CS447 Literature Review: Can convolutional networks replace recurrent networks (RNN, LSTM, GRU) in encoder-decoder architecture for Neural Machine Translation?

Pallaw Kumar

December 13, 2022

1 Introduction

Natural Language Processing (a.k.a., NLP) is a sub-field of Artificial Intelligence that has close links with Machine Learning. In NLP, the research areas in focus are Machine Translation (MT), Semantic Parsing, Sentiment Analysis, Topic Modeling, Named Entity Recognition (NER), Text Analytics, etc. Deep Learning techniques with Natural Language Processing, such as Neural Machine Translation (NMT) find a lot of applications in domains such as Healthcare, Finance, Manufacturing, Education, Retail, and customer service Kalyanathaya et al. (2019). NMT is a new highly active approach for machine translation, which has shown promising results and due to its success, it has attracted many researchers in the field. Language processing-related impacts research is driving fuel to innovate many tools to build industrial applications.

In this paper, the question of interest is related to neural machine translation. In its vanilla form, a neural machine translation is sequence-to-sequence modeling using an encoder-decoder architecture where an input sentence is first encoded by the encoder, later using this information, the decoder translates to another form, to achieve translation between language pairs such as English-German and vice-versa. In existing research works the recurrent networks such as Recurrent Neural Networks (RNNs) & Long-Short Time Memory (LSTM) are extensively used for building encoder-decoder foundational structures but, recent advancements Meng et al. (2015), Su et al. (2016), Gehring et al. (2017a), Gehring et al. (2017b), Singh et al. (2017), Wang and Xu (2017), Watanabe et al. (2017) are shown using convolutional networks based structures in neural machine translation tasks given better results not only in terms of accuracy but also reducing learning and decoding time drastically. Thereby, such developments raise a pertinent question, ”Can convolutional networks (CNNs) replace recurrent networks (RNN, LSTM, Gated recurrent units (GRU)) in encoder-decoder structures for Neural Machine Translation tasks?”

In this paper, we are going to discuss the above seven paper methodologies, findings, and relevance in detail to understand why it seems promising to query more on the research question of interest. The paper is organized in the following sequences - Section 2 explains the motivation factors behind asking the research question of interest, Section 3 briefs existing neural networks architectures, Section 4 explains the motivation behind the included paper, Section 5 summarizes the above seven research papers, Section 6 provides a work assessment and Section 7 highlights their links with the research question if any, and finally, Section 8 concludes the findings.

2 Motivation

Convolution neural networks are extensively used network types for computer vision problems because of its convolution properties that emphasize more on local features seen in visual data. Using CNNs-based architectures is seen to have better processing speed due to its simplicity to work with hardware such as CUDAS - found to be time efficient to work on big data as this architecture is designed to support parallel processing. Architectures such as RNN, LSTM, and GRU are useful for working with sequence-to-sequence models to solve natural language translation however, it does not support parallel processing. As the result, the existing neural machine translator takes longer training time as they have recurrent networks based encoder-decoder combination. Studying papers discussed in section 4 indicates existing NMT architectures are not fully utilized on large datasets due to their slow processing power. It is hopeful that convolutional networks-based encoder-decoder architectures will achieve faster machine translation. Thus, the motivation is to study if all existing RNNs based architectures can be mechanized using only CNNs would be a breakthrough advancement in neural machine translation.

3 Background

3.1 Convolution Neural Network (CNN)

Convolution neural network (CNN) is the most popular neural network architecture that has gained the most success for computer vision tasks and predictive classification of visual data across various domains ranging from finance, to radiology Yamashita et al. (2018). CNNs are the most utilized deep learning network type and are consumed as foundational components in building the most powerful architectures such as the AlexNet network and High-Resolution network (HR.Net) Alzubaidi et al. (2021). CNNs are fused with the latest transformers to work with challenging vision datasets with considerably less computational cost Guo et al. (2022). A typical CNN architecture is a combination of multi-layers convolution layers followed by max-pooling layers (for down-sampling, optional) and non-linear units such as GLU or ReLU with a flat layer before the output layer. Fig 1. shows the ImageNet classification architecture built using convolutional neural networks.

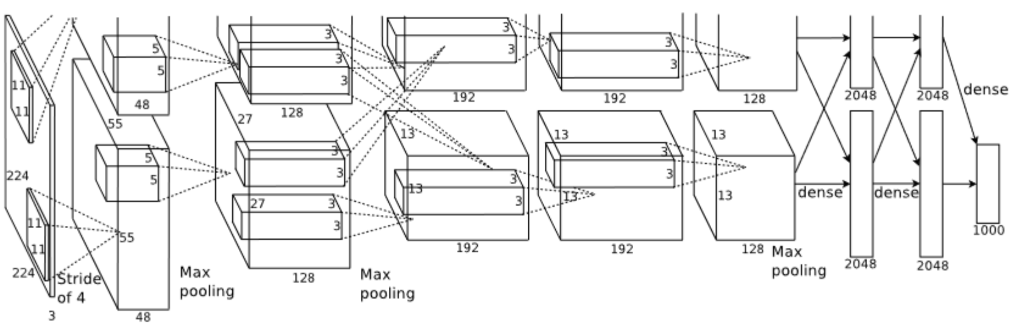

Figure 1: Source: Krizhevsky et al. (2012): ImageNet Classification Architecture - an example of CNN architecure.

3.2 Recurrent Neural Network (RNN)

Recurrent neural networks (RNNs) are a class of neural networks that allow previous outputs to be used as inputs while having hidden states. RNNs are a vanilla form of the recurrent network family (RNN, LSTM, GRU). This family of networks is extensively used for natural language processing. Fig. 2 shows a typical RNN network with inputs X_k, outputs y_k, and hidden units h_k where suffix k represents time information.

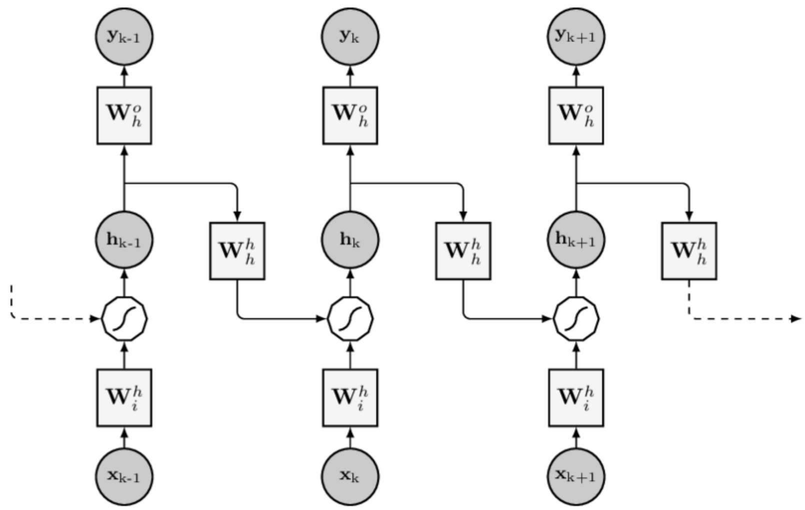

Figure 2: Source: Bianchi et al. (2017): A typical RNN architecture with sequence-to-sequence inputs X, nodes weights W and output Y.

3.3 Neural Machine Translation (NMT)

Neural Machine Translation (NMT) is a sub-branch of machine translation (MT) techniques used to model natural language. Encoder and decoder are the two components that makeup NMT models. The task of the encoder is to create a true vector representation of the sentence termed the summary vector or context vector, which captures all of the vital aspects of the sentence. The perfect context vector would be able to accurately represent all the details in the source sentence as real vectors. This context vector is parsed by the decoder to create the target language sentence word by word, transferring all of the meaning from the source phrase. Translation can be modeled at different levels such as document, sentence, word, etc. NTM architectures are varied based on application by adding additional layers in a combination of encoder and decoder architectures. Fig. 3 shows a typical autoregressive NMT model using the encoder-decoder framework with an additional embedding layer.

4 The motivation of included papers

The motivation for including the seven papers mentioned in Section 1 is that these papers are related to natural language processing (NLP) and worked using convolutional neural networks which is related to the research question asked in this paper. The papers have worked on either improving the existing approaches or proposing new mechanisms using convolutional neural networks. Table

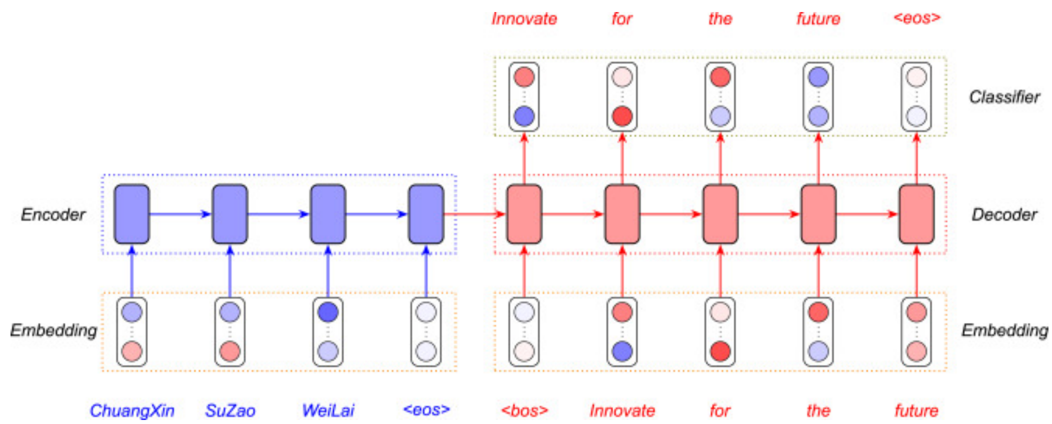

Figure 3: Source: Tan et al. (2020): An overview of the NMT architecture, which consists of embedding layers, a classification layer, an encoder network, and a decoder network. We use different colors to distinguish different languages.

S.N.	Included Research Papers	Natural Language Processing Research Areas
1	Meng et al. (2015)	Neural Machine Translation
2	Su et al. (2016)	Bilingual Semantic Modeling
3	Gehring et al. (2017a)	Neural Machine Translation
4	Gehring et al. (2017b)	Neural Machine Translation
5	Singh et al. (2017)	Neural Machine Translation
6	Wang and Xu (2017)	Word Segmentation
7	Watanabe et al. (2017)	Neural Machine Translation

Table 1: List of papers linked with the research question in this article. Five of them are on Neural Machine Translation (NMT) research areas, one of them on unsupervised learning-based word segmentation and one focuses on semantic parsing area. The commonness across all these papers is that they have some sort of improvements related to Convolutional Neural Networks (CNNs).

1 listed the paper discussed in this article - five of the research papers focus on is Neural Machine Translation (NMT) research area by making convolutional networks-based improvements, and the other two papers are related to areas: word segmentation and bilingual semantic modeling which is a field of Statistics Machine Translation (SMT) also proposing CNN based improvements. Therefore, all of these papers propose convolutional network-based solutions which is the main focus of the research question.

5 Summarization of Papers

Paper-1: Gehring et al. (2017a) presents a faster and relatively simple convolution neural network-based model for neural linguistic translation prior modeled using bi-directional LSTM. The proposed architecture speeds up the decoding process more than twice, achieving similar accuracy on the WMT'16 English-Romanian dataset and better accuracy on the WMT'15 English-Germany dataset compared to the start-of-art.