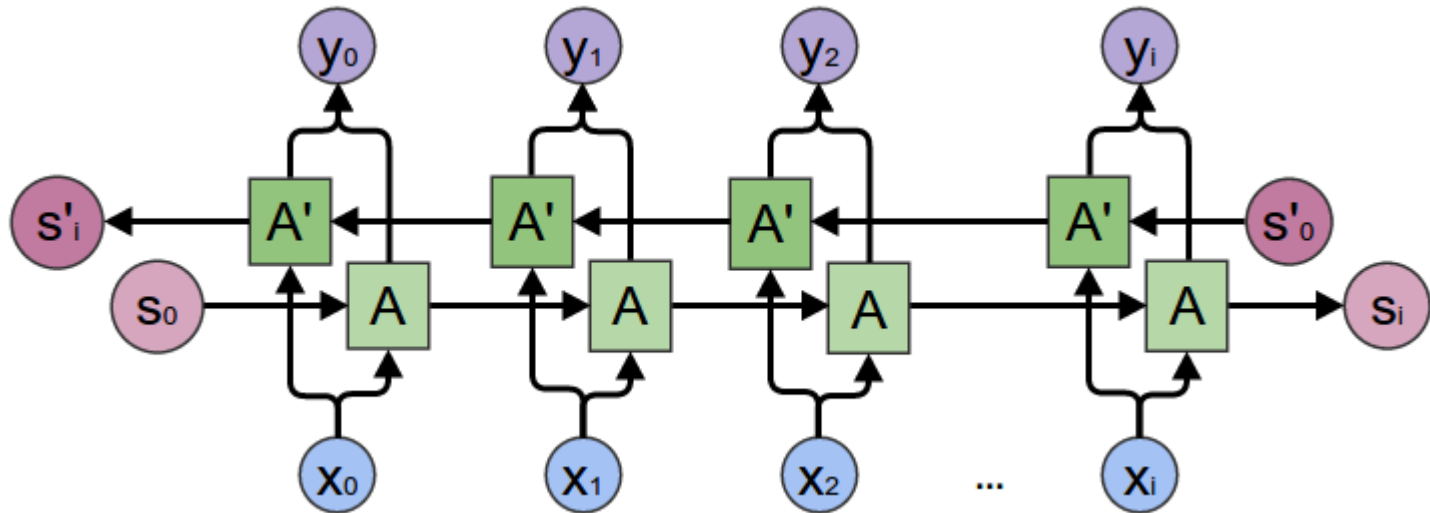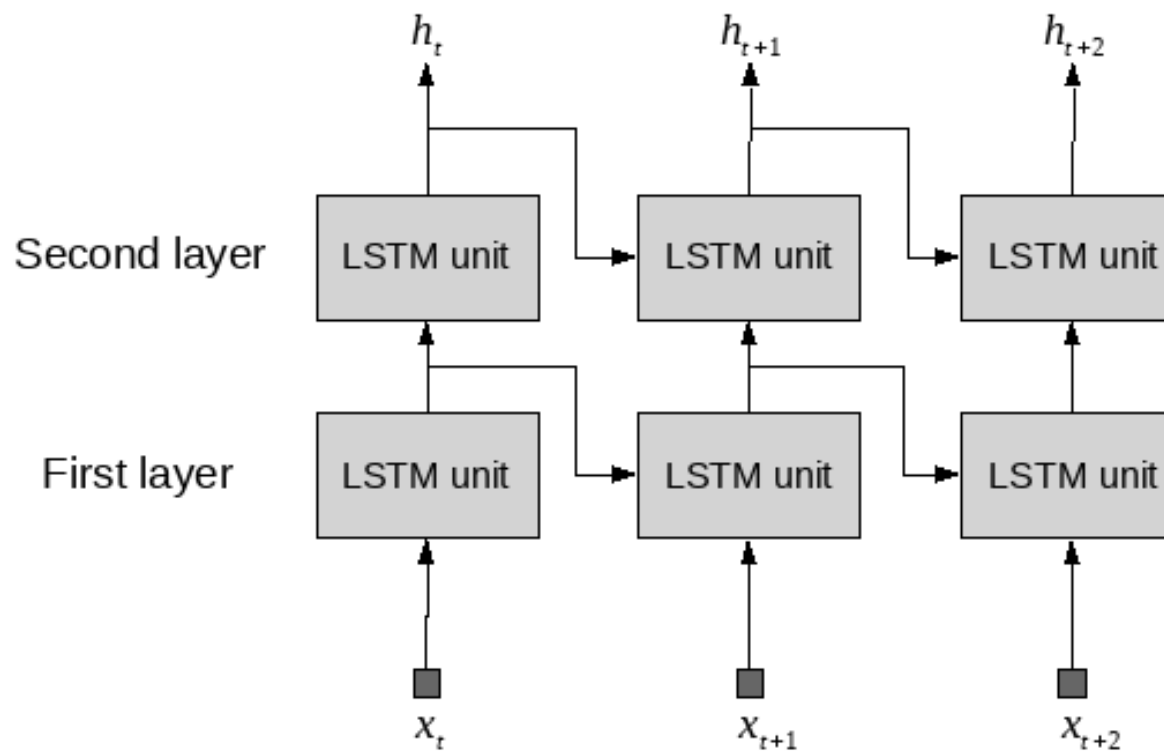# Bidirectional LSTMs

Putting two independent LSTMs together, one for the ==forward sequence== and one for the ==reverse sequence==.

Useful when the context of the input is needed.

Multilayer LSTMs

CNN LSTM architectures

These architectures combine the CNN for dealing with images and LSTM for the sequence part. Examples:

Activity recognition: image sequence to classification
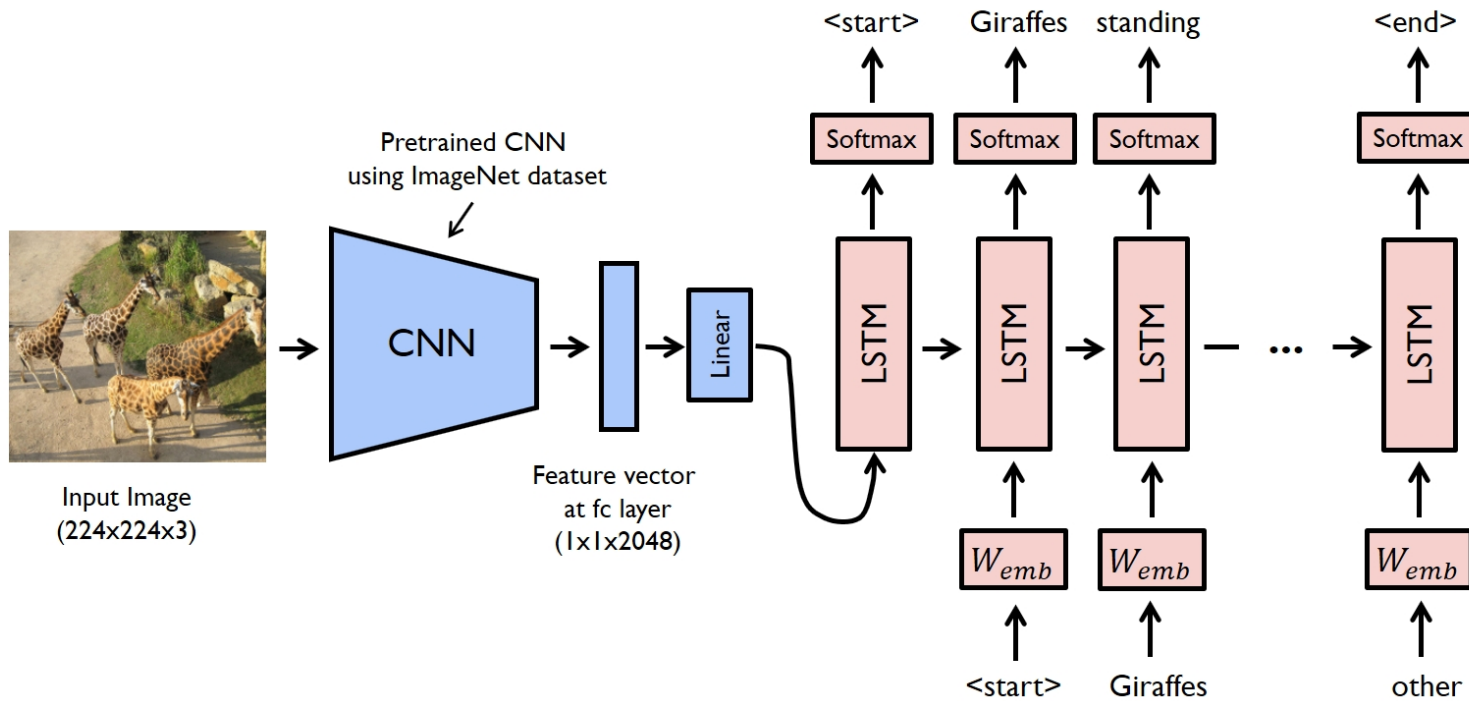**Image description:** image to text sequence (caption)
Video description: video sequence to sequence (description)

Two "early" papers:

Show and Tell: A Neural Image Caption Generator (Link)

Long-term Recurrent Convolutional Networks for
Visual Recognition and Description (Link)

Example of a caption generator:

A person riding a
motorcycle on a dirt road.

Two dogs play in the grass.

A skateboarder does a trick
on a ramp.

A dog is jumping to catch a
frisbee.

A group of young people
playing a game of frisbee.

Two hockey players are
fighting over the puck.

A little girl in a pink hat is
blowing bubbles.

A refrigerator filled with lots of
food and drinks.

A herd of elephants walking
across a dry grass field.

A close up of a cat laying
on a couch.

A red motorcycle parked on the
side of the road.

A yellow school bus parked
in a parking lot.

| Describes without errors | Describes with minor errors | Somewhat related to the image | Unrelated to the image |