DEEP LEARNING

SESSION NO:

TOPIC: ATTENTION NETWORKS

INTRODUCTION

- Attention networks, also known as attention mechanisms or attention mechanisms in neural networks, are a crucial component in modern deep learning models, especially in tasks involving sequences or sets of data.
- They allow the model to focus on different parts of the input when making predictions or generating output.
- The key idea behind attention networks is to assign different weights to different elements of the input sequence, indicating their importance or relevance for the current task.
- These weights are learned during the training process and are used to compute a weighted sum of the input elements, which is then used for further processing.

HOW ATTENTION MECHANISM WAS INTRODUCED IN DEEP LEARNING

- The attention mechanism was introduced in the context of deep learning to address the limitations of traditional neural networks in handling sequential or variable-length data, such as sentences or time series.
- It was first proposed in the domain of machine translation.
- Before the introduction of attention, traditional sequence-to-sequence models (like the ones based on Recurrent Neural Networks or Long Short-Term Memory networks) would encode an entire input sequence into a fixed-size context vector, which would then be used to generate the output sequence.
- This approach had limitations, particularly when dealing with long sequences, as it was difficult for the network to capture all the relevant information in a single fixed-size vector.

ATTENTION MECHANISMS

- **The problem:** Long input sequences and images are often difficult for models to process accurately.
- Each element of an input sequence is turned into a hidden state in an encoder to be fed into the next element. During the decoding process, only the last hidden state with some weighted component is used to set the context for the corresponding element of the output sequence.

The solution:

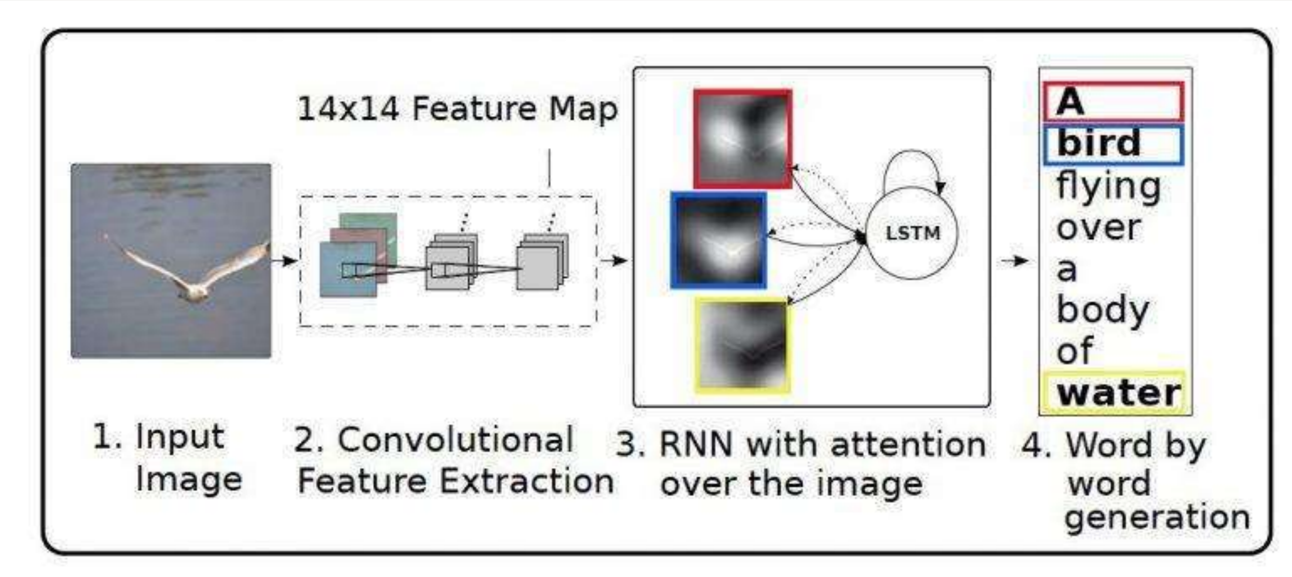
- With an attention model, the hidden states of the input sequence are all retained and utilized during the decoding process.
- A unique mapping is created between each time step of the decoder output and the encoder input.
- Each element of the output sequence coming out of the decoder has access to the entire input sequence to select the appropriate elements for the output.

ATTENTION MECHANISM OVER IMAGES (IMAGE CAPTIONING)

- Instead of compressing a complete image into a static form, the Attention technique dynamically brings important elements to the forefront when they are needed.
- **The processing:** The encoder-decoder image captioning system would encode the image a hidden state. An LSTM would then use it to decode this hidden state and generate a caption.
- **The comparison**: However, because RNNs are computationally expensive to train and assess, memory is typically limited to a few components. By picking the most relevant elements from an input image, attention models can help solve this challenge.

ATTENTION MECHANISM OVER IMAGES (IMAGE CAPTIONING)

- **The method:** The image is first divided into n pieces with an Attention method, and then we compute an image representation for each part. The attention mechanism focuses on the appropriate region of the image when the RNN generates a new word, so the decoder only uses certain sections of the image.


14x14 Feature Map
LSTM
A
bird
flying
over
a
body
of
water
1. Input Image
2. Convolutional Feature Extraction
3. RNN with attention over the image
4. Word by word generation

SUMMARY:

- In this session, we explored the concept of attention networks and their importance in improving the performance of deep learning models.
- We discussed various types of attention mechanisms and their applications in different domains.

SELF-ASSESSMENT QUESTIONS

1. What is the main purpose of attention mechanisms in neural networks?

A. Enhancing model interpretability
B. Allowing the model to focus on specific parts of the input
C. Reducing the computational complexity of the model
D. None of the above

2. Which type of attention allows the model to weigh the importance of different words in a sentence with respect to each other?

A. Soft Attention
B. Hard Attention
C. Self-Attention
D. Multi-Head Attention

TERMINAL QUESTIONS

1. Explain the difference between soft attention and hard attention. Provide an example of a task where each type of attention mechanism would be suitable.
2. How does multi-head attention improve the performance of neural network models? Provide a real-world application where multi-head attention has shown significant benefits.

REFERENCE BOOKS AND WEBLINKS

- Text Books:

1. "Attention is All You Need" by Ashish Vaswani et al.
2. "Neural Machine Translation by Jointly Learning to Align and Translate" by Dzmitry Bahdanau et al.

- Reference Books:

1. "Deep Learning" by Ian Goodfellow, Yoshua Bengio, and Aaron Courville
2. "Natural Language Processing in Action" by Lane, Howard, and Hapke