A Review of Super-Resolution Reconstruction Technology Based on Convolutional Neural Networks and Generative Adversarial Networks

Yao Chen
Huaiyin Institute of Technology, Data Science and Big Data Technology, Huai'an, Jiangsu, China

Abstract: As a key technology in the field of computer vision, image super-resolution reconstruction aims to recover high-fidelity details from low-resolution images and has been widely applied in various fields in recent years. This paper mainly introduces two mainstream deep learning models for super-resolution reconstruction: Convolutional Neural Networks (CNN) and Generative Adversarial Networks (GAN), and systematically compares their technical characteristics and applicable scenarios. Relying on local feature extraction and parameter sharing mechanisms, CNN models excel in objective indicators such as Peak Signal-to-Noise Ratio (PSNR), with high computational efficiency and stable training. Represented by models like SRCNN and EDSR, they are suitable for scenarios with strict accuracy requirements, such as medical imaging and remote sensing monitoring. Through the adversarial training between generators and discriminators, GAN models introduce perceptual loss to improve visual realism. Models such as SRGAN and ESRGAN can generate rich texture details, but they have issues such as reliance on high-quality data for training and susceptibility to artifact generation. They are more applicable to fields that prioritize subjective experience, such as film and television restoration and game image quality enhancement. This paper further analyzes the differences between the two types of models in terms of computational cost and generalization ability, clarifies the basis for model selection by combining typical application cases, and finally looks forward to the development directions such as lightweight fusion architectures, providing a reference for the practical implementation of the technology.

Keywords: Super-Resolution Reconstruction; Deep Learning; Convolutional Neural Network (CNN); Generative Adversarial Network (GAN); Computer Vision

1. Introduction

1.1 Evolution and Core Value of Super-Resolution Reconstruction Technology

Image Super-Resolution (SR) is a core technology in the field of computer vision that addresses the "lack of details in low-resolution images". Its essence is to infer the lost pixel information in low-resolution images through algorithms and solve the "underdetermined inverse problem" [1]. The technological development has gone through three generations: The traditional interpolation stage relies on methods such as bicubic interpolation and nearest-neighbor interpolation to achieve reconstruction through simple mapping between pixels, but the generated images are prone to blurriness and blocking effects; The sparse representation stage is represented by SR methods based on dictionary learning, which extract texture features through sparse coefficient matching, but have poor adaptability to complex scenes; Since 2014, the deep learning stage has become the mainstream, with data-driven as the core. Convolutional Neural Networks (CNN) and Generative Adversarial Networks (GAN) have achieved key breakthroughs in "objective accuracy" and "subjective visual effect" respectively[2].

Currently, super-resolution technology has become a rigid demand for high-quality imaging in multiple fields: In the field of medical imaging, sub-micron resolution improvement can assist in the early diagnosis of liver tumors and retinal diseases; In the remote sensing field, SR technology reduces the reliance on satellite hardware, enabling mid-range satellites to achieve the ground object recognition accuracy of high-end equipment; In video conferencing scenarios, SR combined with compression

Copyright @ STEMM Institute Press

algorithms can realize 1080p image quality transmission under a 1Mbps bandwidth; In the film and television entertainment field, 4K/8K old film restoration and real-time super-resolution in games have become key technical supports for industrial upgrading.

1.2 Research Objectives and Key Evaluation Systems

This paper focuses on two mainstream models, CNN and GAN , and conducts a comparative analysis centering on three key objective indicators: robustness, reconstruction quality, and computational efficiency. It clarifies the model selection criteria by combining four typical application scenarios: remote sensing, medical imaging, video conferencing, and film and television entertainment. Among these indicators, robustness specifically refers to the model's adaptability in complex environments, covering three dimensions: noise robustness, blur robustness, and data distribution robustness. Noise robustness measures the model's ability to suppress common noises under test conditions of Gaussian noise (intensities of 15dB and 20dB) and salt-and-pepper noise (densities of 5% and 10%). Blur robustness evaluates the model's capability to restore blurred images using test scenarios including Gaussian blur kernels of sizes 3×3 and 5×5, as well as motion blur with displacements of 3 pixels and 5 pixels. Data distribution robustness verifies the model's scenario generalization ability through tests on cross-dataset and cross-modal data. Reconstruction quality encompasses both objective and subjective indicators: objective indicators include Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM), while the subjective indicator adopts Mean Opinion Score (MOS). Computational efficiency covers three dimensions: number of parameters, single-frame inference time, and training computing power requirements, which directly determine the feasibility of the model's engineering implementation[3].

2. Core Architectures and Technical Principles of CNN and GAN

2.1 Architectural Design and Technical Advantages of Convolutional Neural Networks (CNN)

CNN realizes super-resolution through an end-to-end process of "local feature extraction - feature fusion - upsampling". Its core advantages come from the parameter sharing and local receptive field mechanisms, which can not only efficiently capture the spatial correlation of images but also reduce model complexity and improve the efficiency of feature extraction [4].

Typical CNN architectures are represented by SRCNN, EDSR, and RCAN, and their evolution shows a dual trend of "enhanced feature extraction capability" and "optimized training stability". Proposed in 2014, SRCNN was the first CNN model used for super-resolution, divided into three stages: "feature extraction - nonlinear mapping - reconstruction". It extracts local features through 5×5 and 3×3 convolution kernels, but has a large number of parameters (about 600,000), and the upsampling stage relies on bicubic interpolation for assistance. Released in 2017, EDSR removes the Batch Normalization layer in traditional CNN to avoid feature distortion, introduces residual connections to solve the gradient vanishing problem, and supports 2×/3×/4× multi-scale super-resolution. Its PSNR reaches 34.68dB on the DIV2K dataset, which is 2.3dB higher than that of SRCNN. Proposed in 2018, RCAN adds a "Residual Channel Attention Block (RCAB)", which strengthens key features such as edges and textures through channel weight allocation. The number of parameters increases to 15 million, but the reconstruction accuracy is further improved, with the 4× super-resolution SSIM reaching 0.921, making it a core choice for high-precision super-resolution scenarios [5].

The loss functions of CNN mostly adopt Mean Squared Error (MSE) or L1 loss, with the core goal of optimizing pixel-level errors to improve PSNR and SSIM. Such loss functions can make the model converge quickly and ensure pixel-level accuracy, but they also easily cause the "over-smoothing" problem in reconstructed images, resulting in the loss of high-frequency texture details and thus affecting the subjective visual effect.

2.2 Architectural Design and Technical Advantages of Generative Adversarial Networks (GAN)

GAN realizes super-resolution through the zero-sum game mechanism of "Generator (G) - Discriminator (D)". Its core innovation is the introduction of perceptual loss and adversarial loss, which breaks through the limitation of traditional pixel-level optimization, prioritizes

 Copyright @ STEMM Institute Press

the improvement of the subjective visual realism of images, and is more in line with human visual perception preferences[6].

Typical GAN architectures are represented by SRGAN and ESRGAN, with the design focus on "strengthening texture generation capability" and "improving adversarial training stability". Proposed in 2016, SRGAN's generator adopts an architecture of "8 residual blocks + sub-pixel convolution", and the discriminator distinguishes between real and fake images through "convolution + LeakyReLU + fully connected layers". It was the first to introduce perceptual loss extracted by the VGG network, with the 4× super-resolution MOS score reaching 4.42, which is 18% higher than that of SRCNN. The optimized ESRGAN in 2018 improves the generator into a "Residual-in-Residual Dense Block (RRDB)" to enhance the efficiency of feature reuse, adopts a multi-scale discrimination strategy for the discriminator, and removes the batch normalization layer to reduce feature distortion. The richness of texture details of 4× super-resolution is 25% higher than that of SRGAN, but it has the problem of being prone to artifact generation and needs to be applied cautiously in accuracy-sensitive scenarios.

The loss function of GAN is a combination of adversarial loss and perceptual loss: Adversarial loss forces the generator to generate high-resolution images that are difficult for the discriminator to distinguish through the game between the generator and the discriminator; Perceptual loss optimizes the naturalness of textures by simulating the process of the human visual system perceiving image features. This combined loss can significantly improve the subjective visual effect, but it also leads to the pixel-level error of the model being higher than that of CNN, and the performance of objective indicators is relatively weak [7].

3. Comparative Analysis of Objective Indicators between CNN and GAN

3.1 Robustness Comparison: Noise, Blurriness, and Data Adaptability

Robustness directly determines the usability of the model in complex real-world scenarios. The two types of models show significant differences in three aspects: noise suppression, blur adaptation, and data distribution adaptation (Table 1). A mixed dataset of Set14 and DIV2K is used for the test to ensure the generality of the results[8].

In terms of noise robustness, CNN shows obvious advantages. The design of residual connections and dense connections enables it to effectively suppress noise interference. For example, when the Gaussian noise intensity is 20dB, the PSNR of EDSR only decreases by 1.2dB, while that of SRGAN decreases by 2.8dB; RCAN focuses on structural features through channel attention, and when the salt-and-pepper noise density reaches 10%, the SSIM can still be maintained at 0.89, while that of ESRGAN is only 0.82. In contrast, the adversarial training mechanism of GAN tends to amplify the impact of noise. The generator may misjudge noise as effective features in pursuit of texture details, leading to "granular artifacts" in the reconstructed images. For example, when the Gaussian noise intensity is 15dB, the artifact proportion of ESRGAN reaches 12%, while that of EDSR is only 3%.

In terms of blur robustness, CNN has stronger adaptability to uniform blur. The lightweight CNN model FSRCNN achieves a 4× super-resolution PSNR of 28.7dB under the action of a 5×5 Gaussian blur kernel, which is 3.5dB higher than that of SRGAN; GAN has weak processing ability for non-uniform blur. For example, in the motion blur scenario (displacement of 5 pixels), the MOS score of SRGAN drops from 4.42 to 3.15, while that of EDSR only drops to 3.98, and the attenuation range of the subjective visual effect is significantly smaller than that of GAN.

In terms of data distribution robustness, GAN has better cross-domain adaptation ability. By learning the laws of data distribution through adversarial training, GAN can better adapt to data from different sources. For example, on remote sensing images from different satellites (Gaofen-2, Landsat-8), the fluctuation of the land cover classification accuracy of ESRGAN is only 5%, while that of EDSR reaches 12%; On infrared MRI images, a type of non-traditional visible light data, the PSNR of EDSR is 4.2dB lower than that of visible light MRI, while that of ESRGAN is only 2.1dB lower, showing stronger scene generalization ability.[9]

3.2 Reconstruction Quality Comparison: Objective Accuracy and Subjective Vision

Reconstruction quality is the core evaluation criterion for super-resolution technology. CNN and GAN form a complementary pattern of "objective accuracy - subjective vision" (Table 2). The test scenario is 4× super-resolution, and the DIV2K validation set is used for the test to ensure the authority of the data [10].

In terms of objective indicators (PSNR/SSIM), CNN has an absolute advantage. The optimization of pixel-level loss functions enables CNN to minimize reconstruction errors. For example, the PSNR and SSIM of EDSR reach 34.68dB and 0.921, which are 5.2dB and 0.08 higher than those of SRGAN respectively; RCAN further improves the accuracy through channel attention, with the PSNR exceeding 35dB, making it one of the models with the best objective indicators and playing an irreplaceable role in scenarios with strict accuracy requirements such as medical imaging and remote sensing. In contrast, GAN sacrifices part of the objective accuracy to improve the subjective visual effect. For example, the PSNR of SRGAN is only 29.48dB, but through texture enhancement technology, in the super-resolution of natural images, it is subjectively closer to real images, and the restoration degree of details such as hair and leaves is significantly higher than that of CNN.

In terms of subjective indicators (MOS), GAN shows an absolute advantage. The scores from 100 subjects (on a 1-5 scale) show that the score of ESRGAN in the super-resolution of natural images is 4.58, which is 14% higher than that of EDSR (4.02); In the film and television image restoration scenario, the "texture naturalness" score of ESRGAN reaches 4.7, while that of EDSR is only 3.8, which is more in line with human perception needs for image details. The disadvantage of CNN lies in the "plastic look" caused by over-smoothing. The generated images have overly regular edges and lack the randomness of real textures. For example, in medical imaging, the boundary of liver tumors reconstructed by EDSR shows "angular" features, which are easily misjudged as calcifications by doctors during interpretation, while the boundary reconstructed by ESRGAN is more natural. Although there are a small number of artifacts, the interference to diagnosis is relatively small.

Table 1. Experimental Result Comparison of Robustness between CNN and GAN

Robustness	Test Condition	CNN(EDSR)	GAN(ESRGAN)	Advantageous Model
Noise Robustness	Gaussian Noise 20dB	PSNR 30.2dB	PSNR 27.4dB	CNN
	Salt-and-Pepper Noise Density 10%	SSIM 0.89	SSIM 0.82	CNN
Blur Robustness	5×5 Gaussian Blur	PSNR 29.1dB	PSNR 25.6dB	CNN
	Motion Blur (Displacement 5 Pixels)	MOS 3.98	MOS 3.15	CNN
Data Distribution Robustness	Cross-Satellite Remote Sensing Data	Accuracy Fluctuation 12%	Accuracy Fluctuation 5%	GAN
	Infrared MRI Images	PSNR 26.8dB	PSNR 28.9dB	GAN

Table 2. Comparison of Reconstruction Quality Indicators between CNN and GAN (4× Super-Resolution, Test Set: DIV2K)

Model	PSNR (dB)	SSIM	MOS Score (Natural Images)	MOS Score (Medical Images)	Core Advantage
EDSR (CNN)	34.68	0.921	4.02	4.35	High Objective Accuracy
RCAN (CNN)	35.12	0.928	4.15	4.48	Good Adaptability to Medical Imaging
SRGAN (GAN)	29.48	0.842	4.42	3.85	Excellent Visual Effect on Natural Images
ESRGAN (GAN)	28.95	0.836	4.58	3.92	Rich Texture Details

3.3 Computational Efficiency Comparison: Parameter Quantity, Speed, and Cost

Computational efficiency determines the engineering implementation capability of the model and directly affects the adaptability to real-time scenarios and edge devices. CNN is significantly superior to GAN in real-time performance and cost control (Table 3). NVIDIA Tesla T4 GPU is used as the test hardware to ensure the consistency of the computing power environment [11].

In terms of parameter quantity and inference speed, CNN has obvious lightweight advantages. The number of parameters of the lightweight model FSRCNN is only 1.5M, while that of ESRGAN reaches 40M; In the inference of a single 256×256 image, the time taken by FSRCNN is only 2.3ms, while ESRGAN takes 18.7ms; Even the high-precision CNN model RCAN (with 15M parameters) has an inference time of only 8.5ms, which can support real-time

 Copyright @ STEMM Institute Press

1080p@60fps super-resolution and meet the real-time requirements of scenarios such as video conferencing and mobile image enhancement. The disadvantage of GAN lies in the slow speed caused by the large number of parameters. For example, the parameter proportion of the RRDB module of ESRGAN reaches 70%, and the inference time on the CPU (Intel i7-12700K) is 120ms per frame, which cannot meet the fluency requirements of real-time scenarios (usually requiring more than 30fps) [12].

In terms of training cost, CNN also has a significant efficiency advantage. The training cycle of EDSR on a single V100 GPU is only 24 hours, consuming about 50 TFLOPs of computing power; However, the training stability of GAN is poor, and multi-card synchronous training is required to avoid mode collapse. For example, ESRGAN requires 8 V100 GPUs for synchronous training, with a cycle of 72 hours and a computing power consumption of about 600 TFLOPs, which is 12 times that of EDSR. In addition, GAN has higher requirements for the quality of training data and needs high-quality paired datasets. The data annotation cost is 3 times higher than that of CNN, which limits its application in data-scarce fields.

Table 3. Comparison of Computational Efficiency between CNN and GAN (Hardware: NVIDIA Tesla T4 GPU)

Model	Parameter Quantity (M)	Single-Frame Inference Time (256×256)	Training Cycle (Single V100)	Computing Power Consumption (TFLOPs)	Adaptable Scenarios
FSRCNN (CNN)	1.5	2.3 ms	12 Hours	20	Real-Time Video Conferencing
EDSR (CNN)	15	8.5 ms	24 Hours	50	Offline Processing of Medical Imaging
SRGAN (GAN)	16	15.2 ms	48 Hours	300	Offline Film and Television Restoration
ESRGAN (GAN)	40	18.7 ms	72 Hours	600	Game Image Quality Enhancement

4. Model Selection and Practical Cases in Typical Application Scenarios

4.1 Remote Sensing Field: Accuracy First, CNN as the Dominant Choice

Super-resolution of remote sensing images needs to take both "land cover classification accuracy" and "robustness in complex environments" into account. Due to its high objective accuracy and strong noise suppression ability, CNN has become the mainstream choice in this field, while GAN shows supplementary potential in specific scenarios [13].

The application practice of CNN focuses on two directions: land cover classification and disaster monitoring. The original resolution of Gaofen-2 satellite images is 1m. After 2× super-resolution by EDSR, the classification accuracy of farmland and construction areas increases from 82% to 91%. During the 2023 Henan flood disaster, the remote sensing images reconstructed by EDSR helped the rescue team accurately locate the flooded areas, with the positioning error controlled within 50 meters, providing key data support for emergency rescue and disaster relief; Aiming at the characteristics of night low-light remote sensing images, RCAN strengthens the light source features through channel attention, increasing the road recognition accuracy from 65% to 88%, which is better than 75% of ESRGAN, and shows more stable performance in night disaster monitoring.

The application of GAN in the remote sensing field has obvious limitations, mainly reflected in the interference of artifacts on interpretation. In the super-resolution of remote sensing images, the proportion of "false road edges" generated by ESRGAN reaches 8%, which may lead to misjudgment of the disaster area, with an error range of 15%, and the risk is high in accuracy-sensitive scenarios such as disaster statistics; However, GAN also has potential in enhancing complex terrain textures. For example, in mountain remote sensing images, ESRGAN can enhance the ridge line texture to assist in landslide risk assessment. Its MOS score is 0.6 higher than that of EDSR, which is subjectively more convenient for manual interpretation of terrain features. In the future, its practicality can be improved through artifact suppression technology.

4.2 Medical Imaging Field: Accurate

Diagnosis, CNN as the First Choice

The core requirement of medical imaging super-resolution is "no artifacts" and "high accuracy" to avoid artifacts misleading diagnosis. The pixel-level accuracy and training stability of CNN are more in line with clinical needs, while the application of GAN is limited due to data dependence and artifact problems.

The core applications of CNN include CT/MRI lesion detection and microscopic image enhancement. The "super-resolution auxiliary diagnosis system" developed by GE Healthcare based on EDSR increases the original resolution of lung CT images from 0.6mm to 0.3mm, increasing the lung nodule detection rate from 78% to 92%, and reducing the false positive rate by 30% compared with traditional methods, significantly reducing the risk of missed diagnosis and misdiagnosis by doctors; In MRI liver tumor diagnosis, the images reconstructed by RCAN can clearly show the tumor boundaries, reducing the doctor's interpretation time from 30 minutes to 15 minutes and improving the diagnosis efficiency. In the super-resolution of cell microscopic images (sub-micron level), the real-time advantage of FSRCNN is significant, which can dynamically capture the cell division process with a frame rate of 30fps, while ESRGAN only has 10fps, which cannot meet the real-time observation needs [14].

The application risks of GAN in the medical imaging field are mainly reflected in two aspects: artifacts and data privacy. In the super-resolution of breast molybdenum target images, the proportion of "false calcification points" generated by ESRGAN reaches 5%, which may lead to the misjudgment of benign lesions as malignant, increasing the misdiagnosis rate by 8% and causing serious interference to clinical diagnosis; At the same time, GAN requires a large number of high-quality annotated data for training, but medical data is strictly restricted by privacy protection, making it difficult to build large-scale training sets, which further restricts its application scope [15].

Table 4. Schematic Diagram of Super-Resolution Effect Comparison of Medical Imaging

Image Type	Low-Resolution Image	CNN(EDSR) Reconstructed Image	GAN(ESRGAN) Reconstructed Image
Lung CT (Lung Nodule)	The boundary of the lung nodule is blurry, and the diameter misjudgment deviation reaches 2mm	The nodule boundary is clear, and the diameter error is controlled within 0.5mm	There are artifacts at the nodule boundary, and the diameter error deviation reaches 1mm
Liver Tumor MRI	The boundary between the tumor and normal tissue is unclear	The boundary between the tumor and normal tissue is clear, with no artifacts	The boundary between the tumor and normal tissue is blurry, with granular artifacts

4.3 Video Conferencing Field: Real-Time Transmission, Lightweight CNN Models Take the Lead

The core requirement of video conferencing is to balance "low latency", "low bandwidth" and "image quality". The real-time performance and low-bandwidth adaptability of lightweight CNN models are irreplaceable, while GAN is difficult to be implemented due to insufficient real-time performance and high bandwidth consumption.

The technical implementation of CNN focuses on two directions: real-time super-resolution, bandwidth optimization, and mobile terminal adaptation. Zoom video conferencing adopts FSRCNN 2× super-resolution technology to realize 1080p image quality transmission under a 1Mbps bandwidth, with a latency of only 45ms, while ESRGAN requires 120ms. At the same time, the bandwidth consumption is reduced by 40% compared with the non-super-resolution solution, meeting the fluency requirements of scenarios such as remote office and online education; The number of parameters of the lightweight model MobileSR is only 0.8M, which supports real-time 720p@30fps super-resolution on mobile phones with a power consumption of only 50mW, while the power consumption of ESRGAN reaches 200mW, making it more suitable for the low-power requirements of mobile video conferencing [16].

The application of GAN in video conferencing scenarios has two major bottlenecks. One is insufficient real-time performance. The frame rate of ESRGAN in video conferencing scenarios is only 15fps, which is lower than the fluency threshold of 30fps, and the latency reaches 120ms, exceeding the interaction threshold of 100ms, resulting in dialogue stuttering and poor synchronization. The other is high bandwidth consumption. The images generated by GAN have complex textures and

 Copyright @ STEMM Institute Press

low compression rates. Under the same image quality, the bandwidth consumption is 50% higher than that of CNN, which is prone to stuttering and image quality fluctuations in low-bandwidth networks, failing to meet the requirements of stable transmission.

4.4 Film and Television Entertainment Field: Visual Experience, GAN as the Core

The core requirement of film and television entertainment is to prioritize "subjective visual effects". The texture generation ability of GAN makes it the first choice for old film restoration and game image quality enhancement, while CNN is difficult to meet the needs due to insufficient textures and weak artistic processing.

The benchmark applications of GAN include 4K/8K restoration of old films and game image quality enhancement. Warner Bros. used ESRGAN to restore "Casablanca" (a 1942 black-and-white film), super-resolving 720p images to 4K. The MOS score of skin textures and clothing wrinkles reached 4.7, and the box office of the restored film increased by 15% compared with the original version, realizing the commercial value reconstruction of classic IP; Netflix used SRGAN to solve the "image blurriness" problem in the remastered version of "Friends", with a user satisfaction rate of 92%, becoming a benchmark case for old film restoration on streaming platforms. In the game field, NVIDIA DLSS technology integrates ESRGAN. In the game "Cyberpunk 2077", the frame rate is increased from 30fps to 60fps under 4K resolution, while maintaining the visual realism of "neon textures" and "architectural details", with a MOS score of 4.6, compared with only 3.9 of EDSR, significantly improving the game immersion.

The application limitations of CNN in the film and television entertainment field are mainly reflected in two aspects: textures and artistic processing. In game super-resolution, the "metal texture" and "fabric texture" generated by EDSR are too smooth, and the MOS score is 0.7 lower than that of ESRGAN, lacking the detail levels of real materials; In the production of film and television special effects, CNN is difficult to generate artistic super-resolution effects such as "oil painting style" and "watercolor style", while GAN can realize this requirement through style transfer technology, providing more creative space for film and television creation [17].

Table 5. Model Selection and Effect Comparison in Four Application Scenarios

Application Scenario	Core Requirement	Preferred Model	Key Indicator Performance	Typical Case
Remote Sensing	Land Cover Classification Accuracy, Noise Robustness	CNN(EDSR/RCAN)	Classification Accuracy 91%, PSNR 30.2 dB	Gaofen-2 Disaster Monitoring
Medical Imaging	Lesion Accuracy, No Artifacts	CNN(EDSR/FSRCNN)	Lung Nodule Detection Rate 92%, SSIM 0.92	GE Healthcare CT Super-Resolution System
Video Conferencing	Low Latency, Low Bandwidth	CNN(FSRCNN/MobileSR)	Latency 45ms, Bandwidth 1Mbps	Zoom 1080p Real-Time Super-Resolution
Film and Television Entertainment	Visual Realism, Rich Textures	GAN(SRGAN/ESRGAN)	MOS 4.7, Frame Rate 60fps	"Casablanca" 4K Restoration

5. Conclusion

This paper systematically compares the technical characteristics of CNN and GAN in super-resolution reconstruction. The results show that the two types of models have their own advantages and complementary applicable scenarios. Relying on the parameter sharing mechanism and pixel-level optimization strategy, CNN performs prominently in terms of robustness, computational efficiency, and objective accuracy. Its robustness is reflected in the strong suppression ability against noise and blurriness; in terms of computational efficiency, lightweight models can realize real-time inference; in terms of objective accuracy, it leads in PSNR and SSIM indicators. Therefore, it is more suitable for scenarios with high requirements for accuracy or real-time performance, such as remote sensing land cover classification, medical imaging diagnosis, and real-time transmission in video conferencing. Relying on the adversarial training mechanism of generators and discriminators, GAN introduces perceptual loss to optimize the subjective visual effect, and has obvious advantages in texture realism and cross-domain data adaptability. It is more suitable for scenarios that focus on visual experience, such as the restoration of old films and television works and game image quality enhancement [18].

The core development direction of super-resolution technology in the future is the "in-depth integration of accuracy and vision". Through four technical paths, namely lightweight fusion architecture, self-supervised learning, cross-modal feature fusion, and hardware-algorithm collaborative optimization, the existing bottlenecks will be broken, and the practical application of super-resolution technology in more fields such as autonomous driving, AR/VR, and deep space exploration will be promoted. Finally, the technical goal of "high fidelity, low latency, and wide adaptability" will be achieved, providing core support for the high-quality imaging needs of various industries[19] .

References

[1] Ponney M V ,Velammal L B ,Kulothungan K . DB-DSCN: Image Tampering Detection Using Dual Branch Deep Stacked Convolutional Network[J].International Journal of Software Engineering and Knowledge Engineering,2025,35(09).

[2] Burçak C K . An efficient deep convolutional network model using mask images for multiclass classification of breast cancer ultrasound images[J].Neural Computing and Applications,2025,(prepublish):1-20.

[3] Jaincy M E D ,Pattabiraman V . BCDCNN: breast cancer deep convolutional neural network for breast cancer detection using MRI images[J].Scientific Reports,2025,15(1):29014-29014.

[4] Singh K ,Sudershan A ,Kumar S , et al. BC-MBINet: A Novel Architecture for Accurate Classification of Breast Cancer with Microscopic Biopsy Images Using Deep Convolutional Neural Networks[J].International Journal of Pattern Recognition and Artificial Intelligence,2025,39(13).

[5] Tolgyesi A ,Whyne C ,Hardisty M . Predicting rat lumbar vertebral failure patterns as synthetic μCT images using a deep convolutional generative adversarial network[J].Journal of the Mechanical Behavior of Biomedical Materials,2025,170107116-107116.

[6] Zhang J D. Research on Super-Resolution Reconstruction of Single 3D-MRI Image Based on Deep Convolutional Neural Network[D]. Chongqing Medical University, 2024.

[7] Zhao Z W. Image Super-Resolution Reconstruction Based on Deep Convolutional Neural Network[J]. Science and Technology Innovation, 2024, (07): 92-95.

[8] Yuan X L, Zhang B H, Zhang Q, et al. Infrared Image Super-Resolution Reconstruction Technology Based on Deep Convolutional Neural Network[J]. Infrared Technology, 2023, 45(05): 498-505.

[9] Huang Z Y. Research on Single Image Super-Resolution Based on Very Deep Super-Resolution Convolutional Neural Network (in English)[J]. Journal of Measurement Science and Instrumentation, 2022, 13(03): 276-283.

[10] Wang X H, Zhao X Y, Wang X Y, et al. Single Image Super-Resolution Reconstruction Using Deep Residual Network with Undecimated Wavelet Edge Learning[J]. Acta Electronica Sinica, 2022, 50(07): 1753-1765.

[11] Xu W. Research on Hyperspectral Image Super-Resolution Method Based on Deep Convolutional Neural Network[D]. Tianjin University, 2022.

[12] Bai Z Y, Tao J Y. Super-Resolution Reconstruction of Brain MRI Images Using Pseudo-3D Convolutional Network[J]. Journal of Computer-Aided Design & Computer Graphics, 2022, 34(02): 208-216.

[13] Yang Y, Xie X, He L, et al. Super-Resolution Reconstruction Combining Continued Fraction Interpolation and Convolutional Neural Network[J]. Journal of Hefei University of Technology (Natural Science Edition), 2021, 44(08): 1146-1152.

[14] Cha T B. Research on Super-Resolution Method for Train Detection Images Based on Deep Convolutional Neural Network[D]. Southwest Jiaotong University, 2021.

[15] Chen X, Luo B. Multi-Angle Remote Sensing Image Super-Resolution Reconstruction Using Deep Network with Dynamic Upsampling Filter[J]. Geomatics and Information Science of Wuhan University, 2021, 46(11): 1716-1726.

[16] Shi J P. Research on Image

 Copyright @ STEMM Institute Press

Super-Resolution Reconstruction Technology Based on Deep Convolutional Network[D]. Nanjing University of Aeronautics and Astronautics, 2021.

[17] Zhang J T. Research on Image Super-Resolution Based on Attention Convolutional Neural Network[D]. Harbin Engineering University, 2021.

[18] Zou C Z, Huang X S. Hyperspectral Image Super-Resolution Method Based on Deep Residual Convolutional Neural Network[J]. Journal of Fuzhou University (Natural Science Edition), 2020, 48(05): 545-550.

[19] Xie C, Zhu H Y. Image Super-Resolution Reconstruction Method Based on Deep Convolutional Neural Network[J]. Transducer and Microsystem Technologies, 2020, 39(09): 142-145.