Editorial | Uncover This Tech Term

eISSN 2005-8330
https://doi.org/10.3348/kjr.2023.0948
Korean J Radiol 2024;25(1):113-115

Korean Journal of Radiology
KJR

Uncover This Tech Term: Transformers

Amit Gupta, Krithika Rangarajan

Department of Radiology, Dr. B.R.A. IRCH, All India Institute of Medical Sciences, New Delhi, India

Keywords: Artificial intelligence; Transformer; ChatGPT; Deep learning; Large language models

What is a Transformer?

First proposed in 2017, the new transformer models have demonstrated unparalleled success and have superseded the recurrent neural networks (RNNs) as the preferred architecture for various natural language tasks [1]. Traditional deep neural networks process data sequentially. Thus, when dealing with long sequences, information may be diluted or lost from the initial content when decoding the later elements.

However, transformers uniquely process input data. First, the input sequence is broken down into smaller units (for example, words in a sentence) called "tokens." Each token is converted into a unique numerical code "feature vector" via embedding. Positional encoding is performed to represent the location of a token in the input sentence. These embeddings and positional encoding capture all the information related to each token. These data vectors then pass through multiple layers of encoders and decoders. At this level, the key component, "attention," is applied, which allows the model to consider how each token in the input is related to the other tokens, helping the model better understand data points' relationships and focus on the most relevant parts.

In the attention process, three metrics are derived for each token. The "query" represents part of the input that the model wants to understand. The "key" labels recognize the important parts and guide the model's attention to these relevant parts. The "value" represents the actual details or context the model uses to understand the data. Thus, attention operations can summarize values relative to a query or determine which values to focus on. Equipped with an attention mechanism, the model can focus on important values (i.e., those with keys having high similarity with the query). There are several types of attention: self-attention captures relationships between different elements in the same sequence (for example, most relevant words in a sentence); cross-attention allows the model to examine two separate sequences and how they are related (for example, how words in one language relate to words in another language in a translation task); multi-head attention focuses on different aspects of the input simultaneously (for example, subject-verb agreements, dialectal expressions, etc.), and various other types of attention exist for specific tasks. After passing through multiple layers of attention, the final layer of the encoder-decoder sequence presents a rich representation of the model's understanding of the input. This output is then used for various downstream tasks such as language translation, text summarization, question-answering, and other natural language tasks. Figure 1 shows a simple analogy explaining the difference between the working of RNNs and transformers.

Recently introduced Vision Transformers (ViTs), based on the same attention mechanism, have also shown similar results compared to Convolutional Neural Networks (CNNs) for various image-based computer vision applications [2]. In essence, in ViTs, images are divided into patches, and these image patches are treated much the same way as word

Received: September 26, 2023 **Revised:** October 18, 2023
Accepted: October 30, 2023
Corresponding author: Krithika Rangarajan, MD, FRCR, Department of Radiology, Dr. B.R.A. IRCH, All India Institute of Medical Sciences, Room No 160D, Ansari Nagar, New Delhi 110029, India
• E-mail: krithikarangarajan86@gmail.com

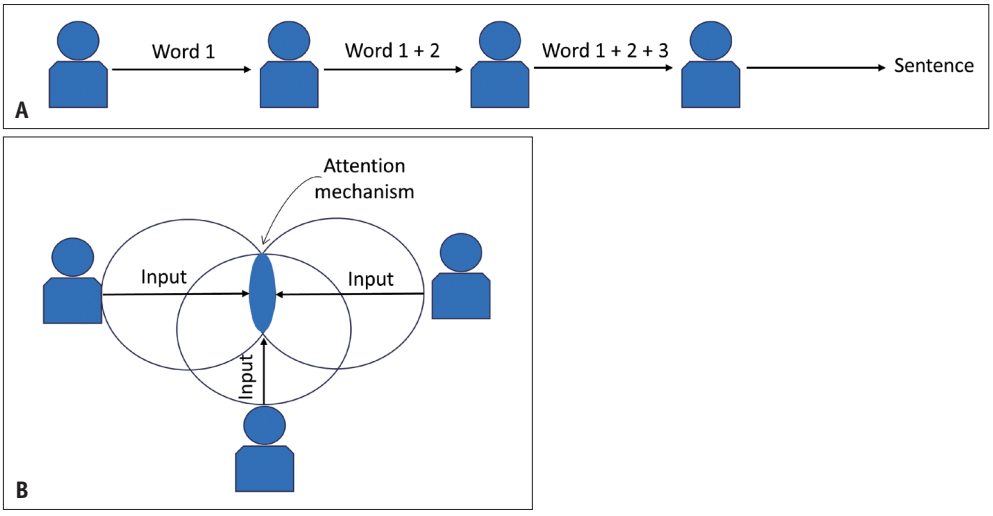

Fig. 1. Analogy explaining the difference between the working of recurrent neural network (RNN) and transformer models. **A:** RNNs can be considered a chain of people passing down a message where each person adds and retains information from previous steps. **B:** Transformer models can be considered a group of experts discussing where each expert considers all parts of the input simultaneously. The attention mechanism then weighs the importance of different elements in the data and combines them.

inputs for language applications.

The greatest advantage of transformers lies in their scalability; unlike previous models whose performance saturates beyond a certain data size threshold, transformers continue showing improvement even with larger datasets. For example, while ViTs may not outperform CNNs in ImageNet-1K database initially, training ViTs on ImageNet-22K followed by fine-tuning them on ImageNet-1K results in superior performance to CNNs. However, one important drawback is that owing to the calculation of attention across the entire latent space, unlike in CNNs, transformer models tend to be larger and slower. While the computation of the attention mechanism is parallelizable and thus efficiently implementable using graphic processing units (GPUs), the transformer itself generally requires more data to reach the same level of performance as other models.

Transformers in Action

Large Language Models (LLMs)

The transformer models were initially applied in natural language processing tasks where they have shown a tremendous potential. Today, we are witnessing a relentless race among the technology giants to develop even more comprehensive and interactive LLMs; some of the most famous ones include ChatGPT (by OpenAI), BERT (by Google), and LLaMA (by Meta). The literature is already replete with studies exploring the various uses of LLMs in radiology. The potential application areas include generating radiology reports, simplifying and translating radiology reports, generating concise report summaries and structured report generation from free-text-based reports, suggesting differential diagnoses, and providing follow-up or guidelines-based recommendations based on imaging findings [3-6]. In addition to generating radiology reports, many researchers have applied ChatGPT for medical writing. Most publishing journals now require a disclaimer if generative artificial intelligence (AI) is used in manuscript preparation. However, journals do not credit LLMs as co-authors because of the lack of accountability and the responsibility for the entire scientific work still resting on human researchers [7,8].

ViTs

Transformer models have also been used for various computer vision problems. There is a growing interest in their use in various aspects of medical imaging, including image segmentation, classification, detection, reconstruction, and generation of new images [9]. Image segmentation tasks comprising multi-organ and organ-specific segmentations have been implemented using transformer models. A breakthrough in this regard was the release of Meta's segment anything model (SAM) and dataset in April 2023, trained on 11 million images, which is expected to have significant implications in automated image segmentation tasks [10]. Image classification transformer models for COVID-19 and tumor diagnosis on radiological imaging have been explored. Other explored applications include object detection (for example, lung nodules on

https://doi.org/10.3348/kjr.2023.0948

Korean Journal of Radiology
KJR

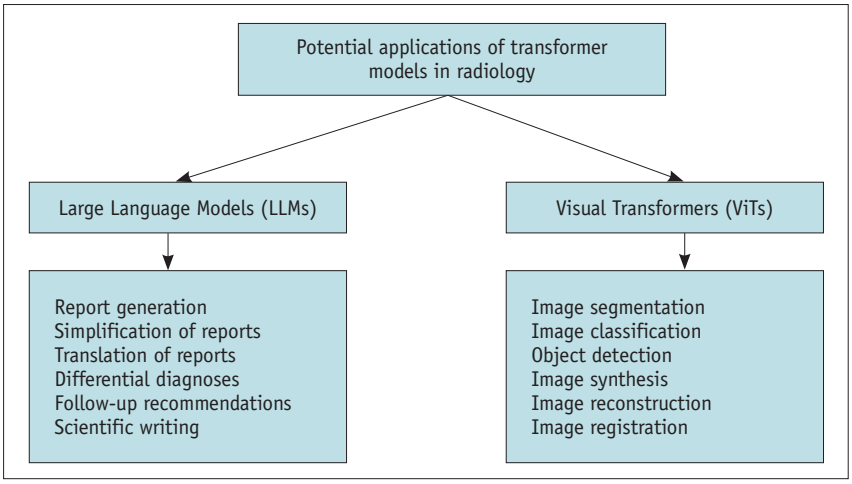

Fig. 2. Various potential applications of transformer models in radiology. These can be divided into being either based on LLMs or ViTs.

chest radiographs), medical image synthesis - intra-modality (like T1-weighted to T2-weighted magnetic resonance imaging [MRI] translation) or inter-modality (like computed tomography [CT] to MRI translation), image reconstruction (for example, obtaining high-quality images from low-dose CT acquisition) and image registration (for example, alignment of scans at different time-points or from different modalities) [9]. Figure 2 summarizes the various potential applications of transformer models in radiology.

Conflicts of Interest

The authors have no potential conflicts of interest to disclose.

Author Contributions

Conceptualization: all authors. Resources: Krithika Rangarajan. Writing—original draft: Amit Gupta. Writing—review & editing: Krithika Rangarajan. Visualization: Amit Gupta.

ORCID IDs

Amit Gupta
https://orcid.org/0000-0003-0192-3512
Krithika Rangarajan
https://orcid.org/0000-0001-5376-6390

Funding Statement

None

REFERENCES

1. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. arXiv:1706.03762 [Preprint]. 2017 [posted Jun 12 2017; cited Sep 14 2023]. Available at: https://doi.org/10.48550/arXiv.1706.03762
2. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An image is worth 16x16 words: transformers for image recognition at scale. arXiv:2010.11929 [Preprint]. 2020 [posted Oct 22 2020; reviced Jun 3 2021; cited Sep 14 2023]. Available at: https://doi.org/10.48550/arXiv.2010.11929
3. Srivastav S, Chandrakar R, Gupta S, Babhulkar V, Agrawal S, Jaiswal A, et al. ChatGPT in radiology: the advantages and limitations of artificial intelligence for medical imaging diagnosis. *Cureus* 2023;15:e41435
4. Adams LC, Truhn D, Busch F, Kader A, Niehues SM, Makowski MR, et al. Leveraging GPT-4 for post hoc transformation of free-text radiology reports into structured reporting: a multilingual feasibility study. *Radiology* 2023;307:e230725
5. Ueda D, Mitsuyama Y, Takita H, Horiuchi D, Walston SL, Tatekawa H, et al. ChatGPT's diagnostic performance from patient history and imaging findings on the diagnosis please quizzes. *Radiology* 2023;308:e231040
6. Jung KH. Uncover this tech term: foundation model. *Korean J Radiol* 2023;24:1038-1041
7. Park SH. Use of generative artificial intelligence, including large language models such as ChatGPT, in scientific publications: policies of KJR and prominent authorities. *Korean J Radiol* 2023;24:715-718
8. Park SH. Authorship policy of the Korean Journal of Radiology regarding artificial intelligence large language models such as ChatGPT. *Korean J Radiol* 2023;24:171-172
9. Shamshad F, Khan S, Zamir SW, Khan MH, Hayat M, Khan FS, et al. Transformers in medical imaging: a survey. *Med Image Anal* 2023;88:102802
10. Kirillov A, Mintun E, Ravi N, Mao H, Rolland C, Gustafson L, et al. Segment anything. arXiv:2304.02643 [Preprint]. 2023 [posted Apr 5 2023; cited Sep 14 2023]. Available at: http://arxiv.org/abs/2304.02643