"Neural Machine Translation by Jointly Learning to Align and Translate"

**Basic Idea of Article**

BILSTM is used as encoder attention layer

Attention Decoder is designed to automatically (soft-)search for parts of a source sentence that are relevant to predicting a target word.

**Limitations**

The paper talks about how traditional translation models struggle with longer sentences because they encode everything into one fixed-length chunk. To fix this, the authors propose a new model that focuses on important parts of the sentence as it translates, instead of cramming everything into a single block.

One problem with this paper is that it mainly looks at translating English to French. While it shows big improvements for this, it doesn't explore how well it works for other languages or translation tasks. It would be better if they tested it on a wider range of languages and situations.

The new model they suggest adds more complexity compared to older ones. It's fancy because it can pick out the important parts of a sentence as it goes, instead of just crunching everything at once. But this fanciness comes with a cost - it needs more computer power and training time. So, while it might make translations better, it's also harder to run.

Another issue is how it deals with words it doesn't know or ones that are rare. Even though the model gets better at handling longer sentences and picking out important bits, it still struggles with words it hasn't seen before. This is a big problem because real-world translations often include words that aren't common or even ones the model hasn't learned.

In short, while the new translation model sounds cool and improves things for English-to-French, it still has some kinks to work out. It needs to be tested on more languages, it's complicated to run, and it still stumbles over new or rare words.