CNNs, RNNs and Transformers

1) Why Transformers ?

Imagine you're trying to understand a story, like the adventures of your favorite toy. Now, there are different ways you can do this.

One way is to read the story from start to finish, remembering each part as you go along. This is a bit like how Recurrent Neural Networks (RNNs) work. They process information step by step, remembering what came before to understand what's happening next.

Another way is to look at pictures in the storybook. Each picture shows a different scene, and you can understand the story by looking at all these scenes together. This is similar to Convolutional Neural Networks (CNNs), which are great at understanding images by looking at different parts and patterns.

Now, imagine you have a magical toy called a Transformer. This toy is super smart! It doesn't need to read the story step by step like an RNN, and it doesn't need to rely only on pictures like a CNN. Instead, it can look at the whole story at once, understanding how all the parts fit together. It's like having the whole storybook open in front of you, and you can see everything at once.

That's what makes Transformers special in the world of artificial intelligence. They can understand information in a more holistic way, without needing to go step by step or focus only on certain parts. So, when we have Transformers, we might not need to use RNNs or CNNs for certain tasks because the Transformers can handle those tasks in a smarter, more efficient way, like a super-smart doggy figuring out the whole story in one glance!

2) IMAGE CLASSIFICATION : CNNs and Transfomers

Traditionally, image classification has been done using Convolutional Neural Networks (CNNs), and they've been quite effective for this task. However, with the advent of Transformers, there has been a shift in the way some image classification tasks are approached.

Transformers were initially developed for natural language processing tasks, but researchers have adapted them for computer vision tasks like image classification as well. These models, often referred to as Vision Transformers (ViTs), apply the transformer architecture to process image data.

Nowadays, both CNNs and Transformers (specifically Vision Transformers) are used for image classification tasks, and which one is better can depend on various factors such as the size of the dataset, the complexity of the images, and computational resources available.

CNNs still tend to perform very well for image classification, especially when working with smaller datasets or when computational resources are limited. They are also well-established and understood, making them a practical choice in many scenarios.

On the other hand, Transformers have shown promising results in image classification tasks, particularly as the scale of datasets and computational resources increase. They have the advantage of being able to capture global context information effectively, which can be beneficial for understanding relationships between different parts of an image.

So, to answer your question, both CNNs and Transformers are being used for image classification tasks today, with each having its own strengths and suitability depending on the specific requirements of the task at hand.

3) SEQUENTIAL DATA : RNNs and Transformers

Just as with image classification, let's discuss the use of Transformers compared to Recurrent Neural Networks (RNNs) for sequential data tasks like text processing and time series analysis. Traditionally, RNNs have been the go-to model for sequential data tasks. They are designed to process sequences step-by-step, maintaining a hidden state that captures information from previous steps. This makes them well-suited for tasks like natural language processing (NLP) and time series prediction, where the order of the elements matters. However, Transformers have emerged as a powerful alternative for handling sequential data. When adapted for sequential tasks, Transformers utilize self-attention mechanisms to capture relationships between different elements in the sequence, allowing them to understand global context more effectively. In the context of NLP, models like the Transformer and its variants (e.g., BERT, GPT) have achieved remarkable performance on tasks such as language understanding, translation, and generation. These models can process entire sentences or paragraphs at once, capturing dependencies between words more efficiently compared to RNNs. Similarly, in time series analysis, Transformers have shown promise for tasks like forecasting and anomaly detection. They can capture long-range dependencies in the data without being constrained by the sequential nature of RNNs, potentially leading to better performance, especially with large-scale datasets. That being said, RNNs still have their place in sequential data tasks, particularly in scenarios where the temporal aspect is crucial and computational resources are limited. Additionally, hybrid models that combine the strengths of both RNNs and Transformers are also being explored, aiming to leverage the benefits of each architecture. So, to sum up, while RNNs have long been the standard for sequential data tasks, Transformers have emerged as a compelling alternative, offering the potential for improved performance and efficiency, especially in tasks like NLP and time series analysis.

Younes Dahami