# A Comprehensive Review of Convolutional Neural Networks for Computer Vision: Architectures, Applications, and Emerging Trends

Sanchit Jakhetia*, Rudransh Chandel, Priyanshu Kumar, Punya Singh Gaur, Pragya Chauhan, Sambhavi Priyadarshani, Pratik Kashyap, Riya Sharma
*Department of Information Technology, Noida Institute of Engineering and Technology, Greater Noida, India*
Email: *sanchitjakhetia@gmail.com

***Abstract*—Convolutional Neural Networks (CNNs) have transformed the field of computer vision by enabling automated feature extraction and delivering remarkable performance across a wide range of visual recognition tasks. The rapid evolution of CNN architectures, coupled with the growing availability of large-scale datasets and high-performance computing platforms, has accelerated their adoption in domains such as image classification, object detection, semantic segmentation, medical image analysis, autonomous driving, precision agriculture, and intelligent surveillance. Despite these advances, the increasing complexity of network designs and the emergence of alternative deep learning paradigms have created the need for a unified and up-to-date assessment of CNN-based developments.**

**This review presents a comprehensive analysis of CNNs for computer vision by systematically examining their architectural evolution, major application areas, and recent research trends. Relevant studies were identified through a structured literature search across leading scientific databases using predefined inclusion and exclusion criteria to ensure the selection of high-quality contributions. The review highlights the progression from early CNN models to modern efficient architectures, compares their strengths and limitations, and summarizes their practical impact across diverse application domains. Furthermore, key challenges, including computational cost, data dependency, model interpretability, robustness, and deployment constraints, are critically discussed. Emerging directions such as lightweight CNNs, hybrid CNN–Transformer frameworks, self-supervised learning, edge intelligence, and explainable artificial intelligence are also explored. By integrating architectural, application-oriented, and future perspectives, this review provides a consolidated reference for researchers and practitioners while identifying promising opportunities for the next generation of computer vision systems.**

***Keywords*—*Convolutional Neural Networks, Computer Vision, Deep Learning, Object Detection, Image Classification, Transfer Learning, Vision Transformers, Medical Imaging***

## I Introduction

Computer vision has become a cornerstone of modern artificial intelligence (AI), enabling machines to interpret and analyze visual information with increasing accuracy and efficiency. The widespread adoption of digital imaging devices, smart sensors, and Internet of Things (IoT) technologies has generated an enormous volume of visual data, creating a growing demand for automated image understanding systems. Applications such as image classification, object detection, semantic segmentation, facial recognition, and activity analysis have become integral to numerous scientific and industrial domains. Traditional computer vision techniques, which relied on handcrafted feature extraction methods such as Scale-Invariant Feature Transform (SIFT) and Histogram of Oriented Gradients (HOG), often exhibited limited robustness when dealing with complex backgrounds, illumination variations, and large-scale datasets [1, 2].

The emergence of deep learning has fundamentally transformed computer vision by enabling data-driven feature learning. Among various deep learning architectures, Convolutional Neural Networks (CNNs) have demonstrated exceptional capability in automatically extracting hierarchical representations from raw images through convolutional operations and shared parameters. The introduction of LeNet established the feasibility of CNNs for image recognition, while the success of AlexNet in the ImageNet challenge significantly accelerated research in deep visual learning. Subsequent architectures, including VGG, GoogLeNet, ResNet, DenseNet, EfficientNet, ConvNeXt, and more recently Vision Transformers, have continuously improved feature representation, computational efficiency, and model scalability for increasingly challenging computer vision tasks [3, 4, 5, 6, 7, 8].

The chronological evolution of representative CNN architectures is illustrated in Figure 1. The progression highlights the transition from shallow networks designed for simple recognition tasks to highly sophisticated models capable of addressing large-scale visual understanding problems.

CNN-based computer vision systems have significantly influenced diverse application domains. In healthcare, deep CNN models assist in disease diagnosis, medical image segmentation, and computer-aided detection. Autonomous vehicles employ CNNs for object recognition, lane detection, and scene understanding to improve navigation safety. Similarly, precision agriculture benefits from CNN-based crop monitoring and plant disease identification, while intelligent surveillance systems utilize CNNs for facial recognition, anomaly detection, and activity monitoring. The adaptability of CNN architectures across these heterogeneous applications demonstrates their importance as a general-purpose framework for visual intelligence [9, 10, 11, 12].

Despite these achievements, several challenges remain. Deep CNN models often require extensive labeled datasets, significant computational resources, and high training costs. Model interpretability, robustness against adversarial perturbations, deployment on resource-constrained devices, and adaptation to unseen environments continue to attract considerable research attention. Furthermore, the rapid emergence of transformer-based vision models and hybrid CNN-transformer frameworks has expanded the research landscape, necessitating


**Fig. 1:** Evolution of representative CNN architectures for computer vision.

a comprehensive understanding of the evolving role of CNNs in modern computer vision [13, 14].

Although numerous survey articles have investigated CNN research, many concentrate either on architectural developments or on specific application domains such as medical imaging or autonomous driving. Existing studies frequently lack a unified discussion that simultaneously examines architectural evolution, practical applications, comparative performance, current challenges, and emerging research trends within a single framework. Consequently, there remains a need for an integrated review that provides a holistic perspective on CNN-based computer vision.

Motivated by these observations, this paper presents a comprehensive review of convolutional neural networks for computer vision. The major contributions of this work are summarized as follows:

- Development of a systematic taxonomy covering the evolution of major CNN architectures.
- Comprehensive analysis of CNN applications across diverse computer vision domains.
- Comparative discussion of representative architectures, highlighting their strengths and limitations.
- Critical examination of current research challenges and practical deployment issues.
- Investigation of emerging trends, including lightweight CNNs, hybrid CNN-transformer models, edge intelligence, and explainable AI.
- Identification of promising future research directions for next-generation computer vision systems.

The remainder of this paper is organized as follows. Section II presents the review methodology. Section III discusses CNN architectures and application domains. Section IV provides a comparative analysis of representative models. Section V highlights research gaps and challenges. Section VI discusses emerging trends and future directions, while the final section concludes the review.

# II Review Methodology

To provide a balanced and comprehensive overview of Convolutional Neural Networks (CNNs) for computer vision, a systematic literature review strategy was adopted. The methodology was designed to identify influential studies, minimize selection bias, and ensure adequate coverage of both foundational architectures and recent advancements. The overall workflow, illustrated in Figure 2, consists of four major stages: literature identification, screening, eligibility assessment, and final selection.

The literature search was conducted using widely recognized scientific databases, including IEEE Xplore, ACM Digital Library, SpringerLink, ScienceDirect, Wiley Online Library, and arXiv. To retrieve relevant publications, combinations of keywords such as “Convolutional Neural Networks,” “Computer Vision,” “Deep Learning,” “Image Classification,” “Object Detection,” and “Semantic Segmentation” were employed. Foundational studies with significant historical impact and recent high-quality contributions published primarily between 2015 and 2026 were considered.

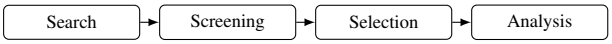


**Fig. 2:** Overview of the review methodology adopted in this study.

**TABLE I:** Summary of the literature selection strategy.

| | |
|---|---|
| Databases | IEEE, ACM, Springer, Elsevier, Wiley, arXiv |
| Period | 2015–2026* |
| Keywords | CNN, Computer Vision, Image Classification, Object Detection |
| Inclusion | Peer-reviewed, CNN-based studies |
| Exclusion | Duplicates, Non-relevant works |

The inclusion criteria comprised peer-reviewed journal and conference papers focusing on CNN architectures and computer vision applications, while duplicate records, non-English articles, and studies unrelated to visual learning were excluded. The selected publications were further examined based on their methodological novelty, application domain, experimental validation, and practical significance. Information extracted from each study included the CNN model, target application, benchmark datasets, evaluation metrics, and major contributions.

The final collection of studies was organized into three broad categories: CNN architectural evolution, application-specific developments, and emerging research trends. This classification facilitates a systematic comparison of existing approaches and provides a coherent foundation for the subsequent analysis presented in this review.

As summarized in Figure 2 and Table I, the adopted methodology ensures a structured and reproducible review process while maintaining a balanced representation of foundational and state-of-the-art CNN research for computer vision.

# III Literature Review

## *A. A. Evolution of CNN Architectures*

The remarkable success of Convolutional Neural Networks (CNNs) in computer vision is primarily attributed to continuous architectural innovations that have improved feature representation, computational efficiency, and model scalability. Since the introduction of the first practical CNN model,

 

researchers have proposed increasingly sophisticated architectures to address challenges such as vanishing gradients, excessive computational cost, and limited feature reuse. Table II summarizes the major CNN architectures and their key contributions.

LeNet, proposed by LeCun *et al.*, is widely recognized as the pioneering CNN architecture for image recognition [19]. The model employed convolutional and pooling layers to learn hierarchical image features and demonstrated remarkable performance on handwritten digit classification. Although relatively shallow by modern standards, LeNet established the fundamental principles of convolutional feature extraction and parameter sharing that continue to underpin contemporary CNN designs.

A significant breakthrough occurred with AlexNet, which achieved outstanding performance in the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) and demonstrated the practical advantages of deep learning for large-scale visual recognition [20]. The architecture introduced Rectified Linear Unit (ReLU) activation, dropout regularization, and GPU-accelerated training, significantly reducing training time while improving classification accuracy.

Subsequent research focused on increasing network depth to improve feature learning. VGG networks adopted a simple yet effective strategy by stacking multiple small convolutional filters, enabling deeper architectures with enhanced representational capability [21]. In contrast, GoogLeNet introduced the Inception module, which employed parallel convolution operations of different sizes to improve computational efficiency while reducing the number of trainable parameters [22].

As network depth increased, optimization difficulties such as vanishing gradients became more pronounced. ResNet addressed this challenge by introducing residual learning through shortcut connections, allowing very deep networks to be trained effectively [23]. Building upon this concept, DenseNet established dense connectivity among layers, encouraging feature reuse and improving gradient propagation while reducing parameter redundancy [24].

Recent CNN architectures have emphasized efficiency and practical deployment. EfficientNet proposed a compound scaling strategy that jointly optimizes network depth, width, and input resolution, achieving superior accuracy with fewer parameters [25]. ConvNeXt further modernized CNN design by incorporating architectural refinements inspired by transformer models while preserving the efficiency of convolutional operations [26]. In parallel, Vision Transformers (ViTs) introduced self-attention mechanisms for image representation learning, offering competitive performance and motivating hybrid CNN-transformer architectures that combine the strengths of both paradigms [27].

Overall, the evolution of CNN architectures reflects a continuous effort to balance accuracy, computational complexity, and scalability. Early models established the foundations of deep visual learning, whereas modern architectures focus on efficient feature extraction and adaptability to diverse computer vision applications. These developments have significantly expanded the applicability of CNNs across numerous real-world domains and continue to influence the design of next-generation intelligent vision systems.

**TABLE II:** Representative CNN architectures and their major contributions.

| Model | Year | Key Contribution |
|---|---|---|
| LeNet | 1998 | Early CNN design |
| AlexNet | 2012 | Deep learning breakthrough |
| VGG | 2014 | Very deep architecture |
| GoogLeNet | 2015 | Inception modules |
| ResNet | 2016 | Residual learning |
| DenseNet | 2017 | Dense connectivity |
| EfficientNet | 2019 | Compound scaling |
| ConvNeXt | 2022 | Modernized CNN |
| Vision Transformer | 2021 | Self-attention learning |

Table II demonstrates that CNN development has progressed from improving classification capability to optimizing computational efficiency and representation learning. The complementary strengths of these architectures have laid the foundation for a wide range of computer vision applications, which are discussed in the following subsection.

### *B. CNN Applications in Computer Vision*

The continuous evolution of CNN architectures has substantially expanded their applicability across diverse computer vision tasks. Their ability to automatically learn hierarchical visual features from raw image data has reduced the dependence on handcrafted descriptors and enabled robust performance in complex real-world environments. Consequently, CNN-based models have become fundamental components of modern intelligent vision systems across healthcare, transportation, agriculture, surveillance, and industrial automation.

Image classification remains one of the earliest and most extensively studied applications of CNNs. Large-scale benchmark datasets such as ImageNet have facilitated the development of highly discriminative models capable of recognizing thousands of object categories with remarkable accuracy [28]. Advanced architectures including ResNet, DenseNet, and EfficientNet have significantly improved classification performance while reducing computational complexity, making them suitable for both cloud and edge computing environments.

Object detection extends image classification by simultaneously identifying object categories and their spatial locations within an image. CNN-based detectors can generally be divided into two-stage and one-stage approaches. Two-stage methods, represented by Faster R-CNN and Mask R-CNN, achieve high detection accuracy through region proposal mechanisms, whereas one-stage detectors such as the YOLO family provide real-time performance by directly predicting object locations and classes [29, 30]. These models have found extensive applications in intelligent transportation systems, robotics, and autonomous navigation.

Another important application area is semantic segmentation, where each pixel in an image is assigned a semantic label. Fully Convolutional Networks (FCNs), U-Net, and DeepLab architectures have demonstrated impressive capability in extracting detailed spatial information for scene understanding

 

**TABLE III:** Representative applications of CNNs in computer vision.

| Application | Primary Objective |
|---|---|
| Image Classification | Object recognition |
| Object Detection | Localization and detection |
| Semantic Segmentation | Pixel-level labeling |
| Medical Imaging | Disease diagnosis |
| Autonomous Driving | Scene understanding |
| Precision Agriculture | Crop monitoring |
| Intelligent Surveillance | Activity analysis |
| Industrial Inspection | Defect detection |

and medical image segmentation [31, 32]. These approaches have significantly improved the precision of image analysis tasks requiring accurate object boundary delineation.

Healthcare represents one of the most impactful application domains of CNNs. Deep learning models assist clinicians in detecting diseases from medical images, including chest X-rays, computed tomography (CT), magnetic resonance imaging (MRI), retinal scans, and dermoscopic images. CNN-based computer-aided diagnosis systems have shown promising results for identifying tumors, diabetic retinopathy, skin lesions, and pulmonary diseases while reducing diagnostic workload [33]. Their ability to process large volumes of medical data efficiently has accelerated the adoption of AI-assisted healthcare technologies.

CNNs have also contributed significantly to autonomous driving systems. Vehicle perception modules rely on deep convolutional models for traffic sign recognition, pedestrian detection, lane identification, obstacle avoidance, and scene understanding. By integrating information from cameras and other sensors, CNN-based frameworks support real-time decision making and enhance the safety and reliability of intelligent transportation systems [34].

In precision agriculture, CNN models facilitate automated crop monitoring, plant disease identification, weed detection, fruit counting, and yield prediction. The combination of aerial imagery, drone technology, and deep learning enables farmers to optimize resource utilization and improve agricultural productivity [35]. Similarly, intelligent surveillance systems employ CNNs for facial recognition, anomaly detection, crowd monitoring, and human activity analysis, contributing to enhanced public safety and security management [36].

Table III summarizes representative computer vision applications of CNNs and their primary objectives. The table highlights the versatility of CNN architectures across multiple domains and demonstrates their broad practical impact.

As summarized in Table III, CNNs have evolved into a versatile computational framework capable of addressing a broad spectrum of computer vision problems. Nevertheless, the suitability of a particular architecture often depends on application-specific requirements such as accuracy, computational efficiency, memory consumption, and real-time processing capability. Consequently, a systematic comparison of representative CNN models is essential for understanding their relative strengths and limitations. The following section presents a comparative analysis of prominent CNN architectures and their practical characteristics.

**TABLE IV:** Comparative analysis of representative CNN architectures.

| Model | Year | Advantage | Limitation |
|---|---|---|---|
| LeNet | 1998 | Simple | Limited depth |
| AlexNet | 2012 | High accuracy | Large model |
| VGG | 2014 | Deep features | High parameters |
| GoogLeNet | 2015 | Efficient | Complex design |
| ResNet | 2016 | Residual learning | High complexity |
| DenseNet | 2017 | Feature reuse | Memory cost |
| EfficientNet | 2019 | Efficient scaling | Tuning required |
| ConvNeXt | 2022 | Modern CNN | Training cost |
| ViT | 2021 | Global attention | Large data demand |

# IV Comparative Analysis

The rapid evolution of CNN architectures has introduced significant improvements in feature extraction, learning capability, and computational efficiency. However, no single architecture is universally optimal for all computer vision tasks, as model selection depends on factors such as accuracy requirements, computational resources, memory constraints, and real-time processing demands. Therefore, a comparative analysis is essential for understanding the relative advantages and limitations of representative CNN models.

Early architectures such as LeNet and AlexNet established the foundation of deep visual learning but were primarily designed for image classification problems with relatively simple structures. VGG improved feature representation by increasing network depth through stacked convolutional layers, although its large number of parameters increased computational cost. GoogLeNet addressed this limitation by introducing Inception modules that efficiently captured multi-scale features while reducing model complexity. Subsequently, ResNet employed residual connections to overcome degradation problems associated with very deep networks, whereas DenseNet further improved feature propagation through dense layer connectivity [37, 38].

Recent architectures have focused on balancing predictive performance and computational efficiency. EfficientNet introduced compound scaling to optimize network depth, width, and input resolution simultaneously, achieving competitive accuracy with fewer parameters. ConvNeXt modernized conventional CNN design by incorporating architectural refinements inspired by transformer models while preserving the advantages of convolutional operations. In parallel, Vision Transformers have demonstrated strong representation learning capabilities through self-attention mechanisms, particularly for large-scale datasets [39, 40].

Table IV provides a concise comparison of representative CNN architectures based on their principal characteristics. The comparison indicates that modern CNN models generally achieve improved accuracy and efficiency while reducing computational overhead, making them suitable for practical deployment across diverse application domains.

As summarized in Table IV, the progression of CNN architectures reflects a shift from maximizing predictive accuracy toward achieving an effective balance between performance, scalability, and computational efficiency. While traditional CNNs remain highly effective for many computer vision

 

tasks, recent developments indicate an increasing integration of convolutional and attention-based mechanisms. Despite these advances, challenges related to data dependency, model interpretability, computational requirements, and robust deployment persist. These limitations motivate further investigation into existing research gaps and open challenges, which are discussed in the following section.

## V Research Gaps and Challenges

Despite the remarkable progress of Convolutional Neural Networks (CNNs) in computer vision, several research challenges continue to limit their widespread adoption and practical deployment. Although modern architectures have achieved impressive performance on benchmark datasets, many models remain highly dependent on large volumes of annotated training data. The collection and labeling of high-quality datasets are often expensive, time-consuming, and domain-specific, particularly in fields such as medical imaging and precision agriculture [41].

Another important challenge is the computational complexity of deep CNN models. State-of-the-art architectures generally require substantial memory, processing power, and energy consumption, making their deployment on edge devices and embedded systems difficult. While lightweight architectures have improved efficiency, achieving an effective balance between accuracy and computational cost remains an active research problem [42].

Model interpretability and robustness also require further investigation. CNNs are frequently regarded as black-box models, making it difficult to understand their decision-making process in safety-critical applications. In addition, adversarial perturbations and variations in real-world environments can significantly degrade model performance, raising concerns regarding reliability and security.

Furthermore, the emergence of Vision Transformers and hybrid CNN-transformer architectures has created new research opportunities while introducing challenges related to model selection, scalability, and integration strategies. Developing efficient, interpretable, and robust vision models capable of learning from limited data and adapting to diverse environments remains an open research direction. Addressing these challenges is essential for advancing next-generation computer vision systems and expanding the practical impact of CNN-based technologies.

## VI Future Research Directions

The future of Convolutional Neural Networks (CNNs) for computer vision is expected to focus on improving efficiency, adaptability, and trustworthiness while addressing existing practical limitations. One promising direction is the development of hybrid CNN–Transformer architectures that combine the local feature extraction capability of convolutional operations with the global contextual modeling provided by self-attention mechanisms. Such frameworks have the potential to achieve superior performance across a wide range of visual recognition tasks.

Another important research trend involves lightweight and edge-deployable CNN models for resource-constrained environments. Model compression, neural architecture search, and knowledge distillation can facilitate real-time inference in mobile devices, autonomous systems, and Internet of Things (IoT) applications. In addition, self-supervised and few-shot learning techniques may reduce the dependence on large labeled datasets, enabling efficient model training in domains where annotated data are limited.

Future investigations should also emphasize explainable and trustworthy AI to improve model transparency and user confidence, particularly in safety-critical applications such as healthcare and autonomous driving. Furthermore, robust learning strategies capable of mitigating adversarial attacks and adapting to dynamic environments remain essential for practical deployment. Collectively, these research directions are expected to enhance the scalability, reliability, and societal impact of CNN-based computer vision systems.

## VII Conclusion

Convolutional Neural Networks (CNNs) have fundamentally transformed computer vision by enabling automated hierarchical feature learning and achieving state-of-the-art performance across numerous visual recognition tasks. This review has presented a comprehensive overview of the evolution of CNN architectures, highlighting the progression from early models such as LeNet to modern efficient architectures and transformer-inspired frameworks. Their applications in image classification, object detection, semantic segmentation, healthcare, autonomous driving, agriculture, and intelligent surveillance demonstrate the versatility and practical significance of CNN-based approaches.

A comparative analysis of representative architectures indicates that recent developments have improved both predictive capability and computational efficiency, although challenges related to data dependency, interpretability, robustness, and deployment complexity remain. Emerging trends, including lightweight CNNs, hybrid CNN–Transformer models, self-supervised learning, edge intelligence, and explainable AI, provide promising opportunities for overcoming these limitations.

Overall, CNNs are expected to remain a fundamental component of intelligent vision systems while evolving alongside emerging deep learning paradigms. The findings presented in this review provide a consolidated reference for researchers and practitioners and may support the development of more efficient, reliable, and adaptable computer vision technologies for future real-world applications.

## References

[1] D. G. Lowe, ”Distinctive image features from scale-invariant keypoints,” Int. J. Comput. Vis., vol. 60, no. 2, pp. 91–110, 2004.

 

[2] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in Proc. CVPR, 2005, pp. 886–893.

[3] Y. LeCun *et al.*, "Gradient-based learning applied to document recognition," Proc. IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.

[4] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," Commun. ACM, vol. 60, no. 6, pp. 84–90, 2017.

[5] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in Proc. ICLR, 2015.

[6] K. He *et al.*, "Deep residual learning for image recognition," in Proc. CVPR, 2016.

[7] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proc. ICML, 2019.

[8] Z. Liu *et al.*, "A ConvNet for the 2020s," in Proc. CVPR, 2022.

[9] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," Med. Image Anal., vol. 42, pp. 60–88, 2017.

[10] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," Comput. Electron. Agric., vol. 147, pp. 70–90, 2018.

[11] W. Rawat and Z. Wang, "Deep convolutional neural networks for image classification: A comprehensive review," Neural Comput., vol. 29, no. 9, pp. 2352–2449, 2017.

[12] A. Garcia-Garcia *et al.*, "A review on deep learning techniques applied to semantic segmentation," 2017.

[13] A. Dosovitskiy *et al.*, "An image is worth 16×16 words: Transformers for image recognition at scale," in Proc. ICLR, 2021.

[14] N. Akhtar and A. Mian, "Threat of adversarial attacks on deep learning in computer vision," IEEE Access, vol. 6, pp. 14410–14430, 2018.

[15] M. J. Page *et al.*, "The PRISMA 2020 statement: An updated guideline for reporting systematic reviews," BMJ, vol. 372, 2021.

[16] B. Kitchenham and S. Charters, "Guidelines for performing systematic literature reviews in software engineering," Keele Univ., Tech. Rep., 2007.

[17] O. Russakovsky *et al.*, "ImageNet large scale visual recognition challenge," Int. J. Comput. Vis., vol. 115, no. 3, pp. 211–252, 2015.

[18] T.-Y. Lin *et al.*, "Microsoft COCO: Common objects in context," in Proc. ECCV, 2014, pp. 740–755.

[19] Y. LeCun *et al.*, "Gradient-based learning applied to document recognition," Proc. IEEE, vol. 86, no. 11, pp. 2278–2324, 1998.

[20] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," Commun. ACM, vol. 60, no. 6, pp. 84–90, 2017.

[21] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in Proc. ICLR, 2015.

[22] C. Szegedy *et al.*, "Going deeper with convolutions," in Proc. CVPR, 2015.

[23] K. He *et al.*, "Deep residual learning for image recognition," in Proc. CVPR, 2016.

[24] G. Huang *et al.*, "Densely connected convolutional networks," in Proc. CVPR, 2017.

[25] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proc. ICML, 2019.

[26] Z. Liu *et al.*, "A ConvNet for the 2020s," in Proc. CVPR, 2022.

[27] A. Dosovitskiy *et al.*, "An image is worth 16×16 words: Transformers for image recognition at scale," in Proc. ICLR, 2021.

[28] O. Russakovsky *et al.*, "ImageNet large scale visual recognition challenge," Int. J. Comput. Vis., vol. 115, no. 3, pp. 211–252, 2015.

[29] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 6, pp. 1137–1149, 2017.

[30] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," arXiv:1804.02767, 2018.

[31] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in Proc. CVPR, 2015.

[32] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in Proc. MICCAI, 2015.

[33] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," Med. Image Anal., vol. 42, pp. 60–88, 2017.

[34] M. Grigorescu *et al.*, "A survey of deep learning techniques for autonomous driving," J. Field Robot., vol. 37, no. 3, pp. 362–386, 2020.

[35] A. Kamilaris and F. X. Prenafeta-Boldú, "Deep learning in agriculture: A survey," Comput. Electron. Agric., vol. 147, pp. 70–90, 2018.

[36] M. Chaudhary, A. Mittal, and G. Battineni, "A review of deep learning approaches for video surveillance," Electronics, vol. 10, no. 19, 2021.

[37] G. Huang *et al.*, "Densely connected convolutional networks," in Proc. CVPR, 2017.

[38] K. He *et al.*, "Deep residual learning for image recognition," in Proc. CVPR, 2016.

[39] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proc. ICML, 2019.

[40] Z. Liu *et al.*, "A ConvNet for the 2020s," in Proc. CVPR, 2022.

[41] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," Nature, vol. 521, no. 7553, pp. 436–444, 2015.

[42] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proc. ICML, 2019.


 