

Article

Computer Vision Algorithms for 3D Object Recognition and Orientation: A Bibliometric Study

Youssef Yahia ^{*,†} , Júlio Castro Lopes [†]  and Rui Pedro Lopes [†] 

Research Center in Digitalization and Intelligent Robotics (CeDRI), Instituto Politécnico de Bragança, 5300-252 Bragança, Portugal; juliolopes@ipb.pt (J.C.L.); rlopes@ipb.pt (R.P.L.)

* Correspondence: youssefyahia@ipb.pt

† These authors contributed equally to this work.

Abstract: This paper consists of a bibliometric study that covers the topic of 3D object detection from 2022 until the present day. It employs various analysis approaches that shed light on the leading authors, affiliations, and countries within this research domain alongside the main themes of interest related to it. The findings revealed that China is the leading country in this domain given the fact that it is responsible for most of the scientific literature as well as being a host for the most productive universities and authors in terms of the number of publications. China is also responsible for initiating a significant number of collaborations with various nations around the world. The most basic theme related to this field is deep learning, along with autonomous driving, point cloud, robotics, and LiDAR. The work also includes an in-depth review that underlines some of the latest frameworks that took on various challenges regarding this topic, the improvement of object detection from point clouds, and training end-to-end fusion methods using both camera and LiDAR sensors, to name a few.

Keywords: 3D object; object detection; object orientation; bibliometric analysis



Citation: Yahia, Y.; Lopes, J.C.; Lopes, R.P. Computer Vision Algorithms for 3D Object Recognition and Orientation: A Bibliometric Study. *Electronics* **2023**, *12*, 4218. <https://doi.org/10.3390/electronics12204218>

Academic Editors: Junhua Ding, Hailhua Chen, Yunhe Feng and Tozammel Hossain

Received: 9 September 2023

Revised: 7 October 2023

Accepted: 9 October 2023

Published: 12 October 2023



Copyright: © 2023 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/>).

1. Introduction

Bibliometric analysis serves as a powerful tool to objectively assess the research landscape. By quantifying publication patterns, citation networks, and trends, bibliometric analysis provides valuable insights into the progression and impact of a scientific field. Considering this, a bibliometric analysis of the current state of the field of 3D object detection was performed. The findings of this study can offer numerous benefits to researchers, practitioners, and decision-makers in the field. By understanding the development of 3D object detection research, stakeholders can identify potential collaborators, prioritize research directions, and gain insights into the current trends. Therefore, this work aims to answer the following questions: which are the main active entities (authors, countries, affiliations) in this field? Which publications influenced the current advancement of the field? What are the most relevant themes related to this concept?

Various techniques and indicators were used to aid them through this process and assist them in presenting a meaningful answer to the aforementioned questions. For instance, hierarchical clustering was used to produce a Dendrogram that helped uncover complex patterns in the data. Additionally, the conducted thematic classification was based on both density and centrality to provide an understandable view of the current state of the main themes related to 3D object detection.

The paper is organized as follows: it commences with the present introduction followed by a brief summary of previous similar works in Section 2. Section 3 outlines the proposed methodology along with the employed tools and the different indicators that were used in the analysis. Sections 4 and 5 were reserved for the performed bibliometric analysis and the in-depth review of a selection of papers presenting a variety of works and

solutions related to the topic of discussion, respectively. Section 6 contains the discussion of the results. The paper ends with the conclusions drawn in Section 7.

2. State-of-the-Art

Bibliometric methods rely on bibliographic data extracted from online databases. According to Ellegaard and Wallin [1], computerized data treatment has significantly enhanced the effectiveness and utility of these methods. By employing these resources, bibliometric analysis enables a thorough examination and a holistic perspective of the scientific field under investigation [2]. In 2021, Lohia et al. [3] presented a study outlining the bibliometrics related to the topic of object detection. It was observed through this work that Chinese universities and researchers are the most active contributors to the domain of object detection. There are also several bibliometric studies concerning other topics that mention object detection in their findings. Chen and Deng [4] conducted a bibliometric analysis regarding the application of Convolutional Neural Networks (CNNs) in CV. Published in 2020, their study concluded that object detection and location are the hottest areas in relation to CNN research.

Liu et al. [5] produced a scientometric visualization analysis that considered image captioning as its main subject of investigation. They indicated that the clustering conducted while studying the development of keywords under the scope of image captioning from 2014 to 2020, included object detection as one of the main concepts. Another study conducted by Khan et al. [6] presented a bibliometric analysis of image processing in domains such as object detection and medical imaging. Their findings state that the interest in these domains has risen in a significant way between 2014 and 2019. Zhang [7] studied the literature on image engineering in China. The computed statistics of 2022 affirm that object detection, among a few other topics related to image engineering, has in fact been the focus of research in the last few years.

3. Methodology

The following section breaks down the science mapping workflow that was followed to conduct this bibliometric study. The workflow includes five meticulous phases: The first phase (1) is the study design, followed by the data collection phase (2) and the analysis phase (3). Phase (4) is reserved for visualizations and the workflow concludes with the interpretation phase (5).

The work started with the study design phase (1) during which the authors defined the research questions to be answered through this study. As a consequence, the Scopus indexer was chosen for retrieving the scientific publications metadata. According to Ref. [8], the Scopus database is a versatile multidisciplinary asset well-suited for researchers in the field of information systems. Based on the work produced by Chadegani et al. [9], it enables scholars to search and organize their search results based on various criteria, including but not limited to the first author, citation count, institution, and more. In addition to that, since its introduction in 2004 by Elsevier, Scopus has become a good alternative to The Web Of Science Database [10].

The next step of this phase entailed the selection of the main keywords to be inserted in the Scopus database. One way to use keywords, according to Chen and Xiao [11], is to employ relevant and significant keywords, allowing for an in-depth analysis of the primary research topics within a specific domain, as well as their interrelationships at a micro-level. The first keyword selected was “3D object”. This keyword was inserted in the Scopus search system and returned 83,317 results. Due to a large number of results, the search query was refined by adding two filters from the filters selection provided by Scopus: the first filter was used to limit the results to the years 2022 and 2023, while the second one was employed to obtain only contributions that are in English, as per the boundaries of this study. The number of publications drastically dropped to 8955 documents. The next step was to further narrow down the search by adding a second keyword, which is “object detection”. The number of remaining documents was 1457, which includes 736 conference

papers, 702 articles, 12 reviews, 6 book chapters, and 1 data paper. In order to widen the research perimeter, it was decided that all these publications will be included in the analysis.

The next phase (2) required the exportation of the .bib file from Scopus containing the results yielded by phase (1). The exportation was done using the exporting feature provided by the Scopus database. This file is essential for phase (3), data analysis. Phase (3) is divided into two steps: Step (a) is the descriptive data analysis, which requires the creation of a matrix containing all the documents using the Bibliometrix functions along with the analysis of authors, their most relevant affiliations, and the frequency of author's keywords. Step (b) consists of the extraction of the conceptual structure, the intellectual structure, and the social structure of the topic under investigation.

Additionally, there is a need for a deeper look into the various techniques and principles established within this sector. This in-depth analysis will later serve as a tool to help the authors of this paper to decide which techniques are the most advanced and interesting for future endeavors.

4. Bibliometric Analysis

Bibliometric analysis is a statistical process that aims to uncover patterns in scientific metadata. For that, specialized software, such as Bibliometrix and Biblioshiny, was used.

4.1. Descriptive Data Analysis

This section delves into the various aspects of the analysis. It starts by examining the authors, commencing with exploring the number of publications they have produced. Table 1 showcases the top 10 most relevant authors, determined by the total number of contributions made from 2022 up to the present year.

Table 1. The top 10 authors based on the number of contributions.

Authors	Number of Articles
LI Y	45
LI X	41
LIU Y	41
LI J	40
WANG Y	38
ZHANG Y	38
LI Z	35
CHEN Y	33
ZHANG X	32
LI H	30

LI Y is the author that has participated in the highest number of papers, with 45 articles, making him the most active author since 2022. Then, in the second rank, there is LIU Y and LI X with 41 articles each. The remaining authors in this table have all been involved in the publication of 30 to 40 works. Although the number of authors is relatively high, only a handful of researchers have participated in more than 10 publications. Table 2 illustrates the number and the ratio of the least productive authors. It states that the number of authors having published at least 1 but no more than 10 articles accounts for 98.7% out of 3739 authors.

Although it is interesting to know the number of publications made by authors, it does not necessarily highlight their impact on the field. In order to take a better look at the impact of authors, Hirsch [12] proposed the H-index, which represents the number of papers with citation numbers greater than H, as a valuable metric for characterizing a researcher's scientific output. Table 3 displays the top influential authors and their local H-index. LI H. and LI J. are in first position with a score of 7 each, meaning that the authors have published at least 7 papers that have each received at least 7 citations. This indicates

that these 2 authors' work has had a noticeable impact within their field of research, as their top 7 papers have been cited at least 7 times each by other scientific publications. Authors LI X and LI Y took the second position with a score of 6 each, which is also notable.

Table 2. Authors with 10 publications or less.

Number of Publications	Number of Authors	Percentage of Authors
10	11	0.003
9	19	0.005
8	18	0.005
7	24	0.006
6	47	0.012
5	46	0.012
4	80	0.021
3	173	0.046
2	436	0.115
1	2892	0.762

Table 3. Local impact of the authors based on their H-index.

Authors	Local H-Index
LI H	7
LI J	7
LI X	6
LI Y	6
CHEN Y	5
LI Z	5
LIU Y	5
WANG H	5
ZHANG Y	5
CHEN X	4

An analysis of the most relevant affiliations was also conducted to illustrate this and extract interesting statistical information. This analysis was based on the total number of articles produced by different universities to which the aforementioned authors belong to. According to Figure 1, Tsinghua University is by far the most productive institution, given the fact that it is responsible for the largest number of publications, which is 75. Tsinghua University produced almost twice as much work as the second-ranked Shanghai Jiao Tong University, which produced 38 papers. The rest of the displayed universities have been involved in at least 20 works. It is to be noted that the universities with the highest production are mostly Chinese. Figure 2 is a 3-field plot linking the 10 most relevant affiliations from the middle field to the countries where they are located on the right field along with the authors related to them on the left field.

In the next step of this bibliometric analysis, the most frequent words were analyzed, based on the author's keywords. These keywords are used by authors to specifically identify the main concepts covered by their papers. The process started by removing words such as "object(s) detection" and "3D object(s) detection" from the list of words, given the fact that they were used at the beginning of this study as the core research keywords on the Scopus database. The remaining words depicted very important strands. Analyzing these keywords unveiled various strands in the field of 3D object detection.

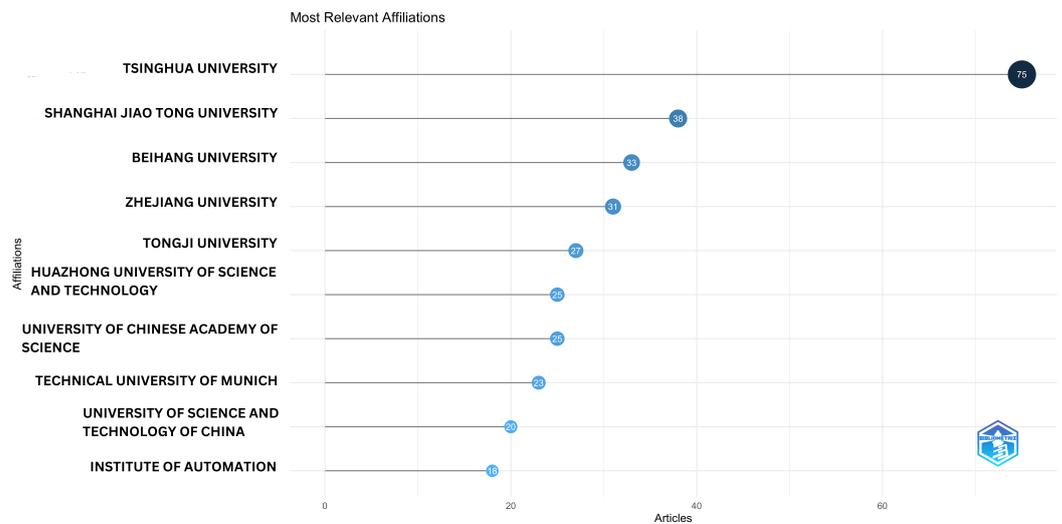


Figure 1. The most relevant affiliations.

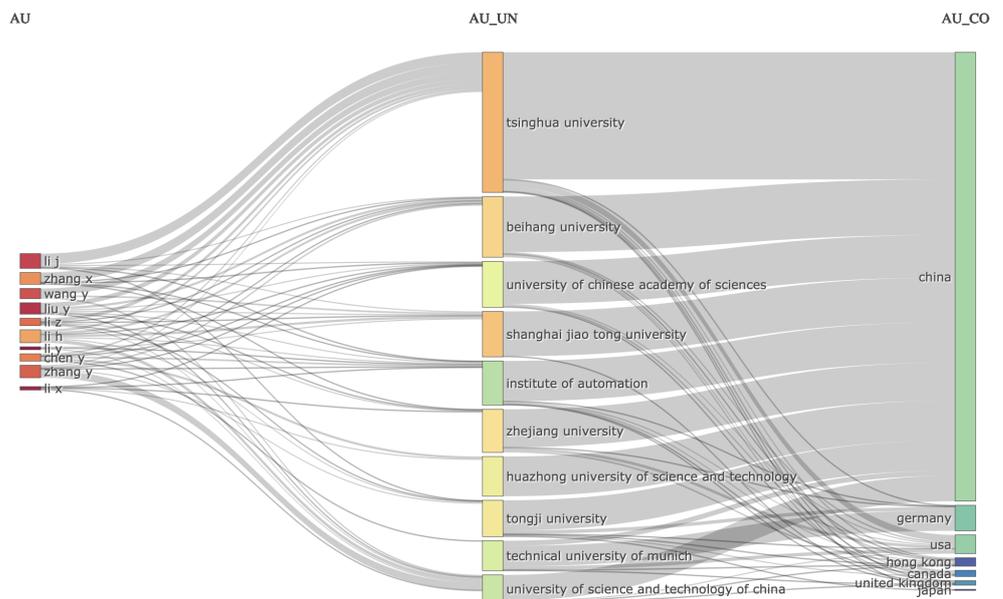


Figure 2. Three-field plot of affiliations and their corresponding countries and notable authors.

Figure 3 represents a word cloud that displays the keywords of 50 authors. Words with a significant size represent the most frequent words. Deep Learning (DL) is the biggest word on the cloud, which shows that it is the most common among researchers and indicates significant interest in its capabilities for 3D object detection, including the ability to identify objects in a given scene using deep networks. Some of the other words, such as point cloud, play a crucial role in the topic of object detection. Point clouds are generated by LiDAR sensors and provide an accurate representation of a given 3D space. Coupled with DL techniques, it can have unlimited applications in fields such as autonomous driving and robotics.

Table 4 contains the top 10 most frequent words used by authors and their respective number of occurrences. When it comes to 3D object detection, DL represents a key element of the topic. Autonomous driving, point cloud, and LiDAR are apparently some of the main manifestations of 3D object detection’s utility. Moreover, Figure 4 depicts a TreeMap containing possible combinations that represent 3D object detection.

4.2. Conceptual Structure

This section aims to identify the conceptual structure of the fields by leveraging the author keywords in a co-word analysis and identifying the underlying relationships between the keywords using clustering and exploring the themes' relevance.

In order to explore the underlying relationship between the author keywords, a hierarchical clustering was performed (Figure 5). Hierarchical clustering offers a valuable approach for detecting and analyzing the intricate structures within the data [13]. As displayed, there are two main blocks to be examined. The first block highlights the association between segmentation and CV algorithms related to visual reasoning, object recognition and understanding, and scene modeling. The second block has far more splits and therefore can be useful in identifying a wide range of areas of interest. The terms detection, recognition, and retrieval are clearly related to interesting subjects such as Autonomous Driving and Robot Vision. Object Tracking seems to be correlated with point cloud compression techniques. These techniques are coupled with detectors such as laser radars and cameras. Three-dimensional displays and feature extraction techniques are part of task analysis for object tracking. The next block features the terms autonomous vehicles and transformer in relation to two main strands. Number 1 is sensor fusion and instance segmentation in autonomous driving in relation to point clouds for depth estimation, which involves semantic segmentation and 3D detection. Number 2 is mainly Machine Learning for classification divided into CV coupled with attention mechanisms and the terms 3D construction and detection. The terms recognition and understanding associated with object detection and the term segmentation are part of the discussion about robotics in correlation with categorization coupled with 3D CV.

The relevance of the various themes related to the topic of discussion was examined by plotting conceptual themes on a bi-dimensional matrix. The factors of analysis were centrality and density. According to Callon et al. [14], centrality measures the significance of a theme within the study field, while density serves as an indicator of a theme's development. As observed in Figure 6, themes are divided into four categories. Basic themes are characterized by high relevance but relatively low development within a given field [15]. In this case they consist of themes that revolve around the words DL, point cloud, and LiDAR. The next category is called motor themes, which are the relevant developed themes regarding this topic. In this case, if they are relevant but less developed, they would be around concepts such as autonomous driving and 3D displays. In the opposite case, if they are more developed and have lower relevance, they would be related to categorization, segmentation, and image recognition and understanding. Niche themes are very specialized themes, and they are not necessarily relevant but are very developed. Themes in this category are related to the terms "recognition: detection", retrieval and "3D for multi-view and sensors". Some themes are described as either emerging or declining as they are characterized by both low relevance and low development. In this category, themes are correlated with words such as "segmentation and categorization" and "deep learning for visual perception".

4.3. Intellectual Structure

In this section, the investigators opted to analyze the intellectual structure of the domain through co-citation analysis. Co-citation analysis allows readers to understand the foundational knowledge behind current works and contributions. Figure 7 represents a co-citation network covering 40 of the top documents referenced by the authors of the contributions gathered for the purposes of this work. The figure shows the different nodes that represent the documents, in which each node corresponds to a single contribution.

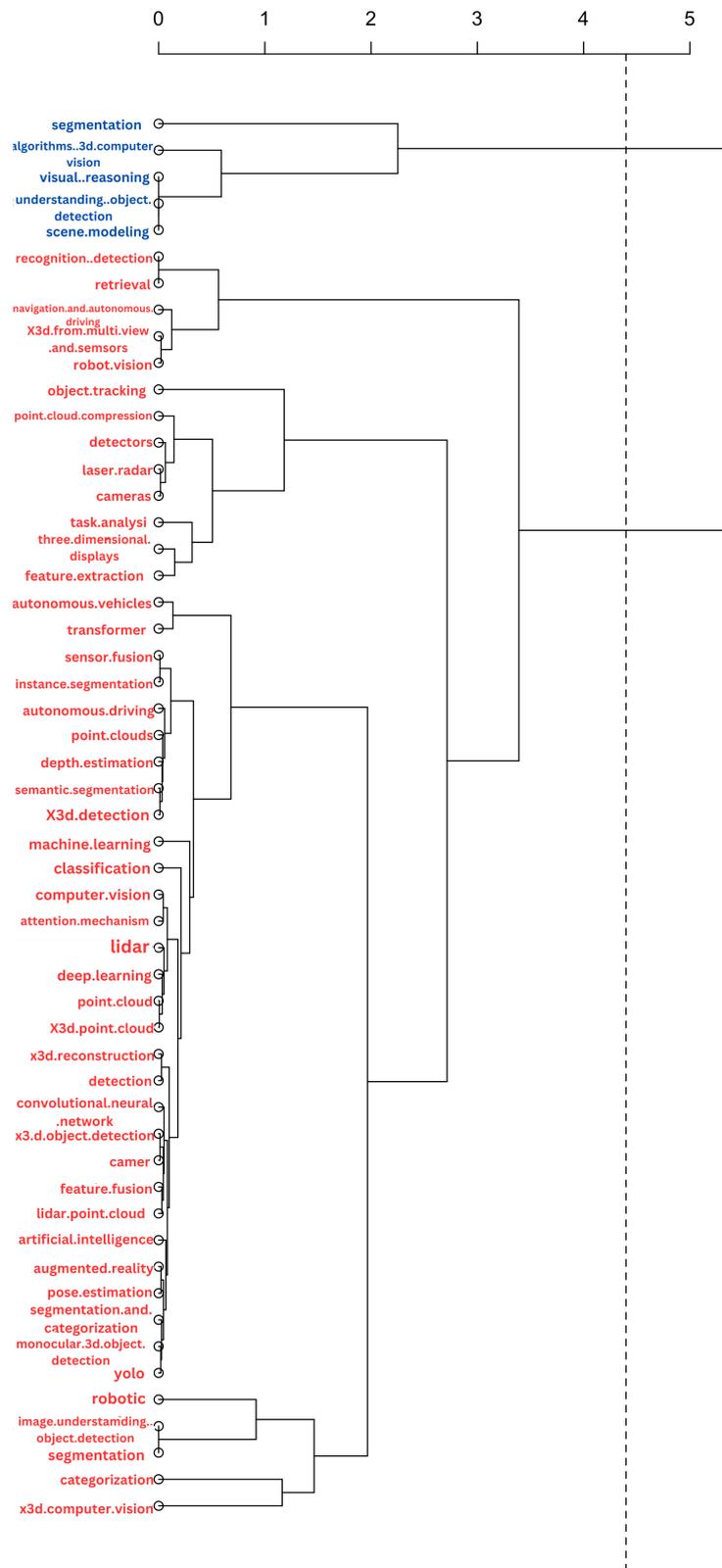


Figure 5. Topic dendrogram.

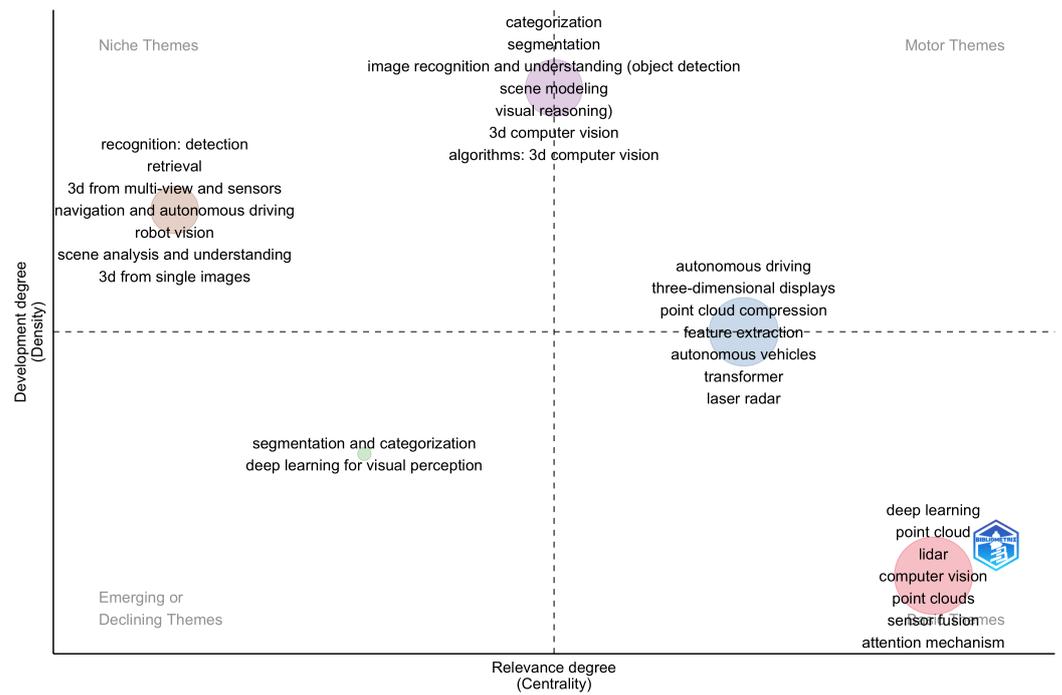


Figure 6. Thematic map based on keywords plus.

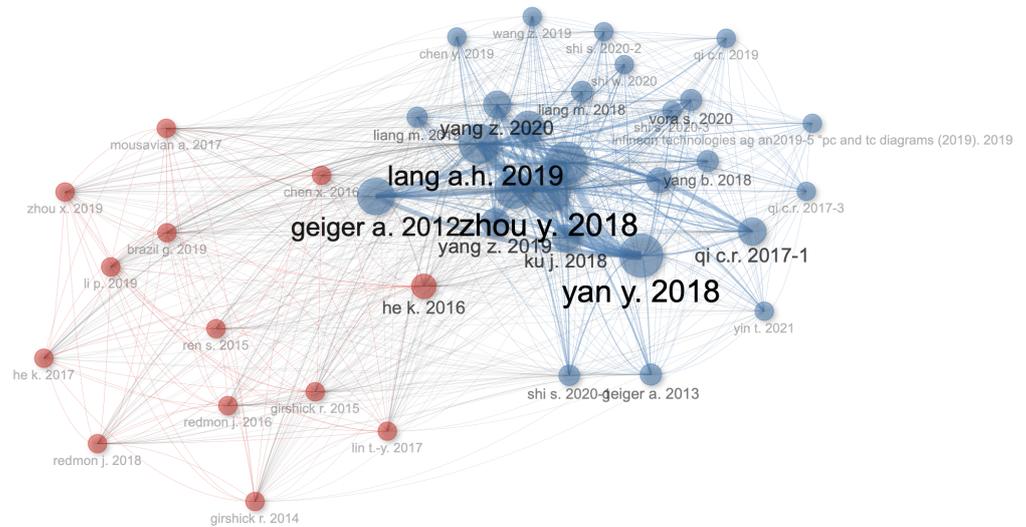


Figure 7. Co-citation network.

To measure the influence of a document, two metrics were used: betweenness centrality, which identifies influential nodes as nodes that are visited by the largest number of shortest paths from all nodes to all others within the network, and closeness centrality, where a node is considered influential if it has the shortest total distance to all other nodes [16]. Tables 5 and 6 outline the top 10 documents based on respectively betweenness and closeness.

Table 5. The 10 most influential documents based on betweenness.

Document	Betweenness
deep residual learning for image recognition (2016)	112.0900843655272
3d bounding box estimation using deep learning and geometry (2017)	33.62872504593668
focal loss for dense object detection (2017)	32.51851645291224
faster r-cnn: towards real-time object detection with region proposal networks (2015)	31.17696480721507
monocular 3d object detection for autonomous driving (2016)	29.91871834524591
fast r-cnn (2015)	29.58210081363794
stereo r-cnn based 3d object detection for autonomous driving (2019)	25.3528614719085
monocular 3d region proposal network for object detection (2019)	20.72914842468325
you only look once: unified real-time object detection (2016)	18.79331822180182
are we ready for autonomous driving? the kitti vision benchmark suite (2012)	13.65817125956712

Table 6. The 10 most influential documents based on closeness.

Document	Closeness
deep residual learning for image recognition (2016)	0.0196078431372549
3d bounding box estimation using deep learning and geometry (2017)	0.0196078431372549
monocular 3d region proposal network for object detection (2019)	0.0196078431372549
faster r-cnn: towards real-time object detection with region proposal networks (2015)	0.01886792452830189
you only look once: unified real-time object detection (2016)	0.01886792452830189
fast r-cnn (2015)	0.01886792452830189
monocular 3d object detection for autonomous driving (2016)	0.01886792452830189
focal loss for dense object detection (2017)	0.01886792452830189
rich feature hierarchies for accurate object detection and semantic segmentation (2014)	0.01886792452830189
mask r-cnn (2017)	0.01886792452830189

Documents with high betweenness centrality in a co-citation network are considered influential, as they have the ability to facilitate the flow of information between different research areas, while documents with high closeness centrality are more central to the co-citation network, as they are well-connected and have a strong influence on the dissemination of research information. Refs. [17,18] are the most influential documents given the fact that they demonstrate both the highest betweenness and closeness. This indicates that these documents not only act as a vital bridge connecting different research clusters but also occupy a central position within the entire co-citation network. The same thing goes for Ref. [19], which holds the fourth position in both Tables 5 and 6. Ref. [20] serves as an influential intermediary in the co-citation network with a betweenness of 32.5. However, it has relatively low closeness compared to other documents (eighth position according to Table 6), which means that it does not have a strong presence in terms of direct co-citation relationships with other documents. Contrarily, Ref. [21] occupies the fourth position based on closeness and the ninth position based on betweenness, making it more of a central node than a bridge node.

4.4. Social Structure

In this section, the authors sought to uncover the frequency of collaboration between countries. This analysis gives a clear image of the research landscape by showcasing the most frequent collaborations. Figure 8 is a country collaboration map. The darker the color of a country, the more documents it has produced and the links between the countries depict the network of collaborations. To further quantify the alliances between nations,

Table 7 displays the origin country of a document (from column) and the country with which the collaboration took place along with the frequency of collaboration.

Country Collaboration Map

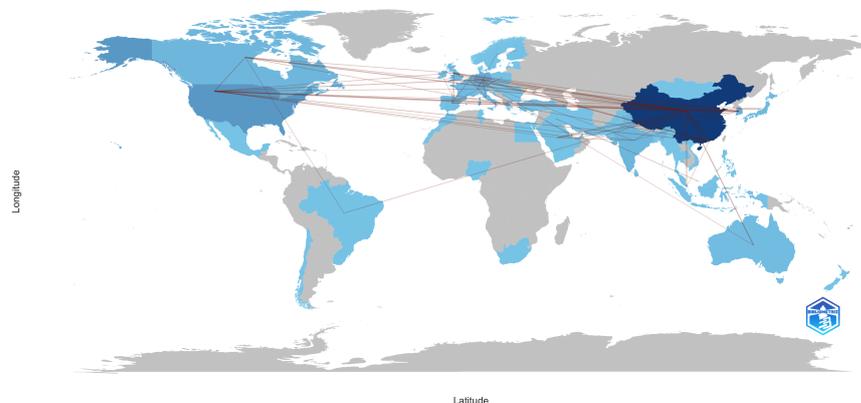


Figure 8. Country collaboration map.

Table 7. The 10 highest country collaborations.

From	To	Frequency of Collaboration
China	Hong Kong	38
China	USA	34
China	Canada	22
China	United Kingdom	21
China	Germany	20
China	Australia	18
China	Japan	10
China	Korea	9
USA	Germany	9
China	Singapore	8

China emerges as the most active country in terms of collaborations, a fact previously mentioned in Section 4.1, where it was observed that a substantial number of the highest document-producing affiliations were Chinese. Notably, China's primary collaborations involve Hong Kong and the USA, with 38 and 34 collaborations, respectively. However, when considering collaborations with other countries such as Canada, the UK, Germany, Australia, and Japan, the number ranges from 10 to 22 collaborations. For the USA, the highest number of collaborations is with Germany, with a total of 9 joint projects.

5. In-Depth Paper Review and Analysis

In this section, 8 papers were selected out of the entire batch of scientific content defined at the beginning. The process of selection started by reading the titles and abstracts of the papers from the aforementioned batch. Based on this preliminary reading, only a handful of papers caught the attention of the authors and were selected for the full reading. The resulting 8 papers present a variety of novel approaches designed for 3D object detection that proved to be effective on various levels along with the datasets that were used to train the various algorithms forming these approaches and the issues taken into consideration by these approaches.

The paper by Ruan et al. [22] introduces GNet, a novel Geometry-Aware Network for 3D object detection from sparse point clouds. Their network uses prior geometric information to improve regression performance and detection accuracy. It employs Point-Nets blocks for voxel-wise feature extraction, a 3D voxel convolution neural network for

proposal generation, an FPN block for the generation of a high resolution feature map, and an OS-loss comprised of ODIoU and Smooth-L1 losses to optimize the regression process. GNet was tested on the KITTI dataset for 3D object detection and Bird's Eye View (BEV) detection, outperforming existing methods. The results show improved accuracy, particularly for the cyclists' category, while maintaining competitive inference speeds.

Yang et al. [23] proposed Mix-Teaching, a semi-supervised learning framework for improving the performance of monocular 3D object detection. The authors addressed the problem of recovering 3D information from 2D images, which results in low precision and recall for current state-of-the-art monocular 3D object detectors. Their Mix-Teaching framework sought to address this issue by decomposing unlabeled samples into high-quality predictions and background images, which were then recombined to train a student model. The framework was tested on the KITTI and nuScenes datasets, and the results show that it outperforms baseline methods in terms of accuracy. Mix-Teaching achieved state-of-the-art results on the KITTI and nuScenes benchmarks, demonstrating its efficacy in semi-supervised monocular 3D object detection.

Wen and Cho [24] demonstrated a novel method for reconstructing 3D scenes with object awareness from a single 2D image. Their method consists of an estimation stage followed by refinement, with the goal of estimating camera parameters, layout bounding boxes, 3D object bounding boxes, and object shapes. To produce more complete and accurate meshes, the method introduces a multitask learning-based mesh reconstruction network that employs two decoders—Local Deep Implicit Functions (LDIFs) and point cloud. A depth-feature generation network was introduced to address scale ambiguity, enhancing depth information in the refined estimation stage. The method outperformed the SUN RGB-D and Pix3D datasets in layout estimation, camera poses estimation, 3D object detection, and mesh reconstruction tasks, demonstrating its potential for improving object-aware 3D scene reconstruction from single 2D images.

Chen et al. [25] proposed a new deep architecture for 3D object detection in self-driving scenarios. MSL3D, their proposed architecture, combines data from multiple sensors, including monocular cameras, stereo systems, and LiDAR, to achieve accurate and robust object detection. The authors addressed the challenge of aligning the feature extraction regions of different modalities and proposed a 2D set abstraction method that unifies the feature extraction regions of image and point cloud data. They also presented a two-stage detection framework that generates high-recall proposals using only LiDAR data in the first stage and fuses image and point cloud data in the second stage for box refinement and confidence prediction. The proposed MSL3D framework was tested on the KITTI 3D object detection dataset, outperforming other LiDAR-only and LiDAR-Camera fusion approaches. The paper emphasizes the advantages of combining multiple sensors to improve 3D object detection accuracy in autonomous driving scenarios.

Beacco et al. [26] presented a methodology for automatically reconstructing 3D objects from frontal RGB images, with a focus on guitars. The method employs sequential weak classifiers for guitar segmentation and classification, distinguishing between frontal and non-frontal views, as well as electric and classical guitar types. The 3D reconstruction is accomplished by the warping depth and normal renders of a 3D template to match the reconstructed silhouette. The method was evaluated using standard metrics, demonstrating its effectiveness in generating realistic 3D guitar models, with implications for virtual reality applications, using their own dataset. The paper discusses challenges such as concavities and occlusions, and its results are competitive with existing methods.

Li et al. [27] presented a novel method for feature fusion between dense 2D images and sparse 3D points for multi-modal 3D object detection in autonomous driving. By converting camera features into LiDAR 3D space, they created a homogeneous structure between the two data types. Their approach, dubbed Homogeneous Multi-modal Feature Fusion and Interaction (HMFI), consists of a Voxel Feature Interaction Module (VFIM) for semantic consistency, an Image Voxel Lifter Module (IVLM) for 2D to 3D feature conversion, and a Query

Fusion Mechanism (QFM) for efficient feature fusion. On the KITTI and Waymo datasets, HMF1 outperformed existing approaches, particularly when it comes to cyclist detection.

Mahmoud et al. [28] introduced a notable methodology that addresses the challenges of training end-to-end fusion methods using both camera and LiDAR sensors. This methodology, termed Dense Voxel Fusion (DVF), is designed to generate multi-scale dense voxel feature representations, specifically enhancing performance in regions characterized by low point density. Complementing this, a novel multi-modal training approach has been introduced which, rather than relying on 2D predictions, employs projected ground truth 3D bounding box labels. When benchmarked, DVF achieved a commendable third position on the KITTI 3D car detection leaderboard. Furthermore, its application to the Waymo Open Dataset showcased a significant leap in performance. Such findings underscore the potential of DVF, while also highlighting the inherent strengths and limitations of camera and LiDAR sensors in the realm of 3D object detection.

Shi et al. [29] introduced a groundbreaking method for the detection of 3D symmetry from single-view RGB-D images, eliminating the need for explicit symmetry supervision. This innovative approach leverages a weakly-supervised network trained to complete shapes based on anticipated symmetry. Central to this methodology is the utilization of a discriminative variational autoencoder, which is adept at learning the shape prior, thereby facilitating the shape completion process. When subjected to benchmark datasets such as ShapeNet and ScanNet, the method achieved a remarkable improvement in F1-score over existing supervised learning methods.

6. Discussion

This section summarizes the findings of this bibliometric study along with the additional in-depth review presented in this paper. It was possible to identify the most relevant authors in terms of the number of publications as well as in terms of influence. It was found that LI Y was the highest contributing author due to his involvement in 45 papers. This is quite an achievement given the fact that more than 98% of authors were only involved in 1 to 10 contributions. In terms of influence based on the H-index, LI H. and LI J. seem to have a considerable impact within their field.

Tsinghua University seems to be the most productive affiliation followed by Shanghai Jiao Tong University. Both universities are located in China along with the majority of the productive universities. This is due to the fact that collaborations originating from China are more significant in size than collaborations initiated by other countries. China appears to have a remarkable global reach that extends to the USA, Canada, the UK, Germany, and even Australia. It is therefore safe to assume that China is a research leader in themes related to the topic of 3D object detection.

By analyzing the frequency of the author's keywords, the investigators noticed that the most used words regarding this topic are DL, autonomous driving, and point clouds. Point clouds are inextricably linked to 3D object detection due to their ability to accurately represent spatial information in 3D space. They are frequently derived from LiDAR sensors and act as an accurate representation of the real-world scene in applications such as robotics and autonomous driving. Point clouds are crucial components of 3D object detection pipelines because they allow for precise localization and feature extraction and are compatible with DL methods.

The analysis of the author's keywords continued while attempting to uncover the conceptual structure of the field. It was possible to identify the four categories of themes: basic, motor, niche and emerging/declining themes. The most relevant and basic theme is DL, followed by point clouds and LiDAR technologies. The topics that are currently highly developed but only moderately relevant are categorization, segmentation, and image recognition and understanding for object detection. Some concepts are only fairly developed; however, their relevance is quite significant compared to the previously mentioned ones, autonomous driving, and 3D displays.

Through co-citation analysis, it was possible to uncover the key documents that served as the foundation of the studies examined in this search, such as Refs. [17,18], to name a few.

By taking a deeper look into a selection of 8 papers chosen and discussed by the investigators, it was possible to identify some novel methods and frameworks that had a significant role in addressing several issues regarding 3D object detection such as improving object detection from point clouds in Gnet [22], which relies on prior geometric information to enhance regression outcomes and boost detection precision. Yang et al. produced a paper [23] that dealt with the issue of 3D information from 2D images by employing a semi-supervised method. In the interest of taking on the issue of aligning the feature extraction regions of different modalities, Chen et al. [25] introduced a two-stage detection framework that consists of generating high-recall proposals using only LiDAR data and then performing box refinement and confidence prediction by fusing image and point cloud data. The KITTI dataset played an important role in the training and evaluation of most of these frameworks.

Some of the previously presented papers also offer an idea about future directions. Ruan et al. [22] suggest exploring multimodal fusion methods that tackle the challenge of improving performance on small objects, particularly pedestrians. In the realm of monocular 3D object detection, the future work of Yang et al. [23] involves continuously improving detectors through the collection of more unlabeled images and considering the adoption of end-to-end training. Wen et al. [24] mention future orientations that include extending their method to more general object reconstruction tasks. Beacco et al. [26] state that their future plans involve addressing challenges such as reconstruction from different viewpoints and conducting perceptual studies for validation. Shi et al. [29] ended their contribution with quite the number of challenges including optimizing object segmentation and symmetry prediction, cross-category generalization, and exploring purely geometric symmetry detection.

7. Conclusions

In conclusion, this paper presents a comprehensive bibliometric analysis focused on the topic of 3D object detection, from 2022 to 2023. The study design, data collection, and analysis phases were meticulously followed to provide a holistic perspective of the research field.

The findings indicate that China is the most active country in terms of collaborations, with significant contributions from Chinese universities and researchers. Collaborations with Hong Kong and the USA are the most prominent, followed by collaborations with other countries such as Canada, the UK, Germany, Australia, and Japan. The analysis of authors and their keywords reveals that DL, autonomous driving, and point clouds are the most frequently used and relevant themes in the field of 3D object detection. The conceptual structure analysis highlights the four categories of themes: basic, motor, niche, and emerging/declining themes. The most relevant and basic themes include DL, point clouds, and LiDAR technologies, while autonomous driving and 3D displays are highly developed but moderately relevant themes. In terms of the intellectual structure, certain influential documents act as critical bridges between research clusters and occupy central positions in the co-citation network.

Overall, this bibliometric study provides valuable insights into the field of 3D object detection, shedding light on the most relevant authors, themes, and influential documents. These findings can guide researchers, institutions, and policymakers in understanding the current state-of-the-art and identifying potential areas for future research and collaboration.

Author Contributions: All the authors have contributed the same. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the Foundation for Science and Technology (FCT, Portugal) for financial support through national funds FCT/MCTES (PIDDAC) to CeDRI (UIDB/05757/2020 and UIDP/05757/2020) and SusTEC (LA/P/0007/2021).

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 conflict of interest.

References

1. Ellegaard, O.; Wallin, J.A. The bibliometric analysis of scholarly production: How great is the impact? *Scientometrics* **2015**, *105*, 1809–1831. [[CrossRef](#)] [[PubMed](#)]
2. Donthu, N.; Kumar, S.; Mukherjee, D.; Pandey, N.; Lim, W.M. How to conduct a bibliometric analysis: An overview and guidelines. *J. Bus. Res.* **2021**, *133*, 285–296. [[CrossRef](#)]
3. Lohia, A.; Kadam, K.; Joshi, R.; Bongale, A. Bibliometric Analysis of One-stage and Two-stage Object Detection. *Libr. Philos. Pract.* **2021**.
4. Chen, H.; Deng, Z. Bibliometric Analysis of the Application of Convolutional Neural Network in Computer Vision. *IEEE Access* **2020**, *8*, 155417–155428. [[CrossRef](#)]
5. Liu, W.; Wu, H.; Hu, K.; Luo, Q.; Cheng, X. A Scientometric Visualization Analysis of Image Captioning Research From 2010 to 2020. *IEEE Access* **2021**, *9*, 156799–156817. [[CrossRef](#)]
6. Khan, U.; Khan, H.U.; Iqbal, S.; Munir, H. Four decades of image processing: A bibliometric analysis. *Libr. Hi Tech* **2022**, *ahead-of-print*. [[CrossRef](#)]
7. Zhang, Y. Image engineering in China: 2022. *J. Image Graph.* **2023**, *28*, 879–892. [[CrossRef](#)]
8. Okoli, C.; Schabram, K. A Guide to Conducting a Systematic Literature Review of Information Systems Research. *SSRN Electron. J.* **2010**. [[CrossRef](#)]
9. Chadegani, A.A.; Salehi, H.; Yunus, M.M.; Farhadi, H.; Fooladi, M.; Farhadi, M.; Ebrahim, N.A. A Comparison between Two Main Academic Literature Collections: Web of Science and Scopus Databases. *Asian Soc. Sci.* **2013**, *9*, 18. [[CrossRef](#)]
10. Vieira, E.S.; Gomes, J.A.N.F. A comparison of Scopus and Web of Science for a typical university. *Scientometrics* **2009**, *81*, 587–600. [[CrossRef](#)]
11. Chen, G.; Xiao, L. Selecting publication keywords for domain analysis in bibliometrics: A comparison of three methods. *J. Inf.* **2016**, *10*, 212–223. [[CrossRef](#)]
12. Hirsch, J.E. An index to quantify an individual's scientific research output. *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 16569–16572. [[CrossRef](#)]
13. Ran, X.; Xi, Y.; Lu, Y.; Wang, X.; Lu, Z. Comprehensive survey on hierarchical clustering algorithms and the recent developments. *Artif. Intell. Rev.* **2023**, *56*, 8219–8264. [[CrossRef](#)]
14. Callon, M.; Courtial, J.P.; Laville, F. Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemistry. *Scientometrics* **1991**, *22*, 155–205. [[CrossRef](#)]
15. Khare, A.; Jain, R. Mapping the conceptual and intellectual structure of the consumer vulnerability field: A bibliometric analysis. *J. Bus. Res.* **2022**, *150*, 567–584. [[CrossRef](#)]
16. Sheikhamadi, A.; Nematbakhsh, M.A.; Shokrollahi, A. Improving detection of influential nodes in complex networks. *Phys. A Stat. Mech. Its Appl.* **2015**, *436*, 833–845. [[CrossRef](#)]
17. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep Residual Learning for Image Recognition, 2015. *arXiv* **2015**, arXiv:1512.03385.
18. Mousavian, A.; Anguelov, D.; Flynn, J.; Kosecka, J. 3D Bounding Box Estimation Using Deep Learning and Geometry. *arXiv* **2016**, arXiv:1612.00496v2.
19. Ren, S.; He, K.; Girshick, R.; Sun, J. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *arXiv* **2015**, arXiv:1506.01497v3.
20. Lin, T.Y.; Goyal, P.; Girshick, R.; He, K.; Dollár, P. Focal Loss for Dense Object Detection. *arXiv* **2017**, arXiv:1708.02002v2.
21. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. *arXiv* **2015**, arXiv:1506.02640v5.
22. Ruan, H.; Xu, B.; Gao, J.; Liu, L.; Lv, J.; Sheng, Y.; Zeng, Z. GNet: 3D Object Detection from Point Cloud with Geometry-Aware Network. In Proceedings of the 2022 IEEE International Conference on Cyborg and Bionic Systems (CBS), Wuhan, China, 24–26 March 2023; pp. 190–195. [[CrossRef](#)]
23. Yang, L.; Zhang, X.; Li, J.; Wang, L.; Zhu, M.; Zhang, C.; Liu, H. Mix-Teaching: A Simple, Unified and Effective Semi-Supervised Learning Framework for Monocular 3D Object Detection. *IEEE Trans. Circuits Syst. Video Technol.* **2023**, *Early Access*. [[CrossRef](#)]
24. Wen, M.; Cho, K. Object-Aware 3D Scene Reconstruction from Single 2D Images of Indoor Scenes. *Mathematics* **2023**, *11*, 403. [[CrossRef](#)]
25. Chen, W.; Li, P.; Zhao, H. MSL3D: 3D object detection from monocular, stereo and point cloud for autonomous driving. *Neurocomputing* **2022**, *494*, 23–32. [[CrossRef](#)]
26. Beacco, A.; Gallego, J.; Slater, M. 3D objects reconstruction from frontal images: An example with guitars. *Vis. Comput.* **2022**. [[CrossRef](#)]

27. Li, X.; Shi, B.; Hou, Y.; Wu, X.; Ma, T.; Li, Y.; He, L. Homogeneous Multi-modal Feature Fusion and Interaction for 3D Object Detection. In Proceedings of the Computer Vision—ECCV 2022, Tel Aviv, Israel, 23–27 October 2022; Lecture Notes in Computer Science; Avidan, S., Brostow, G., Cissé, M., Farinella, G.M., Hassner, T., Eds.; Springer Nature: Cham, Switzerland, 2022; pp. 691–707. [[CrossRef](#)]
28. Mahmoud, A.; Hu, J.S.K.; Waslander, S.L. Dense Voxel Fusion for 3D Object Detection. In Proceedings of the 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 2–7 January 2023; pp. 663–672. [[CrossRef](#)]
29. Shi, Y.; Xu, X.; Xi, J.; Hu, X.; Hu, D.; Xu, K. Learning to Detect 3D Symmetry From Single-View RGB-D Images With Weak Supervision. *IEEE Trans. Pattern Anal. Mach. Intell.* **2022**, *45*, 4882–4896. [[CrossRef](#)] [[PubMed](#)]

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.