HPCI magazine
富岳百景
FUGAKU HYAKKEI

Developing Generative AI Methods for a Secure, Safe, and Convenient Society with Fugaku

Rio Yokota
Professor, Global Scientific Information and Computing Center, Tokyo Institute of Technology

As of 2023, the proliferation of Generative AI is accelerating in Japan. Major IT companies such as Microsoft and Google in the United States have been aggressively investing substantial funds a few years prior to this, intensifying the competition in development. In this context, Dr. Yokota's research group is advancing the research and development of generative AI methods using the supercomputer Fugaku.

What is Generative AI?

Generative AI refers to Artificial Intelligence (AI) that generates new content such as text, images, audio, and videos based on input data. In contrast to traditional AI, which learns from given data to perform tasks like classification and detection, generative AI stands out for its ability to apply learned knowledge to generate images or text. Particularly noteworthy are “language-generative AI,” which produces fluent text, and “image-generative AI,” which creates images based on input text. OpenAI, an AI startup based in the United States, gained significant attention in November 2022 with the release of ChatGPT, a language-generative AI.

Generative AI generates text and images using a “model” that represents relationships between data. While there are various methods for constructing models, they often involve training on a large amount of labeled data, known as training data (e.g. a set of labels like “dog” and corresponding images of dogs). Deep learning requires training of layered connections of an artificial neural network, which models neural circuits in the brain, to input and learn from data. Such neural networks are commonly referred to as a “model.” For example, those used for image classification are called vision models, while those for language tasks are referred to as language models.

“In the past, people trained different models for tasks such as image classification, anomaly detection, summarizing text, and automatic translation. However, recently it has become common to train a single large model on various modes of data and then fine-tune it for different tasks. This approach, involving extensive training on a large dataset inadvance, is called ‘pre-training’,” Dr. Yokota explains.

Improving the Performance of Image Classification Models through Pre-training with Synthetic Images

The first challenge undertaken by Dr. Yokota and his team with Fugaku was the pre-training of image classification models. Dr. Yokota points out that the development of image classification models involves numerous challenges. “The performance of the model tends to increase empirically with the amount of data used in pre-training. For image classification models, it is necessary to input a large quantity of image data as training data. However, the task of collecting and labeling image data posted on the internet is inherently advantageous for companies operating search engines, such as Google. While Google has created several image datasets, they are not accessible to researchers from other companies or academia, posing a significant barrier to the advancement of image processing,” Dr. Yokota explains.

To overcome this situation, Dr. Yokota and his team took on the challenge of large-scale pre-training of image classification models using synthetic images. “Research on synthetic images was initially pioneered by the team at the National Institute of Advanced Industrial Science and Technology (AIST)*1. However, the scale of their work was relatively small. Therefore, in 2021, we started creating a large-scale dataset of synthetic images using only formulas, and in the 2022 Fugaku General Access (Super Large Scale Pre-training of Vision Transformers Using Synthetic Images, hp220028), we conducted pre-training of image classification models on Fugaku,” Dr. Yokota says.

The distinctive feature of Dr. Yokota’s synthetic images is that they are created solely using formulas without the use of real images, depicting

images such as fractals (Fig. 1). Using this dataset of 21 million synthetic images, they performed pre-training on an image classification model called “Vision Transformer.” The effectiveness of this pre-training was nearly equivalent to the results obtained by Google when using their own dataset of 300 million real images. Furthermore, when compared to the commonly used real image dataset “ImageNet-21k” for pre-training (approximately 14 million images), their synthetic image dataset showed higher learning effectiveness. Dr. Yokota reflects on the success, stating, “The effectiveness of synthetic images in pre-training has been validated, giving us significant confidence.”

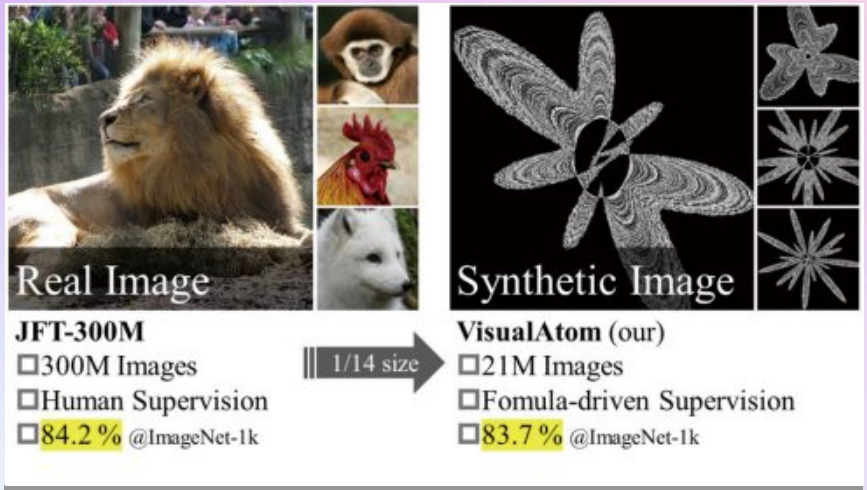

Fig 1 : Real vs. synthetic images
On the left are examples of real images, while on the right are examples of synthetic images. The learning effectiveness of the Vision Transformer using the JFT-300M dataset (Google’s private dataset of approximately 300 million real images) is shown, achieving an image recognition accuracy of 84.2%. The learning effectiveness using Visual Atom (the set of 21 million synthetic images created by Dr. Yokota and his team) is 83.7%, which is nearly equivalent. The “@ImageNet-1k” indicates the result of fine-tuning the pre-trained model on a small dataset of real images called ImageNet-1k.

Dr. Yokota highlights four advantages of synthetic images. The first is the ability to generate an unlimited quantity, the second is the absence of copyright concerns, the third is the lack of need to consider personally identifiable information, and the fourth is the avoidance of societal biases related to gender, race, etc.

“Currently, the internet is filled with biased images. For instance, if you search for ‘nurse,’ you might predominantly find images of women, creating gender biases based on professions. The use of such real images in pre-training poses significant problems. However, synthetic images completely sidestep these issues,” Dr. Yokota explains.

However, a question arises: Can synthetic images, which may not immediately convey their meaning, achieve the same learning effectiveness as real images? Yokota’s response to this question is intriguing: “This can be understood by comparing it to a fetus inside the mother’s womb. The fetal optic nerves react to light even within the mother’s womb. The fetus likely perceives vague shapes and forms even before birth. The rapid recognition of shapes after birth is attributed to the pre-training on light that occurred in the womb. In other words, synthetic images can be likened to what the fetus experiences as light inside the womb.”

Establishing the Research and Development Environment for AI with Fugaku

Why did Dr. Yokota choose Fugaku, which has all CPU nodes and no GPUs, for this large-scale pre-training? He explains: “Certainly, for deep learning, GPUs offer faster computational speed and are more suitable than CPUs. In fact, the ABCI supercomputer specialized for AI at AIST uses GPUs. However, Japanese AI researchers are flocking to ABCI, resulting in overcrowding and shortage of resources. Therefore, we

wanted to enable AI research and development even on supercomputers with only CPUs. Additionally, for the large-scale pre-training we conducted this time, having a large number of nodes available for an extended period was more important than computational speed. In this regard, Fugaku had a significant advantage in the allocated computational resources compared to ABCI (Fig. 2)."

On the other hand, Dr. Yokota acknowledges the challenges he faced with Fugaku as it was not originally designed for AI research and development. He explains, "First of all, the software infrastructure for performing deep learning quickly was not in place, and we had to develop it from scratch. When loading a large amount of image data, individually reading each image could impose a burden on input/output operations. Therefore, we implemented strategies such as compressing a large amount of image data into a single file to reduce input/output frequency. Additionally, to reduce memory consumption, we developed a new method for distributed parallel processing of image data and models. While we faced various challenges, this will enable many researchers to accelerate AI research and development using Fugaku in the future."

Fig 2 : Fugaku, left, and AIST's AI-devoted ABCI
The processing speed of one GPU on ABCI, specialized for AI, is 50 times that of one CPU on Fugaku. However, while ABCI has 4,000 GPUs, Fugaku boasts a massive 160,000 CPUs. Therefore, in a simple calculation, using all the CPUs on Fugaku would allow for computational capabilities nearly equivalent to ABCI. Furthermore, if Fugaku can be utilized for longer durations than ABCI, it holds a significant advantage

The Essential Need for the Development of Generative AI Methods

Building upon these achievements, Dr. Yokota has embarked on a new project, the Government-Initiated Category of Supercomputer Fugaku (Development of Distributed Training Method for Large Language Models on Fugaku, hp230254). In this project, a collaborative research team comprising the Tokyo Institute of Technology, Tohoku University, Fujitsu, RIKEN, CyberAgent, Kotoba Technologies, and Nagoya University aims to develop methods for large-scale distributed parallel processing on Fugaku to efficiently train large language models (models used in language generative AI).

“In the United States, massive investments are being made in generative AI. For instance, ChatGPT surpassed 100 million users in just two months, becoming an internationally adopted societal infrastructure. Given this situation, some argue that it may be too late to develop generative AI methods as we are already playing catch-up. However, the development of generative AI is primarily undertaken by startup companies, and there are concerns about the continuity of their businesses. Relying on such companies entails high risks. Additionally, generative AI is a foundational technology for Society 5.0*2, and it will be present in various fields such as healthcare, industry, and education. This is not a discussion about entering a specific industry after determining its ‘advantage.’ Just as Japan developed power and transportation networks during the high economic growth period not because Japanese power lines and road technologies had a global ‘advantage,’ but because investments in infrastructure with overwhelming productivity were imperative. Failure to invest in such infrastructure can have a devastating impact on the international competitiveness of the industries built on it,” Dr. Yokota emphasizes.

“As generative AI is expected to be widely utilized in various industrial sectors in Japan in the future, investments in computational resources, the accumulation of technical expertise and knowledge in AI, and the cultivation of advanced AI talent are becoming increasingly crucial,” he adds.

Aiming for Generative AI Capable of Handling Both Images and Language

Furthermore, Dr. Yokota and his team are actively involved in the development of generative AI itself. In the 2023 Fugaku General Access (Performance Optimization of Transformers and Their Application to Vision & Language, hp230119), they are working on optimizing the performance of Transformer on the A64FX processor developed by Fujitsu for Fugaku to make it capable of handling both images and language. Transformer is the neural network architecture that served as the basis for the Vision Transformer described above.

“Image generative AI creates images based on text input. Images and language do not exist independently; both need to be learned. Therefore, we aim to develop a high-performance generative AI that can handle both images and language. The goal is to develop a generative AI that can generate correct images, not just amusing misinterpretations like a ramen bowl in the shape of a house, as seen in the ‘Family Ramen’*3 image (Fig. 3),” Dr. Yokota says with a smile.

Amusing images created by generative AI, not limited to family ramen, often appear on social media, becoming popular topics. Dr. Yokota points out, “The misunderstanding arises not because generative AI does not understand Japanese correctly, but because it lacks knowledge of Japanese culture and trends. In the future, we aim to develop generative AI that can accurately respond to Japanese society, not just the Japanese language.”

In the future, with the development of generative AI capable of

Fig 3 : Image of "Family Ramen" created by generative AI
(Note: This image is not the actual image posted on social media.)

handling both images and language, it will become possible, for example, for a cooking robot equipped with a camera to convert image data obtained through the camera into words. The robot can then ask the generative AI what to do next, and the generative AI responds. This enables the robot to autonomously cook without pre-instruction, bringing us closer to a society where such interactions are common.

Addressing the Growing Issue of Fake News, Regulation Is Crucial

In pursuing these challenges, Dr. Yokota and his team are advancing the development of generative AI methodologies while collaborating with other research groups in the country. However, as robots equipped with generative AI become more autonomous, capable of making judgments and taking actions without human intervention, some people may harbor concerns and fears. In response, Dr. Yokota says, "One thing I fear is the spread of fake images created by generative AI. Deepfakes in particular pose a significant problem. Deepfake is a portmanteau of 'deep learning' and 'fake,' originally a technology used in the film and gaming industries that has become widely accessible and is now being maliciously exploited for fake news and information manipulation. Consequently, leaders from various countries are convening conferences to discuss regulations."

Indeed, during the G7 Hiroshima Summit held in May 2023, the creation of the "Hiroshima AI Process" was included. This framework aims to promote international rule-making regarding the use and development of generative AI, as well as its regulation, with a focus on technologies like ChatGPT. Furthermore, in November 2023, the world''s first international conference on the safe use of AI, the "AI Safety Summit," took place in the U.K. High-ranking officials from governments, major AI companies, and leaders discussed the risks associated with the potential misuse and uncontrollable nature of advanced AI.

"In addition to these high-level meetings for regulations, I believe it is essential to hold gatherings that involve AI researchers and technologists. Through ongoing research and development of generative AI, I hope to contribute to the realization of a future society that is secure, safe, and convenient," Dr. Yokota adds.

---

*1 Reference for synthetic image creation and image distinction mechanism through pre-training on synthetic images (Japanese only): https://www.aist.go.jp/aist_j/magazine/20221130.html
*2 Society 5.0 was proposed in the 5th Science and Technology Basic Plan drawn up by Japanese government as a future society that Japan should aspire to. It envisions a human-centered society where the cyber (virtual) and physical (real-world) spaces are highly integrated, aiming to achieve both economic development and solutions to societal challenges.
*3 Family Ramen (家系ラーメン) refers to a style of ramen and a group of ramen shops often identified by names ending with "家" ("ya" or "house"). It gained the name "family" because many of these shops have "家" in their name.

About the Researcher

Dr. Yokota was engaged in experimental research on fluid as an undergraduate student and undertook fluid simulation in his master's course. He started using GPUs for simulations during his doctoral program, when GPUs were just starting to become available. "At the time, we connected about 200 GPUs for gaming to create something like the world's cheapest supercomputer, and in 2009 we won the Association for Computing Machinery's (ACM) Gordon Bell Prize (Low Price / Performance Category)," Dr. Yokota says.

After this experience, he became seriously interested in supercomputers. He began researching machine learning using AI after arriving at Tokyo Institute of Technology in 2015, in response to his students' interest. Looking back, he says, "Although my research interests have changed, the focus on distributed parallel processing has remained consistent since my graduate school days."

In his time off, Dr. Yokota pursues his hobbies, including diving and programming, and he is also a qualified diving instructor. "Recently I've been so busy that I haven't had time for any hobbies, so it's been a little tough," he laughs.

Associated Research Projects:

- Super Large Scale Pre-training of Vision Transformers Using Synthetic Images (hp220028)
- Performance Optimization of Transformers and Their Application to Vision & Language (hp230119)
  Principal Investigator: Rio Yokota, Global Scientific Information and Computing Center, Tokyo Institute of Technology
- Development of Distributed Training Method for Large Language Models on Fugaku (hp230254)
  Project Representative: Seiichi Shimasaki, Director of Research Promotion Bureau, Ministry of Education, Culture, Sports, Science and Technology (MEXT)

HPCI magazine
富岳百景
FUGAKU HYAKKEI

Registered Institution for Facilities Use Promotion /
The Representative for HPCI Operation
Research Organization
for Information Science and Technology
1-5-2, Minatojimaminami-machi, Chuo-ku, Kobe, Hyogo, 650-0047 JAPAN
Tel: +81-78-599-9511

HPCI Portal site
https://www.hpci-office.jp/en
HPCI PR site
https://fugaku100kei.jp/e/

Issued: January 2024