https://doi.org/10.2352/EI.2025.37.12.HPCI-176

Synthetic Dataset Pre-training for Precision Medical Segmentation Using Vision Transformers

Edgar Josafat Martinez-Noriega[1], Peng Chen[2], Truong Thao Nguyen[1], and Rio Yokota[3]; National Institute of Advanced Industrial Science and Technology[1], RIKEN Center for Computational Science (RIKEN-CCS)[2], and Tokyo Institute of Technology[3]; Japan

Abstract

In medical segmentation, the acquisition of high-quality labeled data remains a significant challenge due to the substantial cost and time required for expert annotations. Variability in imaging conditions, patient diversity, and the use of different imaging devices further complicate model training. The high dimensionality of medical images also imposes considerable computational demands, while small lesions or abnormalities can create class imbalance, hindering segmentation accuracy. Pre-training on synthetic datasets in medical imaging may enable Vision Transformers (ViTs) to develop robust feature representations, even during the fine-tuning phase, when high-quality labeled data is limited. In this work, we propose integrating Formula-Driven Supervised Learning (FDSL) synthetic datasets with medical imaging to enhance pre-training for segmentation tasks. We implemented a custom fractal dataset, Style Fractals, capable of generating high-resolution images, including those measuring 8k x 8k pixels. Our results indicate improved performance when using the SAM model for segmentation, in conjunction with robust augmentation techniques, increasing performance from 62.30% to 63.68%. This was followed by fine-tuning on the PAIP dataset, a high-resolution, real-world pathology dataset focused on liver cancer. Additionally, we present results using another synthetic dataset, SegRCDB, for comparative analysis.

Introduction

Pre-training Vision Transformers (ViTs) [1], [2] typically involves a two-stage process: initially, the model is exposed to large-scale labeled or synthetic data to learn foundational visual representations, followed by fine-tuning on more specialized tasks. Vision Transformers, being data-intensive architectures, require substantial datasets for effective training, with state-of-the-art performance in transfer learning often necessitating pre-training on over 100 million images [3]. Formula-Driven Supervised Learning [4] has emerged as an alternative to traditional supervised training on real images. Formula-driven techniques encompass methods for generating synthetic images from mathematical formulas, including fractals [5], geometric patterns [4], polygons [6], and other fundamental shapes. Attributes such as complexity, smoothness, brightness, texture, fill-rate, and others can be customized to generate labeled datasets of arbitrary size without the need for human intervention. Moreover, formula-driven synthetic image generation helps mitigate ethical concerns such as societal bias or copyright infringement [7, 8, 9].

In medical segmentation using deep learning, key challenges persist, particularly in obtaining high-quality labeled data. This process remains difficult due to the significant cost and time required for expert annotations. Variability in imaging conditions, patient diversity, and differing imaging devices further complicate model training. The high dimensionality of medical images imposes substantial computational demands, and small lesions or abnormalities often introduce class imbalance, hindering segmentation accuracy.

Using synthetic datasets for pre-training in medical imaging enables ViTs to develop robust feature representations, even with limited high-quality labeled data. This approach facilitates the effective generalization of Vision Transformers to high-resolution segmentation tasks, enabling the capture of subtle details that are critical in medical contexts. By refining learned representations, pre-trained ViTs significantly enhance segmentation accuracy, reduce the need for extensive annotations, and improve overall performance, making them especially valuable for high-precision applications in healthcare. We propose integrating these two domains by utilizing a custom Fractal dataset to pre-train Vision Transformer models on large datasets. As previously mentioned, Vision Transformers yield favorable results when trained on extensive datasets such as LION-5B [10], YFCC-100M [11], or JFT-300M [12], each containing over hundred million images. These models are not only data-hungry but also require robust augmentation techniques such as AutoAugment [13], CutMix [14] and MixUp [15] during each training epoch, as the attention mechanism they employ benefits from variations in the original training images.

In this study, we implemented a custom fractal dataset capable of generating high-resolution images, such as those measuring 8K x 8K pixels or larger. Our results indicate that performance improves when utilizing the SAM model [16] for segmentation combined with strong augmentation techniques, achieving a score of 63.68% compared to 62.30% from the baseline. After fine-tuning on the PAIP dataset, a high-resolution, real-world pathology dataset focused on liver cancer, we achieved a performance of up to 70.58%, compared to 72.54% from the baseline results. Although we did not surpass the baseline, we believe these results are promising, given that we performed pre-training using only half the size of the real images provided in the baseline results.

Related Work

Medical image processing plays a crucial role in medical analysis, enhancing diagnostic capabilities through a variety of tasks, including cell counting, classification, detection, and segmentation. Among these, medical image segmentation is the most frequently applied task in clinical diagnosis. Inspired by the successful utilization of Vision Transformers in various medical domains, recent studies have proposed ViT-based models for skin lesion segmentation [17, 18, 19]. Another notable application involves retinal vessel segmentation and related tasks [20, 21, 22].

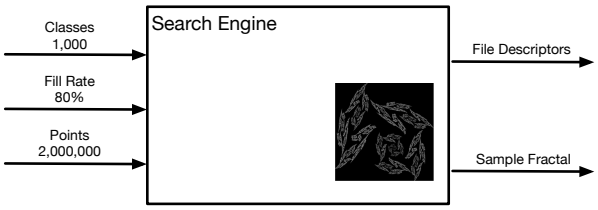


Figure 1. *Search engine for Style Fractal dataset.*

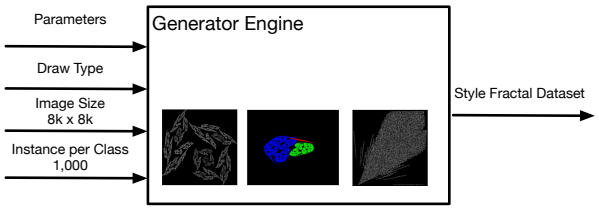


Figure 2. *Generator engine for Style Fractal dataset.*

Furthermore, a novel multi-class prediction approach for skin lesion classification has been introduced, combining ViT and ViT-GAN techniques [23].

On another hand, Formula-Driven Supervised Learning is an innovative deep learning approach that leverages synthetic images generated from mathematical formulas, along with their corresponding labels. Initially proposed by Kataoka *et al.* [24], this method presents a novel pathway for the creation of large, diverse, and dynamic datasets tailored for pre-training vision models. The images generated through this approach encompass a variety of shapes and patterns, such as polygons, geometric configurations, and fractals [24, 4, 5, 6, 25]. A comparable approach to ours is the one proposed by Shinoda *et al.* in their work on SegRCDB [26], where they implemented basic polygonal and radial shapes for segmentation tasks. However, their implementation is primarily aimed at general segmentation tasks and does not focus on medical or high-resolution images. In contrast, FDSL addresses significant issues, such as societal biases and the handling of sensitive information, including personal data and copyrights [8, 7, 9]. By utilizing synthetic datasets derived from mathematical constructs, FDSL bypasses these challenges, ensuring a secure and controlled environment for model training while maintaining higher ethical standards.

Method

This section provides a comprehensive overview of medical image segmentation and its significance as a digital tool in medical diagnostics. Additionally, we introduced our proposed dataset, Style Fractals, which is a variant of Formula-Driven Supervised Learning (FDSL). Moreover, we presented the SAM model, which served as the backbone for our segmentation experiments. Furthermore, we highlighted the critical role of image augmentation in training vision transformers and discussed state-of-the-art augmentation techniques.

Medical Segmentation

Diagnostic imaging has become an indispensable tool in modern medicine. With the rapid growth in the size and volume of medical images, the integration of computational methods for their processing and analysis has become essential. Medical image segmentation involves the automatic identification and labeling of regions of interest within medical images, including modalities such as CT, MRI, and ultrasound. This process partitions an image into semantically meaningful segments, such as organs, tissues, tumors, or other anatomical structures. Image segmentation is fundamental to a wide range of biomedical imaging applications, including tissue volume quantification, anatomical structure analysis, partial volume correction of functional imaging data, and computer-integrated surgical procedures to mentioned a few examples [27, 28, 29, 30].

Style Fractals Dataset

In this work, we propose an enhanced version of the original FractalDB to better adapt this synthetic dataset for large resolution image segmentation in medical applications. To provide context, we first describe the original FractalDB. The dataset comprises 1 million images organized into 1,000 categories, with each category containing 1,000 images [24]. Formally, the dataset can be defined in a metric space $\mathscr{P}$ as follows:

$$\text{IFS} = \{\mathscr{P}; w_1, w_2, \cdots, w_N; p_1, p_2, \cdots, p_N\}. \tag{1}$$

where $w_i : \mathscr{P} \to \mathscr{P}$ are transformations, p_i are probabilities, and N is the transformations. A fractal $S = \{x_t\}_{t=0}^{\infty} \in \mathscr{P}$ can be generated $\mathscr{P} = \mathbb{R}^2$. Each transformation is a type of affine transformation or augmentation. These transformations are characterized by six parameters $\theta_i = (a_i, b_i, c_i, d_i, e_i, f_i)$ including shifting and rotation:

$$w_i(x; \theta_i) = \begin{bmatrix} a_i & b_i \\ c_i & d_i \end{bmatrix} x + \begin{bmatrix} e_i \\ f_i \end{bmatrix}. \tag{2}$$

The parameters $(a_i, b_i, c_i, d_i, e_i, f_i, p_i)$ are randomly initialized and retained if the fill ratio, defined as the proportion of fractal dot pixels to the total image pixels, exceeds a predefined threshold. Parameter adjustments are achieved by multiplying one of the six Iterated Function System (IFS) parameters by specific weights, resulting in modified images that maintain the overall category shape while altering finer details. At this stage, we highlight the key differences introduced in our Style Fractal dataset. The original FractalDB implementation is written in Python, utilizing a serial execution model and a NumPy-based rendering engine, which leads to suboptimal performance. To enhance computational efficiency, we ported the code to a C++ engine and implemented the IFS iteration using OpenMP. As shown in Equation 2, the computation of points for each fractal is inherently sequential. However, the generation of each fractal instance can be parallelized by distributing the workload across multiple threads, thereby significantly accelerating the overall rendering process.

Style Fractal dataset creation process is divided into two major steps: the search of CSV file descriptors and the generation instances for each category. Both steps utilize the IFS function to generate fractal images. First, we generate fractal images with a predefined number of points and assess the fill ratio. This is shown in Figure 1. In the second stage, we use the parameters obtained from the first stage and apply the aforementioned affine transformations shown in Figure 2.

We introduce three key modifications to the original FractalDB, which form the basis of our proposed Style Fractals dataset. First, we incorporate the capability to render high-resolution images. Although this may initially appear straightforward, the search space and perspective projection processes

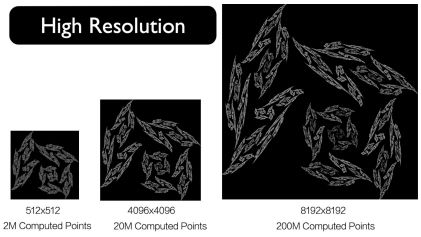


Figure 3. *Style Fractals dataset using high resolution for pre-traning.*

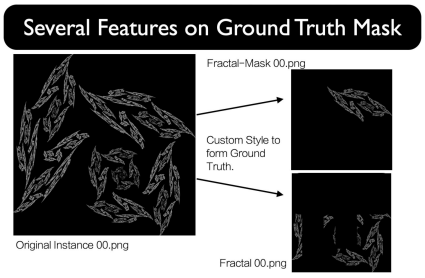


Figure 4. *Augmentation like style to render Style Fractal dataset using several effects on the ground truth mask.*

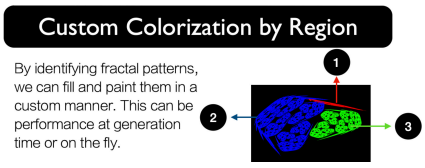


Figure 5. *Augmentation like style to render Style Fractal dataset using colors on different fractal pattern.*

require significant adjustments. For instance, rendering at a resolution of 512×512 pixels necessitates approximately 2 million fractal points. However, increasing the canvas size demands a broader search space to maintain image fidelity, as illustrated in Figure 3. Second, we introduce the functionality to occlude or partially render specific regions of the fractal. This enables the generation of custom ground truth masks, facilitating the training of more sophisticated segmentation models. An overview of these augmentation strategies is provided in Figure 4. Lastly, we enhance the dataset by assigning a unique color space to each fractal pattern. The color is chosen randomly while preserving the consistency of the RGB channels. These enhancements collectively define the Style Fractals dataset, enriching its variability and applicability in image segmentation tasks.

Segment Anything Model - SAM

The Segment Anything Model (SAM) [16] is a vision model utilizing a ViT-H/16 architecture, which incorporates 14x14 windowed attention and four equally spaced global attention blocks, tailored for segmentation tasks. Introduced by Meta AI, SAM was trained on an extensive dataset comprising over a billion segmentation masks, which endows it with robust capabilities. SAM is particularly notable for its zero-shot generalization, as it can segment objects in images without requiring task-specific fine-tuning, making it highly versatile. Thanks to its massive training dataset and transformer-based architecture, SAM produces highly accurate segmentation masks with minimal user input. However, SAM is not without limitations. It may struggle with highly complex scenes or occluded objects, sometimes missing fine-grained details or intricate boundaries. Additionally, the segmentation accuracy is influenced by the quality and placement of the prompts provided. In this study, we leveraged SAM for pre-training using the Style Fractal dataset, adapting the sequence length as suggested by Enzhi *et al.* [31]. Moreover, we specifically utilized model b for this work.

Augmentation Techniques used for better Pre-Training

It is well established that effective pre-training of Vision Transformer models requires the use of strong augmentations [2]. These augmentations typically involve affine transformations such as rotations, occlusions, and the composition of multiple images. By incorporating these transformations, the model is exposed to a diverse set of data in each epoch, enhancing its ability to generalize. State-of-the-art augmentation techniques, including AutoAugment [13], MixUp [15], and CutMix [14], have shown substantial improvements in model performance. In this work, we propose to employ AutoAugment for pre-training our Style Fractal dataset with the SAM model. AutoAugment is a data augmentation approach designed to improve deep learning model performance by automatically discovering the optimal set of augmentation policies for a specific task. Introduced by Cubuk *et al.* [13], AutoAugment searches for the combination of augmentation policies that maximizes model performance on a given dataset. This technique is particularly beneficial for computer vision tasks, where data augmentation plays a crucial role in improving generalization.

Evaluation

In this section, we present the results of our experiments evaluating the pre-training of the SAM model on the SegRCDB and Style Fractal datasets. Following the pre-training phase, we fine-tuned the model on the PAIP dataset. We provide an overview of the experimental setup and the environment in which the experiments were conducted. Additionally, we describe the Dice score metric used in our evaluation. Our analysis focuses on the application of stronger augmentation techniques, comparing the performance of SAM on the PAIP dataset. Specifically, we utilized AutoAugment policies for this comparison. Furthermore, we report the performance of the pre-training process when applied to large image resolutions.

Experimental Environment

We utilized the AI Bridging Cloud Infrastructure (ABCI) supercomputer [32], which is recognized for its optimization in AI computing tasks. The supercomputer features V100 GPUs. The nodes equipped with Volta architecture are composed of 1,088 compute nodes, each integrating two Intel Xeon Gold 6148 CPUs (totaling 40 cores), 384 GiB of DRAM, four NVIDIA V100 GPUs, and InfiniBand EDR NICs. Additionally, each node is provisioned with 1.6TB of local storage and has shared access to a 35PB Lustre parallel file system.

PAIP Dataset

The PAIP (Precision Medicine in Image Processing) 2019 Liver Cancer dataset is a comprehensive collection of medical images designed to support research in liver cancer segmentation and diagnosis. Its primary objective is to improve the precision of medical imaging in the context of liver cancer by providing high-quality annotated data for the training and evaluation of seg-

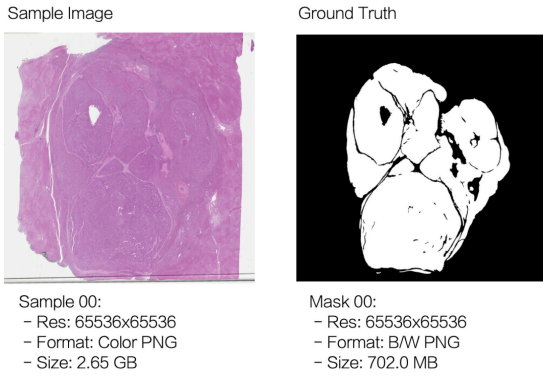


Figure 6. *A sample instance and ground truth from PAIP dataset.*

mentation algorithms. The dataset consists of annotated CT scans from liver cancer patients, with liver tumors clearly marked by medical experts, thereby serving as ground truth for segmentation tasks. The PAIP dataset includes a total of 2,457 Whole-Slide Images (WSIs). For testing purposes, the images are downsampled to square formats, with 8k resolution used for larger test samples. During training, the dataset is divided into three sets: 70% of the images are used for training, 10% for validation, and 20% for testing. To ensure the robustness of model performance, the dataset is shuffled at the beginning of each epoch. A sample of the PAIP dataset is shown in Figure 6.

Metric used for Evaluation

For the quantitative evaluation we choose the Dice score. This metric measures the similarity between a predicted mask and the ground truth mask. We can define this metric as:

$$\text{Dice(X,Y)} = \frac{2 \times |X \cap Y|}{|X| + |Y|}$$

where X and Y are the two sets being compared. $|X \cap Y|$ represents the intersection of sets X and Y. $|X|$ and $|Y|$ represent the cardinality of sets X and Y respectively. A score of 100% indicates identical similarity between the prediction and the ground truth.

Experiments on Augmentation

In this experiment, we evaluated the performance on the PAIP dataset by applying several AutoAugment policies. The results are presented in Figure 7. We further elaborate on the different levels of augmentation, categorized as primary, medium, and strong.

- Raw: No Augmentation.
- Primary: RandomResize + CenterCrop
- AutoAug low: Policy level 3
- AutoAug medium: Policy level 9
- AutoAug high: Policy level 15

For this experiment, we set the image resolution to 512 and the number of epochs to 100. The SAM model size used was set to ”b”. The baseline settings were adapted from Enzhi *et al.* [31]. As shown, the baseline achieved a Dice score of 62.3% when no

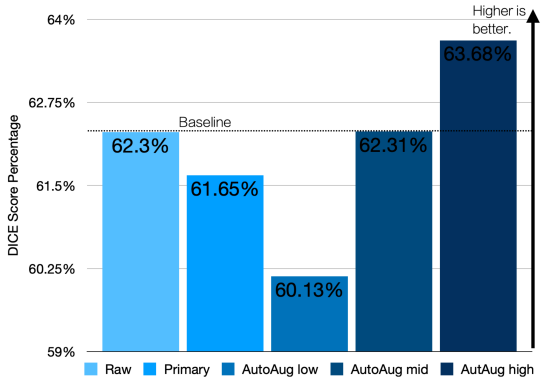


Figure 7. *DICE score on different augmentation techniques on PAIP dataset.*

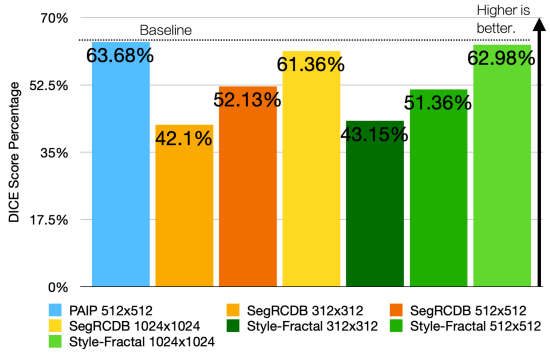


Figure 8. *DICE score on PAIP dataset when pre-trained with synthetic datasets. Low resolution.*

augmentation was applied. With the inclusion of primary augmentations, the Dice score decreased by 0.65%. Furthermore, applying smaller AutoAugment policies led to a further reduction of 2.17%, which may be attributed to insufficient transformations that potentially confused the network, hindering its training. In contrast, applying medium and high AutoAugment policies improved performance, with the strongest policy yielding a Dice score of 63.68%.

Experiments on Large Image Resolutions

For the subsequent set of experiments, we extended the settings from the previous tests using strong AutoAugment policies. The objective of this experiment was to investigate the impact of pre-training on various image sizes across the two synthetic datasets, SegRCDB and our proposed Style Fractal. We only included baseline results for an image size of 512 from the PAIP dataset. However, the results for SegRCDB and Style Fractal are based on fine-tuning to the PAIP dataset using the same image size. The number of pre-training epochs was set to 50, and fine-tuning was also performed for 50 epochs to align with the baseline results of 100 epochs. Figure 8 presents the results for lower resolution image sizes. The image sizes were incrementally increased from 312 to 512 and then to 1024. For the SegRCDB dataset, we employed the original settings proposed by the authors. As observed, the baseline, which incorporated AutoAugment policies, achieved a performance of 63.68%. It is evident that neither SegRCDB nor Style Fractal surpassed the baseline performance. However, a clear trend emerges in which accuracy increases substantially as the image size is scaled up. Notably, Style Fractal

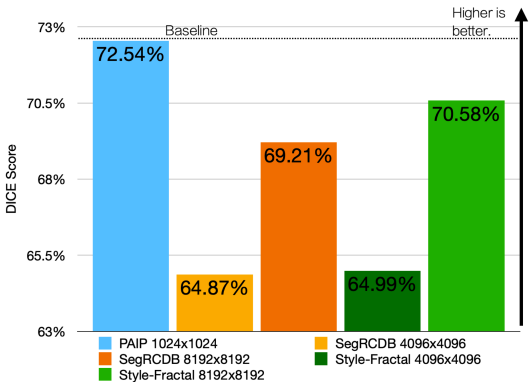


Figure 9. *DICE score on PAIP dataset when pre-trained with synthetic datasets. High resolution.*

consistently outperformed SegRCDB in all three cases. Although the performance did not exceed the baseline, we consider the 62.98% achieved by Style Fractal to be a promising result, especially given that pre-training was conducted on synthetic datasets for only 50 epochs.

Finally, we present the results of pre-training using large image resolutions, as shown in Figure 9. In this experiment, fine-tuning on the PAIP dataset with a resolution of 1024 pixels resulted in an accuracy of 72.54%. We further increased the resolution of the synthetic datasets to 4096 and 8192 pixels. As observed in previous experiments, the baseline performance could not be surpassed. Additionally, the performance of the models diverged further from the baseline results when compared to the low-resolution cases. It is also evident that Style Fractal continues to outperform SegRCDB. The best result achieved was 70.58%, which is only a 1.96% gap from the baseline. This suggests that further improvements could be achieved through better augmentation strategies, extended training durations, and the use of larger models.

Conclusion

In this work, we proposed the use of synthetic datasets for pre-training the SAM model in medical segmentation tasks. We introduced a new synthetic dataset, Style Fractals, capable of generating custom ground truth fractals that can be leveraged to assist in fine-tuning segmentation tasks for medical applications. We evaluated our performance against baseline results and compared it to the state-of-the-art synthetic dataset, SegRCDB. Our experiments demonstrated comparable performance using synthetic datasets, particularly when image sizes were increased and strong augmentation techniques were applied, achieving an accuracy of 70.58%, close to the 72.54% accuracy obtained by the baseline. Future work includes a more detailed exploration of ground truth masks and local feature transformations, larger pre-training on synthetic datasets, and a comprehensive study on the high-resolution fine-tuning process.

Acknowledgements

This paper is based on results obtained from a project, JPNP20006, subsidized by the New Energy and Industrial Technology Development Organization (NEDO).

References

[1] A. Dosovitskiy, L. Beyer, A. Kolesnikov, D. Weissenborn, X. Zhai, T. Unterthiner, M. Dehghani, M. Minderer, G. Heigold, S. Gelly, J. Uszkoreit, and N. Houlsby, "An image is worth 16x16 words: Transformers for image recognition at scale," in *International Conference on Learning Representations*, 2021.

[2] H. Touvron, M. Cord, M. Douze, F. Massa, A. Sablayrolles, and H. Jegou, "Training Data-efficient Image Transformers & Distillation through Attention," in *International Conference on Machine Learning*, vol. 139, pp. 10347–10357, July 2021.

[3] X. Zhai, A. Kolesnikov, N. Houlsby, and L. Beyer, "Scaling vision transformers," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 12104–12113, 2022.

[4] H. Kataoka, R. Hayamizu, R. Yamada, K. Nakashima, S. Takashima, X. Zhang, E. J. Martinez-Noriega, N. Inoue, and R. Yokota, "Replacing labeled real-image datasets with auto-generated contours," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 21232–21241, 2022.

[5] K. Nakashima, H. Kataoka, A. Matsumoto, K. Iwata, N. Inoue, and Y. Satoh, "Can vision transformers learn without natural images?," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 36, pp. 1990–1998, 2022.

[6] H. Kataoka, A. Matsumoto, R. Yamada, Y. Satoh, E. Yamagata, and N. Inoue, "Formula-driven supervised learning with recursive tiling patterns," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 4098–4105, 2021.

[7] K. Yang, K. Qinami, L. Fei-Fei, J. Deng, and O. Russakovsky, "Towards fairer datasets: Filtering and balancing the distribution of the people subtree in the imagenet hierarchy," in *Proceedings of the 2020 conference on fairness, accountability, and transparency*, pp. 547–558, 2020.

[8] Y. M. Asano, C. Rupprecht, A. Zisserman, and A. Vedaldi, "Pass: An imagenet replacement for self-supervised pretraining without humans," *arXiv preprint arXiv:2109.13228*, 2021.

[9] N. Carlini, J. Hayes, M. Nasr, M. Jagielski, V. Sehwag, F. Tramer, B. Balle, D. Ippolito, and E. Wallace, "Extracting training data from diffusion models," in *32nd USENIX Security Symposium (USENIX Security 23)*, pp. 5253–5270, 2023.

[10] C. Schuhmann, R. Beaumont, R. Vencu, C. Gordon, R. Wightman, M. Cherti, T. Coombes, A. Katta, C. Mullis, M. Wortsman, *et al.*, "Laion-5b: An open large-scale dataset for training next generation image-text models," *Advances in Neural Information Processing Systems*, vol. 35, pp. 25278–25294, 2022.

[11] B. Thomee, D. A. Shamma, G. Friedland, B. Elizalde, K. Ni, D. Poland, D. Borth, and L.-J. Li, "Yfcc100m: the new data in multimedia research," *Communications of the ACM*, vol. 59, p. 64–73, Jan. 2016.

[12] C. Sun, A. Shrivastava, S. Singh, and A. Gupta, "Revisiting unreasonable effectiveness of data in deep learning era," in *Proceedings of the IEEE international conference on computer vision*, pp. 843–852, 2017.

[13] E. D. Cubuk, B. Zoph, D. Mane, V. Vasudevan, and Q. V. Le, "Autoaugment: Learning augmentation strategies from data," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 113–123, 2019.

[14] S. Yun, D. Han, S. J. Oh, S. Chun, J. Choe, and Y. Yoo, "Cutmix: Regularization strategy to train strong classifiers with localizable features," in *Proceedings of the IEEE/CVF international conference*

on computer vision, pp. 6023–6032, 2019.

[15] H. Zhang, M. Cisse, Y. N. Dauphin, and D. Lopez-Paz, "mixup: Beyond empirical risk minimization," *arXiv preprint arXiv:1710.09412*, 2017.

[16] A. Kirillov, E. Mintun, N. Ravi, H. Mao, C. Rolland, L. Gustafson, T. Xiao, S. Whitehead, A. C. Berg, W.-Y. Lo, P. Dollár, and R. Girshick, "Segment anything," *arXiv:2304.02643*, 2023.

[17] H. Wu, S. Chen, G. Chen, W. Wang, B. Lei, and Z. Wen, "Fat-net: Feature adaptive transformers for automated skin lesion segmentation," *Medical image analysis*, vol. 76, p. 102327, 2022.

[18] X. He, E.-L. Tan, H. Bi, X. Zhang, S. Zhao, and B. Lei, "Fully transformer network for skin lesion analysis," *Medical Image Analysis*, vol. 77, p. 102357, 2022.

[19] G. S. Krishna, K. Supriya, M. Sorgile, *et al.*, "Lesionaid: Vision transformers-based skin lesion generation and classification," *arXiv preprint arXiv:2302.01104*, 2023.

[20] D. Chen, W. Yang, L. Wang, S. Tan, J. Lin, and W. Bu, "Pcat-unet: Unet-like network fused convolution and transformer for retinal vessel segmentation," *PloS one*, vol. 17, no. 1, p. e0262689, 2022.

[21] H. Yu, J.-h. Shim, J. Kwak, J. W. Song, and S.-J. Kang, "Vision transformer-based retina vessel segmentation with deep adaptive gamma correction," in *ICASSP 2022-2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1456–1460, IEEE, 2022.

[22] H. Zhang, X. Zhong, Z. Li, Y. Chen, Z. Zhu, J. Lv, C. Li, Y. Zhou, and G. Li, "Tim-net: transformer in m-net for retinal vessel segmentation," *Journal of Healthcare Engineering*, vol. 2022, no. 1, p. 9016401, 2022.

[23] K. Lee, H. Chang, L. Jiang, H. Zhang, Z. Tu, and C. Liu, "Vitgan: Training gans with vision transformers. arxiv 2021," *arXiv preprint arXiv:2107.04589*.

[24] H. Kataoka, K. Okayasu, A. Matsumoto, E. Yamagata, R. Yamada, N. Inoue, A. Nakamura, and Y. Satoh, "Pre-training without natural images," in *Proceedings of the Asian Conference on Computer Vision*, 2020.

[25] M. F. Barnsley and Mathematics, *Fractals Everywhere*. USA: Dover Publications, Inc., 2012.

[26] R. Shinoda, R. Hayamizu, K. Nakashima, N. Inoue, R. Yokota, and H. Kataoka, "Segrcdb: Semantic segmentation via formula-driven supervised learning," in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 20054–20063, 2023.

[27] S. M. Lawrie and S. S. Abukmeil, "Brain abnormality in schizophrenia: a systematic and quantitative review of volumetric magnetic resonance imaging studies," *The British Journal of Psychiatry*, vol. 172, no. 2, pp. 110–120, 1998.

[28] A. J. Worth, N. Makris, V. S. Caviness Jr, and D. N. Kennedy, "Neuroanatomical segmentation in mri: technological objectives," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 11, no. 08, pp. 1161–1187, 1997.

[29] H. W. Müller-Gärtner, J. M. Links, J. L. Prince, R. N. Bryan, E. McVeigh, J. P. Leal, C. Davatzikos, and J. J. Frost, "Measurement of radiotracer concentration in brain gray matter using positron emission tomography: Mri-based correction for partial volume effects," *Journal of Cerebral Blood Flow & Metabolism*, vol. 12, no. 4, pp. 571–583, 1992.

[30] W. E. L. Grimson, G. Ettinger, T. Kapur, M. E. Leventon, W. M. Wells III, and R. Kikinis, "Utilizing segmented mri data in image-guided surgery," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 11, no. 08, pp. 1367–1397, 1997.

[31] E. Zhang, I. Lyngaas, P. Chen, X. Wang, J. Igarashi, Y. Huo, M. Munetomo, and M. Wahib, "Adaptive patching for high-resolution image segmentation with transformers," in *SC24: International Conference for High Performance Computing, Networking, Storage and Analysis*, pp. 1–16, IEEE, 2024.

[32] AIST, "National Institute of Advanced Industrial Science and Technology abci 2.0.." `https://abci.ai/en`. Accessed: 2024-11-01.

Author Biography

Edgar Josafat Martinez-Noriega obtained his Doctorate in Computer Science from the University of Electro-Communications, Tokyo in 2020. Following this, he has been employed as a Post-Doctoral Researcher at the National Institute of Advanced Industrial Science and Technology (AIST), working on the application of synthetic datasets for large-scale deep learning. His research focuses on parallel computing, computer graphics, and deep learning.

Peng Chen is a senior scientist at the RIKEN Center for Computational Science (RIKEN-CCS), Japan. He received his B.E. degree in Navigation from Dalian Maritime University, China, in 2005, followed by an M.E. degree in Traffic Information Engineering and Control from Shanghai Maritime University, China, in 2007. He earned his Ph.D. from the Tokyo Institute of Technology, Japan, in 2020. His research interests include high-performance computing (HPC), parallel computing, image processing, and machine learning.

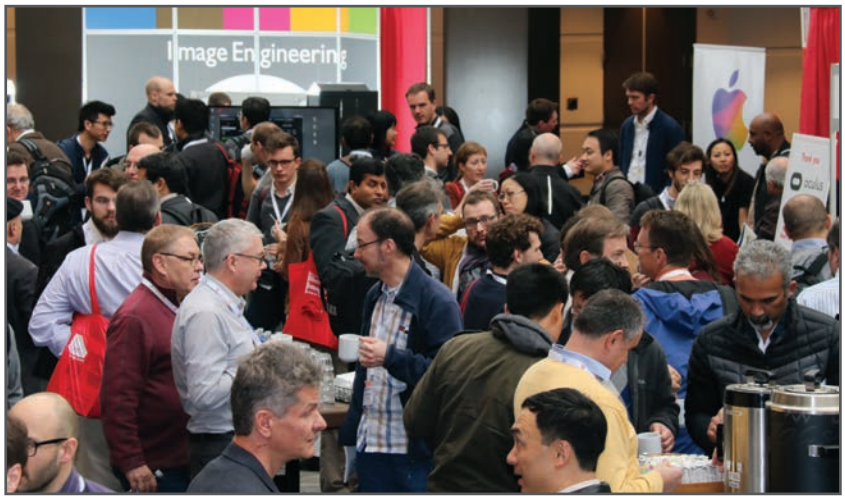
Truong Thao Nguyen received the BE and ME degrees from Hanoi University of Science and Technology, Hanoi, Vietnam, in 2011 and 2014, respectively. He received the Ph.D. in Informatics from the Graduate University for Advanced Studies, Japan in 2018. He is currently working at Digital Architecture Research Center, at National Institute of Advanced Industrial Science and Technology (AIST), where he focuses on the topics of High Performance Computing system, Distributed Deep Learning and beyond.

Rio Yokota is a professor at the Global Scientific Information and Computing Center, Tokyo Institute of Technology. His research focuses on high performance computing, linear algebra, and machine learning. He has developed several libraries, including ExaFMM for fast multipole methods, and Hatrix for hierarchical low-rank algorithms. He has received the Gordon Bell prize in 2009 using the first GPU supercomputer. Rio is a member of ACM, IEEE, and SIAM.

JOIN US AT THE NEXT EI!

electronic IMAGING

Imaging across applications . . . Where industry and academia meet!



• SHORT COURSES • EXHIBITS • DEMONSTRATION SESSION • PLENARY TALKS •
• INTERACTIVE PAPER SESSION • SPECIAL EVENTS • TECHNICAL SESSIONS •

www.electronicimaging.org

ADVANCING IMAGING SCIENCE 75 1947 IS&T 2022 imaging.org