International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

Exploring Computer Vision: From Traditional Techniques to AI - Driven Innovations

Shubham Sarang[1], Flavia Gonsalves[2]

[1]Student, Institute of Computer Science, Mumbai Educational Trust - MET ICS, Mumbai University, India
Email: *mca22_1345ics[at]met.edu*

[2]Professor, Institute of Computer Science, Mumbai Educational Trust - MET ICS, Mumbai University, India
Email: *flaviag_ics[at]met. edu*

Abstract: ***This paper presents a thorough investigation into the expansive realm of computer vision, meticulously charting its progression from foundational traditional techniques to the forefront of AI - driven innovations. By meticulously dissecting the historical trajectory, current applications, and impending challenges of computer vision, this study unveils a rich tapestry of insights. Through critical analysis, the paper unveils the transformative influence of artificial intelligence in catapulting the capabilities of computer vision, heralding a new era of perception and understanding. Furthermore, this examination serves to illuminate prospective pathways for research and development, propelling the field towards unprecedented frontiers of discovery and innovation.***

Keywords: Computer vision, AI - driven innovations, Traditional Techniques, Applications

1. Introduction

History of Computer Vision

1.1 Early Establishments (1950s - 1960s):

The creation of computer vision can be followed back to the 1950s and 1960s when researchers began to explore methods for automatic pattern recognition and image analysis. Some early projects include the development of the "Perceptron" by Frank Rosenblatt in 1957, which laid the groundwork for neural network - based approaches to visual pattern recognition. The **perceptron** is a simple computational model inspired by the structure and function of biological neurons. It consists of an input layer, one or more layers of artificial neurons (also known as perceptrons), and an output layer. Each perceptron takes input signals, applies weights to them, the sum of the weighted input is passed to the result through an activation function to produce the relevant output.

1.2 Pattern Recognition and Image Processing (1960s - 1970s):

During the 1960s and 1970s, researchers focused on developing algorithms for basic image processing tasks such as edge detection, feature extraction, and template matching. The "Template Matching" technique, introduced by Peter Kovesi, involved comparing a template image with a target image to locate instances of the template within the target.

1.3 Early Applications and Systems (1970s - 1980s):

In the 1970s and 1980s, computer vision techniques began to be applied to practical problems in industries such as manufacturing, robotics, and medical imaging. Projects like the "PUMA Unimation" system developed by Joseph Engelberger in the 1980s demonstrated the use of computer vision for robotic assembly tasks, where robots could visually inspect and manipulate objects in a factory setting.

1.4 Advancements in Machine Learning (2000s - present):

The 2000s witnessed significant advancements in machine learning, particularly with the rise of deep learning techniques such as convolutional neural networks (CNNs). Breakthroughs in deep learning, combined with the availability of large datasets and powerful computing resources, revolutionized computer vision research and enabled unprecedented accuracy in tasks such as image classification, object detection, and semantic segmentation.

1.5 Real - World Applications and Integration (Present):

In recent years, computer vision technologies have found widespread adoption across diverse industries and domains, including healthcare, automotive, retail, security, and entertainment. Real - world applications of computer vision range from medical imaging analysis, autonomous vehicles to facial recognition systems and augmented reality experiences.

2. Traditional Techniques

2.1 Image Processing

Image processing techniques involve manipulating digital images to enhance their quality or extract useful information. This may include operations such as noise reduction, contrast enhancement, and image resizing. Common image processing algorithms include filters like Gaussian blur, median filter, and edge detection filters, which are used to highlight important features or suppress noise in images.

Some key aspects of image processing:

a) **Enhancement**: Image processing techniques are used to improve the quality of images, such as adjusting brightness, contrast, and color balance. This can be particularly important in fields like medical imaging, where clear and accurate visuals are essential.
b) **Preprocessing**: Before computer vision algorithms can analyze an image, it often needs to be preprocessed. This

can involve tasks like noise reduction, image resizing, and edge detection, which help to simplify the image and make it easier for computer vision systems to understand. r for computer vision systems to analyze and understand the content of the image.

c) **Compression**: Image processing techniques can be used to reduce the size of digital images without significantly affecting their quality. This is important for applications such as video streaming and image storage, where large image files consume significant amounts of bandwidth and storage space.

d) **Intersection over Union (IoU):** This equation is commonly used to evaluate the accuracy of bounding box predictions in object detection tasks. It can illustrate how well the predicted bounding boxes align with the ground truth annotations.

$$\text{IoU} = \frac{\text{Area of Overlap}}{\text{Area of Union}} \qquad \textit{Equation 1}$$

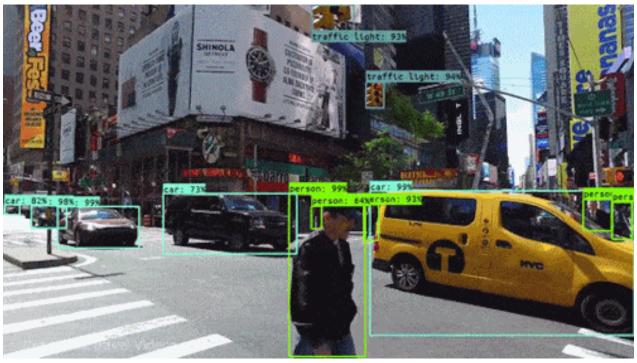

Figure 1: Image object detection and tagging

2.2 Feature Extraction

Feature extraction is the process of identifying and representing distinctive visual features in images that are relevant to the task at hand. These features could be edges, corners, blobs, or texture patterns. Traditional feature extraction methods include techniques like Harris corner detection, Scale - Invariant Feature Transform (SIFT), and Speeded Up Robust Features (SURF).

These methods aim to identify key points in an image that can be used for tasks such as object recognition and image matching. In feature extraction HOG (Histogram of Oriented Gradients) is used where Gradient Magnitude and Orientation equation is applied.

$$G = \sqrt{(G_x^2 + G_y^2)} \text{ and } \theta = \arctan\left(\frac{G_y}{G_x}\right)$$

Equation 2

Then Gx and Gy are the slants in the vertical (x) and perpendicular (y) directions, independently. They're reckoned using secondary pollutants like Sobel or Prewitt drivers.

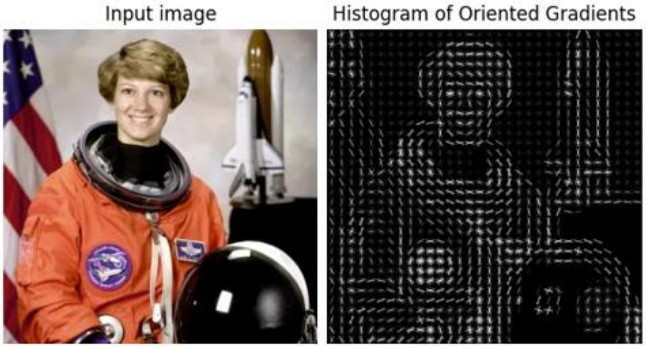

Figure 2: Feature extraction using HOG

2.3 Template Matching

Template matching involves comparing a small template image with regions of a larger target image to locate instances of the template within the target. This technique is used for object detection and recognition in images. The basic idea is to slide the template over the target image and compute a similarity measure at each position to determine the best match. Techniques like cross - correlation and normalized cross - correlation are often used for template matching.

2.4 Edge Detection

Edge detection is a fundamental operation in computer vision that aims to identify abrupt changes in intensity or color in an image, which typically correspond to object boundaries or other important features. Traditional edge detection algorithms include methods like the Sobel operator, Prewitt operator, and Canny edge detector. These algorithms compute gradients in the image to detect regions of high spatial intensity variation, which are indicative of edges.

Harris corner detection:

The Harris Corner Detector identifies important points in images that have significant intensity changes in multiple directions. It helps distinguish between corners and edges, making it a reliable method for corner detection. This algorithm is based on the gradients of the image in both x and y directions (Ix and Iy) will both be active in the corner region. By multiplying Ix2 and Iy2, we can identify regions on the image that have a change in both x and y directions at the same time, which is a characteristic of corners.

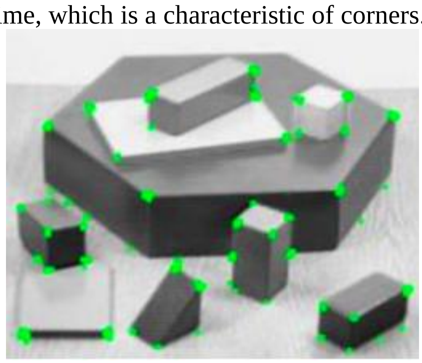

Figure 3: Edge detection

2.4 Segmentation

Based on certain criteria, such as color, texture, or intensity the image segmentation process partitions an image into multiple segments or regions. This is typically used to

Paper ID: SR24611174902 DOI: https://dx.doi.org/10.21275/SR24611174902

separate objects from the background or group similar pixels together. Traditional segmentation techniques include thresholding, region growing, and clustering algorithms like k - means clustering and mean shift clustering. These techniques aim to divide an image into coherent regions that can then be analyzed separately. In image classification the softmax function equation is used where P (y=j|X) is the probability of class j.

$$P(y=j \mid \mathbf{x}) = \frac{e^{\mathbf{x}^{\mathsf{T}}\mathbf{w}_j}}{\sum_{k=1}^{K} e^{\mathbf{x}^{\mathsf{T}}\mathbf{w}_k}}$$ *Equation 3*

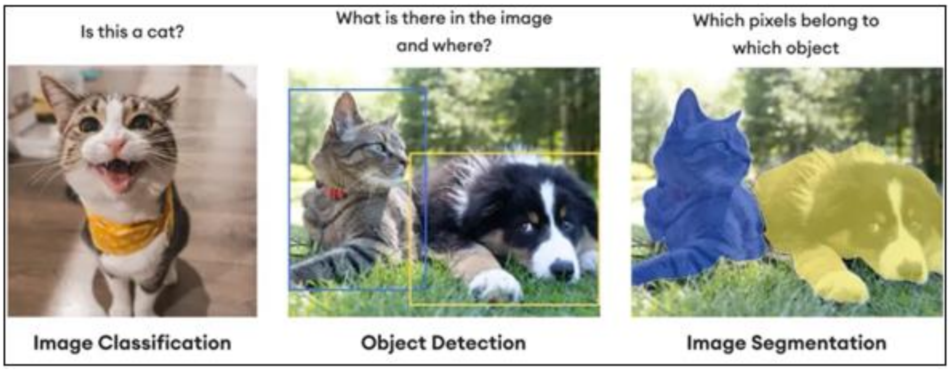

Figure 4: Classification, object detection and segmentation

3. AI driven Techniques

AI - driven innovations in computer vision utilize advanced machine learning techniques, particularly deep learning, to enable computers to learn directly from raw data without the need for explicit feature engineering.

3.1 Convolutional Neural Networks:

CNNs are a class of deep neural networks specifically designed for processing structured grid - like data, such as images. They consist of multiple layers of convolutional, pooling, and fully connected layers. CNNs have revolutionized computer vision by automatically learning hierarchical features directly from raw pixel data. Image classification, object detection, and semantic segmentation are the tasks they excel at.

Some key aspects of CNNs in computer vision:

- **Automatic feature extraction:** CNNs can automatically learn and extract meaningful features from images, reducing the need for manual feature engineering.
- **Hierarchical feature extraction:** CNNs learn features at different levels of abstraction, allowing them to capture both low - level details and high - level semantic information.
- **Robustness to image transformations:** CNNs are generally more robust to image transformations, such as rotation and scaling, compared to traditional computer vision algorithms.
- **Applicability to diverse fields:** CNNs have been successfully applied in various domains, including medicine, autonomous driving, and security systems.

3.2 Object Detection and Localization:

Object detection systems based on CNNs, such as Region - based CNNs (R - CNN), Faster R - CNN, and Single Shot MultiBox Detector (SSD), have significantly improved the accuracy and speed of object detection in images and videos. These systems can not only identify objects within an image but also localize them with bounding boxes, enabling applications like autonomous vehicles, surveillance systems, and augmented reality.

Localization: Localization in object detection refers to the process of precisely identifying the location of objects within an image or video frame. Object localization involves predicting the bounding box coordinates (typically represented as (x, y, w, h)) that tightly enclose the detected object within an image. Here the coordinates denote the position of the bounding box's top - left corner (x, y) (w) is width and (h) height.

3.3 Semantic Segmentation:

Semantic segmentation involves assigning a class label to each pixel in an image, thus segmenting the image into meaningful regions corresponding to different object categories or semantic concepts. Deep learning models like Fully Convolutional Networks (FCNs) and U - Net have achieved remarkable results in semantic segmentation tasks, enabling applications such as medical image analysis, scene understanding, and augmented reality effects.

3.4 Generative Adversarial Networks (GANs):

GANs are a class of deep learning models consisting of two neural networks, a generator and a discriminator, trained in a competitive manner. In computer vision, GANs have been used for tasks such as image generation, image - to - image translation, and super - resolution. They have applications in generating synthetic images for data augmentation, creating realistic artwork, and enhancing image quality.

3.5Attention Mechanisms:

Attention mechanisms, inspired by human visual attention, allow neural networks to focus on relevant parts of an input image while ignoring irrelevant regions. These mechanisms

Paper ID: SR24611174902 DOI: https://dx.doi.org/10.21275/SR24611174902

have been integrated into deep learning models for tasks such as image captioning, visual question answering, and fine - grained image classification, leading to improved performance and interpretability.

4. Application of Computer Vision

4.1 Using Traditional Techniques:

4.1.1 Surveillance and Security:

In surveillance systems, traditional techniques such as background subtraction, motion detection, and simple object tracking are often employed for basic monitoring tasks. These methods may struggle with complex scenarios or variations in lighting conditions. Various techniques can be applied in surveillance and security to enhance its advantages of keeping the location protected and detect any danger.

4.1.2 Barcode and QR Code Scanning:

Traditional computer vision techniques are often used for barcode and QR code scanning applications, where algorithms are used to analyze the images to detect and decode the encoded information. It uses techniques like edge detection, image segmentation, and template matching that may be employed for accurate code recognition.

Figure 5: QR Code scanning

4.1.3 Medical Imaging

Traditional techniques like edge detection, region growing, and template matching are still used in some medical imaging applications for tasks such as image segmentation and feature extraction. However, these methods may lack the robustness and accuracy of AI - driven approaches in challenging medical imaging tasks.

4.1.4 Industrial Automation:

Traditional techniques like image processing and rule - based algorithms are commonly used for quality inspection, defect detection, and object recognition in industrial automation settings.

These methods rely on predefined rules and handcrafted features to analyze images captured from manufacturing processes.

4.2 Using AI - driven Techniques:

4.2.1 Autonomous Vehicles:

These vehicles are capable of navigating and operating on roads without human intervention, relying on a combination of sensors, cameras, radar, lidar, GPS, and sophisticated algorithms to perceive their surroundings and make driving decisions in real - time. Cameras provide visual data that is processed by computer vision algorithms for tasks such as object detection, classification, and tracking. Lidar and radar sensors complement camera data by providing additional depth and distance information. Autonomous vehicles utilize simultaneous localization and mapping (SLAM) algorithms to accurately determine their position and orientation relative to their surroundings.

Decision Making: Based on the perception and localization information, autonomous vehicles employ decision - making algorithms to plan and execute driving actions in real - time. These algorithms consider factors such as traffic laws, road conditions, vehicle dynamics, and safety constraints to generate optimal trajectories and maneuvers for the vehicle to follow.

4.2.2 Gesture Recognition:

AI - driven approaches, including deep learning models, have significantly advanced gesture recognition systems by learning complex spatial and temporal patterns from input data. These models can accurately interpret hand movements and gestures, enabling natural and intuitive human - computer interaction. Multiple students in college/universities have made these kinds of projects where the students' gestures would be recognized for example sign language recognition.

Figure 6: Hand gesture recognition

4.2.3 Retail and E - commerce:

AI - driven computer vision models are employed in retail and e - commerce for advanced tasks such as object detection, instance segmentation, and image - based recommendation systems.

Deep learning techniques enable accurate product recognition, visual search, and personalized shopping experiences.

4.2.4 Augmented Reality (AR):

AR applications often utilize AI - driven computer vision algorithms for tasks such as real - time object tracking, pose estimation, and scene understanding. These techniques enable seamless integration of virtual content with the real world, enhancing user experiences in AR environments. One of its best examples will be Apple's recently launched vision pro.

International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

4.2.5 VAR:

Figure 7: Offside checking using VAR

Video Assistant Referee (VAR) is a technology introduced in football to assist match officials in making more accurate decisions during games. It uses video footage and technology to review key incidents and provide additional information to the on - field referees. VAR aims to minimize errors in crucial decisions, such as goals, penalty kicks, red card offenses, and cases of mistaken identity. It serves as a supplementary tool to assist match officials in making fair and accurate judgments. VAR systems typically consist of multiple cameras positioned around the stadium, including high - definition and ultra - slow - motion cameras. These cameras capture various angles of the game to provide comprehensive coverage of key incidents. Video footage from the cameras is transmitted to a central VAR hub, where trained video assistant referees review the footage and communicate with the on - field match officials via a headset communication system. VAR can be used to review four main types of incidents: goals, penalty kicks, direct red card offenses, and cases of mistaken identity.

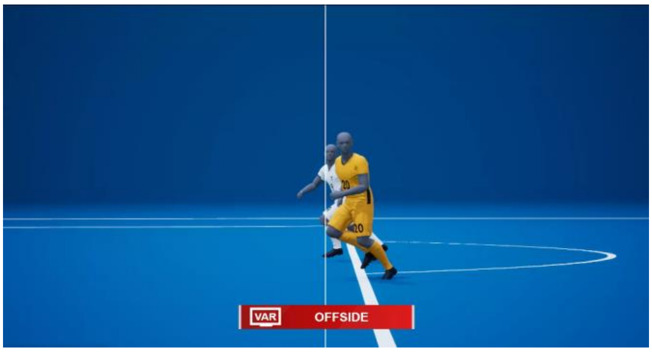

Figure 8: Offside checking using computer vision

5. Survey Analysis

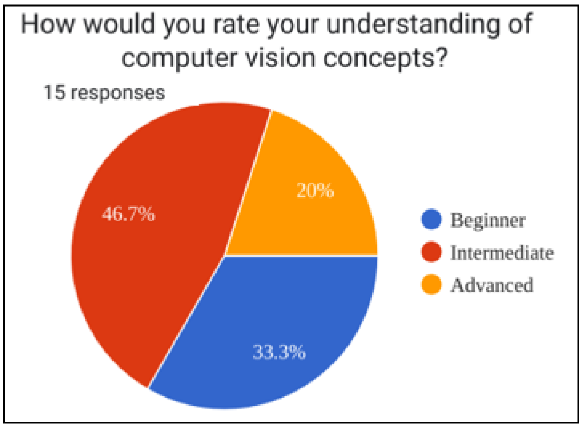

Figure 9: Rate of understanding of computer vision concepts

From the above statistics it is found that 46.7% of people have intermediate level of understanding of computer vision concepts and 33.3% of people have beginner level of understanding, this may be because most people in the IT field have a good familiarity with computer vision while other may have heard of it but not explored it in detail.

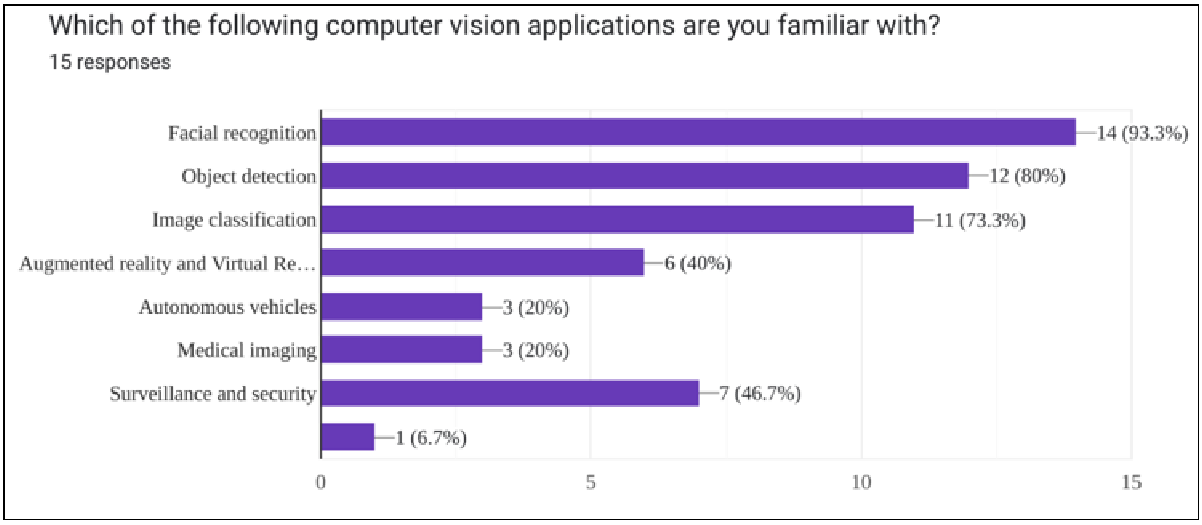

Figure 10: Computer vision application familiarity

According to the survey 93.3% of the people were familiar with facial recognition and about 80% were familiar with object detection. Most people have a basic understanding of computer vision applications.

International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

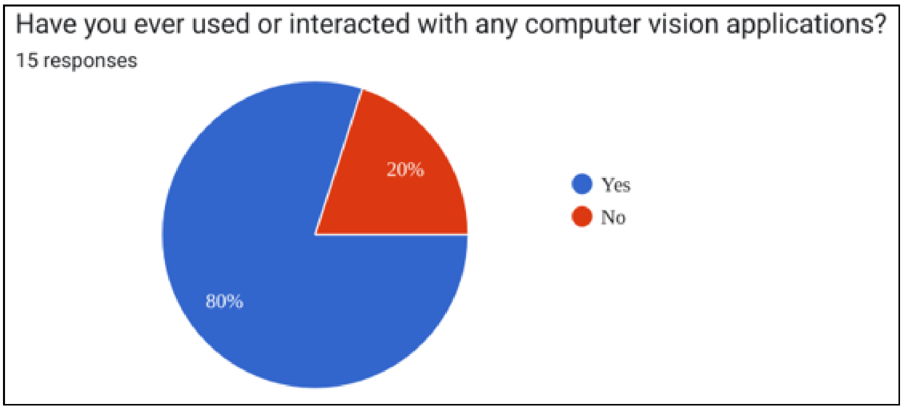

Figure 11: Interaction with computer vision applications

About 80% of people have interacted with computer vision applications like facial recognition, object detection, security and surveillance, QR code scanning and so on and about 20% haven't interacted with computer vision applications.

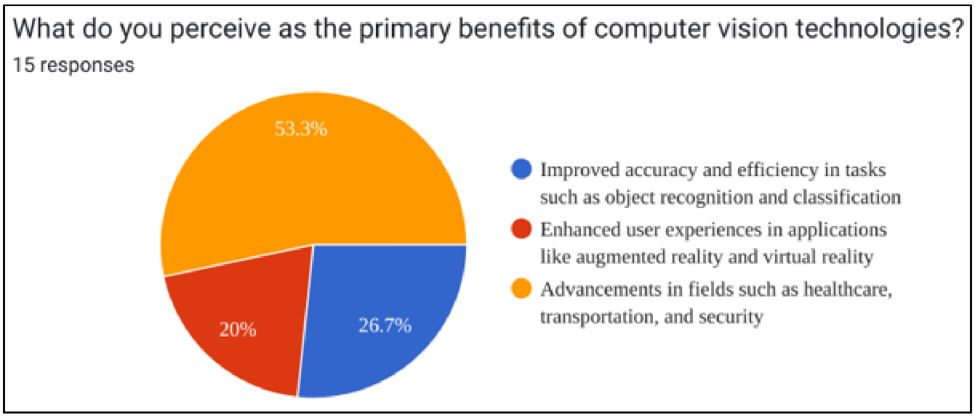

Figure 12: Benefits of computer vision

Benefits of computer vision in daily life, most people believe that it would lead to advancement in fields such as medicine, transportation and security about 53.3% agreed and 20% believe that it would improve accuracy and efficiency in tasks such as object detection and classification.

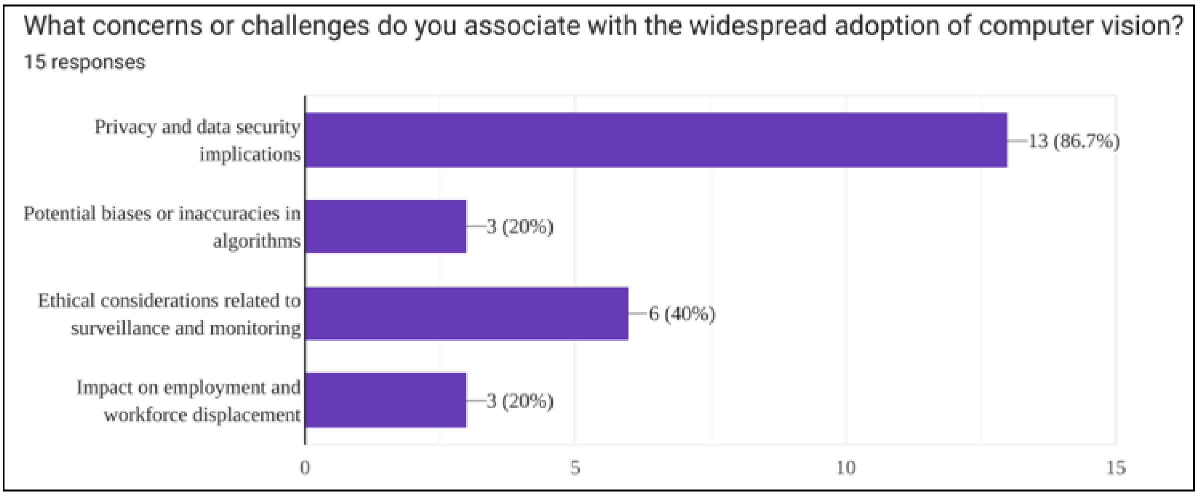

Figure 13: Challenges faced computer vision according to the survey

Challenges faced by computer vision, 86.7% of people believe that computer vision might face privacy and data security problems as its biggest challenge, and 40% believe that surveillance and monitoring problems would be a concern of computer vision.

6. Conclusion

This research paper has explored the evolution of computer vision from traditional techniques to AI - driven innovations, mentioning its applications in various fields such as autonomous vehicles, sports analysis, and everyday technologies. By examining the historical development, current capabilities, and future potential of computer vision, we have identified significant advancements and ongoing challenges. The impact of artificial intelligence has been a critical driver in enhancing computer vision capabilities, enabling more accurate and efficient analysis of visual data. As computer vision continues to evolve, further research and development are essential to address ethical considerations, improve algorithmic fairness, and expand its applications, ensuring that this technology can be harnessed for the benefit of society and the world.

References

[1] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems, 25, 1097 - 1105.
[2] Ren, S., He, K., Girshick, R., & Sun, J. (2015). Faster R - CNN: Towards real - time object detection with region proposal networks. In Proceedings of the 28th International Conference on Neural Information Processing Systems (pp.91 - 99).
[3] Long, J., Shelhamer, E., & Darrell, T. (2015). Fully convolutional networks for semantic segmentation. IEEE Conference on Computer Vision and Pattern Recognition (CVPR).
[4] Shapiro, L. G., & Stockman, G. C. (2001). Computer Vision. Prentice Hall.
[5] Forsyth, D. A., & Ponce, J. (2012). Computer Vision: A Modern Approach. Pearson.
[6] LeCun, Y. (2016, November 21). The future of computer vision. Facebook Research. https: //research. fb. com/the - future - of - computer - vision/