Visual Informatics 8 (2024) 13–22

Contents lists available at ScienceDirect

Visual Informatics

journal homepage: www.elsevier.com/locate/visinf

Review article

# Computer Vision in Augmented, Virtual, Mixed and Extended Reality environments—A bibliometric review

Júlio Castro Lopes *, Rui Pedro Lopes

*Research Center in Digitalization and Intelligent Robotics (CeDRI), Instituto Politécnico de Bragança (IPB), Bragança, Portugal*

ARTICLE INFO

*Article history:*
Received 14 August 2023
Received in revised form 25 October 2024
Accepted 13 November 2024
Available online 12 December 2024

*Keywords:*
Augmented Reality
Virtual Reality
Deep Learning
Computer Vision
Machine Learning

ABSTRACT

This work describes a bibliometric analysis of the literature on the use of computer vision algorithms in Augmented Reality (AR), Virtual Reality (VR), Mixed Reality (MR), and Extended Reality (XR) environments. The analysis aims to highlight the evolution, trends, and effects of research in this field. This review provides an overview of immersive technologies and their applications, as well as the role of computer vision algorithms in enabling these technologies and the potential benefits of using such algorithms. This study identifies important authors, institutions, and research themes by using bibliometric indicators such as citation counts, co-citation analysis, and network analysis. The analysis also identifies gaps and opportunities for additional research in this area, as well as a critical assessment of the quality and relevance of the publications.



## 1. Introduction

Computer Vision (CV) is a subfield of Artificial Intelligence (AI) and computer science that aims to develop and train models to accept and interpret visual data. CV analyzes and interprets visual information from digital images or videos using algorithms, statistical models, and Machine Learning (ML) techniques. This can include tasks like object recognition (Francies et al., 2022), facial expression recognition (Zeng et al., 2022), image segmentation (Hatamizadeh et al., 2022), human action recognition (Chi et al., 2022), and so on. CV can be used in the context of Augmented Reality (AR), Virtual Reality (VR), Mixed Reality (MR), and Extended Reality (XR) to create immersive digital experiences that can seamlessly blend virtual and physical elements (Deshpande and Sanghavi, 2019; González Izard et al., 2019; Cīrulis et al., 2020). While AR involves overlaying digital information on real-world objects, VR creates a completely artificial, computer-generated environment that simulates real or imaginary environments. With MR, users can manipulate and modify virtual objects as if they were in the real world, by combining real and virtual elements. XR is a broader term that encompasses these immersive technologies, as well as other related technologies, such as 360-degree videos. There are many areas in which these technologies have demonstrated great potential, such as education (Sala, 2021; Zhang et al., 2020), entertainment (Hung et al., 2021; Saint-Louis and Hamam, 2021), rehabilitation (Berton et al., 2020; Patil et al., 2022; Postolache et al., 2021), and manufacturing (Ababsa, 2020; Botto et al., 2020), among others (Cruz et al., 2019; Zafar and Zachar, 2020; Blanco-Pons et al., 2019). Interacting with computer-generated objects and environments creates immersive experiences that can enhance learning and entertainment (Rebbani et al., 2021). Tracking and interpreting user interactions in real-time, however, is essential to enable such experiences.

CV assisted with ML and Deep Learning (DL) can play a significant role in the development of VR, AR, and XR technologies, as it enables these environments with the most recent advances in the field, towards better recognition and understanding of visual information (Nalbant and Uyanik, 2021). In these scenarios, object recognition is frequently used (Hbali et al., 2013; Ghasemi et al., 2022), as these algorithms try to recognize real-world objects and integrate them into virtual or augmented environments, recognize specific objects in real-time and track their movements. ML, for instance, can be used in AR games where the player must follow and interact with virtual objects placed in the real world (Samir et al., 2018).

This type of environment, in particular games and serious games, can also benefit from Dynamic Difficulty Adjustment (DDA) methods, which dynamically adapt the game difficulty based on the player's performance (Lopes and Lopes, 2022). For example, if the player is having difficulty completing a task, the game can automatically reduce the difficulty level to make it easier. DDA can help keep users engaged, motivated, and successful in achieving their goals by offering personalized experiences tailored to the user's needs and capabilities. This has a wide range of applications, including entertainment (Li et al., 2022), education (Chrysafiadi et al., 2023), and rehabilitation

* Corresponding author.
*E-mail addresses:* juliolopes@ipb.pt (J.C. Lopes), rlopes@ipb.pt (R.P. Lopes).

https://doi.org/10.1016/j.visinf.2024.11.002

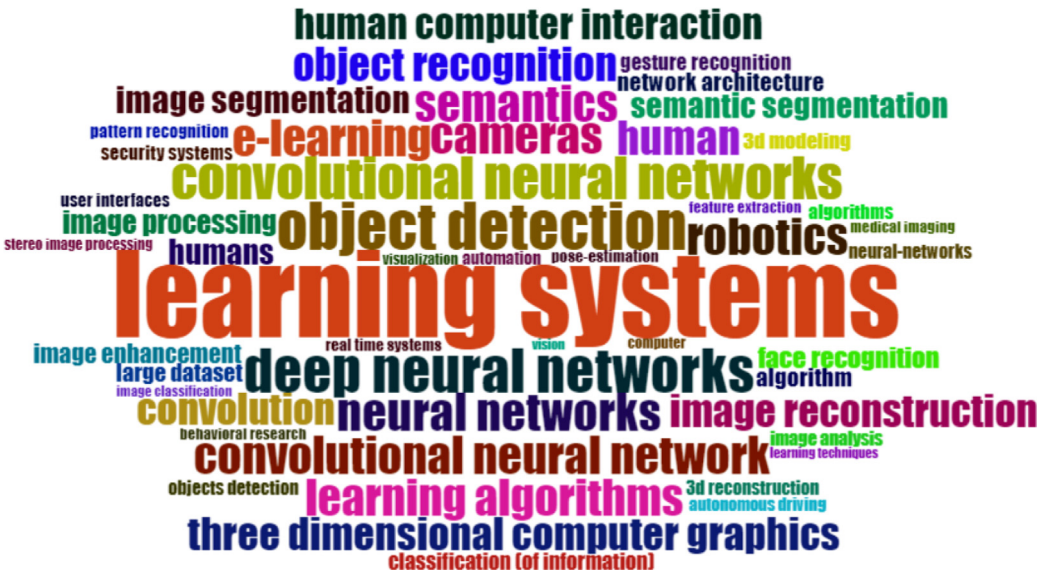


**Fig. 1.** Word Cloud of CV in VR, AR, MR and XR environments.

games (Kamikokuryo et al., 2022). This adaptation can benefit from CV, by the real-time analysis of players' facial expressions (Rodrigues et al., 2022) and body posture (Lopes et al., 2021), among others.

This paper describes a bibliometric analysis of the literature on the use of CV algorithms in AR, VR, MR, and XR environments. The review aims to highlight the evolution, trends, and effects of research in this field. It provides an overview of immersive technologies and their applications, the function of CV algorithms in enabling these technologies, and the potential advantages of using these algorithms. In order to identify important authors, institutions, and research themes, this review examines a wide range of publications. It does this by using different bibliometric indicators, including citation counts, co-citation analysis, and network analysis. This paper also identifies gaps and opportunities for further research in this area and provides a critical analysis of the publications' quality and relevance.

## 2. Methodology

CV algorithms are used in digital technologies in a variety of ways, and this review aims to revise and understand their connections and application areas. To provide a clear analysis of the use of CV algorithms in VR, AR, MR and XR environments, all the keywords used in the string query were removed from World Cloud, Tree Map, and Word Frequency Over Time, since it is expected that they show high relevance in all papers. With the removal of all "redundant" keywords, the application of these algorithms in these environments is more easily highlighted.

Listing 1 was used in the Scopus database.

**Listing 1** Query used to retrieve the Scopus metadata

```
TITLE-ABS-KEY ( ( vr OR "Virtual Reality" OR ar OR "
    ↪ Augmented Reality" OR "Mixed Reality" OR xr OR "
    ↪ Extended Reality" ) AND "computer vision" AND (
    ↪ dl OR "Deep Learning" OR ml OR "Machine Learning
    ↪ " ) ) AND PUBYEAR > 2011 AND PUBYEAR < 2024
    ↪ AND ( LIMIT-TO ( LANGUAGE , "English" ) )
```

As a peer-reviewed literature abstract and citation database, Scopus is the most comprehensive and viable, since it mostly overlaps the results of Web of Science (Al-Khoury et al., 2022). This query aggregates all studies that use VR, AR, MR or XR technologies, focusing on the use of CV and ML/DL in their application. It should be noted that the term MR was not included in the string query because it falls under a different theme that is not the focus of this review (Magnetic Resonance Imaging). Only papers that were written in English were considered for this analysis. In view of the fact that research on this field of application did not have a high level of relevance and intensity up until 2012, then the search was limited by starting in 2012 to 2024. As a result, this bibliometric review considered 750 documents.

Biblioshiny and ElasticSearch were used for the analysis provided in this paper (Aria and Cuccurullo, 2017; Kuc and Rogozinski, 2013). Biblioshiny provides a graphical way to analyze titles, abstracts, keywords, and geographic information collected from the bibliography metadata. However, there are details that can only be retrieved by analyzing the full content of the articles. For this, the full set of PDF files were downloaded and stored locally for subsequent text mining. ElasticSearch was configured to index all the files, allowing the generation of detailed queries over the content of the text. The main purpose of this analysis was to infer the algorithms, programming languages and techniques most frequently used.

## 3. Bibliometric analysis

This section analyzes the information processed in biblioshiny, taking into account all the information in the metadata, excluding the text content. It shows the most frequent keywords, the evolution of the appearance of the most frequent keywords, the production of the most relevant countries (taking into account the number of publications), the collaboration of authors from different countries, the most relevant affiliations, the most relevant sources and also the most cited papers.

Fig. 1 shows the word cloud, which results from keyword analysis, that highlights the most frequently occurring keywords in the selected studies.

Learning systems, e-learning, and learning algorithms are clearly associated with this bibliometric search, along with human–computer interaction, robotics, and three-dimensional computer graphics. Regarding the usage of these technologies in learning platforms, for example, Zhang (2020) used DL to create a VR-AR mobile platform and game model for gesture-based VR game interaction. His game establishes interaction rules, creates gesture

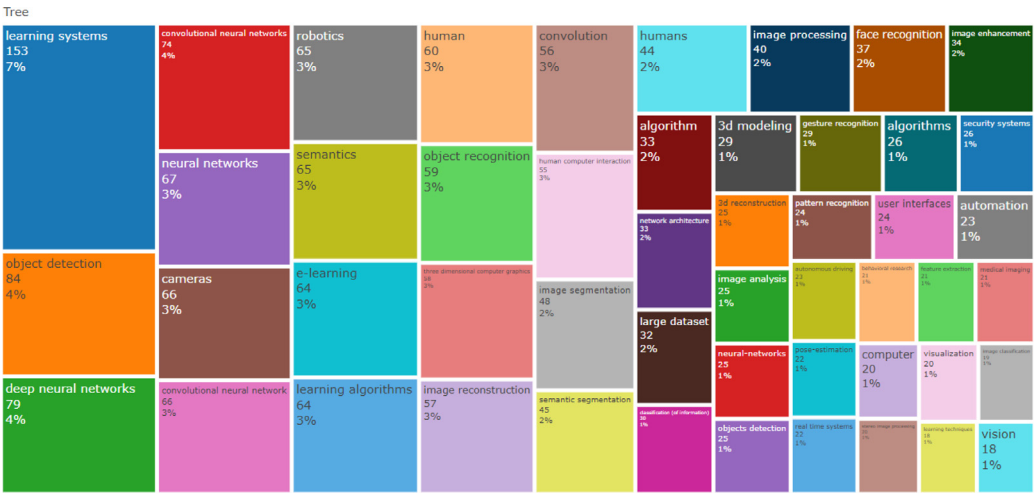


**Fig. 2.** Tree map of CV in VR, AR, MR and XR environments.

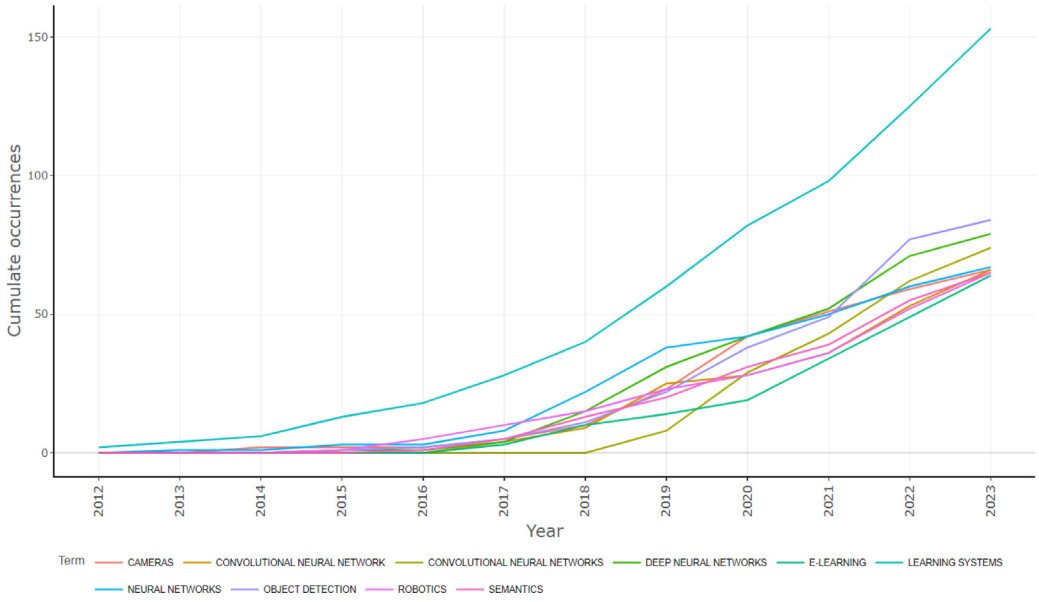


**Fig. 3.** Words frequency over time.

models, uses computer vision for natural gesture recognition, and offers a DL-supported learning platform. Poonja et al. (2023) proposed a CV system, to enhance student engagement in online learning using an AR environment. Tao et al. (2020) proposed a ML-based research framework to estimate trainees' confidence about their decisions in immersive learning environments. Zhao et al. (2018) developed an AR system for plant learning, providing a didactic module that enhances learners' interest in learning.

As this word cloud demonstrates, algorithms and highly related content with CV and DL applications are also evident (object detection, object recognition, neural networks, image reconstruction, etc.), as they connect these environments with the application of AI algorithms. For example, Cozzolino et al. (2022) and Wang et al. (2021), both focus on the application of object detection methods in AR mobile devices, while Lin et al. (2018) deals with object recognition. Neural networks are a topic that is also involved in several papers analyzed in this study, due to its powerful application in CV algorithms (Zhou et al., 2022; Sorokin et al., 2020; Epishkina and Zapechnikov, 2016). Cameras are also mentioned, as they are frequently used to acquire images (Amparore et al., 2022; Liu et al., 2021).

As an overview of the application of these algorithms in these environments, the Tree Map highlights the combination of keywords that are relevant to our search query (Fig. 2).

The analysis of the tree map allows us to quantitatively evaluate how these keywords are related to the query. It is worth noting that learning algorithms and e-learning also have a significant percentage, indicating that learning systems may have gained more prominence, though some authors may consider these two terms interchangeable (Zhao et al., 2021).

CV is still a hot topic as it has been in growing interest and applications since 2016 (Fig. 3). In the last 5 years (since 2018), learning systems have been drastically increasing in relation to these topics.

Aside from learning systems, object detection, deep neural networks, and Convolutional Neural Networks (CNNs), the most frequently used keywords are all related to DL algorithms, with some of them also associated with CV (object recognition and object detection), demonstrating the usefulness and frequency with which all of these terms are associated with digital environments and CV.

Fig. 4 depicts the evolution of production in six of the most active countries currently engaged in research on these topics.

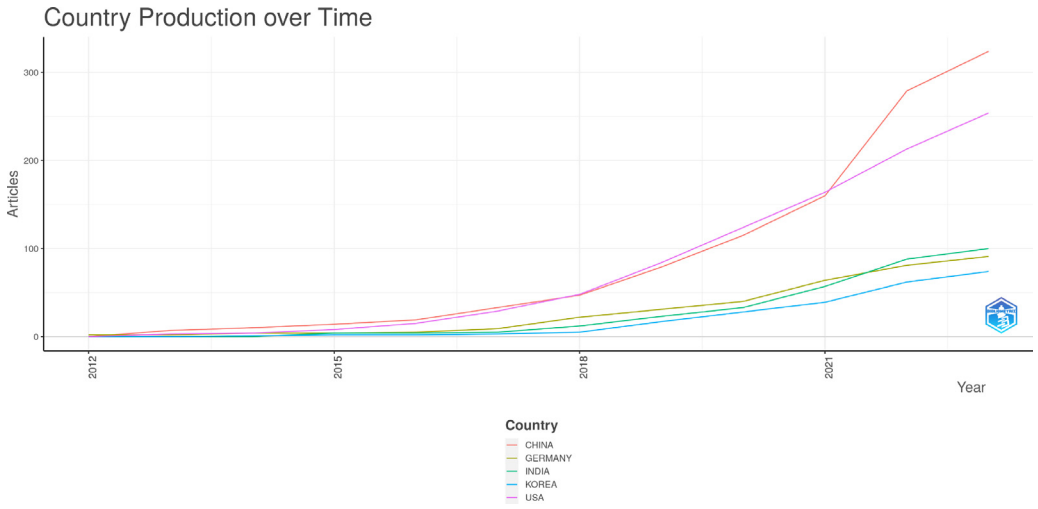


**Fig. 4.** Country production over time.

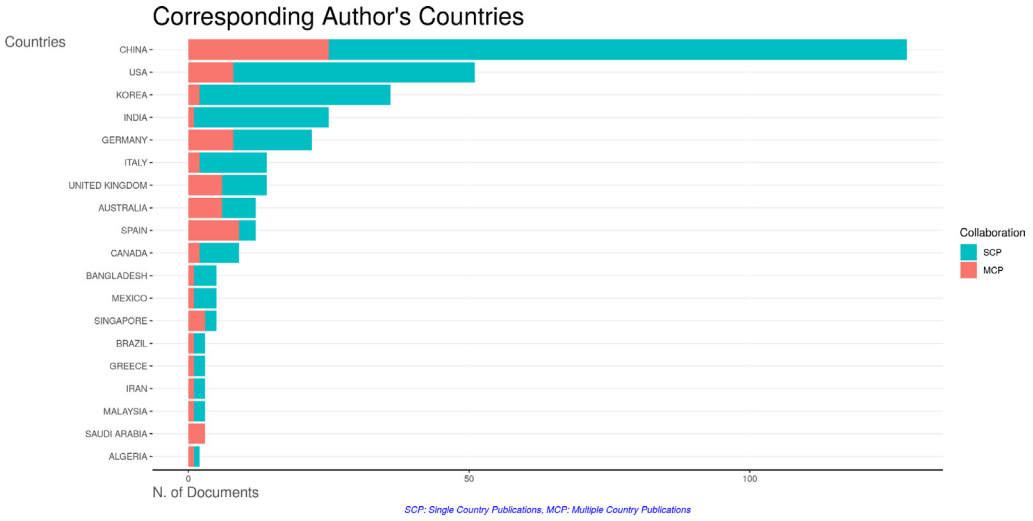


**Fig. 5.** Corresponding author's countries.

By 2016, China and the United States were already ahead in the number of papers published. Until 2021, both of these countries' production was very similar. China began rapidly expanding its research into these topics starting in 2021, and it is currently the country with the highest number of publications. Although the United States produced the most in 2020, China has been very active in the development of new research due to advances in these emerging technologies. Fig. 5 shows the countries of the corresponding authors, as well as whether they have been written only by authors from the same country or a multi-country development.

Collaboration between countries in research can be critical to improving research and has several advantages (Dini et al., 2022). It is an important factor in increasing global impact because it ensures that research findings are relevant and applicable to a global audience, influencing citations, visibility, and replicability of results (Lancho Barrantes et al., 2012). These collaborations also aid in the provision of opportunities for future collaborations, the expansion of the researchers' professional networks, the increase of funding opportunities, and the promotion of cultural exchange (Hou et al., 2021).

In terms of joint publications, the United States has the most associations around the world (Fig. 6). They have shared publications with several European countries and Canada, as well as from Asia, specifically China, demonstrating that the two countries that produced the most are also associated in several publications, which was undoubtedly a key factor in being so distinct in terms of the number of publications.

It is worth noting that the most relevant affiliations (Fig. 7), with 15 publications, are from University of California, it is followed by Technical University of Munich, followed by Beihang University and "National Taiwan University", tied with 9 publications.

Fig. 8 illustrates the evolution of publication in journals and conference proceedings over time in terms of where the majority of works are published.

Springer conferences (Lecture Notes in Computer Science) have the most publications since 2017, and the number of publications has been increasing in recent years. Most of the published papers in a conference were presented at the IEEE Computer Society Conference on Computer Vision and Pattern Recognition. The journal with the most publications is Sensors, which is edited by MDPI.

It is important to evaluate and analyze the publications with more citations, as they are the most relevant fonts for the research community. Table 1 shows the 10 most globally cited papers.

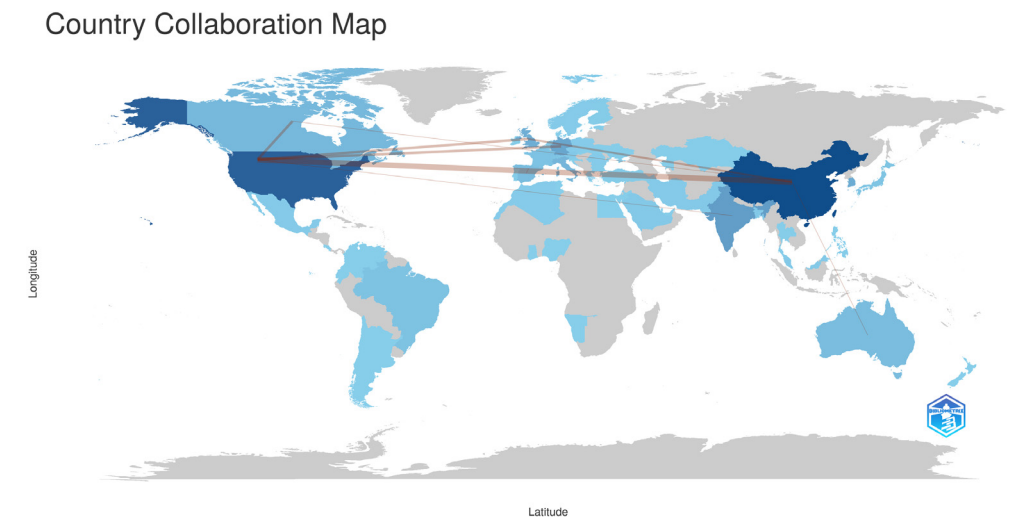


**Fig. 6.** Country collaboration map.

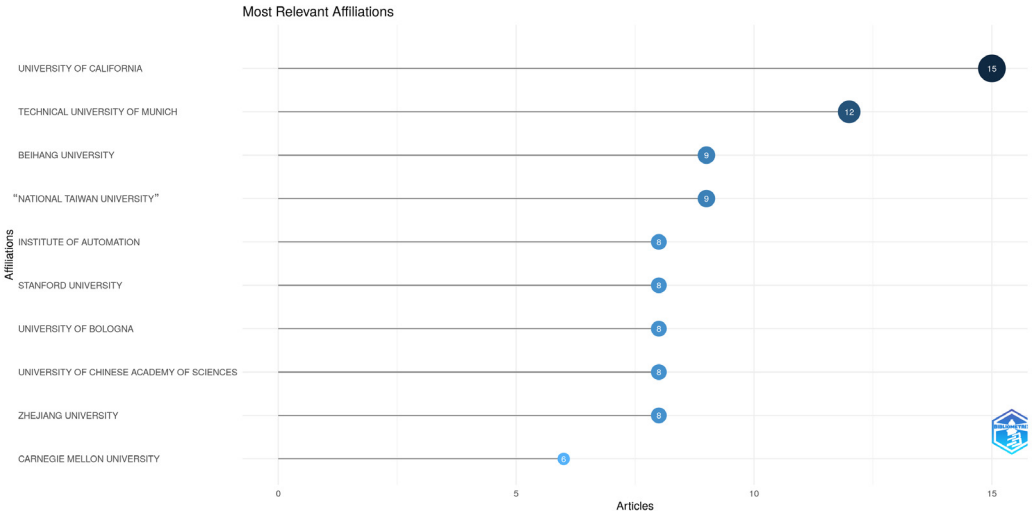


**Fig. 7.** Most Relevant Affiliations.

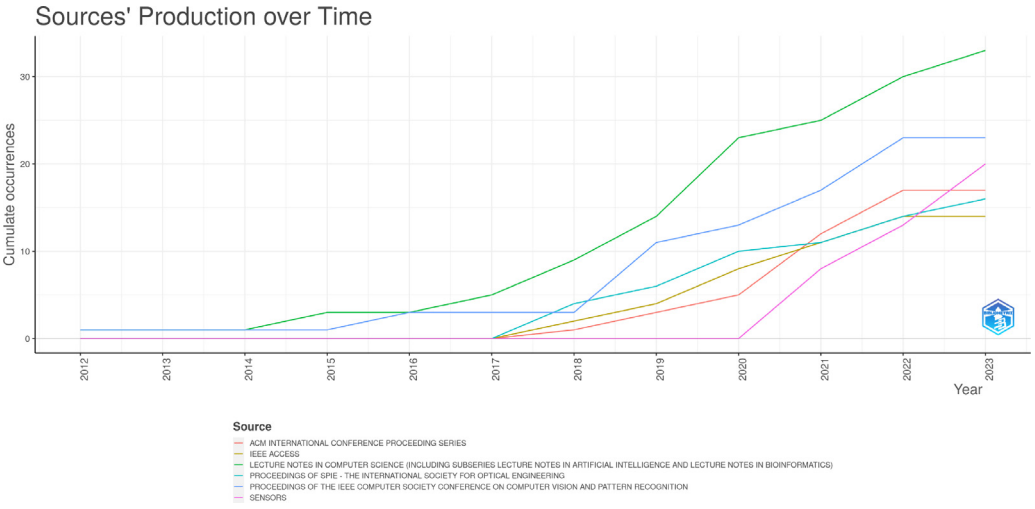


**Fig. 8.** Publication sources.

**Table 1**
Most global cited papers.

| Title | Year | Citations |
|---|---|---|
| VirtualWorlds as Proxy for Multi-object Tracking Analysis (Gaidon et al., 2016) | 2016 | 615 |
| Image Segmentation Using Deep Learning: A Survey (Minaee et al., 2022) | 2021 | 588 |
| A survey on deep learning techniques for image and video semantic segmentation (Garcia-Garcia et al., 2018) | 2018 | 529 |
| On Human Motion Prediction Using Recurrent Neural Networks (Martinez et al., 2017) | 2017 | 443 |
| From Coarse to Fine: Robust Hierarchical Localization at Large Scale (Sarlin et al., 2019) | 2019 | 310 |
| Normalized Object Coordinate Space for Category-Level 6D Object Pose and Size Estimation (Wang et al., 2019) | 2019 | 268 |
| Visual SLAM and Structure from Motion in Dynamic Environments: A Survey Saputra et al. (2018) | 2018 | 248 |
| Augmented Reality Meets Computer Vision: Efficient Data Generation for Urban Driving Scenes (Abu Alhaija et al., 2018) | 2018 | 237 |
| Driving in the Matrix: Can virtual worlds replace human-generated annotations for real world tasks? (Johnson-Roberson et al., 2017) | 2017 | 233 |
| Understanding the Limitations of CNN-Based Absolute Camera Pose Regression (Sattler et al., 2019) | 2019 | 199 |

"Virtual Worlds as a Proxy for Multi-Object Tracking Analysis" (Gaidon et al., 2016) describes a method for creating photo-realistic virtual worlds that serve as proxies for real-world multi-object tracking evaluation by leveraging computer graphics advancements. They demonstrated the similarity of real and virtual worlds for DL models, proposed virtual pre-training to improve performance, and highlighted the impact of varying weather and imaging conditions on recognition accuracy, emphasizing the need for more robust models in a variety of scenarios.

"Image Segmentation Using Deep Learning: A Survey" (Minaee et al., 2022) examines over 100 DL techniques for image segmentation, including fully convolutional networks, encoder–decoder architectures, and attention-based models. It investigates dataset utilization, network designs, training methodologies, and contributions, but only mentions that image segmentation can be very useful in AR environments without providing any studies or examples of applications.

The paper, "A Survey on Deep Learning Techniques for Image and Video Semantic Segmentation" (Garcia-Garcia et al., 2018), investigates the rapid integration of DL approaches in semantic segmentation, a critical aspect of CV. They provide essential resources, insights, and a roadmap for future advancements in deep learning-enabled semantic segmentation through this pioneering work, significantly contributing to the research community's understanding and progress in this vital field, despite only mentioning that image semantic segmentation is widely used in VR and AR environments.

The paper, titled On "Human Motion Prediction Using Recurrent Neural Networks" (Martinez et al., 2017), delves into the challenges of human motion modeling using deep recurrent neural networks (RNNs) for tasks such as short-term motion prediction. The authors investigate cutting-edge methods, propose a novel architecture based on sampling-based loss, and emphasize the importance of large-scale datasets. The paper contributes valuable insights for advancing human motion prediction using RNNs by addressing shortcomings in previous approaches and achieving improved short-term prediction performance, even though it only references human motion modeling as applicable in VR and AR environments, their approach can be possibly applied in these environments.

"From Coarse to Fine: Robust Hierarchical Localization at Large Scale" (Sarlin et al., 2019) describes a robust and efficient hierarchical localization approach for accurate 6-DoF camera pose estimation in large-scale environments. To achieve real-time operation, their method employs a global-to-local strategy based on learned descriptors and keypoints. The paper demonstrates significant improvements in accuracy and efficiency over existing methods, especially in difficult conditions such as night-time queries, and establishes new benchmarks for large-scale localization. Although it does not focus on AR or VR, it mentions the usage of 6-DoF camera pose estimation in this type of environment.

"Normalized Object Coordinate Space for Category-Level 6D Object Pose and Size Estimation" (Wang et al., 2019) describes how to estimate the 6D pose and dimensions of unseen object instances in RGB-D images. It introduces a shared canonical representation for different objects within a category called Normalized Object Coordinate Space (NOCS), which enables robust pose and size estimation. To achieve accurate 6D pose and size estimation, the method employs a region-based neural network to infer correspondences between observed pixels and the NOCS representation, and it combines predictions with depth maps. They conclude that their approach has important applications in areas such as AR, robotics, and 3D scene understanding.

The paper "Visual SLAM and Structure from Motion in Dynamic Environments: A Survey" (Saputra et al., 2018) provides an in-depth examination of Structure from Motion (SfM) and visual Simultaneous Localization and Mapping (SLAM) techniques in dynamic environments. It tackles motion segmentation, dynamic object segmentation and tracking, and joint motion segmentation and reconstruction. The survey categorizes existing approaches, including deep learning-based methods, and discusses their benefits and limitations, as well as the application of these techniques in AR environments.

"Augmented Reality Meets Computer Vision: Efficient Data Generation for Urban Driving Scenes" (Abu Alhaija et al., 2018) describes an innovative data augmentation approach for improving CV tasks in AR and other environments. The authors propose a method for combining real-world images with virtual objects, with a focus on semantic instance segmentation and object detection. The authors demonstrate improved model performance by augmenting real images with realistically rendered virtual objects, overcoming the limitations of purely synthetic or limited real-world data. The proposed method achieves cutting-edge results and emphasizes the importance of data diversity and realism in training deep neural networks for complex tasks.

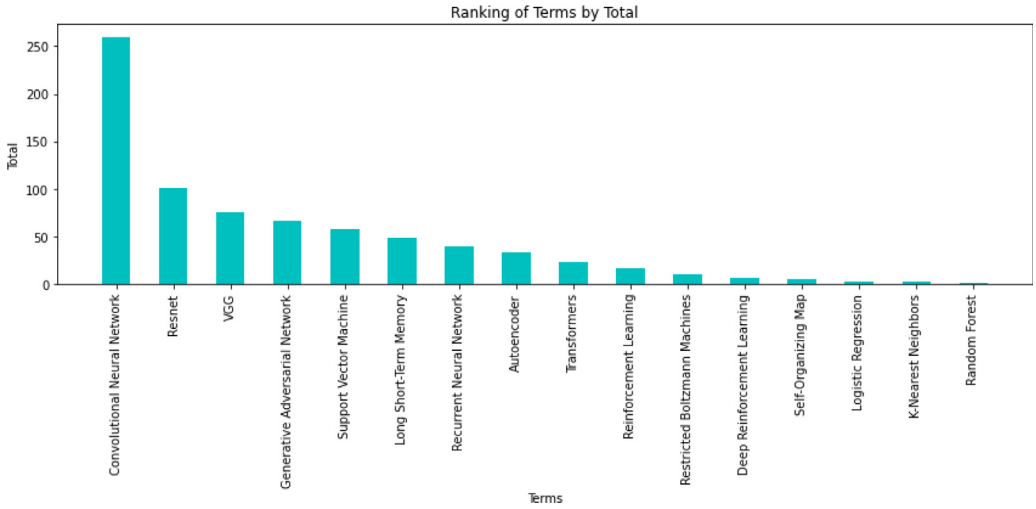


**Fig. 9.** Frequency of Deep and Machine Learning algorithms.

"Driving in the Matrix: Can virtual worlds replace human-generated annotations for real world tasks?" (Johnson-Roberson et al., 2017) suggests using synthetic annotated data from a simulation engine to improve DL for CV. When compared to architectures trained solely on human-annotated real-world data, the approach outperforms them in vehicle detection. The study suggests potential acceleration of deep learning applications for sensor-based classification tasks, such as self-driving cars, by using photo-realistic simulation images, and provides insights into addressing data set bias and improving object classification through simulation-generated data.

The paper "Understanding the Limitations of CNN-Based Absolute Camera Pose Regression" (Sattler et al., 2019) deals with visual localization. It creates a theoretical model to analyze absolute pose regression (APR) methods, highlighting their limitations and similarities to image retrieval techniques. The study found that current APR approaches do not consistently outperform handcrafted retrieval baselines, highlighting the need for additional research before APR algorithms can compete effectively with accurate 3D structure-based methods in practical applications. Regarding the scope of this review, it only refers to some visual localization applications in AR.

To get more detailed information, a content analysis through data mining was also performed with the assistance of ElasticSearch.

## 4. Content analysis

The content analysis performed in the articles retrieved from the previous step is mostly quantitative. Not to be mistaken with the approach followed in social sciences, where a set of categories and sub-categories are inferred from content analysis (Bardin, 2013). Instead, all the PDF files were indexed in ElasticSearch and the following set of queries were performed: cnn, resnet, vgg, gan, svm, lstm, rnn, autoencoder, transformers, reinforcementlear*, restrictedboltzmannmachine, deepreinforcementlear*, som, logisticreg*, nearestnei*, randomfore* and python, tensorflow, pytorch, keras, matlab, java, scala.

The purpose of these is to try to assess the algorithms and techniques of machine learning. Fig. 9, shows the number of papers where DL and ML algorithms were mentioned.

It is clear that DL algorithms have an explicit domain, with only SVM going above some of the most common DL methods, ranking in the top-5. Aiming at understanding which languages/frameworks are related to the topics of this review, Fig. 10 illustrates its ranking.

Python and Tensorflow show their importance in the application of the topics of this review, being the most popular programming language and frameworks respectively.

## 5. Discussion

CV is a field of study that deals with enabling machines to perceive, analyze and understand visual data from the world around us. Virtual reality (VR) and augmented reality (AR) have made significant strides in recent years, and CV plays a crucial role in enhancing the immersive experience.

The findings of this bibliometric review highlight the various environments in which CV algorithms are used. The most common applications involve object detection, object recognition, and learning systems. Particularly in educational and training contexts, learning systems that make use of CV techniques have grown in popularity, improving user engagement and knowledge acquisition. Real-time visual data processing by CV algorithms helps to create dynamic and engaging learning environments. The seamless integration of virtual and physical elements depends on object detection and recognition, which is made possible by CV algorithms. This is particularly clear in applications for AR and VR, where real-world objects are recognized and tracked to create interactive and immersive digital overlays. The use of DL techniques in these applications shows the accuracy and productivity-boosting potential of neural networks.

The fact that research in this area is conducted all over the world emphasizes how crucial interdisciplinary cooperation is. Collaboration between researchers from various nations speeds up the creation of novel solutions, improves the robustness of research findings, and facilitates knowledge exchange. Additionally, the blending of insights from CV with those from other fields, such as psychology, human–computer interaction, and the arts, has the potential to reveal fresh frameworks for immersive experiences that have a lasting impact on users. In terms of authorship, it was found that researchers from North America and China dominate the field, with China being the leading country in terms of the number of publications. This trend is not surprising given that these regions are home to much of the world's rapidly developing AI research in recent years.

The results of this bibliometric study show that the number of publications on these topics has been steadily increasing over the past decade, which reflects the growing interest and potential of these technologies in several fields. The majority of the publications are conference papers, which suggests that the research is

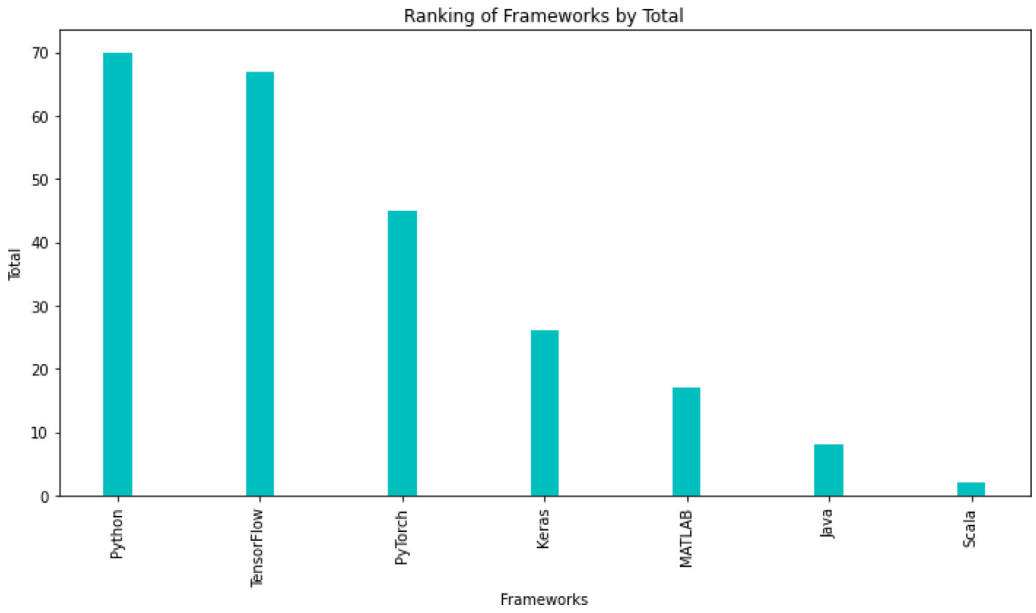


**Fig. 10.** Frequency of programming languages and frameworks.

still very active, and there is much room for further development and refinement. The review by Rojas et al. (Rojas-Sánchez et al., 2023) is the most similar to our study, though it focuses on learning systems that use VR technologies. Instead, our research focuses on understanding the applications of digital technologies (AR, VR, and XR) through CV algorithms, and it was discovered that most of these studies are centered on learning systems.

The content analysis shed light on the frequency and relative importance of DL and ML algorithms mentioned in the papers. DL methods are more frequently mentioned, of the analyzed algorithms, with only Support Vector Machine (SVM) securing a spot in the top-5. This observation highlights the explicit domain of DL algorithms in the context of the papers under consideration. Furthermore, the analysis investigates the relationship between programming languages and frameworks in relation to the topics under consideration. Python, the most popular programming language, emerges as a critical tool for applying the discussed topics, while Tensorflow stands out as a prominent framework. These findings emphasize the significance of Python and Tensorflow in facilitating the implementation and development of DL and ML algorithms.

The results of this study advance CV research more generally as well as the development of immersive technologies. Pushing the limits of object recognition, image segmentation, and scene understanding is tested on testbeds such as the applications of CV algorithms in AR, VR, MR, and XR environments. The knowledge gained from these applications may be applied to conventional CV research, leading to improved methodologies and algorithms that help a variety of industries.

## 6. Conclusions

In conclusion, this study provides a comprehensive overview of the current state of research on the use of CV algorithms in immersive technologies. It highlights the potential applications of these technologies, identifies the most relevant research themes, and provides insights into the geographic distribution of research and the main contributors to the field. The findings of this study can guide future research and development of CV algorithms for immersive technologies, ultimately leading to enhanced user experiences and a better understanding of the potential of these technologies.

The findings of this bibliometric analysis reveal a growing interest in the application of CV algorithms in AR, VR, MR, and XR environments. The number of publications in this field has steadily increased over the past decade, indicating the potential and significance of these technologies in various domains. The analysis highlights the role of CV algorithms in enabling immersive experiences, such as image processing, object detection, object recognition and neural networks. These algorithms facilitate the integration of virtual and physical elements, creating immersive digital experiences.

This paper also illuminates the global research landscape in this field. North American and Chinese authors produce a large amount of research, with China emerging as the country with the most publications. Collaboration between countries has been critical in driving research and advancing the field, particularly between the United States and China. DL algorithms appear typically in the publications examined, with DL methods dominating the rankings. In the context of these technologies, Python and TensorFlow have been identified as critical tools for implementing and developing DL and ML algorithms.

Overall, this bibliometric analysis provides valuable insights into the current state of research on CV algorithms in immersive technologies. The identified trends, gaps, and opportunities can guide future research directions and foster advancements in this exciting field.

## CRediT authorship contribution statement

**Júlio Castro Lopes:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Software, Visualization, Writing – original draft, Writing – review & editing. **Rui Pedro Lopes:** Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – review & editing.

## Ethical approval

This study does not contain any studies with Human or animal subjects performed by any of the authors.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgments

This work is funded by the European Regional Development Fund (ERDF) through the Regional Operational Program North 2020, within the scope of Project GreenHealth - Digital strategies in biological assets to improve well-being and promote green health, Norte-01-0145-FEDER-000042.

This work has been supported by FCT – Fundação para a Ciência e Tecnologia within the Project Scope: UIDB/05757/2020.

## References

Ababsa, F., 2020. Augmented reality application in manufacturing industry: Maintenance and non-destructive testing (NDT) use cases. In: De Paolis, L.T., Bourdot, P. (Eds.), Augmented Reality, Virtual Reality, and Computer Graphics. In: Lecture Notes in Computer Science, Springer International Publishing, Cham, pp. 333–344. http://dx.doi.org/10.1007/978-3-030-58468-9_24.

Abu Alhaija, H., Mustikovela, S.K., Mescheder, L., Geiger, A., Rother, C., 2018. Augmented reality meets computer vision: Efficient data generation for urban driving scenes. Int. J. Comput. Vis. 126 (9), 961–972. http://dx.doi.org/10.1007/s11263-018-1070-x.

Al-Khoury, A., Hussein, S.A., Abdulwhab, M., Aljuboori, Z.M., Haddad, H., Ali, M.A., Abed, I.A., Flayyih, H.H., 2022. Intellectual capital history and trends: A bibliometric analysis using scopus database. Sustainability 14 (18), 11615. http://dx.doi.org/10.3390/su141811615, URL: https://www.mdpi.com/2071-1050/14/18/11615, Number: 18 Publisher: Multidisciplinary Digital Publishing Institute.

Amparore, D., Checcucci, E., Piazzolla, P., Piramide, F., De Cillis, S., Piana, A., Verri, P., Manfredi, M., Fiori, C., Vezzetti, E., Porpiglia, F., 2022. Indocyanine green drives computer vision based 3D augmented reality robot assisted partial nephrectomy: The beginning of "automatic" overlapping era. Urology 164, e312–e316. http://dx.doi.org/10.1016/j.urology.2021.10.053.

Aria, M., Cuccurullo, C., 2017. bibliometrix: An R-tool for comprehensive science mapping analysis. J. Informetr. 11 (4), 959–975. http://dx.doi.org/10.1016/j.joi.2017.08.007, URL: https://www.sciencedirect.com/science/article/pii/S1751157717300500.

Bardin, L., 2013. L'analyse de contenu, 2e édition "Quadrige Quadrige. Manuels, Presses universitaires de France, Paris.

Berton, A., Longo, U.G., Candela, V., Fioravanti, S., Giannone, L., Arcangeli, V., Alciati, V., Berton, C., Facchinetti, G., Marchetti, A., Schena, E., De Marinis, M.G., Denaro, V., 2020. Virtual reality, augmented reality, gamification, and telerehabilitation: Psychological impact on orthopedic patients' rehabilitation. J. Clin. Med. 9 (8), 2567. http://dx.doi.org/10.3390/jcm9082567, URL: https://www.mdpi.com/2077-0383/9/8/2567, Number: 8 Publisher: Multidisciplinary Digital Publishing Institute.

Blanco-Pons, S., Carrión-Ruiz, B., Luis Lerma, J., Villaverde, V., 2019. Design and implementation of an augmented reality application for rock art visualization in Cova dels Cavalls (Spain). J. Cult. Herit. 39, 177–185. http://dx.doi.org/10.1016/j.culher.2019.03.014, URL: https://www.sciencedirect.com/science/article/pii/S1296207418305697.

Botto, C., Cannavò, A., Cappuccio, D., Morat, G., Sarvestani, A.N., Ricci, P., Demarchi, V., Saturnino, A., 2020. Augmented reality for the manufacturing industry: The case of an assembly assistant. In: 2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops. VRW, pp. 299–304. http://dx.doi.org/10.1109/VRW50115.2020.00068.

Chi, H.-g., Ha, M.H., Chi, S., Lee, S.W., Huang, Q., Ramani, K., 2022. InfoGCN: Representation learning for human skeleton-based action recognition. pp. 20186–20196, URL: https://openaccess.thecvf.com/content/CVPR2022/html/Chi_InfoGCN_Representation_Learning_for_Human_Skeleton-Based_Action_Recognition_CVPR_2022_paper.html.

Chrysafiadi, K., Kamitsios, M., Virvou, M., 2023. Fuzzy-based dynamic difficulty adjustment of an educational 3D-game. Multimedia Tools Appl. http://dx.doi.org/10.1007/s11042-023-14515-w.

Cīrulis, A., Brigmanis-Briģis, K., Zvejnieks, G., 2020. Analysis of suitable natural feature computer vision algorithms for augmented reality services. Balt. J. Mod. Comput. 8 (1), http://dx.doi.org/10.22364/bjmc.2020.8.1.10, URL: http://www.bjmc.lu.lv/fileadmin/user_upload/lu_portal/projekti/bjmc/Contents/8_1_10_Cirulis.pdf.

Cozzolino, V., Tonetto, L., Mohan, N., Ding, A., Ott, J., 2022. Nimbus: Towards latency-energy efficient task offloading for AR services. IEEE Trans. Cloud Comput. http://dx.doi.org/10.1109/TCC.2022.3146615.

Cruz, E., Orts-Escolano, S., Gomez-Donoso, F., Rizo, C., Rangel, J.C., Mora, H., Cazorla, M., 2019. An augmented reality application for improving shopping experience in large retail stores. Virtual Real. 23 (3), 281–291. http://dx.doi.org/10.1007/s10055-018-0338-3.

Deshpande, V., Sanghavi, J., 2019. Augmented reality: Technology merging computer vision and image processing by experimental techniques. Int. J. Innov. Technol. Explor. Eng. 8, 534–537.

Dini, C., Pereira, M.M.A., Souza, J.G.S., de Avila, E.D., Barão, V.A.R., 2022. Mapping the trends and impact of research collaboration between countries in oral implantology publications: A bibliometric analysis from 1999 to 2019. J. Prosthet. Dent. http://dx.doi.org/10.1016/j.prosdent.2022.10.009, URL: https://www.sciencedirect.com/science/article/pii/S0022391322006539.

Epishkina, A., Zapechnikov, S., 2016. Towards Approximation of Human's Perceptive Space on Photos, Videos and 3D Scenes, vol. 88, pp. 324–329. http://dx.doi.org/10.1016/j.procs.2016.07.443, ISSN: 1877-0509.

Francies, M.L., Ata, M.M., Mohamed, M.A., 2022. A robust multiclass 3D object recognition based on modern YOLO deep learning algorithms. Concurr. Comput.: Pract. Exper. 34 (1), e6517. http://dx.doi.org/10.1002/cpe.6517, URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/cpe.6517, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cpe.6517.

Gaidon, A., Wang, Q., Cabon, Y., Vig, E., 2016. VirtualWorlds as proxy for multi-object tracking analysis. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition. CVPR, pp. 4340–4349. http://dx.doi.org/10.1109/CVPR.2016.470, ISSN: 1063-6919.

Garcia-Garcia, A., Orts-Escolano, S., Oprea, S., Villena-Martinez, V., Martinez-Gonzalez, P., Garcia-Rodriguez, J., 2018. A survey on deep learning techniques for image and video semantic segmentation. Appl. Soft Comput. 70, 41–65. http://dx.doi.org/10.1016/j.asoc.2018.05.018, URL: https://www.sciencedirect.com/science/article/pii/S1568494618302813.

Ghasemi, Y., Jeong, H., Choi, S.H., Park, K.-B., Lee, J.Y., 2022. Deep learning-based object detection in augmented reality: A systematic review. Comput. Ind. 139, 103661. http://dx.doi.org/10.1016/j.compind.2022.103661, URL: https://www.sciencedirect.com/science/article/pii/S0166361522000586.

González Izard, S., Juanes Méndez, J.A., Ruisoto Palomera, P., García-Peñalvo, F.J., 2019. Applications of virtual and augmented reality in biomedical imaging. J. Med. Syst. 43 (4), 102. http://dx.doi.org/10.1007/s10916-019-1239-z.

Hatamizadeh, A., Tang, Y., Nath, V., Yang, D., Myronenko, A., Landman, B., Roth, H.R., Xu, D., 2022. UNETR: Transformers for 3D medical image segmentation. pp. 574–584, URL: https://openaccess.thecvf.com/content/WACV2022/html/Hatamizadeh_UNETR_Transformers_for_3D_Medical_Image_Segmentation_WACV_2022_paper.html.

Hbali, Y., Sadgal, M., El Fazziki, A., 2013. Object detection based on HOG features: Faces and dual-eyes augmented reality. In: 2013 World Congress on Computer and Information Technology. WCCIT, pp. 1–4. http://dx.doi.org/10.1109/WCCIT.2013.6618716.

Hou, L., Pan, Y., Zhu, J.J.H., 2021. Impact of scientific, economic, geopolitical, and cultural factors on international research collaboration. J. Informetr. 15 (3), 101194. http://dx.doi.org/10.1016/j.joi.2021.101194, URL: https://www.sciencedirect.com/science/article/pii/S1751157721000651.

Hung, S.-W., Chang, C.-W., Ma, Y.-C., 2021. A new reality: Exploring continuance intention to use mobile augmented reality for entertainment purposes. Technol. Soc. 67, 101757. http://dx.doi.org/10.1016/j.techsoc.2021.101757, URL: https://www.sciencedirect.com/science/article/pii/S0160791X21002323.

Johnson-Roberson, M., Barto, C., Mehta, R., Sridhar, S.N., Rosaen, K., Vasudevan, R., 2017. Driving in the matrix: Can virtual worlds replace human-generated annotations for real world tasks? In: 2017 IEEE International Conference on Robotics and Automation. ICRA, pp. 746–753. http://dx.doi.org/10.1109/ICRA.2017.7989092.

Kamikokuryo, K., Haga, T., Venture, G., Hernandez, V., 2022. Adversarial autoencoder and multi-armed bandit for dynamic difficulty adjustment in immersive virtual reality for rehabilitation: Application to hand movement. Sensors 22 (12), 4499. http://dx.doi.org/10.3390/s22124499, URL: https://www.mdpi.com/1424-8220/22/12/4499, Number: 12 Publisher: Multidisciplinary Digital Publishing Institute.

Kuc, R., Rogozinski, M., 2013. Elasticsearch Server. Packt Publishing Ltd, Google-Books-ID: PEFK3MuwBsIC.

Lancho Barrantes, B.S., Guerrero Bote, V.P., Rodríguez, Z.C., de Moya Anegón, F., 2012. Citation flows in the zones of influence of scientific collaborations. J. Am. Soc. Inf. Sci. Technol. 63 (3), 481–489. http://dx.doi.org/10.1002/asi.21682, URL: https://onlinelibrary.wiley.com/doi/abs/10.1002/asi.21682, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/asi.21682.

Li, X., Wira, M., Thawonmas, R., 2022. Toward dynamic difficulty adjustment with audio cues by Gaussian process regression in a first-person shooter. In: Göbl, B., van der Spek, E., Baalsrud Hauge, J., McCall, R. (Eds.), Entertainment Computing. ICEC 2022, In: Lecture Notes in Computer Science, Springer International Publishing, Cham, pp. 154–161. http://dx.doi.org/10.1007/978-3-031-20212-4_12.

Lin, C.-H., Chung, Y., Chou, B.-Y., Chen, H.-Y., Tsai, C.-Y., 2018. A novel campus navigation APP with augmented reality and deep learning. ISBN: 978-1-5386-4342-6, pp. 1075–1077. http://dx.doi.org/10.1109/ICASI.2018.8394464.

Liu, W., Lai, B., Wang, C., Cai, G., Su, Y., Bian, X., Li, Y., Chen, S., Li, J., 2021. Ground camera image and large-scale 3-D image-based point cloud registration based on learning domain invariant feature descriptors. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 14, 997–1009. http://dx.doi.org/10.1109/JSTARS.2020.3035359.

Lopes, R.P., Barroso, B., Deusdado, L., Novo, A., Guimarães, M., Teixeira, J.P., Leitão, P., 2021. Digital technologies for innovative mental health rehabilitation. Electronics 10 (18), 2260. http://dx.doi.org/10.3390/electronics10182260, URL: https://www.mdpi.com/2079-9292/10/18/2260, Number: 18 Publisher: Multidisciplinary Digital Publishing Institute.

Lopes, J.C., Lopes, R.P., 2022. A review of dynamic difficulty adjustment methods for serious games. In: Pereira, A.I., Košir, A., Fernandes, F.P., Pacheco, M.F., Teixeira, J.a.P., Lopes, R.P. (Eds.), Optimization, Learning Algorithms and Applications. In: Communications in Computer and Information Science, Springer International Publishing, Cham, pp. 144–159. http://dx.doi.org/10.1007/978-3-031-23236-7_11.

Martinez, J., Black, M.J., Romero, J., 2017. On human motion prediction using recurrent neural networks. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition. CVPR, pp. 4674–4683. http://dx.doi.org/10.1109/CVPR.2017.497, ISSN: 1063-6919.

Minaee, S., Boykov, Y., Porikli, F., Plaza, A., Kehtarnavaz, N., Terzopoulos, D., 2022. Image segmentation using deep learning: A survey. IEEE Trans. Pattern Anal. Mach. Intell. 44 (7), 3523–3542. http://dx.doi.org/10.1109/TPAMI.2021.3059968, Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence.

Nalbant, K.G., Uyanik, Ş., 2021. Computer vision in the metaverse. J. Metaverse 1 (1), 9–12, URL: https://dergipark.org.tr/en/pub/jmv/issue/67581/1051377, Number: 1 Publisher: İzmir Academy Association.

Patil, V., Narayan, J., Sandhu, K., Dwivedy, S.K., 2022. Integration of virtual reality and augmented reality in physical rehabilitation: A state-of-the-art review. In: Subburaj, K., Sandhu, K., Ćuković, S.s. (Eds.), Revolutions in Product Design for Healthcare: Advances in Product Design and Design Methods for Healthcare. In: Design Science and Innovation, Springer, Singapore, pp. 177–205. http://dx.doi.org/10.1007/978-981-16-9455-4_10.

Poonja, H.A., Shirazi, M.A., Khan, M.J., Javed, K., 2023. Engagement detection and enhancement for STEM education through computer vision, augmented reality, and haptics. Image Vis. Comput. 136, 104720. http://dx.doi.org/10.1016/j.imavis.2023.104720, URL: https://www.sciencedirect.com/science/article/pii/S026288562300094X.

Postolache, O., Monge, J., Alexandre, R., Geman, O., Jin, Y., Postolache, G., 2021. Virtual reality and augmented reality technologies for smart physical rehabilitation. In: Kanoun, O., Derbel, N. (Eds.), Advanced Systems for Biomedical Applications. In: Smart Sensors, Measurement and Instrumentation, Springer International Publishing, Cham, pp. 155–180. http://dx.doi.org/10.1007/978-3-030-71221-1_8.

Rebbani, Z., Azougagh, D., Bahatti, L., Bouattane, O., 2021. Definitions and applications of augmented/virtual reality: A survey. Int. J. Emerg. Trends Eng. Res. 9, 279–285. http://dx.doi.org/10.30534/ijeter/2021/21932021.

Rodrigues, A.S.F., Lopes, J.C., Lopes, R.P., Teixeira, L.F., 2022. Classification of facial expressions under partial occlusion for VR games. In: Pereira, A.I., Koˇsir, A., Fernandes, F.P., Pacheco, M.F., Teixeira, J.a.P., Lopes, R.P. (Eds.), Optimization, Learning Algorithms and Applications. In: Communications in Computer and Information Science, Springer International Publishing, Cham, pp. 804–819. http://dx.doi.org/10.1007/978-3-031-23236-7_55.

Rojas-Sánchez, M.A., Palos-Sánchez, P.R., Folgado-Fernández, J.A., 2023. Systematic literature review and bibliometric analysis on virtual reality and education. Educ. Inf. Technol. 28 (1), 155–192. http://dx.doi.org/10.1007/s10639-022-11167-5.

Saint-Louis, C., Hamam, A., 2021. Survey of haptic technology and entertainment applications. In: SoutheastCon 2021. pp. 01–07. http://dx.doi.org/10.1109/SoutheastCon45413.2021.9401939, ISSN: 1558-058X.

Sala, N., 2021. Virtual reality, augmented reality, and mixed reality in education: A brief overview. ISBN: 978-1-79984-961-2, http://dx.doi.org/10.4018/978-1-7998-4960-5.ch003.

Samir, M., Hanie, A., Aboulgheit, A., Hossam, K., Atia, A., ElMasry, N., 2018. The wanderer: Implementing markerless augmented reality with object position awareness. In: 2018 First International Workshop on Deep and Representation Learning. IWDRL, pp. 31–40. http://dx.doi.org/10.1109/IWDRL.2018.8358212.

Saputra, M.R.U., Markham, A., Trigoni, N., 2018. Visual SLAM and structure from motion in dynamic environments: A survey. ACM Comput. Surv. 51 (2), 37:1–37:36. http://dx.doi.org/10.1145/3177853.

Sarlin, P.-E., Cadena, C., Siegwart, R., Dymczyk, M., 2019. From coarse to fine: Robust hierarchical localization at large scale. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR, pp. 12708–12717. http://dx.doi.org/10.1109/CVPR.2019.01300, ISSN: 2575-7075.

Sattler, T., Zhou, Q., Pollefeys, M., Leal-Taixé, L., 2019. Understanding the limitations of CNN-based absolute camera pose regression. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition. CVPR, pp. 3297–3307. http://dx.doi.org/10.1109/CVPR.2019.00342, ISSN: 2575-7075.

Sorokin, M., Zhdanov, D., Zhdanov, A., 2020. Restoration of Lighting Sources for Virtual Reality Systems Using Convolutional Neural Networks, Computer Vision Algorithms, and Disparity Maps, vol. 11550, http://dx.doi.org/10.1117/12.2575396, ISSN: 0277-786X.

Tao, Y., Coltey, E., Wang, T., Alonso, M., Shyu, M.-L., Chen, S.-C., Alhaffar, H., Elias, A., Bogosian, B., Vassigh, S., 2020. Confidence estimation using machine learning in immersive learning environments. In: 2020 IEEE Conference on Multimedia Information Processing and Retrieval. MIPR, pp. 247–252. http://dx.doi.org/10.1109/MIPR49039.2020.00058.

Wang, H., Sridhar, S., Huang, J., Valentin, J., Song, S., Guibas, L.J., 2019. Normalized object coordinate space for category-level 6D object pose and size estimation. http://dx.doi.org/10.48550/arXiv.1901.02970, arXiv:1901.02970 [cs].

Wang, Q., Wang, Z., Li, B., Wei, D., 2021. An improved YOLOv3 object detection network for mobile augmented reality. In: 2021 IEEE 7th International Conference on Virtual Reality. ICVR, pp. 332–339. http://dx.doi.org/10.1109/ICVR51878.2021.9483829, ISSN: 2331-9569.

Zafar, S., Zachar, J.J., 2020. Evaluation of HoloHuman augmented reality application as a novel educational tool in dentistry. Eur. J. Dental Ed. 24 (2), 259–265. http://dx.doi.org/10.1111/eje.12492, URL: https://onlinelibrary.wiley.com/doi/abs/10.1111/eje.12492, _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/eje.12492.

Zeng, D., Lin, Z., Yan, X., Liu, Y., Wang, F., Tang, B., 2022. Face2Exp: Combating data biases for facial expression recognition. pp. 20291–20300, URL: https://openaccess.thecvf.com/content/CVPR2022/html/Zeng_Face2Exp_Combating_Data_Biases_for_Facial_Expression_Recognition_CVPR_2022_paper.html.

Zhang, G., 2020. Design of virtual reality augmented reality mobile platform and game user behavior monitoring using deep learning. Int. J. Electr. Eng. Educ. 0020720920931079. http://dx.doi.org/10.1177/0020720920931079, Publisher: SAGE Publications Ltd STM.

Zhang, H., Cui, Y., Shan, H., Qu, Z., Zhang, W., Tu, L., Wang, Y., 2020. Hotspots and trends of virtual reality, augmented reality and mixed reality in education field. In: 2020 6th International Conference of the Immersive Learning Research Network. ILRN, pp. 215–219. http://dx.doi.org/10.23919/iLRN47897.2020.9155170.

Zhao, G., Liu, S., Zhu, W.-J., Qi, Y.-H., 2021. A lightweight mobile outdoor augmented reality method using deep learning and knowledge modeling for scene perception to improve learning experience. Int. J. Hum.-Comput. Interact. 37 (9), 884–901. http://dx.doi.org/10.1080/10447318.2020.1848163, Publisher: Taylor & Francis.

Zhao, G., Zhang, Q., Chu, J., Li, Y., Liu, S., Lin, L., 2018. Augmented reality application for plant learning. In: 2018 IEEE 9th International Conference on Software Engineering and Service Science. ICSESS, pp. 1108–1111. http://dx.doi.org/10.1109/ICSESS.2018.8663953, ISSN: 2327-0594.

Zhou, X., Liu, Y., Lian, X., Guo, H., Lee, S., 2022. Fast 3D Reconstruction of Ancient Buildings Based on Virtual Reality and Convolution Network, vol. 12176, http://dx.doi.org/10.1117/12.2636462, ISSN: 0277-786X.