Proceedings of the 4th International Conference on Software Engineering and Machine Learning
DOI: 10.54254/2755-2721/2026.34232

Analysis of Artificial Intelligence Applications in Graphic Recognition

Cheng Zhang

School of Science and Technology, University of Macau, Macau, China
gary79709@outlook.com

Abstract. With the rapid development of artificial intelligence, graphic recognition has become a core and indispensable branch of computer vision, serving as a key support for machines to perceive and understand visual information. This paper systematically combs the whole development process of graphic recognition, including traditional machine learning with manual feature extraction, Convolutional Neural Network, Transformer and emerging multimodal large models. It expounds the connotation and research value of core tasks such as image classification, object detection, and image segmentation, and summarizes representative mainstream algorithms and their technical characteristics. Meanwhile, this paper deeply analyzes typical application scenarios of graphic recognition in security monitoring, medical imaging, automatic driving, industrial inspection, smart agriculture and e-commerce retail. The results show that the overall performance of graphic recognition technology has been continuously improved in accuracy and generalization, but there are still prominent bottlenecks such as labeled data scarcity, weak environmental robustness and insufficient model interpretability. Finally, this paper summarizes the full research content and prospects the future development directions including few-shot learning, lightweight model optimization and multimodal deep fusion.

Keywords: Graphic Recognition, Artificial Intelligence, Deep Learning, Computer Vision, Multimodal Model

1. Introduction

In the current era of rapid iteration of artificial intelligence technology, graphic recognition, as a core research direction in the field of computer vision, is a key support for machines to "understand the world". Its essence is to analyze, process, and interpret image information through computer algorithms, and complete related tasks such as target recognition, localization, and segmentation. Graphic recognition consists of three closely related core tasks: image classification is used to determine the category of an image, object detection is used to achieve the recognition and localization of target objects, and image segmentation is used to complete pixel level region division. These three tasks progress layer by layer and together form the core system of graphic recognition technology.

The rapid development of graphic recognition is attributed to the synergistic effect of the three core driving forces of computing power, data, and algorithms: the popularity of dedicated computing

© 2026 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: 10.54254/2755-2721/2026.34232

chips such as GPUs and TPUs, which provide hardware support for large-scale data processing and model iteration; Internet technology promotes the emergence of large-scale labeled image data sets, providing data guarantee for algorithm training and optimization; The evolution from manual feature extraction to deep learning, and then to multimodal large models, has driven a leapfrog improvement in technical accuracy and generalization ability.

This technology has penetrated into multiple fields such as security, healthcare, autonomous driving, e-commerce, etc., achieving intelligent upgrading and digital transformation. Therefore, it is of great theoretical significance and practical value to systematically sort out its development trajectory, explore its current application status and future trends.

The development of graphic recognition technology can be roughly divided into four stages: from 2000 to 2012, it was the stage of manual feature+traditional machine learning, relying on manually designed SIFT, HOG and other feature operators and SVM models, with limited ability to adapt to complex scenes; The period from 2012 to 2017 was the CNN stage, during which AlexNet ushered in the era of deep learning [1], and models such as VGG and ResNet promoted the large-scale application of technology [2]; The period from 2017 to 2021 is the Transformer stage, where ViT breaks the limitations of CNN [3] and Swin Transformer and other variants optimize performance [4]; Since 2021, it has been the stage of multimodal large-scale modeling, where CLIP, SAM, and other technologies have achieved multimodal fusion and expanded task boundaries [5, 6].

Horizontally speaking, the three core tasks are advancing synchronously, and the current research hotspots are focused on visual model adaptation, small sample learning, and other directions, facing bottlenecks such as insufficient generalization ability and weak interpretability. Foreign countries started early in model architecture and multimodal integration, while China focuses on lightweighting large models, localization adaptation, and industry implementation. Although there are application achievements in the medical and security fields, there is still significant research space overall.

This article adopts a dual framework of "horizontal task dimension+vertical technology evolution" to systematically sort out the development of graphic recognition, focusing on analyzing algorithm innovation and performance breakthroughs in different technological stages, filling the gap of existing single dimensional reviews, and improving the technical system. The core contribution is to construct a dual dimensional framework and clarify the rules of technology and task adaptation; Supplement research on visual macro models; Integrate literature, summarize application experience and shortcomings, and provide guidance for subsequent research and implementation.

2. Basic theory and technological evolution of graphic recognition

2.1. Definition of core task system

The task hierarchy of graphic recognition progresses from shallow to deep, forming a complete closed loop of "classification detection segmentation extension task". Image classification is a fundamental task that focuses on determining what it is, extending from coarse-grained classification to fine-grained recognition. It is the foundation of complex tasks and requires high model generalization ability. Object detection combines classification with positioning function to solve the problem of "what+where". It requires balancing positioning accuracy and inference speed, and adapting to real-time scenarios such as security and autonomous driving. Image segmentation achieves pixel level fine resolution, including semantic, instance, and panoramic segmentation, with extremely high precision requirements. It is the core technology for medical imaging and industrial

DOI: 10.54254/2755-2721/2026.34232

quality inspection. Extended tasks include keypoint detection, target tracking, etc., further expanding the application boundaries.

2.2. Stages of technological development and core principles

The iteration of graphic recognition technology is essentially a transition from "manual feature extraction" to "automatic feature learning". The traditional manual feature stage (before 2000 to 2012) relied on manually designed feature operators, combined with traditional machine learning models, which were lightweight but had poor generalization ability. During the CNN phase (2012-2017), end-to-end automatic feature extraction was achieved, with classic models continuously optimized and various specialized algorithms derived, promoting the practical application of technology [1, 2]. The Transformer stage (2017-2021) relies on attention mechanism to achieve global modeling, solving the limitations of CNN local modeling and significantly improving accuracy [3, 4]. The stage of multimodal large model (2021 present) integrates multimodal information, possesses zero sample learning ability, and promotes the advancement of technology towards generalization and intelligence [5, 6].

3. Classification and technical characteristics of mainstream algorithms for graphic recognition

3.1. Classic algorithm system based on CNN

CNN algorithm has strong stability and low deployment cost, and is still the mainstream in the industry. In image classification networks, AlexNet laid the foundation for deep learning framework, improving the accuracy of ImageNet image classification to over 83% [1]; ResNet solves the problem of deep network training through residual connections, achieving a Top-1 accuracy of around 92% in similar tasks, significantly better than early network models [2]. Object detection algorithms are divided into two-stage and single-stage. Two stage algorithms (Faster R-CNN, etc.) have high accuracy and are used on the COCO dataset mAP@0.5 Up to 78% or more, suitable for high-precision scenarios [7]; Single stage algorithms (such as YOLO series) have fast inference speed and frame rates of up to 55-60 FPS, which can meet real-time detection requirements [8]. In image segmentation algorithms, FCN achieves end-to-end segmentation [9], U-Net has a Dice coefficient of over 89% in medical image segmentation [10], DeepLab improves edge segmentation accuracy through dilated convolution, and mIoU exceeds 80% in urban road scene segmentation.

3.2. Frontier algorithms based on transformer

Visual Transformer relies on global attention mechanism to overcome the shortcomings of CNN's local modeling. ViT breaks down images into patch sequences and achieves higher classification accuracy than CNN of the same scale on large-scale datasets [3]; Swin Transformer introduces hierarchical sliding windows to reduce computational complexity, achieving an accuracy of over 93% in image classification tasks [4]. DETR abandons candidate box design, simplifies the detection process, and achieves better detection accuracy than traditional two-stage models on the COCO dataset [11]; The Mask2Former unified segmentation framework takes into account semantic, instance, and panoramic segmentation, and has reached a leading level in both general and medical segmentation datasets, with strong generality [12].

DOI: 10.54254/2755-2721/2026.34232

3.3. Lightweight and large model adaptation algorithms

For scenarios where computing power is limited on mobile and edge devices, model lightweighting is achieved through techniques such as pruning and quantization. MobileNet and other lightweight models have only 1/10 of the parameter count of traditional CNNs, significantly improving inference speed with minimal accuracy loss and stable operation on low-power devices [13]. Visual big models reduce deployment costs and achieve rapid adaptation of general models to industry scenarios through techniques such as parameter efficient fine-tuning (PEFT), while maintaining high recognition accuracy even under small sample conditions [5, 6].

4. Implementation analysis of typical application scenarios for graphic recognition

4.1. Security monitoring field

Security monitoring is the core scenario where graphic recognition technology was first commercialized on a large scale. Relying on mature algorithms for object detection, facial recognition, pedestrian re identification, and behavior analysis, modern security systems can intelligently monitor public areas such as urban roads, park communities, transportation hubs, and rail transit around the clock. By combining the YOLO series lightweight detection model with facial feature extraction network, the device can quickly capture and compare targets in complex environments such as complex lighting, crowded crowds, and long-distance shooting, achieving automatic detection of risk events such as stranger warning, blacklist personnel recognition, cross-border intrusion, fireworks hazards, abnormal gatherings, and violent behavior. Compared with traditional manual patrol monitoring, graphic recognition technology effectively solves problems such as manual duty fatigue, high missed detection rate, and delayed response. With millisecond level recognition speed and lower false alarm rate, it greatly improves the automation level and prevention and control efficiency of public safety management, providing key technical support for smart city construction, social security control, and regional security governance.

4.2. Medical imaging analysis field

In the modern smart healthcare system, graphic recognition has become the core technology for medical imaging assisted diagnosis. Medical imaging includes various types such as CT, MRI, X-ray, ultrasound, and pathological sections. It is commonly characterized by small lesion areas, blurred boundaries, complex backgrounds, and scarce samples. Based on specialized segmentation networks such as U-Net and DeepLab, as well as optimized feature extraction algorithms, graphic recognition can complete refined tasks such as organ region segmentation, lesion localization, tumor screening, lesion tissue annotation, fundus lesion recognition, and cellular pathological analysis [10]. The model can accurately distinguish between normal tissues and abnormal lesion structures, quantify lesion size, shape, and distribution range, assist radiologists and pathologists in quickly reading images, and shorten diagnostic time. At the same time, artificial intelligence graphic recognition technology can effectively alleviate the practical problems of insufficient primary healthcare resources and a large shortage of professional physicians, reduce misdiagnosis and missed diagnosis caused by subjective judgments, and have important clinical value in early disease screening, preoperative evaluation, postoperative follow-up, and long-term monitoring of chronic diseases, promoting the comprehensive popularization of precision medicine and smart medicine.

DOI: 10.54254/2755-2721/2026.34232

4.3. Autonomous driving field

Autonomous driving and intelligent in vehicle perception systems highly rely on graphic recognition technology to achieve environmental perception and scene understanding, and are key underlying technologies for the development of intelligent transportation. The vehicle's onboard visual camera continuously collects real-time images of the road and completes various visual tasks through high-performance graphic recognition algorithms, including traffic signal recognition, traffic sign classification, lane line segmentation, road obstacle detection, pedestrian and non motorized vehicle recognition, vehicle target tracking, and complex road panoramic semantic segmentation [7, 8]. Faced with practical challenges such as low light at night, severe weather conditions such as rain, snow, and fog, backlit reflections, target occlusion, and long-distance recognition of small targets, combined with the strong generalization ability of models such as Swin Transformer and multi-scale YOLO, the system can stably output high-precision perception results and ensure the operational requirements of real-time inference for onboard devices [4]. Graphic recognition provides reliable visual basis for advanced assisted driving, automatic obstacle avoidance, lane keeping, automatic parking, and even full scene autonomous driving. It is an important technical foundation for ensuring driving safety, regulating road traffic order, and reducing the incidence of traffic accidents.

4.4. Industrial intelligent manufacturing and quality inspection field

The standardization and process oriented characteristics of industrial manufacturing scenarios provide favorable conditions for the deep implementation of graphic recognition technology. Traditional industrial quality inspection has long relied on manual visual inspection, which has drawbacks such as inconsistent testing standards, long-term work fatigue, high defect omission rates, and high labor costs. After introducing graphic recognition technology, automated appearance inspection can be carried out for industrial components, electronic products, metal sheets, plastic products, textile fabrics and other industrial products, quickly identifying minor defects such as surface scratches, dents, cracks, deformations, color differences, missing materials, and assembly misalignment. By relying on high-precision image segmentation and defect feature comparison algorithms, the equipment can stably adapt to complex lighting, high-speed assembly line movement, and large-scale repetitive detection working environment of factories, achieving 24-hour uninterrupted detection [9, 10]. Graphic recognition not only significantly improves industrial quality inspection efficiency and detection accuracy, unifies quality judgment standards, but also reduces labor costs for enterprises, and helps the manufacturing industry achieve automation upgrading, quality improvement, efficiency enhancement, and intelligent transformation.

4.5. In the field of modern agriculture

The rapid development of smart agriculture has promoted the widespread application of graphic recognition technology throughout the entire agricultural production process. In the field management process, graphic recognition algorithms can identify crop types, monitor vegetation coverage, analyze leaf status, and accurately determine crop growth cycles and conditions through aerial images and field footage. In terms of pest and disease control, the model can automatically identify common crop diseases and pest types based on leaf texture, color changes, and disease spot characteristics, achieve early warning, help farmers use pesticides accurately, and reduce resource waste and ecological pollution caused by blind pesticide spraying. At the same time, graphic recognition can also be applied to automatic sorting and grading of agricultural products, quickly

DOI: 10.54254/2755-2721/2026.34232

classifying and screening fruit size, color, and appearance defects, and improving the commercial quality of agricultural products. Combining mobile and edge lightweight models, graphic recognition technology is adapted to complex rural work environments, helping traditional agriculture transform into a modern agricultural model that is refined, digitized, and intelligent.

4.6. E-commerce retail and other livelihood fields

In daily life scenarios such as e-commerce retail, content review, and cultural entertainment, graphic recognition also has mature and widespread applications. E-commerce platforms rely on product image recognition, image search, and similar product matching technologies to optimize user shopping experience, achieve fast product retrieval and intelligent recommendation; In the warehousing and logistics scenario, graphic recognition can complete cargo sorting, package barcode recognition, material classification and inventory, and improve the level of warehousing automation management. At the same time, based on image content recognition and semantic understanding algorithms, network platforms can achieve intelligent review of images and short video content, automatically identify illegal images, bad images, and sensitive content, reduce manual review pressure, and standardize the network environment. In addition, graphic recognition derived applications such as image restoration, image quality enhancement, facial beautification, and style transfer have also deeply integrated into people's daily lives, demonstrating rich scene extensibility and strong practical value for civilian use.

5. Current challenges and future prospects of graphic recognition

Despite the breakthrough in graphic recognition technology, there are still many challenges. In terms of data bottlenecks, accuracy drops sharply in small sample scenarios, annotation costs are high, data in special fields is scarce, and uneven data distribution can easily lead to algorithm bias. In terms of algorithm shortcomings, the model has poor adversarial robustness, insufficient interpretability, and weak generalization ability in complex extreme scenarios. In terms of implementation challenges, the demand for computing power in visual large-scale models is high, and it is difficult to deploy at the edge. The fragmentation of industry scenarios leads to high customization costs, and data privacy and ethical issues urgently need to be regulated. Based on long-term development, graphic recognition technology will continue to optimize and upgrade around existing bottlenecks. In the future, technologies such as small sample learning and self supervised learning will continue to improve, which can effectively reduce the dependence of models on massive annotated data and alleviate the problem of scarce data in special industries. The research on interpretability algorithms and adversarial robustness optimization continues to advance, gradually breaking the black box limitations of deep learning and improving the stability of models in complex environments and high-risk scenarios. At the same time, the optimization of lightweight models and the continuous maturity of end-to-end cloud collaborative deployment solutions can effectively reduce the deployment threshold and computational cost of visual large-scale models [13]. With the deep integration of multimodal fusion, the Internet of Things, and industry digitization, graphic recognition will upgrade from single perception recognition to intelligent analysis and autonomous decision-making, providing more stable, efficient, and secure technical support for the intelligent transformation of various industries [5, 6].

DOI: 10.54254/2755-2721/2026.34232

6. Conclusion

This article systematically reviews the overall development status and application value of graphic recognition under the background of artificial intelligence, and comprehensively discusses the four major development stages, core task system, and mainstream algorithm architecture of graphic recognition. The article sequentially analyzes the technical principles and iterative advantages of traditional machine learning, convolutional neural networks, visual transformers, and multimodal large models , categorizes and introduces the characteristics, advantages, and applicable scenarios of various algorithms, and combines multiple fields such as security, medical care, autonomous driving, industrial quality inspection, and modern agriculture to expound the actual implementation mode and application effectiveness of graphic recognition.

Research has shown that the overall trend of graphic recognition technology is evolving from manual features to automatic features, from single fields to general scenarios, and from single vision to multimodal fusion. CNN is still the mainstream solution for industrial implementation, with a good balance between detection accuracy and inference speed; Transformers and visual models have further broken through the upper limit of accuracy and have significant advantages in complex scenes and cross domain tasks; The lightweight algorithm effectively ensures the deployment requirements of edge devices and achieves a reasonable balance between accuracy and computing power.

At present, there are still problems with data shortage, weak model robustness, insufficient interpretability, and high deployment costs in graphic recognition. In the future, it is necessary to combine small sample learning, model lightweighting, and multimodal fusion technology to continuously balance the contradiction between accuracy and computing power, and promote the intelligent application of graphic recognition technology in various industries that is safer, more efficient, and scalable.

References

[1] Krizhevsky, A., Sutskever, I., & Hinton, G. E. (2012). ImageNet classification with deep convolutional neural networks. Advances in Neural Information Processing Systems, 25, 1097–1105.
[2] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. 2016 IEEE Conference on Computer Vision and Pattern Recognition, 770–778.
[3] Dosovitskiy, A., Beyer, L., Kolesnikov, A., et al. (2021). An image is worth 16×16 words: Transformers for image recognition at scale. International Conference on Learning Representations.
[4] Liu, Z., Lin, Y., Cao, Y., et al. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. 2021 IEEE/CVF International Conference on Computer Vision, 9992–10002.
[5] Radford, A., Kim, J. W., Hallacy, R., et al. (2021). Learning transferable visual models from natural language supervision. International Conference on Machine Learning, 8748–8763.
[6] Kirillov, A., Mintun, E., Ravi, N., et al. (2023). Segment anything. 2023 IEEE/CVF International Conference on Computer Vision, 4015–4026.
[7] Wu, X., Sahoo, D., & Hoi, S. C. H. (2020). Recent advances in deep learning for object detection. Neurocomputing, 396, 309–327.
[8] Redmon, J., Divvala, S., Girshick, R., & Farhadi, A. (2016). You only look once: Unified, real-time object detection. 2016 IEEE Conference on Computer Vision and Pattern Recognition, 779–788.
[9] Li, X., Ding, H., Yuan, H., et al. (2024). Transformer-based visual segmentation: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence, 46(12), 10138-10158.
[10] Zhang, L., Wang, Y., & Liu, J. (2022). High-resolution Swin Transformer for automatic medical image segmentation. Sensors, 23(7), 3420.
[11] Carion, N., Massa, F., Synnaeve, G., et al. (2021). End-to-end object detection with transformers. European Conference on Computer Vision, 213–229.

DOI: 10.54254/2755-2721/2026.34232

[12] Cheng, B., Misra, I., Schwing, A. G., et al. (2022). Masked-attention mask transformer for universal image segmentation. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 1290–1299.
[13] Xin, Z., Chen, S., Wu, T., et al. (2024). Few-shot object detection: research advances and challenges. Information Fusion, 107, 102307.