NIRAV PATEL

Generative Artificial Intelligence and Remote Sensing

A perspective on the past and the future

The first phase of 2023 has been marked with an explosion of interest around generative AI systems, which generate content. This type of machine learning promises to enable the creation of synthetic data and outputs in many different modalities. OpenAI's ChatGPT has certainly taken the world by storm and opened discourse on how the technology should be used.

Historically, generative models are certainly not new, dating back to the 1950s, with hidden Markov models and Gaussian mixture models [1], [2], [3]. The recent development of deep learning has allowed for generative models' utility. In the early days of deep generative models, N-gram language modeling was utilized to generate sentences in natural language processing (NLP) [4]. This modeling did not scale well to generating long sentences, and hence, recurrent neural networks (RNNs) were introduced to deal with longer dependencies [5]. RNNs were followed by the development of long short-term memory [6] and gated recurrent unit methods, which leveraged gating mechanisms to control memory usage during training [7].

OPENAI'S CHATGPT HAS CERTAINLY TAKEN THE WORLD BY STORM AND OPENED DISCOURSE ON HOW THE TECHNOLOGY SHOULD BE USED.

In the computer vision arena (more aligned with remote sensing), traditional image generation algorithms utilized techniques such as texture mapping [8] and texture synthesis [9]. These methods were very limited and could not generate complex and diverse images. The introduction of generative adversarial networks (GANs) [10] and variational autoencoders [11] in the past decade or so has allowed for more control over the image generation process to generate high-resolution images.

Generative models in different modalities felt the advancement of the field in its totality with the introduction of the transformer architecture [12]. Large language models, such as the generative pretrained transformer (GPT), adopt this architecture as the primary building block, which initially had significant utility in the NLP world before later modifications to this architecture allowed for application to image-based streams of information [13], [14], [15], [16], [17]. Transformers consist of an encoder and a decoder, where the encoder takes in an input sequence and generates hidden representations, while the decoder has a multihead attention and feedforward NN [1]. See Figure 1 for an NLP example of a sentence being translated from English to Japanese.

The emergence of these techniques has allowed for the creation of foundation models, which are the technical scaffolding behind generative AI capabilities. Foundation models, such as ChatGPT, learn from unlabeled datasets, which saves a significant amount of time and the expense of manual annotation and human attention. However, there is a reason why the most well-resourced companies in the world have made an attempt at generating these models [19].

First, you need the best computer scientists and engineers to maintain and tweak foundation models, and second, when these models are training data from the whole Internet, the computational cost is not insignificant. OpenAI's GPT-3 was trained on roughly 45 TB of text data (equivalent to 1 million feet of bookshelf space), which cost several million dollars (estimated) [19].

With remote sensing applications, anecdotally, I have witnessed the rise of the use of GANs over the past few years. This deep learning technique, as mentioned

Digital Object Identifier 10.1109/MGRS.2023.3275984
Date of current version: 30 June 2023

before, is an NN architecture that conducts the training process as a competition between a generator and a discriminator to produce new data conforming to learned patterns. Since GANs are able to learn from remote sensing data without supervision, some applications that the community has found useful include (but are not limited to) data generation/augmentation, superresolution, panchromatic sharpening, haze removal and restoration, and cloud removal [20], [21], [22], [23]. I strongly believe that the ever-increasing availability of remotely sensed data and the availability of relatively robust computational power in local and distributed (i.e., cloud)-based environments will make GANs only more useful to the remote sensing community in the coming years and may even lead to some bespoke foundation models, especially with the open source remote sensing efforts that Google [25], Microsoft [26], and Amazon [27] are funding.

In other remote sensing areas, such as image segmentation, the foundation models are already here (within days of writing this piece). In what could be the example for other remote sensing foundation models, Meta AI released Segment Anything [28], which is a new task, model, and dataset for image segmentation. Meta claims to have "built the largest segmentation dataset to date, with over 1 billion masks on 11M licensed and privacy respected images." Social media has many remote sensing companies, scientists, and enthusiasts alike ingesting satellite imagery into the model and yielding results, with varying utility. Meta's paper provides more technical detail on how the foundation model is architected (Figure 2), but in my opinion, the true uniqueness and value lie in how massive the dataset is and how well labeled it is in comparison to other image segmentation datasets of its kind.

The authors of the Segment Anything admit that their model can "miss fine structures, hallucinates small disconnected components at times, and does not produce boundaries as crisply as more computationally intensive methods." They posit that more dedicated interactive segmentation methods would outperform their model when many more points are provided.

My prediction for the future is that as we see the computer vision world make more investments in foundation models related to image processing, the remote sensing and geosciences world will stand to benefit from large investments by the world's well-resourced tech companies.

Advancements in computer vision models, due to the development of foundation models, however, will not always be tailored toward the needs of remote sensing. Closely examining the data being fed into these foundation models and how exactly data are being labeled within these models will allow for discerning remote sensing practitioners get the most value out of using such computer vision models.

Hence, a major caution to users of foundation models for remote sensing applications is the same caution that applies for applications of foundation models to other types of machine learning applications: the limits of utility for outputs are tied closely to the quantity and quality of the labeled data associated with the model. Even the most sophisticated foundation models cannot escape the maxim of "garbage in, garbage out."

Well-resourced technology companies also have their monetary interests that ultimately influence the foundation models that they create. It is important for remote sensing practitioners to understand this dynamic. For example,

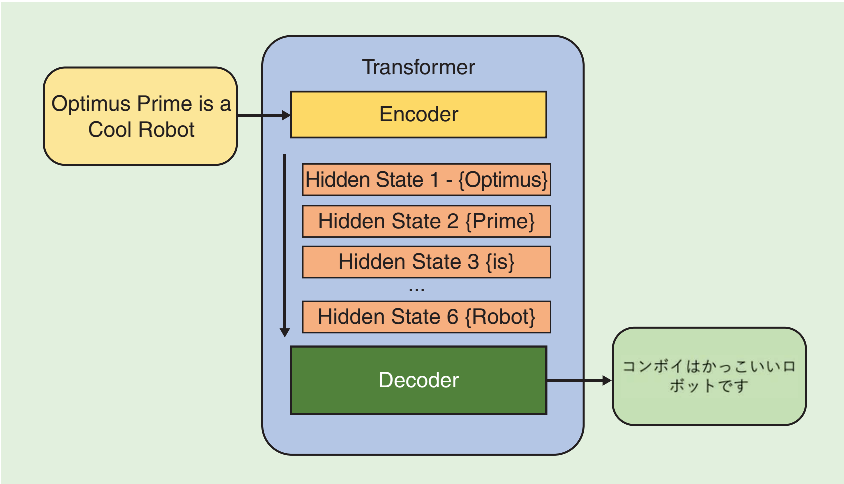

FIGURE 1. An NLP translation of a sentence from English to Japanese [18].

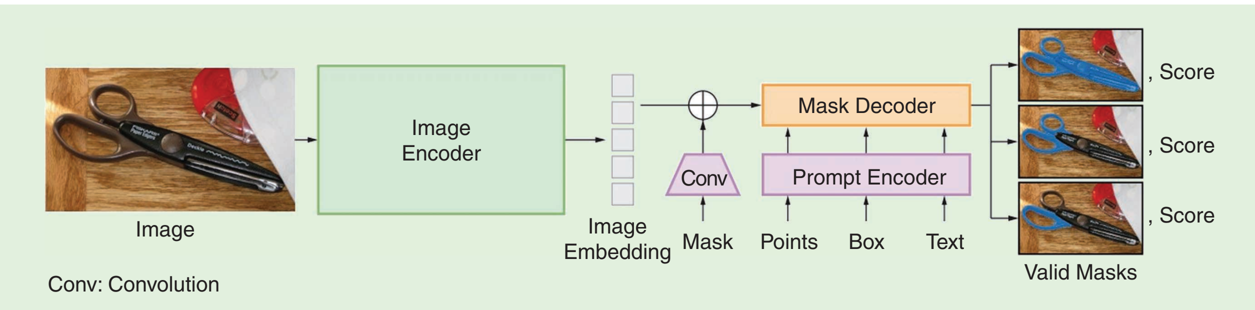

FIGURE 2. Technical detail of the Segment Anything architecture [28].

in exchange for providing access to lightweight and easy-to-access interfaces, such as ChatGPT, all the data that are put in by the user can ultimately be utilized by OpenAI for other purposes. While the service does not cost any money for the user, ChatGPT still will gain insight from your inquiry to make itself better. Indeed, nothing truly comes for free, especially with the use of foundation models and the user interfaces associated with them.

Finally, it is worth discussing the nefarious use cases that this technology can be used for, especially in the context of remote sensing. Synthetic data generation could be utilized, for example, to create fake satellite images that could provide the impression to an undiscerning user of information and evidence of something that doesn't exist, and it could hide potential evidence. Consider an example of a country trying to hide changes around an area (land surface changes) to mask human rights violations. Synthetic data could be provided in the same place as a real satellite image was supposed to be provided in a data feed that is accessed by the public, giving a false sense of what the reality of the situation is.

It is, thus, extremely important that the uses of synthetic data are also well defined and regulated by the community of remote sensing practitioners. Creating methods to identify synthetic remote sensing data would be the most effective in the near term, in my opinion. I also believe that synthetic data will be extremely useful in combination with real remote sensing data to train remote sensing models that aim at "few-shot" circumstances (i.e., detecting rare objects).

Ultimately, the adoption of an extremely novel and effective technology in its nascent stages within a community requires a focus on the ethical implications of the use of the technology in each circumstance. The same holds true for our field of remote sensing, and I have confidence in our community to set the appropriate guardrails on the limits of use of this technology.

ACKNOWLEDGEMENT

None of this article was written by generative artificial intelligence (AI).

AUTHOR INFORMATION

Nirav Patel (niravpatel@ufl.edu) is an affiliate faculty member in the Department of Geography at the University of Florida, Gainesville, FL 32611 USA. He is also a senior scientist in the Office of the Secretary of Defense at the U.S. Defense Innovation Unit, and a program manager for open-source/non-International Traffic in Arms Regulations (ITAR) restricted remote sensing efforts.

REFERENCES

[1] Y. Cao et al., "A comprehensive survey of AI-generated content (AIGC): A history of generative AI from GAN to chatGPT," 2023, *arXiv:2303.04226*.

[2] K. Knill and S. Young, "Hidden Markov models in speech and language processing," in *Corpus-Based Methods in Language Speech Processing*, S. Young and G. Bloothooft, Eds. Dordrecht, The Netherlands: Springer, 1997, pp. 27–68.

[3] D. A. Reynolds, "Gaussian mixture models," in *Encyclopedia Biometrics*, vol. 741, S. Z. Li and A. Jain, Eds. Boston, MA, USA: Springer, 2009, pp. 659–663.

[4] Y. Bengio, R. Ducharme, and P. Vincent, "A neural probabilistic language model," in *Proc. Adv. Neural Inf. Process. Syst.*, 2000, vol. 13, pp. 932–938.

[5] T. Mikolov, S. Kombrink, L. Burget, J. Černocký, and S. Khudanpur, "Extensions of recurrent neural network language model," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, May 2011, pp. 5528–5531, doi: 10.1109/ICASSP.2011.5947611.

[6] A. Graves, "Long short-term memory," in *Supervised Sequence Labelling With Recurrent Neural Networks*, Berlin, Germany: Springer-Verlag, 2012, pp. 37–45.

[7] R. Dey and F. M. Salem, "Gate-variants of gated recurrent unit (GRU) neural networks," in *Proc. 60th IEEE Int. Midwest Symp. Circuits Syst. (MWSCAS)*, Aug. 2017, pp. 1597–1600, doi: 10.1109/MWSCAS.2017.8053243.

[8] P. S. Heckbert, "Survey of texture mapping," *IEEE Comput. Graph. Appl.*, vol. 6, no. 11, pp. 56–67, Nov. 1986, doi: 10.1109/MCG.1986.276672.

[9] A. A. Efros and T. K. Leung, "Texture synthesis by non-parametric sampling," in *Proc. 7th IEEE Int. Conf. Comput. Vision*, Sep. 1999, vol. 2, pp. 1033–1038, doi: 10.1109/ICCV.1999.790383.

[10] I. Goodfellow et al., "Generative adversarial networks," *Commun. ACM*, vol. 63, no. 11, pp. 139–144, Oct. 2020, doi: 10.1145/3422622.

[11] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," 2013, *arXiv:1312.6114*.

[12] A. Vaswani et al., "Attention is all you need," in *Proc. 30th Adv. Neural Inf. Process. Syst.*, 2017, pp. 6000–6010.

[13] T. Brown et al., "Language models are few-shot learners," in *Proc. Adv. Neural Inf. Process. Syst.*, 2020, vol. 33, pp. 1877–1901.

[14] A. Ramesh et al., "Zero-shot text-to-image generation," in *Proc. Int. Conf. Mach. Learn.*, PMLR, Jul. 2021, pp. 8821–8831.

[15] M. Lewis et al., "BART: Denoising sequence-to-sequence pre-training for natural language generation, translation, and comprehension," 2019, *arXiv:1910.13461*.

[16] A. Dosovitskiy et al., "An image is worth 16x16 words: Transformers for image recognition at scale," 2020, *arXiv:2010.11929*.

[17] Z. Liu et al., "Swin transformer: Hierarchical vision transformer using shifted windows," in *Proc. IEEE/CVF Int. Conf. Comput. Vision*, 2021, pp. 10,012–10,022.

[18] D. J. Rogel-Salazar. "Transformers models in machine learning: Self-attention to the rescue." Domino. Accessed: Apr. 13, 2023. [Online]. Available: https://www.dominodatalab.com/blog/transformers-self-attention-to-the-rescue

[19] "What is generative AI?" McKinsey & Company. Accessed: Apr. 13, 2023. [Online]. Available: https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-generative-ai

[20] Y. Weng et al., "Temporal co-attention guided conditional generative adversarial network for optical image synthesis," *Remote Sens.*, vol. 15, no. 7, Mar. 2023, Art. no. 1863, doi: 10.3390/rs15071863.

(continued on p. 100)

- Prof. Avik Bhattacharya, professor, the Centre of Studies in Resources Engineering, Indian Institute of Technology Bombay, Mumbai, India
- Prof. Alejandro Frery, professor of statistics and data science, the Victoria University of Wellington, New Zealand
- Dr. Dipankar Mandal, postdoctoral fellow, Department of Agronomy, Kansas State University, USA
- Dr. Shashi Kumar, scientist, the Indian Institute of Remote Sensing, Indian Space Research Organisation, Dehradun, India
- Prof. Sylvain Lobry, associate professor, the Université Paris Cité, the Laboratoire d'Informatique de Paris Descartes, France
- Dr. Michele Ronco, postdoctoral researcher, the Image Processing Laboratory, the University of Valencia.

IEEE GRSS IADF SCHOOL: FIND OUT THE NEXT EDITION!

After the successful first edition of the IEEE GRSS IADF school, a second one will be announced soon. It will follow the same theme as the 2022 edition, i.e., CV4EO. It will be an in-person event and take place at the University of Sannio, Benevento, Italy, 13–15 September 2023. We look forward to seeing you in Benevento! Please stay tuned!

CONCLUSION

We would like to thank the GRSS and the IADF for their support, and all the lecturers who gave so freely of their time and expertise. A survey among the participants conducted after the school clearly showed that the event received high attention and provided an exciting experience. All the comments have been collected and will be used to improve the format of the next editions.

AUTHOR INFORMATION

Gemine Vivone (gemine.vivone@imaa.cnr.it) is with the National Research Council - Institute of Methodologies for Environmental Analysis, 85050 Tito Scalo, Italy, and National Biodiversity Future Center, 90133 Palermo, Italy. He is a Senior Member of IEEE.

Dalton Lunga (lungadd@ornl.gov) is with the Oak Ridge National Laboratory, Oak Ridge, TN 37830 USA. He is a Senior Member of IEEE.

Francescopaolo Sica (francescopaolo.sica@unibw.de) is with the Institute of Space Technology & Space Applications, University of the Bundeswehr Munich, 85577 Neubiberg, Germany. He is a Member of IEEE.

Gülşen Taşkın (gulsen.taskin@itu.edu.tr) is with the Institute of Disaster Management, Istanbul Technical University, Istanbul 34469, Turkey. She is a Senior Member of IEEE.

Ujjwal Verma (ujjwal.verma@manipal.edu) is with the Department of Electronics and Communication Engineering, Manipal Institute of Technology Bengaluru, Manipal Academy of Higher Education, Manipal 576104, India. He is a Senior Member of IEEE.

Ronny Hänsch (ronny.haensch@dlr.de) is with the DLR, 82234 Weßling, Germany. He is a Senior Member of IEEE.

GRS

PERSPECTIVES *(continued from p. 88)*

[21] W. Boulila, M. K. Khlifi, A. Ammar, A. Koubaa, B. Benjdira, and I. R. Farah, "A hybrid privacy-preserving deep learning approach for object classification in very high-resolution satellite images," *Remote Sens.*, vol. 14, no. 18, Sep. 2022, Art. no. 4631, doi: 10.3390/rs14184631.

[22] S. Zhang, X. Zhang, T. Li, H. Meng, X. Cao, and L. Wang, "Adversarial representation learning for hyperspectral image classification with small-sized labeled set," *Remote Sens.*, vol. 14, no. 11, May 2022, Art. no. 2612, doi: 10.3390/rs14112612.

[23] S. Yang, M. Sun, X. Lou, H. Yang, and H. Zhou, "An unpaired thermal infrared image translation method using GMA-CycleGAN," *Remote Sens.*, vol. 15, no. 3, Jan. 2023, Art. no. 663, doi: 10.3390/rs15030663.

[24] M. Casey. "Foundation models 101: A guide with essential FAQs." Snorkel AI. Accessed: Apr. 13, 2023. [Online]. Available: https://snorkel.ai/foundation-models/

[25] N. Gorelick, M. Hancher, M. Dixon, S. Ilyushchenko, D. Thau, and R. Moore, "Google earth engine: Planetary-scale geospatial analysis for everyone," *Remote Sens. Environ.*, vol. 202, pp. 18–27, Dec. 2017, doi: 10.1016/j.rse.2017.06.031.

[26] T. Augspurger, "Scalable sustainability with the planetary computer," presented at the AGU Fall Meeting Abstracts, New Orleans, LA, USA, Dec. 2021.

[27] "Earth on AWS." Amazon. Accessed: Apr. 13, 2023. [Online]. Available: https://aws.amazon.com/earth/

[28] A. Kirillov et al., "Segment anything," 2023, *arXiv:2304.02643*.

GRS