Computer Modeling in Engineering & Sciences

Tech Science Press

https://doi.org/10.32604/cmes.2026.083726

EDITORIAL

Introduction to the Special Issue on Recent Advances in Signal Processing and Computer Vision

Bo Yang[1,*] **and Chao Liu**[2]

[1]School of Automation Engineering, University of Electronic Science and Technology of China, Chengdu, China
[2]LIRMM, UMR5506, University of Montpellier-CNRS, Montpellier, France

*Corresponding Author: Bo Yang. Email: boyang@uestc.edu.cn

Received: 09 April 2026; Accepted: 14 April 2026; Published: 27 May 2026

Over the past decade, artificial intelligence, particularly deep learning, has fundamentally reshaped the fields of signal processing and computer vision. As mentioned in the introduction to this special issue, we are witnessing a significant paradigm shift. AI has evolved from recognizing the world through classification and detection to simulating it through generative models and synthesis. Most recently, AI has begun to impact the real world through embodied intelligence and robotic interaction. This special issue reflects this trajectory, presenting original research articles, reviews and methodological advances in areas such as multimodal learning, 3D vision, generative modelling, medical image analysis and autonomous systems. The 12 contributions included here reflect the current frontiers of AI, shedding light on the challenges and opportunities involved in bridging the gap between virtual intelligence and physical reality.

Reliable and interpretable AI is of paramount importance in medical applications. This issue contains several papers that address various aspects of this challenge, ranging from signal-enhanced diagnosis to surgical perception and 3D reconstruction.

In the context of ECG-based arrhythmia classification, Zhou et al. [1] proposed a dual-path multimodal framework (DM-EHC) that fuses 1D ECG temporal features with 2D time-frequency representations. Addressing class imbalance and limited feature expressiveness, this model achieves robust performance on the MIT-BIH database, thereby demonstrating the advantages of integrating multimodal signals for automated diagnosis. Talaat et al. [2], in the context of brain tumor detection, proposed a three-module BTD model that combines feature extraction using GLCM, HOG, LBP and Tamura with improved Grey Wolf Optimisation (IGWO) for feature selection, as well as a weighted majority voting ensemble of XGBoost and ANN. This model achieves an accuracy of 98.8%. Crucially, integrating LIME-based explainable AI (XAI) provides transparency in decision-making, which is a critical requirement for clinical adoption. For the early detection of mild cognitive impairment, Grigas and Maskeliunas [3] introduced a fully data-adaptive, non-learned 3D enhancement framework combining Laplacian-based local contrast modulation and a gradient-gated difference-of-Gaussians (DoG) detail injector. Unlike deep learning methods, which may introduce artefacts, or classical filters, which amplify noise, this hybrid design sharpens anatomical boundaries while maintaining a low noise gain of 1.01. On a large ADNI cohort (N = 1928), the method improves the accuracy of MCI classification from 92.63% to 95.79% (MobileNetV4, axial plane), reducing misclassification by 43%. This work demonstrates that careful signal preprocessing can substantially enhance the performance of deep learning models.

Two papers address the challenges of minimally invasive procedures for surgical perception and calibration. Xu et al. [4] reframed disparity estimation as a latent-space optimization problem using a pre-trained StyleGAN generator. An encoder network extracts disparity-relevant features from stereo image pairs and predicts an increment to the latent code, which is then refined via photometric loss. This method can recover high-fidelity disparity maps from stereo-endoscopic videos without the need for further training, thus bridging the gap between generative modelling and 3D perception in dynamic tissue environments. Yang et al. [5] proposed a unified joint calibration model for surgical needle tips and ultrasound probes. They achieve submillimeter accuracy in needle tip localization by optimizing a template coordinate system via gradient descent. They also developed an N-line-based ultrasound image registration method that enables precise spatial localization of the ultrasound image plane and its pixels.

Toan et al. [6] proposed an interpretable, algorithmic alternative to deep learning for 3D reconstruction in plastic surgery. Using four 2D facial views (front, left, right, and bottom) with pre-marked landmarks, their method uses 3D Morphable Models, image processing, and deformation techniques on a set of landmarks and sub-landmarks to reconstruct 3D nose models. This method achieves high accuracy without relying on neural networks (mean landmark error: 0.631 mm; mean boundary shape error: 1.738 mm) and demonstrates the value of transparent, non-black-box approaches in safety-critical medical contexts.

Generative models are essential for simulating realistic environments, particularly in the contexts of autonomous driving and scene understanding. For autonomous driving video synthesis, Yu and Wang [7] introduced GenScene, a world model that generates front-view driving videos based on vehicle trajectories. Unlike reconstruction-based approaches such as NeRF and 3D Gaussian Splatting, which have limited generalization capabilities and require substantial inputs, and unlike pure 2D generative models, which lack temporal coherence, GenScene uses a pretrained video diffusion backbone called Stable Video Diffusion as a strong prior to accelerate learning. A novel temporal module and an innovative attention mechanism that relates pixels within each frame to those in the initial frame significantly improve video consistency and realism. GenScene outperforms several state-of-the-art baselines. This work enables on-demand simulation to accelerate algorithm development.

Beyond medicine and generative applications, several papers advance core perception capabilities ranging from low-level image restoration to high-level recognition and cross-modal understanding. At the lowest level of vision, image restoration remains fundamental. Saxena et al. [8] challenged the conventional constant-scattering-coefficient assumption in single-image dehazing. Their proposed transmission map incorporates spatially varying scattering information and uses a linearized intensity-saturation ratio to improve edge preservation and color fidelity. This approach provides a more physically grounded solution for image restoration.

Moving on to mid-level perception, reliably detecting objects in occluded environments remains challenging. Ouardirhi et al. [9] systematically compared recent 2D and 3D object detection models across varying occlusion levels. They reveal that under heavy occlusion, 2D detectors experience an average AP drop of 10%–15%, whereas 3D models (e.g., VoxelNet) degrade by 12%–15%. This confirms that while depth cues mitigate the effects of occlusion, they do not eliminate them. The paper introduces FuDensityNet, a multimodal framework that dynamically selects between sensor-based depth input (LiDAR, ToF, and stereo) and monocular depth estimation when only 2D data is available. This offers a scalable alternative to expensive 3D sensors. Preliminary estimates suggest that this approach could reduce sensor costs by 50%–60% without significantly compromising performance when handling occlusions. This paves the way for practical, cost-effective, occlusion-aware perception.

Robust activity recognition is essential for high-level behavior understanding, especially in surveillance. Bukht et al. [10] introduced a quantum-integrated convolutional neural network (QI-CNN) for human

activity recognition in smart surveillance that combines preprocessing (filtering, tracking, and segmentation) with fuzzy optimization. Their model achieves 93.02% accuracy on the D3D-HOI dataset and 97.38% accuracy on the SYSU 3D HOI dataset by encoding classical features into an eight-qubit quantum state and processing them through a parameterized quantum circuit. Though reliant on simulators, this work points toward quantum-augmented deep learning for complex activity recognition.

Regarding CAPTCHA recognition, Derea et al. [11] proposed a novel two-layer attention framework with guided visual attention (GVA). Adapted from image captioning tasks, this method uses dual LSTM layers and achieves high accuracy (96.70% on BoC and 95.92% on Weibo) across various datasets. These results demonstrate that sophisticated attention mechanisms can effectively bypass text-based CAPTCHAs, highlighting the ongoing arms race between security and AI.

Image captioning, which connects vision and language, is a quintessential cross-modal task. Thobhani et al. [12] thoroughly examined several key techniques for image captioning, including visual attention, exploitation of semantic information, multi-caption generation, neural architecture search, few-shot learning, multi-phase learning, and cross-modal embedding. These techniques improve performance. Rather than providing an exhaustive historical review, this work identifies transformative strategies that advance the state of the art, serving as a valuable guide for researchers.

Together, the 12 papers in this special issue demonstrate the evolution of AI from an observational tool to an active agent capable of simulation, intervention, and interaction. Contributions to healthcare show how multimodal signal processing, explainability, and generative models can enhance diagnosis and surgical procedures. Generative and 3D vision papers expand the scope of realistic world simulation for autonomous driving and scene comprehension. Multimodal and attention-driven approaches continue to bridge the gap between vision and language. Meanwhile, emerging ideas, such as quantum integration, hint at future leaps in computational efficiency. Nevertheless, several cross-cutting challenges remain. These include discrepancies between simulated and real-world performance; the necessity of sample-efficient, robust learning; the interpretability of deep models in high-stakes settings; and the integration of symbolic and sub-symbolic representations. As AI extends to hardware signal processing and embodied physical interaction, we hope this special issue provides a relevant overview and serves as an inspiring reference for researchers at the intersection of signal processing, computer vision, and intelligent systems.

We thank the authors for their high-quality contributions, the reviewers for their rigorous and constructive feedback, and the editorial team for their support in preparing this special issue.

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

References

1. Zhou Y, Tian J, Kang K. Multimodal signal processing of ECG signals with time-frequency representations for arrhythmia classification. Comput Model Eng Sci. 2026;146(2):35. doi:10.32604/cmes.2026.077373.
2. Talaat FM, Salem M, Shehata M, Shaban WM. An efficient explainable AI model for accurate brain tumor detection using MRI images. Comput Model Eng Sci. 2025;144(2):2325–58. doi:10.32604/cmes.2025.067195.
3. Grigas O, Maskeliunas R. Hybrid laplacian-DoG: noise-preserving 3D FDG-PET contrast enhancement for improved MCI detection. Comput Model Eng Sci. 2026. doi:10.32604/cmes.2026.077324.
4. Xu G, Xu S, Lu S, Liu Y, Yang B, Lyu J, et al. Encoder-guided latent space search based on generative networks for stereo disparity estimation in surgical imaging. Comput Model Eng Sci. 2025;145(3):4037–53. doi:10.32604/cmes.2025.074901.
5. Yang B, Zhou Y, Tian J, Zhang X, Guo F, Liu S. A computational modeling approach for joint calibration of low-deviation surgical instruments. Comput Model Eng Sci. 2025;145(2):2253–76. doi:10.32604/cmes.2025.072031.

6. Toan NK, Tuan HNA, Thinh NT. Non-neural 3D nasal reconstruction: a sparse landmark algorithmic approach for medical applications. Comput Model Eng Sci. 2025;143(2):1273–95. doi:10.32604/cmes.2025.064218.
7. Yu B, Wang D. A trajectory-guided diffusion model for consistent and realistic video synthesis in autonomous driving. Comput Model Eng Sci. 2026;146(1):35. doi:10.32604/cmes.2026.076439.
8. Saxena G, Napte K, Shukla NK, Parihar S. Optimizing haze removal: a variable scattering approach to transmission mapping. Comput Model Eng Sci. 2025;144(2):2307–23. doi:10.32604/cmes.2025.067530.
9. Ouardirhi Z, Zbakh M, Mahmoudi SA. Bridging 2D and 3D object detection: advances in occlusion handling through depth estimation. Comput Model Eng Sci. 2025;143(3):2509–71. doi:10.32604/cmes.2025.064283.
10. Bukht TFN, Wu Y, Almujally NA, AItarbi SS, Rahman H, Jalal A, et al. Novel quantum-integrated CNN model for improved human activity recognition in smart surveillance. Comput Model Eng Sci. 2025;145(3):4013–36. doi:10.32604/cmes.2025.071850.
11. Derea Z, Zou B, Kui X, Thobhani A, Abdussalam A. A dual-layer attention based CAPTCHA recognition approach with guided visual attention. Comput Model Eng Sci. 2025;142(3):2841–67. doi:10.32604/cmes.2025.059586.
12. Thobhani A, Zou B, Kui X, Abdussalam A, Asim M, Shah S, et al. A survey on enhancing image captioning with advanced strategies and techniques. Comput Model Eng Sci. 2025;142(3):2247–80. doi:10.32604/cmes.2025.059192.