



Editorial

# Developments of Computer Vision and Image Processing: Methodologies and Applications

Manuel J. C. S. Reis <sup>1,2</sup>

<sup>1</sup> Engineering Department, University of Trás-os-Montes e Alto Douro, Room E2.10, Quinta de Prados, 5001-801 Vila Real, Portugal; mcabral@utad.pt

<sup>2</sup> IEETA, 3810-193 Aveiro, Portugal

The rapid advancement of technology has enabled a vast and ever-growing number of computer applications in real scenarios of our daily life.

This Special Issue, titled “Developments of Computer Vision and Image Processing: Methodologies and Applications”, aimed to highlight the recent advances in the development of methodologies, algorithms, techniques and applications in the field of Computer Vision and Image Processing. Exploratory, experimental and theoretical results were considered, as well as review papers. The processing of the data, the used methodologies and algorithms, and the consequent extraction of useful information were also included in the topics of this Special Issue. There were no restrictions on the length of the papers.

The number of submitted manuscripts directly reflects the high interest of this topic to the research community: a total of fifteen manuscripts have been submitted; six high-quality papers were accepted/published. An additional interesting note is the fact that all of the published papers are directly related to computer vision and image processing methodologies and applications, and not indirectly related to some specific marginal topic. As usual, the Future Internet journal standards conducted all the submitted manuscripts through a rigorous peer-review process.

In the following presentation, I will use the authors’ own words, so that no distortions or misinterpretations are introduced and a better presentation of the contributions of each paper is conceived.

The work by Jian et al. [1] was the first published paper in this Special Issue. Because light rays are bent by unknown amounts, leading to complex geometric distortions, imaging through a wavy water–air interface (WAI) is challenging. Jian et al. presented an image recovery model via structured light projection, considering the restoration of instantaneous distorted images. The method is composed of two separate parts: in the first part, an algorithm for the determination of the instantaneous shape of the water surface via structured light projection is used; in the second part, they synchronously recover the distorted airborne scene image through reverse ray tracing. The experimental results show that, compared to the state-of-the-art methods, the proposed method not only can overcome the influence of changes in natural illumination conditions for WAI reconstruction, but also can significantly reduce the distortion and achieve better performance.

To achieve an understanding of the emotional changes of a specific target group, Facial Expression Recognition (FER) can be used. Two of the main challenges that researchers face are the relatively small dataset related to facial expression recognition and the lack of high accuracy of expression recognition. Recently, more and more convolutional neural networks have been developed to be used in FER research. However, mostly due to expression-independent intra-class differences, many of the convolutional neural performances are not good enough when dealing with the problems of overfitting from too-small datasets and noise. In [2] the authors propose a Dual Path Stacked Attention Network (DPSAN) to better handle the above challenges. In a first step of the developed methodology, the features of key regions in faces are extracted using segmentation, and irrelevant regions are ignored, which effectively suppresses intra-class differences. Then, in a second step, the overfitting



**Citation:** Reis, M.J.C.S. Developments of Computer Vision and Image Processing: Methodologies and Applications. *Future Internet* **2023**, *15*, 233. <https://doi.org/10.3390/fi15070233>

Received: 26 June 2023  
Accepted: 27 June 2023  
Published: 30 June 2023



**Copyright:** © 2023 by the author. 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/>).

problem of the deep network due to a lack of data can be effectively mitigated, by providing the global image and segmented local image regions as training data for the integrated dual path model. Finally, in the third step, the authors designed a stacked attention module to weight the fused feature maps, according to the importance of each part for expression recognition. The authors adopted a cropping method based on the fixed four regions of the face image, to segment out the key image regions and to ignore the irrelevant regions, so as to improve the efficiency of the algorithm computation. The experimental results on two public datasets, CK+ and FERPLUS, demonstrate the effectiveness of the presented methodology; the accuracy reached the level of current state-of-the-art methods on both CK+ and FERPLUS, with 93.2% and 87.63% accuracy on the CK+ dataset and FERPLUS dataset, respectively.

In ref. [3], Yan, Wang and Tan presented the You Only Look Once (YOLO) Dependency Fusing Attention Network (DFAN) detection algorithm, improved based on the lightweight network YOLOv4-tiny. This approach combines the advantages of fast speed of the traditional lightweight networks and the high precision of traditional heavyweight networks, so that it is very suitable for real-time detection of high-altitude safety belts in embedded equipment. In response to the problem of extracting the features of an object with a low effective pixel ratio (which is an object with a low ratio of actual area to detection anchor area in the YOLOv4-tiny network) the authors made three major improvements to the baseline network. First, they introduced the atrous spatial pyramid pooling network after the extraction step of the lightweight backbone network CSPDarkNet-tiny. Second, they proposed the DFAN itself. Third, they introduced the Path Aggregation Network (PANet) to replace the Feature Pyramid Network (FPN) of the original network and fuse it with the DFAN. According to the experimental results in the high-altitude safety belt dataset, YOLO-DFAN improved the accuracy by 5.13%, when compared to the original network, and its detection speed met the real-time demand. The algorithm also exhibits a good improvement on the Pascal voc07+12 dataset.

Object detection can be view as a computer vision task for detecting instances of objects of a certain class, identifying types of objects, determining their location, and accurately labeling them in an input image or a video. In ref. [4], Kalgaonkar and El-Sharkawy proposed an object detection network, called NextDet, to efficiently detect objects of multiple classes, which utilizes CondenseNeXt. This lightweight image classification convolutional neural network algorithm with a reduced number of FLOPs and parameters as the backbone can be used to efficiently extract and aggregate image features at different granularities in addition to other novel and modified strategies, such as attentive feature aggregation in the head, to perform object detection and draw bounding boxes around the detected objects. Extensive experiments and ablation tests were performed on the Argoverse-HD and COCO datasets, which provide numerous temporarily sparse to dense annotated images, demonstrating that the proposed object detection algorithm with CondenseNeXt as the backbone resulted in an increase in mean Average Precision (mAP) performance and interpretability on Argoverse-HD's monocular ego-vehicle camera captured scenarios by up to 17.39% as well as COCO's large set of images of everyday scenes of real-world common objects by up to 14.62%.

Human pose estimation (HPE) has become a prevalent research topic in computer vision for many years, and the areas of application include video surveillance, medical assistance, and sport motion analysis. Due to the ever-growing demand for HPE, many libraries have been developed in the last two decades, and the number of skeleton-based HPE algorithms that have been developed and packaged into libraries to provide ease of use for researchers has also seen a huge increment. The performance of these libraries is very important when researchers intend to integrate them into real-world applications, such as video surveillance, medical assistance, and sport motion analysis. In ref. [5], the authors investigate the strengths and weaknesses of four popular state-of-the-art skeleton-based HPE libraries for human pose detection, including OpenPose, PoseNet, MoveNet, and MediaPipe Pose. They also present a comparative analysis of these libraries based on

images and videos. The percentage of detected joints (PDJ) was used as the evaluation metric in all comparative experiments to reveal the performance of the HPE libraries. They concluded that MoveNet has the best performance for detecting different human poses in static images and videos.

Finally, as posed by Abesinghe, Kankanamge, Yigitcanlar, and Pancholi's own-words [6], "The image of a city represents the sum of beliefs, ideas, and impressions that people have of that city". In fact, most of the city images are assessed through direct or indirect interviews and cognitive mapping exercises. However, such methods consume more time and effort and are limited to a small number of samples (people). Yet, more recently, people tend to use social media to express their thoughts and experiences of a place. As a consequence, Abesinghe, Kankanamge, Yigitcanlar, and Pancholi attempt to explore city images through social media big data, considering Colombo, Sri Lanka, as the testbed. The aim of the study was to examine the image of a city through Lynchian elements (i.e., landmarks, paths, nodes, edges, and districts), by using community sentiments expressed and images posted on social media platforms. They have used descriptive, image processing, sentiment, popularity, and geo-coded social media analyses. The study findings revealed that: (a) the community sentiments toward the same landmarks, paths, nodes, edges, and districts change over time; (b) decisions related to locating landmarks, paths, nodes, edges, and districts have a significant impact on community cognition in perceiving cities; and (c) geo-coded social media data analytics is an invaluable approach to capture the image of a city.

As a final note, I would personally like to thank all the authors and reviewers contributing to this Special Issue: the former for their original ideas and solutions, and the latter for their time and free precious improvement suggestions. Their excellent work has allowed the Future Internet journal to present novel and interesting contributions in the "Developments of Computer Vision and Image Processing: Methodologies and Applications" field. Thank you very much to all of you!

**Conflicts of Interest:** The author declares no conflict of interest.

## References

1. Jian, B.; Ma, C.; Zhu, D.; Sun, Y.; Ao, J. Seeing through Wavy Water-Air Interface: A Restoration Model for Instantaneous Images Distorted by Surface Waves. *Future Internet* **2022**, *14*, 236. [[CrossRef](#)]
2. Zhu, H.; Xu, H.; Ma, X.; Bian, M. Facial Expression Recognition Using Dual Path Feature Fusion and Stacked Attention. *Future Internet* **2022**, *14*, 258. [[CrossRef](#)]
3. Yan, W.; Wang, X.; Tan, S. YOLO-DFAN: Effective High-Altitude Safety Belt Detection Network. *Future Internet* **2022**, *14*, 349. [[CrossRef](#)]
4. Kalgaonkar, P.; El-Sharkawy, M. NextDet: Efficient Sparse-to-Dense Object Detection with Attentive Feature Aggregation. *Future Internet* **2022**, *14*, 355. [[CrossRef](#)]
5. Chung, J.L.; Ong, L.Y.; Leow, M.C. Comparative Analysis of Skeleton-Based Human Pose Estimation. *Future Internet* **2022**, *14*, 380. [[CrossRef](#)]
6. Abesinghe, S.; Kankanamge, N.; Yigitcanlar, T.; Pancholi, S. Image of a City through Big Data Analytics: Colombo from the Lens of Geo-Coded Social Media Data. *Future Internet* **2023**, *15*, 32. [[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.