

Figure 36-15. Gated recurrent unit

Recurrent Neural Networks Applied to Sequence Problems

Recurrent neural networks have many application areas for using LSTM models for sequence tasks. A couple of problems under this domain include sentiment analysis, machine translation, image captioning, video captioning, and voice recognition. As mentioned earlier, these problems can be modeled as a one-to-many model, a many-to-one model, or a many-to-many model. The section will survey a few LSTM architectures for tackling/modeling sequence problems:

- Long-term recurrent convolutional neural network, also known as CNN LSTM
- Encoder-Decoder LSTMs
- Bidirectional recurrent neural networks

Long-Term Recurrent Convolutional Network (LRCN)

The long-term recurrent convolutional network (LRCN) is a unique neural network architecture for generating descriptions of images and videos (which is seen as a sequence of images). These problems can be termed as visual time series modeling. The LRCN architecture combines the ability of the convolutional neural network (CNN) to extract image features together with a recurrent network for learning sequences or long-term dependencies. The LRCN passes visual inputs into a CNN to retrieve image features as outputs. These outputs are then passed into a recurrent LSTM network layer to generate the natural language descriptions. The recurrent layer can contain stacked LSTMs.

One core advantage of LRCN for modeling sequential vision problems such as image captioning and video captioning is that the network is not constrained to fixed lengths of inputs and outputs. Hence, it can be used to model sequential data with different lengths such as textual data and videos.

The following illustrations show how LRCN is applied to a variety of sequence problems:

1. **Image captioning:** Image captioning can be seen as a one-to-many sequence problem. The input is an image and therefore a static input, and the output is a sequence of text that describes the objects in the image; this is a sequential output. The use of LRCN for image captioning is illustrated in Figure 36-16.

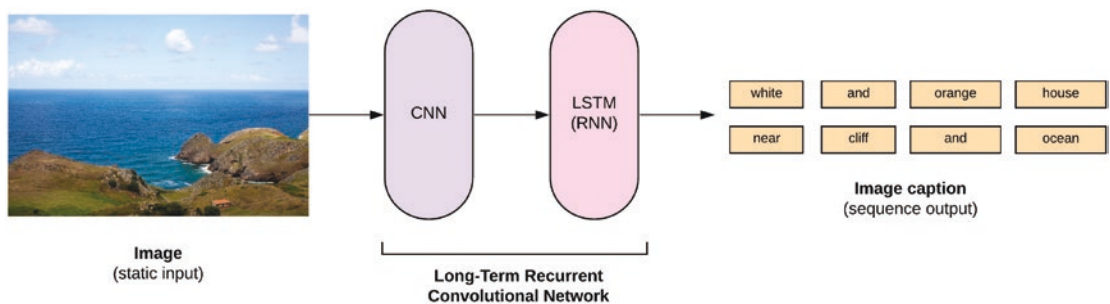


Figure 36-16. Image captioning (photo by Daniel Llorente on Unsplash)