IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, October 2026
Impact Factor: 8.2

Modern Techniques in Computer Vision

Kanthali Harshada[1], Mr. A. A. Dalvi[2], Prof. J. R. Mahajan[3]
Student, Artificial lntelligence and Data Science[1]
Guide, Artificial lntelligence and Data Science[2]
HOD, Artificial lntelligence and Data Science[3]
Adsul Technical Campus, Ahilyanagar, India

Abstract: *Computer vision enables computer systems to extract useful meaning from images and video. Recent advances in artificial intelligence, particularly deep learning, have improved the ability of machines to recognise objects, interpret scenes, and identify visual patterns. This paper presents a structured review of AI techniques used in computer vision, with emphasis on convolutional neural networks (CNNs), their main processing stages, practical applications, and continuing limitations. It explains why manually designed visual features can be unreliable when images vary in lighting, scale, viewpoint, or background, and how learned representations help address these problems. Applications in healthcare, transportation, astronomy, and industrial inspection are discussed alongside concerns involving data quality, bias, privacy, explainability, and computational cost. The paper is a conceptual review based on established literature and does not claim a new experiment or measured performance result. It concludes that responsible deployment requires not only accurate models, but also representative data, careful evaluation, human oversight, and monitoring after deployment.*

Keywords: Artificial Intelligence, Computer Vision, Deep Learning, Convolutional Neural Networks, Image Classification, Object Detection

I. INTRODUCTION

People interpret a scene by combining visual details with experience: they can notice a person, distinguish a road from a footpath, or identify an unusual object. A computer does not understand an image in the same direct way. It receives numerical pixel values and must learn or be programmed to connect those values with meaningful patterns. Computer vision is the field concerned with building systems that perform such visual tasks, while artificial intelligence provides methods for learning patterns and making predictions from data.

Earlier vision systems often depended on hand-designed rules, such as edge detectors, colour statistics, and texture descriptors. These methods remain useful for some controlled tasks, but their performance can drop when the environment changes. Deep learning offers another approach: a model learns multiple levels of representation from examples. In image analysis, convolutional neural networks became especially influential because they use local filters and shared parameters to detect visual structures at different levels [1], [2].

This paper gives an accessible overview of the role of AI in computer vision. It outlines the core problem, compares conventional and learning-based methods, describes the basic CNN pipeline, and reviews representative application areas. It also considers the practical risks that must be addressed before a model can be trusted outside a laboratory.

II. LITERATURE SURVEY

[1] Conventional Computer Vision

Traditional computer-vision techniques mainly depended on manually selected features to identify objects and interpret images. Common techniques included edge detection, corner detection, colour histograms, and texture analysis. The extracted features were then provided to a classification algorithm to recognise objects or categories. Although these methods can produce satisfactory results in controlled environments, their performance may be affected by changes in

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, October 2026
Impact Factor: 8.2

lighting, camera angles, object sizes, and background conditions. These limitations encouraged researchers to develop more flexible approaches capable of learning visual features directly from image data.

[2] Deep Learning and Convolutional Neural Networks

Deep learning introduced a significant improvement in computer vision by allowing models to learn useful features automatically from large datasets. LeCun, Bengio, and Hinton explained how deep learning models develop multiple levels of data representation, enabling them to identify increasingly complex patterns [2]. A major advancement came from Krizhevsky, Sutskever, and Hinton, who demonstrated the effectiveness of AlexNet for large-scale image classification [1]. Convolutional Neural Networks (CNNs) learn features through trainable filters, beginning with simple patterns and gradually identifying more complex structures. This ability makes CNNs suitable for various image-recognition and visual-analysis tasks.

[3] Development of CNN Architectures

Researchers have introduced several CNN architectures to improve feature extraction, training efficiency, and prediction performance. GoogLeNet incorporated Inception modules, enabling the network to process information at different scales while controlling computational requirements [4]. ResNet introduced residual connections that help information and gradients pass through deeper network layers, making the training of very deep models more practical [5]. DenseNet further improved feature reuse by connecting each layer with subsequent layers, allowing previously learned information to contribute to later feature extraction [6]. These developments demonstrate that selecting an appropriate architecture involves balancing prediction accuracy, training stability, memory usage, and computational cost.

[4] Performance Evaluation and Applications

Evaluating a computer-vision model requires more than measuring its overall accuracy. In datasets with unequal class distributions, accuracy alone may not accurately represent the model's performance. Metrics such as precision, recall, F1-score, and intersection over union (IoU), together with confusion matrices, provide a clearer understanding of classification and detection results [3]. Deep-learning-based vision systems are increasingly relevant to areas such as medical-image analysis, intelligent transportation, industrial quality inspection, and scientific research. However, their practical use requires appropriate testing, reliable data, and human supervision, particularly when incorrect predictions could have serious consequences.

[5] Emerging Trends and Research Challenges

Recent developments have expanded computer vision beyond conventional CNN-based methods. Vision transformers use attention mechanisms to model relationships between different parts of an image, while self-supervised learning helps models learn useful representations from data without requiring every image to have a manually assigned label [7], [8]. Despite these advances, several challenges remain. Limited training data, biased datasets, privacy concerns, difficulties in explaining model decisions, and performance variations in unfamiliar environments can affect reliability. In addition, the computational resources required to train and deploy complex models may restrict their use on devices with limited processing power. Future research should therefore focus on developing computer-vision systems that are accurate, efficient, robust, and suitable for real-world applications.

Research Gap

Although AI-based computer-vision systems have achieved significant progress, several challenges still need attention.

- Robustness: Models may not perform reliably when lighting, weather, camera angles, or image quality change.
- Data Quality: Incorrect labels, imbalanced datasets, and data leakage can reduce the reliability of model predictions.
- Computational Cost: Complex models may require more memory, processing power, and energy, making real-time implementation difficult.
- Trustworthy AI: Privacy, fairness, explainability, and human supervision are essential for building reliable and responsible systems.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, October 2026
Impact Factor: 8.2

Proposed Focus: This paper proposes a conceptual framework for evaluating computer-vision systems by considering accuracy, robustness, data quality, computational efficiency, and responsible use.

III. METHODOLOGY

This paper uses a conceptual literature review to explore the development and applications of AI in computer vision.

1. Defining the scope: Identify the key concepts and major applications of AI in computer vision.
2. Reviewing existing research: Study previous research papers and commonly used computer-vision models.
3. Comparing approaches: Understand the differences between traditional image-processing techniques and modern deep-learning methods.
4. Examining evaluation metrics: Review measures such as precision, recall, F1-score, and intersection over union (IoU) to understand how model performance is assessed.
5. Identifying challenges: Explore issues related to data quality, changing environmental conditions, computational costs, privacy, and fairness.
6. Summarising the findings: Highlight the main limitations of existing approaches, identify research gaps, and suggest areas for future research.

IV. AI-BASED APPROACH AND CNN ARCHITECTURE

In a learning-based vision pipeline, images are collected, checked, prepared, and supplied to a model during training. The model adjusts its parameters to reduce the difference between its predictions and the training labels. Its performance is then assessed on separate validation and test data. The exact pipeline depends on whether the goal is image classification, object detection, semantic segmentation, tracking, or another visual task.

A. Convolutional Layer

A convolutional layer moves small learnable filters across an image or feature map. Early filters may respond to simple patterns such as edges and colour transitions; later layers can combine these responses into more complex shapes or object parts. Reusing the same filter across locations reduces the number of parameters compared with connecting every pixel independently to every hidden unit.

B. Activation and Pooling

Activation functions introduce non-linearity, allowing a network to model more complex relationships than a stack of purely linear operations. Pooling or strided operations can reduce the spatial size of feature maps and make later computation more manageable. Such reduction must be chosen carefully because excessive downsampling can remove small but important details.

C. Fully Connected or Task-Specific Output Layers

For basic image classification, the extracted features may be passed to a classification head that estimates the likelihood of each class. Detection and segmentation require additional output structures that predict object locations or assign labels to image regions. Therefore, a CNN is not a single fixed design; it is a family of building blocks that can be arranged for different visual objectives.

D. Training and Evaluation

During training, the model compares predictions with known labels and updates its weights through optimisation. Data augmentation—such as modest cropping, flipping, or brightness changes when appropriate—can help the model cope with expected variation. Evaluation should use data that were not used to fit the model and should include metrics suited to the task. Accuracy alone can be misleading when classes are imbalanced; precision, recall, F1-score, intersection-over-union, and confusion matrices may provide additional insight [3].

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, October 2026
Impact Factor: 8.2

V. SYSTEM WORKFLOW

The workflow of AI in Computer Vision consists of Image Collection → Preprocessing → Feature Extraction → AI Analysis → Object Detection → Risk Assessment → Automated Response → Continuous Monitoring.

- Data Collection: Collects images, videos, and camera feeds for analysis.
- Preprocessing: Improves image quality and removes unwanted noise.
- Feature Extraction: Identifies important shapes, edges, colours, and patterns.
- AI Analysis: Uses machine learning, deep learning, and CNNs to recognize objects and classify images.
- Risk Assessment: Evaluates detected objects and identifies potential risks.
- Automated Response: Tracks objects, generates alerts, and records results.
- Continuous Monitoring: Monitors new data and helps improve system performance.

Copyright to IJARSCT
www.ijarsct.co.in
DOI: 10.48175/IJARSCT-38889

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, October 2026
Impact Factor: 8.2

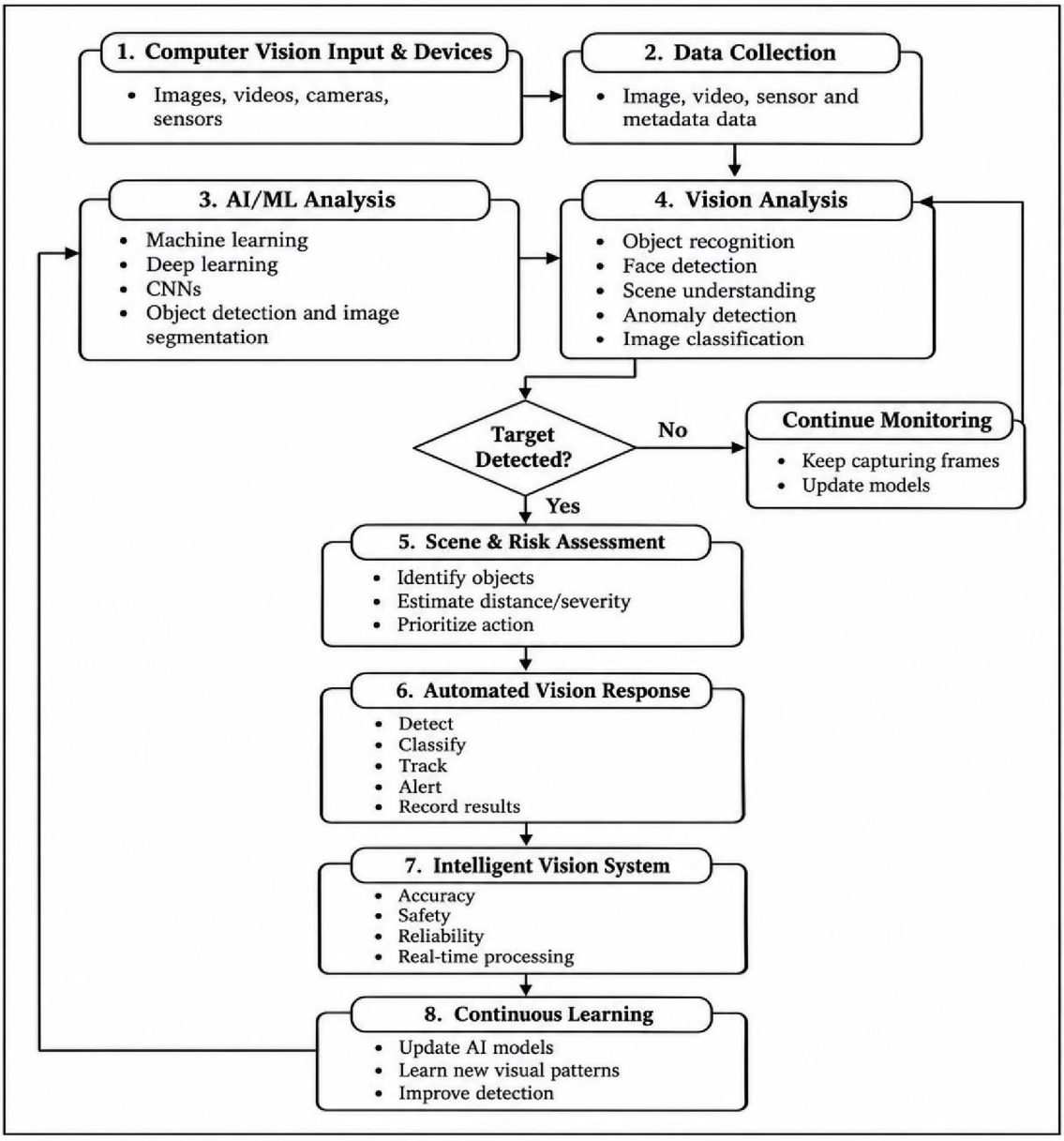


Figure 1: Workflow of AI in Computer Vision

VI. DEVELOPMENT OF CNN ARCHITECTURES

CNN research has progressed through designs that address different limits in image recognition. LeNet demonstrated the usefulness of convolutional processing for digit recognition. AlexNet showed the impact of deeper networks, large labelled datasets, and GPU computation in the 2012 ImageNet competition [1]. GoogLeNet introduced Inception modules to process features at multiple scales with more efficient computation [4]. ResNet used residual connections to make the optimisation of very deep networks more practical [5], while DenseNet connected feature maps across layers

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, October 2026
Impact Factor: 8.2

to encourage feature reuse [6]. These milestones illustrate continuing efforts to balance representational power, training stability, and computational cost.

No architecture is best for every task. The choice should consider dataset size, image resolution, latency, available hardware, memory limits, model interpretability, and the consequences of errors. In some projects, a smaller model or a pretrained model adapted to a specific dataset may be more practical than training a very large network from scratch.

VII. CHALLENGES AND RESPONSIBLE USE

Data quality is a primary concern. Incorrect labels, duplicated images, class imbalance, and insufficient representation of real-world conditions can distort model behaviour. A dataset split should prevent near-duplicate or related samples from leaking between training and test sets. When data come from different sites or devices, evaluation should test whether performance remains stable across those settings.

Fairness and privacy also matter. If some groups or conditions are poorly represented, errors may be distributed unevenly. Images can contain identifying or sensitive information, so collection, storage, and sharing must follow appropriate consent, security, and legal requirements. In high-impact contexts, teams should document intended use, limitations, and evaluation procedures, and retain meaningful human oversight.

Deployment creates additional constraints. High-performing models may require substantial processing power, memory, and energy. Real-time systems must meet latency requirements, while devices in the field may face network or power limitations. Models can also degrade when cameras, environments, or user behaviour change. Monitoring, periodic re-evaluation, and a safe fallback process should therefore be planned from the beginning.

VIII. OBJECTIVES OF AI-BASED COMPUTER VISION

The main objective is to enable machines to interpret visual information with less dependence on manually written feature rules. More specifically, AI-based vision systems aim to: (1) learn meaningful patterns from representative image data; (2) classify, locate, or segment relevant objects; (3) provide reliable results across realistic changes in the environment; (4) reduce repetitive inspection effort where automation is suitable; and (5) support human decision-making while making limitations and uncertainty clear. These are design goals, not guaranteed outcomes; each must be verified for the intended application.

IX. APPLICATION AREAS

A. Healthcare and Medical Imaging

AI methods can assist with analysing X-rays, CT scans, MRI images, retinal photographs, and digital pathology slides. Depending on the task and available evidence, a model may flag regions for review, segment anatomical structures, or help prioritise cases. Such systems should be treated as decision-support tools unless they have been appropriately validated and authorised for independent use. Image quality, differences among hospitals, patient privacy, and the cost of false negatives all require careful attention.

B. Transportation and Driver Assistance

Vehicles and roadside systems use visual perception to identify lanes, pedestrians, signs, vehicles, and possible hazards. In practical systems, camera information may be combined with radar, lidar, maps, or other sensors.

Performance must be evaluated across weather, lighting, road types, and unusual events. A vision model can make errors, so safety-critical driving functions require layered safeguards rather than relying on one prediction alone.

C. Industrial Inspection

Factories can use vision systems to check product surfaces, read labels, verify assembly, and identify visible defects. Automation may improve inspection speed and consistency, especially for repetitive tasks. Nevertheless, the model

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, October 2026
Impact Factor: 8.2

needs examples of relevant defects and acceptable variation. New materials, changed production settings, or rare defect types can lead to missed faults, so ongoing quality checks and a clear human escalation process remain important.

D. Astronomy and Scientific Imaging

Large scientific projects generate more images than researchers can reasonably inspect one by one. Machine-learning methods can help sort objects, identify unusual patterns, and direct experts toward promising observations. Their outputs should be interpreted in the context of the instruments, sampling methods, and scientific questions that produced the data. Automated classifications are useful aids, but they do not replace scientific verification.

X. PROBLEM STATEMENT AND MOTIVATION

An image may contain millions of pixel values, yet a useful system must identify a comparatively small set of meaningful objects or events. The task becomes harder when the same object appears at different sizes, under different lighting, or partly hidden by another object. Background clutter, motion blur, low image quality, and differences between data collected in the laboratory and data collected in the real world can further reduce reliability.

A second difficulty is that visual data are not always labelled consistently. Training examples may be limited, imbalanced, or drawn from only one population or environment. A model can therefore learn accidental correlations rather than the visual evidence needed for the intended task. The central problem is to build computer-vision systems that learn useful visual features, generalise to new situations, and communicate uncertainty where mistakes could have serious consequences.

XI. CONVENTIONAL METHODS AND THEIR LIMITATIONS

Traditional computer-vision pipelines commonly separate feature extraction from classification. A developer first selects descriptors—such as edges, corners, colour histograms, or texture measurements—and then supplies those features to a classifier. This separation can make the system easier to inspect and may work well when the environment is stable and the relevant patterns are already known.

However, a hand-crafted feature set may not represent every important variation in a complex scene. A change in camera angle, illumination, surface appearance, or object position can weaken the relationship between the selected features and the target class. Designing and maintaining specialised rules also takes time and domain knowledge. These limitations motivated methods that learn feature representations from data instead of relying entirely on manually specified descriptors. This does not mean conventional techniques are obsolete; in small, well-defined, or resource-constrained tasks, they can still be appropriate.

XII. FUTURE SCOPE

Future work is likely to focus on systems that learn from fewer labelled examples, combine information from images and other modalities, and adapt more safely to changing environments. Self-supervised and vision-transformer approaches have broadened the range of methods available beyond conventional CNNs [7], [8]. Lightweight networks, edge AI, privacy-preserving learning, and methods that estimate uncertainty are also important directions. For practical adoption, progress in accuracy must be matched by transparent evaluation, robustness testing, accessibility, and clear responsibility for decisions made with model assistance.

XIII. CONCLUSION

Artificial intelligence has expanded computer vision from collections of hand-designed image rules toward systems that learn visual representations from data. CNNs remain an important foundation because their local filters and shared weights can capture spatial patterns efficiently, while newer architectures provide alternative ways to represent complex images. Applications in medicine, transportation, manufacturing, and astronomy show the potential value of automated visual analysis.

IJARSCT
International Journal of Advanced Research in Science, Communication and Technology
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
ISSN: 2581-9429
Volume 6, Issue 2, October 2026
Impact Factor: 8.2

REFERENCES

[1] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," in Advances in Neural Information Processing Systems, vol. 25, 2012.
[2] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, pp. 436–444, 2015.
[3] M. Sokolova and G. Lapalme, "A systematic analysis of performance measures for classification tasks," Information Processing & Management, vol. 45, no. 4, pp. 427–437, 2009.
[4] C. Szegedy et al., "Going Deeper with Convolutions," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2015.
[5] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016.
[6] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely Connected Convolutional Networks," in Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017.
[7] A. Dosovitskiy et al., "An Image is Worth 16x16 Words: Transformers for Image Recognition at Scale," in International Conference on Learning Representations, 2021.
[8] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, "Momentum Contrast for Unsupervised Visual Representation Learning," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020.