

Advancing Image Classification and Segmentation Through Machine Learning: Architectures and Applications

Fan Liu *  and Jian Huo

College of Computer Science and Software Engineering, Hohai University, Nanjing 210098, China;
huojian@hhu.edu.cn

* Correspondence: fanliu@hhu.edu.cn

1. Introduction

Image classification and segmentation constitute the bedrock of modern computer vision, empowering systems to delineate precise object boundaries with pixel-level accuracy. As digital imagery permeates scientific and industrial domains, traditional image processing has been largely superseded by deep learning [1]. Convolutional Neural Networks (CNNs) initially established a robust baseline for extracting semantic meaning from complex visual data [2]. More recently, Vision Transformers (ViTs) and their hierarchical variants, such as the Swin Transformer, have revolutionized the field by capturing long-range dependencies and closing the gap between classification and dense prediction tasks [3,4]. Furthermore, foundational models like the Segment Anything Model (SAM) have redefined zero-shot and open-vocabulary segmentation, demonstrating remarkable generalization across diverse visual domains [5].

Simultaneously, the demand for these advanced architectures spans highly specialized and resource-constrained fields. In healthcare, transformer-based and hybrid networks enable precise anatomical modeling for medical image segmentation [6], while environmental monitoring relies on high-resolution deep learning for robust land cover analysis [7]. Despite these leaps, deploying powerful models in open-vocabulary settings without massive computational overhead [8], or operating them on resource-limited edge devices [9], remains a critical challenge. To overcome these bottlenecks, novel architectures like State Space Models (SSMs)—including Vision Mamba—have emerged as efficient alternatives to traditional attention mechanisms, offering linear computational complexity without sacrificing accuracy [10].

This Special Issue, “Application of Machine Learning to Image Classification and Image Segmentation”, brings together a diverse collection of research that highlights the versatility of modern machine learning. The contributions span a wide spectrum of applications and address critical theoretical challenges, underscoring a pivotal trend: the transition from generic models to highly specialized, efficient, and domain-aware architectures capable of operating in real-world constraints.

2. Overview of the Published Articles

A primary focus of this collection is the advancement of fundamental architectures to address persistent limitations in computational efficiency and generalization. In the realm of semantic segmentation, Ungur and Popa introduce a novel framework that integrates State Space Models (SSMs) into open-vocabulary tasks. By generating high-level guidance maps to assist CLIP feature pooling, their approach addresses the bottleneck of adapting vision-language models to dense prediction tasks, achieving superior performance with reduced



Received: 3 March 2026

Accepted: 10 March 2026

Published: 18 March 2026

Copyright: © 2026 by the authors.

Licensee MDPI, Basel, Switzerland.

This article is an open access article distributed under the terms and

conditions of the [Creative Commons Attribution \(CC BY\)](https://creativecommons.org/licenses/by/4.0/) license.

memory consumption (Contribution 1). Similarly addressing the need for efficiency, Xu et al. propose a fast normalization method for bilinear pooling. By regularizing the magnitude and distribution of eigenvalues, they achieve robust feature covariance normalization without the high computational cost of matrix square roots, thereby enhancing fine-grained image classification performance (Contribution 3).

The application of these advanced techniques to medical imaging represents another cornerstone of this Special Issue. Bolocan et al. present a comparative study on the segmentation of adrenal glands from CT scans, a task complicated by the small size and variable morphology of the organs. Their evaluation of U-Net, SegNet, and NablNet architectures highlights the specific challenges of small-organ segmentation and establishes U-Net as a robust baseline for this clinical application (Contribution 2). Expanding the scope of machine learning in healthcare, Jiménez-Gaona et al. developed a radiomic-based web tool designed to predict Vitamin D deficiency levels. By leveraging machine learning classifiers such as Random Forest on clinical and sociodemographic data, their work illustrates the potential of AI-driven diagnostic tools to complement traditional screening methods (Contribution 4).

Beyond the laboratory and clinic, this Special Issue also explores the deployment of segmentation models in unconstrained environments. Tariku et al. address the specific needs of precision agriculture with a deep learning framework for land cover mapping in vineyard ecosystems. Their study, which introduces the LICAI dataset, demonstrates how pre-trained backbones in models like DeepLabV3 can effectively monitor biodiversity and land use in complex agricultural landscapes (Contribution 5). In the critical domain of search and rescue, Zhao et al. propose a lightweight personnel segmentation network. By integrating MobileNetV2 with H-Swish activation and Convolutional Block Attention Modules (CBAMs), they achieve high-precision segmentation of individuals in disaster scenarios, proving that deep learning models can be both accurate and computationally efficient enough for mobile deployment (Contribution 6).

3. Conclusions

The research presented in this Special Issue illustrates the maturing landscape of computer vision. We are moving beyond the proof-of-concept phase into an era where machine learning models are rigorously optimized for specific tasks, whether it be distinguishing fine-grained categories, segmenting small anatomical structures, or operating on resource-constrained devices in the field. The integration of novel mechanisms, such as State Space Models and attention gates, alongside the refinement of established architectures like U-Net and PSPNet, points to a future where image analysis is not only more accurate but also more accessible and adaptable. We hope this collection serves as a catalyst for further research into efficient, explainable, and application-specific machine learning solutions.

Acknowledgments: The Guest Editor sincerely thanks all the authors for their valuable contributions and the reviewers for their constructive feedback, which significantly enhanced the quality of this Special Issue. Gratitude is also extended to the editorial team at *Applied Sciences* for their professional support throughout the publication process.

Conflicts of Interest: The authors declare no conflicts of interest.

List of Contributions

1. Ungur, V.; Popa, C.A. OpenMamba: Introducing state space models to open-vocabulary semantic segmentation. *Appl. Sci.* **2025**, *15*, 9087.
2. Bolocan, V.O.; Nicu-Canareica, O.; Mitoi, A.; Costache, M.G.; Manolescu, L.S.C.; Medar, C.; Jinga, V. Deep learning for adrenal gland segmentation: Comparing accuracy and efficiency across three convolutional neural network models. *Appl. Sci.* **2025**, *15*, 5388.

3. Xu, S.; Dong, H.; Zhang, C.; Wang, C. Fast normalization for bilinear pooling via eigenvalue regularization. *Appl. Sci.* **2025**, *15*, 4155.
4. Jiménez-Gaona, Y.; Vivanco-Galván, O.; Castillo-Malla, D.; Vivanco-Gualán, I.; Díaz-Guzmán, P. VITA-D: A radiomic web tool for predicting vitamin D deficiency levels. *Appl. Sci.* **2025**, *15*, 1798.
5. Tariku, G.; Ghiglieno, I.; Sanchez Morchio, A.; Facciano, L.; Birolleau, C.; Simonetto, A.; Serina, I.; Gilioli, G. Deep-learning-based land cover mapping in Franciacorta wine growing area. *Appl. Sci.* **2025**, *15*, 871.
6. Zhao, D.; Zhang, W.; Wang, Y. Research on personnel image segmentation based on MobileNetV2 H-Swish CBAM PSPNet in search and rescue scenarios. *Appl. Sci.* **2024**, *14*, 10675.

References

1. Minaee, S.; Boykov, Y.; Porikli, F.; Plaza, A.; Kehtarnavaz, N.; Terzopoulos, D. Image segmentation using deep learning: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2021**, *44*, 3523–3542. [[CrossRef](#)] [[PubMed](#)]
2. Alzubaidi, L.; Zhang, J.; Humaidi, A.J.; Al-Dujaili, A.; Duan, Y.; Al-Shamma, O.; Santamaría, J.; Fadhel, M.A.; Al-Amidie, M.; Farhan, L. Review of deep learning: Concepts, CNN architectures, challenges, applications, future directions. *J. Big Data* **2021**, *8*, 53. [[CrossRef](#)] [[PubMed](#)]
3. Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. An image is worth 16 × 16 words: Transformers for image recognition at scale. In Proceedings of the International Conference on Learning Representations, Vienna, Austria, 3–7 May 2021.
4. Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; Guo, B. Swin transformer: Hierarchical vision transformer using shifted windows. In Proceedings of the 2021 IEEE/CVF International Conference on Computer Vision (ICCV), Montreal, QC, Canada, 10–17 October 2021; pp. 10012–10022.
5. Kirillov, A.; Mintun, E.; Ravi, N.; Mao, H.; Roll, C.; Gustafson, L.; Xiao, T.; Whitehead, S.; Berg, A.C.; Lo, W.-Y.; et al. Segment anything. In Proceedings of the 2023 IEEE/CVF International Conference on Computer Vision (ICCV), Paris, France, 1–6 October 2023; pp. 4015–4026.
6. Hatamizadeh, A.; Tang, Y.; Nath, V.; Yang, D.; Myronenko, A.; Landman, B.; Roth, H.R.; Xu, D. Unetr: Transformers for 3d medical image segmentation. In Proceedings of the 2022 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 3–8 January 2022; pp. 574–584.
7. Yuan, Q.; Shen, H.; Li, T.; Li, Z.; Li, S.; Jiang, Y.; Xu, H.; Tan, W.; Yang, Q.; Wang, J.; et al. Deep learning in environmental remote sensing: Achievements and challenges. *Remote Sens. Environ.* **2020**, *241*, 111716. [[CrossRef](#)]
8. Liang, F.; Wu, B.; Dai, X.; Li, K.; Zhao, Y.; Zhang, H.; Zhang, P.; Vajda, P.; Marculescu, D. Open-vocabulary semantic segmentation with mask-adapted clip. In Proceedings of the 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Vancouver, BC, Canada, 17–24 June 2023; pp. 7061–7070.
9. Maaz, M.; Shaker, A.; Cholakkal, H.; Khan, S.; Zamir, S.W.; Anwer, R.M.; Shahbaz Khan, F. Edgenext: Efficiently amalgamated cnn-transformer architecture for mobile vision applications. In *European Conference on Computer Vision*; Springer Nature: Cham, Switzerland, 2022; pp. 3–20.
10. Han, D.; Wang, Z.; Xia, Z.; Han, Y.; Pu, Y.; Ge, C.; Song, J.; Song, S.; Zheng, B.; Huang, G. Demystify mamba in vision: A linear attention perspective. In Proceedings of the 38th International Conference on Neural Information Processing Systems, Vancouver, BC, Canada, 10–15 December 2024; Volume 37, pp. 127181–127203.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.