applied sciences

MDPI

Editorial

Closing Editorial for Computer Vision and Pattern Recognition Based on Deep Learning

Hui Yuan

School of Control Science and Engineering, Shandong University, No. 17923 Jingshi Road, Jinan 250061, China; huiyuan@sdu.edu.cn

1. Introduction

Deep learning has demonstrated unparalleled performance in various industries. Artificial intelligence technologies centered on deep learning are also revolutionizing people's ways of production and life. Computer vision and pattern recognition are some of the most widely used applications of deep learning, and they are also the focus of attention in artificial intelligence technology. This Special Issue focuses on the theory and methods of computer vision and pattern recognition based on deep learning, with a total of 31 papers accepted, including 6 papers related to image and video processing, 8 papers related to object detection, 8 papers related to object and scene recognition, and 4 papers related to visual application technologies, as well as studies on classification, segmentation, compression, and other related topics.

2. Image and Video Processing

Peng et al. [1] proposed GP-Net, combining Transformer and CNN in parallel, to address the challenge of image manipulation localization by efficiently building global context and capturing low-level details, with an effective fusion module named TcFusion, outperforming existing methods in manipulation detection and localization.

Yi et al. [2] introduced a novel cycle generative adversarial network (CGAN) method with gradient normalization for generating high-quality infrared images from visible images. By employing a residual network in the generator and integrating channel and spatial attention mechanisms, the method enhances feature perception and detail generation in infrared images. Furthermore, the introduction of gradient normalization in the discriminator stabilizes the training process, reducing model collapse.

Yang et al. [3] proposed a model designed to address model collapse in generative adversarial networks (GANs). Inspired by the relationship between Hessian and Jacobian matrices, the proposed framework focuses on disentanglement and mitigating model collapse.

Chéles et al. [4] developed an image processing protocol using well-established techniques to segment the image of blastocysts and extract variables of interest. A total of 33 variables were automatically generated by digital image processing, each representing a different aspect of the embryo and describing a different characteristic of the blastocyst.

Ma et al. [5] simplified the existing Siamese convolutional network by reducing the number of network parameters and proposing an efficient CNN-based structure, namely, adaptive deconvolution-based disparity matching net (ADSM net), by adding deconvolution layers to learn how to enlarge the size of input feature map for the following convolution layers.

Li et al. [6] addressed the challenge of predicting video memorability by analyzing and experimentally verifying the most impactful factors influencing memorability. The proposed Adaptive Multi-modal Ensemble Network framework integrates temporal 3D information, spatial information, and semantics derived from video, image, and caption. It utilizes three individual base learners (ResNet3D, Deep Random Forest, and Multi-Layer Perception) within a weighted ensemble framework to predict video memorability.

Citation: Yuan, H. Closing Editorial for Computer Vision and Pattern Recognition Based on Deep Learning. *Appl. Sci.* **2024**, *14*, 3660. https://doi.org/10.3390/app14093660

Received: 12 March 2024
Accepted: 15 April 2024
Published: 25 April 2024

3. Object Detection

Kang [7] proposed SSDLiteX, a modified version of SSDLite that enhances small object detection by prioritizing higher-resolution feature maps and increasing the number of layers in the base CNN, resulting in a 1.5 percent point improvement in average precision (AP) for small objects in the MS COCO dataset.

Florez et al. [8] proposed a drowsiness detection method focusing on the eye region using Mediapipe for extraction, and evaluated three deep learning networks—InceptionV3, VGG16, and ResNet50V2—on the NITYMED dataset. The results indicate high accuracy across all networks.

Zhao et al. [9] introduced YOLOv5s-Z, a lightweight tennis ball detection algorithm. It leverages a G-Backbone and G-Neck network to reduce parameters and computations, integrates convolutional coordinate attention for location information enhancement, employs a modified concat module for efficient feature fusion, introduces EIOU Loss and Focal-EIOU Loss to handle aspect ratio imbalance, and incorporates Meta-ACON activation for improved accuracy.

Sun et al. [10] proposed Auto-T-YOLO, a detection network consisting of three stages: preattention, attention, and prediction. The experimental results verify the practicality, validity, and robustness of the proposed model.

Zhou et al. [11] proposed a kitchen standard dress detection method, leveraging the YOLOv5s embedded model to swiftly and accurately identify whether a chef is wearing a hat and a mask. By constructing a comprehensive kitchen scene dataset and introducing images of chefs wearing masks and hats, the method mitigates the reliability issue associated with single-object detection. Additionally, the integration of YOLOv5 and DeepStream SDK on Jetson Xavier NX facilitates the real-time detection and early warning of non-standard dress in kitchen environments.

Huang et al. [12] proposed a novel insulator defect detection algorithm based on YOLOv5 to enhance accuracy and speed. It addresses challenges such as background interference and small fault areas by constructing a lightweight backbone network, increasing the detection layer for small targets, and designing a receptive field module.

Lu et al. [13] introduced a multi-channel MSER (Maximally Stable Extreme Regions) method and an enhanced Feature Pyramid Network (FPN) for street sign text detection in complex backgrounds. The multi-channel MSER method effectively reduces background and light interference by leveraging color information. The enhanced FPN incorporates a Feature Pyramid Route Enhancement (FPRE) module and a High-Level Feature Enhancement (HLFE) module to exploit low-level and high-level semantic information, improving text localization and detection across various shapes, sizes, and orientations.

Crespo et al. [14] introduced a compound convolutional neural network (CNN) architecture for face mask detection. The proposed architecture combines two computer vision tasks: object localization to detect faces using RetinaFace, followed by image classification using ResNet-18 architecture to categorize face mask usage as correct, incorrect, or absent. Furthermore, the authors have released the dataset used for model training and the computer vision pipeline to the public, optimized for deployment on embedded systems.

4. Recognition

Monteiro et al. [15] proposed an innovative framework merging advanced Optical Character Recognition (OCR) and object detection (OD) technologies for automating visual inspection processes in industrial settings. The system enhances industrial workflows by extracting supplementary information such as barcodes and QR codes, thus reducing manual labor demands and revolutionizing industrial processes.

Li et al. [16] proposed a novel approach for accurate temporal modeling in action recognition, employing a Local Spatiotemporal Extraction module (LSTE) and a Channel Time Excitation module (CTE) to capture temporal information in video sequences. Experimental results on the Something-Something V1 and V2 datasets demonstrate the effectiveness of the proposed approach.

Khosrobeigi et al. [17] proposed Bina, a specialized Optical Character Recognition (OCR) framework tailored for Persian text, addressing challenges such as character continuity, semicircles, dots, and oblique characters. Bina employs Convolutional Neural Network (CNN) and deep bidirectional Long-Short Term Memory (BLSTM) networks to handle the complexities of Persian text.

Li et al. [18] introduced a fast gunshot-type recognition method based on knowledge distillation to address the challenges of large model size and insufficient real-time detection in urban combat scenarios. It preprocesses muzzle blast and shock wave signals, enhancing dataset quality with corresponding Log-Mel spectra. A teacher network, composed of 10 two-dimensional residual modules, and a student network, employing depth-wise separable convolution, are constructed.

Gao et al. [19] introduced a variable rate IndRNN network to address the challenge of different sampling rates in skeleton-based action recognition, leveraging the well-behaved gradient backpropagation through time of IndRNN by processing samples with variable lengths and time steps, and implementing a learning rate adjustment method based on gradient behavior.

Gao et al. [20] introduced DA-IndRNN, a novel attention-based deep learning model for skeleton-based action recognition, integrating a deep IndRNN for feature extraction and a multi-layered attention network for reliable attention weight estimation, trained with a new triplet loss function to guide attention learning across different action categories.

Niu and Li [21] introduced a traffic light detection and recognition method leveraging YOLOv5s for target detection and AlexNet for image classification, addressing issues of lower accuracy and limited detection types. It enhances recognition rates for small targets and optimizes datasets using the ZeroDCE low-light enhancement algorithm.

Guo et al. [22] proposed a method for improving the accuracy of environmental sound recognition by employing multi-feature parameters and a time–frequency attention module. The approach begins with pretreatment using multi-feature parameters to extract sound, enhancing input feature expressiveness by supplementing lost phase information from the Log-Mel spectrogram. A time–frequency attention module with multiple convolutions is then employed to extract attention weights from the input feature spectrogram, reducing interference from background noise and irrelevant frequency bands.

5. Application

Niu et al. [23] developed a method integrating improved YOLOv5s target detection with Anylogic emergency evacuation simulation to enhance the efficiency of emergency evacuations in field stations. It leverages YOLOv5s with an SE attention mechanism to detect pedestrians and determines their locations based on head detection, considering crowded conditions. Anylogic employs closest distance evacuation principles to guide pedestrians to the nearest exit.

Suarez Baron et al. [24] present the application of supervised learning and image classification for the early detection of late blight disease in potato using convolutional neural network and support vector machine (SVM). Initially, a dataset of crop images is collected and pre-processed to extract disease characteristics. Subsequently, classification models are trained and evaluated using various performance metrics to identify healthy and infected potato plants.

Bi and Li et al. [25] addressed the challenge of separating overlapped space-based ADS-B signals, crucial for air traffic control, by proposing a deep learning approach using a Multi-Scale Conv-TasNet (MConv-TasNet). Specifically, a Multi-Scale Convolutional Separation (MCS) network is introduced to fuse temporal features from overlapping ADS-B signals, enabling accurate signal separation.

Su et al. [26] explored the application of deep learning in cooperative vehicle-infrastructure systems (CVISs) for video applications, highlighting its ability to handle high-dimensional datasets and improve performance compared to traditional approaches.

6. Compression

Chu et al. [27] proposed RBENN, a Quality Enhancement Network dedicated to improving reference blocks in inter prediction for hybrid video coding frameworks. It operates on luma reference blocks pre-motion compensation, enhancing them for better coding efficiency.

Wang et al. [28] proposed an optimization strategy for video codecs in cloud gaming by incorporating deep learning networks. Specifically, it introduces CMGNet, a camera motion-guided network, for enhancing reference frames in cloud gaming videos, thereby reducing bitrate consumption. CMGNet utilizes camera motion information to improve frame alignment and fusion, enhancing the quality of reference frames for better compression efficiency.

Hou et al. [29] introduce an end-to-end learning-based approach for compressing point cloud attributes (PCACs) to reduce transmission and storage costs. It utilizes a sparse convolution-based variational autoencoder (VAE) structure and incorporates an attention mechanism, specifically a non-local attention module, to capture local and global correlations in both spatial and channel dimensions. Additionally, a modulation network enables variable rate compression within a single network, eliminating the need to store multiple networks for different bitrates.

7. Segmentation

Jung et al. [30] were the first to evaluate and compare the performance of state-of-the-art instance segmentation models by focusing on their inference time in a fixed experimental environment. They proposed the accuracy and speed of the models in a fixed hardware environment for quantitative and qualitative analyses.

Shi and Zuo [31] proposed an enhanced MaskRCNN framework for the semantic segmentation of satellite images, specifically targeting the detection and segmentation of shadow cumulus clouds, which play a crucial role in environmental and climate analysis. The proposed approach incorporates two deep neural network architectures leveraging auxiliary loss and feature fusion functions, aimed at improving segmentation accuracy.

8. Classification

Zhang et al. [32] proposed an Edge Continuity Distortion-Aware Block (ECDAB) to mitigate discontinuities and distortion at image edges, along with a Convolutional Row-Column Attention Block (CRCAB) to capture global dependencies for stronger feature representation, to address the challenges of processing 360° omnidirectional images. Additionally, an Improved CRCAB (ICRCAB) reduces memory overhead by adjusting row–column vector numbers.

Mafukidze et al. [33] introduced a novel approach to maize leaf disease quantification. The proposed system leverages deep learning models for disease classification and extracts regions of interest using an adaptive thresholding technique from class activation maps, without prior knowledge.

Conflicts of Interest: The author declare no conflict of interest.

References

1. Peng, J.; Liu, C.; Pang, H.; Gao, X.; Cheng, G.; Hao, B. GP-Net: Image Manipulation Detection and Localization via Long-Range Modeling and Transformers. *Appl. Sci.* **2023**, *13*, 12053. [CrossRef]
2. Yi, X.; Pan, H.; Zhao, H.; Liu, P.; Zhang, C.; Wang, J.; Wang, H. Cycle Generative Adversarial Network Based on Gradient Normalization for Infrared Image Generation. *Appl. Sci.* **2023**, *13*, 635. [CrossRef]
3. Yang, G.; Qu, Y.; Fang, X. Editable Image Generation with Consistent Unsupervised Disentanglement Based on GAN. *Appl. Sci.* **2022**, *12*, 5382. [CrossRef]
4. Chéles, D.S.; Ferreira, A.S.; de Jesus, I.S.; Fernandez, E.I.; Pinheiro, G.M.; Dal Molin, E.A.; Alves, W.; de Souza, R.C.M.; Bori, L.; Meseguer, M.; et al. An Image Processing Protocol to Extract Variables Predictive of Human Embryo Fitness for Assisted Reproduction. *Appl. Sci.* **2022**, *12*, 3531. [CrossRef]
5. Ma, X.; Zhang, Z.; Wang, D.; Luo, Y.; Yuan, H. Adaptive Deconvolution-Based Stereo Matching Net for Local Stereo Matching. *Appl. Sci.* **2022**, *12*, 2086. [CrossRef]

6. Li, J.; Guo, X.; Yue, F.; Xue, F.; Sun, J. Adaptive Multi-Modal Ensemble Network for Video Memorability Prediction. *Appl. Sci.* **2022**, *12*, 8599. [CrossRef]
7. Kang, H.-J. SSDLiteX: Enhancing SSDLite for Small Object Detection. *Appl. Sci.* **2023**, *13*, 12001. [CrossRef]
8. Florez, R.; Palomino-Quispe, F.; Coaquira-Castillo, R.J.; Herrera-Levano, J.C.; Paixão, T.; Alvarez, A.B. A CNN-Based Approach for Driver Drowsiness Detection by Real-Time Eye State Identification. *Appl. Sci.* **2023**, *13*, 7849. [CrossRef]
9. Zhao, Y.; Lu, L.; Yang, W.; Li, Q.; Zhang, X. Lightweight Tennis Ball Detection Algorithm Based on Robomaster EP. *Appl. Sci.* **2023**, *13*, 3461. [CrossRef]
10. Sun, B.; Wang, X.; Oad, A.; Pervez, A.; Dong, F. Automatic Ship Object Detection Model Based on YOLOv4 with Transformer Mechanism in Remote Sensing Images. *Appl. Sci.* **2023**, *13*, 2488. [CrossRef]
11. Zhou, Z.; Zhou, C.; Pan, A.; Zhang, F.; Dong, C.; Liu, X.; Zhai, X.; Wang, H. A Kitchen Standard Dress Detection Method Based on the YOLOv5s Embedded Model. *Appl. Sci.* **2023**, *13*, 2213. [CrossRef]
12. Huang, Y.; Jiang, L.; Han, T.; Xu, S.; Liu, Y.; Fu, J. High-Accuracy Insulator Defect Detection for Overhead Transmission Lines Based on Improved YOLOv5. *Appl. Sci.* **2022**, *12*, 12682. [CrossRef]
13. Lu, M.; Leng, Y.; Chen, C.-L.; Tang, Q. An Improved Differentiable Binarization Network for Natural Scene Street Sign Text Detection. *Appl. Sci.* **2022**, *12*, 12120. [CrossRef]
14. Crespo, F.; Crespo, A.; Sierra-Martínez, L.M.; Peluffo-Ordóñez, D.H.; Morocho-Cayamcela, M.E. A Computer Vision Model to Identify the Incorrect Use of Face Masks for COVID-19 Awareness. *Appl. Sci.* **2022**, *12*, 6924. [CrossRef]
15. Monteiro, G.; Camelo, L.; Aquino, G.; Fernandes, R.d.A.; Gomes, R.; Printes, A.; Torné, I.; Silva, H.; Oliveira, J.; Figueiredo, C. A Comprehensive Framework for Industrial Sticker Information Recognition Using Advanced OCR and Object Detection Techniques. *Appl. Sci.* **2023**, *13*, 7320. [CrossRef]
16. Li, S.; Wang, X.; Shan, D.; Zhang, P. Action Recognition Network Based on Local Spatiotemporal Features and Global Temporal Excitation. *Appl. Sci.* **2023**, *13*, 6811. [CrossRef]
17. Khosrobeygi, Z.; Veisi, H.; Hoseinzade, E.; Shabanian, H. Persian Optical Character Recognition Using Deep Bidirectional Long Short-Term Memory. *Appl. Sci.* **2022**, *12*, 11760. [CrossRef]
18. Li, J.; Guo, J.; Sun, X.; Li, C.; Meng, L. A Fast Identification Method of Gunshot Types Based on Knowledge Distillation. *Appl. Sci.* **2022**, *12*, 5526. [CrossRef]
19. Gao, Y.; Li, C.; Li, S.; Cai, X.; Ye, M.; Yuan, H. Variable Rate Independently Recurrent Neural Network (IndRNN) for Action Recognition. *Appl. Sci.* **2022**, *12*, 3281. [CrossRef]
20. Gao, Y.; Li, C.; Li, S.; Cai, X.; Ye, M.; Yuan, H. A Deep Attention Model for Action Recognition from Skeleton Data. *Appl. Sci.* **2022**, *12*, 2006. [CrossRef]
21. Niu, C.; Li, K. Traffic Light Detection and Recognition Method Based on YOLOv5s and AlexNet. *Appl. Sci.* **2022**, *12*, 10808. [CrossRef]
22. Guo, J.; Li, C.; Sun, Z.; Li, J.; Wang, P. A Deep Attention Model for Environmental Sound Classification from Multi-Feature Data. *Appl. Sci.* **2022**, *12*, 5988. [CrossRef]
23. Niu, C.; Wang, W.; Guo, H.; Li, K. Emergency Evacuation Simulation Study Based on Improved YOLOv5s and Anylogic. *Appl. Sci.* **2023**, *13*, 5812. [CrossRef]
24. Suarez Baron, M.J.; Gomez, A.L.; Diaz, J.E.E. Supervised Learning-Based Image Classification for the Detection of Late Blight in Potato Crops. *Appl. Sci.* **2022**, *12*, 9371. [CrossRef]
25. Bi, Y.; Li, C. Multi-Scale Convolutional Network for Space-Based ADS-B Signal Separation with Single Antenna. *Appl. Sci.* **2022**, *12*, 8816. [CrossRef]
26. Su, B.; Ju, Y.; Dai, L. Deep Learning for Video Application in Cooperative Vehicle-Infrastructure System: A Comprehensive Survey. *Appl. Sci.* **2022**, *12*, 6283. [CrossRef]
27. Chu, Y.; Hui, Y.; Jiang, S.; Fu, C. Neural Network-Based Reference Block Quality Enhancement for Motion Compensation Prediction. *Appl. Sci.* **2023**, *13*, 2795. [CrossRef]
28. Wang, Y.; Wang, H.; Wang, K.; Zhang, W. Cloud Gaming Video Coding Optimization Based on Camera Motion-Guided Reference Frame Enhancement. *Appl. Sci.* **2022**, *12*, 8504. [CrossRef]
29. Huo, X.; Zhang, S.; Yang, F. Variable Rate Point Cloud Attribute Compression with Non-Local Attention Optimization. *Appl. Sci.* **2022**, *12*, 8179. [CrossRef]
30. Jung, S.; Heo, H.; Park, S.; Jung, S.-U.; Lee, K. Benchmarking Deep Learning Models for Instance Segmentation. *Appl. Sci.* **2022**, *12*, 8856. [CrossRef]
31. Shi, G.; Zuo, B. CloudRCNN: A Framework Based on Deep Neural Networks for Semantic Segmentation of Satellite Cloud Images. *Appl. Sci.* **2022**, *12*, 5370. [CrossRef]
32. Zhang, X.; Yang, D.; Song, T.; Ye, Y.; Zhou, J.; Song, Y. Classification and Object Detection of 360° Omnidirectional Images Based on Continuity-Distortion Processing and Attention Mechanism. *Appl. Sci.* **2022**, *12*, 12398. [CrossRef]
33. Mafukidze, H.D.; Owugumisha, G.; Otim, D.; Nechibvute, A.; Nyamhere, C.; Mazunga, F. Adaptive Thresholding of CNN Features for Maize Leaf Disease Classification and Severity Estimation. *Appl. Sci.* **2022**, *12*, 8412. [CrossRef]