electronics

MDPI

Editorial

Object Detection, Segmentation and Categorization in Artificial Intelligence

Hao Li [1,*], Fei Xie [2], Jianbo Zhou [3] and Jieyi Liu [1]

[1] Key Laboratory of Collaborative Intelligence Systems, School of Electronic Engineering, Ministry of Education, Xidian University, Xi'an 710071, China; jieyiliu@xidian.edu.cn
[2] Academy of Advanced Interdisciplinary Research, Xidian University, Xi'an 710071, China; fxie@xidian.edu.cn
[3] Shaanxi Key Laboratory of Underwater Information Technology, School of Marine Science and Technology, Northwestern Polytechnical University, Xi'an 710072, China; jbzhou@nwpu.edu.cn
* Correspondence: haoli@xidian.edu.cn

In the field of computer vision, three basic tasks are particularly important: object detection [1], segmentation [2] and categorization [3]. These tasks form the basis for understanding and interpreting visual data and are widely used in key areas such as autonomous driving, medical imaging, and surveillance. They not only promote the progress of computer vision technology, but also provide solid technical support for various application scenarios, enabling machines to interpret and analyze visual information more accurately. The continuous development of these technologies improves the performance and efficiency of the machine systems and promotes their widespread implementation in all walks of life [4].

Traditional object detection, segmentation and categorization algorithms have been instrumental to the early development of computer vision [5–7]. However, with the rise of deep learning methods, the shortcomings of traditional methods have become clear. Most traditional methods require a great deal of domain knowledge to design and select features, and the generalization ability of these features is limited, making it difficult to apply these methods to large-scale datasets.

Recently, artificial intelligence and deep learning have flourished, and a large number of outstanding artificial intelligence algorithms have emerged for use in object detection, segmentation, and categorization tasks [8–14]. Compared with traditional methods, these algorithms usually have significant advantages such as automatic feature learning, end-to-end training, high precision and robustness, ability to process massive data, flexibility to adapt to multiple tasks and efficient feature expression.

Moreover, artificial intelligence algorithms often achieve excellent real-time performance, especially in applications such as real-time monitoring and autonomous driving [15]. The fast reasoning ability of artificial intelligence methods enables the system to identify and respond to changes and events in the environment, such as traffic signs, pedestrians and obstacles, in a timely manner, ensuring the safety and stability of the system.

Indoor scenes are important urban spaces, and logos are very important for the accurate operation of mobile robots in indoor environment. Therefore, in Contribution 1, a logos detection method named MobileNetV2-YOLOv4-UP is proposed. This method combines unsupervised learning and few-shot learning to obtain a potential feature representation of the logo by pre-training the autoencoder on a public unlabeled logo dataset, and then incorporating a small number of labeled indoor scene logo datasets into the training to update the weight of the logos detection network. The related research on logo detection, few-shot learning and unsupervised learning is also reviewed in this paper.

In recent years, multi-objective and multi-task evolutionary algorithms have attracted a great deal of attention [16]. Evolutionary algorithms can tackle complex nonlinear optimization problems and find the global optimal solution; they can also optimize multiple

check for updates

Citation: Li, H.; Xie, F.; Zhou, J.; Liu, J. Object Detection, Segmentation and Categorization in Artificial Intelligence. *Electronics* **2024**, *13*, 2650. https://doi.org/10.3390/electronics13132650

Received: 20 June 2024
Accepted: 4 July 2024
Published: 5 July 2024

Copyright: © 2024 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) license (https://creativecommons.org/licenses/by/4.0/).

targets at the same time and find a solution to balance each target through the Pareto optimal solution set. In light of this, a multi-objective automatic clustering algorithm based on evolutionary multi-task optimization is introduced (Contribution 2). It introduces multi-task learning in multi-objective optimization, realizes knowledge sharing among different clustering tasks, and improves the performance and efficiency of clustering.

A novel method called causal semantic representation learning with domain intervention (CSRDN) is proposed (Contribution 3). Based on the data generation process in the natural environment, it uses the structural causal model (SCM) to construct an inclusive causal directed acyclic graph (DAG) in order to simulate domain generalization tasks. Proxy domain variables are introduced into the causality diagram to explain the limitations of generalization performance and simulate domain shift. Meanwhile, the domain change is regarded as an intervention, the confounding effect is eliminated by controlling the proxy domain variables, and the unstable bias information flow is blocked. Finally, counterfactual reasoning is used to generate counterfactual representations that are consistent with the original input semantics but are influenced by different domains to improve the generalization ability of deep learning models in unknown environments.

Unlike natural images, hyperspectral images (HSIs) typically contain tens to hundreds of bands. Each band corresponds to a specific wavelength range that captures the detailed spectral properties of an object. Endmember extraction plays an important role in HSI analysis by extracting representative spectra to facilitate more accuracy when performing feature classification, mineral identification, agricultural monitoring, environmental monitoring, and urban planning tasks. An endmember extraction method based on evolutionary competitive multi-task optimization is proposed (Contribution 4), which considers the problem of endmember extraction with different numbers of endmembers as a set of optimization tasks and assumes that there is a competitive relationship between these tasks. Online resource allocation is then used to allocate appropriate computational resources to these tasks. Finally, experiments are conducted on simulated and real hyperspectral datasets to verify the effectiveness of the proposed method.

A composite infrared dim and small target detection model, USES-Net, is proposed (Contribution 5). Based on a typical encoder–decoder structure, it introduces the self-attention mechanism of Swin Transformer to extract potential features related to dim and small targets with a larger sensory field, and designs a bottom-up cross-layer feature fusion module (AFM), which can reconstruct the acquired feature information on different scales of the target feature information. The model effectively improves the detection accuracy.

In the field of rotated object detection, existing regression- and classification-based methods suffer from multiple-solution issues and boundary issues in predicting rotation angles, which limit the upper bound of detector performance. To address these issues, an improved rotation object detection method is introduced (Contribution 6), which proposes a method based on the circular Gaussian distribution to predict angles. This method converts the marked angles into a discrete circular Gaussian distribution, covering a minimum positive period, allowing the model to predict distribution parameters instead of directly regressing or classifying angles. The experimental results on multiple public datasets show that the proposed method is effective and achieves a superior performance.

In order to improve the performance of underwater object detection, an improved model YOLOv5-CFDSDSE based on YOLOv5s is proposed (Contribution 7). This model combines C3 and FasterNet structures to improve model accuracy, reduce the number of parameters, and improve scale perception, spatial perception, and task perception, enhancing the model's detection ability for multi scale and multi class objects. In addition, the small object detection (SD) layer has been optimized to effectively combine the feature information of different scales, retain more detailed information, and enhance the detection ability of small objects. Numerous experiments have shown that YOLOv5-CFDSDSE achieves an excellent performance in underwater object detection.

A database specially designed for autonomous vehicle was introduced (Contribution 8). Based on the KITTI visual benchmark suite, the proposed database contains 7481 scenes,

each of which has a corresponding top-view image, label file, and segmentation map. These segmentation maps are designed specifically for environmental perception tasks of autonomous vehicles, including road, background, car, truck, and truck segmentation. Detailed methods for creating databases were introduced, including explanations of how to convert 3D point clouds into 2D top-view images, how to generate detection labels on a plane, and how to implement neural networks to generate segmentation maps for each scene. The road segmentation performance indicators in the database include an F1 score of 95.77, average accuracy (AP) of 92.54, accuracy (ACC) of 97.53, accuracy (PRE) of 94.34, and recall rate (REC) of 97.25. These indicators demonstrate the superior performance of the database in road segmentation tasks. This study also included a review of research related to top-view segmentation of autonomous vehicles, including semantic segmentation using LiDAR sensor data, as well as existing research databases on autonomous vehicles.

In order to accurately detect and identify different types of aircraft in high-resolution remote sensing images, a novel algorithm Fine-grained S^2ANet (FS^2ANet) is proposed (Contribution 9). It is based on the improved Single-shot Alignment Network (S^2ANet) and uses data augmentation techniques to address the imbalanced number of aircraft instances in the dataset. This model uses ResNet101 as the backbone network and combines the space-to-depth (SPD) module to improve the feature pyramid network (FPN) structure, constructing the FPN-SPD module. An aircraft fine feature focusing module (AF^3M) was constructed at the detection head of the network to reduce the loss of fine-grained information during feature extraction and improve the network's ability to extract small aircraft features. Finally, the challenges that need to be addressed in future work were pointed out, including applications in different remote sensing scenarios, as well as designing algorithms with higher-resolution inputs and faster speeds.

A frame segmentation method based on neural network training is proposed for detecting abnormal eye-blinking events in female drivers in Saudi Arabia (Contribution 10). This method segments the eye area through variational pixel extraction and identifies different frames using texture changes. The model adopts a two-layer convolutional neural network. The first layer recognizes the differences between continuous and discrete frames, while the second layer is responsible for segmenting the eye area based on texture changes. Finally, experimental analysis was conducted using the Niqab dataset to verify that the proposed method improves the accuracy of eye-blink detection and reduces detection time and errors.

To address the challenge of small sample underwater noise data, a typical underwater noise generation model (UNGM) was established (Contribution 11) and increased noise samples by generating noise with specific kurtosis, simulating the spectral and statistical characteristics of actual noise. In addition, underwater noise classification models based on UNGM and convolutional neural networks (CNN) have been developed. UNGM-CNN effectively solves the problem of small sample underwater noise classification and demonstrates the potential application of combining physical principles and machine learning techniques in the field of marine acoustics.

A multi-channel multi-location attention mechanism is proposed for a multi-modal phased transfer side-scan sonar target recognition model (Contribution 12). At the model level, different stages of the backbone network are trained using SAR and ImageNet datasets to improve the adaptability to side scan sonar data. Additionally, an attention mechanism network (AMN) and recognition network are combined, using side-scan sonar data to learn network parameters and obtaining important features of the target through different channel attention factors at different locations. The experiment shows that the proposed model has better accuracy and robustness in side-scan sonar target recognition.

Funding: This work was supported in part by Key R&D program of China (No. 2022YFB4300700), in part by The Key R&D programs of Shaanxi Province (No. 2021ZDLGY02-06), in part by Qin Chuangyuan project (No. 2021QCYRC4-49), in part by Qinchuangyuan Scientist+Engineer (No.

2022KXJ-169), and in part by National Defense Science and Technology Key Laboratory Fund Project (No. 6142101210202).

Data Availability Statement: No new data were created or analyzed in this study. Data sharing is not applicable to this article.

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

List of Contributions

1. Yin, C.; Ye, Q.; Zhang, S.; Yang, Z. Detecting Logos for Indoor Environmental Perception Using Unsupervised and Few-Shot Learning. *Electronics* **2024**, *13*, 2246.
2. Wang, Y.; Dang, K.; Yang, R.; Li, L.; Li, H.; Gong, M. Multi-Objective Automatic Clustering Algorithm Based on Evolutionary Multi-Tasking Optimization. *Electronics* **2024**, *13*, 1987.
3. Shao, Y.; Wang, S.; Zhao, W. A Causality-Aware Perspective on Domain Generalization via Domain Intervention. *Electronics* **2024**, *13*, 1891.
4. Shang, Y.; Wang, Q.; Zhu, W.; Xie, F.; Wang, H.; Li, L. Evolutionary Competition Multitasking Optimization with Online Resource Allocation for Endmemeber Extraction of Hyperspectral Images. *Electronics* **2024**, *13*, 1424.
5. Li, L.; Liu, L.; He, Y.; Zhong, Z. USES-Net: An Infrared Dim and Small Target Detection Network with Embedded Knowledge Priors. *Electronics* **2024**, *13*, 1400.
6. Xu, H.; Liu, X.; Ma, Y.; Zhu, Z.; Wang, S.; Yan, C.; Dai, F. Rotated Object Detection with Circular Gaussian Distribution. *Electronics* **2023**, *12*, 3265.
7. Wang, J.; Li, Y.; Wang, J.; Li, Y. An Underwater Dense Small Object Detection Model Based on YOLOv5-CFDSDSE. *Electronics* **2023**, *12*, 3231.
8. Ortega-Gomez, J.I.; Morales-Hernandez, L.A.; Cruz-Albarran, I.A. A Specialized Database for Autonomous Vehicles Based on the KITTI Vision Benchmark. *Electronics* **2023**, *12*, 3165.
9. Guan, Q.; Liu, Y.; Chen, L.; Zhao, S.; Li, G. Aircraft Detection and Fine-Grained Recognition Based on High-Resolution Remote Sensing Images. *Electronics* **2023**, *12*, 3146.
10. Al-Razgan, M.S.; Alruwaly, I.; Ali, Y.A. Eye-Blink Event Detection Using a Neural-Network-Trained Frame Segment for Woman Drivers in Saudi Arabia. *Electronics* **2023**, *12*, 2699.
11. Song, G.; Guo, X.; Zhang, Q.; Li, J.; Ma, L. Underwater Noise Modeling and Its Application in Noise Classification with Small-Sized Samples. *Electronics* **2023**, *12*, 2669.
12. Wang, J.; Li, H.; Huo, G.; Li, C.; Wei, Y. Multi-Mode Channel Position Attention Fusion Side-Scan Sonar Transfer Recognition. *Electronics* **2023**, *12*, 791.

References

1. Zou, Z.; Chen, K.; Shi, Z.; Guo, Y.; Ye, J. Object detection in 20 years: A survey. *Proc. IEEE* **2023**, *111*, 257–276. [CrossRef]
2. 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]
3. Rawat, W.; Wang, Z. Deep convolutional neural networks for image classification: A comprehensive review. *Neural Comput.* **2017**, *29*, 2352–2449. [CrossRef] [PubMed]
4. Dhillon, A.; Verma, G.K. Convolutional neural network: A review of models, methodologies and applications to object detection. *Prog. Artif. Intell.* **2020**, *9*, 85–112. [CrossRef]
5. Viola, P.; Jones, M. Rapid object detection using a boosted cascade of simple features. In Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2001, Kauai, HI, USA, 8–14 December 2001; Volume 1, pp. 511–518.
6. Comaniciu, D.; Meer, P. Mean shift: A robust approach toward feature space analysis. *IEEE Trans. Pattern Anal. Mach. Intell.* **2002**, *24*, 603–619. [CrossRef]
7. Lowe, D.G. Distinctive image features from scale-invariant keypoints. *Int. J. Comput. Vis.* **2004**, *60*, 91–110. [CrossRef]
8. Xie, X.; Cheng, G.; Wang, J.; Yao, X.; Han, J. Oriented R-CNN for object detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 3520–3529.
9. Tan, M.; Pang, R.; Le, Q.V. Efficientdet: Scalable and efficient object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 10781–10790.
10. Joseph, K.; Khan, S.; Khan, F.S.; Balasubramanian, V.N. Towards open world object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 5830–5840.
11. Kirillov, A.; Mintun, E.; Ravi, N.; Mao, H.; Rolland, C.; Gustafson, L.; Xiao, T.; Whitehead, S.; Berg, A.C.; Lo, W.Y.; et al. Segment anything. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Paris, France, 2–6 October 2023; pp. 4015–4026.

12. Zou, X.; Yang, J.; Zhang, H.; Li, F.; Li, L.; Wang, J.; Wang, L.; Gao, J.; Lee, Y.J. Segment everything everywhere all at once. In Proceedings of the International Conference on Neural Information Processing Systems, New Orleans, LA, USA, 10–16 December 2023; pp. 19769–19782.
13. Li, Y.; Wu, C.Y.; Fan, H.; Mangalam, K.; Xiong, B.; Malik, J.; Feichtenhofer, C. Mvitv2: Improved multiscale vision transformers for classification and detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 4804–4814.
14. Ibrahim Daradkeh, Y.; Gorokhovatskyi, V.; Tvoroshenko, I.; Al-Dhaifallah, M. Classification of Images Based on a System of Hierarchical Features. *Comput. Mater. Contin.* **2002**, *72*, 1785–1797. [CrossRef]
15. Zhao, Y.; Lv, W.; Xu, S.; Wei, J.; Wang, G.; Dang, Q.; Liu, Y.; Chen, J. Detrs beat yolos on real-time object detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 17–21 June 2024; pp. 16965–16974.
16. Li, H.; Wan, F.; Gong, M.; Qin, A.; Wu, Y.; Xing, L. Privacy-enhanced multitasking particle swarm optimization based on homomorphic encryption. *IEEE Trans. Evol. Comput.* **2023**. [CrossRef]

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.