Proceedings of CONF-SPML 2026 Symposium: The 2nd Neural Computing and Applications Workshop 2025
DOI: 10.54254/2755-2721/2026.TJ29438

Different Task Paradigms and Representative Methods in Image Recognition

Kaifei Zhang

Information and Computer Science, Jinan University, Guangzhou, China
zhangkf261106@163.com

Abstract. Image recognition has a wide range of applications and is essential to many facets of contemporary life, including surveillance and healthcare. In three basic tasks, this work attempts to fully review image recognition paradigms: image classification, object detection, and image segmentation. In each task, this paper describes the core objectives and technical challenges and focuses on exploring representative methods—including convolutional neural network architectures of models like AlexNet and ResNet for image classification, the YOLO algorithm for object detection, and models for image segmentation. Furthermore, this paper makes comparisons to illustrate the respective advantages and disadvantages of traditional methods and deep learning methods, providing useful references for technical selection in different scenarios. The integration of deep learning and image recognition has facilitated the improvement of people's quality of life and is expected to become even more essential in the future.

Keywords: Image Recognition, Image Classification, Object Detection, Image Segmentation.

1. Introduction

In the current wave of the digital revolution, automatic image recognition has become a crucial technology. It enables computers to analyze and interpret digital images through a process that is highly similar to the visual perception mechanism of humans. Its application scenarios have widely covered multiple fields, including healthcare, transportation, and industrial automation. In healthcare, DNA molecular computing-based Convolutional Neural Networks (CNNs) enable parallel recognition of medical images [1]. In the transportation sector, vehicle detection on edge devices is supported by lightweight models [2]. In industrial automation, power equipment recognition networks boost the effectiveness of inspections. These uses highlight the fundamental benefits of the technology in intelligent evolution [3].

The growing demand for automated, effective, and intelligent perception systems is the main factor driving the quick development and broad use of image recognition technology. Nevertheless, there are many inherent difficulties in the process of going from raw pixel data to high-level semantic interpretation. The inherent complexity of object shapes, the conceptual divide between low-level features and high-level concepts, and the notable differences in viewpoint, illumination, scale, and occlusion within images are some of these difficulties. The field has had a methodological

© 2025 The Authors. This is an open access article distributed under the terms of the Creative Commons Attribution License 4.0 (https://creativecommons.org/licenses/by/4.0/).

DOI: 10.54254/2755-2721/2026.TJ29438

evolution to get over these challenges. Early image recognition systems mainly relied on hand-crafted characteristics—like Scale-Invariant Feature Transform (SIFT) and Histogram of Oriented Gradients (HOG) [4][5]. Although these techniques established an excellent foundation, they frequently had limited representational and generalization capabilities. The revival of deep learning, particularly the advancement of Convolutional Neural Networks (CNNs), brought about a profound paradigm change. In numerous benchmark tests, CNNs have shown an incredible capacity to autonomously learn structures of features from data, resulting in unprecedented efficiency improvements. In addition to replacing many conventional approaches, the deep learning revolution created a vast and varied ecosystem that includes network designs, learning paradigms, and optimization strategies suited to various recognition tasks. A complicated and fragmented landscape has been created, despite the fact that the proliferation of approaches demonstrates the vibrancy of this research field. It can be difficult for academics and practitioners to sort through this abundance of methods, comprehend how they relate to one another, and ascertain which ones are most appropriate for particular roles and uses. Consequently, there is a pressing need for a structured synthesis that organizes this knowledge, providing a clear roadmap of the fundamental tasks and a balanced analysis of the evolution from classical to contemporary methods.

Therefore, the primary contribution of this paper lies in providing a detailed and systematic taxonomy of fundamental image recognition tasks. It offers a comprehensive examination from three critical perspectives: image classification, object detection, and image segmentation. For every assessment, the study conducts a thorough comparative analysis, meticulously outlining the strengths and limitations of traditional algorithms versus modern, prevalent approaches, while also highlighting key innovative breakthroughs. Furthermore, the research synthesizes the discussion by illustrating practical applications across diverse domains, thereby contextualizing the technological advancements. This structured overview aims to serve as a valuable reference framework, contributing to the foundation for future exploration and development within the field of image recognition.

2. Image classification

2.1. Definition and challenges

Image classification refers to the task in which a computer assigns a predefined category to an input digital image. More precisely, given a digital image composed of pixels, the system produces one or more categorical labels, each accompanied by a confidence score. There are some challenges related to image classification, like viewpoint variation, illumination condition, occlusion, background clutter, and intra-class variation.

2.2. Methods

Most image classification implementations follow a fundamental workflow, with each stage closely connected to guarantee the final system's precision and dependability. This workflow typically includes data preparation and preliminary processing, extraction of features, model building and learning, and assessment of model, each detailed below.

The process begins with data preparation and preprocessing, which involves constructing large-scale labeled datasets and applying standardization techniques. Pixels are normalized and images are shrunk to a predetermined size. Sometimes, to prevent overfitting, training images are randomly transformed, such as by rotation, brightness adjustment, and the inclusion of a tiny bit of noise.

DOI: 10.54254/2755-2721/2026.TJ29438

Feature extraction represents a crucial stage, aiming to reduce unnecessary data and extract high-level information or features from raw pixels that reflect image content. For instance, A fundamental feature extraction method called Scale-Invariant Feature Transform (SIFT) was created to accurately locate and characterize keypoints in images under various circumstances [4].

Model construction forms the core of the image classification systems. The model is the brain of the entire system, responsible for extracting features and making judgments. Currently, deep learning methods are becoming mainstream, with the core approach being to use convolutional neural networks (CNNs) to automatically acquire high-level to low-level layers' hierarchical features. AlexNet demonstrated breakthrough performance through its systematic integration of ReLU and Dropout, which became industry standards for improving training efficiency and generalization [6]. An advancement came with ResNet, which introduced the residual block, which reframes the learning objective through a simple skip connection to facilitate the gradient problem and enable the stable training of very deep networks [7].

The model training phase is an iterative learning process aimed at finding optimal values for the model parameters. Configuring core model parameters, such as the loss function and optimizer, controls the speed and direction of the training process.

Use a test set that the model has never seen before to assess its performance. Then, apply the trained model to real-world applications to classify new images.

All things considered, image classification allows machines to learn and comprehend an image's essential content. The foundation of more complex vision tasks like object detection and image segmentation is this well-established and extensively utilized technique.

3. Object detection

3.1. Definition and challenges

The limitations of conventional image classification, which only predicts one label for the entire image, were addressed by object detection. The object detection task facilitates an organized comprehension of image content by identifying object categories and determining their spatial locations. Because of this feature, object detection is a fundamental technique in fields including surveillance, autonomous driving, and medical imaging. Its main drawbacks, however, are its inability to identify small objects well and the inherent trade-off between accuracy and detection speed.

3.2. Methods

The multi-stage pipeline based on the multi-scale sliding window approach was the defining feature of the traditional object detection paradigm. In order to suggest possible object regions, windows of preset sizes are applied at various locations and scales in an exhaustive search over the image. A foundation for its robustness in encoding shape and edge information is the Histogram of Oriented Gradients (HOG) [5], which can define each of the generating candidate windows using hand-crafted features. To accomplish object-background separation, these feature vectors are then categorized. In the final phase, overlapping detection findings are filtered away using specific methods.

Since the breakthrough of AlexNet on ImageNet in 2012, deep learning has become the dominant approach in object detection. Convolutional neural networks (CNNs) automatically learn high-level semantic characteristic and enable end-to-end optimization, integrating feature extraction and

DOI: 10.54254/2755-2721/2026.TJ29438

candidate region proposal as well as classification and regression. As indicated in Table 1, deep learning-based detection techniques have roughly two paradigms: two-stage detection and one-stage detection, depending on whether an explicit candidate proposal phase is included.

Both Two-Stage and One-Stage Detection techniques have established classic implementations within the field. Faster R-CNN [8], which has long dominated object detection, is a notable example of the two-stage system in action. To create candidate regions, it initially introduces a Region Proposal Network (RPN). A different deep network is then used to categorize and improve these recommendations. In contrast, aiming for higher computational efficiency, one-stage detectors streamline the process by predicting both bounding boxes and class probabilities, removing the need for a separate region proposal step. The fundamental advancement of these deep learning-based detectors is their capacity to perform end-to-end detection by leveraging convolutional neural networks. This allows them to automatically learn multi-scale, hierarchical feature representations—a capability that significantly surpasses the limitations of handcrafted features in traditional methods.

Table 1. Difference between two-stage detection and one-stage detection

School	Core procedure
Two-Stage Detection	High-quality candidate regions are initially produced using the "rough selection first, then fine-tuning" approach. Following this, the candidate regions are categorized and bounding box regression is carried out.
One-Stage Detection	Abandoning the candidate region generation step, the image network is directly densely predicted through CNN, and each grid simultaneously outputs the target category probability and bounding box coordinates.

In general, object detection inherits the feature extraction and classification capabilities of image classification and provides a basis for image segmentation and some complicated assessments. Today, with the continuous upgrading of models, object detection will further break through the limitations of scenarios and play a greater role in more practical products.

4. Image segmentation

4.1. Definition and challenges

Image segmentation is a technique of splitting a picture into distinct areas or objects, each carrying specific semantic information. Unlike basic cutting out of image parts, giving each pixel in the picture a meaningful label—which could stand for a class, an instance, or another type of identification—is the aim. Depending on the required detail, there are three primary categories for image segmentation. Semantic segmentation involves classifying each pixel without separating different individual objects belonging to the same category. Instance segmentation also separates distinct occurrences of objects within the same class. Global segmentation integrates the previous two approaches, which not only differentiates object instances but also identifies background elements, making it a more thorough and demanding segmentation task.

4.2. Methods

The evolution of image segmentation technology reflects a shift from traditional algorithms to deep learning frameworks. Traditional approaches primarily partition images based on low-level visual features of pixels, such as intensity and texture. Threshold-based segmentation separates the foreground from the background by comparing pixel intensity values against one or more predefined

DOI: 10.54254/2755-2721/2026.TJ29438

thresholds. A key limitation of this approach is its heavy reliance on the histogram's shape and its use of grayscale information alone. A representative technique is the Maximum Between-Class Variance Method (Otsu's method), which determines the ideal threshold through variance maximization between the foreground and background classes. This algorithm is automatic, requires no manual intervention, and is computationally efficient [9]. Another category of methods treats image segmentation as a data clustering problem in a feature space, where each pixel is considered a data point. Algorithms like K-Means are employed to group these points. K initial cluster centers are chosen at random at the beginning of the K-Means algorithm. In order to accomplish segmentation, it then allocates each pixel to the closest center and repeatedly recalculates the centers until convergence. A significant drawback is the need to predefine the number of clusters K, and the results are often sensitive to the initial selection of centers [10].

Deep learning has changed image segmentation. These methods typically employ an encoder-decoder architecture. Fully Convolutional Networks (FCN) replaced all fully connected layers in traditional networks with convolutional layers, allowing the network to receive input photos of any size and produce correspondingly sized segmentation maps. In addition, it used deconvolutional layers for upsampling to counteract the significant reduction in feature map size caused by pooling. FCN also introduced skip connections, which combine deep, semantically rich feature maps with shallow feature maps that possess precise spatial information. This fusion helps improve the accuracy of segmentation boundaries while retaining high-level semantics. However, a limitation of FCN is that the upsampling process can still result in relatively coarse segmentation results [11]. U-Net builds upon FCN's fully convolutional concept and skip connections but introduces important enhancements. Its encoder-decoder structure is symmetric, making the architecture clear, elegant, and easy to implement. Furthermore, U-Net's skip connections differ from those in FCN; instead of simple summation, it concatenates feature channels from each encoder stage with the corresponding decoder stage. This provides the decoder with more extensive and complete low-level feature information during upsampling, leading to significantly better boundary recovery capabilities compared to FCN [12].

Overall, Image segmentation is a basic technology in computer domain, enabling machines to progress from seeing an image to understanding each pixel of the image. The advancement of picture segmentation technology, from conventional techniques based on low-level features to current end-to-end models centered on deep learning, has significantly aided the practical implementation of artificial intelligence.

5. Application

5.1. Medical field

Image recognition can focus on the parallel recognition of multimodal medical images and the privacy protection in medical scenarios. For example, X-ray images of illnesses like breast cancer, skin cancer, and pneumonia can be recognized using a convolutional neural network that has a DNA-level that utilizes strand displacement reaction. Its high parallelism supports the synchronous analysis of multimodal images (such as X-ray and ultrasound images), providing a technical basis for the joint diagnosis of multiple diseases [1]. In addition, in healthcare settings, patients' facial information can be used for identity matching during medical consultations, but there exist risks of privacy breaches. Therefore, an ε-noise-based privacy-preserving system is proposed to address the risk by adding ε-noise, satisfying real-time identity verification needs in healthcare while complying with privacy regulations.

DOI: 10.54254/2755-2721/2026.TJ29438

5.2. Surveillance system

The application of image recognition in surveillance scenarios lies in the efficient processing of large-scale surveillance data and the protection of facial privacy. For example, surveillance systems in regions such as urban road monitoring often generate millions or even billions of frames of images. Distributed training and cloud resources meet the needs of real-time surveillance on the big image data such as tracking suspicious vehicles. This application not only solves the problem of processing lag caused by the large amount of data in surveillance scenarios but also is suitable for scenarios such as large transportation hubs and urban real-time monitoring [2].

5.3. Retail and dynamic recognition

Image recognition technology offers innovative solutions for dynamic environments with frequently updated requirements, such as smart retail. When adding new product categories to the traditional model, it incurs high annotation costs and forgets the old products. By using Spatial Global Enhancement to capture semantic features, Spatial Local Enhancement to retain detailed features, and Multi-Head Interaction Block to fuse features, a two-dimensional feature memory is constructed. This application effectively alleviates catastrophic forgetting and provides a feasible solution for low-data scenarios, such as medical imaging and rare object recognition [13].

6. Conclusion

With the combination of deep learning and image recognition, image recognition is undergoing a rapid and significant evolution. This review introduces the basic concepts of image recognition and its tasks classification. It also lists the benefits and drawbacks of both deep learning and conventional approaches, and compares them. Regarding different algorithms, this paper briefly describes their roles and significance in the domain of image recognition and the reasons why they can achieve advanced performance in various tasks. In addition, for different fields, this paper discusses several common applications of image recognition in real life, especially in the medical and surveillance fields. By combining with deep learning or other advanced models, there is still room for further innovation in the research of image recognition in more fields, which can further transform the capabilities of image recognition and give rise to more significant and meaningful applications.

References

[1] Li, S.Y., Yin, Z.X., Chen, C.L., et al. (2025) A DNA-Level Convolutional Neural Network Based on Strand Displacement Reaction for Image Recognition. IEEE Access, 13, 139705-139718.
[2] Puisamlee, W. and Chawuthai, R. (2025) Minimizing Model Size of CNN-Based Vehicle Make Recognition for Frontal Vehicle Images. IEEE Access, 13, 97409-97420.
[3] Lin, S. (2025) Power Equipment Image Recognition Method Based on Feature Extraction and Deep Learning. IEEE Access, 13, 130431-130444.
[4] Lowe, D.G. (2004) Distinctive image features from scale-invariant keypoints. International Journal of Computer Vision, 60(2), 91-110.
[5] Dalal, N. and Triggs, B. (2005) Histograms of Oriented Gradients for Human Detection. 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 886-893.
[6] Krizhevsky, A., Sutskever, I. and Hinton, G.E. (2012) ImageNet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems, 25, 1097-1105.
[7] He, K., Zhang, X., Ren, S., et al. (2016) Deep residual learning for image recognition. Proceedings of the IEEE conference on computer vision and pattern recognition, 770-778.

DOI: 10.54254/2755-2721/2026.TJ29438

[8] Ren, S., He, K., Girshick, R., et al. (2015) Faster r-cnn: Towards real-time object detection with region proposal networks. Advances in neural information processing systems, 28, 91-99
[9] Otsu, N. (1979) A threshold selection method from gray-level histograms. IEEE Transactions on Systems, Man, and Cybernetics, 9(1), 62-66.
[10] MacQueen, J.B. (1967) Some methods for classification and analysis of multivariate observations. Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, Le Cam, L.M. and Neyman, J. (Eds.), University of California Press, Berkeley, 281-297.
[11] Long, J., Shelhamer, E. and Darrell, T. (2015) Fully Convolutional Networks for Semantic Segmentation. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), IEEE, 3431-3440.
[12] Ronneberger, O., Fischer, P. and Brox, T. (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation. Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, Springer, Cham, 234-241.
[13] Wu, H., Ng, Z.Y., Zheng, Z.J., et al. (2025) Few-Shot Incremental Learning With Context-Aware Spatial Enhancement for Image Recognition. IEEE Access, 13, 110569-110583.