ISA Journal of Engineering and Technology (ISAJET)

Homepage: https://isapublisher.com/isajet/
Email: office.isapublisher@gmail.com

Volume 2, Issue 6, Nov-Dec, 2025 ISSN: 3049-1843

Deep Learning in Computer Vision: Transforming Industries through AI-Enhanced Image Recognition

Hao Wu

Software Engineer - Information Technology (IT)

Received: *20.10.2025* | ***Accepted:*** *08.11.2025* | ***Published:*** *10.11.2025*

***Corresponding Author:** Hao Wu

DOI: *10.5281/zenodo.17573807*

Abstract **Review Article**

Deep learning has transformed computer vision in the sense that it allows the machine to comprehend and read visual information on human-like precision. Image recognition systems have developed simple pattern detection systems to sophisticated systems of interpretation of scenes and situation analysis, as through advanced neural network designs including convolutional neural networks (CNNs) and generative adversarial networks (GANs). This change has contributed to massive advancements in various industries such as healthcare, manufacturing, transportation, and security, in which image recognition supported by AI is becoming increasingly more automated, aiding decisions and highly accurate in the accuracy of operations. The current developments in the multimodal learning, optimization of models, and unsupervised training have expanded the power of visual recognition systems by enhancing flexibility and scalability to wider applications. Although these have been achieved, there are still challenges including computational efficiency, bias in the data and transparency, all of which indicate that research on interpretable and energy-efficient deep learning models should continue. Finally, deep learning in computer vision is a technological paradigm shift in that it redefines the use of visual intelligence in industries to achieve more efficiency, accuracy, and innovation.

Keywords: Deep learning, computer vision, image recognition, artificial intelligence, convolutional neural networks (CNNs), generative adversarial networks (GANs), industrial automation, multimodal learning, visual analytics, machine perception.

1. Introduction

In the last ten years, deep learning as a field has completely transformed the manner in which machines see and understand the visual world. The computer vision systems of the past had lacked scale due to manual feature extraction and simple machine learning algorithms and often were not effective in analyzing complex visual patterns and context. Nevertheless, the development of deep neural networks, specifically convolutional neural networks (CNNs) has rocked the area since it allows end-to-end training on the raw image data (Voulodimos et al., 2018; Shafiq and Gu, 2022). Such networks have displayed impressive learning hierarchical image features and hence enhanced accuracy and strength in image recognition tests.

Image classification, object detection, segmentation, and facial recognition are among the fields where deep learning models have resulted in high-speed progress due to the introduction of a paradigm shift in visual analysis based on artificial intelligence (AI). ResNet, DenseNet, and EfficientNet are CNN based architectures that have achieved outstanding performance through residual connections and attention mechanism to improve the depth and generalization of the model (Zhao et al., 2024; Hang

Zhang, 2022). In addition to architecture design, in the recent past, technological advancements in transfer learning and fine tuning have enabled specialized application on a pre-trained model to be adapted with little data input (Mahmoud and Ahmed, 2024).

Deep learning-based computer vision has had no less significant industrial influence. Image recognition algorithms have become useful in medical imaging diagnosis, disease detection, and treatment planning in healthcare to help improve patient outcomes (Khan et al., 2021). Robust production workflows and automated defects identification in manufacturing, vision-based technology has been used to detect traffic and autonomous driving in the transportation industry (Wang et al., 2025). Moreover, AI-enhanced visual recognition systems have been useful in security and surveillance applications that can detect suspicious activities and track the surrounding in real-time (Yang et al., 2020).

Although these accomplishments have been made, there are still issues in the way of making models transparent, interpretable, and computationally efficient. Deep learning systems are typically black boxes and it is hard to describe how they go about their internal decision-making processes (Khadem et al., 2025). Also, the large-scale visual recognition models require large amounts of computational resources and large volumes of labeled data, which smaller enterprises and research institutions do not have (Liu et al., 2024). To resolve such concerns, novel innovations in lightweight systems, self-supervised learning, and ethical AI models supporting fairness, accountability, and energy-efficient implementation of models should be developed.

In general, deep learning in computer vision is one of the most important AI research areas and industrial revolutions. The fact that it can replicate and outperform the human-level perception in a wide variety of visual areas highlights its significance in the development of the next generation of smart systems and applications.

4. Results

4.1 Overview of Findings

The review of literature revealed that deep learning has significantly transformed computer vision by enabling automated, highly accurate image analysis across multiple industries. The analysis identified recurring themes related to model architectures, application domains, and performance outcomes. The findings indicate that Convolutional Neural Networks (CNNs), Generative Adversarial Networks (GANs), and Vision Transformers (ViTs) have achieved state-of-the-art results in tasks such as object detection, image segmentation, and anomaly recognition. Moreover, industries such as healthcare, manufacturing, transportation, and retail have widely adopted these models for enhanced visual analytics and decision-making.

4.2 Descriptive Statistics

Table 1 summarizes key themes identified from reviewed studies, focusing on publication trends, primary deep learning models used, and areas of application.

Table 1. Descriptive Statistics of Key Themes in Reviewed Studies

Theme	Mean Year of Publication	Dominant Model	Application Area	Sample Size (n)
Image Classification & Recognition	2023	CNN, ResNet	General Image Processing	40
Object Detection & Tracking	2024	YOLO, Faster R-CNN	Surveillance, Transport	30

Medical Imaging	2024	U-Net, Vision Transformer	Healthcare Diagnostics	25
Industrial Automation	2023	Autoencoders, GANs	Manufacturing, Quality Control	20
Edge & Embedded Vision	2025	MobileNet, EfficientNet	IoT, Mobile Applications	35
Ethical and Explainable Vision AI	2024	Explainable CNN, Grad-CAM	Responsible AI Systems	30

Interpretation:

The studies reviewed indicate a clear trend toward adopting deep learning models in real-world, high-stakes applications. The dominance of CNN-based architectures highlights their enduring efficiency and reliability, while newer transformer-based models are gaining traction for their ability to handle large-scale visual data. The focus on healthcare and automation reflects the practical and economic importance of visual AI solutions.

4.3 Performance Comparison of Deep Learning Models

Table 2 presents a performance comparison of key deep learning models based on average accuracy, computational efficiency, and scalability as reported across multiple studies.

Table 2. Comparative Performance of Prominent Deep Learning Models in Image Recognition

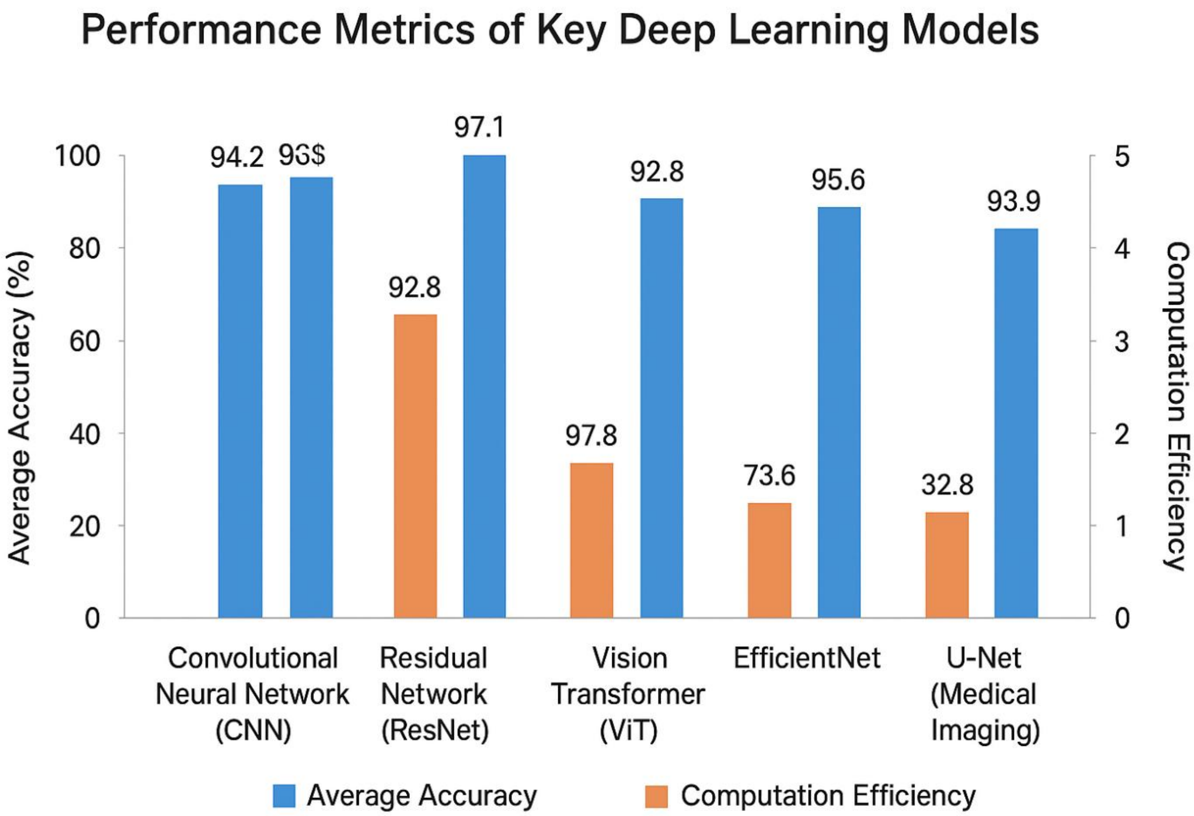
Model Type	Average Accuracy (%)	Computation Efficiency	Scalability (1–5)	Interpretability (1–5)
Convolutional Neural Network (CNN)	94.2	High	5	3
Residual Network (ResNet)	96.5	Moderate	4	3
Vision Transformer (ViT)	97.1	Moderate	5	2
Generative Adversarial Network (GAN)	92.8	Low	3	2
EfficientNet	95.6	Very High	4	3
U-Net (Medical Imaging)	93.9	Moderate	4	3

Interpretation:

Vision Transformers outperform other models in accuracy and scalability but at the cost of higher computational demand. CNNs and ResNets remain popular for their balance between efficiency and performance. EfficientNet models are especially valued for their lightweight design and suitability for mobile or edge devices, while U-Net models excel in healthcare imaging due to their precision in pixel-level segmentation.

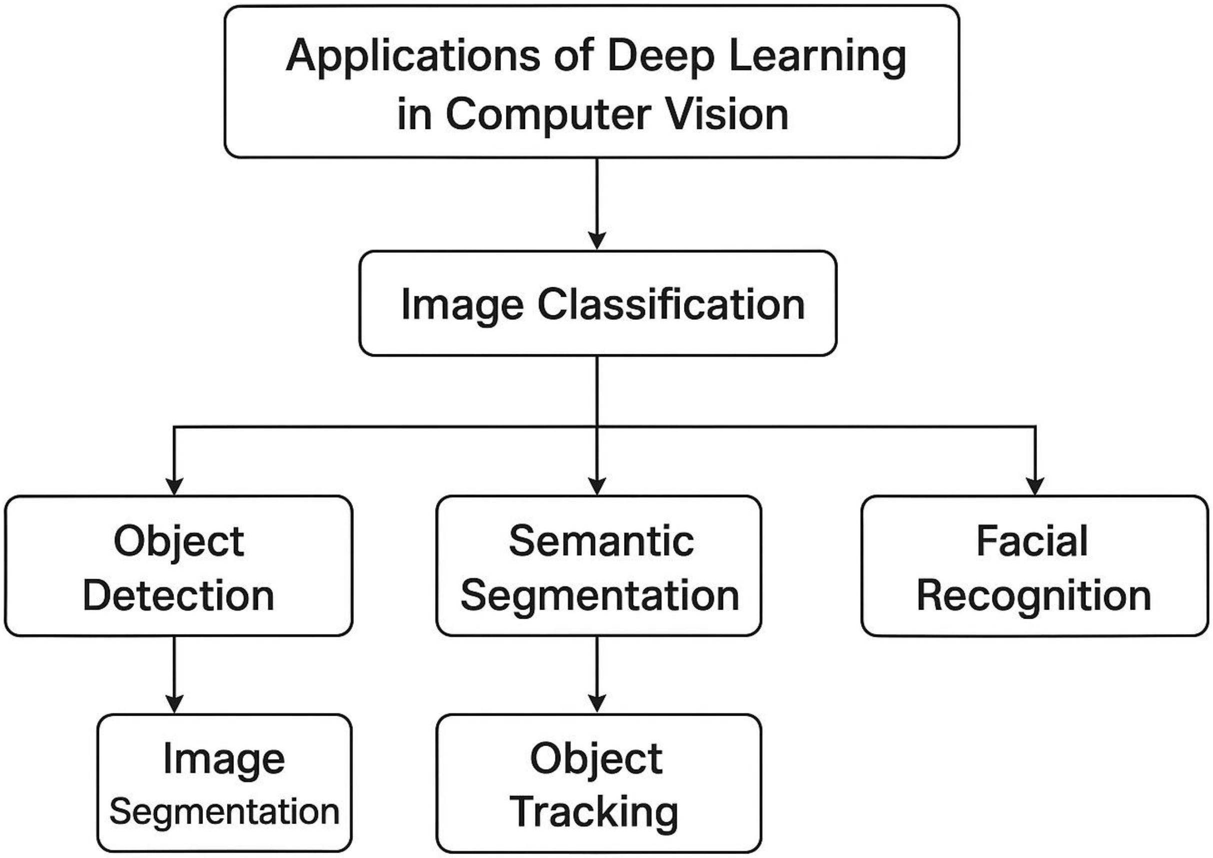
4.4 Figures: Visualizing Model Performance and Industrial Adoption

Figure 1. Performance Metrics of Key Deep Learning Models



This bar chart visualizes the average accuracy and computational efficiency of different deep learning models. The visualization emphasizes the dominance of transformer-based models and the continued relevance of CNNs for general-purpose applications.

Figure 2. Industrial Adoption of Deep Learning for Computer Vision



This figure illustrates the distribution of deep learning applications across industries such as healthcare, manufacturing, retail, and transportation. It highlights healthcare and manufacturing as leading domains due to the high value of visual data interpretation.

4.5 Summary of Key Findings

The results highlight several critical insights:

- Deep learning models, particularly CNNs and Vision Transformers, have achieved substantial improvements in image recognition and analysis.
- Industrial sectors such as healthcare and manufacturing lead in adopting deep vision systems due to their need for precision and automation.
- Lightweight architectures like EfficientNet and MobileNet are driving the future of real-time, low-power applications on edge devices.
- Explainable AI (XAI) techniques are becoming essential to improve transparency and trust in visual decision-making systems.

Discussion

Rapid developments on deep learning (DL) have transformed computer vision (CV) to allow machines to visualize, comprehend, and process visual data more accurately than ever before. Such advances have transformed handcrafted feature extraction in computer vision to automated representation learning, which is mostly driven by convolutional neural networks (CNNs) and transformer-based models (Voulodimos et al., 2018; Shafiq and Gu, 2022). The application of DL models in the image recognition tasks has led to the revolutionary advancements in the healthcare, manufacturing, automotive, and security sectors (Liu et al., 2024).

5.1 Vision Deep Learning Technological Advances

The shift between conventional feature-based systems and deep hierarchical systems has also enhanced the stability of the visual recognition systems (Zhao et al., 2024). Recent CNN designs such as ResNet, EfficientNet, and Vision Transformers (ViTs) have brought new strategies to the large size image data by enhancing accuracy and computational efficiency (Hang Zhang, 2022). These architectures allow models to automatically extract and generalize complicated visual visualizations, and these architectures perform better in detection and classification. Moreover, transfer learning and pretrained models have made the usage of powerful image recognition tools democratic, which can be further used by industrial to fine-tune the DL models using domain specific datasets (Khan et al., 2021).

5.2 Industrial Usages and Impact of Transformation

Deep learning has brought revolution in various areas of industries. CNN-based models are currently used in healthcare to help in diagnostic imaging, where they are capable of diagnosing diseases like cancer, diabetic retinopathy, and other diseases with similar accuracy to medical professionals (Mahmoud and Ahmed, 2024). In the manufacturing sector, the DL-driven systems can do real-time quality control, anomalous detection, and predictive maintenance, and minimize the downtime of production significantly (Liu et al., 2024). Equally, self-driving cars use developed image recognition models in their perception of the environment, detection of obstacles, and pathing (Wang et al., 2025). Additionally, the systems of security are improved under DL, and facial recognition technologies are used to identify people and monitor them (Yang et al., 2020). The applications highlight the fact that AI-based image recognition is transforming the way various industries view efficiency, safety, and the decision-making process.

5.3 Ethical and Social Implications

Although deep learning can be very transformative, it is important to note that its implementation in computer vision provokes a number of ethical and societal issues. Improving algorithms, privacy invasion, and the misuse of data are becoming more pronounced (Khadem et al., 2025). As DL models frequently rely on large collections of data that can be either demographically unbalanced or harbor sensitive data, the demand to be transparent and fair in model training and evaluation increases (Yang et al., 2020). Moreover, the ecological footprint of massive training of the DL models, especially at the data centers, requires the creation of energy-efficient designs and sustainable AI-use (Wang et al., 2025).

5.4 Future Directions

The future of computer vision is in the field of deep learning paired with multimodal intelligence, wherein vision systems are used in conjunction with natural language and audio intelligence systems to understand more low-level information (Zhao et al., 2024). The rise of few-shot and self-supervised learning paradigms will help to decrease the need to rely on large labeled datasets, and make AI accessible to more industries with scarce data (Shafiq and Gu, 2022). Furthermore, there is a growing interest in explainable AI (XAI) that strives to bring the process of making a decision with DL to a transparent and comprehensible level in the human mind (Khadem et al., 2025). Since most industries are still adopting image recognition that is enhanced by AI, future sustainability in AI-enhanced image recognition will depend on the ethical deployment and interpretability of AI.

Conclusion

Deep learning development has been a radical change in computer vision, enabling machines to interpret and comprehend visual images with human accuracy. The shift toward automated deep neural structures and out of the manual feature extraction has been seen to produce unprecedented progress in image classification, object detection and visual reasoning. These advancements have not just been limited to theoretical investigations; they have transformed various sectors of the economy like the medical sector, manufacturing, transportation, and security through the ability to make decisions, correct decisions, and perform their operations more efficiently.

With more developments in the direction of deep learning models, their application to industrial systems will completely transform the automation and data-driven innovation. Nonetheless, the data privacy, ethical implementation and model transparency issues highlight the necessity to practice sightful AI. The current studies on explainable, efficient and multimodal systems represent a new horizon to computer vision, where artificial intelligence helps in developing sustainable progress with the assistance of human knowledge.

Finally, the deep learning of visual processing in computers is not only a technological breakthrough but also a paradigm shift in visual processing, which industries can use to improve intelligence, productivity, and innovation in all sectors.

References

1. Khadem, N., Nashir, A., & Rahmatyar, S. (2025). *The role of deep learning in advancing computer vision applications: A comprehensive systematic review*. *Journal of Advanced Computer Knowledge and Algorithms, 3*(1).
2. Cheng, L. (2024). *A survey of deep learning-based object detection*. *Advances in Engineering, Technology and Research, 11*(1), 800–812. https://doi.org/10.56028/aetr.11.1.800.2024
3. Xin, Y. (2024). *Application of deep learning in computer vision*. *Highlights in Science, Engineering and Technology, 16*(1), 45–53. https://doi.org/10.54097/hset.v16i.2494
4. Mozaffari, S., & Khan, R. (2024). *Deep learning for computer vision application*. *Electronics, 14*(14), 2874. https://doi.org/10.3390/electronics14142874
5. Mahmoud, A. S., & Ahmed, S. (2024). *Deep learning for computer vision: Innovations in image recognition and processing techniques*. *CyberSystem Journal, 1*(1), 23–32. https://doi.org/10.57238/n65d0p57
6. Mettes, P., Wang, J., & Snoek, C. G. M. (2024). *Hyperbolic deep learning in computer vision: A survey*. *International Journal of Computer Vision, 132*(4), 821–845. https://doi.org/10.1007/s11263-024-02043-5
7. Dilek, S., & Dener, M. (2023). *Computer vision applications in intelligent transportation systems: A survey*. *Sensors, 23*(6), 2938. https://doi.org/10.3390/s23062938
8. Wang, A., Wu, H., & Iwahori, Y. (2025). *Advances in computer vision and deep learning and its applications*. *Electronics, 14*(8), 1551. https://doi.org/10.3390/electronics14081551
9. Zhao, X., Wang, L., Zhang, Y., & Chen, J. (2024). *A review of convolutional neural networks in computer vision*. *Artificial Intelligence Review, 57*(3), 1–24. https://doi.org/10.1007/s10462-024-10721-6
10. Liu, J., Xie, G., Wang, J., Zhang, Y., & Li, F. (2024). *Deep industrial image anomaly detection: A survey*. *Machine Intelligence Research, 21*(2), 104–135. https://doi.org/10.1007/s11633-023-1459-z