Ancient Book Image Restoration Using Generative Adversarial Networks

Nan Wang *, JinRan Zhu, Shang Shi

School of Statistics and Data Science, Ningbo University of Technology, Zhejiang, China

* Corresponding Author Email: 1906368022@qq.com

Abstract. As an important carrier of historical and cultural inheritance, the restoration of ancient books is of great significance to the protection of cultural relics and cultural inheritance. However, the traditional repair methods have some problems, such as low efficiency and insufficient precision. In this paper, a deep learning-based restoration method for ancient books is proposed, which is divided into two steps: structure reconstruction and color correction. The structure reconstruction network (SRN) uses line drawing information to ensure the authenticity and structural stability of large-scale content, and the color correction network (CCN) makes local color adjustments to missing pixels, reducing color bias and edge hopping problems. The experimental results show that this method effectively improves the restoration efficiency and image quality, and provides a new technical support for the protection and inheritance of ancient books.

Keywords: Restoration of ancient books, Deep learning, Convolutional neural network, Adversarial generating network, Line guide.

1. Introduction

Artifact restoration today faces both complexity and significance, as growing global awareness of cultural heritage preservation and rapid technological advancements increase the urgency of conservation efforts. Governments are dedicating significant resources, such as the 18 million yuan allocated to Ningbo's Tianyi Pavilion Museum, to support artifact restoration. Artifacts hold immense historical and cultural value, serving as carriers of memory and national identity, making their preservation vital. Traditional restoration techniques, though skilled, are inefficient for large-scale damage. However, the rise of digital technologies has introduced preventive conservation and digitalization as complementary approaches. Preventive conservation protects artifacts from potential damage, while digitalization enables their preservation and dissemination in broader formats. Together, these methods provide essential technical support for protecting and utilizing cultural heritage effectively. In this context, leveraging modern tools and continuing innovation are key to advancing artifact conservation and restoration.

The rapid development of computer technology has brought new opportunities for the restoration of ancient books, especially in preserving intangible cultural heritage. Technological advancements, such as deep learning and convolutional neural networks, offer innovative solutions for more efficient restoration, reducing the workload and making preservation more convenient.

Fangfang (2011) addressed the issues of slow convergence and susceptibility to local minima in traditional BP neural networks for power load forecasting, and proposed an improved BP neural network method [1]. This method introduces momentum factors and an adaptive learning rate during the training process, which enhances the training speed and prediction accuracy of the network [2-4]. Xu et al. (2019) presented an image enhancement algorithm based on a generative adversarial network, combined with an improved game adversarial loss mechanism, effectively improving image quality [5]. Sreekumar et al. (2024) utilized a Wasserstein deep convolutional generative adversarial network based on an optimized vision transformer to achieve multi-label classification of fundus images [6]. Li et al. (2024) proposed a radiomics-supervised generative adversarial network for the generation of lung lesion images [7]. Kala et al. (2024) employed a deep edge region-based generative adversarial network to accomplish the inpainting of brain magnetic resonance images [8]. Vashistha et al. (2024) introduced a modular generative adversarial network for the reconstruction of positron emission tomography images [9]. Anantharajan et al. (2024) presented an attention-based dual residual generative adversarial network classifier that combines multilayer dense net feature extraction and hyper-parameter tuning for brain tumor classification [10].

The restoration of ancient books has become a key research area in cultural relic protection and digital inheritance. Many ancient books have suffered physical damage due to aging, stains, and tears, impacting readability and preservation. With advancements in deep learning, new solutions for image restoration have emerged. This study proposes a two-step approach called line drawing guided step-by-step restoration, involving structural reconstruction and color correction. The structural reconstruction network (SRN) ensures the stability and authenticity of the content, while the color correction network (CCN) adjusts missing pixels to address color deviations. Photos of damaged and restored books are used to train convolutional neural networks, employing the U-Net algorithm for image segmentation and repair. Pre-trained deep neural networks are used for reconstructing large holes in images, with natural language processing aiding in restoring missing text. After restoration, results are optimized for long-term preservation. Additionally, a desktop program or website is proposed to allow users to input images for restoration, facilitating broader access to these advanced techniques. This method significantly improves restoration accuracy and opens new possibilities for ancient book preservation.

2. Design of Generative Adversarial Networks with optimized strategies

The application of convolutional neural networks (CNNs) in the field of ancient book restoration is advancing rapidly and playing an increasingly vital role in preserving and restoring priceless historical texts. The restoration of ancient books is a labor-intensive process, requiring the handling of damaged or decaying manuscripts and documents to ensure their survival. CNNs, as a powerful deep learning technique, have demonstrated immense potential in automating and enhancing these efforts.

CNNs are particularly effective in repairing damaged sections or filling in missing parts of ancient texts. Through model training, CNNs can learn contextual information about damaged areas and attempt to reconstruct the missing content. Additionally, CNNs excel in character recognition, enabling accurate identification of characters within ancient books. Leveraging large datasets of historical text, CNNs can automatically recognize and transcribe ancient manuscripts, significantly improving restoration efficiency and accuracy while reducing the need for manual labor.

2.1. Input Layer

The input layer of a CNN processes multidimensional data, typically representing images. For instance, a 1D CNN may receive a one-dimensional or two-dimensional array, where the array could represent time-series or spectral data. A 2D CNN receives two-dimensional or three-dimensional arrays, often including multiple channels. The 3D CNN accepts four-dimensional arrays, commonly used in tasks involving spatiotemporal data. Given the widespread application of CNNs in computer vision, most studies presume three-dimensional input data, representing 2D pixel images with RGB color channels.

Similar to other neural networks, input features for CNNs must be normalized to facilitate efficient gradient descent learning. Before being fed into the CNN, data—whether pixel values or other numerical arrays—should be normalized to enhance the learning process. For pixel-based data, normalization ensures that the values fall within a certain range, improving the network's convergence and performance.

The input layer typically handles images in the form of a matrix with dimensions [width x height x depth]. For instance, an image input of [16x16x3] represents a 16x16 pixel image with 3 channels (RGB). The input dimensions should generally be divisible by 2, such as 32, 64, 128, 224, 384, or 512.

2.2. Hidden layer

CNN hidden layers consist of various types, including convolutional, pooling, and fully connected layers. In modern architectures, advanced constructions like Inception modules or residual blocks may also appear. The convolutional and pooling layers are exclusive to CNNs, whereas fully connected layers are found in traditional feedforward networks as well. Convolution layers contain learnable weight coefficients, while pooling layers do not, serving more as feature reducers rather than independent learning components. An example of a common CNN architecture, LeNet-5, typically follows this layer sequence: input → convolution → pooling → fully connected → output.

2.3. Pooling layer

After feature extraction in the convolutional layer, the output feature map is sent to the pooling layer for dimensionality reduction, feature selection, and information filtering. The pooling function, typically predefined, aggregates feature map information from a local neighborhood. Pooling is controlled by the region size, step size, and padding.

(1) Lp pooling

Lp pooling is a class of pooling models inspired by the hierarchical structure in the visual cortex, which is generally expressed as:

$$o = \left(\frac{1}{N}\sum_{i=1}^{N} V_i^P\right)^{\frac{1}{p}} \tag{1}$$

o Indicates the output of the pooling operation. N Indicates the number of elements in the pooling window. v_i denotes the value of the ith element within the pooling window. P is a real index that controls the type of pooling operation

The meaning of step size and pixel in the formula is the same as that of the convolutional layer, and they are pre-specified parameters. When p=1, Lp pooling takes the mean value in the pooling region, which is called mean pooling. As p approaches infinity, Lp pooling takes a maximum value in the region and is called maximum pooling. Mean pooling and maximum pooling are long-standing pooling methods used in the design of convolutional neural networks. Both retain the background and texture information of the image at the cost of losing some information or size of the feature map. In addition, Lp pooling at p=2 is also used in some work.

(2) Hybrid pooling and random pooling

Hybrid pooling combines mean and max pooling, while random pooling introduces stochasticity, selecting a value within the pooling region based on a specific probability distribution. Both methods offer regularization benefits, helping reduce overfitting (Figure 1).

The mathematical expression for Mixed Pooling (MP) can be written as:

$$P_{mixed} = \lambda \times P\,max + (1-\lambda) \times P_{avg} \tag{2}$$

- P_{mixed} is the result of hybrid pooling.
- P_{max} is the result of the largest pool.
- $P_{avg}$is the result of average pooling.
- λ is a random variable that takes the value 0 or 1.

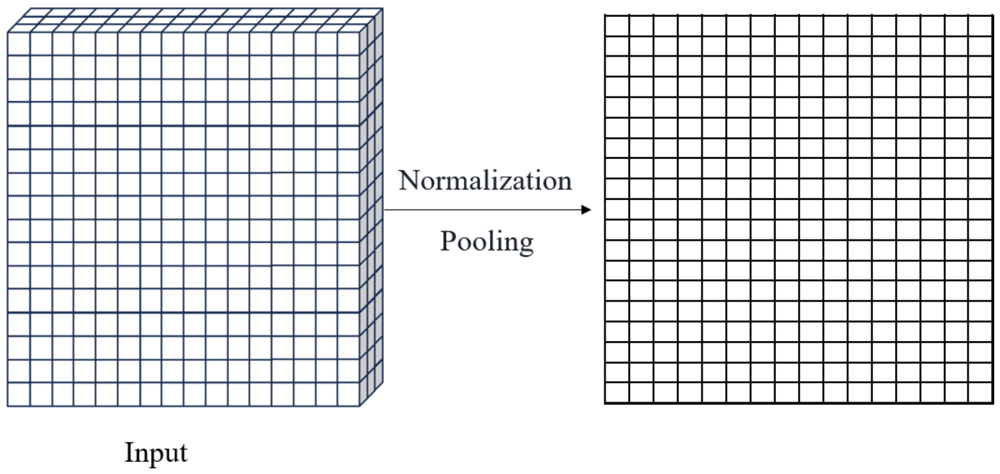

Figure 1. Data Preprocessing

2.4. Fully connected layer

The fully connected layer in convolutional neural networks is equivalent to the hidden layer in traditional feedforward neural networks. The fully connected layer is the last part of the hidden layer of the convolutional neural network and only transmits signals to other fully connected layers. The feature graph loses its spatial topology in the fully connected layer, is expanded as a vector and passes through the excitation function.

From the perspective of representation learning, the convolutional layer and the pooling layer in the convolutional neural network can extract features from the input data, while the function of the fully connected layer is to combine the extracted features in a non-linear way to obtain the output. In other words, the fully connected layer is not expected to have feature extraction capability itself, but tries to use the existing higher-order features to complete the learning goal (Figure 2).

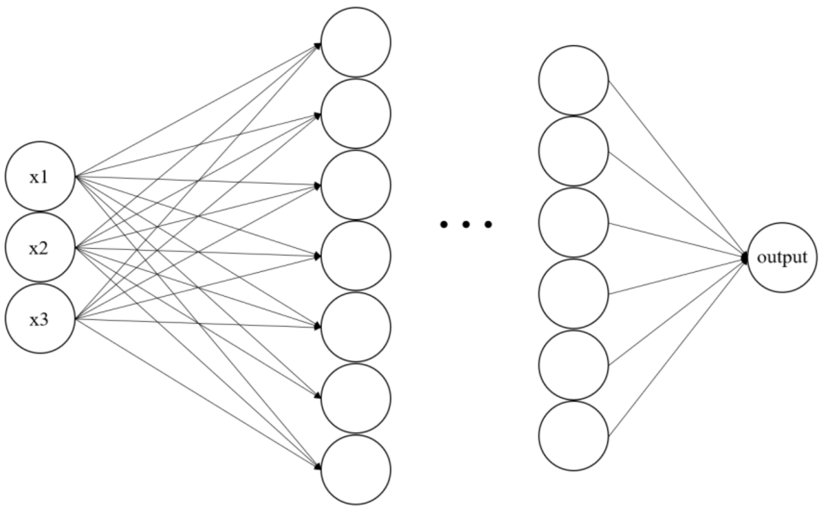

Figure 2. FC Layer

In the figure above, the feature mapping matrix will be converted to vectors (x1, x2, x3...). With a fully connected layer, we combine these functions together to create a model. Finally, we divide the output into target classes through an activation function such as softmax or sigmoid.

2.5. Output layer

The upstream of the output layer in a convolutional neural network is usually a fully connected layer, so its structure and working principle are the same as the output layer in a traditional feedforward neural network. For image classification problems, the output layer outputs classification labels using logical functions or normalized exponential functions. In object recognition problems, the output layer can be designed to output the central coordinates, size, and classification of the object. In image semantic segmentation, the output layer directly outputs the classification results of each pixel.

2.6. Image recognition

Convolutional neural networks have long been one of the core algorithms in the field of image recognition, and have stable performance when the learning data is sufficient. For general large-scale image classification problems, convolutional neural networks can be used to construct hierarchical classifiers, and in fine-classification recognition to extract the discriminant features of images for other classifiers to learn. For the latter, feature extraction can artificially input different parts of the image into the convolutional neural network, or it can be extracted by the convolutional neural network through unsupervised learning.

For image restoration techniques for text, Convolutional Neural Networks as well as Adversarial Neural Networks are used for numerical prediction of pixel points. Where the generative network is responsible for repairing the damaged image and the discriminator network provides feedback that helps the generative network to generate a higher quality image. Through this adversarial training, the generative network learns how to generate more realistic and high quality repaired images.

The generative network includes a coarse repair network as well as a fine repair network. Coarse restoration network is used to generate rough restoration images. It contains multiple convolutional layers, normalization layers, activation functions and residual blocks. The input of the network is the corrupted image and the output is the coarsely repaired image. The fine repair network further refines the repair results on the basis of the coarse repair network and outputs a more finely repaired image. And discriminative network is a network used to discriminate whether the input image is real or generated. It contains multiple convolutional layers and LeakyReLU activation function, and finally outputs a probability value indicating the probability that the input image is a real image.

By this combination of chopping neural network and convolutional neural network thus making the image with better reproduction.

2.7. Detailed explanation of method

2.7.1 GAN Network

Adversarial neural networks (Gans, also known as generative adversarial neural networks) are deep learning models that consist of two neural networks: a generative network and a discriminant network.

The role of the generative network is to generate false data similar to the real data, while the role of the discriminant network is to determine whether the input data is real or false data. Through repeated iterative training, the generative network can generate more realistic false data, and the discriminant network can judge the authenticity of the input data more accurately.

The training process of adversarial neural network usually includes the following steps: Firstly, the generating network generates false data and inputs it into the discriminant network for judgment; Then, the discriminant network classifies the input data and calculates the classification error. Then, according to the classification error, the parameters of the discriminant network are updated to improve the classification accuracy. Finally, the generated network updates its own parameters according to the feedback information of the discriminant network to generate more realistic fake data. This process is iterated until a preset number of training rounds or convergence conditions are reached.

Adversarial neural networks are widely used in many fields, such as image generation, image restoration, image denoising, image style conversion and so on. In addition, it can also be applied to text generation, machine translation, speech synthesis and other fields. Adversarial neural networks have been widely studied in the field of deep learning because of their strong generalization ability and robustness.

In this study, our GAN network, as a larger framework, has made some adjustments and optimizes on the D and G of GAN respectively for the successful completion of our restoration project. The generator network is composed of two networks, namely SRN network and CCN network.

2.7.2 SRN network

SRN (Simple Recurrent Network) is a very basic recurrent neural network. Its main feature is that it has a hidden layer, and the feedback connection between this hidden layer is added. This structure enables SRN to process sequential data and capture time dependencies in the data.

In a two-layer feedforward neural network, connections usually exist only between adjacent layers, and there is no connection between the nodes of the hidden layer. SRN networks, however, break this limitation by providing feedback connections from hidden layers to hidden layers, allowing information to circulate within the network. This structure makes the SRN network more powerful when processing sequence data, as it is able to remember previous information and use it for subsequent calculations.

SRN networks are widely used in many fields, including natural language processing, time series analysis and so on. In natural language processing, SRN networks can improve the accuracy of text processing by remembering previous words or sentences to understand the meaning of current words or sentences. In time series analysis, SRN networks can capture time dependencies in sequence data to predict future data trends or patterns.

2.7.3 CCN network

The Color Correction Network (CCN) addresses color imbalances in restored images, specifically targeting mural restoration. Traditional color correction methods often struggle with complex color distributions, but CCN leverages deep learning to learn and adjust image color features adaptively. CCN improves the visual realism of restored images by making localized adjustments, enhancing both color balance and edge continuity. CCN has broad applications in image processing, from digital photography to medical imaging and autonomous driving (Figure 3).

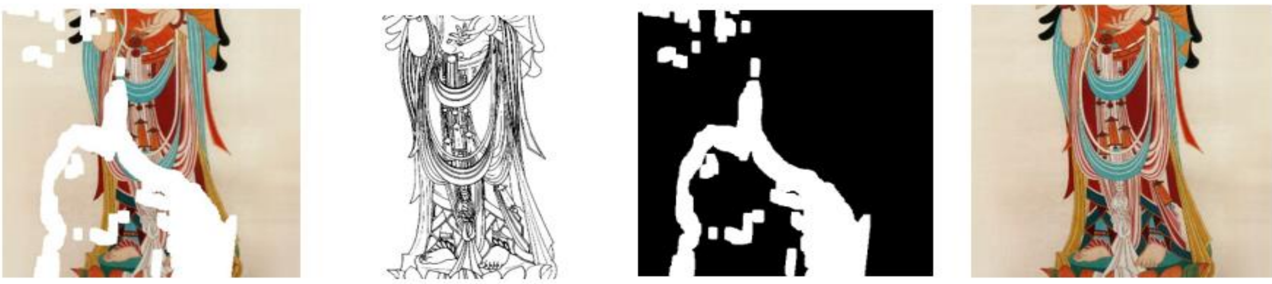

Figure 3. Restored Mural Image

3. Ancient Book Image Restoration

3.1. Data preparation

The model requires a lot of post-masks, bar, and erased original as input, and then will send out the three images in one generated map (Figure 4).

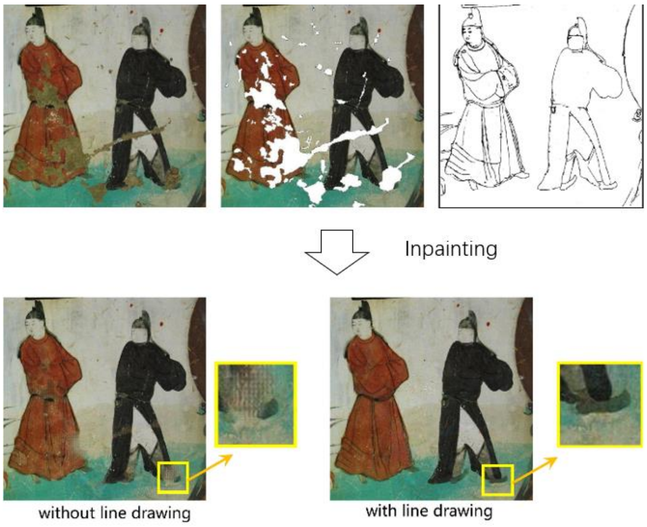

Figure 4. Obtained repaired image.

3.2. Dataset Introduction

The images are sourced partially from portrait photographs in ancient books from the Tianyi Pavilion and some images of Dunhuang murals.

3.3. Data Preprocessing

(1) Image Size Transformation: The original ancient book illustrations and mural photographs are transformed into a uniform size of 512x512 pixels. This step ensures that all image data have the same dimensions, facilitating subsequent image processing and analysis.

(2) Data Augmentation: Techniques such as rotation, scaling, and cropping are applied to increase the diversity of the dataset and improve the model's generalization ability.

(3) Artificial Blurring: Manually creating damaged areas or other content that users do not wish to display. In image restoration tasks, this is typically done to create a mask that instructs the model which regions need to be repaired.

(4) Generating Masks: The mask is a binary matrix of the same size as the image. In this task, the mask is used to indicate which regions are damaged and need to be restored during the image restoration process.

3.4. Network Architectures and Numerical Results

The network architecture of ancient book image restoration is shown in Figure 5. The results are shown in Table 1. The comparison before and after restoration is shown in Figure 6.

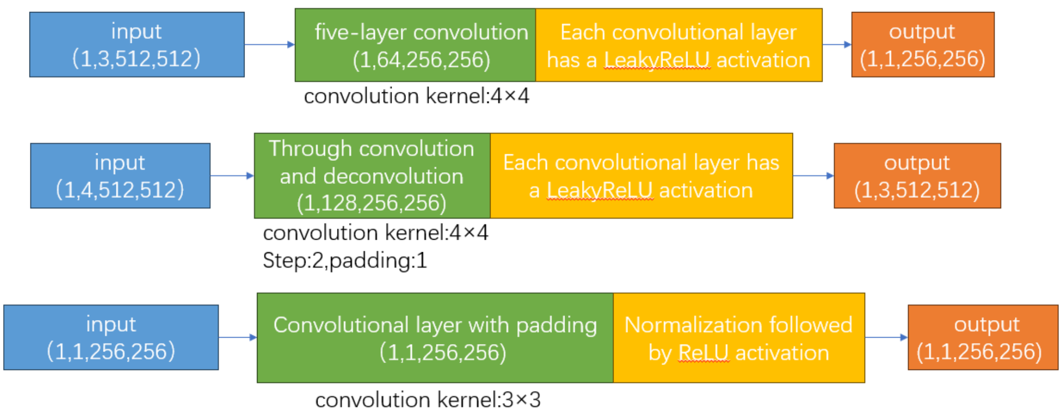

Figure 5. The network architectures for Ancient Book Image Restoration

Table 1 Results Presentation

Image Number	PSNR	SSIM
000041.jpo	21.691696021076844	0.7485775947570801
000042.jpg	21.070404913758786	0.7749924063682556
000043.jpg	24.849733127833122	0.810878574848175
000044.jpg	23.98868657749216	0.8034065365791321
000045.jpg	22.459452559198866	0.7767946124076843
000046.jpg	20.53085237235869	0.8288459777832031
000047.jpg	27.41286970275249	0.8483255505561829
000048.jpg	22.063539674432242	0.7483582496643066
000049.jpg	25.474197979212057	0.7363462448120117
000050.jpg	23.237704031755918	0.703643262386322

Figure 6 Comparison Chart Before and After Restoration

4. Conclusion

The restoration method of ancient books based on deep learning proposed in this paper divides the restoration process into two steps: structural reconstruction and color correction, which effectively solves the shortcomings of traditional restoration methods. SRN network uses line drawing information to ensure the authenticity and stability of the structure, and CCN network makes local color adjustment to make the restoration result more natural and realistic. This method shows advantages in the restoration of ancient books, which can effectively improve the restoration efficiency and image quality, and provide new technical support for the protection and inheritance of ancient books. In the future, with the continuous development of deep learning technology, the method is expected to be applied and promoted in more fields, and make greater contributions to cultural heritage protection and cultural inheritance.

References

[1] Fangfang. Research on power load forecasting based on Improved BP neural network [D]. Harbin Institute of Technology, 2011.

[2] Amjady N. Short-term hourly load forecasting using time series modeling with peak load estimation capability [J]. IEEE Transactions on Power Systems, 2001, 16(4): 798-805.

[3] Ma Kunlong. Short term distributed load forecasting method based on big data [D]. Changsha: Hunan University, 2014.

[4] SHI Biao, LI Yu Xia, YU Xhua, YAN Wang. Short-term load forecasting based on modified particle swarm optimizer and fuzzy neural network model [J]. Systems Engineering-Theory and Practice, 2010, 30(1): 158-160.

[5] Xu C ,Cui Y ,Zhang Y , et al.Image enhancement algorithm based on generative adversarial network in combination of improved game adversarial loss mechanism[J].Multimedia Tools and Applications,2019,79(prepublish):1-16.

[6] Sreekumar K ,Vimina R E .Multi-Label Classification of Fundus Images with Optimized Vision Transformer Based Wasserstein Deep Convolutional Generative Adversarial Network[J].SN Computer Science,2024,5(7):842-842.

[7] Li J ,Pan S ,Zhang X , et al.Generative Adversarial Networks with Radiomics Supervision for Lung Lesion Generation.[J].IEEE transactions on bio-medical engineering,2024,PP

[8] Kala R ,Chandrasekaran R ,Ahilan A , et al.Brain Magnetic Resonance Image Inpainting via Deep Edge Region-based Generative Adversarial Network[J].Journal of Electrical Engineering & Technology,2024,(prepublish):1-12.

[9] Vashistha R ,Vegh V ,Moradi H , et al.Modular GAN: positron emission tomography image reconstruction using two generative adversarial networks[J].Frontiers in Radiology,2024,41466498-1466498.

[10] Anantharajan S ,Gunasekaran S ,Sujana J A J .Brain tumor classification for combining the advantages of multilayer dense net-based feature extraction and hyper-parameters tuned attentive dual residual generative adversarial network classifier using wild horse optimization.[J].NMR in biomedicine,2024,e5246.