https://doi.org/10.1038/s41551-022-00997-w

# Contextual learning is nearly all you need

Check for updates

**Graph neural networks and transformers taking advantage of contextual information and large unannotated multimodal datasets are redefining what is possible in computational medicine.**

Graph neural networks are a type of machine learning algorithm that use graph structures to encode spatial relationships between objects. In the context of pathology and radiology, these algorithms are being used to leverage spatial context for various applications, including image segmentation, disease classification, and tissue analysis.

One of the main challenges in medical imaging is accurately identifying and segmenting objects of interest, such as tumors or organs. Traditional machine learning algorithms often struggle with this task due to the complex spatial relationships within medical images. Graph neural networks, on the other hand, are able to explicitly encode these relationships, allowing for more accurate object segmentation.

For example, a graph neural network may be used to segment an image of the brain into different tissue types, such as gray matter and white matter. The algorithm would first create a graph structure, where each node represents a pixel in the image and each edge represents the spatial relationship between two pixels. The algorithm would then use this graph structure to learn the spatial relationships between pixels and identify the different tissue types.

In addition to image segmentation, graph neural networks are also being used for disease classification. For example, a graph neural network may be trained to identify whether an image of the lungs contains a tumor or not. The algorithm would create a graph structure that encodes the spatial relationships between different regions of the lungs, and then use this structure to learn the spatial patterns associated with healthy and diseased tissue.

Another application of graph neural networks in pathology and radiology is tissue analysis. This involves using the spatial relationships within an image to understand the underlying biology of a tissue sample. For example, a graph neural network may be used to analyze the spatial organization of cells within a tissue sample, allowing for more accurate diagnoses and prognoses of diseases.

Overall, graph neural networks are leveraging spatial context in pathology and radiology to improve upon traditional machine learning algorithms. By explicitly encoding spatial relationships within images, these algorithms are able to improve the accuracy of tasks such as image segmentation, disease classification, and tissue analysis. This has the potential to greatly improve the ability of medical professionals to diagnose and treat diseases, and ultimately improve patient outcomes.

The earlier six paragraphs of main text were not written by the editors of *Nature Biomedical Engineering*. They were instantly generated by ChatGPT (chat generative pre-training) — a prototype general-purpose chatbot based on the GTP3.5 transformer model[1], a large language model with an encoder–decoder architecture and about 200 billion parameters — by inputting the prompt "Discuss, in about 600 words, how graph neural networks are leveraging spatial context for uses in pathology and radiology."

Transformer models are powerful. ChatGPT has been trained on internet corpora available until 2021, by using self-supervised learning tasks, such as masked-language modelling, and by learning the meaning of words via contextual clues through self-attention mechanisms[2] (methods that enhance parts of the input data so as to improve their overall representation; for example, self-attention makes the algorithm consider the context of a word in an input sequence for clues that help better convert the word into a numeric representation that preserves the relationships between words in the sentence). Before applying self-attention, transformers map tokenized text into embeddings (continuous low-dimensional vector representations of discrete variables) and use positional encodings to account for word order. Decoding the output involves cross-attention (that is, the calculation of attention scores for one type of data via information from another type of data, such as other embeddings) and multi-head attention (or multiple attention mechanisms running in parallel).

Hence, transformers learn the meaning of words and the structure of language via contextual clues. By recurrently predicting the next word on the basis of an input query, they can produce new text, summarize stories, re-order scrambled text, identify topics in a discussion, write code, and a myriad other tasks[3], all in perfect prose. Because transformers can learn representations of arbitrary input data, they have also been used to build models that generate images from textual prompts. The quality of the output

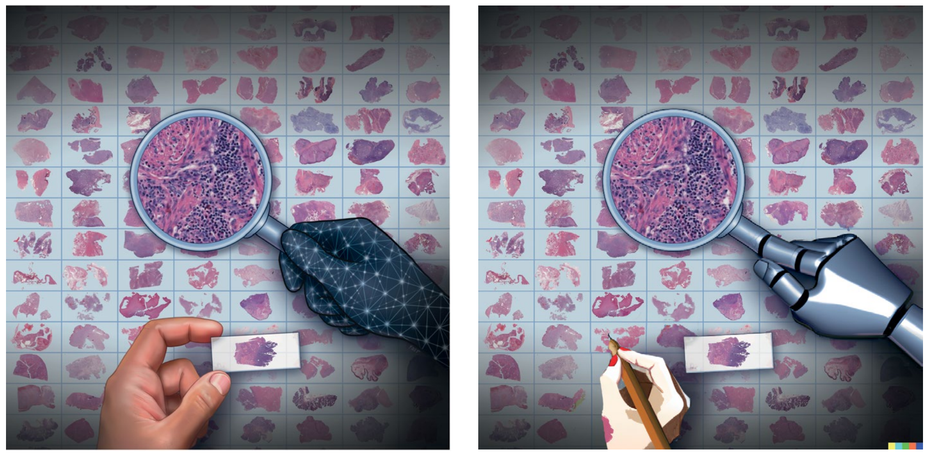

**Fig. 1 | Searching for a matching whole-slide image.** The image on the right corresponds to one output of the OpenAI's transformer model DALL-E 2 when asking it to successively modify the hands on the image on the left according to the prompts "A real female hand." and "A hand from an AI robot."

CREDIT: KATIE YOST AND OPENAI.

of any such models largely depends on the quality of the data they have been trained with. Spurious associations in the data can make the models hallucinate information; they can be perceived as confidently providing plausible yet incorrect information, and can easily doctor existing datasets (Fig. 1).

Because one main advantage of transformers is that they do not need explicit annotations in the data they are trained with, they can also be used to build models to classify images without explicit supervision, as Pranav Rajpurkar and colleagues show (E. Tiu et al. *Nat. Biomed. Eng*. https://doi.org/10.1038/s41551-022-00936-9; 2022) for the detection of the presence of pathologies in unlabelled chest X-ray images. In this case, the model uses a text transformer and a vision transformer to learn pathology features from the raw radiology report associated with each X-ray image by using contrastive learning (a method for the prediction of whether pairs of samples are related or unrelated). Hence, the radiology reports act as a natural source of contextual clues for the model to learn pathology features in the X-ray images.

As the output of ChatGPT relayed above explains, spatial context can improve the performance of machine-learning tasks in pathology. Graph representations — that is, sets of interconnected nodes and edges — are particularly suitable to any system that can be mapped as a network, such as networks of cells or tissue patches in medical images, and networks of medical codes and of concepts in electronic health records, as discussed by Marinka Zitnik and colleagues in a Perspective article (M. M. Li et al. *Nat. Biomed. Eng*. https://doi.org/10.1038/s41551-022-00942-x; 2022). This issue of *Nature Biomedical Engineering* also includes three examples of the advantages of applying graph neural networks to tasks in pathology.

In one article, James Zou, Aaron Mayer and Alexandro Trevino show (Z. Wu et al. *Nat. Biomed. Eng*. https://doi.org/10.1038/s41551-022-00951-w; 2022) that spatial protein profiles, obtained via multiplexed immunofluorescence, in tissue specimens from patients with head-and-neck and colorectal cancers can be used to train a graph neural network to model the tumour microenvironment so that spatial motifs can be associated with cancer recurrence and with patient survival after treatment. In another article, Sunghoon Kwon and colleagues show (Y. Lee et al. *Nat. Biomed. Eng*. https://doi.org/10.1038/s41551-022-00923-0; 2022) that, in whole-slide images of tumours, graph deep learning can be used to derive interpretable contextual histopathological features in the tumour microenvironment that are predictive of the prognosis of the patients. As explained by Faisal Mahmood and co-authors in an associated News & Views (G. Jaume et al. *Nat. Biomed. Eng*. https://doi.org/10.1038/s41551-022-00924-z; 2022), context-aware graph neural networks can be thought of as sitting in between models that do not have access to any context and models (such as vision transformers) that access the full spatial context, and include a degree of inductive bias or assumptions that allow the model to make generalizations. In another research article in this issue, Mahmood and colleagues show another application of self-supervised learning: searching and retrieving gigapixel whole-slide images (Fig. 1) at speeds that are independent of the size of the repository (C. Chen et al. *Nat. Biomed. Eng*. https://doi.org/10.1038/s41551-022-00929-8; 2022). To search for a tissue patch, rather than querying against every slide in the dataset, a variational autoencoder (a probabilistic generative model that learns latent representations of the data) is trained to represent select patches from each slide as a set of codes in a manner that the patches with the highest chances of matching the query can be retrieved by leveraging uncertainty-based ranking and a tree data structure for speed efficiency and scalability.

For applications in medicine and health, openly releasing pre-trained models can facilitate a shift from model building to model deployment in healthcare settings, as argued by Joseph Wu and colleagues in a Review article (A. Zhang et al. *Nat. Biomed. Eng*. https://doi.org/10.1038/s41551-022-00898-y; 2022) in this issue. This is because transformers, graph representation learning and many other self-supervised deep-learning models require large unlabelled datasets for pre-training, which improves the performance of downstream tasks. Indeed, as Rayan Krishnan, Pranav Rajpurkar and Eric Topol conclude in a Review article (R. Krishnan et al. *Nat. Biomed. Eng*. https://doi.org/10.1038/s41551-022-00914-1; 2022), "self-supervised learning leveraging multimodal data will enable the creation of high-performing models that better 'understand' the underlying physiology." The trained chatbot would seem to agree with that, and adds that "quality context is crucial for ChatGPT to perform well".

Published online: 23 December 2022

## References


1. Floridi, L. & Chiriatti, M. *Minds and Machines* **30**, 681–694 (2020).
2. Vaswani, A. et al. In *Proc. 31st International Conference on Neural Information Processing Systems (NIPS)* 6000–6010 (Curran Associates, 2017).
3. *ChatGPT: Optimizing Language Models for Dialogue* (OpenAI, 2022); https://openai.com/blog/chatgpt/