www.ijecs.in
International Journal Of Engineering And Computer Science
Volume 12 Issue 11 November 2023, Page No. 25900-25907
ISSN: 2319-7242 DOI: 10.18535/ijecs/v12i11.4770

Computer Vision and AI: Bridging the Gap between Perception and Understanding in Machines

Chandrima Sarkar[1], Soumargha Kar[2], Anirban Bhar[3], Suman Kumar Bhattacharyya[4]

[1,2] B. Tech student, Department of Information Technology, Narula Institute of Technology, Kolkata, India.
[3,4] Assistant Professor, Department of Information Technology, Narula Institute of Technology, Kolkata, India.

Abstract
The study of giving computers the ability to "see" is known as "computer vision." The overarching goal of computer vision problems is to draw some kind of conclusion about the real world from raw data about images. It spans several fields and disciplines, but can be roughly placed in the realm of AI and ML, where both specialized and more generalized learning approaches find application. Because it draws from many different branches of engineering and computer science, interdisciplinary research can sometimes give the impression of being chaotic. One vision problem might be easily resolved with a custom-built statistical approach, while another might require a complex and extensive collection of off-the-shelf machine learning techniques. The field of computer vision is at the forefront of modern research. Computer Vision and AI work together to enable machines to see and understand the visual world. Computer Vision focuses on developing algorithms and techniques that allow computers to extract meaningful information from images or videos. Conversely, artificial intelligence entails the development of smart machines with enhanced cognitive abilities. By combining these fields, we can create powerful systems that can recognize objects, understand scenes, and even interpret human emotions. Computer Vision and AI have the potential to revolutionize industries like healthcare, transportation, and entertainment, making our lives more convenient, efficient, and enjoyable.

Keywords: Computer Vision, Artificial Intelligence, Neural Network, Deep Learning, Machine Learning.

1. Introduction

It is the goal of computer vision, a subfield of AI, to enable computers and other systems to infer relevant information from digital photos, videos, and other visual inputs, and then act upon or recommend that data. Just as AI lets machines reason, computer vision helps computers perceive, observe, and comprehend.

Similar to human vision, computer vision has its limitations, especially when compared to the human starting point. Human vision has the advantage of being trained over many years to distinguish between objects, determine their distance, detect motion, and detect anomalies in an image. Using cameras, data, and algorithms in place of retinas, optic nerves, and a visual cortex, computer vision educates robots to carry out these tasks in a significantly shorter amount of time. A system designed to check items or monitor a manufacturing asset can evaluate thousands of products or processes every minute, allowing it to fast outpace human abilities by picking up on even the smallest faults or difficulties. As a result of its usefulness in fields as diverse as medicine, entertainment, manufacturing, and autonomous vehicles, computer vision has recently become increasingly mainstream. Visual recognition tasks including image ordering,

constraining, and identifying are essential to many of these applications. Convolutional neural networks (CNNs) have recently advanced, exhibiting their power with amazing results in best-in-class image recognition assignments and frameworks. The result is that in computer vision, convolutional neural networks (CNNs) have become the fundamental building blocks of deep learning computations.

Deep Neural Networks (DNN) are a type of neural network that is commonly used in calculations involving computer vision because of their superior picture identification abilities. Visual sign decoding typically makes use of Convolutional Neural Networks (CNN or ConvNet), a subtype of Deep Neural Networks (DNNs). It's also utilized in the fields of computer vision and natural language processing (NLP) to help structure data. A convolutional neural network can be built from a variety of building elements. In this post, we will briefly go over these building pieces, which comprise convolution layers, pooling layers, and fully linked layers. The author then moves on to discuss Deep Learning and other neural network methods. In addition, the book covers Convolutional Neural Networks, their development, and their applications in numerous sectors, including medicine and engineering.

Using image identification and object detection technologies, deep learning is a technique for realizing computer vision. Since its inception, computer vision has undergone rapid evolution thanks to the advent of deep learning, greatly improving the precision with which images can be recognized. In addition, an expert system can mimic and recreate the reasoning and decision-making process carried out by human experts. In a way that was impossible with traditional expert systems, machine learning—specifically deep learning—has made it possible to "acquire the tacit knowledge of experts." Using massive data and multiple measurements of a phenomenon, machine learning "systematises tacit knowledge." In this article, we describe some knowledge-based computer vision techniques that incorporate deep learning.

2. Literature Survey

"Computer vision" is one of the numerous subfields of research that go under the umbrella of "image processing," and it has been receiving a lot of attention [1–15]. It is both an academic field that studies the "realization of vision using computers" and an artificial intelligence (AI) field that enables computers and systems to derive meaningful information from digital images, videos, and other visual data and act and make recommendations (allowing humans to do so, as well) based on that information. It is a field that enables computers and systems to derive meaningful information from digital images, videos, and other visual data. If AI makes it possible for computers to think, then computer vision makes it possible for them to see, understand, and observe. This operates in a manner comparable to that of human eyesight and is anticipated to provide humans with an advantage. As a result, the objective of 'computer vision' is to equip computers with capabilities that are analogous to those of the human eye, or to actualize the concept of 'computer vision'. To be more specific, the objective is to develop software for computers that can perform similarly to or even better than human vision by making use of data derived from still images or videos. Up until this point, the most common form of image processing that computers have been capable of is known as computer graphics, or CG. Computer graphics (CG) is used for projecting and displaying 3D objects on a 2D display, but computer vision is used to derive 3D information from 2D picture data. This distinction is what distinguishes computer graphics from computer vision. Despite the fact that they represent two distinct technologies, they are complementary to one another and help to further the development of emerging technologies like virtual reality and augmented reality (AR). In augmented reality (AR), three-dimensional computer graphics (3DCG) are superimposed on top of real-world backgrounds, with additional data being supplied by a computer. It is an example of the combination of computer vision, which observes the real environment, with computer graphics, which represent a made-up world [16–20]. The human visual system has the advantage of gathering contextual knowledge, which enables it to differentiate between items,

determine the distance between objects, assess whether or not an object is moving, and determine whether or not something is wrong with an image [21,22]. Using cameras, data, and algorithms rather than the retina, optic nerve, and visual brain, computer vision is the process of teaching robots to do these functions at a rate that is far faster than that of humans. Because they can instantly analyze thousands of items and processes, systems that have been trained to inspect products and monitor production assets can swiftly surpass the capabilities of humans. This is because the systems are able to spot flaws that are imperceptible to the human eye. Because of this, computer vision requires enormous amounts of data, which are then analyzed in an iterative fashion until the computer is able to recognize features and, ultimately, an image. Deep learning, a type of machine learning, and convolutional neural networks (CNNs), another type of neural network, are the two primary technologies that are utilized to accomplish this goal [23–27].

3. Perception in Computer Vision

3.1. Visual information perceived by machine:

Computer vision is a multidisciplinary field that enables machines to interpret and understand visual information from the world, much like humans do. It involves the development of algorithms, models, and systems that enable computers to gain a high-level understanding of images or video. Here are some fundamental concepts and components of computer vision:

Image Formation: Images are composed of pixels (picture elements), which are the smallest units of an image. The number of pixels in an image determines its resolution, with higher resolution providing more detail.

Image Processing: Before analysis, images often undergo preprocessing steps such as resizing, normalization, and noise reduction. Techniques like convolution are used for filtering to enhance or extract specific features from an image.

Feature Extraction: Detection of edges is a common feature extraction technique to identify boundaries in an image. Corner detection helps identify distinctive points in an image. Extracting meaningful information from regions of interest, such as SIFT or SURF descriptors.

Object Recognition: Identifying and locating objects within an image or video. Object Classification: Assigning labels to objects based on their features.

Image Segmentation: Dividing an image into segments or regions to simplify the representation of an object or scene.

Depth Perception: Determining the distance of objects from the camera, often through techniques like stereo vision or depth sensing.

Motion Analysis: Tracking and understanding the movement of objects in images or videos.

Machine Learning in Computer Vision: Supervised Learning: Training models on labeled datasets for tasks like object recognition.

Unsupervised Learning: Discovering patterns and structures in data without explicit labels.

Deep Learning: Using neural networks, particularly convolutional neural networks (CNNs), for feature learning and representation.

Neural Networks and Deep Learning: Convolutional Neural Networks (CNNs) are especially popular in computer vision for tasks like image classification, object detection, and segmentation.

Understanding how machines perceive visual information involves grasping these fundamental concepts and techniques. The field is continually evolving, with ongoing research and advancements in technology contributing to the development of more robust and versatile computer vision systems.

3.2. Capturing visual data:

The role of sensors, cameras, and other data sources is crucial in capturing visual data for computer vision applications. These devices serve as the eyes of the computer system, providing the necessary input for analysis and interpretation. Here's a discussion of their roles:

Cameras: Cameras capture visual information in the form of images. They come in various types, including digital cameras, webcams, and specialized cameras for specific applications. The resolution of a camera determines the level of detail in the captured images. Higher resolution cameras provide more information but may also require more processing power. For video applications, the frame rate of a camera is crucial. Higher frame rates result in smoother videos and are essential for tasks like motion analysis and object tracking.

Depth Sensors: Cameras equipped with depth sensors, such as Time-of-Flight (ToF) or structured light sensors, can capture depth information along with color. This is valuable for tasks like 3D reconstruction and understanding the spatial arrangement of objects.

Infrared Sensors: Infrared sensors capture infrared radiation, which is invisible to the human eye. They find applications in night vision, thermal imaging, and other scenarios where detecting heat or radiation is important.

LiDAR (Light Detection and Ranging): LiDAR sensors use laser light to measure distances and create detailed, three-dimensional maps of the environment. LiDAR is commonly used in autonomous vehicles, robotics, and environmental monitoring.

RGB-D Cameras: These cameras combine traditional RGB (color) information with depth data. They provide a richer representation of the scene and are widely used in applications requiring both color and depth information.

Motion Sensors: Accelerometers and gyroscopes can provide information about the movement and orientation of a device. This data is useful for tasks such as image stabilization, gesture recognition, and tracking dynamic objects.

Microphones: While not visual sensors, microphones capture audio data, which can complement visual information. Audio-visual fusion is used in applications like speech recognition, lip reading, and context-aware computer vision.

Multispectral and Hyperspectral Sensors: These sensors capture information across multiple wavelengths, allowing for the analysis of spectral characteristics. They find applications in agriculture, environmental monitoring, and material identification.

Data Fusion: Integrating data from multiple sensors can enhance the overall understanding of a scene. Sensor fusion techniques combine information from different sources, improving accuracy and robustness.

Wireless and IoT Devices: Cameras and sensors integrated into Internet of Things (IoT) devices contribute to the growing field of edge computing, where data is processed locally before being sent to centralized servers. This is particularly useful for real-time applications and reducing latency.

In summary, sensors, cameras, and other data sources play a pivotal role in capturing diverse visual data for computer vision applications. The choice of sensors depends on the specific requirements of the task, such as the need for color information, depth perception, or environmental awareness. The continuous advancement of sensor technologies contributes to the ongoing progress in the field of computer vision.

4. Bridging the Gap by AI

4.1. HealthCare:

Through the examination of X-rays, magnetic resonance imaging (MRI), and several other types of medical images, computer vision plays a significant role in the process of illness diagnosis. It has been demonstrated

to be just as compelling as the conventional human specialists in the field when it comes to the correctness of their findings. Pneumonia, cerebrum tumors, diabetes, Parkinson's disease, and malignant uterine growth are just some of the conditions that may be accurately diagnosed by computer vision on a regular basis, and the technology is only becoming better. Early detection of any potentially harmful diseases will be possible thanks to cutting-edge picture processing technology and cutting-edge computer vision methodologies. In this way, treatment might be delivered at an inconvenient point throughout the course of the disease, or, failing that, the likelihood of their recurrence might be reduced.

4.2. Automobile:

Because it is designed to grasp the driving conditions, including spotting barriers, persons on footpaths, and possible accident ways, computer vision is becoming increasingly important to the automobile industry. This is particularly the case given the increased visibility of individuals driving their own vehicles. More and more companies are looking for innovative ways to put more electric vehicles onto the road, which has led to the gradual introduction of self-driving cars onto the market. The advancement of computer vision technology makes it possible for self-driving vehicles to "see" the earth, and artificial intelligence calculations make the "minds" that assist computer vision in translating the objects that are near the vehicle. The newest generation of self-driving cars are outfitted with a plethora of cameras, each of which can provide a full 360-degree view of the natural environment within a range of several meters. For example, the automaker Tesla equips its vehicles with something on the order of 8 all-encompassing cameras to achieve this goal. In addition to the cameras, a front-facing radar that enables the recognition of different vehicles even in the presence of precipitation or mist, as well as twelve ultrasonic sensors that can distinguish between hard and soft objects that may be found in the environment, have been included. A standard personal computer won't be adequate to handle the deluge of data that will be pushed into the vehicle because of the amount of information that will be urged into the vehicle. Because of this, every autonomous vehicle has a locally accessible personal computer equipped with computer vision features developed using artificial intelligence. It is the responsibility of the cameras and sensors to identify and collect protests that take place in natural settings, such as groups of people walking. Quick consideration must be given to the dimensions of the items being driven, including their area, thickness, shape, and depth, so that the rest of the driving framework can make the appropriate decisions. All of these computations are only made possible by the combination of AI and deep neural systems, which gives rise to features such as the recognition of a person walking.

4.3. Automobile:

Photon estimates, which are mostly made up of photographs of the universe, provide the foundation of our comprehensive knowledge of the cosmos. Because our universe is so massive, and because the one natural rule that governs our world predicts that the data acquired will be just as large, this paves the way for the possibility of employing computer vision in the field of astronomy. It is not possible for the stargazer or anybody else to physically comprehend this information in its entirety in its entirety in its entirety in its entirety. Because of the capabilities of computer vision, we can understand all of the data in a relatively short amount of time. To put it another way, computer vision is currently being used to locate new planets and large bodies. This technique is applied in applications such as imaging of exoplanets, the grouping of stars and cosmic systems, and other activities that are very similar.

4.4. Industrial:

Computer vision is used on the mechanical production systems in industries for monitoring groups, identifying damaged parts, and for the examination of finished products. This helps ensure quality control and reduces costs. In this context, Machine Vision equipment is utilized to assist in locating minute level defects in products, the likes of which are essentially indiscernible to the naked eye of a human observer.

Reading a product's scanning tag or QR code is essential for assembly tasks since these types of codes provide a unique identifier that may be used to track an item. The task of reviewing a large number of standardized identifications in a single day is not an easy one for people, but it is one that can be accomplished quickly and accurately in a matter of minutes with the help of computer vision.

5. Conclusion

In terms of both market size and industry adoption, computer vision and AI technologies are the fastest expanding fields. In particular, spatial computer vision and edge AI are employed not only for complex operations but also for the enhancement and automation of routine activities. This new reality, along with the decreasing cost of necessary technology and the advancements in depth perception and machine learning, has led to the creation of workable solutions in edge computer vision and artificial intelligence systems. Edge AI and spatial computer vision allow for the deployment of depth-based apps and the processing of images locally on a device. Important advancements in software and machine learning workflows are taking place as technology becomes more widely available. Tools that enable AI and computer vision to learn their own models have made them more accessible, but they remain highly specialized, and numerous technical challenges persist. Meanwhile, the widespread adoption and use of edge computing continues to be a challenge for traditional machine learning pipelines and workflows. Another major obstacle is minimizing the time and money needed to develop and enhance machine learning models for practical use. The problem is how to coordinate all these tools and set up a system for steady progress. Additional constraints, such as the need for an application to be lightweight and performant, must be taken into account when designing the final model that will be installed on a device.

References

1. Al-Oraiqat, A.M.; Smirnova, T.; Drieiev, O.; Smirnov, O.; Polishchuk, L.; Khan, S.; Hasan, Y.M.Y.; Amro, A.M.; AlRawashdeh, H.S. Method for Determining Treated Metal Surface Quality Using Computer Vision Technology. Sensors 2022, 22, 6223.
2. Gumbs, A.A.; Grasso, V.; Bourdel, N.; Croner, R.; Spolverato, G.; Frigerio, I.; Illanes, A.; Abu Hilal, M.; Park, A.; Elyan, E. The Advances in Computer Vision That Are Enabling More Autonomous Actions in Surgery: A Systematic Review of the Literature. Sensors 2022, 22, 4918.
3. Dudek, P.; Richardson, T.; Bose, L.; Carey, S.; Chen, J.; Greatwood, C.; Liu, Y.; Mayol-Cuevas,W. Sensor-level computer vision with pixel processor arrays for agile robots. Sci. Robot. 2022, 7, eabl7755.
4. Abellanas, M.; Elena, M.J.; Keane, P.A.; Balaskas, K.; Grewal, D.S.; Carreño, E. Artificial Intelligence and Imaging Processing in Optical Coherence Tomography and Digital Images in Uveitis. Ocul. Immunol. Inflamm. 2022, 30, 675–681.
5. Kitaguchi, D.; Takeshita, N.; Hasegawa, H.; Ito, M. Artificial intelligence-based computer vision in surgery: Recent advances and future perspectives. Ann. Gastroenterol. Surg. 2021, 6, 29–36.
6. Hellsten, T.; Karlsson, J.; Shamsuzzaman, M.; Pulkkis, G. The Potential of Computer Vision-Based Marker-Less Human Motion Analysis for Rehabilitation. Rehabil. Process Outcome 2021, 10, 11795727211022330.
7. Hassan, H.; Ren, Z.; Zhao, H.; Huang, S.; Li, D.; Xiang, S.; Kang, Y.; Chen, S.; Huang, B. Review and classification of AI-enabled COVID-19 CT imaging models based on computer vision tasks. Comput. Biol. Med. 2022, 141, 105123.
8. D'Antoni, F.; Russo, F.; Ambrosio, L.; Vollero, L.; Vadalà, G.; Merone, M.; Papalia, R.; Denaro, V. Artificial Intelligence and Computer Vision in Low Back Pain: A Systematic Review. Int. J. Environ.

Res. Public Health 2021, 18, 10909.
9. Wang, J.; Zhu, H.; Liu, J.; Li, H.; Han, Y.; Zhou, R.; Zhang, Y. The application of computer vision to visual prosthesis. Artif. Organs 2021, 45, 1141–1154.
10. Victória Matias, A.; Atkinson Amorim, J.G.; Buschetto Macarini, L.A.; Cerentini, A.; Casimiro Onofre, A.S.; De Miranda Onofre, F.B.; Daltoé, F.P.; Stemmer, M.R.; von Wangenheim, A. What is the state of the art of computer vision-assisted cytology? A Systematic Literature Review. Comput. Med. Imaging Graph 2021, 91, 101934.
11. Wu, Z.; Chen, Y.; Zhao, B.; Kang, X.; Ding, Y. Review of Weed Detection Methods Based on Computer Vision. Sensors 2021, 21, 3647.
12. Louis, C.M.; Erwin, A.; Handayani, N.; Polim, A.A.; Boediono, A.; Sini, I. Review of computer vision application in in vitro fertilization: The application of deep learning-based computer vision technology in the world of IVF. J. Assist. Reprod. Genet. 2021, 38, 1627–1639.
13. Kang, X.; Zhang, X.D.; Liu, G. A Review: Development of Computer Vision-Based Lameness Detection for Dairy Cows and Discussion of the Practical Applications. Sensors 2021, 21, 753.
14. Fernandes, A.F.A.; Dórea, J.R.R.; Rosa, G.J.M. Image Analysis and Computer Vision Applications in Animal Sciences: An Overview. Front. Vet. Sci. 2020, 7, 551269.
15. Patel, K.; Parmar, B. Assistive device using computer vision and image processing for visually impaired; review and current status. Disabil. Rehabil. Assist. Technol. 2022, 17, 290–297.
16. Minaee, S.; Liang, X.; Yan, S.Modern Augmented Reality: Applications, Trends, and Future Directions. arXiv 2022, arXiv:2202.09450.
17. Sutherland, J.; Belec, J.; Sheikh, A.; Chepelev, L.; Althobaity, W.; Chow, B.J.W.; Mitsouras, D.; Christensen, A.; Rybicki, F.J.; La Russa, D.J. Applying Modern Virtual and Augmented Reality Technologies to Medical Images and Models. J. Digit. Imaging 2019, 32, 38–53.
18. Lungu, A.J.; Swinkels, W.; Claesen, L.; Tu, P.; Egger, J.; Chen, X. A review on the applications of virtual reality, augmented reality and mixed reality in surgical simulation: An extension to different kinds of surgery. Expert. Rev. Med. Devices 2021, 18, 47–62.
19. Lex, J.R.; Koucheki, R.; Toor, J.; Backstein, D.J. Clinical applications of augmented reality in orthopaedic surgery: A comprehensive narrative review. Int. Orthop. 2022, in press.
20. Tanzer, M.; Laverdière, C.; Barimani, B.; Hart, A. Augmented Reality in Arthroplasty: An Overview of Clinical Applications, Benefits, and Limitations. J. Am. Acad. Orthop. Surg. 2022, 30, e760–e768.
21. Maier, M.; Blume, F.; Bideau, P.; Hellwich, O.; Abdel Rahman, R. Knowledge-augmented face perception: Prospects for the Bayesian brain-framework to align AI and human vision. Conscious Cogn. 2022, 101, 103301.
22. Fooken, J.; Kreyenmeier, P.; Spering, M. The role of eye movements in manual interception: A mini-review. Vision Res. 2021, 183, 81–90.
23. Statsenko, Y.; Habuza, T.; Talako, T.; Pazniak, M.; Likhorad, E.; Pazniak, A.; Beliakouski, P.; Gelovani, J.G.; Gorkom, K.N.; Almansoori, T.M.; et al. Deep Learning-Based Automatic Assessment of Lung Impairment in COVID-19 Pneumonia: Predicting Markers of HypoxiaWith Computer Vision. Front. Med. 2022, 9, 882190.
24. Balasubramanian, S.B.; Jagadeesh, K.R.; Prabu, P.; Venkatachalam, K.; Trojovský, P. Deep fake detection using cascaded deep sparse auto-encoder for effective feature selection. PeerJ Comput. Sci. 2022, 8, e1040.
25. Zhang, Y.; Zhang, S.; Li, Y.; Zhang, Y. Single- and Cross-Modality Near Duplicate Image Pairs Detection via Spatial Transformer Comparing CNN. Sensors 2021, 21, 255.
26. Xia, C.; Pan, Z.; Li, Y.; Chen, J.; Li, H. Vision-based melt pool monitoring for wire-arc additive manufacturing using deep learning method. Int. J. Adv. Manuf. Technol. 2022, 120, 551–562.

27. Li, W.; Zhang, L.; Wu, C.; Cui, Z.; Niu, C. A new lightweight deep neural network for surface scratch detection. Int. J. Adv. Manuf. Technol. 2022, 123, 1999–2015.