An attentional unit can also function as an interface between networks, for instance between CNN and RNN.

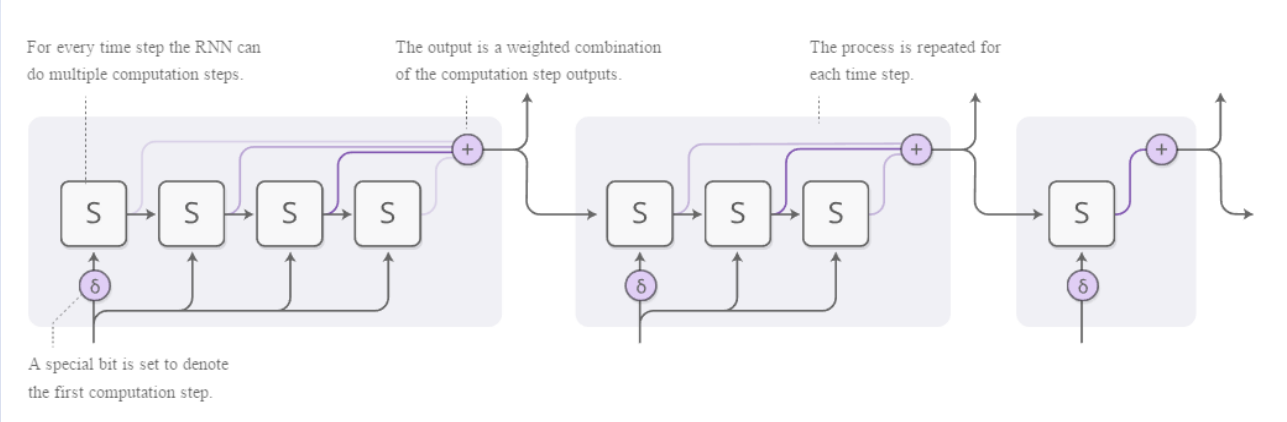
One of the common uses of attentional interfaces is for captioning. First, a CNN extracts relevant regions of an image and their features. Then, an RNN focuses on the regions and features extracted by the CNN to produce a description of the image.

Adaptive Computation Time (ACT)

This is an algorithm enabling RNNs to learn how many computational steps should be performed between receipt of an input and producing an output.

The motivation is the understanding that not every input or every step of processing (correlated to vector locations) justifies or requires the same computational effort. This is intuitive to human thinking: the more difficult the task, the more thought and/or effort it requires.

The idea is simple: for the network to learn how many computational steps should be undertaken, we want the number of steps to be represented by a differentiable function. We do this, yet again, using the same trick; rather than deciding what the number of steps is, we perform the computation on a variety of numbers-of-steps, and the output will be the weighted average of the values (each sub-network performed a different number of steps).



The only open source implementation of Adaptive Computation Time at the moment seems to be Mark Neumann's (TensorFlow).