Academic Journal of Science and Technology
ISSN: 2771-3032 | Vol. 13, No. 2, 2024

Advancements in Computer Vision: A Comprehensive Survey of Image Processing and Interdisciplinary Applications

Wen Gendy[1, *], Dularia Patel[2]

[1]Department of software engineering, Beykent University, Istanbul, Turkey
[2]Computer Science & Engineering, DEPSTAR, Changa, Gujarat 388421, India
[*]Corresponding author: Wen Gendy

Abstract: Computer vision and image processing are rapidly evolving fields with broad applications across numerous domains, including healthcare, autonomous driving, surveillance, and entertainment. These fields have transformed from simple data recording techniques into sophisticated systems that incorporate digital image processing, pattern recognition, machine learning, and computer graphics. This evolution has prompted interdisciplinary interest, pushed the technology's boundaries and expanded its practical uses. This paper offers a comprehensive survey of recent advancements in computer vision, focusing on image processing and its applications across various fields. It delves into the theoretical foundations and technologies that make computer vision a valuable tool for interpreting images and videos, extracting relevant information, recognizing patterns, and understanding events. The ability of computer vision to analyze large datasets across multiple application domains makes it instrumental in tasks such as object identification, facial recognition, scene understanding, and even real-time action prediction. This versatility has established computer vision as a key driver of data-driven insights in both scientific and commercial sectors. The study categorizes computer vision into four main areas: image processing, object recognition, machine learning, and computer graphics. Each of these categories is essential to the functionality of modern computer vision systems. Image processing involves techniques for enhancing image quality and extracting important features. Object recognition and machine learning enable the identification of specific elements within images and allow systems to learn from large datasets, enhancing accuracy over time. Computer graphics, on the other hand, aid in visualizing and interpreting processed data. By offering insights into the latest techniques and evaluating their performance, this survey highlights the current state of computer vision while shedding light on future trends. Computer vision's expanding utility across various fields underscores its critical role in driving interdisciplinary innovation and addressing complex challenges.

Keywords: Computer vision, image processing, Machine Learning, Object Recognition, Interdisciplinary Innovation.

1. Introduction

Under the development of artificial intelligence based on neural network[1], such as face-action recognition[2, 3], multimedia system network[4-6], LLM & generative models[7-11], auto-driving navigation[12, 13], but also in Medical care[14], Spectroscopy[15, 16], human health[14, 17-20], environmental protection[21], scientific research[18, 22], and design optimization[23] etc., computer vision has emerged as a multifaceted field, evolving from simple raw data recording to sophisticated techniques for extracting patterns and interpreting information from images and videos. In engineering applications, progress was made by computer vision in surgery of liver transplantation[24-28]. Recent achievements in social Learning [29-36] makes even great insight in real-world application. As one of the most outstanding application field, the very biggest progress has been achieved in computer vision[37-43] with powerful software[44-46] and hardware support[47-51] like millimeter Wave technology and Robotics[52]. Nowadays, computer vision serves a wide array of applications that involve obtaining and analyzing visual information from digital inputs. Fundamentally, computer vision aims to replicate aspects of human vision, enabling machines to perceive, interpret, and make decisions based on visual data. This process requires techniques for feature extraction and event detection, which vary depending on the application domain and the nature of the data being analyzed.

Computer vision integrates two key areas: image processing and pattern recognition[38-41, 43]. While image processing primarily focuses on the manipulation and enhancement of images to improve quality or extract useful features, pattern recognition involves identifying and categorizing objects or patterns within images. Together, these fields enable the creation of algorithms that can interpret spatial data, ultimately leading to what is known as "image understanding." The development of computer vision techniques is heavily inspired by human visual capabilities, although achieving a fully human-like vision system remains beyond current technology due to limitations in machine interpretation and processing capabilities. The goals of computer vision and image processing, while related, are distinct. The primary objective of computer vision is to create models that extract relevant data and interpret it, allowing for image understanding. In contrast, image processing involves computational transformations such as sharpening, contrast adjustment, and other enhancements that optimize visual clarity. While closely related, these two fields sometimes overlap with Human-Computer Interaction (HCI)[53], which focuses on the design and interaction between humans and computers. HCI has developed as an interdisciplinary field that explores how humans interact with technology, considering user-centered design, interface ergonomics, and the overall effectiveness of human-computer interactions.

However, unlike HCI, computer vision focuses specifically on interpreting visual data.

Functionally, both computer vision and human vision are aimed at interpreting spatial data, allowing for an understanding of objects, scenes, and actions. However, the performance of computer vision systems remains limited compared to the human visual system. Despite advancements, computer vision cannot fully replicate the nuances of human sight, as the systems lack the same adaptability and contextual understanding. Numerous challenges remain in developing algorithms that match the accuracy, flexibility, and complexity of human perception. Key difficulties include sensitivity to parameter tuning, the robustness of algorithms under varied conditions, and ensuring accurate results across diverse scenarios. The performance evaluation of computer vision systems is complex, requiring rigorous testing to measure attributes like accuracy, robustness, and extensibility. This involves examining the core behaviors of algorithms under different conditions, as well as their ability to adapt and maintain performance. Given these challenges, there is significant effort among scholars to enhance computer vision algorithms and categorize them into specialized areas of application. These applications range from automation in manufacturing and assembly lines to remote sensing, robotics, human-computer communication, and assistive tools for the visually impaired.

Ultimately, computer vision's advancement depends on improvements in computer technology, from processing power to algorithmic sophistication. As technology progresses, the field continues to expand, offering new ways to interpret and analyze visual information. This survey highlights computer vision's essential role in bridging the gap between human visual capabilities and machine-based analysis, fostering innovations that impact various industries and pushing the boundaries of what machines can perceive and understand.

2. Recent Research Progress

Computer vision employs advanced algorithms and optical sensors to mimic aspects of human vision[47], enabling the automated extraction of valuable information from objects. Unlike traditional methods, which often require extensive time and complex laboratory analysis, computer vision integrates artificial intelligence to achieve faster and more effective visual data interpretation. This technology is frequently paired with sophisticated lighting systems to optimize image acquisition and enhance the accuracy of image analysis. By capturing digital images and improving their quality through preprocessing, computer vision systems can isolate relevant objects from backgrounds, measure significant features, and then interpret the data with precision.

Recent advancements in image processing have enabled the creation of highly accurate digital recognition systems, transforming how visual data is analyzed across various fields[54]. These developments allow for streamlined data collection and more sophisticated image interpretation, opening up new possibilities for applications such as automated quality control, medical imaging, and autonomous navigation[55]. Computer vision has thus evolved into a powerful tool that can rapidly analyze images and extract meaningful information, enhancing traditional visual methods by providing a structured, efficient approach to image-based analysis, for example:

Image processing in digital sound systems: Reconstructs audio from phonograph recordings using precision metrology and digital image processing to measure groove shapes without direct contact.

Image processing in food analysis: Involves feature selection, extraction, and classification to analyze food and beverage imagery, focusing on role portrayal and industry impact.

Convolutional Neural Networks (CNN) for object detection[56]: Uses artificial neural networks (ANN) for edge detection in image processing tasks.

Digital image processing has become a prominent area within computer vision, driven by advances in several theoretical fields, such as mathematics, linear algebra, statistics, scientific computing, and computational neuroscience. These fields provide the foundations and methodologies that support the development and refinement of image processing techniques. Key areas of focus include methods for depth map estimation, where Bayesian techniques[57] are used to restore 3D scene structures, showing strong depth estimation performance with training data but challenges with natural images. Similarly, image quality assessment for retargeting methods is enhanced by top-down approaches, offering consistent results based on specific quality metrics.

Research also extends into psychophysical experiments that examine human visual perception under varying lighting conditions, achieved through dynamic display and HDR technology, allowing for pixel differentiation across a wide luminance range. Additionally, in applied settings such as driver assistance systems, contrast sensitivity techniques help develop algorithms that monitor driver visibility and offer real-time mentoring for speed adjustment in low-visibility situations. Another key area of development is image-based illumination enhancement, where color pixel correction and decomposition methods provide a more refined approach than traditional histogram equalization techniques, offering improved visual clarity in low-light or high-contrast environments. These advancements collectively highlight the evolving landscape of digital image processing, emphasizing improved methodologies and application-specific adaptations.

Pattern recognition, a fundamental area within computer vision, is dedicated to identifying objects through various image transformation techniques that improve quality and enable precise interpretation. This process facilitates information extraction and decision-making based on sensor-captured images, with the overarching goal of enabling machines to "see" in a manner similar to human vision. Computer vision accomplishes this by following a structured workflow that generally includes stages such as image acquisition, preprocessing, feature extraction, segmentation or detection, high-level processing, and decision-making.

As the most popular method, the convolutional neural network-based object detection are widely used in many applications and research[58, 59]. While, the emphasis is on two phases detectors such as the Region-based Convolutional Neural Network, like R-CNN family[60]. These stages help create an organized pipeline for visual information processing and object identification. Commonly, computer vision frameworks leverage two primary methodologies: 3D morphological analysis and pixel optimization. While 3D morphological analysis has established itself as a standard approach for image processing and pattern recognition, pixel optimization involves an in-depth analysis of pixel structures, including their morphology and internal characteristics, to enhance understanding of vector functions. To gain a holistic perspective, these approaches are typically applied to large datasets covering multiple layers of geometric composition, making efficient and precise algorithms essential for extracting relevant quantitative information and understanding complex color clusters.

Integrating 3D morphological analysis with artificial intelligence methods[38]—such as fuzzy logic, artificial neural networks, and genetic algorithms—further enhances the accuracy and efficiency of computer vision algorithms, especially in scenarios requiring large-scale data processing. These methods allow for the completion of intricate tasks that benefit from automated visual insights, which would otherwise require significant human input. In practical applications, computer vision employs two major approaches for working with image data: segmentation and retrieval. Segmentation involves dividing an image into distinct regions, each of which is comprised of pixels with similar characteristics, such as color, texture, or gray level. The segmentation process is crucial for accurately detecting and interpreting objects, as it creates regions that can be analyzed independently, making it a cornerstone in applications where object recognition and classification are required. Retrieval, in contrast, involves using these segmented regions to enable the search for similar images, supporting systems like content-based image retrieval and enhancing the functionality of image-based search engines. Together, segmentation and retrieval allow computer vision systems to process, categorize, and utilize visual data effectively, supporting a wide range of applications from automated quality inspection to advanced search functionalities.

In image segmentation, popular techniques include methods based on intensity, color, and shape, as well as edge or border detection, all of which contribute to enhanced object recognition. In the literature, segmentation accuracy is often demonstrated on a small sample of images, while large-scale image databases require specific parameter settings for accurate classification. Advanced segmentation approaches may utilize techniques like gradient texture analysis, feature space exploration, and unsupervised clustering to localize objects accurately and define boundaries more precisely. Segmentation's ultimate objective is to create a resemblance map derived from prominent object detection models or hierarchical segmentation processes. This approach supports a salience mapping model that aims to highlight prominent areas of the image. Figure1 shows the different stages in image segmentation. A model for this purpose would require the computation of pixel salience values mapped to specific locations within a hierarchical salience map. Some researchers propose an aggregation model using standard saliency methods to assign salience scores across all pixels and segments, labeling them into prominent clusters.

However, challenges remain with such aggregation models, as they may lack a nuanced approach to interactions between neighboring pixels, which can affect segmentation accuracy in densely packed or complex visual data. For instance, while pixel-wise aggregation helps establish model parameters, it may overlook the local dependencies that play a crucial role in defining object edges and transitions in detailed image structures. Addressing these limitations, current research in computer vision segmentation is exploring more sophisticated models that take into account the relational dynamics between pixels, thus improving the reliability and depth of visual analysis across various applications. As computer vision continues to evolve, these innovations in segmentation and pattern recognition are laying the groundwork for more adaptive and context-aware visual systems capable of navigating and interpreting real-world environments with increasing precision.

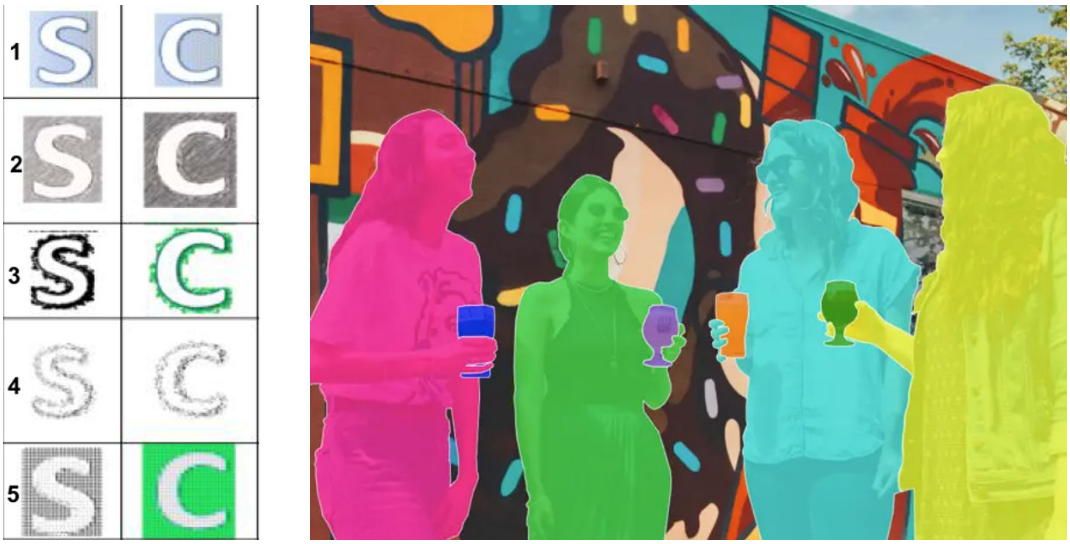

Figure 1. Segmentation in image processing: 1. Input image, 2. segmented map before integration, 3. Edge map before integration, 4. Segmented map and edge map after combination, 5. Pixel clustering

To address these challenges, Khan proposed the use of Conditional Random Fields (CRF) to combine calibration maps from multiple methods and incorporate values from neighboring pixels. The CRF aggregation model enhances parameter optimization during training, as the reliability of each pixel achieves a higher prominence when trained within this framework. Data extraction involves capturing objects photographed by cameras, sensors, or satellite devices in the form of single images or image sequences. The primary goal of this extraction is to separate background elements from foreground objects. This process can result in three types of outputs: (a) the objects retain their original color, (b) the objects are transformed to black and white, or (c) the objects are rendered transparent.

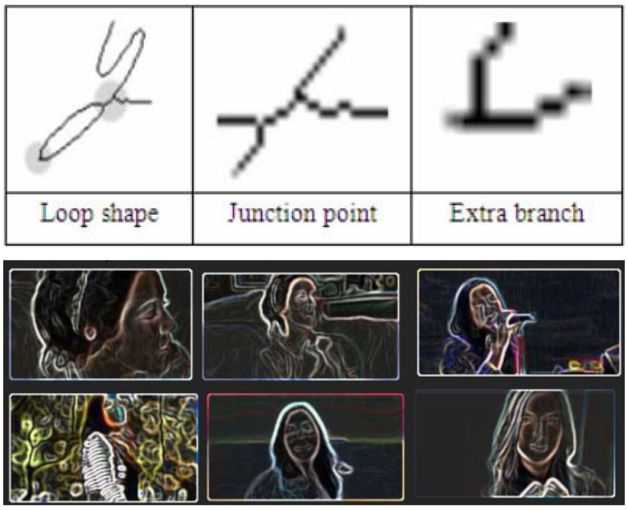

Figure 2. Shape feature extraction from simple sketch to complex human image.

As illustrated in Figure2, the extraction process in computer vision involves several steps: (a) converting object instances to black and white, (b) adjusting object sizes based on a scale factor, (c) applying transparency or color combinations to certain elements, and (d) scaling and repositioning objects, often resulting in a different appearance from their original form. Pixels play a crucial role in determining object sharpness within an image, making pixel optimization essential for tasks like object detection, segmentation, and recognition.

In boundary-based techniques, an edge detector is used to identify object boundaries by detecting rapid intensity changes along region edges. For color segmentation, this detection is performed on each RGB color channel, creating edges that can be combined to form the final edge image. Local-based techniques, on the other hand, group pixels according to uniformity criteria, such as in region-growing and split-and-merge methods. In region-growing, pixels expand from core points into larger areas if they share similar features, such as color or gray values. Split-and-merge techniques begin by dividing an image into smaller regions, which are then combined based on specific criteria. Despite their utility, these region-based techniques have two main limitations: (1) they rely heavily on initial global criteria, which affects regional growth, and (2) the process depends on initial segments and original pixel values, impacting object detection performance. Object detection itself is used to search for and identify objects within both recorded and real-time datasets, but it often has a margin of error when objects deviate from the pattern specified by the algorithm. To enhance accuracy, additional algorithms are frequently implemented to detect smaller features for greater detail. For example, in face detection, algorithms are employed to identify lower-resolution facial elements, such as eyes, eyebrows, and mouth, thereby improving machine accuracy. However, peripheral features like ears and neck are less commonly studied in detail, underscoring a selective focus in facial recognition processing.

A bitmap, also known as a raster image, is an image format that represents an image as a collection of pixels on a computer screen, with each pixel assigned a specific color. Bitmap images are widely used in photographs and digital images, but their quality degrades when enlarged. If a bitmap image is scaled up, for instance by a factor of four, the pixels themselves are enlarged, leading to a blurred or pixelated appearance. Key terms when working with bitmap images include resolution, which refers to the number of pixels in the image, and color depth, which indicates the range of colors each pixel can display. Bitmap images are often generated using scanners, digital cameras, or video capture devices and are susceptible to various forms of noise. To address this, bitmap templates, which are standardized and easily processed by computers, are often used as benchmarks. In contrast, raw bitmap images may contain acquisition errors, resulting in unstructured pixel values that do not accurately represent the actual scene.

Noise can enter a bitmap image in several ways, depending on how the image was captured. For example, when a photograph is scanned from film, the film grain can contribute to noisy pixels, as can potentially damage to the film or interference from the scanner itself. When images are acquired directly in a digital format, noise may be introduced by the data collection mechanisms, such as CCD detectors, or through electronic data transmission, which can degrade image quality. Figure3 shows some examples for image retrievals from big open database. Image processing research aims to develop machine learning and computing methods capable of recognizing patterns in increasingly varied objects. Machine learning, which intersects with computational statistics, is essential in applications like spam filtering, optical character recognition, search engines, and computer vision. Numerous algorithms, such as Gaussian-based linear filtering, have been developed to reduce noise. These algorithms effectively reduce grain noise by averaging pixel values within a local region, which helps diminish local variations caused by noise, ultimately producing clearer, more accurate images.

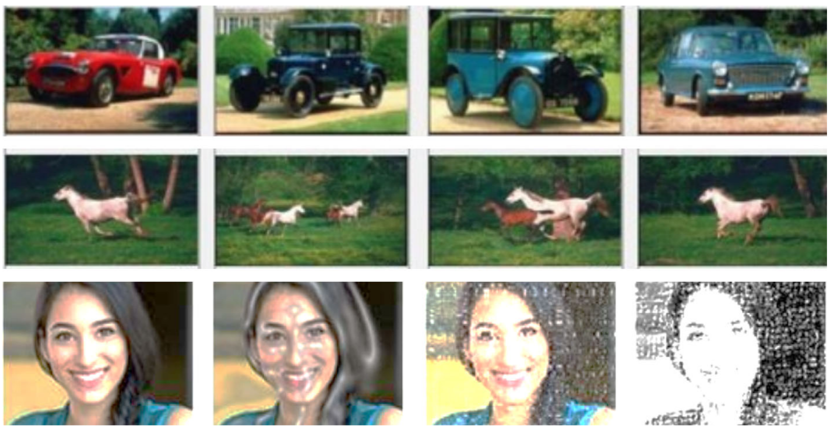

Figure 3. Example of image retrievals using query image from big datasets of PASCAL MTH, MSD etc.

3. Conclusion and Outlook

In conclusion, computer vision, anchored in image processing and machine learning, has significantly advanced numerous technological fields by enabling sophisticated analysis and interpretation of visual data. With roots in image processing, computer vision has become essential in disciplines like geographical remote sensing, robotics, healthcare, and satellite communication, where it enables efficient feature extraction and predictive analysis. By analyzing images and videos, researchers can study patterns, detect behaviors, and make data-driven predictions, broadening the impact of computer vision on both individual and environmental applications. Looking ahead, the continued integration of machine learning and improved image processing algorithms will drive further innovation in computer vision, enhancing both accuracy and adaptability. In healthcare, it offers promising advancements in diagnostics and patient monitoring, while in autonomous systems like self-driving vehicles, it supports navigation and safety. Environmental applications, such as climate monitoring and disaster response, also stand to benefit from improved computer vision capabilities. As the field evolves, future research will likely address challenges like enhancing accuracy under varied conditions, real-time processing, and reducing the computational demands of intensive algorithms. Combined with emerging technologies like augmented reality, the Internet of Things, and edge computing, computer vision is set to become a foundational tool across industries, empowering new ways of understanding and acting on visual information.

References

[1] Y. Wang, Y. Guo, R. Kumar, M. Swaminathan, Order Reduction Using Laguerre-FDTD with Embedded Neural Network, 2024 IEEE/MTT-S International Microwave Symposium-IMS 2024, IEEE, 2024, pp. 473-476.

[2] R. Li, S. Sun, M. Elhoseiny, P. Torr, OxfordTVG-HIC: Can Machine Make Humorous Captions from Images?, Proceedings of the IEEE/CVF International Conference on Computer Vision, 2023, pp. 20293-20303.

[3] X. Chen, K. He, W. Liu, X. Liu, Z.-J. Zha, T. Mei, CLaM: An Open-Source Library for Performance Evaluation of Text-driven Human Motion Generation, Proceedings of the 32nd ACM International Conference on Multimedia, 2024, pp. 11194-11197.

[4] X. Chen, W. Liu, X. Liu, Y. Zhang, T. Mei, A cross-modality and progressive person search system, Proceedings of the 28th ACM International Conference on Multimedia, 2020, pp. 4550-4552.

[5] X. Chen, X. Liu, K. Liu, W. Liu, T. Mei, A baseline framework for part-level action parsing and action recognition, arXiv preprint arXiv:2110.03368 (2021).

[6] X. Chen, X. Liu, W. Liu, K. Liu, D. Wu, Y. Zhang, T. Mei, Part-level Action Parsing via a Pose-guided Coarse-to-Fine Framework, 2022 IEEE International Symposium on Circuits and Systems (ISCAS), IEEE, 2022, pp. 419-423.

[7] M. Qu, X. Chen, W. Liu, A. Li, Y. Zhao, ChatVTG: Video Temporal Grounding via Chat with Video Dialogue Large Language Models, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 1847-1856.

[8] L. Yang, Z. Zhang, J. Han, B. Zeng, R. Li, P. Torr, W. Zhang, Semantic Score Distillation Sampling for Compositional Text-to-3D Generation, arXiv preprint arXiv:2410.09009 (2024).

[9] Z. Gui, S. Sun, R. Li, J. Yuan, Z. An, K. Roth, A. Prabhu, P. Torr, kNN-CLIP: Retrieval Enables Training-Free Segmentation on Continually Expanding Large Vocabularies, arXiv preprint arXiv:2404.09447 (2024).

[10] B. Wang, H. Duan, Y. Feng, X. Chen, Y. Fu, Z. Mo, X. Di, Can LLMs Understand Social Norms in Autonomous Driving Games?, arXiv preprint arXiv:2408.12680 (2024).

[11] H. Liu, X. Chen, X. Liu, X. Gu, W. Liu, AnimateAnywhere: Context-Controllable Human Video Generation with ID-Consistent One-shot Learning, Proceedings of the 5th International Workshop on Human-centric Multimedia Analysis, 2024, pp. 41-43.

[12] M. Yin, T. Li, H. Lei, Y. Hu, S. Rangan, Q. Zhu, Zero-Shot Wireless Indoor Navigation through Physics-Informed Reinforcement Learning, 2024 IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2024, pp. 5111-5118.

[13] K. Huang, X. Chen, X. Di, Q. Du, Dynamic driving and routing games for autonomous vehicles on networks: A mean field game approach, Transportation Research Part C: Emerging Technologies 128 (2021) 103189.

[14] J. Huo, H. Li, J. Roveda, S.F. Quan, A. Li, A Multi-task Deep Learning Algorithm for Sleep Stage Scoring and Sleep Arousal Detection, Authorea Preprints (2023).

[15] H. Guo, A.B. Tikhomirov, A. Mitchell, I.P.J. Alwayn, H. Zeng, K.C. Hewitt, Real-time assessment of liver fat content using a filter-based Raman system operating under ambient light through lock-in amplification, Biomedical Optics Express 13(10) (2022) 5231-5245.

[16] H. Guo, B.L. Gala-Lopez, I.P. Alwayn, K.C. Hewitt, Liver discard rate due to conservative estimations of steatosis: an inference-based approach, medRxiv (2023) 2023.12. 04.23299406.

[17] J. Huo, Machine Learning Application in Sleep Disorder Analysis, The University of Arizona, 2023.

[18] C. Ding, T. Yao, C. Wu, J. Ni, Deep Learning for Personalized Electrocardiogram Diagnosis: A Review, arXiv preprint arXiv:2409.07975 (2024).

[19] J. Huo, S.F. Quan, J. Roveda, A. Li, BASH-GN: a new machine learning–derived questionnaire for screening obstructive sleep apnea, Sleep and Breathing 27(2) (2023) 449-457.

[20] J. Yang, Research on the propagation model of COVID-19 based on virus dynamics, Second International Conference on Biological Engineering and Medical Science (ICBioMed 2022), SPIE, 2023, pp. 962-967.

[21] J. Yang, Predicting water quality through daily concentration of dissolved oxygen using improved artificial intelligence, Scientific Reports 13(1) (2023) 20370.

[22] J. Huo, Y. Wang, N. Wang, W. Gao, J. Zhou, Y. Cao, Data-driven design and optimization of ultra-tunable acoustic metamaterials, Smart Materials and Structures 32(5) (2023) 05LT01.

[23] Y. Guo, O.W. Bhatti, M. Swaminathan, Training Set Optimization with Uncertainty Quantification for Machine Learning Models of Electromagnetic Structures, 2022 IEEE Electrical Design of Advanced Packaging and Systems (EDAPS), IEEE, 2022, pp. 1-3.

[24] H. Guo, A.E. Stueck, J.B. Doppenberg, Y.S. Chae, A.B. Tikhomirov, H. Zeng, M.A. Engelse, B.L. Gala-Lopez, A. Mahadevan-Jansen, I.P. Alwayn, Evaluation of minimum-to-

severe global and macrovesicular steatosis in human liver specimens: a portable ambient light-compatible spectroscopic probe, medRxiv (2023) 2023.12. 04.23299259.

[25] H. Guo, V.S. Zions, B.A. Law, K.C. Hewitt, Potential of Raman-Reflectance Combination in Quantifying Liver Steatosis and Fat Droplet Size: Evidence From Monte Carlo Simulations and Phantom Studies, Journal of Biophotonics (2024) e202400156.

[26] H. Guo, A.E. Stueck, J.B. Doppenberg, Y.S. Chae, A.B. Tikhomirov, H. Zeng, B.L. Gala-Lopez, A. Mahadevan-Jansen, M.A. Engelse, I.P. Alwayn, Assessment of liver steatosis using an ambient light-compatible Raman system: enhancing specificity with supplementary reflectance information, Biomedical Vibrational Spectroscopy 2024: Advances in Research and Industry, SPIE, 2024, p. PC128390B.

[27] H. Guo, A.E. Stueck, A.B. Tikhomirov, H. Zeng, I.P. Alwayn, B.L. Gala-Lopez, A. Mahadevan-Jansen, A.K. Locke, K.C. Hewitt, Evaluation of Steatosis in Human Liver Specimens Using an Ambient Light-compatible Raman Spectroscopy Approach, Bio-Optics: Design and Application, Optica Publishing Group, 2023, p. JTu4B. 26.

[28] H. Guo, A.E. Stueck, J.B. Doppenberg, Y.S. Chae, A.B. Tikhomirov, H. Zeng, M.A. Engelse, B.L. Gala-Lopez, A. Mahadevan-Jansen, I.P. Alwayn, Evaluation of Minimum-To-Severe Global and Macrovesicular Steatosis in Human Liver Specimens: A Portable Ambient Light-Compatible Spectroscopic Probe, Journal of Biophotonics (2023) e202400292.

[29] Z. Shou, X. Chen, Y. Fu, X. Di, Multi-agent reinforcement learning for Markov routing games: A new modeling paradigm for dynamic traffic assignment, Transportation Research Part C: Emerging Technologies 137 (2022) 103560.

[30] S. Liu, Y. Wang, X. Chen, Y. Fu, X. Di, SMART-eFlo: An integrated SUMO-gym framework for multi-agent reinforcement learning in electric fleet management problem, 2022 IEEE 25th International Conference on Intelligent Transportation Systems (ITSC), IEEE, 2022, pp. 3026-3031.

[31] X. Chen, S. Liu, X. Di, A hybrid framework of reinforcement learning and physics-informed deep learning for spatiotemporal mean field games, In Proceedings of the 20th International Conference on Autonomous Agents and Multiagent Systems, ACM DIgital Library, 2023.

[32] F. Zhou, C. Zhang, X. Chen, X. Di, Graphon Mean Field Games with A Representative Player: Analysis and Learning Algorithm, arXiv preprint arXiv:2405.08005 (2024).

[33] X. Chen, S. Liu, X. Di, Learning Dual Mean Field Games on Graphs, ECAI, 2023, pp. 421-428.

[34] S. Liu, X. Chen, X. Di, Scalable Learning for Spatiotemporal Mean Field Games Using Physics-Informed Neural Operator, Mathematics 12(6) (2024) 803.

[35] X. Chen, Z. Li, X. Di, Social learning in Markov games: Empowering autonomous driving, 2022 IEEE Intelligent Vehicles Symposium (IV), IEEE, 2022, pp. 478-483.

[36] X. Chen, X. Di, Z. Li, Social Learning for Sequential Driving Dilemmas, Games 14(3) (2023) 41.

[37] Z. Hu, Y. Sun, Y. Yang, Switch to generalize: Domain-switch learning for cross-domain few-shot classification, International Conference on Learning Representations, 2022.

[38] Z. Hu, Y. Sun, Y. Yang, Suppressing the heterogeneity: A strong feature extractor for few-shot segmentation, The Eleventh International Conference on Learning Representations, 2023.

[39] X. Chen, X. Liu, W. Liu, X.-P. Zhang, Y. Zhang, T. Mei, Explainable person re-identification with attribute-guided metric distillation, Proceedings of the IEEE/CVF international conference on computer vision, 2021, pp. 11813-11822.

[40] Z. Hu, Y. Sun, Y. Yang, J. Zhou, Divide-and-regroup clustering for domain adaptive person re-identification, Proceedings of the AAAI Conference on Artificial Intelligence, 2022, pp. 980-988.

[41] X. Chen, W. Liu, X. Liu, Y. Zhang, J. Han, T. Mei, MAPLE: Masked pseudo-labeling autoencoder for semi-supervised point cloud action recognition, Proceedings of the 30th ACM International Conference on Multimedia, 2022, pp. 708-718.

[42] Z. Hu, Y. Sun, J. Wang, Y. Yang, DAC-DETR: Divide the attention layers and conquer, Advances in Neural Information Processing Systems 36 (2024).

[43] X. Chen, W. Liu, Q. Bao, X. Liu, Q. Yang, R. Dai, T. Mei, Motion Capture from Inertial and Vision Sensors, arXiv preprint arXiv:2407.16341 (2024).

[44] Z. Hu, J. Ye, Y. Zhang, X. Wang, Seeing is Not Always Believing: An Empirical Analysis of Fake Evidence Generators, 2024 IEEE 9th European Symposium on Security and Privacy (EuroS&P), IEEE, 2024, pp. 560-579.

[45] Y. Zhang, Z. Hu, X. Wang, Y. Hong, Y. Nan, X. Wang, J. Cheng, L. Xing, Navigating the Privacy Compliance Maze: Understanding Risks with {Privacy-Configurable} Mobile {SDKs}, 33rd USENIX Security Symposium (USENIX Security 24), 2024, pp. 6543-6560.

[46] M. Yin, Data security and privacy preservation in big data age, 2nd International Conference on Mechatronics Engineering and Information Technology (ICMEIT 2017), Atlantis Press, 2017, pp. 387-391.

[47] M. Yin, A.K. Veldanda, A. Trivedi, J. Zhang, K. Pfeiffer, Y. Hu, S. Garg, E. Erkip, L. Righetti, S. Rangan, Millimeter wave wireless assisted robot navigation with link state classification, IEEE Open Journal of the Communications Society 3 (2022) 493-507.

[48] V. Semkin, M. Yin, Y. Hu, M. Mezzavilla, S. Rangan, Drone detection and classification based on radar cross section signatures, 2020 International Symposium on Antennas and Propagation (ISAP), IEEE, 2021, pp. 223-224.

[49] M. Yin, Millimeter Wave Wireless Assisted Indoor Robot Navigation, New York University Tandon School of Engineering, 2024.

[50] K. Pfeiffer, Y. Jia, M. Yin, A.K. Veldanda, Y. Hu, A. Trivedi, J. Zhang, S. Garg, E. Erkip, S. Rangan, Path planning under uncertainty to localize mmWave sources, 2023 IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2023, pp. 3461-3467.

[51] Y. Hu, M. Yin, W. Xia, S. Rangan, M. Mezzavilla, Multi-frequency channel modeling for millimeter wave and thz wireless communication via generative adversarial networks, 2022 56th Asilomar Conference on Signals, Systems, and Computers, IEEE, 2022, pp. 670-676.

[52] S. Cao, J. Xiao, Human-Robot Complementary Collaboration for Flexible and Precision Assembly, 2024 IEEE International Conference on Robotics and Automation (ICRA), IEEE, 2024, pp. 12971-12977.

[53] T. Kosch, J. Karolus, J. Zagermann, H. Reiterer, A. Schmidt, P.W. Woźniak, A survey on measuring cognitive workload in human-computer interaction, ACM Computing Surveys 55(13s) (2023) 1-39.

[54] L. Liu, W. Ouyang, X. Wang, P. Fieguth, J. Chen, X. Liu, M. Pietikäinen, Deep learning for generic object detection: A survey, International journal of computer vision 128 (2020) 261-318.

[55] E. Imani, G. Zhang, R. Li, J. Luo, P. Poupart, P.H. Torr, Y. Pan, Label Alignment Regularization for Distribution Shift, Journal of Machine Learning Research 25(247) (2024) 1-32.

[56] Y. Guo, X. Li, M. Swaminathan, 2D spectral transposed convolutional neural network for S-parameter predictions, 2022 IEEE 31st Conference on Electrical Performance of Electronic Packaging and Systems (EPEPS), IEEE, 2022, pp. 1-3.

[57] M. Swaminathan, O.W. Bhatti, Y. Guo, E. Huang, O. Akinwande, Bayesian learning for uncertainty quantification, optimization, and inverse design, IEEE Transactions on Microwave Theory and Techniques 70(11) (2022) 4620-4634.

[58] Y. Fu, A. Jain, X. Di, X. Chen, Z. Mo, DriveGenVLM: Real-world Video Generation for Vision Language Model based Autonomous Driving, arXiv preprint arXiv:2408.16647 (2024).

[59] X. Chen, F. Yongjie, S. Liu, X. Di, Physics-informed neural operator for coupled forward-backward partial differential equations, 1st Workshop on the Synergy of Scientific and Machine Learning Modeling@ ICML2023, 2023.

[60] S. Sun, R. Li, P. Torr, X. Gu, S. Li, Clip as rnn: Segment countless visual concepts without training endeavor, Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2024, pp. 13171-13182.