Conventional encoder-decoder approach for neural machine learning encodes a whole input sentence into a fixed-length vector and then a translation is decoded from it. It becomes problematic when the sentence is very long. In this paper, a novel architecture is proposed to address this issue by allowing a model (soft-search) for a set of *annotations* (mapped from input words). Thus, the model need not to encode a whole source sentence into a fixed-length vector. Moreover, it focuses on information relevant to the generation of the next target word. By *jointly training alignment and translation*, the ability to translate longer sentence is improved.

The proposed model utilizes an RNN encoder-decoder as the underlying framework. On one hand, the encoder reads the whole input sentence (a sequence of word vectors) into a context vector c, where hidden states and some nonlinear functions are included for computation. On the other hand, the decoder is usually trained to predict the next word given the context vector c and all previously predicted words. And it tries to maximize the probability over the transition as the serial product of conditional probabilities.

As an extension, the proposed model develops a novel architecture learning to align and translate at the same time, which frees the itself from encoding the whole source sentence into a fixed-length context vector. The encoder utilizes *bidirectional* RNN units, and the decoder emulates searching through the source sentence.

(1) The encoder maps input words into a sequence of annotations h, which summarizes both preceding and following words by concatenating bidirectional hidden states. Thanks to the tendency of RNNs to better represent recent inputs, the annotation hj will be focused on the words *around* word xj.

(2) The decoder computes the conditional probability of yi given previously predicted words with the source sentences, which intakes si (RNN hidden state for time i) and context vector ci. Notice that the context vector here is distinct for each yi and dependent on the annotation sequence, which is different from the conventional model. the context vector ci for yi (the output at position i) is computed as a weight sum of annotations. To decide the value of weights, an alignment model is used. Based on annotation hj of the source sentence and RNN hidden state si-1, the alignment model scores how well the inputs around position j and the output at position i *match*. Indeed, it implements a mechanism of attention. By parametrizing as a feedforward neural network, the alignment model is *jointly* *trained* with the whole transition model.

Experiments are conducted to compared conventional RNN model (RNNencdec) and the proposed model (RNNsearch). The task is to translate English to French. The proposed RNNsearch outperforms RNNencdec for different source sentence lengths. Also, the proposed model is more robust to sentence length. Moreover, qualitative analysis shows that the model is capable to align target word with the relevant annotations/words by visualizing the learned (soft-)alignment. Furthermore, the model performance is comparable to phrase-based statistical machine translation, which indicates a promising step toward better understanding of natural languages.