Proceedings of the 5th International Conference on Modern Medicine and Global Health
DOI: 10.54254/2753-8818/2027.39489

From Traditional Transformation to Deep Learning: The Technical Evolution of Medical Image Compression and the Challenge of Diagnostic Fidelity

Zhuo Chen

Maynooth International Engineering College, Fuzhou University, Fuzhou, China
2791599281@qq.com

Abstract. With the development of Computed Tomography (CT), Magnetic Resonance Imaging (MRI) and other medical imaging technologies, the image data is becoming more and more high-definition, three-dimensional, and the amount of data is increasing, which has brought pressure to hospital storage, transmission and remote diagnosis and treatment. Unlike ordinary image compression, medical image compression should not only pursue smaller files, but also ensure that the details related to diagnosis are not lost, that is, to ensure "diagnostic fidelity". This paper reviews the development process of medical image compression from traditional coding, region of interest compression, three-dimensional compression to deep learning compression, analyzes the characteristics and applicable scenarios of various methods, summarizes the cutting-edge directions of high-level deep data processing and cross scene adaptation combined with recent research, and finally points out the problems of current evaluation standards are not unified, the model is not universal enough, and gives the development ideas for clinical diagnosis.

Keywords: Medical Image Compression, 3D Image, Deep Learning, Lossless Compression, Diagnostic Fidelity

1. Introduction

Now the images of CT, MRI and other examinations are getting clearer and clearer. Hundreds of continuous three-dimensional slices can be generated in a single examination, and the amount of data is also increasing. The hospital needs to keep patient images for a long time for follow-up and consultation; Remote diagnosis and AI assisted diagnosis also need to quickly transmit and read images. Therefore, how to reduce the file size without losing useful information has become an important issue.

There are essential differences between medical image compression and ordinary image compression. Ordinary pictures sacrifice a small amount of details. As long as they look clear to the naked eye, they are generally acceptable; However, in medical images, small lesions, subtle density changes and tissue edge morphology may directly affect the judgment of doctors. The loss of information that is difficult to detect by the naked eye may not only cause misdiagnosis, but also reduce the accuracy of AI diagnostic model. Therefore, medical image compression should not only

© 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: 10.54254/2753-8818/2027.39489

look at the image quality score, but also consider whether the compressed image can continue to support reliable diagnosis.

The existing medical image compression technology has gradually developed from JPEG 2000 and other traditional coding methods [1] to 3D compression methods using the correlation between slices and within slices [2, 3]. In recent years, research has begun to focus on high-level deep image processing [4], cross data source adaptability [5] and three-dimensional data sampling strategy [6]. At the same time, researchers have gradually realized that compression will not only change the image quality, but also affect the analysis results of AI model [7]. Some studies have therefore included disease classification, lesion identification and expert evaluation in the compression effect assessment [8], and some have also studied improving the reconstruction quality of CT images through wavelet transform, attention mechanism and deep learning [9].

However, there are still some deficiencies in the existing research. Firstly, the data set, compression ratio and evaluation index used in different studies are not unified, and the experimental results are difficult to compare directly; Secondly, many methods still mainly rely on PSNR, SSIM and other image quality indicators, lacking the actual verification of AI tasks, doctor reading and long-term storage; In addition, the stability of the deep learning model under different hospitals, equipment and scanning parameters is still unclear, which needs further research. It can be seen that how to establish a unified evaluation method of diagnostic fidelity is the key problem to be solved for the clinical application of medical image compression.

Around the above background, this paper mainly studies three aspects: first, how the medical image compression technology is gradually developed, and which scenarios are suitable for each method; Second, what is "diagnostic fidelity" and how to evaluate it; Third, what core issues should this technology solve in the future.

This paper takes CT and MRI three-dimensional images as the main research object, and sorts out the representative research results in this field from 2000 to 2026. According to the order of technology development, this paper analyzes the existing methods into four categories: traditional coding, region of interest compression, three-dimensional compression and deep learning compression, and summarizes the frontier directions and existing challenges on this basis. The performance analysis in this paper is based on published or public research results. Independent model experiments are not carried out, and other imaging types such as ultrasound and pathological sections are not discussed in depth.

2. Technical evolution of medical image compression

2.1. Traditional transform coding: represented by JPEG 2000

Early medical image compression mainly relied on fixed mathematical transformation and coding technology to reduce data redundancy. The most commonly used standard was JPEG 2000. The standard uses discrete wavelet transform for multi-resolution decomposition, and then combines quantization and rate control to complete coding. It supports both lossy compression and reversible lossless compression [1].

JPEG 2000 has the advantages of mature technology and good compatibility. It can encode and decode quickly without special hardware. It also supports progressive transmission, that is, when viewing the film remotely, you can first transmit the low-definition version, and then gradually load the high-definition details. It is very suitable for long-term archiving, image exchange and remote browsing in hospitals [1]. However, it uses a fixed transformation mode and has limited adaptability to complex human structure and diverse lesions; At high compression ratio, it is easy to lose subtle

DOI: 10.54254/2753-8818/2027.39489

texture and edge information. Previous reviews have pointed out that JPEG and JPEG 2000 compression may also change the input distribution and task performance of the deep learning model, so it is impossible to directly judge whether the compressed image retains the diagnostic value only by the overall image quality indicators such as PSNR and SSIM [7].

2.2. ROI compression: key areas for protection and diagnosis

The importance of the whole image in medical images is different, so the idea of "region of interest (ROI) compression" is formed, that is, the focus, key organs, blood vessels and other diagnostic core areas are delimited, and the quality is guaranteed by low compression or even lossless compression; The background and non-critical areas are reduced with a higher compression ratio. JPEG 2000 itself provides ROI encoding mechanism, which can give priority to ensuring the reconstruction quality of the specified region in the same code stream [1].

The core of this method is to design the compression strategy around the diagnostic requirements, which can achieve a more flexible balance between file size and key information. However, its effect completely depends on the accuracy of region division. If the potential lesions are not divided into key regions, they may be over compressed and cause information loss. Moreover, the key areas of different diseases and different examinations are different. It is difficult to use them automatically on a large scale because the workload of manual annotation is too large. Although it is conducive to automation to let the model automatically identify the important areas in the image, it will transfer the error to the compression process. Therefore, when ROI method is applied in clinic, it is necessary to verify the two links of region recognition and compression reconstruction at the same time, and not only compare the local image quality.

2.3. 3D medical image compression: using the similarity between slices

CT and MRI are not independent pictures, but continuous three-dimensional data. The human structure of adjacent sections is very similar, and there are a lot of duplicate information. If only a single slice is compressed separately, this layer-to-layer correlation is wasted. Learning lossless compression in recent years therefore, the relationship between slices and within slices is also modeled to improve the coding efficiency of 3D medical images [2, 3].

These methods utilize 3D correlation through context modeling, residual coding or volume data sampling; In the lossless mode, the decompressed data can be consistent with the original data, which is suitable for clinical archiving and scientific research analysis with high traceability requirements [2, 3, 6]. At the same time, 3D data brings higher overhead of display memory, operation and delay, and the training effect will also be affected by blocking or sampling strategies [6]. In view of the differences in the distribution of high-level depth images and data from multiple hospitals and devices, recent studies have respectively proposed special high-level depth modeling methods and multi-source domain generalization methods [4, 5], indicating that bit depth adaptation and cross data source stability have become important issues in 3D compression.

2.4. Deep learning compression: automatic learning of image features

With the development of deep learning, compression methods based on neural network are becoming more and more popular. It uses an encoder to convert the original image into a more compact feature representation, then compresses and encodes it, and finally uses a decoder to restore the image. Different from traditional methods, it does not need to manually design fixed

DOI: 10.54254/2753-8818/2027.39489

transformation rules, and can automatically learn the characteristic laws more suitable for medical images through a large number of data training. Attention mechanism, Gaussian mixture probability model and wavelet transform are also used to improve the rate allocation and reconstruction effect [9, 10].

The advantage of deep learning compression is that it can optimize the rate distortion relationship for specific data distribution, and can incorporate task indicators such as classification, detection or segmentation into training and evaluation, so as to explore "diagnostic lossless" rather than just pixel level lossless [7, 8]. However, its effect highly depends on the consistency of training data and deployment data. Multi-source domain generalization research shows that the distribution changes across hospitals, devices or scanning protocols need to be specially handled [5]. In addition, some lossy compression models may also generate details that "look reasonable but actually do not exist", which is very dangerous in medical diagnosis. Therefore, in high-risk diagnosis scenarios, the safety boundary must be verified in combination with downstream AI tasks and clinical evaluation [7, 8].

2.5. Overall comparison of four methods

Because there is no unified standard test set in the field of medical imaging at present, and the data sets, equipment and parameters used in different studies are different, it is difficult to make accurate numerical comparison. Here is only a qualitative overall summary.

The differences between the four methods can be summarized in Table 1. On the whole, they are not a simple replacement relationship, but correspond to different requirements such as standardized exchange, key area protection, 3D volume data archiving and task-oriented compression. In the subsequent application, the appropriate scheme should be selected according to the image type, compression purpose and diagnostic risk.

Table 1. Overall comparison of four types of medical image compression methods

Method Category	Representative Method	Applicable Scenarios	Advantage	Limit
Traditional transform coding	JPEG 2000	PACS archiving; Image exchange; Remote browsing	Mature standards; Support lossy/lossless; Good compatibility	Fixed transformation adaptation is limited; High compression ratio and easy to lose details
ROI compression	ROI code; Manual or automatic zoning	Protection of lesions, organs, blood vessels and other key areas	Priority protection of diagnostic area; Flexible compression strategy	Rely on ROI accuracy; High-risk of missing bids
3D medical image compression	Intra slice/inter slice redundancy modeling; Lossless volume data compression	CT and MRI volume data archiving; Scientific research analysis	Use the correlation between layers; Suitable for continuous 3D data	High computing and video memory costs; Complex sampling and cross source adaptation
Deep learning compression	Encoder decoder; Task-oriented compression	AI assisted diagnosis; Task specific joint optimization	Learnable image distribution; Easy to combine AI task indicators	Rely on training data; Cross domain stability and security to be verified

In addition, the four methods all face a common problem: the evaluation indexes are still mainly focused on the data or image quality level such as compression ratio, PSNR and SSIM, while the verification of the actual impact on AI diagnosis and doctor's film reading is insufficient. Relevant

DOI: 10.54254/2753-8818/2027.39489

reviews have summarized the impact of compression on deep learning tasks [7], and "diagnostic lossless compression" research has also begun to include disease classification, lesion identification or expert evaluation in the assessment [8]. This shows that medical image compression can not only compare image quality, but also make a comprehensive judgment by combining pixel error, task performance and clinical film reading results.

3. Diagnostic fidelity: the core evaluation standard of medical image compression

Diagnostic fidelity is that the compressed image can still retain enough key information to support clinical diagnosis and subsequent analysis. It is not equal to visual acuity, nor can it be completely measured by conventional image quality indicators such as peak signal-to-noise ratio (PSNR) and structural similarity (SSIM). Even if the compressed image is similar to the original image as a whole, low-contrast lesions, fine textures or tissue edges may still change, which will affect the judgment of doctors or AI models.

Diagnostic fidelity mainly includes four requirements: first, the size, shape, edge and internal texture of the lesion cannot be changed, which are the core basis for judging the nature of the disease; Second, the spatial position and structural relationship of human tissue should not be disordered, especially in three-dimensional images, small changes in single slices may affect judgment when viewed together; Third, the compressed image can not significantly affect the diagnostic results of AI model, such as the accuracy of lesion segmentation and disease classification can not be significantly reduced; Fourth, to match the fault tolerance of different scenarios: ordinary browsing and preliminary screening can accept slightly lossy compression. Preoperative planning and accurate measurement must use lossless or near lossless compression.

Diagnostic fidelity is closely related to the practical application of medical imaging. Hospitals need to preserve patients' images in PACS for a long time. The duration of image preservation and the storage pressure caused by it are the practical problems faced by medical institutions [11]; Telemedicine also needs to reduce the storage space and the amount of data transferred by compressing medical image files such as NIfTI [12]. Therefore, medical image compression should not only improve the storage and transmission efficiency, but also ensure that the compressed image can still be used for reliable diagnosis.

Previous studies have begun to incorporate expert evaluation and performance changes such as AUC and accuracy of AI diagnostic model before and after compression into compression effect evaluation [7, 8]. This evaluation method based on the actual diagnostic task constitutes an important basis for the evaluation of diagnostic fidelity, which also shows that a high image quality index does not necessarily mean that the diagnostic information has been completely retained.

At present, there is no unified quantitative standard for diagnostic fidelity. Based on the evaluation ideas of existing studies, this paper summarizes the relatively complete evaluation of diagnostic fidelity into three levels: the first level is the data level, including PSNR, SSIM and the changes of lesions and tissue structure. PSNR mainly reflects the size of pixel error, while SSIM pays more attention to structural similarity, but both can only explain the overall difference of the image, and can not fully judge whether the low-contrast lesions, edge texture and three-dimensional spatial continuity are preserved. The second is the task level, including the changes of AI classification, detection and segmentation performance before and after compression. For example, previous reviews have focused on the impact of compression on the input distribution of deep learning models and the performance of diagnostic tasks [7], and diagnostic lossless compression studies have also included disease classification, lesion identification or expert judgment in the evaluation [8]. If the AUC, accuracy or segmentation results after compression decrease

DOI: 10.54254/2753-8818/2027.39489

significantly, even if the image looks clear, it may affect the reliability of AI aided diagnosis. The third is the clinical level, including the doctor's lesion detection, diagnostic accuracy and film reading consistency. Clinically, the safety boundary can be judged by blind reading of images before and after compression, ROC analysis and consistency comparison of multiple doctors. In the future, grading evaluation criteria should also be established in PACS archiving, remote consultation, AI assisted diagnosis and other actual scenarios. The "three-tier evaluation system" here is the induction and synthesis of the existing evaluation methods in this paper, rather than directly quoting the unified standard proposed by a certain research. Only by combining these three dimensions can people comprehensively judge whether a compression method really achieves diagnostic fidelity.

4. Frontier development direction and clinical application challenges

In recent years, medical image compression has gradually developed in the direction of high-level and deep data processing, cross data source adaptation, 3D model lightweight, flexible bit rate control and joint optimization of diagnostic tasks. However, these methods still need to solve the problems of evaluation criteria, model universality, computational cost and data security in order to be truly applied in clinical practice.

First, lossless compression of high-level deep 3D images. Now the gray level accuracy of CT and MRI images is getting higher and higher, and subtle gray level changes may contain important tissue and lesion information. In view of this feature, existing studies have begun to design a lossless compression model specifically for high-level deep volume data to make full use of the numerical distribution and correlation between slices of high-level deep image [4]. However, such methods deal with a large amount of data and have high computational and storage overhead. In the future, it is necessary to improve the encoding and decoding efficiency through block processing, model simplification and hardware acceleration on the basis of ensuring the integrity of gray information.

Second, enhance the universal ability of the model across data sources. The model trained by single hospital data is not easy to use in another hospital, which is the biggest obstacle to the compression and landing of deep learning. Multi-source domain generalization research attempts to use images from different data sources for training, so that the model can adapt to more complex data differences [5]. However, most of the existing studies are still focused on the limited data set, and it is not clear whether it can remain stable in the real clinical environment. Subsequently, multi-center and multi equipment training should be strengthened, and external verification should be carried out through the data of independent institutions.

Third, reduce the computational cost of 3D compression model. Three dimensional methods can make use of the correlation between slices and within slices at the same time, but they usually need to deal with the whole volume data, and have high requirements for video memory, computing power and encoding and decoding time. Previous studies have shown that different data segmentation and sampling strategies will directly affect the training effect and computational efficiency of 3D compressed models [6]. Therefore, by optimizing the sampling method, block processing, lightweight network and hardware acceleration, the running cost of the model can be reduced, making it more suitable for hospital servers and remote diagnosis and treatment equipment.

Fourth, develop flexible compression framework from lossy to lossless. The quality requirements of clinical images in the scenarios of fast browsing, remote transmission, long-term archiving and preoperative planning are different. Previous studies have proposed lossy to lossless three-dimensional compression methods, so that the same image can be used for transmission and preview with a small amount of data, and further restored to lossless state when necessary [13]. This method helps to meet the use needs of different scenarios, but it is still necessary to clarify the scope of

DOI: 10.54254/2753-8818/2027.39489

application of different compression stages, and verify whether the damaged version will affect the judgment of doctors and the results of AI analysis.

Fifth, promote the joint optimization of compression and diagnosis tasks. Traditional compression only pursues the restoration of image quality, but the real concern in clinical practice is whether it can be diagnosed. Previous studies have begun to incorporate disease classification, lesion recognition, image segmentation and expert evaluation into the design and verification of compression methods [7, 8]. In the future, the requirements of the diagnostic task can be added to the training goal of the compressed model, so that the model can keep the useful information for diagnosis while reducing the file. However, it should also be noted that the information of a certain task cannot be retained, resulting in the loss of other potential diagnostic information. Therefore, lossless or strictly verified near lossless compression should be preferred for high-risk clinical scenes, and the original image, model version and compression parameters should be retained for traceability and review.

In addition to the above problems, the lack of unified evaluation criteria for diagnostic fidelity is still a common challenge for various methods. The data sets, compression ratio and diagnostic tasks used in different studies are inconsistent, and the experimental results are difficult to compare directly. The follow-up study should report the compression efficiency, objective image quality, AI task performance and doctor's film reading results under the unified test conditions, so as to determine whether the compression method can improve the storage and transmission efficiency and stably retain the clinical diagnostic value.

5. Conclusions

On the whole, medical image compression has experienced the development process from traditional transform coding, ROI compression, 3D modeling to deep learning compression. Each of the four methods has its own advantages and disadvantages and applicable scenarios. In the future, medical image compression may adopt the technical route of complementing traditional coding and learning methods for a long time.

The core of the development of medical image compression should not be the blind pursuit of higher compression ratio, but should turn to the core of diagnostic fidelity, and retain the information required for reliable diagnosis while reducing the data volume. In the future, it is necessary to further improve the evaluation method of diagnostic fidelity, improve the adaptability of the model across hospitals and equipment, reduce the computational cost of three-dimensional model, and strengthen the collaborative optimization between compression method and diagnostic task. Only after the objective image quality, AI tasks, clinical film reading and other multi-level verification, can the medical image compression technology be reliably applied in long-term archiving, telemedicine, AI assisted diagnosis and other scenarios.

References

[1] Rabbani, M., & Joshi, R. (2002). An overview of the JPEG 2000 still image compression standard. *Signal Processing: Image Communication, 17*(1), 3–48. https://doi.org/10.1016/S0923-5965(01)00024-8

[2] Chen, Z., Gu, S., Lu, G., & Xu, D. (2022). Exploiting intra-slice and inter-slice redundancy for learning-based lossless volumetric image compression. *IEEE Transactions on Image Processing, 31*, 1697–1707. https://doi.org/10.1109/TIP.2022.3140608

[3] Liu, X., Wang, M., Wang, S., & Kwong, S. (2024). Bilateral context modeling for residual coding in lossless 3D medical image compression. *IEEE Transactions on Image Processing, 33*, 2502–2513. https://doi.org/10.1109/TIP.2024.3378910

DOI: 10.54254/2753-8818/2027.39489

[4] Wang, K., Bai, Y., Li, D., Zhai, D., Jiang, J., & Liu, X. (2025). Learning lossless compression for high bit-depth volumetric medical image. *IEEE Transactions on Image Processing, 34*, 4896–4907. https://doi.org/10.1109/TIP.2024.3513156

[5] Xue, D., Wu, S., Li, L., Liu, D., & Li, Z. (2025). Multi-source domain generalization for learned lossless volumetric biomedical image compression. *IEEE Transactions on Image Processing*. https://doi.org/10.1109/TIP.2025.3592549

[6] Nagoor, O. H., Whittle, J., Deng, J., Mora, B., & Jones, M. W. (2022). Sampling strategies for learning-based 3D medical image compression. *Machine Learning with Applications, 8*, 100273. https://doi.org/10.1016/j.mlwa.2022.100273

[7] Urbaniak, I. A. (2024). Using compressed JPEG and JPEG2000 medical images in deep learning: A review. *Applied Sciences, 14*(22), 10524. https://doi.org/10.3390/app142210524

[8] Van der Sluijs, R., Varma, M., Prince, J., Langlotz, C., & Chaudhari, A. (2023). Diagnostically lossless compression of medical images. In *ICML 2023 Workshop on Neural Compression: From Information Theory to Applications*. https://openreview.net/forum?id=ZiNFhNFxMf

[9] Dai, F. (2025). Deep learning based medical image compression using cross attention learning and wavelet transform. *Scientific Reports, 15*, 40008. https://doi.org/10.1038/s41598-025-23582-y

[10] Cheng, Z., Sun, H., Takeuchi, M., & Katto, J. (2020). Learned image compression with discretized Gaussian mixture likelihoods and attention modules. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 7939–7948). https://doi.org/10.1109/CVPR42600.2020.00796

[11] Behnam, F., Khajouei, R., & Ahmadian, L. (2024). The retention duration of digital images in picture archiving and communication systems. *Heliyon, 10*(6), e27847. https://doi.org/10.1016/j.heliyon.2024.e27847

[12] Elhadad, A. A., Jamjoom, M., & Abulkasim, H. (2024). Reduction of NIFTI files storage and compression to facilitate telemedicine services based on quantization hiding of downsampling approach. *Scientific Reports, 14*, 5168. https://doi.org/10.1038/s41598-024-54820-4

[13] Rossinelli, D., Fourestey, G., Schmidt, F., Busse, B., & Kurtcuoglu, V. (2021). High-throughput lossy-to-lossless 3D image compression. *IEEE Transactions on Medical Imaging, 40*(2), 607–620. https://doi.org/10.1109/TMI.2020.3033456