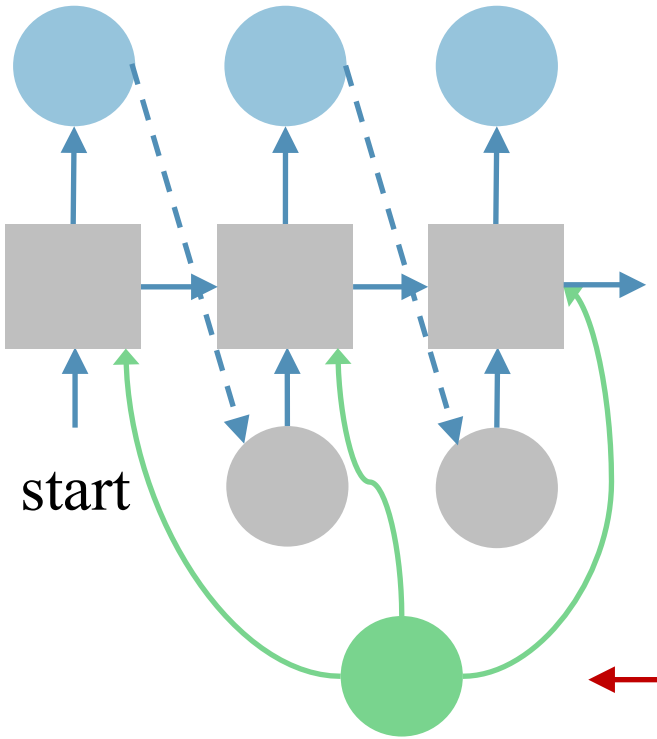
Input some object

Output sequence

Tasks

- Speech generation
- Handwriting generation
- Image captioning
- …

Conditional handwriting generation



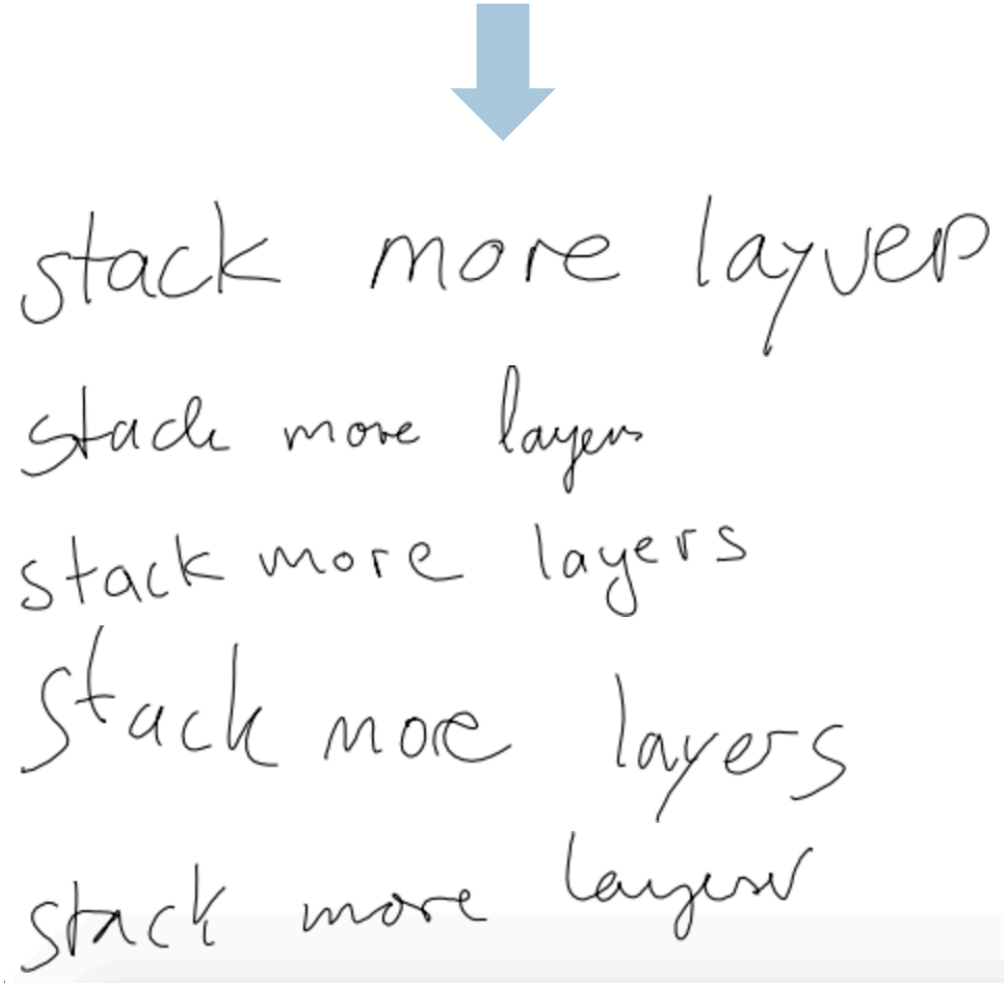
Next pen position

Current pen position

Window: which part of text we are writing now

Conditional handwriting generation

stack more layers



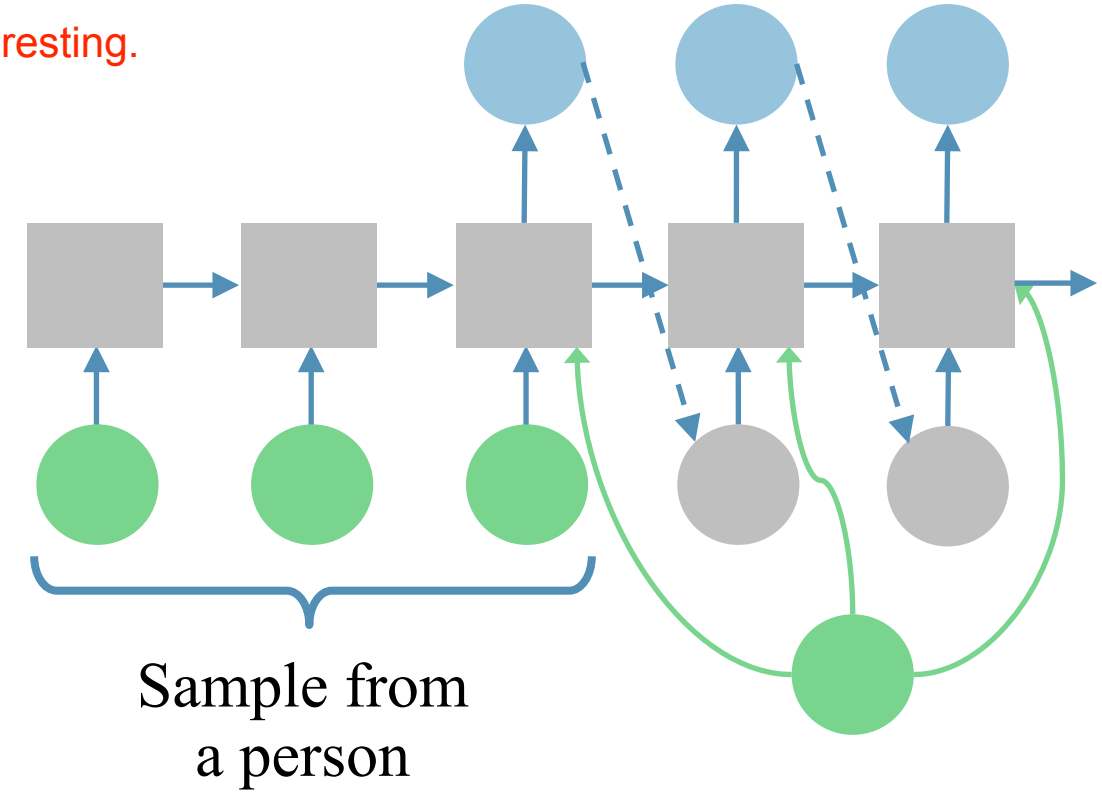
Alex Graves, http://www.cs.toronto.edu/~graves/handwriting.html

Conditional handwriting generation

We can 'prime' the generator to sample some person's handwriting, and then perhaps 'copy' their writing style.

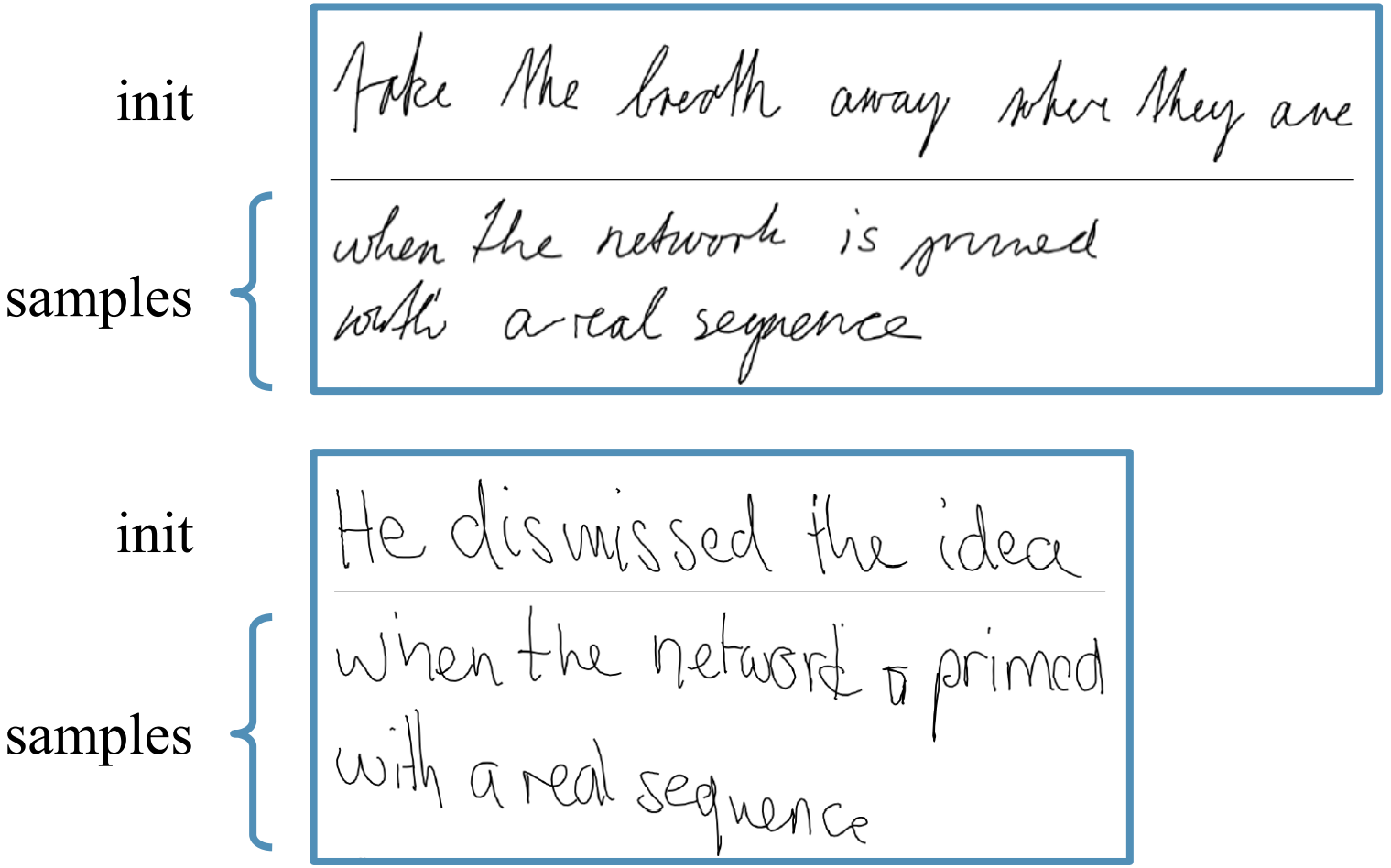
VERY interesting.

Primed sampling



Conditional handwriting generation

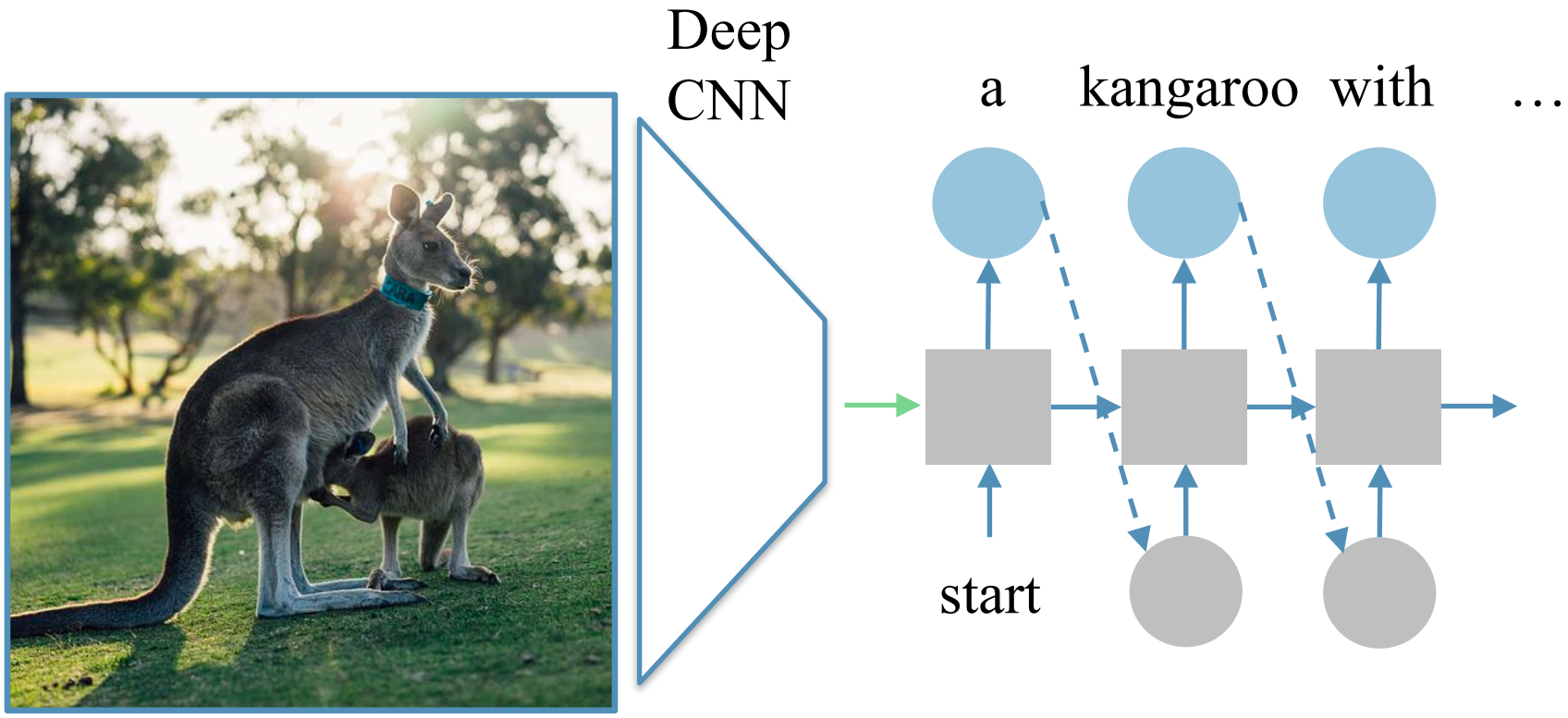
Primed sampling



Alex Graves, https://arxiv.org/pdf/1308.0850.pdf

Image Captioning

This prompts the Generative RNN to create some caption.



^ InceptionV3 Chopped

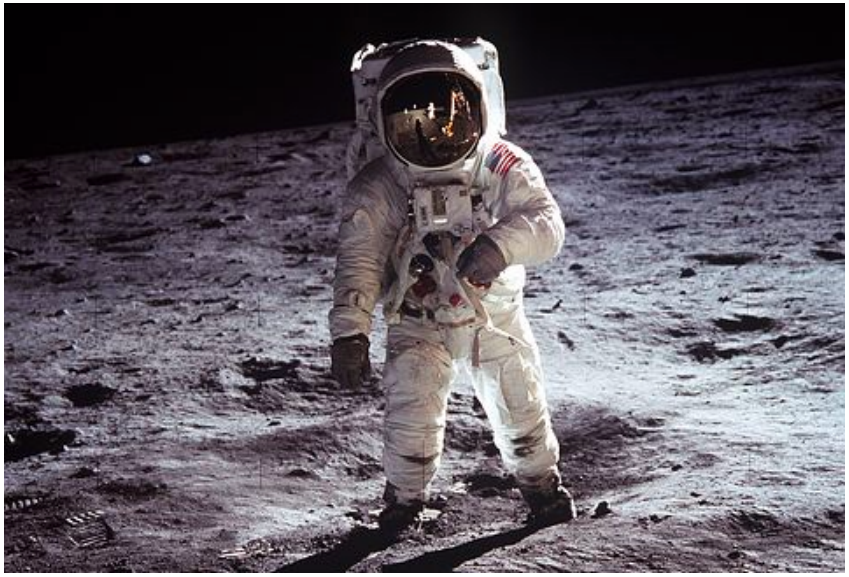
Image Captioning: good examples

a man riding a wave on a surfboard

a large brown bear walking across a river

Image Captioning: bad examples

a man riding on the back of a boat



a man riding a snowboard

Image Captioning: bad examples

a man is holding a tennis racket on a tennis court

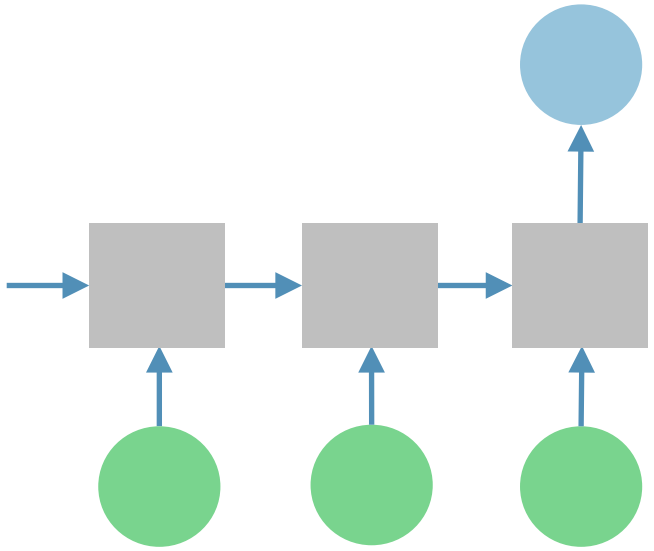
a man is holding a kite in the air

Sequence classification

predict one whole label for the entire sequence.

Input sequence

Output one label



One output at the end

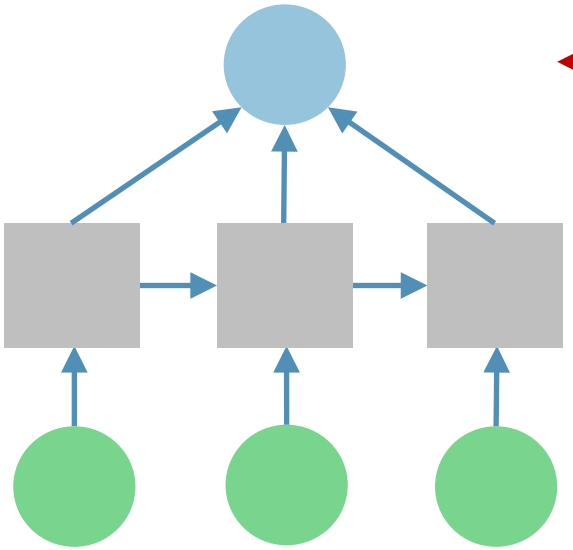
Tasks

- Sentiment analysis
- …

Sequence classification

Input sequence

Output one label



Dynamic mean or max pooling
+ Attention mechanism

^ an attention layer may be useful.
Read up more about this.

Tasks

- Sentiment analysis
- …

Sequence translation

Input sequence
Output sequence

Tasks

- Handwriting to text / text to handwriting
- Speech to text / text to speech
- Machine translation

interesting

Input and output …

- are NOT synchronized
- may have different length
- may have different order

Sequence translation

Input sequence

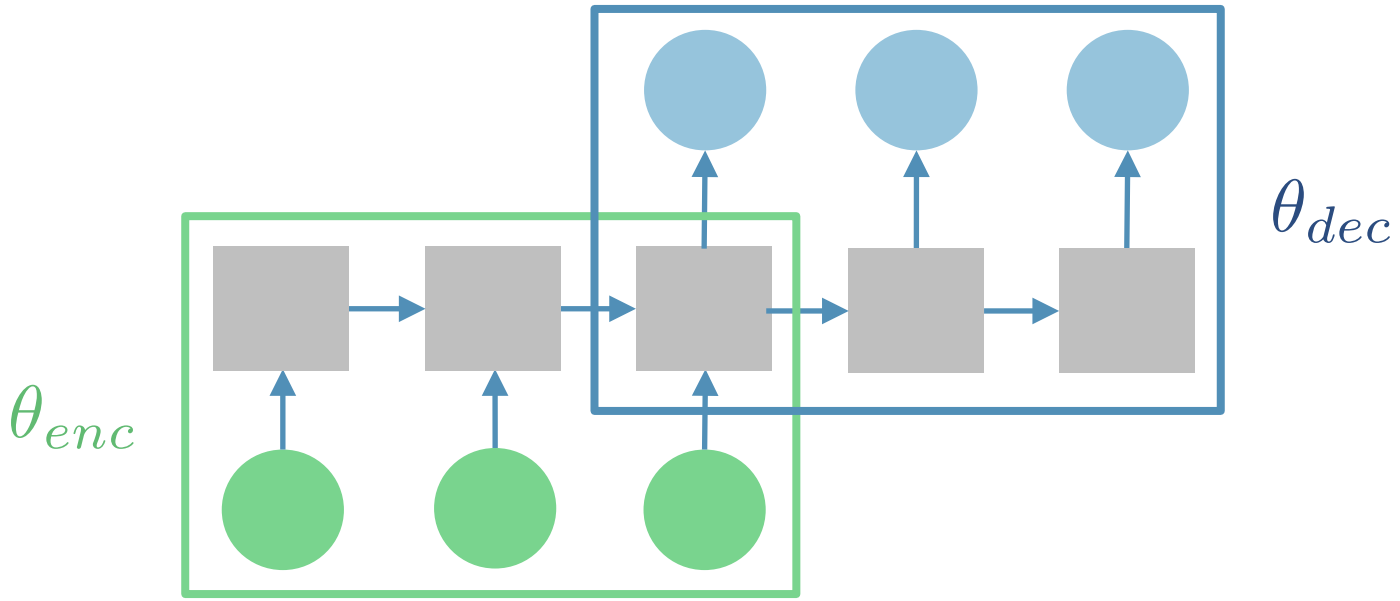
Output sequence

Tasks

- Machine translation

enc and dec should have different WEIGHTS as they are doing different things.

Input and output have different order



Summary

We have learned how to use recurrent networks for:

- Element-wise sequence classification
- Sequence generation: unconditional and conditional
- Sequence classification
- Sequence translation

Good luck with the final project!