**Neural Machine Translation By Jointly Learning To Align And Translate**

1. **Problem Addressed**: Traditional neural machine translation (NMT) models used a fixed-length vector to encode the source sentence, which limited their performance, especially on long sentences.
2. **Main Contribution**: The paper introduces an attention mechanism that allows the model to focus on different parts of the source sentence while generating each word in the translation.
3. **Soft Alignment**: Unlike hard alignment in traditional systems, the model learns a soft alignment, where it calculates a weighted average over all source words to determine which parts to focus on.
4. **Architecture Overview**: It uses a bidirectional RNN as the encoder and a decoder that, at each step, looks back at the encoder outputs using attention to inform word generation.
5. **Dynamic Context Vector**: For every target word, the decoder dynamically creates a context vector based on the most relevant source words instead of relying on a single fixed vector.
6. **Performance**: The model outperforms standard encoder-decoder NMT on English–French translation tasks and achieves results competitive with traditional phrase-based translation systems.
7. **Visualization**: The paper shows attention maps that visually represent how the model aligns source and target words during translation, demonstrating interpretability.
8. **Better Long Sentence Handling**: Because it doesn’t compress the entire input into one vector, the model is particularly better at translating longer and more complex sentences.
9. **Impact**: This attention mechanism laid the foundation for future models like the Transformer, which later removed RNNs entirely and relied solely on attention for translation and other tasks.